
HANDBOOK
of
NUMERICAL ANALYSIS

P. G. CIARLET • Editor

Volume
XV

Special Volume
**Mathematical Modeling
and Numerical Methods
in Finance**

A. BENSOUSSAN
Q. ZHANG
Guest Editors

NORTH-HOLLAND

Handbook of Numerical Analysis

General Editor:

P.G. Ciarlet

*Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie
4 Place Jussieu
75005 PARIS, France*

and

*Department of Mathematics
City University of Hong Kong
Tat Chee Avenue
KOWLOON, Hong Kong*



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier



Volume XV

Special Volume:
Mathematical Modeling and
Numerical Methods in Finance

Guest Editors:

Alain Bensoussan

*International Center for Decision and Risk Analysis (ICDRiA),
School of Management, University of Texas at Dallas,
SM 30, Richardson, TX 75083-0688, USA*

Qiang Zhang

*Department of Mathematics and Department Economics and Finance,
City University of Hong Kong,
83 Tat Chee Avenue, Kowloon, Hong Kong*



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier



North-Holland is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

Copyright © 2009 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at <http://elsevier.com/locate/permissions>, and selecting *Obtaining permission to use Elsevier material*.

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978-0-444-51879-8

For information on all North-Holland publications
visit our website at elsevierdirect.com

Printed and bound in Great Britain

08 09 10 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

General Preface

In the early eighties, when Jacques-Louis Lions and I considered the idea of a *Handbook of Numerical Analysis*, we carefully laid out specific objectives, outlined in the following excerpts from the “General Preface” which has appeared at the beginning of each of the volumes published so far:

During the past decades, giant needs for ever more sophisticated mathematical models and increasingly complex and extensive computer simulations have arisen. In this fashion, two indissociable activities, *mathematical modeling* and *computer simulation*, have gained a major status in all aspects of science, technology and industry.

In order that these two sciences be established on the safest possible grounds, mathematical rigor is indispensable. For this reason, two companion sciences, *Numerical Analysis* and *Scientific Software*, have emerged as essential steps for validating the mathematical models and the computer simulations that are based on them.

Numerical Analysis is here understood as the part of *Mathematics* that describes and analyzes all the numerical schemes that are used on computers; its objective consists in obtaining a clear, precise, and faithful, representation of all the “information” contained in a mathematical model; as such, it is the natural extension of more classical tools, such as analytic solutions, special transforms, functional analysis, as well as stability and asymptotic analysis.

The various volumes comprising the *Handbook of Numerical Analysis* will thoroughly cover all the major aspects of Numerical Analysis, by presenting accessible and in-depth surveys, which include the most recent trends.

More precisely, the Handbook will cover the *basic methods of Numerical Analysis*, gathered under the following general headings:

- Solution of Equations in $\mathbb{R}^n$,
- Finite Difference Methods,
- Finite Element Methods,
- Techniques of Scientific Computing.

It will also cover the *numerical solution of actual problems of contemporary interest in Applied Mathematics*, gathered under the following general headings:

- Numerical Methods for Fluids,
- Numerical Methods for Solids.

In retrospect, it can be safely asserted that Volumes I to IX, which were edited by both of us, fulfilled most of these objectives, thanks to the eminence of the authors and the quality of their contributions.

After Jacques-Louis Lions' tragic loss in 2001, it became clear that Volume IX would be the last one of the type published so far, i.e., edited by both of us and devoted to some of the general headings defined above. It was then decided, in consultation with the publisher, that each future volume will instead be devoted to a single "*specific application*" and called for this reason a "*Special Volume*". "*Specific applications*" will include Mathematical Finance, Meteorology, Celestial Mechanics, Computational Chemistry, Living Systems, Electromagnetism, Computational Mathematics etc. It is worth noting that the inclusion of such "specific applications" in the *Handbook of Numerical Analysis* was part of our initial project.

To ensure the continuity of this enterprise, I will continue to act as Editor of each Special Volume, whose conception will be jointly coordinated and supervised by a Guest Editor.

P.G. CIARLET
July 2002

Model Risk in Finance: Some Modeling and Numerical Analysis Issues

Denis Talay, INRIA

2004 Route des Lucioles, B.P. 93, 06902 Sophia-Antipolis, France.

1. Introduction

The impact of erroneous models and measurements is an important issue in all scientific and technological fields: equations and measurement devices provide approximate descriptions of our real world so that one needs to estimate and possibly control the effects of misspecifications during the modeling and calibration process.

In fields such as physics, conservation laws constrain the models and the values of the model parameters, even when a part of stochasticity is involved to take uncertainties into account. As well, to solve numerically a partial differential equation (PDE) describing macroscopic quantities whose state space is unbounded, one needs to introduce artificial boundary conditions that allow one to compute the solution within a bounded domain; the design of these boundary conditions is a difficult issue, but one may be helped by intuitive considerations on the physical phenomenon under study; to give an example, if one desires to compute turbulent flows around airplane wings, one may assume that, away from the airplane, the velocity of the flow is equal to the wind velocity, and one thus may derive reasonable approximate Dirichlet conditions from a reasonable physical model.

In finance, modeling issues are much more complex than in physics for, at least, the following reasons.

First, no physical law helps the modeler to choose a particular dynamics to describe the time evolution of market prices or indices. The real market is incomplete and arbitrages occur. Moreover, no stationarity argument can help justify that parameters estimated from historical data will keep the same values in the next future. Therefore, the modeller has a high degree of freedom to mathematically describe the market in order to compute

optimal portfolio allocations or risk measures. For example, authors propose to model the volatility of a stock as a deterministic function of the stock (and possibly of exogeneous factors) or as a stochastic process; the stochastic differential equations involved in the models may be driven by Brownian motions or by discontinuous Lévy process; the bond market is modeled by short-term dynamics or by Heath–Jarrow–Morton (HJM) equations. In addition, to compute price options and deltas, practitioners and quants find it convenient to suppose that the no arbitrage and completeness hypotheses prevail: in diffusion models, this assumption constrains the dimension and the algebraic structure of the volatility matrix so that the model used to hedge may not exactly fit the market data.

Second, statistical procedures issued from the theory of statistics of random processes and based upon historical data may be extremely inaccurate because of the lack of data. For example, an accurate parametric estimation of a volatility matrix requires that the asset price is observed at very high frequencies. As well, the parametric estimators of a drift parameter may need long-time observations to provide reliable results (see our illustration in Section 2.1). In such a case, one needs to assume that, during the whole period, the model remains relevant and its parameters remain constant. Of course, it would be unclever to use historical data only to calibrate financial models: in order to calibrate a stock price model, the practitioners not only actually consider the past prices of the stock only but also use other available information such as past prices of derivatives on this stock (see, e.g., papers and references in AVELLANEDA [2001]). However, the stationarity of the market during the observation period remains questionable, and error estimates for complex calibration methods are not available in the literature.

Third, in finance one neither can use data issued from experiments repeated independently nor assume a kind of ergodicity in order to increase the set of available observations. The modeler needs to design and calibrate models using one single history of the market.

Finally, model uncertainties also occur in the numerical resolution of PDEs related to option pricing or optimal portfolio allocation. Commonly used stochastic models in finance actually lead to consider processes whose time marginal laws have unbounded supports. Consequently, the PDEs are posed in unbounded domains, and artificial boundary conditions are necessary. The situation is quite different from the above example in fluid mechanics: usually one has a little knowledge on the behavior of the solution when the norm of the state variable increases: usually one finds estimates by working with simplified models. For an example of a rigorous procedure to design artificial boundary conditions for European options, see COSTANTINI, GOBET and KAROUI [2006]; for an analysis of the error induced by misspecified boundary conditions on American option prices, see BERTHELOT, BOSSY and TALAY [2004].

Consequently, model misspecifications cannot be avoided, which leads to model risk. The specificity and definitions of model risk are not universally admitted (see the extended introduction in CONT [2006] for an interesting discussion on this point and an extended list of references). In the present notes, we limit ourselves to a particular restricted family of questions: how to evaluate — and possibly control — the impact of certain model uncertainties on profit and losses (P&Ls) of hedging portfolios or on portfolio management strategies? We do not examine axiomatic questions on risk measures at all, for which we refer to CHERIDITO, DELBEAN and KUPPER [2005], BARRIEU and EL KAROUI [2005],

and FÖLLMER and SCHIED [2002]. We rather adopt a pragmatic point of view and seek computational means to evaluate the impact of model uncertainties.

We start with illustrating the difficulty in constructing a reliable market model by presenting recent results on one of the very first steps of the modeling process, namely, the design of the driving noise of the dynamics of the assets under consideration. We then present some results concerning the numerical approximation of measures of model risk such as Values at Risk (VaR) in diffusion environments. We also present a stochastic game problem related to model risk control. Finally, we propose a tentative methodology to compare the performances of financial strategies derived from (misspecified) mathematical models and strategies, which, derived from technical analysis, avoid modeling and calibration issues.

2. Limitations of statistical procedures based on historical data

In the literature, one can find a huge number of papers that propose and analyze parametric and nonparametric estimators for the coefficients of stochastic differential equations. A more specific literature also exists on the statistics of stochastic models in finance (for a survey, see, AÏT-SAHALIA and KIMMEL [2007]). Our purpose here is not to provide a summary of these works, even partially: we limit ourselves to refer to PRAKASA RAO [1999a] and PRAKASA RAO [1999b] and the references therein for the reader interested by an overview on the subject, and to JACOD [2000] for an advanced result on the identification of the volatility function with kernel estimators. In the latter reference, it is shown that, if a diffusion process is observed at times i/n and if the diffusion coefficient has regularity r , then the accuracy of the estimator is of order $1/n^{r/(1+2r)}$, pointwise and uniformly on compact subsets of $\mathbb{R}$. Such a convergence rate is low and illustrates that the design of stochastic models for asset prices or indices from historical data necessarily leads to model risk. We give a few other illustrations below: we will start by an elementary observation that shows that the time scales that are necessary to calibrate stochastic models with good accuracies are often incoherent with the time scales at which the market evolves. We will then examine two questions involved, which, to our knowledge, were recently only tackled in the literature in spite of the fact that they should arise before calibration. They concern the driving noise, more precisely, its continuous or discontinuous nature, and (in the Brownian case) its dimension.

2.1. Cramer–Rao lower bounds

Our elementary example concerns maximum likelihood estimators for drift parameters of diffusion processes and therefore the calibration of historical probability measures (e.g., in order to solve optimal portfolio management problems or to simulate benchmark histories of the market).

We are given an open set $\Theta \subset \mathbb{R}$ and a family of real-valued functions $\{b(\theta, \cdot), \theta \in \Theta\}$. Suppose that, for each $\theta \in \Theta$, the function $b(\theta, \cdot)$ is Lipschitz and consider the model

$$X_t^\theta = X_0 + \int_0^t b(\theta, X_s^\theta) ds + B_t, \quad (2.1)$$

where (B_t) is a standard one-dimensional Brownian motion. Up to a transformation by means of the function $\int_0^x \frac{1}{\sigma(z)} dz$, our situation covers the models with a strictly positive continuous volatility function $\sigma(x)$.

Let $\mathbb{P}^{X^\theta}$ be the law of $(X_t^\theta, 0 \leq t \leq T)$, and let $\mathbb{E}^{X^\theta}$ denote the corresponding expectation. Suppose that the function $b(\theta, x)$ is continuously differentiable w.r.t. θ for all x and that

$$I_T(\theta) := \mathbb{E}^{X^\theta} \int_0^T \left| \frac{\partial b}{\partial \theta}(\theta, \pi_s) \right|^2 ds < \infty \quad \text{for all } \theta \in \Theta.$$

Under weak additional conditions, for all unbiased estimator $\hat{\theta}_T$ of θ based upon an observation between times 0 and T such that the function

$$Q_T(\theta) := \mathbb{E}^{X^\theta} (\hat{\theta}_T - \theta)^2 \quad (2.2)$$

is bounded on compact sets, the quadratic estimation error is bounded from below:

$$\mathbb{E}^{X^\theta} (\hat{\theta}_T - \theta)^2 \geq \frac{1}{I_T(\theta)} \quad \text{for all } \theta \in \Theta.$$

The right-hand side is the Cramer–Rao lower bound. For a proof of this classical result, see KUTOYANTS [1984].

For example, consider the model

$$dS_t^\theta = \mu S_t^\theta dt + \sigma S_t^\theta dB_t.$$

Set $X_t^\theta := \frac{1}{\sigma} \log(S_t^\theta)$, that is,

$$dX_t^\theta = \theta dt + dB_t$$

with

$$\theta := \frac{\mu - \frac{\sigma^2}{2}}{\sigma}.$$

The Cramer–Rao lower bound implies that all estimator of θ based upon the observation of one trajectory of (S_t^θ) — equivalently, of (X_t^θ) — in the time interval $[0, T]$, has a quadratic estimation error larger than $\frac{1}{T}$. If the unit of time is 1 year and if one observes the stock prices during 1 year, then the standard deviation of the error cannot be lower than σ .

2.2. Testing whether the noise has jumps

In an impressive recent paper, AÏT-SAHALIA and JACOD [2008] constructed and analyzed a rule to decide whether a price process observed at discrete times is continuous or jumps at least once during the observation time interval. Their paper substantially improves previous works mentioned in its list of references.

The observed process (X_t) is supposed to belong to a fairly general class of models, namely, it is supposed to satisfy

$$\begin{aligned} X_t = X_0 &+ \int_0^t b_s ds + \int_0^t \sigma_s dB_s + \int_0^t \int_{\mathbb{R}} \kappa \circ \delta(s, x) (\mu - \nu)(ds, dx) \\ &+ \int_0^t \int_{\mathbb{R}} (\delta(s, x) - \kappa \circ \delta(s, x)) \mu(ds, dx). \end{aligned} \quad (2.3)$$

Here, B is a Brownian motion, μ is a Poisson random measure with an intensity measure of the form $\nu(ds, dx) = ds \otimes dx$; the function κ is continuous and locally equal to x around the origin; the processes (b_s) and (σ_s) are optional, and the random function $\delta(s, \cdot)$ is predictable and uniformly bounded in ω and time by a deterministic function γ such that $\int_{\mathbb{R}} \min(\gamma(x)^2, 1) dx < \infty$. The authors require a few technical conditions that are not limitative for applications in finance (e.g., the process (σ_t) is supposed to be of the same type as (X_t) itself).

Now, denote by Δ_n a sequence of observation time steps decreasing to 0. Aït-Sahlia and Jacod's test statistics is

$$\hat{C}(p, \Delta_n)_t := \frac{\sum_{i=1}^{t/\Delta_n} |X_{2i\Delta_n} - X_{2(i-1)\Delta_n}|^p}{\sum_{k=1}^{t/\Delta_n} |X_{i\Delta_n} - X_{(i-1)\Delta_n}|^p}.$$

THEOREM 2.1. *Under the above assumptions, for all $t > 0$ and $p > 2$, the variables $\hat{C}(p, \Delta_n)_t$ converge in probability when n goes to infinity to*

$$\mathbb{I}_{\{\omega; s \rightarrow X_s(\omega) \text{ is continuous on } [0, t]\}} + 2^{\frac{p}{2}-1} \mathbb{I}_{\{\omega; s \rightarrow X_s(\omega) \text{ is discontinuous on } [0, t]\}}.$$

Therefore, the decision rule consists in accepting the hypothesis “the process (X_t) is discontinuous” if $\hat{C}(p, \Delta_n)_t < \frac{1+2^{p/2-1}}{2}$, and rejecting it if $\hat{C}(p, \Delta_n)_t \geq \frac{1+2^{p/2-1}}{2}$.

The authors prove several limit theorems that allow them to construct levels of tests based on their tests statistics. In particular, they show the following theorem.

THEOREM 2.2. *For $p > 3$, $\Delta_n^{-1/2}(\hat{C}(p, \Delta_n)_t - 1)$, when restricted to the set of discontinuous paths, converges stably in law.*

If X is continuous, for $p \geq 2$, $\Delta_n^{-1/2}(\hat{C}(p, \Delta_n)_t - 2^{p/2-1})$ converges stably in law.

In both cases, the limits are constructed on an extension of the original probability space, but their conditional distribution w.r.t. the original filtration is Gaussian; the two conditional variances are explicated in terms of respectively,

$$\frac{\int_{s \leq t} |X_s - X_{s-}|^{2p-2} (\sigma_s^2 + \sigma_{s-}^2)}{\left(\int_{s \leq t} |X_s - X_{s-}|^p \right)^2}$$

and

$$\frac{\int_0^t |\sigma_s|^{2p} ds}{\left(\int_0^t |\sigma_s|^p ds\right)^2}.$$

These two asymptotic variances can be estimated by means of the discrete time observations of X .

It is consequently possible to construct real tests for the null hypothesis that X is discontinuous as well as for the null hypothesis that X is continuous. For precise critical regions, asymptotic levels and power functions, we refer to AÏT-SAHALIA and JACOD [2008]. Simulation studies reported in the paper illustrate that observations at high frequencies actually allow one to discriminate continuous and discontinuous models. Similarly, when applied to real historical data (Dow Jones Industrial Average stock prices in 2005), observations each 5 seconds lead to the conclusion that most of the prices should be modeled by models with jumps. However, as predicted by the theoretical results, observations each 30 seconds do not allow one to get a significant information from the test.

In conclusion, although Brownian models are commonly used to compute prices and deltas, it seems that driving noises with jumps should also be considered, especially for prices or physical variables observed at low frequencies since, in such a case, it is impossible to test the (dis)continuity hypothesis.

2.3. *The explicative Brownian dimension of a stochastic model*

Suppose now that one observes prices of a basket of d assets and that these prices are Itô processes driven by a q -dimensional Brownian motion. If no arbitrage and completeness are assumed, then $d = q$. However, it sometimes is useless to calibrate a volatility matrix of dimension d : for example, some components of the noise may play a very small role in the dynamics of the price and, consequently, considering that they are null may not change much the prices of options on the basket under consideration. More generally, one may have to calibrate models for families of processes that do not model prices but indices, meteorological or economical variables, etc, for which the number of random sources is not constrained by no arbitrage or completeness conditions. In all cases, by eliminating “small” noises in the dynamics, one simplifies the calibration of the volatility matrix and decreases the number of operations in the simulations of the model.

JACOD, LEJAY and TALAY [2008] have tackled the question of estimating the “explicative Brownian dimension” of an Itô process from a discrete time observation. By “explicative Brownian dimension r_B ,” we (informally) mean that a model driven by r_B -dimensional Brownian motion satisfyingly fits the information conveyed by the observed path, whereas increasing the Brownian dimension does not bring a better fit.

More precisely, suppose that we observe a path of the process

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s,$$

where B is a standard q -dimensional Brownian Motion, (b_s) is a predictable $\mathbb{R}^d$ -valued locally bounded process, σ is a $d \times q$ matrix-valued adapted and càdlàg processes. Set $c_s := \sigma_s \sigma_s^*$. Our aim is to estimate the maximal explicative rank of c_s on the basis of the observation of $X_{iT/n}$ for $i = 0, 1, \dots, n$. Of course, a natural candidate should resemble the integer such that, if $\lambda(1)_s, \dots, \lambda(d)_s$ are the eigenvalues of c_s in decreasing order, then $\lambda(r_B)_s$ is significantly larger than $\lambda(r_B + 1)_s$. However, this sole definition does not lead to a tractable test since one observes a trajectory of (X_t) and not of (c_t) ; in particular, this implies that we cannot hope to approximate the eigenvalues of c_s with a good accuracy. Therefore, we need to define estimators of the maximal explicative rank or tests based upon observations of (X_t) . Notice also that, as in the preceding section, these observations are at discrete times only.

We start with a linear algebra observation. Let $\mathcal{A}_r$ be the family of all subsets of $\{1, \dots, d\}$ with r elements. For all $K \in \mathcal{A}_r$ and $d \times d$ symmetric nonnegative matrix Σ , let $\text{determinant}_K(\Sigma)$ be the determinant of the $r \times r$ submatrix $(\Sigma^{kl} : k, l \in K)$ and set

$$\text{determinant}(r; \Sigma) := \sum_{K \in \mathcal{A}_r} \text{determinant}_K(\Sigma).$$

It is easy to prove that the eigenvalues

$$\lambda(1) \geq \dots \lambda(d) \geq 0$$

of Σ satisfy for all $r = 1, \dots, d$:

$$\frac{1}{d(d-1) \dots (d-r+1)} \text{determinant}(r; \Sigma) \leq \lambda(1)\lambda(2) \dots \lambda(r) \leq \text{determinant}(r; \Sigma).$$

In addition,

$$1 \leq r \leq d \implies \begin{cases} r \leq \text{rank}(\Sigma) \implies \text{determinant}(r; \Sigma) > 0 \\ r > \text{rank}(\Sigma) \implies \text{determinant}(r; \Sigma) = 0, \end{cases}$$

and

$$\begin{aligned} 2 \leq r \leq d &\implies \frac{r!}{d!} \frac{\text{determinant}(r; \Sigma)}{\text{determinant}(r-1; \Sigma)} \leq \lambda(r) \\ &\leq \frac{d!}{(r-1)!} \frac{\text{determinant}(r; \Sigma)}{\text{determinant}(r-1; \Sigma)}. \end{aligned}$$

Now, set

$$\mathcal{L}(r)_t := \int_0^t \text{determinant}(r; c_s) ds.$$

In view of the preceding inequalities, for choosing an explicative Brownian dimension, this quantity plays a role similar to

$$\bar{L}(r)_t := \int_0^t \lambda(1)_s \dots \lambda(r)_s ds.$$

We approximate $\mathcal{L}(r)_t$ by means of our observations of X : denoting by $[x]$ the integer part of x , we set

$$\mathcal{L}(r)_t^n := \frac{n^{r-1}}{T^{r-1}} \sum_{i=1}^{[nt/T]-r+1} \text{determinant}(r; \zeta(r)_i^n),$$

where

$$\zeta(r)_i^n = \sum_{j=1}^r (\Delta_{i+j-1}^n X) (\Delta_{i+j-1}^n X)^*, \text{ with } \Delta_\ell^n X = X_{\ell T/n} - X_{(\ell-1)T/n}.$$

THEOREM 2.3. *The variables $\mathcal{L}(r)_t^n$ converge in probability to $\mathcal{L}(r)_t$ uniformly in $t \in [0, T]$. The processes*

$$V(r)_t^n := \sqrt{n} (\mathcal{L}(r)_t^n - \mathcal{L}(r)_t)$$

converge stably in law to a limiting process $(V(r)_t)_{1 \leq r \leq d}$, which is defined on, an extension of the original space and is a nonhomogeneous Wiener process with an “explicit” quadratic variation process.

Set

$$R(\omega)_t := \sup_{s \in [0, t]} \text{rank}(c_s(\omega)).$$

We define a scale invariant estimator of R_t by

$$\mathcal{R}_{n,t} := \inf \left(r \in \{0, \dots, d-1\} : \mathcal{L}(r+1)_t^n < \rho_n t^{-1/r} (\mathcal{L}(r)_t^n)^{(r+1)/r} \right).$$

The preceding theorem allows one to propose a test based on a scale invariant relative threshold for which we have the following consistency result under reasonably weak assumptions on the coefficients (b_s) and (σ_s) (more or less similar to those made in the preceding section):

THEOREM 2.4. *For all r, r' in $\{1, \dots, d\}$, provided $\mathbb{P}(R_t = r') > 0$, we have*

$$\mathbb{P}(\mathcal{R}_{n,t} \neq r \mid R_t = r') \longrightarrow \begin{cases} 1 & \text{if } r \neq r', \\ 0 & \text{if } r = r'. \end{cases}$$

Empirical studies for this test and a couple of other tests applied to simulations of models with stochastic volatilities are reported in JACOD, LEJAY and TALAY [2008]. They illustrate that, under circumstances such as observations at low frequencies or systems with strongly oscillating components, the tests may lead to very erroneous conclusions. In any case, the transformation of the real Brownian dimension into an explicative one induces a specific model risk.

3. On calibration methods in finance

Practitioners do not only use estimators based on historical observations of primary assets but also use all the information available on the market, for example, prices of derivatives on the asset under consideration, prices of correlated assets, and forward contracts. Their data set is thus a sample χ of a random vector ξ , which represents market prices of all such products.

Various approaches have been developed by various authors: inverse problem techniques applied to the PDEs for option prices, numerical resolution of Dupire's PDE for the volatility function, optimization techniques to fit the data, entropy minimization techniques, etc.

We first briefly describe Avellaneda–Friedman–Holmes–Samperi's approach for the calibration of volatilities (for more details on this approach and other approaches, see AVELLANEDA, FRIEDMAN, HOLMES and SAMPERI [1997] and the volume edited by AVELLANEDA [2001], and references therein). Consider an asset whose volatility process (σ_t) is progressively measurable and satisfies

$$0 < \underline{\sigma} \leq \sigma_t \leq \bar{\sigma}$$

for some deterministic constants $\underline{\sigma}$ and $\bar{\sigma}$. The set of all such processes is denoted by $\mathcal{H}$. Suppose that the market is complete and that various European options are priced on the market, all the maturities belong to the time interval $[0, T]$. Avellaneda's approach consists in choosing a smooth and strictly convex function H defined on $\mathbb{R}_+$ with minimal value 0 at a given value σ_0 (resulting from statistics based on historical data) and then searching the process (σ_t) , which solves

$$\sup_{(\sigma_t) \in \mathcal{H}} -\mathbb{E}^\sigma \int_0^T \exp(-r\theta) H((\sigma_\theta)^2) d\theta.$$

Denote the observed option prices by P_k , their maturities by T_k , and their payoff functions by Φ_k . Then, set

$$f(\sigma.) := -\mathbb{E}^\sigma \int_0^T \exp(-r\theta) H((\sigma_\theta)^2) d\theta,$$

$$g_k(\sigma.) := \mathbb{E}^\sigma (\exp(-rT_k) \Phi_k(S_{T_k})).$$

The calibration procedure consists in solving

$$\sup_{(\sigma_t) \in \mathcal{H}} \inf_{\mu_k} (f(\sigma.) + \sum_k \mu_k (g_k(\sigma.) - P_k)).$$

For a discussion on the corresponding numerical procedures and a survey on other numerical techniques for calibration, see ACHDOU and PIRONNEAU [2005].

Another direction has been followed by El Karoui and Hounkpatin (see HOUNKPATIN [2002]) to calibrate risk premia rather than volatilities. The El Karoui–Hounkpatin's method is based on a variant of the selection of models by minimizing

entropies as introduced in AVELLANEDA, FRIEDMAN, HOLMES and SAMPERI [1997]. Let $\mathbb{X}$ be the state space of a random vector ξ , which represents market prices of products related to the asset under consideration (e.g., forward contracts, derivatives, ...). We observe one sample χ of this random vector. Define the set A of calibration measures as

$$\mathcal{P}_\chi := \left\{ \mathbb{Q} \text{ probability on } \mathbb{X} \text{ equivalent to } \mathbb{P}, \mathbb{E}^{\mathbb{Q}}[\xi] = \chi \right\}.$$

How to choose an “optimal” element of $\mathcal{P}_\chi$? Consider the entropy

$$H(\mathbb{Q}, \mathbb{P}) := \int \log \left\{ \frac{d\mathbb{Q}}{d\mathbb{P}} \right\} d\mathbb{Q} \quad \text{if } \mathbb{Q} \ll \mathbb{P}, \quad +\infty \text{ otherwise.}$$

Observe that $H(\mathbb{Q}, \mathbb{P})$ is positive, and that $H(\mathbb{Q}, \mathbb{P}) = 0$ iff $\mathbb{P} = \mathbb{Q}$.

Suppose that the asset price solves

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t.$$

For a vector $\xi := (\xi^i)$ of the form $x^i = \phi_i(X_T)$, set

$$h(t, x, \lambda) := \mathbb{E}^{\mathbb{P}} \left[\frac{\exp(\sum_{i=1}^N \lambda_i \xi^i)}{\mathbb{E}^{\mathbb{P}} \exp(\sum_{i=1}^N \lambda_i \xi^i)} \middle| X_t = x \right].$$

Using results of CSISZAR [1975, theorem 3.1], El Karoui and Hounkpatin have shown that there exists a unique $\mathbb{Q}^*$ in the set of calibration measures such that

$$H(\mathbb{Q}^*, \mathbb{P}) = \inf_{\mathbb{Q} \in A} H(\mathbb{Q}, \mathbb{P}),$$

and the dynamics of (X_t) under $\mathbb{Q}^*$ is

$$dX_t = (b(t, X_t) + \sigma(t, X_t)^2 \partial_x \log h(t, X_t, \lambda^*))dt + \sigma(t, X_t)dB_t^*, \quad t \leq T,$$

where (B_t^*) is a Brownian motion under $\mathbb{Q}^*$, and λ^* solves

$$\max_{\lambda \in \mathbb{R}^N} \left\{ \sum_{i=1}^N \lambda_i \chi_i - \log \mathbb{E}^{\mathbb{P}} \exp \left(\sum_{i=1}^N \lambda_i \xi^i \right) \right\}.$$

The numerical approximation of λ^* , $h(t, x, \lambda^*)$, and $\frac{\partial h}{\partial x}(t, x, \lambda^*)$ is theoretically possible owing to Monte Carlo methods. It is a challenging and interesting question to design an efficient algorithm.

It actually appears that all calibration measures lead to difficult numerical problems that getting good accuracies is questionable. We also emphasize that the family of calibration probability measures is arbitrarily chosen. For these reasons, calibration measures, as statistical procedures, cannot cancel model risk.

4. On Monte Carlo approximations of the VaR of model risk P&Ls

4.1. Model risk P&Ls for misspecified hedging strategies in Markovian markets

In this subsection, we follow BOSSY, GIBSON, LHABITANT, PISTRE and TALAY [2006] where numerical results for the P&L function below are discussed.

Consider a primary asset with price processes S and a saving account (or, more generally, a numéraire) with price process F defining a no arbitrage and complete market. Denote by S_t^F the price of the primary assets expressed in this numéraire. We suppose that, up to a change in probability from $\mathbb{P}$ to a new probability $\mathbb{P}^F$, the price S_t^F is a $\mathcal{F}$ -martingale and defines a no arbitrage and complete market.

Consider a trader who needs to hedge a European option on the primary assets with maturity T^O and payoff function ϕ . At all time, $0 \leq t \leq T^O$ the perfectly hedging portfolio consists in H_t^0 units of the saving account and the H_t units of the primary assets:

$$V_t = H_t^0 F_t + H_t S_t.$$

Its value expressed in the numéraire F is

$$V_t^F = H_t^0 + H_t S_t^F.$$

The self-financing condition writes

$$V_t^F = V_0^F + \int_0^t H_\theta dS_\theta^F.$$

As the martingale S^F is supposed to define a complete market and thus to satisfy the martingale representation property, the preceding equality provides a characterization of the process H . However, this characterization generally is only implicit, even when Clark–Ocone formula (see NUALART [2006]) applies, and thus when H_t can be expressed by means of conditional expectations of Malliavin derivatives, and its numerical approximation is quite difficult. Thus, would the market asset prices be a general semi-martingale, even if the trader would perfectly know and measure the model, he/she would nevertheless use a simpler model that would allow him/her to easily, and in short computational time, get numerical values for the delta.

To this end, most often traders consider Markov models: they are given a filtration $\overline{\mathcal{F}}$ and a probability $\overline{\mathbb{P}}^F$, a $\overline{\mathcal{F}}$ -Markov–Feller process (ρ_t) , and functions g and h , and they admit that $S_t^F = g(t, \rho_t)$, $F_t = h(t, \rho_t)$, and

$$dS_t^F = \Psi(t, \rho_t) dB_t^F, \quad (4.1)$$

$$d\rho_t = \beta(\rho_t) dt + \gamma(t, \rho_t) dB_t^F, \quad (4.2)$$

for some functions Ψ , β , and γ , and for some Brownian motion B^F . The process ρ may be S^F itself or the instantaneous rate if S^F is a forward bond price in a one-factor

short-term rate model or a vector of factors. The model is constrained in such a way that there exists a smooth function $\bar{\pi}(t, x)$ solution to

$$\frac{\partial \bar{\pi}}{\partial t}(t, x) + \bar{\mathcal{L}}_t^\rho \bar{\pi}(t, x) = 0, \quad T < T^O, \quad (4.3)$$

where $\bar{\mathcal{L}}_t^\rho$ is the infinitesimal operator of ρ under $\bar{\mathbb{P}}^F$. The boundary condition¹ is

$$\bar{\pi}(T^O, x) = \frac{\phi(h(T^O, x)g(T^O, x))}{h(T^O, x)}.$$

With the above notation, would the true world be actually governed by (ρ_t) , then risk could be eliminated through the delta-hedge actually used by the trader, that is, $\bar{H}_t = \frac{\partial \bar{\pi}}{\partial x}(t, S_t^F)$. Therefore, the self-financed “pseudoreplicating” portfolio has value

$$d\bar{V}_t^F = \bar{H}_t dS_t^F.$$

The model risk P&L function is defined as

$$\text{P\&L}_t^F = \bar{V}_t^F - V_t^F. \quad (4.4)$$

Suppose that, in the true world, the process (ρ_t) is a (not necessarily Markov) Itô process satisfying under $\mathbb{P}^F$:

$$d\rho_t = \beta_t dt + \gamma_t dB_t^F$$

for some adapted processes β and γ . Set

$$\mathcal{L}_t^\rho \bar{\pi}(t, \rho_t) := \beta_t \frac{\partial \bar{\pi}}{\partial x}(t, \rho_t) + \frac{1}{2}(\gamma_t)^2 \frac{\partial^2 \bar{\pi}}{\partial x^2}(t, \rho_t). \quad (4.5)$$

A simple calculation shows that, at maturity T^O ,

$$\text{P\&L}_{T^O}^F = \bar{V}_{T^O}^F - V_{T^O}^F = \bar{V}_0^F - \bar{\pi}(0, \rho_0) + \int_0^{T^O} (\bar{\mathcal{L}}_\theta^\rho - \mathcal{L}_\theta^\rho) \bar{\pi}(\theta, \rho_\theta) d\theta. \quad (4.6)$$

Notice that, if (ρ_t) is a Markov process, that is, if $\beta_t = \beta(t, \rho_t)$ and $\gamma_t = \gamma(t, \rho_t)$ for some functions β and γ , then $\mathcal{L}_t^\rho$ is the classical infinitesimal generator of (ρ_t) and $V_t^F = \pi(t, \rho_t)/F_t$, where $\pi(t, x)$ solves a PDE similar to (4.3) with $\mathcal{L}_t^\rho$ replacing $\bar{\mathcal{L}}_t^\rho$.

4.2. Approximation of quantiles of diffusion processes

The discussion in the preceding subsection can obviously be extended to multidimensional market models, where the prices $dS^{F,i}$ are deterministic functions of a vector of Markov factors (ρ_t^i) , and basket options based on the prices S^i and with maturity T^O .

¹ Observe that

$$V_T^F = H_T^0 + H_T S_T^F = \frac{1}{F_T} \phi(F^T S_T^F).$$

One desires to get statistical information on $\text{P\&L}_{T^0}^F$. A commonly used statistics is its VaR. This leads to the question of approximating quantiles of the law of one component of multidimensional Markov diffusion processes such as $(S_t^{F,i}, \rho_t^j, \text{P\&L}_t)$.

Consider the fairly general stochastic differential equation

$$X_t(x) = x + \int_0^t A_0(s, X_s(x))ds + \sum_{i=1}^r \int_0^t A_i(s, X_s(x))dB_s^i,$$

and suppose that the law of the last component, $X_T^d(x)$, has a density with respect to Lebesgue's measure. Here, (B_s) is an r -dimensional Brownian motion, and the functions $A_0, A_1, \dots, A_r$ are smooth with bounded derivatives.

The Euler scheme with step T/n is defined as

$$\begin{aligned} X_{(p+1)T/n}^n(x) &= X_{pT/n}^n(x) + A_0(pT/n, X_{pT/n}^n(x))\frac{T}{n} \\ &\quad + \sum_{i=1}^r A_i(pT/n, X_{pT/n}^n(x))(B_{(p+1)T/n}^i - B_{pT/n}^i). \end{aligned}$$

We add a small perturbation to $X_T^n(x)$ in order to get a random variable whose law has a density

$$\tilde{X}_T^n(x) = X_T^n(x) + B_{T+T/n} - B_T.$$

We aim to get error estimates on the approximation, by means of Monte Carlo simulations of $\tilde{X}_T^n(x)$, of the quantile of level δ , $\rho(x, \delta)$, of the law of $X_T^d(x)$.

When approximating quantities of the type $\mathbb{E}f(X_T)$, where T is fixed, we have the following result: for functions f with polynomial growth at infinity,

$$\mathbb{E}f(X_T) - \mathbb{E}f(X_T^n) = C_f(T, x)\frac{T}{n} + Q_n(f, T, x)\frac{1}{n^2}, \quad (4.7)$$

where

$$|C_f(T, x)| + \sup_n |Q_n(f, T, x)| \leq C(1 + \|x\|^Q)\frac{1 + K(T)}{T^q}$$

for some positive real numbers C, q , and Q and some increasing function K (see TALAY and TUBARO [1990] for smooth functions f , BALLY and TALAY [1995] under a uniform hypoellipticity condition on the fields A_i , KOHATSU-HIGA [2001] and GOBET and MUNOS [2005] for only measurable functions under nondegeneracy conditions on the Malliavin covariance matrix of $X_T(x)$). Thus, Romberg extrapolation techniques can be used to get higher convergence rates (see TALAY and TUBARO [1990]). For extensions to barrier options, see GOBET and MENOZZI [2004].

For the quantile approximation problem, the estimates are slightly different. We summarize results in TALAY and ZHENG [2002]. Suppose first that the stochastic differential equation (SDE) for (X_t) has time homogeneous coefficients:

$$X_t(x) = x + \int_0^t A_0(X_s(x))ds + \sum_{i=1}^r \int_0^t A_i(X_s(x))dB_s^i. \quad (4.8)$$

For multiindices $\alpha = (\alpha_1, \dots, \alpha_k) \in \{0, 1, \dots, r\}^k$, set $A_i^\emptyset = A_i$ and for $0 \leq j \leq r$, $A_i^{(\alpha, j)} := [A_j, A_i^\alpha]$. Also set

$$V_L(x, \eta) := \sum_{i=1}^r \sum_{|\alpha| \leq L-1} < A_i^\alpha(x), \eta >^2$$

and

$$V_L(x) := 1 \wedge \inf_{\|\eta\|=1} V_L(x, \eta).$$

Suppose

(UH) $C_L := \inf_{x \in \mathbb{R}^d} V_L(x) > 0$ for some integer L ,

(C) The coefficients A_i^j , $i = 0, \dots, r$, $j = 1, \dots, d$ are of class $\mathcal{C}_b^\infty(\mathbb{R}^d)$ (the A_i^j 's may be unbounded).

Under (UH) and (C), the law of $X_T(x)$ has a smooth density $p_T(x, x')$, so that the d -th marginal distribution of $X_T(x)$ also has a smooth density $p_T^d(x, y)$, which is strictly positive at all point y in the interior of its support (cf. NUALART [2006]).

For $0 < \delta < 1$, set

$$\rho(x, \delta) := \inf\{\rho \in \mathbb{R}; \mathbb{P}[X_T^d(x) \leq \rho] = \delta\}$$

and

$$\tilde{\rho}^n(x, \delta) := \inf\{\rho \in \mathbb{R}; \mathbb{P}[\tilde{X}_T^{n,d}(x) \leq \rho] = \delta\}.$$

The discretization error on the quantile $\rho(x, \delta)$ is described by the following theorem.

THEOREM 4.1. *Under conditions (UH) and (C), we have*

$$|\rho(x, \delta) - \tilde{\rho}^n(x, \delta)| \leq \frac{K(T)}{T^q} \cdot \frac{1 + \|x\|^Q}{\bar{p}_T^d(\rho(x, \delta))} \cdot \frac{1}{n}, \quad (4.9)$$

where

$$\bar{p}_T^d(\rho(x, \delta)) = \inf_{y \in (\rho(x, \delta) - 1, \rho(x, \delta) + 1)} p_T^d(x, y).$$

In practice, $\tilde{\rho}^n(x, \delta)$ is estimated by sampling N copies of $\tilde{X}_T^N$ (for variance reduction techniques, see KOHATSU-HIGA and PETTERSON [2002]). Taking the corresponding

Monte Carlo error into account, roughly speaking, the global error on the quantile is of order

$$\mathcal{O}\left(\frac{1}{\bar{p}_T^d(\rho(x, \delta))n}\right) + \mathcal{O}\left(\frac{1}{\tilde{p}_T^{n,d}(x, \rho(x, \delta))\sqrt{N}}\right),$$

where $\tilde{p}_T^{n,d}(x, \xi)$ denotes the density of $\tilde{X}_T^{n,d}(x)$.

One has (see BALLY and TALAY [1995], KOHATSU-HIGA [2001]) that $\tilde{p}_T^{n,d}(x, \xi) - p_T^d(x, \xi)$ is of order $1/n$. For practical applications, one thus needs accurate estimates from below $p_T^d(x, \rho(x, \delta))$. Such estimates are available when the generator of (X_t) is strictly uniform elliptic (see AZENCOTT [1984]), but this hypothesis is too stringent in our context: notice that the law of the above vector $(S_t^{F,i}, \rho_t^j, \text{P\&L}_t)$ may not have a density since all its components are driven by the Brownian processes driving the ρ^j s. Therefore, we now do not suppose that the Malliavin covariance matrix of $(X_t(x))$ is invertible and return to general inhomogeneous stochastic differential equation. Let $(X_s^t(x'), 0 \leq s \leq T - t)$ be a smooth version of the flow solution to

$$X_s^t(x') = x' + \int_0^s A_0(t + \theta, X_\theta^t(x'))d\theta + \sum_{i=1}^r \int_0^s A_i(t + \theta, X_\theta^t(x'))dB_{t+\theta}^i.$$

We denote by $M(t, s, x')$ the Malliavin covariance matrix of $X_s^t(x')$. We now suppose

(C') The functions A_i^j , $i = 0, \dots, r$, $j = 1, \dots, d$ are of class $\mathcal{C}_b^\infty([0, T] \times \mathbb{R}^d)$ (the A_i^j 's may be unbounded).

(M) For all $p \geq 1$, there exists a nondecreasing function K , a positive real number r , and a positive Borel measurable function Ψ such that

$$\left\| \frac{1}{M_d^d(t, s, x')} \right\|_p \leq \frac{K(T)}{s^r} \Psi(t, x')$$

for all t in $[0, T]$ and s in $(0, T - t]$. In addition, Ψ satisfies: for all $\lambda \geq 1$, there exists a function Ψ_λ such that

$$\sup_{t \in [0, T]} \mathbb{E}[\Psi(t, X_t(x))^\lambda] < \Psi_\lambda(x)$$

and

$$\sup_{n > 0} \sup_{t \in [0, T]} \mathbb{E}[\Psi(t, X_t^n(x))^\lambda] < \Psi_\lambda(x).$$

Under condition (M), the d -th marginal distribution of $X_T(x)$ has a smooth density. $p_T^d(x, y)$ is strictly positive at all point y in the interior of its support, and we have the following error estimate.

THEOREM 4.2. *Under conditions (M) and (C'), we have*

$$|\rho(x, \delta) - \tilde{\rho}^n(x, \delta)| \leq \frac{K(T)}{T^q} \cdot \frac{1 + \|x\|^Q}{\bar{p}_T^d(\rho(x, \delta))} \cdot \Psi_\lambda(x) \cdot \frac{1}{n},$$

where

$$\bar{p}_T^d(\rho(x, \delta)) = \inf_{y \in (\rho(x, \delta) - 1, \rho(x, \delta) + 1)} p_T^d(x, y).$$

In practice, one needs to check that condition (M) is satisfied. We here give two examples.

THEOREM 4.3. *Suppose that $\sum_{i=1}^r |A_i^d(t, x)|^2 \geq a > 0$ for some t in $[0, T]$ and x in $\mathbb{R}^d$. Then, the d -th marginal law of $X_t(x)$ has a smooth density, and condition (M) is satisfied.*

Our second example concerns a model risk problem. The trader wants to hedge a European option $\Phi(B(T^O, T))$ on a bond price $B(T^O, T)$, where T^O is the option maturity and $T > T^O$ is the bond maturity. To hedge, the trader uses bonds with maturities T^O and T . Suppose that the bond market is an HJM model. When the HJM model is governed by a deterministic function σ , the delta of the option can be expressed in terms of the solution π_σ to the PDE

$$\begin{cases} \frac{\partial \pi_\sigma}{\partial t}(t, x) + \frac{1}{2} x^2 (\sigma^*(t, T^O) - \sigma^*(t, T))^2 \frac{\partial^2 \pi_\sigma}{\partial x^2}(t, x) = 0, \\ \pi_\sigma(T, x) = \Phi(x). \end{cases}$$

Suppose that the trader chooses an erroneous deterministic model structure $\bar{\sigma}(s, T)$. Then, for suitable functions $u_1(s)$, $u_2(s)$, and $\varphi(s)$, the forward value of the trader's P&Ls satisfies an SDE of the type

$$d\text{P\&L}_t = \varphi(t, Y_t) Y_t u_1(t) dt + \varphi(t, Y_t) Y_t u_2(t) dB_t,$$

where (Y_t) satisfies

$$dY_t = Y_t u_1(t) dt + Y_t u_2(t) dB_t.$$

If

$$|\varphi(t, y) u_2(t)| \geq a > 0 \quad \forall t, \quad \forall y > 0,$$

then condition (M) is satisfied, and one can get an explicit lower bound estimate for the marginal density.

5. A stochastic game to face model risk

Consider the market model

$$\begin{cases} dS_t^i = S_t^i [b_t^i dt + \sum_{j=1}^d \sigma_t^{ij} dB_t^j] & \text{for } 0 \leq i \leq n, \\ dP_t = P_t \sum_{i=1}^n \pi_t^i \left[b_t^i dt + \sum_{j=1}^d \sigma_t^{ij} dB_t^j \right] + rP_t \left(1 - \sum_{i=1}^n \pi_t^i \right) dt. \end{cases}$$

Here $\{\pi^i\}$ = set of prescribed strategies. Consider $u(\cdot) := (b(\cdot), \sigma(\cdot))$ as the market's control process.

CVITANIĆ and KARATZAS [1999] have studied the dynamic measure of risks

$$\inf_{\pi(\cdot) \in \mathcal{A}(x)} \sup_{v \in \mathcal{D}} \mathbb{E}_v(F(X^{x,\pi}(T))),$$

where $\mathcal{A}(x)$ denotes the class of admissible portfolio strategies issued from the initial wealth x , and $\mathbb{E}_v$ denotes the expectation under the probability $\mathbb{P}_v$ for all v in a suitable set. All the measures $\mathbb{P}_v$ have the same risk-neutral equivalent martingale measure, which implies that the trader (or the regulator) is concerned by model risk on stock appreciation rates. For numerical methods related to this approach, see GAO, LIM and NG [2004].

An axiomatic approach to model risk is developed by CONT [2006], who proposes to measure model uncertainty risk by means of a coherent risk measure compatible with market prices of derivatives or of a convex risk measure. The author studies several examples, among them the case where the “real” noise is a linear combination of Poisson and Brownian processes, whereas the trader uses a Brownian model only.

We now present a somewhat different approach, based on a PDE and aimed to compute the minimal amount of money and dynamic strategies that allow the financial institution to (approximately) contain the worst possible damage due to model misspecifications for volatilities, stock appreciation rates, and yield curves. Within this approach, we consider that the trader acts as a minimizer of the risk, whereas the market systematically acts as a maximizer of the risk. Thus, the model risk control problem can be set up as a two-player zero-sum stochastic differential game problem. Given a suitable function F , the cost function is

$$J(t, x, p, \Pi, u(\cdot)) := \mathbb{E}_{t,x,p} F(S_T, P_T),$$

and the value function is

$$V(t, x, p) := \inf_{\Pi \in \text{Ad}_{\Pi}(t)} \sup_{u(\cdot) \in \text{Ad}_u(t)} J(t, x, p, \Pi, u(\cdot)).$$

The next theorem shows that this model risk value function solves an Hamilton–Jacobi–Bellman–Issacs equation.

THEOREM 5.1. *Under an appropriate locally Lipschitz condition on F , the value function $V(t, x, p)$ is the unique viscosity solution in the space*

$$S := \{\varphi(t, x, p) \text{ is continuous on } [0, T] \times \mathbb{R}^n \times \mathbb{R}; \exists \bar{A} > 0, \\ \lim_{|p|^2 + x^2 \rightarrow \infty} \varphi(t, x, p) \exp(-\bar{A} |\log(|p|^2 + x^2)|^2) = 0 \text{ for all } t \in [0, T]\}$$

to the Hamilton–Jacobi–Bellman–Isaacs equation

$$\begin{cases} \frac{\partial v}{\partial t}(t, x, p) + \mathcal{H}^-(D^2 v(t, x, p), Dv(t, x, p), x, p) = 0 & \text{in } [0, T) \times \mathbb{R}^{n+1}, \\ v(T, x, p) = F(x, p), \end{cases}$$

where

$$\mathcal{H}^-(A, z, x, p) := \max_{u \in K_u} \min_{\pi \in K_\pi} \left[\frac{1}{2} \text{Tr}(a(x, p, \sigma, \pi)A) + z \cdot q(x, p, b, \pi) \right].$$

For a proof, see TALAY and ZHENG [2002]. The numerical resolution of the PDE allows one to compute approximate reserve amounts of money to control model risk. Numerical investigations, undone so far, are necessary to evaluate how large are these provisions.

6. Model risk and technical analysis

The practitioners use various rules to rebalance their portfolios. These rules usually come from fundamental economic principles, mathematical approaches derived from mathematical models, or technical analysis approaches. Technical analysis, which provides decision rules based on past prices behavior, avoids model specification and thus model risk (for a survey, see ACHELIS [2001]). PASTUKHOV [2004] has studied mathematical properties of volatility indicators used in technical analysis. BLANCHET, DIOP, GIBSON, TALAY and ANRE [2007] proposed a framework allowing one to compare the performances obtained by strategies derived from erroneously calibrated mathematical models and the performances obtained by technical analysis techniques.

Consider an asset whose instantaneous expected rate of return changes at an unknown random time, and a trader who aims to maximize his/her utility of wealth by selling and buying the asset. The benchmark performance results from a strategy that is optimal when the model is perfectly specified and calibrated. To this benchmark we can compare the performances resulting from optimal rules but erroneous parameters, and the performances resulting from technical analysis indicators.

The real market is described by

$$\begin{cases} dS_t^0 = S_t^0 r dt, \\ dS_t = S_t (\mu_2 + (\mu_1 - \mu_2) \mathbb{I}_{(t \leq \tau)}) dt + \sigma S_t dB_t. \end{cases}$$

Here, the Brownian motion (B_t) and the change time τ are independent, and τ follows an exponential law with parameter λ . One has

$$S_t = S^0 \exp\left(\sigma B_t + \left(\mu_1 - \frac{\sigma^2}{2}\right)t + (\mu_2 - \mu_1) \int_0^t \mathbb{I}_{(\tau \leq s)} ds\right) =: S^0 \exp(R_t),$$

where

$$R_t = \sigma B_t + \left(\mu_1 - \frac{\sigma^2}{2}\right)t + (\mu_2 - \mu_1) \int_0^t \mathbb{I}_{(\tau \leq s)} ds.$$

Suppose

$$\mu_1 - \frac{\sigma^2}{2} < r < \mu_2 - \frac{\sigma^2}{2}.$$

We start with describing one of the technical analysis rules that are applied in the context of instantaneous rates of return changes. Denote by $\pi_t \in \{0, 1\}$ the proportion of the agent's wealth invested in the risky asset at time t , and by M_t^δ the moving average indicator of the prices. Therefore,

$$M_t^\delta = \frac{1}{\delta} \int_{t-\delta}^t S_u du.$$

Given a finite set of decision times t_n , at each t_n the agent invests all his/her wealth into the risky asset if $S_{t_n} > M_{t_n}^\delta$. Otherwise, he/she invests all the wealth into the riskless asset. Consequently,

$$\pi_{t_n} = \mathbb{I}_{(S_{t_n} \geq M_{t_n}^\delta)},$$

and the wealth at time t_{n+1} is

$$W_{t_{n+1}} = W_{t_n} \left(\frac{S_{t_{n+1}}}{S_{t_n}} \pi_{t_n} + \frac{S_{t_{n+1}}^0}{S_{t_n}^0} (1 - \pi_{t_n}) \right),$$

from which, for $T = t_M$,

$$W_T = W_0 \prod_{n=0}^{M-1} [\pi_{t_n} (\exp(R_{t_{n+1}} - R_{t_n}) - \exp(r\Delta t)) + \exp(r\Delta t)].$$

The logarithmic utility of W_T can be explicited in terms of the density of $(\int_0^t \exp(2B_s) ds, B_t)$: its explicit expression, according to Yor [2001], is interesting by itself: let $\sigma > 0$ and ν be real numbers, and let V be the geometric Brownian motion

$$V_t = e^{\sigma^2 \nu t + \sigma B_t}.$$

Then,

$$\mathbb{P}\left(\int_0^t V_s ds \in dy; V_t \in dz\right) = \frac{z^{v-1}}{2y} e^{-\frac{v^2 \sigma^2 t}{2} - \frac{(1+z^2)}{2\sigma^2 y}} i_{\frac{\sigma^2 t}{2}}\left(\frac{z}{\sigma^2 y}\right) dy dz, \quad (6.1)$$

where

$$i_y(z) := \frac{z e^{\pi^2/4y}}{\pi \sqrt{\pi y}} \int_0^\infty e^{-z \cosh(u) - u^2/4y} \sinh(u) \sin(\pi u/2y) du.$$

The performance of the technical analysis strategy is compared to the benchmark performance: the optimal wealth of a trader who perfectly knows the parameters μ_1 , μ_2 , λ , and σ . We impose constraints: as a technical analyst is only allowed to invest all his/her wealth in the stock or the bond, the proportions of the benchmark trader's wealth invested in the stock are constrained to lie within the interval $[0, 1]$. In addition, the trader's strategy is constrained to be adapted with respect to the filtration

$$\mathcal{F}_t^S := \sigma(S_u, 0 \leq u \leq t)$$

generated by (S_t) , which because of τ , is different from the filtration generated by (B_t) .

Let π_t be the proportion of the trader's wealth invested in the stock at time t ; $W^{x,\pi}$ denotes the corresponding wealth process. Let $\mathcal{A}(x)$ denote the set of admissible strategies, that is,

$$\begin{aligned} \mathcal{A}(x) &:= \{\pi. - \mathcal{F}_t^S - \text{progressively measurable process such that} \\ &W_0^{x,\pi} = x, W_t^{x,\pi} > 0 \text{ for all } t > 0, \pi. \in [0, 1]\}. \end{aligned}$$

The value function is

$$V(x) := \sup_{\pi. \in \mathcal{A}(x)} \mathbb{E} U(W_T^\pi).$$

As in KARATZAS and SHREVE [1998], we introduce an auxiliary unconstrained market defined as follows. Let $\mathcal{D}$ the subset of the $\{\mathcal{F}_t^S\}$ -progressively measurable processes $v : [0, T] \times \Omega \rightarrow \mathbb{R}$ such that

$$\mathbb{E} \int_0^T v^-(t) dt < \infty, \text{ where } v^-(t) := -\inf(0, v(t)).$$

The bond price process $S^0(v)$ and the stock price $S(v)$ satisfy

$$\begin{aligned} S_t^0(v) &= 1 + \int_0^t S_u^0(v)(r + v^-(u)) du, \\ S_t(v) &= S_0 + \int_0^t S_u(v) ((\mu_1 + (\mu_2 - \mu_1)F_u + v(u)^- + v(u)) du + \sigma d\bar{B}_u), \end{aligned}$$

where $\bar{B}_t$ is the innovation process, that is, the $\mathcal{F}_t^S$ Brownian motion defined as

$$\bar{B}_t = \frac{1}{\sigma} \left(R_t - (\mu_1 - \frac{\sigma^2}{2})t - (\mu_2 - \mu_1) \int_0^t F_s ds \right), \quad t \geq 0;$$

here, F is the conditional a posteriori probability (given the observation of S) that τ has occurred within $[0, t]$:

$$F_t := \mathbb{P}(\tau \leq t / \mathcal{F}_t^S).$$

For each auxiliary unconstrained market driven by a process v , the value function is

$$V(v, x) := \sup_{\pi \in \mathcal{A}(v, x)} \mathbb{E}_x U(W_T^\pi(v)),$$

where

$$dW_t^\pi(v) = W_t^\pi(v) \left((r + v^-(t))dt + \pi_t(v(t)dt + (\mu_2 - \mu_1)F_t dt + (\mu_1 - r)dt + \sigma d\bar{B}_t) \right).$$

Let the exponential likelihood ratio process $(L_t)_{t \geq 0}$ be defined by

$$L_t = \exp \left\{ \frac{\mu_2 - \mu_1}{\sigma^2} R_t - \frac{1}{2\sigma^2} \left((\mu_2 - \mu_1)^2 + 2(\mu_2 - \mu_1)(\mu_1 - \frac{\sigma^2}{2}) \right) t \right\}.$$

KARATZAS and SHREVE [1998] have proven the following result.

THEOREM 6.1. *If there exists $\tilde{v}$ such that*

$$V(\tilde{v}, x) = \inf_{v \in \mathcal{D}} V(v, x),$$

then there exists an optimal portfolio π^* for which the optimal wealth (for the constrained admissible strategies) is

$$W_t^* = W_t^{\pi^*}(\tilde{v}).$$

An optimal portfolio allocation strategy is

$$\pi_t^* := \sigma^{-1} \left(\frac{\mu_1 - r + (\mu_2 - \mu_1)F_t + \tilde{v}(t)}{\sigma} + \frac{\phi_t}{H_t^{\tilde{v}} W_t^* e^{-rt - \int_0^t \tilde{v}^-(s) ds}} \right),$$

where $H_t^{\tilde{v}}$ is the exponential process

$$\begin{aligned} H_t^{\tilde{v}} = \exp & \left(- \int_0^t \left(\frac{\mu_1 - r + \tilde{v}(s)}{\sigma} + \frac{(\mu_2 - \mu_1)F_s}{\sigma} \right) d\bar{B}_s \right. \\ & \left. - \frac{1}{2} \int_0^t \left(\frac{\mu_1 - r + \tilde{v}(s)}{\sigma} + \frac{(\mu_2 - \mu_1)F_s}{\sigma} \right)^2 ds \right), \end{aligned}$$

and ϕ is a $\mathcal{F}_t^S$ -adapted process, which satisfies

$$\mathbb{E} \left(H_t^{\tilde{v}} e^{-rT - \int_0^T \tilde{v}^-(t) dt} (U')^{-1} (v H_t^{\tilde{v}} e^{-rT - \int_0^T \tilde{v}^-(t) dt}) / \mathcal{F}_t^S \right) = x + \int_0^t \phi_s d\bar{B}_s.$$

Here, v is the Lagrange multiplier, which makes the expectation of the left-hand side equal to x for all x . In addition, F_t satisfies

$$F_t = \frac{\lambda e^{\lambda t} L_t \int_0^t e^{-\lambda s} L_s^{-1} ds}{1 + \lambda e^{\lambda t} L_t \int_0^t e^{-\lambda s} L_s^{-1} ds}.$$

The optimal strategies for the constrained problem are the projections on $[0, 1]$ of the optimal strategies for the unconstrained problem. In addition, using again Yor's Eq. (6.1), one can explicit $W_t^{*,x}$ and π_t^* in the case of the logarithmic utility.

For general utilities, the optimal strategy cannot be explicited. It, thus, is worth considering the case of a trader who chooses to reinvest the portfolio only once, namely at the time when the change time τ is optimally detected owing to the price history. We suppose that the reinvestment rule is the same as the technical analyst's one: at the detected change time from μ_1 to μ_2 , all the portfolio is reinvested in the risky asset. The stopping rule Θ^K , which minimizes the expected miss $\mathbb{E}|\Theta - \tau|$ over all the stopping rules Θ with $\mathbb{E}(\Theta) < \infty$, is as follows:

$$\Theta^K = \inf \left\{ t \geq 0 \mid \lambda e^{\lambda t} L_t \int_0^t e^{-\lambda s} L_s^{-1} ds \geq \frac{p^*}{1 - p^*} \right\},$$

where p^* is the unique solution in $(\frac{1}{2}, 1)$ of the equation

$$\int_0^{1/2} \frac{(1-2s)e^{-\beta/s}}{(1-s)^{2+\beta}} s^{2-\beta} ds = \int_{1/2}^{p^*} \frac{(2s-1)e^{-\beta/s}}{(1-s)^{2+\beta}} s^{2-\beta} ds$$

with $\beta = 2\lambda\sigma^2/(\mu_2 - \mu_1)^2$ (see SHIRYAEV [2004] and references therein). Up to a numerical approximation of p^* , this rule can easily be applied.

In practice, even if we would be able to estimate μ_1 and σ with a good accuracy, the value of μ_2 cannot be determined a priori, and the number of observations of τ may be too small to well estimate λ . Therefore, traders believe that the stock price is

$$dS_t = S_t (\bar{\mu}_2 + (\bar{\mu}_1 - \bar{\mu}_2) \mathbb{I}_{t \leq \bar{\tau}}) dt + \bar{\sigma} S_t dB_t,$$

where the law of $\bar{\tau}$ is exponential with parameter $\bar{\lambda}$. The above decision rules are then governed by

$$\begin{aligned} \bar{L}_t &= \exp \left\{ \frac{1}{\bar{\sigma}^2} (\bar{\mu}_2 - \bar{\mu}_1) R_t - \frac{1}{2\bar{\sigma}^2} \left((\bar{\mu}_2 - \bar{\mu}_1)^2 + 2(\bar{\mu}_2 - \bar{\mu}_1)(\bar{\mu}_1 - \frac{\bar{\sigma}^2}{2}) \right) t \right\}, \\ \bar{F}_t &= \frac{\bar{\lambda} e^{\bar{\lambda} t} \bar{L}_t \int_0^t e^{-\bar{\lambda} s} \bar{L}_s^{-1} ds}{1 + \bar{\lambda} e^{\bar{\lambda} t} \bar{L}_t \int_0^t e^{-\bar{\lambda} s} \bar{L}_s^{-1} ds}. \end{aligned}$$

Actually, the value of a misspecified optimal allocation strategy is

$$\bar{\pi}_t^* = \text{proj}_{[0,1]} \frac{(\bar{\mu}_1 - r + (\bar{\mu}_2 - \bar{\mu}_1)\bar{F}_t)}{\bar{\sigma}^2},$$

and the corresponding wealth is

$$\bar{W}_t^* = e^{rt} \exp\left(\int_0^t \bar{\pi}_u^* d(e^{-ru} S_u)\right).$$

Similarly, the erroneous stopping rule is

$$\bar{\Theta}^K = \inf\left\{t \geq 0, \bar{\lambda} e^{\bar{\lambda} t} \bar{L}_t \int_0^t e^{-\bar{\lambda} s} \bar{L}_s^{-1} ds \geq \frac{\bar{p}^*}{1 - \bar{p}^*}\right\},$$

where $\bar{p}^*$ is the unique solution in $(\frac{1}{2}, 1)$ of

$$\int_0^{1/2} \frac{(1-2s)e^{-\bar{\beta}/s}}{(1-s)^{2+\bar{\beta}}} s^{2-\bar{\beta}} ds = \int_{1/2}^{\bar{p}^*} \frac{(2s-1)e^{-\bar{\beta}/s}}{(1-s)^{2+\bar{\beta}}} s^{2-\bar{\beta}} ds,$$

with $\bar{\beta} = 2\bar{\lambda}\bar{\sigma}^2/(\bar{\mu}_2 - \bar{\mu}_1)^2$. The value of the corresponding portfolio is

$$\bar{W}_T = W_0 S_{\bar{\Theta}^K}^0 \frac{S_T}{S_{\bar{\Theta}^K}} \mathbb{I}_{(\bar{\Theta}^K \leq T)} + W_0 S_T^0 \mathbb{I}_{(\bar{\Theta}^K > T)}.$$

In view of the technical analysis technique and misspecified strategies, it is natural to compare them to the benchmark optimal strategy and to study the following question: Is it better to invest according to a mathematical strategy based a misspecified model or according to a strategy based on technical analysis rules? It appears that, even in the logarithmic utility case, the explicit formulae for the different wealths are too complex to allow analytical comparisons. However, Monte Carlo simulations on study cases show that the technical analyst may overperform misspecified optimal allocation strategies even when for relatively small misspecifications, for example, when the parameter λ is underestimated. Simulations also show that a single misspecified parameter is not sufficient to allow the technical analyst to overperform the traders who use erroneous stopping rules. One can also observe that, when the ratio μ_2/μ_1 decreases, the performances of well-specified and misspecified strategies based upon stopping rules decrease.

7. Conclusion

We have shown that statistical and calibration procedures can hardly reduce model uncertainties in finance. We have also emphasized that model uncertainties appear in the numerical resolution of PDE related to option pricing or optimal allocation problems. We have reviewed a few approaches to evaluate model risk indicators and to control model risk. We have discussed the accuracy of Monte Carlo methods to approximate

VaR statistics in diffusion models. Finally, we have proposed a mathematical framework to compare technical analysis techniques and strategies derived from misspecified mathematical models.

Most of the results that are stated above are recent and open new challenging perspectives in financial mathematics and in numerical analysis. Decreasing and controlling model risk is actually an important issue to make financial strategies more reliable.

References

- ACHDOU, Y., PIRONNEAU, O. (2005). *Computational Methods for Option Pricing, Frontiers in Applied Mathematics* (SIAM, Philadelphia, PA).
- ACHELIS, S. (2001). *Technical Analysis from A to Z* (McGraw Hill).
- AÏTHSAHLIA, Y., JACOD, J. (2008). Testing for jumps in a discretely observed process. *Ann. Stat.*, forthcoming.
- AÏT-SAHALIA, Y., KIMMEL, R. (2007). Maximum likelihood estimation of stochastic volatility models. *J. Financ. Econ.* **83**, 413–452.
- AVELLANEDA, M. (ed.) (2001). *Quantitative Analysis in Financial Markets, Collected Papers of the New York University Mathematical Finance Seminar, Vol. II* (World Scientific Publishing Co., Inc., River Edge, NJ).
- AVELLANEDA, M., FRIEDMAN, M., HOLMES, H., SAMPERI, D. (1997). Calibrating volatility surfaces via relative entropy minimization. *Appl. Math. Financ.* **4** (1), 37–64.
- AZENCOTT, R. (1984). *Densité des diffusions en temps petit: développements asymptotiques, Seminar on probability XVIII, Lecture Notes in Math. vol. 1059* (Springer, Berlin, Germany) pp. 402–498.
- BALLY, V., TALAY, D. (1995). The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function. *Probab. Theory Rel.* **104**, 43–60.
- BALLY, V., TALAY, D. (1996). The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density. *Monte Carlo Methods Appl.* **2**, 93–128.
- BARRIEU, P., EL KAROUI, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Financ. Stoch.* **9** (2), 269–298.
- BERTHELOT, C., BOSSY, M., TALAY, D. (2004). Numerical analysis and misspecifications in finance: from model risk to localization error estimates for nonlinear PDEs. In: Akahori, J., Ogawa, S., Watanabe, S. (eds.), *Proceedings of 2003 Ritsumeikan Symposium on Stochastic Processes and its Applications to Mathematical Finance* (World Scientific Publishing Co., Singapore), pp. 1–25.
- BLANCHET-SCALLIET, C., DIOP, A., GIBSON, R., TALAY, D., TANRE, E. (2007). Technical analysis compared to mathematical models based methods under parameters mis-specification. *J. Bank. Financ.* **31** (5), 1351–1373.
- BOSSY, M., GIBSON, R., LHABITANT, F.-S., PISTRE, N., TALAY, D. (2006). Model misspecification analysis for bond options and Markovian hedging strategies. *Rev. Derivatives Res.* **9** (2), 109–135.
- CHERIDITO, P., DELBAEN, F., KUPPER, M. (2005). Coherent and convex monetary risk measures for unbounded càdlàg processes. *Financ. Stoch.* **9** (3), 369–387.
- CONT, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Math. Financ.* **16** (3), 519–547.
- COSTANTINI, C., GOBET, E., EL KAROUI, N. (2006) Boundary sensitivities for diffusion processes in time dependent domains. *Appl. Math. Optim.* **54** (2), 159–187.
- CISZAR, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3**, 146–158.
- CVITANIĆ, J., KARATZAS, I. (1999). On dynamic measures of risk. *Financ. Stoch.* **3** (4), 451–482.
- FÖLLMER, H., SCHIED, A. (2002). Convex measures of risk and trading constraints. *Financ. Stoch.* **6** (4), 429–447.
- GAO, Y., LIM, K.G., NG, K.H. (2004). An approximation pricing algorithm in an incomplete market: a differential geometric approach. *Financ. Stoch.* **8** (4), 501–523.
- GOBET, E., MENOZZI, S. (2004). Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme. *Stoch. Proc. Appl.* **112** (2), 201–223.

- GOBET, E., MUNOS, R. (2005). Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control. *SIAM J. Control Optim.* **43** (5), 1676–1713.
- HOUNKPATIN, O. (2002). Volatilité du Taux de Swap et Calibrage d'un Processus de Diffusion, thèse de l'université Paris 6, 2002.
- JACOD, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Stat.* **27** (1), 83–96.
- JACOD, J., LEJAY, A., TALAY, D. (2008). Estimation of the Brownian dimension of a continuous Itô process. *Bernoulli*, **14** (2), 469–498.
- KARATZAS, I., SHREVE, S.E. (1998). *Methods of Mathematical Finance, Applications of Mathematics*, vol. 39 (Springer-Verlag, New York, NY).
- KOHATSU-HIGA, A. (2001). Weak approximations: a Malliavin calculus approach. *Math. Comput.* **70** (233), 135–172.
- KOHATSU-HIGA, A., PETTERSSON, R. (2002). Variance reduction methods for simulation of densities on Wiener space. *SIAM J. Numer. Anal.* **40** (2), 431–450.
- KUTOYANTS, Y. (1984). *Parameter Estimation for Stochastic Processes (translated and edited by B.L.S. Prakasa Rao)*, *Research and Exposition in Mathematics*, vol. 6 (Heldermann Verlag, Berlin, Germany).
- NUALART, D. (2006). *The Malliavin Calculus and Related Topics, Probability and its Applications (New York)*, second ed. (Springer-Verlag, Berlin, Germany).
- PASTUKHOV, S.V. (2004). On some probabilistic-statistical methods in technical analysis. *Teor. Veroyatn. Primen.* **49** (2), 297–316; translation in *Theor. Probab. Appl.* **49** (2), 2005, 245–260.
- PRAKASA RAO, B.L.S. (1999a). *Semimartingales and Their Statistical Inference* (Chapman and Hall, Boca Raton, FL).
- PRAKASA RAO, B.L.S. (1999b). *Statistical Inference for Diffusion Type Processes* (Arnold, London, UK).
- SHIRYAEV, A.N. (2004). A remark on the quickest detection problems. *Stat. Decis.* **22**, 79–82.
- TALAY, D., TUBARO, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** (4), 94–120.
- TALAY, D., ZHENG, Z. (2002). Worst case model risk management. *Financ. Stoch.* **6** (4), 517–537.
- TALAY, D., ZHENG, Z. (2004). Approximation of quantiles of components of diffusion processes. *Stoch. Proc. Appl.* **109**, 23–46.
- YOR, M. (2001). *Exponential Functionals of Brownian Motion and Related Processes*, *Springer Finance* (Springer, Berlin, Germany).

Robust Preferences and Robust Portfolio Choice

Alexander Schied

School of ORIE, Cornell University, 232 Rhodes Hall, Ithaca, NY 14853, USA
E-mail address: schied@orie.cornell.edu

Hans Föllmer

*Institut für Mathematik, Humboldt-Universität, Unter den Linden 6,
10099 Berlin, Germany*
E-mail address: foellmer@math.hu-berlin.de

Stefan Weber

School of ORIE, Cornell University, 279 Rhodes Hall, Ithaca, NY 14853, USA
E-mail address: sweber@orie.cornell.edu

1. Introduction

Financial markets offer a variety of financial positions. The net result of such a position at the end of the trading period is uncertain, and it may thus be viewed as a real-valued function X on the set of possible scenarios. The problem of portfolio choice consists in choosing, among all the available positions, a position that is affordable, given the investor's wealth w , and which is optimal with respect to the investor's preferences.

In its classical form, the problem of portfolio choice involves preferences of von Neumann-Morgenstern type, and a position X is affordable if its price does not exceed the initial capital w . More precisely, preferences are described by a utility functional $E_Q[U(X)]$, where U is a concave utility function and Q is a probability measure on the set of scenarios, which models the investor's expectations. The price of a position X is of the form $E^*[X]$, where P^* is a probability measure equivalent to Q . In this

classical case, the optimal solution can be computed explicitly in terms of U , Q , and P^* . Recent research on the problem of portfolio choice has taken a much wider scope. On the one hand, the increasing role of derivatives and of dynamic hedging strategies has led to a more flexible notion of affordability. On the other hand, there is, nowadays, a much higher awareness of model uncertainty, and this has led to a robust formulation of preferences beyond the von Neumann–Morgenstern paradigm of expected utility.

In Section 2 (Robust Preferences and Monetary Risk Measures) we review the theory of robust preferences as developed by SCHMEIDLER [1989], GILBOA and SCHMEIDLER [1989], and MARINACCI, RUSTICHINI and MARINACCI [2006]. Such preferences admit a numerical representation in terms of utility functionals $\mathcal{U}$ of the form

$$\mathcal{U}(X) = \inf_{Q \in \mathcal{Q}} (E_Q[U(X)] + \gamma(Q)). \quad (1.1)$$

This may be viewed as a robust approach to the problem of model uncertainty. The agent considers a whole class of probabilistic models specified by probability measures Q on the given set of scenarios, but different models Q are taken more or less seriously, and this is made precise in terms of the penalty $\gamma(Q)$. In evaluating a given financial position, the agent then takes a worst-case approach by taking the infimum of expected utilities over the suitably penalized models. There is an obvious analogy between such robust utility functionals and convex risk measures. In fact, we show in Section 2.3 how the representation (1.1) of preferences which are characterized in terms of a robust extension of the von Neumann–Morgenstern axioms can be reduced to the robust representation of convex risk measures. This is the reason why we begin this section with a brief review of the basic properties of convex risk measures.

Suppose that the underlying financial market is modeled by a multidimensional semimartingale, which describes the price fluctuation of a number of liquid assets. Affordability of a position X given the investor's wealth w can then be defined by the existence of some dynamic trading strategy such that the value of the portfolio generated from the initial capital w up to the final time T is at least equal to X . This is equivalent to the constraint

$$\sup_{P^* \in \mathcal{P}} E^*[X] \leq w,$$

where $\mathcal{P}$ denotes the class of equivalent martingale measures. If the preferences of the investor are given by a robust utility functional of the form (1.1), then the problem of optimal portfolio choice involves the two classes of probability measures $\mathcal{P}$ and $\mathcal{Q}$. In many situations, the solution will consist in identifying two measures $\hat{P}^* \in \mathcal{P}$ and $\hat{Q} \in \mathcal{Q}$ such that the solution to the robust problem is given by the solution of the classical problem defined in terms of U , $\hat{P}^*$, and $\hat{Q}$.

In Section 3 (Robust Portfolio Choice), we consider several approaches to the optimal investment problem for an economic agent who uses a robust utility functional (1.1) and who can choose between risky and riskless investment opportunities in a financial market. In Section 3.1, we formulate the corresponding optimal investment problem in a general setup and introduce standing assumptions for the subsequent sections. In Section 3.2, we show how methods from robust statistics can be used to obtain explicit

solutions in a complete market model when the robust utility functional is coherent, that is, the penalty function γ in (1.1) takes only the values 0 and ∞ . The relations of this approach to capacity theory are analyzed in Section 3.3, together with several concrete examples. In Section 3.4 we develop the general duality theory for robust utility maximization. These duality techniques are then applied in Section 3.5, where optimal investment strategies for incomplete stochastic factor models are characterized in terms of the unique classical solutions of quasilinear partial differential equations. Instead of this analytical approach, one can also use backward stochastic differential equations to characterize optimal strategies, and this technique is briefly discussed in Section 3.6.

In Section 4 (Portfolio Choice under Robust Constraints), we discuss the problem of portfolio optimization under risk constraints. These constraints have a robust representation if they are formulated in terms of convex risk measures. Research on optimization problems under risk constraints provides a further perspective on risk measures that are used to regulate financial institutions. The axiomatic theory of risk measures does not take into account their impact on the behavior of financial agents who are subject to regulation and thus does not capture the effect of capital requirements on portfolios, market prices, and volatility. In order to deal with such issues, we discuss static and semi-dynamic risk constraints in an equilibrium setting. In Section 4.1, we analyze the corresponding partial equilibrium problem. The general equilibrium is discussed in Section 4.2. The literature on portfolio choice under risk constraints is currently far from complete, and we point to some directions for future research.

2. Robust preferences and monetary risk measures

The goal of this section is to characterize investor preferences that are robust in the sense that they account for uncertainty in the underlying models. The main results are presented in Section 2.3. There it is shown in particular that robust preferences can numerically be represented in terms of robust utility functionals, which involve concave monetary utility functionals. Therefore, we first provide two preliminary sections on concave monetary utility functionals and convex risk measures. In Section 2.1 we discuss the dual representation theory in terms of the penalty function of a concave monetary utility functional. In Section 2.2, we present some standard examples of concave monetary utility functionals.

2.1. Risk measures and monetary utility functionals

In this section, we briefly recall the basic definitions and properties of convex risk measures and monetary utility functionals. We refer to chapter 4 of FÖLLMER and SCHIED [2004] for a more comprehensive account.

One of the basic tasks in finance is to quantify the risk associated with a given financial position, which is subject to uncertainty. Let Ω be a fixed set of scenarios. The profits and losses (P&L) of such a financial position are described by a mapping $X : \Omega \rightarrow \mathbb{R}$, where $X(\omega)$ is the discounted net worth of the position at the end of the trading period if the scenario $\omega \in \Omega$ is realized. The goal is to determine a real number $\rho(X)$ that quantifies the risk and can serve as a capital requirement, that is, as the minimal amount of capital

which, if added to the position and invested in a risk-free manner, makes the position acceptable. The following axiomatic approach to such risk measures was initiated in the coherent case by ARTZNER et al. [1999] and later independently extended to the class of convex risk measures by HEATH [2000], FÖLLMER and SCHIED [2002a], and FRITTELLI and ROSAZZA GIANIN [2002].

DEFINITION 2.1. Let $\mathcal{X}$ be a linear space of bounded functions containing the constants. A mapping $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is called a convex risk measure if it satisfies the following conditions for all $X, Y \in \mathcal{X}$:

- *Monotonicity*: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.
- *Cash invariance*: If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.
- *Convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)$, for $0 \leq \lambda \leq 1$.

The convex risk measure ρ is called a coherent risk measure if it satisfies the condition of

- *Positive homogeneity*: If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda \rho(X)$.

The financial meaning of monotonicity is clear. Cash invariance is also called translation invariance. It is the basis for the interpretation of $\rho(X)$ as a capital requirement: if the amount m is added to the position and invested in a risk-free manner, the capital requirement is reduced by the same amount. In particular, cash invariance implies $\rho(X + \rho(X)) = 0$, that is, the accumulate position consisting of X and the risk-free investment $\rho(X)$ is acceptable. While the axiom of cash invariance is best understood in its relation to the interpretation of $\rho(X)$ as a capital requirement for X , it is often convenient to reverse signs and to put emphasis on the *utility* of a position rather than on its risk. This leads to the following concept.

DEFINITION 2.2. A mapping $\phi : \mathcal{X} \rightarrow \mathbb{R}$ is called a concave monetary utility functional if $\rho(X) := -\phi(X)$ is a convex risk measure. If ρ is coherent, then ϕ is called a coherent monetary utility functional.

We now assume that P&Ls are described by random variables X on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. More precisely, we consider the case in which $\mathcal{X} = L^\infty$, where for $0 \leq p \leq \infty$, we denote by L^p the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$. This choice implicitly assumes that concave monetary utility functionals respect $\mathbb{P}$ -nullsets in the sense that

$$\phi(X) = \phi(Y) \quad \text{whenever } X = Y \text{ } \mathbb{P}\text{-a.s.} \quad (2.1)$$

DEFINITION 2.3. The *minimal penalty function* of the concave monetary utility functional ϕ is given for probability measures $Q \ll \mathbb{P}$ by

$$\gamma(Q) := \sup_{X \in L^\infty} (\phi(X) - E_Q[X]). \quad (2.2)$$

The following theorem was obtained by DELBAEN [2002] in the coherent case and later extended by FÖLLMER and SCHIED [2002a] to the general concave case. It provides

the basic representation for concave monetary utility functionals in terms of probability measures under the condition that certain continuity properties are satisfied. Without these continuity properties, one only gets a representation in terms of finitely additive probability measures (see FÖLLMER and SCHIED [2004, section 4.2]).

THEOREM 2.1. *For a concave monetary utility functional ϕ with minimal penalty function γ , the following conditions are equivalent.*

(i) *For $X \in L^\infty$*

$$\phi(X) = \inf_{Q \ll \mathbb{P}} (E_Q[X] + \gamma(Q)). \quad (2.3)$$

(ii) *ϕ is continuous from above: if $X_n \searrow X$ $\mathbb{P}$ -a.s., then $\phi(X_n) \searrow \phi(X)$.*

(iii) *ϕ has the Fatou property: for any bounded sequence $(X_n) \subset L^\infty$ that converges in probability to some X , we have $\phi(X) \geq \limsup_n \phi(X_n)$.*

Moreover, under these conditions, ϕ is coherent if and only if γ takes only the values 0 and ∞ . In this case, (2.3) becomes

$$\phi(X) = \inf_{Q \in \mathcal{Q}} E_Q[X], \quad X \in L^\infty, \quad (2.4)$$

where $\mathcal{Q} = \{Q \ll \mathbb{P} \mid \gamma(Q) = 0\}$, called maximal representing set of ϕ , is the maximal set of probability measures for which the representation (2.4) holds.

PROOF. See, for instance, FÖLLMER and SCHIED [2004, theorem 4.31 and corollary 4.34]. $\square$

The theorem shows that every concave monetary utility functional that is continuous from above arises in the following manner. We consider any probabilistic model $Q \ll \mathbb{P}$, but these models are taken more or less seriously according to the size of the penalty $\gamma(Q)$. Thus, the value $\phi(X)$ is computed as the worst-case expectation taken over all models $Q \ll \mathbb{P}$ and penalized by $\gamma(Q)$.

THEOREM 2.2. *For a concave monetary utility functional ϕ with minimal penalty function γ , the following conditions are equivalent.*

(i) *For any $X \in L^\infty$,*

$$\phi(X) = \min_{Q \ll \mathbb{P}} (E_Q[X] + \gamma(Q)), \quad (2.5)$$

where the minimum is attained in some $Q \ll \mathbb{P}$.

(ii) *ϕ is continuous from below: if $X_n \nearrow X$ $\mathbb{P}$ -a.s., then $\phi(X_n) \nearrow \phi(X)$.*

(iii) *ϕ has the Lebesgue property: for any bounded sequence $(X_n) \subset L^\infty$ that converges in probability to some X , we have $\phi(X) = \lim_n \phi(X_n)$.*

(iv) *For each $c \in \mathbb{R}$, the level set $\{dQ/d\mathbb{P} \mid \gamma(Q) \leq c\}$ is weakly compact in $L^1(\mathbb{P})$.*

PROOF. The equivalence of (b) and (c) follows from FÖLLMER and SCHIED [2004, remark 4.23]. That (b) implies (a) follows from FÖLLMER and SCHIED [2004, proposition 4.21], and that (b) implies (d) follows from FÖLLMER and SCHIED [2004, lemma 4.22] and the Dunford-Pettis theorem. The proof that (a) implies (d) relies on James' theorem as shown by DELBAEN [2002] in the coherent case. It was recently generalized to the general case by JOUINI et al. [2006]. See also JOUINI et al. [2006, theorem 5.2] and KRÄTSCHMER [2005] for alternative proofs of the other implications. $\square$

REMARK 2.1. Note that it follows from the preceding theorems that continuity from below implies continuity from above. It can be shown that the condition of continuity from above is automatically satisfied if the underlying probability space is standard and ϕ is *law invariant* in the sense that $\phi(X) = \phi(Y)$ whenever the $\mathbb{P}$ laws of X and Y coincide (see JOUINI, SCHACHERMAYER and TOUZI [2006]). Several examples for law-invariant concave monetary utility functionals are provided in the next section. Continuity from above also holds as soon as ϕ extends to a concave monetary utility functional on L^p for some $p \in [1, \infty]$ (see CHERIDITO, DELBAEN and KUPPER [2004, proposition 3.8]).

2.2. Examples of monetary utility functionals

In this section, we briefly present some popular choices for concave monetary utility functionals on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. One of the best studied examples is the entropic monetary utility functional,

$$\phi_\theta^{\text{ent}}(X) = -\frac{1}{\theta} \log \mathbb{E}[e^{-\theta X}], \quad (2.6)$$

where θ is a positive constant. One easily checks that it satisfies the conditions of Definition 2.2. Moreover, ϕ_θ^{ent} is clearly continuous from below so that it can be represented as in (2.5) by its minimal penalty function $\gamma_\theta^{\text{ent}}$. Due to standard duality results, this minimal penalty function is given by $\gamma_\theta^{\text{ent}}(Q) = \frac{1}{\theta} H(Q|\mathbb{P})$, where

$$H(Q|\mathbb{P}) = \sup_{X \in L^\infty} (E_Q[X] - \log \mathbb{E}[e^X]) = \mathbb{E} \left[\frac{dQ}{d\mathbb{P}} \log \frac{dQ}{d\mathbb{P}} \right]$$

is the relative entropy of $Q \ll \mathbb{P}$ (see, FÖLLMER and SCHIED [2004, sections 3.2 and 4.9]).

More generally, let $U : \mathbb{R} \rightarrow \mathbb{R}$ be concave, increasing, and nonconstant and take x in the interior of $U(\mathbb{R})$. Then,

$$\phi_U(X) := \sup \{ m \in \mathbb{R} \mid \mathbb{E}[U(X - m)] \geq x \}, \quad X \in L^\infty \quad (2.7)$$

defines a concave monetary utility functional. When considering the corresponding risk measure, $\rho := -\phi_U$, the emphasis is on losses rather than on utility, and so it is natural to consider instead of U the convex increasing loss function $\ell(x) := -U(-x)$. In terms

of ρ , formula (2.7) then becomes

$$\rho(X) := \inf \left\{ m \in \mathbb{R} \mid \mathbb{E}[\ell(-X - m)] \leq -x \right\}, \quad X \in L^\infty. \quad (2.8)$$

The risk measure ρ is called utility-based shortfall risk measure and was introduced by FÖLLMER and SCHIED [2002a]. When choosing $U(x) = -e^{-\theta x}$ (or, equivalently, $\ell(x) = e^{\theta x}$), we obtain the entropic monetary utility functional (2.6) as a special case. It is easy to check that ϕ_U is always continuous from below and hence admits the representation (2.5). Moreover, the minimal penalty function is given by

$$\gamma_U(Q) = \inf_{\lambda > 0} \frac{1}{\lambda} \left(-x + \mathbb{E} \left[\tilde{U} \left(\lambda \frac{dQ}{d\mathbb{P}} \right) \right] \right), \quad Q \ll \mathbb{P},$$

where $\tilde{U}(y) = \sup_x (U(x) - xy)$ denotes the convex conjugate function of U (see FÖLLMER and SCHIED [2002a, theorem 10] or FÖLLMER and SCHIED [2004, theorem 4.106]).

To introduce another closely related class of concave monetary utility functionals, let $g : [0, \infty[\rightarrow \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous convex function satisfying $g(1) < \infty$ and the superlinear growth condition $g(x)/x \rightarrow +\infty$ as $x \uparrow \infty$. Associated to it is the *g-divergence*

$$I_g(Q|\mathbb{P}) := \mathbb{E} \left[g \left(\frac{dQ}{d\mathbb{P}} \right) \right], \quad Q \ll \mathbb{P}, \quad (2.9)$$

as introduced by CSISZAR [1963, 1967]. The *g-divergence* $I_g(Q|\mathbb{P})$ can be interpreted as a statistical distance between the hypothetical model Q and the reference measure $\mathbb{P}$, and so $\gamma_g(Q) := I_g(Q|\mathbb{P})$ is a natural choice for a penalty function. The level sets $\{dQ/d\mathbb{P} \mid I_g(Q|\mathbb{P}) \leq c\}$ are convex and weakly compact in $L^1(\mathbb{P})$ due to the superlinear growth condition. Hence, it follows that $I_g(Q|\mathbb{P})$ is indeed the minimal penalty function of the concave monetary utility functional

$$\phi_g(X) := \inf_{Q \ll \mathbb{P}} (E_Q[X] + I_g(Q|\mathbb{P})). \quad (2.10)$$

Moreover, weak compactness of the level sets guarantees that ϕ_g is continuous from below. One can show that ϕ_g satisfies the variational identity

$$\phi_g(X) = \sup_{z \in \mathbb{R}} (\mathbb{E}[U(X - z)] + z), \quad X \in L^\infty, \quad (2.11)$$

where $U(x) = \inf_{z > 0} (xz + g(z))$ is the concave conjugate function of g . This formula was obtained by BEN-TAL and TEBoulLE [1987] for $\mathbb{R}$ -valued g and extended to the general case by BEN-TAL and TEBoulLE [2007] and SCHIED [2007a] (see also CHERNY and KUPPER [2007] for further properties). The resulting concave monetary utility functionals were called optimized certainty equivalents by BEN-TAL and TEBoulLE [2007]. Note that the particular choice $g(x) = x \log x$ corresponds to the relative entropy $I_g(Q|\mathbb{P}) = H(Q|\mathbb{P})$, and so ϕ_g coincides with the entropic monetary utility functional. Another important example is provided by taking $g(x) = 0$ for $x \leq \lambda^{-1}$ and

$g(x) = \infty$ otherwise so that the corresponding coherent monetary utility functional is given by

$$\phi_\lambda(X) := \inf_{Q \in \mathcal{Q}_\lambda} E_Q[X] \quad \text{for} \quad \mathcal{Q}_\lambda := \left\{ Q \ll \mathbb{P} \mid \frac{dQ}{d\mathbb{P}} \leq \frac{1}{\lambda} \right\}. \quad (2.12)$$

This shows that $-\phi_\lambda(X)$ is equal to the coherent risk measure *average value at risk*, $AVaR_\lambda(X) = \frac{1}{\lambda} \int_0^\lambda VaR_\gamma(X) d\gamma$, which is also called *expected shortfall*, *conditional value at risk*, or *tail value at risk*, see, e.g., FÖLLMER and SCHIED [2004]. In this case, we have $U(x) = 0 \wedge x/\lambda$ and hence get the classical duality formula

$$AVaR_\lambda(X) = \frac{1}{\lambda} \inf_{z \in \mathbb{R}} (E[(z - X)^+] - \lambda z) \quad (2.13)$$

as a special case of (2.11).

All examples discussed so far in this section are law invariant in the sense that $\phi(X) = \phi(Y)$ whenever the $\mathbb{P}$ -laws of X and Y coincide. One can show that every law-invariant concave monetary utility functional ϕ on L^∞ can be represented in the following form:

$$\phi(X) = \inf_{\mu} \left(\int_{(0,1]} \phi_\lambda(X) \mu(d\lambda) + \beta(\mu) \right), \quad (2.14)$$

where the supremum is taken over all Borel probability measures μ on $[0, 1]$, ϕ_λ is as in (2.12), and $\beta(\mu)$ is a penalty for μ . Under the additional assumption of continuity from above, this representation was obtained in the coherent case by KUSUOKA [2001] and later extended by KUNZE [2003], DANA [2005], FRITTELLI and ROSAZZA-GIANIN [2005], and FÖLLMER and SCHIED [2004, section 4.5]. More recently, JOUINI, SCHACHERMAYER and TOUZI [2006] showed that the condition of continuity from above can actually be dropped. More examples of concave and coherent monetary utility functionals will be provided at the beginning of Section 3.5.

2.3. Robust preferences and their numerical representation

In this section, we describe how robust utility functionals appear naturally as numerical representations of investor preferences in the face of model uncertainty as developed by SCHMEIDLER [1989], GILBOA and SCHMEIDLER [1989], and MACCHERONI, RUSTICHINI and MARINACCI [2006].

The general aim of the theory of choice is to provide an axiomatic foundation and corresponding representation theory for a normative decision rule by means of which an economic agent can reach decisions when presented with several alternatives. A fundamental example is the von Neumann–Morgenstern theory, in which the agent can choose between several monetary bets with known success probabilities. Such a monetary bet can be regarded as a Borel probability measure on $\mathbb{R}$ and is often called a lottery. More specifically, we will consider here the space $\mathcal{M}_{1,c}(S)$ of Borel probability measures with compact support in some given nonempty interval $S \subset \mathbb{R}$. The decision rule is usually taken as a preference relation or preference order $>$ on $\mathcal{M}_{1,c}(S)$, that is, $>$ is a binary

relation on $\mathcal{M}_{1,c}(S)$ that is asymmetric,

$$\mu \succ \nu \Rightarrow \nu \not\succ \mu,$$

and negatively transitive,

$$\mu \succ \nu \text{ and } \lambda \in \mathcal{M}_{1,c}(S) \Rightarrow \mu \succ \lambda \text{ or } \lambda \succ \nu.$$

The corresponding weak preference order, $\mu \succeq \nu$, is defined as the negation of $\nu \succ \mu$. If both $\mu \succeq \nu$ and $\nu \succeq \mu$ hold, we will write $\mu \overset{\circ}{\sim} \nu$. Dealing with a preference order is greatly facilitated if one has a numerical representation, that is, a function $\mathcal{U} : \mathcal{M}_{1,c}(S) \rightarrow \mathbb{R}$ such that

$$\mu \succ \nu \iff \mathcal{U}(\mu) > \mathcal{U}(\nu).$$

VON NEUMANN and MORGENSTERN [1944] formulated a set of axioms that are necessary and sufficient for the existence of a numerical representation $\mathcal{U}$ of von Neumann–Morgenstern form, that is,

$$\mathcal{U}(\mu) = \int U(x) \mu(dx) \tag{2.15}$$

for a function $U : \mathbb{R} \rightarrow \mathbb{R}$. The two main axioms are

- *Archimedean axiom*: for any triple $\mu \succ \lambda \succ \nu$ in $\mathcal{M}_{1,c}(S)$, there are $\alpha, \beta \in [0, 1]$ such that $\alpha\mu + (1 - \alpha)\nu \succ \lambda \succ \beta\mu + (1 - \beta)\nu$.
- *Independence axiom*: for all $\mu, \nu \in \mathcal{M}_{1,c}(S)$, the relation $\mu \succ \nu$ implies $\alpha\mu + (1 - \alpha)\lambda \succ \alpha\nu + (1 - \alpha)\lambda$ for all $\lambda \in \mathcal{M}_{1,c}(S)$ and all $\alpha \in [0, 1]$.

These two axioms are equivalent to the existence of an affine numerical representation $\mathcal{U}$. To obtain an integral representation (2.15) for this affine functional on $\mathcal{M}_{1,c}(S)$, one needs some additional regularity condition such as monotonicity with respect to first-order stochastic dominance or topological assumptions on the level sets of $\succ$ (see KREPS [1988] and FÖLLMER and SCHIED [2004, Chapter 2]; see also HERSTEIN and MILNOR [1953] for a relaxation of these axioms in a generalized setting).

The monetary character of lotteries suggests the further requirement that $\delta_x \succ \delta_y$ for $x > y$, which is equivalent to the fact that U is strictly increasing. In addition, the preference order is called risk averse if for every nontrivial lottery $\mu \in \mathcal{M}_{1,c}(S)$, the certain amount $m(\mu) := \int x \mu(dx)$ is strictly preferred over the lottery μ itself, that is, $\delta_{m(\mu)} \succ \mu$. Clearly, risk aversion is equivalent to the fact that U is strictly concave. If U is both increasing and strictly concave, it is called a utility function.

In the presence of *model* uncertainty or ambiguity, sometimes also called Knightian uncertainty, the economic agent only has imperfect knowledge of the success probabilities of a financial bet. Mathematically, such a situation is often modeled by making lotteries contingent on some external source of randomness. Thus, let $(\Omega, \mathcal{F})$ be a given measurable space and consider the class $\tilde{\mathcal{X}}$ defined as the set of all Markov kernels $\tilde{X}(\omega, dy)$ from $(\Omega, \mathcal{F})$ to S for which there exists a compact set $K \subset S$ such

that $\tilde{X}(\omega, K) = 1$ for all $\omega \in \Omega$. In mathematical economics, the elements of $\tilde{\mathcal{X}}$ are sometimes called acts or horse race lotteries.

Now consider a given preference order $\succ$ on $\tilde{\mathcal{X}}$. The space of standard lotteries, $\mathcal{M}_{1,c}(S)$, has a natural embedding into $\tilde{\mathcal{X}}$ via the identification of $\mu \in \mathcal{M}_{1,c}(S)$ with the constant Markov kernel $\tilde{X}(\omega) = \mu$, and this embedding induces a preference order on $\mathcal{M}_{1,c}(S)$, which we also denote by $\succ$. We assume that the preference order on $\tilde{\mathcal{X}}$ is *monotone* with respect to the embedding of $\mathcal{M}_{1,c}(S)$ into $\tilde{\mathcal{X}}$:

$$\tilde{Y} \succeq \tilde{X} \quad \text{if } \tilde{Y}(\omega) \succeq \tilde{X}(\omega) \quad \text{for all } \omega \in \Omega. \quad (2.16)$$

We will furthermore assume the following three axioms, of which the first two are suitable extensions of the two main axioms of classical von Neumann–Morgenstern theory to our present setting.

- *Archimedean axiom*: if $\tilde{X}, \tilde{Y}, \tilde{Z} \in \tilde{\mathcal{X}}$ are such that $\tilde{Z} \succ \tilde{Y} \succ \tilde{X}$, then there are $\alpha, \beta \in [0, 1]$ with

$$\alpha \tilde{Z} + (1 - \alpha) \tilde{X} \succ \tilde{Y} \succ \beta \tilde{Z} + (1 - \beta) \tilde{X}.$$

- *Weak certainty independence*: if for $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$ and for some $\nu \in \mathcal{M}_{1,c}(S)$ and $\alpha \in [0, 1]$ we have $\alpha \tilde{X} + (1 - \alpha) \nu \succ \alpha \tilde{Y} + (1 - \alpha) \nu$, then

$$\alpha \tilde{X} + (1 - \alpha) \mu \succ \alpha \tilde{Y} + (1 - \alpha) \mu \quad \text{for all } \mu \in \mathcal{M}_{1,c}(S).$$

- *Uncertainty aversion*: if $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$ are such that $\tilde{X} \sim \tilde{Y}$, then

$$\alpha \tilde{X} + (1 - \alpha) \tilde{Y} \succeq \tilde{X} \quad \text{for all } \alpha \in [0, 1].$$

These axioms were formulated by GILBOA and SCHMEIDLER [1989], with the exception that instead of weak certainty independence, they originally used the stronger concept of full certainty independence, which we will explain below. The relaxation of full certainty independence to weak certainty independence was suggested by MACCHERONI, RUSTICHINI and MARINACCI [2006].

REMARK 2.2. In order to motivate the term *uncertainty aversion*, consider the following simple example. For $\Omega := \{0, 1\}$, define

$$\tilde{Z}_i(\omega) := \delta_{1000} \cdot \mathbb{I}_{\{i\}}(\omega) + \delta_0 \cdot \mathbb{I}_{\{1-i\}}(\omega), \quad i = 0, 1.$$

Suppose that an agent is indifferent between the choices $\tilde{Z}_0$ and $\tilde{Z}_1$, both of which involve the same kind of uncertainty. In the case of uncertainty aversion, the convex combination $\tilde{Y} := \alpha \tilde{Z}_0 + (1 - \alpha) \tilde{Z}_1$ is weakly preferred over both $\tilde{Z}_0$ and $\tilde{Z}_1$. It takes the form

$$\tilde{Y}(\omega) = \begin{cases} \alpha \delta_{1000} + (1 - \alpha) \delta_0 & \text{for } \omega = 1, \\ \alpha \delta_0 + (1 - \alpha) \delta_{1000} & \text{for } \omega = 0. \end{cases}$$

This convex combination now allows for upper and lower probability bounds in terms of α , and this means that model uncertainty is reduced in favor of risk. For $\alpha = 1/2$, the resulting lottery $\tilde{Y}(\omega) \equiv \frac{1}{2}(\delta_{1000} + \delta_0)$ is independent of the scenario ω , that is, model uncertainty is completely eliminated.

REMARK 2.3. The Archimedean axiom and weak certainty independence imply that the restriction of $\succ$ to $\mathcal{M}_{1,c}(S)$ satisfies the independence axiom of von Neumann–Morgenstern theory and hence admits an affine numerical representation $\mathcal{U} : \mathcal{M}_{1,c}(S) \rightarrow \mathbb{R}$.

PROOF. We need to derive the independence axiom on $\mathcal{M}_{1,c}(S)$. To this end, take $v \in \mathcal{M}_{1,c}(S)$ such that $\mu \succ v$. We claim that $\mu \geq \frac{1}{2}\mu + \frac{1}{2}v \geq v$. Indeed, otherwise we would, for instance, have that $\frac{1}{2}\mu + \frac{1}{2}v \succ \mu = \frac{1}{2}\mu + \frac{1}{2}\mu$. Weak certainty independence now yields $\frac{1}{2}v + \frac{1}{2}v \succ \frac{1}{2}v + \frac{1}{2}\mu$ and in turn $v \succ \mu$, a contradiction.

Iterating the preceding argument now yields $\mu \geq \alpha\mu + (1 - \alpha)v \geq v$ for every dyadic rational number $\alpha \in [0, 1]$. Applying the Archimedean axiom completes the proof. $\square$

In addition to the axioms listed above, we assume henceforth that the affine numerical representation $\mathcal{U} : \mathcal{M}_{1,c}(S) \rightarrow \mathbb{R}$ of Remark 2.3 is actually of von Neumann–Morgenstern form (2.15) for some function $U : S \rightarrow \mathbb{R}$. For simplicity, we also assume that U is a utility function with unbounded range $U(S)$ containing zero. The following theorem is an extension of the main result of GILBOA and SCHMEIDLER [1989]. In this form, it is due to MACCHERONI, RUSTICHINI and MARINACCI [2006].

THEOREM 2.3. *Under the above conditions, there exists a unique extension of $\mathcal{U}$ to a numerical representation $\tilde{\mathcal{U}} : \tilde{\mathcal{X}} \rightarrow \mathbb{R}$, and $\tilde{\mathcal{U}}$ is of the form*

$$\tilde{\mathcal{U}}(\tilde{X}) = \phi(\mathcal{U}(\tilde{X})) = \phi\left(\int U(x) \tilde{X}(\cdot, dx)\right)$$

for a concave monetary utility functional ϕ defined on the space of bounded measurable functions on $(\Omega, \mathcal{F})$.

PROOF. The proof is a variant of the original proofs by GILBOA and SCHMEIDLER [1989] and MACCHERONI, RUSTICHINI and MARINACCI [2006].

Step 1: We prove that there exists a unique extension of $\mathcal{U}$ to a numerical representation $\tilde{\mathcal{U}} : \tilde{\mathcal{X}} \rightarrow \mathbb{R}$. By definition of $\tilde{\mathcal{X}}$, for every $\tilde{X} \in \tilde{\mathcal{X}}$, there exists some real number a such that $[-a, a] \subset S$ and $\tilde{X}(\omega, [-a, a]) = 1$ for all ω . Monotonicity (2.16) and the fact that U is strictly increasing, thus, imply $\delta_a \geq \tilde{X} \geq \delta_{-a}$. Standard arguments hence yield the existence of a unique $\alpha \in [0, 1]$ such that $\tilde{X} \stackrel{\circ}{\sim} \alpha\delta_a + (1 - \alpha)\delta_{-a}$ (see FÖLLMER and SCHIED [2004, lemma 2.83]). It follows that

$$\tilde{\mathcal{U}}(\tilde{X}) := \tilde{\mathcal{U}}(\alpha\delta_a + (1 - \alpha)\delta_{-a}) = \alpha U(a) + (1 - \alpha)U(-a)$$

must be the desired numerical representation.

Step 2: For $\mu \in \mathcal{M}_{1,c}(S)$, let $c(\mu) := U^{-1}(\mathcal{U}(\mu))$ denote the corresponding certainty equivalent. For $\tilde{X} \in \tilde{\mathcal{X}}$, monotonicity (2.16) then implies that $\tilde{X} \overset{\circ}{\sim} \delta_{c(\tilde{X})}$, where $c(\tilde{X})$ is the bounded S -valued measurable function defined as the ω -wise certainty equivalent of $\tilde{X}$. It is, therefore, enough to show that there exists a concave monetary utility functional ϕ such that

$$\tilde{\mathcal{U}}(\delta_X) = \phi(U(X)) \quad (2.17)$$

for every S -valued measurable function X with compact range. Indeed, for $\tilde{X} \in \tilde{\mathcal{X}}$, we then have

$$\tilde{\mathcal{U}}(\tilde{X}) = \tilde{\mathcal{U}}(\delta_{c(\tilde{X})}) = \phi(U(c(\tilde{X}))) = \phi(\mathcal{U}(\tilde{X})). \quad (2.18)$$

Step 3: We now show that there exists a concave monetary utility functional ϕ such that (2.17) holds for every S -valued measurable function X with compact range. We first note that (2.17) uniquely defines a functional ϕ on the set $\mathcal{X}_U$ of bounded $U(S)$ -valued measurable functions. Moreover, ϕ is monotone due to our monotonicity assumptions. We now prove that ϕ satisfies the translation property on $\mathcal{X}_U$. To this end, we first assume that $U(S) = \mathbb{R}$ and take a bounded measurable function X and some $z \in \mathbb{R}$. We then let $X_0 := U^{-1}(2X)$, $z_0 := U^{-1}(2z)$, and $y := U^{-1}(0)$. Taking a such that $a \geq X_0(\omega) \geq -a$ for each ω , we see as in Step 1 that there exists $\beta \in [0, 1]$ such that

$$\frac{1}{2}\delta_{X_0} + \frac{1}{2}\delta_y \overset{\circ}{\sim} \frac{\beta}{2}(\delta_a + \delta_y) + \frac{1-\beta}{2}(\delta_{-a} + \delta_y) = \frac{1}{2}\mu + \frac{1}{2}\delta_y,$$

where $\mu = \beta\delta_a + (1-\beta)\delta_{-a}$. Using weak certainty independence, we may replace δ_y with δ_{z_0} and obtain $\frac{1}{2}(\delta_{X_0} + \delta_{z_0}) \overset{\circ}{\sim} \frac{1}{2}(\mu + \delta_{z_0})$. Hence, by using (2.18)

$$\begin{aligned} \phi(X+z) &= \phi\left(\frac{1}{2}U(X_0) + \frac{1}{2}U(z_0)\right) = \phi\left(\mathcal{U}\left(\frac{1}{2}\delta_{X_0} + \frac{1}{2}\delta_{z_0}\right)\right) \\ &= \phi\left(\mathcal{U}\left(\frac{1}{2}\mu + \frac{1}{2}\delta_{z_0}\right)\right) = \tilde{\mathcal{U}}\left(\frac{1}{2}\mu + \frac{1}{2}\delta_{z_0}\right) = \mathcal{U}\left(\frac{1}{2}\mu + \frac{1}{2}\delta_{z_0}\right) \\ &= \frac{1}{2}\mathcal{U}(\mu) + \frac{1}{2}\mathcal{U}(z_0). \end{aligned}$$

The translation property now follows from $U(z_0) = 2z$ and the fact that

$$\begin{aligned} \frac{1}{2}\mathcal{U}(\mu) &= \frac{1}{2}(\mathcal{U}(\mu) + \mathcal{U}(\delta_y)) = \mathcal{U}\left(\frac{1}{2}\mu + \frac{1}{2}\delta_y\right) = \tilde{\mathcal{U}}\left(\frac{1}{2}\mu + \frac{1}{2}\delta_y\right) \\ &= \tilde{\mathcal{U}}\left(\frac{1}{2}\delta_{X_0} + \frac{1}{2}\delta_y\right) = \phi\left(\frac{1}{2}U(X_0) + \frac{1}{2}U(y)\right) = \phi(X). \end{aligned}$$

Here, we have again applied (2.18). If $U(S)$ is not equal to $\mathbb{R}$, it is sufficient to consider the cases in which $U(S)$ contains $[0, \infty]$ or $[-\infty, 0]$ and to work with positive or negative quantities X and z , respectively. Then, the preceding argument establishes the translation

property of ϕ on the spaces of positive or negative bounded measurable functions, and ϕ can be extended by translation to the entire space of bounded measurable functions.

We now prove the concavity of ϕ by showing $\phi(\frac{1}{2}(X + Y)) \geq \frac{1}{2}\phi(X) + \frac{1}{2}\phi(Y)$. This is enough since ϕ is Lipschitz continuous due to monotonicity and the translation property. Let $X_0 := U^{-1}(X)$ and $Y_0 := U^{-1}(Y)$ and suppose first that $\phi(X) = \phi(Y)$. Then, $\delta_{X_0} \stackrel{\sim}{\sim} \delta_{Y_0}$ and uncertainty aversion implies that $\tilde{Z} := \frac{1}{2}\delta_{X_0} + \frac{1}{2}\delta_{Y_0} \geq \delta_{X_0}$. Hence, by using (2.18), we get

$$\phi\left(\frac{1}{2}(X + Y)\right) = \tilde{U}(\tilde{Z}) \geq \tilde{U}(\delta_{X_0}) = \phi(X) = \frac{1}{2}\phi(X) + \frac{1}{2}\phi(Y).$$

If $\phi(X) \neq \phi(Y)$, then we let $z := \phi(X) - \phi(Y)$ so that $Y_z := Y + z$ satisfies $\phi(Y_z) = \phi(X)$. Hence,

$$\begin{aligned} \phi\left(\frac{1}{2}(X + Y)\right) + \frac{1}{2}z &= \phi\left(\frac{1}{2}(X + Y)_z\right) \geq \frac{1}{2}\phi(X) + \frac{1}{2}\phi(Y_z) \\ &= \frac{1}{2}\phi(X) + \frac{1}{2}\phi(Y) + \frac{1}{2}z. \end{aligned}$$

□

Instead of weak certainty independence, GILBOA and SCHMEIDLER [1989] consider the stronger axiom of

- *full certainty independence*: for all $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$, $\mu \in \mathcal{M}_{1,c}(S)$, and $\alpha \in [0, 1]$, we have

$$\tilde{X} \succ \tilde{Y} \implies \alpha\tilde{X} + (1 - \alpha)\mu \succ \alpha\tilde{Y} + (1 - \alpha)\mu.$$

As we have seen in Remark 2.3 and its proof, the axioms of full and weak certainty independence extend the independence axiom for preferences on $\mathcal{M}_{1,c}(S)$ to our present setting, but only under the restriction that the replacing act is certain, that is, it is given by a lottery μ that does not depend on the scenario $\omega \in \Omega$. There are good reasons for not requiring full independence for all $\tilde{Z} \in \tilde{\mathcal{X}}$. As an example, take $\Omega = \{0, 1\}$ and define $\tilde{X}(\omega) = \delta_\omega$, $\tilde{Y}(\omega) = \delta_{1-\omega}$, and $\tilde{Z} = \tilde{X}$. An agent may prefer $\tilde{X}$ over $\tilde{Y}$, thus expressing the implicit view that Scenario 1 is somewhat more likely than Scenario 0. At the same time, the agent may like the idea of hedging against the occurrence of Scenario 0, and this could mean that the certain lottery

$$\frac{1}{2}(\tilde{Y} + \tilde{Z}) \equiv \frac{1}{2}(\delta_0 + \delta_1)$$

is preferred over the contingent lottery

$$\frac{1}{2}(\tilde{X} + \tilde{Z}) \equiv \tilde{X},$$

thus violating the independence assumption in its unrestricted form. In general, the role of $\tilde{Z}$ as a hedge against scenarios unfavorable for $\tilde{Y}$ requires that $\tilde{Y}$ and $\tilde{Z}$ are not comonotone, where comonotonicity means

$$\tilde{Y}(\omega) \geq \tilde{Y}(\tilde{\omega}) \iff \tilde{Z}(\omega) \geq \tilde{Z}(\tilde{\omega}).$$

Thus, the wish to hedge would still be compatible with the following stronger version of certainty independence, called

- *comonotonic independence*: For $\tilde{X}, \tilde{Y}, \tilde{Z} \in \tilde{\mathcal{X}}$ and $\alpha \in [0, 1]$,

$$\tilde{X} \succ \tilde{Y} \iff \alpha \tilde{X} + (1 - \alpha) \tilde{Z} \succ \alpha \tilde{Y} + (1 - \alpha) \tilde{Z}$$

whenever $\tilde{Y}$ and $\tilde{Z}$ are comonotone.

THEOREM 2.4. *In the setting of Theorem 2.3, full certainty independence holds if and only if ϕ is coherent. Moreover, comonotonic independence is equivalent to the fact that ϕ is comonotonic, that is, $\phi(X + Y) = \phi(X) + \phi(Y)$, whenever X and Y are comonotone.*

PROOF. See GILBOA and SCHMEIDLER [1989] AND SCHMEIDLER [1989] or FÖLLMER and SCHIED [2004, sections 2.5 and 4.7]. $\square$

The representation theorem for concave monetary utility functionals, Theorem 2.1, suggests that ϕ from Theorem 2.3 typically admits a representation of the form

$$\phi(X) = \inf_{Q \in \mathcal{Q}} (E_Q[X] + \gamma(Q))$$

for some set $\mathcal{Q}$ of probability measures on $(\Omega, \mathcal{F})$ and some penalty function $\gamma : \mathcal{Q} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then, the restriction of $\succ$ to bounded measurable functions X on $(\Omega, \mathcal{F})$, regarded as elements of $\tilde{\mathcal{X}}$ via the identification with δ_X , admits a numerical representation of the form

$$X \mapsto \inf_{Q \in \mathcal{Q}} (E_Q[U(X)] + \gamma(Q)). \quad (2.19)$$

It is this representation in which we are really interested. Note, however, that it is necessary to formulate the axiom of uncertainty aversion on the larger space of uncertain lotteries. But even without its axiomatic foundation, the representation of preferences in the face of model uncertainty by a subjective utility assessment (2.19) is highly plausible as it stands. It may, in fact, be viewed as a robust approach to the problem of model uncertainty: The agent penalizes every possible probabilistic view $Q \in \mathcal{Q}$ in terms of the penalty $\gamma(Q)$ and takes a worst-case approach in evaluating the payoff of a given financial position. The resulting preference structures for entropic penalties $\gamma_\theta^{\text{ent}}(Q) = \frac{1}{\theta} H(Q|\mathbb{P})$ are sometimes called multiplier preferences in economics (see HANSEN and SARGENT [2001] and MACCHERONI, RUSTICHINI and MARINACCI [2006]).

3. Robust portfolio choice

In this section, we consider the optimal investment problem for an economic agent who is averse against both risk and ambiguity and who can choose between risky and riskless investment opportunities in a financial market. Payoffs generated by investment choices are modeled as random variables X defined on the probability space of some underlying

market model. By the theory developed in the preceding chapter, it is natural to assume that the utility derived from such a payoff X is given by

$$\inf_{Q \in \mathcal{Q}} (E_Q[U(X)] + \gamma(Q)) \quad (3.1)$$

for a utility function U , a penalty function γ , and an appropriate set $\mathcal{Q}$ of probability measures. The goal of the investor is, thus, to maximize this expression over the class of achievable payoffs. If the penalty function vanishes on $\mathcal{Q}$, the expression (3.1) reduces to

$$\inf_{Q \in \mathcal{Q}} E_Q[U(X)], \quad (3.2)$$

which often greatly simplifies the complexity of the required mathematics. We will, therefore, often resort to this reduced setting and refer to it as the coherent case.

In the next section, we formulate the corresponding optimal investment problem in a rather general setup, which we will restrict later according to the particular requirements of each method. In the subsequent section, we show how methods from robust statistics can be used to obtain explicit solutions for a class of coherent examples in a complete market model. The relations of this approach to capacity theory are analyzed in Section 3.3 along with several concrete examples. In Section 3.4, we develop the general duality theory for robust utility maximization. These duality techniques are then applied in Section 3.5, where optimal investment strategies for incomplete stochastic factor models are characterized in terms of the unique classical solutions of quasilinear PDE. Instead of PDE, one can also use backward stochastic differential equations to characterize optimal strategies, and this approach is discussed in Section 3.6.

3.1. Problem formulation and standing assumptions

We start by describing the underlying financial market model. The discounted price process of d assets is modeled by a stochastic process $S = (S_t)_{0 \leq t \leq T}$, which is assumed to be a d -dimensional semimartingale on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ satisfying the usual conditions. We assume furthermore that $\mathcal{F}_0$ is $\mathbb{P}$ -trivial. A self-financing trading strategy can be regarded as a pair (x, ξ) , where $x \in \mathbb{R}$ is the initial investment and $\xi = (\xi_t)_{0 \leq t \leq T}$ is a d -dimensional predictable and S -integrable process. The value process X associated with (x, ξ) is given by $X_0 = x$ and

$$X_t = X_0 + \int_0^t \xi_r dS_r, \quad 0 \leq t \leq T.$$

For $x > 0$ given, we denote by $\mathcal{X}(x)$ the set of all value processes X that satisfy $X_0 \leq x$ and are admissible in the sense that $X_t \geq 0$ for $0 \leq t \leq T$. We assume that our model is arbitrage free in the sense that $\mathcal{P} \neq \emptyset$, where $\mathcal{P}$ denotes the set of measures equivalent to $\mathbb{P}$ under which each $X \in \mathcal{X}(1)$ is a local martingale. If S is locally bounded, then a measure $Q \sim \mathbb{P}$ belongs to $\mathcal{P}$ if and only if S is a local Q -supermartingale (see DELBAEN and SCHACHERMAYER [2006]).

We now describe the robust utility functional of the investor. The utility function is a strictly increasing and strictly concave function $U : [0, \infty] \rightarrow \mathbb{R}$. The utility of a payoff, that is, of a random variable $X \in L^0(\mathbb{P})$, shall be assessed in terms of a robust utility functional of the form

$$X \mapsto \inf_Q (E_Q[U(X)] + \gamma(Q)). \quad (3.3)$$

Here, we assume that γ is bounded from below and equal to the minimal penalty function of the concave monetary utility functional $\phi : L^\infty(\mathbb{P}) \rightarrow \mathbb{R}$ that is defined by

$$\phi(Y) := \inf_{Q \ll \mathbb{P}} (E_Q[Y] + \gamma(Q)), \quad Y \in L^\infty(\mathbb{P}) \quad (3.4)$$

and assumed to satisfy the Fatou property. We may suppose without loss of generality that ϕ is normalized in the sense that $\phi(0) = \inf_Q \gamma(Q) = 0$. We also assume that ϕ is sensitive in the sense that every nonzero $Y \in L_+^\infty$ satisfies $\phi(Y) > 0$. Sensitivity is also called relevance.

Note, however, that the utility functional (3.3) cannot be represented as $\phi(U(X))$ unless the random variable $U(X)$ is bounded because ϕ is a priori only defined on $L^\infty(\mathbb{P})$. Moreover, if the utility function U is not bounded from below, we must be particularly careful even in defining the expression $\inf_Q (E_Q[U(X)] + \gamma(Q))$. First, it is clear that probabilistic models with an infinite penalty $\gamma(Q)$ should not contribute to the value of the robust utility functional. We, therefore, restrict the infimum to models Q in the domain

$$\mathcal{Q} := \{Q \ll \mathbb{P} \mid \gamma(Q) < \infty\}$$

of γ . That is, we make (3.3) more precise by writing

$$X \mapsto \inf_{Q \in \mathcal{Q}} (E_Q[U(X)] + \gamma(Q)).$$

Second, we have to address the problem that the Q -expectation of $U(X)$ might not be well defined in the sense that $E_Q[U^+(X)]$ and $E_Q[U^-(X)]$ are both infinite. This problem will be resolved by extending the expectation operator $E_Q[\cdot]$ to the entire set L^0 :

$$E_Q[F] := \sup_n E_Q[F \wedge n] = \lim_{n \uparrow \infty} E_Q[F \wedge n] \quad \text{for arbitrary } F \in L^0.$$

It is easy to see that in doing so we retain the concavity of the functional $X \mapsto E_Q[U(X)]$ and hence of the robust utility functional.

Thus, our main problem can be stated as follows:

$$\text{Maximize } \inf_{Q \in \mathcal{Q}} (E_Q[U(X_T)] + \gamma(Q)) \text{ over all } X \in \mathcal{X}(x). \quad (3.5)$$

REMARK 3.1. Let us comment on the assumptions made on the robust utility functional. First, the assumption that ϕ is defined on $L^\infty(\mathbb{P})$ is equivalent to either of the facts that ϕ respects $\mathbb{P}$ -nullsets in the sense of (2.1) and that $\gamma(Q)$ is finite only if $Q \ll \mathbb{P}$.

Clearly, our problem (3.3) would not be well defined without this assumption as the value of the stochastic integral used to define X_T is only defined $\mathbb{P}$ -a.s. (see DENIS and MARTINI [2006]). Second, by Theorem 2.1, the Fatou property of ϕ is equivalent to the fact that ϕ admits a representation of the form (3.4). Third, the assumption of sensitivity is economically natural since true payoff possibilities should be rewarded with a nonvanishing utility. In the coherent case, sensitivity and the first assumption together are equivalent to the requirement

$$\mathbb{P}[A] = 0 \iff Q[A] = 0, \quad \text{for all } Q \in \mathcal{Q}. \quad (3.6)$$

The fourth assumption is that γ is equal to the minimal penalty function of ϕ . This is a technical assumption, which we can always make without loss of generality.

EXAMPLE 3.1 (Entropic penalties). As discussed in Section 2.2, a popular choice for γ is taking $\gamma(Q) = \gamma_\theta^{\text{ent}} = \frac{1}{\theta} H(Q|\mathbb{P})$, where $H(Q|\mathbb{P})$ is the relative entropy of Q with respect to $\mathbb{P}$. According to (2.6), this choice corresponds to the utility functional

$$\inf_{Q \in \mathcal{Q}} (E_Q[U(X_T)] + \gamma(Q)) = -\frac{1}{\theta} \log \mathbb{E}[e^{-\theta U(X_T)}]$$

of the terminal wealth, which clearly satisfies the assumptions made in this section. Its maximization is equivalent to the maximization of the ordinary expected utility $\mathbb{E}[\tilde{U}(X_T)]$, where $\tilde{U}(x) = -e^{-\theta U(x)}$ is strictly concave and increasing. New types of problems appear, however, if instead of the terminal wealth of an investment strategy, an intertemporal quantity, such as the intertemporal utility from a consumption-investment strategy, is maximized. The maximization of the corresponding entropic utility functionals is also known as risk-sensitive control. We refer, for instance, to FLEMING and SHEU [2000, 2002], HANSEN and SARGENT [2001], BARRIEU and EL KAROUI [2005], BORDIGONI, MATOUSSI and SCHWEIZER [2005], and the references therein.

3.2. Projection techniques for coherent utility functionals in a complete market

In this section, we assume that the underlying market model is complete in the sense that the set $\mathcal{P}$ consists of the single element P^* . We assume, moreover, that the monetary utility functional ϕ is coherent with maximal defining set $\mathcal{Q}$ so that (3.5) becomes

$$\text{Maximize } \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] \text{ over all } X \in \mathcal{X}(x). \quad (3.7)$$

The following definition has its origins in robust statistical test theory (see HUBER and STRASSEN [1973]).

DEFINITION 3.1. $Q_0 \in \mathcal{Q}$ is called a least favorable *measure* with respect to P^* if the density $\pi = dP^*/dQ_0$ (taken in the sense of the Lebesgue decomposition) satisfies

$$Q_0[\pi \leq t] = \inf_{Q \in \mathcal{Q}} Q[\pi \leq t], \quad \text{for all } t > 0.$$

REMARK 3.2. If a least favorable measure Q_0 exists, then it is automatically equivalent to $\mathbb{P}$. To see this, note first that $\mathcal{Q}$ is closed in total variation by our assumption that γ is the minimal penalty function. Hence, according to (3.6) and the Halmos–Savage theorem, $\mathcal{Q}$ contains a measure $Q_1 \sim P^*$. We get

$$1 = Q_0[\pi < \infty] = \lim_{t \uparrow \infty} Q_0[\pi \leq t] = \lim_{t \uparrow \infty} \inf_{Q \in \mathcal{Q}} Q[\pi \leq t] \leq Q_1[\pi < \infty].$$

Hence, also $P^*[\pi < \infty] = 1$ and in turn $P^* \ll Q_0$.

A number of examples for least favorable measures are given in Section 3.3. We next state a characterization of least favorable measures that is a variant of Theorem 3.4 in HUBER and STRASSEN [1973] and in this form taken from SCHIED [2005, proposition 3.1].

PROPOSITION 3.1. *For $Q_0 \in \mathcal{Q}$ with $Q_0 \sim P^*$ and $\pi := dP^*/dQ_0$, the following conditions are equivalent:*

- (a) Q_0 is a least favorable measure for P^* .
- (b) For all decreasing functions $f : [0, \infty] \rightarrow \mathbb{R}$ such that $\inf_{Q \in \mathcal{Q}} E_Q[f(\pi) \wedge 0] > -\infty$,

$$\inf_{Q \in \mathcal{Q}} E_Q[f(\pi)] = E_{Q_0}[f(\pi)].$$

- (c) Q_0 minimizes the g -divergence

$$I_g(Q|P^*) = E_{P^*}\left[g\left(\frac{dQ}{dP^*}\right)\right]$$

among all $Q \in \mathcal{Q}$, for all continuous convex functions $g : [0, \infty] \rightarrow \mathbb{R}$ such that $I_g(P^*|Q)$ is finite for some $Q \in \mathcal{Q}$.

Sketch of proof. According to the definition, Q_0 is a least favorable measure if and only if $Q_0 \circ \pi^{-1}$ stochastically dominates $Q \circ \pi^{-1}$ for all $Q \in \mathcal{Q}$. Hence, the equivalence of (a) and (b) is just the standard characterization of stochastic dominance (see FÖLLMER and SCHIED [2004, Theorem 2.71]). Here and in the next step, some care is needed if f is unbounded or discontinuous.

For showing the equivalence of (b) and (c), let the continuous functions f and g be related by $g(x) = \int_1^x f(t/t) dt$. Then, g is convex if and only if f is decreasing. For $Q_1 \in \mathcal{Q}$, we let $Q_t := tQ_1 + (1-t)Q_0$ and $h(t) := I_g(Q_t|P^*)$. The right-hand derivative of h is given by $h'_+(0) = E_{Q_1}[f(\pi)] - E_{Q_0}[f(\pi)]$, which shows that (b) is the first-order condition for the minimization problem in (c).

The following result from SCHIED [2005] reduces the robust utility maximization problem to a standard utility maximization problem and the computation of a least favorable measure, which is independent of the utility function.

THEOREM 3.1. *Suppose that $\mathcal{Q}$ admits a least favorable measure Q_0 . Then the robust utility maximization problem (3.5) is equivalent to the standard utility maximization*

problem with subjective measure Q_0 , that is, to the problem (3.7) with the choice $\mathcal{Q} = \{Q_0\}$. More precisely, $X_T^* \in \mathcal{X}(x)$ solves the robust problem (3.5) if and only if it solves the standard problem for Q_0 and the corresponding value functions are equal, whether there exists a solution or not:

$$\sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] = \sup_{X \in \mathcal{X}(x)} E_{Q_0}[U(X_T)], \quad \text{for all } x.$$

Idea of proof: For simplicity, we only consider the situation in which the corresponding standard problem for Q_0 admits a unique solution X^0 . By standard theory, the final terminal wealth is of the form $X_T^0 = I(\lambda\pi)$, where λ is a positive constant and I is the inverse of the function U' (see FÖLLMER and SCHIED [2004, section 3.3]). We then have for any $X \in \mathcal{X}(x)$ that is not identical to X^0 :

$$\inf_{Q \in \mathcal{Q}} E_Q[U(X_T)] \leq E_{Q_0}[U(X_T)] < E_{Q_0}[U(X_T^0)] = \inf_{Q \in \mathcal{Q}} E_Q[U(X_T^0)]. \quad (3.8)$$

Here, we have used Proposition 3.1 in the last step. This proves that X^0 is also the unique solution to the robust problem. In the general case, one needs additional arguments (see SCHIED [2005]).

The preceding result has the following economic consequence. Let $\succ$ denote the preference order induced by our robust utility functional, that is,

$$X \succ Y \iff \inf_{Q \in \mathcal{Q}} E_Q[U(X)] > \inf_{Q \in \mathcal{Q}} E_Q[U(Y)].$$

Then, although $\succ$ does not satisfy the axioms of (subjective) expected utility theory, optimal investment decisions with respect to $\succ$ are still made in accordance with von Neumann–Morgenstern expected utility, provided that we take Q_0 as the subjective probability measure. The surprising part is that this subjective measure neither depends on the initial investment $x = X_0$ nor on the choice of the utility function U . If $\mathcal{Q}$ does not admit a least favorable measure, then it is still possible that the robust problem is equivalent to a standard utility maximization problem with a subjective measure Q , which then, however, will depend on x and U . We also have the following converse to Theorem 3.1:

THEOREM 3.2. *Suppose $Q_0 \in \mathcal{Q}$ is such that for all utility functions and all $x > 0$, the robust utility maximization problem (3.5) is equivalent to the standard utility maximization problem with respect to Q_0 . Then Q_0 is a least favorable measure in the sense of Definition 3.1.*

PROOF. See SCHIED [2005]. □

With some additional care, the argument combined in the proofs of Proposition 3.1 and Theorem 3.1 extends to the case in which there exists no least favorable measure for $\mathcal{Q}$ in the sense of Definition 3.1. To explain this extension, which was carried out by GUNDEL [2005], let us assume for simplicity that each $Q \in \mathcal{Q}$ is equivalent to $\mathbb{P}$ and

admits a unique $X^Q \in \mathcal{X}(x)$ that solves the standard problem for Q and is such that $E_{P^*}[X_T^Q] = x$. The goal is to find some $Q_0 \in \mathcal{Q}$ for which

$$E_{Q_0}[U(X_T^{Q_0})] = \inf_{Q \in \mathcal{Q}} E_Q[U(X_T^{Q_0})],$$

for then we can conclude as in (3.8) that X^{Q_0} must be the solution to the robust problem. It is well known that the final terminal wealth of X^Q is of the form $X_T^Q = I(\lambda_Q dP^*/dQ)$, where λ_Q is a positive constant depending on Q , and I is the inverse of the function U' (see FÖLLMER and SCHIED [2004, section 3.3]). If U satisfies the Inada conditions, that is, $U'(0+) = \infty$ and $U'(\infty-) = 0$, then

$$U(X_T^Q) = U\left(I\left(\lambda_Q \frac{dP^*}{dQ}\right)\right) = \tilde{U}\left(\lambda_Q \frac{dP^*}{dQ}\right) + \lambda_Q \frac{dP^*}{dQ} \cdot X_T^Q,$$

where $\tilde{U}(y) := \sup_{x>0} (U(x) - xy)$ denotes the convex conjugate of U . Hence,

$$E_Q[U(X_T^Q)] = E_Q\left[\tilde{U}\left(\lambda_Q \frac{dP^*}{dQ}\right)\right] + \lambda_Q x.$$

Using the standard fact that λ_Q is the minimizer of the right-hand side when regarded as a function of $\lambda = \lambda_Q$ (see KRAMKOV and SCHACHERMAYER [1999, Theorem 2.0]), we thus obtain the following result from GUNDEL [2005].

THEOREM 3.3. *In addition to the preceding assumptions, suppose that Q_0 is a measure in $\mathcal{Q}$ attaining the infimum of the function*

$$Q \longmapsto \inf_{\lambda} \left(E_Q\left[\tilde{U}\left(\lambda \frac{dP^*}{dQ}\right)\right] + \lambda x \right), \quad Q \in \mathcal{Q}. \quad (3.9)$$

Then X^{Q_0} solves problem (3.7).

REMARK 3.3. Suppose λ_0 is a minimizer of the function

$$\lambda \longmapsto \inf_{Q \in \mathcal{Q}} E_Q\left[\tilde{U}\left(\lambda \frac{dP^*}{dQ}\right)\right] + \lambda x.$$

Then, the function (3.9) is equal to

$$E_Q\left[\tilde{U}\left(\lambda_0 \frac{dP^*}{dQ}\right)\right] + \lambda_0 x = E_{P^*}\left[\frac{dQ}{dP^*} \tilde{U}\left(\lambda_0 \frac{dP^*}{dQ}\right)\right] + \lambda_0 x = I_g(Q|P^*) + \lambda_0 x,$$

where $I_g(Q|P^*)$ is the g -divergence associated with the convex function $g(x) = x\tilde{U}(\lambda_0/x)$ (see (2.9)). The measure Q_0 in Theorem 3.3, therefore, can be characterized as the minimizer in $\mathcal{Q}$ of $I_g(\cdot|P^*)$ for this particular choice of g . This fact and Proposition 3.1 provide the connection to the solution to the problem via least favorable measures. Note that in the present context, g typically depends on both U and x , and so does Q_0 unless it is a least favorable measure. Theorem 3.3 can be extended to an

incomplete market model by considering $P^* \in \mathcal{P}$ as an additional argument in (3.9). For details, we refer to GUNDEL [2005]. From a probabilistic point of view, the problem of robust portfolio choice can, in fact, be regarded as a new version of a classical projection problem: We are looking for a pair $(\hat{P}^*, \hat{Q})$ that minimizes a certain divergence functional on the product $\mathcal{P} \times \mathcal{Q}$ of two convex sets of probability measures. For a systematic discussion of this robust projection problem and of a more flexible version where the class $\mathcal{P}$ of equivalent martingale measures is replaced by a larger class of extended martingale measures, we refer to FÖLLMER and GUNDEL [2006] (see also Remark 4.1 below).

3.3. Least favorable measures and their relation to capacity theory

In the preceding section, it was shown that least favorable measures in the sense of Definition 3.1 provide the solution to the robust utility maximization problem (3.7) in a complete market model. In this section, we discuss a general existence result for least favorable measures in the context of capacity theory, namely, the Huber–Strassen theorem. We also provide a number of explicit examples. This connection between Huber–Strassen theory and robust utility maximization was derived by SCHIED [2005].

In Theorem 2.4 we have discussed the assumption of comonotonic independence, which is reasonable insofar as comonotonic positions cannot act as mutual hedges and which is equivalent to the fact that ϕ is comonotonic. It is easy to see that every comonotonic concave monetary utility functional is coherent (see FÖLLMER and SCHIED [2004, lemma 4.77]). Let $\mathcal{Q}$ be the corresponding maximal representing set. Then, comonotonicity is equivalent to the fact that the nonadditive set function

$$\hat{\kappa}(A) := \phi(\mathbb{I}_A) = \inf_{Q \in \mathcal{Q}} Q[A], \quad A \in \mathcal{F}_T,$$

is supermodular in the sense of Choquet:

$$\hat{\kappa}(A \cup B) + \hat{\kappa}(A \cap B) \geq \hat{\kappa}(A) + \hat{\kappa}(B) \quad \text{for } A, B \in \mathcal{F}_T.$$

In this case, $\phi(X)$ can be expressed as the Choquet integral of X with respect to $\hat{\kappa}$, that is,

$$\phi(X) = \int_0^\infty \hat{\kappa}(X > t) dt, \quad \text{for } X \geq 0.$$

These results are due to CHOQUET [1953/54]. We refer to FÖLLMER and SCHIED [2004, theorem 4.88] for a proof in terms of the set function

$$\kappa(A) := 1 - \hat{\kappa}(A^c) = \sup_{Q \in \mathcal{Q}} Q[A],$$

which is submodular in the sense of Choquet:

$$\kappa(A \cup B) + \kappa(A \cap B) \leq \kappa(A) + \kappa(B) \quad \text{for } A, B \in \mathcal{F}_T.$$

In fact, it will be convenient to work with κ in the sequel. Let us introduce the following technical assumption:

$$\begin{aligned} & \text{There exists a Polish topology on } \Omega \text{ such that} \\ & \mathcal{F}_T \text{ is the corresponding Borel field and } \mathcal{Q} \text{ is compact.} \end{aligned} \quad (3.10)$$

It guarantees that κ is a capacity in the sense of Choquet. Assuming that κ is submodular, let us consider the submodular set function

$$w_t(A) := t\kappa(A) - P^*[A], \quad A \in \mathcal{F}_T. \quad (3.11)$$

It is shown in lemmas 3.1 and 3.2 of HUBER and STRASSEN [1973] that under condition (3.10), there exists a decreasing family $(A_t)_{t>0} \subset \mathcal{F}_T$ such that A_t minimizes w_t and such that the continuity condition $A_t = \bigcup_{s>t} A_s$ is satisfied.

DEFINITION 3.2. The function

$$\frac{dP^*}{d\kappa}(\omega) = \inf \{ t \mid \omega \notin A_t \}, \quad \omega \in \Omega$$

is called the Radon–Nikodym derivative of P^* with respect to the Choquet capacity κ .

The terminology *Radon–Nikodym derivative* comes from the fact that $dP^*/d\kappa$ coincides with the usual Radon–Nikodym derivative dP^*/dQ in case where $\mathcal{Q} = \{Q\}$ (see HUBER and STRASSEN [1973]). Let us now state the celebrated Huber–Strassen theorem in a form in which it will be needed here.

THEOREM 3.4 (Huber–Strassen). *If κ is submodular and (3.10) holds, then $\mathcal{Q}$ admits a least favorable measure Q_0 with respect to any probability measure R on $(\Omega, \mathcal{F}_T)$. Moreover, if $R = P^*$ and $\mathcal{Q}$ satisfies (3.6), then Q_0 is equivalent to P^* and given by*

$$dQ_0 = \left(\frac{dP^*}{d\kappa} \right)^{-1} dP^*.$$

PROOF. See HUBER and STRASSEN [1973]. One also needs the fact that $\mathbb{P}[0 < dP^*/d\kappa < \infty] = 1$ (see SCHIED [2005, Lemma 3.1]). $\square$

Together with Theorem 3.1, we get a complete solution to the robust utility maximization problem within the large class of utility functionals that arise from comonotonic coherent monetary utility functionals under assumption (3.10). Before discussing particular examples, let us state the following converse of the Huber–Strassen theorem in order to clarify the role of comonotonicity. For finite probability spaces, Theorem 3.5 is due to HUBER and STRASSEN [1973]. In the form stated above, it was proved by LEMBCKE [1988]. An alternative formulation was given by BEDNARSKI [1982].

THEOREM 3.5. *Suppose (3.10) is satisfied. If $\mathcal{Q}$ is a convex set of probability measures closed in total variation distance such that every probability measure on $(\Omega, \mathcal{F}_T)$ admits a least favorable measure $Q_0 \in \mathcal{Q}$, then $\kappa(A) = \sup_{Q \in \mathcal{Q}} Q[A]$ is submodular.*

PROOF. See LEMBCKE [1988]. $\square$

In Theorem 3.5, it is crucial to require the existence of a least favorable measure with respect to every probability measure on $(\Omega, \mathcal{F}_T)$. We encounter a situation in which least favorable measures exist for certain but not for all probability measures on $(\Omega, \mathcal{F})$, and the corresponding set function κ will not be submodular.

Let us now turn to the discussion of particular examples. The following example class was first studied by BEDNARSKI [1981] under slightly different conditions than reported here. These examples also play a role in the theory of law-invariant risk measures (see KUSUOKA [2001] and sections 4.4 through 4.7 in FÖLLMER and SCHIED [2004]).

EXAMPLE 3.2. The following class of submodular set functions arises in the “dual theory of choice under risk” as proposed by YAARI [1987]. Let $\psi : [0, 1] \rightarrow [0, 1]$ be an increasing concave function with $\psi(0) = 0$ and $\psi(1) = 1$. In particular, ψ is continuous on $[0, 1]$. We define κ by

$$\kappa(A) := \psi(P[A]), \quad A \in \mathcal{F}.$$

Then, κ is submodular and gives rise to a comonotonic monetary utility functional defined as the Choquet integral of $\hat{\kappa}(A) := 1 - \kappa(A)$, and the corresponding maximal representing set $\mathcal{Q}$ can be described in terms of ψ (see CARLIER and DANA [2003] for the case in which ψ is C^1 and FÖLLMER and SCHIED [2004, theorem 4.73 and corollary 4.74] for the general case). If $(\Omega, \mathcal{F}_T)$ is a standard Borel space, then there exists a compact metric topology on Ω whose Borel field is $\mathcal{F}_T$. For such a topology, $\mathcal{Q}$ is weakly compact, and so (3.10) is satisfied. Consequently, $\mathcal{Q}$ admits a least favorable measure Q_0 . It can be explicitly determined in the case in which $\psi(t) = (t\lambda^{-1}) \wedge 1$ for some $\lambda \in [0, 1]$, which corresponds to (2.10) and hence to the risk measure $AVaR_\lambda$. To state this result, we assume that the price density $Z := P^*/d\mathbb{P}$ has a continuous distribution $F_Z(x) = \mathbb{P}[Z \leq x]$. By q_Z , we denote a corresponding quantile function, that is, a generalized inverse of the increasing function F_Z . With this notation, the Radon–Nikodym derivative of P^* with respect to κ is given by

$$\pi = \frac{dP^*}{d\kappa} = c \cdot (Z \vee q_Z(t_\lambda)),$$

where c is the normalizing constant and t_λ is the unique maximizer of the function

$$t \mapsto \frac{(t - 1 + \lambda)^+}{\int_0^t q_Z(s) ds}$$

(see SCHIED [2004, 2005]).

EXAMPLE 3.3 (Weak information). Let Y be a measurable function on $(\Omega, \mathcal{F}_T)$ and denote by μ its law under P^* . For $\nu \sim \mu$ given, let

$$\mathcal{Q} := \left\{ Q \ll P^* \mid Q \circ Y^{-1} = \nu \right\}.$$

The robust utility maximization problem for this set $\mathcal{Q}$ was studied by BAUDOIN [2002], who coined the terminology *weak information*. The interpretation behind the set $\mathcal{Q}$ is that an investor has full knowledge about the pricing measure P^* but is uncertain about the true distribution P of market prices and only knows that a certain functional Y of the stock price has distribution ν . Define Q_0 by

$$dQ_0 = \frac{d\nu}{d\mu}(Y) dP^*.$$

Then $Q_0 \in \mathcal{Q}$ and the law of $\pi := dQ_0/dP^* = d\mu/d\nu(Y)$ is the same for all $Q \in \mathcal{Q}$. Hence, Q_0 satisfies the definition of a least favorable measure. The same procedure can be applied to any measure $R \sim P^*$. Using this fact and Theorem 3.5, one can show that $\mathcal{Q}$ fits into the framework of the Huber–Strassen theory, that is, $\kappa(A) := \sup_{Q \in \mathcal{Q}} Q[A]$ is submodular (see SCHIED [2005, proposition 3.4]).

In the 1970s and 1980s, explicit formulas for Radon–Nikodym derivatives with respect to capacities were found in a number of examples such as sets $\mathcal{Q}$ defined in terms of ε -contamination or via probability metrics like total variation or Prohorov distance; we refer to chapter 10 in the book by HUBER [1981] and the references therein. But, unless Ω is finite, these examples fail to satisfy either implication in (3.6). Nevertheless, they are still interesting for discrete-time market models.

We now study a situation in which a least favorable measure exists although the Huber–Strassen theorem does not apply. To this end, we consider a Black–Scholes market model with d risky assets $S_t = (S_t^1, \dots, S_t^d)$ that satisfy a stochastic differential equations (SDE) of the form

$$dS_t^i = S_t^i \sum_{j=1}^d \sigma_t^{ij} dW_t^j + \alpha_t^i S_t^i dt \quad (3.12)$$

for a d -dimensional Brownian motion $W = (W^1, \dots, W^d)$ and a volatility matrix σ_t that has full rank. Now suppose the investor is uncertain about the future drift $\alpha_t = (\alpha_t^1, \dots, \alpha_t^d)$ in the market: any drift α is possible that is adapted to the filtration generated by W and satisfies $\alpha_t \in C_t$, where C_t is a nonrandom bounded closed convex subset of $\mathbb{R}^d$. Let us denote by $\mathcal{A}$ the set of all such processes α . This uncertainty in the choice of the drift can be expressed by the set

$$\mathcal{Q} := \left\{ Q \mid S \text{ has drift } \alpha^Q \in \mathcal{A} \text{ under } Q \right\}.$$

Under P^* , the drift α in (3.12) vanishes. We denote by α_t^0 the element in C_t that minimizes the norm $|\sigma_t^{-1}x|$ among all $x \in C_t$.

THEOREM 3.6. *Suppose that σ_t is deterministic and that both α_t^0 and σ_t are continuous in t . Then $\mathcal{Q}$ admits a least favorable measure Q_0 with respect to P^* , which is characterized by having the drift α^0 .*

PROOF. In SCHIED [2005, proposition 3.2], the problem is solved by transforming it into a problem for uncertain volatility as studied by EL KAROUI, JEANBLANC–PICQUÉ and SHREVE [1998]. $\square$

An obvious question is whether the strong condition that the volatility σ_t and the drift α^0 are deterministic can be relaxed. One case of interest is, for instance, a local volatility model in which the Eq. (3.12) is replaced with the one-dimensional SDE

$$dS_t = \sigma(t, S_t)S_t dW_t + \alpha_t S_t dt. \quad (3.13)$$

In this case, however, it may occur that it is no longer optimal to take the drift that is closest to the riskneutral case $\alpha \equiv 0$. The reason is that the utility of an investment can be reduced by both a small drift and a large volatility, and these two requirements may be competing with each other. This effect may also destroy the existence of a least favorable measure (see HERNÁNDEZ–HERNÁNDEZ and SCHIED [2006, example 2.7] for the discussion of a related trade-off effect). Furthermore, SCHIED [2005, proposition 3.3] discusses the examples in which no least favorable measure exist due to the fact that either the coefficient σ or the least favorable drift α_t^0 is not deterministic.

3.4. Duality techniques in incomplete markets

In this section, we discuss the general duality theory for robust portfolio choice in a very general setting and under rather weak assumptions. The results presented here build on the corresponding results for ordinary utility maximization as obtained by KRAMKOV and SCHACHERMAYER [1999, 2003]. The duality theory for coherent robust utility functionals was first developed by QUENEZ [2004] and later extended by SCHIED and WU [2005] and SCHIED [2007a]. Our exposition follows the latter article. Recently, WITTMÜSS [2006] further extended these results to cover also the cases of consumption-investment strategies and random endowment (see also BURGERT and RÜSCHENDORF [2005] for some earlier results in that direction). Related questions arise for the problem of efficiently hedging a contingent claim when risk is measured in terms of a convex risk measure (see CVITANIC and KARATZAS [2001], KIRCH [2000], KIRCH and RUNGGALDIER [2005], FAVERO [2001], FAVERO and RUNGGALDIER [2002], SCHIED [2004, 2006], RUDLOFF [2006], SEKINE [2004], KLÖPPEL and SCHWEIZER [2007]).

The main importance of the duality method lies in the fact that the dual problem is often simpler than the primal one. Therefore, it can be advantageous to combine duality with another optimization technique such as optimal stochastic control. This is already true for the maximization of classical von Neumann–Morgenstern utility. But for robust utility maximization, the duality method has the additional advantage that the dual problem simply involves the minimization of a convex functional. The primal problem, however, requires to find a saddlepoint of a functional that is concave in one argument and convex in the other. This fact will become important in Section 3.5, where stochastic control techniques are used to solve the dual rather than the primal problem.

In addition to the assumptions stated in Section 3.1, we assume that the utility function $U : [0, \infty) \rightarrow \mathbb{R}$ is continuously differentiable and satisfies the Inada conditions

$$U'(0+) = +\infty \quad \text{and} \quad U'(\infty-) = 0.$$

We also assume that the concave monetary utility functional,

$$\phi(Y) = \inf_{Q \in \mathcal{Q}} (E_Q[Y] + \gamma(Q)), \quad Y \in L^\infty(\mathbb{P}), \quad (3.14)$$

is continuous from below as defined in Theorem 2.2.

The value function of the robust problem is defined as

$$u(x) := \sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} (E_Q[U(X_T)] + \gamma(Q)).$$

We also need the value function of the optimal investment problem for an investor with subjective measure $Q \in \mathcal{Q}$:

$$u_Q(x) := \sup_{X \in \mathcal{X}(x)} E_Q[U(X_T)].$$

Next, we define the convex conjugate function $\tilde{U}$ of U by

$$\tilde{U}(y) := \sup_{x > 0} (U(x) - xy), \quad y > 0.$$

With this notation, KRAMKOV and SCHACHERMAYER [1999, theorem 3.1] states that for $Q \sim \mathbb{P}$ with finite value function u_Q ,

$$u_Q(x) = \inf_{y > 0} (\tilde{u}_Q(y) + xy) \quad \text{and} \quad \tilde{u}_Q(y) = \sup_{x > 0} (u_Q(x) - xy), \quad (3.15)$$

where the dual value function $\tilde{u}_Q$ is given by

$$\tilde{u}_Q(y) = \inf_{Y \in \mathcal{Y}_Q(y)} E_Q[\tilde{U}(Y_T)],$$

and the space $\mathcal{Y}_Q(y)$ is defined as the set of all positive Q supermartingales such that $Y_0 = y$ and XY is a Q supermartingale for all $X \in \mathcal{X}(1)$. Note that this definition also makes sense for measures $Q \ll \mathbb{P}$ that are not equivalent to $\mathbb{P}$ although in this case the duality relations (3.15) need not hold. We next define the dual value function of the robust problem by

$$\tilde{u}(y) := \inf_{Q \in \mathcal{Q}} (\tilde{u}_Q(y) + \gamma(Q)) = \inf_{Q \in \mathcal{Q}} \inf_{Y \in \mathcal{Y}_Q(y)} (E_Q[\tilde{U}(Y_T)] + \gamma(Q)).$$

DEFINITION 3.3. Let $y > 0$ be such that $\tilde{u}(y) < \infty$. A pair (Q, Y) is a *solution to the dual problem* if $Q \in \mathcal{Q}$, $Y \in \mathcal{Y}_Q(y)$ and $\tilde{u}(y) = E_Q[\tilde{U}(Y_T)] + \gamma(Q)$.

Let us finally introduce the set $\mathcal{Q}_e$ of measures in $\mathcal{Q}$ that are equivalent to $\mathbb{P}$:

$$\mathcal{Q}_e := \{Q \in \mathcal{Q} \mid Q \sim \mathbb{P}\}.$$

The facts that γ is the minimal penalty function of ϕ and ϕ is sensitive guarantee that $\mathcal{Q}_e$ is always nonempty. This follows from the Halmos–Savage theorem similarly to the argument in Remark 3.2.

THEOREM 3.7. *In addition to the above assumptions, let us assume that*

$$u_{Q_0}(x) < \infty \text{ for some } x > 0 \text{ and some } Q_0 \in \mathcal{Q}_e \quad (3.16)$$

and that

$$\tilde{u}(y) < \infty \text{ implies } \tilde{u}_{Q_1}(y) < \infty \text{ for some } Q_1 \in \mathcal{Q}_e. \quad (3.17)$$

Then the robust value function u is concave, takes only finite values, and satisfies

$$u(x) = \sup_{X \in \mathcal{X}(x)} \inf_{Q \in \mathcal{Q}} (E_Q[U(X_T)] + \gamma(Q)) = \inf_{Q \in \mathcal{Q}} \sup_{X \in \mathcal{X}(x)} (E_Q[U(X_T)] + \gamma(Q)).$$

Moreover, the two robust value functions u and $\tilde{u}$ are conjugate to another:

$$u(x) = \inf_{y > 0} (\tilde{u}(y) + xy) \quad \text{and} \quad \tilde{u}(y) = \sup_{x > 0} (u(x) - xy). \quad (3.18)$$

In particular, $\tilde{u}$ is convex. The derivatives of u and $\tilde{u}$ satisfy

$$u'(0+) = \infty \quad \text{and} \quad \tilde{u}'(\infty-) = 0. \quad (3.19)$$

Furthermore, if $\tilde{u}(y) < \infty$, then the dual problem admits a solution $(\hat{Q}, \hat{Y})$ that is maximal in the sense that any other solution (Q, Y) satisfies $Q \ll \hat{Q}$ and $Y_T = \hat{Y}_T$ Q -a.s.

PROOF. See SCHIED [2007a, theorem 2.3]. □

It is possible that the maximal $\hat{Q}$ is not equivalent to $\mathbb{P}$ (see SCHIED [2007a, example 3.2]). If this happens, then $\hat{Q}$ considered a financial market model on its own may admit arbitrage opportunities. In this light, one also has to understand the conditions (3.16) and (3.17): they exclude the possibility that the value functions u_Q and $\tilde{u}_Q$ are only finite for some degenerate model $Q \in \mathcal{Q}$, for which the duality relations (3.15) need not hold.

The situation simplifies considerably if we assume that all measures in $\mathcal{Q}$ are equivalent to $\mathbb{P}$. In this case, condition (3.17) is always satisfied, and (3.16) can be replaced with the assumption that $u(x) < \infty$ for some $x > 0$. Moreover, the optimal $\hat{Y}$ is then $\mathbb{P}$ -almost surely unique. Despite this fact, however, and in contrast to the situation in standard utility maximization, it can happen that the dual-value function $\tilde{u}$ is not strictly convex—even if all measures in $\mathcal{Q}$ are equivalent to $\mathbb{P}$ (see SCHIED [2007a, example 3.1]). Equivalently, the value function u may fail to be continuously differentiable. A sufficient condition for the strict convexity of $\tilde{u}$ and the continuous differentiability of u is given in the next result. It applies, in particular, to entropic penalties and to penalty functions defined in terms of many other statistical distance functions as described in Section 2.2.

PROPOSITION 3.2. *Suppose that the assumptions of Theorem 3.7 are satisfied and γ is strictly convex on $\mathcal{Q}$. Then, u is continuously differentiable and $\tilde{u}$ is strictly convex on its domain.*

PROOF. See SCHIED [2007a, proposition 2.4]. $\square$

Our next aim is to get existence results for optimal strategies. In the classical case $\mathcal{Q} = \{\mathbb{P}\}$, it was shown by KRAMKOV and SCHACHERMAYER [2003] that a necessary and sufficient condition for the existence of optimal strategies at each initial capital is the finiteness of the dual-value function $\tilde{u}_{\mathbb{P}}$. This condition translates as follows to our robust setting:

$$\tilde{u}_Q(y) < \infty, \quad \text{for all } y > 0 \text{ and each } Q \in \mathcal{Q}_e. \quad (3.20)$$

It was shown by KRAMKOV and SCHACHERMAYER [2003, note 2] that (3.20) holds as soon as u_Q is finite for all $Q \in \mathcal{Q}_e$ and the asymptotic elasticity of the utility function U is strictly less than 1:

$$AE(U) = \limsup_{x \uparrow \infty} \frac{xU'(x)}{U(x)} < 1.$$

THEOREM 3.8. *In addition to the assumptions of Theorem 3.7, let us assume (3.20). Then both the value functions u and $\tilde{u}$ take only finite values and satisfy*

$$u'(\infty-) = 0 \quad \text{and} \quad \tilde{u}'(0+) = -\infty. \quad (3.21)$$

The robust value function u is strictly concave, and the dual-value function $\tilde{u}$ is continuously differentiable. Moreover, for any $x > 0$, there exists an optimal strategy $\hat{X} \in \mathcal{X}(x)$ for the robust problem. If $y > 0$ is such that $\tilde{u}'(y) = -x$ and $(\hat{Q}, \hat{Y})$ is a solution to the dual problem, then

$$\hat{X}_T = I(\hat{Y}_T) \quad \hat{Q}\text{-a.s.} \quad (3.22)$$

for $I := -\tilde{U}'$, and $(\hat{Q}, \hat{X})$ is a saddlepoint for the robust problem

$$u(x) = \inf_{Q \in \mathcal{Q}} (E_Q[U(\hat{X}_T)] + \gamma(Q)) = E_{\hat{Q}}[U(\hat{X}_T)] + \gamma(\hat{Q}) = u_{\hat{Q}}(x) + \gamma(\hat{Q}).$$

Furthermore, $\hat{X}\hat{Y}\hat{Z}$ is a martingale under $\mathbb{P}$, where $(\hat{Z}_t)_{0 \leq t \leq T}$ is the density process of $\hat{Q}$ with respect to $\mathbb{P}$.

PROOF. See SCHIED [2007a, theorem 2.5]. $\square$

REMARK 3.4. In the preceding theorem, let us take $(\hat{Q}, \hat{Y})$ as a maximal solution to the dual problem as constructed in Theorem 3.7. Then, the solution $\hat{X}_T$ will be $\mathbb{P}$ -a.s. unique as soon as $\hat{Q} \sim \mathbb{P}$. This equivalence holds trivially if all measures in $\mathcal{Q}$ are equivalent to $\mathbb{P}$. In the general case, however, $\hat{Q}$ need not be equivalent to $\mathbb{P}$ so that

(3.22) cannot guarantee the $\mathbb{P}$ -a.s. uniqueness of $\hat{X}_T$ (see SCHIED [2007a, example 3.2]). Nevertheless, we can construct an optimal strategy from a given solution to the dual problem by superhedging an appropriate contingent claim $H \geq 0$. To this end, suppose that the assumptions of Theorem 3.8 hold. Let $(\hat{Q}, \hat{Y})$ be a solution to the dual problem at level $y > 0$ and consider the contingent claim

$$H := I(\hat{Y}_T) \mathbb{I}_{\{\hat{Z} > 0\}},$$

where $d\hat{Q} = \hat{Z} d\mathbb{P}$. Then, $x = -\tilde{u}'(y)$ is the minimal initial investment $x' > 0$ for which there exists some $X \in \mathcal{X}(x')$ such that $X_T \geq H$ $\mathbb{P}$ -a.s. Furthermore, if $\hat{X} \in \mathcal{X}(x)$ is such a strategy, then it is a solution to the robust utility maximization problem at initial capital x (see SCHIED [2007a, corollary 2.6]).

REMARK 3.5. Instead of working with the terminal values of processes in the space $\mathcal{Y}_Q(y)$, it is sometimes more convenient to work with the densities of measures in the set $\mathcal{P}$ of equivalent local martingale measures. In fact, one can show that the dual value function satisfies

$$\tilde{u}(y) = \inf_{P^* \in \mathcal{P}} \inf_{Q \in \mathcal{Q}_e} \left(E_Q \left[\tilde{U} \left(y \frac{dP^*}{dQ} \right) \right] + \gamma(Q) \right) \quad (3.23)$$

(see SCHIED [2007a, remark 2.7]). Since the infimum in (3.23) need not be attained, it is often not possible to represent the optimal solution $\hat{X}_T$ in terms of the density of an equivalent martingale measure. However, FÖLLMER and GUNDEL [2006] recently observed that the elements of $\mathcal{Y}_Q(1)$ can be interpreted as density processes of extended martingale measures, as explained in Remark 4.1.

3.5. Solution with stochastic control techniques

Stochastic control techniques for solving robust utility maximization problems were used by HANSEN and SARGENT [2001], TALAY and ZHENG [2002], KORN and WILMOTT, [2002], KORN and MENKENS [2005], KORN and STEFFENSEN [2006], HERNÁNDEZ-HERNÁNDEZ and SCHIED [2006, 2007a, 2007b], SCHIED [2007b], and DOKUCHAEV [2007]. Here, we consider an incomplete market model with a risky asset, whose volatility and long-term trend are driven by an external stochastic factor process. The robust utility functional is defined in terms of a hyperbolic absolute risk aversion (HARA) utility function with risk-aversion parameter $\alpha \in \mathbb{R}$ and a dynamically consistent concave or coherent monetary utility functional, which allows for model uncertainty in the distributions of both the asset price dynamics and the factor process. The exposition follows HERNÁNDEZ-HERNÁNDEZ and SCHIED [2006, 2007a], and SCHIED [2007b], and the main idea is to apply stochastic control techniques to the dual rather than the primal problem. This has the advantage that the dual problem is a pure minimization problem, while the original primal problem is a minimax problem so that the associated nonlinear PDE would be of Hamilton–Jacobi–Bellman–Isaacs type. This idea is well known in nonlinear optimization. In the context of robust utility maximization, it was first used by QUENEZ [2004] to facilitate the use of backward stochastic

differential equations (BSDE) techniques (cf. Section 3.6; see CASTAÑEDA-LEYVA and HERNÁNDEZ-HERNÁNDEZ [2005] for a related control approach to the dual problem of a standard utility maximization problem and we refer to FLEMING and SONER [1993] for an introduction to stochastic control).

We first describe the financial market model. Under the reference measure $\mathbb{P}$, the risky asset is defined through the SDE of the following factor model:

$$dS_t = S_t b(Y_t) dt + S_t \sigma(Y_t) dW_t^1 \quad (3.24)$$

with deterministic initial condition S_0 . Here, W^1 is a standard $\mathbb{P}$ -Brownian motion, and Y denotes an external economic factor process modeled by the SDE

$$dY_t = g(Y_t) dt + \rho_1(Y_t) dW_t^1 + \rho_2(Y_t) dW_t^2 \quad (3.25)$$

for a standard $\mathbb{P}$ -Brownian motion W^2 , which is independent of W^1 under $\mathbb{P}$. We suppose that the economic factor can be observed but cannot be traded directly so that the market model is typically incomplete. Models of this type have been widely used in finance and economics, the case of a mean-reverting factor process with the choice $g(y) := -\kappa(\mu - y)$ being particularly popular (see FOUQUE, PAPANICOLAOU and SIRCAR [2000], FLEMING and HERNÁNDEZ-HERNÁNDEZ [2003], and the references therein). We assume that g belongs to $C^2(\mathbb{R})$, with derivative $g' \in C_b^1(\mathbb{R})$, and b , σ , ρ_1 , and ρ_2 belong to $C_b^2(\mathbb{R})$, where $C_b^k(\mathbb{R})$ denotes the class of bounded functions with bounded derivatives up to order k . We will also assume that

$$\sigma(y) \geq \sigma_0 \text{ and } a(y) := \frac{1}{2}(\rho_1^2(y) + \rho_2^2(y)) \geq \sigma_1^2 \text{ for some constants } \sigma_0, \sigma_1 > 0. \quad (3.26)$$

The market price of risk with respect to the reference measure $\mathbb{P}$ is defined via the function

$$\theta(y) := \frac{b(y)}{\sigma(y)}.$$

The assumption of time-independent coefficients is for convenience only. It is also easy to extend our results to a d -dimensional stock market model replacing the one-dimensional SDE (3.24).

REMARK 3.6. By taking $\rho_2 \equiv 0$, $\rho_1(y) = \sigma(y)$, $g(y) = b(y) - \frac{1}{2}\sigma^2(y)$, and $Y_0 = \log S_0$, it follows that Y coincides with $\log S$. Hence, S solves the SDE of a local volatility model:

$$dS_t = S_t \tilde{b}(S_t) dt + S_t \tilde{\sigma}(S_t) dW_t^1, \quad (3.27)$$

where $\tilde{b}(x) = b(\log x)$ and $\tilde{\sigma}(x) = \sigma(\log x)$. Thus, our analysis includes the study of the robust optimal investment problem for local volatility models given by (3.27).

To define $\gamma(Q)$, we assume henceforth that $(\Omega, \mathcal{F}, (\mathcal{F}_t))$ is the canonical path space of $W = (W^1, W^2)$. Then, every probability measure $Q \ll \mathbb{P}$ admits a progressively measurable process $\eta = (\eta_1, \eta_2)$ such that

$$\frac{dQ}{d\mathbb{P}} = \mathcal{E} \left(\int_0^T \eta_{1t} dW_t^1 + \int_0^T \eta_{2t} dW_t^2 \right)_T \quad Q\text{-a.s.},$$

where $\mathcal{E}(M)_t = \exp(M_t - \langle M \rangle_t/2)$ denotes the Doleans–Dade exponential of a continuous semimartingale M . Such a measure Q will receive a penalty

$$\gamma(Q) := E_Q \left[\int_0^T h(\eta_t) dt \right], \quad (3.28)$$

where $h : \mathbb{R}^2 \rightarrow [0, \infty]$ is convex and lower semicontinuous. For simplicity, we suppose that $h(0) = 0$ so that $\gamma(\mathbb{P}) = 0$. We also assume that h is continuously differentiable on its effective domain $\text{dom } h := \{\eta \in \mathbb{R}^2 \mid h(\eta) < \infty\}$ and satisfies the coercivity condition

$$h(x) \geq \kappa_1 |x|^2 - \kappa_2, \quad \text{for some constants } \kappa_1, \kappa_2 > 0. \quad (3.29)$$

Again, our assumption that h does not depend on time is for notational convenience only. Let us also introduce the concave monetary utility functional

$$\phi(X) := \inf_{Q \ll \mathbb{P}} (E_Q[X] + \gamma(Q)), \quad X \in L^\infty.$$

REMARK 3.7. The choice $h(x) = |x|^2/2$ corresponds to the entropic penalty function $\gamma(Q) = \gamma_1^{\text{ent}}(Q) = H(Q|\mathbb{P})$ (see also Section 2.2). Hence, the coercivity condition (3.29) implies that also in the general case, γ can be bounded by the relative entropy $H(\cdot|\mathbb{P})$. This easily yields that ϕ is sensitive in the sense that $\phi(X) > 0$ for any nonzero $X \in L_+^\infty$ because the entropic monetary utility functional (2.6) is obviously sensitive. Moreover, since the level sets $\{dQ/d\mathbb{P} \mid H(Q|\mathbb{P}) \leq c\}$ are weakly compact (this follows, e.g., by combining Theorem 2.2 with the straightforward fact that the entropic monetary utility functional (2.6) is continuous from below), also γ must have weakly relatively compact level sets. In fact, one can show that the level sets of γ are weakly closed (see DELBAEN [2006] for the coherent and HERNÁNDEZ–HERNÁNDEZ and SCHIED [2007a, Lemma 4.1] for the general case) so that γ is the minimal penalty function of ϕ , and ϕ is continuous from below. In particular, ϕ and γ satisfy the assumptions of Sections 3.1 and 3.4. Delbaen recently showed that the coercivity condition (3.29) is not only sufficient but also necessary for ϕ to be continuous from below.

An important particular case occurs if for some compact convex set $\Gamma \subset \mathbb{R}^2$,

$$h(x) = \begin{cases} 0 & \text{if } x \in \Gamma, \\ \infty & \text{if } x \notin \Gamma. \end{cases} \quad (3.30)$$

In this case, ϕ is coherent with maximal representing set

$$\mathcal{Q} := \left\{ Q \sim \mathbb{P} \mid \frac{dQ}{d\mathbb{P}} = \mathcal{E} \left(\int_0^\cdot \eta_{1t} dW_t^1 + \int_0^\cdot \eta_{2t} dW_t^2 \right)_T, \eta = (\eta_1, \eta_2) \in \mathcal{C} \right\}, \quad (3.31)$$

where $\mathcal{C}$ denotes the set of all progressively measurable processes $\eta = (\eta_1, \eta_2)$ such that, $d\mathbb{P}$ -almost everywhere, $\eta_t \in \Gamma$. Note that according to Novikov's theorem, we have a one-to-one correspondence between measures $Q \in \mathcal{Q}$ and processes $\eta \in \mathcal{C}$ (up to $d\mathbb{P}$ -nullsets).

REMARK 3.8. Let us introduce the conditional penalty functions

$$\gamma_t(Q) := E_Q \left[\int_t^T h(\eta_u) du \mid \mathcal{F}_t \right], \quad t \geq 0,$$

and the corresponding family of conditional concave monetary utility functionals,

$$\phi_t(X) := \operatorname{ess\,inf}_{Q \ll \mathbb{P}} (E_Q[X] + \gamma_t(Q)), \quad t \geq 0 \text{ and } X \in L^\infty.$$

This family is dynamically consistent in the sense that

$$\phi_0(\phi_t(X)) = \phi_0(X), \quad \text{for all } X \in L^\infty, \quad (3.32)$$

and this property greatly facilitates the use of our class of concave monetary utility functionals. Indeed, dynamic consistency corresponds to the Bellman principle in dynamic programming and is the essential ingredient for the application of control methods. Recently, the dynamic consistency (3.32) of risk measures has been the subject of ongoing research (see ARTZNER et al. [2007], RIEDEL [2004], CHERIDITO et al. [2004, 2005, 2006], DETLEFSEN and SCANDOLO [2005], FRITTELLI and ROSAZZA GIANIN [2003], WEBER [2006], TUTSCH [2006], FÖLLMER and PENNER [2006]). Note, however, that with the exception of the entropic monetary utility functional, the conditional versions of most of the examples in Section 2.2 are not dynamically consistent (see SCHIED [2007a, section 3] for some examples and discussion).

Let $\mathcal{A}$ denote the set of all progressively measurable process π such that $\int_0^T \pi_s^2 ds < \infty$ $\mathbb{P}$ -a.s. For $\pi \in \mathcal{A}$, we define

$$X_t^{x,\pi} := x \cdot \exp \left(\int_0^t \pi_s \sigma(Y_s) dW_s^1 + \int_0^t \left[\pi_s b(Y_s) - \frac{1}{2} \sigma^2(Y_s) \pi_s^2 \right] ds \right). \quad (3.33)$$

Then, $X^{x,\pi}$ satisfies

$$X_t^{x,\pi} = x + \int_0^t \frac{X_s^{x,\pi} \pi_s}{S_s} dS_s$$

and thus describes the evolution of the wealth process $X^{x,\pi}$ of an investor with initial endowment $x > 0$ investing the fraction π_s of the current wealth into the risky asset at

time $s \in [0, T]$. That is, $X^{x,\pi}$ can be represented as the value process of the admissible strategy $\xi_s = X_s^{x,\pi} \pi_s / S_s$ and hence belongs to the set $\mathcal{X}(x)$. Conversely, any strictly positive process in $\mathcal{X}(x)$ can be described as in (3.34).

The objective of the investor consists in

$$\text{maximizing } \inf_{Q \ll \mathbb{P}} (E_Q[U(X_T^{x,\pi})] + \gamma(Q)) \quad \text{over } \pi \in \mathcal{A}, \quad (3.34)$$

where the utility function $U :]0, \infty[\rightarrow \mathbb{R}$ is henceforth specified as a HARA utility function with constant relative risk aversion $\alpha \in \mathbb{R}$, that is,

$$U(x) = \begin{cases} \frac{x^\alpha}{\alpha} & \text{if } \alpha \neq 0, \\ \log x & \text{if } \alpha = 0. \end{cases} \quad (3.35)$$

Such utility functions are also called constant relative risk aversion (CRRA) utility functions. For $\alpha \neq 0$, we define the conjugate exponent β by

$$\beta := \frac{\alpha}{1 - \alpha}.$$

The following theorem combines the main results of HERNÁNDEZ-HERNÁNDEZ and SCHIED [2006] and SCHIED [2007b] into a single statement. It can be extended to cover also the optimization of consumption-investment strategies (see SCHIED [2007b]) for the case $\alpha > 0$). Recall that $a = \frac{1}{2}(\rho_1^2 + \rho_2^2)$.

THEOREM 3.9 (Coherent case, $\alpha \neq 0$). *Suppose $\alpha \neq 0$ and h is given by (3.30) so that ϕ is coherent with maximal representing set (3.31). Then there exists a unique strictly positive and bounded solution $v \in C^{1,2}([0, T] \times \mathbb{R}) \cap C([0, T] \times \mathbb{R})$ of the quasilinear PDE*

$$\begin{aligned} w_t = & aw_{yy} + (g + \beta\rho_1\theta)w_y + \frac{1}{2}(1 - \alpha\rho_2^2)w_y^2 + \\ & + \inf_{\eta \in \Gamma} \left[(\rho_1(1 + \beta)\eta_1 + \beta\rho_2\eta_2)w_y + \frac{\beta(1 + \beta)}{2}(\eta_1 + \theta)^2 \right] \end{aligned} \quad (3.36)$$

with initial condition

$$w(0, \cdot) \equiv 0, \quad (3.37)$$

and the value function of the robust utility maximization problem (3.34) can then be expressed as

$$u(x) = \sup_{\pi \in \mathcal{A}} \inf_{Q \in \mathcal{Q}} E_Q \left[U(X_T^{x,\pi}) \right] = \frac{x^\alpha}{\alpha} e^{(1-\alpha)w(T, Y_0)}. \quad (3.38)$$

If $\eta^*(t, y)$ is a measurable Γ -valued function that realizes the maximum in (3.36), then an optimal strategy $\hat{\pi} \in \mathcal{A}$ can be obtained by letting $\hat{\pi}_t = \pi^*(T - t, Y_t)$ for

$$\pi^*(t, y) = \frac{1}{\sigma(y)} \left[(1 + \beta)(\eta_1^*(t, y) + \theta(y)) + \rho_1(y) \frac{v_y(t, y)}{v(t, y)} \right].$$

Moreover, by defining a measure $\hat{Q} \in \mathcal{Q}$ via

$$\frac{d\hat{Q}}{d\mathbb{P}} = \mathcal{E}\left(\int_0^T \eta_1^*(T-t, Y_t) dW_t^1 + \int_0^T \eta_2^*(T-t, Y_t) dW_t^2\right)_T,$$

we obtain a saddlepoint $(\hat{\pi}, \hat{Q})$ for the maximin problem (3.34).

Idea of Proof. The theorem was obtained by HERNÁNDEZ-HERNÁNDEZ and SCHIED [2006] for $\alpha < 0$ and deterministic coefficients ρ_1 , and ρ_2 , and by SCHIED [2007b] in the general case with $\alpha > 0$. In both cases, the main idea is to apply stochastic control techniques to the dual rather than the primal problem. First, it follows from Remark 3.7 that the results in Section 3.4 are applicable. Let us denote by $\mathcal{M}$ the set of all progressively measurable processes v such that $\int_0^T v_t^2 dt < \infty$ $\mathbb{P}$ -a.s., and define

$$Z_t^v := \mathcal{E}\left(-\int \theta(Y_s) dW_s^1 - \int v_s dW_s^2\right)_t.$$

Then, Z^v belongs to the space $\mathcal{Y}_{\mathbb{P}}(1)$ as defined in Section 3.4, and the density process of every $P^* \in \mathcal{P}$ is of this form. As before, we denote by $\tilde{U}(z) = \sup_{x \geq 0} (U(x) - zx)$ the convex conjugate function of U . By (3.23), the dual-value function of the robust utility maximization problem is given by

$$\tilde{u}(z) := \inf_{\eta \in \mathcal{C}} \inf_{v \in \mathcal{M}} \mathbb{E}\left[D_T^\eta \tilde{U}\left(\frac{z Z_T^v}{D_T^\eta}\right)\right], \quad (3.39)$$

where $D_t^\eta = \mathcal{E}(\int_0^t \eta_s dW_s)_t$. Due to (3.18), the primal value function u can then be obtained as

$$u(x) = \min_{z > 0} (\tilde{u}(z) + zx). \quad (3.40)$$

Moreover, Theorem 3.8 yields that if $\hat{z} > 0$ minimizes (3.40) and there are control processes $(\hat{\eta}, \hat{v})$ minimizing (3.39) for $z = \hat{z}$, then $X_T^{x, \hat{\pi}} = I(\hat{z} Z_T^{\hat{v}} / D_T^{\hat{\eta}})$ is the terminal wealth of an optimal strategy $\hat{\pi}$. In our specific setting (3.35), we have $\tilde{U}(z) = z^{-\beta} / \beta$. Thus, we can simplify the duality formula (3.40) as follows. First, the expectation in (3.39) equals

$$\mathbb{E}\left[D_T^\eta \tilde{U}\left(\frac{z Z_T^v}{D_T^\eta}\right)\right] = \frac{z^{-\beta}}{\beta} \mathbb{E}[(D_T^\eta)^{1+\beta} (Z_T^v)^{-\beta}] =: \frac{z^{-\beta}}{\beta} \Lambda_{\eta, v}.$$

Optimizing over $z > 0$ then yields that

$$\min_{z > 0} \left(\frac{z^{-\beta}}{\beta} \Lambda_{\eta, v} + zx\right) = \frac{1+\beta}{\beta} x^{\beta/(1+\beta)} \Lambda_{\eta, v}^{1/(1+\beta)} = \frac{x^\alpha}{\alpha} \Lambda_{\eta, v}^{1-\alpha},$$

where the optimal z is given by $\hat{z} = (\Lambda_{\eta, v}/x)^{1-\alpha}$. Using (3.39) and (3.40) yields

$$u(x) = \frac{x^\alpha}{\alpha} \left(\inf_{v \in \mathcal{M}} \inf_{\eta \in \mathcal{C}} \Lambda_{\eta, v}\right)^{1-\alpha}. \quad (3.41)$$

Our next aim is to further simplify $\Lambda_{\eta, \nu}$. To this end, note that

$$\begin{aligned} & (D_T^\eta)^{1+\beta} (Z_T^\nu)^{-\beta} \\ &= \mathcal{E} \left(\int ((1+\beta)\eta_{1s} + \beta\theta(Y_s)) dW_s^1 + \int ((1+\beta)\eta_{2s} + \beta\nu_s) dW_s^2 \right)_T \\ & \quad \times \exp \left(\int_0^T q(Y_s, \eta_s, \nu_s) ds \right), \end{aligned} \quad (3.42)$$

where the function $q : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R} \rightarrow [0, \infty[$ is given by

$$q(y, \eta, \nu) = \frac{\beta(1+\beta)}{2} [(\eta_1 + \theta(y))^2 + (\eta_2 + \nu)^2] + \beta r(y).$$

The Doleans–Dade exponential in (3.42) will be denoted by $\Delta_t^{\eta, \nu}$. If $\int_0^T \nu_t^2 dt$ is bounded, then $\mathbb{E}[\Delta_T^{\eta, \nu}] = 1$. In general, however, we may have $\mathbb{E}[\Delta_T^{\eta, \nu}] < 1$, and this fact creates some technical difficulties. Our aim is to minimize $\Lambda_{\eta, \nu}$ over $\eta \in \mathcal{C}$ and $\nu \in \mathcal{M}_0$. To this end, we introduce the function

$$J(t, y, \eta, \nu) := \mathbb{E} \left[(D_t^\eta)^{1+\beta} (Z_t^\nu)^{-\beta} \right] = \mathbb{E} \left[\Delta_t^{\eta, \nu} \exp \left(\int_0^t q(Y_r, \eta_r, \nu_r) dr \right) \right]$$

so that $J(T, Y_0, \eta, \nu) = \Lambda_{\eta, \nu}$. The minimization of $J(t, y, \eta, \nu)$ is now carried out by stochastic control methods. Let us denote

$$\tilde{g}(y) := g(y) + \beta \rho_1(y) \theta(y).$$

If we have a (sufficiently bounded) classical solution v to the HJB equation

$$\begin{aligned} v_t &= av_{yy} + \tilde{g}(y)v_y + \inf_{\nu \in \mathbb{R}} \inf_{\eta \in \Gamma} \left([\rho_1(1+\beta)\eta_1 + \rho_2((1+\beta)\eta_2 + \beta\nu)]v_y, \right. \\ & \quad \left. + q(\cdot, \eta, \nu)v \right), \end{aligned}$$

$$v(0, y) = 1,$$

then standard verification arguments yield that $v(t, y) = \inf_{\nu \in \mathcal{M}} \inf_{\eta \in \mathcal{C}} J(t, y, \eta, \nu)$. Moreover, $w := \log v$ solves (3.36). It remains to prove existence of classical solutions to the preceding HJB equation. This is carried out by using a priori estimates in conjunction with approximation arguments. The details are beyond the scope of this survey, and we refer to HERNÁNDEZ–HERNÁNDEZ and SCHIED [2006] for the case $\alpha < 0$ and to SCHIED [2007b] for the case $\alpha > 0$. It should be noted that the methods for obtaining classical solutions in these two cases are rather different. $\square$

We now turn to the case of a general penalty function γ given by (3.28). We also specify the risk-aversion parameter α as zero, that is,

$$U(x) = \log x.$$

This choice has the advantage that the portfolio optimization no longer depends on the initial capital x , resulting in a dimension reduction. Our goal is to characterize the value function

$$u(x) = \sup_{\pi \in \mathcal{A}} \inf_{Q \ll \mathbb{P}} (E_Q[\log X_T^{x,\pi}] + \gamma(Q))$$

of the robust utility maximization problem (3.34) in terms of the solution v to the quasilinear parabolic initial value problem

$$\begin{cases} v_t = av_{yy} + \Psi(v_y) + gv_y \\ v(0, \cdot) = 0, \end{cases} \quad (3.43)$$

where the nonlinearity $\Psi(v_y) = \Psi(y, v_y(t, y))$ is given by

$$\Psi(y, z) := \psi(y, (\rho_1(y), \rho_2(y))z), \quad y, z \in \mathbb{R},$$

for the function

$$\psi(y, x) := \inf_{\eta \in \mathbb{R}^2} \left\{ \eta \cdot x + \frac{1}{2}(\eta_1 + \theta(y))^2 + h(\eta) \right\}, \quad y \in \mathbb{R}, x \in \mathbb{R}^2.$$

Here, $\eta \cdot x$ denotes the inner product of η and x . We note that similar results as in Theorems 3.10 and 3.11 hold also for the robust optimization of consumption-investment strategies (see HERNÁNDEZ-HERNÁNDEZ and SCHIED [2007b]).

THEOREM 3.10. *Suppose that $\text{dom } h$ is compact. Then there exists a unique classical solution v to (3.42) within the class of functions in $C^{1,2}([0, T] \times \mathbb{R}) \cap C([0, T] \times \mathbb{R})$ satisfying a polynomial growth condition. The value function u of the robust utility maximization problem is given by*

$$u(x) = \log x + v(T, Y_0).$$

Suppose furthermore that $\eta^* : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function such that $\eta^*(t, y)$ belongs to the supergradient of the concave function $x \mapsto \psi(y, x)$ at $x = (\rho_1(y), \rho_2(y))v_y(t, y)$. Then an optimal strategy $\hat{\pi}$ for the robust problem can be obtained by letting

$$\hat{\pi}_t = \frac{\eta_1^*(T - t, Y_t) + \theta(Y_t)}{\sigma(Y_t)}, \quad 0 \leq t \leq T.$$

Moreover, by defining a measure $\hat{Q} \sim \mathbb{P}$ via

$$\frac{d\hat{Q}}{d\mathbb{P}} = \mathcal{E} \left(\int_0^\cdot \eta^*(T - t, Y_t) dW_t \right)_T, \quad (3.44)$$

we obtain a saddlepoint $(\hat{\pi}, \hat{Q})$ for the maximin problem (3.34).

PROOF. The strategy of the proof is similar to the one of Theorem 3.9. (see HERNÁNDEZ-HERNÁNDEZ and SCHIED [2007a]). $\square$

The problem becomes more difficult when $\text{dom } h$ is noncompact because then we can no longer apply standard theorems on the existence of classical solutions to (3.42). Other problems appear when $\text{dom } h$ is not only noncompact but also unbounded. For instance, we then may have $\gamma(Q) < \infty$ even if Q is not equivalent but merely absolutely continuous with respect to $\mathbb{P}$, and this can lead to difficulties as pointed out in Section 3.4. Moreover, since the optimal η^* takes values in the unbounded set $\text{dom } h$, one needs an additional argument to ensure that the stochastic exponential in (3.43) is a true martingale and so defines a probability measure $\hat{Q} \ll \mathbb{P}$. To deal with this case, we assume for simplicity that ρ_1 and ρ_2 are constant. We also need an additional condition on the shape of the function ψ . Note that g is unbounded if, for example, Y is an Ornstein–Uhlenbeck process.

DEFINITION 3.4. Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be an upper semicontinuous concave function. We say that f satisfies a radial growth condition in direction $x \in \mathbb{R}^2 \setminus \{0\}$ if there exist positive constants p_0 and C such that

$$\max \{ |z| \mid z \in \partial f(px) \} \leq C(1 + |\partial_p^+ f(px)| \vee |\partial_p^- f(px)|) \quad \text{for } p \in \mathbb{R}, |p| \geq p_0,$$

where $\partial f(px)$ denotes the supergradient of f at px and $\partial_p^+ f(px)$ and $\partial_p^- f(px)$ are the right-hand and left-hand derivatives of the concave function $p \mapsto f(px)$.

Note that if f is of the form $f(x) = f_0(|x|)$ for some convex increasing function f_0 , then the radial growth condition is satisfied in any direction $x \neq 0$ with constant $C = 1/|x|$.

THEOREM 3.11. Suppose that ρ_1 and ρ_2 are constants, $|\Psi(y, p)/p| \rightarrow \infty$ as $|p| \rightarrow \infty$, and assume that $\psi(y, \cdot)$ satisfies a radial growth condition in direction (ρ_1, ρ_2) , uniformly in y . Then there exists a unique classical solution v to (3.42) within the class of polynomially growing functions in $C^{1,2}([0, T] \times \mathbb{R}) \cap C([0, T] \times \mathbb{R})$ whose gradient satisfies a growth condition of the form

$$|\partial_p^- \Psi(y; v_y(t, y))| \vee |\partial_p^+ \Psi(y; v_y(t, y))| \leq C_1(1 + |y|)$$

for some constant C_1 . The value function u of the robust utility maximization problem satisfies $u(x) = \log x + v(T, Y_0)$, and also the conclusions on the optimal strategy $\hat{\pi}$ and the measure $\hat{Q}$ in Theorem 3.9 remains true.

PROOF. The proof relies on Theorem 3.10 and PDE arguments (see HERNÁNDEZ-HERNÁNDEZ and SCHIED [2007a]). $\square$

REMARK 3.9. For numerical solutions of the HJB equations in this section, one can use, for example, a multigrid Howard algorithm as explained by AKIAN [1990] and KUSHNER and DUPUIS [2001]. For convergence results of such numerical schemes see KUSHNER

and DUPUIS [2001], KRYLOV [2000], BARLES and JAKOBSEN [2005], and the references therein.

3.6. BSDE approach

In the preceding section, we used stochastic control methods to characterize the solution to our optimization problem in terms of a quasilinear PDE, which then can be solved numerically. Instead of PDE, one can also use backwards stochastic differential equations (BSDEs), and in this section, we discuss some possible approaches. An early result in this direction is due to QUENEZ [2004], where, as in Section 3.5, BSDE techniques are applied to the dual rather than the primal problem. A direct BSDE approach to the primal problem was given by MÜLLER [2005]. Related problems arise in the maximization of recursive utilities in the sense of DUFFIE and EPSTEIN [1992] (see EL KAROUI et al. [2001], LAZRAK and QUENEZ [2003], and the references therein). For the general notion of a BSDE and its applications to finance, we refer to EL KAROUI et al. [1997].

The market model we consider in this section is similar to the ones used at the end of Sections 3.3 and 3.5. It consists of m risky assets $S_t = (S_t^1, \dots, S_t^m)$ that satisfy an SDE of the form

$$dS_t^i = S_t^i \sum_{j=1}^d \sigma_t^{ij} dW_t^j + b_t^i S_t^i dt, \quad i = 1 \dots, m,$$

for a d -dimensional Brownian motion $W = (W^1, \dots, W^d)$, a drift vector process $b = (b^1, \dots, b^m)$, and a volatility matrix process σ . Both b and σ are assumed to be bounded and adapted to the natural filtration $(\mathcal{F}_t)$ of W . In addition, we suppose that $d \geq m$, that σ has full rank $dt \otimes d\mathbb{P}$ -a.e., and that

$$\theta_t := \sigma_t' (\sigma_t \sigma_t')^{-1} b_t$$

is bounded. Here and in the sequel, a' denotes the transpose of a vector or a matrix a . Similarly as in (3.31), model uncertainty is described in terms of the set

$$\mathcal{Q} := \left\{ Q \ll \mathbb{P} \mid \frac{dQ}{d\mathbb{P}} = D_T^\eta, \eta \in \mathcal{C} \right\},$$

where for a predictable family (C_t) of uniformly bounded closed convex subsets of $\mathbb{R}^d$,

$$\mathcal{C} = \left\{ \eta \mid \eta \text{ is predictable and } \eta_t \in C_t \text{ } dt \otimes d\mathbb{P}\text{-a.e.} \right\}$$

and

$$D_t^\eta = \mathcal{E} \left(\int_0^t \eta_s' dW_s \right)_t, \quad 0 \leq t \leq T.$$

The utility function U is assumed to be a logarithmic utility function. To formulate the dual problem, we introduce the set

$$\mathcal{M} := \left\{ v \mid v \text{ is predictable, } \mathbb{R}^d\text{-valued, and } \sigma_t v_t = 0 \text{ } dt \otimes d\mathbb{P}\text{-a.e.} \right\}$$

and the local martingales

$$Z_t^v = \mathcal{E} \left(- \int_0^t (\theta_s + v_s)' dW_s \right), \quad 0 \leq t \leq T, \quad v \in \mathcal{M}.$$

Then Z^v belongs to the space $\mathcal{Y}_{\mathbb{P}}(1)$ as defined in Section 3.4, and the density process of every $P^* \in \mathcal{P}$ is of this form. As before, we denote by $\tilde{U}(z) = \sup_{x \geq 0} (U(x) - zx)$ the convex conjugate function of U . By (3.23), the dual-value function of the robust utility maximization problem is given by

$$\tilde{u}(z) := \inf_{\eta \in \mathcal{C}} \inf_{v \in \mathcal{M}} \mathbb{E} \left[D_T^\eta \tilde{U} \left(\frac{z Z_T^v}{D_T^\eta} \right) \right], \quad (3.45)$$

where $D_t^\eta = \mathcal{E}(\int_0^t \eta_s dW_s)_t$. Due to (3.18), the primal value function u can then be obtained as

$$u(x) = \min_{z > 0} (\tilde{u}(z) + zx). \quad (3.46)$$

Moreover, Theorem 3.8 yields that if $\hat{z} > 0$ minimizes (3.45) and there are control processes $(\hat{\eta}, \hat{v})$ minimizing (3.44) for $z = \hat{z}$, then $X_T^{x, \hat{\pi}} = I(\hat{z} Z_T^{\hat{v}} / D_T^{\hat{\eta}})$ is the terminal wealth of an optimal strategy $\hat{\pi}$. We, therefore, concentrate on solving the dual problem of finding minimizers $(\hat{\eta}, \hat{v})$ in (3.44). The following result is taken from QUENEZ [2004].

THEOREM 3.12. *Suppose $U(x) = \log x$ and, for*

$$f(t, z) := \text{ess inf}_{\eta' \in \mathcal{C}, v \in \mathcal{M}} \left(\eta'_t z + \frac{1}{2} |\theta_t + \eta_t + v_t|^2 \right), \quad z \in \mathbb{R}^d, \quad (3.47)$$

let (Y, Z) be the solution to the BSDE $-dY_t = f(t, Z_t) dt - Z'_t dW_t$ with terminal condition $Y_T = 0$. Then there exists a pair $(\hat{\eta}, \hat{v}) \in \mathcal{C} \times \mathcal{M}$ such that $\hat{v}$ is bounded, $f(t, Z_t) = \hat{\eta}'_t Z_t + \frac{1}{2} |\theta_t + \hat{\eta}_t + \hat{v}_t|^2$, and $(\hat{\eta}, \hat{v})$ solves the dual problem (3.44) for any z .

Sketch of Proof. In the logarithmic case, we have $\tilde{U}(z) = -1 - \log z$ and hence

$$\mathbb{E} \left[D_T^\eta \tilde{U} \left(\frac{z Z_T^v}{D_T^\eta} \right) \right] = -1 - \log z + \mathbb{E} \left[D_T^\eta \log \frac{D_T^\eta}{D_T^v} \right].$$

It is possible to show that the rightmost expectation is equal to

$$\frac{1}{2} E_Q \left[\int_0^T |\theta_s + \eta_s + v_s|^2 ds \right]$$

(see HERNÁNDEZ-HERNÁNDEZ and SCHIED [2007a, lemma 3.4]). Letting

$$J_t^{\eta, v} := \frac{1}{2} \mathbb{E} \left[\int_t^T \frac{D_s^\eta}{D_t^\eta} \cdot |\theta_s + \eta_s + v_s|^2 ds \mid \mathcal{F}_t \right]$$

there exists a square-integrable process $Z^{\eta, \nu}$ such that $(J^{\eta, \nu}, Z^{\eta, \nu})$ solves the BSDE

$$-dJ_t^{\eta, \nu} = (\eta_t' Z_t^{\eta, \nu} + \frac{1}{2} |\theta_t + \eta_t + \nu_t|^2) dt - (Z_t^{\eta, \nu})' dW_t, \quad dJ_T^{\eta, \nu} = 0.$$

Once the existence of $(\hat{\eta}, \hat{\nu})$ as minimizer in (3.46) has been established, the result follows (see QUENEZ [2004, section 7.4] for details).

4. Portfolio choice under robust constraints

The measurement and management of the downside risk of portfolios is a key issue for financial institutions and regulatory authorities. The regulator is concerned with the stability of the financial system and intends to minimize the risk of financial crises by imposing rules on financial institutions. An important regulatory tool are capital constraints, as has often been stressed by regulatory authorities: “Capital regulation is the cornerstone of bank regulators’ efforts to maintain a safe and sound banking system, a critical element of overall financial stability” (BERNANKE [2006]). Capital constraints restrict the risk that banks can take on. The rules specify the amount of capital that banks need to hold to safeguard their solvency and long-run viability.

Regulatory rules for financial institutions have been revised in recent years: new minimum standards for capital adequacy are described in the Basel II framework that national supervisory authorities are currently implementing. This new regulatory framework seeks to improve the previous rules and to provide at the same time a more flexible framework that can better adjust to the evolution of financial markets. The goal of regulation is to maintain overall financial stability. While the aims of the new Basel II framework are well justified, it remains an open question to what extent and in which circumstances the new rules will actually enhance the stability of financial markets. Recent research hints that capital constraints can also lead to adverse effects in certain economic situations (see Section 4.1). While it is an important first step to better understand the impact of the Basel II framework on financial markets, ultimately only the design, evaluation, and implementation of alternative risk measurement techniques and associated capital constraints can lead to better and possibly even optimal regulatory standards.

Regulation requires appropriate ways to measure risk. The properties of the risk measures that are used for this purpose directly influence the impact of regulation on the economic stability of individual banks and the overall financial system. It is, therefore, important to thoroughly understand risk measurement schemes and the corresponding capital constraints. Different methodologies can be applied for this purpose. From a mathematical point of view, risk measures are functionals on spaces of random variables, stochastic processes, or more general measurable functions, which model financial positions. The recently very popular axiomatic approach to risk measurement first specifies desirable features and then characterizes functionals that satisfy these properties (see Section 2.1 for a detailed discussion). The foundation to this systematic investigation of risk measures was provided in the seminal paper by ARTZNER, DELBAEN, EBER and HEATH [1999]. Their work was motivated by the serious deficiencies of the industry

standard Value at Risk (VaR) as a measure of the downside risk. VaR penalizes diversification in many situations and does not take into account the size of very large losses exceeding the value at risk.

While such axiomatic results are an important first step toward better risk management, an analysis of the economic implications of different approaches to risk measurement is indispensable. If risk measures are used as the basis of regulatory capital constraints, they distort the incentives of financial institutions that are subject to regulation. This impact of capital requirements on portfolio holdings cannot be inferred from the axiomatic theory on risk measures. The resulting feedback effects on portfolios, market prices, and volatility need to be taken into account.

The analysis of the virtues and drawbacks of risk measurement schemes requires models in which the investment decisions of financial agents be explicitly modeled. Regulatory authorities force financial institutions to abide by risk constraints whose formulation is based on risk measurement procedures. Financial institutions try to optimize their portfolios under these constraints. Different modeling approaches are available to capture these economic realities. A first approach consists in the analysis of the portfolio optimization problem for a single agent in a financial market in which primary security prices are modeled as exogenous stochastic processes (partial equilibrium). A second approach focuses on market equilibrium models with multiple agents in which prices are formed endogenously under risk constraints (general equilibrium).

Since specific models can only be caricatures of reality, good risk management techniques should work well for a large number of models, that is, they should be robust. This includes general stochastic market state processes and general classes of preferences. In the sections below, we will review relevant contributions to the theory of portfolio choice under risk constraints. This will illustrate that the current understanding of optimal regulation is far from complete. Partial equilibrium models are considered in Section 4.1, and general equilibrium models in Section 4.2.

4.1. Partial equilibrium

The formulation of risk constraints requires the specification of risk measurement procedures. So far, the literature has considered two measurement schemes that were suggested by BASAK and SHAPIRO [2001] and CUOCO, HE and ISAENKO [2007]. A third approach could use dynamic risk measures but has not been investigated in models of portfolio choice so far.

4.1.1. Static risk constraints

The first risk measurement scheme, suggested by BASAK and SHAPIRO [2001], works as follows. Consider a financial institution that intends to maximize its wealth at a finite time horizon T . The institution has to respect its budget constraint. In addition, BASAK and SHAPIRO [2001] assume that final wealth at time T needs to satisfy some risk constraint which can be specified in terms of a risk measure or another ad hoc risk measurement functional.

To be more specific, consider a market over a finite time horizon T that consists of $d + 1$ assets, one bond, and d stocks. We suppose that the bond price is constant. The

price processes of the stocks are given by an $\mathbb{R}^d$ -valued semimartingale S on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, R)$ with $\mathcal{F} = \mathcal{F}_T$ satisfying the usual conditions. An $\mathcal{F}$ -measurable random variable will be interpreted as the value of a financial position at maturity T or, equivalently, as the terminal wealth of an agent. Positions that are R -almost surely equal can be identified.

An investor with initial capital x intends to maximize his/her utility from terminal wealth at time horizon T by choosing an optimal admissible strategy. A trading strategy with initial value x is a d -dimensional predictable, S -integrable process $(\xi_t)_{0 \leq t \leq T}$, which specifies the amount of each asset in the portfolio. In order to exclude doubling strategies, it is usually required for admissible trading strategies that the corresponding value process

$$X_t := x + \int_0^t \xi_s dS_s \quad (0 \leq t \leq T) \quad (4.1)$$

is bounded from below by some constant (which may depend on ξ). $\mathcal{X}(x)$ denotes the set of admissible wealth processes with initial value less than or equal to x . In the absence of a risk constraint, the investor can choose a self-financing admissible trading strategy with corresponding wealth process $X \in \mathcal{X}(x)$ to optimize terminal wealth X_T according to his/her preferences.

Preferences are commonly represented in terms of a utility functional $\mathcal{U}$ (see Section 2.3). Particular examples include expected and robust expected utility. Letting $U : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$ be a utility function, the robust expected utility of wealth X_T at maturity T is given by

$$\mathcal{U}(X_T) = \inf_{Q \in \mathcal{Q}_0} E_Q[U(X_T)], \quad (4.2)$$

where $\mathcal{Q}_0$ is a set of subjective probability measures. If the cardinality of $\mathcal{Q}_0$ is 1, the utility functional reduces to classical expected utility. For a discussion of numerical representations of preference orders, see Section 2.3.

A risk constraint in the sense of BASAK and SHAPIRO [2001] amounts to requiring the agent to satisfy $\rho(X_T) \leq z$ for some risk measurement functional ρ , for example, a risk measure (see Section 2.1) and a threshold level z . The agent's optimization problem is in this case:

$$\begin{aligned} &\text{Maximize } \mathcal{U}(X_T) \text{ over all } X \in \mathcal{X}(x), \\ &\text{which satisfy } \rho(X_T) \leq z. \end{aligned} \quad (4.3)$$

Observe that the risk constraint is imposed at the initial date and not reevaluated later. This is a serious disadvantage of the risk measurement procedure in (4.3). In addition, optimal stochastic dynamic trading strategies and portfolio wealth processes need to be interpreted as commitment solutions, which are specified at the initial date for all future contingencies by the optimizing financial agent.

The partial equilibrium behavior of the single agent problem (4.3) has been discussed on different levels of generality. A general model framework is important to ensure the

robustness of the results. BASAK and SHAPIRO [2001], GABIH, GRECKSCH and WUNDERLICH [2004], and GABIH, GRECKSCH and WUNDERLICH [2005] analyze the economic impact of the risk constraints in a complete financial market, which is driven by Brownian motions. Risk constraints are formulated in terms of VaR and an additional risk functional. Solutions are conjectured by duality considerations, but these articles do not verify that these satisfy the constraints and hence exist.

In contrast to the one-dimensional case involving only a budget constraint, precise conditions for existence constitute the most difficult part of the analysis. This gap in the literature is closed by GUNDEL and WEBER [2008], who, in addition, formulate the risk constraint in terms of convex risk measures and do not stick to a Brownian world. Instead, GUNDEL and WEBER [2008] and GUNDEL and WEBER [2007] provide a complete solution to the problem in a semimartingale setting. GUNDEL and WEBER [2007] investigate the problem of portfolio choice under robust risk constraints in an incomplete market for agents whose preferences can be represented by general robust utility functionals (see Section 2.3). We will, first, review the general results and techniques of GUNDEL and WEBER [2007] and then discuss the economic implications, which are investigated for specific examples by BASAK and SHAPIRO [2001], GABIH, GRECKSCH and WUNDERLICH [2005], GABIH, GRECKSCH and WUNDERLICH [2004] and GUNDEL and WEBER [2008].

GUNDEL and WEBER [2007] focus on the optimization problem (4.3) for a robust utility functional

$$\mathcal{U}(X_T) := \inf_{Q_0 \in \mathcal{Q}_0} E_{Q_0}[U(X_T)]. \quad (4.4)$$

Downside risk is measured by utility-based shortfall risk (UBSR), a convex risk measure in the sense of Definition 2.1, which was already introduced in Section 2.2. Let $\ell : \mathbb{R} \rightarrow [0, \infty]$ be a loss function, that is, an increasing function that is not constant. The level x_1 shall be a point in the interior of the range of ℓ . Let Q_1 be a fixed subjective probability measure equivalent to R , which we will use for risk management. The space of financial positions $\mathcal{D}$ consists of random variables X for which the integral $\int \ell(-X) dQ_1$ is well defined. The UBSR ρ_{Q_1} of a position X is defined by

$$\rho_{Q_1}(X) = \inf \{m \in \mathbb{R} : E_{Q_1}[\ell(-X - m)] \leq x_1\} \quad (4.5)$$

(see also (2.8)). If there is no model uncertainty, the shortfall risk constraint is given by $\rho_{Q_1}(X) \leq 0$. A financial position X that satisfies this constraint is acceptable from the point of view of the risk measure ρ_{Q_1} . This is equivalent to $E_{Q_1}[\ell(-X)] \leq x_1$. In the case of model uncertainty, the probability measure Q_1 is unknown, and one considers a whole set $\mathcal{Q}_1$ of subjective measures, which are equivalent to the reference measure R . The corresponding robust UBSR constraint is given by $\sup_{Q_1 \in \mathcal{Q}_1} \rho_{Q_1}(X) \leq 0$. That is, any financial position must be acceptable from the point of view of all risk measures ρ_{Q_1} ($Q_1 \in \mathcal{Q}_1$). This corresponds to choosing $\rho = \sup_{Q_1 \in \mathcal{Q}_1} \rho_{Q_1}$ and $z = 0$ in problem (4.3) and is equivalent to

$$\sup_{Q_1 \in \mathcal{Q}_1} E_{Q_1}[\ell(-X)] \leq x_1. \quad (4.6)$$

GUNDEL and WEBER [2007] show that the dynamic robust optimization problem (4.3) with robust risk constraint (4.6) can be reduced to a static optimization problem. Letting $\mathcal{P}$ be the set of equivalent martingale measures and

$$\mathcal{I} = \left\{ X \geq 0 : X \in L^1(P) \text{ for all } P \in \mathcal{P} \text{ and } U(X)^- \in L^1(Q_0) \text{ for all } Q_0 \in \mathcal{Q}_0 \right\} \quad (4.7)$$

be the set of terminal financial positions with well-defined utility and prices, the corresponding static problem is given by

$$\begin{aligned} & \text{Maximize } \inf_{Q_0 \in \mathcal{Q}_0} E_{Q_0}[U(X)] \text{ over all } X \in \mathcal{I} \\ & \text{that satisfy } \sup_{Q_1 \in \mathcal{Q}_1} E_{Q_1}[\ell(-X)] \leq x_1 \text{ and } \sup_{P \in \mathcal{P}} E_P[X] \leq x. \end{aligned} \quad (4.8)$$

THEOREM 4.1. *Let S be locally bounded, and assume that the essential domain of the utility function U is bounded from below.*

The optimization problem (4.8) admits a solution if and only if the optimization problem (4.3) with risk constraint (4.6) admits a solution.

If X^* is a solution to problem (4.8), then there exists a solution $(\hat{X}_t) \in \mathcal{X}(x_0)$ to (4.3) with risk constraint (4.6) with $\hat{X}_T \geq X^*$ R -almost surely. In this case, $\hat{X}_T = X^*$ R -almost surely if the solution to (4.8) is R -almost surely unique. If, conversely, $(\hat{X}_t) \in \mathcal{X}(x_0)$ is a solution to (4.3) with risk constraint (4.6), then $\hat{X}_T$ is a solution to (4.8).

Theorem 4.1 reduces the original dynamic problem to the static problem (4.8) and a replication problem. Observe that under the conditions of this theorem, the optimal solution can always be replicated by an admissible trading strategy. GUNDEL and WEBER [2008] and GUNDEL and WEBER [2007] characterize the optimal solution to problem (4.8).

GUNDEL and WEBER [2008] provide the solution to an auxiliary problem without model uncertainty. This provides the basis for the complete solution to problem (4.8) in the general case. Consider first the special case that the set of subjective probability measures $\mathcal{Q}_0$ and $\mathcal{Q}_1$ and the set of martingale measures $\mathcal{P}$ are singletons. Under suitable integrability assumptions, the unique solution to the constrained maximization problem (4.8) can be written in the form

$$x^* \left(\lambda_1^* \frac{dQ_1}{dQ_0}, \lambda_2^* \frac{dP}{dQ_0} \right), \quad (4.9)$$

where $x^* : [0, \infty[\times]0, \infty[\rightarrow \mathbb{R}$ is a continuous deterministic function. λ_1^* and λ_2^* are suitable real parameters, which need to be chosen in such a way that the budget and risk constraint are satisfied. $\frac{dQ_1}{dQ_0}$ and $\frac{dP}{dQ_0}$ signify the Radon–Nikodym densities of Q_1 and P with respect to Q_0 . The function x^* is obtained as the solution to a family of deterministic maximization problems and can explicitly be characterized.

The solution to the auxiliary problem corresponds to a dual problem, which is also key to characterization of the optimal solution in the general case. Consider the function

$$(\lambda_1, \lambda_2) \mapsto \tilde{U}_{\lambda_1, \lambda_2}(P|Q_1|Q_0) = E_R \left[\tilde{U} \left(\lambda_2 \frac{dP}{dR}, \lambda_1 \frac{dQ_1}{dR}, \frac{dQ_0}{dR} \right) \right]$$

with $\tilde{U}(p, q_1, q_0) = \sup_{x \in \mathbb{R}} (q_0 U(x) - q_1 \ell(-x) - xp)$. The parameters $(\lambda_1^*, \lambda_2^*)$ in (4.9) can be identified as the minimizers of the function

$$(\lambda_1, \lambda_2) \mapsto \tilde{U}_{\lambda_1, \lambda_2}(P|Q_1|Q_0) + \lambda_1 x_1 + \lambda_2 x_2.$$

In the general case of an incomplete market and model uncertainty, under technical conditions described by GUNDEL and WEBER [2007], the optimal solution takes the same form as before:

$$X^* := x^* \left(\lambda_1^* \frac{dQ_1^*}{dQ_0^*}, \lambda_2^* \frac{dP^*}{dQ_0^*} \right).$$

However, the subjective probability measures $Q_0^* \in \mathcal{Q}_0$, $Q_1^* \in \mathcal{Q}_1$, the real parameters λ_1^* and λ_2^* , and a finite measure P^* , which is equivalent to the reference measure R , need to be chosen appropriately. It is interesting to observe that the positive measure P^* is not necessarily a probability measure but might have total mass strictly less than 1.

The quantities Q_0^* , Q_1^* , and P^* , and λ_1^* and λ_2^* can be characterized through the dual formulation of the original problem. Letting

$$\tilde{U}_{\lambda_1, \lambda_2}(P|Q_1|Q_0) = E_R \left[\tilde{U} \left(\lambda_2 \frac{dP}{dR}, \lambda_1 \frac{dQ_1}{dR}, \frac{dQ_0}{dR} \right) \right],$$

there exists a minimizer $(\lambda_1^*, \lambda_2^*, Q_0^*, Q_1^*, P^*) \in (\mathbb{R}_+)^2 \times \mathcal{Q}_0 \times \mathcal{Q}_1 \times \mathcal{P}^T$ of

$$\tilde{U}_{\lambda_1, \lambda_2}(P|Q_1|Q_0) + \lambda_1 x_1 + \lambda_2 x_2.$$

In the dual problem, the set of martingale measures $\mathcal{P}$ is replaced with appropriate projections $\mathcal{P}^T$ of extended martingale measures, which are introduced in Remark 4.1. The utility of the optimal claim X^* is given by

$$\inf_{Q_0 \in \mathcal{Q}_0} E_{Q_0}[U(X^*)] = \tilde{U}_{\lambda_1^*, \lambda_2^*}(P^*|Q_1^*|Q_0^*) + \lambda_1^* x_1 + \lambda_2^* x_2.$$

The measures Q_0^* , Q_1^* , and P^* , which are obtained from the solution to the dual problem, can be characterized as worst-case measures. If the expectation of the optimal wealth or claim X^* with respect to a measure $P \in \mathcal{P}^T$ is interpreted as the “ P price” of X^* , then X^* is most expensive under the pricing measure P^* , that is,

$$E_{P^*}[X^*] = \sup_{P \in \mathcal{P}^T} E_P[X^*].$$

At the same time, the subjective probability measures Q_1^* and Q_0^* assign to the optimal claim X^* the highest risk and the lowest von Neumann–Morgenstern utility among all measures in $\mathcal{Q}_1$ and $\mathcal{Q}_0$, respectively:

$$E_{Q_1^*}[\ell(-X^*)] = \sup_{Q_1 \in \mathcal{Q}_1} E_{Q_1}[\ell(-X^*)],$$

$$E_{Q_0^*}[U(X^*)] = \inf_{Q_0 \in \mathcal{Q}_0} E_{Q_0}[U(X^*)].$$

The robust solution X^* turns out to be the classical solution under these worst-case measures.

Observe, however, that P^* is not necessarily a probability measure but could also have mass strictly less than 1. It is, therefore, useful to formulate the solution to the robust utility maximization problem under a joint budget and risk constraint in terms of the dual set of nonnegative supermartingales

$$\mathcal{Y}_R(1) = \{Y \geq 0 : Y_0 = 1, XY \text{ } R\text{-supermartingale} \quad \forall X \in \mathcal{X}(1)\}$$

(see Section 3.4).

REMARK 4.1. FÖLLMER and GUNDEL [2006] show that the elements $Y \in \mathcal{Y}_R(1)$ can be identified with extended martingale measures $\bar{P}^Y$ on the product space $\bar{\Omega} = \Omega \times]0, \infty]$ endowed with the predictable sigma-algebra $\bar{\mathcal{F}}$. More precisely, under suitable regularity assumptions on the underlying filtration, any nonnegative supermartingale Y with $Y_0 = 1$ induces a unique probability measure $\bar{P}^Y$ on $(\bar{\Omega}, \bar{\mathcal{F}})$ such that

$$\bar{P}^Y[A \times]t, \infty] = E_R[Y_t; A] \quad (A \in \mathcal{F}_t, t \geq 0),$$

in analogy to Doob's classical construction of conditional Brownian motions induced by superharmonic functions (cf. FÖLLMER [1973]). The property $Y \in \mathcal{Y}_R(1)$ translates into the condition that the value process (X_t) of any admissible trading strategy, viewed as a process $\bar{X}_t(\omega) = X_t(\omega)\mathbf{1}_{]t, \infty]}(s)$ on the product space, is a supermartingale with respect to $\bar{P}^Y$ and the predictable filtration $(\bar{\mathcal{F}}_t)_{t \geq 0}$ on $(\bar{\Omega}, \bar{\mathcal{F}})$. This condition defines the class of extended martingale measures, introduced by FÖLLMER and GUNDEL [2006].

Let us now discuss economic implications of downside risk constraints. Specific examples suggest that VaR might actually increase extreme risks in comparison to the unconstrained optimal strategy. This has first been pointed out in the seminal paper by BASAK and SHAPIRO [2001]. For a detailed mathematical derivation of the results, the reader is referred to GABIH, GRECKSCH and WUNDERLICH [2005] and GABIH, GRECKSCH and WUNDERLICH [2004]. BASAK and SHAPIRO [2001] consider a model with just one risky asset in a Black–Scholes market, that is, the price S of the single stock is modeled by a geometric Brownian motion. Economic agents solve the maximization problem (4.3) under a VaR constraint for $\rho = \text{VaR}_p$, $p \in]0, 1[$. The utility functional takes the form $\mathcal{U}(X_T) = E_R[U(X_T)]$, where R denotes the statistical measure and $U(x) = \frac{x^\alpha}{\alpha}$, $\alpha < 1$, is a utility function for agents with CRRA. These functions are also called HARA

utility functions, which refers to hyperbolic absolute risk aversion. The case $\alpha = 0$ corresponds to logarithmic utility. Compared with an unconstrained portfolio, a VaR constraint reduces, of course, the overall utility an investor can achieve; positive gains of the optimal claim decrease for good states of the economy. For intermediate states of the economy, a VaR investor behaves like a portfolio insurer to keep the final wealth level above $-z$. However, in those worst states of the world, which occur with probability p , the losses of the VaR investor are larger than for an investor who does not face any constraint. Compared with no constraint, the VaR investor reduces his/her holding of the stock for large stock prices S . However, for small values of S , which correspond to low wealth, the VaR investor adopts a gambling strategy and increases his/her exposure to the risky asset. It has been pointed out by BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] that this behavior resembles strategies of investors who choose their investments according to prospect theory (see KAHNEMAN and TVERSKY [1979] and KAHNEMAN and TVERSKY [1992]). These exhibit risk-averse behavior over gains but are risk seeking over losses.

In contrast to VaR, in the simple Black–Scholes market setting of BASAK and SHAPIRO [2001], alternative risk constraints lead to a significant reduction of the downside risk. This has been verified for UBSR by GUNDEL and WEBER [2008]. Properties of this risk measure are discussed by FÖLLMER and SCHIED [2004], WEBER [2006], DUNKEL and WEBER [2007], and GIESECKE, SCHMIDT and WEBER [2005]. BASAK and SHAPIRO [2001] and GABIH, GRECKSCH and WUNDERLICH [2005] choose $\rho : L^1 \rightarrow \mathbb{R}$, $X \mapsto E_{\tilde{R}}[(X - q)^-]$ to define the risk constraint in (4.3). Here, $q \in \mathbb{R}$ and $\tilde{R}$ is chosen either as the unique equivalent martingale measure (BASAK and SHAPIRO [2001]) or as the statistical measure (GABIH, GRECKSCH and WUNDERLICH [2005]). Observe that ρ is not a cash invariant and thus not a risk measure in the sense of Definition 2.1. But its risk constraint can be reformulated in terms of a UBSR measure, which can be interpreted as a limiting case of GUNDEL and WEBER [2008] (see GABIH, SASS and WUNDERLICH [2007]).

Although the specific examples above already hint at which risk measures can successfully be employed to contain risk, more case studies are necessary to obtain robust characterization results. However, there are more fundamental reasons why one needs to move away from the setup of BASAK and SHAPIRO [2001]. While GUNDEL and WEBER [2007] provide a very general solution to the portfolio optimization problem (4.3) under risk constraints, all five papers, BASAK and SHAPIRO [2001], GABIH, GRECKSCH and WUNDERLICH [2005], GABIH, GRECKSCH and WUNDERLICH [2004], GUNDEL and WEBER [2008], and GUNDEL and WEBER [2007], use the risk measurement scheme (4.3), which is imposed at the initial date and not reevaluated later. These papers might be an important first step in understanding the behavioral impact of regulatory capital requirements. However, they need to be complemented by models that incorporate fully dynamic risk measurement techniques. Risk measurement values should be revised as additional information becomes available.

4.1.2. Semidynamic risk constraints

An alternative risk measurement scheme has been suggested by CUOCO, HE and ISAENKO [2007]. It provides a more realistic and semidynamic model of risk constraints. The

scope of the original paper by CUOCO, HE and ISAENKO [2007] is limited. It investigates a complete financial market whose primary security price processes follow a geometric Brownian motion and focuses on only a few risk constraint specifications. However, the basic modeling idea, which resembles current industry practice in the special case of VaR, can be extended. In combination with results from the axiomatic theory of risk measures, the approach of CUOCO, HE and ISAENKO [2007] has significant potential as a starting point for future research in general market settings. Since CUOCO, HE and ISAENKO [2007] focus only on the simplest special cases, we give here a stylized description generalizing their approach.

At each point in time t , investors assess their risk on the basis of all available information. Risk is measured for the time window $[t, t + \tau]$ with $\tau > 0$ using a distribution-invariant static risk measure ρ (or other risk measurement functional). The risk measure is applied to the conditional distribution of projected changes in wealth.

In this context, projected wealth is an auxiliary quantity in the risk measurement procedure. Given a portfolio strategy at time t of the investor, wealth is projected to time $t + \tau$ under the counter to fact assumption that the proportion of wealth invested in each asset in the portfolio (relative exposure) and the market coefficients do not change in the time interval $[t, t + \tau]$. The dynamic risk measurement at time t is obtained by applying the static risk measure ρ to the conditional distribution of the projected change in wealth.

Let us emphasize that this quantity does not represent the risk of the true change of wealth over the time window from t to $t + \tau$ in terms of the risk measure ρ . First, market coefficients change over time. Second, investors are allowed to modify their trading strategies continuously. The dynamic risk measurement procedure is rather a scheme that is easily implementable and, at the same time, sensitive to new information.

Consider, for example, a financial market with d primary assets $S^1, \dots, S^d$, which are modeled by a d -dimensional Itô process and a money market account S^0 with constant interest rate r :

$$\begin{aligned} dS_t^0 &= S_t^0 r dt \\ dS_t^i &= S_t^i \left(\mu_t^i dt + \sum_{j=1}^m \sigma_t^{ij} dW_t^j \right), \quad i = 1, 2, \dots, d \end{aligned}$$

with mean rate of return process μ and variance–covariance process σ . Letting $\pi = (\pi_t)_{t \in [0, \infty)}$ be the fraction of current wealth X_t^π invested in each of the d assets, $t \in [0, \infty]$, the SDE of the wealth process is given by

$$dX_t^\pi = X_t^\pi \left((r + \pi_t^* \mu_t) dt + \pi_t^* \sigma_t dW_t \right),$$

where v^* denotes the transpose of a vector $v \in \mathbb{R}^d$.

The fictitious projected change in wealth at time t for the time interval $[t, t + \tau]$ is given by

$$P_t^\pi = X_t^\pi \cdot \exp \left((r + \pi_t^* \mu_t - \frac{1}{2} |\pi_t^* \sigma_t|^2) \tau + \pi_t^* \sigma_t (W_{t+\tau} - W_t) \right) - X_t^\pi. \quad (4.10)$$

The risk measurement at time t is obtained by applying ρ to the conditional distribution $\mathcal{L}(P_t^\pi | \mathcal{F}_t)$ of P_t^π given the information $\mathcal{F}_t$ at time t .

The risk constraint is now specified as follows. A trading strategy is feasible at time t if the risk of the projected change of wealth (4.10) measured by the risk measure ρ does not exceed a fixed threshold level. The objective of the financial investor is to invest optimally according to some criterion while at the same time satisfying the risk constraint.

There are certain variants of the latter model that focus on relative quantities instead of absolute quantities. Alternatively, when projected wealth changes are calculated, one could assume that instead of wealth proportions the number of shares is fictitiously held constant or that market coefficients are not fixed but vary stochastically. In any case, given such a model, the optimal trading strategy and wealth process need to be characterized and the impact on the downside risk needs to be evaluated.

CUOCO, HE and ISAENKO [2007] investigate a complete market model where asset price processes follow geometric Brownian motions. The objective of the investors is to maximize the von Neumann–Morgenstern utility of terminal wealth in the finite time-horizon economy. Absolute and relative risk constraints are specified in terms of VaR and average value at risk (AVaR). CUOCO, HE and ISAENKO [2007] characterize the optimal trading strategy and terminal wealth in terms of a Hamilton–Jacobi–Bellman equation. The optimal trading strategy is a multiple of the classical Merton proportion (the unconstrained optimal strategy) with a factor of at most 1. The equivalence of VaR and AVaR is demonstrated, and numerical case studies for CRRA/HARA utility illustrate the model. CUOCO, HE and ISAENKO [2007] claim that a dynamic version of VaR can successfully be used for regulation in a market driven by a multidimensional geometric Brownian motion. Similar results have also been obtained by PIRVU and ZITKOVIC [2007] who investigate growth-optimal investment in a market driven by Itô processes under dynamic risk constraints when projected wealth is calculated under the assumption of fixed market coefficients. However, it remains open whether these findings are robust. In alternative or more general settings, different risk measures might be appropriate, but this issue requires substantial further investigation.

4.1.3. Further contributions

GUNDY [2005] investigates the problem (4.3) under risk constraints that are specified in terms of VaR, expected shortfall, and AVaR. Under certain conditions, the dynamic problem corresponds to a static utility maximization under risk constraints. GUNDY [2005] characterizes the existence, uniqueness, and structure of the solutions in the static case. The dynamic problem is studied for a complete financial market that is driven by Brownian motion. EMMER, KORN and KLÜPPELBERG [2001] investigate the optimal portfolio problem (4.3) in a complete multidimensional Black–Scholes market under a capital at risk constraint. The capital at risk at level $p \in]0, 1[$ of a random variable is the difference between the mean and the VaR at level p . EMMER, KORN and KLÜPPELBERG [2001] solve the optimization problem under the strong assumption that the fraction of wealth invested in each asset is held constant over time. KLÜPPELBERG and PERGAMENCHTCHIKOV [2007] investigate optimal utility of consumption and terminal wealth

for investors with power utility functions under downside risk constraints in a generalized complete multidimensional Black–Scholes market where the interest rate, the mean rate of return process, and the variance–covariance process must be deterministic but may be time dependent. Downside risk constraints are uniform versions of VaR and AVaR constraints. As in BASAK and SHAPIRO [2001], these are imposed at time 0 and not reevaluated later. BOYLE and TIAN [2007], GABIH, GRECKSCH, RICHTER and WUNDERLICH [2006], and BASAK, SHAPIRO and TEPLA [2006] solve versions of problem (4.3) if investors compare their performances to a random benchmark at the time horizon T . GABIH, GRECKSCH, RICHTER and WUNDERLICH [2006] focus on a Black–Scholes market with limits on the expected utility loss and derive explicit results. Generalizing VaR, BOYLE and TIAN [2007] impose limits on the probability that terminal wealth lies below the benchmark. For a complete market driven by Brownian motion, the existence and structure of the solution are characterized, and special cases are discussed explicitly. For a Black–Scholes market, the portfolio optimization problem of an investor with CRRA/HARA utility is considered by BASAK, SHAPIRO and TEPLA [2006]. Economic implications for special cases are discussed in detail. For contributions to strict portfolio insurance, we refer to BRENNAN and SCHWARTZ [1989], BASAK [1995], GROSSMAN and ZHOU [1996], JENSEN and SORESENSEN [2001], and LAKNER and NYGREN [2006].

CUOCO and LIU [2006] emphasize that the actual values of risk measures cannot be observed by regulators. Instead, the Basel Committee’s Internal Model Approach (IMA) requires financial institutions to self-report VaR measurements. Capital constraints are based on these self-reported numbers. The IMA mechanism creates an adverse selection problem since banks have an incentive to underreport the true VaR to reduce capital constraints. The Basel Committee suggested to address this problem by “backtesting”: regulators should record actual profit and loss distributions and evaluate the frequency of exceptions, which exceed the reported VaR; banks should be penalized if inconsistencies are observed. CUOCO and LIU [2006] provide a model for IMA and investigate the optimal reporting and portfolio selection problem in a complete, multidimensional Black–Scholes market. The optimal trading strategy can be recovered from the dual-value function, which is characterized in terms of a Hamilton–Jacobi–Bellman equation. Based on numerical case studies, CUOCO and LIU [2006] claim that IMA effectively bounds portfolio risk and induces risk revelation in their model framework.

4.2. General equilibrium

Single-agent models (partial equilibrium) specify prices exogenously and constitute one possible approach to analyze the impact of downside risk constraints on the behavior of economic agents and to assess the virtues of risk measures; another approach are market equilibrium models with multiple agents in which prices are formed endogenously under risk constraints (general equilibrium). General equilibrium models provide a framework to study feedback effects of regulation on prices, which are neglected in the partial equilibrium case.

4.2.1. Static risk constraints

So far, the literature on general equilibrium models that incorporate risk constraints is very limited, and only special cases have been studied. BASAK and SHAPIRO [2001]

provide a first characterization of general equilibrium effects in their risk management setting for agents with intertemporal consumption and logarithmic utility for the case that instantaneous aggregate consumption follows a geometric Brownian motion. BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] base their analysis on BASAK and SHAPIRO [2001] and LUCAS [1978] and provide a more detailed analysis in a model with economic agents with constant relative risk aversion.

In their model, agents maximize the utility of consumption and terminal wealth over a finite time horizon T . The risk constraint is imposed at time 0 on terminal wealth at time T . The total consumption rate in the economy equals an exogenous dividend rate process that is modeled as a geometric Brownian motion. The equilibrium price and consumption processes are derived for an economy with two types of traders: unregulated and VaR-constrained traders.

To be more specific, BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] consider a pure exchange economy in a finite horizon $[0, T]$ with agents with CRRA/HARA utility. The utility functions of all agents are assumed to be identical. Agents consume a single perishable consumption good. The aggregate endowment of the economy with this good is modeled by a geometric Brownian motion:

$$d\delta_t = \mu_\delta \delta_t dt + \sigma_\delta \delta_t dB_t,$$

where μ_δ and σ_δ are constant drift and volatility coefficients and B is Brownian motion. The information is modeled by the augmented Brownian filtration generated by B . All processes are assumed to be adapted. Two financial assets are traded in the financial market, a money market account with price process β that is in zero supply and a stock with price S that is in constant net supply of 1. These processes follow the SDE:

$$d\beta_t = r_t \beta_t dt,$$

$$d(S_t + \delta_t) = S_t (\mu_t dt + \sigma_t dB_t),$$

where the interest rate process r , the drift process μ , and the volatility process σ are not exogenously given but determined in equilibrium. The dividends δ of the stock correspond to the perishable consumption good.

For agents $i = 1, 2$, let $H^i, U^i : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\}$ denote appropriate utility functions. At time t , agents of type i hold ξ_t^i stocks and ψ_t^i bonds such that their wealth equals $W_t^i = \xi_t^i S_t + \psi_t^i \beta_t$. Their consumption rate process is denoted by $(c_t^i)_{t \in [0, T]}$. All agents are assumed to be small and act as price takers.

Unregulated agents solve the standard optimization problem under a budget constraint, that is,

$$\begin{aligned} \max_{c^i, \xi^i, \psi^i} \quad & E \left[\int_0^T U^i(c_s^i) ds + \rho^i H^i(W_T^i) \right] \\ \text{s.t.} \quad & W_0^i = w^i, \\ & dW_t^i = \xi_t^i d(S_t + \delta_t) + \psi_t^i d\beta_t - c_t^i dt, \\ & W_t^i \geq 0, \quad \text{for } \forall t \in [0, T], \end{aligned}$$

where $i = 1$ denotes the type of the agents, $\rho^i > 0$ is the weight of the relative importance of consumption and final wealth at time T in the utility functional, and $w^i > 0$ denotes initial wealth of agents of type i .

Regulated agents solve the same problem for $i = 2$ under an additional VaR constraint at level p with threshold q , that is,

$$P[W_T^2 \geq q] \geq 1 - p.$$

In order to determine the price processes in equilibrium, the following conditions are imposed on the optimal solutions, $\hat{c}^i$, $\hat{\xi}^i$, $\hat{\psi}^i$, $i = 1, 2$, of the utility maximization problems:

(i) *Clearing of the commodity market:*

$$\hat{c}_t^1 + \hat{c}_t^2 = \delta_t, \quad 0 \leq t \leq T.$$

(ii) *Clearing of the stock market:*

$$\hat{\xi}_t^1 + \hat{\xi}_t^2 = 1, \quad 0 \leq t \leq T.$$

(iii) *Clearing of the money market:*

$$\hat{\psi}_t^1 + \hat{\psi}_t^2 = 0, \quad 0 \leq t \leq T.$$

For an introduction to the equilibrium problem for small investors in financial markets and mathematical solution techniques, we refer to chapter 4 in KARATZAS and SHREVE [1998].

BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] find that the results of BASAK and SHAPIRO [2001] derived in partial equilibrium still hold in general equilibrium. In addition, the presence of VaR risk managers typically reduces stock volatility in general equilibrium but may increase it in bad states of the economy, that is for high values of the state price density. In some cases, it can also increase the probability of extremely negative returns. BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] conclude that VaR risk management has a stabilizing effect on the economy for normal and good states. It might, however, worsen catastrophic states that occur with small probability since VaR managers adopt gambling strategies and increase their stock holdings in these circumstances.

While BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] provide many interesting insights for risk management in a general equilibrium framework, they restrict attention to VaR constraints, CRRA utility and dividend rates that follow a geometric Brownian motion. At the same time, BERKELAAR, CUMPERAYOT and KOUWENBERG [2002] stick to the risk measurement setup (4.3) of BASAK and SHAPIRO [2001], GABIH, GRECKSCH and WUNDERLICH [2005], GABIH, GRECKSCH and WUNDERLICH [2004], GUNDEL and WEBER [2008], and GUNDEL and WEBER [2007] in which the risk constraint on terminal wealth is imposed at time 0 and not reevaluated later. Future research needs to incorporate general dynamic risk measure constraints, utility functionals, and dividend rate processes.

4.2.2. *Semidynamic risk constraints*

LEIPPOLD, TROJANI and VANINI [2006] investigate a general equilibrium model similar to BERKELAAR, CUMPERAYOT and KOUWENBERG [2002]. In contrast to the latter paper, they impose dynamic wealth-dependent VaR limits, which are similar to those in CUOCO, HE and ISAENKO [2007]. Instantaneous aggregate consumption does not necessarily follow a geometric Brownian motion but is driven by a stochastic factor process; the risk aversion of agents is heterogeneous. When analyzing the model, LEIPPOLD, TROJANI and VANINI [2006] use a perturbation approximation.

Their analysis suggests that VaR constraints have ambiguous effects on equity volatility and equity expected returns. The consequences of VaR regulation on economic variables are hardly predictable. Their paper and the literature review above demonstrate that the design of robust regulatory standards with an unambiguous and desirable impact across a large number of economic models is an important open problem; the current regulatory standard VaR seems deficient in many respects.

4.2.3. *Further contributions*

General equilibrium models of portfolio insurance are provided by BRENNAN and SCHWARTZ [1989], BASAK [1995], GROSSMAN and ZHOU [1996], and VANDEN [2006]. BARRIEU and EL KAROUI [2005] investigate optimal risk transfer and the design of financial instruments aimed to hedge risk which is not traded on financial markets. The issuer minimizes a risk measure under the constraint imposed by the buyer who enters the transaction only if his/her risk level remains below a given threshold. The problem is reduced to an inf-convolution problem involving a transformation of the risk measure.

References

- AKIAN, M. (1990). Méthodes multigrilles en contrôle stochastique. Thesis, Université de Paris IX (Paris-Dauphine), Paris, 1990 (Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, France).
- ANSCOMBE, F.J., AUMANN, R.J. (1963). A definition of subjective probability. *Ann. Math. Stat.* **34**, 199–205.
- ARTZNER, P., DELBAEN, F., EBER, J.-M., HEATH, D. (1999). Coherent measures of risk. *Math. Financ.* **9** (3), 203–228.
- ARTZNER, P., DELBAEN, F., EBER, J.-M., HEATH, D., KU, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Ann. Oper. Res.* **152**, 5–22.
- BARILLAS, F., HANSEN, L., SARGENT, T. (2007). Doubts or variability? Working paper, University of Chicago and New York University.
- BARLES, G., JAKOBSEN, E. (2005). Error bounds for monotone approximation schemes for parabolic Hamilton–Jacobi–Bellman equations. *SIAM J. Numer. Anal.* **43** (2), 540–558.
- BARRIEU, P., EL KAROUI, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Financ. Stoch.* **9** (2), 269–298.
- BASAK, S. (1995). A general equilibrium model of portfolio insurance. *Rev. Financ. Stud.* **8** (4), 1059–1090.
- BASAK, S., SHAPIRO, A. (2001). Value-at-risk based risk management: optimal policies and asset prices. *Rev. Financ. Stud.* **14**, 371–405.
- BASAK, S., SHAPIRO, A., TEPLA, L. (2006). Risk management with benchmarking. *Manage. Sci.* **52** (4), 542–557.
- BAUDOIN, F. (2002). Conditioned stochastic differential equations: theory, examples and application to finance. *Stoch. Proc. Appl.* **100**, 109–145.
- BEDNARSKI, T. (1981). On solutions of minimax test problems for special capacities. *Z. Wahrsch. Verw. Gebiete* **58**, 397–405.
- BEDNARSKI, T. (1982). Binary experiments, minimax tests and 2-alternating capacities. *Ann. Stat.* **10**, 226–232.
- BENSOUSSAN, A. (1984). On the theory of option pricing. *Acta Appl. Math.* **2** (2), 139–158.
- BEN-TAL, A., TEBOULLE, M. (1987). Penalty functions and duality in stochastic programming via ϕ -divergence functionals. *Math. Oper. Res.* **12**, 224–240.
- BEN-TAL, A., TEBOULLE, M. (2007). An old-new concept of convex risk measures: the optimized certainty equivalent. *Math. Financ.* **17** (3), 449–476.
- BERKELAAR, A., CUMPERAYOT, P., KOUWENBERG, R. (2002). The effect of VaR-based risk management on asset prices and volatility smile. *Eur. Financ. Manage.* **8** (2), 139–164.
- BERNANKE, B.S. (2006). Banking regulation and supervision: balancing benefits and costs. Remarks before the Annual Convention of the American Bankers Association, Phoenix, AZ.
- BORDIGONI, G., MATOUSSI, A., SCHWEIZER, M. (2005). A stochastic control approach to a robust utility maximization problem. To appear in Proceedings of Abel Symposium 2005, Springer.
- BOYLE, P., TIAN, W. (2007). Portfolio management with constraints. *Math. Financ.* **17** (3), 319–343.
- BRENNAN, M.J., SCHWARTZ, E.S. (1989). Portfolio insurance and financial market equilibrium. *J. Bus.* **62** (4), 455–472.
- BURGERT, C., RÜSCHENDORF, L. (2005). Optimal consumption strategies under model uncertainty. *Stat. Decis.* **23** (1), 1–14.
- CARLIER, G., DANA, R.A. (2003). Core of convex distortions of a probability. *J. Econom. Theory* **113** (2), 199–222.

- CARR, P., GEMAN, H., MADAN, D. (2001). Pricing and hedging in incomplete markets. *J. Financ. Econom.* **62**, 131–167.
- CASTAÑEDA-LEYVA, N., HERNÁNDEZ-HERNÁNDEZ, D. (2005). Optimal consumption-investment problems in incomplete markets with stochastic coefficients. *SIAM J. Control Optim.* **44** (4), 1322–1344.
- CHERIDITO, P., DELBAEN, F., KUPPER, M. (2004). Coherent and convex monetary risk measures for bounded càdlàg processes. *Stoch. Proc. Appl.* **112**, 1–22.
- CHERIDITO, P., DELBAEN, F., KUPPER, M. (2005). Coherent and convex monetary risk measures for unbounded càdlàg processes. *Financ. Stoch.* **9**, 1713–1732.
- CHERIDITO, P., DELBAEN, F., KUPPER, M. (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electron. J. Probab.* **11**, 57–106.
- CHERNY, A. (2006). Weighted VaR and its properties. *Financ. Stoch.* **10** (3), 367–393.
- CHERNY, A. (2007a). Equilibrium with coherent risk. *Theory Probab. Appl.* **52** (4), 34.
- CHERNY, A. (2007b). Pricing and hedging European options with discrete-time coherent risk. *Financ. Stoch.* **11**, (4), 537–569.
- CHERNY, A., GRIGORIEV, P. (2007). Dilatation monotone risk measures are law invariant. *Financ. Stoch.* **11** (2), 291–298.
- CHERNY, A., KUPPER, M. (2007). Divergence utilities Preprint (Moscow State University, Moscow, Russia).
- CHOQUET, G. (1953) Theory of capacities. *Ann. Inst. Fourier* **5**, 131–295.
- CONT, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Math. Financ.* **16**, 519–542.
- CSISZAR, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffischen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8**, 85–108.
- CSISZAR, I. (1967). On topological properties of f -divergences. *Studia. Sci. Math. Hungarica* **2**, 329–339.
- CUOCO, D., HE, H., ISAENKO, S. (2007). Optimal dynamic trading strategies with risk limits. *Oper. Res.* To appear.
- CUOCO, D., LIU, H. (2006). An analysis of VaR-based capital requirements. *J. Financ. Intermed.* **15**, 362–394.
- CVITANIC, J., KARATZAS, I. (2001). Generalized Neyman-Pearson lemma via convex duality. *Bernoulli* **7**, 79–97.
- DANA, R.-A. (2005). A representation result for concave Schur concave functions. *Math. Financ.* **15**, 613–634.
- DELBAEN, F. (2000). *Coherent Risk Measures, Cattedra Galileiana* (Scuola Normale Superiore, Classe di Scienze, Pisa, Italy).
- DELBAEN, F. (2002). Coherent measures of risk on general probability spaces. In: *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann* (Springer-Verlag), pp. 1–37.
- DELBAEN, F. (2006). The structure of m -stable sets and in particular of the set of riskneutral measures. In: Yor, M., Émery, M. (eds.), *In Memoriam Paul-André Meyer - Séminaire de Probabilités XXXIX* (Springer, Berlin, Germany, Heidelberg, Germany, New York, NY), pp. 215–258.
- DELBAEN, F., SCHACHERMAYER, W. (2006). *The Mathematics of Arbitrage, Springer Finance* (Springer-Verlag, Berlin, Germany).
- DENIS, L., MARTINI, C. (2006). A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *Ann. Appl. Probab.* **16** (2), 827–852.
- DENNEBERG, D. (1994). *Non-Additive Measure and Integral*. Theory decision library series B: mathematical and statistical methods Volume 27. (Kluwer Academic Publishers, Dordrecht, Netherlands).
- DETLEFSEN, K., SCANDOLO, G. (2005). Conditional and dynamic convex risk measures. *Financ. Stoch.* **9** (4), 539–561.
- DOKUCHAEV, N. (2007). Maximin investment problems for discounted and total wealth. To appear in IMA Journal of Management Mathematics.
- DUFFIE, D., EPSTEIN, L. (1992). Stochastic differential utility. With an appendix by the authors and C. Skiadas. *Econometrica* **60** (2), 353–394.
- DUNKEL, J., WEBER, S. (2007). Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models, *Proceedings of the 2007 Winter Simulation Conference*, pp. 958–966, 2007.
- EICHHORN, A., RÖMISCH, W. (2005). Polyhedral risk measures in stochastic programming. *SIAM J. Optim.* **16** (1), 69–95.

- EL KAROUI, N., JEANBLANC-PICQUÉ, M., SHREVE, S. (1998). Robustness of the black and scholes formula. *Math. Financ.* **8**, 93–126.
- EL KAROUI, N., PENG, S., QUENEZ, M.C. (1997). Backward stochastic differential equations in finance. *Math. Financ.* **7** (1), 1–71.
- EL KAROUI, N., PENG, S., QUENEZ, M.C. (2001). A dynamic maximum principle for the optimization of recursive utilities under constraints. *Ann. Appl. Probab.* **11** (3), 664–693.
- EMMER, S., KORN, R., KLÜPPELBERG, C. (2001). Optimal portfolios with bounded capital at risk. *Math. Financ.* **11** (4), 365–384.
- FAVERO, G. (2001). Shortfall risk minimization under model uncertainty in the binomial case: adaptive and robust approaches. *Math. Methods Oper. Res.* **53** (3), 493–503.
- FAVERO, G., RUNGGLADIER, W. (2002). A robustness result for stochastic control. *Syst. Control Lett.* **46** (2), 91–97.
- FLEMING, W., HERNÁNDEZ-HERNÁNDEZ, D. (2003). An optimal consumption model with stochastic volatility. *Financ. Stoch.* **7** (2), 245–262.
- FLEMING, W., SONER, M. (1993). *Controlled Markov Processes and Viscosity Solutions* (Springer-Verlag, New York, NY).
- FLEMING, W.H., SHEU, S.J. (2000). Risk-sensitive control and an optimal investment model. INFORMS applied probability conference (Ulm, 1999). *Math. Financ.* **10** (2), 197–213.
- FLEMING, W.H., SHEU, S.J. (2002). Risk-sensitive control and an optimal investment model, II. *Ann. Appl. Probab.* **12** (2), 730–767.
- FÖLLMER, H. (2001). Probabilistic Aspects of Financial Risk. Plenary Lecture at the Third European Congress of Mathematics. In: *Proceedings of the European Congress of Mathematics, Barcelona 2000* (Birkhäuser, Basel, Switzerland).
- FÖLLMER, H., LEUKERT, P. (2000). Efficient hedging: cost versus shortfall risk. *Financ. Stoch.* **4**, 117–146.
- FÖLLMER, H., PENNER, I. (2006). Convex risk measures and the dynamics of their penalty functions. *Stat. Decis.* **24** (1), 61–96.
- FÖLLMER, H., SCHIED, A. (2002a). Convex measures of risk and trading constraints. *Financ. Stoch.* **6**, 429–447.
- FÖLLMER, H., SCHIED, A. (2002b). *Robust Preferences and Convex Measures of Risk*. Advances in Finance and Stochastics (Springer, Berlin, Germany).
- FÖLLMER, H., SCHIED, A. (2004). *Stochastic Finance: An Introduction in Discrete Time, 2nd Revised and Extended Edition* (Walter de Gruyter & Co., Berlin, Germany), de Gruyter Studies in Mathematics 27, 2004.
- FÖLLMER, H. (1973). On the representation of semimartingales. *Ann. Probab.* **1** (4), 580–589.
- FÖLLMER, H., GUNDEL, A. (2006). Robust projections in the class of martingale measures. *Illinois J. Math.* **50** (2), 439–472.
- FOUQUE, J.-P., PAPANICOLAOU, G., SIRCAR, K.R. (2000). *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge University Press, Cambridge, MA).
- FRITTELLI, M., ROSAZZA GIANIN, E. (2002). Putting order in risk measures. *J. Bank. Financ.* **26**, 1473–1486.
- FRITTELLI, M., ROSAZZA GIANIN, E. (2003). Dynamic convex risk measures. In: Szegö, G. (ed.), *New Risk Measures in Investment and Regulation* (John Wiley & Sons, New York, NY).
- FRITTELLI, M., ROSAZZA GIANIN, E. (2005). Law-invariant convex risk measures. *Adv. Math. Econ.* **7**, 33–46.
- GABIH, A., GRECKSCH, W., RICHTER, M., WUNDERLICH, R. (2006). Optimal portfolio strategies benchmarking the stock market. *Math. Method. Oper. Res.* **64**, 211–225.
- GABIH, A., GRECKSCH, W., WUNDERLICH, R. (2004). Optimal portfolios with bounded shortfall risks. In: 'Tagungsband zum Workshop Stochastic Analysis' (TU Chemnitz, Chemnitz, Germany), pp. 21–41. (Available at: <http://archiv.tu-chemnitz.de/pub/2004/0120>).
- GABIH, A., GRECKSCH, W., WUNDERLICH, R. (2005). Dynamic portfolio optimization with bounded shortfall risks. *Stoch. Anal. Appl.* **3** (23), 579–594.
- GABIH, A., SASS, J., WUNDERLICH, R. (2007). Utility maximization under bounded expected loss, RICAM report. (Available at: <http://www.ricam.oeaw.ac.at/publications/reports/06/rep06-24.pdf>).
- GIESECKE, K., SCHMIDT, T., and WEBER, S. (2005). Measuring the risk of large losses. *Journal of Investment Management*, **6**(4), 2008.

- GILBOA, I., SCHMEIDLER, D. (1989). Maxmin expected utility with non-unique prior. *J. Math. Econ.* **18**, 141–153.
- GROSSMAN, S.J., ZHOU, Z. (1996). Equilibrium analysis of portfolio insurance. *J. Financ.* **51** (4), 1379–1403.
- GUNDEL, A. (2005). Robust utility maximization in complete and incomplete market models. *Financ. Stoch.* **9** (2), 151–176.
- GUNDEL, A. (2006). Robust utility maximization, f -projections, and risk constraints, Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin, Germany.
- GUNDEL, A., WEBER, S. (2007). Robust utility maximization with limited downside risk in incomplete markets. *Stoch. Proc. Appl.* **117** (11), 1663–1688.
- GUNDEL, A., WEBER, S. (2008). Utility maximization under a shortfall risk constraint. To appear in *Journal of Mathematical Economics*.
- GUNDY, R. (2005). Portfolio optimization with risk constraints, PhD thesis (Universität Ulm, Ulm, Germany) Available at: <http://vts.uni-ulm.de/doc.asp?id=5427>.
- HANSEN, L., SARGENT, T. (2001). Robust control and model uncertainty. *Am. Econ. Rev.* **91**, 60–66.
- HEATH, D. (2000). Back to the Future. Plenary lecture. In: *First World Congress of the Bachelier Finance Society, Paris, France*.
- HEATH, D., KU, H. (2004). Pareto equilibria with coherent measures of risk. *Math. Financ.* **14** (2), 163–172.
- HERNÁNDEZ-HERNÁNDEZ, D., SCHIED, A. (2006). Robust utility maximization in a stochastic factor model. *Stat. Decis.* **24** (3), 109–125.
- HERNÁNDEZ-HERNÁNDEZ, D., SCHIED, A. (2007a). A control approach to robust utility maximization with logarithmic utility and time-consistent penalties. *Stoch. Proc. Appl.* **117** (8), 980–1000.
- HERNÁNDEZ-HERNÁNDEZ, D., SCHIED, A. (2007b). Robust maximization of consumption with logarithmic utility. In: *Proceedings of the 2007 American Control Conference* pp. 1120–1123.
- HERSTEIN, I., MILNOR, J. (1953). An axiomatic approach to measurable utility. *Econometrica* **21**, 291–297.
- HU, Y., IMKELLER, P., MÜLLER, M. (2005). Utility maximization in incomplete markets. *Ann. Appl. Probab.* **15** (3), 1691–1712.
- HUBER, P. (1981). *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics (Wiley, New York, NY).
- HUBER, P., STRASSEN, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Stat.* **1**, 251–263.
- JENSEN, B.A., SORESENSEN, C. (2001). Paying for minimum interest rate guarantees: who should compensate whom. *Eur. Financ. Manage.* **7** (2), 183–211.
- JOUINI, E., SCHACHERMAYER, W., TOUZI, N. (2006). *Law Invariant Risk Measures Have the Fatou Property*. Advances in Mathematical Economics Volume 9 (Springer, Tokyo, Japan).
- KAHNEMAN, D., TVERSKY, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291.
- KAHNEMAN, D., TVERSKY, A. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertainty* **5**, 297–323.
- KARATZAS, I., ŽITKOVIĆ, G. (2003). Optimal consumption from investment and random endowment in incomplete semimartingale markets. *Ann. Probab.* **31** (4), 1821–1858.
- KARATZAS, I., SHREVE, S.E. (1998). *Methods of Mathematical Finance* (Springer, New York, NY).
- KIRCH, M. (2000). Efficient hedging in incomplete markets under model uncertainty, PhD thesis (Humboldt-Universität zu Berlin, Berlin, Germany).
- KIRCH, M., RUNGALDIER, W. (2005). Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optim.* **43** (4), 1174–1195.
- KLÖPPEL, S., SCHWEIZER, M. (2007). Dynamic indifference valuation via convex risk measures. To appear in *Mathematical Finance*.
- KLÖPPELBERG, C., PERGAMENCHTCHIKOV, S. (2007). Optimal consumption and investment with bounded capital-at-risk for power utility functions Preprint (TU München, Munich, Germany).
- KORN, R., MENKENS, O. (2005). Worst-case scenario portfolio optimization: a new stochastic control approach. *Math. Methods Oper. Res.* **62** (1), 123–140.
- KORN, R., STEFFENSEN, M. (2006). On worst case portfolio optimization Preprint (TU Kaiserslautern, Kaiserslautern, Germany).

- KORN, R., WILMOTT, P. (2002). Optimal portfolios under the threat of a crash. *Int. J. Theor. Appl. Financ.* **5** (2), 171–187.
- KRÄTSCHEMER, V. (2005). Robust representation of convex risk measures by probability measures. *Financ. Stoch.* **9**, 597–608.
- KRAMKOV, D., SCHACHERMAYER, W. (1999). The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Ann. Appl. Probab.* **9** (3), 904–950.
- KRAMKOV, D., SCHACHERMAYER, W. (2003). Necessary and sufficient conditions in the problem of optimal investment in incomplete markets. *Ann. Appl. Probab.* **13** (4).
- KREPS, D. (1988). *Notes on the Theory of Choice* (Westview Press, Boulder, CO).
- KRYLOV, N.V. (2000). On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients. *Probab. Theory Rel.* **117**, 1–16.
- KUNZE, M. (2003). *Verteilungsinvariante konvexe Risikomaße. Diplomarbeit* (Humboldt-Universität zu Berlin, Berlin, Germany).
- KUSHNER, H., DUPUIS, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time, Second Edition*. Applications of mathematics (New York), 24. Stochastic modelling and applied probability (Springer-Verlag, New York, NY).
- KUSUOKA, S. (2001). On law invariant coherent risk measures. *Adv. Math. Econ.* **3**, 83–95.
- LAKNER, P., NYGREN, L.M. (2006). Portfolio optimization with downside risk constraints. *Math. Financ.* **16** (2), 283–299.
- LAZRAK, A., QUENEZ, M.-C. (2003). A generalized stochastic differential utility. *Math. Oper. Res.* **28** (1), 154–180.
- LEIPPOLD, M., TROJANI, F., VANINI, P. (2006). Equilibrium impact of value-at-risk regulation. *J. Econ. Dyn. Control* **30**, 1277–1313.
- LEMBCKE, J. (1988). The necessity of strongly subadditive capacities for Neyman-Pearson minimax tests. *Monatsh. Math.* **105**, 113–126.
- LUCAS, R.E. (1978). Asset pricing in an exchange economy. *Econometrica* **46** (6), 1429–1445.
- MÜLLER, M. (2005). Market completion and robust utility maximization, PhD thesis (Humboldt-Universität zu Berlin, Berlin, Germany) (Available at: <http://edoc.hu-berlin.de/docviews/abstract.php?id=26287>).
- MACCHERONI, F., RUSTICHINI, A., MARINACCI, M. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* **74**, 1447–1498.
- PIRVU, T., ZITKOVIC, G. (2007). Maximizing the growth rate under risk constraints. Working paper, to appear in Mathematical Finance.
- QUENEZ, M. (2004). *Optimal Portfolio in a Multiple-Priors Model*. Seminar on stochastic analysis, random fields and applications IV, 291–321, In: Progress in Probability, volume 58 (Birkhäuser, Basel, Switzerland).
- RIEDEL, F. (2004). Dynamic coherent risk measures. *Stoch. Proc. Appl.* **112** (2), 185–200.
- RUDLOFF, B. (2006). Hedging in incomplete markets and testing compound hypotheses via convex duality, PhD thesis (University of Halle-Wittenberg, Halle-Wittenberg, Germany).
- RUNGALDIER, W. (2001). Adaptive and robust control procedures for risk minimization under uncertainty. In: Menaldi, J.L., Rofman, E., Sulem, A. (eds.), *Optimal control and Partial Differential Equations*. Volume in Honour of Prof. Alain Bensoussan's 60th Birthday (IOS Press), pp. 549–557.
- RUNGALDIER, W. (2003). On stochastic control in finance. In: *Mathematical Systems Theory in Biology, Communications, Computation, and Finance* IMA Volumes in Mathematics and its Applications, Volume 134 (Springer, New York, NY), pp. 317–344.
- RUSZCZYŃSKI, A., SHAPIRO, A. (2006a). Conditional risk mappings. *Math. Oper. Res.* **31** (3), 544–561.
- RUSZCZYŃSKI, A., SHAPIRO, A. (2006b). Optimization of convex risk functions. *Math. Oper. Res.* **31** (3), 433–452.
- SAVAGE, L.J. (1954). *The Foundations of Statistics* (John Wiley and Sons, New York, NY).
- SCHIED, A. (2004). On the Neyman-Pearson problem for law-invariant risk measures and robust utility functionals. *Ann. Appl. Probab.* **14**, 1398–1423.
- SCHIED, A. (2005). Optimal investments for robust utility functionals in complete market models. *Math. Oper. Res.* **30** (3), 750–764.
- SCHIED, A. (2006). Risk measures and robust optimization problems. *Stoch. Models* **22**, 753–831.

- SCHIED, A. (2007a). Optimal investments for risk-and ambiguity-averse preferences: a duality approach. *Financ. Stoch.* **11** (1), 107–129.
- SCHIED, A. (2007b). Robust optimal control for a consumption-investment problem. To appear in *Mathematical Methods of Operations Research*.
- SCHIED, A., STADJE, M. (2007). Robustness of Delta hedging for path-dependent options in local volatility models. To appear in *Journal of Applied Probability* 44, no. 4.
- SCHIED, A., WU, C.-T. (2005). Duality theory for optimal investments under model uncertainty. *Stat. Decis.* **23** (3), 199–217.
- SCHMEIDLER, D. (1989). Subjective probability and expected utility without additivity. *Econometrica* **57** (3), 571–587.
- SEKINE, J. (2004). Dynamic minimization of worst conditional expectation of shortfall. *Math. Financ.* **14**, 605–618.
- TALAY, D., ZHENG, Z. (2002). Worst case model risk management. *Financ. Stoch.* **6**, 517–537.
- TUTSCH, S. (2006). Konsistente und konsequente dynamische Risikomasse und das Problem der Aktualisierung, Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin, Germany.
- VANDEN, J.M. (2006). Portfolio insurance and volatility regime switching. *Math. Financ.* **16** (2), 387–417.
- VON NEUMANN, J., MORGENTERN, O. (1944). *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ).
- WEBER, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Math. Financ.* **16**, 419–442.
- WITTMÜSS, W. (2006). Robust optimization of consumption with random endowment. To appear in *Stochastics*.
- YAARI, M. (1987). The dual theory of choice under risk. *Econometrica* **55**, 95–116.

Stochastic Portfolio Theory: an Overview

Ioannis Karatzas

Department of Mathematics, Columbia University, New York, NY 10027, USA

E-mail address: ik@math.columbia.edu

Robert Fernholz

INTECH, One Palmer Square, Princeton, NJ 08542, USA

E-mail address: bob@enhanced.com

Abstract

Stochastic Portfolio Theory is a flexible framework for analyzing portfolio behavior and equity market structure. This theory was introduced by Fernholz in the papers (*Journal of Mathematical Economics*, 1999; *Finance & Stochastics*, 2001) and in the monograph *Stochastic Portfolio Theory* (Springer 2002). It was further developed by FERNHOLZ, KARATZAS and KARDARAS (*Finance & Stochastics*, 2005), FERNHOLZ & KARATZAS (*Annals of Finance*, 2005), BANNER, FERNHOLZ and KARATZAS (*Annals of Applied Probability*, 2005), and KARATZAS and KARDARAS (*Finance & Stochastics*, 2007). This theory is descriptive, as opposed to normative; it is consistent with observable characteristics of actual portfolios and markets, and it provides a theoretical tool which is useful for practical applications.

As a theoretical tool, this framework offers fresh insights into questions of stock market structure and arbitrage, and can be used to construct portfolios with controlled behavior. As a practical tool, stochastic portfolio theory has been applied to the analysis and optimization of portfolio performance and has been the basis of successful investment strategies for over a decade.

Contents

CHAPTER I	95
1. Markets and portfolios	95
2. The market portfolio	100
3. Some useful properties	102
4. Portfolio optimization	106
CHAPTER II	111
5. Diversity	111
6. Relative arbitrage and its consequences	113
7. Diversity leads to arbitrage	118
8. Mirror portfolios, short-horizon arbitrage	122
9. A diverse market model	126
10. Hedging and optimization without EMM	127
CHAPTER III	135
11. Portfolio-generating functions	135
CHAPTER IV	149
12. Volatility-stabilized markets	149
13. Rank-based models	155
14. Some concluding remarks	164

Introduction

Stochastic Portfolio Theory (SPT), as we currently think of it, appeared in 1995 in the manuscript “On the Diversity of Equity Markets,” which eventually appeared as a paper FERNHOLZ [1999] in the *Journal of Mathematical Economics*. Since then, SPT has evolved into a flexible framework for analyzing portfolio behavior and equity market structure, with both theoretical and practical applications. As a theoretical methodology, this framework provides insight into questions of market behavior and arbitrage and can be used to construct portfolios with controlled behavior under quite general conditions. As a practical tool, SPT has been applied to the analysis and optimization of portfolio performance and has been the basis of successful equity investment strategies for over a decade.

SPT is a descriptive theory, which studies and attempts to explain observable phenomena that take place in equity markets. This orientation is quite different from that of the well-known modern portfolio theory of dynamic asset pricing (DAP), in which market structure is analyzed under strong normative assumptions regarding the behavior of market participants. It has long been suggested that the distinction between descriptive and normative theories separates the natural sciences from the social sciences; if this dichotomy is valid, then one might argue that SPT resides with the natural sciences.

SPT descends from the “classical portfolio theory” of HARRY MARKOWITZ [1952], as does much of mathematical finance. At the same time, it represents a rather significant departure from some important aspects of the current theory of DAP. DAP is a normative theory that grew out of the general equilibrium model of mathematical economics for financial markets, evolved through the capital asset pricing models, and is currently predicated on the absence of arbitrage and on the existence of equivalent martingale measure(s) (EMM). SPT, by contrast, is applicable under a wide range of assumptions and conditions that may hold in actual equity markets. Unlike dynamic asset pricing, it is consistent with either equilibrium or disequilibrium, with either arbitrage or no-arbitrage, and is not predicated on the existence of EMM.

While SPT has been developed with equity markets in mind, a reasonable portion of the theory is valid for general financial assets, as long as the asset values remain positive. For such general assets, the “market” can be replaced by an arbitrary passive portfolio with positive holdings in each of the assets. Although some concepts related to equity markets may not be meaningful in these general applications, other concepts would appear to carry over without significant modification.

This survey reviews the central ideas of SPT and presents examples of portfolios and markets with a wide variety of different properties. SPT is a fast-evolving field, so we also present a number of research problems that remain open, at least at the time of this writing. Proofs for some of the results are included here, but at other times, simply a reference is given.

The survey is separated into four chapters. Chapter I, Basics, introduces the concepts of markets and portfolios, in particular, the market portfolio, the most important portfolio of them all. In this first chapter we also encounter the excess growth rate process, a quantity that pervades SPT. Chapter II, Diversity & Arbitrage, introduces market diversity and shows how diversity can lead to relative arbitrage in an equity market. Historically, these were among the first phenomena analyzed using SPT. Portfolio generating functions are versatile tools for constructing portfolios with particular properties, and these functions are discussed in Chapter III, Functionally Generated Portfolios. Here, we also consider stocks identified by rank, as opposed to by name, and discuss implications regarding the size effect. Roughly speaking, these first three chapters of the survey outline the techniques that historically have comprised SPT; the fourth chapter looks toward the future.

Chapter IV, Abstract Markets, is devoted to the area of much of the current research in SPT. Abstract markets are models of equity markets that show certain characteristics of real stock markets, but for which the precise mathematical structure is known (since we can define them as we wish!). Here, we see volatility-stabilized markets that are not diverse but nevertheless allow arbitrage, and we also look at rank-based markets that have stability properties similar to those of real stock markets. Several problems regarding these abstract markets are proposed.

Basics

SPT uses the logarithmic representation for stocks and portfolios rather than the arithmetic representation used in “classical” mathematical finance. In the logarithmic representation, the classical rate of return is replaced by the growth rate, sometimes referred to as the geometric rate of return or the logarithmic rate of return. The logarithmic and arithmetic representations are equivalent, but nevertheless, the different perspectives bring to light distinct aspects of portfolio behavior. The use of the logarithmic representation in no way implies the use of a logarithmic utility function: indeed, SPT is not concerned with expected utility maximization at all.

We introduce here the basic structures of SPT, stocks and portfolios, and discuss that most important portfolio of them all, the market portfolio. We show that the growth rate of a portfolio depends not only on the growth rates of the component stocks but also on the excess growth rate which is determined by the stocks’ variances and covariances. Finally, we consider a few optimization problems in the logarithmic setting.

Most of the material in this chapter can be found in FERNHOLZ [2002].

1. Markets and portfolios

We shall place ourselves in a model $\mathcal{M}$ for a financial market of the form

$$\begin{aligned} dB(t) &= B(t)r(t)dt, \quad B(0) = 1, \\ dX_i(t) &= X_i(t) \left(b_i(t)dt + \sum_{v=1}^d \sigma_{iv}(t)dW_v(t) \right), \\ X_i(0) &= x_i > 0, \quad i = 1, \dots, n, \end{aligned} \tag{1.1}$$

consisting of a money market $B(\cdot)$ and of n stocks, whose prices $X_1(\cdot), \dots, X_n(\cdot)$ are driven by the d -dimensional Brownian motion $W(\cdot) = (W_1(\cdot), \dots, W_d(\cdot))'$, with $d \geq n$. Contrary to a usual assumption imposed on such models, here it is not crucial that the filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{0 \leq t < \infty}$, which represents the “flow of information” in the market, be the one generated by the Brownian motion itself. Thus, until further notice, we shall consider $\mathbb{F}$ to contain (possibly strictly) this Brownian filtration $\mathbb{F}^W = \{\mathcal{F}^W(t)\}_{0 \leq t < \infty}$, where $\mathcal{F}^W(t) := \sigma(W(s), 0 \leq s \leq t) \subseteq \mathcal{F}(t), \forall t \in [0, \infty)$.

We shall assume that the interest-rate process $r(\cdot)$ for the money market, the vector-valued process $b(\cdot) = (b_1(\cdot), \dots, b_n(\cdot))'$ of rates of return for the various stocks, and the $(n \times d)$ -matrix-valued process $\sigma(\cdot) = (\sigma_{iv}(\cdot))_{1 \leq i \leq n, 1 \leq v \leq d}$ of stock-price volatilities, are all $\mathbb{F}$ -progressively measurable and satisfy for every $T \in (0, \infty)$ the integrability conditions

$$\int_0^T |r(t)| dt + \sum_{i=1}^n \int_0^T \left(|b_i(t)| + \sum_{v=1}^d (\sigma_{iv}(t))^2 \right) dt < \infty. \quad \text{a.s.} \quad (1.2)$$

This setting admits a rich class of continuous-path Itô processes with very general distributions: in particular, *no Markovian or Gaussian assumption is imposed*. In fact, it is possible to extend the scope of the theory to very general semimartingale settings (see KARDARAS [2003] for details).

We shall introduce the notation

$$a_{ij}(t) := \sum_{v=1}^d \sigma_{iv}(t) \sigma_{jv}(t) = (\sigma(t) \sigma'(t))_{ij} = \frac{d}{dt} \langle \log X_i, \log X_j \rangle(t) \quad (1.3)$$

for the nonnegative definite matrix-valued covariance process $a(\cdot) = (a_{ij}(\cdot))_{1 \leq i, j \leq n}$ of the stocks in the market, and

$$\gamma_i(t) := b_i(t) - \frac{1}{2} a_{ii}(t), \quad i = 1, \dots, n. \quad (1.4)$$

Then, we may use Itô's rule to solve (1.1) in the form

$$d \log X_i(t) = \gamma_i(t) dt + \sum_{v=1}^d \sigma_{iv}(t) dW_v(t), \quad i = 1, \dots, n, \quad (1.5)$$

or equivalently

$$X_i(t) = x_i \exp \left\{ \int_0^t \gamma_i(u) du + \sum_{v=1}^d \int_0^t \sigma_{iv}(u) dW_v(u) \right\}, \quad 0 \leq t < \infty.$$

Eq. (1.5) is called the logarithmic representation of the stock price process, and we shall refer to the quantity of (1.4) as the growth rate of the i th stock, because of the a.s. relationship

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\log X_i(T) - \int_0^T \gamma_i(t) dt \right) = 0. \quad (1.6)$$

This is valid when the individual stock variances $a_{ii}(\cdot)$ do not increase too quickly, for example if we have

$$\lim_{T \rightarrow \infty} \left(\frac{\log \log T}{T^2} \int_0^T a_{ii}(t) dt \right) = 0, \quad \text{a.s.}; \quad (1.7)$$

then, (1.6) follows from the law of the iterated logarithm and from the representation of (local) martingales as time-changed Brownian motions.

DEFINITION 1.1. A portfolio $\pi(\cdot) = (\pi_1(\cdot), \dots, \pi_n(\cdot))'$ is an $\mathbb{F}$ -progressively measurable process, bounded uniformly in (t, ω) , with values in the set

$$\bigcup_{\kappa \in \mathbb{N}} \{(\pi_1, \dots, \pi_n) \in \mathbb{R}^n \mid \pi_1^2 + \dots + \pi_n^2 \leq \kappa^2, \pi_1 + \dots + \pi_n = 1\}.$$

A long-only portfolio $\pi(\cdot) = (\pi_1(\cdot), \dots, \pi_n(\cdot))'$ is a portfolio that takes values in the set

$$\Delta^n := \{(\pi_1, \dots, \pi_n) \in \mathbb{R}^n \mid \pi_1 \geq 0, \dots, \pi_n \geq 0 \text{ and } \pi_1 + \dots + \pi_n = 1\}.$$

For future reference, we shall introduce also the notation

$$\Delta_+^n := \{(\pi_1, \dots, \pi_n) \in \Delta^n \mid \pi_1 > 0, \dots, \pi_n > 0\}.$$

Thus, a portfolio can sell one or more stocks short (though certainly not all) but is never allowed to borrow from, or invest in, the money market, whereas a long-only portfolio sells no stocks short at all. The interpretation is that $\pi_i(t)$ represents the proportion of wealth $V^{w,\pi}(t)$ invested at time t in the i th stock, so the quantities

$$h_i(t) = \pi_i(t) V^{w,\pi}(t), \quad i = 1, \dots, n \quad (1.8)$$

are the dollar amounts invested at any given time t in the individual stocks.

The wealth process $V^{w,\pi}(\cdot)$, which corresponds to a portfolio $\pi(\cdot)$ and initial capital $w > 0$, satisfies the stochastic equation

$$\begin{aligned} \frac{dV^{w,\pi}(t)}{V^{w,\pi}(t)} &= \sum_{i=1}^n \pi_i(t) \frac{dX_i(t)}{X_i(t)} = \pi'(t) [b(t)dt + \sigma(t) dW(t)] \\ &= b_\pi(t) dt + \sum_{v=1}^d \sigma_{\pi v}(t) dW_v(t), \quad V^{w,\pi}(0) = w, \end{aligned} \quad (1.9)$$

where

$$b_\pi(t) := \sum_{i=1}^n \pi_i(t) b_i(t), \quad \sigma_{\pi v}(t) := \sum_{i=1}^n \pi_i(t) \sigma_{iv}(t) \quad \text{for } v = 1, \dots, d. \quad (1.10)$$

These quantities are the rate of return and the volatility coefficients associated with the portfolio $\pi(\cdot)$, respectively.

By analogy with (1.5) we can write the solution of the Eq. (1.9) as

$$d \log V^{w,\pi}(t) = \gamma_\pi(t) dt + \sum_{v=1}^d \sigma_{\pi v}(t) dW_v(t), \quad V^{w,\pi}(0) = w \quad (1.11)$$

or equivalently

$$V^{w,\pi}(t) = w \exp \left\{ \int_0^t \gamma_\pi(u) du + \sum_{v=1}^d \int_0^t \sigma_{\pi v}(u) dW_v(u) \right\}, \quad 0 \leq t < \infty$$

Here,

$$\gamma_\pi(t) := \sum_{i=1}^n \pi_i(t) \gamma_i(t) + \gamma_\pi^*(t) \quad (1.12)$$

is the *growth rate* of the portfolio $\pi(\cdot)$, and

$$\gamma_\pi^*(t) := \frac{1}{2} \left(\sum_{i=1}^n \pi_i(t) a_{ii}(t) - \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t) \right) \quad (1.13)$$

is the *excess growth rate* of the portfolio $\pi(\cdot)$. As we shall see in Lemma 3.3, for a long-only portfolio this excess growth rate is always nonnegative and is strictly positive for such portfolios that do not concentrate their holdings in just one stock.

Again, the terminology “growth rate” is justified by the a.s. property

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\log V^{w,\pi}(T) - \int_0^T \gamma_\pi(t) dt \right) = 0, \quad (1.14)$$

valid under the analogue

$$\lim_{T \rightarrow \infty} \left(\frac{\log \log T}{T^2} \int_0^T \|a(t)\| dt \right) = 0, \quad \text{a.s.} \quad (1.15)$$

of condition (1.7). Clearly, this condition is satisfied when all eigenvalues of the covariance matrix process $a(\cdot)$ of (1.3) are uniformly bounded away from infinity: that is, when

$$\xi' a(t) \xi = \xi' \sigma(t) \sigma'(t) \xi \leq K \|\xi\|^2, \quad \forall t \in [0, \infty) \quad \text{and} \quad \xi \in \mathbb{R}^n \quad (1.16)$$

holds almost surely, for some constant $K \in (0, \infty)$. We shall refer to (1.16) as the *uniform boundedness* condition on the volatility structure of $\mathcal{M}$.

Without further comment, we shall write $V^\pi(\cdot) \equiv V^{1,\pi}(\cdot)$ for initial wealth $w = \$1$. Let us also note the following analog of (1.11), namely,

$$d \log V^\pi(t) = \gamma_\pi^*(t) dt + \sum_{i=1}^n \pi_i(t) d \log X_i(t). \quad (1.17)$$

DEFINITION 1.2. We shall use the reverse-order statistics notation for the weights of a portfolio $\pi(\cdot)$, ranked at time t in decreasing order, from largest down to smallest

$$\max_{1 \leq i \leq n} \pi_i(t) =: \pi_{(1)}(t) \geq \pi_{(2)}(t) \geq \dots \geq \pi_{(n-1)}(t) \geq \pi_{(n)}(t) := \min_{1 \leq i \leq n} \pi_i(t). \quad (1.18)$$

For an arbitrary portfolio $\pi(\cdot)$, and with e_i denoting the i th unit vector in $\mathbb{R}^n$, let us introduce the quantities

$$\begin{aligned}\tau_{ij}^\pi(t) &:= \sum_{v=1}^d (\sigma_{iv}(t) - \sigma_{\pi v}(t))(\sigma_{jv}(t) - \sigma_{\pi v}(t)) \\ &= (\pi(t) - e_i)' a(t) (\pi(t) - e_j) = a_{ij}(t) - a_{\pi i}(t) - a_{\pi j}(t) + a_{\pi\pi}(t)\end{aligned}\quad (1.19)$$

for $1 \leq i, j \leq n$ and set

$$a_{\pi i}(t) := \sum_{j=1}^n \pi_j(t) a_{ij}(t), \quad a_{\pi\pi}(t) := \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t) = \sum_{v=1}^d (\sigma_{\pi v}(t))^2. \quad (1.20)$$

It is seen from (1.11) that this last quantity is the variance of the portfolio $\pi(\cdot)$.

We shall call the matrix-valued process $\tau^\pi(\cdot) = (\tau_{ij}^\pi(\cdot))_{1 \leq i, j \leq n}$ of (1.19) the process of individual stocks' covariances relative to the portfolio $\pi(\cdot)$. It satisfies the elementary property

$$\sum_{j=1}^n \tau_{ij}^\pi(t) \pi_j(t) = 0, \quad i = 1, \dots, n. \quad (1.21)$$

Trading strategies: For completeness of exposition and for later use in this survey, let us go briefly beyond portfolios and recall the notion of trading strategies: these are allowed to invest in (or borrow from) the money market. Formally, they are $\mathbb{F}$ -progressively measurable, $\mathbb{R}^n$ -valued processes $h(\cdot) = (h_1(\cdot), \dots, h_n(\cdot))'$ that satisfy the integrability condition

$$\sum_{i=1}^n \int_0^T \left(|h_i(t)| |b_i(t) - r(t)| + h_i^2(t) a_{ii}(t) \right) dt < \infty, \quad \text{a.s.}$$

for every $T \in (0, \infty)$. The interpretation is that the real-valued, $\mathcal{F}(t)$ -measurable random variable $h_i(t)$ stands for the dollar amount invested by the strategy $h(\cdot)$ at time t in the i th stock. If we denote by $\mathcal{V}^{w,h}(t)$ the wealth at time t corresponding to this strategy $h(\cdot)$ and to an initial capital $w > 0$, then $\mathcal{V}^{w,h}(t) - \sum_{i=1}^n h_i(t)$ is the amount invested in the money market, and we have

$$\begin{aligned}d\mathcal{V}^{w,h}(t) &= \left(\mathcal{V}^{w,h}(t) - \sum_{i=1}^n h_i(t) \right) r(t) dt \\ &\quad + \sum_{i=1}^n h_i(t) \left(b_i(t) dt + \sum_{v=1}^d \sigma_{iv}(t) dW_v(t) \right)\end{aligned}$$

or equivalently

$$\frac{\mathcal{V}^{w,h}(t)}{B(t)} = w + \int_0^t \frac{h'(s)}{B(s)} \left((b(s) - r(s)\mathbf{I}) ds + \sigma(s) dW(s) \right), \quad 0 \leq t < \infty. \quad (1.22)$$

Here, $\mathbf{I} = (1, \dots, 1)'$ is the n -dimensional column vector with 1 in all entries. Again, without further comment, we shall write $\mathcal{V}^h(\cdot) \equiv \mathcal{V}^{1,h}(\cdot)$ for initial wealth $w = \$1$.

As mentioned already, all quantities $h_i(\cdot)$, $1 \leq i \leq n$, and $\mathcal{V}^{w,h}(t) - h'(\cdot)\mathbf{I}$ are allowed to take negative values. This possibility opens the door to the notorious doubling strategies of martingale theory (e.g. KARATZAS and SHREVE [1998], chapter 1). In order to rule these out on a given time horizon $[0, T]$, we shall confine ourselves here to trading strategies $h(\cdot)$ that satisfy

$$\mathbb{P} \left(\mathcal{V}^{w,h}(t) \geq 0, \quad \forall 0 \leq t \leq T \right) = 1. \quad (1.23)$$

Such strategies will be called *admissible* for the initial capital $w > 0$ on the time horizon $[0, T]$; their collection will be denoted $\mathcal{H}(w; T)$, and we shall set $\mathcal{H}(w) := \bigcap_{T>0} \mathcal{H}(w; T)$.

We shall also find useful to look at the collection $\mathcal{H}_+(w; T) \subset \mathcal{H}(w; T)$ of *strongly admissible* strategies, with $\mathbb{P}(\mathcal{V}^{w,h}(t) > 0, \quad \forall 0 \leq t \leq T) = 1$. Similarly, we shall set $\mathcal{H}_+(w) := \bigcap_{T>0} \mathcal{H}_+(w; T)$.

Each portfolio $\pi(\cdot)$ generates, via (1.8), a trading strategy $h(\cdot) \in \mathcal{H}_+(w)$ we have $\mathcal{V}^{w,h}(\cdot) \equiv V^{w,\pi}(\cdot)$. It is not difficult to see from (1.9) that the trading strategy generated by a portfolio $\pi(\cdot)$ is self-financing (see DUFFIE [1992] for a discussion).

2. The market portfolio

Suppose we normalize so that each stock has always just one share outstanding; then, the stock price $X_i(t)$ can be interpreted as the capitalization of the i th company at time t , and the quantities

$$X(t) := X_1(t) + \dots + X_n(t) \quad \text{and} \quad \mu_i(t) := \frac{X_i(t)}{X(t)}, \quad i = 1, \dots, n \quad (2.1)$$

as the total capitalization of the market and the relative capitalizations of the individual companies, respectively. Clearly, $0 < \mu_i(t) < 1$, $\forall i = 1, \dots, n$ and $\sum_{i=1}^n \mu_i(t) = 1$, so we may think of the vector process $\mu(\cdot) = (\mu_1(\cdot), \dots, \mu_n(\cdot))'$ as a portfolio that invests the proportion $\mu_i(t)$ of current wealth in the i th asset at all times. Equivalently, this portfolio holds the same constant number of shares in all assets at all times. The resulting wealth process $V^{w,\mu}(\cdot)$ satisfies

$$\frac{dV^{w,\mu}(t)}{V^{w,\mu}(t)} = \sum_{i=1}^n \mu_i(t) \frac{dX_i(t)}{X_i(t)} = \sum_{i=1}^n \frac{dX_i(t)}{X(t)} = \frac{dX(t)}{X(t)},$$

in accordance with (2.1) and (1.9). In other words,

$$V^{w,\mu}(\cdot) \equiv \frac{w}{X(0)} X(\cdot); \quad (2.2)$$

investing in the portfolio $\mu(\cdot)$ is tantamount to ownership of the entire market, in proportion of course, to the initial investment. For this reason, we shall call $\mu(\cdot)$ of (2.1) the *market portfolio*, and the processes $\mu_i(\cdot)$ the *market weight processes*.

By analogy with (1.11), we have

$$d \log V^{w, \mu}(t) = \gamma_\mu(t) dt + \sum_{v=1}^d \sigma_{\mu v}(t) dW_v(t), \quad V^{w, \mu}(0) = w, \quad (2.3)$$

and comparison of Eq. (2.3) with (1.5) gives the dynamics of the market weights

$$d \log \mu_i(t) = (\gamma_i(t) - \gamma_\mu(t)) dt + \sum_{v=1}^d (\sigma_{iv}(t) - \sigma_{\mu v}(t)) dW_v(t) \quad (2.4)$$

in (2.1) for all stocks $i = 1, \dots, n$ in the notation of (1.10) and (1.12); equivalently,

$$\frac{d\mu_i(t)}{\mu_i(t)} = \left(\gamma_i(t) - \gamma_\mu(t) + \frac{1}{2} \tau_{ii}^\mu(t) \right) dt + \sum_{v=1}^d (\sigma_{iv}(t) - \sigma_{\mu v}(t)) dW_v(t). \quad (2.5)$$

We are recalling here the quantities

$$\tau_{ij}^\mu(t) := \sum_{v=1}^d (\sigma_{iv}(t) - \sigma_{\mu v}(t)) (\sigma_{jv}(t) - \sigma_{\mu v}(t)) = \frac{d\langle \mu_i, \mu_j \rangle(t)}{\mu_i(t) \mu_j(t) dt}, \quad 1 \leq i, j \leq n \quad (2.6)$$

of (1.19) for the market portfolio $\pi(\cdot) \equiv \mu(\cdot)$, namely, the covariances of the individual stocks relative to the entire market.

REMARK 2.1. Coherence: We say that the market model $\mathcal{M}$ of (1.1) and (1.2) is *coherent* if the relative capitalizations of (2.1) satisfy

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \mu_i(T) = 0 \quad \text{almost surely, for each } i = 1, \dots, n \quad (2.7)$$

(i.e., if none of the stocks decline too rapidly with respect to the market as a whole). Under the condition (1.15) on the covariance structure, it can be shown that coherence is equivalent to each of the following two conditions:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\gamma_i(t) - \gamma_\mu(t)) dt = 0 \quad \text{a.s., for each } i = 1, \dots, n, \quad (2.8)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\gamma_i(t) - \gamma_j(t)) dt = 0 \quad \text{a.s., for each pair } 1 \leq i, j \leq n. \quad (2.9)$$

See FERNHOLZ [2002], pp 26–27 for details.

3. Some useful properties

In this section, we collect together some useful properties of the relative covariance process in (1.19), for ease of reference in future usage. For any given stock i and portfolio $\pi(\cdot)$, the *relative return process* of the i th stock versus $\pi(\cdot)$ is the process

$$R_i^\pi(t) := \log \left(\frac{X_i(t)}{V^{w,\pi}(t)} \right) \Big|_{w=X_i(0)}, \quad 0 \leq t < \infty. \quad (3.1)$$

LEMMA 3.1. *For any portfolio $\pi(\cdot)$, and for all $1 \leq i, j \leq n$ and $t \in [0, \infty)$, we have almost surely*

$$\tau_{ij}^\pi(t) = \frac{d}{dt} \langle R_i^\pi, R_j^\pi \rangle(t), \quad \text{in particular,} \quad \tau_{ii}^\pi(t) = \frac{d}{dt} \langle R_i^\pi \rangle(t) \geq 0, \quad (3.2)$$

for the relative covariances of (1.19); and the matrix $\tau^\pi(t) = (\tau_{ij}^\pi(t))_{1 \leq i, j \leq n}$ is a.s. nonnegative definite. Furthermore, if the covariance matrix $a(t)$ is positive definite, then the relative covariance matrix $\tau^\pi(t)$ has rank $n - 1$, and its null space is spanned by the vector $\pi(t)$, almost surely.

PROOF. Comparing (1.5) with (1.11), we get the analogue

$$dR_i^\pi(t) = (\gamma_i(t) - \gamma_\pi(t)) dt + \sum_{v=1}^d (\sigma_{iv}(t) - \sigma_{\pi v}(t)) dW_v(t),$$

of (2.4), from which the first two claims follow.

Now, suppose that $a(t)$ is positive definite. For any $x \in \mathbb{R}^n \setminus \{0\}$ and with $\eta := \sum_{i=1}^n x_i$, we compute from (2.6), (1.19):

$$x' \tau^\pi(t) x = x' a(t) x - 2\eta x' a(t) \pi(t) + \eta^2 \pi'(t) a(t) \pi(t).$$

If $\sum_{i=1}^n x_i = 0$, then $x' \tau^\pi(t) x = x' a(t) x > 0$. If on the other hand $\eta := \sum_{i=1}^n x_i \neq 0$, we consider the vector $y := x/\eta$ that satisfies $\sum_{i=1}^n y_i = 1$ and observe that $\eta^{-2} x' \tau^\pi(t) x$ is equal to

$$y' \tau^\pi(t) y = y' a(t) y - 2y' a(t) \pi(t) + \pi'(t) a(t) \pi(t) = (y - \pi(t))' a(t) (y - \pi(t)),$$

thus zero if and only if $y = \pi(t)$ or equivalently $x = \eta \pi(t)$. □

LEMMA 3.2. *For any two portfolios $\pi(\cdot)$ and $\rho(\cdot)$, we have*

$$d \log \left(\frac{V^\pi(t)}{V^\rho(t)} \right) = \gamma_\pi^*(t) dt + \sum_{i=1}^n \pi_i(t) d \log \left(\frac{X_i(t)}{V^\rho(t)} \right). \quad (3.3)$$

In particular, we get the dynamics

$$\begin{aligned} d \log \left(\frac{V^\pi(t)}{V^\mu(t)} \right) &= \gamma_\pi^*(t) dt + \sum_{i=1}^n \pi_i(t) d \log \mu_i(t) \\ &= (\gamma_\pi^*(t) - \gamma_\mu^*(t)) dt + \sum_{i=1}^n (\pi_i(t) - \mu_i(t)) d \log \mu_i(t) \end{aligned} \quad (3.4)$$

for the relative return of an arbitrary portfolio $\pi(\cdot)$ with respect to the market.

PROOF. Eq. (3.3) follows from (1.17), and the first equality in (3.4) is the special case of (3.3) with $\rho(\cdot) \equiv \mu(\cdot)$. The second equality in (3.4) follows upon observing from (2.4) that

$$\sum_{i=1}^n \mu_i(t) d \log \mu_i(t) = \sum_{i=1}^n \mu_i(t) (\gamma_i(t) - \gamma_\mu(t)) dt = -\gamma_\mu^*(t) dt. \quad \square$$

LEMMA 3.3. For any two portfolios $\pi(\cdot)$ and $\rho(\cdot)$, we have the numéraire-invariance property

$$\gamma_\pi^*(t) = \frac{1}{2} \left(\sum_{i=1}^n \pi_i(t) \tau_{ii}^\rho(t) - \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) \pi_j(t) \tau_{ij}^\rho(t) \right). \quad (3.5)$$

In particular, recalling (1.21), we obtain the representation

$$\gamma_\pi^*(t) = \frac{1}{2} \sum_{i=1}^n \pi_i(t) \tau_{ii}^\pi(t) \quad (3.6)$$

for the excess growth rate, as a weighted average of the individual stocks' variances $\tau_{ii}^\pi(\cdot)$ relative to the portfolio $\pi(\cdot)$, as in (1.19). From (3.6), (3.2), and Definition 1.1, we get for any long-only portfolio $\pi(\cdot)$ the property

$$\gamma_\pi^*(t) \geq 0. \quad (3.7)$$

PROOF. From (1.19), we obtain

$$\sum_{i=1}^n \pi_i(t) \tau_{ii}^\rho(t) = \sum_{i=1}^n \pi_i(t) a_{ii}(t) - 2 \sum_{i=1}^n \pi_i(t) a_{\rho i}(t) + a_{\rho\rho}(t)$$

and

$$\sum_{i=1}^n \sum_{j=1}^n \pi_i(t) \tau_{ij}^\rho(t) \pi_j(t) = \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t) - 2 \sum_{i=1}^n \pi_i(t) a_{\rho i}(t) + a_{\rho\rho}(t),$$

and (3.5) follows from (1.13). $\square$

For the market portfolio, Eq. (3.6) becomes

$$\gamma_\mu^*(t) = \frac{1}{2} \sum_{i=1}^n \mu_i(t) \tau_{ii}^\mu(t); \quad (3.8)$$

the summation on the right-hand side is the average, according to the market weights of individual stocks, of these stocks' variances relative to the market. Thus, (3.8) gives an interpretation of the excess growth rate of the market portfolio, as a measure of the market's "intrinsic" volatility.

REMARK 3.1. Note that (3.4), in conjunction with (2.4), (2.5) and the numéraire-invariance property (3.5), implies that for any portfolio $\pi(\cdot)$, we have the relative return formula $d(V^\pi(t)/V^\mu(t)) = (V^\pi(t)/V^\mu(t)) \sum_{i=1}^n (\pi_i(t)/\mu_i(t)) d\mu_i(t)$, or equivalently, in conjunction with (2.6):

$$d \log \left(\frac{V^\pi(t)}{V^\mu(t)} \right) = \sum_{i=1}^n \frac{\pi_i(t)}{\mu_i(t)} d\mu_i(t) - \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n \pi_i(t) \pi_j(t) \tau_{ij}^\mu(t) \right) dt. \quad (3.9)$$

LEMMA 3.4. Assume that the covariance process $a(\cdot)$ of (1.3) satisfies the following strong nondegeneracy condition: there exists a constant $\varepsilon \in (0, \infty)$ such that

$$\xi' a(t) \xi = \xi' \sigma(t) \sigma'(t) \xi \geq \varepsilon \|\xi\|^2, \quad \forall t \in [0, \infty) \text{ and } \xi \in \mathbb{R}^n \quad (3.10)$$

holds almost surely (all eigenvalues are bounded away from zero). Then, for every portfolio $\pi(\cdot)$ and all $0 \leq t < \infty$, we have in the notation of (1.18) the inequalities

$$\varepsilon (1 - \pi_i(t))^2 \leq \tau_{ii}^\pi(t), \quad i = 1, \dots, n, \quad (3.11)$$

almost surely. If the portfolio $\pi(\cdot)$ is long only, we also have

$$\frac{\varepsilon}{2} (1 - \pi_{(1)}(t)) \leq \gamma_\pi^*(t). \quad (3.12)$$

PROOF. With e_i denoting the i th unit vector in $\mathbb{R}^n$, we have

$$\tau_{ii}^\pi(t) = (\pi(t) - e_i)' a(t) (\pi(t) - e_i) \geq \varepsilon \|\pi(t) - e_i\|^2 = \varepsilon \left((1 - \pi_i(t))^2 + \sum_{j \neq i} \pi_j^2(t) \right)$$

from (1.19) and (3.10), thus (3.11) follows. Back into (3.6), and with $\pi_i(t) \geq 0$ valid for all $i = 1, \dots, n$, this lower estimate gives

$$\gamma_\pi^*(t) \geq \frac{\varepsilon}{2} \sum_{i=1}^n \pi_i(t) \left((1 - \pi_i(t))^2 + \sum_{j \neq i} \pi_j^2(t) \right)$$

$$\begin{aligned}
&= \frac{\varepsilon}{2} \left(\sum_{i=1}^n \pi_i(t) (1 - \pi_i(t))^2 + \sum_{j=1}^n \pi_j^2(t) (1 - \pi_j(t)) \right) \\
&= \frac{\varepsilon}{2} \sum_{i=1}^n \pi_i(t) (1 - \pi_i(t)) \geq \frac{\varepsilon}{2} (1 - \pi_{(1)}(t)). \quad \square
\end{aligned}$$

LEMMA 3.5. Assume that the uniform boundedness condition (1.16) holds; then, for every long-only portfolio $\pi(\cdot)$ and for $0 \leq t < \infty$, we have in the notation of (1.18) the a.s. inequalities

$$\tau_{ii}^\pi(t) \leq K(1 - \pi_i(t))(2 - \pi_i(t)), \quad i = 1, \dots, n \quad (3.13)$$

$$\gamma_\pi^*(t) \leq 2K(1 - \pi_{(1)}(t)). \quad (3.14)$$

PROOF. By analogy with the previous proof, we get

$$\begin{aligned}
\tau_{ii}^\pi(t) &\leq K \left((1 - \pi_i(t))^2 + \sum_{j \neq i} \pi_j^2(t) \right) \leq K \left((1 - \pi_i(t))^2 + \sum_{j \neq i} \pi_j(t) \right) \\
&= K(1 - \pi_i(t))(2 - \pi_i(t))
\end{aligned}$$

as claimed in (3.13), and bringing this estimate into (3.6) leads to

$$\begin{aligned}
\gamma_*^\pi(t) &\leq K \sum_{i=1}^n \pi_i(t) (1 - \pi_i(t)) \\
&= K \left(\pi_{(1)}(t) (1 - \pi_{(1)}(t)) + \sum_{k=2}^n \pi_{(k)}(t) (1 - \pi_{(k)}(t)) \right) \\
&\leq K \left((1 - \pi_{(1)}(t)) + \sum_{k=2}^n \pi_{(k)}(t) \right) = 2K(1 - \pi_{(1)}(t)). \quad \square
\end{aligned}$$

REMARK 3.2. **Portfolio diversification and market volatility as drivers of growth:** Suppose that the market $\mathcal{M}$ of (1.1) and (1.2) satisfies the strong nondegeneracy condition (3.10). Consider a long-only portfolio $\pi(\cdot)$ for which $\pi_{(1)}(t) := \max_{1 \leq i \leq n} \pi_i(t) < 1$ holds for all $t \geq 0$; that is, which never concentrates its holdings in just one asset. The growth rate of such a portfolio will dominate strictly the average of the individual assets' growth rates: we have almost surely

$$\gamma_\pi(t) - \sum_{i=1}^n \pi_i(t) \gamma_i(t) = \gamma_\pi^*(t) \geq \frac{\varepsilon}{2} (1 - \pi_{(1)}(t)) > 0, \quad 0 \leq t < \infty, \quad (3.15)$$

thanks to (1.12) and (3.12). (In particular, if all growth rates $\gamma_i(\cdot) \equiv \gamma(\cdot)$, $i = 1, \dots, n$ are the same, then the growth rate of such a portfolio will dominate strictly this common growth rate.) The more volatile the market (i.e., the higher the $\varepsilon > 0$ in (3.10)) and the more diversified the portfolio (to wit, the higher the lower bound $\eta > 0$ in $1 - \pi_{(1)}(t) \geq$

$\eta, 0 \leq t < \infty$), the bigger the lower bound of (3.15). In other words, as FERNHOLZ and SHAY [1982] were the first to observe: in the presence of sufficient market volatility, even minimal portfolio diversification can significantly enhance growth.

To see how significant such an enhancement can be, let us consider **any fixed-proportion, long-only portfolio** $\pi(\cdot) \equiv \pi$, for some vector $\pi \in \Delta^n$ with $1 - \pi_{(1)} = 1 - \max_{1 \leq i \leq n} \pi_i =: \eta > 0$.

(i) From (3.4) and (3.15) we have the a.s. comparisons

$$\frac{1}{T} \log \left(\frac{V^\pi(T)}{V^\mu(T)} \right) - \sum_{i=1}^n \frac{\pi_i}{T} \log \mu_i(T) = \frac{1}{T} \int_0^T \gamma_\pi^*(t) dt \geq \frac{\varepsilon \eta}{2} > 0,$$

$$\forall T \in (0, \infty).$$

If the market is *coherent* as in Remark 2.1, we conclude from these comparisons that the wealth corresponding to any such fixed-proportion, long-only portfolio grows exponentially and at a rate strictly higher than that of the overall market:

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V^\pi(T)}{V^\mu(T)} \right) \geq \frac{\varepsilon \eta}{2} > 0, \quad \text{a.s.} \quad (3.16)$$

(ii) Similarly, if the long-term growth rates $\lim_{T \rightarrow \infty} (1/T) \log X_i(T) = \gamma_i$ exist a.s. for every $i = 1, \dots, n$, then (1.17) gives the a.s. comparisons

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \log V^\pi(T) \geq \sum_{i=1}^n \pi_i \gamma_i + \frac{\varepsilon \eta}{2} > \sum_{i=1}^n \pi_i \gamma_i.$$

4. Portfolio optimization

We can formulate already some fairly interesting optimization problems.

PROBLEM 4.1 (Quadratic criterion, linear constraint (Markowitz [1952])). Minimize the portfolio variance $a_{\pi\pi}(t) = \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t)$, among all portfolios $\pi(\cdot)$ with rate of return $b_\pi(t) = \sum_{i=1}^n \pi_i(t) b_i(t) \geq b_0$ greater than, or equal to, a given constant $b_0 \in \mathbb{R}$.

PROBLEM 4.2 (Quadratic criterion, quadratic constraint). Minimize the portfolio variance

$$a_{\pi\pi}(t) = \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t)$$

among all portfolios $\pi(\cdot)$ with growth rate at least equal to a given constant γ_0 , namely,

$$\sum_{i=1}^n \pi_i(t) \left(\gamma_i(t) + \frac{1}{2} a_{ii}(t) \right) \geq \gamma_0 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) a_{ij}(t) \pi_j(t).$$

PROBLEM 4.3. *Maximize, over long-only portfolios $\pi(\cdot)$, the probability of reaching a given “ceiling” $\mathfrak{c}$ before reaching a given “floor” $\mathfrak{f}$, with $0 < \mathfrak{f} < w < \mathfrak{c} < \infty$. More specifically, maximize the probability $\mathbb{P}[\mathfrak{T}_{\mathfrak{c}}^{\pi} < \mathfrak{T}_{\mathfrak{f}}^{\pi}]$, with the notation $\mathfrak{T}_{\xi}^{\pi} := \inf\{t \geq 0 \mid V^{w,\pi}(t) = \xi\}$ for $\xi \in (0, \infty)$.*

In the case of constant coefficients γ_i and a_{ij} , the solution to this problem comes in the following simple form: one looks at the mean-variance, or *signal-to-noise*, ratio

$$\frac{\gamma_{\pi}}{a_{\pi\pi}} = \frac{\sum_{i=1}^n \pi_i (\gamma_i + \frac{1}{2} a_{ii})}{\sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} \pi_j} - \frac{1}{2},$$

and finds a vector $\pi \in \Delta^n$ that maximizes it (PESTIEN and SUDDERTH, [1985]).

PROBLEM 4.4. *Minimize, over long-only portfolios $\pi(\cdot)$, the expected time $\mathbb{E}(\mathfrak{T}_{\mathfrak{c}}^{\pi})$ until a given “ceiling” $\mathfrak{c} \in (w, \infty)$ is reached.*

Again with constant coefficients, it turns out that it is enough to maximize the drift in the equation for $\log V^{w,\pi}(\cdot)$, namely

$$\gamma_{\pi} = \sum_{i=1}^n \pi_i \left(\gamma_i + \frac{1}{2} a_{ii} \right) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} \pi_j,$$

the portfolio growth rate (HEATH, OREY, PESTIEN and SUDDERTH, 1987), over vectors $\pi \in \Delta^n$.

PROBLEM 4.5. *Maximize, over portfolios $\pi(\cdot)$, the probability $\mathbb{P}[\mathfrak{T}_{\mathfrak{c}}^{\pi} < T \wedge \mathfrak{T}_{\mathfrak{f}}^{\pi}]$ of reaching a given “ceiling” $\mathfrak{c}$ before reaching a given “floor” $\mathfrak{f}$ with $0 < \mathfrak{f} < w < \mathfrak{c} < \infty$, by a given “deadline” $T \in (0, \infty)$.*

Always with constant coefficients, suppose there is a vector $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_n)'$ that maximizes both the signal-to-noise ratio and the variance,

$$\frac{\gamma_{\pi}}{a_{\pi\pi}} = \frac{\sum_{i=1}^n \pi_i (\gamma_i + \frac{1}{2} a_{ii})}{\sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} \pi_j} - \frac{1}{2} \quad \text{and} \quad a_{\pi\pi} = \sum_{i=1}^n \sum_{j=1}^n \pi_i a_{ij} \pi_j,$$

respectively, over all vectors $(\pi_1, \dots, \pi_n)$ that satisfy $\sum_{i=1}^n \pi_i = 1$ (as well as $\pi_i \geq 0, \dots, \pi_n \geq 0$ if we restrict ourselves to long-only portfolios). Then the resulting constant-proportion portfolio $\hat{\pi}(\cdot) \equiv \hat{\pi}$ is optimal for the above criterion (SUDDERTH and WEERASINGHE, [1989]).

This is a big assumption; it is satisfied, for instance, under the (very stringent and unnatural, etc.) condition that for some real number $b \leq 0$, we have

$$b_i = \gamma_i + \frac{1}{2} a_{ii} = b, \quad \text{for all } i = 1, \dots, n.$$

As far as the authors are aware, nobody seems to have solved this problem when such simultaneous maximization is not possible.

PROBLEM 4.6 (The growth-optimal portfolio). Suppose we can find a portfolio $\hat{\pi}(\cdot)$ such that with probability one: for each $t \in [0, \infty)$, the vector $\hat{\pi}(t)$ maximizes the expression

$$\sum_{i=1}^n x_i \gamma_i(t) + \frac{1}{2} \left(\sum_{i=1}^n x_i a_{ii}(t) - \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij}(t) x_j \right) = x' b(t) - \frac{1}{2} x' a(t) x \quad (4.1)$$

over all vectors $(x_1, \dots, x_n) \in \mathbb{R}^n$ with $\sum_{i=1}^n x_i = 1$. In particular, this vector has to satisfy the first-order condition associated with this maximization, namely,

$$(x - \hat{\pi}(t))' (b(t) - a(t)\hat{\pi}(t)) \leq 0, \quad \text{for every vector } (x_1, \dots, x_n) \in \mathbb{R}^n \\ \text{with } \sum_{i=1}^n x_i = 1. \quad (4.2)$$

It is clear then that for any portfolio $\pi(\cdot)$, we have the a.s. comparison

$$\gamma_\pi(t) \leq \gamma_{\hat{\pi}}(t), \quad \forall 0 \leq t < \infty \quad (4.3)$$

of growth rates. If (1.15) is satisfied (e.g., if (1.16) holds), then the consequence

$$d \log \left(\frac{V^\pi(t)}{V^{\hat{\pi}}(t)} \right) = (\gamma_\pi(t) - \gamma_{\hat{\pi}}(t)) dt + \sum_{v=1}^d (\sigma_{\pi v}(t) - \sigma_{\hat{\pi} v}(t)) dW_v(t) \quad (4.4)$$

of (1.11) leads to the growth-optimality property

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V^\pi(T)}{V^{\hat{\pi}}(T)} \right) \leq 0 \quad \text{a.s., for every portfolio } \pi(\cdot); \quad (4.5)$$

and if for some $\mathbb{F}$ -stopping time T we have $\mathbb{E} \int_0^T \|a(t)\| dt < \infty$, then (4.3) and (4.4) lead to the *log-optimality property*

$$\mathbb{E}(\log V^\pi(T)) \leq \mathbb{E}(\log V^{\hat{\pi}}(T)), \quad \text{for every portfolio } \pi(\cdot). \quad (4.6)$$

There is more one can say: denoting by $\hat{\mathcal{R}}^\pi(\cdot) := V^\pi(\cdot)/V^{\hat{\pi}}(\cdot)$ the process of (4.4), an application of Itô's rule gives

$$\begin{aligned} \frac{d\hat{\mathcal{R}}^\pi(t)}{\hat{\mathcal{R}}^\pi(t)} &= \left[\gamma_\pi(t) - \gamma_{\hat{\pi}}(t) + \frac{1}{2} \sum_{v=1}^d (\sigma_{\pi v}(t) - \sigma_{\hat{\pi} v}(t))^2 \right] dt \\ &\quad + \sum_{v=1}^d (\sigma_{\pi v}(t) - \sigma_{\hat{\pi} v}(t)) dW_v(t) \\ &= (\pi(t) - \hat{\pi}(t))' \left[(b(t) - a(t)\hat{\pi}(t)) dt + \sigma(t) dW(t) \right]. \end{aligned}$$

In conjunction with the first-order condition of (4.2), this semimartingale decomposition shows that $\hat{\mathcal{R}}^\pi(\cdot)$ is a local supermartingale. Because it is positive, this process is, therefore, a supermartingale, by Fatou's lemma. We obtain the numéraire property of the growth-optimal portfolio $\hat{\pi}(\cdot)$:

$$\hat{\mathcal{R}}^\pi(\cdot) = V^\pi(\cdot) / V^{\hat{\pi}}(\cdot) \quad \text{is a supermartingale, for every portfolio } \pi(\cdot). \quad (4.7)$$

Diversity & Arbitrage

Roughly speaking, a market is *diverse* if it avoids concentrating all its capital into a single stock, and the *diversity* of a market is a measure of how uniformly the capital is spread among the stocks. These concepts were introduced by FERNHOLZ [1999]; it was shown by FERNHOLZ [2002], section 3.3, and by FERNHOLZ, KARATZAS and KARDARAS [2005] that market diversity gives rise to arbitrage. Diversity is a concept that is meaningful for equity markets but probably not for more general classes of assets. Nevertheless, some of the results in this chapter may be relevant for passive portfolios comprising more general types of assets.

Unlike classical mathematical finance, SPT is not averse to the existence of arbitrage in markets but rather studies the market characteristics that imply the existence of arbitrage. Moreover, it shows that the existence of arbitrage does not preclude the development of option pricing theory or certain types of utility maximization. These and other related ideas are presented in this chapter.

5. Diversity

The notion of diversity for a financial market corresponds to the intuitive (and descriptive) idea that no single company can ever be allowed to dominate the entire market in terms of relative capitalization. To make this notion precise, let us say that the model $\mathcal{M}$ of (1.1), (1.2) is *diverse* on the time horizon $[0, T]$, with $T > 0$ a given real number, if there exists a number $\delta \in (0, 1)$ such that the quantities of (2.1) satisfy almost surely

$$\max_{1 \leq i \leq n} \mu_i(t) =: \mu_{(1)}(t) < 1 - \delta, \quad \forall 0 \leq t \leq T \quad (5.1)$$

in the order-statistics notation of (1.18). In a similar vein, we say that $\mathcal{M}$ is *weakly diverse* on the time horizon $[0, T]$, if for some $\delta \in (0, 1)$, we have

$$\frac{1}{T} \int_0^T \mu_{(1)}(t) dt < 1 - \delta, \quad \text{a.s.} \quad (5.2)$$

We say that $\mathcal{M}$ is *uniformly weakly diverse* on $[T_0, \infty)$, for some real number $T_0 > 0$, if there exists a number $\delta \in (0, 1)$ such that (5.2) holds for every $T \in [T_0, \infty)$.

It follows directly from (3.14) of Lemma 3.5 that, under the uniform boundedness condition (1.16), the model $\mathcal{M}$ of (1.1), (1.2) is diverse (respectively, weakly diverse)

on the time-horizon $[0, T]$ if there exists a number $\zeta > 0$ such that

$$\gamma_\mu^*(t) \geq \zeta, \quad \forall 0 \leq t \leq T \quad \left(\text{respectively, } \frac{1}{T} \int_0^T \gamma_\mu^*(t) dt \geq \zeta \right) \quad (5.3)$$

holds almost surely. And (3.12) of Lemma 3.4 shows that, under the strong nondegeneracy condition (3.10), the first (respectively, the second) inequality of (5.3) is satisfied if diversity (respectively, weak diversity) holds on the time interval $[0, T]$.

As we shall see in Section 9, diversity can be ensured by a strongly negative rate of growth for the largest stock, resulting in a sufficiently strong repelling drift (e.g., a log-pole-type singularity) away from an appropriate boundary, as well as nonnegative growth rates for all the other stocks.

If all the stocks in $\mathcal{M}$ have the same growth rate, $(\gamma_i(\cdot) \equiv \gamma(\cdot), \forall 1 \leq i \leq n)$, and (1.15) holds, then we have almost surely:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \gamma_\mu^*(t) dt = 0. \quad (5.4)$$

In particular, such an equal-growth-rate market $\mathcal{M}$ cannot be diverse, even weakly, over long time horizons, provided that (3.10) is also satisfied.

Here is a quick argument for these claims: recall that for $X(\cdot) = X_1(\cdot) + \cdots + X_n(\cdot)$, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\log X(T) - \int_0^T \gamma_\mu(t) dt \right) = 0, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \left(\log X_i(T) - \int_0^T \gamma(t) dt \right) = 0$$

a.s., from (1.14), (1.6), and $\gamma_i(\cdot) \equiv \gamma(\cdot)$ for all $1 \leq i \leq n$. But then, we have also

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\log X_{(1)}(T) - \int_0^T \gamma(t) dt \right) = 0, \quad \text{a.s.}$$

for the biggest stock $X_{(1)}(\cdot) := \max_{1 \leq i \leq n} X_i(\cdot)$, and note the inequalities $X_{(1)}(\cdot) \leq X(\cdot) \leq nX_{(1)}(\cdot)$. Therefore,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\log X_{(1)}(T) - \log X(T) \right) = 0, \quad \text{thus} \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\gamma_\mu(t) - \gamma(t)) dt = 0,$$

almost surely. But $\gamma_\mu(t) = \sum_{i=1}^n \mu_i(t) \gamma(t) + \gamma_\mu^*(t) = \gamma(t) + \gamma_\mu^*(t)$ because of the assumption of equal growth rates, and (5.4) follows. If (3.10) also holds, then (3.12) and (5.4) imply

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (1 - \mu_{(1)}(t)) dt = 0$$

almost surely, so weak diversity fails on long-time horizons: once in a while, a single stock dominates the entire market, then recedes; sooner or later another stock takes its place as absolutely dominant leader, and so on.

REMARK 5.1. If all the stocks in the market $\mathcal{M}$ have constant (though not necessarily the same) growth rates and if (1.16) and (3.10) hold, then $\mathcal{M}$ cannot be diverse, even weakly, over long-time horizons.

6. Relative arbitrage and its consequences

The notion of arbitrage is of paramount importance in mathematical finance. We present in this section an allied notion, that of relative arbitrage, and explore some of its consequences. In later sections, we shall encounter specific, descriptive conditions on market structure that lead to this form of arbitrage. Relative arbitrage, although discussed here in the context of equity markets, is a concept that remains meaningful for general classes of assets.

DEFINITION 6.1. Given any two portfolios $\pi(\cdot)$ and $\rho(\cdot)$ with the same initial capital $V^\pi(0) = V^\rho(0) = 1$, we shall say that $\pi(\cdot)$ represents an arbitrage opportunity (respectively, a strong arbitrage opportunity) relative to $\rho(\cdot)$ over the time horizon $[0, T]$, with $T > 0$ a given real number, if

$$\mathbb{P}(V^\pi(T) \geq V^\rho(T)) = 1 \quad \text{and} \quad \mathbb{P}(V^\pi(T) > V^\rho(T)) > 0 \quad (6.1)$$

(respectively, if $\mathbb{P}(V^\pi(T) > V^\rho(T)) = 1$) holds. We shall say that $\pi(\cdot)$ represents a superior long-term growth opportunity relative to $\rho(\cdot)$ if

$$\mathcal{L}^{\pi, \rho} := \liminf_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V^\pi(T)}{V^\rho(T)} \right) > 0 \quad \text{holds a.s.} \quad (6.2)$$

(Recall here the comparison of (3.16).)

REMARK 6.1. The definition of relative arbitrage has historically included the condition that there exists a constant $q = q_{\pi, \rho, T} > 0$ such that

$$\mathbb{P}(V^\pi(t) \geq qV^\rho(t), \forall 0 \leq t \leq T) = 1. \quad (6.3)$$

However, if one can find a portfolio $\pi(\cdot)$ that satisfies the domination properties (6.1) relative to some other portfolio $\rho(\cdot)$, then there exists another portfolio $\tilde{\pi}(\cdot)$ that satisfies both (6.3) and (6.1) relative to the same $\rho(\cdot)$. The construction involves a strategy of investing a portion $w \in (0, 1)$ of the initial capital \$1 in $\pi(\cdot)$, and the remaining portion $1 - w$ in $\rho(\cdot)$. This observation is due to KARDARAS [2006].

6.1. Strict local martingales

Let us place ourselves now, and for the remainder of this section, within the market model $\mathcal{M}$ of (1.1) under the conditions (1.2). We shall assume further that there exists

a market price of risk (or “relative risk”) $\theta : [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$; namely, an $\mathbb{F}$ -progressively measurable process with

$$\sigma(t)\theta(t) = b(t) - r(t)\mathbf{I}, \quad \forall 0 \leq t \leq T \quad \text{and} \quad \int_0^T \|\theta(t)\|^2 dt < \infty \quad (6.4)$$

valid almost surely, for each $T \in (0, \infty)$. (If the volatility matrix $\sigma(\cdot)$ has full rank, namely, n , we can take, for instance, $\theta(t) = \sigma'(t)(\sigma(t)\sigma'(t))^{-1}[b(t) - r(t)\mathbf{I}]$ in (6.4).)

In terms of this process $\theta(\cdot)$, we can define the exponential local martingale and supermartingale

$$Z(t) := \exp\left\{-\int_0^t \theta'(s) dW(s) - \frac{1}{2} \int_0^t \|\theta(s)\|^2 ds\right\}, \quad 0 \leq t < \infty \quad (6.5)$$

(a martingale, if and only if $\mathbb{E}(Z(T)) = 1$, $\forall T \in (0, \infty)$) and the shifted Brownian motion

$$\hat{W}(t) := W(t) + \int_0^t \theta(s) ds, \quad 0 \leq t < \infty. \quad (6.6)$$

PROPOSITION 6.1. A Strict Local Martingale: *Under the assumptions of this subsection, as well as (1.16), suppose that for some real number $T > 0$ and for some portfolio $\rho(\cdot)$ there exists arbitrage relative to $\rho(\cdot)$ on the time horizon $[0, T]$. Then, the process $Z(\cdot)$ of (6.5) is a strict local martingale: $\mathbb{E}(Z(T)) < 1$.*

PROOF. Assume, by way of contradiction, that $\mathbb{E}(Z(T)) = 1$. Then, from the Girsanov theorem (KARATZAS and SHREVE [1991], Section 3.5), the recipe $\mathbb{Q}_T(A) := \mathbb{E}[Z(T) 1_A]$, $A \in \mathcal{F}(T)$ defines a probability measure, equivalent to $\mathbb{P}$, under which the process $\hat{W}(t)$, $0 \leq t \leq T$ as in (6.6) is Brownian motion.

Under this probability measure $\mathbb{Q}_T$, the discounted stock prices $X_i(\cdot)/B(\cdot)$, $i = 1, \dots, n$ are positive martingales on $[0, T]$, because of

$$d(X_i(t)/B(t)) = (X_i(t)/B(t)) \sum_{v=1}^d \sigma_{iv}(t) d\hat{W}_v(t)$$

and of the uniform boundedness condition (1.16). As usual, we express this by saying that $\mathbb{Q}_T$ is then an EMM for the model on the given time horizon $[0, T]$.

More generally, for any portfolio $\pi(\cdot)$, we get from (6.6) and (1.9),

$$d(V^\pi(t)/B(t)) = (V^\pi(t)/B(t)) \pi'(t)\sigma(t) d\hat{W}(t), \quad V^\pi(0) = 1, \quad (6.7)$$

and from (1.16), the discounted wealth process $V^\pi(t)/B(t)$, $0 \leq t \leq T$ is a positive martingale under $\mathbb{Q}_T$. Thus, the difference $\Delta(t) := (V^\pi(t) - V^\rho(t))/B(t)$, $0 \leq t \leq T$ is a martingale under $\mathbb{Q}_T$ for any other portfolio $\rho(\cdot)$ with $V^\rho(0) = 1$; consequently, $\mathbb{E}^{\mathbb{Q}_T}(\Delta(T)) = \Delta(0) = 0$. But this conclusion is inconsistent with (6.1), which mandates $\mathbb{Q}_T(\Delta(T) \geq 0) = 1$ and $\mathbb{Q}_T(\Delta(T) > 0) > 0$. $\square$

Now let us consider the deflated stock price and wealth processes

$$\begin{aligned}\hat{X}_i(t) &:= \frac{Z(t)}{B(t)} X_i(t), \quad i = 1, \dots, n, \quad \hat{X}(t) := \frac{Z(t)}{B(t)} X(t) \quad \text{and} \\ \hat{V}^{w,h}(t) &:= \frac{Z(t)}{B(t)} V^{w,h}(t)\end{aligned}\tag{6.8}$$

for $0 \leq t < \infty$ for an arbitrary trading strategy $h(\cdot) \in \mathcal{H}(w)$ admissible for the initial capital $w > 0$. These processes satisfy, respectively, the dynamics

$$d\hat{X}_i(t) = \hat{X}_i(t) \sum_{v=1}^d (\sigma_{iv}(t) - \theta_v(t)) dW_v(t), \quad \hat{X}_i(0) = x_i,\tag{6.9}$$

$$\begin{aligned}d\hat{X}(t) &= \hat{X}(t) \sum_{v=1}^d (\sigma_{\mu v}(t) - \theta_v(t)) dW_v(t), \quad \hat{X}(0) = \sum_{i=1}^n x_i, \\ d\hat{V}^{w,h}(t) &= \left(\frac{Z(t)h'(t)}{B(t)} \sigma(t) - \hat{V}^{w,h}(t)\theta'(t) \right) dW(t), \quad \hat{V}^{w,h}(0) = w\end{aligned}\tag{6.10}$$

in conjunction with (1.1), (1.22) and (6.5). In particular, these processes are nonnegative local martingales (and supermartingales) under $\mathbb{P}$.

In other words, the ratio $Z(\cdot)/B(\cdot)$ continues to play its usual role as deflator of prices in such a market, even when $Z(\cdot)$ is just a local martingale.

REMARK 6.2. Strict Local Martingales Galore: In the setting of Proposition 6.1 with $\rho(\cdot) \equiv \mu(\cdot)$, the market portfolio, it can be shown from (6.9), (6.10) that the deflated stock-price processes $\hat{X}_i(t)$, $0 \leq t \leq T$ of (6.8) are all strict local martingales and (strict) supermartingales:

$$\mathbb{E}(\hat{X}_i(T)) < x_i \quad \text{holds for every } i = 1, \dots, n.\tag{6.11}$$

We shall prove this property based on a more general result. *Under the assumptions of this subsection, suppose that for some real number $T > 0$ and for some portfolio $\rho(\cdot)$, there exists arbitrage relative to $\rho(\cdot)$, on the time horizon $[0, T]$. Then, the process $\hat{V}^{w,\rho}(t) := Z(t)V^{w,\rho}(t)/B(t)$, $0 \leq t \leq T$, defined as in (6.8), is a strict local martingale and a strict supermartingale, namely,*

$$\mathbb{E}(\hat{V}^{w,\rho}(T)) < w.\tag{6.12}$$

PROPOSITION 6.2. Non existence of Equivalent Martingale Measure: *In the context of Proposition 6.1, no EMM can exist for the model $\mathcal{M}$ of (1.1) on $[0, T]$, if the filtration is generated by the driving Brownian motion $W(\cdot) : \mathbb{F} = \mathbb{F}^W$.*

PROOF. If $\mathbb{F} = \mathbb{F}^W$, and if the probability measure $\mathbb{Q}$ is equivalent to $\mathbb{P}$ on $\mathcal{F}(T)$, the martingale representation property of the Brownian filtration gives $(d\mathbb{Q}/d\mathbb{P})|_{\mathcal{F}(t)} =$

$Z(t)$, $0 \leq t \leq T$ for some process $Z(\cdot)$ of the form (6.5) and some progressively measurable $\theta(\cdot)$ with $\int_0^T \|\theta(t)\|^2 dt < \infty$ a.s. Then, Itô's rule leads to the extension

$$\frac{d\hat{X}_i(t)}{\hat{X}_i(t)} = \left(b_i(t) - r(t) - \sum_{v=1}^d \sigma_{iv}(t)\theta_v(t) \right) dt + \sum_{v=1}^d (\sigma_{iv}(t) - \theta_v(t)) dW_v(t)$$

of (6.9) for the deflated stock-prices of (6.8).

But if $\mathbb{Q}$ is an EMM (i.e., if all the $X_i(\cdot)/B(\cdot)$'s are $\mathbb{Q}$ -martingales on $[0, T]$), then the $\hat{X}_i(\cdot)$'s are all $\mathbb{P}$ -martingales on $[0, T]$, and this leads to the first property $\sigma(t)\theta(t) = b(t) - r(t)\mathbf{I}$, $0 \leq t \leq T$ in (6.4). We repeat now the argument of Proposition 6.1 and arrive at a contradiction with (6.1), the existence of relative arbitrage on $[0, T]$. $\square$

6.2. On "Beating the Market"

Let us introduce now the nonincreasing, right-continuous function

$$f(t) := \frac{1}{X(0)} \cdot \mathbb{E} \left(\frac{Z(t)}{B(t)} X(t) \right), \quad 0 \leq t < \infty. \quad (6.13)$$

If relative arbitrage exists on the time horizon $[0, T]$, with $T > 0$, a real number, then we know $f(0) = 1 > f(T) > 0$ from Remark 6.2.

REMARK 6.3. With Brownian filtration $\mathbb{F} = \mathbb{F}^W$, with $n = d$ (equal numbers of stocks and driving Brownian motions), and with an invertible volatility matrix $\sigma(\cdot)$, consider the maximal relative return

$$\mathfrak{R}(T) := \sup \{ r > 0 \mid \exists h(\cdot) \in \mathcal{H}(1; T) \text{ s.t. } (\mathcal{V}^h(T)/V^\mu(T)) \geq r, \text{ a.s.} \} \quad (6.14)$$

in excess of the market that can be obtained by trading strategies over the interval $[0, T]$. It can be shown that this quantity is computed in terms of the function of (6.13), as $\mathfrak{R}(T) = 1/f(T)$.

REMARK 6.4. **The shortest time to beat the market by a given amount:** Let us place ourselves again under the assumptions of Remark 6.3 and assume that relative arbitrage exists on $[0, T]$ for every $T \in (0, \infty)$ (see Section 8 for elaboration). For a given "exceedance level" $r > 1$, consider the shortest length of time

$$\mathbf{T}(r) := \inf \{ T \in (0, \infty) \mid \exists h(\cdot) \in \mathcal{H}(1; T) \text{ s.t. } (\mathcal{V}^h(T)/V^\mu(T)) \geq r, \text{ a.s.} \} \quad (6.15)$$

required to guarantee a return of at least r times the market. It can be shown that this quantity is given by the inverse of the decreasing function $f(\cdot)$ of (6.13) evaluated at $1/r$:

$$\mathbf{T}(r) = \inf \{ T \in (0, \infty) \mid f(T) \leq 1/r \}. \quad (6.16)$$

A detailed argument is presented at the end of subsection 10.1.

Question: Can the counterparts of (6.14) and (6.15) be computed when one is not allowed to use general strategies $h(\cdot) \in \mathcal{H}(1; T)$, but rather long-only portfolios $\pi(\cdot)$?

REMARK 6.5. It is not possible to construct arbitrage relative to the growth-optimal portfolio $\hat{\pi}(\cdot)$ of Problem 4.6 in Section 4, on any given time horizon $[0, T]$, with $T > 0$ a real number. For if such relative arbitrage $\pi(\cdot)$ existed, we would have

$$\mathbb{P}[\hat{\mathcal{R}}^\pi(T) \geq 1] = 1 \quad \text{and} \quad \mathbb{P}[\hat{\mathcal{R}}^\pi(T) > 1] > 0$$

in the notation of (4.7), thus also $\mathbb{E}[\hat{\mathcal{R}}^\pi(T)] > 1$; but this contradicts the numéraire property (4.7) of the growth-optimal portfolio, which implies $\mathbb{E}[\hat{\mathcal{R}}^\pi(T)] \leq 1$ for every real number $T > 0$. We owe this observation to KARDARAS [2006].

In a similar vein, suppose that $u : [0, \infty) \rightarrow [0, \infty)$ is a strictly increasing function and that, for some real number $T > 0$ and some portfolio $\rho(\cdot)$, we have the comparison

$$\mathbb{E}[u(V^\pi(T))] \leq \mathbb{E}[u(V^\rho(T))] \quad \text{for every portfolio } \pi(\cdot). \quad (6.17)$$

Then, it is not possible to construct arbitrage relative to this $\rho(\cdot)$ on the given time horizon $[0, T]$; for otherwise there would exist a portfolio $\tilde{\pi}(\cdot)$ with the properties of (6.1), thus also with the property $\mathbb{E}[u(V^{\tilde{\pi}}(T))] > \mathbb{E}[u(V^\rho(T))]$ which contradicts (6.17).

Proof of (6.12). We shall employ the usual notation $V^{w,\rho}(\cdot) = wV^\rho(\cdot)$ and $\hat{V}^{w,\rho}(\cdot)$ for the wealth and the deflated wealth, respectively, of our given portfolio $\rho(\cdot)$ with initial capital $w > 0$. Setting

$$h(\cdot) := V^{w,\rho}(\cdot)\rho(\cdot) \quad \text{and} \quad \theta^\rho(\cdot) := \sigma'(\cdot)\rho(\cdot) - \theta(\cdot),$$

the equation (6.10) takes the form $d\hat{V}^{w,\rho}(t) = \hat{V}^{w,\rho}(t)(\theta^\rho(t))' dW(t)$, or equivalently

$$\hat{V}^{w,\rho}(t) = w \cdot \hat{V}^\rho(t) = w \cdot \exp\left\{\int_0^t (\theta^\rho(s))' dW(s) - \frac{1}{2} \int_0^t \|\theta^\rho(s)\|^2 ds\right\}, \quad 0 \leq t \leq T. \quad (6.18)$$

On the other hand, introducing the process

$$\tilde{W}^{(\rho)}(t) := W(t) - \int_0^t \theta^\rho(s) ds = \hat{W}(t) - \int_0^t \sigma'(s)\rho(s) ds, \quad 0 \leq t \leq T, \quad (6.19)$$

we obtain

$$(\hat{V}^{w,\rho}(t))^{-1} = w^{-1} \cdot \exp\left\{-\int_0^t (\theta^\rho(s))' d\tilde{W}^{(\rho)}(s) - \frac{1}{2} \int_0^t \|\theta^\rho(s)\|^2 ds\right\}. \quad (6.20)$$

We shall argue (6.12) by contradiction: let us assume that it fails, namely, that $\hat{V}^{w,\rho}(\cdot)$ of (6.18) is a martingale. From (6.18) and the Girsanov theorem, the process $\tilde{W}^{(\rho)}(\cdot)$ of (6.19) is then a Brownian motion under the probability measure

$\tilde{\mathbb{P}}_T^{(\rho)}(A) := \mathbb{E}(\hat{V}^{w,\rho}(T) 1_A) / w$, $A \in \mathcal{F}(T)$, which is equivalent to $\mathbb{P}$. Then, Itô's rule gives

$$d\left(\frac{V^\pi(t)}{V^\rho(t)}\right) = \left(\frac{V^\pi(t)}{V^\rho(t)}\right) \cdot \sum_{k=1}^n \sum_{v=1}^d (\pi_k(t) - \rho_k(t)) \sigma_{kv}(t) d\tilde{W}_v^{(\rho)}(t) \quad (6.21)$$

for any portfolio $\pi(\cdot)$, in conjunction with (6.7), (6.20), and (6.19); the ratio $V^\pi(\cdot)/V^\rho(\cdot)$ is seen to be a positive local martingale and supermartingale, under $\tilde{\mathbb{P}}_T^{(\rho)}$. In particular, we obtain $\tilde{\mathbb{E}}_T^{(\rho)}(V^\pi(T)/V^\rho(T)) \leq 1$, where $\tilde{\mathbb{E}}_T^{(\rho)}$ denotes expectation with respect to $\tilde{\mathbb{P}}_T^{(\rho)}$.

Now consider any portfolio $\pi(\cdot)$ that satisfies the conditions of (6.1) on the time horizon $[0, T]$, relative to $\rho(\cdot)$; such a portfolio exists by assumption. The first condition in (6.1) gives the comparison $\tilde{\mathbb{P}}_T^{(\rho)}(V^\pi(T) \geq V^\rho(T)) = 1$. In conjunction with the inequality $\tilde{\mathbb{E}}_T^{(\rho)}(V^\pi(T)/V^\rho(T)) \leq 1$ just proved, we obtain the equality $\tilde{\mathbb{P}}_T^{(\rho)}(V^\pi(T) = V^\rho(T)) = 1$ or equivalently

$$\mathbb{P}(V^\pi(T) = V^\rho(T)) = 1 \quad \text{for every portfolio } \pi(\cdot) \text{ that satisfies the first condition in (6.1).}$$

But this contradicts the second condition $\mathbb{P}(V^\pi(T) > V^\rho(T)) > 0$ of (6.1). $\square$

Proof of (6.11). From what has already been shown (for (6.12), now applied to the market portfolio), the process $V^{x,\mu}(\cdot) \equiv \hat{X}(\cdot) = \hat{X}_1(\cdot) + \cdots + \hat{X}_n(\cdot)$ is a strict local martingale and a strict supermartingale. Now, each $\hat{X}_i(\cdot)$ is a positive local (and super-) martingale, so there must exist at least one $j \in \{1, \dots, n\}$ for which $\hat{X}_j(\cdot)$ is a strict local martingale and a strict supermartingale.

We shall argue once again by contradiction: suppose that (6.11) fails, to wit, that $\hat{X}_i(\cdot)$ is a martingale for some $i \neq j$. Then (6.21) with $\rho(\cdot) \equiv e_i$ and $\pi(\cdot) \equiv e_j$ gives

$$d\left(\frac{X_j(t)}{X_i(t)}\right) = \left(\frac{X_j(t)}{X_i(t)}\right) \cdot \sum_{v=1}^d (\sigma_{jv}(t) - \sigma_{iv}(t)) d\tilde{W}_v^{(e_i)}(t),$$

so condition (1.16) implies that $X_j(\cdot)/X_i(\cdot)$ is a $\tilde{\mathbb{P}}_T^{(e_i)}$ -martingale on $[0, T]$. In particular, we get

$$\frac{x_j}{x_i} = \tilde{\mathbb{E}}_T^{(e_i)} \left[\frac{X_j(T)}{X_i(T)} \right] = \mathbb{E} \left[\frac{Z(T) X_j(T)}{B(T) x_i} \right],$$

which contradicts the strict supermartingale property of $\hat{X}_j(\cdot) = Z(\cdot)X_j(\cdot)/B(\cdot)$ and proves (6.11). $\square$

7. Diversity leads to arbitrage

We provide examples that demonstrate the following principle: *If the model $\mathcal{M}$ of (1.1) and (1.2) is weakly diverse over the time horizon $[0, T]$, and if (3.10) holds, then $\mathcal{M}$*

contains strong arbitrage opportunities relative to the market portfolio, at least for sufficiently large real numbers $T > 0$.

The first such examples involve heavily the diversity-weighted portfolio $\mu^{(p)}(\cdot) = (\mu_1^{(p)}(\cdot), \dots, \mu_n^{(p)}(\cdot))'$ defined, for some arbitrary but fixed $p \in (0, 1)$, in terms of the market portfolio $\mu(\cdot)$ of (2.1) by

$$\mu_i^{(p)}(t) := \frac{(\mu_i(t))^p}{\sum_{j=1}^n (\mu_j(t))^p}, \quad \forall i = 1, \dots, n. \quad (7.1)$$

Compared to $\mu(\cdot)$, the portfolio $\mu^{(p)}(\cdot)$ in (7.1) decreases the proportion(s) held in the largest stock(s) and increases those placed in the smallest stock(s), while preserving the relative rankings of all stocks (see (7.7)). It does this in a systematic and “passive” way, that involves neither parameter estimation nor optimization. The actual performance of this portfolio relative to the S&P 500 index over a 33-year period is discussed in detail by FERNHOLZ [2002], chapter 7).

We show below that if the model $\mathcal{M}$ is weakly diverse on a time horizon $[0, T]$, with $T > 0$ a given real number, then the value process $V^{\mu^{(p)}}(\cdot)$ of the diversity-weighted portfolio in (7.1) satisfies

$$V^{\mu^{(p)}}(T) > V^\mu(T) \left(n^{-1/p} e^{\varepsilon \delta T/2} \right)^{1-p} \quad (7.2)$$

almost surely. In particular,

$$\mathbb{P} \left(V^{\mu^{(p)}}(T) > V^\mu(T) \right) = 1, \quad \text{provided that } T \geq \frac{2}{p\varepsilon\delta} \log n, \quad (7.3)$$

and $\mu^{(p)}(\cdot)$ is a strong arbitrage opportunity relative to the market $\mu(\cdot)$, in the sense of (6.1). The significance of such a result for practical long-term portfolio management cannot be overstated.

Proof of (7.3). Let us start by introducing the function

$$\mathbf{G}_p(x) := \left(\sum_{i=1}^n x_i^p \right)^{1/p}, \quad x \in \Delta_+^n, \quad (7.4)$$

which we shall interpret as a “measure of diversity” (see below). An application of Itô’s rule to the process $\{\mathbf{G}_p(\mu(t)), 0 \leq t < \infty\}$ leads after some computation, and in conjunction with (3.9) and the numéraire-invariance property (3.5), to the expression

$$\log \left(\frac{V^{\mu^{(p)}}(T)}{V^\mu(T)} \right) = \log \left(\frac{\mathbf{G}_p(\mu(T))}{\mathbf{G}_p(\mu(0))} \right) + (1-p) \int_0^T \gamma_{\mu^{(p)}}^*(t) dt, \quad \text{a.s.} \quad (7.5)$$

for the wealth $V^{\mu^{(p)}}(\cdot)$ of the diversity-weighted portfolio $\mu^{(p)}(\cdot)$ of (7.1) (see also Section 11, particularly (11.2) and its proof). One big advantage of the expression (7.5)

is that it is free of stochastic integrals and thus lends itself to pathwise (almost sure) comparisons.

For the function of (7.4), we have the simple bounds

$$1 = \sum_{i=1}^n \mu_i(t) \leq \sum_{i=1}^n (\mu_i(t))^p = (\mathbf{G}_p(\mu(t)))^p \leq n^{1-p}.$$

In other words, the minimum of $\mathbf{G}_p(\mu(t))$ occurs when the entire market is concentrated in one stock ($\mu_j(t) = 1$ for some $j \in \{1, \dots, n\}$), and its maximum when all stocks have the same capitalization ($\mu_1(t) = \dots = \mu_n(t) = 1/n$); this justifies considering the function of (7.4) as a measure of diversity. We deduce the comparison

$$\log\left(\frac{\mathbf{G}_p(\mu(T))}{\mathbf{G}_p(\mu(0))}\right) \geq -\frac{1-p}{p} \log n, \quad \text{a.s.} \quad (7.6)$$

which, coupled with (7.5) and (3.7), shows that $V^{\mu^{(p)}}(\cdot)/V^\mu(\cdot)$ is bounded from below by the constant $n^{-(1-p)/p}$. In particular, (6.3) is satisfied for $\rho(\cdot) \equiv \mu(\cdot)$ and $\pi(\cdot) \equiv \mu^{(p)}(\cdot)$.

On the other hand, we have already remarked that the biggest weight of the portfolio $\mu^{(p)}(\cdot)$ in (7.1) does not exceed the largest market weight:

$$\mu_{(1)}^{(p)}(t) := \max_{1 \leq i \leq n} \mu_i^{(p)}(t) = \frac{(\mu_{(1)}(t))^p}{\sum_{k=1}^n (\mu_{(k)}(t))^p} \leq \mu_{(1)}(t). \quad (7.7)$$

The reverse inequality holds for the smallest weights: $\mu_{(n)}^{(p)}(t) := \min_{1 \leq i \leq n} \mu_i^{(p)}(t) \geq \mu_{(n)}(t)$.

We have assumed that the market is weakly diverse over $[0, T]$, namely, that there is some $0 < \delta < 1$ for which $\int_0^T (1 - \mu_{(1)}(t)) dt > \delta T$ holds almost surely. From (3.12) and (7.7), this implies

$$\int_0^T \gamma_{\mu^{(p)}}^*(t) dt \geq \frac{\varepsilon}{2} \int_0^T (1 - \mu_{(1)}^{(p)}(t)) dt \geq \frac{\varepsilon}{2} \int_0^T (1 - \mu_{(1)}(t)) dt > \frac{\varepsilon}{2} \delta T$$

a.s. In conjunction with (7.6), this leads to (7.2) and (7.3) via

$$\log\left(\frac{V^{\mu^{(p)}}(T)}{V^\mu(T)}\right) > (1-p)\left(\frac{\varepsilon T}{2} \delta - \frac{1}{p} \log n\right). \quad (7.8)$$

□

If $\mathcal{M}$ is uniformly weakly diverse and strongly nondegenerate over an interval $[T_0, \infty)$, then (7.8) implies that the market portfolio will lag rather significantly behind the diversity-weighted portfolio over long-time horizons. To wit, that (6.2) will hold:

$$\mathcal{L}^{\mu^{(p)}, \mu} = \liminf_{T \rightarrow \infty} \frac{1}{T} \log\left(V^{\mu^{(p)}}(T)/V^\mu(T)\right) \geq (1-p)\varepsilon\delta/2 > 0, \quad \text{a.s.}$$

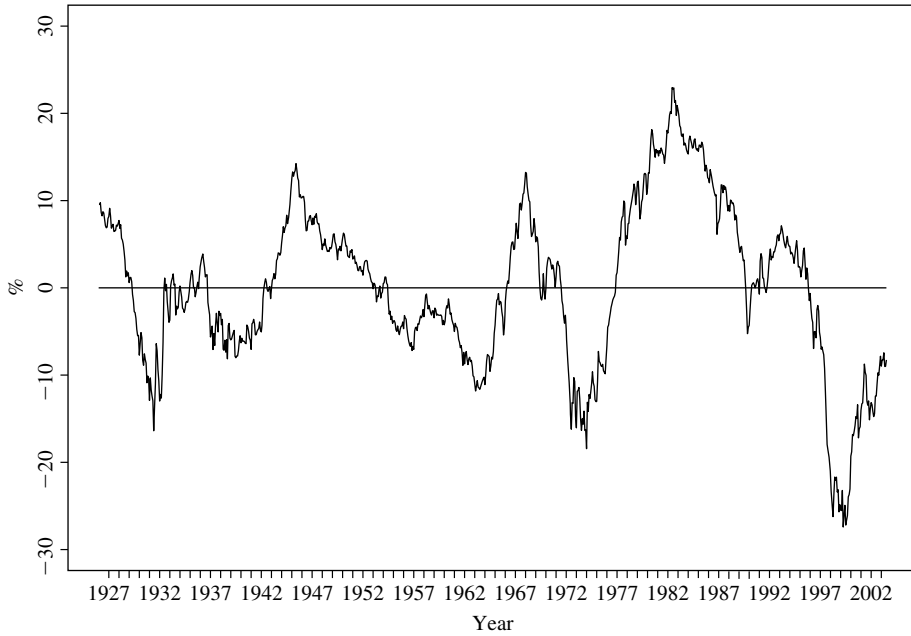


FIG. 3.1 Cumulative change in market diversity, 1927–2004.

In Fig. 3.1, we see the cumulative changes in the diversity of the U.S. stock market over the period from 1927 to 2004, measured by $\mathbf{G}_p(\cdot)$ with $p = 1/2$. The chart shows the cumulative changes in diversity due to capital gains and losses, rather than absolute diversity, which is affected by changes in market composition and corporate actions. Considering only capital gains and losses has the same effect as adjusting the “divisor” of an equity index. The values used in Fig. 3.1 have been normalized so that the average over the whole period is zero. We can observe from the chart that diversity appears to be mean reverting over the long term, with intermediate trends of 10–20 years. The extreme lows for diversity seem to accompany bubbles: the Great Depression, the nifty fifty era of the early 1970s, and the “irrational exuberance” period of the late 1990s.

REMARK 7.1. (FERNHOLZ [2002]): Under the conditions of this section, consider the portfolio with weights

$$\pi_i(t) = \left(\frac{2 - \mu_i(t)}{\mathbf{G}(\mu(t))} - 1 \right) \mu_i(t), \quad 1 \leq i \leq n, \quad \text{where} \quad \mathbf{G}(x) := 1 - \frac{1}{2} \sum_{i=1}^n x_i^2$$

for $x \in \Delta^n$. It can be shown that this portfolio leads to arbitrage relative to the market, over sufficiently long-time horizons $[0, T]$, namely, with $T \geq (2n/\varepsilon\delta^2) \log 2$. In this case, we also have $\pi_i(t) \leq 3\mu_i(t)$, for all $t \in [0, T]$, a.s., so, with appropriate initial conditions, there is no risk that this $\pi(\cdot)$ will hold more of a stock than the market holds.

REMARK 7.2. Statistical Arbitrage and Enhanced Indexing. With $p = 1$, the portfolio $\mu^{(p)}(\cdot)$ of (7.1) corresponds to the market portfolio; with $p = 0$, it gives the equally weighted portfolio, namely, $\varphi_i(\cdot) := \mu_i^{(0)}(\cdot) \equiv 1/n$ for all $i = 1, \dots, n$.

The market portfolio $\mu(\cdot)$ buys at time $t = 0$ the same number of shares in all companies of the market and holds them until the end $t = T$ of the investing horizon. It represents the quintessential “buy-and-hold” strategy.

The equally weighted portfolio $\varphi(\cdot)$ maintains equal weights in all stocks at all times; it accomplishes this by selling those stocks whose price rises relative to the rest, and by buying stocks whose price falls relative to the others. Because of this built-in aspect of “buying-low-and-selling-high”, the equally weighted portfolio can be used as a simple prototype for studying systematically the performance of statistical arbitrage strategies in equity markets; see FERNHOLZ and MAGUIRE [2006] for details. Of course, implementing such a strategy necessitates very frequent trading and can incur substantial transaction costs for an investor who is not a broker/dealer. It can also involve considerable risk: whereas the second term on the right-hand side of

$$\log V^\varphi(T) = \frac{1}{n} \log \left(\frac{X_1(T) \dots X_n(T)}{X_1(0) \dots X_n(0)} \right) + \int_0^T \gamma_\varphi^*(t) dt, \quad (7.9)$$

or of

$$\log \left(\frac{V^\varphi(T)}{V^\mu(T)} \right) = \frac{1}{n} \log \left(\frac{\mu_1(T) \dots \mu_n(T)}{\mu_1(0) \dots \mu_n(0)} \right) + \int_0^T \gamma_\varphi^*(t) dt, \quad (7.10)$$

is increasing in T , the first terms on the right-hand sides of these expressions can fluctuate quite a bit. These equations are obtained by reading (1.17), (1.13), (3.9) with $\pi_i(\cdot) \equiv \varphi_i(\cdot) \equiv 1/n$ for all $i = 1, \dots, n$, thus with excess growth rate

$$\gamma_\varphi^*(t) = \frac{1}{2n} \left(\sum_{i=1}^n a_{ii}(t) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{ij}(t) \right). \quad (7.11)$$

The diversity-weighted portfolios $\mu^{(p)}(\cdot)$ of (7.1) with $0 < p < 1$ stand between these two extremes, of capitalization weighting (as in S&P 500) and of equal weighting (as in the Value-Line Index); they try to capture some of the “buy-low/sell-high” characteristics of equal weighting, but without deviating too much from the market capitalizations and without incurring a lot of trading costs or excessive risk. They can be viewed as “enhanced market portfolios” or “enhanced indices”, in this sense.

8. Mirror portfolios, short-horizon arbitrage

In the previous section, we saw that in weakly diverse markets which satisfy the strict nondegeneracy condition (3.10), one can construct explicitly simple long-only portfolios that lead to strong arbitrages relative to the market over sufficiently long time horizons. The purpose of this section is to demonstrate that, under these same conditions, such arbitrages exist indeed over arbitrary time horizons, no matter how small.

For any given portfolio $\pi(\cdot)$ and real number $q \neq 0$, define the q -mirror image of $\pi(\cdot)$ with respect to the market portfolio, as

$$\tilde{\pi}^{[q]}(\cdot) := q\pi(\cdot) + (1 - q)\mu(\cdot).$$

This is clearly a portfolio; it is long only if $\pi(\cdot)$ itself is long only and $0 < q < 1$. If $q = -1$, we call $\tilde{\pi}^{[-1]}(\cdot) = 2\mu(\cdot) - \pi(\cdot)$ the “mirror image” of $\pi(\cdot)$ with respect to the market.

By analogy with (1.19), let us define the relative covariance of $\pi(\cdot)$ with respect to the market, as

$$\tau_{\mu\mu}^{\pi}(t) := (\pi(t) - \mu(t))' a(t) (\pi(t) - \mu(t)), \quad 0 \leq t \leq T.$$

REMARK 8.1. Recall from (1.21) the fact $\tau^{\mu}(t)\mu(t) \equiv 0$ and establish the elementary properties $\tau_{\mu\mu}^{\pi}(t) = \pi'(t)\tau^{\mu}(t)\pi(t) = \tau_{\pi\pi}^{\mu}(t)$ and $\tau_{\tilde{\pi}^{[q]}\tilde{\pi}^{[q]}}^{\mu}(t) = q^2\tau_{\pi\pi}^{\mu}(t)$.

REMARK 8.2. The wealth of $\tilde{\pi}^{[q]}(\cdot)$ relative to the market can be computed as

$$\log\left(\frac{V^{\tilde{\pi}^{[q]}}(T)}{V^{\mu}(T)}\right) = q \log\left(\frac{V^{\pi}(T)}{V^{\mu}(T)}\right) + \frac{q(1-q)}{2} \int_0^T \tau_{\pi\pi}^{\mu}(t) dt.$$

Indeed, let us write the second equality in (3.4) with $\pi(\cdot)$ replaced by $\tilde{\pi}^{[q]}(\cdot)$, and recall $\tilde{\pi}^{[q]} - \mu = q(\pi - \mu)$. From the resulting expression, let us subtract the second equality in (3.4), now multiplied by q ; the result is

$$\frac{d}{dt} \left(\log \frac{V^{\tilde{\pi}^{[q]}}(t)}{V^{\mu}(t)} - q \log \frac{V^{\pi}(t)}{V^{\mu}(t)} \right) = (q-1)\gamma_{\mu}^*(t) + (\gamma_{\tilde{\pi}^{[q]}}^*(t) - q\gamma_{\mu}^*(t)).$$

But from the equalities of Remark 8.1 and Lemma 3.3, we obtain

$$\begin{aligned} 2(\gamma_{\tilde{\pi}^{[q]}}^*(t) - q\gamma_{\pi}^*(t)) &= \sum_{i=1}^n (\tilde{\pi}^{[q]}(t) - q\pi_i(t)) \tau_{ii}^{\mu}(t) - \tau_{\tilde{\pi}^{[q]}\tilde{\pi}^{[q]}}^{\mu}(t) + q\tau_{\pi\pi}^{\mu}(t) \\ &= (1-q) \sum_{i=1}^n \mu_i(t) \tau_{ii}^{\mu}(t) + q\tau_{\pi\pi}^{\mu}(t) - q^2\tau_{\pi\pi}^{\mu}(t) \\ &= (1-q)(2\gamma_{\mu}^*(t) + q\tau_{\pi\pi}^{\mu}(t)). \end{aligned}$$

The desired equality now follows.

REMARK 8.3. Suppose that the portfolio $\pi(\cdot)$ satisfies

$$\mathbb{P}(V^{\pi}(T)/V^{\mu}(T) \geq \beta) = 1 \quad \text{or} \quad \mathbb{P}(V^{\pi}(T)/V^{\mu}(T) \leq 1/\beta) = 1$$

and

$$\mathbb{P}\left(\int_0^T \tau_{\pi\pi}^{\mu}(t) dt \geq \eta\right) = 1$$

for some real numbers $T > 0$, $\eta > 0$, and $0 < \beta < 1$. Then, there exists another portfolio $\hat{\pi}(\cdot)$ with $\mathbb{P}(V^{\hat{\pi}}(T) < V^{\mu}(T)) = 1$.

To see this, suppose first that we have $\mathbb{P}(V^{\pi}(T)/V^{\mu}(T) \leq 1/\beta) = 1$; then, we can just take $\hat{\pi}(\cdot) \equiv \tilde{\pi}^{[q]}(\cdot)$ with $q > 1 + (2/\eta) \log(1/\beta)$, because Remark 8.2 gives

$$\log \left(\frac{V^{\tilde{\pi}^{[q]}}(T)}{V^{\mu}(T)} \right) \leq q \left(\log(1/\beta) + \frac{1-q}{2} \eta \right) < 0, \quad \text{a.s.}$$

If, on the other hand, $\mathbb{P}(V^{\pi}(T)/V^{\mu}(T) \geq \beta) = 1$ holds, then similar reasoning shows that it suffices to take $\hat{\pi}(\cdot) \equiv \tilde{\pi}^{[q]}(\cdot)$ with $q \in (0, 1 - (2/\eta) \log(1/\beta))$.

8.1. A “seed” portfolio

Now let us consider $\pi = e_1 = (1, 0, \dots, 0)'$ and the market portfolio $\mu(\cdot)$; we shall fix a real number $q > 1$ in a moment, and define the portfolio

$$\hat{\pi}(t) := \tilde{\pi}^{[q]}(t) = qe_1 + (1-q)\mu(t), \quad 0 \leq t < \infty \quad (8.1)$$

which takes a long position in the first stock and a short position in the market. In particular, $\hat{\pi}_1(t) = q + (1-q)\mu_1(t)$ and $\hat{\pi}_i(t) = (1-q)\mu_i(t)$ for $i = 2, \dots, n$. Then, we have

$$\log \left(\frac{V^{\hat{\pi}}(T)}{V^{\mu}(T)} \right) = q \log \left(\frac{\mu_1(T)}{\mu_1(0)} \right) - \frac{q(q-1)}{2} \int_0^T \tau_{11}^{\mu}(t) dt \quad (8.2)$$

from Remark 8.2. But taking $\beta := \mu_1(0)$, we have $(\mu_1(T)/\mu_1(0)) \leq 1/\beta$; if the market is weakly diverse on $[0, T]$ and satisfies the strict nondegeneracy condition (3.10), we obtain from (3.11) and the Cauchy–Schwarz inequality

$$\int_0^T \tau_{11}^{\mu}(t) dt \geq \varepsilon \int_0^T (1 - \mu_{(1)}(t))^2 dt > \varepsilon \delta^2 T =: \eta. \quad (8.3)$$

Recalling Remark 8.3, we see that the market portfolio represents then a strong arbitrage opportunity with respect to the portfolio $\hat{\pi}(\cdot)$ of (8.1), provided that for any given real number $T > 0$ we select

$$q > q(T) := 1 + (2/\varepsilon \delta^2 T) \log(1/\mu_1(0)). \quad (8.4)$$

The portfolio $\hat{\pi}(\cdot)$ of (8.1) can be used as a “seed” to create long-only portfolios that outperform the market portfolio $\mu(\cdot)$, over any time horizon $[0, T]$ with given real number $T > 0$. The idea is to immerse $\hat{\pi}(\cdot)$ in a sea of market portfolio, swamping the short positions while retaining the essential portfolio characteristics. Crucial in these constructions is following the a.s. comparison, a consequence of (8.2):

$$V^{\hat{\pi}}(t) \leq \left(\frac{\mu_1(t)}{\mu_1(0)} \right)^q V^{\mu}(t), \quad 0 \leq t < \infty. \quad (8.5)$$

8.2. Relative arbitrage on arbitrary time horizons

To implement this idea, consider a strategy $h(\cdot)$ that, at time $t = 0$, invests $q/(\mu_1(0))^q$ dollars in the market portfolio, goes one dollar short in the portfolio $\hat{\pi}(\cdot)$ of (8.1), and makes no change thereafter. The number $q > 1$ is chosen again as in (8.4). The wealth generated by this strategy, with initial capital $z := q/(\mu_1(0))^q - 1 > 0$, is

$$\mathcal{V}^{z,h}(t) = \frac{qV^\mu(t)}{(\mu_1(0))^q} - V^{\hat{\pi}}(t) \geq \frac{V^\mu(t)}{(\mu_1(0))^q} (q - (\mu_1(t))^q) > 0, \quad 0 \leq t < \infty, \quad (8.6)$$

thanks to (8.5) and $q > 1 > (\mu_1(t))^q$. This process $\mathcal{V}^{z,h}(\cdot)$ coincides with the wealth $V^{z,\eta}(\cdot)$ generated by a portfolio $\eta(\cdot)$ with weights

$$\eta_i(t) = \frac{1}{\mathcal{V}^{z,h}(t)} \left(\frac{q\mu_i(t)}{(\mu_1(0))^q} V^\mu(t) - \hat{\pi}_i(t) V^{\hat{\pi}}(t) \right), \quad i = 1, \dots, n \quad (8.7)$$

that satisfy $\sum_{i=1}^n \eta_i(t) = 1$. Now, we have $\hat{\pi}_i(t) = -(q-1)\mu_i(t) < 0$ for $i = 2, \dots, n$, so the quantities $\eta_2(\cdot), \dots, \eta_n(\cdot)$ are strictly positive. To check that $\eta(\cdot)$ is a long-only portfolio, we have to verify $\eta_1(t) \geq 0$; but the dollar amount invested by $\eta(\cdot)$ in the first stock at time t , namely,

$$\frac{q\mu_1(t)}{(\mu_1(0))^q} V^\mu(t) - [q - (q-1)\mu_1(t)] V^{\hat{\pi}}(t),$$

dominates $\frac{q\mu_1(t)}{(\mu_1(0))^q} V^\mu(t) - [q - (q-1)\mu_1(t)] \left(\frac{\mu_1(t)}{\mu_1(0)} \right)^q V^\mu(t)$, or equivalently

$$\frac{V^\mu(t)\mu_1(t)}{(\mu_1(0))^q} \left((q-1)(\mu_1(t))^q + q[1 - (\mu_1(t))^{q-1}] \right) > 0,$$

again thanks to (8.5) and $q > 1 > (\mu_1(t))^{q-1}$. Thus, $\eta(\cdot)$ is indeed a long-only portfolio.

On the other hand, $\eta(\cdot)$ outperforms at $t = T$ a market portfolio that starts with the same initial capital at $t = 0$; this is because $\eta(\cdot)$ is long in the market $\mu(\cdot)$ and short in the portfolio $\hat{\pi}(\cdot)$, which underperforms the market at $t = T$. Indeed, from Remark 8.3, we have

$$V^{z,\eta}(T) = \frac{q}{(\mu_1(0))^q} V^\mu(T) - V^{\hat{\pi}}(T) > zV^\mu(T) = V^{z,\mu}(T), \quad \text{a.s.}$$

Note, however, that as $T \downarrow 0$, the initial capital $z(T) = q(T)/(\mu_1(0))^{q(T)} - 1$ required to do all of this, increases without bound: it may take a huge amount of initial investment to realize the extra basis point's worth of relative arbitrage over a short-time horizon. This confirms, if confirmation is needed, the old adage that *time is money*...

9. A diverse market model

The careful reader might have been wondering whether the theory we have developed so far may turn out to be vacuous. Do there exist market models of the form (1.1) and (1.2) that are diverse, at least weakly? This is, of course, a very legitimate question.

Let us mention then, rather briefly, an example of such a market model $\mathcal{M}$ that is diverse over any given time horizon $[0, T]$ with real $T > 0$. For the details of this construction, we refer to [FERNHOLZ, KARATZAS and KARDARAS] [2005].

With given $\delta \in (1/2, 1)$, equal numbers of stocks and driving Brownian motions (that is, $d = n$), constant volatility matrix σ that satisfies (3.10), and nonnegative numbers $g_1, \dots, g_n$, we take a model

$$dx \log X_i(t) = \gamma_i(t) dt + \sum_{v=1}^n \sigma_{iv} dx W_v(t), \quad 0 \leq t \leq T \quad (9.1)$$

in the form (1.5) for the vector $\mathfrak{X}(\cdot) = (X_1(\cdot), \dots, X_n(\cdot))'$ of stock prices. With the usual notation $X(t) = \sum_{j=1}^n X_j(t)$, its growth rates are specified as

$$\gamma_i(t) := g_i 1_{\mathcal{Q}_i^c}(\mathfrak{X}(t)) - \frac{M}{\delta} \frac{1_{\mathcal{Q}_i}(\mathfrak{X}(t))}{\log((1-\delta)X(t)/X_i(t))}. \quad (9.2)$$

In other words, $\gamma_i(t) = g_i \geq 0$ if $\mathfrak{X}(t) \notin \mathcal{Q}_i$ (the i th stock does not have the largest capitalization) and

$$\gamma_i(t) = -\frac{M}{\delta} \frac{1}{\log((1-\delta)/\mu_i(t))}, \quad \text{if } \mathfrak{X}(t) \in \mathcal{Q}_i \quad (9.3)$$

(the i th stock does have the largest capitalization). We are setting here

$$\mathcal{Q}_1 := \left\{ x \in (0, \infty)^n \mid x_1 \geq \max_{2 \leq j \leq n} x_j \right\}, \quad \mathcal{Q}_n := \left\{ x \in (0, \infty)^n \mid x_n > \max_{1 \leq j \leq n-1} x_j \right\},$$

and

$$\mathcal{Q}_i := \left\{ x \in (0, \infty)^n \mid x_i > \max_{1 \leq j \leq i-1} x_j, x_i \geq \max_{i+1 \leq j \leq n} x_j \right\} \quad \text{for } i = 2, \dots, n-1.$$

With the specification (9.2) and (9.3), all stocks but the largest behave like geometric Brownian motions (with growth rates $g_i \geq 0$ and variances $a_{ii} = \sum_{v=1}^n \sigma_{iv}^2$), whereas the log price of the largest stock is subjected to a log-pole-type singularity in its drift, away from an appropriate right boundary.

One can then show that the resulting system of stochastic differential equations has a unique, strong solution (so the filtration $\mathbb{F}$ is now the one generated by the driving n -dimensional Brownian motion), and that the diversity requirement (5.1) is satisfied on any given time horizon. Such models can be modified appropriately, to create ones that are weakly diverse but not diverse (see [FERNHOLZ and KARATZAS] [2005] for details).

Slightly more generally, in order to guarantee diversity, it is enough to require

$$\min_{2 \leq k \leq n} \gamma_{(k)}(t) \geq 0 \geq \gamma_{(1)}(t), \quad \min_{2 \leq k \leq n} \gamma_{(k)}(t) - \gamma_{(1)}(t) + \frac{\varepsilon}{2} \geq \frac{M}{\delta} F(Q(t)),$$

where $Q(t) := \log((1 - \delta)/\mu_{(1)}(t))$.

Here the function $F : (0, \infty) \rightarrow (0, \infty)$ is taken to be continuous and such that the associated scale function

$$U(x) := \int_1^x \exp\left\{-\int_1^y F(z) dz\right\} dy, \quad x \in (0, \infty) \quad \text{satisfies} \quad U(0+) = -\infty;$$

for instance, we have $U(x) = \log x$ when $F(x) = 1/x$ as above. Under these conditions, it can then be shown that the process $Q(\cdot)$ satisfies $\int_0^T (Q(t))^{-2} dt < \infty$ a.s., and this leads to the a.s. square integrability

$$\sum_{i=1}^n \int_0^T (b_i(t))^2 dt < \infty \quad (9.4)$$

of the induced rates of return of the individual stocks

$$b_i(t) = \frac{1}{2} a_{ii} + g_i 1_{Q_i^c}(\mathfrak{X}(t)) - \frac{M}{\delta} \frac{1_{Q_i}(\mathfrak{X}(t))}{\log((1 - \delta)X(t)/X_i(t))}, \quad i = 1, \dots, n.$$

The square-integrability property (9.4) is, of course, crucial: it guarantees that the market price of risk process $\theta(\cdot) := \sigma^{-1}b(\cdot)$ is square-integrable a.s., exactly as posited in (6.4), so the exponential local martingale $Z(\cdot)$ of (6.5) is well defined (we are assuming $r(\cdot) \equiv 0$ in all this). Thus, the results of Propositions 6.1, and 6.2, and Remark 6.2 are applicable to this model.

For additional examples, and for an interesting probabilistic construction of diverse markets that leads to arbitrage, see OSTERRIEDER and RHEINLÄNDER [2006].

10. Hedging and optimization without EMM

Let us broach now the issue of hedging contingent claims in a market such as that of subsection 6.1, and over a time horizon $[0, T]$ with a real number $T > 0$ satisfying (6.1).

Consider first a European contingent claim, that is, an $\mathcal{F}(T)$ -measurable random variable $Y : \Omega \rightarrow [0, \infty)$ with

$$0 < y := \mathbb{E}(YZ(T)/B(T)) < \infty \quad (10.1)$$

in the notation of (6.5). From the point of view of the seller of the contingent claim (e.g. stock option), this random amount represents a liability that has to be covered with the right amount of initial funds at time $t = 0$ and the right trading strategy during the interval $[0, T]$, so that at the end of the time horizon (time $t = T$) the initial funds have

grown enough to cover the liability without risk. Thus, the seller is interested in the so-called upper hedging price

$$\mathcal{U}^Y(T) := \inf \{ w > 0 \mid \exists h(\cdot) \in \mathcal{H}(w; T) \text{ such that } \mathcal{V}^{w,h}(T) \geq Y, \text{ a.s.} \}, \quad (10.2)$$

the smallest amount of initial capital that makes such riskless hedging possible.

The standard theory of mathematical finance assumes that $\mathfrak{M}$, the set of EMMs for the model $\mathcal{M}$, is nonempty, and then shows that $\mathcal{U}^Y(T)$ can be computed as

$$\mathcal{U}^Y(T) = \sup_{\mathbb{Q} \in \mathfrak{M}} \mathbb{E}^{\mathbb{Q}}(Y/B(T)), \quad (10.3)$$

the supremum of the claim's discounted expected values over this set of probability measures. In our context, no EMM exists (i.e., $\mathfrak{M} = \emptyset$), so the approach breaks down and the problem seems hopeless.

Not quite, though; there is still a long way one can go, simply by using the availability of the strict local martingale $Z(\cdot)$ (and of the associated “deflator” $Z(\cdot)/B(\cdot)$), as well as the properties (6.9), (6.10) of the processes in (6.8). For instance, if the set on the right-hand side of (10.2) is not empty, then for any $w > 0$ in this set and for any $h(\cdot) \in \mathcal{H}(w; T)$, the local martingale $\hat{\mathcal{V}}^{w,h}(\cdot)$ of (6.8) is nonnegative, thus a supermartingale. This gives

$$w \geq \mathbb{E}(\mathcal{V}^{w,h}(T)Z(T)/B(T)) \geq \mathbb{E}(YZ(T)/B(T)) = y,$$

and because $w > 0$ is arbitrary we deduce $\mathcal{U}^Y(T) \geq y$. This inequality holds trivially if the set on the right-hand side of (10.2) is empty, since then we have $\mathcal{U}^Y(T) = \infty$.

10.1. Completeness without EMM

To obtain the reverse inequality, we shall assume that $n = d$, that is, we have exactly as many sources of randomness as there are stocks in the market $\mathcal{M}$, and that the filtration $\mathbb{F}$ is generated by the driving Brownian motion $W(\cdot)$ in (1.1): $\mathbb{F} = \mathbb{F}^W$.

With these assumptions, one can represent the nonnegative martingale

$$M(t) := \mathbb{E}(YZ(T)/B(T) | \mathcal{F}(t)), \quad 0 \leq t \leq T$$

as a stochastic integral

$$M(t) = y + \int_0^t \psi'(s) dW(s), \quad 0 \leq t \leq T \quad (10.4)$$

for some progressively measurable and a.s. square-integrable process $\psi : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ and with the notation of (10.1). Setting

$$V_*(\cdot) := M(\cdot)B(\cdot)/Z(\cdot) \quad \text{and} \quad h_*(\cdot) := (B(\cdot)/Z(\cdot)) a^{-1}(\cdot) \sigma(\cdot) (\psi(\cdot) + M(\cdot)\theta(\cdot)),$$

then comparing (6.10) with (10.4), we observe that $V_*(0) = y$, $V_*(T) = Y$, and $V_*(\cdot) \equiv \mathcal{V}^{y,h_*}(\cdot) \geq 0$ hold almost surely.

Therefore, the trading strategy $h_*(\cdot)$ is in $\mathcal{H}(y; T)$ and satisfies the exact replication property $\mathcal{V}^{y,h_*}(T) = Y$ a.s. This implies that y belongs to the set on the right-hand side

of (10.2), and so $y \geq \mathcal{U}^Y(T)$. But we have already established the reverse inequality, actually in much greater generality, so recalling (10.1) we get the *Black–Scholes-type formula*

$$\mathcal{U}^Y(T) = \mathbb{E}(YZ(T)/B(T)) \quad (10.5)$$

for the upper hedging price of (10.2), under the assumptions of the first paragraph in this subsection.

In particular, we see that a market $\mathcal{M}$ that is weakly diverse, hence without an equivalent probability measure under which discounted stock prices are (at least local) martingales, can nevertheless be complete.

Similar observations have been made by LOWENSTEIN and WILLARD [2000a,b] and by PLATEN [2002, 2006].

REMARK 10.1. Put-Call Parity. In the context of this subsection, suppose $L_1(\cdot)$ and $L_2(\cdot)$ are positive, continuous, and adapted processes, representing the values of two different financial instruments in the market. For instance, $L_1(\cdot) = V^{w_1, \pi_1}(\cdot)$ and $L_2(\cdot) = V^{w_2, \pi_2}(\cdot)$ for two different portfolios $\pi_1(\cdot)$ and $\pi_2(\cdot)$ and real numbers $w_1 > 0$ and $w_2 > 0$. Consider the contingent claims

$$Y_1 := (L_1(T) - L_2(T))^+ \quad \text{and} \quad Y_2 := (L_2(T) - L_1(T))^+.$$

According to (10.5), the quantity $\mathcal{U}_1 = \mathbb{E}[Z(T)Y_1/B(T)]$ is the upper hedging price at $t = 0$ of a contingent claim that confers to its holder the right, though not the obligation, to exchange instrument 2 for instrument 1 at time $t = T$; ditto for $\mathcal{U}_2 = \mathbb{E}[Z(T)Y_2/B(T)]$, with the rôles of instruments 1 and 2 interchanged. Of course,

$$\mathcal{U}_1 - \mathcal{U}_2 = \mathbb{E}[Z(T)(L_1(T) - L_2(T))/B(T)];$$

we say that the two instruments are in put-call parity, if $\mathcal{U}_1 - \mathcal{U}_2 = L_1(0) - L_2(0)$. This will be the case, for instance, if $Z(\cdot)(L_1(\cdot) - L_2(\cdot))/B(\cdot)$ is a martingale.

Put-call parity can fail when relative arbitrage of the type (6.1) exists. For example, take $L_1(\cdot) \equiv V^\pi(\cdot)$ and $L_2(\cdot) \equiv V^\rho(\cdot)$ and observe that (6.1) leads to

$$\mathcal{U}_1 - \mathcal{U}_2 = \mathbb{E}[Z(T)(V^\pi(T) - V^\rho(T))/B(T)] > 0 = V^\pi(0) - V^\rho(0).$$

Proof of (6.16). We can provide now a proof for the claim (6.16) in Remark 6.4. Let us denote by $\mathbf{T}$ the right-hand side of this equation, and note that the inequality $\mathbf{T} \leq \mathbf{T}(r)$ is automatically satisfied if the set in (6.15) is empty (its infimum is then $+\infty$); if the set in (6.15) is not empty, pick any element $T \in (0, \infty)$ and an arbitrary trading strategy $h(\cdot) \in \mathcal{H}(1; T)$ that satisfies $\mathcal{V}^h(T) \geq r \cdot V^\mu(T)$ a.s. The supermartingale property of $Z(\cdot)\mathcal{V}^h(\cdot)/B(\cdot)$ gives then

$$1 \geq \mathbb{E}[Z(T)\mathcal{V}^h(T)/B(T)] \geq r \cdot \mathbb{E}[Z(T)V^\mu(T)/B(T)] = r \cdot f(T),$$

which means that this $T \in (0, \infty)$ belongs to the set of (6.16); thus, the inequality $\mathbf{T} \leq \mathbf{T}(r)$ holds again.

For the reverse inequality, consider the number $y := f(\mathbf{T})$ and observe $0 < y \leq 1/r$ (the right-continuity of $f(\cdot)$); from what we just proved, there exists a trading strategy $h_*(\cdot) \in \mathcal{H}(1; \mathbf{T})$ with which the contingent claim $Y := X(\mathbf{T})/X(0)$ can be replicated exactly at time $t = \mathbf{T}$, in the sense $y \mathcal{V}^{h_*}(\mathbf{T}) = Y$ a.s., since $\mathbb{E}[Z(\mathbf{T})Y/B(\mathbf{T})] = y$. Therefore,

$$(1/r) \cdot \mathcal{V}^{h_*}(\mathbf{T}) \geq y \cdot \mathcal{V}^{h_*}(\mathbf{T}) = Y = X(\mathbf{T})/X(0) = V^\mu(\mathbf{T}) \quad \text{holds a.s.,}$$

and this means that $\mathbf{T}$ belongs to the set of (6.16); thus the inequality $\mathbf{T} \geq \mathbf{T}(r)$ holds as well. $\square$

10.2. Ramifications and open problems

EXAMPLE 10.1. A European call option. Consider the contingent claim $Y = (X_1(T) - q)^+$: this is a European call option on the first stock with strike $q \in (0, \infty)$ and expiration $T \in (0, \infty)$. Let us assume also that the interest-rate process $r(\cdot)$ is bounded away from zero, namely, that $\mathbb{P}[r(t) \geq r, \forall t \geq 0] = 1$ holds for some $r > 0$ and that the market $\mathcal{M}$ is weakly diverse on all sufficiently large time horizons $T \in (0, \infty)$.

Then, for the hedging price $\mathcal{U}^Y(T)$ of this contingent claim, we have from Remark 6.2, (10.5), Jensen's inequality, and $\mathbb{E}(Z(T)) < 1$:

$$\begin{aligned} X_1(0) &> \mathbb{E}(Z(T)X_1(T)/B(T)) \geq \mathbb{E}(Z(T)(X_1(T) - q)^+/B(T)) = \mathcal{U}^Y(T) \\ &\geq \left(\mathbb{E}(Z(T)X_1(T)/B(T)) - q \mathbb{E}(Z(T)e^{-\int_0^T r(t)dt}) \right)^+ \\ &\geq \left(\mathbb{E}(Z(T)X_1(T)/B(T)) - q e^{-rT} \mathbb{E}[Z(T)] \right)^+ \\ &\geq \left(\mathbb{E}(Z(T)X_1(T)/B(T)) - q e^{-rT} \right)^+, \end{aligned}$$

thus,

$$0 \leq \mathcal{U}^Y(\infty) := \lim_{T \rightarrow \infty} \mathcal{U}^Y(T) = \lim_{T \rightarrow \infty} \downarrow \mathbb{E}(Z(T)X_1(T)/B(T)) < X_1(0). \quad (10.6)$$

The upper hedging price of the option is strictly less than the capitalization of the underlying stock at time $t = 0$ and tends to $\mathcal{U}^Y(\infty) \in [0, X_1(0))$ as the time horizon increases without limit.

If $\mathcal{M}$ is weakly diverse uniformly over some $[T_0, \infty)$, then the limit in (10.6) is actually zero: The hedging price of a European call option that can never be exercised is equal to zero. Indeed, for every fixed $p \in (0, 1)$ and $T \geq \left(\frac{2 \log n}{p\epsilon\delta}\right) \vee T_0$, and with the normalization $X(0) = 1$, the quantity

$$\mathbb{E}\left(\frac{Z(T)}{B(T)}X_1(T)\right) \leq \mathbb{E}\left(\frac{Z(T)}{B(T)}V^\mu(T)\right) \leq \mathbb{E}\left(\frac{Z(T)}{B(T)}V^{\mu^{(p)}}(T)\right) n^{\frac{1-p}{p}} e^{-\epsilon\delta(1-p)T/2}$$

is dominated by $n^{\frac{1-p}{p}} e^{-\varepsilon\delta(1-p)T/2}$ from (7.2), (2.2), and the supermartingale property of the process $Z(\cdot)V^{\mu^{(p)}}(\cdot)/B(\cdot)$. Letting $T \rightarrow \infty$, we obtain $\mathcal{U}^Y(\infty) = 0$.

REMARK 10.2. Note the sharp difference between this case and the situation where an EMM exists on every finite time horizon, namely, when both $Z(\cdot)$ and $Z(\cdot)X_1(\cdot)/B(\cdot)$ are martingales. Then we have $\mathbb{E}(Z(T)X_1(T)/B(T)) = X_1(0)$ for all $T \in (0, \infty)$, and $\mathcal{U}^Y(\infty) = X_1(0)$: as the time horizon increases without limit, the hedging price of the call option approaches the stock price at $t = 0$ (see KARATZAS and SHREVE [1998], pp 62).

REMARK 10.3. The above theory extends to the case $d > n$ of incomplete markets, and more generally to closed, convex constraints on portfolio choice as in chapter 5 of KARATZAS and SHREVE [1998], under the conditions of (6.4). The paper by KARATZAS and KARDARAS [2007] can be consulted for a treatment of these issues in a general semimartingale setting.

In particular, the Black–Scholes-type formula (10.5) can be generalized, in the spirit of (10.3), to the case $d > n$ and to a filtration $\mathbb{F}$ not necessarily equal to the Brownian filtration $\mathbb{F}^W$. Let Θ be the set of $\mathbb{F}$ –progressively measurable processes $\theta(\cdot)$ that satisfy the requirements of (6.4); for each $\theta(\cdot) \in \Theta$, let us denote by $Z_\theta(\cdot)$ the process of (6.5). Then, the upper hedging price of (10.2) is given as

$$\mathcal{U}^Y(T) = \sup_{\theta(\cdot) \in \Theta} \mathbb{E}(YZ_\theta(T)/B(T)). \quad (10.7)$$

REMARK 10.4. Open Question: Develop a theory for pricing *American contingent claims* under the assumptions of the present section. As KARDARAS [2006] observes, in the absence of an EMM it is not optimal to exercise an American call option (written on a non-dividend-paying stock) only at maturity $t = T$. Can one then characterize or compute the optimal exercise time?

10.3. Utility maximization in the absence of EMM

Suppose we are given initial capital $w > 0$, a time horizon $[0, T]$ for some real $T > 0$, and a utility function $u : (0, \infty) \rightarrow \mathbb{R}$ (strictly increasing, strictly concave, of class C^1 , with $u'(0) := \lim_{x \downarrow 0} u'(x) = \infty$, $u'(\infty) := \lim_{x \rightarrow \infty} u'(x) = 0$ and $u(0) := \lim_{x \downarrow 0} u(x)$). The problem is to compute the maximal expected utility from terminal wealth

$$\mathfrak{U}(w) := \sup_{h(\cdot) \in \mathcal{H}(w; T)} \mathbb{E}[u(\mathcal{V}^{w, h}(T))],$$

to decide whether the supremum is attained, and if so, to identify a strategy $\hat{h}(\cdot) \in \mathcal{H}(w; T)$ that attains it. We place ourselves under the assumptions of the present section, including those of subsection 10.1 ($d = n$, $\mathbb{F} = \mathbb{F}^W$).

REMARK 10.5. The solution to this question is given by the replicating strategy $\hat{h}(\cdot) \in \mathcal{H}_+(w; T)$ for the contingent claim

$$\Upsilon = I(\Xi(w)D(T)), \quad \text{where} \quad D(t) := Z(t)/B(t) \quad \text{for } 0 \leq t \leq T,$$

in the sense $\mathcal{V}^{w, \hat{h}}(T) = \Upsilon$ a.s. Here $Z(\cdot)$ is the exponential local martingale of (6.5), $I : (0, \infty) \rightarrow (0, \infty)$ is the inverse of the strictly decreasing marginal utility function $u' : (0, \infty) \rightarrow (0, \infty)$, and $\Xi : (0, \infty) \rightarrow (0, \infty)$ is the inverse of the strictly decreasing function $\mathcal{W}(\cdot)$ given by

$$\mathcal{W}(\xi) := \mathbb{E}[D(T)I(\xi D(T))], \quad 0 < \xi < \infty,$$

which we are assuming to be $(0, \infty)$ -valued.

In the case of the *logarithmic utility function* $u(x) = \log x$, $x \in (0, \infty)$, it is easily shown that the “log-optimal” trading strategy $h^*(\cdot) \in \mathcal{H}_+(w; T)$ and its associated wealth process $V_*(\cdot) \equiv \mathcal{V}^{w, h^*}(\cdot)$ are given, respectively, by

$$h^*(t) = V_*(t)a^{-1}(t)[b(t) - r(t)\mathbf{I}], \quad V_*(t) = w/D(t) \quad (10.8)$$

for $0 \leq t \leq T$. The discounted log-optimal wealth process satisfies

$$d(V_*(t)/B(t)) = (V_*(t)/B(t))\theta'(t)[\theta(t)dt + dW(t)], \quad (10.9)$$

an equation whose solution is readily seen to be $V_*(t)/B(t) = w/Z(t)$, $0 \leq t \leq T$.

Note that no assumption is been made regarding the existence of an EMM; to wit, $Z(\cdot)$ does not have to be a martingale. (See KARATZAS, LEHOCZKY, SHREVE and XU [1991] for more information on this problem and on its much more interesting incomplete market version $d > n$, under the assumption that the volatility matrix $\sigma(\cdot)$ is of full (row) rank and without assuming the existence of EMM).

Note also that the deflated optimal wealth process is constant: $\hat{V}_*(\cdot) \equiv V_*(\cdot)Z(\cdot)/B(\cdot) = w$. This should be contrasted to (6.12) of Remark 6.2 in the light of Remark 6.5.

The log-optimal trading strategy of (10.8) has some obviously desirable features, discussed in the next remark. But unlike the diversity-weighted portfolio of (7.1) or, more generally, the functionally generated portfolios of the next section, it needs for its implementation knowledge of the covariance structure and of the mean rates of return; these are quite hard to estimate in practice.

REMARK 10.6. **The “Numéraire” property:** Assume that the log-optimal strategy $h^*(\cdot) \in \mathcal{H}_+(w)$ of (10.8) is defined for all $0 \leq t < \infty$; it has then the following numéraire property

$$\mathcal{V}^{w, h}(\cdot)/\mathcal{V}^{w, h^*}(\cdot) \quad \text{is a supermartingale, } \forall h(\cdot) \in \mathcal{H}_+(w), \quad (10.10)$$

and from this, one can derive the asymptotic growth optimality property

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{\mathcal{V}^{w, h}(T)}{\mathcal{V}^{w, h^*}(T)} \right) \leq 0 \quad \text{a.s., } \forall h(\cdot) \in \mathcal{H}_+(w).$$

These are the same notions we encountered in Problem 6 of Section 4, in the setup of portfolios (as opposed to trading strategies). For a detailed study of these issues in a far more general context, see KARATZAS and KARDARAS [2007].

REMARK 10.7. (PLATEN [2006]): The equation for $\Psi(\cdot) := V_*(\cdot)/B(\cdot) = w/Z(\cdot)$ in (10.9) is

$$d\Psi(t) = \alpha(t) dt + \sqrt{\Psi(t)\alpha(t)} d\mathfrak{B}(t), \quad \Psi(0) = w$$

with $\mathfrak{B}(\cdot)$ a one-dimensional Brownian motion and $\alpha(t) := \Psi(\cdot)\|\theta(\cdot)\|^2$.

Then, $\Psi(\cdot)$ is a time-changed and scaled squared Bessel process in dimension 4 (sum of squares of four independent Brownian motions), that is, $\Psi(\cdot) = \mathfrak{X}(A(\cdot))/4$, where

$$A(\cdot) := \int_0^\cdot \alpha(s) ds \quad \text{and} \quad \mathfrak{X}(u) = 4(w + u) + 2 \int_0^u \sqrt{\mathfrak{X}(v)} db(v), \quad u \geq 0$$

in terms of yet another standard, one-dimensional Brownian motion $b(\cdot)$.

REMARK 10.8. It might be useful to note at this point that, just as for the optimization problems of this subsection, no assumption regarding the existence of EMM was necessary for any of the Problems 1–6 of Section 4.

Functionally Generated Portfolios

Functionally generated portfolios were introduced by FERNHOLZ [1999a] and generalize broadly the diversity-weighted portfolios of Section 7. For this new class of portfolios, one can derive a decomposition of their relative return analogous to that of (7.5), and this proves useful in the construction and study of arbitrages relative to the market. Just like (7.5), this new decomposition (11.2) does not involve stochastic integrals and opens the possibility for making probability-one comparisons over given fixed time horizons. Functionally generated portfolios can be constructed for general classes of assets, with the market portfolio replaced by an arbitrary passive portfolio of the assets under consideration.

11. Portfolio-generating functions

Certain real-valued functions of the market weights $\mu_1(t), \dots, \mu_n(t)$ can be used to construct dynamic portfolios that behave in a controlled manner. The portfolio-generating functions that interest us most fall into two categories: smooth functions of the market weights and smooth functions of the ranked market weights. Those portfolio-generating functions that are smooth functions of the market weights can be used to create portfolios with returns that satisfy almost sure relationships relative to the market portfolio and hence can be applied to situations in which arbitrage might be possible. Those functions that are smooth functions of the ranked market weights can be used to analyze the role of company size in portfolio behavior.

Suppose we are given a function $\mathbf{G} : U \rightarrow (0, \infty)$ which is defined and of class $\mathcal{C}^2$ on some open neighborhood U of Δ_+^n , and such that the mapping $x \mapsto x_i D_i \log \mathbf{G}(x)$ is bounded on U for all $i = 1, \dots, n$. Consider also the portfolio $\pi(\cdot)$ with weights

$$\pi_i(t) = \left(D_i \log \mathbf{G}(\mu(t)) + 1 - \sum_{j=1}^n \mu_j(t) D_j \log \mathbf{G}(\mu(t)) \right) \cdot \mu_i(t), \quad 1 \leq i \leq n. \quad (11.1)$$

We call this the *portfolio generated by $\mathbf{G}(\cdot)$* . It can be shown that the relative wealth process of this portfolio, with respect to the market, is given by the master formula

$$\log \left(\frac{V^\pi(T)}{V^\mu(T)} \right) = \log \left(\frac{\mathbf{G}(\mu(T))}{\mathbf{G}(\mu(0))} \right) + \int_0^T \mathbf{g}(t) dt, \quad 0 \leq T < \infty, \quad (11.2)$$

where the so-called drift process $\mathbf{g}(\cdot)$ is given by

$$\mathbf{g}(t) := \frac{-1}{2\mathbf{G}(\mu(t))} \sum_{i=1}^n \sum_{j=1}^n D_{ij}^2 \mathbf{G}(\mu(t)) \mu_i(t) \mu_j(t) \tau_{ij}^\mu(t). \quad (11.3)$$

The portfolio weights of (11.1) depend only on the market weights $\mu_1(t), \dots, \mu_n(t)$, not on the covariance structure of the market. Thus, the portfolio of (11.1) can be implemented, and its associated wealth process $V^\pi(\cdot)$ observed through time, only in terms of the evolution of these market weights over $[0, T]$.

The covariance structure enters only in the computation of the drift term in (11.3). But the remarkable thing is that in order to compute the cumulative effect $\int_0^T \mathbf{g}(t) dt$ of this drift, there is no need to know or estimate this covariance structure at all; (11.2) does this for us in the form $\int_0^T \mathbf{g}(t) dt = \log(V^\pi(T)\mathbf{G}(\mu(0))/V^\mu(T)\mathbf{G}(\mu(T)))$ and in terms of quantities that are observable.

The proof of the very important “master formula” (11.2) is given below, at the very end of the present section. It can be skipped on first reading.

REMARK 11.1. Suppose the function $\mathbf{G}(\cdot)$ is concave, or, more precisely, its Hessian $D^2\mathbf{G}(x) = (D_{ij}^2\mathbf{G}(x))_{1 \leq i, j \leq n}$ has at most one positive eigenvalue for each $x \in U$ and, if a positive eigenvalue exists, the corresponding eigenvector is orthogonal to Δ_+^n . Then, the portfolio $\pi(\cdot)$ generated by $\mathbf{G}(\cdot)$ as in (11.1) is long-only (i.e., each weight $\pi_i(\cdot)$ is nonnegative), and the drift term $\mathbf{g}(\cdot)$ is nonnegative; if $\text{rank}(D^2\mathbf{G}(x)) > 1$ holds for each $x \in U$, then $\mathbf{g}(\cdot)$ is positive.

For instance,

1. $\mathbf{G}(\cdot) \equiv w$, a positive constant, generates the market portfolio;
2. $\mathbf{G}(x) = w_1x_1 + \dots + w_nx_n$ generates the passive portfolio that buys at time $t = 0$ and holds up until time $t = T$, a fixed number of shares w_i in each stock $i = 1, \dots, n$ (the market portfolio corresponds to the special case $w_1 = \dots = w_n = w$ of equal numbers of shares across assets);
3. $\mathbf{G}(x) \equiv \mathbf{G}_p(x) := (x_1^p + \dots + x_n^p)^{1/p}$, for some $0 < p < 1$, generates the diversity-weighted portfolio $\mu^{(p)}(\cdot)$ of (7.1), with drift process $\mathbf{g}(\cdot) \equiv (1-p)\gamma_{\mu^{(p)}}^*(\cdot)$;
4. $\mathbf{G}(x) \equiv \mathbf{F}(x) := (x_1 \dots x_n)^{1/n}$ generates the equally weighted portfolio $\varphi_i(\cdot) \equiv 1/n, i = 1, \dots, n$ introduced in Remark 7.2, with drift $\mathbf{g}^\varphi(\cdot) \equiv \gamma_\varphi^*(\cdot)$ as in (7.11).

In a similar manner, $\mathbf{F}_c(x) := c + \mathbf{F}(x)$, for $c \in (0, \infty)$, generates the convex combination

$$\varphi_i^c(t) := \frac{\mathbf{F}(\mu(t))}{c + \mathbf{F}(\mu(t))} \cdot \frac{1}{n} + \frac{c}{c + \mathbf{F}(\mu(t))} \cdot \mu_i(t), \quad i = 1, \dots, n \quad (11.4)$$

of the equally weighted portfolio and the market, with associated drift rate

$$\mathbf{g}^{\varphi^c}(t) = \frac{\mathbf{F}(\mu(t))}{c + \mathbf{F}(\mu(t))} \gamma_\varphi^*(t). \quad (11.5)$$

5. Consider now the entropy function $\mathbf{H}(x) := -\sum_{i=1}^n x_i \log x_i$, $x \in \Delta_+^n$ and, for any given $c \in (0, \infty)$, its modification

$$\mathbf{H}_c(x) := c + \mathbf{H}(x), \quad \text{which satisfies} \quad c < \mathbf{H}_c(x) \leq c + \log n, \quad x \in \Delta_+^n. \quad (11.6)$$

This modified entropy function generates an *entropy-weighted* portfolio $\pi^c(\cdot)$ with weights and drift process given, respectively, as

$$\pi_i^c(t) = \frac{\mu_i(t)}{\mathbf{H}_c(\mu(t))} (c - \log \mu_i(t)), \quad 1 \leq i \leq n \quad \text{and} \quad \mathbf{g}^c(t) = \frac{\gamma_\mu^*(t)}{\mathbf{H}_c(\mu(t))}. \quad (11.7)$$

To obtain some idea about the behavior of one of these portfolios with actual stocks, we ran a simulation of a diversity-weighted portfolio using the stock database from the Center for Research in Securities Prices (CRSP) at the University of Chicago. The data included 50 years of monthly values from 1956 to 2005 for exchange-traded stocks, after the removal of closed-end funds, REITs, and ADRs not included in the S&P 500 Index. From this universe, we considered a cap-weighted large-stock index consisting of the largest 1000 stocks in the database. Against this index, we simulated the performance of the corresponding diversity-weighted portfolio, generated by $\mathbf{G}_p$ of Remark 11.1, Example 11.1 above, with $p = 1/2$. No trading costs were included.

The results of the simulation are presented in Fig. 11.1: Curve 1 is the change in the generating function, Curve 2 is the cumulative drift process $\int_0^\cdot \mathbf{g}(t) dt$, and Curve 3 is the relative return. Each curve shows the cumulative value of the monthly changes induced in the corresponding process by capital gains or losses in the stocks, so the curves are unaffected by monthly changes in the composition of the database. As can be seen, Curve 3 is the sum of Curves 1 and 2. The cumulative drift process $\int_0^\cdot \mathbf{g}(t) dt$ was the dominant term over the period, with a total contribution of about 40 percentage points to the relative return. The drift process $\mathbf{g}(\cdot)$ was quite stable over the 50-year period, with the possible exception of the period around 2000, when “irrational exuberance” increased the volatility of the stocks as well as the intrinsic volatility of the entire market and, hence, increased the value of $\mathbf{g}(\cdot) \equiv (1 - p)\gamma_{\mu^{(p)}}^*(\cdot)$. The cumulative drift process $\int_0^\cdot \mathbf{g}(t) dt$ here has been adjusted to account for “leakage” (see Remark 11.9).

11.1. Sufficient intrinsic volatility leads to arbitrage

Broadly accepted practitioner wisdom upholds that sufficient volatility *creates* growth opportunities *in a financial market*.

We have already encountered an instance of this phenomenon in Remark 3.2; we saw there that, in the presence of a strong nondegeneracy condition on the market’s covariance structure, “reasonably diversified” long-only portfolios with constant weights represent superior long-term growth opportunities relative to the overall market.

We shall examine in Example 11.1 another instance of this phenomenon. We shall try again to put the above intuition on a precise quantitative basis by identifying now the

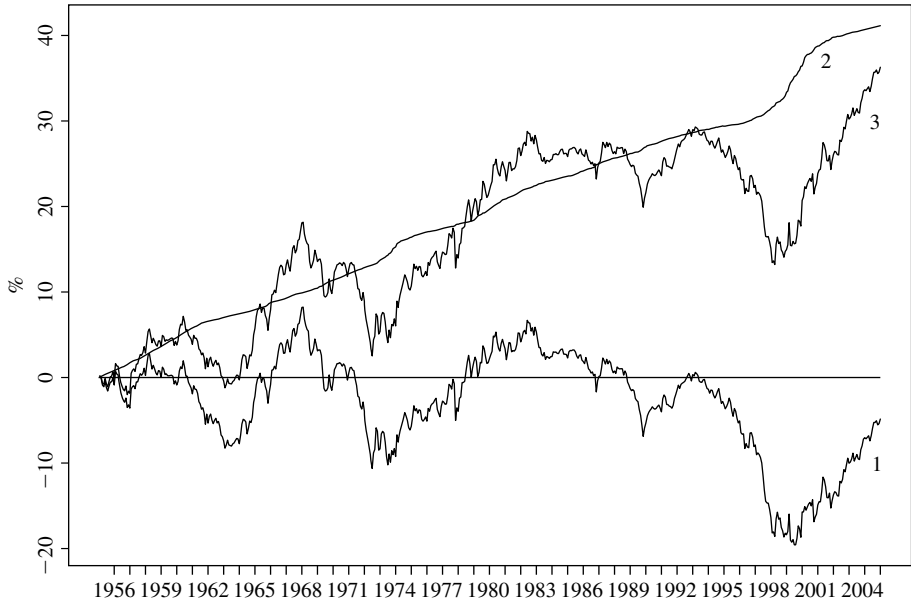


FIG. 11.1 Simulation of a $\mathbf{G}_p$ -weighted portfolio, 1956–2005. 1 generating function; 2 drift process; 3 relative return.

excess growth rate of the market portfolio, which also measures the market's intrinsic volatility, according to (3.8) and the discussion following it, as a *driver of growth*; to wit, as a quantity whose “availability” or “sufficiency” (boundedness away from zero) can lead to opportunities for strong arbitrage and for superior long-term growth relative to the market.

EXAMPLE 11.1. Suppose now that in the market $\mathcal{M}$ there exist real constants $\zeta > 0$, $T > 0$ such that

$$\frac{1}{T} \int_0^T \gamma_\mu^*(t) dt \geq \zeta \quad (11.8)$$

holds almost surely. For instance, this is the case when the excess growth rate of the market portfolio is bounded away from zero: that is, when we have almost surely

$$\gamma_\mu^*(t) \geq \zeta, \quad \forall 0 \leq t \leq T. \quad (11.9)$$

Consider again the entropy-weighted portfolio $\pi^c(\cdot)$ of (11.7), namely,

$$\pi_i^c(t) = \frac{\mu_i(t)(c - \log \mu_i(t))}{\sum_{j=1}^n \mu_j(t)(c - \log \mu_j(t))}, \quad i = 1, \dots, n, \quad (11.10)$$

now written in a form that makes plain its over weighting of the small capitalization stocks relative to the market portfolio. From (11.2), (11.7), and the inequalities of (11.6),

one sees that the portfolio $\pi^c(\cdot)$ in (11.7) satisfies

$$\begin{aligned} \log\left(\frac{V^{\pi^c}(T)}{V^\mu(T)}\right) &= \log\left(\frac{\mathbf{H}_c(\mu(T))}{\mathbf{H}_c(\mu(0))}\right) + \int_0^T \frac{\gamma_\mu^*(t)}{\mathbf{H}_c(\mu(t))} dt \\ &> -\log\left(\frac{c + \mathbf{H}(\mu(0))}{c}\right) + \frac{\zeta T}{c + \log n} \end{aligned} \quad (11.11)$$

almost surely. Thus, for every time horizon $[0, T]$ of length

$$T > \mathcal{T}_*(c) := \frac{1}{\zeta} (c + \log n) \log\left(\frac{c + \mathbf{H}(\mu(0))}{c}\right),$$

or for that matter every

$$T > \mathcal{T}_* = \frac{1}{\zeta} \mathbf{H}(\mu(0)) \quad (11.12)$$

(since $\lim_{c \rightarrow \infty} \mathcal{T}_*(c) = \mathcal{T}_*$), and for $c > 0$ sufficiently large, the portfolio $\pi^c(\cdot)$ of (11.7) satisfies the condition $\mathbb{P}(V^{\pi^c}(T) > V^\mu(T)) = 1$ for strong arbitrage relative to the market $\mu(\cdot)$, on the given time horizon $[0, T]$. It is straightforward that (6.3) is also satisfied, with $q = c/(c + \mathbf{H}(\mu(0)))$.

In particular, with the notation of (6.2), we have almost surely $\mathcal{L}^{\pi^c, \mu} \geq \zeta/(c + \log n) > 0$ (the condition for superior long-term growth for $\pi^c(\cdot)$ relative to the market $\mu(\cdot)$), provided that (11.9) holds for *all* sufficiently long time horizons $T > 0$.

It should also be noted that we have not imposed in the discussion of Example 11.1 **any** assumption on the volatility structure of the market (such as (1.15), (1.16), or (3.10)) beyond the absolutely minimal condition of (1.2).

Figure 11.2 shows the cumulative excess growth $\int_0^\cdot \gamma_\mu^*(t) dt$ for the U.S. equities market over most of the 20th century. Note the conspicuous bumps in the curve, first in the Great Depression period in the early 1930s, then again in the “irrational exuberance” period at the end of the century. The data used for this chart come from the monthly stock database of the CRSP at the University of Chicago. The market we construct consists of the stocks traded on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the NASDAQ Stock Market after the removal of all REITs, all closed-end funds, and those ADRs not included in the S&P 500 Index. Until 1962, the CRSP data included only NYSE stocks. The AMEX stocks were included after July 1962, and the NASDAQ stocks were included at the beginning of 1973. The number of stocks in this market varies from a few hundred in 1927 to about 7500 in 2005.

This computation for Fig. 11.2 does not need any estimation of covariance structure. From (11.11), we can express this cumulative excess growth

$$\int_0^\cdot \gamma_\mu^*(t) dt = \int_0^\cdot \mathbf{H}_c(\mu(t)) d \log\left(\frac{V^{\pi^c}(t) \mathbf{H}_c(\mu(0))}{V^\mu(t) \mathbf{H}_c(\mu(t))}\right)$$

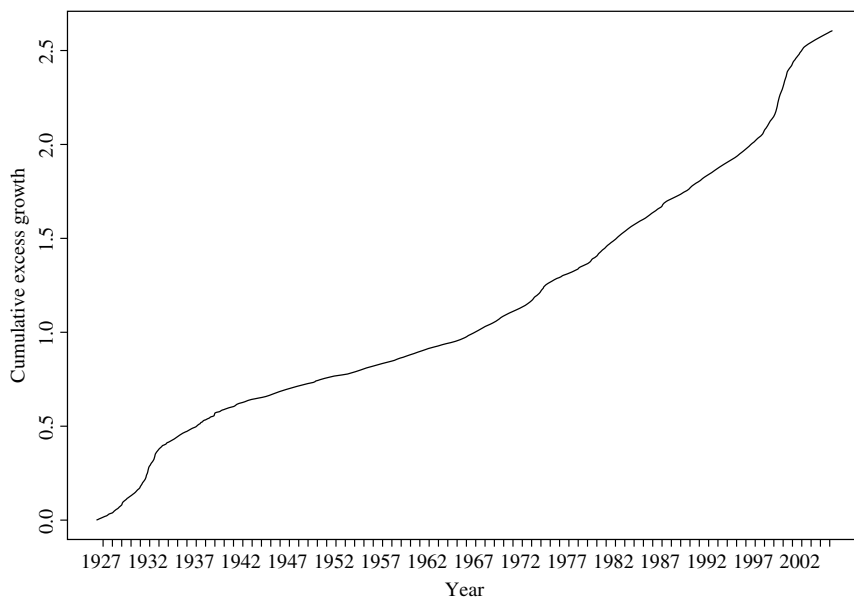


FIG. 11.2 Cumulative excess growth $\int_0^t \gamma_\mu^*(t) dt$. U.S. market, 1927–2005.

just in terms of quantities that are observable in the market. The plot suggests that the U.S. market has exhibited a strictly increasing cumulative excess growth over this period.

REMARK 11.2. Let us recall here our discussion of the conditions in (5.3): if the covariance matrix $a(\cdot)$ has all its eigenvalues bounded away from both zero and infinity, then the condition (11.9) (respectively, (11.8)) is equivalent to diversity (respectively, weak diversity) on $[0, T]$. The point of these conditions is that they guarantee the existence of strong arbitrage relative to the market, even when volatilities are unbounded and diversity fails. In the next section, we shall study a concrete example of such a situation.

REMARK 11.3. Open Question: From (11.11), it is not difficult to see that if we are allowed to start with the market arbitrarily close to the “boundary”, that is, if $\mu(0)$ can be chosen such that $\mathbf{H}(\mu(0))$ is arbitrarily small, then condition (11.9) will assure the existence of short-term arbitrage (as opposed to arbitrage over sufficiently long time intervals). Suppose now that the market can reach a point arbitrarily close to the boundary in an arbitrarily short time with positive probability. We could then use the strategy of holding the market portfolio until we arrive close enough to the boundary—which will occur, at least with positive probability—and then switch to the arbitrage portfolio, so short-term arbitrage will again be possible. However, strong arbitrage, in the sense that

$$\mathbb{P}[V^\pi(T) > V^\mu(T)] = 1$$

in (6.1), cannot be assured by this argument. Indeed, it seems to be an open problem whether or not condition (11.9) implies strong arbitrage relative to the market over arbitrarily short time periods.

REMARK 11.4. Example and open questions: For $0 < p \leq 1$, the quantity

$$\gamma_{\pi,p}^*(t) := \frac{1}{2} \sum_{i=1}^n (\pi_i(t))^p \tau_{ii}^\pi(t) \quad (11.13)$$

generalizes the excess growth rate of a portfolio $\pi(\cdot)$, in the sense that $\gamma_{\pi,1}^*(\cdot) \equiv \gamma_\pi^*(\cdot)$. With $0 < p < 1$, consider the a.s. requirement

$$\Gamma(T) = \int_0^T \gamma_{p,\mu}^*(t) dt < \infty, \quad \forall 0 \leq T < \infty, \quad (11.14)$$

for some continuous, strictly increasing function $\Gamma : [0, \infty) \rightarrow [0, \infty)$, with $\Gamma(0) = 0$, $\Gamma(\infty) = \infty$. As shown in proposition 3.8 of FERNHOLZ and KARATZAS [2005], the condition (11.14) guarantees that the portfolio

$$\pi_i(t) := p \cdot \frac{(\mu_i(t))^p}{\sum_{j=1}^n (\mu_j(t))^p} + (1-p) \cdot \mu_i(t), \quad i = 1, \dots, n \quad (11.15)$$

is a strong arbitrage opportunity relative to the market, namely, that $\mathbb{P}[V^\pi(T) > V^\mu(T)] = 1$ holds over sufficiently long time horizons: $T > \Gamma^{-1}((1/p)n^{1-p} \log n)$.

Note that the portfolio of (11.15) is a convex combination, with fixed weights $1-p$ and p , of the market and of its diversity-weighted index $\mu^{(p)}(\cdot)$ in (7.1), respectively.

Some questions suggest themselves:

- Does (11.14) guarantee the existence of relative arbitrage opportunities over arbitrary time horizons?
- Is there a result on the existence of relative arbitrage that generalizes both Example 11.1 and the result outlined in (11.14) and (11.15)?
- What quantity or quantities might then be involved in place of the market excess growth or its generalization (11.14)? Is there a “best” result of this type?

EXAMPLE 11.2. Equal Weighting: Recall the computation (7.11) for the excess growth rate of the equally weighted portfolio $\varphi_i(\cdot) \equiv 1/n$, $i = 1, \dots, n$, and suppose that

$$(\mu_1(t) \dots \mu_n(t))^{1/n} \gamma_\varphi^*(t) \geq \zeta, \quad 0 \leq t \leq T \quad (11.16)$$

holds a.s. for some real constant $\zeta > 0$.

Recall also the modification $\varphi^c(\cdot)$ of this portfolio, as in (11.4); this is generated by the function $\mathbf{F}_c(x) = c + \mathbf{F}(x)$, with $c > 0$ and $\mathbf{F}(x) := (x_1 \dots x_n)^{1/n} \in (0, n^{-1/n}]$, $x \in \Delta_+^n$. From (11.5) and (11.2), we deduce the a.s. comparisons

$$\log \left(\frac{V^{\varphi^c}(T)}{V^\mu(T)} \right) = \log \left(\frac{c + \mathbf{F}(\mu(T))}{c + \mathbf{F}(\mu(0))} \right) + \int_0^T \frac{\mathbf{F}(\mu(t)) \gamma_\varphi^*(t)}{c + \mathbf{F}(\mu(t))} dt$$

and

$$\log\left(\frac{V^{\varphi^c}(T)}{V^\mu(T)}\right) \geq \log\left(\frac{c}{c + n^{-1/n}}\right) + \frac{\zeta T}{c + n^{-1/n}} \quad (11.17)$$

for the portfolio $\varphi^c(\cdot)$ of (11.4). Therefore, we have $\mathbb{P}(V^{\varphi^c}(T) > V^\mu(T)) = 1$, provided that $T > \frac{1}{\zeta} (c + n^{-1/n}) \cdot \log(c/(c + n^{-1/n}))$. Consequently, if the time horizon is sufficiently long, to wit,

$$T > T_* := \frac{1}{\zeta} n^{-1/n},$$

there exists a number $c \in (0, \infty)$ such that the market-modulated equally weighted portfolio $\varphi^c(\cdot)$ of (11.4) is a strong arbitrage relative to the market.

REMARK 11.5. Open question: We have presented a few portfolios that lead to arbitrage relative to the market; they are all functionally generated. Is there a “best” such example within that class? Are there similar examples of portfolios that are *not* functionally generated or trivial modifications thereof? How representative (or “dense”) in this context is the class of functionally generated portfolios?

REMARK 11.6. Open question: Generalize the theory of functionally generated portfolios to the case of a market with a countable infinity ($n = \infty$) of assets or to some other model with a variable, unbounded number of assets.

REMARK 11.7. Open question: What, if any, is the connection of functionally generated portfolios with the “universal portfolios” of COVER [1991] and JAMSHIDIAN [1992]?

11.2. Rank, leakage, and the size effect

An important generalization of the ideas and methods in this section concerns generating functions that record market weights not according to their name (or index) i , but according to their rank. To present this generalization, let us start by recalling the order statistics notation of (1.18), and consider for each $0 \leq t < \infty$ the random permutation $(p_t(1), \dots, p_t(n))$ of $(1, \dots, n)$ with

$$\mu_{p_t(k)}(t) = \mu_{(k)}(t) \quad \text{and} \quad p_t(k) < p_t(k+1) \quad \text{if} \quad \mu_{(k)}(t) = \mu_{(k+1)}(t) \quad (11.18)$$

for $k = 1, \dots, n$. In words, $p_t(k)$ is the name (index) of the stock that occupies the k th rank in terms of relative capitalization at time t ; ties are resolved by resorting to the lowest index.

Using Itô’s rule for convex functions of semimartingales (KARATZAS and SHREVE [1991], Section 3.7), one can obtain the following analogue of (2.5) for the ranked

market weights

$$\begin{aligned} \frac{d\mu_{(k)}(t)}{\mu_{(k)}(t)} &= \left(\gamma_{p_t(k)}(t) - \gamma^\mu(t) + \frac{1}{2} \tau_{(kk)}^\mu(t) \right) dt + \frac{1}{2} \left(d\mathfrak{L}^{k,k+1}(t) - d\mathfrak{L}^{k-1,k}(t) \right) \\ &\quad + \sum_{v=1}^d \left(\sigma_{p_t(k)v}(t) - \sigma_v^\mu(t) \right) dW_v(t) \end{aligned} \quad (11.19)$$

for each $k = 1, \dots, n-1$. Here, the quantity $\mathfrak{L}^{k,k+1}(t) \equiv \Lambda_{\Xi_k}(t)$ is the semimartingale local time at the origin accumulated by the nonnegative process

$$\Xi_k(t) := \log(\mu_{(k)}/\mu_{(k+1)})(t) \quad 0 \leq t < \infty \quad (11.20)$$

up to the calendar time t ; it measures the cumulative effect of the changes that have occurred during the time interval $[0, t]$ between ranks k and $k+1$. We are also setting $\mathfrak{L}^{0,1}(\cdot) \equiv 0$, $\mathfrak{L}^{m,m+1}(\cdot) \equiv 0$, and $\tau_{(k\ell)}^\mu(\cdot) := \tau_{p_t(k)p_t(\ell)}^\mu(\cdot)$.

A derivation of this result, under appropriate conditions that we choose not to broach here, can be found on pp 76–79 of FERNHOLZ [2002]; see also BANNER and GHOMRASNI [2008] for generalizations.

With this setup, we have then the following generalization of the “master equation” (11.2): Consider a function $\mathbf{G} : U \rightarrow (0, \infty)$ exactly as assumed there, written in the form

$$\mathbf{G}(x_1, \dots, x_n) = \mathfrak{G}(x_{(1)}, \dots, x_{(n)}), \quad \forall \quad x \in U$$

for some $\mathfrak{G} \in \mathcal{C}^2(U)$ and U an open neighborhood of Δ_+^n . Introduce the shorthand

$$\begin{aligned} x_{(\cdot)} &:= (x_{(1)}, \dots, x_{(n)})', \quad \mu_{(\cdot)}(t) := (\mu_{(1)}(t), \dots, \mu_{(n)}(t))', \\ \tau_{(k\ell)}^\mu(t) &:= \tau_{p_t(k)p_t(\ell)}^\mu(t) \end{aligned}$$

as well as the notation

$$\begin{aligned} \Gamma(T) &:= - \int_0^T \frac{1}{2\mathfrak{G}(\mu_{(\cdot)}(t))} \sum_{k=1}^n \sum_{\ell=1}^n D_{k\ell}^2 \mathfrak{G}(\mu_{(\cdot)}(t)) \mu_{(k)}(t) \mu_{(\ell)}(t) \tau_{(k\ell)}^\mu(t) dt \\ &\quad + \frac{1}{2} \sum_{k=1}^{n-1} (\pi_{p_t(k+1)}(t) - \pi_{p_t(k)}(t)) d\mathfrak{L}^{k,k+1}(t). \end{aligned} \quad (11.21)$$

Then, it can be shown that the performance of the portfolio $\pi(\cdot)$ given as

$$\pi_{p_t(k)}(t) = \left(D_k \log \mathfrak{G}(\mu_{(\cdot)}(t)) + 1 - \sum_{\ell=1}^n \mu_{(\ell)}(t) D_\ell \log \mathfrak{G}(\mu_{(\cdot)}(t)) \right) \mu_{(k)}(t) \quad (11.22)$$

for $1 \leq k \leq n$, relative to the market, is

$$\log\left(\frac{V^\pi(T)}{V^\mu(T)}\right) = \log\left(\frac{\mathfrak{G}(\mu_{(\cdot)}(T))}{\mathfrak{G}(\mu_{(\cdot)}(0))}\right) + \Gamma(T), \quad 0 \leq T < \infty. \quad (11.23)$$

We say that $\pi(\cdot)$ is the portfolio generated by the function $\mathfrak{G}(\cdot)$. The detailed proof can be found in FERNHOLZ [2002], pp 79–83.

For instance, $\mathfrak{G}(x_{(\cdot)}) = x_{(1)}$ generates the portfolio $\pi_{p_t(k)}(t) = \delta_{1k}$, $k = 1, \dots, n$, $0 \leq t < \infty$ that invests only in the largest stock at all times. The relative performance

$$\log\left(\frac{V^\pi(T)}{V^\mu(T)}\right) = \log\left(\frac{\mu_{(1)}(T)}{\mu_{(1)}(0)}\right) - \frac{1}{2} \mathfrak{L}^{1,2}(T), \quad 0 \leq T < \infty$$

of this portfolio will suffer in the long run, if there are many changes in leadership in order for the biggest stock to do well relative to the market, it must crush all competition!

EXAMPLE 11.3. The Size Effect: This is the tendency of small stocks to have higher long-term returns relative to their larger brethren. The formula of (11.23) offers a simple, structural explanation of this observed phenomenon, as follows.

Fix an integer $m \in \{2, \dots, n-1\}$ and consider the functions $\mathcal{G}_L(x) = x_{(1)} + \dots + x_{(m)}$ and $\mathcal{G}_S(x) = x_{(m+1)} + \dots + x_{(n)}$. These generate, respectively, a large-stock portfolio

$$\zeta_{p_t(k)}(t) = \frac{\mu_{(k)}(t)}{\mathcal{G}_L(\mu(t))}, \quad k = 1, \dots, m \quad \text{and} \quad \zeta_{p_t(k)}(t) = 0, \quad k = m+1, \dots, n \quad (11.24)$$

and a small-stock portfolio

$$\eta_{p_t(k)}(t) = \frac{\mu_{(k)}(t)}{\mathcal{G}_S(\mu(t))}, \quad k = m+1, \dots, n \quad \text{and} \quad \eta_{p_t(k)}(t) = 0, \quad k = 1, \dots, m. \quad (11.25)$$

According to (11.23) and (11.22), the performances of these portfolios, relative to the market, are given by

$$\log\left(\frac{V^\zeta(T)}{V^\mu(T)}\right) = \log\left(\frac{\mathcal{G}_L(\mu(T))}{\mathcal{G}_L(\mu(0))}\right) - \frac{1}{2} \int_0^T \zeta_{(m)}(t) d\mathfrak{L}^{m,m+1}(t), \quad (11.26)$$

$$\log\left(\frac{V^\eta(T)}{V^\mu(T)}\right) = \log\left(\frac{\mathcal{G}_S(\mu(T))}{\mathcal{G}_S(\mu(0))}\right) + \frac{1}{2} \int_0^T \eta_{(m)}(t) d\mathfrak{L}^{m,m+1}(t), \quad (11.27)$$

respectively. Therefore,

$$\log\left(\frac{V^\eta(T)}{V^\zeta(T)}\right) = \log\left(\frac{\mathcal{G}_S(\mu(T))\mathcal{G}_L(\mu(0))}{\mathcal{G}_L(\mu(T))\mathcal{G}_S(\mu(0))}\right) + \int_0^T \frac{\zeta_{(m)}(t) + \eta_{(m)}(t)}{2} d\mathfrak{L}^{m,m+1}(t). \quad (11.28)$$

If there is “stability” in the market, in the sense that the ratio of the relative capitalization of small to large stocks remains stable over time, then the first term on the right-hand side of (11.28) does not change much, whereas the second term keeps increasing and accounts for the better relative performance of the small stocks. Note that this argument does not need to invoke any assumption about the putative greater riskiness of the smaller stocks at all.

FERNHOLZ and KARATZAS [2006] studied conditions under which such stability in relative capitalizations prevails and further discussed the “liquidity premium” for equities.

REMARK 11.8. Estimation of Local Times: Hard as this might be to have guessed from the outset, the local times $\mathfrak{L}^{k,k+1}(\cdot) \equiv \Lambda_{\Xi_k}(\cdot)$ appearing in (11.19) and (11.21) can be estimated in practice quite accurately; indeed, (11.26) gives

$$\mathfrak{L}^{m,m+1}(\cdot) = \int_0^\cdot \frac{2}{\zeta_{(m)}(t)} d\log\left(\frac{\mathcal{G}_L(\mu(t))}{\mathcal{G}_L(\mu(0))} \frac{V^\mu(t)}{V^\zeta(t)}\right), \quad m = 1, \dots, n-1, \quad (11.29)$$

and the quantity on the right-hand side is completely observable.

REMARK 11.9. Leakage in a Diversity-Weighted Index of Large Stocks: With the integer m and the large-stock portfolio $\zeta(\cdot)$ as in Example 11.3, and a fixed number $r \in (0, 1)$, consider the diversity-weighted, large-stock portfolio

$$\mu_{p_t(k)}^\sharp(t) = \frac{(\mu_{(k)}(t))^r}{\sum_{\ell=1}^m (\mu_{(\ell)}(t))^r}, \quad 1 \leq k \leq m \quad \text{and} \quad \mu_{p_t(k)}^\sharp(t) = 0, \quad m+1 \leq k \leq n \quad (11.30)$$

generated by the function $\mathbb{G}_r(x) = \left(\sum_{\ell=1}^m (x_{(\ell)})^r\right)^{1/r}$, by analogy with (7.4) and (7.1). Then,

$$\begin{aligned} \log\left(\frac{V^{\mu^\sharp}(T)}{V^\mu(T)}\right) &= \log\left(\frac{\mathbb{G}_r(\mu(T))}{\mathbb{G}_r(\mu(0))}\right) + (1-r) \int_0^T \gamma_{\mu^\sharp}^*(t) dt \\ &\quad - \int_0^T \frac{\mu_{(m)}^\sharp(t)}{2} d\mathfrak{L}^{m,m+1}(t) \end{aligned}$$

gives the performance of the portfolio in (11.30) relative to the market, and

$$\begin{aligned} \log \left(\frac{V^{\mu^\sharp}(T)}{V^\zeta(T)} \right) &= \log \left(\frac{\mathbf{G}_r(\zeta_{(1)}(T), \dots, \zeta_{(m)}(T))}{\mathbf{G}_r(\zeta_{(1)}(0), \dots, \zeta_{(m)}(0))} \right) + (1-r) \int_0^T \gamma_{\mu^\sharp}^*(t) dt \\ &\quad - \frac{1}{2} \int_0^T (\mu_{(m)}^\sharp(t) - \zeta_{(m)}(t)) d\mathfrak{L}^{m,m+1}(t) \end{aligned} \quad (11.31)$$

gives the performance of (11.30) relative to the large-stock portfolio $\zeta(\cdot)$ of (11.24). We have used here the scale-invariance property

$$\frac{\mathbf{G}_r(x_1, \dots, x_n)}{x_1 + \dots + x_n} = \mathbf{G}_r \left(\frac{x_1}{x_1 + \dots + x_n}, \dots, \frac{x_n}{x_1 + \dots + x_n} \right)$$

of the diversity function $\mathbf{G}_r(\cdot)$ in (7.4) for $0 < r < 1$, which implies the reduction

$$\frac{\mathbb{G}_r(\mu(t))}{\mathcal{G}_L(\mu(t))} = \mathbf{G}_r(\zeta_{(1)}(t), \dots, \zeta_{(m)}(t)).$$

Since $\mu_{(m)}^\sharp(\cdot) \geq \zeta_{(m)}(\cdot)$ from (7.7) and the remark following it, the last term in (11.31) is monotonically increasing in T . It measures the “leakage” that occurs when a capitalization-weighted portfolio is contained inside a larger market, and stocks cross-over (“leak”) from the cap-weighted to the market portfolio. For details of these derivations, see FERNHOLZ [2002], pp. 84–88.

Proof of the “Master Equation” (11.2). To ease notation, we set

$$g_i(t) := D_i \log \mathbf{G}(\mu(t)) \quad \text{and} \quad N(t) := 1 - \sum_{j=1}^n \mu_j(t) g_j(t),$$

so (11.1) reads $\pi_i(t) = (g_i(t) + N(t))\mu_i(t)$, $i = 1, \dots, n$. Then, the terms on the right-hand side of (3.9) become

$$\sum_{i=1}^n \frac{\pi_i(t)}{\mu_i(t)} d\mu_i(t) = \sum_{i=1}^n g_i(t) d\mu_i(t) + N(t) \cdot d \left(\sum_{i=1}^n \mu_i(t) \right) = \sum_{i=1}^n g_i(t) d\mu_i(t)$$

and

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \pi_i(t) \pi_j(t) \tau_{ij}^\mu(t) &= \sum_{i=1}^n \sum_{j=1}^n (g_i(t) + N(t))(g_j(t) + N(t)) \mu_i(t) \mu_j(t) \tau_{ij}^\mu(t) \\ &= \sum_{i=1}^n \sum_{j=1}^n g_i(t) g_j(t) \mu_i(t) \mu_j(t) \tau_{ij}^\mu(t), \end{aligned}$$

the latter thanks to (1.21) and Lemma 3.1. Thus, (3.9) gives

$$d \log \left(\frac{V^\pi(t)}{V^\mu(t)} \right) = \sum_{i=1}^n g_i(t) d\mu_i(t) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n g_i(t) g_j(t) \mu_i(t) \mu_j(t) \tau_{ij}^\mu(t) dt. \quad (11.32)$$

On the other hand, we have

$$D_{ij}^2 \log \mathbf{G}(x) = (D_{ij}^2 G(x)/G(x)) - D_i \log \mathbf{G}(x) \cdot D_j \log \mathbf{G}(x),$$

so we get

$$\begin{aligned} d \log \mathbf{G}(\mu(t)) &= \sum_{i=1}^n g_i(t) d\mu_i(t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij}^2 \log \mathbf{G}(\mu(t)) d\langle \mu_i, \mu_j \rangle(t) \\ &= \sum_{i=1}^n g_i(t) d\mu_i(t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{D_{ij}^2 \mathbf{G}(\mu(t))}{\mathbf{G}(\mu(t))} - g_i(t) g_j(t) \right) \\ &\quad \cdot \mu_i(t) \mu_j(t) \tau_{ij}^\mu(t) dt \end{aligned}$$

by Itô's rule in conjunction with (2.6). Comparing this last expression with (11.32) and recalling (11.3), we deduce (11.2), namely, $d \log \mathbf{G}(\mu(t)) = d \log (V^\pi(t)/V^\mu(t)) - g(t)dt$. $\square$

Abstract Markets

The basic market model in (1.1) is too general for us to be able to draw many interesting conclusions. Hence, we would like to consider a more restricted class of models that still capture certain aspects of real equity markets, but are more analytically tractable than the general model (1.1). Abstract markets are relatively simple stochastic equity market models that exhibit selected characteristics of real equity markets, so that an understanding of these models will provide some insight into the behavior of actual markets. In particular, there are two classes of abstract markets that we shall discuss here: volatility-stabilized markets introduced by FERNHOLZ and KARATZAS [2005], and rank-based models exemplified by Atlas models and their generalizations, which first appeared in FERNHOLZ [2002], with further development by BANNER, FERNHOLZ and KARATZAS [2005].

12. Volatility-stabilized markets

Volatility-stabilized market models are remarkable because in these models the market itself behaves in a rather sedate fashion, viz. (exponential) Brownian motion with drift, while the individual stocks are going all over the place (in a rigorously defined manner, of course). These models reflect the fact that in real markets, the smaller stocks tend to have greater volatility than the larger stocks.

Let us consider the abstract market model $\mathcal{M}$ with

$$d \log X_i(t) = \frac{\alpha}{2\mu_i(t)} dt + \frac{1}{\sqrt{\mu_i(t)}} dW_i(t), \quad i = 1, \dots, n, \quad (12.1)$$

where $\alpha \geq 0$ is a given real constant. The theory developed by BASS and PERKINS [2002] shows that the resulting system of stochastic differential equations, for $i = 1, \dots, n$,

$$dX_i(t) = \frac{1 + \alpha}{2} (X_1(t) + \dots + X_n(t)) dt + \sqrt{X_i(t)(X_1(t) + \dots + X_n(t))} dW_i(t), \quad (12.2)$$

determines the distribution of the Δ_+^n -valued diffusion process $\mathfrak{X}(\cdot) = (X_1(\cdot), \dots, X_n(\cdot))'$ uniquely, and that the conditions of (1.2), (6.4) are satisfied by the processes

$$b_i(\cdot) = (1 + \alpha)/2\mu_i(\cdot), \quad \sigma_{iv}(t) = (\mu_i(t))^{-1/2}\delta_{iv}, \quad r(\cdot) \equiv 0, \quad \text{and} \\ \theta_v(\cdot) = (1 + \alpha)/2\sqrt{\mu_v(\cdot)}$$

for $1 \leq i, v \leq n$. The reader might wish to remark that condition (3.10) is satisfied in this case, in fact with $\varepsilon = 1$; but (1.16) fails.

The model of (12.1) assigns to all stocks log-drifts $\gamma_i(t) = \alpha/2\mu_i(t)$, covariances $a_{ij}(t) = 0$ for $j \neq i$, and variances $a_{ii}(t) = 1/\mu_i(t)$, $i = 1, \dots, n$ that are largest for the smallest stocks and smallest for the largest stocks. Not surprisingly then, individual stocks fluctuate rather widely in a market of this type; in particular, diversity fails on every $[0, T]$ (see Remarks 12.2 and 12.3).

Yet despite these fluctuations, the overall market has quite stable behavior. We call this phenomenon stabilization by volatility in the case $\alpha = 0$, and stabilization by both volatility and drift in the case $\alpha > 0$.

Indeed, the quantities $a_{\mu\mu}(\cdot)$, $\gamma_\mu^*(\cdot)$, and $\gamma_\mu(\cdot)$ are computed from (1.20) and (1.13), (1.12) as

$$a_{\mu\mu}(\cdot) \equiv 1, \quad \gamma_\mu^*(\cdot) \equiv \gamma^* := \frac{n-1}{2} > 0, \quad \text{and} \quad \gamma_\mu(\cdot) \equiv \gamma := \frac{(1+\alpha)n-1}{2} > 0. \quad (12.3)$$

This, in conjunction with (2.2), computes the total market capitalization

$$X(t) = X_1(t) + \dots + X_n(t) = X(0)e^{\gamma t + \mathcal{W}(t)}, \quad 0 \leq t < \infty \quad (12.4)$$

as the exponential of the standard, one-dimensional Brownian motion $\mathcal{W}(\cdot) := \sum_{v=1}^n \int_0^\cdot \sqrt{\mu_v(s)} dW_v(s)$, plus drift $\gamma t > 0$. In particular, the overall market and the largest stock $X_{(1)}(\cdot) = \max_{1 \leq i \leq n} X_i(\cdot)$ grow at the same constant rate:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log X(T) = \lim_{T \rightarrow \infty} \frac{1}{T} \log X_{(1)}(T) = \gamma, \quad \text{a.s.} \quad (12.5)$$

On the other hand, according to Example 11.1, there exist in this model portfolios that lead to strong arbitrage opportunities relative to the market, at least on time horizons $[0, T]$ with $T \in (\mathcal{T}_*, \infty)$, where

$$\mathcal{T}_* := \frac{2\mathbf{H}(\mu(0))}{n-1} \leq \frac{2\log n}{n-1}. \quad (12.6)$$

To wit, strong relative arbitrage can exist in non-diverse markets with unbounded volatilities.

The last upper bound in the above expression (12.6) becomes small as the number of stocks in the market increases. In fact, BANNER and FERNHOLZ [2007] provided recently an elaborate construction which shows that strong arbitrage exists, relative to the market described by (12.1), over *arbitrary* time horizons.

12.1. Bessel processes

The crucial observation now is that the solution of the system (12.1) can be expressed in terms of the squares of independent Bessel processes $\mathfrak{R}_1(\cdot), \dots, \mathfrak{R}_n(\cdot)$ in dimension $\kappa := 2(1+\alpha) \geq 2$ and of an appropriate time change:

$$X_i(t) = \mathfrak{R}_i^2(\Lambda(t)), \quad 0 \leq t < \infty, \quad i = 1, \dots, n, \quad (12.7)$$

where

$$\Lambda(t) := \frac{1}{4} \int_0^t X(u) du = \frac{X(0)}{4} \int_0^t e^{\gamma s + \mathcal{W}(s)} ds, \quad 0 \leq t < \infty \quad (12.8)$$

and

$$\mathfrak{R}_i(u) = \sqrt{X_i(0)} + \frac{\kappa - 1}{2} \int_0^u \frac{d\xi}{\mathfrak{R}_i(\xi)} + \mathfrak{W}_i(u), \quad 0 \leq u < \infty. \quad (12.9)$$

Here, the driving processes $\mathfrak{W}_i(\cdot) := \int_0^{\Lambda^{-1}(\cdot)} \sqrt{\Lambda'(t)} dW_i(t)$, $i = 1, \dots, n$ are independent, standard one-dimensional Brownian motions (KARATZAS and SHREVE [1991], pp 157–162). In a similar vein, we have the representation

$$X(t) = \mathfrak{R}^2(\Lambda(t)), \quad 0 \leq t < \infty$$

of the total market capitalization, in terms of the Bessel process

$$\mathfrak{R}(u) = \sqrt{X(0)} + \frac{n\kappa - 1}{2} \int_0^u \frac{d\xi}{\mathfrak{R}(\xi)} + \mathfrak{W}(u), \quad 0 \leq u < \infty \quad (12.10)$$

in dimension $n\kappa$, and of yet another one-dimensional Brownian motion $\mathfrak{W}(\cdot)$.

This observation provides a wealth of structure, which can be used then to study the asymptotic properties of the model (12.1).

REMARK 12.1. For the case $\alpha > 0$ ($\kappa > 2$), we have for each $i = 1, \dots, n$ the ergodic property

$$\lim_{u \rightarrow \infty} \frac{1}{\log u} \int_0^u \frac{d\xi}{\mathfrak{R}_i^2(\xi)} = \frac{1}{\kappa - 2} = \frac{1}{2\alpha}, \quad \text{a.s.}$$

(a consequence of the Birkhoff ergodic theorem and of the strong Markov property of the Bessel process), as well as the Lamperti representation

$$\mathfrak{R}_i(u) = \sqrt{x_i} e^{\alpha\theta + \mathfrak{B}_i(\theta)} \Big|_{\theta = \int_0^u \frac{d\xi}{\mathfrak{R}_i^2(\xi)}}, \quad 0 \leq u < \infty$$

for the Bessel process $\mathfrak{R}_i(\cdot)$ in terms of the exponential of a standard Brownian motion $\mathfrak{B}_i(\cdot)$ with positive drift $\alpha > 0$. From these considerations, one can deduce the a.s. properties

$$\lim_{u \rightarrow \infty} \frac{\log \mathfrak{R}_i(u)}{\log u} = \frac{1}{2}, \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log X_i(t) = \gamma, \quad (12.11)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a_{ii}(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{dt}{\mu_i(t)} = \frac{2\gamma}{\alpha} = n + \frac{n-1}{\alpha}, \quad (12.12)$$

for each $i = 1, \dots, n$ (see pp. 174–175 in FERNHOLZ and KARATZAS [2005] for details). In particular, all stocks grow at the same asymptotic rate $\gamma > 0$ of (12.3), as does the entire market; the model of (12.1) is coherent in the sense of Remark 2.1; and the conditions (1.6) and (1.7) hold.

REMARK 12.2. In the case $\alpha = 0$ ($\kappa = 2$), it can be shown that

$$\lim_{u \rightarrow \infty} \frac{\log \mathfrak{R}_i(u)}{\log u} = \frac{1}{2} \quad \text{holds in probability,} \quad (12.13)$$

but that we have almost surely

$$\limsup_{u \rightarrow \infty} \frac{\log \mathfrak{R}_i(u)}{\log u} = \frac{1}{2}, \quad \liminf_{u \rightarrow \infty} \frac{\log \mathfrak{R}_i(u)}{\log u} = -\infty. \quad (12.14)$$

It follows from this and (12.5) that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log X_i(t) = \gamma \quad \text{holds in probability,} \quad (12.15)$$

and also that

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log X_i(t) = \gamma, \quad \liminf_{t \rightarrow \infty} \frac{1}{t} \log X_i(t) = -\infty \quad (12.16)$$

hold almost surely, for each $i = 1, \dots, n$. To wit, individual stocks can “crash” in this case, despite the overall stability of the market, and coherence now fails, as does the condition (1.6).

(Note: The claim (12.13) comes from the observation

$$\mathfrak{R}_i(u) = \|\mathfrak{R}_i(0) + \mathfrak{b}_i(u)\| = \sqrt{u} \|\mathfrak{R}_i(0)/\sqrt{u} + \mathfrak{b}_i(1)\| \quad \text{in distribution,}$$

where $\mathfrak{R}_i(\cdot)$ and $\mathfrak{b}_i(\cdot)$ are Brownian motions on the plane and on the real line, respectively; thus, we have $\lim_{u \rightarrow \infty} (\log \mathfrak{R}_i(u) - (1/2) \log u) = \log \|\mathfrak{b}_i(1)\|$ in distribution and (12.13) follows.

As for (12.14), its first claim follows from the law of the iterated logarithm for Brownian motion on the real line whereas the second claim is obtained from the following result:

For a decreasing function $h(\cdot)$, we have

$$\mathbb{P}(\mathfrak{R}_i(u) \geq u^{1/2} h(u) \text{ for all } u > 0 \text{ sufficiently large}) = 1 \text{ or } 0,$$

depending on whether the series $\sum_{k \in \mathbb{N}} (k |\log h(k)|)^{-1}$ converges or diverges.

This zero-one law is due to SPITZER [1958]; details of the argument can be found on pp. 176–177 of FERNHOLZ and KARATZAS [2005].)

REMARK 12.3. In the case $\alpha = 0$ ($\kappa = 2$), it can be shown that

$$\lim_{u \rightarrow \infty} \mathbb{P}(\mu_i(\Lambda^{-1}(u)) > 1 - \delta) = \delta^{n-1}$$

holds for every $i = 1, \dots, n$ and $\delta \in (0, 1)$; here $\Lambda^{-1}(\cdot) = 4 \int_0^\cdot \mathfrak{R}^{-2}(\xi) d\xi$ is the inverse of the time change $\Lambda(\cdot)$ in (12.8), and $\mathfrak{R}(\cdot)$ is the Bessel process in (12.10). It follows that this model is not diverse on $[0, \infty)$.

REMARK 12.4. The exponential strict local martingale of (6.5) can be computed as

$$Z(T) = \exp \left\{ \frac{\alpha^2 - 1}{8} \sum_{i=1}^n \int_0^T \frac{X_1(t) + \dots + X_n(t)}{X_i(t)} dt \right\} \\ \cdot \left(\frac{X_1(0) \dots X_n(0)}{X_1(T) \dots X_n(T)} \right)^{(1+\alpha)/2}.$$

Thus, the log-optimal trading strategy $h_*(\cdot)$ and its associated wealth process $V_*(\cdot) \equiv \mathcal{V}^{1, h^*}(\cdot)$ of Remark 10.5, are given as $V_*(\cdot) = 1/Z(\cdot)$ and $h_i^*(\cdot) = (1 + \alpha)V_*(\cdot)/2$, $i = 1, \dots, n$.

For $\alpha > 0$, we deduce from this and from (12.11) and (12.12) that we have the following a.s. growth rates:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log V_*(T) = n\gamma(1 + \alpha)^2/4\alpha$$

and therefore

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V_*(T)}{V^\mu(T)} \right) = \left(\frac{n(1 + \alpha)^2}{4\alpha} - 1 \right) \gamma \\ = \left(\frac{n(1 + \alpha)^2}{4\alpha} - 1 \right) \frac{(1 + \alpha)n - 1}{2}. \quad (12.17)$$

EXAMPLE 12.1. **Diversity Weighting:** In the context of the volatility-stabilized model of this section with $p = 1/2$, the diversity-weighted portfolio

$$\mu_i^{(p)}(t) = \frac{\sqrt{\mu_i(t)}}{\sum_{j=1}^n \sqrt{\mu_j(t)}}, \quad i = 1, \dots, n$$

of (7.1) represents a strong arbitrage relative to the market portfolio, namely,

$$\mathbb{P}[V^{\pi^{(p)}}(T) > V^\mu(T)] = 1, \quad \text{at least on time horizons } [0, T] \text{ with } T > \frac{8 \log n}{n - 1}.$$

Furthermore, this diversity-weighted portfolio outperforms considerably the market over long time horizons

$$\mathcal{L}^{\mu^{(p)}, \mu} := \liminf_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V^{\mu^{(p)}}(T)}{V^\mu(T)} \right) \\ = \liminf_{T \rightarrow \infty} \frac{1}{2T} \int_0^T \gamma_{\mu^{(p)}}^*(t) dt \geq \frac{n - 1}{8}, \quad \text{a.s.}$$

Question: Do the indicated limits exist? Can they be computed in closed form?

EXAMPLE 12.2. Equal Weighting: With a covariance structure of the form $a_{ij}(t) = (1/\mu_i(t)) \delta_{ij}$, as in the volatility-stabilized model of the present section, the excess growth rate $\gamma_\varphi^*(\cdot)$ in (7.11) for the equally weighted portfolio $\varphi(\cdot)$ of Remark 7.2 takes the form

$$\gamma_\varphi^*(\cdot) = \frac{n-1}{2n^2} \sum_{i=1}^n \frac{1}{\mu_i(t)}.$$

The geometric-mean/harmonic-mean inequality now implies that the condition (11.16) is satisfied by the constant $\zeta = (n-1)/2n$; thus, according to Example 11.2, the market-modulated, equally weighted portfolio $\varphi^c(\cdot)$ of (11.4) is a strong arbitrage opportunity relative to the market, over time horizons $[0, T]$ with $T > 2n^{1-(1/n)}/(n-1)$, provided that $c > 0$ is chosen sufficiently large in (11.4).

How much better is equal weighting, relative to the volatility-stabilized market of this section with $\alpha > 0$, over very large time horizons? In conjunction with (7.10) and the coherence property of this market, the strong law of large numbers (12.12) implies that the limit

$$\mathcal{L}^{\varphi, \mu} := \lim_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{V^\varphi(T)}{V^\mu(T)} \right) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \gamma_\varphi^*(t) dt$$

of (6.2) exists a.s., and equals

$$\mathcal{L}^{\varphi, \mu} = \frac{n-1}{2} \left(1 + \frac{n-1}{n\alpha} \right). \quad (12.18)$$

In other words, equal weighting, with its built-in “buying low and selling high” features, outperforms *considerably* this drift- and volatility-stabilized market over long time horizons.

EXAMPLE 12.3. Growth Optimality: For the volatility-stabilized model of this section with $0 < \alpha < 1$ and $\lambda := \gamma + (1/2) = n(1 + \alpha)/2 \geq 1$, the portfolio

$$\hat{\pi}_i(t) := \frac{1 + \alpha}{2} - \left(\frac{n}{2}(1 + \alpha) - 1 \right) \mu_i(t) = \lambda \varphi_i(t) - (\lambda - 1) \mu_i(t), \quad i = 1, \dots, n \quad (12.19)$$

maximizes pointwise the growth rate as in (4.1) of Problem 4.6, Section 4: it is the growth-optimal portfolio for this model. Its excess growth rate is computed as

$$\gamma_{\hat{\pi}}^*(t) = \frac{\lambda(n - \lambda)}{2n^2} \sum_{i=1}^n \frac{1}{\mu_i(t)} - \frac{\lambda - 1}{2} (n - \lambda - 1).$$

Note that $\hat{\pi}(\cdot)$ is long in the equally weighted portfolio $\varphi(\cdot)$ of Example 12.2, and short in the market portfolio $\mu(\cdot)$. Using the structure of these two simple portfolios, it is

relatively straightforward to compute the performance of $\hat{\pi}(\cdot)$ relative to the market, namely,

$$\log\left(\frac{V^{\hat{\pi}}(T)}{V^{\mu}(T)}\right) = \frac{\lambda}{n} \log\left(\frac{\mu_1(T) \dots \mu_n(T)}{\mu_1(0) \dots \mu_n(0)}\right) + \int_0^T \left(\gamma_{\hat{\pi}}^*(t) + (\lambda - 1) \gamma_{\mu}^*(t)\right) dt.$$

Recalling the coherence of this model, the asymptotic property (12.12), and the computation $\gamma_{\mu}^*(t) = (n - 1)/2$, we deduce

$$\begin{aligned} \mathcal{L}^{\hat{\pi}, \mu} &:= \lim_{T \rightarrow \infty} \frac{1}{T} \log\left(\frac{V^{\hat{\pi}}(T)}{V^{\mu}(T)}\right) = \frac{\lambda(n - \lambda)}{n} \cdot \frac{\gamma}{\alpha} + \frac{\lambda(\lambda - 1)}{2} \\ &= \frac{n^2}{8}(1 + \alpha) \left[1 + \alpha + \frac{1}{2n} + (1 - \alpha) \left(1 + \frac{n - 1}{\alpha n}\right)\right]. \end{aligned} \quad (12.20)$$

A comparison with (12.18) shows that *shorting* the market portfolio as in (12.19) improves the performance of equal weighting by an entire order of magnitude in terms of market size n . The quantity of (12.20) is smaller than that of (12.17), as of course it should be, but has the same order of magnitude in terms of market size.

REMARK 12.5. Open Question: For the entropy-weighted portfolio $\pi_i^c(\cdot)$ of (11.10), compute in the context of the volatility-stabilized model the expression

$$\mathcal{L}^{\pi^c, \mu} := \liminf_{T \rightarrow \infty} \frac{1}{T} \log\left(\frac{V^{\pi^c}(T)}{V^{\mu}(T)}\right) = \liminf_{T \rightarrow \infty} \frac{\gamma^*}{T} \int_0^T \frac{dt}{c + \mathbf{H}(\mu(t))}$$

of (6.2), using (11.10) and (12.3). But note already from these expressions that

$$\mathcal{L}^{\pi^c, \mu} \geq \frac{n - 1}{2(c + \log n)} > 0 \quad \text{a.s.,}$$

suggesting again a significant outperformance of the market over long time horizons. Do the indicated limits exist, as one would expect?

REMARK 12.6. Open Questions: For fixed $t \in (0, \infty)$, determine the distributions of $\mu_i(t)$, $i = 1, \dots, n$ and of the largest $\mu_{(1)}(t) := \max_{1 \leq i \leq n} \mu_i(t)$ and smallest $\mu_{(n)}(t) := \min_{1 \leq i \leq n} \mu_i(t)$ market weights.

What can be said about the behavior of the averages $\frac{1}{T} \int_0^T \mu_{(k)}(t) dt$, particularly for the largest ($k = 1$) and the smallest ($k = n$) stocks?

13. Rank-based models

Size is one of the most important descriptive characteristics of financial assets. One can understand a lot about equity markets by observing, and trying to make sense of, the continual ebb and flow of small-, medium-, and large-capitalization stocks in their

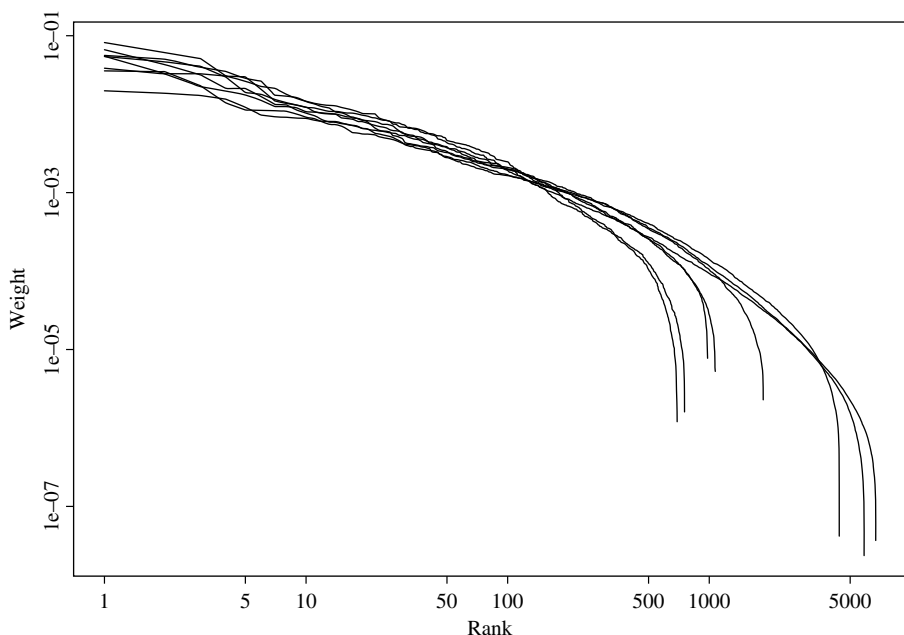


FIG. 13.1 Capital distribution curves: 1929–1999. The later the period, the longer the curve.

midst. A particularly convenient way to study this feature is by looking at the evolution of the capital distribution curve $\log k \mapsto \log \mu_{(k)}(t)$; that is, the logarithms of the market weights arranged in descending order versus the logarithms of their respective ranks (see also (13.14) below for a steady-state counterpart of this quantity). As shown in Fig. 13.1 of FERNHOLZ [2002], reproduced here as Fig. 13.1, this log-log plot has exhibited remarkable stability over the decades of the last century.

It is of considerable importance, then, to have available models that describe this flow of capital and exhibit stability properties for capital distribution that are in at least broad agreement with these observations.

The simplest model of this type assigns growth rates and volatilities to the various stocks, not according to their names (the indices i) but according to their ranks within the market's capitalization. More precisely, let us pick real numbers $\gamma, g_1, \dots, g_n$ and $\sigma_1 > 0, \dots, \sigma_n > 0$, satisfying conditions that will be specified in a moment, and prescribe growth rates $\gamma_i(\cdot)$ and volatilities $\sigma_{iv}(\cdot)$ as

$$\gamma_i(t) = \gamma + \sum_{k=1}^n g_k 1_{\{X_i(t)=X_{p_t(k)}(t)\}} \quad \sigma_{iv}(t) = \delta_{iv} \cdot \sum_{k=1}^n \sigma_k 1_{\{X_i(t)=X_{p_t(k)}(t)\}} \quad (13.1)$$

for $1 \leq i, v \leq n$ with $d = n$. We are using here the random permutation notation of (11.18), and we shall denote again by $\mathfrak{X}(\cdot) = (X_1(\cdot), \dots, X_n(\cdot))'$ the vector of stock capitalizations.

It is intuitively clear that if such a model is to have some stability properties, it has to assign considerably higher growth rates to the smallest stocks than to the biggest ones. It turns out that the right conditions for stability are

$$g_1 < 0, \quad g_1 + g_2 < 0, \dots, \quad g_1 + \dots + g_{n-1} < 0, \quad \text{and} \quad g_1 + \dots + g_n = 0. \quad (13.2)$$

These conditions are satisfied in the simplest model of this type, the *Atlas model* that assigns

$$\gamma = g > 0, \quad g_k = -g \quad \text{for } k = 1, \dots, n-1, \quad \text{and} \quad g_n = (n-1)g, \quad (13.3)$$

thus $\gamma_i(t) = ng \, 1_{\{X_i(t)=X_{p_t(n)}(t)\}}$ in (13.1): zero growth rate goes to all the stocks but the smallest, which then becomes responsible for supporting the entire growth of the market.

In addition to the drift condition (13.2), we shall impose a condition on the variances of the model

$$\sum_{k=1}^n \sigma_k^2 > 2 \cdot \max_{1 \leq k \leq n} \sigma_k^2, \quad 0 \leq \sigma_2^2 - \sigma_1^2 \leq \sigma_3^2 - \sigma_2^2 \leq \dots \leq \sigma_n^2 - \sigma_{n-1}^2.$$

Making these specifications amounts to postulating that the log capitalizations $Y_i(\cdot) := \log X_i(\cdot)$ $i = 1, \dots, n$ satisfy the system of stochastic differential equations

$$dY_i(t) = \left(\gamma + \sum_{k=1}^n g_k 1_{\mathcal{Q}_i^{(k)}}(\mathfrak{Y}(t)) \right) dt + \sum_{k=1}^n \sigma_k 1_{\mathcal{Q}_i^{(k)}}(\mathfrak{Y}(t)) dW_i(t), \quad (13.4)$$

with $Y_i(0) = y_i = \log x_i$. Here, $\{\mathcal{Q}_i^{(k)}\}_{1 \leq i, k \leq n}$ is a collection of polyhedral domains in $\mathbb{R}^n$, with the properties

$\{\mathcal{Q}_i^{(k)}\}_{1 \leq i \leq n}$ is a partition of $\mathbb{R}^n$, for each fixed k ,

$\{\mathcal{Q}_i^{(k)}\}_{1 \leq k \leq n}$ is a partition of $\mathbb{R}^n$, for each fixed i ,

and the interpretation

$\mathfrak{Y} = (Y_1, \dots, Y_n) \in \mathcal{Q}_i^{(k)}$ means that Y_i is ranked k th among $Y_1, \dots, Y_n$.

As long as the vector of log-capitalizations $\mathfrak{Y}(\cdot) = (Y_1(\cdot), \dots, Y_n(\cdot))'$ is in the polyhedron $\mathcal{Q}_i^{(k)}$, the Eq. (13.3) posits that the coördinate process $Y_i(\cdot)$ evolves like a Brownian motion with drift $\gamma + g_k$ and variance σ_k^2 . (Ties are resolved by resorting to the lowest index i ; for instance, $\mathcal{Q}_i^{(1)}$, $1 \leq i \leq n$ corresponds to the partition $\mathcal{Q}_i$ of $(0, \infty)^n$ of Section 9, right below (9.3), and so on.)

The theory of BASS and PARDOUX [1987] guarantees that this system has a weak solution, which is unique in distribution; once this solution has been constructed, we obtain stock capitalizations as $X_i(\cdot) = e^{Y_i(\cdot)}$ that satisfy (1.4) with the specifications of (13.1).

REMARK 13.1. **Research Problem:** There is a natural generalization of (13.4) to

$$dY_i(t) = \left(\gamma_i + \sum_{k=1}^n g_k 1_{Q_i^{(k)}}(\mathfrak{Y}(t)) \right) dt + \sum_{k=1}^n \sigma_k 1_{Q_i^{(k)}}(\mathfrak{Y}(t)) dW_i(t) + \rho_i dB_i(t), \quad (13.5)$$

where $(B_1(\cdot), \dots, B_n(\cdot))$ is a Brownian motion independent of $(W_1(\cdot), \dots, W_n(\cdot))$, and the γ_i and ρ_i are constants. In this case, it can be shown that the system is stable if and only if, besides (13.2), we have $\gamma_1 + \dots + \gamma_n = 0$ and

$$\sum_{k=1}^{\ell} (g_k + \gamma_{\pi(k)}) < 0, \quad \ell = 1, \dots, n-1,$$

for any permutation π of $\{1, 2, \dots, n\}$. The model (13.5) is known as the *hybrid model*, since the growth rates and variances depend on both rank and name, i.e., index. These models provide a simplification of the general market model of (1.1), but nevertheless one that may be both tractable enough and ample enough to allow meaningful insight into the behavior of real equity markets. Be that as it may, at this writing, there remain many open research questions regarding these hybrid models.

An immediate observation from (13.3) is that the sum $Y(\cdot) := \sum_{i=1}^n Y_i(\cdot)$ of log-capitalizations satisfies

$$Y(t) = y + n\gamma t + \sum_{k=1}^n \sigma_k B_k(t), \quad 0 \leq t < \infty$$

with $y := \sum_{i=1}^n y_i$ and $B_k(\cdot) := \sum_{i=1}^n \int_0^\cdot 1_{Q_i^{(k)}}(\mathfrak{Y}(s)) dW_i(s)$, $k = 1, \dots, n$ independent scalar Brownian motions. Thus, the strong law of large numbers implies

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^n Y_i(T) = n\gamma, \quad \text{a.s.}$$

Then it takes a considerable amount of work (see appendix in BANNER, FERNHOLZ and KARATZAS [2005]), in order to strengthen this result to

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log X_i(T) = \lim_{T \rightarrow \infty} \frac{Y_i(T)}{T} = \gamma \quad \text{a.s., for every } i = 1, \dots, n; \quad (13.6)$$

to wit, all the stocks have the same asymptotic growth-rate γ in this model. Using (13.6), it can be shown that the model specified by (1.5), (13.1) is coherent in the sense of Remark 2.1.

REMARK 13.2. **Taking Turns in the Various Ranks.** From (13.4), (13.6), and the strong law of large numbers for Brownian motion, we deduce that the quantity

$\sum_{k=1}^n g_k \left(\frac{1}{T} \int_0^T 1_{\mathcal{Q}_i^{(k)}}(\mathfrak{Y}(t)) dt \right)$ converges a.s. to zero, as $T \rightarrow \infty$. For the Atlas model in (13.3), this expression becomes $g \left(\frac{n}{T} \int_0^T 1_{\mathcal{Q}_i^{(n)}}(\mathfrak{Y}(t)) dt - 1 \right)$, and we obtain

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T 1_{\mathcal{Q}_i^{(n)}}(\mathfrak{Y}(t)) dt = \frac{1}{n} \quad \text{a.s., for every } i = 1, \dots, n.$$

Namely, each stock spends roughly $(1/n)$ th of the time, acting as “Atlas.”

Again with considerable work, this is strengthened by BANNER, FERNHOLZ and KARATZAS [2005] to the statement

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T 1_{\mathcal{Q}_i^{(k)}}(\mathfrak{Y}(t)) dt = \frac{1}{n}, \quad \text{a.s., for every } 1 \leq i, k \leq n, \quad (13.7)$$

valid not just for the Atlas model, but under the more general conditions of (13.2). Thanks to the symmetry inherent in this model, each stock spends roughly $(1/n)$ th of the time in any given rank (see proposition 2.3 by BANNER, FERNHOLZ and KARATZAS [2005]).

13.1. Ranked capitalization processes

For many purposes in the study of these models, it makes sense to look at the ranked log-capitalization processes

$$Z_k(t) := \sum_{i=1}^n Y_i(t) \cdot 1_{\mathcal{Q}_i^{(k)}}(\mathfrak{Y}(t)), \quad 0 \leq t < \infty \quad (13.8)$$

for $1 \leq k \leq n$. From these, we get the ranked capitalizations via $X_{(k)}(t) = e^{Z_k(t)}$, with notation similar to (1.18). Using an extended Tanaka-type formula, as we did in (11.19), it can be seen that the processes of (13.8) satisfy

$$Z_k(t) = Z_k(0) + (g_k + \gamma)t + \sigma_k B_k(t) + \frac{1}{2} \left(\mathfrak{L}^{k,k+1}(t) - \mathfrak{L}^{k-1,k}(t) \right), \quad 0 \leq t < \infty \quad (13.9)$$

in that notation. Here, as in subsection 11.2, the continuous and increasing process $\mathfrak{L}^{k,k+1}(\cdot) := \Lambda_{\Xi_k}(\cdot)$ is the semimartingale local time at the origin of the continuous, nonnegative process

$$\Xi_k(\cdot) = Z_k(\cdot) - Z_{k+1}(\cdot) = \log \left(\mu_{(k)}(\cdot) / \mu_{(k+1)}(\cdot) \right)$$

of (11.20) for $k = 1, \dots, n-1$; we make again the convention $\mathfrak{L}^{0,1}(\cdot) \equiv \mathfrak{L}^{n,n+1}(\cdot) \equiv 0$.

These local times play a big rôle in the analysis of this model. The quantity $\mathfrak{L}^{k,k+1}(T)$ represents again the cumulative amount of change between ranks k and $k+1$ that occurs over the time interval $[0, T]$. Of course, in a model such as the one studied here, the intensity of changes for the smaller stocks should be higher than for the larger stocks.

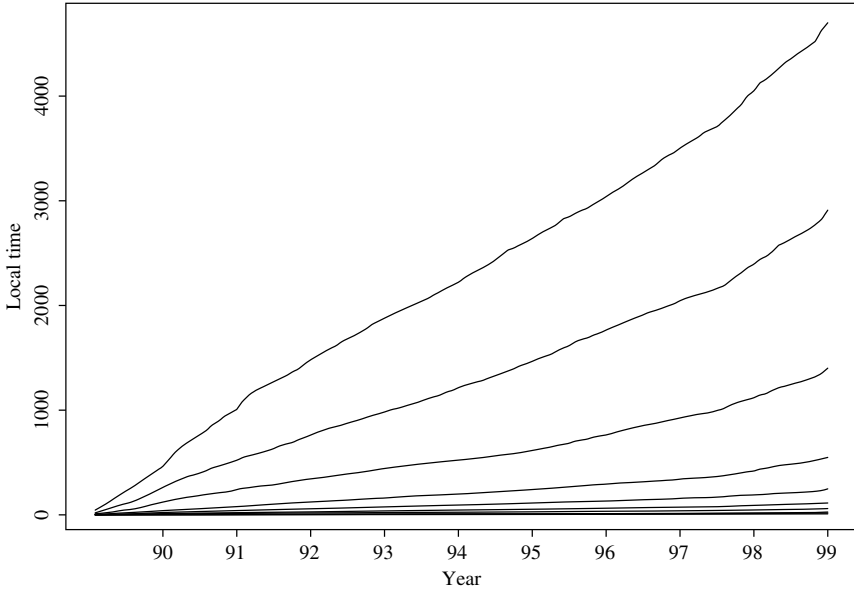


FIG. 13.2 $\mathfrak{L}^{k,k+1}(\cdot)$, $k = 10, 20, 40, \dots, 5120$.

This is borne out by experiment: as we saw in Remark 11.8, it turns out, somewhat surprisingly, that these local times can be estimated based only on observations of relative market weights and of the performance of simple portfolios over $[0, T]$, and that they exhibit a remarkably linear increase, with positive rates that grow with k , as we see in Fig. 13.2, reproduced from FERNHOLZ [2002], Fig. 13.2.

The analysis of the present model agrees with these observations: it follows from (13.6) and the dynamics of (13.9) that, for $k = 1, \dots, n - 1$, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathfrak{L}^{k,k+1}(T) = \lambda_{k,k+1} := -2(g_1 + \dots + g_k) > 0, \quad \text{a.s.} \quad (13.10)$$

Our stability condition guarantees that these partial sums are positive – as indeed the limits on the right-hand side of (13.10) ought to be; in typical examples, such as the Atlas model of (13.3) where $\lambda_{k,k+1} = kg$, they do increase with k , as suggested by Fig. 13.2.

13.2. Some asymptotics

A slightly more careful analysis of these local times reveals that the nonnegative semimartingale $\Xi_k(\cdot)$ of (11.20) can be cast in the form of a Skorohod problem

$$\Xi_k(t) = \Xi_k(0) + \Theta_k(t) + \Lambda_{\Xi_k}(t), \quad 0 \leq t < \infty,$$

as the reflection, at the origin, of the semimartingale

$$\Theta_k(t) = (g_k - g_{k+1})t - \frac{1}{2} \left(\mathfrak{L}^{k-1,k}(t) + \mathfrak{L}^{k+1,k+2}(t) \right) + \mathbf{s}_k \tilde{W}^{(k)}(t),$$

where $\mathbf{s}_k := (\sigma_k^2 + \sigma_{k+1}^2)^{1/2}$ and $\tilde{W}^{(k)}(\cdot) := (\sigma_k B_k(\cdot) - \sigma_{k+1} B_{k+1}(\cdot))/\mathbf{s}_k$ is standard Brownian motion.

As a result of these observations and of (13.10), we conclude that the process $\Xi_k(\cdot)$ behaves asymptotically like Brownian motion with drift

$$g_k - g_{k+1} - \frac{1}{2}(\lambda_{k-1,k} + \lambda_{k,k+1}) = -\lambda_{k,k+1} < 0,$$

variance $\mathbf{s}_k^2$, and reflection at the origin. Consequently,

$$\lim_{t \rightarrow \infty} \log \left(\frac{\mu_{(k)}(t)}{\mu_{(k+1)}(t)} \right) = \lim_{t \rightarrow \infty} \Xi_k(t) = \xi_k, \quad \text{in distribution} \quad (13.11)$$

where for each $k = 1, \dots, n-1$, the random variable ξ_k has an exponential distribution

$$\begin{aligned} \mathbb{P}(\xi_k > x) &= e^{-r_k x}, \quad x \geq 0 \quad \text{with parameter} \quad r_k := \frac{2\lambda_{k,k+1}}{\mathbf{s}_k^2} \\ &= -\frac{4(g_1 + \dots + g_k)}{\sigma_k^2 + \sigma_{k+1}^2} > 0. \end{aligned} \quad (13.12)$$

As ICHIBA [2006] observes, the theory of HARRISON and WILLIAMS [1987a,b] implies that the random variables $\xi_1, \dots, \xi_n$ are independent when the variances are of the form $\sigma_k^2 = \sigma^2 + ks^2$ for some real numbers $\sigma^2 > 0$ and $s^2 \geq 0$, that is, are either constant or grow linearly with rank.

13.3. The steady-state capital distribution curve

We also have from (13.11), the strong law of large numbers

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(\Xi_k(t)) dt = \mathbb{E}(g(\xi_k)), \quad \text{a.s.}$$

for every rank k and every measurable function $g: [0, \infty) \rightarrow \mathbb{R}$ with $\int_0^\infty |g(x)|e^{-r_k x} dx < \infty$ (see KHAS'MINSKII [1960]). In particular,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \log \left(\frac{\mu_{(k)}(t)}{\mu_{(k+1)}(t)} \right) dt = \mathbb{E}(\xi_k) = \frac{1}{r_k} = \frac{\mathbf{s}_k^2}{2\lambda_{k,k+1}}, \quad \text{a.s.} \quad (13.13)$$

This observation provides a tool for studying the steady-state capital distribution curve

$$\log k \longmapsto \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \log \mu_{(k)}(t) dt =: \mathbf{m}(k), \quad k = 1, \dots, n-1 \quad (13.14)$$

alluded to at the beginning of this section (more on the existence of this limit in the next subsection). To estimate the slope $\mathbf{q}(k)$ of this curve at the point $\log k$, we use (13.13), and the estimate $\log(k+1) - \log k \approx 1/k$, to obtain in the notation of (13.12):

$$\mathbf{q}(k) \approx \frac{\mathbf{m}(k) - \mathbf{m}(k+1)}{\log k - \log(k+1)} = -\frac{k}{r_k} = \frac{k(\sigma_k^2 + \sigma_{k+1}^2)}{4(g_1 + \dots + g_k)} < 0. \quad (13.15)$$

Consider now an Atlas model as in (13.3). With equal variances $\sigma_k^2 = \sigma^2 > 0$, this slope is the constant $q(k) \approx -\sigma^2/2g$ and the steady-state capital distribution curve can be approximated by a straight Pareto line.

On the other hand, with variances of the form $\sigma_k^2 = \sigma^2 + ks^2$ for some $s^2 > 0$, growing linearly with rank, we get for large k the approximate slope

$$q(k) \approx -\frac{1}{2g}(\sigma^2 + ks^2), \quad k = 1, \dots, n-1.$$

Such linear growth is suggested in fig. 5.5 in FERNHOLZ [2002], which is reproduced here as Fig. 13.3. This would imply a decreasing and concave steady-state capital distribution curve, whose (negative) slope becomes more and more pronounced in magnitude with increasing rank, much in accord with the features of Fig. 13.1.

We see, in other words, that even such a simplistic model as that of (1.5) and (13.1), which has features such as (13.6) and (13.7) that are not particularly realistic, is able to capture asymptotic stability properties observed in real markets, such as those exhibited in Figs. 4–6. It is possible to modify the model of the present section in ways that remove the ‘simplistic’ features (13.6) and (13.7), but retain the good asymptotic properties already mentioned. One is, thus, led to the “hybrid” models of Remark 13.1 that prescribe growth rates and covariances based on both name (the index i) and rank; as already mentioned, such models are the subject of very active current research.

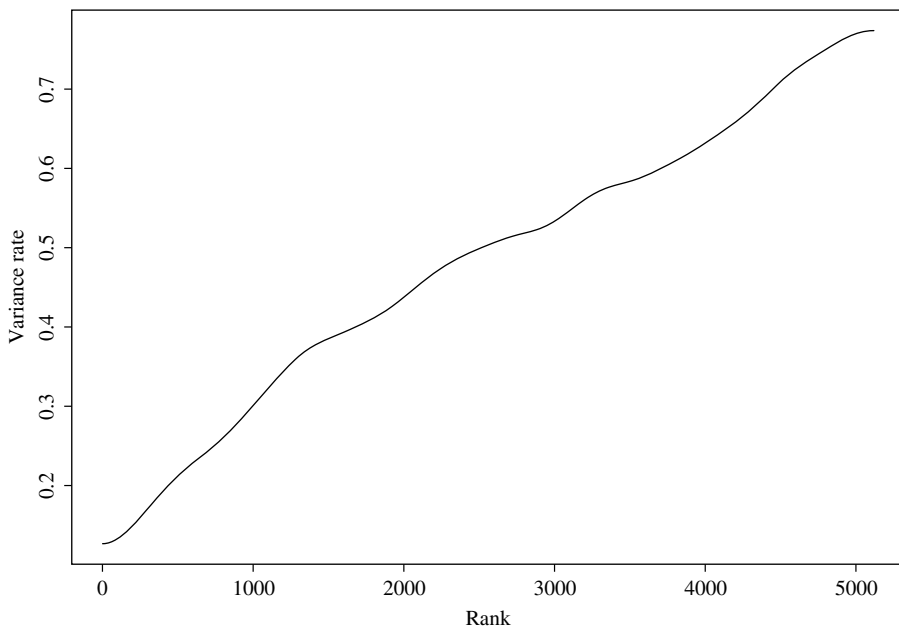


FIG. 13.3 Smoothed annualized values of $\hat{s}_k^2$ for $k = 1, \dots, 5119$. Calculated from 1990–1999 data.

REMARK 13.3. Estimation of Parameters in this Model. Let us remark that (13.10) provides a method for obtaining estimates $\hat{\lambda}_{k,k+1}$ of the parameters $\lambda_{k,k+1}$ from the observable random variables $\mathfrak{L}^{k,k+1}(T)$ that measure cumulative change between ranks k and $k+1$ (recall Remark 11.8 once again). Then, estimates of the parameters g_k follow, as $\hat{g}_k = (\hat{\lambda}_{k-1,k} - \hat{\lambda}_{k,k+1})/2$, and the parameters $s_k^2 = \sigma_k^2 + \sigma_{k+1}^2$ can be estimated from (13.13) and from the increments of the observable capital distribution curve of (13.14), namely $\hat{s}_k^2 = 2\hat{\lambda}_{k,k+1}(\mathfrak{m}(k) - \mathfrak{m}(k+1))$. For the decade 1990–1999, these estimates are presented in Fig. 13.3.

Finally, we make the following selections for estimating the variances

$$\hat{\sigma}_k^2 = \frac{1}{4} (\hat{s}_{k-1}^2 + \hat{s}_k^2), \quad k = 2, \dots, n-1, \quad \text{and} \quad \hat{\sigma}_1^2 = \frac{1}{2} \hat{s}_1^2, \quad \hat{\sigma}_n^2 = \frac{1}{2} \hat{s}_{n-1}^2.$$

13.4. Stability of the capital distribution

Let us now go back to (13.11); it can be seen that this leads to the convergence of the ranked market weights

$$\lim_{t \rightarrow \infty} (\mu_{(1)}(t), \dots, \mu_{(n)}(t)) = (M_1, \dots, M_n), \quad \text{in distribution} \quad (13.16)$$

to the random variables

$$M_n := (1 + e^{\xi_{n-1}} + \dots + e^{\xi_1 + \dots + \xi_{n-1}})^{-1} \quad \text{and} \quad M_k := M_n e^{\xi_k + \dots + \xi_{n-1}} \quad (13.17)$$

for $k = 1, \dots, n-1$. These are the long-term (steady-state) relative weights of the various stocks in the market, ranked from largest, M_1 , to smallest, M_n . Again, we have from (13.16) the strong law of large numbers

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\mu_{(1)}(t), \dots, \mu_{(n)}(t)) dt = \mathbb{E}(f(M_1, \dots, M_n)), \quad \text{a.s.} \quad (13.18)$$

for every bounded and measurable $f : \Delta_+^n \rightarrow \mathbb{R}$. Note that (13.13) is a special case of this result, and that the function $\mathfrak{m}(\cdot)$ of (13.14) takes the form

$$\mathfrak{m}(k) = \mathbb{E}(\log(M_k)) = \sum_{\ell=k}^{n-1} \frac{1}{r_\ell} - \mathbb{E}(\log(1 + e^{\xi_{n-1}} + \dots + e^{\xi_1 + \dots + \xi_{n-1}})). \quad (13.19)$$

This is the good news; the bad news is that we do not know, in general, the joint distribution of the exponential random variables $\xi_1, \dots, \xi_{n-1}$ in (13.11), so we cannot find that of $M_1, \dots, M_n$ either. In particular, we cannot pin down the steady-state capital distribution function of (13.19), though we do know precisely its increments $\mathfrak{m}(k+1) - \mathfrak{m}(k) = -(1/r_k)$ and thus are able to estimate the slope of the steady-state capital distribution curve, as indeed we did in (13.15). In BANNER, FERNHOLZ and KARATZAS [2005] a simple, certainty-equivalent approximation of the steady-state

ranked market weights of (13.17) is carried out and is used to study in detail the behavior of simple portfolios in such a model.

REMARK 13.4. Open Question: What can be said about the joint distribution of the long-term (steady-state) relative market weights of (13.17)? Can it be characterized, computed, or approximated in a good way? What can be said about the fluctuations of the random variables $\log(M_k)$ with respect to their means $m(k)$ in (13.19)?

For answers to some of these questions for equal variances and large numbers of assets (in the limit as $n \rightarrow \infty$), see the important recent work of PAL and PITMAN [2007] and CHATTERJEE and PAL [2007].

REMARK 13.5. Research Question and Conjecture: Study the steady-state capital distribution curve of the volatility-stabilized model in (12.1). With $\alpha > 0$, check the validity of the following conjecture: the slope

$$q(k) \approx \frac{m(k) - m(k+1)}{\log k - \log(k+1)}$$

of the capital distribution $m(\cdot)$ at $\log k$ should be given as

$$q(k) \approx -4\gamma k \eta_k, \quad \eta_k := \mathbb{E} \left(\frac{\log Q_{(k)} - \log Q_{(k+1)}}{Q_{(1)} + \cdots + Q_{(n)}} \right),$$

where $Q_{(1)} \geq \cdots \geq Q_{(n)}$ are the order statistics of a random sample from the chi-square distribution with $\kappa = 2(1 + \alpha)$ degrees of freedom.

If this conjecture is correct, does $k\eta_k$ increase with k ?

14. Some concluding remarks

We have surveyed a framework, called *Stochastic Portfolio Theory*, for studying the behavior of portfolio rules and for modeling and analyzing equity market structure. We have also exhibited simple conditions, such as “diversity” and “availability of intrinsic volatility,” which can lead to arbitrages relative to the market.

These conditions are descriptive in nature and can be tested from the predictable characteristics of the model posited for the market. In contrast, familiar assumptions, such as the existence of an EMM, are normative in nature; they cannot be decided on the basis of predictable characteristics in the model. In this vein, the Example 4.7, pp. 469–470 of KARATZAS and KARDARAS [2007] is quite instructive.

The existence of such relative arbitrage is not the end of the world. Under reasonably general conditions, one can still work with appropriate “deflators” for the purposes of hedging contingent claims and of portfolio optimization, as we have tried to illustrate in Section 10.

Considerable computational tractability is lost, as the marvelous tool that is the EMM goes out the window. Nevertheless, big swaths of the field of mathematical finance

remain totally or mostly intact; completely new areas and issues, such as those of the “Abstract Markets” in Chapter IV of this survey, thrust themselves onto the scene.

Acknowledgments

We are indebted to Professor Alain Bensoussan for suggesting to us that we write this survey.

The survey is an expanded version of the Lukacs Lectures, given by one of us at Bowling Green University in May–June 2006. We are indebted to our hosts at Bowling Green, Ohio, for the invitation to deliver the lectures, for their hospitality, their interest, and their incisive comments during the lectures; these helped us sharpen our understanding and improved the exposition of the chapter.

We are also indebted to our seminar audiences at MIT, Boston, Texas-Austin, Yale, Carnegie-Mellon, Charles University in Prague; at the Columbia University Mathematical Finance Practitioners’ Seminar; at a Summer School on the island of Chios, organized by the University of the Aegean; at a Morgan-Stanley seminar; and at the Risk Magazine Conferences in July, October, and November 2006, as well as in June 2007 and July 2008, for their comments and suggestions.

Many thanks are due to Constantinos Kardaras for going over an early version of the manuscript and offering many valuable suggestions; to Adrian Banner for his comments on a later version; and to Mihai Sîrbu for helping us simplify and sharpen some of our results and for catching several typos in the near-final version of the chapter.

References

- BANNER, A., FERNHOLZ, D. (2007). Short-term arbitrage in volatility-stabilized markets. *Ann. Financ.* to appear.
- BANNER, A., FERNHOLZ, R., KARATZAS, I. (2005). On Atlas models of equity markets. *Ann. Appl. Probab.* **15**, 2296–2330.
- BANNER, A., GHOMRASNI, R. (2008). Local times of ranked continuous semimartingales. *Stoch. Proc. Appl.* **118**, 1244–1253.
- BASS, R., PARDOUX, E. (1987). Uniqueness of diffusions with piecewise constant coefficients. *Probab. Theory. Rel. Fields* **76**, 557–572.
- BASS, R., PERKINS, E. (2002). Degenerate stochastic differential equations with Hölder-continuous coefficients and super-Markov chains. *Trans. Am. Math. Soc.* **355**, 373–405.
- CHATTERJEE, S., PAL, S. (2007). A phase-transition behavior for Brownian motions interacting through their ranks, Preprint.
- COVER, T. (1991). Universal portfolios. *Math. Financ.* **1**, 1–29.
- DUFFIE, D. (1992). *Dynamic Asset Pricing Theory* (Princeton University Press, Princeton, NJ).
- FERNHOLZ, E.R. (1999). On the diversity of equity markets. *J. Math. Econ.* **31**, 393–417.
- FERNHOLZ, E.R. (1999a). Portfolio generating functions. In: Avellaneda, M. (ed.), *Quantitative Analysis in Financial Markets* (World Scientific, River Edge, NJ).
- FERNHOLZ, E.R. (2001). Equity portfolios generated by functions of ranked market weights. *Financ. Stoch.* **5**, 469–486.
- FERNHOLZ, E.R. (2002). *Stochastic Portfolio Theory* (Springer-Verlag, New York, NY).
- FERNHOLZ, E.R., KARATZAS, I. (2005). Relative arbitrage in volatility-stabilized markets. *Ann. Financ.* **1**, 149–177.
- FERNHOLZ, E.R., KARATZAS, I. (2006). The implied liquidity premium for equities. *Ann. Financ.* **2**, 87–99.
- FERNHOLZ, E.R., KARATZAS, I., KARDARAS, C. (2005). Diversity and arbitrage in equity markets. *Financ. Stoch.* **9**, 1–27.
- FERNHOLZ, E.R., MAGUIRE, C. (2007). The statistics of ‘statistical arbitrage’. *Financial Analysts J.* **63**, 46–52.
- FERNHOLZ, E.R., SHAY, B. (1982). Stochastic portfolio theory and stock market equilibrium. *J. Financ.* **37**, 615–624.
- HARRISON, M., WILLIAMS, R. (1987a). Multi-dimensional reflected Brownian motions having exponential stationary distributions. *Ann. Probab.* **15**, 115–137.
- HARRISON, M., WILLIAMS, R. (1987b). Brownian models of open queuing networks with homogeneous customer populations. *Stochastics* **22**, 77–115.
- HEATH, D., OREY, S., PESTIEN, V., SUDDERTH, W.D. (1987). Maximizing or minimizing the expected time to reach zero. *SIAM J. Control. Optim.* **25**, 195–205.
- ICHIBA, T. (2006). Personal communication.
- JAMSHIDIAN, F. (1992). Asymptotically optimal portfolios. *Math. Financ.* **3**, 131–150.
- KARATZAS, I., KARDARAS, C. (2007). The numéraire portfolio and arbitrage in semimartingale markets. *Financ. Stoch.* **11**, 447–493.
- KARATZAS, I., LEHOCZKY, J.P., SHREVE, S.E., XU, G.L. (1991). Martingale and duality methods for utility maximization in an incomplete market. *SIAM J. Control. Optim.* **29**, 702–730.
- KARATZAS, I., SHREVE, S.E. (1991). *Brownian Motion and Stochastic Calculus*, Second ed. (Springer-Verlag, New York, NY).
- KARATZAS, I., SHREVE, S.E. (1998). *Methods of Mathematical Finance* (Springer-Verlag, New York, NY).

- KARDARAS, C. (2003). *Stochastic Portfolio Theory in Semimartingale Markets* (Unpublished Manuscript, Columbia University).
- KARDARAS, C. (2006). Personal communication.
- KHAS'MINSKII, R.Z. (1960). Ergodic properties of recurrent diffusion processes, and stabilization of the solution to the Cauchy problem for parabolic equations. *Theor. Probab. Appl.* **5**, 179–196.
- LOWENSTEIN, M., WILLARD, G.A. (2000a). Local martingales, arbitrage and viability. *Econ. Theor.* **16**, 135–161.
- LOWENSTEIN, M., WILLARD, G.A. (2000b). Rational equilibrium asset-pricing bubbles in continuous trading models. *J. Econ. Theor.* **91**, 17–58.
- MARKOWITZ, H. (1952). Portfolio selection. *J. Financ.* **7**, 77–91.
- OSTERIEDER, J., RHEINLÄNDER, Th. (2006). A note on arbitrage in diverse markets. *Ann. Financ.* **2**, 287–301.
- PAL, S., PITMAN, J. (2007). One-dimensional Brownian particle systems with rank-dependent drifts, Preprint.
- PESTIEN, V., SUDDERTH, W.D. (1985). Continuous-time red-and-black: how to control a diffusion to a goal. *Math. Oper. Res.* **10**, 599–611.
- PLATEN, E. (2002). Arbitrage in continuous complete markets. *Adv. Appl. Probab.* **34**, 540–558.
- PLATEN, E. (2006). A benchmark approach to finance. *Math. Financ.* **16**, 131–151.
- SPITZER, F. (1958). Some theorems concerning two-dimensional Brownian motion. *Trans. Am. Math. Soc.* **87**, 187–197.
- SUDDERTH, W.D., WEERASINGHE, A. (1989). Controlling a process to a goal in finite time. *Math. Oper. Res.* **14**, 400–409.

Asymmetric Variance Reduction for Pricing American Options

Chuan-Hsiang Han²

*Department of Quantitative Finance, National Tsing-Hua University, Hsinchu,
Taiwan 30013, ROC
E-mail address: chhan@mx.nthu.edu.tw*

Jean-Pierre Fouque¹

*Department of Statistics and Applied Probability, University of California,
Santa Barbara, CA 93106-3110, USA
E-mail address: fouque@pstat.ucsb.edu*

Abstract

Based on the dual formulation by ROGERS [2002], Monte Carlo algorithms to estimate the high-biased and low-biased estimates for American option prices are proposed. Bounds for pricing errors and the variance of biased estimators are shown to be dependent on hedging martingales. These martingales are applied to (1) simultaneously reduce the error bound and the variance of the high-biased estimator and (2) reduce the variance of the low-biased estimator while preserving its biased level. For a class of stochastic volatility models, projected hedging martingales are constructed based on an application of asymptotic expansion of option prices introduced in Fouque [3]. These martingales are easy to compute. Numerical results demonstrate the robustness and effectiveness of these projected hedging martingales.

1. Introduction

The right to early exercise a contingent claim is an important feature for derivative trading. An American option offers its holder, not the seller, the right but not the obligation

¹ Work supported by NSF grant DMS-0455982.

² This work is supported by NSC grant 95-2115-M-007-017-MY2, Taiwan, C.-H. Han is grateful for discussions with Professor Sheunn-Jhi Sheu at the Institute for Mathematics, Academia Sinica, Taiwan.

to exercise the contract at any time prior to maturity during its contract lifetime. Based on the no-arbitrage argument, the American option price at time 0, denoted by P_0 , with maturity $T < \infty$ is considered an optimal stopping problem ROGERS [2002]. That is, under the risk-neutral probability space $(\Omega, \mathcal{F}, \mathbb{P}^*, (\mathcal{F}_t)_{t \in [0, T]})$,

$$P_0 = \sup_{0 \leq \tau \leq T} \mathbb{E}^* \{Z_\tau | \mathcal{F}_0\}, \quad (1.1)$$

where the supremum is taken over all the stopping times τ bounded by T , the discounted payoff is denoted by $Z_t = D_t \tilde{Z}_t$ in which D_t is the discount factor and $\tilde{Z}_t$ is the payoff at time t , both $(\mathcal{F}_t)$ adapted. We assume that Z_t satisfies the uniformly integrable condition $\sup_{0 \leq t \leq T} |Z_t| \in L^p$ for some $p > 1$, and Z is right continuous.

During the last decade, methods of Monte Carlo simulations have made a great progress in solving the American option pricing problem. Among these, primal methods and dual methods provide upper solutions and lower solutions for the American option price, respectively. Primal methods such as in LONGSTAFF and SCHWARTZ [2001] and TSITSIKLIS and VAN ROY [2001] address the optimal stopping problem (1.1) by approximating the free boundary or the optimal stopping rule, while dual methods such as in HAUGH and KOGAN [2004] and ROGERS [2002] address a stochastic minimization problem by approximating the optimal supermartingale or martingale, respectively. As a result, the primal method induces a low-biased estimate and the dual method induces a high-biased estimate for the American option price. Results from these two methods are useful for practical trading activities: the option seller is typically interested in the high-biased estimate as the hedging strategy is related to (super)martingale, while the option holder is interested in low-biased estimate as the time to early exercise can be simulated.

Based on Rogers' dual formulation ROGERS [2002], this paper proposes and analyzes methods to compute high-biased estimates and low-biased estimates for the American option price. The low-biased estimator is naturally equipped with a variance reduction feature. Primal and dual representations for the American option price by martingales are characterized. The price gap and variance of biased estimators are sensitive to zero-centered martingales. Because the martingales are associated with hedging strategies, we refer them as hedging martingales. It remains a task to search for hedging martingales.

There exists an enormous literature on American option price approximations, which are typical in closed form or analytic form. Unfortunately, the process obtained from an approximate discounted price may not be a martingale or a supermartingale. For instance, in an example given in Lemma A.1 in the Appendix, we find that the discounted quadratic approximation in BARONE-ADESI and WHALEY [1987] does not possess the (super) martingale property. However, if the delta of an option price approximation is easy to compute, then the corresponding stochastic integral-type martingale becomes easy to construct. This can be considered as one advantage. Another advantage is that the integral-type martingale represents a continuous-time trading activity of dynamic hedge. Therefore, the variance of the biased-price estimator represents the quadratic measure of associated hedging errors. In this chapter, we consider hedging martingales being in stochastic integral type rather than a discounted approximate price as in ROGERS [2002].

For Monte Carlo simulations, variance reduction methods are important to improve the precision of estimates (see Glasserman [6] for a general background). In our

formulation, the low-biased estimator comprises the sample mean of a discounted pay-off less a hedging martingale at a stopping time. Hence, based on the optional sampling theorem, the hedging martingale is a natural candidate to play the role of a linear control to reduce the variance of low-biased estimators. For the high-biased estimator of the American option price, the hedging martingale can be understood as a nonlinear control. Theorem 2.2 in this chapter guarantees that a hedging martingale inducing a smaller high-biased estimate will induce a smaller variance, at least in a neighborhood of the optimal hedging martingale. In other words, for high-biased estimates, one can reduce the bias and the variance of the estimator at the same time. This effect contradicts typical variance reduction methods for unbiased estimators as used for low-biased estimates. That is, given a stopping rule, the hedging martingale will only affect the variance without changing the mean level. This asymmetric variance reduction effect produced by hedging martingales to estimate upper and lower solutions can be observed numerically in Section 3 under Black–Scholes model and in Section 4 under stochastic volatility models.

The organization of the paper is as follows. In Section 2, the high-biased and low-biased estimates for the American option price are proposed. We deduce representations for primal and dual formulations. The asymmetric behavior between the price bias and the variance is analyzed. In Section 3, the Black–Scholes model is considered. A characterization of the optimal stopping time is obtained. Two dual formulations, hedging martingale by ROGERS [2002] and hedging supermartingale by HAUGH and KOGAN [2004], are shown to be equivalent. Numerical results to estimate American option prices are demonstrated. In Section 4, multiscale stochastic volatility models are considered. Based on an asymptotic expansion for the option price as in FOUQUE, PAPANICOLAOU, SIRCAR and SOLNA [2003], the construction of a projected hedging martingale is proposed and some numerical examples are demonstrated. We conclude this paper in Section 5.

2. Primal and dual formulations of American option prices

ROGERS [2002] obtained a dual formulation for the American option problem (1.1) by solving an *inf-sup* problem over martingales:

$$P_0 = \inf_{M \in H_0^1} \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t) | \mathcal{F}_0 \right\}, \quad (2.1)$$

where the space of martingales is

$$H_0^1 = \left\{ (M_t)_{0 \leq t \leq T} : \text{martingales with } \sup_{0 \leq t \leq T} |M_t| \in L^1 \text{ and } M_0 = 0 \right\}.$$

The proof is based on the Doob–Meyer decomposition of a supermartingale process, and, in fact, the infimum in (2.1) is attained so that

$$P_0 = \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t^*) | \mathcal{F}_0 \right\}, \quad (2.2)$$

where the optimal martingale M^* is the unique martingale obtained from the Doob–Meyer decomposition. The following result shows that the American option price

is bounded above by a lookback-style option price based on the dual approach and bounded below by a barrier-style option price based on the primal approach.

2.1. High-biased and low-biased estimates

PROPOSITION 2.1. *Given an integrable martingale $M \in H_0^1$ and a stopping time $0 \leq \tau \leq T$, the high-biased estimate and the low-biased estimate of the American option price are obtained:*

$$\mathbb{E}^* \{Z_\tau - M_\tau | \mathcal{F}_0\} \leq P_0 \leq \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t) | \mathcal{F}_0 \right\}.$$

PROOF. From (2.1), it is easy to obtain an upper-bound solution or a high-biased estimate of the American option price

$$P_0 \leq \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t) | \mathcal{F}_0 \right\} \quad (2.3)$$

for any given integrable martingale $M \in H_0^1$. On the other hand, for any bounded stopping time $0 \leq \tau \leq T$, it is readily seen that

$$Z_\tau - M_\tau \leq \sup_{0 \leq t \leq T} (Z_t - M_t),$$

such that after taking an expectation, the left-hand side is equal to

$$\mathbb{E}^* \{Z_\tau - M_\tau | \mathcal{F}_0\} = \mathbb{E}^* \{Z_\tau | \mathcal{F}_0\} = \mathbb{E}^* \{Z_\tau - \tilde{M}_\tau | \mathcal{F}_0\}$$

for any other integrable martingale $\tilde{M} \in H_0^1$ due to the optional sampling theorem. Therefore, a lower bound solution or a low-biased estimate of the American option price is deduced

$$\mathbb{E}^* \{Z_\tau - M_\tau | \mathcal{F}_0\} \leq P_0. \quad (2.4)$$

Note that the hedging martingales used to compute the high-biased estimate and the low-biased estimate can be different. This proposition indicates that whenever one computes a high-biased estimate, it is possible that one can calculate a corresponding low-biased estimate as long as a stopping rule can be realized from, for instance, the least squares method LONGSTAFF and SCHWARTZ [2001]. Though the lower bound estimate is exactly equal to $\mathbb{E}^* \{Z_\tau\}$, we prefer to keep the stopped-martingale term M_τ in order to emphasize its hedging feature and its application to variance reduction. The next result provides two representations for the American option price. They lay the foundation to estimate the price gap and variance of biased estimators in Section 2.2. $\square$

THEOREM 2.1. *Let M^* denote the optimal martingale from the dual formulation (2.2) and τ^* denote an optimal stopping time from the primal formulation (1.1). Then*

- (i) $P_0 = \sup_{0 \leq t \leq T} (Z_t - M_t^*)$.
- (ii) $P_0 = Z_{\tau^*} - M_{\tau^*}^*$.

PROOF.

(i) We first introduce the Snell's envelop process

$$P_t = \text{ess sup}_{t \leq \tau \leq T} \mathbb{E}^* \{Z_\tau | \mathcal{F}_t\},$$

which is a supermartingale of class (D) KARATZAS and SHREVE [2000]. Based on the Doob–Meyer decomposition, for any time t , $0 \leq t \leq T$, we have

$$P_t = P_0 + M_t^* - A_t^*, \quad (2.5)$$

where $A_t^* \geq 0$ is a non-decreasing predictable process vanishing at time zero. Using $P_t \geq Z_t$ and the above decomposition, we have

$$\begin{aligned} Z_t - M_t^* &\leq P_t - M_t^* \\ &= P_0 - A_t^* \\ &\leq P_0 \end{aligned}$$

since $A_t^* \geq 0$. Taking the supremum over time $0 \leq t \leq T$, we see that $\sup_{0 \leq t \leq T} (Z_t - M_t^*) \leq P_0$. But from Proposition 2.1 by substituting M^* , we ensure that almost surely

$$P_0 = \sup_{0 \leq t \leq T} (Z_t - M_t^*). \quad (2.6)$$

(ii) On the other hand, the low-biased estimate becomes the American option price when an optimal stopping time τ^* is chosen such that

$$\mathbb{E}^* \{Z_{\tau^*} - M_{\tau^*}^* | \mathcal{F}_0\} = \mathbb{E}^* \{Z_{\tau^*} | \mathcal{F}_0\} = P_0. \quad (2.7)$$

From Eq. (2.2) and (2.7) and by the fact that $\sup_{0 \leq t \leq T} (Z_t - M_t^*) \geq Z_{\tau^*} - M_{\tau^*}^*$, these two random variables have to be equal to the price almost surely

$$Z_{\tau^*} - M_{\tau^*}^* = \sup_{0 \leq t \leq T} (Z_t - M_t^*) = P_0. \quad (2.8)$$

□

2.2. Price and variance errors in American option price estimation

Based on Proposition 2.1, for any given martingale $M \in H_0^1$, one can compute a high-biased and a low-biased estimate for the American option price. We show next that the price error between the high-biased estimate and the American option price, and the variance of high-biased estimator are both highly dependent on the choice of martingales.

THEOREM 2.2. Let M^* denote the optimal martingale as in Theorem 2.1, and $M \in H_0^1$ be any given martingale.

- (i) The price error between the high-biased estimate $\bar{P}_0 = \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t) \mid \mathcal{F}_0 \right\}$ and P_0 is bounded above, namely

$$\bar{P}_0 - P_0 \leq 2 \sqrt{\text{Var} (M_T^* - M_T)}.$$

- (ii) The variance of $\sup_{0 \leq t \leq T} (Z_t - M_t)$ vanishes if and only if the martingale M is optimal.

PROOF.

(i) By definition

$$\begin{aligned} \bar{P}_0 &= \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (Z_t - M_t^* + M_t^* - M_t) \mid \mathcal{F}_0 \right\} \\ &\leq P_0 + \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (M_t^* - M_t) \mid \mathcal{F}_0 \right\}. \end{aligned}$$

By $\sup_{0 \leq t \leq T} (M_t^* - M_t) \leq \sqrt{(\sup_{0 \leq t \leq T} |M_t^* - M_t|)^2}$ and Jensen's inequality, we deduce

$$\begin{aligned} \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (M_t^* - M_t) \mid \mathcal{F}_0 \right\} &\leq \sqrt{\mathbb{E}^* \left\{ \left(\sup_{0 \leq t \leq T} |M_t^* - M_t| \right)^2 \mid \mathcal{F}_0 \right\}} \\ &\leq 2 \sqrt{\mathbb{E}^* \left\{ (M_T^* - M_T)^2 \mid \mathcal{F}_0 \right\}}. \end{aligned}$$

The last inequality is obtained from Doob's maximal inequality KARATZAS and SHREVE [2000].

(ii) ($\Rightarrow$) Since the variance is zero, let $\sup_{0 \leq t \leq T} (Z_t - M_t) = C < +\infty$ almost surely for a constant C . Then C is not smaller than the price P_0 based on the dual formulation. On the other hand, let τ^ε be the first entry time of $Z_t - M_t$ in the region $[C - \varepsilon, C]$, then $C - \varepsilon \leq Z_{\tau^\varepsilon} - M_{\tau^\varepsilon} \leq C$. Let $\lim_{\varepsilon \rightarrow 0} \tau^\varepsilon = \tau$ be the limiting stopping time, by dominated convergence theorem,

$$C = \mathbb{E}^* \{ Z_\tau - M_\tau \mid \mathcal{F}_0 \} = \mathbb{E}^* \{ Z_\tau \mid \mathcal{F}_0 \}.$$

Hence, from the primal formulation, C is not larger than the price. Therefore, C is equal to the option price P_0 and by the uniqueness of M^* in Theorem 2.1(ii) M must be M^* .

($\Leftarrow$) Follows directly from Theorem 2.1(i).

For high-biased estimates, Theorem 2.2 points out that a martingale closer to the optimal hedging martingale possibly induces a lower upper-bound estimate for the

option price and a smaller variance for the high-biased estimator. This property will be illustrated by numerical results implemented in Sections 3 and 4.

On the other hand, for the low-biased estimate, the variance of the optimally stopped payoff Z_{τ^*} is $\text{Var}\{M_{\tau^*}^*\}$ as seen from Theorem 2.1(ii). We show next that this variance can potentially be reduced by considering an unbiased control variate $Z_{\tau^*} - M_{\tau^*}$ given a hedging martingale control $M \in H_0^1$. $\square$

PROPOSITION 2.2. *Given an optimal stopping time $0 \leq \tau^* \leq T$ and for any given integrable martingale $M \in H_0^1$, the variance of the low-biased estimate satisfies*

$$\text{Var}\{Z_{\tau^*} - M_{\tau^*}\} \leq \text{Var}\{M_T^* - M_T\}.$$

PROOF.

$$\begin{aligned} \text{Var}\{Z_{\tau^*} - M_{\tau^*}\} &= \mathbb{E}^* \left\{ (Z_{\tau^*} - M_{\tau^*} - P_0)^2 \mid \mathcal{F}_0 \right\} \\ &= \mathbb{E}^* \left\{ (M_{\tau^*}^* - M_{\tau^*})^2 \mid \mathcal{F}_0 \right\} \\ &\leq \text{Var}\{M_T^* - M_T\}. \end{aligned}$$

When an arbitrary stopping time is used, the error bound and the variance between its low-biased estimate and the American option price are not given explicitly here. However, an asymptotic result on the least squares method LONGSTAFF and SCHWARTZ [2001] shows that the optimal stopping rule or the free boundary can be realized when the number of simulated trajectory and the number of basis used to estimate the continuation value go to infinity.

To summarize, we observe an asymmetric effect for pricing American options from the point of view of variance reduction. A better hedging martingale provides a smaller variance for high- and low-biased estimators. But it preserves the bias for the low-biased estimate while shrinking the high-bias price gap as the hedging martingale approaches the optimal martingale. $\square$

3. Numerical results I: one-dimensional case

This section concerns a typical American put option pricing problem under the Black–Scholes model. That is, under the risk-neutral probability measure, the underlying risky stock price S_t is governed by the geometric Brownian motion

$$dS_t = r S_t dt + \sigma S_t dW_t,$$

where r is the risk-free interest rate and W_t is a Brownian motion.

The American put option price at time t is given as an optimal stopping problem

$$P(t, S_t) = \text{ess sup}_{t \leq \tau \leq T} \mathbb{E}^* \left\{ e^{-r(\tau-t)} (K - S_\tau)^+ \mid S_t \right\}, \quad (3.1)$$

with τ being a bounded stopping time between the current time t and the maturity T and where we have used the Markov property of S_t .

PROPOSITION 3.1. *The optimal stopping time $0 \leq \tau^* \leq T$ of the American option price $P(0, S_0)$ is the first time that maximizes the hedging error, namely, for any time $0 \leq u < \tau^*$*

$$e^{-ru}(K - S_u)^+ - M_u^* < \sup_{0 \leq t \leq T} (e^{-rt}(K - S_t)^+ - M_t^*),$$

but the equality holds when $t = \tau^*$, namely,

$$\tau^* = \inf \{0 \leq t \leq T, e^{-rt}(K - S_t)^+ - M_t^* = P(0, S_0)\}.$$

PROOF. For any time $0 \leq t < \tau^*$, the exercise payoff must be less than the American option price $(K - S_t)^+ < P(t, S_t)$ and

$$e^{-rt}(K - S_t)^+ - M_t^* < e^{-rt}P(t, S_t) - M_t^* = P(0, S_0) - A_t^* \leq P(0, S_0) \quad (3.2)$$

by the Doob–Meyer decomposition (2.5). We see that the discounted payoff $e^{-rt}(K - S_t)^+$ is superhedged by the hedging portfolio $P(0, S_0) + M_t^*$ at any time prior to the optimal stopping time. Combining with Theorem 2.1(ii), we conclude that τ^* is the first time maximizing the hedging error $Z_t - M_t^*$. If $\tau^* = 0$, it is a trivial case. $\square$

3.1. Hedging martingales

It is known that there is no closed-form solution for the American option price $P(t, x)$ given by (3.1). ROGERS [2002] introduced the counterpart European put option price, denoted by P_E and constructed the hedging martingale $e^{-rt}P_E(t, S_t) - P_E(0, S_0)$. This choice is useful because P_E admits a closed-form solution, known as the Black–Scholes formula for put options. Instead, we write an equivalent integral representation of that hedging martingale as

$$M(P_E; t) = \int_0^t e^{-rs} \frac{\partial P_E}{\partial x}(s, S_s) \sigma S_s dW_s, \quad (3.3)$$

obtained by an application of Ito's lemma to $e^{-rt}P_E(t, S_t)$. The main advantage of (3.3) is that any approximate American option price $\tilde{P}$ can constitute an integral martingale $M(\tilde{P}; t)$ in addition to P_E , without requiring that $e^{-rt}\tilde{P}_t$ be a martingale. Algorithms to compute the high- and low-biased estimates for the American put option are based on Proposition 2.1. The Monte Carlo estimator for the high-biased estimator is

$$\frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} \left(e^{-rt}(K - S_t^{(i)})^+ - M^{(i)}(\tilde{P}; t) \right) \quad (3.4)$$

and for the low-biased estimator is

$$\frac{1}{N} \sum_{i=1}^N [e^{-r\tau}(K - S_\tau^{(i)})^+ - M^{(i)}(\tilde{P}; \tau)], \quad (3.5)$$

where the approximation $\tilde{P}$ will be easy to compute, for example, the counterpart European option price P_E or the quadratic approximation P_{BAW} , introduced by BARONE-ADESI and WHALEY [1987]. The total number of i.i.d. trajectories is denoted by N , the superscript (i) denoting the i -th replication, and τ denoting a stopping rule, obtained by least squares method LONGSTAFF and SCHWARTZ [2001].

Based on the solution of an elliptic-type variational inequalities shown in Eq. (A.1) in the Appendix, the approximation P_{BAW} admits the following analytic solution

$$P_{BAW}(t, x) = \begin{cases} \lambda x^\alpha + P_E(t, x), & x > x^*(t), \\ K - x, & x \leq x^*(t), \end{cases}$$

where $P_E(t, x)$ denotes the counterpart European put option price, and where the approximate free boundary $x^*(t)$ solves a nonlinear algebraic equation

$$x^*(t) = |\alpha| \frac{K - P_E(t, x^*)}{\frac{\partial P_E(t, x^*(t))}{\partial x} + 1 + |\alpha|},$$

with parameters

$$\alpha = \frac{1 - \frac{2r}{\sigma^2} - \sqrt{(1 - \frac{2r}{\sigma^2})^2 + \frac{8(\kappa r + 1)}{\kappa \sigma^2}}}{2}$$

$$\lambda = \frac{K - x^*(t) - P_E(t, x^*(t))}{(x^*(t))^\alpha}.$$

It is shown in Lemma A.1 in the Appendix that the discounted approximate price $e^{-rt} P_{BAW}$ is not a martingale or a supermartingale. It cannot be used in Proposition 2.1 to estimate American option prices. However, the stochastic integral $M(P_{BAW}; t)$ is guaranteed to be a martingale. The martingale property of stochastic integrals not only provides a larger class for the computational purpose but also it is a clear demonstration of delta hedging strategy used in dynamic trading.

We are now ready to compare these hedging martingales $M(P_E; t)$ and $M(P_{BAW}; t)$ when estimating high- and low-biased solutions for American option prices. Parameters of one-dimensional American put options are as follows: the strike price $K = 100$, the risk-free interest rate $r = 6\%$, maturity $T = 0.5$ year, and the volatility $\sigma = 0.4$ (Table 3.1). The initial stock price S_0 is varying from 80 to 120. We run $N = 5000$ sample paths, and for each trajectory, we use the discretized time step $\Delta t = 0.001$. The true prices shown in column 5 in Table 3.1 are identical to the example in ROGERS [2002].

Low-biased estimates and their standard errors for American option prices are illustrated between the second column and the fourth column. Results in column 2 are calculated from the least squares algorithm LONGSTAFF and SCHWARTZ [2001] where there are no hedging martingales within the price estimator. Columns 3 and 4 illustrate

TABLE 3.1

Numerical results I. Comparisons of high-biased price estimates (columns 6–8), low-biased price estimates (columns 2–4), and actual American option prices (column 5). Two hedging martingales $M_t(P_E)$ and $M_t(P_{BAW})$ are constructed from the counterpart European option price: P_E and the quadratic approximation P_{BAW} , respectively. Model parameters are chosen as in ROGERS [2002]: $K = 100$, $r = 0.06$, $T = 0.5$, and $\sigma = 0.4$, with various initial stock prices ranging from 80 to 120. Monte Carlo simulations are implemented under the sample size $N = 5000$ and 500 discrete time steps corresponding to $\Delta t = .001$

S_0	LSM	$M_t(P_E)$	$M_t(P_{BAW})$	True price	$M_t(P_{BAW})$	$M_t(P_E)$	SM_t
80	21.522 (0.1507)	21.513 (0.0131)	21.592 (0.0108)	21.606	21.754 (0.0097)	21.947 (0.0107)	22.637 (0.0092)
85	17.907 (0.1631)	17.952 (0.0138)	17.999 (0.0125)	18.037	18.203 (0.0121)	18.325 (0.0128)	18.793 (0.0093)
90	14.817 (0.1706)	14.874 (0.0155)	14.845 (0.0139)	14.919	15.073 (0.0129)	15.132 (0.0143)	15.482 (0.0085)
95	12.141 (0.1640)	12.163 (0.0153)	12.202 (0.0155)	12.231	12.371 (0.0138)	12.391 (0.0148)	12.649 (0.0075)
100	9.993 (0.1585)	9.868 (0.0158)	9.880 (0.0150)	9.946	10.090 (0.0144)	10.147 (0.0153)	10.270 (0.0066)
105	8.214 (0.1497)	8.023 (0.0166)	8.026 (0.0154)	8.028	8.140 (0.0146)	8.181 (0.0151)	8.275 (0.0056)
110	6.205 (0.1304)	6.355 (0.0160)	6.433 (0.0153)	6.435	6.564 (0.0143)	6.612 (0.0149)	6.625 (0.0048)
115	5.126 (0.1219)	5.085 (0.0157)	5.055 (0.0150)	5.127	5.256 (0.0135)	5.269 (0.0141)	5.280 (0.0041)
120	4.230 (0.1162)	4.029 (0.0147)	4.039 (0.0143)	4.061	4.184 (0.0128)	4.198 (0.0134)	4.180 (0.0033)

effects of martingale controls $M(P_E; t)$ and $M(P_{BAW}; t)$, respectively, under the least squares method. We make the following observations.

- First, these control variates are unbiased to least squares estimators, but the standard errors with martingales are greatly reduced compared with the least squares estimators. The variance reduction ratios are roughly between 60 and 200. As the low-biased estimates should behave, sample means within column 2–column 4 are all smaller than the true prices shown in column 5.
- Second, the algorithm using $M(P_{BAW}; t)$ improves the precision of the low-biased estimates obtained from $M(P_E; t)$ as the variance produced by $M(P_{BAW}; t)$ is smaller than $M(P_E; t)$ except when $S_0 = 95$. Columns 6 and 7 illustrate high-biased estimates based on the algorithm using the martingales $M(P_{BAW}; t)$ and $M(P_E; t)$, respectively. Compared with the true price in column 5, sample means obtained from these martingales are all high biased, and the price gap in column 6 is smaller than those in column 7. Moreover, the standard errors in column 6 are all smaller than those in column 7. This justifies the asymmetric property between the bias and variance in Theorem 2.2, namely, a small variance implies a small bias.

We do not report mean absolute deviation (MAD) from the mean defined in ROGERS [2002] as we now have both high- and low-biased price estimates for which the actual option price is in between. By cross-comparison between columns 3–7 and columns 4–6, we find that P_{BAW} does provide a better approximation than P_E as P_{BAW} produces smaller variances than P_E does.

3.2. Errors in delta approximations

As suggested in Theorem 2.2, a martingale close to the optimal one will induce a smaller price gap and a smaller variance for the high-biased estimate. We measure the distance between two martingales by using the second moment or the variance. It is shown in FOUQUE and HAN [2007] that the variance is bounded above

$$\mathbb{E}^* \left\{ (M_T^* - M(\tilde{P}; T))^2 | \mathcal{F}_0 \right\} \leq C \sqrt{\int_0^T \mathbb{E}^* \left\{ \left(\frac{\partial P}{\partial x} - \frac{\partial \tilde{P}}{\partial x} \right)^2 (t, S_t) | \mathcal{F}_0 \right\} dt}, \quad (3.6)$$

where the constant C depends only on the initial stock price S_0 and the volatility σ . The mean square of the delta difference $\frac{\partial P}{\partial x} - \frac{\partial \tilde{P}}{\partial x}$ is crucial to control the distance between hedging martingales. There is no guarantee that a better price approximation provides a better delta approximation. The study of delta approximation for European option prices can be found in FOUQUE and HAN [2007] under multiscale stochastic volatility models. It remains a challenging task to study delta approximation for American options. At least from numerical results, one can see a strong empirical support that on average the approximate price P_{BAW} provides a better delta approximation than the European option price P_E . Note that these comparisons are useful to justify the effectiveness of price approximations.

3.3. Hedging supermartingales

We should mention an important result from HAUGH and KOGAN [2004]. Rather than martingales, they used supermartingales to obtain high-biased estimates for the American option price

$$P(0, S_0) = \inf_{\bar{M} \in \bar{H}^1} \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (e^{-rt} (K - S_t)^+ - \bar{M}_t) | \mathcal{F}_0 \right\} + \bar{M}_0, \quad (3.7)$$

where $\bar{H}^1 = \{(\bar{M}_t)_{0 \leq t \leq T} : \text{supermartingale with } \sup_{0 \leq t \leq T} |\bar{M}_t| \in L^1\}$. It is shown in HAUGH and KOGAN [2004] that the infimum can be obtained by choosing $\bar{M}_t^* = e^{-rt} P(t, S_t)$ such that

$$P(0, S_0) = \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} (e^{-rt} (K - S_t)^+ - (\bar{M}_t^* - \bar{M}_0^*)) | \mathcal{F}_0 \right\}. \quad (3.8)$$

PROPOSITION 3.2. *For the American option pricing problem (3.1), the supermartingale characterization (3.7) by HAUGH and KOGAN [2004], and the martingale characterization (2.1) by ROGERS [2002] are the same in the following sense: at the optimal stopping time τ^* , defined in Proposition 3.1, the optimizer of supermartingale $\overline{M}_{\tau^*}^* - \overline{M}_0^*$ is equal to the optimizer of martingale $M_{\tau^*}^*$.*

PROOF.

1. We first show that the hedging supermartingale representation holds almost surely

$$P(0, S_0) = \sup_{0 \leq t \leq T} \left(e^{-rt} (K - S_t)^+ - (e^{-rt} P(t, S_t) - P(0, S_0)) \right). \quad (3.9)$$

By Doob-Meyer decomposition as in (2.5), we obtain

$$e^{-rt} P(t, S_t) - P(0, S_0) = M_t^* - A_t^* \quad (3.10)$$

such that

$$\begin{aligned} & \sup_{0 \leq t \leq T} \left(e^{-rt} (K - S_t)^+ - (e^{-rt} P(t, S_t) - P(0, S_0)) \right) \\ & \geq \sup_{0 \leq t \leq T} \left(e^{-rt} (K - S_t)^+ - M_t^* \right) = P(0, S_0) \end{aligned}$$

by $A_t^* \geq 0$ and Theorem 2.1(ii). The supremum of hedging errors by supermartingales $e^{-rt} P(t, S_t)$ is no less than the true price $P(0, S_0)$. This contradicts Eq. (3.8) unless the supremum is the price almost surely. Thus, we obtain Eq. (3.9).

2. Substituting the decomposition (3.10) in (3.9), we deduce

$$P(0, S_0) = \sup_{0 \leq t \leq T} \left(e^{-rt} (K - S_t)^+ - (M_t^* - A_t^*) \right).$$

By Proposition 3.1, the optimal stopping time τ^* is the first time such that

$$P(0, S_0) = e^{-r\tau^*} (K - S_{\tau^*})^+ - M_{\tau^*}^*,$$

with $A_{\tau^*}^* = 0$. Hence $\overline{M}_{\tau^*}^* - \overline{M}_0^* = e^{-r\tau^*} P(t, S_t) - P(0, S_0) = M_{\tau^*}^*$.

□

Given a supermartingale $\overline{M}_t \in \overline{H}^1$, one can calculate a high-biased estimate for the American option price

$$P(0, S_0) \leq \mathbb{E}^* \left\{ \sup_{0 \leq t \leq T} \left(e^{-rt} H(S_t) - (\overline{M}_t - \overline{M}_0) \right) \right\}.$$

As revealed from Lemma A.1, the approximate early exercise premium with a weighted discount factor is a supermartingale. So, we propose the supermartingale

$SM_t = e^{-rt} P_E(t, S_t) + e^{-(r+\frac{1}{\kappa(t)})t} V(S_t)$, where the early exercise premium approximation is given by $V(S_t) = P_{BAW}(t, S_t) - P_E(t, S_t)$, and we construct the following supermartingale control:

$$SM_t - SM_0 = e^{-rt} P_E(t, S_t) + e^{-(r+\frac{1}{\kappa(t)})t} V(S_t) - P_{BAW}(0, S_0).$$

Based on the high-biased estimate

$$E^* \left\{ \sup_{0 \leq t \leq T} (e^{-rt} H(S_t) - (SM_t - SM_0)) \right\},$$

numerical results are shown in the last column of Table 3.1. Because of the supermartingale property of $e^{-(r+\frac{1}{\kappa(t)})t} V(S_t)$, the bias estimated is larger than those obtained from martingale control $M(P_E; t)$. But it is surprising to see that the standard errors obtained from the supermartingale SM_t algorithm are the smallest compared with those obtained from $M(P_{BAW}; t)$ and $M(P_E; t)$. Though this phenomenon is not in contradiction with Theorem 2.2 for martingales, it remains to investigate further how to construct suitable supermartingale estimators in order to reduce the price gap while keeping small variance.

4. Numerical results II: stochastic volatility

4.1. Multiscale stochastic volatility models

Following FOUQUE, PAPANICOLAOU, SIRCAR and SOLNA [2003], we consider the following class of multiscale stochastic volatility models, under a risk-neutral pricing probability measure $\mathbb{P}^*$ parametrized by the combined market prices of volatility risk (Λ_1, Λ_2) :

$$\begin{aligned} dS_t &= rS_t dt + \sigma_t S_t dW_t^{(0)*}, \\ \sigma_t &= f(Y_t, Z_t), \\ dY_t &= \left[\frac{1}{\varepsilon} c_1(Y_t) + \frac{g_1(Y_t)}{\sqrt{\varepsilon}} \Lambda_1(Y_t, Z_t) \right] dt \\ &\quad + \frac{g_1(Y_t)}{\sqrt{\varepsilon}} \left(\rho_1 dW_t^{(0)*} + \sqrt{1 - \rho_1^2} dW_t^{(1)*} \right), \\ dZ_t &= \left[\delta c_2(Z_t) + \sqrt{\delta} g_2(Z_t) \Lambda_2(Y_t, Z_t) \right] dt \\ &\quad + \sqrt{\delta} g_2(Z_t) \left(\rho_2 dW_t^{(0)*} + \rho_{12} dW_t^{(1)*} + \sqrt{1 - \rho_2^2 - \rho_{12}^2} dW_t^{(2)*} \right), \end{aligned} \tag{4.1}$$

where S_t is the underlying asset price process with a constant risk-free interest rate r . The random stochastic volatility σ_t is driven by two stochastic processes Y_t and Z_t , varying on the time scales ε and $1/\delta$, respectively (ε is intended to be a short time scale, while $1/\delta$ is thought as a longer time scale). The vector

$(W_t^{(0)*}, W_t^{(1)*}, W_t^{(2)*})$ consists of three independent standard Brownian motions. The instant correlation coefficients ρ_1 , ρ_2 , and ρ_{12} satisfy $|\rho_1| < 1$ and $|\rho_2^2 + \rho_{12}^2| < 1$. The volatility function f is assumed to be bounded and bounded away from zero to avoid degeneracy though these assumptions are not crucial and can be relaxed to accommodate, for instance, Heston-type models with a Cox-Ingersoll-Ross (CIR) stochastic volatility factor. The coefficient functions of Y_t , namely, c_1 and g_1 , are assumed to be such that under the physical probability measure ($\Lambda_1 = \Lambda_2 = 0$), Y_t is ergodic. The Ornstein–Uhlenbeck process is a typical example by defining $c_1(y) = m_1 - y$ and $g_1(y) = v_1\sqrt{2}$ such that $1/\varepsilon$ is the rate of mean reversion, m_1 is the long-run mean, and v_1 is the long-run standard deviation. Its invariant distribution is $\mathcal{N}(m_1, v_1^2)$.

The coefficient functions of Z_t , namely, c_2 and g_2 , are assumed to be smooth enough in order to satisfy existence and uniqueness conditions for diffusions. The combined risk premia Λ_1 and Λ_2 are assumed to be smooth, bounded, and dependent on the variables y and z only. Within this setup, the joint process (S_t, Y_t, Z_t) is Markovian. We refer to FOUQUE, PAPANICOLAOU, SIRCAR and SOLNA [2003] for a detailed discussion on this class of models.

Under the stochastic volatility models considered, the American option price at time 0 with an integrable payoff function H is given by

$$P^{\varepsilon, \delta}(t, x, y, z) = \text{ess sup}_{t \leq \tau \leq T} \mathbb{E}^* \{ e^{-r(\tau-t)} H(S_\tau) | S_t = x, Y_t = y, Z_t = z \}, \quad (4.2)$$

where τ denotes any stopping time greater than or equal to t , bounded by T , and is adapted to the completion of the natural filtration generated by Brownian motions $(W_t^{(0)*}, W_t^{(1)*}, W_t^{(2)*})$. We consider a typical American put option pricing problem, namely, $H(x) = (K - x)^+$.

4.2. Projected hedging martingales from asymptotic expansion

As shown in Proposition 2.1 one needs to construct a martingale in order to calculate the high- and low-biased estimates for the American option price. Under the Black–Scholes model, the volatility is assumed to be constant. We have observed in previous section that the use of counterpart-discounted European option price, which admits a closed-form solution, as a martingale is adequate. Under stochastic volatility models, there no longer exists a closed-form solution for the European option price. A martingale being a discounted European option price must be computed by, for example, another Monte Carlo simulation. This computation of Monte Carlo on Monte Carlo is typically very time consuming. To overcome this difficulty, the authors, in FOUQUE and HAN [2007], proposed the following: first apply Ito’s lemma to $e^{-rt} P(t, S_t, Y_t, Z_t)$ and integrate from time 0 to τ^* . Then, a hedging martingale consists of three parts $M_t(\tilde{P}) = M_0(\tilde{P}; t) + M_1(\tilde{P}; t) + M_2(\tilde{P}; t)$, where $\tilde{P}(s, S_s, Y_s, Z_s)$ denotes any approximation to the true model price $P^{\varepsilon, \delta}(s, S_s, Y_s, Z_s)$ given by (4.2), the three martingales being given by

$$M_0(\tilde{P}; t) = \int_0^t e^{-rs} \frac{\partial \tilde{P}}{\partial x}(s, S_s, Y_s, Z_s) f(Y_s, Z_s) S_s dW_s^{(0)*}, \quad (4.3)$$

$$M_1(\tilde{P}; t) = \frac{1}{\sqrt{\varepsilon}} \int_0^t e^{-rs} \frac{\partial \tilde{P}}{\partial y}(s, S_s, Y_s, Z_s) g_1(Y_s) d\tilde{W}_s^{(1)*}, \quad (4.4)$$

$$M_2(\tilde{P}; t) = \sqrt{\delta} \int_0^t e^{-rs} \frac{\partial \tilde{P}}{\partial z}(s, S_s, Y_s, Z_s) g_2(Z_s) d\tilde{W}_s^{(2)*}, \quad (4.5)$$

where the Brownian motions are defined by

$$\begin{aligned} \tilde{W}_s^{(1)*} &= \rho_1 W_s^{(0)*} + \sqrt{1 - \rho_1^2} W_s^{(1)*}, \\ \tilde{W}_s^{(2)*} &= \rho_2 W_s^{(0)*} + \rho_{12} W_s^{(1)*} + \sqrt{1 - \rho_1^2 - \rho_{12}^2} W_s^{(2)*}. \end{aligned}$$

In general, the hedging martingale can include control parameters $\lambda_0, \lambda_1, \lambda_2$ such that

$$M_t(\tilde{P}; \lambda_0, \lambda_1, \lambda_2) = \lambda_0 M_0(\tilde{P}; t) + \lambda_1 M_1(\tilde{P}; t) + \lambda_2 M_2(\tilde{P}; t). \quad (4.6)$$

A projected martingale considered here is constructed from a combination of hedging martingales and asymptotic methods.

We now focus on an approximation of the American option price under stochastic volatility models. When the time scales $1/\varepsilon$ and δ are well separated, namely, $0 < \varepsilon, \delta \ll 1$, the American option price $P^{\varepsilon, \delta}(t, S_t, Y_t, Z_t)$ admits an asymptotic expansion following the arguments in FOUQUE, PAPANICOLAOU and SIRCAR [2001], FOUQUE, PAPANICOLAOU, SIRCAR and SOLNA [2003]. The leading order term in the expansion is given by

$$P_0(t, S_t; \bar{\sigma}(Z_t)) = \text{ess sup}_{t \leq \tau \leq T} E \left\{ e^{-r(\tau-t)} H(\bar{S}_\tau) | \bar{S}_t = S_t \right\}, \quad (4.7)$$

where the homogenized stock price $\bar{S}_t$ follows a geometric Brownian motion with the averaged volatility $\bar{\sigma}(z) = \sqrt{\langle f^2(y, z) \rangle_Y}$, and $\langle \cdot \rangle_Y$ denotes the averaging with respect to the invariant distribution of the fast varying process Y . Note that because the homogenized American option price $P_0(t, S_t; \bar{\sigma}(Z_t))$ does not depend on the Y process, $M_1(P_0; t)$ shown in (4.4) is omitted (in fact, it can be shown that the next term of order $\sqrt{\varepsilon}$ in the expansion is also independent of y so that $M_1(P_0; t)$ would only contribute to the order $\sqrt{\varepsilon}$, which justifies this omission). Since $M_2(P_0; t)$ in (4.5) is of small order $\sqrt{\delta}$, this martingale is also neglected. As a result, the hedging martingale (4.6) is reduced to $M_t(P_0) = \lambda_0 M_0(P_0; t)$.

As an American option under constant volatility, the homogenized American option $P_0(t, x; \bar{\sigma}(Z_t))$ does not admit a closed-form solution. We follow the discussion in Section 3 and use approximations to $P_0(t, x)$ in order to construct hedging martingales as stochastic integrals such as $M_t = M_0(P_E; t)$ or $M_t = M_0(P_{BAW}; t)$ in which we do not pursue the optimal λ_0 but simply take $\lambda_0 = 1$ as it is found near one in ROGERS [2002] under the Black–Scholes model. As a result, we can use the same algorithm in (3.4, 3.5) to estimate American option prices under stochastic volatility models, though a stopping rule must be calculated by, for example, the least squares method.

We consider American put options under two-factor stochastic volatility models, specified in Table 4.1 and Table 4.2. Results of high- and low-biased estimates for

TABLE 4.1
Parameters used in the two-factor stochastic volatility model (4.1)

r	m_1	m_2	ν_1	ν_2	ρ_1	ρ_2	ρ_{12}	Λ_1	Λ_2	$f(y, z)$
10%	-1	-1	1	1	-0.3	-0.3	0	0	0	$\exp(y + z)$

TABLE 4.2
Initial conditions and American
put option parameters

$\$S_0$	Y_0	Z_0	$\$K$	T years
90	-1	-1	100	1

TABLE 4.3
Numerical results: comparison of low-biased estimates and high-biased estimates with some projected hedging martingales under different sets of time scales

$1/\varepsilon$	δ	$LSM(\text{primal})$	$M_t(P_E)(\text{primal})$	$M_t(P_{BAW})(\text{primal})$	$M_t(P_{BAW})(\text{dual})$	$M_t(P_E)(\text{dual})$
100	0.01	21.83 (0.241)	21.70 (0.034)	21.69 (0.025)	22.29 (0.024)	22.89 (0.037)
75	0.1	21.69 (0.238)	21.57 (0.034)	21.57 (0.027)	22.33 (0.027)	22.86 (0.039)
50	1	21.90 (0.242)	21.53 (0.040)	21.51 (0.033)	22.37 (0.033)	22.91 (0.042)
25	10	21.10 (0.267)	21.38 (0.055)	21.31 (0.048)	22.29 (0.043)	22.94 (0.051)

price American put options are illustrated in Table 4.3 with various time scale parameters ε and δ . The discrete time step size is $\Delta t = .001$ and the total sample size is $N = 5000$.

Observations from numerical results in Table 4.3 can be made similar to those from Table 3.1. Low-biased estimates are all unbiased to the estimator using the least squares method. We see that the control $M_t(P_{BAW})$ provides slightly better variance reduction ratios than $M_t(P_E)$ does. High-biased estimates obtained from $M_t(P_{BAW})$ outperform those from $M_t(P_E)$ because they provide both smaller biases and errors.

5. Conclusion

We have shown that hedging martingales are crucial for both primal and dual approaches to estimating American option prices by Monte Carlo simulations. The hedging martingales can be constructed from any price approximation to American option prices.

We uncovered the following asymmetric relation between the biases and variances for primal approach and dual approach: the dual approach ensures that a good hedging martingale induces a lower high-biased estimate with a smaller variance, while the primal approach ensures that a good hedging martingale reduces the variance for a low-biased estimate given a stopping time. Moreover, under more realistic multifactor stochastic volatility models, we propose a projected hedging martingale obtained by an asymptotic expansion. Numerical results demonstrate the robustness and efficiency of this method.

Appendix A

The approximate American option price $P_{BAW}(t, x)$ is equal to the sum of the counterpart European option price denoted by $P_E(t, x)$ and an approximate early exercise premium $V(x; t)$, where $V(x; t)$ solves an elliptic-type variational inequalities

$$\begin{cases} \mathcal{A}_{BS}(\sigma)V(x; t) - \left(r + \frac{1}{\kappa(t)}\right)V(x; t) \leq 0 \\ V(x; t) \geq (K - x)^+ - P_E(t, x) \\ \left(\mathcal{A}_{BS}(\sigma)V(x; t) - \left(r + \frac{1}{\kappa(t)}\right)V(x; t)\right) \cdot (V(x; t) - (K - x)^+ + P_E(t, x)) = 0, \end{cases} \quad (\text{A.1})$$

with the differential operator $\mathcal{A}_{BS}(\sigma) = \frac{\sigma^2 x^2}{2} \frac{\partial^2}{\partial x^2} + rx \frac{\partial}{\partial x} - r$.

LEMMA A.1.

- (i) $e^{-rt}V(S_t)$ is not a supermartingale or $e^{-rt}P_{BAW}(t, S_t)$.
- (ii) $e^{-\left(r + \frac{1}{\kappa(t)}\right)t}V(S_t)$ is a supermartingale, where $\kappa(t) = \frac{e^{r(T-t)} - 1}{r} \geq 0$ for each $t \in [0, T]$.

PROOF.

1. For any time $t < \tau^*$,

$$\begin{aligned} d[e^{-rt}V(S_t)] &= e^{-rt}[\mathcal{A}_{BS}(\sigma)V(S_t) - rV(S_t)]dt + e^{-rt}\frac{\partial V}{\partial x}(S_t)\sigma S_t dW_t \\ &= e^{-rt}\frac{V(S_t)}{\kappa(t)}dt + e^{-rt}\frac{\partial V}{\partial x}(S_t)\sigma S_t dW_t. \end{aligned}$$

Since the drift term is greater than zero, $e^{-rt}V(S_t)$ is not a supermartingale. Because $e^{-rt}P_{BAW}(t, S_t) = e^{-rt}P_E(t, S_t) + e^{-rt}V(S_t)$, it cannot be a supermartingale as well.

2. By an application of Ito's lemma to $e^{-(r+\frac{1}{\kappa(s)})s} V(S_s)$, we obtain

$$\begin{aligned} d \left[e^{-(r+\frac{1}{\kappa(s)})s} V(S_s) \right] &= e^{-(r+\frac{1}{\kappa(s)})s} \left\{ \left(\frac{\sigma^2 S_s^2}{2} \frac{\partial^2 V}{\partial x^2}(S_s) + r S_s \frac{\partial V}{\partial x}(S_s) \right. \right. \\ &\quad \left. \left. - \left(r + \frac{1}{\kappa(s)} \right) V(S_s) \right) ds + \frac{\kappa'(s)s}{\kappa^2(s)} V(S_s) ds \right\} \\ &\quad + e^{-(r+\frac{1}{\kappa(s)})s} \frac{\partial V}{\partial x}(S_s) \sigma S_s dW_s. \end{aligned}$$

Since for any positive x , $\frac{\sigma^2 x^2}{2} \frac{\partial^2 V}{\partial x^2}(x) + r x \frac{\partial V}{\partial x}(x) - (r + \frac{1}{\kappa(s)}) V(x) \leq 0$, $\kappa'(s) \leq 0$, and $V(x) \geq 0$, the coefficient in the ds term above is negative or zero, and the supermartingale property of the process $e^{-(r+\frac{1}{\kappa(t)})t} V(S_t)$ follows.

□

References

- BARONE-ADESI, G., WHALEY, R.E. (1987). Efficient analytic approximation of American option values. *J. Financ.* **XLII** (2), 301–320.
- FOUQUE, J.-P., HAN, C.-H. (2007). A martingale control variate method for option pricing with stochastic volatility. *ESAIM Probabil. Stat.* **11**, 40–54.
- FOUQUE, J.-P., PAPANICOLAOU, G., SIRCAR, R. (2000). *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge University Press).
- FOUQUE, J.P., PAPANICOLAOU, G., SIRCAR, R. (2001). From the implied volatility skew to a Robust correction to Black-Scholes American option prices. *Int. J. Theoretical Appl. Financ.* **4** (4), 651–675.
- FOUQUE, J.-P., PAPANICOLAOU, G., SIRCAR, R., SOLNA, K. (2003). Multiscale stochastic volatility asymptotics. *SIAM J. Multiscale Model. Sim.* **2** (1), 22–42.
- GLASSERMAN, P. (2003). *Monte Carlo Methods in Financial Engineering* (Springer Verlag).
- HAUGH, M.B., KOGAN, L. (2004). Pricing American options: a duality approach. *Oper. Res.* **52** (2), 258–270.
- LONGSTAFF, F., SCHWARTZ, E. (2001). Valuing American options by simulation: a simple least-squares approach. *Rev. Financ. Stud.* **14**, 113–147.
- KARATZAS, I., SHREVE, S.E. (2000). *Brownian Motion and Stochastic Calculus*, 2/e (Springer).
- ROGERS, L.C.G. (2002). Monte carlo valuation of American options. *Math. Financ.* **12**, 271–286.
- TSITSIKLIS, J., VAN ROY, B. (2001). Regression methods for pricing complex American-style options. *IEEE T. Neural. Networ.* **12**, 694–703.

Downside and Drawdown Risk Characteristics of Optimal Portfolios in Continuous Time

Dennis Yang

ATMIF LLC, New Jersey, USA

E-mail address: dennis.yang@atmif.com

Minjie Yu

Department of Mathematics, City University of Hong Kong,

Hong Kong, China

E-mail address: minjie.yu@hotmail.com

Qiang Zhang¹

Department of Economics and Finance, City University of Hong Kong,

Hong Kong, China

E-mail address: mazq@cityu.edu.hk

Abstract

Downside risk and drawdown risk measures are two important measures that qualify the risk characteristics of a portfolio. In this chapter, we consider three well-known optimal dynamic strategies and examine in detail their risk characteristics in long-term investments and portfolio frontiers under various downside and drawdown risk measures. We determine which strategy among the three performs best in various parameter regions for a given downside or drawdown risk measure. An investigation on the correlation among different risk measures has also been carried out.

1. Introduction

Risk measure and optimal portfolio selection are both important issues in modern finance. In recent years, several continuous-time optimal dynamic strategies have been developed

¹The work of Q. Zhang was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, Project CityU 103205.

for attaining various goals over an investment horizon, and various risk measures have been proposed in the literature. However, the issue of comparing performance of these continuous-time strategies under various risk measures has not received much attention. The aim of this chapter is to fill this gap.

For a given risk measure, ideally the investor should use the optimal portfolio strategy, which can obtain the maximum expected return rate under this risk. However, a recent work by JIN, YAN and ZHOU [2005] showed that for downside risk measures, the mean-risk problem admits no optimal solution in the continuous-time setting. It is also known that the optimal strategies under various drawdown risk measures have not been found except for certain special cases. Therefore, comparing the performance of various existing continuous-time portfolio strategies under various proposed downside and drawdown risk measures in the literature becomes desirable and important. We consider the following situation: an investor has several portfolio strategies that he/she may use for the investment, then an important problem he/she will face is the selection of strategy for a given downside or drawdown risk measure. This issue certainly has practical importance, as a fund manager needs to know which strategy will perform better for a given downside or drawdown risk measure.

In this chapter, we will investigate and compare three existing well-known continuous-time portfolio strategies: modified mean-variance (MMV), shortfall probability minimization (SPM), and power utility maximization (PUM) under various downside and drawdown risk measures. We consider three downside risk measures, below-mean semi variance (SV), value at risk (VaR), and conditional value at risk (CVaR), which is also known as expected shortfall, and two drawdown risk measures, average-percentage drawdown (Add) and maximum-percentage drawdown (Mdd). We will determine that, for a given downside or drawdown risk measure, which one among the three continuous-time portfolio strategies performs best in various parameter regions, that is, for different values of drift and volatility of the stock, risk-free interest rate, expected return rate, and investment horizon. For comparison, the performance under the variance risk measure is also presented.

The outline of the chapter is as follows: In Section 2, we review previous works on risk measures and optimal dynamic strategies, chapter. In Section 3, we state the financial market model used, introduce dimensionless parameters, and summarize portfolio strategies. In Section 4, we present the definitions of the risk measures. From Sections 6 to 10, we examine the risk characteristics and portfolio frontier of three strategies under various downside and drawdown risk measures. The question regarding which strategy performs best for each given risk measure will be addressed. In Section 11, we examine the correlations among different risk measures. Section 12 concludes. All derivations and proofs are given in the Appendix.

2. Literature review

Variance was proposed as the first risk measure in the pioneering work of MARKOWITZ [1959, 1987]. However, a substantial amount of arguments have shown that variance is not a proper risk measure since investor's concerns were different between downside losses and upside gains. In fact, MARKOWITZ [1959] advocated to use SV, rather

than variance, as a measure of risk because SV weights downside losses differently from upside gains. Consequently, various downside risk measures were proposed by FISHBURN [1977], SORTINO and VAN DER MEER [1991], JORION [1997] and NAWROCKI [1999]. In the case of single-period investment, several mean-downside-risk models have been studied: for example, the mean-semivariance model studied by MARKOWITZ [1959], the mean-semideviation model by OGRYCAK and RUSZCZYNSKI [1989], the mean-VaR model by CAMPBELL, HUISMAN and KOEDIJK [2001], and the mean-CVaR model by ROCKAFELLAR and URYASEV [2000, 2002], and KROKHMAL, PALMQUIST and URYASEV [2001]. Comparisons among these models are also available in the literature (see ORTOBELLI, RACHEV, STOYANOV, FABOZZI and BIGLOVA [2005] and JARROW and ZHAO [2006]).

Naturally, one would like to find a continuous-time strategy that is optimal under a given downside risk measure. However, JIN, YAN and ZHOU [2005] proved recently that the mean-semivariance problem admits no optimal solution in the continuous-time setting. They further extended this conclusion to a general mean-downside-risk model. Therefore, a comparison of performances among various known dynamic strategies under downside risk measures becomes necessary and important. We will carry out such study in detail for three downside risk measures: below-mean SV, VaR, and CVaR. They are only related to the final distribution of the wealth, that is, two different portfolios with the same distribution function at the end of investment horizon will have the same risk. These risk measures work well in single-period models. But in continuous-time portfolio management, another type of risk measure called drawdown risk measure plays an important role as an index for historical performance. GROSSMAN and ZHOU [1993] proposed the maximum drawdown measures and argued that a reasonably low drawdown is critical to the success of any fund. The problem of maximizing the growth rate over the infinite horizon for certain maximum drawdown has been studied by GROSSMAN and ZHOU [1993] in one-dimensional case and then generalized by CVITANIĆ and KARATZAS [1995] to multidimensions. CHEKHLOV, URYASEV and ZABARANKIN [2003] proposed and solved the mean-conditional-drawdown model with the assumption that portfolio weights are static overtime. However, without making any special assumptions, it is not easy to derive analytical expressions for the solutions to the optimal strategies under these drawdown risk measures. In this chapter, we will focus on two important drawdown risk measures: Add and Mdd.

In developing optimal strategies in continuous-time setting, two approaches are commonly used: expected utility theory based on the pioneering work of VON NEUMANN and MORGENSTERN [1947] and mean-risk approach developed by MARKOWITZ [1952, 1959]. ORTOBELLI, RACHEV, STOYANOV, FABOZZI and BIGLOVA [2005] stated that the linkage between these two approaches is generally represented by the consistency of the risk measure in the latter approach with a stochastic dominance order that relates to utility functions of certain qualitative behavior in the former approach. This property allows to define three types of strategies as follows.

The first type is the optimal strategy for a risk-averse investor with a concave utility function. This is consistent with the Rothschild–Stiglitz (R-S) stochastic dominance order. The most well-known strategy for this type of investors is the one that maximizes the mean of final wealth of the portfolio for a given variance. This strategy was first

proposed in the single-period setting by MARKOWITZ [1952, 1959] in his pioneering work, then generalized to multiperiod settings by HAKANSSON [1971] and by GRAUER and HAKANSSON [1993], and to a continuous-time setting by ZHOU and LI [2000].

The second type is the optimal strategy for a nonsatiable investor with a nondecreasing utility function. This is consistent with the first stochastic dominance (FSD) order. A well-known strategy designed for this type of investors is the SPM strategy studied by BROWNE [1999]. It aims to minimize the probability of the portfolio value falling below a specified wealth level at a given investment horizon.

The third type is the optimal strategy for a nonsatiable risk-averse investor with a nondecreasing concave utility function, which is consistent with the second stochastic dominance (SSD) order. The most well-known strategy designed for this type of investors is the portfolio selection based on the PUM, introduced by MERTON [1971] in his famous work.

However, as YU, ZHANG and YANG [2006] showed, mean-variance strategy, which belongs to the class of strategies of no lower bound in the value of the portfolio, will lead to a sure bankruptcy in long-term investments. Therefore, we consider the MMV strategy proposed by BIELECKI, JIN, PLISKA and ZHOU [2005], rather than the mean-variance strategy, in our comparison study. The MMV strategy imposes the nonnegative wealth restriction to rule out arbitrage possibilities that exist in the original mean-variance portfolio selection strategy.

3. Summary of dynamic strategies

In this section, we first present the financial market model adopted in this chapter. Then, based on this model, we review the analytical formulas for the MMV, SPM, and PUM strategies.

The model we use for a financial market is same as that in MERTON [1971], which consists of n log-normal distributed risky assets governed by

$$dS_i(t) = S_i(t) \left[\mu_i dt + \sum_{j=1}^n \sigma_{ij} dW_t^{(j)} \right] \quad (i = 1, \dots, n) \quad (3.1)$$

and a money market with a constant risk-free interest rate r . Here, μ_i and σ_{ij} are growth rate and volatility of risky asset S_i , respectively, and $\mathbf{W}_t := (W_t^{(1)}, \dots, W_t^{(n)})'$ is a standard n -dimension Brownian motion. For all aforementioned strategies, the n -risky-asset problem is equivalent to a one-risky-asset problem (KHANNA and KULLDORFF [1999], YU, ZHANG and YANG [2006]). Therefore, for simplicity, our presentation below will be based on market with single risky asset. All results presented in this chapter can be easily transformed to the market with multirisky assets by a simple substitution. The detail of substitution can be found in Section 5 of YU, ZHANG and YANG [2006].

The price process of the equivalent single risky asset is governed by

$$dS(\tilde{t}) = S(\tilde{t})(\tilde{\mu}d\tilde{t} + \tilde{\sigma}dW(\tilde{t})), \quad (3.2)$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are constants, $\tilde{\mu} > r$, $\tilde{\sigma} > 0$, and $W(\tilde{t})$ is a standard Brownian motion.

Let $X(\tilde{t})$ denote the present value of an investor's wealth at time $\tilde{t}$, which is discounted by the factor $\exp(-r\tilde{t})$, and $\tilde{\pi}(\tilde{t})$ be the total discounted wealth invested in the risky asset at time $\tilde{t}$. Then, the discounted wealth process $X(\cdot)$ obeys

$$\begin{aligned} dX(\tilde{t}) &= \tilde{\pi}(\tilde{t})(\tilde{\mu} - r)d\tilde{t} + \tilde{\pi}(\tilde{t})\tilde{\sigma}dW(\tilde{t}) \\ X(0) &= x. \end{aligned} \quad (3.3)$$

We use the symbol “~” to indicate that these quantities are dimensional.

Let $\tilde{R}$ be the expected return rate adjusted by the discount rate over an investment horizon $\tilde{T}$, that is, $E(X(\tilde{T})) = xe^{\tilde{R}\tilde{T}}$. Different values of $\tilde{R}$ lead to a different wealth allocation to the risky asset for each given strategy. Therefore, $\tilde{R}$ is a free-varying parameter and leads to a one-parameter family of portfolio choices for each strategy.

We introduce the following dimensionless quantities

$$R := \left(\frac{\tilde{\mu} - r}{\tilde{\sigma}} \right)^{-2} \tilde{R}, \quad (3.4)$$

$$t := \left(\frac{\tilde{\mu} - r}{\tilde{\sigma}} \right)^2 \tilde{t}, \quad (3.5)$$

and a scaled quantity

$$\pi(\cdot) := \left(\frac{\tilde{\mu} - r}{\tilde{\sigma}^2} \right)^{-1} \tilde{\pi}(\cdot). \quad (3.6)$$

Obviously, the dimensionless investment horizon T can be expressed as $T = \left(\frac{\tilde{\mu} - r}{\tilde{\sigma}} \right)^2 \tilde{T}$ according to Eq. (3.5), which depends not only on the dimensional investment horizon $\tilde{T}$ but also on the dimensional drift $\tilde{\mu}$, interest rate r , and volatility $\tilde{\sigma}$.

In terms of these quantities, the wealth process (3.3) can be rewritten as

$$\begin{aligned} dX(t) &= \pi(t)dt + \pi(t)dW(t) \\ X(0) &= x. \end{aligned} \quad (3.7)$$

Now, we redefine $S(t)$ to be the discounted stock price, that is, $S(t) = e^{-r\tilde{t}}S(\tilde{t})$, then the dynamic equation for discounted stock, expressed in terms of the dimensionless time t , is

$$dS(t) = S(t)\alpha(dt + dW(t)), \quad (3.8)$$

where $\alpha := \sigma^2/(\tilde{\mu} - r)$. We comment that the dimensionless parameter α will not appear in the rest of the chapter since the following derivations are based only on Eq. (3.7) in which α is absorbed into the definition of $\pi(\cdot)$ (see Eq. (3.6)).

Unless otherwise specified, dimensionless and scaled quantities will be used in the rest of this chapter. It should be noted that the essential factors controlling the portfolio management are only two dimensionless parameters R and T , instead of five dimensional parameters r , $\tilde{\mu}$, $\tilde{\sigma}$, $\tilde{R}$, and $\tilde{T}$ in the original statement of the problem. The effect of all

other factors is simply brought in by inverting the above transformation, that is, by a straightforward arithmetical calculation from Eqs. (3.4)–(3.6), which can also recover the corresponding results in unscaled dimensional quantities.

We study following three continuous-time portfolio strategies.

3.1. MMV strategy

It is known that the wealth process $X(t)$ of an optimal continuous-time portfolio based on mean variance can become negative within the investment horizon. To overcome this problem, BIELECKI, JIN, PLISKA and ZHOU [2005] studied the mean-variance portfolio selection problem under the restriction that the wealth cannot be negative over the entire investment horizon.

The formulation of this problem is

$$\begin{aligned} \text{Goal: } \min \quad & \text{Var}(X(T)) \\ \text{such that } \quad & E(X(T)) = xe^{R_1 T} \\ & X(t) \geq 0 \text{ a.s., } \forall t \in [0, T], \end{aligned}$$

which has the solution

$$\pi_1(t) = \omega_1 \Phi(-d_-(t, y(t))) - X(t), \quad (3.9)$$

$$X(t) = \omega_1 \Phi(-d_-(t, y(t))) - \Phi(-d_+(t, y(t))) y(t), \quad (3.10)$$

where

$$y(t) = \omega_2 e^T e^{-\frac{3}{2}t - W(t)} \quad (3.11)$$

$$d_+(t, y) = \frac{\ln(y/\omega_1) + \frac{1}{2}(T - t)}{\sqrt{T - t}} \quad (3.12)$$

$$d_-(t, y) = d_+(t, y) - \sqrt{T - t}. \quad (3.13)$$

Here, $W(t)$ is the standard Brownian motion and (ω_1, ω_2) is the unique solution to the following equations:

$$\begin{cases} \omega_1 \Phi\left(\frac{\ln(\omega_1/\omega_2) - \frac{1}{2}T}{\sqrt{T}}\right) - \omega_2 e^T \Phi\left(\frac{\ln(\omega_1/\omega_2) - \frac{3}{2}T}{\sqrt{T}}\right) = x \\ \omega_1 \Phi\left(\frac{\ln(\omega_1/\omega_2) + \frac{1}{2}T}{\sqrt{T}}\right) - \omega_2 \Phi\left(\frac{\ln(\omega_1/\omega_2) - \frac{1}{2}T}{\sqrt{T}}\right) = x e^{R_1 T} \end{cases}, \quad (3.14)$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function.

We would like to comment that, by definition, the target return rate R_1 in Eq. (3.9) is also the expected return rate of the wealth for this strategy.

From Eq. (3.10), after some calculations, one can obtain the following expression for $X(T)$

$$X(T) = \max(0, \omega_1 - \omega_2 e^{-\frac{1}{2}T - W(T)}) := \left(\omega_1 - \omega_2 e^{-\frac{1}{2}T - W(T)}\right)^+. \quad (3.15)$$

3.2. SPM strategy

This strategy is designed to minimize the probability of the value of portfolio below a specified wealth level at a given investment horizon. This probability is known as shortfall probability.

This problem can be expressed as

$$\text{Goal: } \min_{\pi} P \left[X(T) < x e^{R'_2 T} \right].$$

BROWNE [1999] studied this problem and gave an explicit solution

$$\pi_2(t) = \frac{x e^{R'_2 T}}{\sqrt{T-t}} \phi \left(\Phi^{-1} \left(\frac{X(t)}{x e^{R'_2 T}} \right) \right) \quad (3.16)$$

$$X(t) = x e^{R'_2 T} \Phi \left(\frac{1}{\sqrt{T-t}} \left[W(t) + t + \sqrt{T} \Phi^{-1} \left(e^{-R'_2 T} \right) \right] \right) \quad (3.17)$$

for $(0 \leq t < T)$, where $\phi(\cdot)$ denotes the density function of a standard normal variable,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (3.18)$$

It should be noted that the target return rate R'_2 is different from the expected return rate R_2 . For a given R'_2 , R_2 can be determined numerically from the following relation

$$E(X(T)) = x e^{R'_2 T} \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R'_2 T} \right) \right) = x e^{R_2 T}. \quad (3.19)$$

Since the comparison among different strategies is only meaningful under the same expected return rate, we will use R_2 not R'_2 when we compare the SPM strategy with other two strategies.

However, it is easy to check that these two rates coincide in the long investment horizon limit provided that $R'_2 < \frac{1}{2}$ due to the following relation:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \ln(E(X(T)/x)) = R'_2, \quad \forall R'_2 < \frac{1}{2}. \quad (3.20)$$

Equation (3.17) shows that the final wealth $X(T)$ satisfies the binomial distribution, that is,

$$X(T) = \begin{cases} x e^{R'_2 T} & \text{if } W(T) + T + \sqrt{T} \Phi^{-1} \left(e^{-R'_2 T} \right) > 0 \\ 0 & \text{if } W(T) + T + \sqrt{T} \Phi^{-1} \left(e^{-R'_2 T} \right) < 0 \end{cases}. \quad (3.21)$$

3.3. PUM strategy

This is a classic optimal investment strategy in continuous-time model introduced by MERTON [1971].

The formulation of this problem is

$$\text{Goal: } \max_{\pi} E \left[\frac{1}{\gamma} (X(T))^{\gamma} \right],$$

which has the solution

$$\pi_3(t) = \frac{1}{1 - \gamma} X(t). \quad (3.22)$$

From the equality,

$$E(X(T)) = x \exp \left(\frac{1}{1 - \gamma} T \right) := x \exp(R_3 T), \quad (3.23)$$

one can obtain the relation between the target return rate R_3 and the relative risk aversion parameter γ ,

$$\gamma = 1 - \frac{1}{R_3}. \quad (3.24)$$

The target return R_3 is also the expected return rate of the wealth. In terms of R_3 , $X(t)$ can be expressed as follows

$$X(t) = x \exp \left(\left(R_3 - \frac{R_3^2}{2} \right) t + R_3 W(t) \right). \quad (3.25)$$

4. Various risk measures

We introduce the definitions of various downside and drawdown risk measures, which will be discussed in detail subsequently.

Let x be the initial wealth of a portfolio, X be the wealth of a portfolio at time T , and $\bar{x} = E(X)$ be the mean of X . We introduce the definitions of three popular downside risk measures:

- (i) Below-mean SV is defined as

$$\text{SV}(X) := \int_{-\infty}^{\bar{x}} (\bar{x} - u)^2 dF(u). \quad (4.1)$$

Here, $F(x)$ is the distribution function of X .

Obviously, the below-mean SV only considers the samples with their final wealth less than their mean.

(ii) VaR is defined as

$$\text{VaR}_\alpha(X) := \bar{x} - Q_\alpha(X), \quad (4.2)$$

$$\text{where } Q_\alpha := \inf\{u : F(u) > \alpha\} \quad \alpha \in (0, 1). \quad (4.3)$$

VaR stands for the minimum loss incurred in the α worst case of the portfolio, where the loss is measured by the downside deviation from the expected final wealth $\bar{x}$. Usually, the values of α in these definitions are small, for example, 0.05 or 0.01.

(iii) CVaR, which is also known as expected shortfall, is defined as

$$\text{CVaR}_\alpha(X) := \bar{x} - C_\alpha(X), \quad (4.4)$$

$$\begin{aligned} \text{where } C_\alpha(X) := & \frac{P(X \leq Q_\alpha(X))}{\alpha} E(X|X \leq Q_\alpha(X)) \\ & + \left(1 - \frac{P(X \leq Q_\alpha(X))}{\alpha}\right) Q_\alpha(X). \end{aligned}$$

CVaR stands for the average loss incurred in the α worst case of the portfolio.

The definitions we adopt here for VaR and CVaR are same as those given by GAIVORONSKI and PFLUG [2004] and LEMUS RODRIGUEZ [1999]. It should be noted that another definition of VaR is given by BASAK and SHAPIRO [2001] and DOWD, BLAKE and CAIRNS [2004], which is also called as capital at risk (CaR) by EMMER, KLUPPELBERG and KORN [2001] and DMITRASINOVIC-VIDOVIC, LARA-LAVASSANI, LI and WARE [2003], where the loss is measured by the downside deviation from the initial wealth. For this type of VaR, EMMER, KLUPPELBERG and KORN [2001] and DOWD, BLAKE and CAIRNS [2004] show that although the value of VaR is bounded by the initial capital x , it has no corresponding lower bound and will fall infinitely as the time horizon continues to rise (note that negative VaR means that the likely worst outcome at the specified level of confidence is a profit, rather than a loss). Therefore, we adopt VaR measured from the mean value to discuss in this chapter.

We should mention that these two types of VaR give the same ranking among strategies since the comparison should be made with the same initial capital and expected final wealth. The only difference between them is the benchmark to measure the risk. All the statements we make about the best strategy under VaR can be translated into equivalent statements about VaR as it is defined by BASAK and SHAPIRO [2001] and DOWD, BLAKE and CAIRNS [2004] or CaR by EMMER, KLUPPELBERG and KORN [2001] and DMITRASINOVIC-VIDOVIC, LARA-LAVASSANI, LI and WARE [2003].

Next, we turn our attention to the path-dependent drawdown risk measures. Let $X_m(t)$ be the maximum wealth of a portfolio before time t , that is, $X_m(t) = \sup_{0 \leq s \leq t} X(s)$, then

we introduce the notion of the current drawdown $Cdd(t)$

$$Cdd(t) := \frac{X_m(t) - X(t)}{X_m(t)} \quad (4.5)$$

and the definitions of two important drawdown risk measures

(i) Add is defined as

$$Add(T) := \frac{1}{T} \int_0^T Cdd(t) dt. \quad (4.6)$$

(ii) Mdd is defined as

$$Mdd(T) := \sup_{0 \leq t \leq T} Cdd(t). \quad (4.7)$$

Notice that the above drawdown risk measures are defined on a sample path of portfolio process. We will choose their expected values as measuring risk, namely,

$$ADD(T) := E(Add(T)) \quad (4.8)$$

$$MDD(T) := E(Mdd(T)). \quad (4.9)$$

YU, ZHANG and YANG [2006] pointed out that for all three strategies mentioned before there is a threshold in the expected return rate for long-term investment, above which bankruptcy will surely happen. Therefore, in this chapter, we will only consider the performance of three strategies under these risk measures when the expected return rate is below the threshold, namely $R < \frac{1}{2}$.

5. More on mean-downside-risk models

Downside risk is an important risk measure, and there is vast literature on the subject of optimal portfolio strategies based on various downside risk measures.

LI and WU [2005] solve the mean-below-target-SV problem, which can be formulated as

$$\min_{\pi} E [\max(L - X(T), 0)]^2 \quad (5.1)$$

$$\text{such that } E(X(T)) = \bar{x} \quad \text{with } L > \bar{x} \quad (5.2)$$

$$E(\xi(T)X(T)) = \xi(0)x, \quad (5.3)$$

where L is the target wealth level and $\xi(t)$ is the state price density process. This problem has also been studied by JIN, YAN and ZHOU [2005], in which they further showed that the Eq. (5.1) is not well defined when $L = 1$, that is, there is no optimal solution under the mean-below-mean SV.

BASAK and SHAPIRO [2001] give a close-form solution to the utility maximization problem with a constrain on the VaR in continuous time, where the loss is measured

by the downside deviation from the initial wealth (as CaR defined in this chapter), for example, $P(x - X(T) \leq \text{VaR}_\alpha) = 1 - \alpha$. They formulate the problem as follows:

$$\max_{\pi} E[u(X(T))] \quad (5.4)$$

$$\text{such that } E(\xi(T)X(T)) \leq \xi(0)x \quad (5.5)$$

$$P(X(T) \geq x_\alpha) \geq 1 - \alpha, \quad (5.6)$$

where x_α satisfies

$$\text{VaR}_\alpha \leq x - x_\alpha. \quad (5.7)$$

GABIH, GRECKSCH and WUNDERLICH [2005] study the following utility maximization problem with a constrain on the expected loss (which is closely related to the definition of CVaR risk measure in this thesis) problem:

$$\max_{\pi} E[u(X(T))] \quad (5.8)$$

$$\text{such that } E(\xi(T)X(T)) \leq \xi(0)x \quad (5.9)$$

$$E[\max(q - X(T), 0)] \leq \epsilon, \quad (5.10)$$

where q is a wealth level, and ϵ is a given bound for the expected loss.

Both BASAK and SHAPIRO [2001] and GABIH, GRECKSCH and WUNDERLICH [2005] give the analytical solution based on the power utility function: $U(X) = \frac{X^{1-\gamma}}{1-\gamma}$ ($\gamma > 0$). However, similar to the situation of the mean-below-mean-SV, the mean-VaR and mean-expected-loss problem are not well defined, that is, the Eqs. (5.4) and (5.8) admit no optimal solutions under $\gamma = 0$.

It should be noted that the suboptimal strategies can be obtained by letting L be close to $\bar{x}$ in the Eq. (5.1) or letting γ be close to 0 in the Eqs. (5.4) and (5.8). However, the optimal strategy does not exist since the problems (5.1), (5.4), and (5.8) are not well defined as $L = \bar{x}$ and $\gamma = 0$.

6. Below-mean SV

In this section, we will investigate in detail the performance of three strategies under the SV, as well as the variance for comparison. As we mentioned earlier, all the comparisons will be made under the same expected return rate, that is, $E(X(T)) = x e^{RT}$.

6.1. Analytical formula

Due to the analytical expressions of final wealth for the MMV, SPM, and PUM strategies, given by Eqs. (3.15), (3.21), and (3.25), respectively, it is easy to derive their final wealth distributions. Then, after some algebraic calculations, we can obtain analytical

expressions of the below-mean SV and variance (Var) for the MMV, SPM, and PUM strategies as follows:

1. MMV strategy:

$$\begin{aligned} \text{SV}(X(T)) = & \omega_1 e^{R_1 T} - \omega_2 - e^{2R_1 T} - \omega_2^2 e^T \Phi \left(\frac{\ln((\omega_1 - e^{R_1 T})/\omega_2) - \frac{3}{2}T}{\sqrt{T}} \right) \\ & - 2\omega_2 (e^{R_1 T} - \omega_1) \Phi \left(\frac{\ln((\omega_1 - e^{R_1 T})/\omega_2) - \frac{1}{2}T}{\sqrt{T}} \right) \\ & - (e^{R_1 T} - \omega_1)^2 \Phi \left(\frac{\ln((\omega_1 - e^{R_1 T})/\omega_2) + \frac{1}{2}T}{\sqrt{T}} \right) \end{aligned} \quad (6.1)$$

$$\text{Var}(X(T)) = \omega_1 e^{R_1 T} - \omega_2 - e^{2R_1 T}, \quad (6.2)$$

where ω_1 and ω_2 are determined by Eq. (3.14).

2. SPM strategy:

$$\begin{aligned} \text{SV}(X(T)) = & x^2 e^{2R_2' T} \Phi^2 \left(\sqrt{T} + \Phi^{-1} \left(e^{-R_2' T} \right) \right) \\ & \times \left(1 - \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R_2' T} \right) \right) \right) \end{aligned} \quad (6.3)$$

$$\begin{aligned} \text{Var}(X(T)) = & x^2 e^{2R_2' T} \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R_2' T} \right) \right) \\ & \times \left(1 - \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R_2' T} \right) \right) \right) \end{aligned} \quad (6.4)$$

3. PUM strategy:

$$\text{SV}(X(T)) = x^2 e^{2R_3 T} \left[e^{R_3^2 T} \Phi \left(-\frac{3R_3 \sqrt{T}}{2} \right) - 2 + 3 \Phi \left(\frac{R_3 \sqrt{T}}{2} \right) \right] \quad (6.5)$$

$$\text{Var}(X(T)) = x^2 e^{2R_3 T} (e^{R_3^2 T} - 1). \quad (6.6)$$

6.2. Long-term asymptotic behavior

In this subsection, we focus on the performance of strategies in the long-term investment under the SV risk measures.

For the MMV and SPM strategies, we have the following binomial distribution for the SV and VaR risk measures.

PROPOSITION 6.1. *For both the MMV and SPM strategies,*

$$\lim_{T \rightarrow \infty} \text{SV}(X(T)) = \lim_{T \rightarrow \infty} \text{Var}(X(T)) = \begin{cases} 0 & \text{if } 0 < R \leq \frac{3}{2} - \sqrt{2} \\ \infty & \text{if } R > \frac{3}{2} - \sqrt{2} \end{cases}. \quad (6.7)$$

For the PUM strategy,

$$\lim_{T \rightarrow \infty} \text{SV}(X(T)) = \lim_{T \rightarrow \infty} \text{Var}(X(T)) = \infty \quad \forall R > 0. \quad (6.8)$$

PROOF. See Appendix B. $\square$

Proposition 6.1 shows that there exists a threshold $R^\sharp = \frac{3}{2} - \sqrt{2}$ in SV and variance when one follows the MMV and SPM strategies for long-term investment. The SV and variance will tend to zero for $R < R^\sharp$ and to infinite for $R > R^\sharp$.

Figure 6.1 shows the SV of the MMV, SPM, and PUM strategies as a function of investment time horizon T for different values of $R/R^\sharp$. The case of $R > R^\sharp$ is plotted in the top row, whereas the case of $R \leq R^\sharp$ is plotted in the bottom row because the vertical scales of the bottom panels are different from the ones of the top panels. It is clear that for the PUM strategy, the SV is an increasing function of T regardless of what R is. Furthermore, for $R > R^\sharp$, the SV is an increasing function of T for all three strategies. However, for the MMV and SPM strategies with $R \leq R^\sharp$, the SV is an increasing function of T for small value of T and a decreasing function of T for large value of T . Therefore, for a fixed R , there exists a maximum semivariance $\text{SV}^\sharp$ located at $T = T^\sharp$. For the investor who only wants to make his or her wealth grow better than sleeping in the bank, he/she may follow the MMV and SPM strategies in a sufficient longtime, for example, $T \gg T^\sharp$, with the expected return rate R less than the threshold $R^\sharp$.

Similar analysis can be carried out for the variance $\text{Var}(X(T))$, and the behavior is also similar to that of SV risk measure.

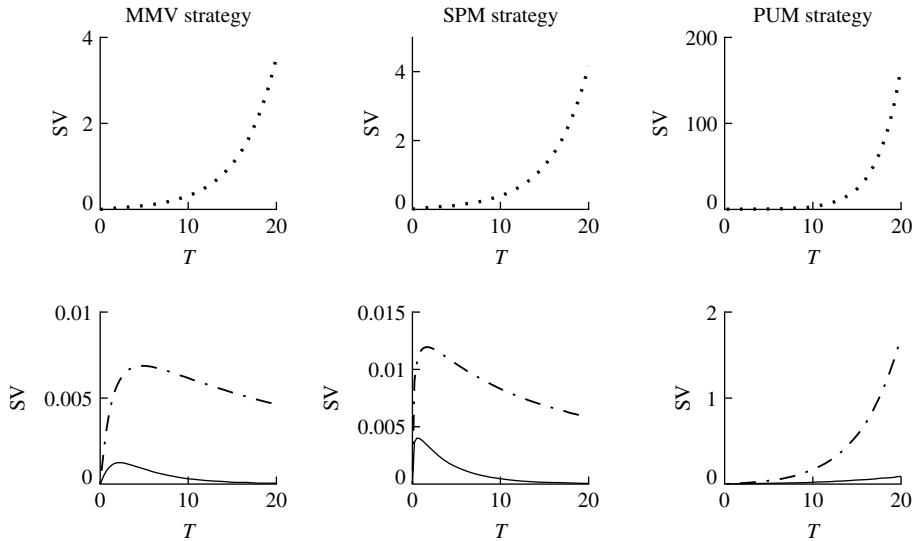


FIG. 6.1 $\text{SV}(X(T))$ as a function of T for different values of R . Solid, dash-dotted, and dotted lines stands for $R = 0.5R^\sharp$, $R^\sharp$, and $2R^\sharp$, respectively.

6.3. Portfolio frontier

An important concept in modern portfolio theory, the efficient frontier, was first defined by MARKOWITZ [1952], which represents variously weighted combinations of the portfolio's assets that yield the maximum possible expected return at any given level of portfolio risk. However, as shown by JIN, YAN and ZHOU [2005], no efficient frontier will exist under the SV risk measure or furthermore, downside risk measure in the continuous-time portfolio management.

Analogous to the analysis for the efficient frontier, we define the excess return of a strategy

$$r^*(T) := \frac{X(T) - X(0)}{X(0)} \quad (6.9)$$

and

$$E(r^*(T)) = e^{RT} - 1. \quad (6.10)$$

Then, we can plot the portfolio frontier in Fig. 6.2: the expected excess return versus standard semideviation for all three strategies to compare their performances, where expected return rate R is in the range $0 < R < \frac{1}{2}$. We observe that when T is small, the PUM strategy has the smallest SV among the three, and when T is large, the MMV strategy has the smallest SV.

6.4. Downside ratio comparisons

To quantify how much downside variance has been contained in the total variance, we define the following downside ratio:

$$\theta = \frac{SV(X(T))}{\text{Var}(X(T))}. \quad (6.11)$$

It is obvious that the smaller this ratio is the more upside gains will be in the total variance. Therefore, $1 - \theta$ provides a measure about how much upside gains are included in the

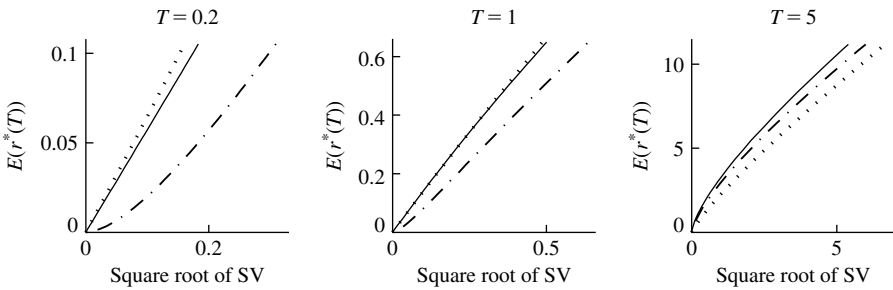


FIG. 6.2 Portfolio frontier: expected excess return versus standard semideviation for different time horizon T . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

variance. All the comparisons about the performance of three strategies are made under the same expected return rate R .

We now examine two special cases: short- and long-term investments, namely, $T \rightarrow 0$ and $T \rightarrow \infty$.

PROPOSITION 6.2. When $0 < R < \frac{1}{2}$,

- (i) $\lim_{T \rightarrow 0} \theta_{\text{PUM}} = \lim_{T \rightarrow 0} \theta_{\text{MMV}} = \frac{1}{2}, \quad \lim_{T \rightarrow 0} \theta_{\text{SPM}} = 1$
- (ii) $\lim_{T \rightarrow \infty} \theta_{\text{PUM}} = 0, \quad \lim_{T \rightarrow \infty} \theta_{\text{MMV}} = \lim_{T \rightarrow \infty} \theta_{\text{SPM}} = 1.$

PROOF. See Appendix B. □

Proposition 6.2 shows that for a short-term investment horizon, that is, $T \rightarrow 0$, 50% of the variance is due to downside variance in the MMV and SPM strategies, but the SPM have 100% downside variance. For long-term investment horizon, that is, $T \rightarrow \infty$, the SPM and MMV strategies have the same limit of 1 for the downside ratio θ , which means that the variance is almost surely due to the downside variance. In contrast, for the PUM strategy, the variance almost surely belong to the upside gains. This point is clearly illustrated in Fig. 6.3, where the final wealth distributions for the three strategies are plotted.

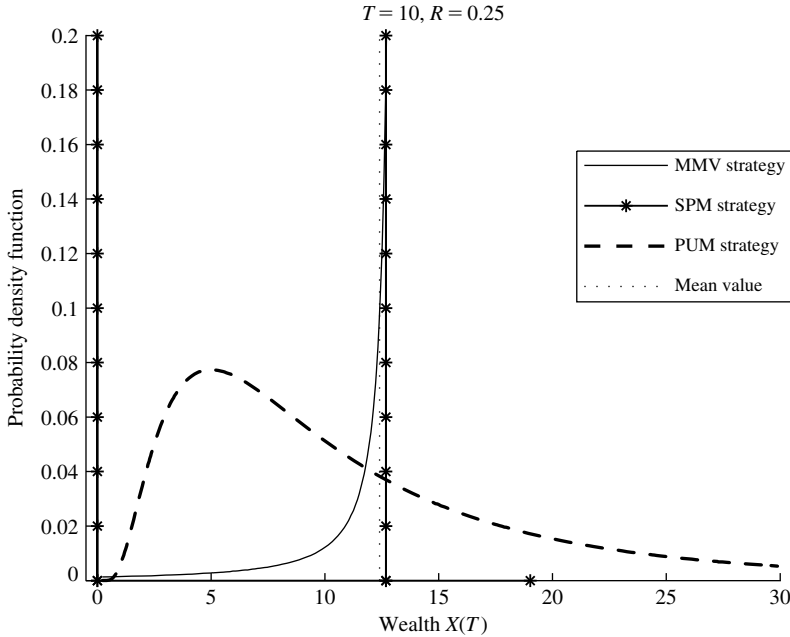


Fig. 6.3 Final wealth distributions for the MMV, SPM, and PUM strategies.

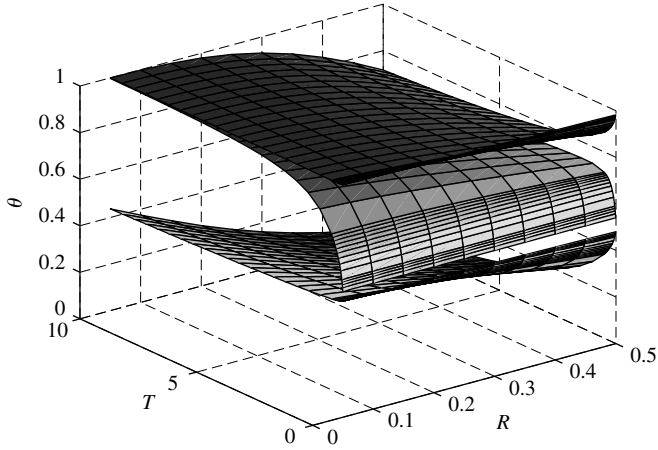


FIG. 6.4 A comparison of downside ratios among the MMV, SPM, and PUM strategies at different time horizon T and expected return rate R . The upper, middle, and lower surfaces stand for downside ratios of the SPM, MMV, and PUM strategies, respectively. It shows that $\theta_{\text{PUM}} \leq \theta_{\text{MMV}} \leq \theta_{\text{SPM}}$ for all values of T and R .

This is due to the facts that the final wealth of these two strategies has continuous distributions and the drift of stock does not play a significant role for very short-time horizon. It follows that the probability of increasing wealth and that of decreasing wealth are equal. Unlike the MMV and PUM strategies, the SPM strategy has binomial distribution, which is discontinuous (see Eq. (3.21)). The wealth paths that end at 0, although with small probability, make main contribution to the variance and consequently lead to a high downside ratio.

Proposition 6.2 also shows that the order of θ_i is the same for both the short-term investment and the long-term investment, that is, for both $T \rightarrow 0$ and $T \rightarrow \infty$,

$$\theta_{\text{PUM}} \leq \theta_{\text{MMV}} \leq \theta_{\text{SPM}}. \quad (6.12)$$

Furthermore, Fig. 6.4 shows that the relation (6.12) holds for all values of T , that is, $0 < T < \infty$. The surface of the MMV strategy always lies between the surface of the PUM strategy and that of the SPM strategy. It coincides with surface of the SPM strategy in the limit $T \rightarrow \infty$ and with that of the PUM strategy in the limit $T \rightarrow 0$.

6.5. Best strategy

We know that the downside ratio of the PUM strategy is much lower than that of the MMV strategy and that of the SPM strategy, as shown in Fig. 6.4, but one cannot conclude from this that the PUM strategy outperforms the MMV and SPM strategies because the downside ratio θ does not contain the information about the size of the downside variance.

Then comes a natural question: for investors who adopt SV as a risk measure, which dynamic strategy they should follow? The answer to this question is presented in Fig. 6.5

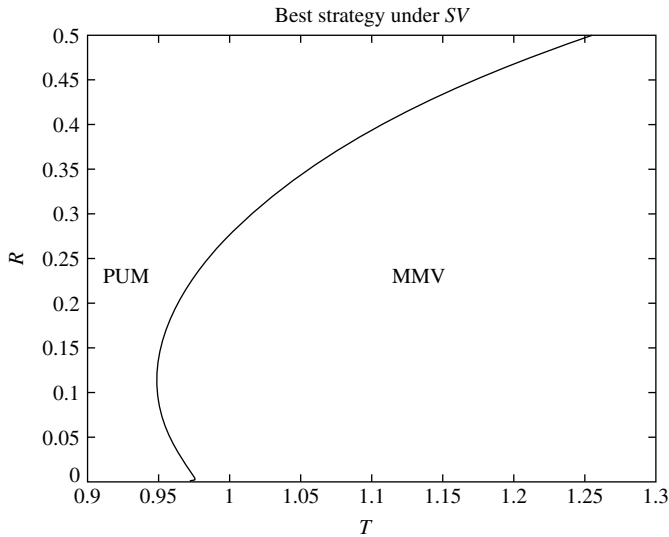


FIG. 6.5 The domain of dominant strategy under SV risk measure in parameter spaces R and T . The dominant strategy is labeled in each domain. The PUM/MMV strategy dominates the short/large investment horizon.

as a function of dimensionless expected return rate R and dimensionless investment horizon T .

In Fig. 6.5, we plot the phase boundary between the domains to show which strategy among the MMV, SPM, and PUM strategies performs best under the SV risk measure: the PUM strategy dominates in the short-term investment, for example, from $T = 0$ to the phase boundary, and the MMV strategy dominates in the long-term investment, for example, from the phase boundary to $T = \infty$. It should be noted that though the MMV strategy has larger downside ratio than the PUM strategy, it still performs best in most regions under the downside variance. This is due to the fact that the MMV strategy is the optimal solution for the mean-variance model and therefore has the smallest variance, which leads to small downside variance even when downside ratio is large. We also observed that the SPM strategy never appears in Fig. 6.5, namely, one should not choose the SPM strategy for investment under SV risk measure. This is because that, relative to the MMV strategy, the SPM strategy has both larger variance and larger downside ratio, which surely lead to larger downside variance. Similarly, it is also easy to verify that in the domain where the PUM strategy dominates, the PUM strategy contains not only less downside variance but also more upside gains in comparison with the MMV strategy. This is because relative to the MMV strategy, the PUM strategy has larger variance but smaller downside ratio, which leads to larger upside variance. It should be noted that Fig. 6.5 is consistent with Fig. 6.2. When T is small, for example, $T = 0.2$, the PUM strategy outperforms the other strategies. When T is large, the MMV strategy outperforms the other strategies. $T = 1$ is near the phase boundary between the PUM strategy and the MMV strategy, therefore the portfolio frontiers of the PUM and MMV strategies are close to each other.

We now provide an example illustrating the use of Fig. 6.5 for selecting the portfolio strategies. Considering the dimensional parameters, interest rate $r = 5\%$, drift of stock $\tilde{\mu} = 15\%$, and volatility of stock $\tilde{\sigma} = 20\%$, if an investor wants to obtain the expected return rate $\tilde{R} = 10\%$ (or $\tilde{R} + r = 15\%$ for undiscounted wealth) for an investment with time horizon $\tilde{T} = 3$ years, then the dimensionless parameters can be computed according to Eqs. (3.4) and (3.5) as $R = 0.4$ and $T = 0.75$, which give a point located in the domain where the PUM strategy dominates. Therefore, the investor will follow the PUM strategy for the investment if he/she adopts the SV as risk measure. If the investor wants to change the investment horizon to $\tilde{T} = 5$ years, with all other parameters remaining the same, the corresponding dimensionless parameters are $R = 0.4$ and $T = 1.25$, which give a point located in the domain where the MMV strategy dominates. Therefore, the investor will follow the MMV strategy in this case.

7. VaR

In this section, we study the performance of three strategies under a popular downside risk measure: VaR. The definition of VaR risk measure is given by Eq. (4.2).

7.1. Analytical formula

VaR risk measure can be determined analytically for the three strategies.

1. MMV strategy

$$\text{VaR}(X(T)) = \begin{cases} e^{R_1 T} & \text{if } \alpha \leq 1 - \Phi_1 \\ e^{R_1 T} - \omega_1 + \omega_2 e^{\sqrt{T} \Phi^{-1}(1-\alpha) - \frac{1}{2}T} & \text{if } \alpha > 1 - \Phi_1 \end{cases}, \quad (7.1)$$

where

$$\Phi_1 := \Phi\left(\frac{1}{\sqrt{T}}\left(\ln\left(\frac{\omega_1}{\omega_2}\right) + \frac{1}{2}T\right)\right). \quad (7.2)$$

2. SPM strategy

$$\text{VaR}(X(T)) = \begin{cases} xe^{R'_2 T} \Phi_2 & \text{if } \alpha \leq 1 - \Phi_2 \\ 0 & \text{if } \alpha > 1 - \Phi_2 \end{cases}, \quad (7.3)$$

where

$$\Phi_2 := \Phi\left(\sqrt{T} + \Phi^{-1}\left(e^{-R'_2 T}\right)\right). \quad (7.4)$$

Here, we redefine the value of VaR equal to 0 when $\alpha > 1 - \Phi_2$. This is due to the fact that the original value of VaR calculated according to Eq. (4.2) is negative, that is, $xe^{R'_2 T}(\Phi_2 - 1) < 0$ when $\alpha > 1 - \Phi_2$. We comment that the redefinition here is reasonable because when $\alpha > 1 - \Phi_2$, the upper level of the α worst case of the portfolio is located at $X(T) = xe^{R'_2 T}$, which is the highest wealth level and

larger than the mean value $E(X(T)) = xe^{R_2T}$ (see Eq. (3.19)), and no risk under VaR will be involved in this case.

3. PUM strategy

$$\text{VaR}(X(T)) = x \exp(R_3T) - x \exp\left(R_3\sqrt{T} \Phi^{-1}(\alpha) + \left(R_3 - \frac{R_3^2}{2}\right)T\right). \quad (7.5)$$

7.2. Long-term asymptotic behavior

We rewrite the definition of VaR given in Eq. (4.2) as follows:

$$\text{VaR}(X) = E(X) \left(1 - \frac{Q_\alpha(X)}{E(X)}\right) := \kappa E(X), \quad (7.6)$$

where κ stands for the risk-reward ratio of $\text{VaR}(X)/E(X)$. We study asymptotic behavior of κ instead of VaR.

Then, it is straightforward to obtain the following proposition after some algebraic calculations.

PROPOSITION 7.1. *When $0 < R < \frac{1}{2}$,*

$$\lim_{T \rightarrow \infty} \kappa_{\text{MMV}}(T) = \lim_{T \rightarrow \infty} \kappa_{\text{SPM}}(T) = 0, \quad \lim_{T \rightarrow \infty} \kappa_{\text{PUM}}(T) = 1.$$

Proposition 7.1 shows that in the long-term investment, $\text{VaR}(X(T))/E(X(T))$ ratio is very small for the MMV and SPM strategies, while for the PUM strategy, VaR risk measure increases at the same exponential rate as the mean value of final wealth. It should be noted that for the MMV and SPM strategies, even the ratio $\kappa(T)$ decreases for large value of T , the VaR risk measure could still increase due to the exponential-increasing term $E(X(T)) = e^{RT}$.

Fig. 7.1 shows the ratio κ of the MMV, SPM, and PUM strategies as a function of investment time horizon T for $R = 0.1, 0.25$, and 0.4 respectively. It is clear that for

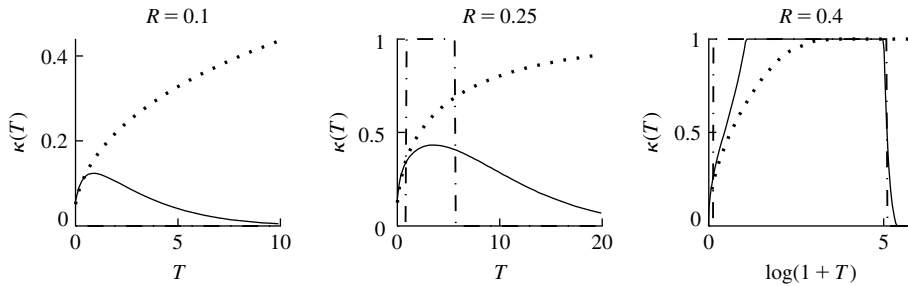


FIG. 7.1 κ as a function of T for different values of R . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively. $\log(1+T)$ has been used instead of T when $R = 0.4$.

the PUM strategy, $\kappa(T)$ is an increasing function of T with limit value of 1 regardless of what R is. However, for the MMV and SPM strategies, κ is an increasing function of T for small value of T and a decreasing function of T for large value of T . Obviously, when applying the MMV and SPM strategies, one should avoid choosing investment horizon near the critical T_c , where κ attains the maximum.

Then, one question arise: How to get the critical time horizon?

For the MMV strategy, the critical time horizon T_c can be solved in the following equations:

$$\frac{\partial \kappa_{\text{MMV}}(T)}{\partial T} = 0 \quad \text{when } \alpha > 1 - \Phi_1 \quad \text{for } \forall T > 0 \quad (7.7)$$

$$\alpha = 1 - \Phi_1 \quad \text{when } \alpha \leq 1 - \Phi_1 \quad \text{if } \exists T > 0. \quad (7.8)$$

When Eq. (7.7) holds, the maximum value of κ is unique and less than 1 (see the first and second panels in Fig. 7.1). When Eq. (7.8) holds, the maximum value of κ is 1, and there is a range of time horizon in which $\kappa_{\text{MMV}} = 1$ if Eq. (7.8) admits two solutions (see the third panel in Fig. 7.1).

We note here that Eq. (7.8) admits at least one solution. This is due to the fact (see proof of Theorem 1 in YU, ZHANG and YANG [2006])

$$\lim_{T \rightarrow 0} \Phi_1 = \lim_{T \rightarrow \infty} \Phi_1 = 1. \quad (7.9)$$

Therefore, $\alpha > 1 - \Phi_1$ as $T \rightarrow 0$ and $T \rightarrow \infty$. If $\exists T > 0$ such that $\alpha < 1 - \Phi_1$, then by Weierstrass intermediate value theorem, there exist two solutions: one between 0 and T , and the other between T and ∞ . Furthermore, considering the quantitative behavior of Φ_1 as a function of T , increasing function for small value of T and a decreasing function for large value of T , *there exist and only exist two solutions*. If $\exists T > 0$ such that $\alpha = 1 - \Phi_1$, T is the exact unique solution.

For the SPM strategy, the critical time horizon T_c can be solved in the following equation:

$$\alpha = 1 - \Phi_2. \quad (7.10)$$

Since Φ_2 as a function of T has same quantitative behavior as Φ_1 , Eq. (7.10) will admit two solutions (see the second and third panels in Fig. 7.1), one solution, or no solution (see the first panel of Fig. 7.1, where $\kappa = 0$ for $\forall T > 0$).

7.3. Portfolio frontier

To study the portfolio frontier under VaR risk measure, we plot Fig. 7.2 to show the expected excess return versus VaR for all three strategies. Notice in Eq. (6.10), $E(r^*(T))$ is an increasing function of expected return rate R for a given time horizon T . Therefore, we observe from Fig. 7.2 that when expected return rate R is small, the SPM strategy outperforms other strategies, and when R is large, the PUM strategy outperforms other strategies.

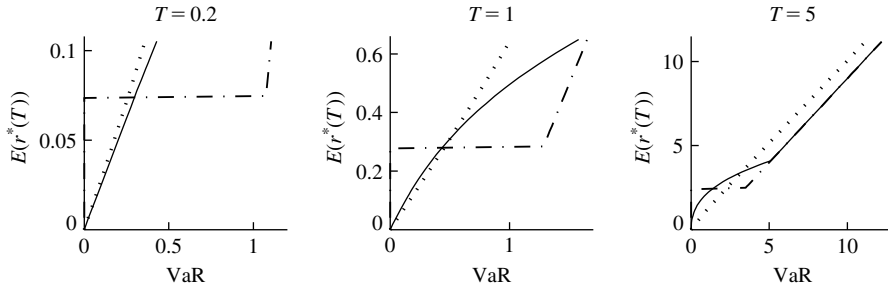


FIG. 7.2 Portfolio frontier: expected excess return versus VaR for different time horizon T . Solid and dash-dotted lines are results for the MMV and SPM strategies, respectively.

We comment that in Fig. 7.2 the dash-dotted curves for the SPM strategy stay at zero initially and then jump to positive values at certain value:

$$R_2^b = -\frac{1}{T} \ln \Phi(\Phi^{-1}(1 - \alpha) - \sqrt{T}) + \frac{1}{T} \ln(1 - \alpha), \quad (7.11)$$

which satisfies the following system of equations:

$$\begin{cases} \alpha = 1 - \Phi(\sqrt{T} + \Phi^{-1}(e^{-R_2^b T})) \\ e^{R_2^b T} = e^{R_2^b T} \Phi(\sqrt{T} + \Phi^{-1}(e^{-R_2^b T})) \end{cases}, \quad (7.12)$$

where the first equation in (7.12) comes from Eqs. (7.1) and (7.2), and the second equation in (7.12) comes from Eq. (3.19). It will be mentioned later that R_2^b also plays an important role in comparison of the best strategy under VaR and CVaR in the asymptotic situation for large T .

7.4. Best strategy

We plot Fig. 7.3 to show which strategy among the MMV, SPM, and PUM strategies performs best under the VaR risk measure in different parameter regions. In order to describe the long-term situation, we use $\log(1 + T)$ scale instead of T in Fig. 7.3. Further illustration and discussion about Fig. 7.3 are presented in next section when comparing with another important downside risk measure: CVaR. Obviously, results in Figs. 7.2 and 7.3 are consistent.

8. Conditional VaR

As noted by ARTZNER, DELBAEN, EBER and HEATH [1999], VaR is not a coherent measure of risk because it fails to be subadditive. In this section, we will carry out the same study for a coherent downside risk measure: CVaR, which is often proposed as an alternative for VaR. The definition of CVaR is given by Eq. (4.4).

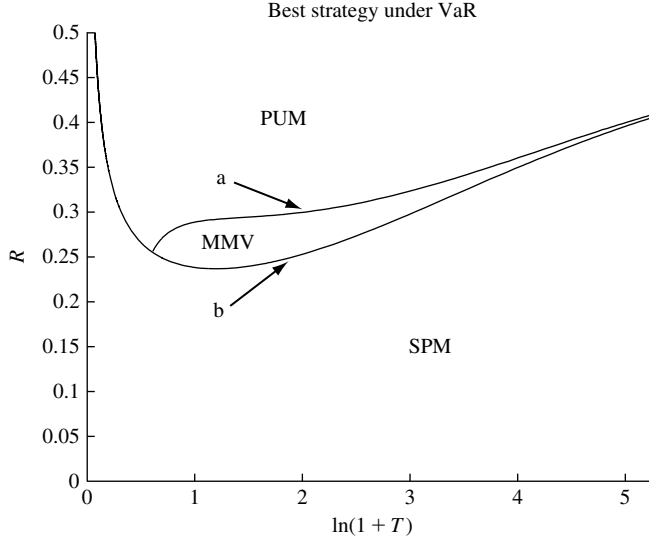


FIG. 7.3 The domain of dominant strategy under VaR in parameter space R and $\log(1+T)$. The dominant strategy is labelled in each domain, and phase boundaries between the MMV strategy and the other two strategies are shown as curves a and b.

8.1. Analytical formula

CVaR risk measure can be determined analytically for the three strategies.

1. MMV strategy

$$\begin{aligned} & \text{CVaR}(X(T)) \\ &= \begin{cases} xe^{R_1 T} & \text{if } \alpha \leq 1 - \Phi_1 \\ xe^{R_1 T} - \frac{\omega_1}{\alpha} \left[\Phi \left(\frac{\ln(\omega_1/\omega_2) + \frac{1}{2}T}{\sqrt{T}} \right) - \Phi \left(\frac{\sqrt{T} \Phi^{-1}(1-\alpha)}{\sqrt{T}} \right) \right] & \text{if } \alpha > 1 - \Phi_1, \end{cases} \end{aligned} \quad (8.1)$$

where

$$\Phi_1 = \Phi \left(\frac{1}{\sqrt{T}} \left(\ln \left(\frac{\omega_1}{\omega_2} \right) + \frac{1}{2}T \right) \right). \quad (8.2)$$

2. SPM strategy

$$\text{CVaR}(X(T)) = \begin{cases} xe^{R'_2 T} \Phi_2 & \text{if } \alpha \leq 1 - \Phi_2 \\ \frac{1-\alpha}{\alpha} xe^{R'_2 T} (1 - \Phi_2) & \text{if } \alpha > 1 - \Phi_2, \end{cases} \quad (8.3)$$

where

$$\Phi_2 = \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R'_2 T} \right) \right). \quad (8.4)$$

3. PUM strategy

$$\text{CVaR}(X(T)) = xe^{R_3 T} \left(1 - \frac{1}{\alpha} \Phi \left(\Phi^{-1}(\alpha) - R_3 \sqrt{T} \right) \right). \quad (8.5)$$

8.2. Long-term asymptotic behavior

As we did for VaR risk measure, we rewrite the definition of CVaR given in Eq. (4.4) as follows:

$$\text{CVaR}(X) = E(X) \left(1 - \frac{C_\alpha(X)}{E(X)} \right) := \nu E(X), \quad (8.6)$$

where ν stands for the risk-reward ratio of $\text{CVaR}(X)/E(X)$. We study asymptotic behavior of ν instead of CVaR.

We can obtain the following proposition after some algebraic calculations.

PROPOSITION 8.1. *When $0 < R < \frac{1}{2}$,*

$$\lim_{T \rightarrow \infty} \nu_{\text{MMV}}(T) = \lim_{T \rightarrow \infty} \nu_{\text{SPM}}(T) = 0, \quad \lim_{T \rightarrow \infty} \nu_{\text{PUM}}(T) = 1. \quad (8.7)$$

Proposition 8.1 describes same things as Proposition 7.1 but in a different framework under CVaR risk measure. Therefore, discussions similar to that for Fig. 7.1 can be repeated for Fig. 8.1.

It should be noted that the quantitative behavior of ratio ν_{SPM} is different from κ_{SPM} , which can take only two values. The equations for solving the critical time horizon for ν_{SPM} are as follows:

$$\frac{\partial \nu_{\text{SPM}}(T)}{\partial T} = 0 \quad \text{when } \alpha > 1 - \Phi_2 \text{ for } \forall T > 0 \quad (8.8)$$

$$\alpha = 1 - \Phi_2 \quad \text{when } \alpha \leq 1 - \Phi_2 \text{ if } \exists T > 0. \quad (8.9)$$

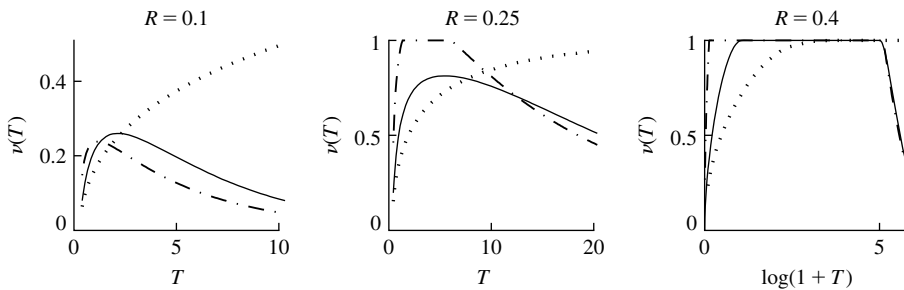


FIG. 8.1 ν as a function of T for different values of R . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively. $\log(1+T)$ has been used instead of T when $R = 0.4$.

8.3. Portfolio frontier

To study the portfolio frontier under CVaR risk measure, we plot Fig. 8.2 to show the expected excess return versus CVaR for all three strategies.

Figure 8.2 shows that when expected return rate R is small, the SPM strategy outperforms other strategies, and when R is large, the PUM strategy outperforms other strategies, which is the same conclusion for VaR risk measure. However, there do exist difference between VaR and CVaR risk measures when T is small. We discuss this issue in next section.

8.4. Best strategy: VaR versus CVaR

We plot Fig. 8.3 to show which strategy performs best under the risk measure CVaR. When comparing Figs. 7.3 and 8.3, we observe that the results are similar for both VaR and CVaR risk measures in the long-term investment but very different in the short-term investment. For VaR, the SPM-dominated domain is larger than the PUM-dominated domain, and for CVaR, the PUM-dominated domain is larger than the SPM-dominated domain. This phenomenon is consistent with the results in Figs. 7.2 and 8.2. When T is small (e.g., $T = 0.2$) and large (e.g., $T = 5$), the SPM or PUM strategy dominates more expected return rate under VaR and CVaR risk measures, respectively.

For comparing shapes of the dominated domains under VaR and CVaR risk measures, we plot Fig. 8.4, which shows that there exists one domain bounded by the curves a and d in which the MMV strategy dominates for both VaR and CVaR risk measures. The domain bounded by the curves c and b becomes increasing by narrower as T increases and eventually disappears as $T \rightarrow \infty$. Therefore, when $T \gg 1$, there is only one phase boundary, which separates the SPM-dominated domain and the PUM-dominated domain for both VaR and CVaR risk measures. This phase boundary is located at

$$R_b(T) = -\frac{1}{T} \ln \Phi \left(\Phi^{-1}(1 - \alpha) - \sqrt{T} \right) + \frac{1}{T} \ln(1 - \alpha), \quad (8.10)$$

which is exactly the value of R_2^b given by Eq. (7.11). Long-term investors will follow the SPM strategy when $R < R_b(T)$ and the PUM strategy when $R > R_b(T)$ for both

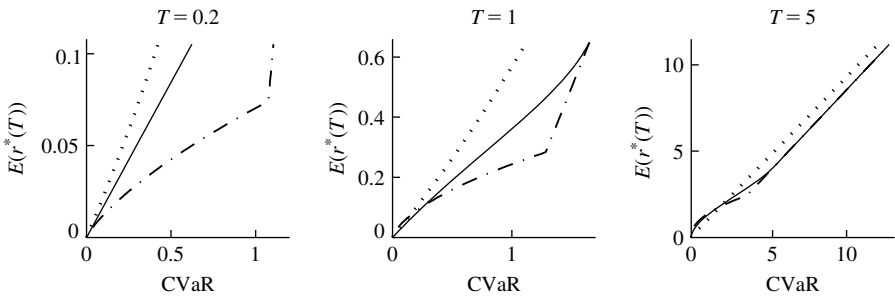


FIG. 8.2 Portfolio frontier: expected excess return vs CVaR for different time horizon T . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

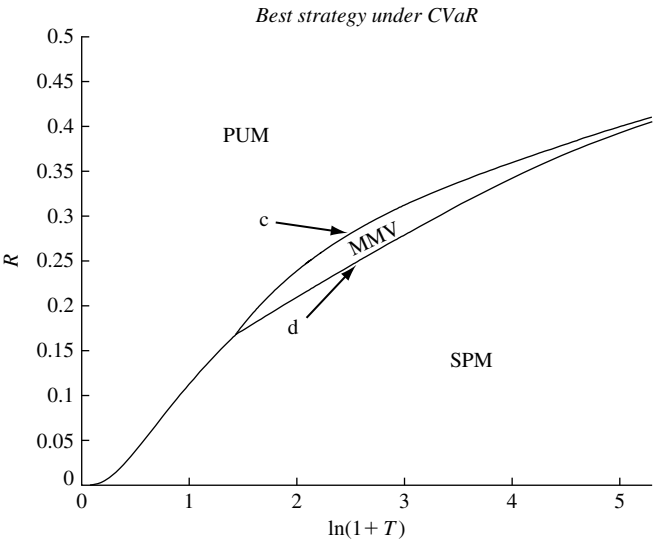


FIG. 8.3 The domain of dominant strategy under CVaR in parameter space R and $\log(1+T)$. The dominant strategy is labelled in each domain, and phase boundaries between the MMV strategy and the other two strategies are shown as curves c and d .

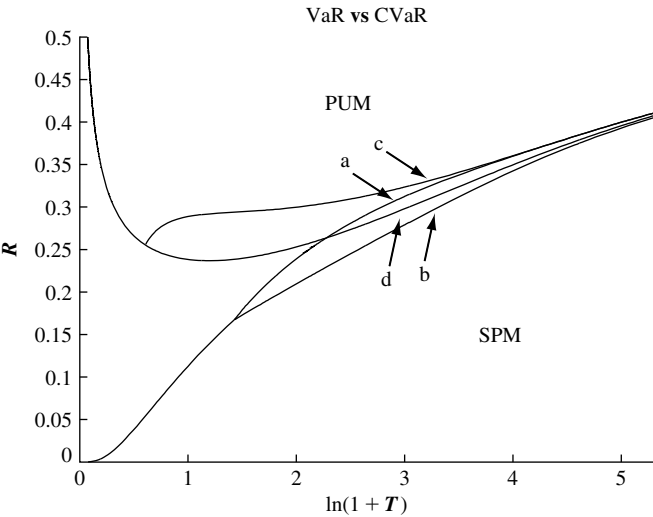


FIG. 8.4 The combined domain of dominant strategy under both VaR and CVaR in parameter space R and $\log(1+T)$.

VaR and CVaR risk measures. It is also easy to show that the limit value of $R_b(T)$ is $\lim_{T \rightarrow \infty} R_b(T) = \frac{1}{2}$.

YU, ZHANG and YANG [2006] show that the MMV and SPM strategies are equivalent in the limit $T \rightarrow \infty$; why the SPM strategy still outperforms the MMV strategy for large T is due to the following fact: relative to the MMV strategy, the SPM strategy was designed to minimize the probability of the portfolio value falling below a specified wealth level, which is closely related to the definition of VaR or CVaR.

9. Average drawdown

In this section, we study the performance of the MMV, SPM, and PUM strategies under an important drawdown risk measures: ADD, which is given by Eq. (4.8). Unlike the downside risk measures, which only relate to final distribution of the portfolio value, drawdown risk measures relate to the whole path of the portfolio wealth and hence more difficult to obtain closed-form expressions. We derived the analytical expressions of ADD risk measure for the PUM strategy, but the analytical expressions of ADD risk measure for the MMV and SPM strategies so far are not available; therefore, we will solve them in the Monte Carlo simulation framework.

9.1. Long-term asymptotic behavior

The asymptotic behavior under $T \rightarrow \infty$ of ADD for the PUM strategy has been studied by YANG [2006], which shows long-term Add as follows:

$$\lim_{T \rightarrow \infty} \text{ADD}(T) = \frac{R}{2} \quad (R \leq 2). \quad (9.1)$$

Let M be the number of sample paths simulated and N be the number of discretization points per sample path, we take $M = 25\,000$ and $N = 500\,000$ in our Monte Carlo simulation experiments. The large number of N is to provide a better discretely measured maximum value, that is, $\max_{0 \leq t \leq T} X(t)$.

Fig. 9.1 shows the values of the ADD risk measure for three strategies. For the PUM strategy, we can observe that the curve is very close to $\text{ADD} = \frac{R}{2}$ for different values of R when T is large. For the MMV and SPM strategies, the ADD risk measure also has a limit value when $T \rightarrow \infty$.

The analytic expression of ADD for the PUM strategy is

$$\begin{aligned} \text{ADD}(T) = 1 - \frac{1}{T} \frac{2a}{b^2 + 2a} \left[T - \frac{b^2}{a^2} + \frac{2(b^2 + a)^2}{a^2(b^2 + 2a)} \right] \Phi\left(\frac{a}{b}\sqrt{T}\right) + \frac{1}{aT} \\ - \frac{4}{T} \frac{b^2 + a}{(b^2 + 2a)^2} e^{\frac{T}{2}(b^2 + 2a)} \Phi\left(-\frac{b^2 + a}{b}\sqrt{T}\right) - \frac{2be^{-\frac{a^2}{2b^2}T}}{\sqrt{2\pi T}(b^2 + 2a)}, \end{aligned} \quad (9.2)$$

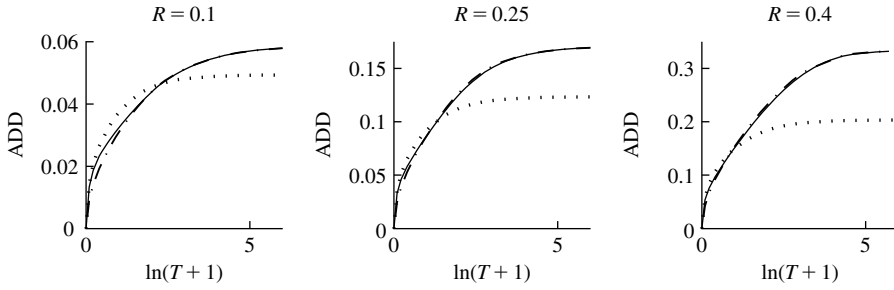


FIG. 9.1 Long-term asymptotic behavior of ADD risk measure for three strategies when $R = 0.1, 0.25$, and 0.4 , respectively. The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

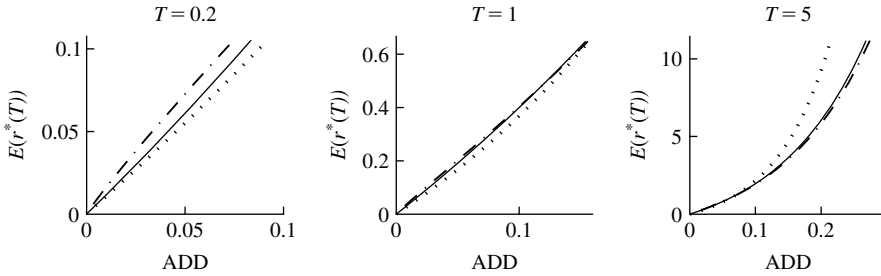


FIG. 9.2 Portfolio frontier: expected excess return versus average-percent drawdown for different time horizon T . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

where $a = R_3 - \frac{R_3^2}{2}$ and $b = R_3$. The corresponding derivation can be found in Appendix A.

9.2. Portfolio frontier

Fig. 9.2 shows the portfolio frontier, expected excess return versus ADD risk measure for all three strategies: when T is small, the SPM strategy has the smallest average drawdown, and when T is large, the PUM strategy has the smallest average drawdown.

9.3. Best strategy

Fig. 9.3 shows which strategy should be adopted in different parameter regions for the ADD risk measures: the investor will choose the SPM strategy when either R or T is small, and the PUM strategy where either R or T is large. The domain in which the MMV strategy dominates is sandwiched between the SPM-dominated and PUM-dominated domains.

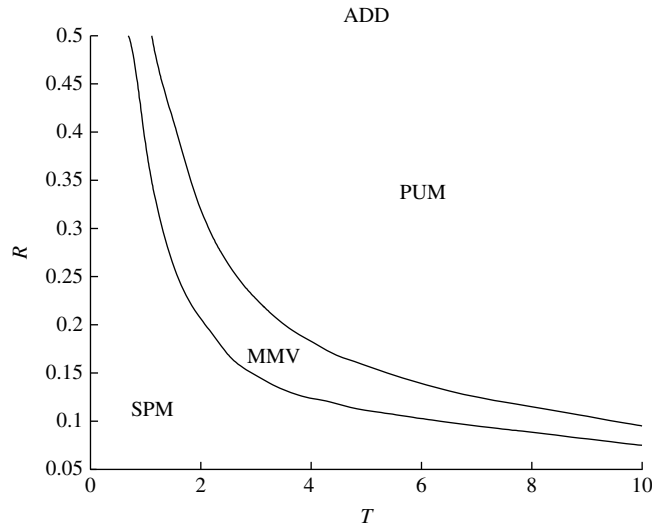


FIG. 9.3 The domain of dominant strategy under average drawdown in parameter spaces R and T . The dominant strategy is labelled in each domain, and phase boundaries between the domains are also shown.

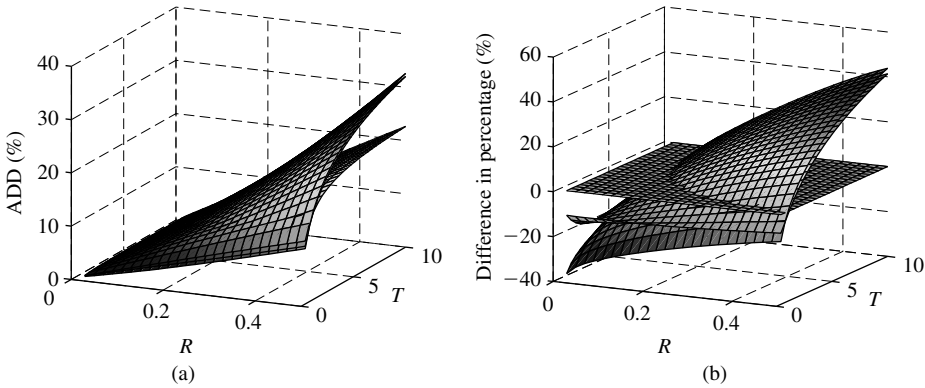


FIG. 9.4 Average drawdown comparisons in parameter spaces R and T .

However, Fig. 9.3 does not tell us how large the drawdown is, which is important information for investors. Therefore, we plot Fig. 9.4(a) to show how the values of drawdown risk measures vary in different parameter regions for three strategies. The relative drawdown difference is also presented in Fig. 9.4(b) when considering the PUM strategy as a benchmark.

We observe that large T and R lead to high drawdown risk measures and small T and R lead to low drawdown risk measures, which fit the intuitions of investors. For

the ADD risk measure, the larger T or R is, the more superiority the PUM strategy has. The MMV and SPM strategies has up to 40% higher average drawdown than the PUM strategy, as shown in Fig. 9.4(b). As T or R is small, the ADD risk measure is small for all three strategies, and the MMV and SPM strategies has -20% and -40% less average drawdown, respectively, compared to the PUM strategy.

10. Maximum drawdown

In this section, we study the performance of the MMV, SPM, and PUM strategies under another popular drawdown risk measures: MDD, which is given by Eq. (4.9).

10.1. Long-term asymptotic behavior

MAGDON-ISMAIL, ATIYA, PRATAP and ABU-MOSTAFA [2004] have studied and derived the expected absolute maximum drawdown for a Brownian motion with drift and its corresponding long-term asymptotic behavior. By taking a log transformation to the wealth portfolio of the PUM strategy,

$$d \ln X(t) = R_3(1 - \frac{R_3}{2})dt + R_3dW(t), \quad (10.1)$$

the expectation of a variable related to MDD can be obtained analytically, that is, $E(\ln(1 - \text{Mdd}(T)))$. However, so far we are unable to obtain analytical expression of MDD for the PUM strategy.

We also plot Fig. 10.1 to show the values of MDD risk measure for long-term horizon. Unlike the ADD risk measure, we observe that the MDD risk measure tends to 1 as T is large enough, for example, the MDD is 100% in long-term investment for all three strategies. It is because the maximum percent drawdown is memoryless, for example, it will always be updated with the recent drawdown, which is larger than before; therefore in certain sense, MDD risk measure is not well defined in long-term horizon.

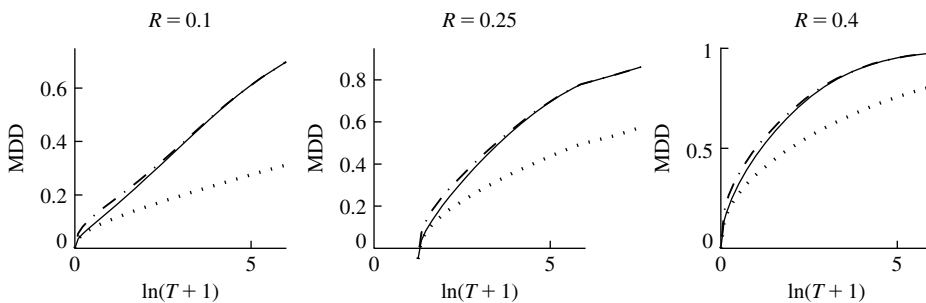


FIG. 10.1 Long-term asymptotic behavior of MDD risk measure for three strategies. The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

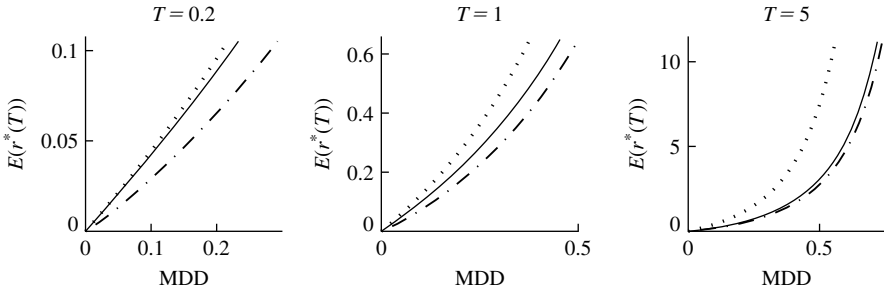


FIG. 10.2 Portfolio frontier: expected excess return vs MDD for different time horizon T . The solid, dash-dotted, and dotted lines are results for the MMV, SPM, and PUM strategies, respectively.

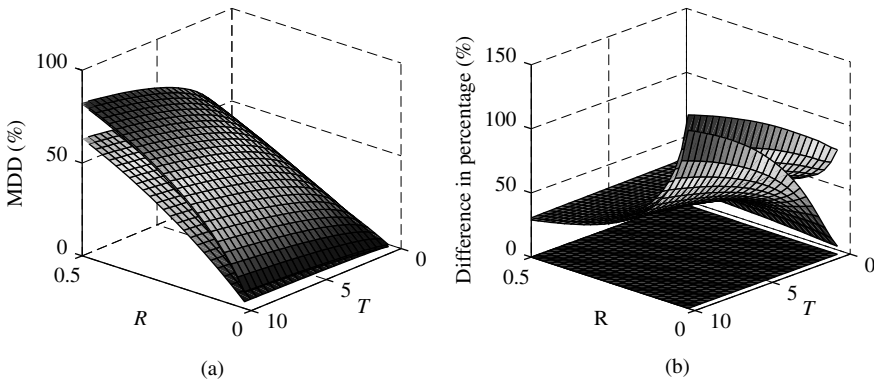


FIG. 10.3 Maximum drawdown comparisons in parameter spaces R and T .

10.2. Portfolio frontier and best strategy

The comparison of the portfolio frontier in Fig. 10.2 shows that the PUM strategy always has the smallest drawdown for both small and large T . In fact, for the MDD risk measure, the whole domain in parameter space R and T is entirely dominated by the PUM strategy. Therefore, the investor will definitely follow the PUM strategy for the investment if he/she adopts the maximum drawdown as a risk measure. Consequently, we do not give the corresponding domain of dominant strategy under maximum drawdown.

We plot Fig. 10.3 to show the maximum drawdown values for three strategies and relative drawdown difference when considering the PUM strategy as a benchmark. For the MDD risk measures, the PUM strategy always dominates the other two strategies, as shown in Fig. 10.3(a). However, the largest relative difference between the PUM strategy and the MMV and SPM strategies does not happen at both large T and R , but at large T and small R , an relative difference up to 150% has been observed in Fig. 10.3(b).

11. Correlations between different risk measures

In the above sections, we have examined the performances of the MMV, SPM, and PUM strategies for a given risk measure. In this section, we will take a very different view: studying correlations among different risk measures for a given portfolio strategy. This is an interesting topic and we will carry out such study in the Monte Carlo simulation framework.

We will consider four risk measures: Var, SV, ADD, and MDD. Each can be expressed as an expectation of a corresponding random variable measurable at the end of investment horizon T . These corresponding random variables are

$$\text{for Var: } X_1 = (X(T) - \bar{x})^2, \quad (11.1)$$

$$\text{for SV: } X_2 = \max((X(T) - \bar{x}), 0)^2, \quad (11.2)$$

$$\text{for ADD: } X_3 = \frac{1}{T} \int_0^T \frac{X_m(t) - X(t)}{X_m(t)} dt, \text{ and} \quad (11.3)$$

$$\text{for MDD: } X_4 = \sup_{0 \leq t \leq T} \frac{X_m(t) - X(t)}{X_m(t)}, \quad (11.4)$$

where

$$\bar{x} = E(X(T)) \quad \text{and} \quad X_m(t) = \sup_{0 \leq s \leq t} X(s). \quad (11.5)$$

Taking expectation of X_1, X_2, X_3 , and X_4 gives the values of Var, SV, ADD, and MDD risk measures, respectively.

The correlation ρ between two random variables X_i and X_j is defined as

$$\rho := \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i) \text{Var}(X_j)}}. \quad (11.6)$$

The value of X_i ($1 \leq i \leq 4$) is the realization of the corresponding risk measure in each sample path generated in the numerical experiment.

We do not consider the VaR and CVaR risk measures in this correlation study since they are related to α -quantile of the final distribution, not the expectation of a corresponding random variable at end of investment horizon.

Table 11.1 shows that for the MMV and SPM strategies, usually the largest correlations exists in the following pairs: between SV and Var and between ADD and MDD. The large correlation phenomenon between two drawdown risk measures ADD and MDD is expected: the risk measures belonging to the same risk type are more closely related than the risk measures of different types. Although Var does not belong to the downside risk measures, we observe a perfect correlation between Var and SV risk measures. This is consistent with the discussions in Section 6, namely, Var contains much more downside variance than upside gains and is almost surely due to the downside variance in long-term investment.

TABLE 11.1
Correlation situation. The correlations above and below the diagonal of the matrix are corresponding to $T = 1$ and $T = 10$, respectively

For the MMV strategy				
	Var	SV	ADD	MDD
Var	1.000	0.998	0.681	0.634
SV	1.000	1.000	0.703	0.658
ADD	0.625	0.629	1.000	0.930
MDD	0.438	0.442	0.847	1.000
For the SPM strategy				
	Var	SV	ADD	MDD
Var	1.000	1.000	0.680	0.660
SV	1.000	1.000	0.680	0.660
ADD	0.611	0.611	1.000	0.887
MDD	0.451	0.451	0.844	1.000
For the PUM strategy				
	Var	SV	ADD	MDD
Var	1.000	0.221	0.006	-0.029
SV	-0.075	1.000	0.803	0.720
ADD	-0.130	0.752	1.000	0.895
MDD	-0.124	0.559	0.832	1.000

Tables 11.1 also shows that for the PUM strategy, the largest correlations always exists between ADD and MDD, for same reason given for the MMV and SPM strategies. Unlike the MMV and SPM strategies, we observe a small correlation between Var and all other three risk measures, and when T is large, the Var risk measure seems not correlated or even negatively correlated with SV, ADD, and MDD risk measures. This phenomenon is also consistent with the discussions given in Section 6: Var contains much more upside variance than downside variance and is almost surely due to the upside gains in long-term investment. This also shows that the variance is not a good risk measure for the PUM strategy. Another noticeable feature in Table 11.1 is that the correlations between two drawdown risk measures decrease as T increases. This is due to the fact that the longer the time horizon is, the more information will be contained in ADD and MDD risk measures, and this increases the difference between them since they have different definitions.

A further investigation on the correlation among different risk measures could be carried out in order to get a more general conclusion.

12. Conclusions

In the past several decades, a variety of risk measures have been proposed in the literature and most of them have been studied thoroughly and led to various one-period

optimal mean-risk strategies. However, recent studies show that in continuous-time setting, many important risk measures such as downside risks and drawdown risks do not admit or hard to find optimal strategies under the mean-risk framework. So, it is not clear how an investor should do when considering some popular and widely adopted risk measures in continuous time. In this chapter, we consider three well-known optimal dynamic strategies and examine in detail their risk characteristics for long-term investments and their corresponding portfolio frontiers under three downside risk measures (below-mean SV, VaR, and CVaR), as well as two drawdown risk measures (average drawdown and maximum drawdown). We determine for a given downside or drawdown risk measure, which strategy among the three performs best under various conditions: drift and volatility of the stock movement, risk-free interest rate, expected return rate, and investment horizon. An investigation on the correlation among different risk measures has also been carried out.

References

- ARTZNER, P., DELBAEN, F., EBER, J.-M., HEATH, D. (1999). Coherent measures of risk. *Math. Financ.* **9**, 203–228.
- BAILEY, B.J.R. (1981). Alternatives to hastings approximation to the inverse of the normal cumulative distribution function. *Appl. Stat.* **30**, 275–276.
- BASAK, S., SHAPIRO, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices. *Rev. Financ. Stud.* **14**, 371–405.
- BIELECKI, T.R., JIN, H.Q., PLISKA, S.R., ZHOU, X.Y. (2005). Continuous-time mean-variance portfolio selection with bankruptcy prohibition. *Math. Financ.* **15**, 213–244.
- BROWNE, S. (1999). Reaching goals by a deadline: digital options and continuous-time active portfolio management. *Adv. Appl. Probab.* **31**, 551–577.
- CAMPBELL, R., HUISMAN, R., KOEDIJK, K. (2001). Optimal portfolio selection in a value-at-risk framework. *J. Bank. Financ.* **25**, 1789–1804.
- CHEKHLOV, A., URYASEV, S.P., ZABARANKIN, M. (2003). Drawdown measure in portfolio optimization. Research Report.
- CVITANIĆ, J., KARATZAS, I. (1995). On portfolio optimization under drawdown constraints. *IMA in Mathematics and its applications* **65**, 35–45.
- DMITRASINOVIC-VIDOVIC, G., LARI-LAVASSANI, A., LI, X., WARE, A. (2003). Dynamic Portfolio Selection under Capital-at-Risk. The Mathematical and Computational Finance Laboratory, University of Calgary. Preprint.
- DOWD, K., BLAKE, D., CAIRNS, A. (2004). Long-term value at risk. *J. Risk Financ.* **5**, 52–57.
- EMMER, S., KLUPPELBERG, C., KORN, R. (2001). Optimal portfolios with bounded capital at risk. *Math. Financ.* **11**, 365–384.
- FISHBURN, P.C. (1977). Mean-risk analysis with risk associated with below-target returns. *Am. Econ. Rev.* **67**, 116–126.
- GABRIH, A., GRECKSCH, W., WUNDERLICH, R. (2005). Dynamic portfolio optimization with bounded shortfall risks. *Stoch. Anal. Appl.* **23**, 579–594.
- GAIVORONSKI, A.A., PFLUG, G. (2004). Value-at-risk in portfolio optimization: Properties and computational approach. *J. Risk* **7**, 1–31.
- GRAUER, R.R., HAKANSSON, N.H. (1993). On the use of mean-variance and quadratic approximations in implementing dynamic investment strategies: A comparison of returns and investment policies. *Manage. Sci.* **39**, 856–871.
- GROSSMAN, S.J., ZHOU, Z.Q. (1993). Optimal investment strategies for controlling drawdowns. *Math. Financ.* **3**, 241–276.
- HAKANSSON, N.H. (1971). Capital growth and the mean-variance approach to portfolio selection. *J. Financ. Quant. Anal.* **6**, 517–557.
- JARROW, R., ZHAO, F. (2006). Downside loss aversion and portfolio management. *Manage. Sci.* **52**, 558–566.
- JIN, H., YAN, J.A., ZHOU, X.Y. (2005). Continuous time mean-risk portfolio selection. *Ann. Inst. Henri Poincaré* **41**, 559–580.
- JORION, P. (1997). Value at Risk: The New Benchmark for Controlling Market Risk (Irwin, Chicago).
- KHANNA, A., KULLDORFF, M. (1999). A generalization of the mutual fund theorem. *Financ. Stoch.* **3**, 167–185.
- KROKHMAL, P., PALMQUIST, J., URYASEV, S. (2001). Portfolio optimization with conditional value-at-risk objective and constraints. *J. Risk* **4**, 43–68.

- LEMUS RODRIGUEZ, G.J. (1999). Portfolio optimization with quantile-based risk measures. Ph.D. thesis, Massachusetts Institute of Technology.
- LI, X., WU, Z.Y. (2005). Dynamic downside risk measure and optimal asset allocation. Preprint.
- MAGDON-ISMAIL, M., ATIYA, A., PRATAP, A., ABU-MOSTAFA, Y. (2004). On the maximum drawdown of a Brownian motion. *J. Appl. Probab.* **41**, 147–161.
- MARKOWITZ, H. (1952). Portfolio selection. *J. Financ.* **7**, 77–91.
- MARKOWITZ, H. (1959). Portfolio selection: Efficient Diversification of Investments (John Wiley & Sons).
- MARKOWITZ, H. (1987). Mean-Variance Analysis in Portfolio Choice and Capital Markets (Basil Blackwell).
- MERTON, R. (1971). Optimum consumption and portfolio rules in a continuous time model. *J. Econ. Theory* **3**, 373–413.
- NAWROCKI, D. (1999). A brief history of downside risk measures. *J. Invest.* **8**, 9–26.
- OGRYCZAK, W., RUSZCZYNSKI, A. (1989). On consistency of stochastic dominance and mean-semideviation model. *Math. Program.* **89**, 217–232.
- ORTOBELLI, S., RACHEV, S.T., STOYANOV, S., FABOZZI, F.J., BIGLOVA, A. (2005). The proper use of risk measures in portfolio theory. *Int. J. Theo. Appl. Financ.* **8**, 1107–1133.
- ROCKAFELLAR, R.T., URYASEV, S. (2000). Optimization of conditional value-at-risk. *J. Risk* **2**, 21–41.
- ROCKAFELLAR, R.T., URYASEV, S. (2002). Conditional value-at-risk for general loss distributions. *J. Bank. Financ.* **26**, 1443–1471.
- SORTINO, F.A., VAN DER MEER, R. (1991). Downside risk. *J. Portfolio. Manage.* **17**, 27–31.
- VON NEUMANN, J., MORGENSTERN, O. (1947). Theory of Games and Economic Behavior (Princeton University Press).
- YANG, D. (2006). Quantitative Strategies for Derivatives Trading (Atmif, New Jersey).
- YU, M.J., ZHANG, Q., YANG, D. (2006). Bankruptcy in long-term investment. *Quant. Financ.*, accepted for publication.
- ZHOU, X.Y., LI, D. (2000). Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Appl. Math. Opt.* **42**, 19–33.

Appendix A: Derivation of average drawdown

We will present the derivation for the ADD risk measure of the PUM strategy.

By introducing the symbols $a = R_3 - \frac{R_3^2}{2}$ and $b = R_3$, $X(t)$ given by Eq. (3.25) can be rewritten as

$$X(t) = xe^{\left(R_3 - \frac{R_3^2}{2}\right)t + R_3 W(t)} \quad (\text{A.1})$$

$$= xe^{b(W(t) + \frac{a}{b}t)}. \quad (\text{A.2})$$

By the Girsanov theorem, $\tilde{W}(t) = W(t) + \frac{a}{b}t$ is a Brownian motion under the probability $\tilde{P}$, where

$$\frac{d\tilde{P}}{dP} = Z(t) = e^{-\frac{a}{b}W(t) - \frac{1}{2}\frac{a^2}{b^2}t} = e^{-\frac{a}{b}\tilde{W}(t) + \frac{1}{2}\frac{a^2}{b^2}t}. \quad (\text{A.3})$$

Using the joint distribution of $(\tilde{W}(t), \inf_{0 \leq s \leq t} \tilde{W}(s))$:

$$\tilde{P}(\tilde{W}(t) \in dx, \inf_{0 \leq s \leq t} \tilde{W}(s) \in dy) = \frac{2(x-2y)}{\sqrt{2\pi t^3}} e^{-\frac{(2y-x)^2}{2t}} dx dy \quad (y \leq x, y \leq 0) \quad (\text{A.4})$$

we can calculate $E(\text{Cdd}(t))$ as follows:

$$E(\text{Cdd}(t)) = 1 - E\left(\frac{X(t)}{X_m(t)}\right) = 1 - \tilde{E}\left(\frac{\frac{X(t)}{X_m(t)}}{Z(t)}\right) \quad (\text{A.5})$$

$$= 1 - \tilde{E}\left(e^{(b+\frac{a}{b})\tilde{W}(t) - b\tilde{W}_m(t) - \frac{1}{2}\frac{a^2}{b^2}t}\right) \quad (\text{A.6})$$

$$= 1 - \frac{2a}{b^2 + 2a} \Phi\left(\frac{a}{b}\sqrt{t}\right) - \frac{2(b^2 + a)}{b^2 + 2a} e^{\frac{t}{2}(b^2 + 2a)} \Phi\left(-\frac{b^2 + a}{b}\sqrt{t}\right). \quad (\text{A.7})$$

Then, we can get

$$\begin{aligned} \text{ADD}(T) = & 1 - \frac{1}{T} \frac{2a}{b^2 + 2a} \left[T - \frac{b^2}{a^2} + \frac{2(b^2 + a)^2}{a^2(b^2 + 2a)} \right] \Phi\left(\frac{a}{b}\sqrt{T}\right) + \frac{1}{aT} \\ & - \frac{4}{T} \frac{b^2 + a}{(b^2 + 2a)^2} e^{\frac{T}{2}(b^2 + 2a)} \Phi\left(-\frac{b^2 + a}{b}\sqrt{T}\right) - \frac{2be^{-\frac{a^2}{2b^2}T}}{\sqrt{2\pi T}(b^2 + 2a)}. \end{aligned} \quad (\text{A.8})$$

Appendix B: Proofs related to SV

Proof of Proposition 6.1:

PROOF. Since the MMV strategy and the SPM strategy have the same behavior as $T \rightarrow \infty$ (see YU, ZHANG and YANG [2006]), we only analyze the SV and variance given by Eqs. (6.3) and (6.4) of SPM strategy due to the simplicity of its formula.

Notice that R_2 and R'_2 coincide in the long-investment horizon limit shown in Eq. (3.20), and we will use R_2 in the following derivation.

First, we calculate the following limit by L'Hopital's rule

$$\lim_{T \rightarrow \infty} e^{aT} \Phi(-b\sqrt{T}) = \lim_{T \rightarrow \infty} \frac{\Phi(-b\sqrt{T})}{e^{-aT}} \quad (\text{B.1})$$

$$= \lim_{T \rightarrow \infty} \frac{-\frac{b}{2\sqrt{T}}\phi(-b\sqrt{T})}{-ae^{-aT}} \quad (\text{B.2})$$

$$= \lim_{T \rightarrow \infty} \frac{b}{2\sqrt{2\pi}a} \frac{e^{(a-\frac{b^2}{2})T}}{\sqrt{T}} \quad (\text{B.3})$$

$$= \begin{cases} 0 & a \leq \frac{b^2}{2} \\ \infty & a > \frac{b^2}{2} \end{cases},$$

where $a > 0$ and $b > 0$.

Using the asymptotic expansions of $\Phi^{-1}(\cdot)$ (see BAILEY [1981]), we have the following expression in the limit $T \rightarrow \infty$

$$\Phi^{-1}(e^{-R_2 T}) = -\sqrt{2R_2 T} + \frac{\ln(4\pi RT)}{2\sqrt{2RT}} + o\left(\frac{1}{T}\right) \quad (\text{B.4})$$

$$= -\sqrt{2R_2 T} + o(\ln T) \quad (\text{B.5})$$

and notice that $R_2 < \frac{1}{2}$, we can obtain

$$\begin{aligned} \lim_{T \rightarrow \infty} \text{SV}(X(T)) &= \lim_{T \rightarrow \infty} e^{2R_2 T} \Phi\left(\left(\sqrt{2R_2} - 1\right)\sqrt{T} + o(\ln T)\right) \\ &= \begin{cases} 0 & \text{if } 2R_2 \leq \frac{1}{2}(1 - \sqrt{2R_2})^2 \\ \infty & \text{if } 2R_2 > \frac{1}{2}(1 - \sqrt{2R_2})^2 \end{cases} \end{aligned} \quad (\text{B.6})$$

$$= \begin{cases} 0 & \text{if } R_2 \leq \frac{3}{2} - \sqrt{2} \\ \infty & \text{if } R_2 > \frac{3}{2} - \sqrt{2} \end{cases}. \quad (\text{B.7})$$

Similarly, we can obtain

$$\lim_{T \rightarrow \infty} \text{Var}(X(T)) = \begin{cases} 0 & \text{if } R_2 \leq \frac{3}{2} - \sqrt{2} \\ \infty & \text{if } R_2 > \frac{3}{2} - \sqrt{2} \end{cases}. \quad (\text{B.8})$$

The results for the PUM strategy is straightforward by using Eqs. (6.5) and (6.6). $\square$

Proof of Proposition 6.2:

PROOF. The equalities $\theta_{\text{PUM}} = \theta_{\text{MMV}}$ as $T \rightarrow 0$ and $\theta_{\text{MMV}} = \theta_{\text{SPM}}$ as $T \rightarrow \infty$ are due to the equivalence between these strategies (see theorem 1 in YU, ZHANG and YANG [2006]). We only need to prove the following relation $\theta_{\text{PUM}} < \theta_{\text{SPM}}$ holds when $T \rightarrow 0$ and $T \rightarrow \infty$.

For the SPM strategy, by Eqs. (6.3) and (6.4), we have

$$\theta_{\text{SPM}} = \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R'_2 T} \right) \right). \quad (\text{B.9})$$

When $T \rightarrow 0$, the target return rate R'_2 and the expected return rate R_2 have the following relation:

$$e^{R'_2 T} \Phi \left(\sqrt{T} + \Phi^{-1} \left(e^{-R'_2 T} \right) \right) = e^{R_2 T}. \quad (\text{B.10})$$

From Eq. (B.10), it is not difficult to verify that for a fixed R'_2 , $\lim_{T \rightarrow 0} R'_2 T = 0$. Then, we can derive

$$\lim_{T \rightarrow 0} \theta_{\text{SPM}} = 1. \quad (\text{B.11})$$

When $T \rightarrow \infty$, from Eq. (3.20), we know that $R_2 = R'_2$, then we can derive

$$\lim_{T \rightarrow \infty} \theta_{\text{SPM}} = 1. \quad (\text{B.12})$$

For the PUM strategy, by Eqs. (6.6) and (6.5), we have

$$\theta_{\text{PUM}} = \frac{e^{R_3^2 T} \Phi \left(-\frac{3R_3 \sqrt{T}}{2} \right) - 2 + 3 \Phi \left(\frac{R_3 \sqrt{T}}{2} \right)}{e^{R_3^2 T} - 1}. \quad (\text{B.13})$$

Then, we can derive

$$\lim_{T \rightarrow 0} \theta_{\text{PUM}} = \frac{1}{2}, \quad \lim_{T \rightarrow \infty} \theta_{\text{PUM}} = 0. \quad (\text{B.14})$$

A comparison among Eqs. (B.11), (B.12), and (B.14) completes the proof. $\square$

Investment Performance Measurement Under Asymptotically Linear Local Risk Tolerance¹

T. Zariphopoulou² and T. Zhou

The University of Texas at Austin

Abstract

Using forward optimality criteria, we analyze a portfolio choice problem when the local risk tolerance is time dependent and asymptotically linear in wealth. This class corresponds to a dynamic extension of the traditional (static) risk tolerances associated with the power, logarithmic, and exponential utilities. We provide explicit solutions for the optimal investment strategies and wealth processes in an incomplete non-Markovian market with asset prices modeled as Ito processes. The methodology allows for measuring the investment performance in terms of a benchmark and alternative market views.

1. Introduction

This chapter is a contribution to optimal portfolio management using the forward performance approach. This approach, developed by the first author and M. Musiela (see MUSIELA and ZARIPHOPOULOU [2003, 2007b]), is based on the martingale properties of the so-called forward performance process, which combines the investor's preferences with market-related inputs. In many aspects, it is similar to the traditional maximal

¹Parts of this work were presented at the 4th World Congress of the Bachelier Finance Society, (Tokyo, August 2006), the Workshop on "Financial Engineering and Actuarial Mathematics", University of Michigan, (Ann Arbor, May 2007) and the Workshop on "Further Developments in Quantitative Finance", ICMS, (Edinburgh, July 2007). The authors thank the participants for their valuable comments. They also thank G. Zitkovic for his suggestions.

²The author acknowledges partial support from the National Science Foundation (NSF Grants DMS-0091946 and DMS-FRG-0456118).

expected utility methodology where the martingality of the solution (value function) is a consequence of the dynamic programming principle. It differs, however, in that the forward performance process is defined endogenously to the market environment and for all times. A direct consequence of these properties is that the forward solution follows the market movements path-by-path and, moreover, can be constructed without references to a specific trading horizon.

Constructing the forward performance process and the associated optimal portfolio strategies poses many difficulties due to the fact that the implicit stochastic optimization problem is posed “forward” in time. A class of such processes was recently constructed by MUSIELA and ZARIPHPOULOU [2006b, 2007b] using the compilation of differential and stochastic inputs. The inputs are given by the solution of a fully nonlinear partial differential equations and a triple of stochastic processes representing a benchmark, alternative market views, and (random) time rescaling. The optimal policies are given as a linear combination of the investor’s optimal wealth and the time-rescaled risk tolerance processes. An important result is that these two processes solve an autonomous system of stochastic differential equations.

In the above analysis, pivotal role is played by the local risk tolerance function. This function is constructed from the investor’s initial risk preferences and the solution to an equation of fast-diffusion type. It is, then, used to solve the aforementioned system and, in turn, to explicitly specify the optimal investment processes in a feedback form. We note that such optimal policies come as a surprise given the non-Markovian nature of the market model.

Motivated by the emerging modeling importance of the local risk tolerance, we concentrate herein on a specific class of such functions. The family we consider corresponds to a dynamic generalization of the popular utilities used in academic works of portfolio management, namely, the power, logarithmic, and exponential ones. However, in contrast to the power and logarithmic cases, the risk tolerances we consider are globally defined (i.e., for positive and negative wealth levels).

The chapter is organized as follows. In Section 2, we introduce the model and review the definition of forward performance process and the main results of MUSIELA and ZARIPHPOULOU [2007b]. In Section 3, we focus on a two-parameter family of risk tolerance functions and construct the related forward performance process. In Section 4, we provide an explicit construction of the associated optimal allocations and wealth processes. We conclude in Section 5 where we concentrate on special limiting choices of the two risk tolerance parameters.

2. The model and its investment performance measurement

The market environment consists of one riskless and k risky securities. The risky securities are stocks, and their prices are modeled as positive and continuous Ito processes, namely, for $i = 1, \dots, k$, the price S^i of the i^{th} risky asset solves

$$dS_t^i = S_t^i \left(\mu_t^i dt + \sum_{j=1}^d \sigma_t^{ji} dW_t^j \right) \quad (2.1)$$

with $S_0^i > 0$. The process $W = (W^1, \dots, W^d)$ is a standard d -dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For simplicity, it is assumed that the underlying filtration, $\mathcal{F}_t$, coincides with the one generated by the Brownian motion, that is, $\mathcal{F}_t = \sigma(W_s : 0 \leq s \leq t)$.

The coefficients μ^i and σ^i , $i = 1, \dots, k$, follow $\mathcal{F}_t$ -adapted processes with values in $\mathcal{R}$ and $\mathcal{R}^d$, respectively. For brevity, we use σ_t to denote the volatility matrix, that is, the $d \times k$ random matrix (σ_t^{ji}) , whose i^{th} column represents the volatility σ_t^i of the i^{th} risky asset. We may, then, alternatively write (2.1) as

$$dS_t^i = S_t^i (\mu_t^i dt + \sigma_t^i \cdot dW_t).$$

The riskless asset, the savings account, has the price process B satisfying

$$dB_t = r_t B_t dt$$

with $B_0 = 1$, and for a nonnegative, $\mathcal{F}_t$ -adapted interest rate process r_t . The market coefficients, μ , σ , and r are taken to be bounded.

It is postulated that there exists an $\mathcal{F}_t$ -adapted process λ , known as the market price of risk, taking values in $\mathcal{R}^d$ and such that the equality

$$\mu_t^i - r_t = \sum_{j=1}^d \sigma_t^{ji} \lambda_t^j = \sigma_t^i \cdot \lambda_t$$

is satisfied for $t \geq 0$, for all $i = 1, \dots, k$. Using vector and matrix notation, the above becomes

$$\mu_t - r_t \mathbf{1} = \sigma_t^T \lambda_t, \quad (2.2)$$

where σ^T stands for the transpose matrix of σ and $\mathbf{1}$ denotes the d -dimensional vector with every component equal to one. It is assumed that, for all $t \geq 0$, $E_{\mathbb{P}} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds < \infty$, where σ^+ denotes the Moore–Penrose pseudoinverse of the volatility matrix (PENROSE [1955]). Recall that the matrix σ^+ exists and is unique even if the market fails to be complete.

Starting at $t = 0$ with an initial endowment $x \in \mathcal{R}$, the investor invests at all future times $t > 0$ in the riskless and risky assets. The present value of the amounts invested is denoted by π_t^0 and π_t^i , $i = 1, \dots, k$.

The present value of investor's aggregate investment is, then, given by $X_t = \sum_{i=0}^k \pi_t^i$. We will refer to X as the discounted wealth. The investment strategies $(\pi_t^0, \pi_t^1, \dots, \pi_t^k)$ will play the role of control processes and are taken to satisfy the standard assumption of being self-financing, that is, for $s \geq 0$,

$$X_s = x + \sum_{i=1}^k \int_0^s \pi_u^i (\mu_u^i - r_u) du + \sum_{i=1}^k \int_0^s \pi_u^i \sigma_u^i \cdot dW_u. \quad (2.3)$$

Writing the above in differential form yields the evolution of the discounted wealth,

$$dX_t = \sum_{i=1}^k \pi_t^i \sigma_t^i \cdot (\lambda_t dt + dW_t) = \sigma_t \pi_t \cdot (\lambda_t dt + dW_t), \quad (2.4)$$

where the (column) vector, $\pi_t = (\pi_t^i; i = 1, \dots, k)$.

The set of admissible strategies, $\mathcal{A}$, consists of all self-financing $\mathcal{F}_t$ -adapted processes π_t such that $E_{\mathbb{P}} \int_0^s |\sigma_t \pi_t|^2 dt < \infty$, for $s > 0$. It is also assumed, in order to preclude arbitrage opportunities, that for each $s > 0$, the associated wealth process, X_t , $0 \leq t \leq s$, is a $\mathbb{Q}|_{\mathcal{F}_s}$ -supermartingale for some equivalent martingale measure $\mathbb{Q}|_{\mathcal{F}_s} \sim \mathbb{P}|_{\mathcal{F}_s}$.

We continue with the definition of the forward performance process. We refer the reader to MUSIELA and ZARIPHOPOULOU [2007a,b] (see also MUSIELA and ZARIPHOPOULOU [2003]) for a detailed analysis on the motivation and modeling considerations that led to the development of the forward performance concept.

DEFINITION 2.1. An $\mathcal{F}_t$ -adapted process $U_t(x)$ is a forward performance if

- i) for each $t \geq 0$, the mapping $x \rightarrow U_t(x)$ is concave and increasing,
- ii) for each $t \geq 0$ and each self-financing strategy, $\pi \in \mathcal{A}$,

$$E_{\mathbb{P}} [U_t(X_t^{\pi})]^+ < \infty,$$

- iii) for each self-financing strategy, $\pi \in \mathcal{A}$,

$$E_{\mathbb{P}} [U_s(X_s^{\pi}) | \mathcal{F}_t] \leq U_t(X_t^{\pi}), \quad s \geq t,$$

- iv) there exists a self-financing strategy, $\pi^* \in \mathcal{A}$, for which

$$E_{\mathbb{P}} [U_s(X_s^{\pi^*}) | \mathcal{F}_t] = U_t(X_t^{\pi^*}), \quad s \geq t,$$

and

- v) it satisfies the initial datum $U_0(x) = u_0(x)$, $x \in \mathcal{R}$, where $u_0 : \mathcal{R} \rightarrow \mathcal{R}$ is a concave and increasing function of wealth.

Related to our work is the recent paper by CHOULLI, STRICKER and LI [2007] in which the authors considered random horizon choices, aiming at alleviating the dependence of the value function on a fixed (and deterministic) horizon. Their model is more general than ours, in terms of the assumptions on the price processes. However, the focus is on horizon effects and not on additional features affecting the form of the forward solution such as numeraire choice, tracking a benchmark, and alternative market views. Horizon issues were also considered by HENDERSON and HOBSON [2007a,b] who proposed the so-called horizon-unbiased utilities in the context of lognormal diffusion models and constructed a deterministic class of solutions. While preparing this work, the authors came across the preprint of BERRIER, ROGERS and TEHRANCHI [2007], where a special case of forward processes is considered in a model similar to ours (see Corollary 2.1).

We mention that forward formulations of optimal control problems have been proposed and analyzed in the past. For deterministic models, we refer the reader, among others, to SEINFELD and LAPIDUS [1968] and chapter 1 in LARSON [1968] (see also VIT [1977]). In stochastic settings, forward optimality has been studied, primarily under Markovian assumptions, by KURTZ [1984] using the associated controlled martingale problems and the construction of the Nisio semigroup (see NISIO [1981]).

Next, we review the results of MUSIELA and ZARIPHOUPOULOU [2007b]. The results consist of three parts, namely, the representation of a family of forward performance processes, the specification of the associated optimal investment strategies and wealth processes, and the construction of an autonomous system of stochastic differential equations that the optimal wealth and risk tolerance processes solve.

THEOREM 2.1. *Let the processes Y and Z solve*

$$dY_t = Y_t \delta_t \cdot (\lambda_t dt + dW_t) \quad (2.5)$$

and

$$dZ_t = Z_t \phi_t \cdot dW_t, \quad (2.6)$$

with $Y_0 = Z_0 = 1$, δ and ϕ being $\mathcal{F}_t$ -adapted and bounded with δ such that $\sigma\sigma^+\delta = \delta$ and $E_{\mathbb{P}} \int_0^t |\sigma_s \sigma_s^+ \phi_s|^2 ds < \infty$. Define the process

$$A_t = \int_0^t |\sigma_s \sigma_s^+ (\lambda_s + \phi_s) - \delta_s|^2 ds, \quad t \geq 0, \quad (2.7)$$

where λ is as in (2.2).

Let $u : \mathcal{R} \times (0, \infty) \rightarrow \mathcal{R}$ be a concave and increasing function of the spatial argument, with $u : C^{3,1}(\mathcal{R} \times (0, \infty))$ satisfying the differential constraint

$$u_t u_{xx} = \frac{1}{2} u_x^2 \quad (2.8)$$

and the initial datum

$$u(x, 0) = u_0(x), \quad (2.9)$$

with $u_0 : \mathcal{R} \rightarrow \mathcal{R}$ be in $C^3(\mathcal{R})$. Then, the process $U_t(x)$ defined by

$$U_t(x) = u\left(\frac{x}{Y_t}, A_t\right) Z_t, \quad t \geq 0 \quad (2.10)$$

is a forward performance.

The process Y , which normalizes the wealth argument, may be thought as a benchmark (or numeraire) with respect to which the investment performance is measured. The process Z refers to changes in the historical probability measure and accommodates alternative views on anticipated market movements. We will refer to Y and Z as the benchmark and market view processes, respectively.

COROLLARY 2.1. *In the special case $\delta_t = \phi_t = 0$, $t \geq 0$, the forward performance process reduces to*

$$U_t(x) = u\left(x, \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds\right). \quad (2.11)$$

If, in addition, the market parameters are constant, the forward solution is given by the deterministic function

$$U_t(x) = u\left(x, |\sigma \sigma^+ \lambda|^2 t\right). \quad (2.12)$$

Forward solutions of form (2.11) [resp. (2.12)] are the ones considered by BERRIER, ROGERS and TEHRANCHI [2007] (resp. HENDERSON and HOBSON [2007a,b]).

We continue with the optimal investment strategies and the wealth they generate. It is worth mentioning that despite the dimensionality and incompleteness of the model, as well as the allowed path dependence of the coefficients, the optimal control policies are given in an explicit feedback form. To our knowledge, this is one of the very few such examples.

For convenience and generality, we work in the benchmarked configuration, namely, we consider the processes

$$\tilde{\pi}_t^* \equiv \frac{1}{Y_t} \pi_t^* \quad \text{and} \quad \tilde{X}_t^* \equiv \frac{X_t^*}{Y_t} \quad (2.13)$$

denoting the benchmarked optimal portfolio and benchmarked optimal wealth, respectively.

A quantity that will play an important role in the analysis herein is the local risk tolerance $r : \mathcal{R} \times [0, \infty) \rightarrow \mathcal{R}^+$, defined as

$$r(x, t) = -\frac{u_x(x, t)}{u_{xx}(x, t)}, \quad (2.14)$$

with u as in (2.10). For its initial value, we will be using the notation

$$r_0(x) = r(x, 0) = -\frac{u'(x)}{u''(x)}. \quad (2.15)$$

The following assumption will be standing throughout.

ASSUMPTION 2.1. *There exist constants K_1 and K_2 such that, for all $t \geq 0$ and $x, \bar{x} \in \mathcal{R}$,*

$$r^2(x, t) \leq K_1 (1 + x^2) \quad \text{and} \quad |r(x, t) - r(\bar{x}, t)| \leq K_2 |x - \bar{x}|. \quad (2.16)$$

Next, we introduce the risk tolerance process (at benchmarked optimal wealth)

$$\tilde{R}_t^* = r(\tilde{X}_t^*, A_t), \quad (2.17)$$

with r as in (2.14) and A being the time-rescaling process defined in (2.7).

THEOREM 2.2. *The optimal benchmarked portfolio $\tilde{\pi}_t^*$, $t > 0$ is given by*

$$\tilde{\pi}_t^* = \Pi_t^*(\tilde{X}_t^*),$$

with

$$\Pi_t^*(x) = x\sigma^+\delta_t + r(x, A_t)\sigma_t^+(\lambda_t + \phi_t - \delta_t), \quad (2.18)$$

where A is as in (2.7) and $\tilde{X}_t^*$, $t > 0$, solving

$$d\tilde{X}_t^* = (\sigma_t^*\tilde{\pi}_t^* - \tilde{X}_t^*\delta_t) \cdot ((\lambda_t - \delta_t)dt + dW_t), \quad (2.19)$$

with $\tilde{\pi}_t^*$ being used.

Equivalently,

$$\tilde{\pi}_t^* = m_t\tilde{X}_t^* + n_t\tilde{R}_t^*, \quad (2.20)$$

with $\tilde{R}_t^*$ as in (2.17) and the portfolio weights given by

$$m_t = \sigma_t^+\delta_t \quad \text{and} \quad n_t = \sigma_t^+(\lambda_t + \phi_t - \delta_t). \quad (2.21)$$

An important consequence of the above theorem is that, under any choice of risk preferences, the optimal investment strategy is represented as a linear combination of two funds, namely,

$$\tilde{\pi}_t^{*,X} = m_t\tilde{X}_t^* \quad \text{and} \quad \tilde{\pi}_t^{*,R} = n_t\tilde{R}_t^*. \quad (2.22)$$

The portfolio $\tilde{\pi}_t^{*,X}$ depends functionally only on current wealth and not on the risk tolerance. The situation, however, is reversed for the second investment strategy, $\tilde{\pi}_t^{*,R}$. Note that the portfolio weights m_t , n_t , and $t > 0$ are affected exclusively by the market. They may take the value zero in which case the relevant optimal allocation vanishes. Such cases are discussed at the end of this section.

Next, we present the autonomous system of stochastic differential equations that the processes $\tilde{X}_t^*$ and $\tilde{R}_t^*$, $t > 0$ solve. Solving this system and using the linear representation result of (2.20) enable us to explicitly construct the optimal allocation vector $\tilde{\pi}_t^*$.

PROPOSITION 2.1. *Let r be the local risk tolerance function, introduced in (2.14), and A be the time-rescaling process given in (2.7). Then, for $t > 0$, the processes $\tilde{X}_t^*$ and $\tilde{R}_t^*$, $t > 0$, representing the (benchmarkd) optimal wealth and risk tolerance, solve the system*

$$\begin{cases} d\tilde{X}_t^* = \tilde{R}_t^*\sigma_t n_t \cdot ((\lambda_t - \delta_t)dt + dW_t) \\ d\tilde{R}_t^* = r_x(\tilde{X}_t^*, A_t) d\tilde{X}_t^*, \end{cases} \quad (2.23)$$

with $\tilde{X}_0^* = x$, $\tilde{R}_0^* = r_0(x)$, and n_t , $t > 0$ as in (2.21).

From (2.23), we see that the solution $(\tilde{X}_t^*, \tilde{R}_t^*)$ is fully specified once the model is chosen and the local risk tolerance function is known. Recall that r is constructed from the function u (cf. (2.14)), obtained from the nonlinear Eq. (2.8) and the initial datum (2.9). The form of the above system, however, motivates us to question whether one should first model the differential input u and, then, specify r (cf. (2.14)) or do the opposite. Herein, we follow the second approach, namely, we first choose a family of risk tolerances and, in turn, recover the associated differential input. A fundamental result used for this construction is that r satisfies an autonomous differential equation. This rather interesting property was shown by MUSIELA and ZARIPHPOULOU [2006b].

PROPOSITION 2.2. *If u satisfies (2.8), the associated local risk tolerance function r , defined in (2.14), satisfies*

$$r_t + \frac{1}{2}r^2 r_{xx} = 0. \quad (2.24)$$

It is easy to see how the differential input, u , is recovered once the local risk tolerance is known. Indeed, choosing the initial condition $r_0(x) = r(x, 0)$ and using (2.15) yield (modulo two constants) the initial datum (2.9). In turn, Eq. (2.24), together with the initial condition r_0 , will give the values $r(x, t)$, for $t > 0$. The function $u(x, t)$, $t > 0$, can be, then, retrieved from (2.14) by successive integration provided certain (time dependent) quantities are correctly specified. Related arguments are found in the proof of Proposition 3.2.

The reader with expertise in nonlinear partial differential equations will find the form of (2.24) familiar. In fact, it is a nonlinear heat equation, frequently called equation of fast-diffusion type. There is a vast literature on this equation and we refer the reader, among others to VASQUEZ [2006]. Note, however, that classical results might not be applicable since the equation is ill posed, a fact that adds various difficulties to the construction of well-defined and stable solutions.

We finish this section by mentioning that there is an alternative way to construct u from r , which could, perhaps, provide more intuition for the evolution of the differential input. Namely, note that (2.8) and (2.14) yield the transport equation

$$u_t + \frac{1}{2}r(x, t)u_x = 0. \quad (2.25)$$

Such first-order equations can be solved by the method of characteristics. In (2.25), these curves have slope equal to one-half the local risk tolerance. The input u is, then, readily constructed through the initial condition u_0 , computed from (2.15), and its propagation along the characteristic curves.

3. Asymptotically linear local risk tolerance functions

We now focus on a specific class of risk tolerance functions. To provide some motivation for our choice, let us recall that the utilities most frequently appearing in academic papers

of portfolio management are the power, logarithmic, and exponential³. In the generic problem of maximizing the expected utility of terminal wealth, these utilities are assigned at the end of the trading horizon, say $[0, T]$, and given, by

$$u^p(x; T) = \frac{1}{\gamma} x^\gamma, \quad x \geq 0, \quad \gamma < 1, \quad \gamma \neq 0, \quad (3.1)$$

$$u^l(x; T) = \log x, \quad x > 0 \quad (3.2)$$

and

$$u^e(x; T) = -e^{-\kappa x}, \quad x \in \mathcal{R}, \quad \kappa > 0. \quad (3.3)$$

The associated risk tolerances (with a slight abuse of notation, we denote them by r but keep the argument T to emphasize their dependence on the horizon choice) are, naturally, time independent and given by

$$r^p(x; T) = \frac{1}{1-\gamma} x, \quad x \geq 0 \quad \text{and} \quad r^l(x; T) = x, \quad x > 0 \quad (3.4)$$

and

$$r^e(x; T) = \frac{1}{\kappa}, \quad x \in \mathcal{R}. \quad (3.5)$$

Notice that in the traditional setting,⁴ risk preferences are chosen exclusively at the single time instant, T . In the forward framework, however, they are set at initial time, $t = 0$, and then specified for all future times $t > 0$. For the family of forward performance processes, we consider, the specification of the future values of r comes from the differential constraint (2.24).

Next, we introduce a rich family of solutions that, on one hand, are appropriate for the new framework and, on the other hand, resemble a dynamic extension of their traditional counterparts (3.4) and (3.5).

PROPOSITION 3.1. *Let $\alpha, \beta > 0$ and $r_0 : \mathcal{R} \rightarrow \mathcal{R}^+$ be given by*

$$r_0(x) = \sqrt{\alpha x^2 + \beta}.$$

Then, the function $r : \mathcal{R} \times [0, \infty) \rightarrow \mathcal{R}^+$,

$$r(x, t; \alpha, \beta) = \sqrt{\alpha x^2 + \beta e^{-\alpha t}}, \quad (3.6)$$

solves (2.24).

³The quadratic utility deserves special attention due to its saturation properties and will be studied separately.

⁴We remind the reader that there is no intermediate consumption and thus no risk preferences are allocated to incoming consumption streams.

It is easy to verify that for fixed $t = T$, $r^p(x; T)$, $r^l(x; T)$, and $r^e(x; T)$ are limiting cases of (3.6) in their respective spatial domains. Indeed,

$$r^p(x; T) = \lim_{\beta \rightarrow 0} r(x, T; \alpha, \beta), \quad x \geq 0 \quad \text{and} \quad \gamma = \frac{\sqrt{\alpha} - 1}{\sqrt{\alpha}}, \quad \alpha \neq 1, \quad (3.7)$$

$$r^l(x; T) = \lim_{\beta \rightarrow 0} r(x, T; \alpha, \beta), \quad x > 0 \quad \text{and} \quad \alpha = 1, \quad (3.8)$$

and

$$r^e(x; T) = \lim_{\alpha \rightarrow 0} r(x, T; \alpha, \beta) \quad \text{and} \quad \beta^2 = \kappa^{-1}. \quad (3.9)$$

It is immediate that the family $r(x, t; \alpha, \beta)$ satisfies Assumption 2.1. Moreover, it is *globally* defined and remains strictly positive at all positive times,

$$r(x, t; \alpha, \beta) > 0, \quad x \in \mathcal{R} \quad \text{and} \quad t > 0.$$

It has a global minimum at the origin, $(0, 0)$. The top panel of Fig. 3.1 provides its graph for $\alpha = 4$ and $\beta = 0.1$.

The family (3.6) will be called asymptotically linear due to its limiting behavior

$$\lim_{x \rightarrow \pm\infty} \frac{r(x, t; \alpha, \beta)}{|x|} = \sqrt{\alpha}, \quad t \geq 0. \quad (3.10)$$

REMARK 3.1. The above class can be readily generalized to the three-parameter family

$$r(x, t; x_0, \alpha, \beta) = \sqrt{\alpha(x - x_0)^2 + \beta e^{-\alpha t}}, \quad t > 0.$$

Since the arguments developed in the sequel can be easily extended to the above case, we choose $x_0 = 0$.

The rest of the chapter is dedicated to the construction of the forward performance process, the optimal investment allocations, and the optimal wealth when the local risk tolerance is given by (3.6). The first step is to identify the differential input that is associated with (3.6), that is, for an increasing and concave function $u(x, t; \alpha, \beta)$ satisfying

$$-\frac{u_x(x, t; \alpha, \beta)}{u_{xx}(x, t; \alpha, \beta)} = \sqrt{\alpha x^2 + \beta e^{-\alpha t}}, \quad x \in \mathcal{R} \quad \text{and} \quad t \geq 0.$$

It is easy to verify that the construction is invariant under affine transformations, namely, if $u(x, t; \alpha, \beta)$ satisfies the above, then, for M and N constants,

$$\bar{u}(x, t; \alpha, \beta) = Mu(x, t; \alpha, \beta) + N \quad (3.11)$$

satisfies it as well. To preserve the desired monotonicity of u , we need to choose $M > 0$.

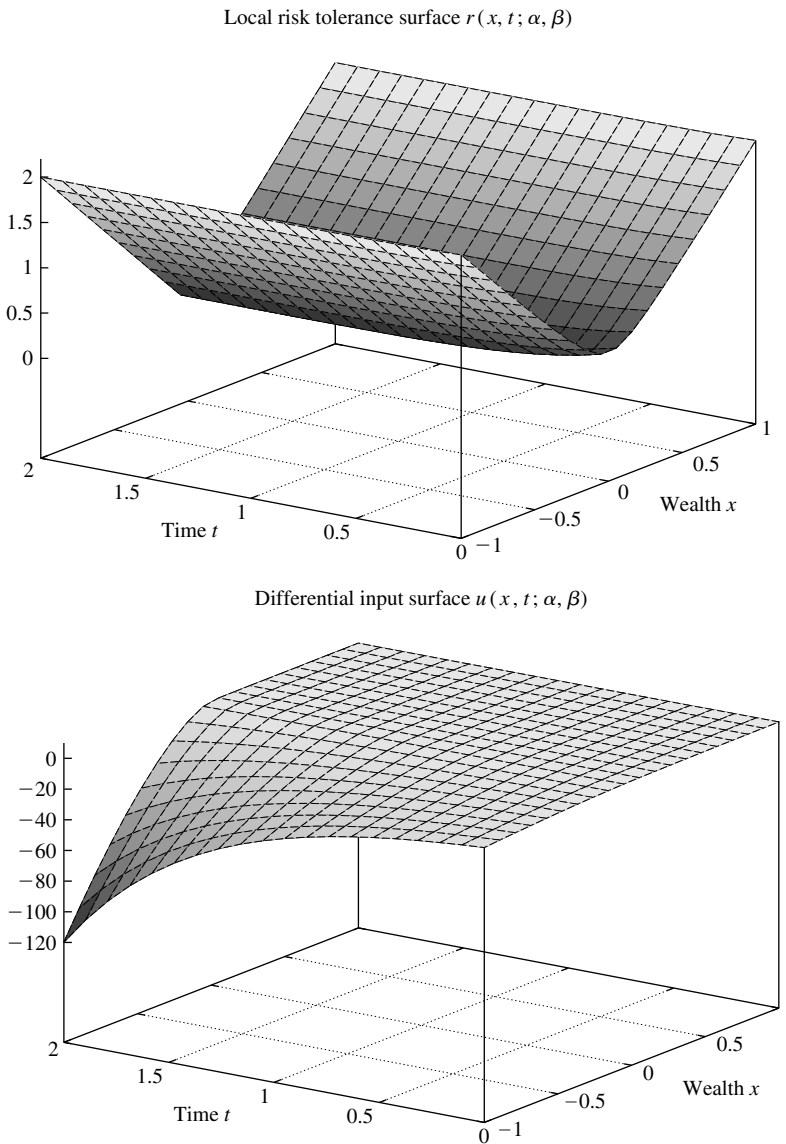


FIG. 3.1 **The risk tolerance and differential input surfaces.** For parameters $\alpha = 4$ and $\beta = 0.1$, this figure presents the local risk tolerance surface $r(x, t; \alpha, \beta) = \sqrt{\alpha x^2 + \beta e^{-\alpha t}}$ (first panel) and the differential input surface $u(x, t; \alpha, \beta)$ given in (3.12), for $M = 1$ and $N = 0$ (second panel).

As it will be clear from the proof of the next proposition, the form of u depends on the range of the parameter α . Specifically, one needs to look at the cases $\alpha = 1$ and $\alpha \neq 1$, separately.

PROPOSITION 3.2. *Let r be given by (3.6) with $\alpha, \beta > 0$. The following statements hold*

i) *If $\alpha \neq 1$, the associated differential input is given, for $x \in \mathcal{R}$ and $t \geq 0$, by*

$$\begin{aligned} u(x, t; \alpha, \beta) &= M \frac{(\sqrt{\alpha})^{1+\frac{1}{\sqrt{\alpha}}}}{\alpha-1} e^{\frac{1-\sqrt{\alpha}}{2}t} \frac{\left(\frac{\beta}{\sqrt{\alpha}}e^{-\alpha t} + (1+\sqrt{\alpha})x\left(\sqrt{\alpha}x + \sqrt{\alpha x^2 + \beta e^{-\alpha t}}\right)\right)}{\left(\sqrt{\alpha}x + \sqrt{\alpha x^2 + \beta e^{-\alpha t}}\right)^{1+\frac{1}{\sqrt{\alpha}}}} + N. \end{aligned} \quad (3.12)$$

ii) *If $\alpha = 1$, then, for $x \in \mathcal{R}$ and $t \geq 0$,*

$$\begin{aligned} u(x, t; 1, \beta) &= \frac{M}{2} \left(\log \left(x + \sqrt{x^2 + \beta e^{-t}} \right) - \frac{e^t}{\beta} x \left(x - \sqrt{x^2 + \beta e^{-t}} \right) - \frac{t}{2} \right) + N. \end{aligned} \quad (3.13)$$

PROOF. Rewriting (2.14) as $(\log u_x(x, t; \alpha, \beta))_x = -r(x, t; \alpha, \beta)^{-1}$ and integrating yields

$$u_x(x, t; \alpha, \beta) = m(t) \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right)^{-\frac{1}{\sqrt{\alpha}}} \quad (3.14)$$

for some function $m : [0, \infty) \rightarrow \mathcal{R}^+$. In turn,

$$u_{xx}(x, t; \alpha, \beta) = -m(t) \frac{\left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right)^{-\frac{1}{\sqrt{\alpha}}}}{\sqrt{\alpha x^2 + \beta e^{-\alpha t}}}.$$

From Eq. (2.8), we, then, deduce that

$$u_t(x, t; \alpha, \beta) = -\frac{1}{2} m(t) \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right)^{-\frac{1}{\sqrt{\alpha}}} \sqrt{\alpha x^2 + \beta e^{-\alpha t}}.$$

Integrating yields, for $\alpha = 1$,

$$\begin{aligned} u(x, t; 1, \beta) &= -\frac{1}{2\beta} m(t) \left(e^t x^2 - e^t x \sqrt{x^2 + \beta e^{-t}} - \beta \log \left(x + \sqrt{x^2 + \beta e^{-t}} \right) \right) \\ &\quad + n(t), \end{aligned}$$

while for $\alpha \neq 1$,

$$u(x, t; \alpha, \beta) = m(t) \frac{\sqrt{\alpha}}{\alpha - 1} \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right)^{-\left(1 + \frac{1}{\sqrt{\alpha}}\right)} \\ \times \left(\frac{\beta}{\alpha} e^{-\alpha t} + (1 + \sqrt{\alpha}) x \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right) \right) + n(t).$$

We analyze only the latter case. Differentiating the above gives

$$u_t(x, t) = n'(t) + \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right)^{-\left(1 + \frac{1}{\sqrt{\alpha}}\right)} \\ \times \left(\beta e^{-\alpha t} \left(\frac{m'(t)}{\sqrt{\alpha}(\alpha - 1)} - \frac{m(t)}{2(\sqrt{\alpha} + 1)} \right) \right. \\ \left. + m'(t) \frac{\sqrt{\alpha}}{\sqrt{\alpha} - 1} x \left(x + \sqrt{x^2 + \frac{\beta}{\alpha} e^{-\alpha t}} \right) \right).$$

Reconciling the above two expressions for $u_t(x, t)$ yields

$$m'(t) = -\frac{\sqrt{\alpha} - 1}{2} m(t) \quad \text{and} \quad n'(t) = 0.$$

Thus, $m(t) = M e^{-\frac{\sqrt{\alpha}-1}{2}t}$ and $n(t) = N$, and (3.12) follows. $\square$

The initial value u_0 , derived from (3.12) and (3.13) for $t = 0$, will be needed for special cases presented in the sequel. For convenience, we write it below, namely, for $x \in \mathcal{R}$, $\alpha > 0$ ($\alpha \neq 1$),

$$u_0(x; \alpha, \beta) = M \frac{(\sqrt{\alpha})^{1 + \frac{1}{\sqrt{\alpha}}} \left(\frac{\beta}{\sqrt{\alpha}} + (1 + \sqrt{\alpha}) x \left(\sqrt{\alpha} x + \sqrt{\alpha x^2 + \beta} \right) \right)}{\alpha - 1 \left(\sqrt{\alpha} x + \sqrt{\alpha x^2 + \beta} \right)^{1 + \frac{1}{\sqrt{\alpha}}}} + N \quad (3.15)$$

while for $\alpha = 1$,

$$u_0(x, 1, \beta) = \frac{M}{2} \left(\log \left(x + \sqrt{x^2 + \beta} \right) - \frac{x}{\beta} \left(x - \sqrt{x^2 + \beta} \right) \right) + N. \quad (3.16)$$

Once the differential input is specified, the construction of the forward performance process is an immediate application of Theorem 2.1.

PROPOSITION 3.3. *Let the local risk tolerance and (Y, Z, A) be as in (3.6), (2.5), (2.6), and (2.7). Then, for $x \in \mathcal{R}$ and $t \geq 0$, the process*

$$U_t(x; \alpha, \beta) = u\left(\frac{x}{Y_t}, A_t; \alpha, \beta\right) Z_t, \quad (3.17)$$

with $u(x, t; \alpha, \beta)$ given in Proposition 3.2, is a forward performance.

REMARK 3.2. It is important to notice that in the classical case, the power and logarithmic utilities u^l and u^p [cf. (3.1) and (3.2)] are not everywhere defined. This restrains the applicability of such preferences especially when we introduce derivatives and liabilities. Note, however, that their time-dependent forward counterparts, (3.12) and (3.13), are spatially globally defined. For this reason, the above process $U_t(x; \alpha, \beta)$ is also globally defined. The situation changes, however, when $\beta \rightarrow 0$ and/or $\alpha \rightarrow 0$. These cases deserve a special attention and are discussed separately (see Section 5).

In the second panel of Fig. 3.1, we provide the graph of the function $u(x, t; \alpha, \beta)$ [cf. (3.12)] for $\alpha = 4$ and $\beta = 0.1$. Also, we provide the cross sections $u(x, t_0; \alpha, \beta)$ and $u(x_0, t; \alpha, \beta)$. The first panel of Fig. 3.2 shows, for fixed time t_0 , the monotonicity and concavity of $u(x, t_0; \alpha, \beta)$, while the second panel shows the monotonicity of $u(x_0, t; \alpha, \beta)$ in terms of time.

4. At the optimum

We provide explicit solutions for the optimal investment policies, the associated wealth and the optimal investment performance. The key ingredients used in the construction of these processes are the autonomous system that the optimal wealth and risk tolerance processes satisfy [cf. (2.23)] together with the specific form of the local risk tolerance function [cf. (3.6)]. We remind the reader that the results are stated in the benchmarked configuration.

THEOREM 4.1. *The processes $\tilde{X}_t^*$ and $\tilde{R}_t^*$, $t > 0$, representing the optimal (benchmarked) wealth and risk tolerance solve the system of linear stochastic differential equations*

$$\begin{cases} d\tilde{X}_t^* = \tilde{R}_t^* \sigma_t n_t \cdot ((\lambda_t - \delta_t) dt + dW_t) \\ d\tilde{R}_t^* = \alpha \tilde{X}_t^* \sigma_t n_t \cdot ((\lambda_t - \delta_t) dt + dW_t), \end{cases} \quad (4.1)$$

with $\tilde{X}_0^* = x$ and $\tilde{R}_0^* = r(x, 0) = \sqrt{\alpha x^2 + \beta}$.

In turn,

$$\tilde{X}_t^* = e^{-\frac{\alpha}{2} \int_0^t |\sigma_s n_s|^2 ds} \left(x \cosh(\sqrt{\alpha} k_t) + \sqrt{x^2 + \frac{\beta}{\alpha}} \sinh(\sqrt{\alpha} k_t) \right) \quad (4.2)$$

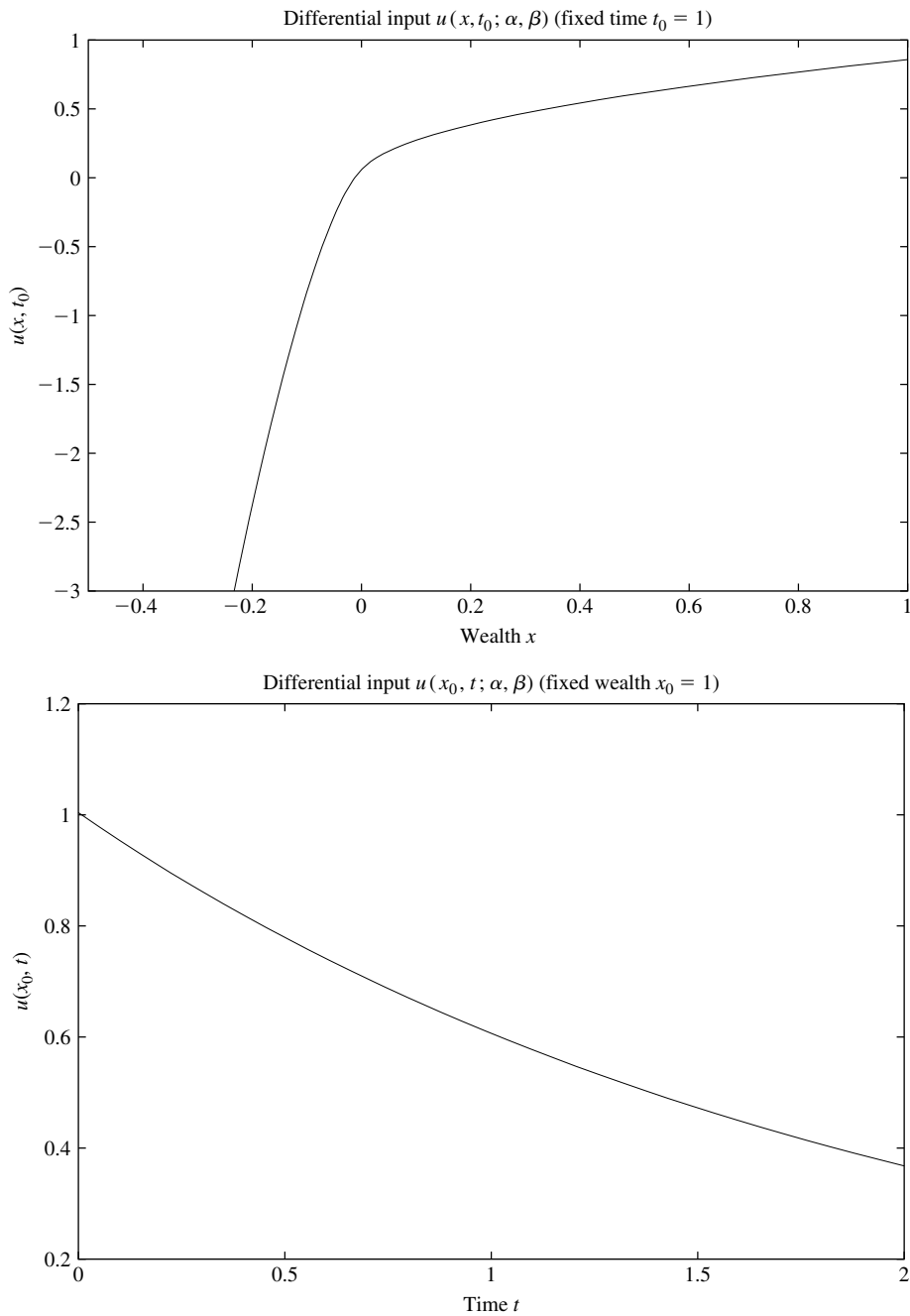


FIG. 3.2 **Cross sections of the differential input.** For parameters $\alpha = 4$ and $\beta = 0.1$, this figure presents the cross sections of the differential input surface $u(x, t; \alpha, \beta)$ given in (3.12), for $M = 1$ and $N = 0$. The first panel corresponds to $u(x, t_0; \alpha, \beta)$, with $t_0 = 1$. The second panel corresponds to $u(x_0, t; \alpha, \beta)$, with $x_0 = 1$.

and

$$\tilde{R}_t^* = e^{-\frac{\alpha}{2} \int_0^t |\sigma_s n_s|^2 ds} \left(\sqrt{\alpha} x \sinh(\sqrt{\alpha} k_t) + \sqrt{\alpha x^2 + \beta} \cosh(\sqrt{\alpha} k_t) \right), \quad (4.3)$$

where $n_t, t > 0$, as in (2.21) and

$$k_t = \int_0^t \sigma_s \sigma_s^+ (\lambda_s + \phi_s - \delta_s) \cdot ((\lambda_s - \delta_s) ds + dW_s). \quad (4.4)$$

The vector of optimal asset allocations is given by

$$\tilde{\pi}_t^* = m_t \tilde{X}_t^* + n_t \tilde{R}_t^*, \quad (4.5)$$

with $\tilde{X}_t^*, \tilde{R}_t^*$ as above and m_t as in (2.21).

PROOF. The coefficients in (4.1) follow from Theorem 2.2 [see (2.18) and (2.19)] and (3.6). The admissibility conditions for the optimal policy follow from the boundedness assumption on the market coefficients. Indeed, one can easily see that the integrability condition $E_{\mathbb{P}} \int_0^s |\pi_t^*|^2 dt < \infty$ holds for $0 \leq t \leq s$ and that the wealth process $X_t^*, 0 \leq t \leq s$, is a $\mathbb{Q}|_{\mathcal{F}_s}$ -martingale, where

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_s} = \exp \left\{ - \int_0^s \lambda_t \cdot dW_t - \frac{1}{2} \int_0^s |\lambda_t|^2 ds \right\}.$$

The arguments in the benchmarked configuration follow easily as well.

Adding and subtracting the equations in (4.1) yields

$$d(\sqrt{\alpha} \tilde{X}_t^* + \tilde{R}_t^*) = \sqrt{\alpha} (\sqrt{\alpha} \tilde{X}_t^* + \tilde{R}_t^*) \sigma_t n_t \cdot ((\lambda_t - \delta_t) dt + dW_t)$$

and

$$d(\sqrt{\alpha} \tilde{X}_t^* - \tilde{R}_t^*) = -\sqrt{\alpha} (\sqrt{\alpha} \tilde{X}_t^* - \tilde{R}_t^*) \sigma_t n_t \cdot ((\lambda_t - \delta_t) dt + dW_t),$$

and we easily conclude. $\square$

For completeness, we provide the optimal allocations π_t^* and wealth X_t^* in the original (nonbenchmarked) formulation. Recall [see (2.13) and (2.5)] that, for $t > 0$,

$$X_t^* = Y_t \tilde{X}_t^* \quad \text{and} \quad \pi_t^* = m_t X_t^* + n_t Y_t r \left(\frac{X_t^*}{Y_t}, A_t \right).$$

PROPOSITION 4.1. *Let $x \in \mathcal{R}$ be the investor's initial endowment. Then, the optimal allocation vector and associated optimal wealth are given, respectively, by*

$$\begin{aligned}\pi_t^* &= e^{\zeta_t} m_t \left(x \cosh(\sqrt{\alpha} k_t) + \sqrt{x^2 + \frac{\beta}{\alpha}} \sinh(\sqrt{\alpha} k_t) \right) \\ &\quad + e^{\zeta_t} n_t \left(\sqrt{\alpha} x \cosh(\sqrt{\alpha} k_t) + \sqrt{\alpha x^2 + \beta} \sinh(\sqrt{\alpha} k_t) \right), \quad t > 0,\end{aligned}\quad (4.6)$$

and

$$X_t^* = e^{\zeta_t} \left(x \cosh(\sqrt{\alpha} k_t) + \sqrt{x^2 + \frac{\beta}{\alpha}} \sinh(\sqrt{\alpha} k_t) \right), \quad t \geq 0, \quad (4.7)$$

where

$$\zeta_t = \int_0^t \left(\delta_s \cdot \lambda_s - \frac{1}{2} |\delta_s|^2 - \frac{\alpha}{2} |\sigma_s n_s|^2 \right) ds + \int_0^t \delta_s \cdot dW_s \quad (4.8)$$

and m_t , n_t and k_t as in (2.21) and (4.4).

Next, we look at the extreme cases $m_t = n_t = 0$, $t > 0$ leading, respectively, to $\tilde{\pi}_t^{*,X} = 0$ and $\tilde{\pi}_t^{*,R} = 0$. It is easy to check that they reduce to $\delta_t = 0$ and $\lambda_t + \phi_t - \delta_t = 0$, $t \geq 0$.

(i) *Absence of benchmark*: $\delta_t = 0$. Then, (2.5) yields $Y_t = Y_0 = 1$, $t \geq 0$.

Then, the first portfolio component vanishes, $\pi_t^{*,X} = 0$, while the second simplifies to

$$\begin{aligned}\pi_t^{*,R} &= e^{-\frac{\alpha}{2} \int_0^t |\sigma_s \sigma_s^+ (\lambda_s + \phi_s)|^2 ds} \sigma_t^+ (\lambda_t + \phi_t) \\ &\quad \times \left(\sqrt{\alpha} x \cosh(\sqrt{\alpha} k'_t) + \sqrt{\alpha x^2 + \beta} \sinh(\sqrt{\alpha} k'_t) \right),\end{aligned}$$

with

$$k'_t = \int_0^t \sigma_s \sigma_s^+ (\lambda_s + \phi_s) \cdot (\lambda_s ds + dW_s). \quad (4.9)$$

The optimal wealth is given by

$$X_t^* = e^{-\frac{\alpha}{2} \int_0^t |\sigma_s \sigma_s^+ (\lambda_s + \phi_s)|^2 ds} \left(x \cosh(\sqrt{\alpha} k'_t) + \sqrt{x^2 + \frac{\beta}{\alpha}} \sinh(\sqrt{\alpha} k'_t) \right).$$

The (sub)case $\lambda_t + \phi_t = 0$ deserves special attention since $\pi_t^{*,R}$ also vanishes. Moreover, $A_t = 0$, $t \geq 0$, leads to the performance process

$$U_t(x, t; \alpha, \beta) = u_0(x; \alpha, \beta) Z_t, \quad (4.10)$$

with u_0 as in (3.15) or (3.16). Moreover,

$$\tilde{\pi}_t^{*,X} = \pi_t^{*,X} = 0 \quad \text{and} \quad \tilde{\pi}_t^{*,R} = \pi_t^{*,R} = 0, \quad t \geq 0$$

and, in turn,

$$\tilde{X}_t^* = X_t^* = x, \quad t \geq 0.$$

At the optimum,

$$U_t^*(x; \alpha, \beta) = U_t(x; \alpha, \beta) = u_0(x; \alpha, \beta) Z_t.$$

The above results show that for the above choice of coefficients ($\lambda_t + \phi_t = 0$ and $\delta_t = 0, t \geq 0$), it is optimal for the investor to invest zero wealth into each risky asset, a result that comes as a surprise given the nonzero returns. Notice that such a solution seems to capture quite accurately the strategy of a derivatives' trader for whom the underlying objective is to hedge as opposed to the asset manager whose objective is to invest. Naturally, under this strategy, the forward performance process is not affected by the time evolution of u . This is a direct consequence of the fact that the time-rescaling process A degenerates.

(ii) *Tracking the benchmark:* $\lambda_t + \phi_t - \delta_t = 0, t \geq 0$.

In this case, the portfolio $\tilde{\pi}_t^{*,R}$ vanishes and thus any dependence on the risk tolerance dissipates. The investor invests the fraction m_t of his/her (benchmarked) wealth to the risky assets and puts the rest in the riskless bond. We have $A_t = 0, t \geq 0$, and thus the performance process is given by (4.10). Moreover,

$$\tilde{\pi}_t^{*,X} = m_t \tilde{X}_t^* \quad \text{and} \quad \tilde{\pi}_t^{*,R} = 0, \quad t > 0.$$

The absolute wealth tracks the benchmark, while the (benchmarked) risk tolerance process remains unchanged:

$$X_t^* = x Y_t \quad \text{and} \quad \tilde{R}_t^* = \tilde{R}_0^* = \sqrt{\alpha x^2 + \beta}.$$

At the optimum,

$$U_t^*(x; \alpha, \beta) = u_0\left(\frac{X_t^*}{Y_t}; \alpha, \beta\right) Z_t = u_0(x; \alpha, \beta) Z_t.$$

REMARK 4.1. The above result shows that the investor allocates in the riskless asset the amount $\tilde{\pi}_t^{*,0} = p_t X_t^*$, with $p_t = 1 - m_t \cdot \mathbf{1}$. Note that depending on the level of the weight process $p_t, t \geq 0$, which is determined only by the market parameters, the investor allocates arbitrarily small or large proportions of the wealth in the riskless asset. In the extreme case, $p_t = 0, t \geq 0$, the investor allocates zero wealth in the riskless asset, while in the other such case, namely, when $p_t = 1, t \geq 0$, in the optimal allocation, the investor allocates all wealth in the riskless asset.

5. Special cases: CARA and CRRA forward performance processes

We now look at the behavior of the solutions when the parameters α and β vanish. Recalling equalities (3.7), (3.8), and (3.9), we anticipate that the limiting risk tolerance and differential input must resemble their classical power, logarithmic, and exponential analogues. Although passing to the limit in (3.6), (3.12), and (3.13) is not difficult from the technical point of view, the emerging limits have some noteworthy properties. To simplify the notation, we skip throughout the parameter notation and use, instead, the superscripts e , p , and l in a self-evident way.

(i) *The case $\alpha = 0$.*

Passing to the limit in (3.6) and (3.12) yields, for $t \geq 0$,

$$\lim_{\alpha \rightarrow 0} r(x, t; \alpha, \beta) = \sqrt{\beta}, \quad x \in \mathcal{R} \quad (5.1)$$

and

$$u^e(x, t) = \lim_{\alpha \rightarrow 0} u(x, t; \alpha, \beta) = -e^{-\frac{x}{\sqrt{\beta}} + \frac{t}{2}}, \quad x \in \mathcal{R}, \quad (5.2)$$

where we chose, for simplicity, $M = (\sqrt{\alpha})^{\frac{1}{\sqrt{\alpha}}}(\sqrt{\beta})^{\frac{1-\sqrt{\alpha}}{\sqrt{\alpha}}}$ and $N = 0^5$; Fig. 5.1 demonstrates this convergence.

One, easily, finds that the limiting local risk tolerance (5.1) leads to an exponential forward performance process. This class of solutions was extensively analyzed by MUSIELA and ZARIPHOPULOU [2006b, 2007a], and we refer the reader therein for detailed arguments.

PROPOSITION 5.1. *For $\alpha = 0$, $\beta > 0$, $t \geq 0$, $x \in \mathcal{R}$, and (Y, Z, A) as in (2.5), (2.6), and (2.7), the process*

$$U_t^e(x) = -\exp\left(-\frac{1}{\sqrt{\beta}}\frac{x}{Y_t} + \frac{A_t}{2}\right)Z_t$$

is a forward performance. Moreover, the optimal (benchmarked) investment strategy and the associated wealth are given by the processes

$$\tilde{\pi}_t^{*,e} = (x + \sqrt{\beta}k_t)m_t + \sqrt{\beta}n_t \quad \text{and} \quad \tilde{X}_t^{*,e} = x + \sqrt{\beta}k_t, \quad (5.3)$$

with n_t, k_t as in (2.21) and (4.4).

⁵For the second limit, we use in (3.12) that for $\beta > 0$, $x \in \mathcal{R}$,

$$\lim_{\alpha \rightarrow 0} \left(\sqrt{\frac{\alpha}{\beta}} + \sqrt{\frac{\alpha}{\beta}x^2 + 1} \right)^{-\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}}} = e^{-\frac{x}{\sqrt{\beta}}}.$$

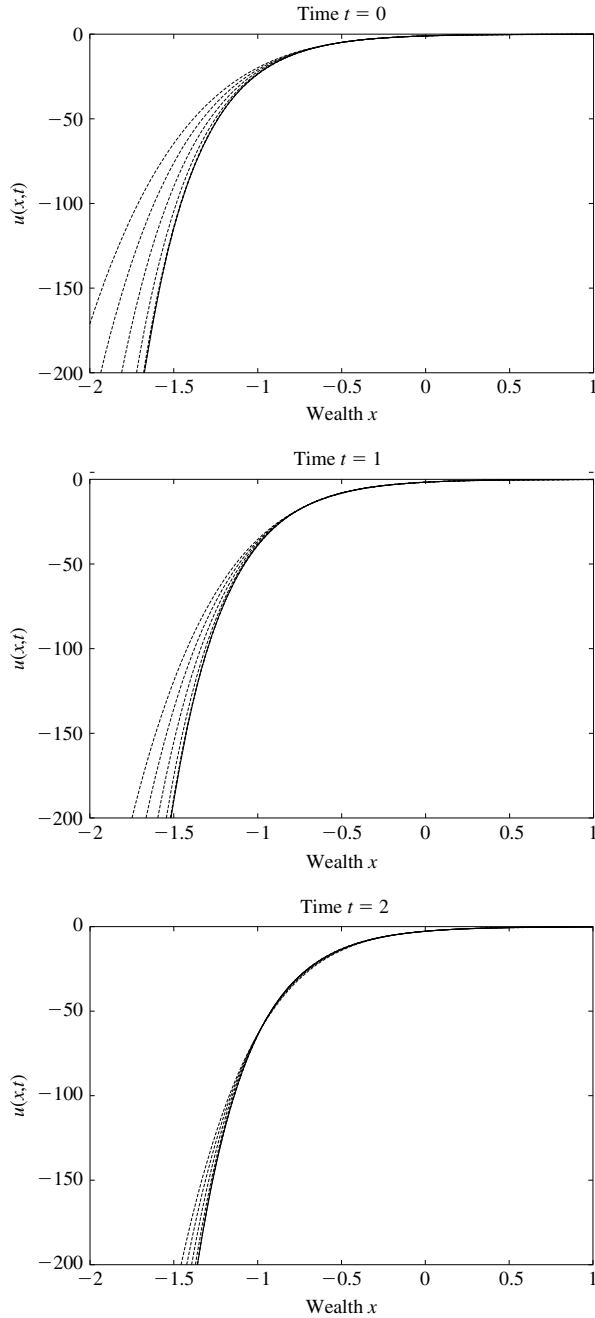


FIG. 5.1 **Convergence to the exponential case.** We choose $\beta = 0.1$. For times $t = 0, 1, 2$, the three panels demonstrate the convergence, as $\alpha \rightarrow 0$, of the differential input $u(x, t; \alpha, \beta)$, given in (3.12), for $M = (\sqrt{\alpha})^{\frac{1}{\sqrt{\alpha}}} (\sqrt{\beta})^{\frac{1-\sqrt{\alpha}}{\sqrt{\alpha}}}$ and $N = 0$. The curve of solid line corresponds to the exponential differential input $u^e(x, t) = \lim_{\alpha \rightarrow 0} u(x, t; \alpha, \beta) = -e^{-\frac{x}{\sqrt{\beta}} + \frac{1}{2}t}$. The curves of dotted lines correspond to $u(x, t; \alpha, \beta)$ for $\alpha = 1 \times 10^{-1}, 6 \times 10^{-2}, 3 \times 10^{-2}, 1 \times 10^{-2}, 1 \times 10^{-3}$ and 1×10^{-4} , respectively.

At the optimum,

$$U_t^e(X_t^*) = -\exp\left(-\frac{x}{\sqrt{\beta}} - k_t + \frac{1}{2} \int_0^t |\sigma_s n_s|^2 ds\right) Z_t.$$

REMARK 5.1. It is interesting to observe that due to the presence of the benchmark, the optimal investment policy depends on the current wealth. This is in contrast to the known results, which yield wealth-independent policies, a fact that is frequently used against the use of exponential preferences in models of investment and (indifference) valuation.

Next, we write the solutions when both the benchmark and the market view process are absent.

COROLLARY 5.1. Let $\delta_t = \phi_t = 0$, $t \geq 0$. Then,

$$U_t^e(x) = -\exp\left(-\frac{1}{\sqrt{\beta}}x + \frac{1}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds\right). \quad (5.4)$$

Moreover,

$$X_t^{*,e} = x + \sqrt{\beta} \int_0^t \sigma_s \sigma_s^+ \lambda_s \cdot (\lambda_s ds + dW_s) \quad \text{and} \quad \pi_t^{*,e} = \sqrt{\beta} \sigma_t^+ \lambda_t.$$

(ii) The case $\beta = 0$.

Passing to the limit in (3.6) yields, for $t \geq 0$,

$$\lim_{\beta \rightarrow 0} r(x, t; \alpha, \beta) = \sqrt{\alpha}|x|, \quad x \in \mathcal{R}. \quad (5.5)$$

In turn, for $\alpha > 1$ ($\alpha < 1$), (3.12) gives

$$u^p(x, t) = \lim_{\beta \rightarrow 0} u(x, t; \alpha, \beta) = \begin{cases} \frac{1}{\gamma} x^\gamma e^{-\frac{1}{2} \frac{\gamma}{1-\gamma} t} & \text{for } x \geq 0 \quad (x > 0) \\ -\infty & \text{for } x < 0 \quad (x \leq 0) \end{cases} \quad (5.6)$$

with

$$\gamma = \frac{\sqrt{\alpha} - 1}{\sqrt{\alpha}}, \quad \alpha > 0, \quad (5.7)$$

and where we chose the constants $M = 2^{\frac{1}{\sqrt{\alpha}}}$ and $N = 0$.

For $\alpha = 1$, (3.13) yields

$$u^l(x, t) = \lim_{\beta \rightarrow 0} u(x, t; 1, \beta) = \begin{cases} \log x - \frac{1}{2}t & \text{for } x > 0 \\ -\infty & \text{for } x \leq 0 \end{cases} \quad (5.8)$$

for the choice $M = 2$ and $N = -(\frac{1}{2} + \log 2)$.

The limiting behavior of the differential inputs $u(x, t; \alpha, \beta)$ and $u(x, t; 1, \beta)$ when $\beta \rightarrow 0$ is shown in Figs. 5.2 and 5.3.

We see that while the local risk tolerance in (5.5) is well defined for all $x \in \mathcal{R}$, the associated differential inputs u^p and u^l explode for nonpositive wealth levels. This impedes us from having globally defined forward performance processes. A well-defined problem may be formulated if we a priori constrain the set of admissible policies to strategies that generate nonnegative wealth. A modification of the proofs of Theorems 2.1 and 2.2 yields the following results.

PROPOSITION 5.2. *Let the local risk tolerance be given by*

$$r(x, t; \alpha, 0) = \sqrt{\alpha}x,$$

for $x \geq 0$ when $\alpha > 1$ and $x > 0$ when $\alpha < 1$ ($\alpha \neq 0$). Let, also, (Y, Z, A) be as in (2.5), (2.6), and (2.7). Then, for $\alpha > 1$ ($\alpha < 1$), the process

$$U_t^p(x) = \frac{1}{\gamma} \left(\frac{x}{Y_t} \right)^\gamma e^{-\frac{1}{2} \frac{\gamma}{1-\gamma} A_t} Z_t, \quad x \geq 0 \quad (x > 0), \quad (5.9)$$

is a forward performance. Moreover, the optimal investment strategy and associated wealth processes are given by

$$\tilde{\pi}_t^{*,p} = x (m_t + \sqrt{\alpha} n_t) \exp \left(-\frac{\alpha}{2} \int_0^t |\sigma_s n_s|^2 ds + \sqrt{\alpha} k_t \right)$$

and

$$\tilde{X}_t^{*,p} = x \exp \left(-\frac{\alpha}{2} \int_0^t |\sigma_s n_s|^2 ds + \sqrt{\alpha} k_t \right),$$

with n_t and k_t as in (2.21) and (4.4).

At the optimum,

$$\begin{aligned} & U_t^p(X_t^{*,p}) \\ &= \frac{1}{\gamma} x^\gamma \exp \left(-\frac{\alpha-1}{2} \int_0^t |\sigma_s n_s|^2 ds + (\sqrt{\alpha}-1)k_t \right) Z_t, \quad \text{for } x \geq 0 \quad (x > 0). \end{aligned}$$

Similar results can be obtained for the logarithmic case.

PROPOSITION 5.3. *Let the local risk tolerance be given by*

$$r(x, t; 1, 0) = x, \quad x > 0.$$

Then, the process

$$U_t^l(x) = \left(\log \frac{x}{Y_t} - \frac{A_t}{2} \right) Z_t, \quad x > 0$$

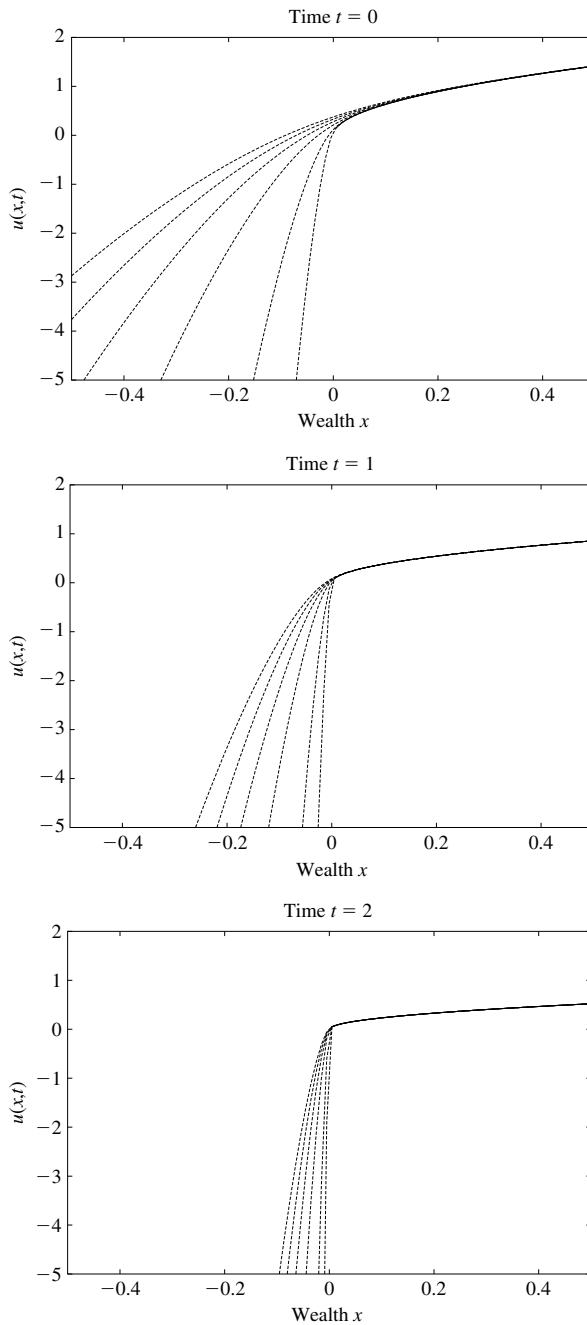


FIG. 5.2 **Convergence to the power case.** We choose $\alpha = 4$. For times $t = 0, 1, 2$, the three panels demonstrate the convergence, as $\beta \rightarrow 0$, of the differential input $u(x, t; \alpha, \beta)$, given in (3.12), for $M = 2^{\frac{1}{\alpha}}$ and $N = 0$. The curve of solid line corresponds to the power differential input $u^P(x, t) = \lim_{\beta \rightarrow 0} u(x, t; \alpha, \beta) = \frac{1}{\gamma} x^\gamma e^{-\frac{1}{2} \frac{\gamma}{1-\gamma} t}$. The curves of dotted lines correspond to $u(x, t; \alpha, \beta)$ for $\beta = 1 \times 10^{-1}, 6 \times 10^{-2}, 3 \times 10^{-2}, 1 \times 10^{-2}, 1 \times 10^{-3}$ and 1×10^{-4} , respectively.

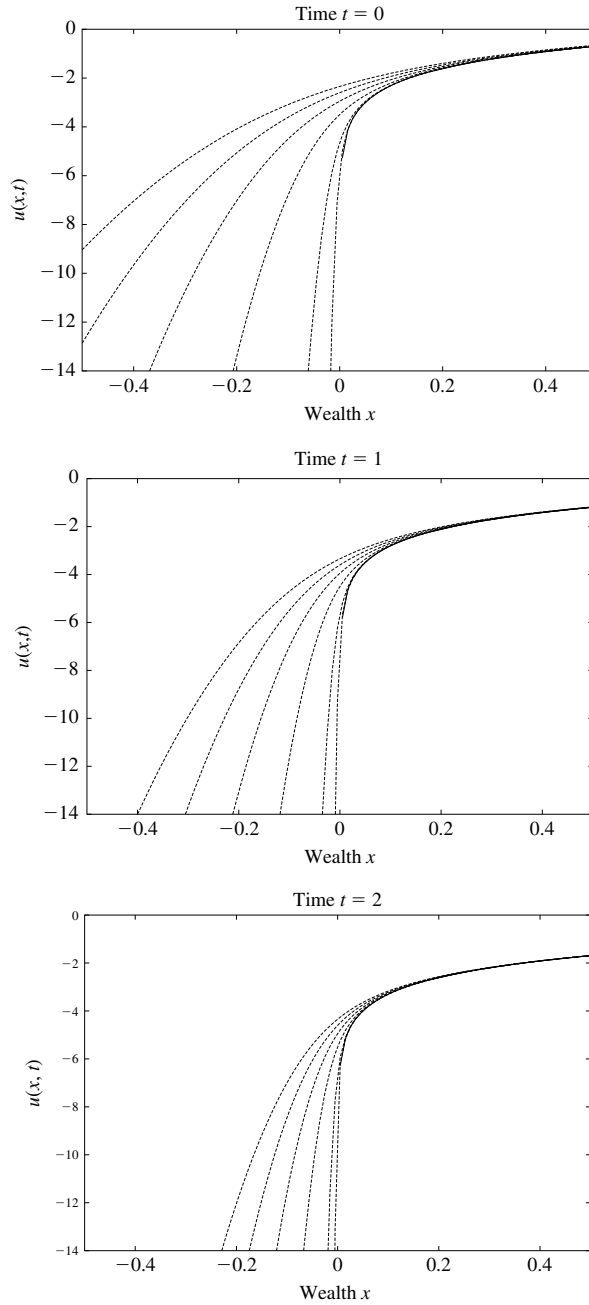


FIG. 5.3 **Convergence to the logarithmic case.** For times $t = 0, 1, 2$, the three panels demonstrate the convergence, as $\beta \rightarrow 0$, of the differential input $u(x, t; \alpha, \beta)$, given in (3.13), for $M = 2$ and $N = -(\frac{1}{2} + \log 2)$. The curve of solid line corresponds to the logarithmic differential input $u^l(x, t) = \lim_{\beta \rightarrow 0} u(x, t; 1, \beta) = \log(x) - \frac{1}{2}t$. The curves of dotted lines correspond to $u(x, t; \alpha, \beta)$ for $\beta = 1 \times 10^{-1}, 6 \times 10^{-2}, 3 \times 10^{-2}, 1 \times 10^{-2}, 1 \times 10^{-3}$ and 1×10^{-4} , respectively.

is a forward performance. Moreover, the optimal investment strategy and associated wealth processes are given by

$$\tilde{\pi}_t^{*,l} = x(m_t + n_t) \exp \left(-\frac{1}{2} \int_0^t |\sigma_s n_s|^2 ds + k_t \right)$$

and

$$\tilde{X}_t^{*,l} = x \exp \left(-\frac{1}{2} \int_0^t |\sigma_s n_s|^2 ds + k_t \right).$$

At the optimum,

$$U_t^l(X_t^{*,l}) = \left(\log x - \int_0^t |\sigma_s n_s|^2 ds + k_t \right) Z_t$$

with n_t and k_t as in (2.21) and (4.4).

In an analogy to Corollary 5.1, we look at the case of no benchmark and no alternative market views.

COROLLARY 5.2. Let $\delta_t = \phi_t = 0$, $t \geq 0$, and $\beta = 0$. Then, for $\alpha > 1$ ($\alpha < 1$),

$$U_t^p(x) = \frac{1}{\gamma} x^\gamma \exp \left(-\frac{\gamma}{2(1-\gamma)} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds \right), \quad x \geq 0 \quad (x > 0). \quad (5.10)$$

Moreover,

$$\pi_t^{*,p} = \sqrt{\alpha} x \sigma_t^+ \lambda_t \exp \left(-\frac{\alpha}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds + \sqrt{\alpha} \int_0^t \sigma_s \sigma_s^+ \lambda_s \cdot (\lambda_s ds + dW_s) \right)$$

and

$$X_t^{*,p} = x \exp \left(-\frac{\alpha}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds + \sqrt{\alpha} \int_0^t \sigma_s \sigma_s^+ \lambda_s \cdot (\lambda_s ds + dW_s) \right).$$

COROLLARY 5.3. Let $\delta_t = \phi_t = 0$, $t \geq 0$ and $\beta = 0$. Then, for $\alpha = 1$,

$$U_t^l(x) = \left(\log x - \frac{1}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds \right), \quad x > 0. \quad (5.11)$$

Moreover,

$$\pi_t^{*,l} = x \sigma_t^+ \lambda_t \exp \left(-\frac{1}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds + \int_0^t \sigma_s \sigma_s^+ \lambda_s \cdot (\lambda_s ds + dW_s) \right)$$

and

$$X_t^{*,l} = x \exp \left(-\frac{1}{2} \int_0^t |\sigma_s \sigma_s^+ \lambda_s|^2 ds + \int_0^t \sigma_s \sigma_s^+ \lambda_s \cdot (\lambda_s ds + dW_s) \right).$$

When the market coefficients are constants, the forward processes $U_t^e(x)$, $U_t^p(x)$ and $U_t^l(x)$ in (5.4), (5.10) and (5.11) reduce to deterministic functions. These special cases can be found in HENDERSON and HOBSON [2007a,b].

References

- BERRIER, F.P.Y.S., ROGERS, L.C.G., TEHRANCHI, M.R. (2007). A characterization of forward utility functions. Preprint.
- CHOULLI, T., STRICKER, C., LI, J. (2007). Minimal Hellinger martingale measures of order q . *Financ. Stoch.* **11** (3), 399–427.
- HENDERSON, V., HOBSON, D. (2007a). Horizon-unbiased utility functions. *Stoch. Proc. Appl.* **117** (11), 1621–1641.
- HENDERSON, V., HOBSON, D. (2007b). Valuing the option to invest in an incomplete market. To appear in *Math. Financ. Econ.*
- KURTZ, T. (1984). Martingale problems for controlled processes. In: Thoma, M., Wyner, A. (eds.), *Stochastic Modeling and Filtering*. In: Lecture Notes in Control and Information Sciences (Springer-Verlag), pp. 75–90.
- LARSON, R.E. (1968). *State Increment Dynamic Programming, Modern Analytic and Computational Methods in Science and Mathematics* (Elsevier).
- MUSIELA, M., ZARIPHPOULOU, T. (2003). Backward and forward utilities and the associated pricing systems: the case study of the binomial model. Preprint.
- MUSIELA, M., ZARIPHPOULOU, T. (2006a). Investments and forward utilities. Preprint.
- MUSIELA, M., ZARIPHPOULOU, T. (2006b). Optimal asset allocation under forward exponential criteria. *Markov Processes and Related Topics: A Festschrift for T. G. Kurtz* In: Lecture Notes–Monograph Series (Institute of Mathematical Statistics). In print.
- MUSIELA, M., ZARIPHPOULOU, T. (2007a). Investment and valuation under backward and forward dynamic exponential utilities in a stochastic factor model, *Dilip Madan's Festschrift*, pp. 303–334.
- MUSIELA, M., ZARIPHPOULOU, T. (2007b). Investment performance measurement, risk tolerance and optimal portfolio choice. Submitted for publication.
- NISIO, M. (1981). *Lectures on Stochastic Control Theory* In: ISI Lecture Notes **9** (Macmillan).
- PENROSE, R. (1955). A generalized inverse for matrices. In: *Proc. Camb. Philol. Soc.* **51**, 406–413.
- SEINFELD, J., LAPIDUS, L. (1968). Aspects of the forward dynamic programming algorithm. *Ind. Eng. Chem. Process Des. Develop.* **7** (3), 475–478.
- VASQUEZ, J.-L. (2006). *The Porous Medium Equation* (Oxford University Press).
- VIT, K. (1977). Forward differential dynamic programming, *J. Optim. Theory Appl.* **21** (4), 487–504.

Malliavin Calculus for Pure Jump Processes and Applications to Finance

Marie-Pierre Bavouzet

INRIA Rocquencourt, projet MATHFI, 78153 Le Chesnay cedex, France.

E-mail address: marie-pierre.bavouzet@inria.fr

Marouen Messaoud

IXIS, 47 quai d'Austerlitz, 75648 Paris cedex 13, France.

E-mail address: mmessaoud@ixis-cib.com

Vlad Bally

*Université de Marne-la-Vallée, laboratoire d'Analyse et de Mathématiques Appliquées,
5 bd Descartes, Cité Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cédex 2,
France.*

E-mail address: vlad.bally@univ-mlv.fr

Abstract

We settle an integration by parts formula of the Malliavin type in an abstract framework, and we apply it to jump-type market models. Then, we give numerical algorithms for sensitivity computations of European options and for pricing American options in a model driven by a compound Poisson process.

1. Introduction

Following the pioneering papers FOURNIÉ, LASRY, LEBOUCHOUX, LIONS and TOUZI [1999] and FOURNIÉ, LASRY, LEBOUCHOUX and LIONS [2001], a lot of work concerning the numerical applications of the stochastic variational calculus (Malliavin calculus) has been done. This mainly concerns applications in mathematical finance: computations of conditional expectations (which appear in the American option pricing) and

of sensitivities (the so-called Greeks). The models at hand are usually log-normal type diffusions, and then one may use the standard Malliavin calculus. But nowadays people are more and more interested in jump-type diffusions (see CONT and TANKOV [2003] for example), and then one has to use the stochastic variational calculus corresponding to Poisson point processes. Such a calculus has already been developed in by BICHTLER, GRAVEREAUX and JACOD [1987] concerning the noise coming from the amplitudes of the jumps and by CARLEN and PARDOUX [1990] concerning the jump times (see also DENIS [2000], PICARD [1996], PICARD [1996], PRIVAULT and WEI [2005], and PRIVAULT and WEI [2004] for more recent developments). Recently, BOULEAU [2003] settled the so-called error calculus based on the Dirichlet form language, and showed that the approaches in both BICHTLER, GRAVEREAUX and JACOD [1987] and CARLEN and PARDOUX [1990] fit in this frame. Moreover, much work concerning the applications in finance has been done: see DAVIS and JOHANSSON [2006], EL KHATIB and PRIVAULT [2004], FORSTER, LTKEBOHMERT and TEICHMANN [2005], and PRIVAULT and WEI [2004]. Another point of view based on chaos decomposition may be found in ØKSENDAL [1996], BIAGINI, ØKSENDAL, SULEM and WALLNER [2004], DI NUNNO, ØKSENDAL and PROSKE [2004], VIVES, LEÓN, UTZET and SOLÉ [2002], and NUALART and VIVES [1990].

In BALLY, BAVOUZET and MESSAOUD [2005], we gave a new approach to this problem. Roughly speaking, we consider functionals of the form $F = f(V_1, \dots, V_n)$, and we assume that the conditional law of V_i (with respect to V_j , $j \neq i$) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and has a density $p_i(\omega, y)$, which is piecewise differentiable with respect to y . Then, using standard integration by parts, we settle a duality formula that is analogous to the one in Malliavin calculus. Then, the standard machinery of Malliavin calculus produces an integration by parts formula, which may be used to compute the sensitivities of financial options. This is done in our previous study BALLY, BAVOUZET and MESSAOUD [2005]. In the framework of stochastic equations driven by a compound Poisson process, the variables V_i , $i = 1, \dots, n$ may be either the amplitudes of the jumps or the times at which the Poisson process jumps. Recently in a study by BAVOUZET [2006], the author uses the integration by parts formula to derive a representation theorem for conditional expectations and use it to price American options. This represents the analogous of the work done by Herve and Lions LIONS and REGNIER [2000] in the case of the Brownian motion. But Bavouzet uses the Malliavin calculus based on the jump amplitudes of a compound Poisson process.

In this chapter, we give a simplified and unitary presentation of the results derived BALLY, BAVOUZET and MESSAOUD [2005] and BAVOUZET [2006] as well as a new approach to the Greeks computations based on the idea of the Bismut–Elworthy formula. It turns out that this approach is much easier to handle than the one based directly on the Malliavin integration by parts formula. Our presentation focuses on algorithms so, we leave out some heavy technical points related to integrability problems, and we refer to BALLY, BAVOUZET and MESSAOUD [2005] and BAVOUZET [2006] for details.

The chapter is organized as follows. In Section 2, we give the general presentation of the Malliavin calculus in an abstract framework. We introduce the differential operators and derive the duality and the integration by parts formula. We stay in a finite dimensional setting that is enough for numerical applications. In Section 3, we consider jump-type diffusions, and we specify the calculus associated with, the amplitude of the jumps

corresponding to the jump times. In Section 4, we present the algorithms for the delta computation of European options when the underlying asset follows two specific models. Our approach is based on the integration by parts formula using the jump amplitudes with respect to the jump times. Finally, we give numerical experiments, which lead to conclusions similar to those for the standard models based on the Brownian motion. For smooth payoffs (as the call option price, for example), the estimators based on the Malliavin approach and finite differences are very close. But for payoffs with singularities (as digital options), the Malliavin approach is much more efficient. It also turns out that it is crucial to use variance reduction techniques based on localization. Finally, we conclude that the calculus based on the amplitudes of the jumps is generally more efficient than the one based on the jump times. Roughly speaking, there is more noise available in the first case (at least in our choice of parameters). In Section 5, we present the approach based on the Bismut-Elworthy formula, which turns out to be much more simpler and easier to implement. In Section 6, we briefly present the algorithm for pricing American options and a few numerical experiments. We refer to BAVOUZET and MESSAOUD [2006] for a more detailed presentation including sensitivity computations for American options.

2. Malliavin calculus for simple functionals

2.1. The framework

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sub σ -algebra $\mathcal{G} \subseteq \mathcal{F}$ and a sequence of random variables $V_i, i \in \mathbb{N}$. We denote $\mathcal{G}_i = \mathcal{G} \vee \sigma(V_j, j \neq i)$. Our aim is to settle an integration by parts formula for functionals of $V_i, i \in \mathbb{N}$, which is analogous to the one in the standard Malliavin calculus. The σ -algebra $\mathcal{G}$ appears to describe all the randomness that is not involved in the differential calculus.

We work on a set $A \in \mathcal{G}$, which will be fixed through this section.

For each $i \in \mathbb{N}$, we consider some $\mathcal{G}_i$ -measurable random variables $a_i(\omega) < b_i(\omega)$, and we denote $B_i(\omega) = (a_i(\omega), b_i(\omega))$. Note that we may take $a_i = -\infty$ and $b_i = +\infty$. We work with functions $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ for some $n \in \mathbb{N}$, and we denote

$$I_i(f)(\omega, y) := f(\omega, V_1, \dots, V_{i-1}, y, V_{i+1}, \dots, V_n). \quad (2.1)$$

Given $n, k \in \mathbb{N}$, we denote by $C_{n,k}$ the class of functions $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that $(y \rightarrow I_i(f)(\omega, y))$ is k times continuously differentiable on $B_i(\omega), i = 1, \dots, n$, and such that the left-hand side and the right-hand side limits of $I_i(f)$ in a_i and b_i exist and are finite, that is,

$$I_i(f)(\omega, a_i+) < \infty \text{ and } I_i(f)(\omega, b_i-) < \infty.$$

In the case $k = 0$, we just assume continuity.

Our basic hypothesis is the following one.

HYPOTHESIS 2.1. *For every $i \in \mathbb{N}$, the conditional law of V_i , given $\mathcal{G}_i$, is absolutely continuous on (a_i, b_i) with respect to the Lebesgue measure. This means that there exists*

a $\mathcal{G}_i \times B(\mathbb{R})$ -measurable function $p_i = p_i(\omega, x)$ such that

$$\mathbb{E}(\Theta \psi(V_i) \mathbf{1}_{(a_i, b_i)}(V_i)) = \mathbb{E} \left(\Theta \int_{\mathbb{R}} \psi(x) p_i(\omega, x) \mathbf{1}_{(a_i, b_i)}(x) dx \right),$$

for every positive $\mathcal{G}_i$ -measurable random variable Θ and every positive and measurable function $\psi : \mathbb{R} \rightarrow \mathbb{R}$.

We assume that

$$p_i \in C_{1,1} \text{ and for all } p \in \mathbb{N}, \mathbb{E} \left[\left| \partial_y \ln p_i(\omega, y) \right|^p \mathbf{1}_A \right] < \infty.$$

In our calculus, we use some weights $(\pi_i)_{i \in \mathbb{N}}$ that we define in the following. For each $i \in \mathbb{N}$, we consider a $\mathcal{G}_i \times B(\mathbb{R})$ -measurable and positive function $\pi_i : \Omega \times \mathbb{R} \rightarrow \mathbb{R}_+$ such that $\pi_i \in C_{1,1}$. We assume the following hypothesis:

HYPOTHESIS 2.2.

- (i) $\pi_i(\omega, y) \mathbf{1}_{(a_i, b_i)^c}(y) = 0$,
- (ii) $\lim_{y \downarrow a_i} \pi_i(\omega, y) = \lim_{y \uparrow b_i} \pi_i(\omega, y) = 0$.

Hypothesis 2.2 (ii) is the reason of being of the weights $(\pi_i)_{i \in \mathbb{N}}$: they are used to cancel the border terms in a_i and b_i of the integration by parts formula. The typical example of weights is

$$\pi_i(\omega, y) = (y - a_i(\omega))^\theta (b_i(\omega) - y)^\theta \mathbf{1}_{(a_i(\omega), b_i(\omega))}(y), \quad (2.2)$$

with $\theta > 0$. In concrete examples, we must also assume that $\theta < 1/2$, if not, the inverse of the Malliavin covariance matrix (see the following section) does not verify suitable integrability conditions.

Note that if p_i is differentiable on the whole $\mathbb{R}$, then we may take $a_i = -\infty$, $b_i = +\infty$ and $B_i = \mathbb{R}$. Thus, we may choose the weights $\pi_i \equiv 1$ (see BAVOUZET and MESSAOUD [2006]).

2.2. The differential operators

In this section, we introduce the differential operators that represent the analogous of the Malliavin derivative and the Skorohod integral.

Simple functionals. A random variable F is called a simple functional if there exists some $n \in \mathbb{N}^*$ and some $\mathcal{G} \times \mathcal{B}(\mathbb{R}^n)$ -measurable function $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$F = f(\omega, V_1, \dots, V_n).$$

We denote by $\mathcal{S}_{(n,k)}$ the space of the simple functionals such that $f \in C_{n,k}$. Simple processes. A simple process of length n is a finite sequence of random variables

$U = (U_i)_{i \leq n}$ such that

$$U_i(\omega) = u_i(\omega, V_1(\omega), \dots, V_n(\omega)),$$

where $u_i : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$, $i \in \mathbb{N}$ are $\mathcal{G} \times \mathcal{B}(\mathbb{R}^n)$ -measurable functions. We denote by $\mathcal{P}_{(n,k)}$ the space of the simple processes of length n such that $u_i \in C_{n,k}$, $i = 1, \dots, n$. Note that if $U \in \mathcal{P}_{(n,k)}$, then $U_i \in \mathcal{S}_{(n,k)}$.

On the space of simple processes, we consider the inner product associated with the weights $(\pi_i)_{i \in \mathbb{N}}$:

$$\langle U, V \rangle_\pi := \sum_{i=1}^n \pi_i(\omega, V_i) U_i(\omega) V_i(\omega).$$

We define now the differential operators.

- **The Malliavin derivative** $D : \mathcal{S}_{(n,1)} \rightarrow \mathcal{P}_{(n,0)}$.

If $F = f(\omega, V_1, \dots, V_n)$, then

$$D_i F := \frac{\partial f}{\partial x_i}(\omega, V_1(\omega), \dots, V_n(\omega)),$$

$$DF = (D_i F)_{i \leq n} \in \mathcal{P}_{(n,0)}.$$

- **The Malliavin covariance matrix.**

Given $F = (F^1, \dots, F^d)$, $F^i = f^i(\omega, V_1, \dots, V_n) \in \mathcal{S}_{(n,1)}$, the Malliavin covariance matrix is

$$\sigma_{\pi, F}^{ij} = \langle DF^i, DF^j \rangle_\pi = \sum_{p=1}^n \pi_p(\omega, V_p) \partial_p f^i \partial_p f^j(\omega, V_1, \dots, V_n).$$

This is a symmetric positive definite matrix.

- **The Skorohod integral (Divergence type operator).**

We define $\delta_\pi : \mathcal{P}_{(n,1)} \rightarrow \mathcal{S}_{(n,0)}$ by

$$\delta_{i,\pi}(U) := - \left(\frac{\partial}{\partial x_i} (\pi_i u_i) + (\pi_i u_i) \partial \ln p_i \right) (\omega, V_1, \dots, V_n), \quad i = 1, \dots, n$$

$$\delta_\pi(U) := \sum_{i=1}^n \delta_{i,\pi}(U).$$

In our framework, the duality between δ_π and D is given by the following proposition.

PROPOSITION 2.1. *Let $F \in \mathcal{S}_{(n,1)}$ and $U \in \mathcal{P}_{(n,1)}$. Suppose that for every $i = 1, \dots, n$*

$$\mathbb{E} \left(|F \delta_{i,\pi}(U)| \mathbf{1}_A \right) + \mathbb{E}(\pi_i(\omega, V_i) |D_i F \times U_i| \mathbf{1}_A) < \infty. \quad (2.3)$$

Then,

$$\mathbb{E}(\langle DF, U \rangle_\pi \mathbf{1}_A) = \mathbb{E}(F \delta_\pi(U) \mathbf{1}_A). \quad (2.4)$$

PROOF. One writes

$$\begin{aligned} & \mathbb{E}(\langle DF, U \rangle_\pi \mathbf{1}_A) \\ &= \mathbb{E} \left(\sum_{i=1}^n \mathbb{E}(\pi_i(\omega, V_i) D_i F \times U_i \mid \mathcal{G}_i) \mathbf{1}_A \right) \\ &= \mathbb{E} \left(\mathbf{1}_A \sum_{i=1}^n \int_{\mathbb{R}} (\pi_i u_i \partial_i f)(\omega, V_1, \dots, V_{i-1}, y, V_{i+1}, \dots, V_n) p_i(\omega, y) dy \right). \end{aligned}$$

Using integration by parts and Hypothesis 2.2, in particular, $\pi_i = 0$ on $(a_i, b_i)^c$ and $\lim_{y \downarrow a_i} \pi_i(\omega, y) = \lim_{y \uparrow b_i} \pi_i(\omega, y) = 0$, we obtain

$$\begin{aligned} \int_{\mathbb{R}} \partial_i f \times (\pi_i u_i) \times p_i &= \int_{a_i}^{b_i} \partial_i f \times (\pi_i u_i) \times p_i \\ &= - \int_{a_i}^{b_i} f \times (\partial_i(\pi_i u_i) \times p_i + (\pi_i u_i) \times \partial p_i) \\ &= - \int_{\mathbb{R}} f \times (\partial_i(\pi_i u_i) + \pi_i u_i \partial \ln p_i) \times p_i. \end{aligned}$$

By Hypothesis (2.3), we have for almost every $\omega \in A$,

$$\begin{aligned} \int_{\mathbb{R}} (|u_i \partial_i f| \pi_i p_i)(\omega, V_1, \dots, V_{i-1}, y, V_{i+1}, \dots, V_n) dy &< \infty, \\ \int_{\mathbb{R}} (|f (\partial_i(\pi_i u_i) + \pi_i u_i \partial \ln p_i)| \times p_i)(\omega, V_1, \dots, V_{i-1}, y, V_{i+1}, \dots, V_n) dy &< \infty, \end{aligned}$$

so the above integrals make sense. Using the definition of p_i , we come back to expectations and we obtain

$$\int_{\mathbb{R}} (\pi_i u_i \partial_i f)(\omega, V_1, \dots, V_{i-1}, y, V_{i+1}, \dots, V_n) p_i(\omega, y) dy = \mathbb{E}(F \delta_{i,\pi}(U_i) \mid \mathcal{G}_i).$$

Summing over i , the proof is complete. $\square$

Let us finally introduce

- **The Ornstein Uhlenbeck operator** $L_\pi := \delta_\pi(D) : \mathcal{S}_{(n,2)} \rightarrow \mathcal{S}_{(n,0)}$.
We define

$$L_\pi := \sum_{i=1}^n L_{i,\pi},$$

where

$$\begin{aligned} L_{i,\pi} F &:= -(\partial_i(\pi_i \partial_i f) + \pi_i \partial_i f \partial \ln p_i)(\omega, V_1, \dots, V_n) \\ &= -((\pi'_i + \pi_i \partial \ln p_i) \partial_i f + \pi_i \partial_i^2 f)(\omega, V_1, \dots, V_n), \end{aligned}$$

We denote by $C_p^k(\mathbb{R}^d)$ the space of functions $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$, which are k times differentiable such that ϕ and its derivatives up to order k have polynomial growth. The standard differential calculus gives the following chain rules.

LEMMA 2.1.

i) Let $\phi \in C_p^1(\mathbb{R}^d)$ and $F = (F^1, \dots, F^d)$, $F^i \in \mathcal{S}_{(n,1)}$. Then, $\phi(F) \in \mathcal{S}_{(n,1)}$ and

$$D\phi(F) = \sum_{k=1}^d \partial_k \phi(F) DF^k. \quad (2.5)$$

ii) If $\phi \in C_p^2(\mathbb{R}^d)$ and $F^i \in \mathcal{S}_{(n,2)}$, then $\phi(F) \in \mathcal{S}_{(n,2)}$ and

$$L_\pi \phi(F) = \sum_{k=1}^d \partial_k \phi(F) L_\pi F^k - \sum_{k,p=1}^d \partial_{k,p}^2 \phi(F) \left\langle DF^k, DF^p \right\rangle_\pi.$$

iii) Let $F \in \mathcal{S}_{(n,1)}$ and $U \in \mathcal{P}_{(n,1)}$. Then, $FU \in \mathcal{P}_{(n,1)}$ and

$$\delta_\pi(FU) = F \delta_\pi(U) - \langle DF, U \rangle_\pi. \quad (2.6)$$

In particular, if $F \in \mathcal{S}_{(n,1)}$ and $G \in \mathcal{S}_{(n,2)}$, then $FDG \in \mathcal{P}_{(n,1)}$ and

$$\delta_\pi(FDG) = F L_\pi G - \langle DF, DG \rangle_\pi. \quad (2.7)$$

2.3. The integration by parts formula

The basic integration by parts formula is the following.

THEOREM 2.1. Let $F = (F^1, \dots, F^d) \in \mathcal{S}_{(n,2)}^d$ and $G \in \mathcal{S}_{(n,1)}$.

Suppose that for all $\omega \in A$, the covariance matrix $\sigma_{\pi,F}(\omega)$ is invertible and denote $\gamma_{\pi,F} := \sigma_{\pi,F}^{-1}$.

We assume that for all $k = 1, \dots, n$ and $i, j, l = 1, \dots, d$,

$$\mathbb{E} \left[\mathbf{1}_A \left(\left| \phi(F) \delta_{k,\pi}(G \gamma_{\pi,F}^{ji} DF^j) \right| + \pi_i(\omega, V_i) \left| G \gamma_{\pi,F}^{ji} D_k F^j D_k F^l \right| \right) \right] < \infty. \quad (2.8)$$

Then, for every $\phi \in \mathcal{C}_p^1(\mathbb{R}^d)$ and for every $i = 1, \dots, d$, one has

$$\begin{aligned} \mathbb{E}(\partial_i \phi(F) G \mathbf{1}_A) &= \mathbb{E} \left[\phi(F) \delta_\pi \left(G \sum_{j=1}^d \gamma_{\pi, F}^{ji} DF^j \right) \mathbf{1}_A \right] \\ &= \mathbb{E}(\phi(F) H_{i, \pi}(F, G) \mathbf{1}_A), \end{aligned} \quad (2.9)$$

with

$$H_{i, \pi}(F, G) = \sum_{j=1}^d \left(G \gamma_{\pi, F}^{ji} L_\pi F^j - \left\langle D(G \gamma_{\pi, F}^{ji}), DF^j \right\rangle_\pi \right). \quad (2.10)$$

PROOF. Using the chain rule (2.5), we get for all $j = 1, \dots, d$

$$\begin{aligned} \langle D\phi(F), DF^j \rangle_\pi &= \sum_{r=1}^n \pi_r(\omega, V_r) D_r \phi(F) D_r F^j \\ &= \sum_{r=1}^n \sum_{i=1}^d \partial_i \phi(F) \pi_r(\omega, V_r) D_r F^i D_r F^j \\ &= \sum_{i=1}^d \partial_i \phi(F) \sigma_{\pi, F}^{ij}, \end{aligned}$$

$$\text{so that } \partial_i \phi(F) = \sum_{j=1}^d \langle D\phi(F), DF^j \rangle_\pi \gamma_{\pi, F}^{ji}.$$

Under the integrability condition (2.8), we can use the duality relation (2.4) to obtain

$$\begin{aligned} \mathbb{E}(\partial_i \phi(F) G \mathbf{1}_A) &= \sum_{j=1}^d \mathbb{E} \left[\langle D\phi(F), DF^j \rangle_\pi \gamma_{\pi, F}^{ji} \mathbf{1}_A \right] \\ &= \sum_{j=1}^d \mathbb{E} \left[\phi(F) \delta_\pi (G \gamma_{\pi, F}^{ji} DF^j) \mathbf{1}_A \right]. \end{aligned}$$

□

Let us give sufficient conditions that imply that the integrability assumption (2.8) is satisfied. Suppose that for all $i, p \in \mathbb{N}$,

$$\mathbb{E} \left[\mathbf{1}_A (|F|^p + |\partial_y \ln p_i(\omega, V_i)|^p + |\pi_i(\omega, V_i)|^p) \right] < \infty, \quad \mathbb{E} \left[\mathbf{1}_A |\pi'_i(\omega, V_i)| \right] < \infty.$$

Assume also that

$$\mathbb{E} \left[\mathbf{1}_A (\det \gamma_{\pi, F})^2 (1 + |\pi'_l|) \right] < \infty. \quad (2.11)$$

Then, the integrability condition (2.8) holds true (we refer to BALLY, BAVOUZET and MESSAOUD [2005] for the proof).

3. Integration by parts formula for pure jump processes

In this section, we apply the integration by parts formula (2.9) to a one-dimensional pure jump diffusion process $(S_t)_{t \in [0, T]}$. Let us precise the model. We construct a compound Poisson process in the following way:

Consider two sequences of independent random variables $(\tau_k)_{k \in \mathbb{N}}$ and $(\Delta_k)_{k \in \mathbb{N}}$ such that for all $k \in \mathbb{N}$, τ_k is exponentially distributed for parameter λ and Δ_k has a law denoted by $\nu(da)$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Define $T_k = \tau_1 + \dots + \tau_k$ and $J_t = \text{Card}\{k : T_k \leq t\}$, which is a Poisson process of intensity λ .

Thus, define the so-called counting measure $N(dt, da)$ on $(\mathbb{R}_+ \times \mathbb{R}, \mathcal{B}(\mathbb{R}_+) \times \mathcal{B}(\mathbb{R}))$ by

$$N((0, t] \times A) = \text{Card}\{k : T_k \leq t, \Delta_k \in A\}.$$

For any measurable function $f : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$, the integral with respect to this measure is given by

$$\int_0^t \int_{\mathbb{R}} f(u, a) N(du, da) = \sum_{T_k \leq t} f(T_k, \Delta_k). \quad (3.1)$$

The measure $N(dt, da)$ is a Poisson point measure, and the stochastic calculus associated with such measures may be found in IKEDA and WATANABE [1989]. But in our framework, we just use elementary properties of the integral (3.1).

We look at $(S_t)_{t \in [0, T]}$ solution of the equation

$$\begin{aligned} S_t &= x + \sum_{i=1}^{J_t} c(T_i, \Delta_i, S_{T_i}^-) + \int_0^t g(r, S_r) dr \\ &= x + \int_0^t \int_{\mathbb{R}} c(s, a, S_{s-}) N(ds, da) + \int_0^t g(r, S_r) dr, \quad 0 \leq t \leq T. \end{aligned} \quad (3.2)$$

We work under the following hypothesis:

HYPOTHESIS 3.1. *The functions $(a, x) \rightarrow c(t, a, x)$ and $x \rightarrow g(t, x)$ are twice differentiable and have bounded derivatives of first and second orders. The function $t \rightarrow c(t, a, x)$ is differentiable with bounded derivative.*

Moreover, we assume that there exists a positive constant K such that

- i) $|c(t, a, x) - c(u, a, y)| \leq K (|t - u| + |x - y|),$
- ii) $|g(t, x) - g(u, y)| \leq K (|t - u| + |x - y|),$
- iii) $|c(t, a, x)| + |g(t, x)| \leq K (1 + |x|).$

As we mentioned in Introduction, in this framework, we may use Malliavin calculus with respect to the jump amplitudes $(\Delta_k)_{k \in \mathbb{N}}$ or to the jump times $(T_k)_{k \in \mathbb{N}}$. Let us first introduce a deterministic calculus that allows us to express S_t as a simple functional and to compute its Malliavin derivatives.

3.1. The deterministic equation

We fix some deterministic times $0 = u_0 < u_1 < \dots < u_n < T$, and we denote $u = (u_1, \dots, u_n)$ and $J_t(u) = k$ if $u_k \leq t < u_{k+1}$. They represent the jump times. We also fix a vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, which represents the amplitudes of the jumps. To these fixed vectors, we associate the deterministic equation

$$s_t = x + \sum_{i=1}^{J_t(u)} c(u_i, a_i, s_{u_i}^-) + \int_0^t g(r, s_r) dr, \quad 0 \leq t \leq T. \quad (3.3)$$

We denote by $s_t(u, a)$ or simply by s_t the solution of this equation. This is the deterministic counterpart of the stochastic Eq. (3.2).

For all $t \in [0, T]$, for all $n \geq 1$, on the set $\{J_t = n\}$, the solution S_t of (3.2) is represented as

$$S_t = s_t(T_1, \dots, T_n, \Delta_1, \dots, \Delta_n).$$

In order to solve (3.3), we introduce the flow $\Phi = \Phi_u(t, x)$, $0 \leq u \leq t$, $x \in \mathbb{R}$, solution of the ordinary integral equation

$$\Phi_u(t, x) = x + \int_u^t g(r, \Phi_u(r, x)) dr, \quad t \geq u.$$

The solution s of Eq. (3.3) is given by

$$\begin{aligned} s_0 &= x, \\ s_t &= \Phi_{u_i}(t, s_{u_i}) \text{ for } u_i \leq t < u_{i+1}, \\ s_{u_{i+1}} &= s_{u_{i+1}}^- + c(u_{i+1}, a_{i+1}, s_{u_{i+1}}^-) \\ &= \Phi_{u_i}(u_{i+1}, s_{u_i}) + c(u_{i+1}, a_{i+1}, \Phi_{u_i}(u_{i+1}, s_{u_i})). \end{aligned} \quad (3.4)$$

Our aim is to compute the derivatives of s with respect to u_j and a_j . We first introduce some notations.

We denote

$$e_{u,t}(x) := \exp\left(\int_u^t \partial_x g(r, \Phi_u(r, x)) dr\right).$$

Since $\Phi_{u_i}(r, s_{u_i}) = s_r$ for $u_i \leq r < u_{i+1}$, we have

$$e_{u_i,t}(s_{u_i}) = \exp\left(\int_{u_i}^t \partial_x g(r, s_r) dr\right), \quad \text{for } u_i \leq t < u_{i+1}.$$

Since

$$\partial_x \Phi_u(t, x) = 1 + \int_u^t \partial_x g(r, \Phi_u(r, x)) \partial_x \Phi_u(r, x) dr,$$

it follows that

$$\partial_x \Phi_u(t, x) = e_{u,t}(x).$$

And since

$$\partial_u \Phi_u(t, x) = -g(u, x) + \int_u^t \partial_x g(r, \Phi_u(r, x)) \partial_u \Phi_u(r, x) dr,$$

we have

$$\partial_u \Phi_u(t, x) = -g(u, x) e_{u,t}(x).$$

We finally denote

$$q(t, \alpha, x) := (\partial_t c + g \partial_x c)(t, \alpha, x) + g(t, x) - g(t, x + c(t, \alpha, x)).$$

LEMMA 3.1. *Suppose that Hypothesis 3.1 holds true. Then, $s_t(u, a)$ is twice differentiable with respect to u_j and a_j , and we have explicit expressions of the derivatives.*

A. Derivatives with respect to u_j

For $t < u_j$, $\partial_{u_j} s_t(u, a) = 0$. Moreover,

$$\begin{aligned} \partial_{u_j} s_{u_j-} &= g(u_j, s_{u_j-}), \\ \partial_{u_j} s_{u_j} &= (\partial_t c + g(1 + \partial_x c))(u_j, a_j, s_{u_j-}). \end{aligned}$$

For $u_j < t < u_{j+1}$,

$$\begin{aligned} \partial_{u_j} s_t &= q(u_j, a_j, s_{u_j-}) e_{u_j,t}(s_{u_j}), \\ \partial_{u_j} s_{u_{j+1}-} &= q(u_j, a_j, s_{u_j-}) e_{u_j, u_{j+1}}(s_{u_j}) \\ \partial_{u_j} s_{u_{j+1}} &= q(u_j, a_j, s_{u_j-}) (1 + \partial_x c(u_{j+1}, a_{j+1}, s_{u_{j+1}-})) e_{u_j, u_{j+1}}(s_{u_j}). \end{aligned} \tag{3.5}$$

Finally, for $p \geq j+1$ and $u_p \leq t < u_{p+1}$, we have the recurrence relations

$$\begin{aligned} \partial_{u_j} s_t &= e_{u_p,t}(s_{u_p}) \partial_{u_j} s_{u_p}, \\ \partial_{u_j} s_{u_{p+1}} &= (1 + \partial_x c(u_{p+1}, a_{p+1}, s_{u_{p+1}-})) e_{u_p, u_{p+1}}(s_{u_p}) \partial_{u_j} s_{u_p}. \end{aligned} \tag{3.6}$$

Let us denote $T(f) := \partial_t f + g \partial_x f$. The second-order derivatives are given by

$$\begin{aligned} \partial_{u_j}^2 s_{u_j-} &= T(g)(u_j, a_j, s_{u_j-}), \\ \partial_{u_j}^2 s_{u_j} &= T(\partial_t c + g(1 + \partial_x c))(u_j, a_j, s_{u_j-}). \end{aligned}$$

We denote

$$\begin{aligned}\rho_j(t) &= \partial_{u_j} e_{u_j, t}(s_{u_j}) \\ &= e_{u_j, t}(s_{u_j}) \left(-\partial_x g(u_j, s_{u_j}) + q(u_j, a_j, s_{u_j-}) \int_{u_j}^t \partial_x^2 g(r, s_r) e_{u_j, r}(s_{u_j}) dr \right).\end{aligned}$$

Then, for $u_j < t < u_{j+1}$,

$$\partial_{u_j}^2 s_t(u, a) = T(q)(u_j, a_j, s_{u_j-}(u, a)) e_{u_j, t}(s_{u_j}) + q(u_j, a_j, s_{u_j-}(u, a)) \rho_j(t),$$

and

$$\begin{aligned}\partial_{u_j}^2 s_{u_{j+1}} &= T(q)(u_j, a_j, s_{u_j-}) (1 + \partial_x c)(u_{j+1}, a_{j+1}, s_{u_{j+1}-}) e_{u_j, u_{j+1}}(s_{u_j}) \\ &\quad + q^2(u_j, a_j, s_{u_j-}) \partial_x^2 c(u_{j+1}, a_{j+1}, s_{u_{j+1}-}) e_{u_j, u_{j+1}}^2(s_{u_j}) \\ &\quad + q(u_j, a_j, s_{u_j-}) (1 + \partial_x c)(u_{j+1}, a_{j+1}, s_{u_{j+1}-}) \rho_j(u_j).\end{aligned}$$

For $p \geq j+1$, we denote

$$\rho_{j, p}(t) = \partial_{u_j} e_{u_p, t}(s_{u_p}) = e_{u_p, t}(s_{u_p}) \partial_{u_j} s_{u_p} \int_{u_p}^t \partial_x^2 g(r, s_r) e_{u_p, r}(s_{u_p}) dr.$$

Then, for $p \geq j$ and $u_p \leq t < u_{p+1}$, we have the recurrence relations

$$\begin{aligned}\partial_{u_j}^2 s_t &= e_{u_p, t}(s_{u_p}) \partial_{u_j}^2 s_{u_p} + \rho_{j, p}(t, u, a) \partial_{u_j} s_{u_p}, \\ \partial_{u_j}^2 s_{u_{p+1}} &= \partial_x^2 c(u_{p+1}, a_{p+1}, s_{u_{p+1}-}) (e_{u_p, u_{p+1}}(s_{u_p}) \partial_{u_j} s_{u_p})^2 \\ &\quad + (1 + \partial_x c)(u_{p+1}, a_{p+1}, s_{u_{p+1}-}) (\rho_{j, p}(u_{p+1}) \partial_{u_j} s_{u_p} \\ &\quad + e_{u_p, u_{p+1}}(s_{u_p}) \partial_{u_j}^2 s_{u_p}).\end{aligned}$$

B. Derivatives with respect to a_j

For $t < u_j$, $\partial_{a_j} s_{u_j}(u, a) = 0$ and for $t \geq u_j$, $\partial_{a_j} s_t(u, a)$ satisfies the following equation:

$$\begin{aligned}\partial_{a_j} s_t &= \partial_a c(u_j, a_j, s_{u_j-}) + \sum_{i=j+1}^{J_t(u)} \partial_x c(u_i, a_i, s_{u_i-}) \partial_{a_j} s_{u_i-} \\ &\quad + \int_{u_j}^t \partial_x g(r, s_r) \partial_{a_j} s_r dr.\end{aligned}\tag{3.7}$$

The second-order derivatives are given by

$$\begin{aligned} \partial_{a_j}^2 s_t &= \partial_a^2 c(u_j, a_j, s_{u_j-}) + \sum_{i=j+1}^{J_t(u)} \partial_x^2 c(u_i, a_i, s_{u_i-}) (\partial_{a_j} s_{u_i-})^2 \\ &\quad + \int_{u_j}^t \partial_x^2 g(r, s_r) (\partial_{a_j} s_r)^2 dr \\ &\quad + \sum_{i=j+1}^{J_t(u)} \partial_x c(u_i, a_i, s_{u_i-}) \partial_{a_j}^2 s_{u_i-} + \int_{u_j}^t \partial_x g(r, s_r) \partial_{a_j}^2 s_r dr, \end{aligned} \quad (3.8)$$

and for $i < j$,

$$\begin{aligned} \partial_{a_j, a_i}^2 s_t &= \partial_{a, x}^2 c(u_j, a_j, s_{u_j-}) + \sum_{k=j+1}^{J_t(u)} \partial_x^2 c(u_k, a_k, s_{u_k-}) \partial_{a_i} s_{u_k-} \partial_{a_j} s_{u_k-} \\ &\quad + \sum_{k=j+1}^{J_t(u)} \partial_x c(u_k, a_k, s_{u_k-}) \partial_{a_j, a_i}^2 s_{u_k-} + \int_{u_j}^t \partial_x g(r, s_r) \partial_{a_j, a_i}^2 s_r dr \\ &\quad + \int_{u_k}^t \partial_x^2 g(r, s_r) \partial_{a_i} s_r \partial_{a_j} s_r dr. \end{aligned}$$

For $i > j$, we derive $\partial_{a_j, a_i}^2 s_t$ by symmetry.

PROOF. We refer to BALLY, BAVOUZET and MESSAOUD [2005] for detailed computations. $\square$

As an immediate consequence of Lemma 3.1, we obtain the following upper bound.

COROLLARY 3.1. *Suppose that Hypothesis 3.1 holds true and that the starting point x satisfies $|x| \leq K$, for some $K > 0$.*

Then, for each $n \in \mathbb{N}$ and $T > 0$, there exists a constant $C_n(K, T)$ such that for every $0 < u_1 < \dots < u_n < T$, $a \in \mathbb{R}^n$ and $0 \leq t \leq T$,

$$\max_{j=1, \dots, n} \left(|s_t| + |\partial_{u_j} s_t| + \left| \partial_{u_j}^2 s_t \right| + |\partial_{a_j} s_t| + \left| \partial_{a_j}^2 s_t \right| \right) (u, a) \leq C_n(K, T). \quad (3.9)$$

3.2. Integration by parts formula using the jump amplitudes

In this section, we look at S_t as a simple functional of the jump amplitudes Δ_i , $i \in \mathbb{N}$. Using the notation of Section 2, this means that $V_i = \Delta_i$, $\mathcal{G} = \sigma\{T_i : i \in \mathbb{N}\}$, and on $A := \{J_t = n\}$, $n \geq 1$. If $\omega \in A$, then we have

$$S_t = s_t(T_1(\omega), \dots, T_n(\omega), \Delta_1, \dots, \Delta_n),$$

where s_t is defined by (3.3).

We consider some $\alpha < \beta$, and we denote $I = (\alpha, \beta)$. Note that we may take $\alpha = -\infty$ and $\beta = +\infty$.

HYPOTHESIS 3.2. *The law of Δ_i is absolutely continuous on I with respect to the Lebesgue measure and has the density $p(y) = e^{\rho(y)} \mathbf{1}_{(\alpha, \beta)}(y)$, that is,*

$$\mathbb{E}(f(\Delta_i) \mathbf{1}_I(\Delta_i)) = \int_I f(y) e^{\rho(y)} dy,$$

for every measurable and positive function f .

The function ρ is assumed to be continuously differentiable and bounded on I .

Since p has discontinuities in α and β , we work with the following weights. Take $\delta \in (0, 1)$ and for $0 \leq s < t \leq T$, define

$$\pi_{(s,t)}^i(\omega, \Delta_i) := \mathbf{1}_{[s,t]}(T_i(\omega)) \pi(\Delta_i), \quad (3.10)$$

with

$$\pi(y) = \begin{cases} (\beta - y)^\delta (y - \alpha)^\delta, & \text{for } y \in (\alpha, \beta), \\ 0, & \text{for } y \in (\alpha, \beta)^c. \end{cases}$$

Note that the indicative function $\mathbf{1}_{[s,t]}(T_i)$ allows us to settle a Malliavin calculus, which involves the jumps occurring between s and t only. In the case $s = 0$ and $t = T$, we thus use all the jump amplitudes Δ_i , $i \in \mathbb{N}$ of $[0, T]$.

EXAMPLE 3.1. Let us give three examples to illustrate how the weights are chosen.

1. Δ_i has a uniform law on $(0, 1)$.

This means that $I = (0, 1)$ and $p(y) = 1$ for all $y \in (0, 1)$. Then, we take $\delta \in (0, 1)$ and

$$\pi(y) = \begin{cases} (1 - y)^\delta y^\delta, & \text{for } y \in (0, 1), \\ 0, & \text{for } y \in (0, 1)^c. \end{cases}$$

2. Δ_i is a standard Gaussian random variable.

This means that $I = \mathbb{R}$ ($\alpha = -\infty$ and $\beta = +\infty$) and $p(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ for all $y \in \mathbb{R}$. Since p is differentiable on the whole $\mathbb{R}$, we take $\pi(y) = 1$ for all $y \in \mathbb{R}$.

3. Δ_i has an exponential law.

This means that $I = (0, \infty)$ and $p(y) = e^{-y}$ for all $y > 0$. Then, we take $\delta \in (0, 1)$ and $\beta > \delta$, and we set

$$\pi(y) = \begin{cases} y^{-\beta} y^\delta, & \text{for } y > 0, \\ 0, & \text{for } y \leq 0. \end{cases}$$

Let $A := \{J_t = n\}$, $n \geq 1$. Since $\delta \in (0, 1)$, elementary computations show that $(\pi_{(s,t)}^i)_{i \in \mathbb{N}}$ satisfies Hypothesis 2.2. Assuming Hypothesis 3.1, (3.9) implies that $(a_1, \dots, a_n) \rightarrow s_t(T_1(\omega), \dots, T_n(\omega), a_1, \dots, a_n)$ is twice continuously differentiable and has bounded derivatives. It follows that for all $t \in [0, T]$, S_t is a twice differentiable simple functional such that S_t and its first and second derivatives have finite moments of any order on A .

In the following, we use the notation (s, t) to indicate that the Malliavin operators are associated with the inner product $\langle \cdot, \cdot \rangle_{\pi_{(s,t)}}$. The differential operators that appear in the integration by parts formula are

$$D_i S_t = \partial_{a_i} s_t(T_1, \dots, T_n, \Delta_1, \dots, \Delta_n), \quad (3.11)$$

$$\begin{aligned} L_{(s,t)} S_t = & - \sum_{i=1}^n \mathbf{1}_{[s,t]}(T_i(\omega)) \times \left(\pi(\Delta_i) \partial_{a_i}^2 s_t(T_1, \dots, T_n, \Delta_1, \dots, \Delta_n) \right. \\ & \left. + (\pi' + \pi \frac{\rho'}{\rho})(\Delta_i) \partial_{a_i} s_t(T_1, \dots, T_n, \Delta_1, \dots, \Delta_n) \right), \end{aligned}$$

$$\begin{aligned} \sigma_t^{(s,t)} := \sigma_{\pi_{(s,t)}, S_t} &= \sum_{i=1}^n \mathbf{1}_{[s,t]}(T_i(\omega)) \pi(\Delta_i) |D_i S_t|^2 \\ &= \sum_{i=1}^n \mathbf{1}_{[s,t]}(T_i(\omega)) \pi(\Delta_i) |\partial_{a_i} s_t(T_1, \dots, T_n, \Delta_1, \dots, \Delta_n)|^2. \end{aligned}$$

All these quantities may be computed using (3.7) and (3.8).

Let us give sufficient conditions of ellipticity type to obtain the nondegeneracy condition (2.11) for the Malliavin covariance matrix of S_t , solution of Eq. (3.2).

PROPOSITION 3.1. *Suppose that Hypotheses 3.1 and 3.2 hold true.*

We assume that there exists a positive constant ϵ such that for every $(t, a, x) \in [0, T] \times \mathbb{R} \times \mathbb{R}$,

$$|\partial_a c(t, a, x)| \geq \epsilon \text{ and } |1 + \partial_x c(t, a, x)| \geq \epsilon. \quad (3.12)$$

Take $\delta \in (0, 1/2)$ in the definition of the weight π .

Then, for all $t \in [0, T]$, S_t satisfies the nondegeneracy condition (2.11) on $A = \{J_t = n\}$, for all $n \geq 1$.

PROOF. We refer to BALLY, BAVOUZET and MESSAOUD [2005]. □

Hence, Theorem 2.1 allows us to settle integration by parts formulas by using the jump amplitudes of S_t . Let us denote

$$U_t^{(s,t)} := \gamma_t^{(s,t)} L_{(s,t)} S_t - \langle DS_t, D\gamma_t^{(s,t)} \rangle_{(s,t)}, \quad (3.13)$$

$$\begin{aligned}
V_{(s,t)} &:= U_s^{(0,s)} - \gamma_s^{(0,s)} \langle DS_s, DS_t \rangle_{(0,s)} U_t^{(s,t)} \\
&\quad + \frac{1}{2} \gamma_s^{(0,s)} \gamma_t^{(s,t)} \langle DS_s, D\sigma_t^{s,t} \rangle_{(0,s)}.
\end{aligned} \tag{3.14}$$

We first give the integration by parts formula, which will be used in Section 4 to compute the Delta of European options.

PROPOSITION 3.2. *Suppose that Hypotheses 3.1 and 3.2 hold true. Assume that hypothesis (3.12) is satisfied and take $\delta \in (0, 1/2)$ in the definition of the weight π . For every function $\phi \in C_p^1(\mathbb{R})$, for all $t \in [0, T]$, we have*

$$E(\phi'(S_t) \partial_x S_t \mathbf{1}_{\{J_t \geq 1\}}) = E(\phi(S_t) H_\pi(S_t, \partial_x S_t) \mathbf{1}_{\{J_t \geq 1\}}), \tag{3.15}$$

where $H_\pi(S_t, \partial_x S_t)$ is given by

$$H_\pi(S_t, \partial_x S_t) = \partial_x S_t U_t^{(0,t)} - \gamma_t^{(0,t)} \langle DS_t, D(\partial_x S_t) \rangle_{(0,t)}.$$

PROOF. Applying Theorem 2.1, we obtain for all $n \geq 1$,

$$E(\phi'(S_t) \partial_x S_t \mathbf{1}_{\{J_t = n\}}) = E(\phi(S_t) H_n \mathbf{1}_{\{J_t = n\}}),$$

where H_n is defined by (2.10). Summing over $n \geq 1$, we get (3.15), where $H_\pi(S_t, \partial_x S_t) \mathbf{1}_{\{J_t \geq 1\}} = \sum_{n=1}^{\infty} H_n \mathbf{1}_{\{J_t = n\}}$. $\square$

In the following proposition, we derive integration by parts formulas which are used to compute conditional expectations.

PROPOSITION 3.3. *Suppose that Hypotheses 3.1 and 3.2 hold true.*

Assume that hypothesis (3.12) is satisfied and take $\delta \in (0, 1/2)$ in the definition of the weight π . For all $0 < s < t \leq T$, for every $\phi, \psi \in C_p^1(\mathbb{R})$, we have

$$E(\phi'(S_s) \psi(S_t) \mathbf{1}_{\{0 < J_s < J_t\}}) = E(\phi(S_s) \psi(S_t) V_{(s,t)} \mathbf{1}_{\{0 < J_s < J_t\}}), \tag{3.16}$$

where $V_{(s,t)}$ is given by (3.14).

PROOF. We denote $A_{s,t} := \{0 < J_s < J_t\}$. Using Theorem 2.1 with the weights $(\pi_{(0,s)}^i)_{i \in \mathbb{N}}$, we obtain

$$E(\phi'(S_s) \psi(S_t) \mathbf{1}_{A_{s,t}}) = E(\phi(S_s) H_1(S_s, S_t) \mathbf{1}_{A_{s,t}}), \tag{3.17}$$

with

$$\begin{aligned}
H_1(S_s, S_t) &= \psi(S_t) \left(\gamma_s^{(0,s)} L_{(0,s)}(S_s) - \langle DS_s, D\gamma_s^{(0,s)} \rangle_{(0,s)} \right) \\
&\quad - \gamma_s^{(0,s)} \psi'(S_t) \langle DS_s, DS_t \rangle_{(0,s)}.
\end{aligned} \tag{3.18}$$

Taking the weight $\pi_{(0,s)}^i$ in Eq. (3.17) means that we use a Malliavin integration by parts formula with respect to the jumps that occur before s . We do not take into account the jumps falling in $]s, t]$.

We now get rid of the derivative of ψ in expectation

$$\mathbb{E} \left(\phi(S_s) \gamma_s^{(0,s)} \psi'(S_t) \langle DS_s, DS_t \rangle_{(0,s)} \mathbf{1}_{A_{s,t}} \right).$$

We apply again Theorem 2.1 using the weights $(\pi_{(s,t)}^i)_{i \in \mathbb{N}}$. This means that we use a Malliavin integration by parts formula with respect to the jumps occurring between s and t . Note that $\phi(S_s) \gamma_s^{(0,s)}$ does not depend on the jumps of $]s, t]$ as if they were constants. We, thus, obtain

$$\mathbb{E} \left(\phi(S_s) \psi'(S_t) \gamma_s^{(0,s)} \langle DS_s, DS_t \rangle_{(0,s)} \mathbf{1}_{A_{s,t}} \right) = \mathbb{E}(\psi(S_t) \mathcal{H}_1(S_s, S_t) \mathbf{1}_{A_{s,t}}),$$

where

$$\begin{aligned} \mathcal{H}_1(S_s, S_t) &= \phi(S_s) \gamma_s^{(0,s)} \langle DS_s, DS_t \rangle_{(0,s)} \left(\gamma_t^{(s,t)} L_{(s,t)}(S_t) - \langle D\gamma_t^{(s,t)}, DS_t \rangle_{(s,t)} \right) \\ &\quad - \gamma_t^{(s,t)} \phi(S_s) \gamma_s^{(0,s)} \langle D(\langle DS_s, DS_t \rangle_{(0,s)}), DS_t \rangle_{(s,t)}. \end{aligned}$$

Since DS_s does not depend on the jumps of $]s, t]$, we have

$$\begin{aligned} D(\langle DS_s, DS_t \rangle_{(0,s)}), DS_t \rangle_{(s,t)} &= \sum_{i,j=1}^{\infty} \pi_{(0,s)}^i(\Delta_i) \pi_{(s,t)}^j(\Delta_j) D_i S_s D_j S_t D_{ij}^2 S_t \\ &= \sum_{i,j=1}^{\infty} \pi_{(0,s)}^i(\Delta_i) \pi_{(s,t)}^j(\Delta_j) D_i S_s \times \frac{1}{2} D_i (D_j S_t^2) \\ &= \sum_{i=1}^{\infty} \pi_{(0,s)}^i(\Delta_i) D_i S_s \times \frac{1}{2} D_i \sigma_t^{(s,t)} \\ &= \frac{1}{2} \langle DS_s, D\sigma_t^{(s,t)} \rangle_{(0,s)}. \end{aligned}$$

So,

$$\begin{aligned} \mathcal{H}_1(S_s, S_t) &= \phi(S_s) \gamma_s^{(0,s)} \langle DS_s, DS_t \rangle_{(0,s)} \left(\gamma_t^{(s,t)} L_{(s,t)}(S_t) - \langle D\gamma_t^{(s,t)}, DS_t \rangle_{(s,t)} \right) \\ &\quad - \frac{1}{2} \gamma_t^{(s,t)} \gamma_s^{(0,s)} \phi(S_s) \langle DS_s, D\sigma_t^{(s,t)} \rangle_{(0,s)}. \end{aligned} \quad (3.19)$$

We plug (3.18) and (3.19) in (3.17) and obtain

$$\mathbb{E}(\phi'(S_s) \psi(S_t) \mathbf{1}_{A_{s,t}}) = \mathbb{E}(\phi(S_s) \psi(S_t) V_{(s,t)} \mathbf{1}_{A_{s,t}}).$$

□

Using the integration by parts formula of Proposition 3.2, we can derive representation formulas for conditional expectations.

COROLLARY 3.2. (Representation formulas)

For all $0 \leq s < t \leq T$, for all $\phi \in \mathcal{C}_p^1(\mathbb{R})$, one has

$$\mathbb{E}(\phi(S_t) \mathbf{1}_{\{0 < J_s < J_t\}} \mid S_s = \alpha) = \frac{\mathbb{T}_{s,t}[\phi](\alpha)}{\mathbb{T}_{s,t}[1](\alpha)} \mathbf{1}_{\{0 < J_s < J_t\}},$$

where

$$\mathbb{T}_{s,t}[f](\alpha) = \mathbb{E}(f(S_t) H(S_s - \alpha) V_{(s,t)} \mathbf{1}_{\{0 < J_s < J_t\}}), \quad (3.20)$$

with $H(z) = \mathbf{1}_{z \geq 0}$, $z \in \mathbb{R}$.

PROOF. The proof is quite similar to the one developed in BALLY, CAMELLINO and ZANETTE [2003] and LIONS and REGNIER [2000]. The main difference is that, in this work, we use a localized integration by parts formula since (3.16) holds true on $A_{s,t} := \{0 < J_s < J_t\}$. We refer to BAVOUZET [2006] for details. $\square$

3.3. Integration by parts formula using the jump times

In this section, we differentiate with respect to the jump times T_i , $i \in \mathbb{N}$. It is well known (see BERTOIN [1996]) that conditionally to $\{J_t = n\}$, the law of the vector $(T_1, \dots, T_n)$ is absolutely continuous with respect to the Lebesgue measure and has the density

$$p(\omega, t_1, \dots, t_n) = \frac{n!}{t^n} \mathbf{1}_{\{0 < t_1 < \dots < t_n < t\}}(t_1, \dots, t_n) \mathbf{1}_{\{J_t(\omega) = n\}}.$$

In particular, for a given $i = 1, \dots, n$, conditionally to $\{J_t = n\}$ and to $\{T_j, j \neq i\}$, T_i is uniformly distributed on $[T_{i-1}(\omega), T_{i+1}(\omega)]$. Hence, T_i has the density (with the convention $T_0 = 0$, $T_{n+1} = t$)

$$p_i(\omega, u) = \frac{1}{T_{i+1}(\omega) - T_{i-1}(\omega)} \mathbf{1}_{[T_{i-1}(\omega), T_{i+1}(\omega)]}(u) \, du, \quad i = 1, \dots, n.$$

Since p_i is not differentiable with respect to u , we use the weights:

$$\pi_i(\omega, u) = (T_{i+1}(\omega) - u)^\alpha (u - T_{i-1}(\omega))^\alpha \mathbf{1}_{[T_{i-1}(\omega), T_{i+1}(\omega)]}(u), \quad i = 1, \dots, n$$

with $\alpha \in (0, 1)$. Then, the weights $(\pi_i)_{i \in \mathbb{N}}$ satisfy Hypothesis 2.2.

Using the notation of Section 2, we take $V_i = T_i$, $\mathcal{G} = \sigma(\Delta_i, i \in \mathbb{N}) \vee \sigma(J_t)$, and $A = \{J_t = n\}$, for $n \in \mathbb{N}^*$. On $\{J_t = n\}$, we have $S_t = s_t(T_1, \dots, T_n, \Delta_1(\omega), \dots, \Delta_n(\omega))$, so that $S_t \in S_{(n,2)}(A)$. The differential operators are

$$\begin{aligned} D_i S_t &= \partial_{u_i} s_t(T_1, \dots, T_n, \Delta_1(\omega), \dots, \Delta_n(\omega)), \\ \sigma_{\pi, S_t} &= \sum_{i=1}^n \pi_i(\omega, T_i) \left| \partial_{u_i} s_t(T_1, \dots, T_n, \Delta_1(\omega), \dots, \Delta_n(\omega)) \right|^2, \\ L_{i, \pi} S_t &= -(\pi'_i \partial_{u_i} s_t + \pi_i \partial_{u_i}^2 s_t)(T_1, \dots, T_n, \Delta_1(\omega), \dots, \Delta_n(\omega)). \end{aligned}$$

All these quantities may be computed using Lemma 3.1.
Recall that we have defined in Section 3.1

$$q(t, \alpha, x) := (\partial_t c + g \partial_x c)(t, \alpha, x) + g(t, x) - g(t, x + c(t, \alpha, x)).$$

Then, the ellipticity-type condition that implies that the Malliavin covariance matrix of S_t is nondegenerated reads as follows:

PROPOSITION 3.4. *Suppose that Hypothesis 3.1 holds true and that for some $\epsilon > 0$, for all $(t, a, x) \in [0, T] \times \mathbb{R} \times \mathbb{R}$,*

$$|q(t, a, x)| \geq \epsilon > 0 \text{ and } |(1 + \partial_x c)(t, a, x)| \geq \epsilon > 0. \quad (3.21)$$

Take $\alpha \in (0, 1/2)$ in the definition of the weights $(\pi_i)_{i \in \mathbb{N}}$.

Then, for all $t \in [0, T]$, S_t satisfies the nondegeneracy condition (2.11) on $\{J_t = n\}$, for all $n \geq 4$.

PROOF. We refer to BALLY, BAVOUZET and MESSAOUD [2005] for the proof. $\square$

REMARK 3.1. We are not able to handle the nondegeneracy problem corresponding to the jump times if there is less than three jumps on $[0, t]$. But note that if the jump intensity λ is large enough, then the probability of the event $\{J_t \leq 3\}$ is very small.

PROPOSITION 3.5. *Suppose that Hypothesis 3.1 holds true and that condition (3.21) is satisfied.*

Take $\alpha \in (0, 1/2)$ in the definition of the weights $(\pi_i)_{i \in \mathbb{N}}$.

Then, for every function $\phi \in C_p^1(\mathbb{R})$, for all $t \in [0, T]$, we have

$$E(\phi'(S_t) \partial_x S_t \mathbf{1}_{\{J_t \geq 4\}}) = E(\phi(S_t) H_\pi(S_t, \partial_x S_t) \mathbf{1}_{\{J_t \geq 4\}}), \quad (3.22)$$

where $H_\pi(S_t, \partial_x S_t)$ is given by

$$\begin{aligned} H_\pi(S_t, \partial_x S_t) &= \partial_x S_t \gamma_{\pi, S_t} L_\pi S_t - \gamma_{\pi, S_t} < DS_t, D(\partial_x S_t) >_\pi \\ &\quad - \partial_x S_t < DS_t, D\gamma_{\pi, S_t} >_\pi. \end{aligned}$$

4. Sensitivity analysis for European options using integration by parts formula

The aim of this section is to show how the integration by parts formulas (3.15) and (3.22) are used to compute the Delta of European options when the asset price is a pure jump process.

4.1. The procedure

Let us compute $\partial_x E(\phi(S_T))$, where $x = S_0$. We write

$$\begin{aligned}\partial_x E(\phi(S_T)) &= E(\phi'(S_T) \partial_x S_T) \\ &= E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T=0\}}) + E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}}).\end{aligned}$$

On $\{J_T \geq 1\}$, we use an integration by parts formula of the Malliavin type, and we obtain

$$E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}}) = E(\phi(S_T) H(S_T, \partial_x S_T) \mathbf{1}_{\{J_T \geq 1\}}),$$

where $H(S_T, \partial_x S_T)$ is a weight involving Malliavin derivatives of S_T and $\partial_x S_T$. Hence, we have

$$\partial_x E(\phi(S_T)) = E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T=0\}}) + E(\phi(S_T) H(S_T, \partial_x S_T) \mathbf{1}_{\{J_T \geq 1\}}).$$

In order to compute the two terms in the right-hand side of the above equality, we proceed as follows.

On $\{J_T = 0\}$, there is no jump on $]0, T]$, thus S_T and $\partial_x S_T$ solve some deterministic integral equation. In the examples that we consider in this section, the solution of these equations are explicit so that this term is explicitly known. Hence, we may use the finite-difference method to compute $E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T=0\}})$.

For the computation of the term $E(\phi(S_T) H(S_T, \partial_x S_T) \mathbf{1}_{\{J_T \geq 1\}})$, we use a Monte Carlo algorithm. We simulate a sample $((T_n^k)_{n \in \mathbb{N}}, (\Delta_n^k)_{n \in \mathbb{N}})$, $k = 1, \dots, M$ of the jump times amplitudes, and we compute the corresponding J_T^k , S_T^k , and $H^k(S_T^k, \partial_x S_T^k)$. Then, we write

$$E(\phi(S_T) H(S_T, \partial_x S_T) \mathbf{1}_{\{J_T \geq 1\}}) \simeq \frac{1}{M} \sum_{k=1}^M \phi(S_T^k) H^k(S_T^k, \partial_x S_T^k) \mathbf{1}_{\{J_T^k \geq 1\}}.$$

Thus,

$$\begin{aligned}\partial_x E(\phi(S_T)) &\simeq \phi'(S_T^k) \partial_x S_T^k \mathbf{1}_{\{J_T^k=0\}} \\ &\quad + \frac{1}{M} \sum_{k=1}^M \phi(S_T^k) H^k(S_T^k, \partial_x S_T^k) \mathbf{1}_{\{J_T^k \geq 1\}},\end{aligned}$$

and we have to compute the Malliavin estimators $H^k(S_T^k, \partial_x S_T^k)$.

Note that the estimator $H(S_T, \partial_x S_T)$ may have a large variance. In order to reduce this variance, we use the same localization method as the one introduced by FOURNIÉ, LASRY, LEBOUCHOUX, LIONS and TOUZI [1999], which means that we use a localization function that vanishes out of an interval $[K - \delta, K + \delta]$, for some $\delta > 0$ (where K is the strike of the option). Let us be more precise.

For $\delta > 0$, we consider the following function (see Fig. 4.1)

$$\begin{aligned} B_\delta(s) &:= 0 && \text{if } s \leq K - \delta, \\ &:= \frac{s - (K - \delta)}{2\delta} && \text{if } s \in [K - \delta, K + \delta], \\ &:= 1 && \text{if } s \geq K + \delta. \end{aligned}$$

Let the function G_δ (see Fig. 4.2) be a primitive of B_δ :

$$\begin{aligned} G_\delta(t) &:= \int_{-\infty}^t B_\delta(s) ds \\ &:= 0 && \text{if } t \leq K - \delta \\ &:= \frac{(t - (K - \delta))^2}{4\delta} && \text{if } t \in [K - \delta, K + \delta] \\ &:= t - K && \text{if } t \geq K + \delta. \end{aligned}$$

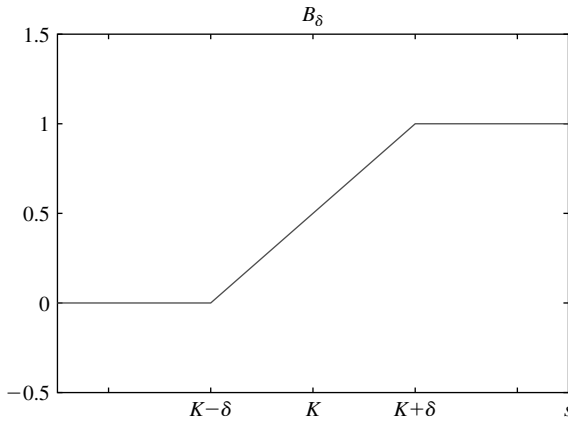


FIG. 4.1 Representation of B for $K = 100$, $\delta = 20$.

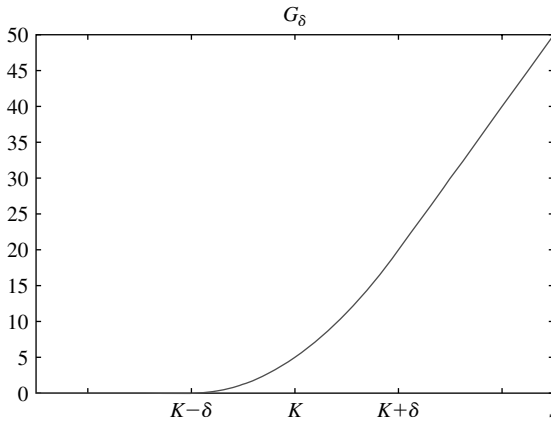
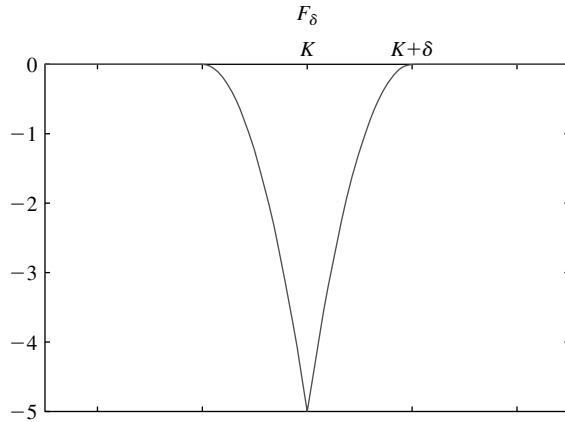


FIG. 4.2 Representation of G for $K = 100$, $\delta = 20$.

FIG. 4.3 Representation of F for $K = 100$, $\delta = 20$.

We then define the localization function (see Fig. 4.3)

$$\begin{aligned}
 F_\delta(t) &:= (t - K)_+ - G_\delta(t) \\
 &:= 0 && \text{if } t \leq K - \delta \\
 &:= -\frac{(t - (K - \delta))^2}{4\delta} && \text{if } t \in [K - \delta, K] \\
 &:= t - K - \frac{(t - (K + \delta))^2}{4\delta} && \text{if } t \in [K, K + \delta] \\
 &:= 0 && \text{if } t \geq K + \delta.
 \end{aligned}$$

Since $F_\delta(S_T) + G_\delta(S_T) = (S_T - K)_+$, we have on $\{J_T \geq 1\}$

$$\begin{aligned}
 \partial_{S_0} \mathbb{E}[(S_T - K)_+] &= \partial_{S_0} \mathbb{E}[G_\delta(S_T)] + \partial_{S_0} \mathbb{E}[F_\delta(S_T)] \\
 &= \mathbb{E}[B_\delta(S_T) \partial_{S_0} S_T] + \mathbb{E}[F_\delta(S_T) H(S_T, \partial_{S_0} S_T)].
 \end{aligned}$$

Since F_δ vanishes out of $[K - \delta, K + \delta]$, the value of the second expectation does not blow up as $H(S_T, \partial_{S_0} S_T)$ increases.

We deal with two examples for modelling the asset $(S_t)_{t \in [0, T]}$.

The first one is motivated by the Vasicek model used for interest rates (but we consider a jump process instead of a Brownian motion):

$$S_t = x - \int_0^t r(S_u - \alpha) du + \sum_{i=1}^{J_t} \sigma \Delta_i. \quad (4.1)$$

And the second one is of the Black–Scholes type:

$$S_t = x + \int_0^t r S_u du + \sigma \sum_{i=1}^{J_t} S_{T_i^-} \Delta_i. \quad (4.2)$$

In both models, we take $\Delta_i \sim \mathcal{N}(0, 1)$, $i \geq 1$. That is, for all $i \geq 1$, Δ_i has the density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{\rho(x)}, \text{ with } \rho(x) = -\frac{x^2}{2}.$$

4.2. Malliavin estimators computation

One may use Lemma 3.1, but in the particular cases that we discuss here, we have explicit solutions. So direct computations of the differential operators are much easier.

- We first study the Vasicek model (4.1). Let us fix $n \geq 1$. We have an explicit expression of S_T on $\{J_T = n\}$:

$$S_T = x e^{-rT} + \alpha (1 - e^{-rT}) + \sigma \sum_{j=1}^n \Delta_j e^{-r(T-T_j)}. \quad (4.3)$$

We may use integration by parts with respect to the jump amplitudes [see (3.15)] or to the jump times [see (3.22)].

- * **Jump amplitudes:** Since Δ_i is Gaussian distributed for all i , we choose the weight $\pi(\omega, \Delta_i) = 1$ (see Example 3.1, 2), and on $\{J_T = n\}$ we get for all $1 \leq i \leq n$,

$$D_i S_T = \sigma e^{-r(T-T_i)} \quad (4.4)$$

$$D_{ii}^2 S_T = 0$$

$$Y_T := \frac{\partial S_T}{\partial x} = e^{-rT} \quad (4.5)$$

$$D_i Y_T = 0,$$

and the covariance matrix is given by :

$$\sigma_T = \sum_{j=1}^n |D_j S_T|^2 = \sigma^2 \sum_{j=1}^n e^{-2r(T-T_j)}.$$

Then, $\gamma_T = \frac{1}{\sigma_T} \Rightarrow D_i \gamma_T = 0$, for all $1 \leq i \leq n$. Since $\frac{\partial \ln p(\Delta)}{\partial \Delta} = -\Delta$, one has

$$L S_T = - \sum_{j=1}^n D_j S_T \frac{\partial \ln p(\Delta_j)}{\partial \Delta_j} = \sum_{j=1}^n \sigma e^{-r(T-T_j)} \Delta_j.$$

We obtain finally on $\{J_T = n\}$ for $n \geq 1$,

$$H_n(S_T, \partial_x S_T) = \frac{\sum_{j=1}^n e^{rT_j} \Delta_j}{\sigma \sum_{j=1}^n e^{2rT_j}}. \quad (4.6)$$

* **Jump times.** We present now the estimators based on the jump times T_k , $k \in \mathbb{N}$. We use the weights $\pi_i(\omega, T_i) = (T_{i+1} - T_i)^\alpha (T_i - T_{i-1})^\alpha$, and we have $\pi'_i = \alpha \delta_{i+1}^{\alpha-1} \delta_i^{\alpha-1} (\delta_{i+1} - \delta_i)$, where $\delta_i = T_i - T_{i-1}$. Differentiating with respect to the jump times in Eq. (4.3), we have

$$D_i S_T = \sigma \Delta_i r e^{-r(T-T_i)}$$

and then on $\{J_T = n\}$,

$$\sigma_{\pi, S_T} = \sum_{i=1}^n \pi_i (\sigma r)^2 \Delta_i^2 e^{-2r(T-T_i)}.$$

On $\{J_T = n\}$, we have $L_\pi(S_T) = - \sum_{i=1}^n L_{i,\pi}(S_T)$, with

$$L_{i,\pi} S_T = -\sigma r \Delta_i e^{-r(T-T_i)} \left(r \pi_i + \alpha (\delta_{i+1} \delta_i)^{\alpha-1} (\delta_{i+1} - \delta_i) \right).$$

Let us denote

$$A_j = \alpha (\delta_{j+1} \delta_j)^{\alpha-1} \Delta_j^2 e^{2r T_j},$$

$$B_j = \Delta_j^2 e^{2r T_j} \left[2r \pi_j + \alpha (\delta_{j+1} \delta_j)^{\alpha-1} (\delta_{j+1} - \delta_j) \right].$$

We then obtain

$$D_j \sigma_{\pi, S_T} = (\sigma r)^2 e^{-2r T} (A_{j-1} \delta_{j-1} - A_{j+1} \delta_{j+2} + B_j).$$

Moreover, $\partial_x S_T = e^{-r T}$ so that $D_i \partial_x S_T = 0$ for all $i = 1, \dots, n$.

We have now the expression of all the terms involved in $H_n(S_T, \partial_x S_T)$. For $n \geq 4$, on $\{J_T = n\}$, we obtain

$$\begin{aligned} H_n(S_T, \partial_x S_T) &= \frac{\sum_{i=1}^n \Delta_i e^{r T_i} (r \pi_i + \alpha (\delta_{i+1} \delta_i)^{\alpha-1} (\delta_{i+1} - \delta_i))}{\sigma r \hat{\sigma}} \\ &\quad - \frac{\sum_{i=1}^n \pi_i \Delta_i e^{r T_i} (A_{i-1} \delta_{i-1} - A_{i+1} \delta_{i+2} + B_i)}{\sigma r \hat{\sigma}^2}, \end{aligned} \quad (4.7)$$

where $\hat{\sigma} = \sum_{i=1}^n \pi_i \Delta_i^2 e^{2r T_i}$.

For $n = 1, 2, 3$, we use integration by parts with respect to the first jump amplitude Δ_1 only. Then, similar computations give on $\{J_T = n\}$ for $1 \leq n \leq 3$

$$H_n(S_T, \partial_x S_T) = \frac{e^{-r T_1}}{\sigma \Delta_1}.$$

- We now study the geometrical model (4.2). Let us fix $n \geq 1$. On $\{J_T = n\}$ we have

$$S_T = x e^{rT} \prod_{j=1}^n (1 + \sigma \Delta_j). \quad (4.8)$$

We may not use integration by parts with respect to the jump times because S_T depends on $T_1, \dots, T_n$ by means of J_t only. So we perform integration by parts formula using the jump amplitudes only. We have for all $1 \leq i \leq n$,

$$D_i S_T = \frac{\sigma S_T}{1 + \sigma \Delta_i} = \sigma \prod_{j=1, j \neq i}^n (1 + \sigma \Delta_j). \quad (4.9)$$

Let us define

$$\tilde{A}_\sigma = \sum_{j=1}^n \frac{1}{(1 + \sigma \Delta_j)^2}, \quad (4.10)$$

$$\tilde{B}_\sigma = \sum_{j=1}^n \frac{\Delta_j}{(1 + \sigma \Delta_j)}, \quad (4.11)$$

$$\tilde{C}_\sigma = \sum_{j=1}^n \frac{1}{(1 + \sigma \Delta_j)^4}. \quad (4.12)$$

We then get for all $1 \leq i \leq n$

$$\begin{aligned} D_{ii}^2 S_T &= 0, \\ Y_T &= \frac{\partial S_T}{\partial x} = \frac{S_T}{S_0}, \end{aligned} \quad (4.13)$$

$$\begin{aligned} D_i Y_T &:= \frac{\sigma S_T}{S_0 (1 + \sigma \Delta_i)}, \\ \sigma_T &= \sigma^2 S_T^2 \sum_{j=1}^n \frac{1}{(1 + \sigma \Delta_j)^2} = \sigma^2 S_T^2 \tilde{A}_\sigma, \end{aligned} \quad (4.14)$$

$$D_i \sigma_T = \left(\frac{2\sigma^3 S_T^2}{1 + \sigma \Delta_i} \right) \left(\tilde{A}_\sigma - \frac{1}{(1 + \sigma \Delta_i)^2} \right),$$

$$D_i \gamma_T = -\frac{D_i \sigma_T}{\sigma_T^2}.$$

Hence, on $\{J_T = n\}$, $n \geq 1$, the Malliavin weight is given by

$$H_n(S_T, \partial_x S_T) = \frac{\tilde{B}_\sigma}{\sigma x \tilde{A}_\sigma} + \frac{1}{x} - \frac{2\tilde{C}_\sigma}{x \tilde{A}_\sigma^2}. \quad (4.15)$$

4.3. Numerical experiments

In this section, we compute the Delta of two European options: call option with payoff $\phi(x) = (x - K)_+$ and digital option with payoff $\phi(x) = \mathbf{1}_{x \geq K}$. We present numerical experiments in order to compare the Malliavin approach to the finite-difference method. We look at two kinds of Malliavin Monte Carlo estimators: those obtained using a localization method or not using a localization method. Moreover, we compute Malliavin estimators using the jump amplitudes or the jump times. For European call options, we use the variance reduction method detailed in Section 4.1.

REMARK 4.1. We choose the parameter σ in the diffusion models (4.1) and (4.2) in the following way:

- For the geometrical model, the variance of S_t is

$$\text{Variance}(S_t) = x^2 e^{2rt} (e^{\sigma^2 \lambda t} - 1).$$

Taking $\lambda = 1$, $r = 0.1$, $T = 5$, and $x = 100$, if $\sigma \in [0.1, 0.6]$, we have $1393.69 \leq \text{Variance}(S_T) \leq 137\,264$. We choose here small values for σ in order to fit the usual values of the volatility taken in the Black–Scholes model.

- For the Vasicek type model, we have

$$\text{Variance}(S_t) = 2\alpha e^{-2rt} (x - \alpha) + \frac{\lambda \sigma^2}{2r} (1 - e^{-2rt}).$$

Taking $\lambda = 1$, $r = 0.1$, $T = 5$, $\alpha = 10$, and $x = 100$, if $\sigma \in [16, 50]$, we have $1471.3 \leq \text{Variance}(S_T) \leq 8563.69$. Note that choosing large values for σ seems to be “sensible” in order to fit the usual values taken by the practitioners in the Vasicek model.

First, we compare the results given by Malliavin calculus (using the jump amplitudes only) and finite-difference method. We also compare the localized and nonlocalized Malliavin estimators.

Let us first present the figures obtained for European options using the Vasicek model.

We now present the results obtained for European options using the geometrical model.

These experiments show that we can numerically compute the greeks for European options with a pure jump underlying process. Moreover, from Figs. 4.4–4.7, we obtain numerical results similar to those in the Wiener case (FOURNIÉ, LASRY, LEBOUCHOUX, LIONS and TOUZI [1999] and FOURNIÉ, LASRY, LEBOUCHOUX and LIONS [2001]).

For European call options (with regular payoff), the Malliavin estimator has larger variance than the finite-difference one (see Figs. 4.5 and 4.7): the finite difference method approximates the first derivative of the payoff, whereas the Malliavin estimator contains a weight (independent of the payoff), which may increase the variance. The localization method detailed in Section 4.1 allows us to reduce the variance of the Malliavin estimator.

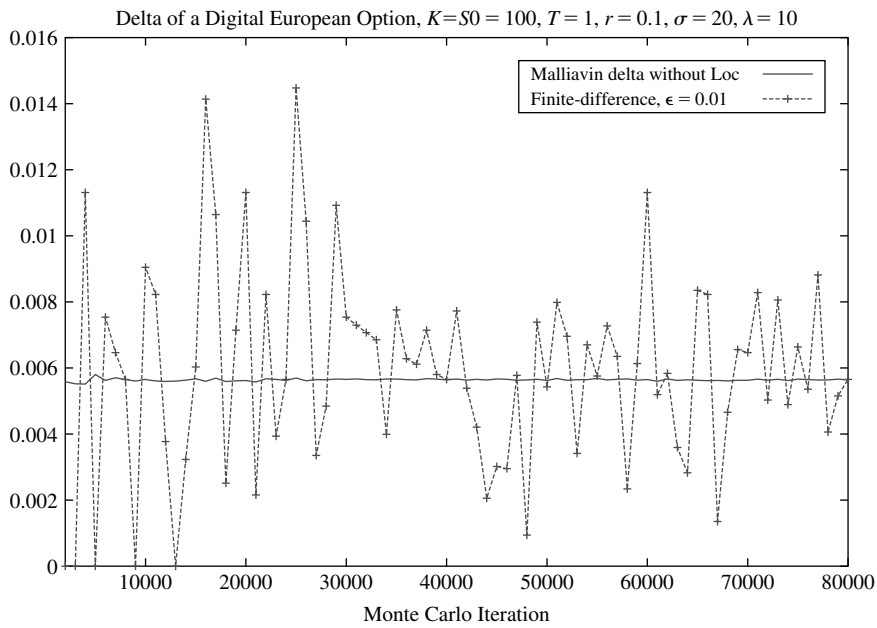


FIG. 4.4 Delta of a European digital option using Malliavin calculus and finite-difference method (Vasicek model).

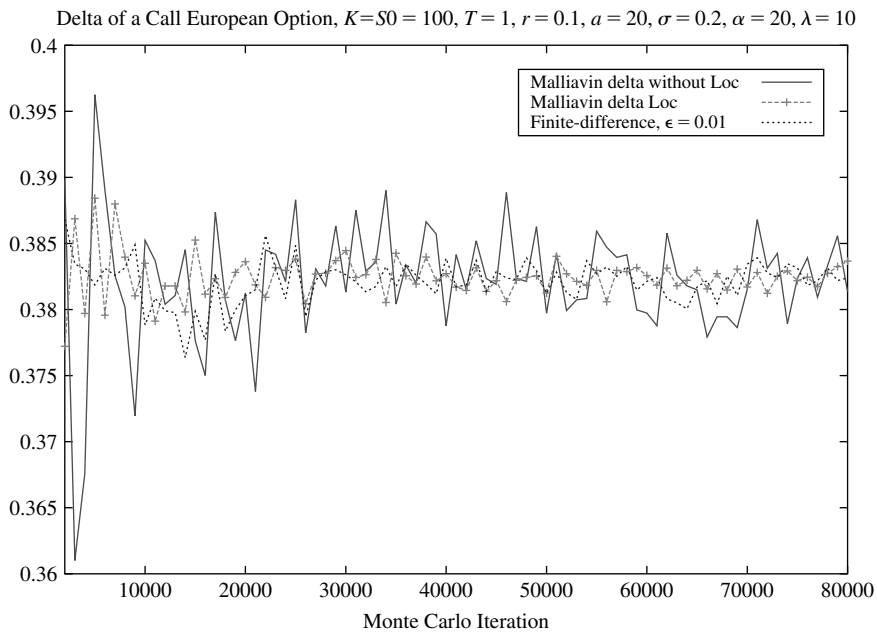


FIG. 4.5 Delta of a European call option using Malliavin calculus and finite-difference method (Vasicek model).

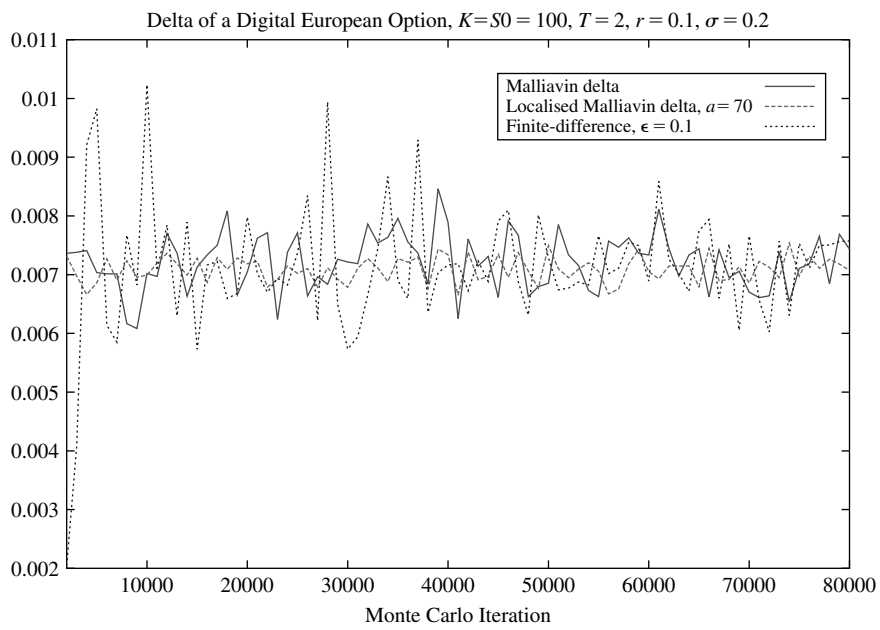


FIG. 4.6 Delta of a European digital option using Malliavin calculus and finite-difference method (Geometrical model).

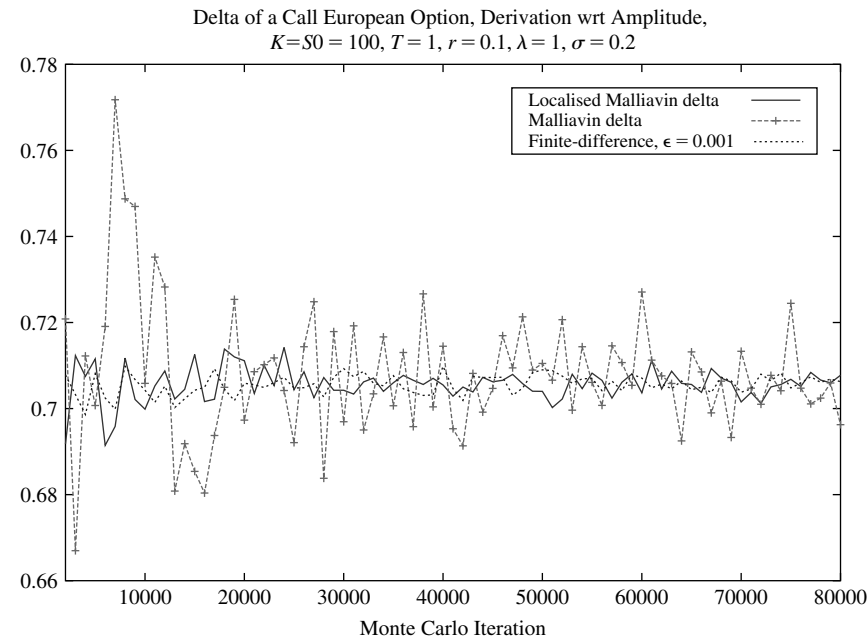


FIG. 4.7 Delta of a European call option using Malliavin calculus and finite-difference method (Geometrical model).

On the opposite, the Malliavin estimator of digital options (with singular payoff) has lower variance than the finite-difference one (see Figs. 4.4 and 4.6) and so does not need to be localized; in this case, the first derivative of the payoff is a Dirac distribution and, contrary to the finite-difference method, the Malliavin calculus allows us to avoid this strong discontinuity.

Finally, note that for both call and digital options, the finite-difference method requires to simulate twice more samples of the asset than the Malliavin method does; the finite-difference method uses the samples starting from S_0 and those starting from $S_0 + \epsilon$. The Malliavin method is, thus less time consuming.

Now, we compare Malliavin estimators using jump amplitudes with those using jump times. In Tables 4.1 and 4.2, we give the empirical variance of these estimators: we denote by Var Mall JT (Var Mall AJ) the variance of the Malliavin estimator based on jump times (jump amplitudes). Moreover, we compare them with the finite-difference estimator, which we denote by Var Diff.

TABLE 4.1
Variance of the Malliavin JT estimator, AJ estimator, and the FD for call option in the Vasicek model

σ	Variance(S_T)	Var Mall JT	Var Mall AJ	Var Diff
15.8114	796.241	0.0285123	0.0106426	0.0300379
16.6667	897.577	0.0417219	0.0115955	0.0298567
17.6777	991.453	0.0400695	0.013123	0.0298904
18.8982	1134.11	0.0410136	0.0144516	0.0299574
20.4124	1313.42	0.0433065	0.0162378	0.029862
22.3607	1584.9	0.0400481	0.0178726	0.0298987
25	1967.53	0.0407136	0.0202055	0.0299007
28.8675	2604.22	0.0362728	0.0224265	0.0299651
35.3553	3961.31	0.0343158	0.0253757	0.0297775
50	7890.4	0.0333298	0.0287716	0.0299749

TABLE 4.2
Vasicek model. Variance of the Malliavin JT estimator, AJ estimator, and the FD for Digital option

σ	Variance(S_T)	Var Mall JT	Var Mall AJ	Var Diff
15.8114	796.241	0.00144622	$7.18878e-5$	0.00514743
16.6667	897.577	0.00254652	$7.3629e-5$	0.00459619
17.6777	991.453	0.0018011	$7.85552e-5$	0.00496369
18.8982	1134.11	0.0109864	$8.14005e-5$	0.00477995
20.4124	1313.42	0.00177648	$8.1627e-5$	0.00386111
22.3607	1584.9	0.00152777	$8.06193e-5$	0.00496369
25	1967.53	0.0013786	$7.94341e-5$	0.0062497
28.8675	2604.22	0.00100181	$7.5835e-5$	0.00551488
35.3553	3961.31	0.000617271	$6.95225e-5$	0.00459619
50	7890.4	0.000373802	$5.64325e-5$	0.00533116

We also mention in Tables 4.1 and 4.2 the *value of the volatility σ that we use* and the corresponding variance of the underlying, denoted by $\text{Variance}(S_t)$,

We use the following abbreviations:

- AJ : jump amplitudes
- JT : jump times
- FD : finite difference
- G : geometrical model
- V : vasicek model
- Call : call option
- Dig : digital option.

Then, (V/Dig/AJ) means that we deal with the Vasicek model (V), with a digital option (Dig) and we use an algorithm based on the amplitudes of the jumps (AJ). (V/Dig/AJ) versus (V/Dig/JT) means that we compare these two estimators.

Let us compare the variance of the Malliavin estimators based on jump times or amplitudes for the Vasicek model.

- (V/Call/AJ) versus (V/Call/JT) versus (V/Call/FD)
- (V/Dig/AJ) versus (V/Dig/JT) versus (V/Dig/FD)

As we can see from Figs. 4.8 and 4.9, the comparison between the finite-difference method and the Malliavin estimator using jump times leads to conclusions similar to

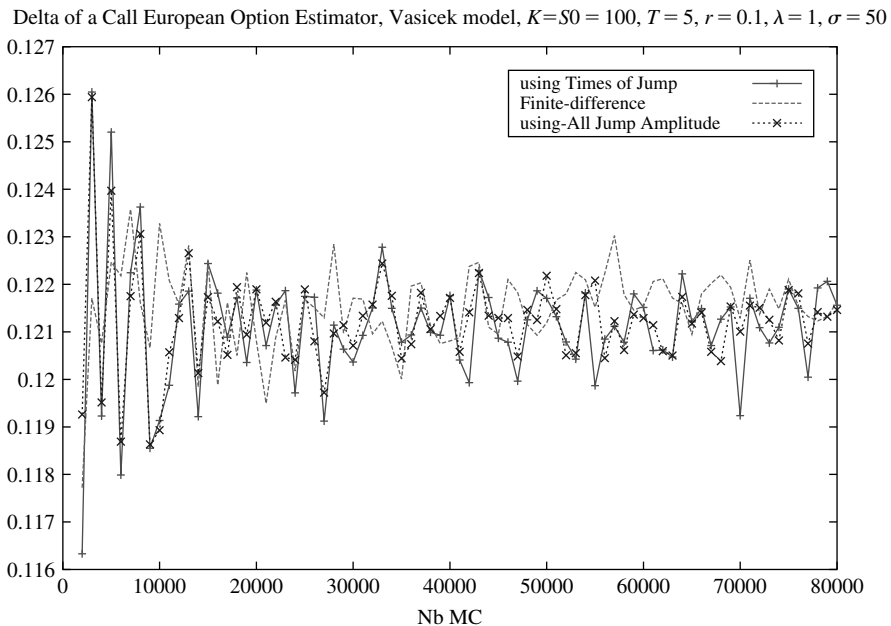


FIG. 4.8 Vasicek model. Delta of a European call option using Malliavin calculus based on jump times, on jump amplitudes, and finite-difference method.

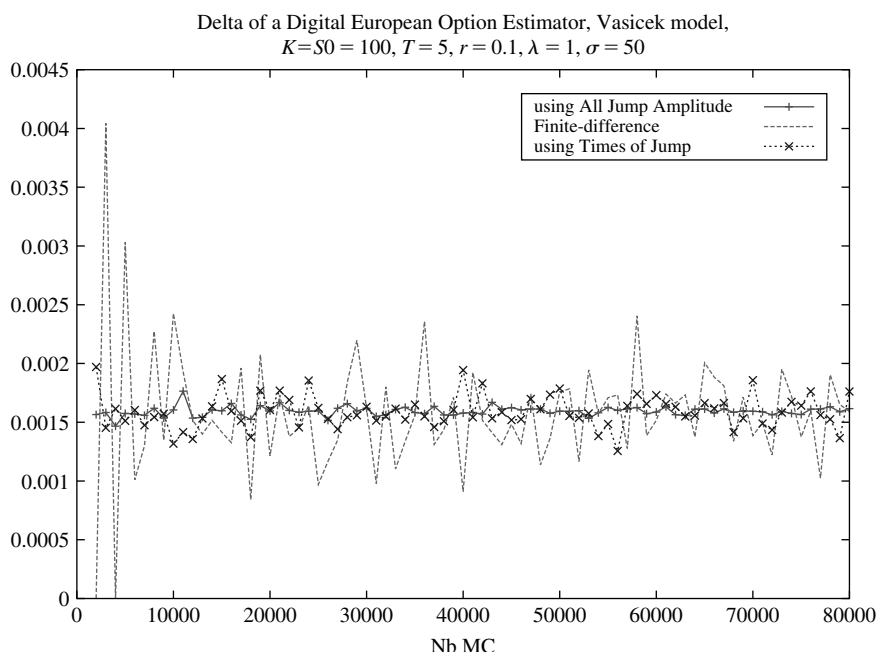


FIG. 4.9 Vasicek model. Delta of a European Digital option using Malliavin calculus based on the jump amplitudes, on the jump times, and finite-difference method.

that the comparison of the Malliavin estimator using jump amplitudes with the finite-difference method, for call options, these estimators are close, but for digital options, the Malliavin one is the most efficient.

On the other hand, if we look at Tables 4.1 and 4.2, we note that $\text{Var Mall JT} \geq \text{Var Mall AJ}$. This means that the use of Malliavin calculus with respect to jump amplitudes leads to estimators with lower variance than those based on jump times.

5. Sensitivity analysis for European options using the Bismut approach

In this section, we present an alternative procedure (for computing sensitivities) to the one developed in Section 4. Here, we use the Bismut–Elworthy formula to compute the Delta of European options, which is quite similar to FOURNIÉ, LASRY, LEBOUCHOUX, LIONS and TOUZI [1999] and FOURNIÉ, LASRY, LEBOUCHOUX and LIONS [2001]. This approach is based on a duality formula of the type (2.4) although using an integration by parts formula of the type (2.9).

5.1. The procedure

Let us first use the variance of constant method in order to give the link between the Malliavin derivative and the tangent flow.

In view of (3.7), we have for all $t \in [0, T]$,

$$\begin{aligned} D_i S_t &= \partial_a c(T_i, \Delta_i, S_{T_i}^-) + \int_{T_i}^t \partial_x g(r, S_r) D_i S_r \, dr \\ &\quad + \int_{T_{i+1}}^t \int_{\mathbb{R}} \partial_x c(u, a, S_{u-}) D_i S_{u-} N(du, da). \end{aligned} \quad (5.1)$$

Let us denote $Y_t := \partial_x S_t$ for all $t \in [0, T]$. Then, Y_t satisfies

$$Y_t = 1 + \int_0^t \partial_x g(r, S_r) Y_r \, dr + \int_0^t \int_{\mathbb{R}} \partial_x c(u, a, S_{u-}) Y_{u-} N(du, da).$$

Suppose that there exists a positive constant ϵ such that

$$\forall (t, a, x) \in [0, T] \times \mathbb{R} \times \mathbb{R}, |1 + \partial_x c(t, a, x)| \geq \epsilon > 0.$$

Then, its lemma implies that Y_t is invertible, and Y_t^{-1} satisfies

$$\begin{aligned} Y_t^{-1} &= 1 - \int_0^t \partial_x g(r, S_r) Y_r^{-1} \, dr \\ &\quad - \int_0^t \int_{\mathbb{R}} \frac{\partial_x c(u, a, S_{u-})}{1 + \partial_x c(u, a, S_{u-})} Y_{u-}^{-1} N(du, da). \end{aligned}$$

Thus,

$$Y_T - Y_{T_i} = \int_{T_i}^T \partial_x g(r, S_r) Y_r \, dr + \int_{T_{i+1}}^T \int_{\mathbb{R}} \partial_x c(u, a, S_{u-}) Y_{u-} N(du, da).$$

Since $Y_{T_i}^{-1} \partial_a c(T_i, \Delta_i, S_{T_i}^-)$ is $\mathcal{F}_{T_i}$ measurable, we get

$$\begin{aligned} &Y_T Y_{T_i}^{-1} \partial_a c(T_i, \Delta_i, S_{T_i}^-) - \partial_a c(T_i, \Delta_i, S_{T_i}^-) \\ &= \int_{T_i}^T \partial_x g(r, S_r) \left(Y_r Y_{T_i}^{-1} \partial_a c(T_i, \Delta_i, S_{T_i}^-) \right) \, dr \\ &\quad + \int_{T_{i+1}}^T \int_{\mathbb{R}} \partial_x c(u, a, S_{u-}) \left(Y_{u-} Y_{T_i}^{-1} \partial_a c(T_i, \Delta_i, S_{T_i}^-) \right) N(du, da). \end{aligned}$$

Using the unicity of (5.1), we obtain finally

$$D_i S_T = Y_T Y_{T_i}^{-1} \partial_a c(T_i, \Delta_i, S_{T_i}^-). \quad (5.2)$$

Let us come back to the Delta computation. As in Section 4, we write

$$\begin{aligned} \partial_x E(\phi(S_T)) &= E(\phi'(S_T) \partial_x S_T) \\ &= E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T=0\}}) + E(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}}). \end{aligned} \quad (5.3)$$

Suppose that there exists a positive constant ϵ such that

$$\forall (t, a, x) \in [0, T] \times \mathbb{R} \times \mathbb{R}, |\partial_a c(t, a, x)| \geq \epsilon > 0.$$

Then, using (5.2), we can write

$$\partial_x S_T = \frac{\partial_x S_{T_i} D_i S_T}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}. \quad (5.4)$$

Then, we multiply (5.4) by $\pi(\omega, \Delta_i)$, and we sum over $i = 1, \dots, J_T$:

$$\phi'(S_T) \partial_x S_T = \frac{1}{\sum_{i=1}^{J_T} \pi(\omega, \Delta_i)} \times \sum_{i=1}^{J_T} \pi(\omega, \Delta_i) \frac{\partial_x S_{T_i} (\phi'(S_T) D_i S_T)}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}.$$

Using the chain rule (2.5), we get

$$\phi'(S_T) \partial_x S_T = \frac{1}{\sum_{i=1}^{J_T} \pi(\omega, \Delta_i)} \times \sum_{i=1}^{J_T} \pi(\omega, \Delta_i) \frac{\partial_x S_{T_i} D_i \phi(S_T)}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}. \quad (5.5)$$

Let us define $U = (U_i)_{i \in \mathbb{N}}$ by

$$U_i := \frac{\partial_x S_{T_i}}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}. \quad (5.6)$$

Denote

$$\theta := \frac{1}{\sum_{i=1}^{J_T} \pi(\omega, \Delta_i)},$$

thus, the duality formula (2.4) gives

$$\begin{aligned} \mathbb{E} (\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}}) &= \mathbb{E} (\theta \langle U, D\phi(S_T) \rangle_\pi \mathbf{1}_{\{J_T \geq 1\}}) \\ &= \mathbb{E} (\phi(S_T) \delta_\pi (\theta U) \mathbf{1}_{\{J_T \geq 1\}}). \end{aligned}$$

Using (2.6), we have

$$\begin{aligned} \delta_\pi (\theta U) &= \theta \delta_\pi (U) - \langle U, D\theta \rangle_\pi \\ &= \theta \delta_\pi (U) + \theta^2 \sum_{i=1}^{J_T} \frac{\pi(\omega, \Delta_i) \pi'(\omega, \Delta_i) \partial_x S_{T_i}}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}. \end{aligned} \quad (5.7)$$

Thus, we have

$$\begin{aligned} \partial_x \mathbb{E}(\phi(S_T)) &= \mathbb{E}(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T=0\}}) + \mathbb{E}(\phi(S_T) \theta \delta_\pi(U) \mathbf{1}_{\{J_T \geq 1\}}) \\ &\quad + \mathbb{E}\left(\phi(S_T) \theta^2 \sum_{i=1}^{J_T} \frac{\pi(\omega, \Delta_i) \pi'(\omega, \Delta_i) \partial_x S_{T_i}}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}\right). \end{aligned}$$

Hence, it remains to compute $\delta_\pi(U)$. Generally, this is not a stochastic integral, but for simple examples as the geometrical and Vasicek models (see Section 5.2 below), it turns out that $\delta_\pi(U)$ is a stochastic integral.

5.2. The Bismut weight computation

Let us compute $\delta_\pi(U)$ when the asset $(S_t)_{t \in [0, T]}$ follows the models (4.1) and (4.2). In both models, the jump amplitudes are standard Gaussian random variables, so we take $\pi(\omega, \Delta_i) = 1$ and (5.5) becomes

$$\phi'(S_T) \partial_x S_T = \frac{\partial_x S_{T_i} D_i \phi(S_T)}{\partial_a c(T_i, \Delta_i, S_{T_i}^-)}.$$

Since $\partial_y \ln p_i(\omega, y) = -y$, we have (see Section 2)

$$\delta(U) = - \sum_{i=1}^{J_T} (\partial_{\Delta_i} U_i - U_i \Delta_i),$$

where

$$U_i = \frac{\partial_x S_T}{D_i S_T}.$$

We consider the same examples as in Section 4.

- **Vasicek model.** By (4.4) and (4.5), we know that $\partial_x S_T = e^{-rT}$ and $D_i S_T = \sigma e^{-r(T-T_i)}$ so that

$$U_i = \frac{e^{-rT_i}}{\sigma}.$$

Hence, $\delta(U) = \frac{1}{\sigma} \sum_{i=1}^{J_T} e^{-rT_i} \Delta_i = \frac{1}{\sigma} \int_0^T \int_{\mathbb{R}} e^{-ru} a N(du, da)$, and we get

$$\mathbb{E}(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}}) = \frac{1}{\sigma} \mathbb{E}\left(\phi(S_T) \mathbf{1}_{\{J_T \geq 1\}} \int_0^T \int_{\mathbb{R}} e^{-ru} a N(du, da)\right).$$

- **Geometrical model.** By (4.9) and (4.13), we know that $\partial_x S_T = \frac{S_T}{x}$ and

$$D_i S_T = \frac{\sigma S_T}{1 + \sigma \Delta_i} \text{ so that}$$

$$U_i = \frac{1 + \sigma \Delta_i}{\sigma x}.$$

Hence,

$$\begin{aligned} \delta(U) &= - \sum_{i=1}^{J_T} \left(\frac{1}{x} - \frac{1 + \sigma \Delta_i}{\sigma x} \Delta_i \right) \\ &= - \frac{J_T}{x} + \frac{1}{\sigma x} \int_0^T \int_{\mathbb{R}} (a + \sigma a^2) N(du, da), \end{aligned}$$

and we get

$$\begin{aligned} \mathbb{E} \left(\phi'(S_T) \partial_x S_T \mathbf{1}_{\{J_T \geq 1\}} \right) &= - \frac{1}{x} \mathbb{E} \left(\phi(S_T) J_T \mathbf{1}_{\{J_T \geq 1\}} \right) \\ &\quad + \frac{1}{\sigma x} \mathbb{E} \left(\phi(S_T) \mathbf{1}_{\{J_T \geq 1\}} \int_0^T \int_{\mathbb{R}} (a + \sigma a^2) N(du, da) \right). \end{aligned}$$

6. American option pricing

Our aim is to evaluate the price $P(0, x)$ of an American option with payoff function ϕ and maturity T when the asset price S is a pure jump diffusion process. We assume that there is no arbitrage opportunity and that the spot rate r is constant. We work under the martingale measure $\mathbb{Q}$ that cancels the drift term of $(\tilde{S}_t)_{t \in [0, T]}$ so that the (risk neutral) dynamic of S_t under $\mathbb{Q}$ is given by

$$\begin{aligned} S_t &= x + \int_0^t g(u, S_u) du + \sum_{i=1}^{J_t} c(T_i, \Delta_i, S_{T_i}^-) \\ &= x + \int_0^t r S_u du + \int_0^t \int_{\mathbb{R}} c(u, a, S_{u-}) \tilde{N}(du, da), \end{aligned} \tag{6.1}$$

where $g(u, S_u) = r S_u - \int_{\mathbb{R}} c(u, a, S_u) \nu(da)$. The function c satisfies Hypothesis 3.1 and 3.12; the jump times $(T_i)_{i \in \mathbb{N}}$ (of intensity λ) and the jump amplitudes $(\Delta_i)_{i \in \mathbb{N}^*}$ are defined in the beginning of Section 3, and the density $\exp(\rho)$ of each amplitude Δ_i satisfies Hypothesis 3.2.

We consider the filtration $(\mathcal{F}_t)$ defined by $\mathcal{F}_t = \sigma(N(s, A), s \leq t, A \in \mathcal{B}(\mathbb{R}))$, where $N(s, A) = \text{Card}(T_i \leq s : \Delta_i \in A)$. Then, the price $P(t, S_t)$ at time t of the American

option of payoff ϕ and maturity T is given by

$$P(t, S_t) = \max_{\tau \in \Gamma_{t,T}} \mathbb{E}_{\mathbb{Q}}(e^{-r(\tau-t)} \phi(S_{\tau}) \mid \mathcal{F}_t), \quad (6.2)$$

where $\Gamma_{t,T}$ denotes the set of stopping times with values in $[t, T]$.

Let us give the dynamic programming equations, which allow us to compute (6.2).

6.1. The algorithm

We fix $L \in \mathbb{N}^*$, and we consider $0 = t_0 < t_1 < \dots < t_L = T$ a discretization grid of $[0, T]$ with step size $\varepsilon_k = t_k - t_{k-1}$.

In order to construct an approximation scheme $\bar{S}_t$ for S_t , we will first construct an approximation scheme for the deterministic curve s_t defined by (3.3). We recall that we are given deterministic times $0 = u_0 < \dots < u_n < T$ and $J_{t_k}(u) = n_k$ if $u_{n_k} \leq t_k < u_{n_k+1}$ and a vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. For $k = 0, \dots, L-1$, we put

$$\bar{s}_{t_k} = x + \sum_{l=1}^k g(t_{l-1}, \bar{s}_{t_{l-1}}) \varepsilon_l + \sum_{l=1}^k \sum_{t_{l-1} < u_i \leq t_l} c(t_{l-1}, a_i, \bar{s}_{t_{l-1}}).$$

Then, we define $\bar{S}_{t_0} = x$ and for all $k = 1, \dots, L$, on $\{J_{t_k} = n_k\}$, $\bar{S}_{t_k} = \bar{s}_{t_k}(T_1, \dots, T_{n_k}, \Delta_1, \dots, \Delta_{n_k})$. We denote $\tau(t) := t_k$ if $t_k < t \leq t_{k+1}$. Then, for all $t \geq 0$, we have

$$\bar{S}_t = x + \int_0^t r \bar{S}_{\tau(s)} ds + \int_0^t \int_{\mathbb{R}} c(\tau(s), a, \bar{S}_{\tau(s)-}) \tilde{N}(ds, da). \quad (6.3)$$

Since $(\bar{S}_{t_k})_{k=0, \dots, L}$ is a Markov chain with respect to $(\mathcal{F}_{t_k})$, the price $P(0, x)$ is approximated by $\bar{P}_0(x)$, where $(\bar{P}_{t_k}(\bar{S}_{t_k})_{k=0, \dots, L})$ is defined by (see NEVEU [1972]):

$$\begin{aligned} \bar{P}_{t_L}(\bar{S}_{t_L}) &= \phi(\bar{S}_T), \\ \bar{P}_{t_k}(\bar{S}_{t_k}) &= \max \{ \phi(\bar{S}_{t_k}), e^{-r \varepsilon_{k+1}} \mathbb{E}_{\mathbb{Q}} [\bar{P}_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mid \bar{S}_{t_k}] \}, k = L-1, \dots, 1 \\ \bar{P}_0(x) &= \max \{ \phi(x), e^{-r \varepsilon_1} \mathbb{E}_{\mathbb{Q}} [\bar{P}_{t_1}(\bar{S}_{t_1})] \}. \end{aligned} \quad (6.4)$$

Using the representation formula for conditional expectations given in Proposition 3.2, we have

$$\mathbb{E}_{\mathbb{Q}} [\bar{P}_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{\{J_{t_{k+1}} - J_{t_k} \geq 1; J_{t_k} \geq 1\}} \mid \bar{S}_{t_k} = \alpha] = \frac{\mathbb{T}[\bar{P}_{t_{k+1}}](\alpha)}{\mathbb{T}[1](\alpha)} \mathbf{1}_{\{J_{t_{k+1}} - J_{t_k} \geq 1; J_{t_k} \geq 1\}},$$

where $\mathbb{T}[f](\alpha)$ is defined by (3.20).

Thus, one may compute $\mathbb{E}_{\mathbb{Q}}[\bar{P}_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mid \bar{S}_{t_k} = \alpha]$ if there is at least one jump on $[0, t_k]$ and at least one jump on $[t_k, t_{k+1}]$. Hence, we approximate the previous algorithm by a localized one: denote

$$A_k := \{J_{t_{k+1}} - J_{t_k} \geq 1; J_{t_k} \geq 1\},$$

and set

$$\begin{aligned} u_{t_L}(\bar{S}_{t_L}) &= \phi(\bar{S}_T), \\ u_{t_k}(\bar{S}_{t_k}) &= \max \left\{ \phi(\bar{S}_{t_k}), e^{-r\varepsilon_{k+1}} \mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{A_k} \mid \bar{S}_{t_k}] \right\}, \\ u_0(x) &= \max \left\{ \phi(x), e^{-r\varepsilon_1} \mathbb{E}_{\mathbb{Q}}[u_{t_1}(\bar{S}_{t_1})] \right\}. \end{aligned} \quad (6.5)$$

Thus, the approximation error between the algorithm (6.4) and the localized one (6.5) is

$$|\bar{P}_0(x) - u_0(x)| \leq C \frac{T}{\underline{\varepsilon}} e^{-\lambda \underline{\varepsilon}},$$

where $C := 2 \max \|\bar{P}_{t_k}\|_{\infty}$ and $\underline{\varepsilon} := \min \varepsilon_k$.

REMARK 6.1. The above evaluation shows that we may not choose a very small $\underline{\varepsilon}$: the intensity λ of the jump process being given, we have to keep $\lambda \underline{\varepsilon}$ large enough to minimize the error coming from the localization. This seems to be intuitive since when λ is large, there are many jumps so that the probability of having at least one jump in a given interval is close to 1.

The conditional expectation $\mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{A_k} \mid \bar{S}_{t_k}]$ will be computed using the representation formula given in Proposition 3.2, by means of suitable empirical means evaluated over N simulated paths. Let us be more precise.

- We fix the intensity of the jumps λ , and we choose a step size $\varepsilon_k = t_k - t_{k-1}$ with respect to λ so that the probability of having at least one jump on $[t_k, t_{k+1}]$ is large (see Remark 6.1).
- We simulate the jump times $(T_i^p)_{i \geq 1}$ (such that $T_i^p - T_{i-1}^p \sim \text{Exp}(\lambda)$) and the jump amplitudes $(\Delta_i^p)_{i \geq 1}$, $p = 1, \dots, N$.
- We then compute the samples $(\bar{S}_k^p, \bar{J}_{t_k}^p)_{k=1, \dots, L}$, $p = 1, \dots, N$.

Let us compute $u_{t_k}(\bar{S}_{t_k}^p)$ given by (6.5).

Using the representation Corollary 3.2, we obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{A_k} \mid \bar{S}_{t_k} = \alpha] \\ = \frac{\mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{\bar{S}_{t_k} \geq \alpha} \bar{V}_{(t_k, t_{k+1})} \mathbf{1}_{A_k}]}{\mathbb{E}_{\mathbb{Q}}[\mathbf{1}_{\bar{S}_{t_k} \geq \alpha} \bar{V}_{(t_k, t_{k+1})} \mathbf{1}_{A_k}]} \mathbf{1}_{A_k} \\ \simeq \frac{\sum_{q=1}^N u_{t_{k+1}}(\bar{S}_{t_{k+1}}^q) \mathbf{1}_{\bar{S}_{t_k}^q \geq \alpha} \bar{V}_{(t_k, t_{k+1})}^q \mathbf{1}_{\{\bar{J}_{t_{k+1}}^q - \bar{J}_{t_k}^q \geq 1; \bar{J}_{t_k}^q > 0\}}}{\sum_{q=1}^N \mathbf{1}_{\bar{S}_{t_k}^q \geq \alpha} \bar{V}_{(t_k, t_{k+1})}^q \mathbf{1}_{\{\bar{J}_{t_{k+1}}^q - \bar{J}_{t_k}^q \geq 1; \bar{J}_{t_k}^q > 0\}}} \mathbf{1}_{A_k}. \end{aligned}$$

We denote by $\Psi_k(\alpha)$ the fraction:

$$\Psi_k(\alpha) := \frac{\sum_{q=1}^N u_{t_{k+1}}(\bar{S}_{t_{k+1}}^q) \mathbf{1}_{\bar{S}_{t_k}^q \geq \alpha} \bar{V}_{(t_k, t_{k+1})}^q \mathbf{1}_{\{\bar{J}_{t_{k+1}}^q - \bar{J}_{t_k}^q \geq 1; \bar{J}_{t_k}^q > 0\}}}{\sum_{q=1}^N \mathbf{1}_{\bar{S}_{t_k}^q \geq \alpha} \bar{V}_{(t_k, t_{k+1})}^q \mathbf{1}_{\{\bar{J}_{t_{k+1}}^q - \bar{J}_{t_k}^q \geq 1; \bar{J}_{t_k}^q > 0\}}}. \quad (6.6)$$

Thus, we have

$$\mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{A_k} \mid \bar{S}_{t_k} = \alpha] \simeq \Psi_k(\alpha) \mathbf{1}_{\{\bar{J}_{t_{k+1}} - \bar{J}_{t_k} \geq 1; \bar{J}_{t_k} > 0\}}.$$

Applying this result to $(\bar{J}_{t_k}^p)_{k=0, \dots, L}$ and $\alpha = \bar{S}_{t_k}^p$, we obtain for $k = L - 1, \dots, 1$,

$$\mathbb{E}_{\mathbb{Q}}[u_{t_{k+1}}(\bar{S}_{t_{k+1}}) \mathbf{1}_{A_k} \mid \bar{S}_{t_k} = \bar{S}_{t_k}^p] \simeq \Psi_k(\bar{S}_{t_k}^p) \mathbf{1}_{\{\bar{J}_{t_{k+1}}^p - \bar{J}_{t_k}^p \geq 1; \bar{J}_{t_k}^p > 0\}}.$$

Hence, we can set up the dynamic programming equation:

$$\hat{u}_{t_L}(\bar{S}_{t_L}^p) = \phi(\bar{S}_T), \text{ and for } k = L - 1, \dots, 1,$$

$$\hat{u}_{t_k}(\bar{S}_{t_k}^p) = \max \left\{ \phi(\bar{S}_{t_k}^p), e^{-r \varepsilon_{k+1}} \Psi_k(\bar{S}_{t_k}^p) \mathbf{1}_{\{\bar{J}_{t_{k+1}}^p - \bar{J}_{t_k}^p \geq 1; \bar{J}_{t_k}^p > 0\}} \right\}, \quad (6.7)$$

$$\text{then, } \hat{u}_0(x) = \max \left\{ \phi(x), e^{-r \varepsilon_1} \frac{1}{N} \sum_{p=1}^N \hat{u}_{t_1}(\bar{S}_{t_1}^p) \right\}.$$

6.2. Numerical results

In this section, we deal with the geometrical model:

$$S_t = x + \int_0^t r S_u du + \int_0^t \int_{\mathbb{R}} \sigma a S_u^- N(du, da), \quad t \in [0, T], \quad (6.8)$$

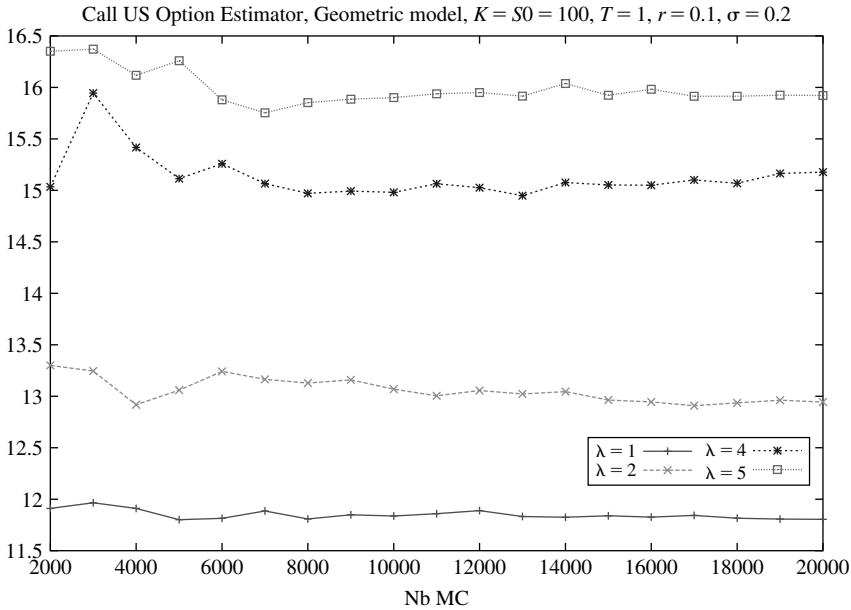


FIG. 6.1 Price of American call options for various jump intensities. Geometrical model.

where $N(t, A) = \text{Card}\{T_i \leq t : \Delta_i \in A\}$. We suppose that $T_i - T_{i-1} \sim \text{Exp}(\lambda)$ for all $i \geq 1$ and that Δ_i has a uniform law on $(0, 1)$. Hence, in view of (3.10), we work with the weights

$$\pi_{(s,t)}(\omega, \Delta_i) := \mathbf{1}_{[s,t]}(T_i) \pi(\Delta_i), \text{ for } 0 \leq s < t \leq T,$$

where

$$\pi(\Delta_i) = (1 - \Delta_i)^{1/4} \Delta_i^{1/4}.$$

Our aim is to perform dynamic programming equation to approximate the price $P(0, x)$. In Eq. (6.6), the function Ψ_k depends on the Malliavin estimator $V_{(s,t)}$ given by (3.14). Hence, we have to compute the Malliavin operators of S_t involved in this expression. Let $(S_t)_{t \in [0, T]}$ be the solution of the geometrical model (6.8). Since we have an explicit expression of S_t (see (4.8)) for all $t \in [0, T]$, the process S can be exactly simulated at each time t_k , and we do not need an approximation $\tilde{S}_{t_k}$ of S_{t_k} . Let us give the expression of $V_{(s,t)}$ (for detailed computations, see BAVOUZET [2006]). For $0 \leq s \leq t$, we denote

$$F_{(s,t)} := \sum_{i=0}^{\infty} \mathbf{1}_{[s,t]}(T_i) \frac{\pi'(\Delta_i)}{1 + \sigma \Delta_i},$$

$$\begin{aligned}
A_{(s,t)} &:= \sum_{i=0}^{\infty} \mathbf{1}_{[s,t]}(T_i) \frac{\pi(\Delta_i)}{(1 + \sigma \Delta_i)^2}, \\
B_{(s,t)} &:= \sum_{i=0}^{\infty} \mathbf{1}_{[s,t]}(T_i) \frac{\pi(\Delta_i) \pi'(\Delta_i)}{(1 + \sigma \Delta_i)^3}, \\
C_{(s,t)} &:= \sum_{i=0}^{\infty} \mathbf{1}_{[s,t]}(T_i) \frac{\pi(\Delta_i)^2}{(1 + \sigma \Delta_i)^4}.
\end{aligned}$$

We then have

$$\begin{aligned}
V_{(s,t)} &= \frac{1}{S_s} + \frac{1}{\sigma S_s} \left(\frac{B_{(0,s)} - 2 \sigma C_{(0,s)}}{A_{(0,s)}^2} - \frac{B_{(s,t)} - 2 \sigma C_{(s,t)}}{A_{(s,t)}^2} \right) \\
&\quad + \frac{1}{\sigma S_s} \left(\frac{F_{(s,t)}}{A_{(s,t)}} - \frac{F_{(0,s)}}{A_{(0,s)}} \right). \tag{6.9}
\end{aligned}$$

Figure and comments.

We compute the price of the American call option of maturity $T = 1$ and strike $K = 100$ when the asset $(S_t)_{t \in [0, T]}$ follows the geometrical model (6.8).

Figure 6.1 shows several values of prices corresponding to different jump intensities $\lambda = 1, 2, 4, 5$.

We can observe that the price increases when the jump intensity increases as well, which seems to be intuitive since the jump intensity λ represents the noise available in the system (see Remark 6.1).

References

- BALLY, V., BAVOUZET, M.P., MESSAOUD, M. (2005). Integration by parts for locally smooth laws and applications to sensitivity computations. Inria research report, RR-5567, Inria, Rocquencourt, France.
- BALLY, V., CARAMELLINO, L., ZANETTE, A. (2003) Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach. Inria research report, RR-4804, Inria, Rocquencourt, France.
- BAVOUZET, M.P. (2006). Minoration de densité pour les diffusions à sauts. Calcul de Malliavin pour processus de sauts purs, applications à la Finance. Thesis, Dauphine university.
- BAVOUZET, M.P., MESSAOUD, M. (2006). Computation of Greeks using Malliavin calculus in jump type market models. *Electron. J. Probab* **11**, 276–300.
- BAVOUZET, M.P., MESSAOUD, M. (2006). Pricing and sensitivity computations of American options in one-dimensional jump type market model. Research report, Inria, Rocquencourt, France.
- BERTOIN, J. (1996). *Lévy Processes*. (Cambridge University Press).
- BIAGINI, F., ØKSENDAL, B., SULEM, A., WALLNER, N. (2004). An introduction to white noise and Malliavin calculus for fractional Brownian motion. *Proc. Roy. Soc., special issue on stochastic analysis and applications* **460**, 347–372.
- BICHTLER, K., GRAVEREAUX, J.B., JACOD, J. (1987). *Malliavin Calculus for Processes with Jumps* (Gordon and Breach).
- BOULEAU, N. (2003). *Error Calculus for Finance and Physics, the Language of Dirichlet Forms* (De Gruyter).
- CARLEN, E.A., PARDOUX, E. (1990). Differential calculus and integration by parts on Poisson space. In: *Stochastics, Algebra and Analysis in Classical and Quantum Dynamics* (Kluwer), pp. 63–73.
- CONT, R., TANKOV, P. (2003). *Financial Modelling with Jump Processes* (Chapman and Hall/CRC Press).
- DAVIS, M.H.A., JOHANSSON, M. (2006). Malliavin Monte Carlo Greeks for jump diffusions. *Stoch. Proc. Appl.* **116**, 101–129.
- DENIS, L. (2000). A criterion of density for solutions of poisson driven SDE's. *Probab. Theory. Rel.* **118**, 406–426.
- FORSTER, B., LTKEBOHMERT, E., TEICHMANN, J. (2005). Calculation of Greeks for jump-diffusions. Submitted.
- FOURNIÉ, E., LASRY, J.M., LEBOUCHOUX, J., LIONS, P.L. (2001). Applications of Malliavin Calculus to Monte Carlo methods in finance II. *Financ. Stoch.* **2**, 73–88.
- FOURNIÉ, E., LASRY, J.M., LEBOUCHOUX, J., LIONS, P.L., TOUZI, N. (1999). Applications of Malliavin calculus to Monte Carlo methods in finance. *Financ. Stoch.* **5** (2), 201–236.
- IKEDA, N., WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, Netherlands).
- EL KHATIB, Y., PRIVAULT, N. (2004). Computation of Greeks in a market with jumps via the Malliavin calculus. *Financ. Stoch.* **8**, 161–179.
- ØKSENDAL, B. (1996). An introduction to Malliavin calculus with applications to Economics. In: *Lecture notes, Norwegian School of Economics and Business Administration*, Bergen, Norway.
- LIONS, P.-L., REGNIER, H. (2000). Calcul du prix et des sensibilités d'une option américaine par une méthode de monte carlo. Technical report, Ceremade, Paris, France.
- NEVEU, J. (1972). *Martingales à Temps Discret* (Masson).
- NUALART, D., VIVES, J. (1990). Anticipative calculus for the Poisson process based on the Fock space. In: *sem Proba. XXIV*. In: *Lecture Notes in Mathematics* **1426** (Springer), pp. 154–165.
- DI NUNNO, G., ØKSENDAL, B., PROSKE, F. (2004). White noise analysis for Lévy processes. *J. Funct. Anal.* **206**, 109–148.

- PICARD, J. (1996). Formules de dualité sur l'espace de Poisson. *Ann. I. H. Poincaré-PR.* **32** (4), 509–548.
- PICARD, J. (1996). On the existence of smooth densities for jump processes. *Probab. Theory. Rel.* **105**, 481–511.
- PRIVAULT, N. (1999). A calculus on Fock space and its probabilistic interpretation. *Bull. Sci. Math.* **123**, 97–114.
- PRIVAULT, N., WEI, X. (2004). A Malliavin calculus approach to sensitivity analysis in insurance. *Insur. Math. Econ.* **35**, 679–690.
- PRIVAULT, N., WEI, X. (2005). Integration by parts for point processes and Monte Carlo estimation, Preprint.
- VIVES, J., LEÓN, J.A., UTZET, F., SOLÉ, J.L. (2002). On Lévy processes, Malliavin calculus and market models with jumps. *Financ. Stoch.* **6** (2), 197–225.

On the Discrete Time Capital Asset Pricing Model

Alain Bensoussan

*International Center for Decision and Risk Analysis, ICDRiA, School of Management,
University of Texas at Dallas, P.O.Box 830688, SM30, Richardson, TX 75083-0688, USA
E-mail: Alain.Bensoussan@utdallas.edu*

Abstract

We give in this chapter a presentation of the capital asset pricing model in discrete time. The presentation is usually done in continuous time. However, the discrete time model is not just a discrete time version of the continuous time model. Some significant differences occur. They are related to the fact that the usual assumption of complete markets is not satisfied in discrete time unless the randomness is modelled by a finite number of events.

1. Introduction

The capital asset pricing model is a major theory in modern finance. The concept of price of state of nature is a concept of stochastic economics introduced by ARROW and DEBREU [1954]. When there are only a finite number of states of nature and when the assumption of complete markets is valid, these prices play a significant role in obtaining the optimal consumption and portfolio investment policies in a general setup. This fact is well known for a single-period problem. A dynamic version of this model, called the intertemporal capital asset pricing model, has been introduced by MERTON [1973]. Curiously, this dynamic version is in continuous time and not in discrete time. The theory has been further developed in continuous time, using properties of martingale generated by Wiener processes (see KARATZAS and SHREVE [1998]). In discrete time, although natural, the corresponding treatment is not widely available in the literature (see SHREVE [2004] for a partial treatment in the case of the binomial model). We present

here a detailed theory of what can be done in discrete time. The situation of a finite number of random events at each period is compatible with an assumption of complete markets. However, the analogue of the continuous time model when the randomness is modelled by a finite number of Wiener processes is not. A natural discretization of the continuous time model introduces a situation of incompleteness. The author is grateful to Steve Shreve for illuminating discussions and information.

2. A probability setup

2.1. General framework

The time is described by $0, 1, \dots, T$. We consider a probability space Ω made of a finite number of single events

$$\omega = (\omega_{j_1}^1, \dots, \omega_{j_T}^T),$$

where

$$j_1, \dots, j_T \in \{1, \dots, n\}.$$

We interpret $\omega_{j_t}^t$, $j_t \in \{1, \dots, n\}$ as the states of nature that can occur at time t . There are n new possibilities at each time. Globally, there are n^T single events ω . To each single event is attached a probability

$$p_{j_1, \dots, j_T} = \text{Prob}\{(\omega_{j_1}^1, \dots, \omega_{j_T}^T)\}.$$

We assume

$$p_{j_1, \dots, j_T} > 0, \forall j_1, \dots, j_T \in \{1, \dots, n\}. \quad (2.1)$$

We naturally have

$$\sum_{j_1, \dots, j_T=1}^n p_{j_1, \dots, j_T} = 1.$$

This probability distribution on Ω is denoted by $P = P^T$. We introduce next

$$\tilde{\omega}^t = (\omega_{j_1}^1, \dots, \omega_{j_t}^t) = \bigcup_{j_{t+1}, \dots, j_T=1}^n (\omega_{j_1}^1, \dots, \omega_{j_T}^T).$$

These subsets of Ω represent events that have a common history up to time t . Define next

$$\mathcal{F}^t = \sigma - \text{algebra on } \Omega \text{ generated by unions of sets } \tilde{\omega}^t.$$

Events belonging to $\mathcal{F}^t$ are the only ones that can be observed at time t . Clearly, $\mathcal{F}^t$ is generated by the coordinates

$$\omega \rightarrow \omega_{j_s}^s, s = 1 \dots t, j_s = 1, \dots, n.$$

We note that $\mathcal{F}^T$ is simply the σ -algebra made of all the subsets of Ω and $\mathcal{F}^0 = \{\Omega, \emptyset\}$. Define the numbers

$$p_{j_1, \dots, j_t} = \sum_{j_{t+1}, \dots, j_T=1}^n p_{j_1, \dots, j_t, j_{t+1}, \dots, j_T},$$

which form a probability distribution on $(\Omega, \mathcal{F}^t)$. We denote this probability by P^t . It coincides with P on the σ -algebra $\mathcal{F}^T$. These concepts reduce for $t = 0$ to a single event

$$\tilde{\omega}^0 = \Omega; \quad p_0 = 1.$$

We can compute the conditional probability

$$\begin{aligned} & \text{Prob}((\omega_{j_1}^1, \dots, \omega_{j_{t+1}}^{t+1}) | (\omega_{j_1}^1, \dots, \omega_{j_t}^t)) \\ &= \frac{p_{j_1, \dots, j_{t+1}}}{p_{j_1, \dots, j_t}}, \end{aligned}$$

which we denote to simplify notation

$$\theta(\omega_{j_{t+1}}^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t).$$

Obviously,

$$\sum_{j_{t+1}=1}^n \theta(\omega_{j_{t+1}}^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t) = 1, \forall \omega_{j_1}^1, \dots, \omega_{j_t}^t.$$

These conditional probabilities generate the full probabilistic setup by induction.

2.2. Binomial model

As a particular case, we consider the binomial model developed by SHREVE [2004]. We have $n = 2$, and we assume

$$\theta(\omega_1^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t) = p,$$

$$(\omega_2^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t) = q = 1 - p,$$

for all values $j_1, \dots, j_t \in \{1, 2\}$. Therefore,

$$p_{j_1, \dots, j_t} = p^{\#1(j_1, \dots, j_t)} q^{\#2(j_1, \dots, j_t)},$$

where

$$j_1, \dots, j_t \in \{1, 2\},$$

$$\#1(j_1, \dots, j_t) = \text{number of 1 in } (j_1, \dots, j_t),$$

$$\#2(j_1, \dots, j_t) = \text{number of 2 in } (j_1, \dots, j_t).$$

3. Description of the market

3.1. Prices of states of nature

The market is made of $n - 1$ assets whose prices at time t are represented by

$$Y(t; \omega) = Y(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t), \mathcal{F}^t \text{ measurable.}$$

The price of cash at time t is simply R^t , where $R = 1 + r$ and r is the discount rate. The price of cash at time 0 is 1. At time 0, the prices of assets $Y(0)$ are deterministic.

We make the following essential assumption of market completeness: for any fixed $\omega_{j_1}^1, \dots, \omega_{j_t}^t$, the matrix

$$\begin{pmatrix} Y_1(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1}) \dots Y_1(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_n^{t+1}) \\ \dots \\ Y_{n-1}(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1}) \dots Y_{n-1}(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_n^{t+1}) \\ 1 \dots 1 \end{pmatrix} \quad (3.1)$$

is invertible.

Therefore, there exists a unique random variable

$$\psi(t; \omega) = \psi(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t)$$

such that

$$\begin{aligned} \sum_{j_{t+1}=1}^n \psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_{j_{t+1}}^{t+1}) Y(t+1; \omega_{j_1}^1, \dots, \omega_{j_{t+1}}^{t+1}) \\ = Y(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t), \forall \omega_{j_1}^1, \dots, \omega_{j_t}^t \end{aligned} \quad (3.2)$$

$$\sum_{j_{t+1}=1}^n \psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_{t+1}}^{t+1}) = \frac{1}{R}$$

Following ARROW and DEBREU [1954] terminology, we call $\psi(t; \omega)$ the *prices of the states of nature* at time t . They are $\mathcal{F}^t$ -measurable real random variables. At time t , $(\omega_{j_1}^1, \dots, \omega_{j_t}^t)$ is known, and there are n new states of nature

$$\omega_{j_{t+1}}^{t+1} \in \{1, \dots, n\}$$

to which correspond prices

$$\psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_{j_{t+1}}^{t+1}).$$

Consider the binomial model. There is a single asset for which we assume the following evolution of prices

$$Y(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1}) = uY(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t),$$

$$Y(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_2^{t+1}) = dY(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t),$$

with

$$d < R < u.$$

The assumption 3.2 is satisfied with

$$\begin{aligned} \psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1}) &= \frac{R-d}{R(u-d)} = \frac{\tilde{p}}{R} \\ \psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_2^{t+1}) &= \frac{u-R}{R(u-d)} = \frac{\tilde{q}}{R}. \end{aligned} \tag{3.3}$$

3.2. Risk-neutral probability

We define next a new probability on $\Omega, \mathcal{F}^T$ as follows

$$\hat{p}_{j_1, \dots, j_T} = R^T \prod_{t=1}^T \psi(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t). \tag{3.4}$$

We call this new probability $\hat{P}$. We associate probabilities on $\Omega, \mathcal{F}^t$, called $\hat{P}^t$ defined by

$$\hat{p}_{j_1, \dots, j_t} = R^t \prod_{s=1}^t \psi(s; \omega_{j_1}^1, \dots, \omega_{j_s}^s) \tag{3.5}$$

Clearly,

$$\frac{\hat{P}_{j_1, \dots, j_{t+1}}}{\hat{P}_{j_1, \dots, j_t}} = R\psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_{t+1}}^{t+1})$$

$$\sum_{j_{t+1}=1}^n \hat{P}_{j_1, \dots, j_{t+1}} = \hat{P}_{j_1, \dots, j_t}.$$

As P^t for P , $\hat{P}^t$, and $\hat{P}$ coincide on the σ -algebra $\mathcal{F}^t$. Finally, we define the process

$$Z(t) = Z(t; \omega) = Z(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t) = \frac{\hat{P}_{j_1, \dots, j_t}}{P_{j_1, \dots, j_t}}. \quad (3.6)$$

We set $Z(0) = 1$.

The process $Z(t)$ can be viewed as the Radon–Nikodym derivative

$$\frac{d\hat{P}}{dP}|_{\mathcal{F}^t} = Z(t).$$

By construction, $Z(t)$ is a $P, \mathcal{F}^t$ martingale. It is easy to check the relation

$$\frac{Z(t+1)}{Z(t)} = \frac{R\psi(t+1; \omega_{j_1}^1, \dots, \omega_{j_{t+1}}^{t+1})}{\theta(\omega_{j_{t+1}}^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t)}. \quad (3.7)$$

Moreover, we can assert

PROPOSITION 3.1. *The following relation holds*

$$\hat{E}[Y(t+1)|\mathcal{F}^t] = RY(t),$$

which can be expressed as: The process

$$\frac{Y(t)}{R^t} \text{ is a } \hat{P}, \mathcal{F}^t \text{ martingale.}$$

The proof is an easy consequence of the definitions of $\hat{P}$ and the prices of states of nature $\psi(t; \omega)$. The probability $\hat{P}$ is called the risk-neutral probability. So, under the risk-neutral probability, the discounted asset prices process is a martingale.

3.3. Replication portfolio

Suppose we have a process scalar $V(t)$ that adapted to the filtration $\mathcal{F}^t$. So,

$$V(t) = V(t; \omega) = V(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t).$$

In addition, we assume

$$\hat{E}[V(t+1)|\mathcal{F}^t] = 0. \quad (3.8)$$

A **replication portfolio** is a $n - 1$ dimensional process $\pi(t)$, which is adapted to the filtration $\mathcal{F}^t$ and satisfies

$$V(t+1) = \pi(t) \cdot (Y(t+1) - RY(t)), \forall t = 0, \dots, T-1. \quad (3.9)$$

We have the following result

PROPOSITION 3.2. *We assume that the matrix 3.1 is invertible (market completeness). For any process $V(t)$ such that 3.8 is satisfied, there exists a unique replication portfolio.*

PROOF. Clearly, condition (3.8) is necessary as a consequence of Proposition 3.1. The result follows from well-known properties of matrices, which we recall as follows.

Let

$$a_{ij}, i = 1, \dots, n-1; j = 1, \dots, n$$

be a matrix such that

$$\begin{pmatrix} a_{1,1} \dots a_{1,n} \\ \dots \\ a_{n-1,1} \dots a_{n-1,n} \\ 1 \dots 1 \end{pmatrix}$$

is invertible. Let p_j be the unique solution of the system

$$\sum_{j=1}^n a_{i,j} p_j = c_i$$

$$\sum_{j=1}^n p_j = 1.$$

Consider next the dual system

$$\sum_{i=1}^{n-1} \pi_i (a_{ij} - c_i) = b_j, j = 1 \dots n.$$

Such a system has one and only one solution π_i , provided the condition

$$\sum_{j=1}^n p_j b_j = 0$$

is satisfied. It is defined by taking an arbitrary vector of R^{n-1} ,

$$\eta_1, \dots, \eta_{n-1}$$

solving

$$\sum_{j=1}^n a_{ij} \tilde{\xi}_j = \eta_i,$$

$$\sum_{j=1}^n \tilde{\xi}_j = 0$$

and setting

$$\sum_{j=1}^n \tilde{\xi}_j b_j = \sum_{i=1}^{n-1} \pi_i \eta_i.$$

Since the η_i is arbitrary, this defines the π_i in a unique way, and we obtain a unique solution. $\square$

Consider the binomial model again. There is one asset, and we have the evolution described in Section 2.2. We write

$$Y(t+1; 1) = Y(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, 1),$$

$$Y(t+1; 2) = Y(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, 2),$$

and

$$Y(t) = Y(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t).$$

We want to find that

$$\pi(t) = \pi(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t)$$

such that

$$\pi(t)(Y(t+1; 1) - RY(t)) = V(t+1; 1),$$

$$\pi(t)(Y(t+1; 2) - RY(t)) = V(t+1; 2),$$

Recalling

$$Y(t+1; 1) = uY(t),$$

$$Y(t+1; 2) = dY(t),$$

we check immediately the necessary condition

$$\frac{V(t+1; 1)}{V(t+1; 2)} = \frac{u - R}{d - R} = -\frac{\tilde{q}}{\tilde{p}},$$

and we obtain the formula

$$\pi(t) = \frac{V(t+1; 2) - V(t+1; 1)}{Y(t+1; 2) - Y(t+1; 1)}.$$

4. Optimal portfolio and consumption

4.1. Setting of the problem

We define two utility functions $U_1(x)$ and $U_2(x)$ satisfying

$$\begin{aligned} U_1, U_2 \text{ nondecreasing and concave} \\ U'_i(0) = \infty; \quad U'_i(\infty) = 0. \end{aligned} \tag{4.1}$$

These functions will be differentiable in $(0, \infty)$. A consumption process $C(t)$ is a positive $\mathcal{F}^t$ -adapted stochastic process. Similarly, a portfolio $\pi(t)$ is an $\mathcal{F}^t$ -adapted stochastic process with values in R^{n-1} . There is also an $\mathcal{F}^t$ -adapted real stochastic process representing the amount of cash at time t . We call it $\pi_f(t)$. This process will, in fact, cancel out from budget considerations. There is no outside income flux in this model.

The wealth process $X(t)$ satisfies the relations

$$\begin{aligned} X(t+1) &= \pi(t).Y(t+1) + \pi_f(t)R^{t+1}, \\ X(t) &= C(t) + \pi(t).Y(t) + \pi_f(t)R^t \end{aligned}$$

with a given initial wealth $X(0)$, which is a deterministic number. These two relations are self-explanatory. The first one tells that the wealth at time $t+1$ corresponds to the portfolio decided at time t and the amount of cash available at t , with changes of values coming from the market. The second relation is the consequence of the decisions taken at time t . One allocates the wealth that is available between consumption expenses and investment decisions.

We can next eliminate $\pi_f(t)$, and we obtain the evolution relation

$$X(t+1) = R(X(t) - C(t)) + \pi(t).(Y(t+1) - RY(t)). \tag{4.2}$$

The problem is stated as follows: maximize

$$J_{X(0)}(C(.), \pi(.)) = E \left[\sum_{t=0}^{T-1} \frac{U_1(C(t))}{R^t} + \frac{U_2(X(T))}{R^T} \right]. \tag{4.3}$$

4.2. Martingale considerations

Define

$$\zeta(t) = \frac{Z(t)}{R^t},$$

where $Z(t)$ has been defined by Eq. (3.6). We introduce the process

$$M(t) = X(t)\zeta(t) + \sum_{s=0}^{t-1} C(s)\zeta(s). \quad (4.4)$$

We have

$$M(0) = X(0).$$

We have the proposition

PROPOSITION 4.1. *The process $M(t)$ is a $P, \mathcal{F}^t$ martingale.*

PROOF. From Proposition 3.1, we know that

$$\frac{Y(t)}{R^t} \text{ is a } \hat{P}, \mathcal{F}^t \text{ martingale,}$$

which implies that

$$Y(t)\zeta(t) \text{ is a } P, \mathcal{F}^t \text{ martingale.}$$

We also have

$$E[\zeta(t+1)|\mathcal{F}^t] = \frac{\zeta(t)}{R}.$$

Using the wealth evolution Eq. (4.2) and the definition of $M(t)$, the property is obtained easily. □

From the martingale property, we derive

$$E[X(T)\zeta(T) + \sum_{t=0}^{T-1} C(t)\zeta(t)] = X(0). \quad (4.5)$$

4.3. Optimality conditions

From Eq. (4.5), we can consider the optimization problem with constraints

$$\begin{aligned} & \text{Maximize} \\ & E \left[\sum_{t=0}^{T-1} \frac{U_1(C(t))}{R^t} + \frac{U_2(X(T))}{R^T} \right] \\ & \text{with the constraint} \\ & E \left[X(T)\zeta(T) + \sum_{t=0}^{T-1} C(t)\zeta(t) \right] = X(0). \end{aligned} \quad (4.6)$$

If we consider in the problem (4.6) $X(T)$ and the consumptions $C(t)$ as the variables to be optimized and as unrelated quantities, we get a larger maximum, since we do not pay attention to the relation describing the wealth evolution (see (4.2)).

This problem is a standard maximization problem with concave payoff and linear constraint. It is easily solved by introducing a Lagrange multiplier λ as follows: maximize

$$E \left[\sum_{t=0}^{T-1} \left(\frac{U_1(C(t))}{R^t} - \lambda C(t)\zeta(t) \right) + \left(\frac{U_2(X(T))}{R^T} - \lambda X(T)\zeta(T) \right) \right].$$

We express that the gradients of this quantity with respect to $X(T)$ and $C(t)$ vanish at the optimum $\hat{X}(T)$ and $\hat{C}(t)$. Recalling the definition of $\zeta(t)$, we deduce easily

$$\begin{aligned} U'_2(\hat{X}(T)) &= \lambda Z(T), \\ U'_1(\hat{C}(t)) &= \lambda Z(t). \end{aligned} \quad (4.7)$$

Define I_1 and I_2 to be the inverses of U'_1 and U'_2 , respectively. I_1 and I_2 are decreasing functions since U_1 and U_2 are concave. The relations (4.7) imply

$$\begin{aligned} \hat{X}(T) &= I_2(\lambda Z(T)), \\ \hat{C}(t) &= I_1(\lambda Z(t)). \end{aligned} \quad (4.8)$$

To get the value of λ , we will use the constraint. Consider

$$\begin{aligned} \hat{M}(T) &= \hat{X}(T)\zeta(T) + \sum_{t=0}^{T-1} \hat{C}(t)\zeta(t), \\ E[I_2(\lambda Z(T))\zeta(T) + \sum_{t=0}^{T-1} I_1(\lambda Z(t))\zeta(t)] &= X(0). \end{aligned} \quad (4.9)$$

From the assumptions on $U'_i(0)$, $U'_i(\infty)$, we can assert that the left-hand side of Eq. (4.9) is a monotone decreasing function of λ , which is equal to ∞ when $\lambda = 0$ and equal to 0 when $\lambda = \infty$. Therefore, the equation has a unique solution $\hat{\lambda}$.

4.4. Solution

The consumption $\hat{C}(t)$ computed by the second Eq. (4.8), with $\lambda = \hat{\lambda}$, will be optimal if one can find a portfolio $\hat{\pi}(t)$ that achieves the wealth $\hat{X}(t)$ at time t . Knowing $\hat{M}(T)$, we compute $\hat{M}(t)$ by the martingale property

$$\hat{M}(t) = E[\hat{M}(T)|\mathcal{F}^t].$$

This relation implies

$$\begin{aligned} \hat{M}(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t) = \\ \sum_{j_{t+1}=1}^n \hat{M}(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_{j_{t+1}}^{t+1}) \theta(\omega_{j_{t+1}}^{t+1} | \omega_{j_1}^1, \dots, \omega_{j_t}^t). \end{aligned} \quad (4.10)$$

Once we know $\hat{M}(t)$ and the optimal consumptions $\hat{C}(t)$, we can compute the expression of $\hat{X}(t)$, the optimal wealth process. We then define

$$\hat{V}(t+1) = \hat{X}(t+1) - R(\hat{X}(t) - \hat{C}(t)).$$

To find the optimal portfolio, we then solve the replication Eq. (3.9)

$$\hat{V}(t+1) = \hat{\pi}(t) \cdot (Y(t+1) - RY(t)),$$

which has a unique solution from Proposition 3.2.

4.5. Application to the binomial problem

We apply the preceding results in the binomial model. From Eq. (3.7), we deduce

$$\begin{aligned} \frac{Z(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1})}{Z(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t)} &= \frac{\tilde{p}}{p}, \\ \frac{Z(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_2^{t+1})}{Z(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t)} &= \frac{\tilde{q}}{q}. \end{aligned}$$

Hence,

$$\begin{aligned} Z(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t) = \\ \left(\frac{\tilde{p}}{p} \right)^{\#1(j_1, \dots, j_t)} \left(\frac{\tilde{q}}{q} \right)^{\#2(j_1, \dots, j_t)}. \end{aligned}$$

Eq. (4.9), giving the optimal value of λ , is written as

$$\begin{aligned} & \sum_{j_1, \dots, j_T=1,2} I_2 \left(\lambda \left(\frac{\tilde{p}}{p} \right)^{\#1(j_1, \dots, j_T)} \left(\frac{\tilde{q}}{q} \right)^{\#2(j_1, \dots, j_T)} \right) \frac{\tilde{p}^{\#1(j_1, \dots, j_T)} \tilde{q}^{\#2(j_1, \dots, j_T)}}{R^T} \\ & + \sum_{t=0}^{T-1} \sum_{j_1, \dots, j_t=1,2} I_1 \left(\lambda \left(\frac{\tilde{p}}{p} \right)^{\#1(j_1, \dots, j_t)} \left(\frac{\tilde{q}}{q} \right)^{\#2(j_1, \dots, j_t)} \right) \\ & \frac{\tilde{p}^{\#1(j_1, \dots, j_t)} \tilde{q}^{\#2(j_1, \dots, j_t)}}{R^t} = X(0). \end{aligned}$$

We then compute

$$\hat{M}(T) = I_2(\hat{\lambda} Z(T)) \frac{Z(T)}{R^T} + \sum_{t=0}^{T-1} I_1(\hat{\lambda} Z(t)) \frac{Z(t)}{R^t}$$

and successively

$$\begin{aligned} \hat{M}(t; \omega_{j_1}^1, \dots, \omega_{j_t}^t) &= \hat{M}(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_1^{t+1})p \\ &+ \hat{M}(t+1; \omega_{j_1}^1, \dots, \omega_{j_t}^t, \omega_2^{t+1})q. \end{aligned}$$

This leads to $\hat{X}(t)$, and the amount invested in the asset Y is

$$\hat{\pi}(t) = \frac{\hat{X}(t+1; 2) - \hat{X}(t+1; 1)}{Y(t+1; 2) - Y(t+1; 1)},$$

where we have omitted to write explicitly $\omega_{j_1}^1, \dots, \omega_{j_t}^t$.

5. Dynamic programming approach

5.1. Notation and setting

We introduce the family of problems

$$J_{x,t}(C(\cdot), \pi(\cdot)) = E \left[\sum_{s=t}^{T-1} \frac{U_1(C(s))}{R^{s-t}} + \frac{U_2(X(T))}{R^{T-t}} \mid X(t) = x, \mathcal{F}^t \right] \quad (5.1)$$

This payoff depends only on $C(t), \pi(t); \dots; C(T-1), \pi(T-1)$ and the events $\omega_{j_1}^1, \dots, \omega_{j_t}^t$. Our original problem corresponds to the case $X(0) = x$, and $Y(0)$ is a deterministic quantity. We are interested in the random function

$$W(x, t) = \max J_{x,t}(C(\cdot), \pi(\cdot)). \quad (5.2)$$

Note that $W(x, t)$ is $\mathcal{F}^t$ measurable.

5.2. Dynamic programming equation

It is convenient to denote by $X_{x,t}(s)$, $s \geq t$ the evolution of the wealth process for a given initial wealth at time t equal to x . An optimal policy of consumption and investment depends also on x, t and is denoted by

$$\hat{C}_{x,t}(s); \hat{\pi}_{x,t}(s),$$

and the corresponding wealth process is denoted by $\hat{X}_{x,t}(s)$. We note the consistency relations

$$\begin{aligned} \hat{C}_{x,t}(s+1) &= \hat{C}_{\hat{X}_{x,t}(s)}(s+1), \\ \hat{X}_{x,t}(s+1) &= \hat{X}_{\hat{X}_{x,t}(s)}(s+1). \end{aligned}$$

Bellman equation writes

$$\begin{aligned} W(x, t) &= \max_{C(t), \pi(t)} \{U_1(C(t)) + \frac{1}{R} E[W(X_{x,t}(t+1), t+1) | \mathcal{F}^t]\} \\ W(x, T) &= U_2(x). \end{aligned} \quad (5.3)$$

5.3. Derivative of the value function

If we call $W(x, t)$ the value function, which is a process adapted to the filtration $\mathcal{F}^t$, then its derivative $W_x(x, t)$ with respect to x will have an interesting interpretation, reminiscent of the Lagrange multiplier introduced in Section 4.3. We first differentiate Eq. (5.3) to obtain

$$W_x(x, t) = E[\hat{W}_x(X_{x,t}(t+1), t+1) | \mathcal{F}^t]. \quad (5.4)$$

Next, we recall that the wealth process satisfies a martingale property as follows:

$$E[X_{x,t}(t+1) \frac{Z(t+1)}{R} | \mathcal{F}^t] + Z(t)C(t) = xZ(t), \quad (5.5)$$

which is adapted from Proposition 4.1. This constraint holds for any pair $C(t), \pi(t)$. We proceed as in Section 4.3 and look for optimal

$$\hat{C}_{x,t}(t), \hat{X}_{x,t}(t+1)$$

in the right-hand side of Bellman equation by minimizing

$$U_1(C(t)) + \frac{1}{R} E[W(X_{x,t}(t+1), t+1) | \mathcal{F}^t]$$

under the constraint (5.5). Therefore, we introduce a Lagrange multiplier $\lambda_{x,t}$, which is $\mathcal{F}^t$ measurable and such that the optimum is obtained by minimizing

$$\frac{1}{R} E \left[W(X_{xt}(t+1), t+1) - \lambda_{xt} X_{xt}(t+1) Z(t+1) | \mathcal{F}^t \right] \\ + U_1(C(t)) - \lambda_{xt} Z(t) C(t)$$

in $C(t)$, $X_{xt}(t+1)$.

The optimal $\pi(t)$ is obtained later as indicated in Section 4.4 using the replication Eq. (3.9). We, thus, write the conditions

$$U'_1(\hat{C}_{xt}(t)) = \lambda_{xt} Z(t), \\ W_x(\hat{X}_{xt}(t+1), t+1) = \lambda_{xt} Z(t+1). \quad (5.6)$$

Taking into account (5.4), we get also

$$W_x(x, t) = Z(t) \lambda_{xt}, \quad (5.7)$$

which connects the Lagrange multiplier to the gradient in x of the value function.

5.4. Obtaining the derivative of the value function

We can get an explicit formula for $W_x(x, t)$ as it has been done in Section 4.4. We first prove the following result:

PROPOSITION 5.1. *The following relations hold*

$$U'_1(\hat{C}_{xt}(s)) = \frac{Z(s) W_x(x, t)}{Z(t)}, \forall s \geq t, \\ U'_2(\hat{X}_{xt}(T)) = \frac{Z(T) W_x(x, t)}{Z(t)}. \quad (5.8)$$

PROOF. From Eq. (5.6), we can write

$$W_x(\hat{X}_{xt+1}(t+2), t+2) = \lambda_{xt+1} Z(t+2) = \frac{W_x(x, t+1) Z(t+2)}{Z(t+1)}.$$

We may apply this relation with

$$x = \hat{X}_{xt}(t+1).$$

We obtain

$$W_x(\hat{X}_{xt}(t+2), t+2) = \frac{W_x(\hat{X}_{xt}(t+1), t+1) Z(t+2)}{Z(t+1)},$$

and using again (5.6) and (5.7), it follows

$$W_x(\hat{X}_{xt}(t+2), t+2) = \frac{W_x(x, t) Z(t+2)}{Z(t)}.$$

Proceeding by induction, we derive the second relation (5.8). The proof of the first relation (5.8) is done in a similar manner.

From the relation (5.8), we deduce

$$\begin{aligned}\hat{C}_{xt}(s) &= I_1 \left(\frac{Z(s)W_x(x, t)}{Z(t)} \right), \forall s \geq t, \\ \hat{X}_{xt}(T) &= I_2 \left(\frac{Z(T)W_x(x, t)}{Z(t)} \right).\end{aligned}\tag{5.9}$$

We finally write the martingale property

$$\begin{aligned}E \left[I_2 \left(\frac{Z(T)W_x(x, t)}{Z(t)} \right) \frac{Z(T)}{Z(t)R^{T-t}} + \right. \\ \left. \sum_{s=t}^{T-1} I_1 \left(\frac{Z(s)W_x(x, t)}{Z(t)} \right) \frac{Z(s)}{Z(t)R^{s-t}} | \mathcal{F}^t \right] = x.\end{aligned}\tag{5.10}$$

This equation defines uniquely the random variable $\mathcal{F}^t$ -measurable $W_x(x, t)$. We can complete the definition of the optimal feedback

$$\hat{C}_{xt}(t) = I_1(W_x(x, t))$$

and $\hat{\pi}_{xt}(t)$ by using the replication formulas. □

6. Markovian framework

6.1. Setting of the framework

In order to be close to the continuous case, we introduce now periods of length h instead of 1. So, in the sequel, any time t is a multiple of h . The value of cash at time t is still $\exp rt$. We will use the notation

$$\delta f(t) = f(t + h) - f(t).$$

The evolution of prices of assets $Y(t)$ is given by

$$Y_i(t + h) = Y_i(t) \exp[(\alpha_i(t) - \frac{1}{2}a_{ii}(t))h + \sum_{j=1}^n \sigma_{ij}(t)\delta w_j(t)],\tag{6.1}$$

where the processes $w_j(t)$ are independent standard Wiener processes built on an appropriate probability space $\Omega, \mathcal{A}, P$.

We call

$$\mathcal{F}^t = \sigma - \text{algebra generated by } \delta w(s), s = 0, \dots, t - h.$$

We assume that

$$\alpha_i(t), \sigma_{ij}(t)$$

are adapted to the filtration $\mathcal{F}^t$. It is convenient to introduce the process

$$\theta(t) = \sigma^{-1}(t)(\alpha(t) - r/1),$$

so we can write

$$\begin{aligned} Y_i(t+h) \exp -r(t+h) &= Y_i(t) \exp -rt \exp[-\frac{1}{2}ha_{ii}(t) \\ &\quad + \sum_{j=1}^n \sigma_{ij}(t)(\delta w_j(t) + \theta_j(t)h)]. \end{aligned}$$

Define

$$\mu_i(t) = \sum_{s=0}^{t-h} [-\frac{1}{2}ha_{ii}(s) + \sum_{j=1}^n \sigma_{ij}(s)(\delta w_j(s) + \theta_j(s)h)],$$

and thus we get

$$\delta(Y_i(t) \exp -rt) = Y_i(t) \exp -rt(\exp \delta \mu_i(t) - 1). \quad (6.2)$$

Introduce next the process $Z(t)$ defined by

$$Z(t+h) = Z(t) \exp[-\theta(t) \cdot \delta w(t) - \frac{1}{2}h|\theta(t)|^2], \quad Z(0) = 1. \quad (6.3)$$

If we denote the wealth at time t by $X(t)$, we can write

$$X(t+h) = \pi_f(t) \exp r(t+h) + \pi(t) \cdot Y(t+h),$$

$$X(t) = C(t)h + \pi_f(t) \exp rt + \pi(t) \cdot Y(t),$$

where $\pi(t)$ defines the portfolio of investments. So we get the evolution equation

$$\delta(X(t) \exp -rt) = \pi(t) \cdot \delta(Y(t) \exp -rt) - C(t)h \exp -rt. \quad (6.4)$$

It is convenient to introduce $\varpi_i(t)$ defined by

$$\pi_i(t)Y_i(t) = \varpi_i(t)X(t).$$

The evolution of wealth is then governed by the relation

$$\begin{aligned} \delta(X(t) \exp -rt) &= X(t) \exp -rt \sum_i \varpi_i(t) (\exp \delta\mu_i(t) - 1) \\ &\quad - C(t)h \exp -rt. \end{aligned} \quad (6.5)$$

We can also compute

$$\begin{aligned} &E[\delta(Y_i(t) \exp -rt) \delta(Y_j(t) \exp -rt) | \mathcal{F}^t] \\ &= Y_i(t) \exp -rt Y_j(t) \exp -rt E(\exp \delta\mu_i(t) - 1)(\exp \delta\mu_j(t) - 1) \\ &= Y_i(t) \exp -rt Y_j(t) \exp -rt (\exp ha_{ij}(t) - 1). \end{aligned}$$

6.2. Martingale properties

The assumption of complete markets in this framework amounts to

$$\text{the matrix } a_{ij}^h(t) = \exp ha_{ij}(t) - 1 \text{ is invertible.} \quad (6.6)$$

The matrix $a^h(t)$ and its inverse are bounded in time. Noting the equality

$$a_{ij}^h(t) = ha_{ij}(t) \int_0^1 \exp \theta ha_{ij}(t) d\theta,$$

we see that for diagonal matrices, the invertibility of $a^h(t)$ is equivalent to the invertibility of $a(t)$.

We find easily that $Z(t)$ and $Y_i(t)Z(t) \exp -rt$ are $P, \mathcal{F}^t$ martingales.

We then consider the process

$$M(t) = X(t)Z(t) \exp -rt$$

for which we deduce easily that

$$M(t) + \sum_{s=0}^{t-h} hC(s)Z(s) \exp -rs \text{ is a } P\mathcal{F}^t \text{ martingale.} \quad (6.7)$$

Indeed, we can compute

$$\begin{aligned} \delta M(t) + hC(t)Z(t) \exp -rt &= \delta Z(t) \exp -rt (X(t) - hC(t)) \\ &\quad + Z(t)X(t) \exp -rt \sum_i \varpi_i(t) \left[\exp \left(\sum_j (\sigma_{ij}(t) - \theta_j(t)) \delta w_j(t) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} h \sum_j (\sigma_{ij}(t) - \theta_j(t))^2 \right) \right. \\ &\quad \left. - \exp(-\theta(t) \cdot \delta w(t) - \frac{1}{2} h |\theta(t)|^2) \right]. \end{aligned}$$

6.3. Risk-neutral probability

Define a probability on $\Omega, \mathcal{A}$ called $\hat{P}$ defined by the Radon–Nikodym derivative

$$\frac{d\hat{P}}{dP}|_{\mathcal{F}^t} = Z(t).$$

This probability is called the *Risk-neutral probability*. Define next

$$\delta\tilde{w}(t) = \delta w(t) + h\theta(t),$$

then the variables $\delta\tilde{w}(t)$ for $t = 0, \dots, T-h$ are independent gaussian, with 0 mean and covariance matrix hI and $\mathcal{F}^t$ measurable. Indeed, for an arbitrary deterministic function $\lambda(t)$, one checks the relation

$$\hat{E}[\exp i\lambda(t).\delta\tilde{w}(t)|\mathcal{F}^t] = \exp -\frac{1}{2}h|\lambda(t)|^2.$$

It follows that

$Y_i(t) \exp -rt$ is a $\hat{P}, \mathcal{F}^t$ martingale,

$X(t) \exp -rt + \sum_{s=0}^{t-h} hC(s) \exp -rs$ is a $\hat{P}, \mathcal{F}^t$ martingale.

We can also check the important relation

$$\begin{aligned} \hat{E} \left[\frac{\delta(Y_i(t) \exp -rt)}{Y_i(t) \exp -rt} \frac{\delta(X(t) \exp -rt) + hC(t) \exp -rt}{X(t) \exp -rt} | \mathcal{F}^t \right] \\ = \sum_j \varpi_j(t) a_{ij}^h(t). \end{aligned} \quad (6.8)$$

6.4. Approximate replication property

Suppose we have a process $X(t)$ adapted to the filtration $\mathcal{F}^t$ such that

$X(t) \exp -rt + \sum_{s=0}^{t-h} hC(s) \exp -rs$ is a $\hat{P}, \mathcal{F}^t$ martingale.

Do we have the replication property

$$\delta(X(t) \exp -rt) + hC(t) \exp -rt = \pi(t)\delta(Y(t) \exp -rt)$$

for a convenient portfolio allocation $\pi(t)$, which is adapted to $\mathcal{F}^t$? We cannot guarantee this, but we can find a portfolio that minimizes

$$\hat{E}[X(T) \exp -rT + h \sum_{t=0}^{T-h} C(t) \exp -rt - \sum_{t=0}^{T-h} \pi(t)\delta(Y(t) \exp -rt)]^2.$$

The corresponding $\varpi(t)$ is solution of the system (6.8). Thanks to the assumption of complete markets, such system has one and only one solution.

6.5. Optimization of portfolio and consumption

For a wealth process evolving as described in Eq. (6.5), we want to maximize

$$\max_{\varpi(\cdot), C(\cdot)} E\left[\sum_{t=0}^{T-h} hU_1(C(t)) \exp -rt + U_2(X(T)) \exp -rT\right].$$

From the martingale property of the process

$$M(t) + \sum_{s=0}^{t-h} hC(s)Z(s) \exp -rs,$$

we can write the constraint

$$E[X(T)Z(T) \exp -rT + \sum_{s=0}^{T-h} hC(s)Z(s) \exp -rs] = X(0).$$

We use the fact that only $C(\cdot)$ and $X(T)$ appear explicitly in the expression of the objective function and of the constraint. So the idea is to find the optimal $\hat{C}(\cdot)$ and $\hat{X}(T)$. They are obtained the following formulas:

$$\begin{aligned}\hat{C}(t) &= I_1(\lambda Z(t)), \\ \hat{X}(T) &= I_2(\lambda Z(T)),\end{aligned}\tag{6.9}$$

where λ is a deterministic parameter, the Lagrange multiplier of the constraint. This parameter is obtained by solving the algebraic equation

$$E[Z(T)I_2(\lambda Z(T)) \exp -rT + \sum_{t=0}^{T-h} hZ(t)I_1(\lambda Z(t)) \exp -rt] = X(0).\tag{6.10}$$

We recall that I_1, I_2 are the inverse of U'_1, U'_2 . Eq. (6.10) has a unique solution since the left-hand side is a monotone decreasing function of λ from $+\infty$ to 0. We then define the optimal wealth $\hat{X}(t)$ at each time by using the martingale property

$$\hat{X}(t) \exp -rt = \hat{E}[\hat{X}(T) \exp -rT + \sum_{s=t}^{T-h} h\hat{C}(s) \exp -rs] | \mathcal{F}^t.\tag{6.11}$$

We can finally obtain the approximate optimal replication portfolio by solving the linear system

$$\begin{aligned} \hat{E} \left[\frac{\delta(Y_i(t) \exp -rt)}{Y_i(t) \exp -rt} \frac{\delta(\hat{X}(t) \exp -rt) + h\hat{C}(t) \exp -rt}{\hat{X}(t) \exp -rt} \middle| \mathcal{F}^t \right] = \\ = \sum_j \hat{\omega}_j(t) a_{ij}^h(t). \end{aligned} \quad (6.12)$$

We cannot find an optimal replication portfolio due to lack of completeness.

7. Bellman equation

7.1. Notation and framework

We assume that $\theta(t)$ and $\sigma(t)$ are deterministic functions. We recall that

$$\delta\mu_i(t) = -\frac{1}{2}ha_{ii}(t) + \sum_{j=1}^n \sigma_{ij}(t)(\delta w_j(t) + \theta_j(t)h),$$

and the wealth evolution equation

$$\begin{aligned} \delta(X_{xt}(s) \exp -rs) = \\ X_{xt}(s) \exp -rs \sum_i \varpi_i(s)(\exp \delta\mu_i(s) - 1) - C(s)h \exp -rs \end{aligned} \quad (7.1)$$

with $X_{xt}(t) = x$.

We want to maximize the objective function

$$\begin{aligned} J_{xt}(C(.), \varpi(.)) \\ = E \left[\sum_{s=t}^{T-h} hU_1(C(s)) \exp -r(s-t) + U_2(X_{xt}(T)) \exp -r(T-t) \right]. \end{aligned}$$

7.2. Functional iteration

We can derive Bellman equation

$$\begin{aligned} W(x, t) = \max_{C, \varpi} \{ & hU_1(C) + \\ & \exp -rh E[W(x \exp rh(1 + \sum_i \varpi_i(\exp \delta\mu_i(t) - 1)) \\ & - Ch \exp rh, t + h)] \} \\ W(x, T) = & U_2(x). \end{aligned} \quad (7.2)$$

This equation does not lead easily to an optimal feedback. We can, however, proceed with a reasoning similar to that of Sections 5.2–5.4. In fact, the reasoning is much facilitated by

considering h small and by making a perturbation argument. We recover the continuous time framework, which we develop in the next section.

8. Continuous time framework

8.1. The model

In continuous time, the difference relations become stochastic differentials. We have successively

$$dZ(s) = -Z(s)\theta(s).dw(s),$$

$$dY_i(s) - rY_i(s)ds = Y_i(s) \sum_j \sigma_{ij}(s)(dw_j(s) + \theta_j(s)ds),$$

$$dX(s) - rX(s)ds = X(s)\varpi(s).\sigma(s)(dw(s) + \theta(s)ds) - C(s)ds.$$

By setting

$$M(s) = X(s)Z(s) \exp -rs,$$

we can obtain the stochastic differential

$$dM(s) = M(s)(\sigma^*(s)\varpi(s) - \theta(s)).dw(s) - C(s)Z(s) \exp -rs.$$

The wealth process corresponding to an initial wealth x at time t is denoted by $X_{xt}(s)$. We define the objective function

$$J_{xt}(C(.), \varpi(.)) = E \left[\int_t^T U_1(C(s)) \exp -r(s-t) + U_2(X_{xt}(T)) \exp -r(T-t) \right].$$

From the expression of $M(s)$, we derive the constraint

$$E \left[X_{xt}(T)Z(T) \exp -r(T-t) + \int_t^T C(s)Z(s) \exp -r(s-t) | \mathcal{F}^t \right] = xZ(t).$$

8.2. Lagrange multiplier

Introducing a Lagrange multiplier λ_{xt} , which is a random variable $\mathcal{F}^t$ measurable depending on x, t , and noting by

$$\hat{C}_{xt}(s), \hat{X}_{xt}(T)$$

the corresponding optimal consumption and final wealth, we deduce as usual

$$\begin{aligned} \hat{C}_{xt}(s) &= I_1(\lambda_{xt}Z(s)), \\ \hat{X}_{xt}(T) &= I_2(\lambda_{xt}Z(T)). \end{aligned}$$

8.3. Connection with dynamic programming

The Bellman equation of dynamic programming writes

$$\begin{aligned} \frac{\partial W}{\partial t} - rW + rx \frac{\partial W}{\partial x} = \\ \max_{C, \varpi} \left[-C \frac{\partial W}{\partial x} + U_1(C) + \frac{\partial W}{\partial x} x \varpi \cdot \sigma \theta + \frac{1}{2} \frac{\partial^2 W}{\partial x^2} x^2 \varpi^* a \varpi \right]. \end{aligned} \quad (8.1)$$

$$W(x, T) = U_2(x).$$

We deduce from Bellman equation feedback rules giving the optimal consumption and portfolio $\hat{C}(x, t)$ and $\hat{\varpi}(x, t)$. We have the relations

$$\begin{aligned} U'_1(\hat{C}(x, t)) &= \frac{\partial W}{\partial x}, \\ x \frac{\partial W}{\partial x} \sigma(t) \theta(t) + x^2 \frac{\partial^2 W}{\partial x^2} a(t) \hat{\varpi}(x, t) &= 0. \end{aligned}$$

Therefore, we get

$$\begin{aligned} \hat{C}(x, t) &= I_1\left(\frac{\partial W}{\partial x}\right), \\ \hat{\varpi}(x, t) &= -\frac{\frac{\partial W}{\partial x}}{x \frac{\partial^2 W}{\partial x^2}} (\sigma^*(t))^{-1} \theta(t). \end{aligned}$$

Noting that $\hat{C}(x, t) = \hat{C}_{xt}(t)$, we deduce the formula

$$\frac{\partial W}{\partial x} = \lambda_{xt} Z(t)$$

Therefore, the optimal consumption process is obtained by the formula

$$\hat{C}_{xt}(s) = I_1\left(\frac{\partial W}{\partial x}(x, t) \frac{Z(s)}{Z(t)}\right), \quad (8.2)$$

and the optimal final wealth is given by

$$\hat{X}_{xt}(T) = I_2\left(\frac{\partial W}{\partial x}(x, t) \frac{Z(T)}{Z(t)}\right). \quad (8.3)$$

We next define the wealth at any time s by conditioning

$$\begin{aligned} \hat{X}_{xt}(s) Z(s) &= E[\hat{X}_{xt}(T) Z(T) \exp -r(T-s) \\ &\quad + \int_s^T \hat{C}_{xt}(\tau) Z(\tau) \exp -r(\tau-s) | \mathcal{F}^s]. \end{aligned} \quad (8.4)$$

It follows that

$$\hat{X}_{xt}(s)Z(s)\exp -rs + \int_t^s \hat{C}_{xt}(\tau)Z(\tau)\exp -r\tau d\tau$$

is a $P, \mathcal{F}^t$ martingale.

8.4. Representation formula

Since martingales can be represented by stochastic integrals, we can write

$$\begin{aligned} \hat{X}_{xt}(s)Z(s)\exp -rs + \int_t^s \hat{C}_{xt}(\tau)Z(\tau)\exp -r\tau d\tau = \\ = xZ(t)\exp -rt + \int_t^s \hat{X}_{xt}(\tau)Z(\tau)\exp -r\tau \hat{\zeta}_{xt}(\tau).dw(\tau), \end{aligned}$$

where $\hat{\zeta}_{xt}(s)$ is a process adapted to the filtration $\mathcal{F}^s$.

Comparing with the expression of $M(s)$, we deduce the relation

$$\hat{\zeta}_{xt}(\tau) = \sigma^*(\tau)\hat{\omega}_{xt}(\tau) - \theta(\tau),$$

where $\hat{\omega}_{xt}(\tau)$ is the optimal portfolio. Recalling that

$$\hat{\omega}_{xt}(t) = \varpi(x, t),$$

the optimal feedback obtained from Bellman equation, we obtain

$$\begin{aligned} \hat{\omega}_{xt}(t) &= (\sigma^*)^{-1}(t)(\theta(t) + \hat{\zeta}_{xt}(t)) \\ &= -\frac{\frac{\partial W}{\partial x}}{x \frac{\partial^2 W}{\partial x^2}}(\sigma^*(t))^{-1}\theta(t). \end{aligned}$$

Finally, we obtain

$$\hat{\zeta}_{xt}(t) = -(1 + \frac{\frac{\partial W}{\partial x}}{x \frac{\partial^2 W}{\partial x^2}})\theta(t),$$

and more generally, we can assert that

$$\hat{\zeta}_{xt}(s) = -\left(1 + \frac{\frac{\partial W}{\partial x}}{x \frac{\partial^2 W}{\partial x^2}}(\hat{X}_{xt}(s), s)\right)\theta(s), \quad (8.5)$$

which connects the representation formula of martingales with the derivatives of the value function.

Note that the portfolio is proportional to $\theta(s)$, which expresses the one-fund theorem. The ratio of investment between the risky portfolio and the risk-less asset is defined by the proportionality factor, which depends on the present wealth.

EXERCISE 8.1. Consider the case

$$U_1(C) = \log C, \quad U_2(x) = \log x$$

show that

$$W(x, t) = \rho(t) \log x + \mu(t),$$

and the optimal feedbacks are given by

$$\hat{C}(x, t) = \frac{x}{\rho(t)}, \quad \hat{\varpi}(x, t) = (\sigma^*)^{-1} \theta(t).$$

References

- ARROW, K., DEBREU, G. (1954). Existence of equilibrium for a competitive economy. *Econometrica* **22**, 265–290.
- KARATZAS, I., SHREVE, S.E. (1998). *Methods of Mathematical Finance* (Springer Verlag, New York, NY).
- MERTON, R. (1973). An intertemporal capital asset pricing model. *Econometrica* **41**, 867–888.
- SHREVE, S.E. (2004). *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, (*Springer Finance*) (Springer Verlag, New York, NY).

Numerical Approximation by Quantization of Control Problems in Finance Under Partial Observations

Huyên Pham

*Laboratoire de Probabilités et, Modèles Aléatoires CNRS, UMR 7599 Université Paris 7,
and Institut Universitaire de France, Paris, France
E-mail address: pham@math.jussieu.fr*

Marco Corsi

*Dipartimento di Matematica Pura ed Applicata, Università degli studi di Padova,
Laboratoire de Probabilités et, Modèles Aléatoires CNRS, UMR 7599 Université Paris 7,
Paris, France
E-mail address: marco.corsi@hotmail.fr*

Wolfgang J. Runggaldier

*Dipartimento di Matematica Pura ed Applicata, Università di Padova, Padova, Italy
E-mail address: runggal@math.unipd.it*

Abstract

We study numerical solutions to discrete time control problems under partial observation when the state of the system is described by (X, Y, V^α) with X signal process, Y observation process, and V^α controlled process. The control process α is required to be adapted with respect to the observation filtration. The structure of the control problem is motivated with a view toward financial applications. In particular, we consider the problem of hedging a future liability in the context of incomplete information. To cope with difficulties arising from partial information, stochastic filtering is used, and the filter process is discretized in order to obtain a feasible numerical solution. This is done by performing a quantization of the pair process filter observation. Dynamic programming is finally applied to solve the approximated filtered control problem. Convergence results are given, and numerical applications

are presented and discussed for the problem of hedging an European put (and call) option with unobservable volatility.

1. Introduction

This chapter concerns numerically feasible approximations to discrete time stochastic control problems under partial observation. Such problems arise naturally in financial market models where some model coefficients (volatility, drift, etc.) may depend on stochastic factors that are not observable. They were investigated in numerous papers, mostly from a theoretical viewpoint. However, numerical tests are rarely performed due to computational difficulties, especially when observations are multiplicative noises and non-Gaussian, like in unobservable stochastic volatility models.

Here, we consider a discrete time model where the signal process X is a Markov chain, which may not be observable and takes value in a set E consisting of a finite number of points $\{x^1, \dots, x^m\}$. The observation process Y takes values in $\mathbb{R}^d$ and is such that the pair (X, Y) is a Markov chain. The control process, denoted by α , is adapted with respect to the observation filtration, and V^α is the controlled process.

The structure of our model is motivated with a view toward financial applications. Consider the case where Y is the price of a risky asset, X is its unobservable volatility or drift, and V^α is the wealth process. The investment strategy is represented by a control process α , which gives the number of risky asset shares held in the portfolio. Denoting by $\mathbb{F}^Y = (\mathcal{F}_k^Y)_k$ the filtration generated by the observation process Y , the filter process Π is given by

$$\Pi_k^i := \mathbb{P}[X_k = x^i | \mathcal{F}_k^Y], \quad k \in \mathbb{N}, \quad i = 1, \dots, m.$$

By using the filter process, the original control problem under partial observation is transformed into an equivalent one under complete observation with observed state process given by the filter Π instead of the unobservable signal X , and we may apply dynamic programming method (see BENSOUSSAN [1992]).

The numerical difficulty of this procedure concerns the filtered problem dimension because the number of values taken by the filter is infinite even though the process X has only a finite number of states. More precisely, as the state space E consists of a finite number m of points $\{x_1, \dots, x_m\}$, the filter is characterized by an m -vector with components $\Pi_k^i := \mathbb{P}[X_k = x^i | \mathcal{F}_k^Y]$, and it takes values in the m simplex K_m of $\mathbb{R}^m$. Therefore, in order to numerically solve the problem, the filter has to be approximated with another process taking only a finite number of values in K_m . A classical approach (see BENSOUSSAN and RUNGGALDIER [1987]) is to discretize the observation process Y by a process $\hat{Y}$ taking a finite number N of values and then approximate for each k the filter Π_k by the filter of X_k given $\hat{Y}_1, \dots, \hat{Y}_k$. The numerical drawback of this approach is that the number of possible values taken by the approximating filter grows exponentially with the time step; in fact, at time n , the approximated filter is identified by a random vector taking N^n possible values.

In this chapter, we suggest an alternative approach, which has been recently developed to numerically solve optimal stopping time problems under partial observations (see

PHAM, RUNGGALDIER and SELLAMI [2004]). The method consists in approximating the Markov pair process (Π, Y) by a process $(\hat{\Pi}, \hat{Y})$ taking at each time step k a finite number of values, N_k , that is arbitrarily assigned. This approach relates to the field of quantization methods, recently developed in numerical probability and applied to solve various financial problems (see PAGÈS, PHAM and PRINTEMPS [2003], PHAM, RUNGGALDIER and SELLAMI [2004], PAGÈS, PHAM and PRINTEMPS [2004], and PAGÈS and PHAM [2005]). In particular, by using results from PHAM, RUNGGALDIER and SELLAMI [2004], it is possible to make an optimal quantization, which for each time step k minimizes the quantity

$$\mathbb{E}\left[|(Y_k, \Pi_k) - (\hat{Y}_k, \hat{\Pi}_k)|^2\right]$$

called quantization error or distortion. The implementation of this optimal quantization is based on a stochastic gradient descent method combined with Monte Carlo simulations of the pair (Π, Y) . Once the problem has been discretized, we can solve it numerically by using dynamic programming, and we prove that when N_k grows, the approximated solution converges toward the real solution with rate dominated by the quantization error.

Finally, we apply the method described above in order to solve a specific financial problem, which consists in the hedging of a European put (and call) option. Since we are in an incomplete market setting, it is not possible to obtain a self-financing and perfect hedging strategy, and we consider as hedging criterion the expected value of a convex function applied to the residual hedging error. In particular, we will focus on the case of the quadratic criterion (see FÖLLMER and SONDERMANN [1986]) and the shortfall risk criterion (see FÖLLMER and LEUKERT [2000]).

The outline of the chapter is as follows. In Section 2, the partial observation discrete-time control problem is formulated. In Section 3, stochastic filtering is used to transform the original control problem into a complete observation problem that can be studied using the dynamic programming method. In Section 4, the numerical approximation by quantization to this control problem is described, and some convergence results are proved. The financial application is presented in Section 5, where we study the problem of hedging a European put (and call) option with unobservable volatility. Some numerical tests are finally performed and discussed.

1.1. Notations

In the sequel, we denote by $|\cdot|_1$ the l^1 norm on $\mathbb{R}^l$, by $|\cdot|$ the Euclidean norm on $\mathbb{R}^l$ and, for any random variable X taking values in $\mathbb{R}^l$, we denote

$$\|X\|_2 := (\mathbb{E}|X|^2)^{\frac{1}{2}} \quad \text{and} \quad \|X\|_1 := \mathbb{E}|X|_1.$$

For any measurable function g from $D \subset \mathbb{R}^l$ into $\mathbb{R}$, we define

$$[g]_{\text{sup}} := \sup_{x \in D} |g(x)| \tag{1.1}$$

and

$$[g]_{\text{Lip}} := \sup_{x, y \in D; x \neq y} \frac{|g(x) - g(y)|}{|x - y|_1}. \tag{1.2}$$

2. Problem setup

Let us consider a discrete time dynamical system over a horizon $\{0, \dots, n\}$ with n fixed and with state at time k ($k = 0, \dots, n$) described by the variables (X_k, Y_k, V_k^α) . In particular, $(X_k)_k$ represents the signal process that may not be observable, $(Y_k)_k$ is the observation process, and $(V_k^\alpha)_k$ is the process controlled by a process α adapted with respect to $(\mathcal{F}_k^Y)$, the filtration generated by $(Y_k)_k$.

In a financial setting, we think of the case where Y is the price of a risky asset, X is its unobservable volatility or drift, and V^α is the wealth process. The investment strategy is represented by a control process α representing the number of risky asset shares held in the portfolio and based on the information derived from the price observations.

We assume that the process $(X_k)_k$ is a finite-state Markov chain taking values in the space $E = \{x^1, \dots, x^m\}$. Its probability transition P_k (from the period $k - 1$ to the period k) and initial law μ are defined by

$$\mu^i = \mathbb{P}[X_0 = x^i], \quad i = 1, \dots, m,$$

$$P_k^{ij} = \mathbb{P}[X_k = x^j | X_{k-1} = x^i], \quad i = 1, \dots, m, \quad j = 1, \dots, m.$$

The process $(Y_k)_k$ takes values in $\mathbb{R}^d$ and is such that the pair $(X_k, Y_k)_k$ is a Markov chain, and the conditional law of Y_k given (X_{k-1}, Y_{k-1}, X_k) admits a (known) bounded density

$$y' \rightarrow g_k(X_{k-1}, Y_{k-1}, X_k, y').$$

For simplicity, we assume that Y_0 is a known deterministic constant, fixed equal to y_0 . The control process is denoted by $(\alpha_k)_{k \geq 0}$, takes values in $A \subset \mathbb{R}^l$, and is supposed to be adapted with respect to the filtration $(\mathcal{F}_k^Y)_k$ generated by (Y_k) . We denote by $\mathcal{A}$ the set of control processes. The controlled process $(V_k^\alpha)_k$ takes values in $\mathbb{R}$ and is governed by dynamics of the form:

$$V_{k+1}^\alpha = H(V_k^\alpha, \alpha_k, Y_k, Y_{k+1}), \quad (2.1)$$

where H is a measurable function.

We are given a running (measurable) cost function f on $E \times \mathbb{R}^d \times \mathbb{R} \times A$ and a terminal (measurable) cost function h on $E \times \mathbb{R}^d \times \mathbb{R}$. Given an initial value v_0 for the controlled process, an admissible control $\alpha \in \mathcal{A}$, the expected cost function is defined by

$$J(v_0, \alpha) = \mathbb{E} \left[\sum_{k=0}^{n-1} f(X_k, Y_k, V_k^\alpha, \alpha_k) + h(X_n, Y_n, V_n^\alpha) \right], \quad (2.2)$$

and the goal is to choose a control process in order to minimize the cost J up to the time horizon n :

$$J_{\text{opt}}(v_0) = \inf_{\alpha \in \mathcal{A}} J(\alpha). \quad (2.3)$$

2.1. Financial example

A typical financial example corresponds to the case where Y represents the price of a risky asset and X is its unobservable volatility. Assume that a riskless n -maturity bond is available for trading, yielding constant interest rate $r = 0$ (for simplicity). We consider an economic agent over an investment time horizon n . At time $k = 0$, the agent starts with an initial wealth v and then at each instant $k = 1, \dots, n$, he rebalances his portfolio holdings by choosing the investment allocations in the bond and in the risky asset. Under the assumption of self-financing, the wealth process V satisfies

$$V_{k+1}^\alpha = V_k^\alpha + \alpha_k [Y_{k+1} - Y_k], \quad (2.4)$$

where α_k represents the number of shares of risky asset held in the portfolio at time k . The process $(\alpha_k)_{k=1, \dots, n}$ is supposed to be adapted with respect to the filtration generated by the price process Y , that is, the investment strategy is selected only on the basis of past observations of the security prices.

Given a loss function $\ell : \mathbb{R} \rightarrow \mathbb{R}$, the hedging criterion for a derivative asset $h(Y_n)$ of maturity n consists in minimizing the expected loss

$$\mathbb{E}[\ell(h(Y_n) - V_n^\alpha)]$$

over all admissible portfolios $\alpha = (\alpha_k)_{k=0, \dots, n}$.

In order to prove convergence results, we shall make some technical assumptions.

H1 The set A is compact.

H2 H is continuous, and there exists some positive constant $[H]_{\text{Lip}}$ s.t. for all (v, a, y, y') and $(\hat{v}, \hat{a}, \hat{y}, \hat{y}') \in \mathbb{R} \times A \times \mathbb{R}^d \times \mathbb{R}^d$:

$$|H(v, a, y, y') - H(\hat{v}, \hat{a}, \hat{y}, \hat{y}')|_1 \leq [H]_{\text{Lip}} (|v - \hat{v}| + |y - \hat{y}|_1 + |y' - \hat{y}'|_1).$$

H3 Functions f and h are bounded and Lipschitz.

H4 There exists some positive constant L_g such that for all $k = 1, \dots, n$

$$\begin{aligned} \sum_{i,j=1}^m P_k^{ij} \int |g_k(x^i, y, x^j, y') - g_k(x^i, \hat{y}, x^j, y')| dy' \\ \leq L_g |y - \hat{y}|_1 \quad \forall y, \hat{y} \in \mathbb{R}^d. \end{aligned}$$

REMARK 2.1. The hypothesis **H2** is verified by (2.4) in the previous example. Concerning the hypothesis **H4**, we will see that it is satisfied for the model analyzed in the numerical application given in the last section.

3. Filtering and dynamic programming

Recalling that the state space of (X_k) consists of a finite number of points and denoting by $(\mathcal{F}_k^Y)$ the filtration generated by the observation process (Y_k) , the filter is

defined as

$$\Pi_k^i = \mathbb{P}[X_k = x^i | \mathcal{F}_k^Y] \quad i = 1, \dots, m \quad \text{and} \quad k = 1, \dots, n$$

and is a random vector process, which takes values in the m -simplex K_m in $\mathbb{R}^m$:

$$K_m = \left\{ \pi = (\pi^i) \in \mathbb{R}^m : \pi^i \geq 0 \quad \text{and} \quad |\pi|_1 = \sum_{i=1}^m \pi^i = 1 \right\}.$$

By using Bayes' formula, the filter process can be calculated in a recursive way as follows (see LIPSTER and SHIRYAEV [1977]):

$$\begin{aligned} \Pi_0 &= \mu \\ \Pi_k &= \bar{G}_k(\Pi_{k-1}, Y_{k-1}, Y_k) = \frac{G P_k(Y_{k-1}, Y_k)^T \Pi_{k-1}}{|G P_k(Y_{k-1}, Y_k)^T \Pi_{k-1}|_1}, \quad k \geq 1, \end{aligned} \quad (3.1)$$

where $G P_k(Y_{k-1}, Y_k)$ is an $m \times m$ random matrix given by

$$G P_k(Y_{k-1}, Y_k)_{ij} = g_k(x_{k-1}^i, Y_{k-1}, x_k^j, Y_k) P_k^{ij}, \quad 1 \leq i, j \leq m,$$

and ^{T} is the transpose. One can also show (see PHAM, RUNGALDIER and SELLAMI [2004]) that the pair $(\Pi_k, Y_k)_k$ is a Markov chain with respect to the filtration $(\mathcal{F}_k^Y)_k$, and the conditional law Q_k of Y_k given (Π_{k-1}, Y_{k-1}) admits a density given by

$$y' \rightarrow q_k(\Pi_{k-1}, Y_{k-1}, y') := \sum_{i,j=1}^m g_k(x^i, Y_{k-1}, x^j, y') P_k^{ij} \Pi_{k-1}^i, \quad (3.2)$$

Relations (3.1) and (3.2) show that, although the probability transition of the Markov chain (Π_k, Y_k) is not explicitly known, it can be simulated. This point is important when one needs Monte Carlo simulations of (Π_k, Y_k) , (see Subsection 4.1.1).

By using the law of iterated conditional expectations, we can rewrite the expected cost function (2.2) as follows:

$$\begin{aligned} J(v_0, \alpha) &= \mathbb{E} \left[\sum_{k=0}^{n-1} \mathbb{E} [f(X_k, Y_k, V_k^\alpha, \alpha_k) | \mathcal{F}_k^Y] + \mathbb{E} [h(X_n, Y_n, V_n^\alpha) | \mathcal{F}_n^Y] \right] \\ &= \mathbb{E} \left[\sum_{k=0}^{n-1} \sum_{i=1}^m f(x^i, Y_k, V_k^\alpha, \alpha_k) \Pi_k^i + \sum_{i=1}^m h(x^i, Y_n, V_n^\alpha) \Pi_n^i \right] \\ &= \mathbb{E} \left[\sum_{k=0}^{n-1} \hat{f}(\Pi_k, Y_k, V_k^\alpha, \alpha_k) + \hat{h}(\Pi_n, Y_n, V_n^\alpha) \right], \end{aligned}$$

where

$$\begin{aligned}\hat{f}(\pi, y, v, a) &:= \int f(x, y, v, a) \pi(dx) = \sum_{i=1}^m f(x^i, y, v, a) \pi^i \\ \hat{h}(\pi, y, v) &:= \int h(x, y, v) \pi(dx) = \sum_{i=1}^m h(x^i, y, v) \pi^i\end{aligned}$$

The original problem (2.3) can now be formulated as a problem under full observation with state variables (Π_k, Y_k, V_k) :

$$J_{\text{opt}}(v_0) = \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\sum_{k=0}^{n-1} \hat{f}(\Pi_k, Y_k, V_k^\alpha, \alpha_k) + \hat{h}(\Pi_n, Y_n, V_n^\alpha) \right]. \quad (3.3)$$

Actually, recalling (2.1) and following the dynamic programming algorithm (see BERTSEKAS [1992]) for solving the filtered problem (3.3), we define the sequence of functions:

$$(DP) \begin{cases} u_n(\pi, y, v) = \hat{h}(\pi, y, v) \\ u_k(\pi, y, v) = \inf_{a \in A} \left\{ \hat{f}(\pi, y, v, a) \right. \\ \quad \left. + \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \mid (\Pi_k, Y_k) = (\pi, y) \right] \right\} \\ k = 0, \dots, n-1. \end{cases}$$

The following result shows that this backward procedure gives the solution for $k = 0$ to the original problem (2.3).

PROPOSITION 3.1. *Assume **H1**, **H2**, and **H3**. Then, the algorithm (DP) provides the solution to problem (2.3), that is,*

$$u_0(\mu, y_0, v_0) = J_{\text{opt}}(v_0).$$

PROOF. See Appendix A. □

4. Approximation by quantization and error analysis

4.1. The numerical approximation method

From a numerical viewpoint, the formula given by the (DP) algorithm is still untractable since the state variable $Z_k^\alpha := (\Pi_k, Y_k, V_k^\alpha)$ takes values in a continuous state space. In order to obtain a numerical solution, the basic idea is to approximate at each time step k the continuous state variable Z_k^α by a discrete state variable $\hat{Z}_k^\alpha$ taking a finite number of values. The main concern is how to discretize in an efficient and feasible way the variables Z_k^α that depend on the control α ?

We deal separately with the approximation of the pair filter observation $W := (\Pi, Y)$ that does not depend on the control and the approximation of the controlled state variable V^α . The approximation of (Π, Y) is obtained following an optimal quantization method as in PHAM, RUNGALDIER and SELLAMI [2004]. The approximation of V^α is obtained by a classical uniform space discretization similar to the Markov chain method as in KUSHNER and DUPUIS [2001].

4.1.1. Optimal quantization of the pair filter observation

In a first step, we discretize for each k the pair (Π_k, Y_k) by approximating it by $(\hat{\Pi}_k, \hat{Y}_k)$ taking a finite number of values. The space discretization (or quantization) of the random vector $W_k = (\Pi_k, Y_k)$ valued in $K_m \times \mathbb{R}^d$ is constructed as follows. At initial time $k = 0$, recall that W_0 is a known deterministic vector equal to $w_0 = (\mu, y_0)$, so we start from the grid with one point in $K_m \times \mathbb{R}^d$:

$$\Gamma_0 = \{w_0 = (\mu, y_0)\}.$$

At time $k \geq 1$, we are given a grid Γ_k of N_k points in $K_m \times \mathbb{R}^d$,

$$\Gamma_k = \left\{ w_k^1 = (\pi_k(1), y_k^1), \dots, w_k^{N_k} = (\pi_k(N_k), y_k^{N_k}) \right\},$$

and we approximate the pair $W_k = (\Pi_k, Y_k)$ by $\hat{W}_k = (\hat{\Pi}_k, \hat{Y}_k)$ valued in Γ_k and defined as the closest neighbour projection,

$$\hat{W}_k = \text{Proj}_{\Gamma_k}(W_k) := \sum_{i=1}^{N_k} w_k^i 1_{C_i(\Gamma_k)}(W_k),$$

where the so-called Voronoi tessellations $C_1(\Gamma_k), \dots, C_{N_k}(\Gamma_k)$ are Borel partitions of $K_m \times \mathbb{R}^d$ satisfying

$$C_i(\Gamma_k) \subset \left\{ w \in K_m \times \mathbb{R}^d : |w - w_k^i| = \min_{j=1, \dots, N_k} |w - w_k^j| \right\}, \quad i = 1, \dots, N_k.$$

The L^2 error induced by this projection, called L^2 quantization error, is equal at time k to $\|W_k - \hat{W}_k\|_2$. As a function of the grid Γ_k identified with the N_k -tuple $(w_k^1, \dots, w_k^{N_k})$ in $K_m \times \mathbb{R}^d$, the square of the L^2 quantization error, called distorsion, is written as

$$D_{N_k}^{W_k}(\Gamma_k) = \|W_k - \text{Proj}_{\Gamma_k}(W_k)\|_2^2 = \mathbb{E} \left[\min_{i=1, \dots, N_k} |W_k - w_k^i|^2 \right]. \quad (4.1)$$

Notice by definition of the closest neighbour projection that the L^2 quantization error is the minimum of the L^2 error $\|W_k - U\|_2$ among all random variables U taking values in the grid Γ_k .

In a second step, we approximate the probability transitions of the Markov chain (W_k) by the following probability transition matrix:

$$\hat{r}_k^{ij} = \mathbb{P} \left[\hat{W}_k = w_k^j \mid \hat{W}_{k-1} = w_{k-1}^i \right] = \frac{\mathbb{P} \left[W_k \in C_j(\Gamma_k), W_{k-1} \in C_i(\Gamma_{k-1}) \right]}{\mathbb{P} \left[W_{k-1} \in C_i(\Gamma_{k-1}) \right]}$$

for all $k = 1, \dots, n, i = 1, \dots, N_{k-1}, j = 1, \dots, N_k$.

The grids Γ_k are optimally chosen so as to minimize at each time k the distortion $D_{N_k}^{W_k}(\Gamma_k)$. This relies on the property that the distortion is differentiable, with a gradient obtained by formal differentiation in (4.1):

$$\nabla D_{N_k}^{W_k}(\Gamma_k) = 2 \left(\mathbb{E} \left[(w_k^i - W_k) 1_{W_k \in C_i(\Gamma_k)} \right] \right)_{1 \leq i \leq N_k}. \quad (4.2)$$

The optimal grids and the associated probability transition matrix are then processed and estimated by a stochastic gradient descent method, known in this context as the Kohonen algorithm and based on the integral representation (with respect to the probability law of W_k) (4.2). This is achieved by Monte Carlo simulations of the Markov chain $(W_k)_k = (\Pi_k, Y_k)_k$ through the following simulation procedure: starting from (Π_{k-1}, Y_{k-1}) ,

- simulate Y_k according to the density given in (3.2)
- compute Π_k by the formula (3.1).

We refer to PHAM, RUNGGALDIER and SELLAMI [2004] for the details and the practical implementation of the optimal grids.

4.1.2. Space discretization of the controlled variable

We fix a bounded uniform grid on the state space $\mathbb{R}$ for the controlled process V^α . Namely, we set

$$\Gamma^V := (2\nu)\mathbb{Z} \cap [-R, R],$$

where ν is the spatial step and R is the grid size. We denote by Proj_{Γ^V} the projection on the grid Γ^V according to the closest neighbor rule. Recalling the dynamics (2.1) of the controlled process, we approximate it as follows: given a control $\alpha \in \mathcal{A}$, we discretize $(\hat{V}_k^\alpha)_k$ by the controlled process valued in Γ^V and evolving according to the dynamics

$$\hat{V}_{k+1}^\alpha = \text{Proj}_{\Gamma^V}(H(\hat{V}_k^\alpha, \alpha_k, \hat{Y}_k, \hat{Y}_{k+1})). \quad (4.3)$$

Here, $\hat{Y}_k$ is the quantization of Y_k obtained in the previous subsection.

4.1.3. Approximation of the control problem

We approximate the sequence of functions (u_k) by the sequence of function $\hat{u}_k$ defined on $\Gamma_k \times \Gamma^V$, $k = 0, \dots, n$, by a dynamic programming type formula:

$$\begin{aligned}\hat{u}_n(\pi, y, v) &= \hat{h}(\pi, y, v) \\ \hat{u}_k(\pi, y, v) &= \inf_{a \in A} \left\{ \hat{f}(\pi, y, v, a) \right. \\ &\quad \left. + \mathbb{E}[\hat{u}_{k+1}(\hat{\Pi}_{k+1}, \hat{Y}_{k+1}, \text{Proj}_{\Gamma^V}(H(v, a, y, \hat{Y}_{k+1}))) | (\hat{\Pi}_k, \hat{Y}_k) = (\pi, y)] \right\}.\end{aligned}$$

From an algorithmic viewpoint, this is computed explicitly as follows:

$$\begin{aligned}\hat{u}_n(w_n^i, v) &= \hat{h}(w_n^i, v), \quad w_n^i = (\pi_n(i), y_n^i) \in \Gamma_n, i = 1, \dots, N_n, v \in \Gamma^V \\ \hat{u}_k(w_k^i, v) &= \inf_{a \in A} \left\{ \hat{f}(w_k^i, v, a) \right. \\ &\quad \left. + \sum_{j=1}^{N_{k+1}} \hat{r}_{k+1}^{ij} \hat{u}_{k+1}(w_{k+1}^j, \text{Proj}_{\Gamma^V}(H(v, a, y_k^i, y_{k+1}^j))) \right\} \\ w_k^i &= (\pi_k(i), y_k^i) \in \Gamma_k, i = 1, \dots, N_k, v \in \Gamma^V.\end{aligned}\tag{4.4}$$

For $v_0 \in \Gamma^V$, the solution $J_{\text{opt}}(v_0) = u_0(\mu, y_0, v_0)$ to our control problem is then approximated by

$$\hat{J}_{\text{quant}}(v_0) = \hat{u}_0(\mu, y_0, v_0).$$

Moreover, this backward dynamic programming scheme allows us to compute at each step $k = 0, \dots, n - 1$, an approximate optimal control $\hat{a}_k(w, v)$, $w = (\pi, y) \in \Gamma_k$, $v \in \Gamma^V$, by taking the infimum in (4.4).

4.2. Error analysis and rate of convergence

We state an error estimation between the optimal cost function J_{opt} and the approximated cost function $\hat{J}_{\text{quant}}$, in terms of

- the quantization errors $\Delta_k = \|W_k - \hat{W}_k\|_2$ for the pair $W_k = (\Pi_k, Y_k)$, $k = 0, \dots, n$
- the spatial step ν and the grid size R for V_k^α , $k = 0, \dots, n$.

THEOREM 4.1. *Under **H1**, **H2**, **H3**, and **H4**, we have for all $v_0 \in \Gamma^V$*

$$\left| J_{\text{opt}}(v_0) - \hat{J}_{\text{quant}}(v_0) \right| \leq C_1(n) \sum_{k=0}^n \sum_{j=0}^k \Psi^k \left(\nu + \frac{C_2}{R} + \Delta_{k-j} \right),\tag{4.5}$$

where $C_1(n) = \sqrt{m+d+1} [2(n\bar{L}_g\bar{f} + \bar{M} + 3\bar{L}_g\bar{h}) \frac{(2\bar{L}_g)^n}{2\bar{L}_g-1} + \bar{f} + \bar{h}]$, $\bar{f} = \max([f]_{\text{sup}}, [f]_{\text{Lip}})$, $\bar{h} = \max([h]_{\text{sup}}, [h]_{\text{Lip}})$, $\bar{L}_g = \max(L_g, 1)$, $\bar{M} = \max([H]_{\text{Lip}}, 1)$, C_2 is the maximum value of H over $\Gamma^V \times A \times \cup_k \Gamma_k \times \cup_k \Gamma_k$, and $\Psi = (2d+1)[H]_{\text{Lip}}$.

PROOF. See Appendix B. $\square$

4.2.1. Convergence of the approximation

As a consequence of Zador's theorem (see GRAF and LUSCHGY [2000]), which gives the asymptotic behavior of the optimal quantization error, when the number of grid points goes to infinity, we can derive the following estimation on the optimal quantization error for the pair filter observation (see PHAM, RUNGGALDIER and SELLAMI [2004]):

$$\limsup_{N_k \rightarrow \infty} N_k^{\frac{2}{m-1+d}} \min_{|\Gamma_k| \leq N_k} \|W_k - \hat{W}_k\|_2^2 \leq C_k(m, d),$$

where $C_k(m, d)$ is a constant depending on m, d , and the marginal density of Y_k . Therefore, the estimation (4.5) provides a rate of convergence for the approximation of J_{opt} of order

$$n^2 \Psi^n C_1(n) \left(\nu + \frac{1}{R} + \frac{1}{N^{\frac{1}{m-1+d}}} \right),$$

when $N_k = N$ is the number of points at each grid Γ_k used for the optimal quantization of $W_k = (\Pi_k, Y_k)$, $k = 1, \dots, n$. We then get the convergence of the approximated cost function $\hat{J}_{\text{quant}}$ to the optimal cost function J_{opt} when ν goes to zero and N and R go to infinity. Moreover, by extending the approximate control $\hat{\alpha}_k$, $k = 0, \dots, n-1$, to the continuous state space $K_m \times \mathbb{R}^d \times \mathbb{R}$ by

$$\hat{\alpha}_k(\pi, y, v) = \hat{\alpha}_k(\text{Proj}_{\Gamma_k}(\pi, y), \text{Proj}_{\Gamma^v}(v)), \quad \forall (\pi, y, v) \in K_m \times \mathbb{R}^d \times \mathbb{R},$$

and by setting (by abuse of notation) $\hat{\alpha}_k = \hat{\alpha}_k(\Pi_k, Y_k, \hat{V}_k^{\hat{\alpha}})$, we get an approximate control $\hat{\alpha} = (\hat{\alpha}_k)_k$ in $\mathcal{A}$, which is ε -optimal for the original control problem (see RUNGGALDIER [1991]) in the sense that for all $\varepsilon > 0$

$$J(v_0, \hat{\alpha}) \leq J_{\text{opt}}(v_0) + \varepsilon,$$

whenever N and R are large enough and ν is small enough.

5. Financial application: European option hedging in a partially observed stochastic volatility model

In this section, we apply the methodologies described above in order to study the problem of hedging an European put (or call) option in the context of incomplete information on the underlying price evolution model. Since we are in an incomplete market setting, the perfect replication of the claim is not possible, and as hedging criterion we choose the expected value of a convex loss function applied to the hedging error. In particular, we will consider the case of the quadratic criterion and that of the shortfall risk criterion.

5.1. The model

We consider a stochastic volatility model where for simplicity we have only one risky asset with observable price (S_k) whose dynamics is given by

$$S_{k+1} = S_k \exp \left[\left(r - \frac{1}{2} X_k^2 \right) \delta + X_k \sqrt{\delta} \epsilon_{k+1} \right], \quad k = 1, \dots, n,$$

$$S_0 = s_0 > 0$$

where $(\epsilon_k)_k$ is a Gaussian white noise sequence, X_k is the unobservable volatility process, $\delta = 1/n$ represents the discretization time step over the interval $[0, 1]$, and r is the riskless interest rate per unit of time.

We denote by S^0 the riskless asset price with dynamics

$$S_{k+1}^0 = S_k^0 e^{r\delta}.$$

Notice that the conditional law of S_{k+1} given (X_k, S_k) has a density given by

$$g(X_k, S_k, s') = \frac{1}{s' \sqrt{2\pi\delta X_k^2}} \exp \left[-\frac{\left(\ln s' - \ln S_k - \left(r - \frac{1}{2} X_k^2 \right) \delta \right)^2}{2X_k^2\delta} \right], \quad s' > 0,$$

and notice that, as the first derivative of g with respect to s' is bounded, the hypothesis **H4** is satisfied.

The volatility (X_k) is described by a Markov chain taking three possible values $x^b < x^m < x^h$ in $(0, \infty)$. Its probability transition matrix is given by

$$P_k = \begin{pmatrix} 1 - (p_{bm} + p_{bh})\delta & p_{bm}\delta & p_{bh}\delta \\ p_{mb}\delta & 1 - (p_{mb} + p_{mh})\delta & p_{mh}\delta \\ p_{hb}\delta & p_{hm}\delta & 1 - (p_{hb} + p_{hm})\delta \end{pmatrix}. \quad (5.1)$$

The volatility (X_k) is a Markov-chain approximation à la Kushner (see KUSHNER and DUPUIS [2001]) of a mean-reverting process

$$dX_t = \lambda(x_0 - X_t)dt + \eta dW_t.$$

Denoting by $\Delta > 0$ the spatial step, this corresponds to a probability transition matrix of the form (5.1) with

$$x^b = x_0 - \Delta, \quad x^m = x_0, \quad x^h = x_0 + \Delta,$$

and

$$\begin{aligned} p_{bm} &= \lambda + \frac{\eta^2}{2\Delta^2}, & p_{bh} &= 0 \\ p_{mb} &= \frac{\eta^2}{2\Delta^2}, & p_{mh} &= \frac{\eta^2}{2\Delta^2} \\ p_{hb} &= 0, & p_{hm} &= \lambda + \frac{\eta^2}{2\Delta^2}, \end{aligned}$$

with the condition that $1 - \lambda - \frac{\eta^2}{2\Delta^2} > 0$ and $1 - \frac{\eta^2}{\Delta^2} > 0$.

In order to hedge the European put option with strike K , we invest an initial capital v_0 in the risky asset following a self-financing strategy. Recall that the wealth process is given by

$$V_{k+1}^\alpha = V_k^\alpha e^{r\delta} + \alpha_k [S_{k+1} - S_k e^{r\delta}], \quad (5.2)$$

where α_k represents the number of shares of asset S_k held in the portfolio at time k . Observe that (5.2) verifies the hypothesis **H2**, and recall that the control process (α_k) is adapted with respect to the filtration $(\mathcal{F}_k^S)$ generated by the observation process.

In what follows, we will work with the log price instead of the price and we set $Y_k = \ln S_k$.

5.2. Hedging of an European put option: quadratic criterion

Using a quadratic loss criterion (see FÖLLMER and SONDERMANN [1986]), an optimal strategy is a solution to the optimization problem:

$$\inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\left((K - e^{Y_n})_+ - V_n^\alpha \right)^2 \right], \quad (5.3)$$

where $\mathcal{A}$ is the control space.

Since the process $(X_k)_{k=1, \dots, n}$ is unobservable, the optimization problem described above is a control problem under partial information and can thus be studied by using stochastic filtering and approximation techniques as shown in the previous sections. An approximated solution is in particular obtained from the following steps:

1. **Quantization.** Denoting by Π_k the filter process, we discretize the pair (Π_k, Y_k) by performing an optimal quantization as explained in Subsection 4.1.1. This procedure provides, for all instants k ,
 - a. An N_k -point grid Γ_k , which is a discretization of the state space of (Π_k, Y_k) . This discretization is optimal in the sense specified in PAGÈS, PHAM and PRINTEMPS [2003].
 - b. A matrix $\{\hat{r}_k^{ij}, i = 1, \dots, N_{k-1}, j = 1, \dots, N_k\}$, which approximates the probability transition of the Markov chain (Π_k, Y_k) .

The controlled one-dimensional process (V_k^α) is discretized using a regular N^V -point grid of $\mathbb{R}$ given by

$$\Gamma^V = (2\nu)\mathbb{Z} \cap [V_{\inf}, V_{\sup}],$$

where ν is some discretization space step and $V_{\inf}$ and $V_{\sup}$ are the bounds of the grid size.

- 2. Dynamic programming.** Once the problem has been discretized, we use the dynamic programming algorithm to calculate an approximated solution:

$$\begin{aligned} \hat{u}_n(w_n^i, v) &= (v - (K - e^{y_n^i})_+)^2 \quad \forall w_n^i = (\pi_n(i), y_n^i) \in \Gamma_n, \quad \forall v \in \Gamma^V \\ \hat{u}_k(w_k^i, v) &= \inf_{a \in A} \sum_{j=1}^{N_{k+1}} \hat{r}_{k+1}^{ij} \hat{u}_{k+1}(w_{k+1}^j, \text{Proj}_{\Gamma^V}(ve^{r\delta} + a(e^{y_{k+1}^j} - e^{y_k^i}e^{r\delta}))) \\ \forall w_k^i &= (\pi_k(i), y_k^i) \in \Gamma_k, \quad \forall v \in \Gamma^V, \quad k = 0, \dots, n-1. \end{aligned}$$

Numerical tests are performed by using the following parameter values:

- Price at time 0 : $S_0 = 110$;
- Strike of the European put option: $K = 110$;
- Riskless interest rate over the interval $[0, 1]$: $r = 0.05$;
- Volatility : $x_0 = 0, 15$, $\Delta = 0, 05$, $\lambda = 0, 1$, and $\eta = 0, 1$.
- Quantization of (Π, Y) : grids have same size N for each time period with step $\delta = \frac{1}{n}$, and they are obtained by using 10^6 iterations of the procedure described in PHAM, RUNGALDIER and SELLAMI [2004];
- Discretization of V^α : we use an N^V -point grid defined by $\Gamma^V = (2\nu)\mathbb{Z} \cap [V_{\inf}, V_{\sup}]$, where ν , $V_{\inf}$, and $V_{\sup}$, determined by performing some preliminary tests, are given by:

$$\nu = \frac{25}{2(N^V - 1)}, \quad V_{\inf} = -10, \quad V_{\sup} = 15;$$

- Approximation of the optimal control : golden search method (see LUENBERGER [1984]) on $A = [-1, 1]$
- When not specified, the number of time steps is $n = 5$.

In order to study the effects of the quantization grid size N and uniform grid size N^V , we plot the graph of $V_0 \mapsto \inf_{\alpha \in A} \mathbb{E}((K - e^{Y_n})_+ - V_n^\alpha)^2$ for different values of N and N^V (Figs. 5.1 and 5.3).

As expected, the global shape of the graph is parabolic, due to the quadratic hedging criterion that we have used. The minimum is reached at $v_{\min}$, which can be considered the quadratic hedging price of our European put option.

Corresponding hedging strategies at time $t = 0$ are given in Tables 5.1 and 5.2, and Fig. 5.2 shows the graph of α_0 as a function of the initial wealth V_0 . We can observe

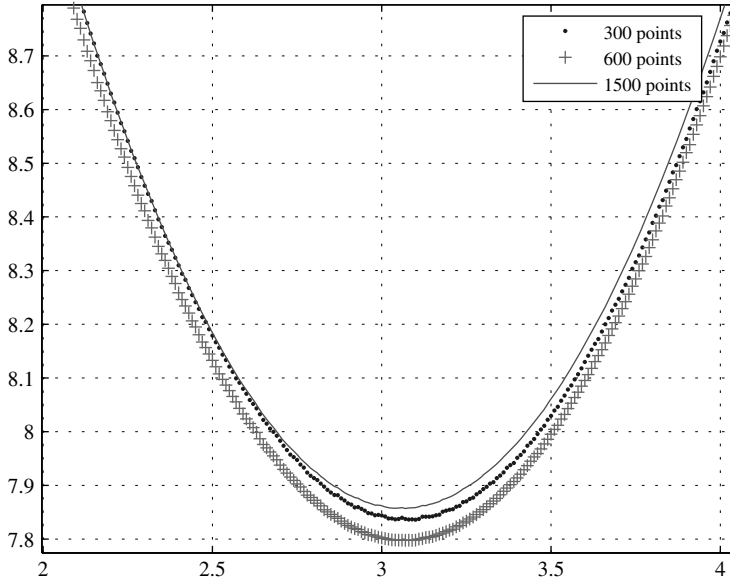


FIG. 5.1 Quadratic hedging of an European put: graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_n^\alpha)^2$ for different quantification grid sizes ($N = 300, 600, 1500$) and a fixed uniform grid size ($N^V = 400$).

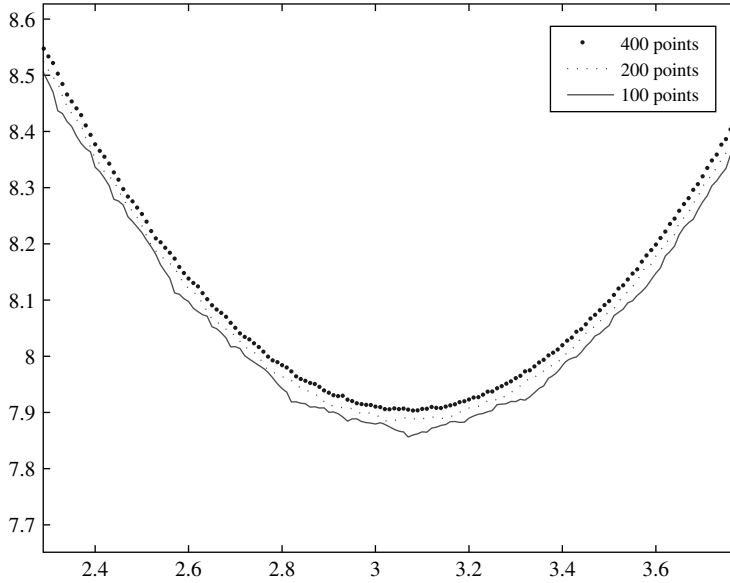


FIG. 5.2 Quadratic hedging of an European put: graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_n^\alpha)^2$ for different fixed uniform grid sizes ($N^V = 50, 100, 200, 400$) and a fixed quantization grid size ($N = 300$).

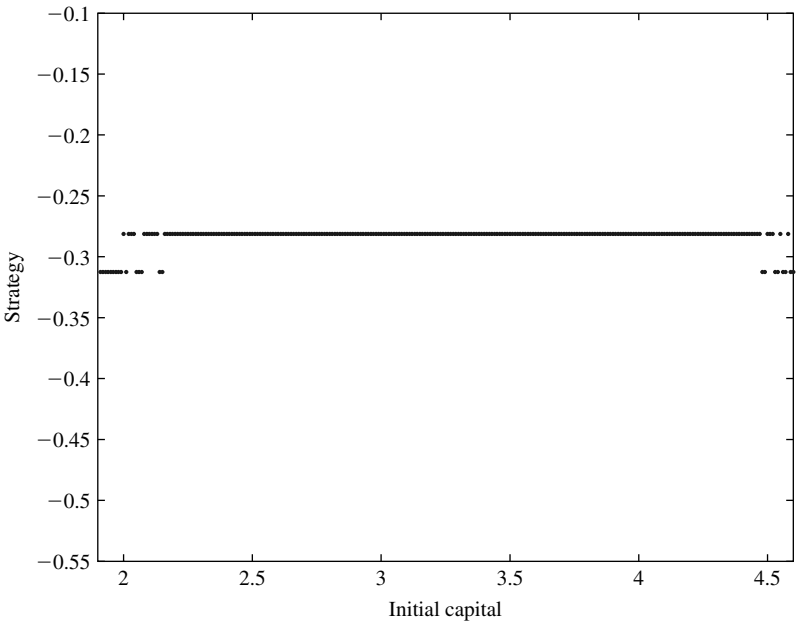


FIG. 5.3 Quadratic hedging of an European put: graph of $V_0 \mapsto \alpha_0(V_0)$ for a quantization grid size of $N = 300$ and a fixed uniform grid size of $N^V = 400$.

TABLE 5.1
Quadratic hedging of an European put: European put price (defined as the initial capital minimizing the risk) and optimal control strategy calculated for different quantization grid sizes ($N=300, 600, 1500$) and a fixed uniform grid size ($N^V = 400$)

N	European put price	Optimal control strategy α_0
300	3.04132	−0.2813
600	3.05965	−0.2813
1500	3.07098	−0.2813

that the strategy is nearly constant for $V_0 \in [2, 4]$, where the nonconstant values may be due to numerical imprecision. This result can be explained¹ by observing that in our example, the discounted price process $\tilde{S}_k = S_k e^{-rk\delta}$, $k = 0, \dots, n$, is a martingale, and by applying the Kunita Watanabe decomposition to the discounted option payoff $F = e^{-r}(K - e^{Y_n})_+$, we get

$$F = \mathbb{E}[F] + \sum_{k=1}^n \alpha_k^F \Delta \tilde{S}_k + R_n^F, \tag{5.4}$$

¹For more details concerning the quadratic hedging in the martingale case, see F         and SONDERMANN [1986].

TABLE 5.2
Quadratic hedging of an European put: European put price (defined as the initial capital minimizing the risk) and optimal control strategy calculated for different fixed uniform grid sizes ($N^V = 50, 100, 200, 400$) and a fixed quantization grid size ($N = 300$)

N^V	European put price	Optimal control strategy α_0
50	2.97501	-0.2813
100	3.04132	-0.2813
300	3.04132	-0.2813
400	3.04132	-0.2813

where $\Delta\tilde{S}_k := \tilde{S}_{k+1} - \tilde{S}_k$, α^F is an admissible control process and R^F is a martingale orthogonal to $\tilde{S}_k$, that is, $E[R_k^F \Delta\tilde{S}_k] = 0$, $k = 0, \dots, n$. Recalling the dynamics (5.2) of the wealth V_n^α , we can write again the objective function as

$$\mathbb{E}\left[\left((K - e^{Y_n})_+ - V_n^\alpha\right)^2\right] = e^{2r}\mathbb{E}\left[\left(F - v_0 - \sum_{k=0}^n \alpha_k \Delta\tilde{S}_k\right)_+^2\right]. \quad (5.5)$$

By combining (5.4) and (5.5) and by exploiting the orthogonality between R^F and $\tilde{S}$, we obtain

$$\begin{aligned} \mathbb{E}\left[\left((K - e^{Y_n})_+ - V_n^\alpha\right)^2\right] &= e^{2r} \left\{ \left(\mathbb{E}[F] - v_0\right)^2 \right. \\ &\quad \left. + \mathbb{E}\left[\left(\sum_{k=1}^n (\alpha_k^F - \alpha_k) \Delta\tilde{S}_k\right)^2\right] + \mathbb{E}\left[(R_n^F)^2\right] \right\}, \end{aligned}$$

which shows that the optimal control is always $\alpha^{\text{opt}} = \alpha^F$ regardless of v .

In Fig. 5.4 and in the Table 5.3, we compare the European put option price under partial and complete observation when we increase the number of observations (i.e., the time step δ decreases to zero). Denoting by $N_{\Pi,Y}$ the number of grid points used in the partial observation case to make an optimal quantization of the pair (Π, Y) , by $N_{X,Y}$ the number of grid points used in the total observation case to make an optimal quantization of the pair (X, Y) , and by $R = V_{\text{sup}} - V_{\text{inf}}$ the grid size in the discretization of the controlled variable V^α , we recall that the discretization error is of order

$$\left(N_{\Pi,Y}^{\frac{-1}{d+m-1}} + v + \frac{1}{R}\right)$$

for the partial observation case. For the total observation case, we have

$$\left(\frac{1}{N_{X,Y}} + v + \frac{1}{R}\right),$$

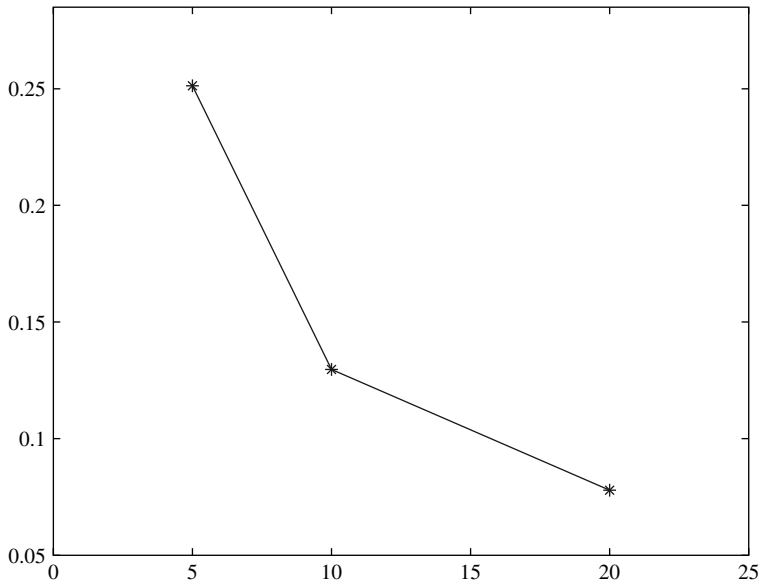


FIG. 5.4 Quadratic hedging of an European put: distance between total and partial observation European put prices (defined as the initial capital minimizing the risk) when we increase the number of observations (axis of abscissae) and consequently the time step δ goes to 0. Size grid for $V^\alpha = 30$ points, and size grid for $(e^Y, \Pi) = 1500$ points, and size grid for $(e^Y, X) = 45$ points.

TABLE 5.3

Quadratic hedging of an European put: comparison between partial and total observation prices (defined as the initial capital minimizing the quadratic risk) and strategies when we increase the number of observations and consequently the time step δ goes to 0. Size grid for $V^\alpha = 30$ points, size grid for $(e^Y, \Pi) = 1500$ points, and size grid for $(e^Y, X) = 45$ points

Time step δ	Partial observation price	Partial observation strategy	Total observation price	Total observation strategy
1\5	2.9933	-0.2813	3.24459	-0.2734
1\10	3.5255	-0.3013	3.65515	-0.2422
1\20	3.9501	-0.3215	4.02799	-0.3614

where $N_{X,Y} = mN_Y$ (see PHAM, RUNGALDIER and SELLAMI [2004]). So, in order to obtain comparable results, given the uniform grid discretizing the variable V^α , we perform an optimal quantization of (Π, Y) and (X, Y) by using grid sizes $N_{\Pi,Y}$ and $N_{X,Y} = mN_Y$ such that

$$N_Y \simeq N_{\Pi,Y}^{\frac{1}{d+m-1}},$$

where $d = 1$ and $m = 3$. Hence, we have chosen $N_{\Pi,Y} = 1500$ and $N_{X,Y} = 45$.

We notice that when the number of observations increases (i.e., $\delta \rightarrow 0$), the partial observation price converges to the complete observation price; this is due to the fact that with observation performed in continuous time, we are able to calculate the volatility given by the quadratic variation of the price process (e^Y).

Figure 5.5 shows that by working in a total observation setting, the quadratic risk associated with a given initial wealth is smaller than the corresponding value obtained in the partial observation case. This is consistent with the fact that the filtration generated by the observation price is included in the full information filtration, and consequently the corresponding optimal cost function in the partial information case is larger than the one in the full information case.

5.3. Hedging of an European put option: shortfall risk criterion

Using the shortfall risk criterion (see FÖLLMER and LEUKERT [2000]), an optimal strategy is a solution to the optimization problem

$$\inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\left((K - e^{Y_n})_+ - V_n^\alpha \right)_+^2 \right], \quad (5.6)$$

where $\mathcal{A}$ is the control space.

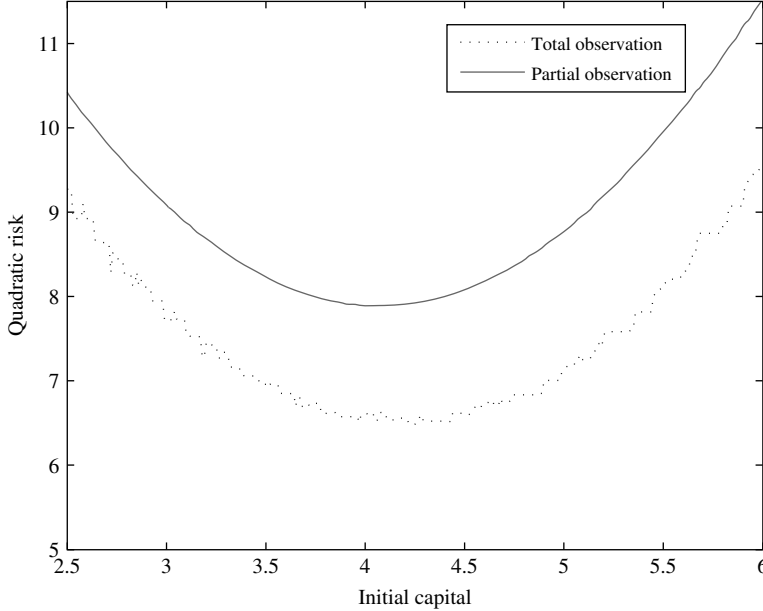


FIG. 5.5 Quadratic hedging of an European put: graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_n^\alpha)^2$ in the partial and total observation cases. Size grid for $V = 100$ points, size grid for $(e^Y, \Pi) = 1500$ points, and size grid for $(e^Y, X) = 45$ points.

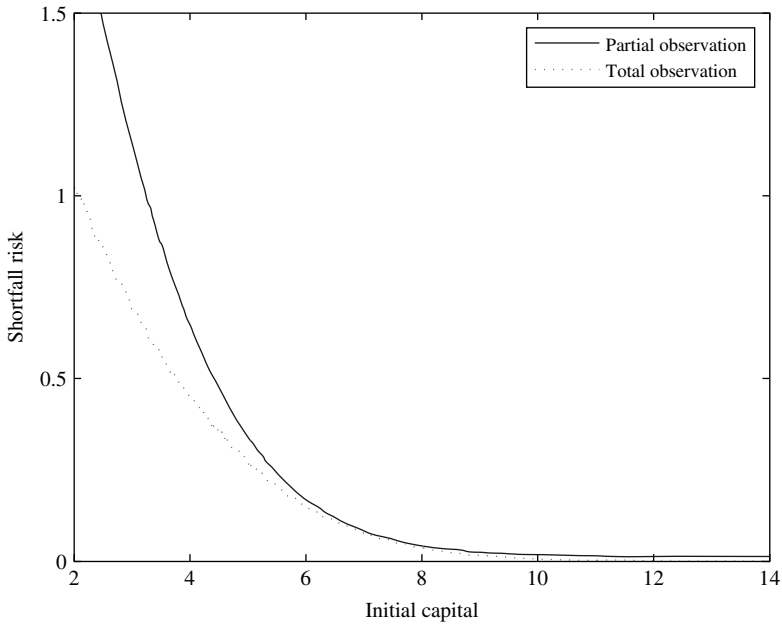


FIG. 5.6 European put option. Shortfall risk criterion: graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_n)_+$ in the partial and total observation cases. Size grid for $V = 100$ points, size grid for $(e^Y, \Pi) = 600$ points, and size grid for $(e^Y, X) = 45$ points.

Fig. 5.6 and the Table 5.4 are obtained by applying the procedure described in the previous section with

$$V_{\inf} = -15 \quad \text{and} \quad V_{\sup} = 25.$$

Figure 5.6 shows the graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_T^\alpha)_+$ in the partial and in the total observation case. Notice that, as expected, the shortfall risk given by $\inf_{\alpha \in \mathcal{A}} \mathbb{E}((K - e^{Y_n})_+ - V_T^\alpha)_+$ decreases with the initial capital and becomes zero for approximately the same value of V_0 in the partial and total observation cases. Notice also that if we tolerate a little risk, we can considerably reduce the requested initial capital.

Moreover, as in the quadratic hedging, for a given initial value V_0 , the shortfall risk obtained in a partial observation setting is greater than the corresponding one in a context of total observation.

In Table 5.4, we compare the initial capital required to minimize the quadratic risk and the shortfall risk associated with our European put option. As expected, the initial capital necessary to minimize the quadratic risk is bounded by the corresponding one in the shortfall risk case, which is actually the superhedging price.

Finally, Fig. 5.7 shows the quadratic and shortfall risks for various values of the initial capital.

TABLE 5.4

European put option: comparison between quadratic hedging and shortfall hedging. $v_{\min}$ is the initial capital requested to minimize the corresponding risk. Size grid for $V^\alpha = 100$ points, and size grid for $(e^Y, \Pi) = 1500$ points, size grid for $(e^Y, X) = 45$ points

Case	Quadratic hedging $v_{\min}$	Shortfall hedging $v_{\min}$	Quadratic hedging strategy	Shortfall hedging strategy
Total observation	3.5750	~ 16	-0.2656	-0.98995
Partial observation	3.07098	~ 17.8	-0.2813	-0.99187

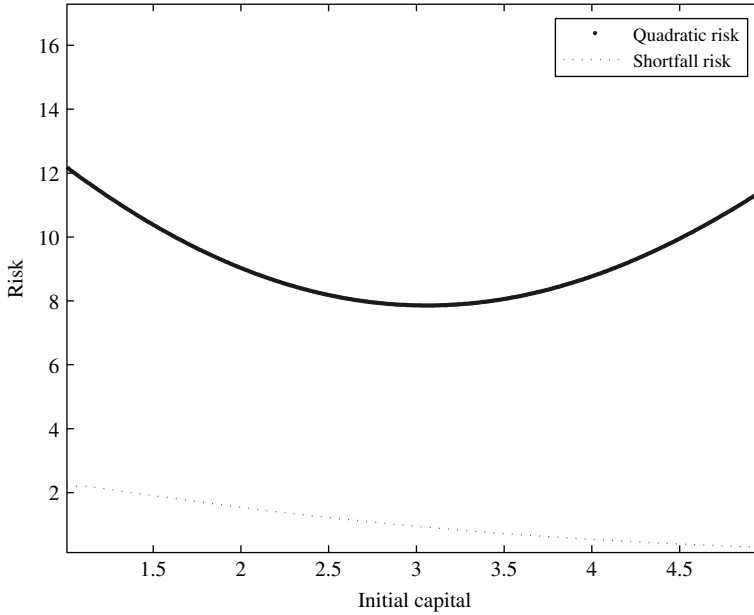


FIG. 5.7 European put option: comparison between quadratic hedging and shortfall hedging: graph of $V_0 \mapsto \inf_{\alpha \in A} \mathbb{E}((K - e^{Y_n})_+ - V_n^\alpha)_+$ in the partial observation case. Size grid for $V^\alpha = 100$ points and size grid for $(e^Y, \Pi) = 600$ points.

5.4. Hedging of an European call option

5.4.1. Quadratic hedging

An optimal strategy is a solution to the optimization problem

$$\inf_{\alpha \in A} \mathbb{E} \left[\left((e^{Y_n} - K)_+ - V_n^\alpha \right)^2 \right], \quad (5.7)$$

where $\mathcal{A}$ is the control space. The procedure described in the previous section has been applied by taking

$$V_{\inf} = -20 \quad \text{and} \quad V_{\sup} = 30.$$

Figure 5.8 shows the graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((e^{Y_n} - K)_+ - V_n^\alpha)^2)$, that is, the quadratic risk as a function of the initial capital V_0 . Notice that as expected the global shape of the graph is parabolic; the initial capital corresponding to the minimum can be interpreted as the quadratic hedging price of the European call option. Notice also that as in the European put case, for a given initial wealth, the corresponding quadratic risk in the partial observation case is greater than in the total observation case.

Figure 5.9 shows the graph of the optimal strategy at time $t = 0$ as a function of the initial capital. As in the put case the optimal strategy is nearly constant.

Finally, in Table 5.5 we compare the initial capital requested to minimize the quadratic risk (quadratic hedging price) with the call price obtained by using the put-call parity relation and the quadratic hedging put price calculated in the previous section. We can observe that the two prices are very close, thus justifying further the expression quadratic hedging price.

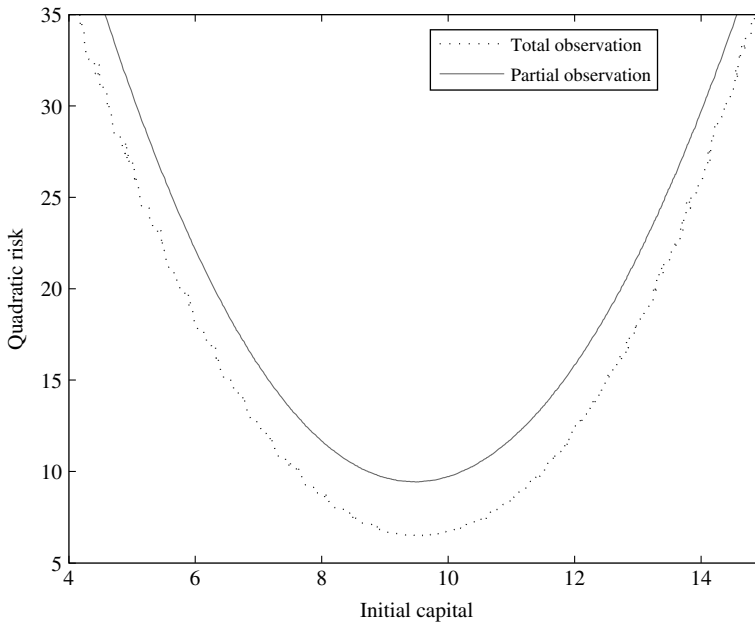


FIG. 5.8 Quadratic hedging of an European call: graph of $V_0 \mapsto \inf_{\alpha \in \mathcal{A}} \mathbb{E}((e^{Y_n} - K)_+ - V_n^\alpha)^2)$. Size grid for $V^\alpha = 100$ points and size grid for $(e^Y, \Pi) = 600$ points

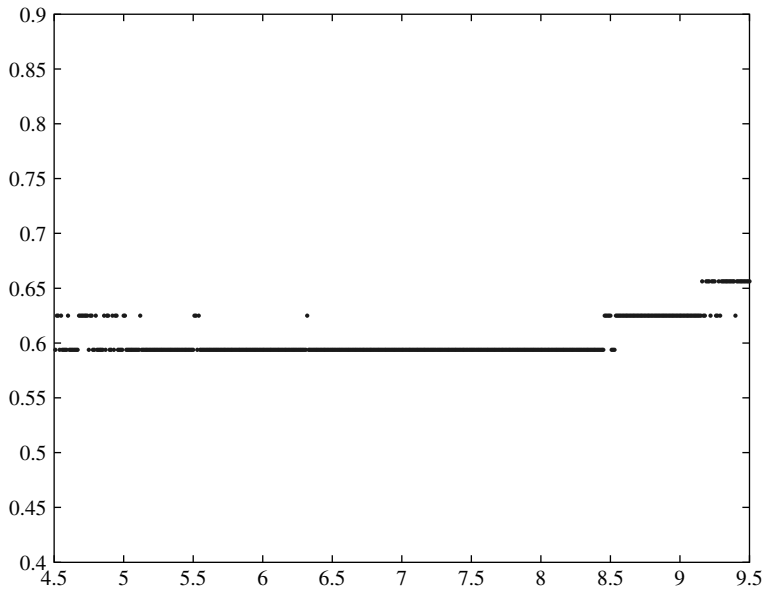


FIG. 5.9 Quadratic hedging of an European call: graph of $V_0 \mapsto \alpha_0(V_0)$. Size grid for $V^\alpha = 100$ points and size grid for $(e^Y, \Pi) = 600$ points

TABLE 5.5

European call option: comparison between quadratic hedging and put-call parity. Size grid for $V^\alpha = 100$ points, size grid for $(e^Y, \Pi) = 300$ points, and size grid for $(e^Y, X) = 45$ points

Case	Call price by quadratic hedging	Call price by call-put parity	Difference
Total observation	8.74596	8.91377	0.1678
Partial observation	8.55202	8.38009	0.1719

5.4.2. Shortfall risk criterion

An optimal strategy is a solution to the optimization problem

$$\inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\left((e^{Y_n} - K)_+ - V_n^\alpha \right)_+ \right], \quad (5.8)$$

where $\mathcal{A}$ is the control space.

Figure 5.10 and Table 5.6 are obtained by applying the procedure described in the previous sections with

$$V_{\inf} = -25 \quad \text{and} \quad V_{\sup} = 35.$$

Appendix A: Proof of Proposition 3.1

We begin with a definition and a preliminary result:

DEFINITION A.1. Let $\alpha = (\alpha_k)_k$ be a fixed control process. Functions u_k^α ($k = 0, \dots, n$) are defined recursively by

$$\begin{cases} u_n^\alpha(\pi, y, v) := \hat{h}(\pi, y, v) \\ u_k^\alpha(\pi, y, v) := \hat{f}(\pi, y, v, \alpha_k) \\ \quad + \mathbb{E} \left[u_{k+1}^\alpha(\Pi_{k+1}, Y_{k+1}, H(v, \alpha_k, y, Y_{k+1})) \mid (\Pi_k, Y_k) = (\pi, y) \right]. \end{cases}$$

LEMMA A.1. Assume **H1**, **H2**, and **H3**. Then, there exists a control process $\tilde{\alpha} = (\tilde{\alpha}_k)_k \in \mathcal{A}$, such that for all $k = 0, \dots, n-1$,

$$u_k(\pi, y, v) = u_k^{\tilde{\alpha}}(\pi, y, v), \quad (\pi, y, v) \in K_m \times \mathbb{R}^d \times \mathbb{R}.$$

PROOF. The function u_k is defined by

$$\begin{aligned} u_k(\pi, y, v) = \inf_{a \in A} \Big\{ & \hat{f}(\pi, y, v, a) \\ & + \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \mid (\Pi_k, Y_k) = (\pi, y) \right] \Big\}, \end{aligned}$$

and we see that the terms in brackets are continuous functions with respect to (v, a, y) . Indeed, $\hat{f}$ is Lipschitz, and the second term can be written as follows:

$$\begin{aligned} F_k(\pi, y, v, a) &:= \\ &= \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \mid (\Pi_k, Y_k) = (\pi, y) \right] \\ &= \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, V_{k+1}^\alpha) \mid \mathcal{F}_k^Y \right] \\ &= \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \mid \mathcal{F}_k^Y \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \mid \mathcal{F}_k \right] \mid \mathcal{F}_k^Y \right] \\ &= \mathbb{E} \left[\int \sum_{i=1}^m u_{k+1}(\bar{G}_{k+1}(\pi, y, y'), y', H(v, a, y, Y_{k+1})) g_{k+1}(X_k, y, x^j, y') \right. \\ &\quad \left. \mathbb{P}[X_{k+1} = x^j \mid X_k] dy' \mid \mathcal{F}_k^Y \right] \\ &= \int \sum_{i,j=1}^m u_{k+1}(\bar{G}_{k+1}(\pi, y, y'), y', H(v, a, y, Y_{k+1})) g_{k+1} \\ &\quad (x^i, y, x^j, y') P_{k+1}^{ij} \Pi_k^i dy', \end{aligned}$$

which is a continuous function with respect to (π, y, v, a) .

By exploiting this fact, we build the requested control process following a backward recursion:

$$\begin{aligned}
 u_n(\pi, y, v) &= \hat{h}(\pi, y, v) \\
 u_{n-1}(\pi, y, v) &= \inf_{a \in A} \left\{ \hat{f}(\pi, y, v, a) \right. \\
 &\quad \left. + \mathbb{E} \left[u_n(\Pi_n, Y_n, H(v, a, y, Y_n)) \middle| (\Pi_{n-1}, Y_{n-1}) = (\pi, y) \right] \right\} \\
 &= \inf_{a \in A} \left[\hat{f}(\pi, y, v, a) + F_{n-1}(\pi, y, v, a) \right].
 \end{aligned}$$

Since A is a compact set and the argument of the infimum is a continuous function with respect to (π, y, v, a) , we deduce the existence of

$$\tilde{\alpha}_{n-1}(\pi, y, v) \in \arg \min_{a \in A} \left[\hat{f}(\pi, y, v, a) + F_{n-1}(\pi, y, v, a) \right]$$

for almost every $(\pi, y, v) \in K_m \times \mathbb{R}^d \times \mathbb{R}^l$, which may be chosen to be Borel measurable by a classical measurable selection theorem (see Proposition 7.33 in BERTSEKAS and SHREVE [1996]). By using the same argument, at the generic time step k , we have

$$\begin{aligned}
 u_k(\pi, y, v) &= \inf_{a \in A} \left[\hat{f}(\pi, y, v, a) \right. \\
 &\quad \left. + \mathbb{E} \left[u_{k+1}(\Pi_{k+1}, Y_{k+1}, H(v, a, y, Y_{k+1})) \middle| (\Pi_k, Y_k) = (\pi, y) \right] \right] \\
 &= \hat{f}(\pi, y, v, \tilde{\alpha}_k(\pi, y, v)) + F_k(\pi, y, v, \tilde{\alpha}_k(\pi, y, v)).
 \end{aligned}$$

Finally, we define the $\mathcal{F}^Y$ -adapted process $\tilde{\alpha}$ as follows:

$$\tilde{\alpha} := (\tilde{\alpha}_k(\Pi_k, Y_k, V_k))_k,$$

and we obtain by construction

$$u_k(\pi, y, v) = u_k^{\tilde{\alpha}}(\pi, y, v) \quad \text{for all } k \geq 0.$$

□

Proof of Proposition 3.1

We shall prove that

$$\inf_{\alpha \in \mathcal{A}} u_0^\alpha(\mu, y_0, v_0) = u_0(\mu, y_0, v_0) = J_{\text{opt}}(v_0). \quad (\text{A.1})$$

First, we easily show by induction that

$$u_k(\pi, y, v) \leq u_k^\alpha(\pi, y, v), \quad k = 0, \dots, n, \quad \alpha \in \mathcal{A}, \quad (\text{A.2})$$

for all (π, y, v) . Now, fix some arbitrary control $\alpha \in \mathcal{A}$. By taking expectation in the definition of u_k^α , we have

$$\begin{aligned} \mathbb{E}[u_k^\alpha(\Pi_k, Y_k, V_k^\alpha)] &= \mathbb{E}\left[\hat{f}(\Pi_k, Y_k, V_k^\alpha, \alpha_k)\right] \\ &\quad + \mathbb{E}\left[u_{k+1}^\alpha(\Pi_{k+1}, Y_{k+1}, V_{k+1}^\alpha)\right], \quad k = 0, \dots, n-1. \end{aligned}$$

By adding up for k running from 0 to $n-1$, we get

$$\begin{aligned} u_0^\alpha(\mu, y_0, v_0) &= \mathbb{E}\left[\sum_{k=0}^{n-1} \hat{f}(\Pi_k, Y_k, V_k^\alpha, \alpha_k) + u_n^\alpha(\Pi_n, Y_n, V_n^\alpha)\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{n-1} \hat{f}(\Pi_k, Y_k, V_k^\alpha, \alpha_k) + \hat{h}(\Pi_n, Y_n, V_n^\alpha)\right] = J(v_0, \alpha). \end{aligned} \quad (\text{A.3})$$

From (A.2) and (A.3), we then get

$$u_0(\mu, y_0, v_0) \leq \inf_{\alpha \in \mathcal{A}} J(v_0, \alpha) = J_{\text{opt}}(v_0). \quad (\text{A.4})$$

Moreover, from Lemma A.1, there exists some $\tilde{\alpha} \in \mathcal{A}$ such that $u_0(\mu, y_0, v_0) = u_0^{\tilde{\alpha}}(\mu, y_0, v_0)$. Together with (A.3) and (A.4), this proves (A.1).

Appendix B: Proof of Theorem 4.1

We first give some estimations on the functions u_k^α defined in (A.1).

LEMMA B.1. *Assume **H2**, then we have for all $k = 0, \dots, n$ and $\alpha \in \mathcal{A}$*

$$[u_k^\alpha]_{\text{sup}} \leq (n-k)\bar{f} + \bar{h},$$

where

$$\bar{f} := \max([f]_{\text{sup}}, [f]_{\text{Lip}}) \quad \text{and} \quad \bar{h} := \max([h]_{\text{sup}}, [h]_{\text{Lip}}).$$

PROOF. By definition of u_k^α , we clearly have

$$[u_k^\alpha]_{\text{sup}} \leq \bar{f} + [u_{k+1}^\alpha]_{\text{sup}}$$

and so by induction

$$[u_k^\alpha]_{\text{sup}} \leq (n-k)\bar{f} + [u_n^\alpha]_{\text{sup}} \leq (n-k)\bar{f} + \bar{h}.$$

□

LEMMA B.2. Assume **H2** and **H4** and set for all $k = 0, \dots, n$, $(\pi, \hat{\pi}, y, \hat{y}, v, \hat{v}) \in K_m \times K_m \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$, $\alpha \in \mathcal{A}$

$$B_1(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) = \int \left| u_{k+1}^\alpha(\bar{G}_{k+1}(\pi, y, y'), y', H(v, \alpha_k, y, y')) \right. \\ \left. - u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) \right| Q_{k+1}(\pi, y, dy'),$$

where $Q_k(\Pi_{k-1}, Y_{k-1}, dy')$ denotes the conditional law of Y_k , given (Π_{k-1}, Y_{k-1}) , and $\bar{G}_k$ is defined in (3.1). Then, we have

$$B_1(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) \leq 2[u_{k+1}^\alpha]_{\text{Lip}} (L_g |y - \hat{y}|_1 + \|\pi - \hat{\pi}\|_1) \\ + [H]_{\text{Lip}} (|v - \hat{v}|_1 + |y - \hat{y}|_1).$$

PROOF. Under assumption **H2**, we have

$$B_1(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) \tag{B.1}$$

$$= \int \left| u_{k+1}^\alpha(\bar{G}_{k+1}(\pi, y, y'), y', H(v, \alpha_k, y, y')) \right. \tag{B.2}$$

$$\left. - u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) \right| Q_{k+1}(\pi, y, dy') \\ \leq [u_{k+1}^\alpha]_{\text{Lip}} \int \left| \bar{G}_{k+1}(\pi, y, y') - \bar{G}_{k+1}(\hat{\pi}, \hat{y}, y') \right|_1 Q_{k+1}(\pi, y, dy') \\ + \int \left| H(v, \alpha_k, y, y') - H(\hat{v}, \alpha_k, \hat{y}, y') \right|_1 Q_{k+1}(\pi, y, dy') \\ \leq [u_{k+1}^\alpha]_{\text{Lip}} \int \left| \bar{G}_{k+1}(\pi, y, y') - \bar{G}_{k+1}(\hat{\pi}, \hat{y}, y') \right|_1 Q_{k+1}(\pi, y, dy') \\ + [H]_{\text{Lip}} (|v - \hat{v}|_1 + |y - \hat{y}|_1). \tag{B.3}$$

Now, from (3.1) and (3.2), we have

$$\int \left| \bar{G}_{k+1}(\pi, y, y') - \bar{G}_{k+1}(\hat{\pi}, \hat{y}, y') \right|_1 Q_{k+1}(\pi, y, dy') \\ = \int \left| \bar{G}_{k+1}(\pi, y, y') - \bar{G}_{k+1}(\hat{\pi}, \hat{y}, y') \right|_1 q_{k+1}(\pi, y, y') dy' \\ \leq \sum_{i,j=1}^m \int \left| \frac{g_{k+1}(x^i, y, x^j, y') P_k^{ij} \pi^i}{q_{k+1}(\pi, y, y')} - \frac{g_{k+1}(x^i, \hat{y}, x^j, y') P_k^{ij} \hat{\pi}^i}{q_{k+1}(\hat{\pi}, \hat{y}, y')} \right| q_{k+1}(\pi, y, y') dy'$$

$$\begin{aligned}
&\leq \sum_{i,j=1}^m P_k^{ij} \hat{\pi}^j \int \left| \frac{g_{k+1}(x^i, y, x^j, y') q_{k+1}(\hat{\pi}, \hat{y}, y')}{q_{k+1}(\pi, y, y')} \right. \\
&\quad \left. - \frac{g_{k+1}(x^i, \hat{y}, x^j, y') q_{k+1}(\pi, y, y')}{q_{k+1}(\hat{\pi}, \hat{y}, y')} \right| dy' + \sum_{i=1}^m |\pi^i - \hat{\pi}^i| \\
&\leq \sum_{i,j=1}^m P_k^{ij} \int |g_{k+1}(x^i, y, x^j, y') - g_{k+1}(x^i, \hat{y}, x^j, y')| dy' \\
&\quad + \int |q_{k+1}(\pi, y, y') - q_{k+1}(\hat{\pi}, \hat{y}, y')| dy' + \sum_{i=1}^m |\pi^i - \hat{\pi}^i| \\
&\leq 2 \sum_{i,j=1}^m P_k^{ij} \int |g_{k+1}(x^i, y, x^j, y') - g_{k+1}(x^i, \hat{y}, x^j, y')| dy' + 2 \sum_{i=1}^m |\pi^i - \hat{\pi}^i|.
\end{aligned} \tag{B.4}$$

Plugging (B.4) into (B.3) and using assumption **(H4)**, we get the required result. $\square$

LEMMA B.3. Assume **H4** and set for all $k = 0, \dots, n$, $(\pi, \hat{\pi}, y, \hat{y}, v, \hat{v}) \in K_m \times K_m \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$, $\alpha \in \mathcal{A}$

$$\begin{aligned}
B_2(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) &= \int \left| u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) \right| \\
&\quad \left[Q_{k+1}(\hat{\pi}, \hat{y}, dy') - Q_{k+1}(\pi, y, dy') \right].
\end{aligned}$$

Then, we have

$$B_2(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) \leq [u_{k+1}^\alpha]_{\sup} L_g |y - \hat{y}|_1 + [u_{k+1}^\alpha]_{\sup} |\pi - \hat{\pi}|_1.$$

PROOF. From (3.2), we have

$$\begin{aligned}
B_2(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha) &\leq [u_{k+1}^\alpha]_{\sup} \int |q_{k+1}(\hat{\pi}, \hat{y}, y') - q_{k+1}(\pi, y, y')| dy' \\
&\leq [u_{k+1}^\alpha]_{\sup} \sum_{i,j=1}^m P_k^{ij} \int |g_{k+1}(x^i, y, x^j, y') \\
&\quad - g_{k+1}(x^i, \hat{y}, x^j, y')| dy' \\
&\quad + [u_{k+1}^\alpha]_{\sup} \sum_{i=1}^m |\pi^i - \hat{\pi}^i|,
\end{aligned}$$

and we conclude with **H4**. $\square$

LEMMA B.4. *Let **H2**, **H3**, and **H4** hold. Then, for all $k = 0, \dots, n$, the function u_k^α is Lipschitz, uniformly with respect to α and*

$$[u_k^\alpha]_{\text{Lip}} \leq \mathcal{L}_k,$$

where

$$\mathcal{L}_k := \left(\bar{L}_g \bar{f}(n-k) + \bar{M} + 3\bar{L}_g \bar{h} \right) \frac{(2\bar{L}_g)^{n-k}}{2\bar{L}_g - 1},$$

and $\bar{f} := \max([f]_{\text{sup}}, [f]_{\text{Lip}})$, $\bar{h} := \max([h]_{\text{sup}}, [h]_{\text{Lip}})$, $\bar{L}_g := \max(L_g, 1)$, and $\bar{M} := \max([H]_{\text{Lip}}, 1)$.

PROOF. We denote

$$z := (\pi, y, v), \quad \hat{z} := (\hat{\pi}, \hat{y}, \hat{v}), \quad Z_k^\alpha := (\Pi_k, Y_k, V_k^\alpha)$$

and we have

$$\begin{aligned} [u_k^\alpha]_{\text{Lip}} &\leq [\hat{f}]_{\text{Lip}} + [\mathbb{E}(u_{k+1}^\alpha(Z_{k+1}^\alpha) \mid Z_k^\alpha = z)]_{\text{Lip}} \\ &= [\hat{f}]_{\text{Lip}} + [I_2]_{\text{Lip}}, \end{aligned} \tag{B.5}$$

where

$$I_2 := \mathbb{E}[u_{k+1}^\alpha(Z_{k+1}^\alpha) \mid Z_k^\alpha = z].$$

We have for all $(\pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, a) \in \times K_m \times K_m \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times A$,

$$\begin{aligned} \left| \hat{f}(\pi, y, v, \alpha_k) - \hat{f}(\hat{\pi}, \hat{y}, \hat{v}, \alpha_k) \right| &= \left| \int f(x, y, v, \alpha_k) \pi(\mathrm{d}x) - \int f(x, \hat{y}, \hat{v}, \alpha_k) \hat{\pi}(\mathrm{d}x) \right| \\ &\leq \int |f(x, y, v, \alpha_k) - f(x, \hat{y}, \hat{v}, \alpha_k)| \pi(\mathrm{d}x) \\ &\quad + \int |f(x, \hat{y}, \hat{v}, \alpha_k)| |\hat{\pi} - \pi|(\mathrm{d}x) \\ &\leq [f]_{\text{Lip}} (|y - \hat{y}|_1 + |v - \hat{v}|_1) + [f]_{\text{sup}} |\hat{\pi} - \pi|_1 \\ &\leq \bar{f} |z - \hat{z}|_1, \end{aligned}$$

where $\bar{f} := \max([f]_{\text{sup}}, [f]_{\text{Lip}})$. Therefore, $[\hat{f}]_{\text{Lip}} \leq \bar{f}$. Let us consider the term I_2 . By definition of Q_{k+1} and V_{k+1}^α , we have

$$\begin{aligned}
 & \left| \mathbb{E}[u_{k+1}^\alpha(Z_{k+1}^\alpha) | Z_k^\alpha = z] - \mathbb{E}[u_{k+1}^\alpha(Z_{k+1}^\alpha) | Z_k^\alpha = \hat{z}] \right| \\
 &= \left| \int u_{k+1}^\alpha(\bar{G}_{k+1}(\pi, y, y'), y', H(v, \alpha_k, y, y')) Q_{k+1}(\pi, y, dy') + \right. \\
 &\quad \left. - \int u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) Q_{k+1}(\hat{\pi}, \hat{y}, dy') \right| \\
 &\leq \int \left| u_{k+1}^\alpha(\bar{G}_{k+1}(\pi, y, y'), y', H(v, \alpha_k, y, y')) \right. \\
 &\quad \left. - u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) \right| Q_{k+1}(\pi, y, dy') \\
 &\quad + \int \left| u_{k+1}^\alpha(\bar{G}_{k+1}(\hat{\pi}, \hat{y}, y'), y', H(\hat{v}, \alpha_k, \hat{y}, y')) \right| \times \\
 &\quad \times \left[Q_{k+1}(\hat{\pi}, \hat{y}, dy') - Q_{k+1}(\pi, y, dy') \right] \\
 &= B_1(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha_k) + B_2(k, \pi, \hat{\pi}, y, \hat{y}, v, \hat{v}, \alpha_k). \tag{B.6}
 \end{aligned}$$

By using Lemmas B.2 and B.3, we then get

$$\begin{aligned}
 & \left| \mathbb{E}[u_{k+1}^\alpha(Z_{k+1}^\alpha) | Z_k^\alpha = z] - \mathbb{E}[u_{k+1}^\alpha(Z_{k+1}^\alpha) | Z_k^\alpha = \hat{z}] \right| \\
 &\leq \left(2[u_{k+1}^\alpha]_{\text{Lip}} + [u_{k+1}^\alpha]_{\text{sup}} \right) (L_g |y - \hat{y}|_1 |\pi - \hat{\pi}|_1) \\
 &\quad + [H]_{\text{Lip}} (|v - \hat{v}|_1 + |y - \hat{y}|_1) \\
 &\leq \left(2[u_{k+1}^\alpha]_{\text{Lip}} + [u_{k+1}^\alpha]_{\text{sup}} \right) |z - \hat{z}|_1 \bar{L}_g + \bar{M} |z - \hat{z}|_1,
 \end{aligned}$$

where $\bar{L}_g := \max(L_g, 1)$ and $\bar{M} := \max([H]_{\text{Lip}}, 1)$, and we deduce that

$$[I_2]_{\text{Lip}} \leq \left(2[u_{k+1}^\alpha]_{\text{Lip}} + [u_{k+1}^\alpha]_{\text{sup}} \right) \bar{L}_g + \bar{M}. \tag{B.7}$$

Plugging (B.7) into (B.5) yields

$$[u_k^\alpha]_{\text{Lip}} \leq \bar{f} + \bar{M} + \bar{L}_g \left(2[u_{k+1}^\alpha]_{\text{Lip}} + [u_{k+1}^\alpha]_{\text{sup}} \right)$$

so that from Lemma B.1:

$$\begin{aligned}
[u_k^\alpha]_{\text{Lip}} &\leq \bar{f} + \bar{M} + 2\bar{L}_g[u_{k+1}^\alpha]_{\text{Lip}} + \bar{L}_g\bar{h} + \bar{L}_g(n-k-1) \\
&\leq 2\bar{L}_g[u_{k+1}^\alpha]_{\text{Lip}} + \bar{M} + \bar{L}_g\bar{h} + \bar{L}_g(n-k)\bar{f} \\
&\leq 2\bar{L}_g \left\{ \bar{L}_g(n-k-1)\bar{f} + \bar{M} + \bar{h}\bar{L}_g + 2\bar{L}_g[u_{k+2}^\alpha]_{\text{Lip}} \right\} \\
&\quad + \bar{M} + \bar{h}\bar{L}_g + \bar{L}_g(n-k)\bar{f} \\
&= \bar{L}_g\bar{f}[(n-k) + 2\bar{L}_g(n-k-1)] + (\bar{M} + \bar{h}\bar{L}_g)[1 + 2\bar{L}_g] \\
&\quad + (2\bar{L}_g)^2[u_{k+2}^\alpha]_{\text{Lip}}.
\end{aligned}$$

By induction, this yields

$$\begin{aligned}
[u_k^\alpha]_{\text{Lip}} &\leq \bar{L}_g\bar{f} \sum_{i=0}^{n-k-1} (2\bar{L}_g)^i(n-k-i) + (\bar{M} + \bar{L}_g\bar{h}) \sum_{i=0}^{n-k-1} (2\bar{L}_g)^i + (2\bar{L}_g)^{n-k}\bar{h} \\
&\leq (\bar{L}_g\bar{f}(n-k) + \bar{M} + \bar{L}_g\bar{h}) \left(\frac{(2\bar{L}_g)^{n-k} - 1}{2\bar{L}_g - 1} \right) + \bar{h}(2\bar{L}_g)^{n-k} \\
&\leq (\bar{L}_g\bar{f}(n-k) + \bar{M} + \bar{L}_g\bar{h} + \bar{h}(2\bar{L}_g - 1)) \left(\frac{(2\bar{L}_g)^{n-k}}{2\bar{L}_g - 1} \right) \\
&\leq (\bar{L}_g\bar{f}(n-k) + \bar{M} + 3\bar{L}_g\bar{h}) \frac{(2\bar{L}_g)^{n-k}}{2\bar{L}_g - 1}.
\end{aligned}$$

Therefore,

$$[u_k^\alpha]_{\text{Lip}} \leq (\bar{L}_g\bar{f}(n-k) + \bar{M} + 3\bar{L}_g\bar{h}) \frac{(2\bar{L}_g)^{n-k}}{2\bar{L}_g - 1}$$

and the required result follows. $\square$

We now study estimations for the approximated cost function. Similarly, as in Definition A.1, we introduce the following sequence of functions.

DEFINITION B.1. Let $\alpha = (\alpha_k)_k$ be a control process in $\mathcal{A}$. Functions $\hat{u}_k^\alpha, k = 0, \dots, n$, are defined recursively by

$$\begin{cases} \hat{u}_n^\alpha(\pi, y, v) := \hat{h}(\pi, y, v) \\ \hat{u}_k^\alpha(\pi, y, v) := \hat{f}(\pi, y, v, \alpha_k) \\ \quad + \mathbb{E} \left[\hat{u}_{k+1}^\alpha \left(\hat{\Pi}_{k+1}, \hat{Y}_{k+1}, \hat{V}_{k+1}^\alpha \right) \mid (\hat{\Pi}_k, \hat{Y}_k, \hat{V}_k^\alpha) = (\pi, y, v) \right], \end{cases}$$

and we notice by same arguments as in Proposition 3.1 (see (A.1)) that

$$\inf_{\alpha \in \mathcal{A}} \hat{u}_0^\alpha(\mu, y_0, v_0) = \hat{u}_0(\mu, y_0, v_0) = \hat{J}_{\text{quant}}(v_0). \quad (\text{B.8})$$

For any $\alpha \in \mathcal{A}$, we denote $Z_k^\alpha = (\Pi_k, Y_k, V_k^\alpha)$ and $\hat{Z}_k^\alpha = (\hat{\Pi}_k, \hat{Y}_k, \hat{V}_k^\alpha)$, $k = 0, \dots, n$.

LEMMA B.5. Assume H1, H2, H3, and H4. Then, we have for all $k = 0, \dots, n$, $\alpha \in \mathcal{A}$

$$\|u_k^\alpha(Z_k^\alpha) - \hat{u}_k^\alpha(\hat{Z}_k^\alpha)\|_1 \leq \mathcal{M}_k^{(\alpha)} \quad (\text{B.9})$$

with

$$\mathcal{M}_k^{(\alpha)} := \sqrt{m+d+q} \sum_{i=k}^n \left[2(\bar{L}_g \bar{f}(n-i) + \bar{M} + 3\bar{L}_g \bar{h}) \frac{(2\bar{L}_g)^{n-i}}{2\bar{L}_g - 1} + \bar{f} + \bar{h} \right] \|Z_i^\alpha - \hat{Z}_i^\alpha\|_2.$$

PROOF.

$$\begin{aligned} & \|u_k^\alpha(Z_k^\alpha) - \hat{u}_k^\alpha(\hat{Z}_k^\alpha)\|_1 \\ & \leq \|u_k^\alpha(Z_k^\alpha) - u_k^\alpha(\hat{Z}_k^\alpha)\|_1 + \|u_k^\alpha(\hat{Z}_k^\alpha) - \mathbb{E}[u_k^\alpha(Z_k^\alpha) | \hat{Z}_k^\alpha]\|_1 \\ & \quad + \|\mathbb{E}[u_k^\alpha(Z_k^\alpha) | \hat{Z}_k^\alpha] - \hat{u}_k^\alpha(\hat{Z}_k^\alpha)\|_1 \\ & \leq 2 \|u_k^\alpha(Z_k^\alpha) - u_k^\alpha(\hat{Z}_k^\alpha)\|_1 + \|\mathbb{E}[u_k^\alpha(Z_k^\alpha) | \hat{Z}_k^\alpha] - \hat{u}_k^\alpha(\hat{Z}_k^\alpha)\|_1 \\ & = I_1 + I_2, \end{aligned} \quad (\text{B.10})$$

with

$$I_1 := 2 \|u_k^\alpha(Z_k^\alpha) - u_k^\alpha(\hat{Z}_k^\alpha)\|_1$$

and

$$I_2 := \|\mathbb{E}[u_k^\alpha(Z_k^\alpha) | \hat{Z}_k^\alpha] - \hat{u}_k^\alpha(\hat{Z}_k^\alpha)\|_1.$$

Consider now the term I_2 :

$$\begin{aligned} I_2 &= \left\| \mathbb{E}[\hat{f}(Z_k^\alpha, \alpha) + \mathbb{E}[u_{k+1}^\alpha | Z_k^\alpha] | \hat{Z}_k^\alpha] - \hat{f}(\hat{Z}_k^\alpha, \alpha) + \mathbb{E}[\hat{u}_{k+1}^\alpha | \hat{Z}_k^\alpha] \right\|_1 \\ &= \left\| \mathbb{E}[\hat{f}(Z_k^\alpha, \alpha) - \hat{f}(\hat{Z}_k^\alpha, \alpha) + u_{k+1}^\alpha(Z_{k+1}^\alpha) - \hat{u}_{k+1}^\alpha(\hat{Z}_{k+1}^\alpha) | \hat{Z}_k^\alpha] \right\|_1 \\ &\leq \|\hat{f}(Z_k^\alpha, \alpha) - \hat{f}(\hat{Z}_k^\alpha, \alpha)\|_1 + \|u_{k+1}^\alpha(Z_{k+1}^\alpha) - \hat{u}_{k+1}^\alpha(\hat{Z}_{k+1}^\alpha)\|_1 \\ &\leq \bar{f} \|\hat{Z}_k^\alpha - Z_k^\alpha\|_1 + \|u_{k+1}^\alpha(Z_{k+1}^\alpha) - \hat{u}_{k+1}^\alpha(\hat{Z}_{k+1}^\alpha)\|_1. \end{aligned} \quad (\text{B.11})$$

Concerning the term I_1 , we have

$$I_1 \leq 2\mathcal{L}_k \left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_1, \quad (\text{B.12})$$

where we have used the Proposition B.4.

Plugging (B.12) and (B.11) into (B.10) yields

$$\begin{aligned} & \left\| u_k^\alpha(Z_k^\alpha) - \hat{u}_k^\alpha(Z_k^\alpha) \right\|_1 \\ & \leq (2\mathcal{L}_k + \bar{f}) \left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_1 + \left\| u_{k+1}^\alpha(Z_{k+1}^\alpha) - \hat{u}_{k+1}^\alpha(\hat{Z}_{k+1}^\alpha) \right\|_1 \\ & \leq (2\mathcal{L}_k + \bar{f}) \left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_1 + (2\mathcal{L}_{k+1} + \bar{f}) \left\| Z_{k+1}^\alpha - \hat{Z}_{k+1}^\alpha \right\|_1 \\ & \quad + \left\| u_{k+2}^\alpha(Z_{k+2}^\alpha) - \hat{u}_{k+2}^\alpha(\hat{Z}_{k+2}^\alpha) \right\|_1 \\ & \leq \sum_{i=k}^{n-1} \left\| Z_i^\alpha - \hat{Z}_i^\alpha \right\|_1 (2\mathcal{L}_i + \bar{f}) + \bar{h} \left\| Z_n - \hat{Z}_n \right\|_1 \\ & \leq \sum_{i=k}^n \left[2(\bar{L}_g \bar{f}(n-i) + \bar{M} + 3\bar{L}_g \bar{h}) \frac{(2\bar{L}_g)^{n-i}}{2\bar{L}_g - 1} + \bar{f} + \bar{h} \right] \left\| Z_i^\alpha - \hat{Z}_i^\alpha \right\|_1, \end{aligned}$$

and the required result is proved by using the Cauchy–Schwarz inequality on $\left\| Z_i^\alpha - \hat{Z}_i^\alpha \right\|_1$. $\square$

The term $\left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_2$ represents the discretization error at time k and is bounded with the following estimation.

LEMMA B.6. *Assume **H2** holds. Then, for each time step $k = 0, \dots, n$, and $\alpha \in \mathcal{A}$, we have*

$$\left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_2 \leq \Psi^k |v_0 - \text{Proj}_{\Gamma^v}(v_0)| + \sum_{i=0}^k \Delta_{k-i} \Psi^i + \left\{ v + \frac{C_2}{R} \right\} \sum_{i=0}^k \Psi^i, \quad (\text{B.13})$$

where $\Psi := [H]_{\text{Lip}}(2d+1)$, C_2 is the maximum value of H over $\Gamma \times A \times \cup_k \Gamma_k \times \cup_k \Gamma_k$, and Δ_k is the L^2 quantization error at the time step k :

$$\Delta_k = \left\| (\hat{\Pi}_k, \hat{Y}_k) - (\Pi_k, Y_k) \right\|_2.$$

PROOF. By Minkowski’s inequality, we have

$$\left\| Z_k^\alpha - \hat{Z}_k^\alpha \right\|_2 \leq \left\| V_k^\alpha - \hat{V}_k^\alpha \right\|_2 + \Delta_k. \quad (\text{B.14})$$

Recalling the dynamics (2.1) and (4.3), we have

$$\left\| V_k^\alpha - \hat{V}_k^\alpha \right\|_2 \leq \left\| H_k^\alpha - \hat{H}_k^\alpha \right\|_2 + \left\| \hat{H}_k^\alpha - \text{Proj}_\Gamma \hat{H}_k^\alpha \right\|_2, \quad (\text{B.15})$$

where $H_k^\alpha := H(V_{k-1}^\alpha, \alpha_{k-1}, Y_{k-1}, Y_k)$ and $\hat{H}_k^\alpha := H(\hat{V}_{k-1}^\alpha, \alpha_{k-1}, \hat{Y}_{k-1}, \hat{Y}_k)$. Under **H2**, and by using Minkowski's and Cauchy–Schwarz' inequalities, we get

$$\begin{aligned} \|H_k^\alpha - \hat{H}_k^\alpha\|_2 &\leq [H]_{\text{Lip}}(2d + q) \left(\|\hat{V}_{k-1}^\alpha - V_{k-1}^\alpha\|_2 + \|\hat{Y}_{k-1} - Y_{k-1}\|_2 + \|\hat{Y}_k - Y_k\|_2 \right) \\ &\leq \Psi \left(\|\hat{V}_{k-1}^\alpha - V_{k-1}^\alpha\|_2 + \Delta_{k-1} + \Delta_k \right), \end{aligned}$$

and so by (B.15)

$$\begin{aligned} \|V_k^\alpha - \hat{V}_k^\alpha\|_2 + \Delta_k &\leq \Psi \left(\|\hat{V}_{k-1}^\alpha - V_{k-1}^\alpha\|_2 + \Delta_{k-1} \right) \\ &\quad + (\Psi + 1)\Delta_k + \|\hat{H}_k^\alpha - \text{Proj}_\Gamma \hat{H}_k^\alpha\|_2. \end{aligned}$$

Hence, a direct backward induction yields

$$\begin{aligned} \|V_k^\alpha - \hat{V}_k^\alpha\|_2 + \Delta_k &\leq \Psi^k |v_0 - \text{Proj}_{\Gamma^v}(v_0)| + \sum_{i=0}^k \Delta_{k-i} \Psi^i \\ &\quad + \sum_{i=0}^k \Psi^i \|\hat{H}_{k-i}^\alpha - \text{Proj}_\Gamma \hat{H}_{k-i}^\alpha\|_2. \end{aligned} \tag{B.16}$$

By noting that $|v - \text{Proj}_{\Gamma^v}(v)| \leq \max(|v| - R, 0) + v$, for all $v \in \mathbb{R}$, we have

$$\|\hat{H}_{k-i}^\alpha - \text{Proj}_\Gamma \hat{H}_{k-i}^\alpha\|_2 \leq v + \|\hat{H}_{k-i}^\alpha \mathbf{1}_{\{\hat{H}_{k-i}^\alpha \geq R\}}\|_2 \tag{B.17}$$

$$\begin{aligned} &\leq v + \frac{1}{R} \|\hat{H}_{k-i}^\alpha\|_2 \\ &\leq v + \frac{1}{R} C_2, \end{aligned} \tag{B.18}$$

where we used Markov inequality. The requested result is proved by plugging (B.18) and (B.16) into (B.14). $\square$

Proof of Theorem 4.1

This follows directly from the estimations (B.9) and (B.13) for $k = 0$ and from the relations (A.1) and (B.8).

References

- BENSOUSSAN, A. (1992). *Stochastic Control of Partially Observable Systems* (Cambridge University Press, New York).
- BENSOUSSAN, A., RUNGALDIER, W.J. (1987). An approximation method for stochastic control problems with partial observation of the state—a method for constructing ϵ -optimal controls. *Acta. Appl. Math.* **10**, 145–170.
- BERTSEKAS, D. (1992). *Dynamic Programming and Stochastic Control* (Academic Press, New York).
- BERTSEKAS, D., SHREVE, S. (1996). *Stochastic Optimal Control: The Discrete-Time Case* (Athena Scientific, Belmont, MA).
- FÖLLMER, H., LEUKERT, P. (2000). Efficient hedging: Cost versus shortfall risk. *Financ Stoch* **4**, 117–146.
- FÖLLMER, H., SONDERMANN, D. (1986). Hedging of non redundant contingent claims. In: Hildebrand, W., Mas-Colell, A. (eds.), *Contributions to Mathematical Economics in Honor of Gérard Debreu* (North-Holland), pp. 205–233.
- GRAF, S., LUSCHGY, H. (2000). *Foundations of Quantization for Random Vectors. Lecture Notes in Mathematics* (Springer, Berlin).
- KUSHNER, H., DUPUIS, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, second ed. (Springer Verlag).
- LIPSTER, R.S., SHIRYAEV, A.N. (1977). *Statistics of Random Processes: I. General Theory* (Springer Verlag, Berlin).
- LUENBERGER, D. (1984). *Linear and Nonlinear Programming* (AddisonWesley).
- PAGÈS, G., PHAM, H. (2005). Optimal quantization methods for nonlinear filtering with discrete time observations. *Bernoulli* **11**, 893–932.
- PAGÈS, G., PHAM, H., PRINTEMS, J. (2003). Optimal quantization methods and applications to numerical problems in finance. In: Rachev, Z. (ed.), *Handbook of Numerical Methods in Finance* (Birkhauser).
- PAGÈS, G., PHAM, H., PRINTEMS, J. (2004). An optimal Markovian quantization algorithm for multidimensional stochastic control problems. *Stoch. Dynam.* **4**, 501–502.
- PHAM, H., RUNGALDIER, W.J., SELLAMI, A. (2004). Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods Appl.* **11**, 57–82.
- RUNGALDIER, W.J. (1991). On the construction of ϵ -optimal strategies in partially observed mdps. *Ann. Oper. Res.* **28**, 81–96.

Recombining Binomial Tree Approximations for Diffusions

John van der Hoek

*School of Mathematics and Statistics, University of South Australia,
GPO Box 2471, Adelaide, South Australia 5001, Australia
E-mail address: john.vanderhoek@unisa.edu.au*

Abstract

In this chapter, we present a novel way to approximate a diffusion by a recombining binomial tree model. The method is obtained by approximating a procedure to find a weak solution of a stochastic differential equation. We shall indicate some theory that provides analysis that the method does indeed provide an approximation. If the original diffusion is expressed in risk neutral terms, then the binomial tree can be used to approximate the value of a wide number of financial derivatives, and if the original diffusion is expressed in real-world probabilities, then the tree could be used to provide approximate simulations that could be used in risk analysis. We present a list of examples of one-dimensional diffusions and an illustrative two-dimensional example.

1. The methodology

We shall first provide a recombining binomial tree model approximation for the solution of the stochastic differential equation:

$$dS(t) = \mu(t, S(t))dt + \sigma(t, S(t))dB(t) \quad (1.1)$$

$$S(0) = S \quad (1.2)$$

on some probability space $(\Omega, \mathcal{F}, P)$, where B is a standard one-dimensional Brownian motion. We will study these equations over a time interval $[0, T]$.

1.1. The weak solution

We construct a weak solution to (1.1) and (1.2) as follows:

Step 1

On a probability space $(\Omega, \mathcal{F}, \bar{P})$, let $\bar{B}$, be a standard one-dimensional Brownian motion and suppose that

$$S(t) = \phi(t, \bar{B}(t)), \quad (1.3)$$

where ϕ solves the differential equation

$$\frac{\partial \phi}{\partial z}(t, z) = \sigma(t, \phi(t, z)) \quad (1.4)$$

$$\phi(0, 0) = S, \quad (1.5)$$

then

$$dS(t) = \bar{m}(t, \bar{B}(t))dt + \sigma(t, S(t))d\bar{B}(t), \quad (1.6)$$

where

$$\bar{m}(t, \bar{B}(t)) = \frac{\partial \phi}{\partial t}(t, \bar{B}(t)) + \frac{1}{2}\sigma(t, \phi(t, \bar{B}(t)))\frac{\partial \sigma}{\partial z}(t, \phi(t, \bar{B}(t))). \quad (1.7)$$

These statements follow from Itô's lemma provided that a solution of (1.4) and (1.5) is smooth enough. These conditions can be checked in any application.

Step 2

We now make a change of probabilities to adjust the drift in (1.6) to coincide with that in (1.1). This can be achieved by setting

$$\left. \frac{dP}{d\bar{P}} \right|_{\mathcal{F}_T} = \Lambda_T = \exp \left[\int_0^T \psi(u) d\bar{B}(u) - \frac{1}{2} \int_0^T \psi(u)^2 du \right] \quad (1.8)$$

for suitable ψ satisfying the Novikov condition, say, when

$$B(t) = \bar{B}(t) - \int_0^t \psi(u) du \quad (1.9)$$

is the standard one-dimensional Brownian motion under P , and $\{\mathcal{F}_t\}$ is the filtration generated by $\bar{B}$.

Under P , Eq. (1.6) becomes

$$dS(t) = \bar{m}(t, \bar{B}(t))dt + \sigma(t, S(t)) [dB(t) + \psi(t)dt], \quad (1.10)$$

and we choose ψ so that

$$\mu(t, S(t)) = \bar{m}(t, \bar{B}(t)) + \sigma(t, S(t))\psi(t) \quad (1.11)$$

or

$$\psi(t) = \frac{\mu(t, \phi(t, \bar{B}(t))) - m(t, \bar{B}(t))}{\sigma(t, \phi(t, \bar{B}(t)))} \equiv \Psi(t, \bar{B}(t)). \quad (1.12)$$

Thus under P , S given in (1.3) provides a (weak) solution of Eq. (1.1).

1.2. The approximations

Let N be a positive integer and let $\Delta t = \frac{T}{N}$. We then define

$$S(0, 0) = S \quad (1.13)$$

$$S(n, j) = \phi(n\Delta t, (2j - n)\sqrt{\Delta t}) \quad (1.14)$$

for $j = 0, 1, \dots, n$ and $n = 0, 1, \dots, N$.

From (n, j) (time n and state j), we can move to either $(n + 1, j + 1)$ or $(n + 1, j)$. In this way, we obtain a recombining binomial tree of values for S . If $(n, j) \rightarrow (n + 1, j + 1)$ and $(n, j) \rightarrow (n + 1, j)$ occur with equal probability, then S in (1.13) and (1.14) provides a numerical approximation to Eq. (1.6). For this, we refer NELSON and RAMASWAMY [1990].

We now assign new probabilities $p(n, j)$ to $(n, j) \rightarrow (n + 1, j + 1)$ and $1 - p(n, j)$ to $(n, j) \rightarrow (n + 1, j)$ so that S in (1.13) and (1.14) provides a numerical approximation to Eq. (1.1). We now motivate the formulas for the $p(n, j)$. The details of the convergence are again provided by NELSON and RAMASWAMY [1990].

Let us set for $t < s$

$$\Lambda_{t,s} = \exp \left[\int_t^s \psi(u) d\bar{B}(u) - \frac{1}{2} \int_t^s \psi(u)^2 du \right] \quad (1.15)$$

and let X be $\mathcal{F}_s$ measurable. Then, using E for expectations under P and $\bar{E}$ for expectations under $\bar{P}$

$$\begin{aligned} E[X|\mathcal{F}_t] &= \frac{\bar{E}[\Lambda_{0,T}X|\mathcal{F}_t]}{\bar{E}[\Lambda_{0,T}|\mathcal{F}_t]} \\ &= \frac{\bar{E}[\Lambda_{t,s}X|\mathcal{F}_t]}{\bar{E}[\Lambda_{t,s}|\mathcal{F}_t]}. \end{aligned}$$

One now applies this calculation with $s = t + \Delta t$ and with $X = X_+ = \mathbf{I}[\bar{B}(t + \Delta t) - \bar{B}(t) = \sqrt{\Delta t}]$ and $X = X_- = \mathbf{I}[\bar{B}(t + \Delta t) - \bar{B}(t) = -\sqrt{\Delta t}]$.

Of course in this we are using the approximation

$$\bar{B}(t + \Delta t) - \bar{B}(t) = \pm\sqrt{\Delta t}$$

with equal probabilities under $\bar{P}$. We are led to the approximation:

$$\begin{aligned}
 E[X_+ | \mathcal{F}_t] &\approx \frac{\frac{1}{2} \exp[\psi(t)\sqrt{\Delta t} - \frac{1}{2}\psi(t)^2\Delta t]}{\frac{1}{2} \exp[\psi(t)\sqrt{\Delta t} - \frac{1}{2}\psi(t)^2\Delta t] + \frac{1}{2} \exp[-\psi(t)\sqrt{\Delta t} - \frac{1}{2}\psi(t)^2\Delta t]} \\
 &= \frac{\frac{1}{2} \exp[\psi(t)\sqrt{\Delta t}]}{\frac{1}{2} \exp[\psi(t)\sqrt{\Delta t}] + \frac{1}{2} \exp[-\psi(t)\sqrt{\Delta t}]} \\
 &= \frac{1}{2} + \frac{1}{2} \tanh[\psi(t)\sqrt{\Delta t}]
 \end{aligned} \tag{1.16}$$

and likewise

$$E[X_- | \mathcal{F}_t] \approx \frac{1}{2} - \frac{1}{2} \tanh[\psi(t)\sqrt{\Delta t}]. \tag{1.17}$$

These heuristic calculations lead to our choices for the $p(n, j)$. We use (1.12) to set

$$p(n, j) = \frac{1}{2} + \frac{1}{2} \tanh[\Psi(n\Delta t, (2j - n)\sqrt{\Delta t})\sqrt{\Delta t}]. \tag{1.18}$$

When $N \rightarrow \infty$, the results of NELSON and RAMASWAMY [1990] show that S in (1.13) and (1.14) converges to a solution to Eq. (1.1).

We can also apply this analysis to a system of d stochastic differential equations driven by d -dimensional Brownian motion using analogous arguments.

We now proceed to illustrations of this approach.

2. One-dimensional examples

EXAMPLE 2.1 (the Black and Scholes equation).

We have $\mu(t, x) = \mu x$ and $\sigma(t, x) = \sigma x$. Then, $\phi(t, z) = S \exp(\sigma z)$, and we set

$$\begin{aligned}
 S(n, j) &= S \exp[(2j - n)\sigma\sqrt{\Delta t}] \\
 &= Su^j d^{n-j},
 \end{aligned}$$

where

$$\begin{aligned}
 u &= \exp[\sigma\sqrt{\Delta t}] \\
 d &= \exp[-\sigma\sqrt{\Delta t}] \\
 \psi(t) &= \frac{\mu}{\sigma} - \frac{1}{2}\sigma
 \end{aligned}$$

and

$$p(n, j) = \frac{1}{2} + \frac{1}{2} \tanh\left[\left(\frac{\mu}{\sigma} - \frac{1}{2}\sigma\right)\sqrt{\Delta t}\right].$$

EXAMPLE 2.2. We have $\mu(t, x) = a - bx$ and $\sigma(t, x) = \sigma x$. Then, ϕ and $S(n, j)$ are as in Example 1. But

$$\begin{aligned}\psi(t) &= \frac{a}{\sigma S(t)} - \frac{b}{\sigma} - \frac{1}{2}\sigma \\ p(n, j) &= \frac{1}{2} + \frac{1}{2} \tanh \left\{ \left[\frac{a}{\sigma S} \exp[(n-2j)\sigma\sqrt{\Delta t}] - \frac{b}{\sigma} - \frac{1}{2}\sigma \right] \sqrt{\Delta t} \right\}\end{aligned}$$

EXAMPLE 2.3 (The CIR equation (COX, INGERSOL and ROSS [1985])).

We have $\mu(t, x) = a - bx$ and $\sigma(t, x) = \sigma\sqrt{x}$. Then,

$$\begin{aligned}\phi(t, z) &\equiv \phi(z) = \begin{cases} [\sqrt{S} + \frac{1}{2}\sigma z]^2 & \text{if } \sqrt{S} + \frac{1}{2}\sigma z \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ S(n, j) &= \phi((2j-n)\sqrt{\Delta t}) \\ \psi(t) &= \left(\frac{a}{\sigma} - \frac{\sigma}{2} \right) \frac{1}{\sqrt{S(t)}} - \frac{b}{\sigma} \sqrt{S(t)} \\ p(n, j) &= \frac{1}{2} + \frac{1}{2} \tanh \left\{ \left[\left(\frac{a}{\sigma} - \frac{\sigma}{2} \right) \frac{1}{\sqrt{S(n, j)}} - \frac{b}{\sigma} \sqrt{S(n, j)} \right] \sqrt{\Delta t} \right\}\end{aligned}$$

and we note that as $S(n, j) \rightarrow 0+$

$$p(n, j) \rightarrow \begin{cases} 1 & \text{if } a > \frac{1}{2}\sigma^2 \\ 0 & \text{if } a < \frac{1}{2}\sigma^2, \end{cases} \quad (2.1)$$

which is why we often assume the second case ($\sigma^2 < 2a$) when S models an interest rate.

The model with $\sigma(t, x) = \sigma x^\beta$ with $0 < \beta < 1$ is treated in a similar way.

EXAMPLE 2.4 (the Ornstein–Uhlenbeck process).

Here we have $\mu(t, x) = \beta(a - x)$ and $\sigma(t, x) = \sigma$. Then,

$$\begin{aligned}\phi(t, z) &\equiv \phi(z) = S + \sigma z \\ \psi(t) &= \frac{\beta(a - S(t))}{\sigma} \\ S(n, j) &= S + \sigma(2j - n)\sqrt{\Delta t} \\ p(n, j) &= \frac{1}{2} + \frac{1}{2} \tanh \left[\frac{\beta(a - S(n, j))}{\sigma} \sqrt{\Delta t} \right],\end{aligned}$$

and we note that

$$p(n, j) \text{ is } \begin{cases} < \frac{1}{2} & \text{if } S(n, j) > a \\ > \frac{1}{2} & \text{if } S(n, j) < a, \end{cases}$$

which supports the mean-reverting property.

The Vasicek interest model uses this process (VASICEK [1977]).

3. A two-dimensional example

We present a result that can easily be derived in a similar way to the one-dimensional case.

EXAMPLE 3.1 (Schwartz and Smith model (SCHWARTZ and SMITH [2000])).

Using a notation similar to this paper, we have

$$\begin{aligned} d\xi(t) &= (\mu_\xi - \lambda_\xi)dt + \sigma_\xi dB_\xi(t) \\ d\chi(t) &= (-\kappa\chi(t) - \lambda_\chi)dt + \sigma_\chi dB_\chi(t), \end{aligned}$$

where $\mu_\xi, \lambda_\xi, \sigma_\xi, \kappa, \lambda_\chi$, and σ_χ are constants and $dB_\xi(t)dB_\chi(t) = \varrho dt$. We write

$$B_\chi(t) = \rho B_\xi(t) + \sqrt{1 - \rho^2} B_\xi^*(t),$$

where B_ξ and B_ξ^* are independent Brownian motions.

$$\begin{aligned} \xi(n, j, k) &= \xi(0) + \sigma_\xi(2j - n)\sqrt{\Delta t} \\ \chi(n, j, k) &= \chi(0) + \sigma_\chi \left[\varrho(2j - n)\sqrt{\Delta t} + \sqrt{1 - \varrho^2}(2k - n)\sqrt{\Delta t} \right] \end{aligned}$$

and now for each node (n, j, k) , we must calculate four probabilities:

$$\begin{aligned} p_1(n, j, k) &\text{ for } (n, j, k) \rightarrow (n + 1, j + 1, k + 1) \\ p_2(n, j, k) &\text{ for } (n, j, k) \rightarrow (n + 1, j + 1, k) \\ p_3(n, j, k) &\text{ for } (n, j, k) \rightarrow (n + 1, j, k + 1) \\ p_4(n, j, k) &\text{ for } (n, j, k) \rightarrow (n + 1, j, k). \end{aligned}$$

Setting

$$\begin{aligned} \psi_1(n, j, k) &= \frac{\mu_\xi - \lambda_\xi}{\sigma_\xi} \\ \psi_2(n, j, k) &= \frac{-\kappa\chi(n, j, k) - \lambda_\chi - \rho\sigma_\chi\psi_1(n, j, k)}{\sqrt{1 - \rho^2}\sigma_\chi} \\ \tau_1(n, j, k) &= \tanh \left[\psi_1(n, j, k)\sqrt{\Delta t} \right] \\ \tau_2(n, j, k) &= \tanh \left[\psi_1(n, j, k)\sqrt{\Delta t} \right] \end{aligned}$$

we use

$$p_1(n, j, k) = \frac{1}{4}(1 + \tau_1(n, j, k))(1 + \tau_2(n, j, k))$$

$$p_2(n, j, k) = \frac{1}{4}(1 + \tau_1(n, j, k))(1 - \tau_2(n, j, k))$$

$$p_3(n, j, k) = \frac{1}{4}(1 - \tau_1(n, j, k))(1 + \tau_2(n, j, k))$$

$$p_4(n, j, k) = \frac{1}{4}(1 - \tau_1(n, j, k))(1 - \tau_2(n, j, k)).$$

It is automatic in this construction that $p_i(n, j, k) \geq 0$ for each $i = 1, 2, 3, 4$ and the probabilities sum to 1. This could suggest that this algorithm is an improvement over another algorithm provided for this example by HAHN and DYER [2004].

References

- COX, J.C., INGERSOL, J.E., ROSS, R.A. (1985). An equilibrium characterization theory of the term structure. *Econometrica* **53**, 385–407.
- HAHN, W.J., DYER, J.S. (2004). *A Discrete-Time Approach for Valuing Real Options With Underlying Mean-Reverting Stochastic Processes* (McCombs School of Business, The University of Texas at Austin).
- NELSON, D.B., RAMASWAMY, K. (1990). Simple binomial processes and diffusion approximations in financial models. *Rev. Fin. Stud.* **3**, 393–430.
- SCHWARTZ, E., SMITH, J.E. (2000). Short-term variations and long-term dynamics in commodity prices. *Manage. Sci.* **46**, 893–911.
- VASICEK, O. (1977). A theory of the term structure of interest rates. *J. Financ. Econ.* **5**, 177–188.

Partial Differential Equations for Option Pricing

Olivier Pironneau

*Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, Boîte courrier 187,
75252 Paris Cedex 05. France
E-mail address: pironneau@ann.jussieu.fr*

Yves Achdou

*UFR Mathématiques, Université Paris 7, Case 7012, 75251 Paris Cedex 05, France and
Laboratoire Jacques-Louis Lions, Université Paris 6, France
E-mail address: achdou@math.jussieu.fr*

Contents

CHAPTER I	377
1. The partial differential equation	377
2. A finite-element method	388
3. Mesh adaptivity	394
CHAPTER II	403
4. European basket options	403
5. Numerical methods for European basket options	414
6. American basket options	437
7. Stochastic volatility	449
CHAPTER III	471
8. Sensitivity	471
9. Calibration	475

Introduction

Option pricing is one of the many problems of financial mathematics or financial engineering as it is called now. It all started in the seventies with the celebrated model of BLACK and SCHOLES [1973], MERTON [1973] and their Nobel consecration later. Some of these ideas were already in the thesis of BACHELIER [1995] in 1900, but everyone forgot because the era of fast electronic transactions had not come.

Today, financial assets (stocks, bonds, commodities etc) are used as a base for thousands of more complex financial products known as financial derivatives. The simplest example may be the European put option on a given asset that enables its holder to sell the asset at a future time T , the maturity, for a price K , the strike. If the asset is worth S_T at time T , the option will be exercised only if $K > S_T$, generating a profit $K - S_T$. In the other case, the option will not be exercised, and the profit will be 0. Therefore, the profit generated by the option at T will be $(K - S_T)^+$. Assuming that the market is liquid and arbitrage is not possible (one cannot make an instantaneous benefit without taking a risk), the price of the option at T will be $(K - S_T)^+$. Option pricing at time $t < T$ is more difficult because S_T is not known. The Black–Scholes model makes the above assumptions and supposes furthermore that the market is made of two assets, the previously mentioned risky asset and a riskless asset whose price evolves with a known interest rate r . It allows for pricing the above-mentioned put option as the expectation of $(K - S_T)^+$ discounted at the interest rate r . Another assumption of the Black–Scholes model is that $S_{t+\delta t}$ evolves from S_t with a mean tendency μ and a random fluctuation of intensity σ , the volatility:

$$S_{t+\delta t} = S_t(1 + \mu\delta t) + \sigma S_t N(0, \delta t),$$

where $N(0, v)$ is a normal distribution with mean zero and variance v , and that $S_{t+\delta t} - S_t$ is independent of the events before t . A very simple set of ideas indeed! But as the model should not depend on the time increment δt , one should use continuous-time stochastic processes; therefore, the European put P_t is priced by

$$P_t = e^{-r(T-t)} \mathbb{E}(K - S_T)^+, \quad \text{with} \quad dS_t = S_t(\mu dt + \sigma dB_t), \quad (0.1)$$

where B_t is a Brownian motion. Since S_t is a Markov process, there exists a two-variable function P , called the pricing function, such that $P_t = P(S_t, t)$, and P solves the partial

differential equation (PDE):

$$\frac{\partial P}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} - rP = 0, \quad (0.2)$$

for $t \in [0, T)$ and $S > 0$.

There are three important classes of numerical methods in financial engineering: Monte Carlo methods, tree-based methods, and deterministic methods based on the PDE (0.2). The goal of the chapter is to focus on the latter.

Of course, it may seem that numerical methods for (0.2) are now well known, but there are special difficulties in finance because traders require a quick and accurate response. Furthermore, there are much more complicated contracts than the one described above, and the PDE may become much more complex, for example, American option pricing involves variational inequalities, while stochastic volatility models lead to multidimensional PDEs; for basket options, the problem can become numerically formidable because set in a space whose dimension is the number of assets in the basket; finally, models involving more general Lévy processes lead to partial integrodifferential equations [PIDE], CONT and TANKOV [2003].

The diversity of the models for financial derivatives has grown to such a point that it is not possible to discuss all of them in one book chapter. Here, only local volatility (the volatility may depend on time and on the price of the underlying asset) and stochastic volatility models will be considered, and models based on jump process will just be briefly described. Similarly, the types of contracts on the markets are too numerous, and we will mostly deal with European, American options with or without barriers, possibly multidimensional (basket options, for example).

Our objective is numerical: what is a good method for the computer implementations?

One may choose between finite-difference methods, RICHTMYER and MORTON [1994], finite-element methods, CIARLET [1978, 1991], ZIENKIEWICZ and TAYLOR [2000], finite-volume methods, EYMARD, GALLOUËT and HERBIN [2000], spectral methods, BERNARDI and MADAY [1997], QUARTERONI [1991] etc. We have decided to work with the finite-element method (FEM) because it is very flexible on the one hand and supported by a strong theory on the other hand. In financial engineering, the PDEs often have a parabolic character (the first-order terms in the equations are usually not dominant), which makes FEMs well adapted.

In this restricted context, what is the best way to implement the methods, namely, what polynomial degree, what mesh, what linear system solver etc? As usual, the answer goes through a mathematical analysis of the variational formulation of the problem, in which one can prove existence, uniqueness, and qualitative properties of the solution. Then, error estimates, especially a posteriori estimates are most useful, and the FEM is best suited for such analysis. The outcome of such a study is that one can guarantee the precision of the calculations, a property appreciated in banking where, in view of the large sums involved, an error greater than 0.1% is often unacceptable.

For clarity, we present the material by increasing order of mathematical complexity. The first chapter deals with plain vanilla European options. The variational formulation

is given and studied, and the posteriori estimates derived by ACHDOU and PIRONNEAU [2005] are restated. A C++ implementation on an arbitrary mesh is given at the end.

The Second chapter deals with higher dimensional models. We consider in particular European and American options on baskets and stochastic volatility models. In this context, we discuss the variational analysis of the boundary value problems. When the dimension of the problem is rather small (≤ 4), FEMs are competitive; we describe several techniques concerning the solution procedure, in particular for American options. We also review on a promising and new class of methods, which may be used for parabolic PDEs when the dimension of the problem lies between 4 and 20, the sparse grid and sparse Galerkin methods.

In the third chapter, we recall the method given by ACHDOU and PIRONNEAU [2005] for computing the sensitivity of the solution with respect to the parameters of the problem, the *Greeks* (called so because practitioners have used Greek letters). The method is based on automatic differentiation of computer programs, GRIEWANK [2000], HASCOET and PASCUAL [2004] a very powerful technique particularly appropriate to financial engineering. The operator overloading feature of the C++ language makes it easy to implement this approach. It is also useful in the context of parameters calibration, where the gradients of the least square functionals with respect to the model parameters are needed. This brings us to another important topic of financial engineering: better models or better parameters? As an example of calibration, we consider the calibration of local volatility. We discuss Dupire's equation DUPIRE [1997] and the use of least squares methods for calibration.

One-Dimensional Partial Differential Equations For Option Pricing

1. The partial differential equation

A *European vanilla call* (respectively, put) option is a contract giving its owner the right to buy (respectively, sell) a share of a specific common stock at a fixed price K at a certain date T . The specific stock is called the underlying asset. The fixed price K is termed *the strike* and T is called *the maturity*. The term *vanilla* is used to notify that this kind of option is the simplest among possibly complicated contracts. The price of the underlying asset at time t will be referred to as the *spot price* and will be noted S_t . Assuming that the market rules out arbitrage (the possibility to make an instantaneous risk-free benefit), it is easy to see that the price of a call (respectively, put) option at maturity is $C_0(S_T) = (S_T - K)^+$, (respectively, $P_0(S_T) = (K - S_T)^+$). The payoff of the option at maturity is a function of S_T , called the *payoff function*. Naturally, other payoff functions than the ones mentioned above are possible and used in practice.

In order to price the option before maturity, some assumptions have to be made on the spot price S_t : the Black–Scholes model assumes the existence of a risk-free asset whose price at time t is $S_t^0 = S_0^0 \exp(\int_0^t r(s)ds)$, where $r(t)$ is the interest rate; the model assumes that the price of the risky asset satisfies the following stochastic differential equation

$$dS_t = S_t(\mu dt + \sigma_t dB_t), \quad (1.1)$$

where B_t is a standard Brownian motion on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Here, σ_t is a positive number called the *volatility*. With the Black–Scholes assumptions, it is possible to prove that the option's price at time t is given by

$$P_t = \exp\left(\int_t^T r(s)ds\right) \mathbb{E}^*(P_0(S_T)|F_t), \quad (1.2)$$

where the expectation $\mathbb{E}^*$ is taken with respect to the so-called risk-neutral probability $\mathbb{P}^*$ (equivalent to $\mathbb{P}$ and under which $dS_t = S_t(rdt + \sigma_t dW_t)$, W_t being a standard Brownian motion under $\mathbb{P}^*$ and F_t being the natural filtration of W_t).

From (1.2) and since S_t is a Markov process, it can be shown that the option's price P_t is a function of t and S_t , that is, that there exists a two-variable function P , called *the pricing function* such that $P_t = P(S_t, t)$.

Assuming that $\sigma_t = \sigma(S_t, t)$, where σ is a smooth-enough function, it can be seen that the pricing function P solves the backward in time parabolic PDE:

$$\frac{\partial P}{\partial t} + \frac{\sigma^2(S, t)S^2}{2} \frac{\partial^2 P}{\partial S^2} + r(t)S \frac{\partial P}{\partial S} - r(t)P = 0 \quad (1.3)$$

for $t \in [0, T)$ and $S > 0$ and satisfies the final time condition

$$P(S, t = T) = P_0(S) \quad (1.4)$$

for $S > 0$. Problems (1.3), and (1.4) are called final value problem.

The volatility is the difficult parameter of the Black–Scholes model. It is convenient to take it to be constant, but then the computed options' prices do not match the prices given by the market. There are essentially three ways for improving the Black–Scholes model with a constant volatility:

- Use a local volatility, that is, assume that the volatility is a function of time and the stock price. Then, one has to calibrate the volatility from the market data, that is, to find a volatility function that permits to recover the prices of the options available on the market.
- assume that the volatility is itself a stochastic process (see FOUQUE, PAPANICOLAOU and SIRCAR [2000], HESTON [1993] and §7).
- generalize the Black–Scholes model by assuming that the spot price is, for example, a Lévy process; (see CONT and TANKOV [2003] and references therein).

There is much discussion among specialists in finance on comparing the merits of the three kinds of models above. In the following paragraph, we will focus on the first one.

1.1. Changes of variables

Several changes of variables and unknown functions can be used.

Step 1

Consider the function v such that $P(S, t) = v(S, t)e^{-\lambda(t)}$, then (1.3) can be written as

$$\frac{\partial P}{\partial t} = -\lambda'(t)e^{-\lambda(t)}v + e^{-\lambda(t)}\frac{\partial v}{\partial t}, \quad \frac{\partial P}{\partial S} = e^{-\lambda(t)}\frac{\partial v}{\partial S}, \quad \frac{\partial^2 P}{\partial S^2} = e^{-\lambda(t)}\frac{\partial^2 v}{\partial S^2}.$$

Choosing $\lambda(t) = -\int_t^T r(s)ds$ leads to

$$\frac{\partial v}{\partial t} + rS \frac{\partial v}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 v}{\partial S^2} = 0.$$

Step 2

Now set $x = \log S$, and check that $\frac{\partial v}{\partial S} = \frac{1}{S} \frac{\partial v}{\partial x}$ and $\frac{\partial^2 v}{\partial S^2} = -\frac{1}{S^2} \frac{\partial v}{\partial x} + \frac{1}{S^2} \frac{\partial^2 v}{\partial x^2}$. We also set $\tau = T - t$ and $w(x, \tau) = v(e^x, T - \tau)$. Calling $\tilde{r}$ and $\tilde{\sigma}$ the functions defined by $\tilde{r}(\tau) = r(t)$ and $\tilde{\sigma}(x, \tau) = \sigma(e^x, t)$, we have

$$\frac{\partial w}{\partial \tau} - \frac{\tilde{\sigma}^2(x, \tau)}{2} \frac{\partial^2 w}{\partial x^2} + \left(\tilde{r}(\tau) - \frac{\tilde{\sigma}^2(x, \tau)}{2} \right) \frac{\partial w}{\partial x} = 0 \quad \text{in } \mathbb{R} \times (0, T). \quad (1.5)$$

Step 3

When σ depends on t only, one may use the change of variable $(x, \tau) \mapsto (y, \tau)$, where $y = x - \int_0^\tau (\tilde{r}(\theta) - \frac{\tilde{\sigma}^2(\theta)}{2}) d\theta$ and set $W(y, \tau) = w(x, \tau)$; it is easy to see that

$$\frac{\partial W}{\partial \tau}(y, \tau) - \frac{\tilde{\sigma}^2(\tau)}{2} \frac{\partial^2 W}{\partial y^2}(y, \tau) = 0, \quad \text{in } \mathbb{R} \times (0, T),$$

and that $W(y, 0) = w(y, 0)$. When σ is a positive constant, this equation is the heat equation.

A similar idea can be used if $x \mapsto \tilde{\sigma}^2(x, \tau)$ is Lipschitz continuous uniformly with respect to τ : we call $X(\theta; x, \tau)$ the solution of the ordinary differential equation

$$\frac{d}{d\theta} X(\theta; x, \tau) = \left(\tilde{r}(\theta) - \frac{\tilde{\sigma}^2(X(\theta; x, \tau), \theta)}{2} \right) \quad \theta \in (0, T), \quad X(\tau; x, \tau) = x.$$

Assuming that $(x, \theta) \mapsto X(\theta; x, \tau)$ is regular enough and introducing $W(x, \theta) = w(X(\theta; x, \tau), \theta)$, we obtain

$$\frac{\partial W}{\partial \theta}(x, \theta) = \frac{\partial w}{\partial t}(X(\theta; x, \tau), \theta) + \left(\tilde{r}(\theta) - \frac{\tilde{\sigma}^2(X(\theta; x, \tau), \theta)}{2} \right) \frac{\partial w}{\partial x}(X(\theta; x, \tau), \theta)$$

and

$$\begin{aligned} \frac{\partial W}{\partial x}(x, \theta) &= \frac{\partial w}{\partial x}(X(\theta; x, \tau), \theta) \frac{\partial X(\theta; x, \tau)}{\partial x}, \\ \frac{\partial^2 W}{\partial x^2}(x, \theta) &= \frac{\partial^2 w}{\partial x^2}(X(\theta; x, \tau), \theta) \left(\frac{\partial X(\theta; x, \tau)}{\partial x} \right)^2 \\ &\quad + \frac{\partial w}{\partial x}(X(\theta; x, \tau), \theta) \frac{\partial^2 X(\theta; x, \tau)}{\partial x^2}. \end{aligned}$$

Taking $\theta = \tau - \delta t$ for δt small, we have that $\frac{\partial X(\tau; x, \tau - \delta t)}{\partial x} \sim 1$ and $\frac{\partial^2 X(\tau; x, \tau - \delta t)}{\partial x^2} \sim 0$. Then, using (1.5), we obtain the following semidiscrete scheme:

$$\frac{1}{\delta t} (W(x, \tau) - W(x, \tau - \delta t)) - \frac{\tilde{\sigma}^2(x, \tau)}{2} \frac{\partial^2 W}{\partial x^2}(x, \tau) \sim 0,$$

that is,

$$\frac{1}{\delta t} (w(x, \tau) - w(X(\tau - \delta t; x, \tau), \tau - \delta t)) - \frac{\tilde{\sigma}^2(x, \tau)}{2} \frac{\partial^2 w}{\partial x^2}(x, \tau) \sim 0,$$

which is known as the method of characteristics and often used in fluid mechanics.

1.2. The Black–Scholes formulas

Calling $P(S, t)$ the price of an option with maturity T and payoff function P_0 and assuming that r and $\sigma > 0$ are constant, the Black–Scholes formula is

$$P(S, t) = e^{-r(T-t)} \mathbb{E}^*(P_0(S e^{r(T-t)} e^{\sigma(W_T - W_t) - \frac{\sigma^2}{2}(T-t)})), \quad (1.6)$$

and since under P^* , $W_T - W_t$ is a centered Gaussian distribution with variance $T - t$,

$$P(S, t) = \frac{1}{\sqrt{2\pi}} e^{-r(T-t)} \int_{\mathbb{R}} P_0(S e^{(r - \frac{\sigma^2}{2})(T-t) + \sigma x \sqrt{T-t}}) e^{-\frac{x^2}{2}} dx. \quad (1.7)$$

When the option is a vanilla European option, noting C the price of the call and P the price of the put, a more explicit formula can be deduced from (1.2). For example, take a call:

$$\begin{aligned} C(S, t) &= \frac{1}{\sqrt{2\pi}} \int_{-d_2}^{+\infty} \left(S e^{-\frac{\sigma^2}{2}(T-t) + \sigma x \sqrt{T-t}} - K e^{-r(T-t)} \right) e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_2} \left(S e^{-\frac{\sigma^2}{2}(T-t) - \sigma x \sqrt{T-t}} - K e^{-r(T-t)} \right) e^{-\frac{x^2}{2}} dx, \end{aligned} \quad (1.8)$$

where

$$d_1 = \frac{\log(\frac{S}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma \sqrt{T-t}. \quad (1.9)$$

Finally, introducing the upper tail of the Gaussian function

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{x^2}{2}} dx, \quad (1.10)$$

and using (1.8) and (1.9), we obtain the Black–Scholes formula.

PROPOSITION 1.1. *When σ and r are constant, the price of the call is given by*

$$C(S, t) = S N(d_1) - K e^{-r(T-t)} N(d_2), \quad (1.11)$$

and the price of the put is given by

$$P(S, t) = -S N(-d_1) + K e^{-r(T-t)} N(-d_2), \quad (1.12)$$

where d_1 and d_2 are given by (1.9) and N is given by (1.10).

REMARK 1.1. If r is a function of time, (1.9) must be replaced with

$$d_1 = \frac{\log(\frac{S}{K}) + \int_t^T r(\tau) d\tau + \frac{\sigma^2}{2}(T-t)}{\sigma \sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma \sqrt{T-t}. \quad (1.13)$$

1.3. Classical solutions

In the previous paragraph, we have seen that if the coefficients are constant (with a positive volatility), then (1.3) (1.4) has a solution given by (1.7).

In this paragraph, we give a classical existence and uniqueness result for the final value problem (1.3) (1.4) in the general case when $r = r(t)$ and $\sigma = \sigma(S, t)$. It is necessary to restrict the growth of the solutions when $S \rightarrow 0$ or $S \rightarrow +\infty$. Here, we will impose that the solution is bounded, but this restriction can be relaxed (e.g., depending on P_0 , one can look for solutions with linear growth as $S \rightarrow +\infty$).

DEFINITION 1.1. We fix a positive number ρ_0 . Let α be a real number such that $0 < \alpha < 1$. We call $\mathcal{C}^{0,\alpha}(\mathbb{R}^d)$ the space of continuous real-valued functions $v \in \mathcal{C}^0(\mathbb{R}^d)$ such that

$$\|v\|_{\mathcal{C}^{0,\alpha}(\mathbb{R}^d)} = \sup_{x \in \mathbb{R}^d} |v(x)| + \sup_{x, y \in \mathbb{R}^d, |x-y| \leq \rho_0} \frac{|v(x) - v(y)|}{|x - y|^\alpha} < +\infty.$$

The space $\mathcal{C}^{0,\alpha}(\mathbb{R}^d)$ endowed with the norm $\|\cdot\|_{\mathcal{C}^{0,\alpha}(\mathbb{R}^d)}$ is a Banach space.

We call $\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R}^d \times [0, T])$ the space of continuous real-valued functions $v \in \mathcal{C}^0(\mathbb{R}^d \times [0, T])$ such that

$$\|v\|_{\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R}^d \times [0, T])} = \sup_{(x,t) \in \mathbb{R}^d \times [0, T]} |v(x, t)| + \sup_{(x,t), (y,s) \in \mathbb{R}^d \times [0, T], (|x-y|^2 + |t-s|)^{\frac{\alpha}{2}} \leq \rho_0} \frac{|v(x, t) - v(y, s)|}{(|x - y|^2 + |t - s|)^{\frac{\alpha}{2}}} < +\infty.$$

The space $\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R}^d \times [0, T])$ endowed with the norm $\|\cdot\|_{\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R}^d \times [0, T])}$ is a Banach space.

THEOREM 1.1. Under the following assumptions on the coefficients and the final value,

1. the real-valued function defined on $\mathbb{R} \times [0, T]$, $(x, t) \mapsto \sigma^2(e^x, t)$ belongs to $\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R} \times [0, T])$,
2. the function $t \mapsto r(t)$ belongs to $\mathcal{C}^{\alpha/2}([0, T])$,
3. the function $\tilde{P}_0$ defined on $\mathbb{R}$ by $\tilde{P}_0(x) = P_0(e^x)$ is such that $\tilde{P}_0$, $\frac{\partial \tilde{P}_0}{\partial x}$, and $\frac{\partial^2 \tilde{P}_0}{\partial x^2}$ belong to $\mathcal{C}^\alpha(\mathbb{R})$,
4. there exists a positive constant $\underline{\sigma}$ such that, for all $x \in \mathbb{R}$, $t \in [0, T]$, $\sigma(t, e^x) \geq \underline{\sigma}$,

the final value problem (1.3) (1.4) has a unique solution P such that, calling $\tilde{P}$ the function defined on $[0, T] \times \mathbb{R}$ by $\tilde{P}(x, t) = P(e^x, t)$, the functions $\tilde{P}$, $\frac{\partial \tilde{P}}{\partial x}$, $\frac{\partial \tilde{P}}{\partial t}$, and $\frac{\partial^2 \tilde{P}}{\partial x^2}$ belong to $\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R} \times [0, T])$.

Under the previous assumptions except the one on $\tilde{P}_0$ and assuming that $\tilde{P}_0$ is a bounded function, (1.3) (1.4) has a unique solution $\tilde{P}$ such that for all $\tau < T$, the functions $\tilde{P}$, $\frac{\partial \tilde{P}}{\partial x}$, $\frac{\partial \tilde{P}}{\partial t}$, and $\frac{\partial^2 \tilde{P}}{\partial x^2}$ belong to $\mathcal{C}^{\alpha,\alpha/2}(\mathbb{R} \times [0, \tau])$.

This theorem is proved by LADYŽENSKAJA, SOLONNIKOV and URAL'CEVA [1967] (see also FRIEDMAN [1964], KRYLOV [1996]).

1.4. Variational framework

1.4.1. Weighted Sobolev norms

The theory of variational formulations of parabolic equations is well known (see the work of LIONS [1969]). It is particularly useful when strong solutions do not exist either because of some singularity in the data or the domain boundary, the coefficients, or nonlinearity. Such situations are very frequent in physics and engineering. Even when the boundary value problem has a classical solution, the variational theory is interesting for several reasons:

- it provides global estimates, often called energy estimates.
- it has strong connections with the finite-element method, which will be advocated below.
- it is the most natural way to study obstacle problems (see the section § 6 devoted to American options on baskets).

Note that there are other theories of weak solutions, in particular, the theory of viscosity solutions, which may also be quite useful in the context of quantitative finance. We will not discuss viscosity solutions here, and we refer the reader to BARLES [1994], CRANDALL, ISHII and LIONS [1992], FLEMING and SONER [1993].

Almost all the proofs of the results below are omitted for brevity; they can be found in ACHDOU and PIRONNEAU [2005].

Variational formulations of parabolic PDE rely on suitable Sobolev spaces. We are going to introduce the Sobolev space useful for the initial value problem (1.3) (1.4) posed in the price variable S .

We denote by $L^2(\mathbb{R}_+)$ the Hilbert space of square integrable functions on $\mathbb{R}_+$ endowed with the norm $\|v\|_{L^2(\mathbb{R}_+)} = (\int_{\mathbb{R}_+} v(S)^2 dS)^{\frac{1}{2}}$ and the inner product $(v, w)_{L^2(\mathbb{R}_+)} = \int_{\mathbb{R}_+} v(S)w(S)dS$. Calling $\mathcal{D}(\mathbb{R}_+)$ the space of the smooth functions with compact support in $\mathbb{R}_+$, we know that $\mathcal{D}(\mathbb{R}_+)$ is dense in $L^2(\mathbb{R}_+)$.

Let us introduce the space

$$V = \left\{ v \in L^2(\mathbb{R}_+) : S \frac{dv}{dS} \in L^2(\mathbb{R}_+) \right\}, \quad (1.14)$$

where the derivative must be understood in the sense of the distributions on $\mathbb{R}_+$. A natural scalar product for V is $(v, w)_V = (v, w) + (S \frac{dv}{dS}, S \frac{dw}{dS})$; the space V endowed with the norm $\|v\|_V = \sqrt{(v, v)_V}$ is a Hilbert space.

We have the following properties (see ACHDOU and PIRONNEAU [2005]).

THEOREM 1.2.

- *The space $\mathcal{D}(\mathbb{R}_+)$ is dense in V .*
- *(Poincaré's inequality) If $v \in V$, then*

$$\|v\|_{L^2(\mathbb{R}_+)} \leq 2 \|S \frac{dv}{dS}\|_{L^2(\mathbb{R}_+)}, \quad (1.15)$$

so the seminorm $|v|_V = \|S \frac{dv}{dS}\|_{L^2(\mathbb{R}_+)}$ is also a norm on V , equivalent to $\|\cdot\|_V$.

- For any $w \in L^2(\mathbb{R}_+)$, the function $S \rightarrow v(S) = \frac{1}{S} \int_0^S w(s) ds$ belongs to V , and $\|v\|_V \leq C \|w\|_{L^2(\mathbb{R}_+)}$ for some positive constant C is independent of w .

We denote by V' the topological dual space of V , and for $w \in V'$, $\|w\|_{V'} = \sup_{v \in V \setminus \{0\}} \frac{(w, v)}{|v|_V}$.

1.4.2. The weak formulation of the Black–Scholes equation

Consider a vanilla put option with maturity T and payoff function u_0 . Let u be the pricing function, that is, the price of the option at time $T - t$ and when the spot price S is $u(S, t)$. The function u solves the initial value problem

$$\frac{\partial u}{\partial t} - \frac{\sigma^2 S^2}{2} \frac{\partial^2 u}{\partial S^2} - rS \frac{\partial u}{\partial S} + ru = 0 \text{ in } \mathbb{R}_+ \times (0, T), \quad u(S, 0) = u_0(S) \text{ in } \mathbb{R}_+. \quad (1.16)$$

Let us multiply (1.16) by a smooth real-valued function w defined on $\mathbb{R}_+$ and integrate in the variable S on $\mathbb{R}_+$. Assuming that integrations by part are allowed, we obtain

$$\frac{d}{dt} \left(\int_{\mathbb{R}_+} u(S, t) w(S) dS \right) + a_t(v, w) = 0,$$

where the bilinear form a_t is defined by

$$\begin{aligned} a_t(v, w) = & \int_{\mathbb{R}_+} \left(\frac{1}{2} S^2 \sigma^2(S, t) \frac{\partial v}{\partial S} \frac{\partial w}{\partial S} + r(t) v w \right) dS \\ & + \int_{\mathbb{R}_+} \left(-r(t) + \sigma^2(S, t) + S \sigma(S, t) \frac{\partial \sigma}{\partial S}(S, t) \right) S \frac{\partial v}{\partial S} w dS. \end{aligned} \quad (1.17)$$

Assume that the coefficient $r \geq 0$ is bounded and σ is sufficiently regular so that the following makes sense.

ASSUMPTION 1.1.

1. There exists two positive constants $\underline{\sigma}$ and $\bar{\sigma}$ such that for all $t \in [0, T]$ and all $S \in \mathbb{R}_+$,

$$0 < \underline{\sigma} \leq \sigma(S, t) \leq \bar{\sigma}. \quad (1.18)$$

2. There exists a positive constant C_σ such that for all $t \in [0, T]$ and all $S \in \mathbb{R}_+$,

$$|S \frac{\partial \sigma}{\partial S}(S, t)| \leq C_\sigma. \quad (1.19)$$

LEMMA 1.1. Under Assumption 1.1, the bilinear form a_t is continuous on V , that is there exists a positive constant μ such that for all $v, w \in V$,

$$|a_t(v, w)| \leq \mu |v|_V |w|_V. \quad (1.20)$$

It also satisfies Gårding's inequality : there exists a nonnegative constant λ such that for all $v \in V$,

$$a_t(v, v) \geq \frac{\sigma^2}{4} |v|_V^2 - \lambda \|v\|_{L^2(\mathbb{R}_+)}^2. \quad (1.21)$$

One associates with the bilinear form a_t the continuous linear operator $A_t: V \rightarrow V'$; for all $v, w \in V$, $(A_t v, w) = a_t(v, w)$. The interpretation of A_t is as follows:

$$A_t v = -\frac{1}{2} \sigma^2(S, t) S^2 \frac{\partial^2 v}{\partial S^2} - r(t) S \frac{\partial v}{\partial S} + r(t) v.$$

We define $\mathcal{C}^0([0, T]; L^2(\mathbb{R}_+))$ as the space of continuous functions on $[0, T]$ with values in $L^2(\mathbb{R}_+)$, and $L^2(0, T; V)$ as the space of square-integrable functions on $(0, T)$ with values in V . Assuming that $u_0 \in L^2(\mathbb{R}_+)$ and following LIONS and MAGENES [1968], it is possible to write a weak formulation for (1.16):

Weak formulation of (1.16)

Find $u \in \mathcal{C}^0([0, T]; L^2(\mathbb{R}_+)) \cap L^2(0, T; V)$ with $\frac{\partial u}{\partial t} \in L^2(0, T; V')$, and

$$u|_{t=0} = u_0 \quad \text{in } \mathbb{R}_+, \text{ and for a.e. } t \in (0, T), \quad (1.22)$$

$$\forall v \in V, \quad \left(\frac{\partial u}{\partial t}(t), v \right) + a_t(u(t), v) = 0. \quad (1.23)$$

THEOREM 1.3. *Under Assumption 1.1 and if $u_0 \in L^2(\mathbb{R}_+)$, the weak formulation (1.22) (1.23) has a unique solution, and we have the estimate, for all t , $0 < t < T$*

$$e^{-2\lambda t} \|u(t)\|_{L^2(\mathbb{R}_+)}^2 + \frac{1}{2} \sigma^2 \int_0^t e^{-2\lambda \tau} |u(\tau)|_V^2 d\tau \leq \|u_0\|_{L^2(\mathbb{R}_+)}^2. \quad (1.24)$$

Note that Theorem 1.3 does not apply to a European call option because the payoff is not a function of $L^2(\mathbb{R}_+)$; one must either use the put-call parity (see § 1.4.4) and deduce the price of the call from that of the put or work with a different Sobolev space with a weight decaying at infinity.

1.4.3. Regularity of the weak solutions

If the interest rate, the volatility, and the payoff are smooth enough, then it is possible to prove additional regularity for the solution to (1.22) (1.23). In particular, for all $t \in [0, T]$ and for λ given in Lemma 1.1, the domain of $A_t + \lambda$ is

$$D = \{v \in V; S^2 \frac{\partial^2 v}{\partial S^2} \in L^2(\mathbb{R}_+)\}. \quad (1.25)$$

Let us assume the following.

ASSUMPTION 1.2. *There exists a positive constant C and $0 \leq \alpha \leq 1$ such that for all $t_1, t_2 \in [0, T]$ and $S \in \mathbb{R}_+$,*

$$|r(t_1) - r(t_2)| + |\sigma(S, t_1) - \sigma(S, t_2)| + S \left| \frac{\partial \sigma}{\partial S}(S, t_1) - \frac{\partial \sigma}{\partial S}(S, t_2) \right| \leq C |t_1 - t_2|^\alpha. \quad (1.26)$$

THEOREM 1.4. *Under Assumptions 1.1 and 1.2, for all $s, 0 < t \leq T$, the solution u to (1.22) (1.23) satisfies $u \in C^0([t, T]; D)$ and $\frac{\partial u}{\partial t} \in C^0([t, T]; L^2(\mathbb{R}_+))$, and there exists a constant C such that for all $t, 0 < t \leq T$,*

$$\|A_t u(t)\|_{L^2(\mathbb{R}_+)} \leq \frac{C}{t}.$$

If $u_0 \in D$, then the solution u of (1.22) (1.23) belongs to $C^0([0, T]; D)$ and $\frac{\partial u}{\partial t} \in C^0([0, T]; L^2(\mathbb{R}_+))$.

Furthermore, if $u_0 \in V$, then the solution to (1.22) (1.23) belongs to $C^0([0, T]; V) \cap L^2(0, T; D)$, $\frac{\partial u}{\partial t} \in L^2(0, T; L^2(\mathbb{R}_+))$, and there exists a nonnegative constant $\tilde{\lambda}$ such that

$$e^{-2\tilde{\lambda}t} \|S \frac{\partial u}{\partial S}(t)\|_{L^2(\mathbb{R}_+)}^2 + \frac{\sigma^2}{2} \int_0^t e^{-2\tilde{\lambda}\tau} |S \frac{\partial u}{\partial S}(\tau)|_V^2 d\tau \leq \|S \frac{\partial u_0}{\partial S}\|_{L^2(\mathbb{R}_+)}^2. \quad (1.27)$$

1.4.4. The maximum principle for weak solutions

We refer to PROTTER and WEINBERGER [1984] for a monograph on the maximum principle. The solutions of (1.16) may not vanish for $S \rightarrow +\infty$; therefore, we are going to state the maximum principle for a class of functions much larger than V , that is,

$$\mathcal{V} = \{v : \forall \epsilon > 0, v(S)e^{-\epsilon \log^2(S+2)} \in V\}. \quad (1.28)$$

Note that the polynomial functions belong to $\mathcal{V}$.

THEOREM 1.5 (Weak maximum principle). *Let $u(S, t)$ be such that for all positive number ϵ ,*

- $ue^{-\epsilon \log^2(S+2)} \in C^0([0, T]; L^2(\mathbb{R}_+)) \cap L^2(0, T; V)$,
- $u|_{t=0} \geq 0$ a.e.,
- $\frac{\partial u}{\partial t} + A_t u \geq 0$ (in the sense of distributions),

then $u \geq 0$ almost everywhere.

Various bounds The maximum principle is an extremely powerful tool for proving estimates on the solutions of elliptic and parabolic PDEs. Here, we give easy examples of its application to option pricing.

PROPOSITION 1.2. *Under Assumption 1.1, let u be the weak solution to (1.16), with $u_0 \in L^2(\mathbb{R}_+)$ being a bounded positive function, that is, $0 \leq \underline{u}_0 \leq u_0(S) \leq \overline{u}_0$. Then, a.e.*

$$\underline{u}_0 e^{-\int_0^t r(\tau) d\tau} \leq u(S, t) \leq \overline{u}_0 e^{-\int_0^t r(\tau) d\tau}. \quad (1.29)$$

PROOF. We know that $\underline{u}_0 e^{-\int_0^t r(\tau) d\tau}$ and $\overline{u}_0 e^{-\int_0^t r(\tau) d\tau}$ are two solutions to (1.16). Therefore, we can apply the maximum principle to $u - \underline{u}_0 e^{-\int_0^t r(\tau) d\tau}$ and to $\overline{u}_0 e^{-\int_0^t r(\tau) d\tau} - u$. $\square$

REMARK 1.2. In the case of a vanilla put option, $u_0(S) = (K - S)^+$, Proposition 1.2 just says that $0 \leq u(S, t) \leq K e^{-\int_0^t r(\tau) d\tau}$, which is certainly not a surprise.

For the vanilla put option as in Remark 1.2, we have more information.

PROPOSITION 1.3. *Under Assumption 1.1, let u be the weak solution to (1.16), with $u_0(S) = (K - S)^+$, then*

$$(K e^{-\int_0^t r(\tau) d\tau} - S)^+ \leq u(S, t) \leq K e^{-\int_0^t r(\tau) d\tau}. \quad (1.30)$$

PROOF. Observe that $K e^{-\int_0^t r(\tau) d\tau} - S$ is a solution to (1.16) and apply the maximum principle to $u(S, t) - (K e^{-\int_0^t r(\tau) d\tau} - S)$. We have $K e^{-\int_0^t r(\tau) d\tau} - S \leq u(S, t)$. Then, (1.30) is obtained by combining this estimate with the one given in Remark 1.2. Note that (1.30) yields $u(0, t) = K e^{-\int_0^t r(\tau) d\tau}$ for all $t \leq T$. $\square$

The Put-Call parity Let u be the pricing function of a vanilla put option with strike K , and consider $C(S, t)$ given by

$$C(S, t) = S - K e^{-\int_0^t r(\tau) d\tau} + u(S, t). \quad (1.31)$$

From the fact that u and $S - K e^{-\int_0^t r(\tau) d\tau}$ satisfy (1.16), it is clear that C is a solution to (1.16), with the Cauchy condition $C(S, 0) = (S - K)^+$. This is precisely the boundary value problem for the European vanilla call option. Furthermore, from the maximum principle, we know that a well-behaved solution of this boundary value problem (in the sense of Theorem 1.5) is unique.

1.4.5. Convexity of u in the Variable S

ASSUMPTION 1.3. *There exists a positive constant C such that*

$$|S^2 \frac{\partial^2 \sigma}{\partial S^2}(S, t)| \leq C, \quad \text{a.e.} \quad (1.32)$$

PROPOSITION 1.4. *Under Assumptions 1.1 and 1.3, let u be the weak solution to (1.16), where $u_0 \in V$ is a convex function such that $\frac{\partial^2 u_0}{\partial S^2}$ has a compact support. Then, for all $t > 0$, $u(S, t)$ is a convex function of S .*

As a consequence, we see that under Assumptions 1.1 and 1.3, the price of a vanilla European put option is convex with respect to S , and thanks to the call-put parity, this is also true for the vanilla European call.

More bounds

We focus on a vanilla put with a local volatility σ . By using Proposition 1.4, it is possible to compare u with the pricing function of vanilla puts with constant volatilities.

PROPOSITION 1.5. *Under Assumption 1.1, we have for all $t \in [0, T]$ and for all $S > 0$,*

$$\underline{u}(S, t) \leq u(S, t) \leq \bar{u}(S, t), \quad (1.33)$$

where $\underline{u}$ (respectively, $\bar{u}$) is the solution to (1.16) with $\sigma = \underline{\sigma}$, (respectively, $\bar{\sigma}$).

Localization Again, we focus on a vanilla put. For a numerical approximation to u , one has to limit the domain in the variable S , that is, to consider only $S \in (0, \bar{S})$ for $\bar{S}$ large enough and to impose some artificial boundary condition at $S = \bar{S}$. Imposing that the new function vanishes on the artificial boundary, we obtain the new boundary value problem:

$$\begin{aligned} \frac{\partial \tilde{u}}{\partial t} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 \tilde{u}}{\partial S^2} - rS \frac{\partial \tilde{u}}{\partial S} + r\tilde{u} &= 0, \quad t \in (0, T], \quad S \in (0, \bar{S}), \\ \tilde{u}(\bar{S}, t) &= 0, \quad t \in (0, T], \end{aligned} \quad (1.34)$$

with the Cauchy data $\tilde{u}(S, 0) = (K - S)^+$ in $(0, \bar{S})$. The theory of LIONS–MAGENES applies to this new boundary value problem, but one has to work in the new Sobolev space:

$$\tilde{V} = \{v, S \frac{\partial v}{\partial S} \in L^2((0, \bar{S})), v(\bar{S}) = 0\}.$$

The theory of weak solutions can be applied to problem (1.34). The question is to estimate the error between u and $\tilde{u}$.

PROPOSITION 1.6. *Under Assumption 1.1, the error $\max_{t \in [0, T], S \in [0, \bar{S}]} |u(S, t) - \tilde{u}(S, t)|$ decays faster than any negative power of $\bar{S}$ as $\bar{S} \rightarrow \infty$, that is, faster than $\bar{S}^{-\eta}$ for any positive number η .*

PROOF. From the maximum principle applied to weak solutions to (1.16) in $(0, \bar{S}) \times (0, T]$, we immediately see that $u \geq \tilde{u}$ in $(0, \bar{S}) \times (0, T)$ because $u(\bar{S}, t) \geq \tilde{u}(\bar{S}, t) = 0$. However, from Proposition 1.5, $u \leq \bar{u}$, which implies that $u(\bar{S}, t) \leq \bar{u}(\bar{S}, t)$. Call $\pi(\bar{S}) = \max_{t \in [0, T]} \bar{u}(\bar{S}, t)$. The maximum principle applied to the function $E(S, t) = \pi(\bar{S}) - u(S, t) + \tilde{u}(S, t)$ yields that $\pi(\bar{S}) \geq u - \tilde{u}$ in $[0, \bar{S}] \times [0, T]$. At this point, we have proved that

$$0 \leq u - \tilde{u} \leq \pi(\bar{S}), \quad \text{in } [0, \bar{S}] \times [0, T].$$

But $\pi(\bar{S})$ can be computed semiexplicitly by the Black–Scholes formula (1.12), and it is easy to see that for all $\eta > 0$, $\lim_{\bar{S} \rightarrow \infty} \pi(\bar{S})\bar{S}^\eta = 0$.

Therefore, $\max_{t \in [0, T], S \in [0, \bar{S}]} |u(S, t) - \tilde{u}(S, t)|$ decays faster than any power $\bar{S}^{-\eta}$ as $\bar{S} \rightarrow \infty$. $\square$

2. A finite-element method

2.1. Description of the method

Consider the boundary value problem

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{1}{2}\sigma^2(S, t)S^2 \frac{\partial^2 u}{\partial S^2} - \alpha(t)S \frac{\partial u}{\partial S} + \beta(t)u &= 0, \quad t \in (0, T), S \in (0, \bar{S}), \\ u(S, 0) &= u_0(S) \quad S \in (0, \bar{S}), \quad u(\bar{S}, t) = 0 \quad t \in (0, T]. \end{aligned} \quad (2.1)$$

This generalization of (1.34) is used for pricing European puts with possibly continuously paid dividends: this corresponds to the choice $\alpha = r(t) - q(t)$ and $\beta = r(t)$, where r is the interest rate and q is the dividend yield. A problem of the form (2.1) also arises when one looks for the option's price as a function of the maturity and the strike at a fixed spot price (the PDE is known as Dupire's equation ACHDOU and PIRONNEAU [2005], DUPIRE [1994, 1997], see 9.5); this corresponds to the choice $\alpha = -r(t) + q(t)$ and $\beta = q(t)$.

To apply the finite-element method of degree 1, we start with the variational formulation introduced in 1.4.2, given in (1.22) and (1.23).

We introduce a partition of the interval $[0, \bar{S}]$ into subintervals $\kappa_i = [S_{i-1}, S_i]$, $1 \leq i \leq N+1$, such that $0 = S_0 < S_1 < \dots < S_N < S_{N+1} = \bar{S}$. We call $h_i = S_i - S_{i-1}$ and $h = \max_{i=1, \dots, N+1} h_i$. We define the mesh $\mathcal{T}_h$ of $[0, \bar{S}]$ as the set $\{\kappa_1, \dots, \kappa_{N+1}\}$. In what follows, we will assume that the strike K coincides with some node of $\mathcal{T}_h$, that is, there $S_{k_0} = K$ for some admissible k_0 .

We define the discrete space V_h by

$$V_h = \left\{ v_h \in C^0([0, \bar{S}]), \quad v_h(\bar{S}) = 0; \quad \forall \kappa \in \mathcal{T}_h, v_h|_\kappa \text{ is affine} \right\}. \quad (2.2)$$

The assumption on the mesh ensures that $u_0 \in V_h$ when $u_0 = (K - S)^+$. The discrete problem obtained by applying the Euler implicit scheme in time reads:

$$\begin{aligned} \text{find } (u_h^m)_{1 \leq m \leq M}, u_h^m \in V_h \text{ with } u_h^0(S_i) &= u_0(S_i), \quad i = 0 \dots N+1, \text{ and,} \\ \text{for } m = 1 \dots M, \forall v_h \in V_h, \quad &\left(u_h^m - u_h^{m-1}, v_h \right) + \delta t_m a_{t_m}(u_h^m, v_h) = 0, \end{aligned} \quad (2.3)$$

where

$$\begin{aligned} a_t(v, w) &= \int_0^{\bar{S}} \frac{1}{2} S^2 \sigma^2(S, t) \frac{\partial v}{\partial S} \frac{\partial w}{\partial S} \\ &+ \int_0^{\bar{S}} \left(-\alpha(t) + \sigma^2(S, t) + S \sigma(S, t) \frac{\partial \sigma}{\partial S}(S, t) \right) S \frac{\partial v}{\partial S} w + \beta(t) \int_0^{\bar{S}} v w. \end{aligned} \quad (2.4)$$

Note that we have a simpler expression for $a_t(v, w)$, for $v, w \in V_h$, when σ is continuous with respect to S :

$$a_t(v, w) = - \sum_{i=1}^N \frac{1}{2} S_i^2 \sigma^2(S_i, t) \left[\frac{\partial v}{\partial S} \right](S_i) w(S_i) - \alpha(t) \int_0^{\bar{S}} S \frac{\partial v}{\partial S} w + \beta(t) \int_0^{\bar{S}} v w, \quad (2.5)$$

where $[\cdot]$ denotes the jump

$$\left[\frac{\partial v}{\partial S} \right](S_i) = \frac{\partial v}{\partial S}(S_i^+) - \frac{\partial v}{\partial S}(S_i^-). \quad (2.6)$$

For $i = 0, \dots, N+1$, let w_i be the piecewise linear function on the mesh that takes the value 1 at S_i and 0 at S_j , $j \neq i$, $j = 0, \dots, N+1$. Then, $(w_i)_{i=0, \dots, N}$ is the nodal basis of V_h and

$$u_h^m(S) = \sum_0^N u_h^m(S_i) w_i(S). \quad (2.7)$$

Let $\mathbf{M}$ and $\mathbf{A}^m$ in $\mathbb{R}^{N \times N}$ be, respectively, the mass and stiffness matrices defined by $\mathbf{M}_{i,j} = (w_i, w_j)$, $\mathbf{A}_{i,j}^m = a_{t_m}(w_j, w_i)$, $0 \leq i, j \leq N$. Calling $\mathbf{u}^m = (u_h^m(S_0), \dots, u_h^m(S_N))^T$, (2.3) is equivalent to

$$(\mathbf{M} + \delta_{t_m} \mathbf{A}^m) \mathbf{u}^m = \mathbf{M} \mathbf{u}^{m-1}. \quad (2.8)$$

The shape functions w_i corresponding to vertex S_i are supported in $[S_{i-1}, S_{i+1}]$. This implies that the matrices $\mathbf{M}$ and $\mathbf{A}^m$ are tridiagonal because when $|i - j| > 1$, the intersection of the supports of w_i and w_j has measure 0. Furthermore, for $i \leq N$,

$$\begin{aligned} w_i(S) &= \frac{S - S_{i-1}}{h_i}, & \frac{\partial w_i}{\partial S} &= \frac{1}{h_i}, & \forall S \in (S_{i-1}, S_i), \\ w_i(S) &= \frac{S_{i+1} - S}{h_{i+1}}, & \frac{\partial w_i}{\partial S} &= -\frac{1}{h_{i+1}}, & \forall S \in (S_i, S_{i+1}), \end{aligned} \quad (2.9)$$

giving

$$\begin{aligned} \int_0^{\bar{S}} w_{i-1} w_i &= \frac{h_i}{6}, & \int_0^{\bar{S}} S w_i \frac{\partial w_{i-1}}{\partial S} &= -\frac{S_{i-1}}{6} - \frac{S_i}{3}, \\ \int_0^{\bar{S}} w_i w_i &= \frac{h_i + h_{i+1}}{3}, & \int_0^{\bar{S}} S w_i \frac{\partial w_i}{\partial S} &= -\frac{1}{2} \int_0^{\bar{S}} w_i^2 = -\frac{h_i + h_{i+1}}{6}, & \text{if } i > 0, \\ \int_0^{\bar{S}} w_0 w_0 &= \frac{h_1}{3}, & \int_0^{\bar{S}} S w_0 \frac{\partial w_0}{\partial S} &= -\frac{1}{2} \int_0^{\bar{S}} w_0^2 = -\frac{h_1}{6}, \\ \int_0^{\bar{S}} w_{i+1} w_i &= \frac{h_{i+1}}{6}, & \int_0^{\bar{S}} S w_i \frac{\partial w_{i+1}}{\partial S} &= \frac{S_{i+1}}{6} + \frac{S_i}{3}. \end{aligned}$$

From this, a few calculations show that the entries of A^m are

$$\begin{aligned}
 A_{i,i-1}^m &= -\frac{S_i^2 \sigma^2(S_i, t_m)}{2h_i} + \frac{\alpha(t_m)S_i}{2} + (\beta(t_m) - \alpha(t_m))\frac{h_i}{6}, \quad 1 \leq i \leq N, \\
 A_{i,i}^m &= \frac{S_i^2 \sigma^2(S_i, t_m)}{2} \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) + \frac{\alpha(t_m)}{2}(h_{i+1} + h_i) \\
 &\quad + (\beta(t_m) - \alpha(t_m))\frac{h_i + h_{i+1}}{3}, \quad 1 \leq i \leq N, \\
 A_{0,0}^m &= \frac{\alpha(t_m)}{2}h_1 + (\beta(t_m) - \alpha(t_m))\frac{h_1}{3}, \\
 A_{i,i+1}^m &= -\frac{S_i^2 \sigma^2(S_i, t_m)}{2h_{i+1}} - \frac{\alpha(t_m)S_i}{2} + (\beta(t_m) - \alpha(t_m))\frac{h_{i+1}}{6}, \quad 0 \leq i \leq N-1.
 \end{aligned}$$

When the mesh is uniform, this matrix is close (but not proportional) to the stiffness matrix obtained by using the finite-difference method with a centered scheme (see ACHDOU and PIRONNEAU [2005]). The entries of M are

$$\begin{aligned}
 M_{i,i-1} &= \frac{h_i}{6}, \quad 1 \leq i \leq N, \\
 M_{i,i} &= \frac{h_i + h_{i+1}}{3}, \quad 1 \leq i \leq N, \quad M_{0,0} = \frac{h_1}{3}, \\
 M_{i,i+1} &= \frac{h_{i+1}}{6}, \quad 0 \leq i \leq N-1.
 \end{aligned}$$

REMARK 2.1. The value of u at $S = 0$ is known for all time because the equation degenerates into $\frac{\partial u}{\partial S} + \beta(t)u = 0$. Therefore, $u(0, t) = u_0(0) \exp(-\int_0^t \beta(s)ds)$. Hence, it is possible to impose that

$$u_0^m = u_0(0) \exp\left(-\int_0^{t_m} \beta(s)ds\right)$$

and plug this into (2.8). In this case, since u_0^m is known, (2.8) can be rewritten as $\forall i = 1, \dots, N$

$$\sum_{j=1}^N (M_{i,j} + \delta t_m A_{i,j}^m) u_j^m = \sum_{j=0}^N M_{i,j} u_j^{m-1} - (M_{i,0} + \delta t_m A_{i,0}^m) u_0^m. \quad (2.10)$$

2.2. A C++ implementation

The following is a simple C++ implementation of the above for a put option with dividend $d(t)$ on a general mesh that may vary at each time step. It can also solve the Dupire equation (see section 9.5).

The boundary condition at zero is implemented as in Remark 2.1. There are two classes, one for the mesh and one for the put option problem. The mesh class has a simple constructor for a mesh that can be refined near the strike and at the origin in time. The calling program is

```

int main()
{
    VarMesh m(50,100,0.5,300.,1.05,0.9,100.,1.02);
    Option p(1,&m,100.,0.05,0.3);
    p.calc();
    ofstream result("u.txt");
    for(int i = 0; i < m.nT; i++)
    {
        for(int j = 0; j < m.nX[i]; j++)
            ff << m.x[i][j] << "\t" << m.t[i] << "\t" << p.u[i][j] << endl;
        ff << endl;
    }
    return 0;
}

```

The mesh is called m ; it has 50 time steps and at most 100 mesh points maximum over $(0, 300) \times (0, 0.5)$. At each time level, the time step is increased by a factor 1.05, and the number of mesh points N_m is decreased by a factor 1.02, namely,

$$\delta t^m = 1.05 \delta t^{m-1}, \quad N_m = \text{int}(N_{m-1}/1.02), \quad m = 1, \dots, 49.$$

Finally, the mesh is not uniform in space, it is refined near the strike $x_S = 100$ by a factor 0.9, namely, on the left of x_S , $\delta x_i = 0.9 \delta x_{i-1}$ and on the right of x_S , $\delta x_{i-1} = 0.9 \delta x_i$.

The put option is called p ; it is solution to Black–Schole’s PDE (hence, the 1 as first parameter in the constructor, 0 being for Dupire’s equation). The strike is 100, and the maturity is obtained from the second parameter in the mesh constructor; the interest rate is constant here and equal to 0.05 and so is the volatility, which is 0.3. The function `calc` solves Black–Scholes or Dupire’s equation by FEM with LU factorization at each time step. The results are in $m.u$ and written in the file `u.txt` in a format readable by `gnuplot`. The following `gnuplot` statement `splot "u.txt" w l` displays something like Fig. 2.1. File `optionhb.hpp` contains the definitions and implementation of the classes `VarMesh` and `Option`.

Inappropriate ioctl for device

```

using namespace std;

// #include "ddouble.h"
typedef double ddouble; // for automatic differentiation

class VarMesh {
public:
    const int nT; // nb time step
    int *nX; // nb of vertices at each time step
    double *t, **x; // mesh points x at times t
    double T, xmax;
    double **xx, *vxx; // holds vertices of 2 mesh levels; holds
        function val
    int kk, *ixx; // total nb vertices; holds origin of vertex (<0
        if from level k-1)
    VarMesh(const int nt, const int nx, const double T1, const double
        xmax1,
    const double tscal=1, const double xc=0, const double xS=1,

```

```

const double xsize=1); // defaults=>uniform mesh
void interpol(ddouble* v, const int k, ddouble* w);
// w=interpol of v[] on x[k-1], to x[k]
void intersect(const int k); // intersect level k and k-1,
// result in xx,ixx,kk
void integral(ddouble* v, const int k, ddouble* w);
// even if v and w are on != mesh
};

```

```

class Option {
public:
    const int s; // s=0 for Dupire and =1 for B&S
    VarMesh *msh; // mesh
    const double S,T; // spot price and Maturity
    double *r, *d; // interest rate and dividend
    ddouble **u, **sigma; // solution and volatility(function of K and T)
    ddouble *w, *am, *bm, *cm; // working arrays for M.u and Gauss fact.

    ddouble u0(const double x); // initial condition
    void factLU(const int nX1);
    void solveLU(const int nX1, ddouble* z);
    void calc();
    Option(const int s1, VarMesh* msh1, const double S1,
    double r1, double sigma1, double d1=0);
    ~Option();
};

```

```

Option::Option(const int s1, VarMesh* msh1, const double S1,
    double r1, double sigma1, double d1): s(s1), msh(msh1), S(S1),
    T(msh1->T){
    VarMesh& m = *msh;
    r = new double[m.nT];
    d = new double[m.nT];
    u = new ddouble*[m.nT];
    sigma = new ddouble*[m.nT];
    int nXmax =0;
    for(int i = 0; i < m.nT; i++)
    { if(nXmax<m.nX[i] ) nXmax=m.nX[i];
      u[i] = new ddouble[m.nX[i]];
      sigma[i] = new ddouble[m.nX[i]];
      r[i]=r1; d[i] = d1;
      for(int j=0;j<m.nX[i];j++)sigma[i][j]=sigma1;
    }
    am = new ddouble[nXmax];
    bm = new ddouble[nXmax];
    cm = new ddouble[nXmax];
    w = new ddouble[nXmax];
}

```

```

ddouble Option::u0(const double x1) { return x1<S ? (S-x1) : 0;}

```

```

void Option::factLU(const int nX1){
    cm[1] /= bm[1];
    for(int i=2;i<nX1-1;i++)
    {
        bm[i] -= am[i]*cm[i-1];
        cm[i] /= bm[i];
    }
}

void Option::solveLU(const int nX1, ddouble *z){
    z[1] /= bm[1];
    for(int i=2;i<nX1-1;i++)
        z[i] = (z[i] - am[i]*z[i-1])/bm[i];
    for(int i=nX1-2;i>0;i--)
        z[i] -= cm[i]*z[i+1];
}

void Option::calc( ){
    VarMesh& m = *msh;
    for(int i=0;i<m.nX[0];i++) u[0][i] = u0(m.x[0][i]);
    const double nml=1./6.; // no mass lumping = 1./6., mass lumping =0
    for(int j=1;j<m.nT;j++) // time loop
    {
        double dt = m.t[j]-m.t[j-1];
        double aux = (r[j]*(4*s-1)+d[j]*(3-4*s))/3,
            auy = (r[j]*(1-s)+d[j]*s)*dt/6;
        m.integral(u[j-1],j,w); // rhs of PDE
        for(int i=1;i<m.nX[j]-1;i++)
        {
            double hi = m.x[j][i]-m.x[j][i-1], hi1 = m.x[j][i+1]-m.x[j][i];
            double xss = m.x[j][i]*sigma[j][i]*sigma[j][i];
            bm[i] =(hi+hi1)*(0.5-nml +dt*(m.x[j][i]*xss/hi/hi1+aux)/2);
            // FEM matrix
            am[i] = nml*hi - dt*m.x[j][i]*(xss/hi - (2*s-1)*(r[j]-d[j]))/2
                + auy*hi;
            cm[i] = nml*hi1- dt*m.x[j][i]*(xss/hi1 + (2*s-1)*(r[j]-d[j]))/2
                + auy*hi1;
        }
        u[j][m.nX[j]-1]=0; // C.L.
        u[j][0]=u[j-1][0]*exp(((1-s)*d[j]-s*r[j])*dt); // C.L.
        double hi = m.x[j][1]-m.x[j][0];
        w[1] -= u[j][0]*(nml*hi - dt*m.x[j][1]*
            ( m.x[j][1]*sigma[j][1]*sigma[j][1]/hi- (2*s-1)*(r[j]-d[j]))/2
            + auy*hi);
        factLU(m.nX[j]);
        solveLU(m.nX[j],w);
        for(int i=1;i<m.nX[j]-1;i++) u[j][i]=w[i];
    }
}

```

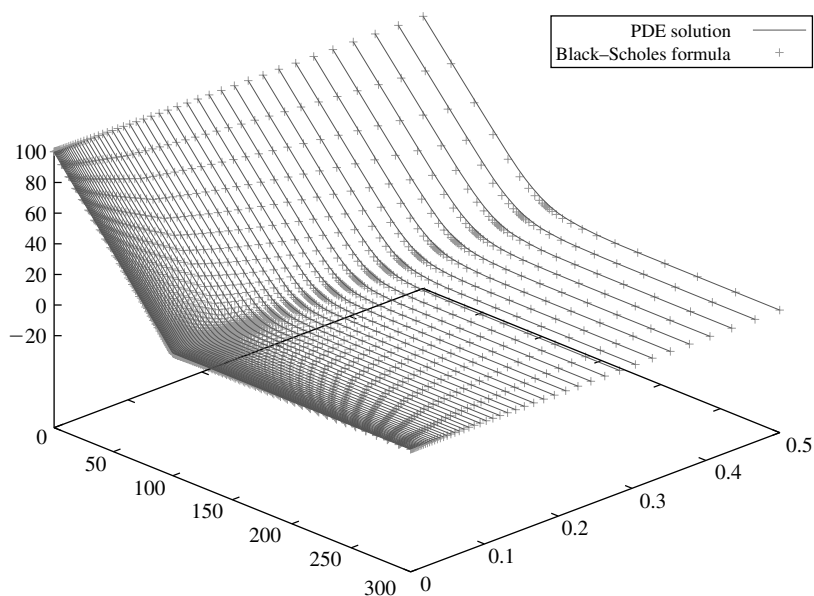


FIG. 2.1 Result of the program of Section 2.2 and comparison with the Black–Scholes formula at the mesh points (crosses).

3. Mesh adaptivity

Here, we discuss a strategy in order to separately adapt the grids in the variables t and S . Moreover, the mesh in the variable S may vary in time. This method was originally proposed by BERGAM, BERNARDI and MGHAZLI [2005]. The goal is to find local error indicators that can be explicitly computed from the solution to the discrete problem such that their Hilbertian sum is equivalent to the global error. These indicators are said to be optimal if the constants of the norm-equivalence inequalities are independent of the error. We consider two families of error indicators, both of residual type. The first family is global with respect to the spot price variable and local with respect to time: it gives relevant information in order to refine the mesh in time. The second family is local with respect to both spot and time variables and provides an efficient tool for mesh adaptivity in the price variable at each time step.

All the results below are proved by ACHDOU and PIRONNEAU [2005].

3.1. Multiple meshes and integration

Consider a parabolic problem in the variables x and t , where t is the time variable. Mesh adaptivity may lead to situations when the mesh used for the variable x varies between two time steps; assume that mesh adaption is performed at time t_m : in this case, u^m is defined on the mesh $\mathcal{T}_H$ used at time t_m and has to be integrated against functions defined

on the mesh $\mathcal{T}_h$ used at time t_{m+1} . Let w_i be a nodal basis function associated with the mesh $\mathcal{T}_h$. To compute the integral of $u^m w_i$, we intersect the meshes $\mathcal{T}_H$ and $\mathcal{T}_h$; the result is a new mesh on which both functions u^m and w_i are piecewise linear; therefore, the integral can be computed without error.

3.2. Error indicators for the semidiscrete problem in time

Our goal is Theorem 3.1 below, but we need first to define appropriate norms.

We go back to problem (1.34). From Assumption 1.1 and from Gårding's inequality (1.21), it is possible to prove that (1.34) admits a unique solution that we write u for simplicity. Moreover, introducing the norm

$$[[v]](t) = \left(e^{-2\lambda t} \|v(t)\|^2 + \frac{1}{2} \sigma^2 \int_0^t e^{-2\lambda \tau} |v(\tau)|_V^2 d\tau \right)^{\frac{1}{2}}, \quad (3.1)$$

where $\|\cdot\|$ stands for the norm in $L^2(0, \bar{S})$, we have, multiplying (1.34) by $u(t)e^{-2\lambda t}$ and integrating in $(0, \bar{S}) \times (0, T)$,

$$[[u]](t) \leq \|u_0\|. \quad (3.2)$$

We introduce a partition of the interval $[0, T]$ into subintervals $[t_{n-1}, t_n]$, $1 \leq n \leq N$, such that $0 = t_0 < t_1 < \dots < t_N = T$. We denote by δt_n the length $t_n - t_{n-1}$ and by δt the maximum of the δt_n , $1 \leq n \leq N$. We also define the regularity parameter $\rho_{\delta t}$:

$$\rho_{\delta t} = \max_{2 \leq n \leq N} \frac{\delta t_n}{\delta t_{n-1}}. \quad (3.3)$$

Given $u^0 = u_0 \in L^2(0, \bar{S})$, the semidiscrete problem arising from an implicit Euler scheme is: find $(u^n)_{1 \leq n \leq N} \in V^N$ satisfying

$$\forall n, 1 \leq n \leq N, \quad \left(u^n - u^{n-1}, v \right) + \delta t_n a_{t_n}(u^n, v) = 0, \quad \forall v \in V, \quad (3.4)$$

where V is the space

$$V = \left\{ v \in L^2(0, \bar{S}) : S \frac{\partial v}{\partial \bar{S}} \in L^2(0, \bar{S}), v(\bar{S}) = 0 \right\}. \quad (3.5)$$

For δt smaller than $1/(2\lambda)$ (where λ is the constant appearing in Gårding's inequality (1.21)), the existence and uniqueness of $(u^n)_{1 \leq n \leq N}$ is a consequence of the Lax–Milgram lemma. We call $u_{\delta t}$ the function that is affine on each interval $[t_{n-1}, t_n]$ and such that $u_{\delta t}(t_n) = u^n$, $0 \leq n \leq N$.

From the standard identity $(a - b, a) = \frac{1}{2}|a|^2 + \frac{1}{2}|a - b|^2 - \frac{1}{2}|b|^2$, a few calculations show that

$$(1 - 2\lambda \delta t_n) \|u^n\|^2 + \frac{1}{2} \delta t_n \sigma^2 |u^n|_V^2 \leq \|u^{n-1}\|^2. \quad (3.6)$$

Consider the discrete norm for the sequence $(v^m)_{1 \leq m \leq n}$:

$$[[v^m]]_n = \left(\left(\prod_{i=1}^n (1 - 2\lambda\delta t_i) \right) \|v^n\|^2 + \frac{\sigma^2}{2} \sum_{m=1}^n \delta t_m \left(\prod_{i=1}^{m-1} (1 - 2\lambda\delta t_i) \right) |v^m|_V^2 \right)^{\frac{1}{2}} \quad (3.7)$$

Multiplying Eq. (3.6) by $\prod_{i=1}^{n-1} (1 - 2\lambda\delta t_i)$ and summing the equations on n , we obtain

$$[[u^m]]_n^2 \leq \|u^0\|^2. \quad (3.8)$$

We will need an equivalence relation between $[[u^m]]_n$ and $[[u_{\delta t}]](t_n)$.

LEMMA 3.1. *There exists a positive real number $\alpha \leq \frac{1}{2}$ such that the following equivalence property holds for $\delta t \leq \frac{\alpha}{\lambda}$ and for any family $(v^n)_{0 \leq n \leq N}$ in V_0^{N+1} ,*

$$\frac{1}{8} [[v^m]]_n^2 \leq [[v_{\delta t}]]^2(t_n) \leq \max(2, 1 + \rho_{\delta t}) [[v^m]]_n^2 + \frac{1}{2} \sigma^2 \delta t_1 |v^0|_V^2. \quad (3.9)$$

From (3.8) and (3.9), we deduce that for all n , $1 \leq n \leq N$,

$$[[u_{\delta t}]](t_n) \leq c(u_0), \quad (3.10)$$

where

$$c(u_0) = \left(\max(2, 1 + \rho_{\delta t}) \|u_0\|^2 + \frac{1}{2} \sigma^2 \delta t_1 |u_0|_V^2 \right)^{\frac{1}{2}}. \quad (3.11)$$

To evaluate $[[u - u_{\delta t}]](t_n)$, we make a further assumption on the coefficients.

ASSUMPTION 3.1. *The function r is Lipschitz continuous in $[0, T]$. The functions σ and $S \frac{\partial \sigma}{\partial S}$ are Lipschitz continuous with respect to t uniformly in the variable S .*

With Assumption 3.1 and the previous set of assumptions, there exist constants L_1 , L_2 , and L_3 such that, for all t and t' in $[0, T]$,

$$\left\| \sigma^2(., t) - \sigma^2(., t') \right\|_{L^\infty(0, \bar{S})} \leq L_1 |t' - t|, \quad |r(t) - r(t')| \leq L_3 |t' - t| \quad (3.12)$$

and a similar inequality for $\| -r(t) + \frac{1}{2} \sigma^2(., t) + S \sigma(., t) \frac{\partial \sigma}{\partial S}(., t) \|_{L^\infty(0, \bar{S})}$ with L_2 on the right-hand side.

LEMMA 3.2. *Assume that $u_0 \in V_{0h}$. Then, there exists a constant $\alpha \leq \frac{1}{2}$ such that if $\delta t \leq \frac{\alpha}{\lambda}$, the following a posteriori error estimate holds between the solutions of (1.34)*

and (3.4):

$$[[u - u_{\delta t}]](t_n) \leq cb$$

$$\text{with } b = \left(\frac{L}{\underline{\sigma}^2} c(u_0) \delta t + \frac{\mu}{\underline{\sigma}^2} (1 + \rho_{\delta t}) [[u_{\delta t} - u_{h, \delta t}]](t_n) + \frac{\mu}{\underline{\sigma}^2} \left(\sum_{m=1}^n \eta_m^2 \right)^{\frac{1}{2}} \right), \quad (3.13)$$

where

$$\eta_m^2 = \delta t_m e^{-2\lambda t_{m-1}} \frac{\sigma^2}{2} |u_h^m - u_h^{m-1}|_V^2, \quad (3.14)$$

and c is a positive constant, $L = 4L_1 + 2L_2 + L_3$, where L_1 , L_2 , and L_3 are given by (3.12), $c(u_0)$ is given by (3.11). Furthermore, if the assumptions of Proposition 3.2 are satisfied, the following a posteriori error estimate holds:

$$\left\| \frac{\partial}{\partial t} (u - u_{\delta t}) \right\|_{L^2(0, t_n, V')} \leq c \left(\frac{\mu + \sigma^2}{\underline{\sigma}} \right) b, \quad (3.15)$$

where μ is the continuity constant of a in (1.20).

3.3. Error indicators for the fully discrete problem

The fully discrete problem has already been defined in (2.3). For each time interval, let $(\mathcal{T}_{nh})$ be the mesh of Ω . Let $h^{(n)}$ denote the maximal size of the intervals in $\mathcal{T}_{nh}$. For a given element $\omega \in \mathcal{T}_{nh}$, let h_ω be the diameter of ω and let $S_{\min}(\omega)$ and $S_{\max}(\omega)$ be the endpoints of ω . We assume that there exists a constant ρ_h such that, for two adjacent elements ω and ω' of $(\mathcal{T}_{nh})$,

$$h_\omega \leq \rho_h h_{\omega'}. \quad (3.16)$$

For each h , we define the discrete spaces by

$$V_{nh} = \{v_h \in V, \forall \omega \in \mathcal{T}_{nh}, v_h|_\omega \in \mathcal{P}_1\}. \quad (3.17)$$

Here, the grids $\mathcal{T}_{nh}$ for different values of n are not independent; indeed, each triangulation $\mathcal{T}_{nh}$ is derived from $\mathcal{T}_{n-1, h}$ by cutting some elements of $\mathcal{T}_{n-1, h}$ into a limited number of smaller intervals or by gluing together a limited number of elements of $\mathcal{T}_{n-1, h}$. This permits (w_h^{n-1}, v_h^n) to be evaluated exactly if $w_h^{n-1} \in V_{n-1, h}$ and $v_h^n \in V_{nh}$.

Assuming that $u_0 \in V_{0h}$, the fully discrete problem reads:

find $(u_h^n)_{1 \leq n \leq N}$, $u_h^n \in V_{nh}$ satisfying

$$\forall v_h \in V_{nh}, \quad \left(u_h^n - u_h^{n-1}, v_h \right) + \delta t_n a_{t_n}(u_h^n, v_h) = 0, \quad (3.18)$$

where $u_h^0 = u_0$. As above, for δt smaller than $1/(2\lambda)$, the existence and uniqueness of $(u_h^n)_{0 \leq n \leq N}$ is a consequence of the Lax–Milgram lemma, and we have the stability

estimate

$$[[u_h^m]]_n \leq \|u^0\|. \quad (3.19)$$

We call $u_{h,\delta t}$ the function that is affine on each interval $[t_{n-1}, t_n]$ and such that $u_{h,\delta t}(t_n) = u_h^n$.

We wish to bound the error $[[u - u_{h,\delta t}]](t_n)$, $1 \leq n \leq N$, by indicators computable from $u_{h,\delta t}$. First, let us separate the time discretization error from the space discretization by the triangular inequality

$$[[u - u_{h,\delta t}]](t_n) \leq [[u - u_{\delta t}]](t_n) + [[u_{\delta t} - u_{h,\delta t}]](t_n).$$

LEMMA 3.3. Assume that $u_0 \in V_{0h}$. Then, the following a posteriori error estimate holds between the solution $(u^n)_{1 \leq n \leq N}$ to problem (3.4) and the solution $(u_h^n)_{1 \leq n \leq N}$ to problem (3.18): there exists a constant c such that, for all t_n , $1 \leq n \leq N$,

$$[[u_{\delta t} - u_{h,\delta t}]]^2(t_n) \leq \frac{c}{\underline{\sigma}^2} \max(2, 1 + \rho_{\delta t}) \sum_{m=1}^n \delta t_m \prod_{i=1}^{m-1} (1 - 2\lambda \delta t_i) \sum_{\omega \in \mathcal{T}_{mh}} \eta_{m,\omega}^2, \quad (3.20)$$

where

$$\eta_{m,\omega} = \left(\frac{h_\omega}{S_{\max}(\omega)} \left\| \frac{u_h^m - u_h^{m-1}}{\delta t_m} - rS \frac{\partial u_h^m}{\partial S} + r u_h^m \right\|_{L^2(\omega)} \right). \quad (3.21)$$

Combining Lemmas 3.2 and 3.3 leads to the full a posteriori error estimate.

THEOREM 3.1. Assume that $u_0 \in V_{0h}$, and that $\lambda \delta t \leq \alpha$ is as in Lemma 3.1. Then, the following a posteriori error estimate holds between the solution u to problem (1.34) and the solution $u_{h,\delta t}$ to problem (3.18): there exists a constant c such that, for all t_n , $1 \leq n \leq N$,

$$[[u - u_{h,\delta t}]](t_n) \leq c \left(\frac{L}{\underline{\sigma}^2} c(u_0) \delta t + \frac{\mu}{\underline{\sigma}^2} \left(\sum_{m=1}^n \eta_m^2 + \frac{\delta t_m}{\underline{\sigma}^2} g(\rho_{\delta t}) \prod_{i=1}^{m-1} (1 - 2\lambda \delta t_i) \sum_{\omega \in \mathcal{T}_{mh}} \eta_{m,\omega}^2 \right)^{\frac{1}{2}} \right), \quad (3.22)$$

where $L = 4L_1 + 2L_2 + L_3$, L_1 , L_2 , and L_3 are given by (3.12), $c(u_0)$ is given by (3.11), η_m is given by (3.14), and $\eta_{m,\omega}$ is given by (3.21), and

$$g(\rho_{\delta t}) = (1 + \rho_{\delta t})^2 \max(2, 1 + \rho_{\delta t}).$$

3.3.1. Conclusion

In (3.13), (3.15), and (3.20), we have bounded the norm of the error produced by the finite-element method by a Hilbert sum involving the error indicators η_m and $\eta_{m,\omega}$ given

in (3.14) and (3.21), which are respectively, local in t and local in t and S . Conversely, in Propositions 3.1 and 3.2 below, we will see that the error indicators can be bounded by local norms of the error. This shows that the error indicators are both reliable and efficient or, in other words, that the error produced by the method is well approximated by these indicators. Furthermore, since the indicators are local, they tell us where the mesh should be refined. It is now possible to build a computer program that adapts the mesh so as to reduce the error to a given number ϵ ; from the result of an initial computation $u_{h,\delta t}$, we can adapt separately the meshes in the variables t and S so as to decrease the Hilbert sum in (3.22). The process may be repeated until the desired accuracy is obtained.

3.4. Upper bounds for the error indicators

We now investigate the efficiency of the indicators in (3.14) and (3.21). For that, we introduce the notation $[[v^n]]$, for $(v^n)_{1 \leq n \leq N}$, $v^n \in V$:

$$[[v^n]]^2 = \frac{\sigma^2}{2} \delta t_n \prod_{i=1}^{n-1} (1 - 2\lambda \delta t_i) |v^n|_V^2. \quad (3.23)$$

PROPOSITION 3.1. *Assume that u^0 belongs to V , and that $\lambda \delta t \leq \alpha$ as in Lemma 3.1. The following estimate holds for the indicator η_n , $2 \leq n \leq N$,*

$$\eta_n \leq c \left(\begin{aligned} & [[u^n - u_h^n]] + \sqrt{\rho_{\delta t}} [[u^{n-1} - u_h^{n-1}]] \\ & + \frac{e^{-\lambda t_{n-1}}}{\underline{\sigma}} \left(\left\| \frac{\partial}{\partial t} (u - u_{\delta t}) \right\|_{L^2(t_{n-1}, t_n; V')} + \|u - u_{\delta t}\|_{L^2(t_{n-1}, t_n; V)} \right) \\ & + \left(\frac{L}{\underline{\sigma}^2} (\max(1, \rho_{\delta t}))^{\frac{1}{2}} + \frac{\lambda \mu}{\underline{\sigma}^2} \right) \delta t_n \|u^0\| \end{aligned} \right), \quad (3.24)$$

and

$$\eta_1 \leq c \left(\begin{aligned} & [[u^1 - u_h^1]] + \frac{1}{\underline{\sigma}} \left(\left\| \frac{\partial}{\partial t} (u - u_{\delta t}) \right\|_{L^2(0, t_1; V')} + \|u - u_{\delta t}\|_{L^2(0, t_1; V)} \right) \\ & + \frac{L + \lambda \mu}{\underline{\sigma}^2} \delta t_1 \|u^0\| + \frac{L}{\underline{\sigma}} (\delta t_1)^{\frac{3}{2}} |u^0|_V \end{aligned} \right), \quad (3.25)$$

where c is a positive constant.

The most important property of estimate (3.24) is that, up to the last term that depends on the data, all the terms in the right-hand side of (3.24) are local in time. More precisely, they involve the solution in the interval $[t_{n-1}, t_n]$.

We need to define a few notations before stating the upper bound result for $\eta_{n,\omega}$.

For $\omega \in \mathcal{T}_{n,h}$, let K_ω be the union of ω and the element that shares a node with ω , and $V_0(K_\omega)$ be the closure of $\mathcal{D}(K_\omega)$ in $V(K_\omega) = \{v \in L^2(K_\omega); S \frac{\partial v}{\partial S} \in L^2(K_\omega)\}$ endowed with the norm $\|v\|_{V(K_\omega)} = (\int_{K_\omega} v^2(S) + S^2 (\frac{\partial v}{\partial S}(S))^2)^{\frac{1}{2}}$. We also define $\|v\|_{V_0(K_\omega)} = (\int_{K_\omega} S^2 (\frac{\partial v}{\partial S}(S))^2)^{\frac{1}{2}}$, for $v \in V_0(K_\omega)$. We denote by $V'_0(K_\omega)$ the dual space of $V_0(K_\omega)$

endowed with dual norm. We also need the assumption that the meshes do not vary too much between two time steps.

ASSUMPTION 3.2. *For $n = 1, \dots, N$, there exists a family of grids $(\mathcal{T}_{n,h}^*)_h$ with property (3.16) such that for all h and n each element of $\mathcal{T}_{n,h}$ and of $\mathcal{T}_{n-1,h}$ is the union of at most s elements of $\mathcal{T}_{n,h}^*$ (where s is bounded independently of h and n).*

PROPOSITION 3.2. *Under Assumption 3.2, the following estimate holds for the indicator $\eta_{n,\omega}$ defined in (3.21), for all $\omega \in \mathcal{T}_{n,h}$, $1 \leq n \leq N$,*

$$\eta_{n,\omega} \leq C \left(\left\| \frac{u^{n-1} - u_h^{n-1} - u^n + u_h^n}{\delta t_n} \right\|_{V'_0(K_\omega)} + \mu \left\| S \frac{\partial(u^n - u_h^n)}{\partial S} \right\|_{L^2(K_\omega)} \right). \quad (3.26)$$

3.5. Computation of the bounds

To compute $\rho_{\delta t}$ and $\eta_{n,\omega}$, we may do as follows:

```
void errorindic(Option& p, double*
rho, double** etamw){
    int nT = p.msh->nT;
    double sigma_m=p.sigma[0][0];
    for(int k=1;k<nT;k++){
        for(int i=0;i<p.msh->nX[k];i++)
            if(sigma_m>p.sigma[k][i]) sigma_m= p.sigma[k][i];
        rho[k]=0;
        double aux = 0.5*(p.msh->t[k]-p.msh->t[k-1])*sqr(sigma_m);
        for(int i=1;i<p.msh->nX[k];i++){
            rho[k] += aux*(
                p.msh->x[k][i]*sqr((p.u[k][i]-p.u[k-1][i]-p.u[k][i-1]+p.u[k-1][i-1])
                    /(p.msh->x[k][i]-p.msh->x[k][i-1])) + sqr(p.u[k][i]-p.u[k-1][i]));
            etamw[k][i] = fabs((p.u[k][i]-p.u[k-1][i])/(p.msh->t[k]-p.msh->t[k-1]))
                *(p.msh->x[k][i]-p.msh->x[k][i-1])/p.msh->x[k][i];
        }
    }
}
```

The graphs in Fig. 3.1 show the two indicators and the actual error. Comparison shows the efficiency of the indicators.

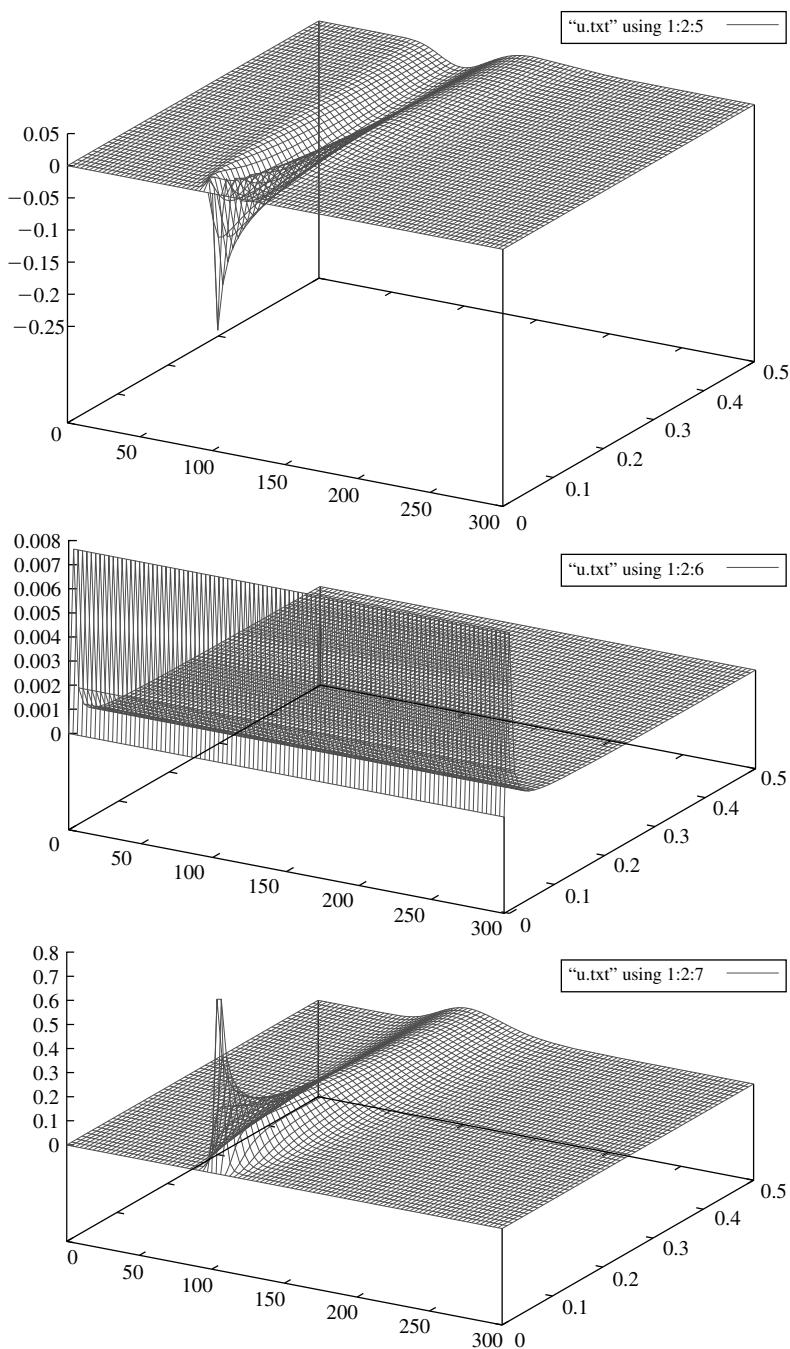


FIG. 3.1 The first graph displays the error between the computed solution on a uniform mesh and one computed with Black–Scholes formula. The second graph shows $\rho_{\delta t}$; the third graphs shows $\eta_{n,\omega}$ as a function of x and t . The parameters are $K = 100$, $T = 0.5$, $r = 0$, $\sigma = 0.3$, 50 time steps and 100 mesh points.

Multidimensional Partial Differential Equations For Option Pricing

4. European basket options

We consider d risky assets whose prices at time t are called $S_{i,t}$, $i = 1, \dots, d$. We assume that for all i , $1 \leq i \leq d$, $S_{i,t}$ satisfies the stochastic differential equation

$$dS_{i,t} = S_{i,t}(\mu_i dt + \sigma_i dW_{i,t}). \quad (4.1)$$

Here,

- $(W_{i,t})$, $1 \leq i \leq d$, are possibly correlated standard Brownian motions on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We call $\rho_{i,j}$ the correlation factor of $(W_{i,t})$ and $(W_{j,t})$. We have $-1 \leq \rho_{i,j} \leq 1$. Of course, $\rho_{i,i} = 1$.
- the volatilities σ_i , $1 \leq i \leq d$, are positive constants.

The Black–Scholes model assumes the existence of a risk-free asset whose interest rate r may be supposed constant for simplicity; yet, it is possible to generalize what follows to the case when the volatilities are sufficiently regular functions of t and of the prices $S_{j,t}$, $j = 1, \dots, d$ and when r is a nonnegative function of t .

In what follows, the notation S is used for the vector $(S_1, \dots, S_d)$. Consider a European option on a basket containing the above mentioned d risky assets, whose maturity is T and whose payoff function is $S \mapsto P_o(S)$. As for an option on a single asset, it is possible to find a risk-neutral probability $\mathbb{P}^*$ under which the price of the option at time t is

$$P_t = e^{-r(T-t)} \mathbb{E}^*(P_o(S_{1,T}, \dots, S_{d,T}) | F_t). \quad (4.2)$$

The payoff functions actually used in finance may be quite complicated. Among the simplest ones, we can list

- the payoff is a function of a weighted sum of the assets prices:
 - call option on a weighted sum: $P_o(S) = \left(\sum_{i=1}^d \alpha_i S_i - K \right)^+$
 - put option on a weighted sum: $P_o(S) = \left(K - \sum_{i=1}^d \alpha_i S_i \right)^+$

Calling P_t the price of the put option and C_t the price of the call option, the put-call parity $C_t - P_t = \sum_{i=1}^d \alpha_i S_{i,t} - Ke^{-r(T-t)}$ can be proved either by using (4.2) or by arguments on the PDEs used for pricing the options (see 4.1).

- the payoff is a function of $\max_{i=1,\dots,d} S_i$: these options are called best-of options.

– best-of call option: $P_o(S) = (\max_{i=1,\dots,d} S_i - K)^+$

– best-of put option: $P_o(S) = (K - \max_{i=1,\dots,d} S_i)^+$.

In contrast to the previous case, there is no put-call parity for these two options. Payoff functions depending on $\min_{i=1,\dots,d} S_i$ are used as well.

4.1. The partial differential equation

DEFINITION 4.1. Let Ω be an open subset of $\mathbb{R}^d$. A function $f : \Omega \times (0, T) \rightarrow \mathbb{R}$, continuous and such that its partial derivatives $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial S_i}$ and $\frac{\partial^2 f}{\partial S_i \partial S_j}$, $i, j = 1, \dots, d$, exist and are continuous on $\Omega \times (0, T)$ is said to belong to the class $\mathcal{C}^{2,1}(\Omega \times (0, T))$. Furthermore if f and the above-mentioned partial derivatives have continuous extensions in $\Omega \times [0, T)$, it is said that $f \in \mathcal{C}^{2,1}(\Omega \times [0, T))$.

It is possible to relate the option's price to the solution of a parabolic PDE with $1 + d$ variables. The partial differential operator appears naturally in the following result.

PROPOSITION 4.1. *Call L the partial differential operator with variable coefficients:*

$$Lf = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \Xi_{i,j} S_i S_j \frac{\partial^2 f}{\partial S_i \partial S_j} + r \sum_{i=1}^d S_i \frac{\partial f}{\partial S_i}, \quad (4.3)$$

with

$$\Xi_{i,j} = \sigma_i \sigma_j \rho_{i,j}. \quad (4.4)$$

For any function $u : (S, t) \mapsto u(S, t)$, $u \in \mathcal{C}^{2,1}(\mathbb{R}_+^d \times [0, T))$ and such that $|S_i \frac{\partial u}{\partial S_i}| \leq C(1 + |S_i|)$, $i = 1, \dots, d$, with C independent of t , the process

$$M_t = e^{-rt} u(S_{1,t}, \dots, S_{d,t}, t) - \int_0^t e^{-r\tau} \left(\frac{\partial u}{\partial t} + Lu - ru \right) (S_{1,\tau}, \dots, S_{d,\tau}, \tau) d\tau$$

is a martingale under $\mathbb{P}^*$. The partial differential operator L is called the infinitesimal generator of the Markov family $(S_{1,t}, \dots, S_{d,t})$.

PROOF. The proof uses the multidimensional Itô's formula (see KARATZAS and SHREVE [1991], LAMBERTON and LAPEYRE [1997]). $\square$

THEOREM 4.1. Consider a continuous function $P \in \mathcal{C}^{2,1}(\mathbb{R}_+^d \times [0, T])$ such that $|S_i \frac{\partial P}{\partial S_i}| \leq C(1 + S_i)$, with C independent of t . Assume that P satisfies

$$\left(\frac{\partial P}{\partial t} + LP - rP \right) (S, t) = 0, \quad t < T, \quad S \in \mathbb{R}_+^d \quad (4.5)$$

and

$$P(S, T) = P_\circ(S), \quad S \in \mathbb{R}_+^d, \quad (4.6)$$

then the price of the European option given by (4.2) satisfies

$$P_t = P(S_{1,t}, \dots, S_{d,t}, t). \quad (4.7)$$

REMARK 4.1. All the preceding results can be generalized to the case when r is a bounded and continuous function on $[0, T]$, and when the drifts and volatilities are functions of the variables S and t such that

1. For all $i = 1, \dots, d$, $(S, t) \mapsto \mu_i(S, t)$ and $(S, t) \mapsto \sigma_i(S, t)$ are bounded and continuous functions.
2. For all $i = 1, \dots, d$, $S \mapsto \mu_i(S, t)S_i$ and $S \mapsto \sigma_i(S, t)S_i$ are Lipschitz continuous with a Lipschitz constant C independent of t .

In this case, $\Xi_{i,j}$ and r in (4.3) are functions instead of parameters: $\Xi_{i,j}(S, t) = \rho_{i,j}\sigma_i(S, t)\sigma_j(S, t)$ and $r = r(t)$.

4.1.1. A first change of variables

It is possible to make the change of variable $x_i = \log(S_i)$, $i = 1, \dots, d$; calling $\mathbf{x}$ the vector $(x_1, \dots, x_d)$ and $\tilde{P}$ the function, $\tilde{P}(\mathbf{x}, t) = P(S, t)$, the final value problem (4.5) (4.6) becomes

$$\begin{aligned} \left(\frac{\partial \tilde{P}}{\partial t} + \tilde{L}\tilde{P} - r\tilde{P} \right) (\mathbf{x}, t) &= 0, \quad t < T, \quad \mathbf{x} \in \mathbb{R}^d, \\ \tilde{P}(\mathbf{x}, T) &= P_\circ(e^{x_1}, \dots, e^{x_d}), \quad \mathbf{x} \in \mathbb{R}^d, \end{aligned} \quad (4.8)$$

where

$$\tilde{L}f = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \Xi_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^d \left(r - \frac{\sigma_i^2}{2} \right) \frac{\partial f}{\partial x_i}. \quad (4.9)$$

One sees that when the volatilities and the interest rate are constant, the operator $\tilde{L}$ has constant coefficients. In that case, calling $\mathbf{v}$ the vector of $\mathbb{R}^d$ such that $v_i = \sigma_i^2/2 - r$

and $\bar{P}(\mathbf{y}, t) = e^{rt} \tilde{P}(\mathbf{y} + t\mathbf{v}, T - t)$, we have

$$\begin{aligned} \frac{\partial \bar{P}}{\partial t} - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \Xi_{i,j} \frac{\partial^2 \bar{P}}{\partial y_i \partial y_j} &= 0, \quad 0 < t \leq T, \quad \mathbf{y} \in \mathbb{R}^d, \\ \bar{P}(\mathbf{y}, 0) &= P_{\circ}(e^{y_1}, \dots, e^{y_d}), \quad \mathbf{y} \in \mathbb{R}^d. \end{aligned} \quad (4.10)$$

We make the following assumption: there exists a unitary matrix Θ such that

$$\begin{aligned} \Theta^T \Xi \Theta &= A \text{ is a diagonal positive matrix:} \\ A &= \text{Diag}(a_1, \dots, a_d). \end{aligned} \quad (4.11)$$

Consider the change of variables

$$\mathbf{z} = \sqrt{2} A^{-\frac{1}{2}} \Theta^T \mathbf{y},$$

then the function $\bar{\bar{P}}(\mathbf{z}, t) = \bar{P}(\mathbf{y}, t)$ satisfies the heat equation

$$\frac{\partial \bar{\bar{P}}}{\partial t} - \Delta \bar{\bar{P}} = 0, \quad 0 < t \leq T, \quad \mathbf{z} \in \mathbb{R}^d. \quad (4.12)$$

From this, calling G the fundamental solution of the heat equation,

$$G(\mathbf{z}, t) = (4\pi t)^{-\frac{d}{2}} e^{-\frac{|\mathbf{z}|^2}{4t}},$$

we have

$$\begin{aligned} e^{r(T-t)} P(\mathbf{S}, t) &= \\ \sqrt{2^d \prod_{i=1}^d a_i} \int_{\mathbb{R}^d} G\left(\sqrt{2} A^{-\frac{1}{2}} \Theta^T (\log(\mathbf{S}) - \mathbf{z} - (T-t)\mathbf{v}), T-t\right) P_{\circ}(e^{z_1}, \dots, e^{z_d}) d\mathbf{z}, \end{aligned} \quad (4.13)$$

where $\log(\mathbf{S}) = (\log(S_1), \dots, \log(S_d))^T$.

Assume that $P_{\circ}$ has a compact support contained in the rectangular domain $[0, \bar{S}_{\circ})^d$, with $\bar{S}_{\circ} > 1$ and that there exists a constant K such that $0 \leq P_{\circ} \leq K$. The identity (4.13) gives information on the decay of $P(\mathbf{S}, t)$ when $t \in [0, T]$ and $\max_{i=1, \dots, d} S_i$ tends to ∞ .

Let us introduce

$$\bar{a} = \max(a_1, \dots, a_d), \quad (4.14)$$

and let us assume that $\max_{i=1,\dots,d} S_i \geq \bar{S} \gg 1$, where $\bar{S}$ is a positive number. Let $\mathbf{z} \in \mathbb{R}^d$ be such that $(e^{z_1}, \dots, e^{z_d})$ belongs to the support of $P_\circ$. Then, $-\infty < z_i \leq \log(\bar{S}_\circ)$ and

$$\begin{aligned} P(\mathbf{S}, t) &\leq K \sqrt{2^d \prod_{i=1}^d a_i} \prod_{i=1}^d \int_{-\infty}^{\log(\bar{S}_\circ)} \frac{e^{-\frac{2(\log(S_i) - z_i - (T-t)v_i)^2}{4a(T-t)}}}{(4\pi(T-t))^{\frac{1}{2}}} dz_i \\ &\leq K \bar{a}^d \prod_{i=1}^d \int_{\frac{\log(S_i) - \log(\bar{S}_\circ) - (T-t)v_i}{\sqrt{a(T-t)}}}^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du. \end{aligned}$$

Let us assume that

$$\log(\bar{S}) > \log(\bar{S}_\circ) + T|\mathbf{v}|_\infty. \quad (4.15)$$

From the assumptions on $\mathbf{S}$, there exists at least one index ℓ , $1 \leq \ell \leq d$, such that $S_\ell \geq \bar{S}$; therefore,

$$\begin{aligned} P(\mathbf{S}, t) &\leq K \bar{a}^d \int_{\frac{\log(\bar{S}_\circ) + (t-T)v_\ell - \log(S_\ell)}{\sqrt{a(T-t)}}}^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \\ &\leq K \bar{a}^d \int_{\frac{\log(\bar{S}) - \log(\bar{S}_\circ) - T|\mathbf{v}|_\infty}{\sqrt{aT}}}^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \\ &\leq \frac{K}{2} \bar{a}^d e^{-\frac{(\log(\bar{S}/\bar{S}_\circ) - T|\mathbf{v}|_\infty)^2}{2aT}}. \end{aligned} \quad (4.16)$$

Solutions of (4.5) (4.6) In the previous paragraph, we have seen that if the coefficients are constant and if assumption (4.11) is satisfied, then (4.5) (4.6) has a solution given by (4.13).

In this paragraph, we give a classical existence and uniqueness result for the final value problem (4.5), (4.6) in the general case when $r = r(t)$ and $\sigma_i = \sigma_i(\mathbf{S}, t)$, $i = 1, \dots, d$. It is necessary to restrict the growth of the solutions when $S_i \rightarrow 0$ or $S_i \rightarrow +\infty$. Here, we will impose that the solution is bounded, but this restriction can be relaxed (e.g., depending on $P_\circ$, one can look for solutions with linear growth as $S_i \rightarrow +\infty$).

THEOREM 4.2. *Under the following regularity assumptions on the coefficients and the final value,*

1. *for $i, j = 0, \dots, d$, the real-valued function $\tilde{\Xi}_{i,j}$ defined on $\mathbb{R}^d \times [0, T]$*

$$\tilde{\Xi}_{i,j}(x_1, \dots, x_d, t) = \Xi_{i,j}(e^{x_1}, \dots, e^{x_d}, t)$$

belongs to $C^{\alpha, \alpha/2}(\mathbb{R}^d \times [0, T])$,

2. *the function $t \mapsto r(t)$ belongs to $C^{\alpha/2}([0, T])$,*
3. *the function $\tilde{P}_\circ$ defined on $\mathbb{R}^d$ by $\tilde{P}_\circ(x_1, \dots, x_d) = P_\circ(e^{x_1}, \dots, e^{x_d})$ is such that $\tilde{P}_\circ, \frac{\partial \tilde{P}_\circ}{\partial x_i}, \frac{\partial^2 \tilde{P}_\circ}{\partial x_i \partial x_j}$, $i, j = 1, \dots, d$ belong to $C^\alpha(\mathbb{R}^d)$,*

and under the following ellipticity assumption, there exists a positive constant c such that, for all $(x_1, \dots, x_d) \in \mathbb{R}^d$, $t \in [0, T]$, $\xi \in \mathbb{R}^d$,

$$\sum_{i=1}^d \sum_{j=1}^d \tilde{\Xi}_{i,j}(x_1, \dots, x_d, t) \xi_i \xi_j \geq c |\xi|^2,$$

the final value problem (4.8) has a unique solution $\tilde{P}$ such that the functions $\tilde{P}$, $\frac{\partial \tilde{P}}{\partial x_i}$, $\frac{\partial \tilde{P}}{\partial t}$, $\frac{\partial^2 \tilde{P}}{\partial x_i \partial x_j}$, $i, j = 1, \dots, d$, belong to $C^{\alpha, \alpha/2}(\mathbb{R}^d \times [0, T])$.

Under the previous assumptions except the one on $\tilde{P}_0$ and assuming that $\tilde{P}_0$ is a bounded function, (4.8) has a unique solution $\tilde{P} \in C^{2,1}(\mathbb{R}^d \times [0, T])$ such that for all $\tau < T$, the functions $\tilde{P}$, $\frac{\partial \tilde{P}}{\partial x_i}$, $\frac{\partial \tilde{P}}{\partial t}$, $\frac{\partial^2 \tilde{P}}{\partial x_i \partial x_j}$ belong to $C^{\alpha, \alpha/2}(\mathbb{R}^d \times [0, \tau])$.

This theorem is proved by LADYŽENSKAJA, SOLONNIKOV and URAL'CEVA [1967] (see also FRIEDMAN [1964], KRYLOV [1996]).

4.1.2. A second change of variables

For basket options with a payoff depending on the weighted sum $\sum_{i=1}^d \alpha_i S_i$, the following change of variables has been proposed by Reisinger REISINGER [2004], REISINGER and WITTUM [2007]:

$$\begin{aligned} y_1 &= \sum_{i=1}^d \alpha_i S_i \in \mathbb{R}_+, \\ y_i &= \frac{\alpha_{i-1} S_{i-1}}{\sum_{k=i-1}^d \alpha_k S_k} \in [0, 1], \quad 2 \leq i \leq d. \end{aligned} \tag{4.17}$$

This mapping is a C^1 diffeomorphism from $\mathbb{R}_+^d$ onto $\mathbb{R}_+ \times (0, 1)^{d-1}$. The inverse change of variables is

$$\begin{aligned} S_1 &= y_1 y_2 / \alpha_1, \\ S_i &= y_1 y_{i+1} / \alpha_i \prod_{k=2}^i (1 - y_k), \quad 2 \leq i \leq d-1, \\ S_d &= y_1 / \alpha_d \prod_{k=2}^d (1 - y_k). \end{aligned} \tag{4.18}$$

After some calculus, one sees that

$$\begin{aligned} S_i \frac{\partial}{\partial S_i} &= y_1 f_{1,i}(y) \frac{\partial}{\partial y_1} - \sum_{j=2}^d y_j (1 - y_j) f_{j,i}(y) \frac{\partial}{\partial y_j} \\ &\quad + y_{i+1} (1 - y_{i+1}) \frac{\partial}{\partial y_{i+1}}, \quad 1 \leq i < d, \\ S_d \frac{\partial}{\partial S_d} &= y_1 f_{1,d}(y) \frac{\partial}{\partial y_1} - \sum_{j=2}^d y_j (1 - y_j) f_{j,d}(y) \frac{\partial}{\partial y_j}, \end{aligned} \tag{4.19}$$

where

$$\begin{aligned}
 f_{j,i}(y) &= y_{i+1} \prod_{k=j+1}^i (1 - y_k), & j < i < d, \\
 f_{j,d}(y) &= \prod_{k=j+1}^d (1 - y_k), & j < d, \\
 f_{i,i}(y) &= y_{i+1}, & i < d, \\
 f_{d,d}(y) &= 1, \\
 f_{j,i}(y) &= 0, & i < j.
 \end{aligned} \tag{4.20}$$

The identities

$$\sum_{k=1}^d f_{i,k} = \sum_{k=i}^d f_{i,k} = 1, \quad \forall i = 1, \dots, d, \tag{4.21}$$

are true as well. One can verify that

$$\sum_{i=1}^d S_i \frac{\partial}{\partial S_i} = \frac{\partial}{\partial y_1}. \tag{4.22}$$

Playing with the identities (4.19), (4.20), (4.21), and (4.22), one obtains that in the new variables $y_1, \dots, y_d$, the PDE (4.5) becomes

$$\left(\frac{\partial \check{P}}{\partial t} + \check{L}\check{P} - r\check{P} \right) (y_1, \dots, y_d, t) = 0, \quad t < T, \quad (y_1, \dots, y_d) \in \mathbb{R}_+ \times (0, 1)^{d-1}, \tag{4.23}$$

where

$$\check{L}f = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \check{\Xi}_{i,j}(y) y_i y_j \frac{\partial^2 f}{\partial y_i \partial y_j} + \sum_{i=1}^d \beta_i(y) y_i \frac{\partial f}{\partial y_i}, \tag{4.24}$$

with

$$\begin{aligned}
 \check{\Xi}_{1,1}(y) &= \sum_{k,l=1}^d \Xi_{k,l} f_{1,k}(y) f_{1,l}(y), \\
 \check{\Xi}_{1,j}(y) &= \check{\Xi}_{j,1}(y) = (1 - y_j) \sum_{k,l=1}^d (\Xi_{k,j-1} - \Xi_{k,l}) f_{1,k}(y) f_{j,l}(y), \quad 1 < j \leq d, \\
 \check{\Xi}_{i,j}(y) &= (1 - y_i)(1 - y_j) \sum_{k,l=1}^d \left(\begin{pmatrix} \Xi_{k,l} + \Xi_{i-1,j-1} \\ -\Xi_{i-1,l} - \Xi_{k,j-1} \end{pmatrix} f_{i,k}(y) f_{j,l}(y) \right), \\
 &\hspace{15em} 1 < i, j \leq d,
 \end{aligned} \tag{4.25}$$

and

$$\begin{aligned} \beta_1(y) &= r, \\ \beta_i(y) &= (1 - y_i) \sum_{k,l=1}^d \left(\left(\begin{array}{c} 2(1 - y_i)\Xi_{k,l} - 2y_i\Xi_{i-1,i-1} \\ +(2y_i - 1)(\Xi_{i-1,l} + \Xi_{k,i-1}) \end{array} \right) f_{i,k}(y)f_{i,l}(y) \right) \\ &\quad 1 < i \leq d. \end{aligned} \tag{4.26}$$

No boundary conditions are necessary on $y_1 = 0$ and on $y_i = 0$ or 1 , $2 \leq i \leq d$, because the PDE is degenerate on these parts of the domain's boundary. The same remark can be made for the boundaries $y_i = 1$, $i = 2, \dots, d$, because $\check{\Xi}_{i,j}$ and $\check{\Xi}_{i,j}$ vanish, for $j = 1, \dots, d$.

The final condition becomes

$$\check{P}(y_1, \dots, y_d, T) = \check{P}_o(y_1).$$

For numerical computations (with d small enough), the main interest of such a change of variables is as follows: if the payoff function has a singularity on the hyperplane $\sum_{i=1}^d \alpha_i S_i = K$, then one needs to refine the mesh in the neighborhood of this hyperplane. This is incompatible with the use of Cartesian grids in the S variables. After the change of variables $S \mapsto y$, one can use a Cartesian grid in the $y_1, \dots, y_d$ variables, with a refinement in the neighborhood of $y_1 = K$. The price to pay is that the operator in the PDE has variable coefficients with rather complicated expressions. As we will see below, a more general and robust alternative is to use finite elements with adaptive mesh refinement.

4.1.3. Consequences of the maximum principle

The maximum principle may be applied to (4.5). This has many consequences.

The superreplication principle holds for basket options: take two European put options on the above-mentioned basket with the same maturity and different payoff functions P_o and Q_o . Call P and Q their respective pricing functions, both of which satisfy (4.5). We assume that P_o and Q_o have reasonable growth at infinity, that is, $\lim_{|S| \rightarrow \infty} \max(P_o(S), Q_o(S))e^{-\epsilon \log^2(|S|+2)} = 0$ for all $\epsilon > 0$. This is always satisfied in practice. Applying the maximum principle to (4.5), one sees that if $P_o(S) \leq Q_o(S)$ for all S , then for all $t \leq T$ and S , $P(S, t) \leq Q(S, t)$.

For example, if P_t is the price of the best-of put option with maturity T and payoff function $(K - \max_{j=1, \dots, d} S_j)^+$, then for all i , $1 \leq i \leq d$, and for all time $t < T$, we have

$$\tilde{Q}_{i,t} \geq P_t \geq Q_t, \tag{4.27}$$

where Q_t is the price of the put option on the weighted sum with the same maturity T and payoff function $(K - \sum_{j=1}^d S_j)^+$, and $\tilde{Q}_{i,t}$ is the price of the put option on the single asset indexed by i , with maturity and with payoff function $(K - S_i)^+$.

The maximum principle also gives information on the behavior of the pricing function at the boundary of $\mathbb{R}_+^d$ where the operator is degenerate. It can be proved that for a

put option on a weighted sum (i.e., $P_o(\mathbf{S}) = (K - \sum_{i=1}^d \alpha_i S_i)^+$), the pricing function satisfies

$$P(S_1, \dots, S_{i-1}, 0, S_{i+1}, \dots, S_d, t) = Q(S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_d, t),$$

where Q is the pricing function of the put option on the basket containing all the previous assets but the one indexed by i , with payoff function

$$Q_o(S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_d) = \left(K - \sum_{j=1, j \neq i}^d \alpha_j S_j \right)^+.$$

The same is true for the best-of put option with payoff function ($P_o(\mathbf{S}) = (K - \max_{i=1, \dots, d} S_i)^+$): we have

$$P(S_1, \dots, S_{i-1}, 0, S_{i+1}, \dots, S_d, t) = Q(S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_d, t),$$

where Q is the pricing function of the best-of put option on the basket containing all the previous assets but the one indexed by i , with payoff function

$$Q_o(S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_d) = \left(K - \max_{j=1, j \neq i}^d S_j \right)^+.$$

4.2. Variational formulation

Here, we assume that the volatilities are functions of the variables $\mathbf{S}$, t , and r is a function of time.

In order to obtain both global energy estimates and a nice framework for studying more complex situations (i.e., for American options on the same basket), we aim at finding a variational formulation for (4.5) (4.6), assuming that the payoff function is square integrable. For more general payoff functions, variational formulation can be proposed as well with suitable weighting functions as $|\mathbf{S}| \rightarrow \infty$. Variational formulations are also the cornerstone for the finite-element method (see Section 6.2).

The change of variable $T - t \rightarrow t$ yields a forward in time parabolic initial value problem:

$$\begin{aligned} \left(\frac{\partial P}{\partial t} - LP + rP \right) (\mathbf{S}, t) &= 0, & 0 < t \leq T, \mathbf{S} \in \mathbb{R}_+^d, \\ P(\mathbf{S}, 0) &= P_o(\mathbf{S}), & \mathbf{S} \in \mathbb{R}_+^d, \end{aligned} \quad (4.28)$$

where L is given by (4.3) with $\Xi_{i,j}(\mathbf{S}, t) = \rho_{i,j} \sigma_i(\mathbf{S}, t) \sigma_j(\mathbf{S}, t)$. Assuming that the coefficients are regular enough (this will be made clear below), we write the operator in divergence form:

$$Lu = \frac{1}{2} \sum_{i=1}^d \frac{\partial}{\partial S_i} \left(\sum_{j=1}^d \Xi_{i,j} S_i S_j \frac{\partial u}{\partial S_j} \right) + \sum_{j=1}^d \left(r(t) S_j - \frac{1}{2} \sum_{i=1}^d \frac{\partial}{\partial S_i} (\Xi_{i,j} S_i S_j) \right) \frac{\partial u}{\partial S_j}. \quad (4.29)$$

Multiplying $-Lu + ru$ by a test function v , integrating on $\mathbb{R}_+^d$ and performing suitable integrations by part, one obtains the bilinear form

$$\begin{aligned} a_t(u, v) = & \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_{\mathbb{R}_+^d} \Xi_{i,j} S_i S_j \frac{\partial u}{\partial S_j} \frac{\partial v}{\partial S_i} \\ & - \sum_{j=1}^d \int_{\mathbb{R}_+^d} \left(r(t) S_j - \frac{1}{2} \sum_{i=1}^d \frac{\partial}{\partial S_i} (\Xi_{i,j} S_i S_j) \right) \frac{\partial u}{\partial S_j} v + r(t) \int_{\mathbb{R}_+^d} u v. \end{aligned} \quad (4.30)$$

We introduce the Hilbert space

$$V = \left\{ v : v \in L^2(\mathbb{R}_+^d), S_i \frac{\partial v}{\partial S_i} \in L^2(\mathbb{R}_+^d), i = 1, \dots, d \right\}, \quad (4.31)$$

with the norm

$$\|v\|_V = \left(\|v\|_{L^2(\mathbb{R}_+^d)}^2 + \sum_{i=1}^d \|S_i \frac{\partial v}{\partial S_i}\|_{L^2(\mathbb{R}_+^d)}^2 \right)^{\frac{1}{2}}, \quad (4.32)$$

one can check the following properties:

- The space $\mathcal{D}(\mathbb{R}_+^d)$ of smooth and compactly supported functions in $\mathbb{R}_+^d$ is dense in V .
- V is separable.
- The seminorm $|\cdot|_V$ defined by $|v|_V^2 = \sum_{i=1}^d \|S_i \frac{\partial v}{\partial S_i}\|_{L^2(\mathbb{R}_+^d)}^2$ is, in fact, a norm on V equivalent to $\|\cdot\|_V$ because $\|v\|_{L^2(\mathbb{R}_+^d)} \leq 2|v|_V$.

We make the following assumptions:

- The functions $\Xi_{i,j}$, $1 \leq i, j \leq d$, are bounded by a positive constant $\bar{\sigma}^2$ independent of S and t :

$$\|\Xi_{i,j}\|_{L^\infty(\mathbb{R}_+^d)} \leq \bar{\sigma}^2. \quad (4.33)$$

- There exists a positive constant $\underline{\sigma}$ such that for all $q \in \mathbb{R}^d$,

$$\sum_{i=1}^d \sum_{j=1}^d \Xi_{i,j} q_i q_j \geq \underline{\sigma}^2 |q|^2. \quad (4.34)$$

- There exists a positive constant M such that, for $i = 1, \dots, d$,

$$\left\| \sum_{i=1}^d \frac{\partial}{\partial S_i} (\Xi_{i,j} S_i) \right\|_{L^\infty(\mathbb{R}_+^d)} \leq M. \quad (4.35)$$

- The function r is nonnegative and bounded by a constant.

LEMMA 4.1. *With the above assumptions, the bilinear form a_t is continuous on $V \times V$, and there exists a constant $\bar{c}$ independent of t such that, for any $v, w \in V$,*

$$a_t(v, w) \leq \bar{c}|v|_V|w|_V. \quad (4.36)$$

Moreover, there is a uniform Gårding's inequality for a_t : there exist a positive constant $\underline{c}$ and a nonnegative constant λ such that, for any $v \in V$,

$$a_t(v, v) \geq \underline{c}|v|_V^2 - \lambda \|v\|_{L^2(\mathbb{R}_+^d)}^2. \quad (4.37)$$

From Lemma 4.1, we deduce the existence and uniqueness of a weak solution to the initial value problem (4.28).

THEOREM 4.3. *Under the above assumptions, for any $P_\circ \in L^2(\mathbb{R}_+^d)$, there exists a unique P in $L^2(0, T; V) \cap C^0([0, T]; L^2(\mathbb{R}_+^d))$, with $\frac{\partial P}{\partial t} \in L^2(0, T; V')$ such that, for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in V$,*

$$-\int_0^T \phi'(t) \left(\int_{\mathbb{R}_+^d} P(t)v \right) dt + \int_0^T \phi(t) a_t(P, v) dt = 0 \quad (4.38)$$

and

$$P(t=0) = P_\circ. \quad (4.39)$$

The mapping $P_\circ \mapsto P$ is continuous from $L^2(\mathbb{R}_+^d)$ to $L^2(0, T; V) \cap C^0([0, T]; L^2(\mathbb{R}_+^d))$, and we have the energy estimate

$$e^{-2\lambda t} \|P(t)\|_{L^2(\mathbb{R}_+^d)}^2 + 2\underline{c} \int_0^t e^{-2\lambda \tau} |P(\tau)|_V^2 d\tau \leq \|P_\circ\|_{L^2(\mathbb{R}_+^d)}^2. \quad (4.40)$$

Naturally, there may be barrier options on baskets. For a barrier independent of time, pricing the option then amounts to solving the boundary value problem

$$\begin{aligned} \left(\frac{\partial P}{\partial t} - LP + rP \right) (S, t) &= 0, & 0 < t \leq T, \quad S \in \Omega, \\ P(S, t) &= 0, & 0 < t \leq T, \quad S \in \partial\Omega \text{ s.t. } S_i > 0, i = 1, \dots, d, \\ P(S, 0) &= P_\circ(S), & S \in \Omega, \end{aligned} \quad (4.41)$$

for a domain Ω of $\mathbb{R}_+^d$. We restrict ourselves to domains whose boundaries are locally the graph of Lipschitz continuous functions. Then, the Sobolev space to work with is the closure of $\mathcal{D}(\Omega)$ in the space $\{v \in L^2(\Omega); S_i \frac{\partial v}{\partial S_i} \in L^2(\Omega), i = 1, \dots, d\}$ equipped with the norm

$$\left(\|v\|_{L^2(\Omega)}^2 + \sum_{i=1}^d \|S_i \frac{\partial v}{\partial S_i}\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}}.$$

5. Numerical methods for European basket options

5.1. Localization

Consider the initial value problem (4.28) for pricing the European basket option discussed in Section 4. For computing a numerical approximation to P , one may

- truncate the domain in the variable S : the pricing function will be computed for $S_i \in (0, \bar{S})$, with $\bar{S}$ large enough.
- impose some boundary condition at the artificial boundaries $S_i = \bar{S}$. Naturally, the choice of these boundary conditions has to depend on the payoff function P_o .

We introduce the rectangular domain $\Omega = (0, \bar{S})^d$ and Γ_0 the part of the boundary of Ω given by the following

$$\Gamma_0 = \{S \in \partial\Omega; \max_{i=1, \dots, d} S_i = \bar{S}\}. \quad (5.1)$$

Let us assume that P_o has a compact support contained in the rectangular domain $[0, \bar{S}_o)^d$ and there exists $K \in \mathbb{R}_+$ such that $0 \leq P_o(S) \leq K$ for $S \in \mathbb{R}_+^d$. This is the case for European put options on weighted sums or for best-of European put options. We choose $\bar{S}$ such that $\bar{S} > \bar{S}_o$. From the estimate (4.16), a natural choice of boundary condition is $P = 0$ on Γ_0 . The new boundary value problem becomes

$$\begin{aligned} \left(\frac{\partial P}{\partial t} - LP + rP\right)(S, t) &= 0, & 0 < t \leq T, \quad S \in \Omega, \\ P(S) &= 0, & 0 < t \leq T, \quad S \in \Gamma_0, \\ P(S, 0) &= P_o(S), & S \in \Omega, \end{aligned} \quad (5.2)$$

with L given by (4.3).

One can introduce a variational formulation for (5.2): the only modification to bring to the content of Section 4.2 is the choice of the space V , which must take into account the change in the domain and the new boundary conditions. We introduce the Hilbert space

$$\tilde{V} = \left\{ v : v \in L^2(\Omega), S_i \frac{\partial v}{\partial S_i} \in L^2(\Omega), i = 1, \dots, d \right\}, \quad (5.3)$$

with the norm $\|v\|_{\tilde{V}} = \left(\|v\|_{L^2(\Omega)}^2 + \sum_{i=1}^d \|S_i \frac{\partial v}{\partial S_i}\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}}$ and V as the completion in $\tilde{V}$ of the space of smooth functions with compact support in Ω . It can be proved that there is a continuous trace operator from $\tilde{V}$ to $L^2(\Gamma_0)$: for a function $v \in \tilde{V}$, we call $v|_{\Gamma_0} \in L^2(\Gamma_0)$ its trace on Γ_0 . The identity

$$V = \{v \in \tilde{V}; v|_{\Gamma_0} = 0\} \quad (5.4)$$

is a consequence of the continuity of the trace operator. Making the same assumptions on the volatilities σ_i and on r as in Section 4.2, we introduce the bilinear

form a_t on $V \times V$:

$$\begin{aligned} a_t(u, v) = & \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_{\Omega} \Xi_{i,j} S_i S_j \frac{\partial u}{\partial S_j} \frac{\partial v}{\partial S_i} \\ & - \sum_{j=1}^d \int_{\Omega} \left(r(t) S_j - \frac{1}{2} \sum_{i=1}^d \frac{\partial}{\partial S_i} (\Xi_{i,j} S_i S_j) \right) \frac{\partial u}{\partial S_j} v + r(t) \int_{\Omega} u v. \end{aligned} \quad (5.5)$$

It can be proved that a_t is continuous on $\tilde{V}$ and there is a Gårding's inequality in $\tilde{V}$ as in (4.37). The variational formulation for (5.2) is to find P in $L^2(0, T; V) \cap C^0([0, T]; L^2(\Omega))$ with $\frac{\partial P}{\partial t} \in L^2(0, T; V')$ such that for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in V$,

$$- \int_0^T \phi'(t) \left(\int_{\Omega} P(t) v \right) dt + \int_0^T \phi(t) a_t(P, v) dt = 0 \quad (5.6)$$

and

$$P(t=0) = P_{\circ}. \quad (5.7)$$

This problem has a unique solution.

5.1.1. Localization error

Of course, the artificial boundary conditions produce an error because the solution P_{exact} of (4.28) does not vanish on Γ_0 .

The maximum principle can be used for proving that

$$P(S, t) < P_{\text{exact}}(S, t), \quad \text{for } S \in \Omega, \quad \text{and } 0 < t \leq T.$$

It can also be proved, at least for the above-mentioned two examples of options, that the maximum of $P_{\text{exact}} - P$ in $\bar{\Omega} \times [0, T]$ is reached on $\Gamma_0 \times (0, T]$.

On the other hand, if the volatilities and the interest rate r are constant, we call v the vector of $\mathbb{R}^d$ whose components are $v_i = \sigma_i^2/2 - r$, and we define $\bar{a}$ by (4.14), (4.11); if $\bar{S}$ satisfies (4.15), then P_{exact} satisfies (4.16) for all $S \in \Gamma_0$. Thus,

$$\|P - P_{\text{exact}}\|_{L^\infty(\Omega \times (0, T))} \leq \|P_{\text{exact}}\|_{L^\infty(\Gamma_0 \times (0, T))} \leq \frac{K}{2} \bar{a}^d e^{-\frac{(\log(\bar{S}/\bar{S}_0) - T|v|_\infty)^2}{2\bar{a}T}}. \quad (5.8)$$

We see that the error between P and P_{exact} can be made arbitrarily small by letting $\bar{S}$ tend to infinity. In particular, the choice

$$\bar{S} \geq \bar{S}_0 \exp \left(T|v|_\infty + \sqrt{2\bar{a}T \log \left(\frac{K\bar{a}^d}{\epsilon} \right)} \right)$$

guarantees an error smaller than ϵ .

For nonconstant coefficients, obtaining such an accurate result is not as easy. Yet, the above formula may be used for a reasonable choice of $\bar{S}$.

For compactly supported payoff functions, the Neumann boundary conditions

$$\frac{\partial P}{\partial S_i} = 0, \quad \text{on } (\Gamma^0 \cap \{S_i = \bar{S}\})$$

can be used as well. Accurate error bounds may be obtained for put options on a weighted sum and for best-of put options if the coefficients are constant. The variational formulation with the Neumann boundary conditions is (5.6) (5.7), but now $V = \tilde{V}$.

For a call option on the weighted sum $\sum_{i=1}^d \alpha_i S_i$, the following conditions can be used:

- Dirichlet conditions: $P(S, t) = \sum_{i=1}^d \alpha_i S_i - Ke^{-rt}$ on $\Gamma^0 \times (0, T)$. With V given by (5.4), the variational formulation is to find P in $L^2(0, T; \tilde{V}) \cap C^0([0, T]; L^2(\Omega))$ with $\frac{\partial P}{\partial t} \in L^2(0, T; V')$ such that for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in V$,

$$-\int_0^T \phi'(t) \left(\int_{\Omega} P(t)v \right) dt + \int_0^T \phi(t) a_t(P, v) dt = 0, \quad (5.9)$$

$P(t)|_{\Gamma^0} = \sum_{i=1}^d \alpha_i S_i - Ke^{-rt}$ for a.e. t , and (5.7).

- Neumann conditions: $\sum_{j=1}^d \Xi_{i,j} S_j \frac{\partial P}{\partial S_j} = \sum_{j=1}^d \Xi_{i,j} S_j \alpha_j$ on $(\Gamma^0 \cap \{S_i = \bar{S}\}) \times (0, T)$. The variational formulation is to find P in $L^2(0, T; \tilde{V}) \cap C^0([0, T]; L^2(\Omega))$, with $\frac{\partial P}{\partial t} \in L^2(0, T; \tilde{V}')$ such that for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in \tilde{V}$,

$$\begin{aligned} & -\int_0^T \phi'(t) \left(\int_{\Omega} P(t)v \right) dt + \int_0^T \phi(t) a_t(P, v) dt \\ & = \frac{\bar{S}}{2} \int_0^T \phi(t) \left(\sum_{i=1}^d \int_{\Gamma^0 \cap \{S_i = \bar{S}\}} \left(\sum_{j=1}^d \Xi_{i,j} S_j \alpha_j \right) v \right) dt. \end{aligned} \quad (5.10)$$

and (5.7).

The error due to artificial boundary conditions can be accurately estimated in the case of constant coefficients by using the previously obtained estimates for the corresponding put options and the put-call parity.

For a best-of call option, finding reasonable boundary condition near the regions $S_i = S_j = \bar{S}$, $i \neq j$ is much more difficult. One may have to use an alternative option to artificial boundary conditions, that is, a change of variables, which maps the unbounded domain to a bounded one; one obtains a new boundary value problem in a bounded domain. The PDE becomes degenerate on the part of the boundary that is sent to infinity by the inverse mapping, thus no boundary condition is needed there. An example of such a program is given in Section 7.2 below in the context of option pricing with stochastic volatility.

5.2. Finite-element methods

Conforming FEM are numerical approximations closely linked to the theory of variational or weak formulations presented in Section 4.2. The first FEM can be attributed to COURANT [1943].

Conforming FEM have the same framework in any dimension of space d : for a weak formulation posed in an infinite-dimensional function space V , for example (5.10)–(5.7), it consists of choosing a finite-dimensional subspace V_h of V , for instance, the space of continuous piecewise affine functions on a triangulation of Ω , and of solving the problem with test and trial functions in V_h instead of V . We speak of conforming methods because $V_h \subset V$. Nonconforming methods, that is, $V_h \not\subset V$ are possible too, but we will not consider this topic here. In the simpler FEM, the construction of the space V_h is done as follows:

- The domain is partitioned into nonoverlapping cells (elements) whose shapes are simple and fixed: for example, intervals in one dimension, triangles or quadrilaterals in two dimensions, tetrahedra, prisms, or hexahedra in three dimensions. The set of the elements is, in general, an unstructured mesh called a triangulation.
- The maximal degree k of the polynomial approximation in the elements is chosen.
- V_h is made of continuous functions of V whose restriction to the elements is polynomial of degree less than k .

Programming the method is also somewhat similar in any dimension, but mesh generation is very much dimension dependent. A nice survey on the FEM, both on the theoretical and practical viewpoints, is proposed by ERN and GUERMOND [2004].

There is a very well-understood theory on error estimates for finite elements. It is possible to distinguish a priori and a posteriori error estimates: in a priori estimates, the error is bounded by some quantity depending on the solution to the continuous problem (which is unknown, but for which estimates are available), whereas in a posteriori estimates, the error is bounded by some quantity depending on the solution to the discrete problem, which is available.

For a priori error estimates, one can see the books of RAVIART and THOMAS [1983], STRANG and FIX [1973], BRAESS [2001], BRENNER and SCOTT [1994], CIARLET [1978, 1991], and THOMÉE [1997] for parabolic problems. By and large, deriving error estimates for FEM consists of

1. establishing the stability of the discretization with respect to some norms related to $\|\cdot\|_V$.
2. Once this is done, one sees (at least in simple cases) that the error depends on some distance of the solution to the continuous problem to the space V_h . This quantity cannot be computed exactly since the solution is unknown. However, it can be estimated from a priori knowledge on the regularity of the solution.

When accurate results on the solution to the continuous problem are available, the a priori estimates give very valuable information on how to choose the discretization a priori (see SCHÖTZAU and SCHWAB [2001], WERDER, GERDES, SCHÖTZAU and SCHWAB [2001]), in the case of homogeneous parabolic problems with smooth coefficients.

A posteriori error estimates are a precious tool since they give practical information that can be used to refine the mesh when needed. The bibliography on a posteriori error estimates for FEM is quite rich: one can see the book of VERFÜRTH [1996] and the references therein. For time-dependent problems, a posteriori error estimates and mesh adaption for space-time finite-element problems have been investigated by ERIKSSON, ESTEP, HANSBO and JOHNSON [1995], ERIKSSON and JOHNSON [1991, 1995]. Another strategy based on decoupled space and time error indicators can be implemented (see BERGAM, BERNARDI and MGHAZLI [2005] and Section 3 for an example with one space variable).

When the space variable is multidimensional, very anisotropic meshes may be useful. A trend in mesh adaptivity consists of building anisotropic meshes by imposing regularity and quasi uniformity with respect to a new metric constructed from the a posteriori error estimates (see GEORGE, HECHT and SALTEL [1991]). We show some examples of anisotropic meshes generated with the open-source software BAMG (a bidimensional anisotropic mesh generator) of GEORGE, HECHT and SALTEL [1991].

Example: the case of a put option on a basket of two assets In the sequel, we deal with a simple implementation of the FEM for approximating the pricing function of an option on a basket containing two assets. Therefore, $d = 2$.

5.2.1. The time semidiscrete problem

We introduce a partition of the interval $[0, T]$ into subintervals $[t_{m-1}, t_m]$, $1 \leq m \leq M$, such that $0 = t_0 < t_1 < \dots < t_M = T$. We denote by δt_m the length $t_m - t_{m-1}$ and by δt the maximum of the δt_m , $1 \leq m \leq M$.

For simplicity, we assume that $P_o \in V$, where V is given by (5.4) with $d = 2$.

We discretize (5.6) by means of an implicit Euler scheme, that is, we look for $P^m \in V$, $m = 0, \dots, M$ such that $P^0 = P_o$, and for $m = 1, \dots, M$, $\forall v \in V$,

$$\frac{1}{\delta t_m} (P^m - P^{m-1}, v)_{L^2(\Omega)} + a_{t_m}(P^m, v) = 0, \quad (5.11)$$

where a_{t_m} is given by (5.5). This scheme is first order.

REMARK 5.1. If P_o does not belong to V , then we first have to approximate P_o by a function in V , at the cost of an additional error.

5.2.2. The full discretization: Lagrange finite elements

Discretization with respect to S_1 and S_2 consists of replacing V with a finite-dimensional subspace $V_h \subset V$. For example, one may choose V_h as a space of continuous piecewise polynomial functions on a triangulation of Ω : for a positive real number h , consider a partition $\mathcal{T}_h$ of Ω into nonoverlapping closed triangles ($\mathcal{T}_h$ is the set of all the triangles forming the partition) such that

- $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K$,
- for all $K \neq K'$, two triangles of $\mathcal{T}_h$, $K \cap K'$ is empty, a vertex of both K and K' , or a whole edge of both K and K' .

REMARK 5.2. If Ω is not polygonal but has a smooth boundary, it is possible to find a set $\mathcal{T}_h$ of nonoverlapping triangles of diameters less than h such that the distance between $\bar{\Omega}$ and $\cup_{K \in \mathcal{T}_h} K$ scales like h^2 .

For a positive integer k , we introduce the spaces

$$W_h = \{w_h \in C^0(\bar{\Omega}) : w_h|_K \in \mathcal{P}^k, \forall K \in \mathcal{T}_h\}, \quad V_h = \{v_h \in W_h, v_h|_{\Gamma_0} = 0\}. \quad (5.12)$$

Then, we focus on the case when $k = 1$, that is, the functions in W_h are piecewise affine. It is clear that V_h is a finite-dimensional subspace of V .

Assuming that $P_\circ \in V_h$, the full discretization of the variational formulation consists of finding $P_h^m \in V_h$, $m = 0, \dots, M$, such that $P_h^0 = P_\circ$ and

$$\forall v_h \in V_h, \quad \frac{1}{\delta t_m} (P_h^m - P_h^{m-1}, v_h)_{L^2(\Omega)} + a_{t_m}(P_h^m, v_h) = 0. \quad (5.13)$$

Here, for simplicity, we assume that $a_{t_m}(u_h, v_h)$ can be computed algebraically for $u_h, v_h \in V_h$, which is the case when the volatilities do not depend on S_1 and S_2 , for example. If this is not the case, then quadrature formulas have to be used, which induce an additional but controlled source of error.

5.2.3. The discrete problem in matrix form

A basis of V_h is chosen, $(w_i)_{i=1, \dots, N}$. Then, for $1, \dots, M$, u_h^m can be written as

$$u_h^m(S_1, S_2) = \sum_{j=1}^N u_j^m w_j(S_1, S_2), \quad (5.14)$$

and using (5.14) in (5.13) with $v_h = w_i$, we obtain a system of linear equations for $\mathbf{U}^m = (u_j^m)_{j=1, \dots, N}^T$:

$$\mathbf{M}(\mathbf{U}^m - \mathbf{U}^{m-1}) + \delta t_m \mathbf{A}^m \mathbf{U}^m = 0, \quad (5.15)$$

where $\mathbf{M}$ and $\mathbf{A}$ are matrices in $\mathbb{R}^{N \times N}$, and assuming that the volatilities do not depend on S_1 and S_2 ,

$$\begin{aligned} M_{ij} &= \int_{\Omega} w_i w_j, \\ A_{i,j}^m &= a(w_j, w_i) = \frac{1}{2} \sum_{\ell=1}^2 \sum_{k=1}^2 \int_{\Omega} \Xi_{\ell,k}(t_m) S_{\ell} S_k \frac{\partial w_j}{\partial S_k} \frac{\partial w_i}{\partial S_{\ell}} \\ &\quad - \sum_{k=1}^2 \int_{\Omega} \left(r(t_m) S_k - \frac{1}{2} \sum_{\ell=1}^2 \frac{\partial}{\partial S_{\ell}} (\Xi_{\ell,k}(t_m) S_{\ell} S_k) \right) \frac{\partial w_j}{\partial S_k} w_i + r(t_m) \int_{\Omega} w_j w_i. \end{aligned} \quad (5.16)$$

The matrix $\mathbf{M}$ is called the mass matrix and $\mathbf{A}^m$ is called the stiffness matrix. It can be proved thanks to estimates (4.36)–(4.37) that if δt is small enough, then $\mathbf{M} + \delta t_m \mathbf{A}^m$ is invertible, and it is possible to solve (5.13).

The Nodal Basis On each triangle $K \in \mathcal{T}_h$, noting by $q^i, i = 1, 2, 3$ the vertices of K , we define for $S \in \mathbb{R}^2$ the barycentric coordinates of S , that is, the solution to

$$\sum_{i=1,2,3} \lambda_i^K(S) q^i = S, \quad \sum_{i=1,2,3} \lambda_i^K(S) = 1.$$

This 3×3 system of linear equations is never singular because its determinant is twice the area of K . It is obvious that the barycentric coordinates λ_i^K are affine functions of S . Furthermore,

- when $S \in K$, $\lambda_i^K \geq 0, i = 1, 2, 3$,
- if $K = [q^{i_1}, q^{i_2}, q^{i_3}]$ and S is aligned with q^{i_1}, q^{i_2} then, $\lambda_{i_3}^K = 0$.

Let v_h be a function in V_h : it is easy to check that, on each triangle $K \in \mathcal{T}_h$,

$$v_h(S) = \sum_{j=1,2,3} v_h(q^{i_j}) \lambda_{i_j}^K(S) \quad \forall S \in K.$$

Therefore, a function in V_h is uniquely defined by its values at the nodes of $\mathcal{T}_h$ not located on Γ_0 .

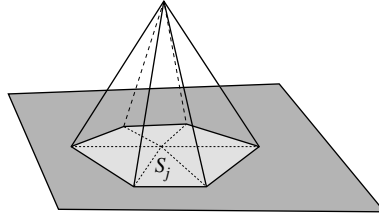
Call $(q^i)_{i=1,\dots,N}$ the nodes of $\mathcal{T}_h$ not located on Γ_0 , and let w^i be the unique function in V_h such that $w^i(q^j) = \delta_{i,j}, \forall j = 1, \dots, N$. For a triangle K such that q^i is a vertex of K , it is clear that w^i coincides in K with one of the three barycentric coordinates attached to triangle K . Therefore, we have the identity

$$v_h = \sum_{i=1}^N v_h(q^i) w^i, \quad (5.17)$$

which shows that $(w^i)_{i=1,\dots,N}$ is a basis of V_h . As shown in Fig. 5.1, the support of w^i is the union of the triangles of $\mathcal{T}_h$ containing the node q^i , so it is very small when the mesh is fine, and the support of two basis functions w^i and w^j intersects if and only if q^i and q^j are the vertices of a same triangle of $\mathcal{T}_h$. Therefore, the matrices $\mathbf{M}$ and $\mathbf{A}^m$ constructed with this basis are sparse. This dramatically reduces the complexity when solving properly (5.15). The basis $(w^i)_{i=1,\dots,N}$ is often called the nodal basis of V_h . The shape functions w^i are sometimes called hat functions. For $v_h \in V_h$, the values $v_i = v_h(q^i)$ are called the degrees of freedom of v_h . If $K = [q^{i_1}, q^{i_2}, q^{i_3}]$ and if b^{i_1} is the point aligned with q^{i_2} and q^{i_3} and such that $b^{i_1} \vec{q}^{i_1} \perp q^{i_2} \vec{q}^{i_3}$, then

$$\nabla \lambda_{i_1}^K = \frac{1}{|b^{i_1} \vec{q}^{i_1}|^2} b^{i_1} \vec{q}^{i_1}, \quad (5.18)$$

and calling n^{i_1} the unit vector orthogonal to $q^{i_2} \vec{q}^{i_3}$ and pointing to q^{i_1} , that is, $n^{i_1} = \frac{1}{|b^{i_1} \vec{q}^{i_1}|} b^{i_1} \vec{q}^{i_1}$ and E^{i_1} the length of the edge of K opposite to q^{i_1} , and using the well-known

FIG. 5.1 The shape function w^j .

identity $E^{i_1} |b^{i_1} \vec{q}^{i_1}| = 2|K|$, we obtain

$$\nabla \lambda_{i_1}^K = \frac{E^{i_1}}{2|K|} n^{i_1}. \quad (5.19)$$

The following integration formula is very important for the numerical implementation of the FEM:

PROPOSITION 5.1. *Calling λ_i , $i = 1, 2, 3$, the barycentric coordinates of the triangle K , and v_1 , v_2 , and v_3 , three nonnegative integers and $|K|$, the measure of K ,*

$$\int_K (\lambda_1^K)^{v_1} (\lambda_2^K)^{v_2} (\lambda_3^K)^{v_3} = 2|K| \frac{v_1! v_2! v_3!}{(v_1 + v_2 + v_3 + 2)!}. \quad (5.20)$$

REMARK 5.3. It may be useful to use other bases than the nodal basis, for example, bases related to wavelet decompositions, in particular, for speeding up the solution of (5.15) (see MATACHE, VON PETERSDOFF and SCHWAB [2004], VON PETERSDOFF and SCHWAB [2004]).

REMARK 5.4. The integral of a quadratic function on a triangle K is one-third the sum of the values of the function on the mid-edges times $|K|$; therefore, (5.20) is simpler when $v_1 + v_2 + v_3 = 2$:

$$\int_K \lambda_i^K \lambda_j^K = \frac{|K|}{12} (1 + \delta_{ij}). \quad (5.21)$$

When the system (5.15) becomes large, iterative methods such as gradient methods, GMRES (generalized minimum residual method) or BICG-stab (stabilized biconjugate gradient method) become attractive. We refer to AXELSSON [1994], GOLUB and VAN LOAN [1989], GREENBAUM [1997], MEURANT [1999], SAAD [1996] for good books on this topic. Iterative methods do not need the matrix $\mathbf{M} + \delta t_m \mathbf{A}^m$ but only a function that implements $\mathbf{U} \rightarrow (\mathbf{M} + \delta t_m \mathbf{A}^m) \mathbf{U}$, that is, which computes

$$\sum_j u_j ((w^j, w^i)_{L^2(\Omega)} + \delta t_m a_{t_m}(w^j, w^i)).$$

Let us show how $A^m U$ should be computed (we take $A^m U$ instead of $(M + \delta t_m A^m)U$ only for simplicity). We use the fact that

$$A^m U = \sum_K A^{m,K} U,$$

where $A^{m,K} U$ is the vector whose entries are $\sum_j u_j a_{i_m}^K(w^j, w^i)$, $i = 1, \dots, N$ and

$$\begin{aligned} a_i^K(u, v) = & \frac{1}{2} \sum_{\ell=1}^2 \sum_{k=1}^2 \int_K \Xi_{\ell,k}(t) S_\ell S_k \frac{\partial u}{\partial S_k} \frac{\partial v}{\partial S_\ell} \\ & - \sum_{k=1}^2 \int_K \left(r(t) S_k - \frac{1}{2} \sum_{\ell=1}^2 \frac{\partial}{\partial S_\ell} (\Xi_{\ell,k}(t) S_\ell S_k) \right) \frac{\partial u}{\partial S_k} v + r(t) \int_K u v. \end{aligned} \quad (5.22)$$

Hence,

$$(A^{m,K} U)_i = \sum_j u_j \sum_K A_{ij}^{m,K}. \quad (5.23)$$

For simplicity only, let us only consider the first term in (5.22), so a_i^K becomes

$$a_i^K(u, v) = \frac{1}{2} \sum_{\ell=1}^2 \sum_{k=1}^2 \int_K \Xi_{\ell,k}(t) S_\ell S_k \frac{\partial u}{\partial S_k} \frac{\partial v}{\partial S_\ell}$$

and

$$A_{ij}^{m,K} = \frac{1}{2} \sum_{\ell=1}^2 \sum_{k=1}^2 \int_K \Xi_{\ell,k}(t_m) S_\ell S_k \frac{\partial w^j}{\partial S_k} \frac{\partial w^i}{\partial S_\ell}.$$

But ∇w^i and ∇w^j are constant on K , and $S_k = \sum_{v=1}^3 S_{k,v} \lambda_v^K$, so from (5.22),

$$\begin{aligned} A_{i,j}^{m,K} &= \frac{1}{2} \sum_{k,\ell=1}^2 \Xi_{\ell,k}(t_m) \frac{\partial w^i}{\partial S_\ell} \frac{\partial w^j}{\partial S_k} \sum_{v_1=1}^3 \sum_{v_2=1}^3 S_{\ell,v_1} S_{k,v_2} \int_K \lambda_{v_1}^K \lambda_{v_2}^K \\ &= \frac{|K|}{24} \sum_{k,\ell=1}^2 \Xi_{\ell,k}(t_m) \frac{\partial w^i}{\partial S_\ell} \frac{\partial w^j}{\partial S_k} \sum_{v_1=1}^3 \sum_{v_2=1}^3 S_{\ell,v_1} S_{k,v_2} (1 + \delta_{v_1 v_2}). \end{aligned} \quad (5.24)$$

The summation (5.23) should not be programmed directly, like

$$\begin{aligned} & \text{for } i = 1..N \\ & \text{for } j = 1..N \\ & \quad \text{for } K \in \mathcal{T}_h \\ & (A^m U)_i += A_{ij}^{m,K} u_j, \end{aligned} \quad (5.25)$$

because the numerical complexity of this loop is of the order of $N^2 N_T$, where N_T is the number of triangles in $\mathcal{T}_h$. One should rather notice that the sums commute, that is,

$$\begin{aligned}
 &\text{for } K \in \mathcal{T}_h \\
 &\text{for } j = 1..N \\
 &\text{for } i = 1..N \\
 &(A^m U)_i += A_{ij}^{m,K} u_j
 \end{aligned} \tag{5.26}$$

and then see that $A_{ij}^{m,K}$ is zero when q^i or q^j is not in K . The loop

$$\begin{aligned}
 &\text{for } K \in \mathcal{T}_h \\
 &\text{for } j_{\text{loc}} = 1, 2, 3 \\
 &\text{for } i_{\text{loc}} = 1, 2, 3 \\
 &(A^m U)_{i_{\text{loc}}} += A_{i_{\text{loc}} j_{\text{loc}}}^{m,K} u_{j_{\text{loc}}}
 \end{aligned} \tag{5.27}$$

has a complexity of the order of $O(N_T)$. This technique is called assembling. It has brought up the fact that vertices of triangle K have global indices (their position in the array that store them) and local indices (their position in the triangle K , that is, 1, 2, or 3). The notation i_{iloc} refers to the map from local to global.

The convergence of iterative methods for solving linear systems depends on the spectral properties of the matrix; for example, if the matrix is symmetric and positive definite, the convergence rate of the conjugate gradient method depends on the condition number of the matrix; for a general matrix, the speed of convergence of the GMRES method depends on the numerical range of the matrix. For the linear systems arising from the discretization of parabolic PDE, it is observed that the convergence deteriorates when the size of the systems increases. Therefore, when solving, for example, the linear system (5.15), one has better solution instead

$$B^{-1}(M + \delta t_m A^m)U^m = B^{-1}MU^{m-1}, \tag{5.28}$$

where B is a matrix such that

- the spectral properties of $B^{-1}(M + \delta t_m A^m)$ are better than those of $M + \delta t_m A^m$. This means that B is in some sense close to $M + \delta t_m A^m$.
- the solution of a linear system of the form $BV = G$ can be achieved at a reasonable computational cost.

Such a matrix B is called a preconditioner for (5.15), and the iterative method applied to (5.28) is called a preconditioned iterative method. The construction of good preconditioners is an important topic in numerical analysis. We again refer to AXELSSON [1994], GOLUB and VAN LOAN [1989], GREENBAUM [1997], MEURANT [1999], SAAD [1996].

REMARK 5.5 (Mass lumping for piecewise linear triangular elements). Let f be a smooth function and consider the following approximation for the integral of f over

$\Omega = \cup_{K \in \mathcal{T}_h} K$, where $\mathcal{T}_h$ is a triangulation of Ω :

$$\int_{\Omega} f = \sum_{K \in \mathcal{T}_h} \int_K f \approx \sum_{K \in \mathcal{T}_h} \frac{|K|}{3} \sum_{i=1}^3 f(q_i^K),$$

where q_1^K, q_2^K , and q_3^K are the three vertices of K . If f is affine, this formula is exact, otherwise, it computes the integral with an error $O(h^2)$.

This approximation is called mass lumping: for two functions u_h and $v_h \in V_h$, we call U and V the vectors of their coordinates in the nodal basis; applying mass lumping, one approximates $\int_{\Omega} u_h v_h$ by $U^T \tilde{M} V$, where $\tilde{M}$ is a diagonal matrix with positive diagonal entries.

Results The discrete method discussed above has been applied to compute the pricing function of a best-of put option on a two assets basket, $P_o(S_1, S_2) = (100 - \max(S_1, S_2))^+$. The artificial boundary Γ_0 is $\{\max(S_1, S_2) = \bar{S} = 200\}$. Homogeneous Dirichlet conditions have been imposed on Γ_0 . Such a choice of $\bar{S}$ may not be enough for a good accuracy; in fact, $\bar{S} = 200$ was chosen to obtain figures with nice proportions.

The parameters of the Black–Scholes model are

$$\sigma_1 = 0.2, \quad \sigma_2 = 0.1, \quad \text{and} \quad r = 0.05.$$

The correlation factor is either -0.3 (Fig. 5.2) or -0.9 (Fig. 5.3).

The first-order implicit Euler scheme has been used with a uniform time step of $1/250$ year. Mesh adaption in the (S_1, S_2) variable has been performed every $1/10$ year. For mesh adaption, we have used the software BAMG (see GEORGE, HECHT and SALTEL [1991]). In Fig. 5.2, the adapted mesh and the contours of the pricing function are plotted 0.2 year to maturity (top) and 1 year to maturity (bottom). The mesh is refined near the lines where the payoff function exhibits singularities. As time to maturity grows, the mesh becomes coarser in these regions. In fact, such a large number of mesh adaptations are not necessary.

It is clearly seen that the pricing function diffuses more in the S_1 variable, which is not surprising, because the volatility of the first asset is higher.

5.2.4. Multigrid methods

Geometric multigrid method can be applied if there is a hierarchy of nested meshes $\mathcal{T}_{h_i}$, $i = 1, \dots, q$, in such a way that the corresponding finite-element spaces satisfy $V_1 \subset \dots \subset V_i \subset V_{i+1} \subset \dots \subset V_q$. The dimensions of these spaces are $N_1 < \dots < N_i < N_{i+1} < \dots < N_q$. The heuristics supporting multigrid methods for elliptic problem is as follows: the first observation is that with common iterative solvers like Jacobi or successive over relaxation (SOR), (see BRIGGS [1987], MEURANT [1999]) the components of the error corresponding to high frequencies are usually decreased much faster than those associated with the lower frequencies. This also explains why the convergence rate of such methods deteriorates as the number of unknowns grows. To summarize, the iterative solver makes the error smooth, and the smooth part of the error has a slow decay. An iterative solver with this property is called a smoother. The second observation is that for

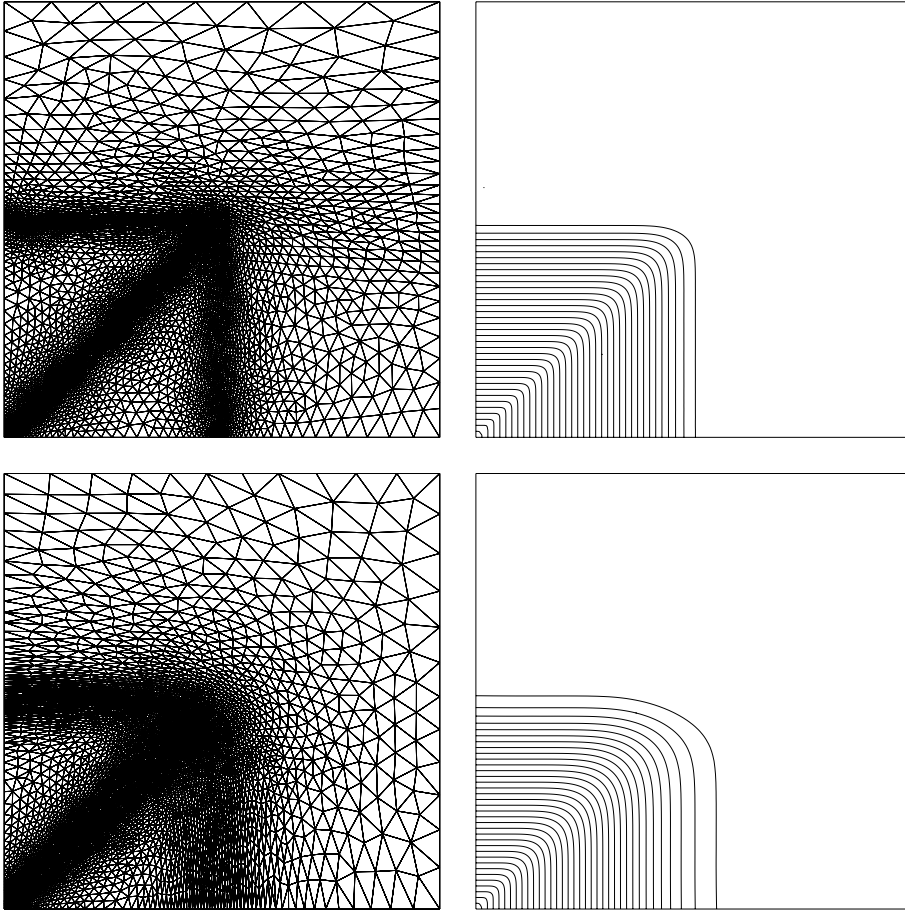


FIG. 5.2 The adapted mesh and the contours of P , at the times to maturity 0.2 year (top) and 1 year (bottom).
 $\sigma_1 = 0.2, \sigma_2 = 0.1, \rho = -0.3$.

a given function $f \in V_{i+1}$, its projection on V_i will be decreased faster by the smoother at level i (operating on V_i) than by the smoother at level $i+1$ (operating on V_{i+1}) because it appears less smooth to the first operator. From these observations, an efficient procedure can be designed by combining the iterations of the smoother with coarse-level corrections. If this idea is also applied to the coarse-level correction, the result is a recursive algorithm.

Assume that a Galerkin method is applied for approximating the solution to a boundary value problem by a function in V_q : the system of linear equation reads $A^{(q)}\mathbf{u}^{(q)} = \mathbf{f}^{(q)}$. Note that it is also possible to define the similar Galerkin discretizations at the lower levels: using the nodal basis, the corresponding system reads

$$A^{(i)}\mathbf{u}^{(i)} = \mathbf{f}^{(i)}, \quad 1 \leq i \leq q. \quad (5.29)$$

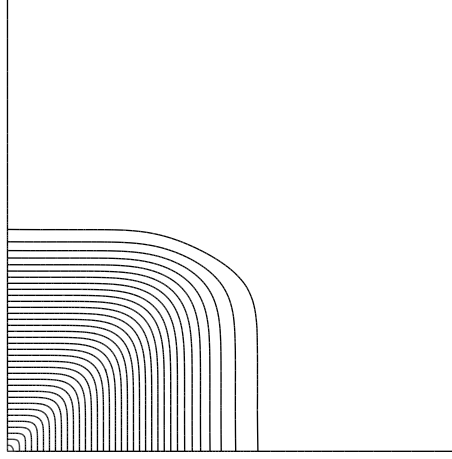


FIG. 5.3 The contours of P , 1 year to maturity. $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.9$.

For simplicity, assume that $A^{(i)}$ are symmetric and positive definite. Denote by $\mathcal{S}^{(i)}$ the smoother at level i : the vector obtained by performing ν iterations of the smoother at level i for solving (5.29) starting from the initial guess $\mathbf{w}$ is written as $\mathcal{S}^{(q)}(\mathbf{f}^{(i)}, \mathbf{w}, \nu)$. An ingredient of the method is the canonical injection from V^i to V^{i+1} : let $\mathbf{I}_i^{i+1}$ stand for its matrix in the nodal bases. Another ingredient is the restriction operator from V_{i+1} to V_i , whose matrix is $\mathbf{I}_{i+1}^i$: a possible choice is to take the Galerkin projection, that is,

$$(A^{(i)} \mathbf{I}_{i+1}^i \mathbf{u}, \mathbf{w}) = (A^{(i+1)} \mathbf{u}, \mathbf{I}_i^{i+1} \mathbf{w}), \quad \forall \mathbf{u} \in \mathbb{R}^{N_{i+1}}, \forall \mathbf{w} \in \mathbb{R}^{N_i}. \quad (5.30)$$

We denote by $MG(\mathbf{f}^{(i)}, \mathbf{w}, i)$ one iteration of the multigrid method at level i for solving (5.29) starting from $\mathbf{w}$.

One of the most commonly used multigrid algorithm is the V cycle.

One V cycle: $MG(\mathbf{f}^{(i)}, \mathbf{w}, i) \rightarrow \mathbf{w}$ If $i = 1$, solve the system (5.29) with a direct method, and let $\mathbf{w}$ be the solution.

Else

1. Perform ν_1 iterations of the smoother at level i : $\mathcal{S}^{(i)}(\mathbf{f}^{(i)}, \mathbf{w}, \nu_1) \rightarrow \mathbf{w}$.
2. Compute the residual $\mathbf{r} \in \mathbb{R}^{N_{i-1}}$ on level $i - 1$ by

$$(\mathbf{r}, \mathbf{z}) = (\mathbf{f}^{(i)} - A^{(i)} \mathbf{w}, \mathbf{I}_{i-1}^i \mathbf{z}), \quad \forall \mathbf{z} \in \mathbb{R}^{N_{i-1}}.$$

Note that $\mathbf{r}$ can be expressed in terms of $A^{(i-1)} \mathbf{I}_i^{i-1} \mathbf{w}$ and of the projection of $\mathbf{f}^i$.

3. Apply the multigrid method at level $i - 1$: $MG(\mathbf{r}, 0, i - 1) \rightarrow \tilde{\mathbf{w}}$.
4. Add the coarse-level correction to $\mathbf{w}$: $\mathbf{w} + \mathbf{I}_{i-1}^i \tilde{\mathbf{w}} \rightarrow \mathbf{w}$.
5. Perform another ν_2 iterations of the smoother at level i : $\mathcal{S}^{(i)}(\mathbf{f}^{(i)}, \mathbf{w}, \nu_2) \rightarrow \mathbf{w}$.

The iterative method consists of computing the sequence $\mathbf{w}_{n+1} = MG(\mathbf{f}^{(q)}, \mathbf{w}_n, q)$ until the residual norm becomes smaller than some tolerance ϵ . Under some reasonable assumptions on the elliptic equation, the mesh, and the smoother (see BRAESS

and HACKBUSCH [1983], YSERENTANT [1993]), it can be proved that the norm $\|\mathbf{u}^q - \mathbf{w}_n\|_A$ decays like ρ^n where $\rho < 1$ does not depend on the mesh parameters.

A very nice introduction to multigrid methods is given by BRIGGS [1987]. Multigrid methods can also be used in the construction of preconditioners (see BRAMBLE, PASCIAK and XU [1990], YSERENTANT [1993]). Finally, the ideas above have been generalized in the so-called algebraic multigrid methods when there is no hierarchy of grids (see RUGE and STÜBEN [1987]). Algebraic multigrid methods are among the most robust and efficient for solving the linear systems arising from the discretization of elliptic and parabolic PDEs. Open-source, libraries are available, such as the library hypre, see http://www.llnl.gov/CASC/linear_solvers/.

5.3. Sparse methods

Consider a boundary value problem in the hypercube $\Omega = (0, 1)^d$. One can think of a Poisson problem $\Delta u = -f$ with the Dirichlet boundary conditions $u = 0$ on $\partial\Omega$. For the variational formulation, we need to use the space $H^1(\Omega)$ equipped with the norm $\|v\|_{H^1(\Omega)} = \sqrt{\|v\|_{L^2(\Omega)}^2 + |v|_{H^1(\Omega)}^2}$, where $|v|_{H^1(\Omega)}^2 = \sum_{i=1}^d \|\frac{\partial v}{\partial x_i}\|_{L^2(\Omega)}^2$, and $H_0^1(\Omega)$, the completion in $H^1(\Omega)$ of the subspace of smooth functions compactly supported in Ω . The previous elliptic problem has a weak or variational formulation in $H_0^1(\Omega)$: find $u \in H_0^1(\Omega)$ such that $\int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v$ for all $v \in H_0^1(\Omega)$.

Assume that the solution to the Poisson problem is approximated by a conforming multilinear FEM on a Cartesian mesh, more precisely with piecewise linear functions of total degree $\leq d$. This is the lowest order FEM on this mesh. Assume that the mesh is uniform and each element is a cube of size n^{-1} . It is easy to see that the dimension of the approximation space is of the order of n^d : the algorithmic complexity grows exponentially with d , which actually forbids the use of this method for $d > 4$. This rapid growth in complexity is known as the curse of dimensionality.

Yet, quite recent developments have shown that it may be possible to use deterministic Galerkin methods or grid-based methods for elliptic or parabolic problems in dimension d , for $4 \leq d \leq 20$: these methods are based on either sparse grids GRIEBEL [1998], GRIEBEL, SCHNEIDER and ZENGER [1992], ZENGER [1991] or sparse tensor product approximation spaces GRIEBEL and OSWALD [1995], VON PETERSDOFF and SCHWAB [2004].

In this paragraph, we aim at rapidly describing the principle of sparse approximations. This presentation heavily relies on the review article by BUNGARTZ and GRIEBEL [2004]. We concentrate on the previously mentioned Dirichlet boundary value problem in Ω . The solution u will be approximated by a Galerkin method, that is, a variational problem posed in a finite-dimensional approximation space V_n instead of $H_0^1(\Omega)$. The goal is to use approximation spaces V_n whose dimensions do not grow too rapidly with d .

The results below are proved in BUNGARTZ and GRIEBEL [2004].

5.3.1. Notations and preliminary results

In this section, bold letters will stand for d -uples: for example, $\mathbf{x} = (x_1, \dots, x_d)$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$. We set $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^d$ and $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$. Take a sufficiently

smooth function f defined on $[0, 1]^d$; if $\alpha \in \mathbb{N}^d$, we call $D^\alpha f$ the partial derivative

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}},$$

where $|\alpha| = \sum_{i=1}^d \alpha_i$. For two multiindices α and β and a scalar λ , we define

$$\alpha \cdot \beta = \sum_{i=1}^d \alpha_i \beta_i, \quad \lambda \alpha = (\lambda \alpha_1, \dots, \lambda \alpha_d), \quad 2^\alpha = (2^{\alpha_1}, \dots, 2^{\alpha_d}).$$

We say that $\alpha \leq \beta$ if $\alpha_i \leq \beta_i$, $i = 1, \dots, d$, and $\alpha < \beta$ if $\alpha \leq \beta$ and $\alpha \neq \beta$.

Let us introduce the function spaces $X^{q,r}(\Omega)$, for $r \in \mathbb{N}$ and $q \in [1, +\infty]$:

$$X^{q,r}(\Omega) = \{u \in L^q(\Omega), \forall \alpha \text{ s.t. } \alpha \leq r\mathbf{1}, D^\alpha u \in L^q(\Omega)\}, \quad (5.31)$$

which are endowed with the seminorms:

$$|u|_{q,\alpha} = \left(\int_{\Omega} |D^\alpha u|^q \right)^{\frac{1}{q}}, \quad \alpha \leq r\mathbf{1}, \text{ if } q < \infty,$$

$$|u|_{\infty,\alpha} = \|D^\alpha u\|_{L^\infty(\Omega)}, \quad \alpha \leq r\mathbf{1}, \text{ if } q = \infty.$$

Note that $X^{q,r}(\Omega)$ is imbedded in the more usual Sobolev space $W^{q,r}(\Omega) = \{u \in L^q(\Omega), \forall \alpha \text{ s.t. } |\alpha| \leq r, D^\alpha u \in L^q(\Omega)\}$.

For a multiindex ℓ , consider the Cartesian meshes $\mathcal{T}_\ell$ of $\overline{\Omega}$ with mesh steps $\mathbf{h}_\ell = 2^{-\ell} = (2^{-\ell_1}, \dots, 2^{-\ell_d})$. The grid nodes of $\mathcal{T}_\ell$ are the points $\mathbf{x}_i = \mathbf{i} \cdot \mathbf{h}_\ell$, $\mathbf{0} \leq \mathbf{i} \leq 2^\ell$.

We note by ϕ the mother hat function,

$$\phi(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1, \end{cases}$$

and by $\phi_{\ell,i}$ the d -dimensional hat function,

$$\phi_{\ell,i}(\mathbf{x}) = \prod_{k=1}^d \phi(2^{\ell_k} x_k - i_k). \quad (5.32)$$

We call V_ℓ

$$V_\ell = \text{span} \left(\phi_{\ell,i}, \mathbf{1} \leq \mathbf{i} \leq 2^\ell - \mathbf{1} \right). \quad (5.33)$$

We also consider the wavelet subspaces:

$$W_k = \text{span} \left\{ \phi_{k,i}, \mathbf{1} \leq \mathbf{i} \leq 2^k - \mathbf{1}, i_j \text{ odd}, 1 \leq j \leq d \right\}. \quad (5.34)$$

We have

$$V_\ell = \bigoplus_{1 \leq k \leq \ell} W_k.$$

The basis of V_ℓ obtained by assembling the previously mentioned bases of W_k $1 \leq k \leq \ell$ is called the hierarchical basis of V_ℓ . Calling $I_\ell = \{\mathbf{i} \leq 2^\ell - \mathbf{1} : i_j \text{ odd}, 1 \leq j \leq d\}$, the hierarchical basis of V_ℓ is $\{\phi_{k,i}, \mathbf{i} \in I_k, k \leq \ell\}$. Note that the completion of $\bigoplus_{1 \leq k} W_k$ with respect to the $H^1(\Omega)$ norm is exactly $H_0^1(\Omega)$.

Rescaling the $\phi_{k,i}$ as follows

$$\psi_{k,i} = -2^{-(k+1) \cdot \mathbf{1}} \phi_{k,i}, \quad \mathbf{i} \in I_k, \quad (5.35)$$

we obtain another basis of W_k .

If a function u is smooth enough, then the coefficients of its expansion in the hierarchical basis are obtained by a simple integral formula.

LEMMA 5.1. *If $u \in H_0^1(\Omega) \cap X^{1,2}(\Omega)$, then*

$$u = \sum_{k \geq 1} \sum_{\mathbf{i} \in I_k} u_{k,i} \phi_{k,i}, \quad \text{where } u_{k,i} = \int_{\Omega} D^2 u \cdot \psi_{k,i}. \quad (5.36)$$

By using Lemma 5.1, one may evaluate the contribution u_k of a subspace W_k to the hierarchical expansion of u .

LEMMA 5.2. *If $u \in H_0^1(\Omega) \cap X^{2,2}(\Omega)$, then the component $u_k \in W_k$ of the expansion of u in the hierarchical representation is such that*

$$\begin{aligned} \|u_k\|_{L^2(\Omega)} &\leq 2^{-2|\mathbf{k}|} 3^{-d} |u|_{2,2}, \\ |u_k|_{H^1(\Omega)} &\leq 2^{-2|\mathbf{k}|} 3^{-d+\frac{1}{2}} \left(\sum_{j=1}^d 2^{2k_j} \right)^{\frac{1}{2}} |u|_{2,2}. \end{aligned} \quad (5.37)$$

5.3.2. Sparse Galerkin methods

It is clear that the dimension of V_ℓ is $\prod_{j=1}^d (2^{\ell_j} - 1)$. In particular, $\dim(V_{n\mathbf{1}}) = (2^n - 1)^d$. As already mentioned, the full tensor product space $V_{n\mathbf{1}}$ is often too large for practical use when $d > 4$.

Let us give an example of a sparse Galerkin method: the discrete space is chosen to be

$$V_n = \bigoplus_{1 \leq k, |\mathbf{k}| \leq n+d-1} W_k, \quad (5.38)$$

instead of the full tensor product space $V_{n\mathbf{1}} = \bigoplus_{1 \leq k \leq n\mathbf{1}} W_k$. One may prove that

$$\dim(V_n) = 2^n \left(\frac{n^{d-1}}{(d-1)!} + O(n^{d-2}) \right). \quad (5.39)$$

Therefore, $\dim(V_n)$ is much smaller than $\dim(V_{n\mathbf{1}})$. It can be seen that a Galerkin method with V_n is feasible for d of the order of 10. In Fig. 5.4, we display the bases of $V_{n\mathbf{1}}$ and V_n .

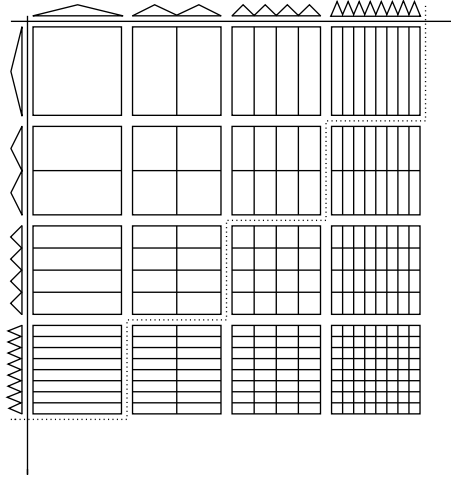


FIG. 5.4 The case $d = 2$: each entry of this array corresponds to a pair of integer $\mathbf{k} = (k_1, k_2)$, $1 \leq k_1, k_2 \leq 4$, and contains the grid corresponding to $W_{\mathbf{k}}$. Each space $W_{\mathbf{k}}$ is the tensor product of two spaces whose bases are plotted on the sides of the array. The full tensor space $V_{n\mathbf{1}}$ is given by $V_{n\mathbf{1}} = \bigoplus_{1 \leq k \leq n\mathbf{1}} W_{\mathbf{k}}$, whereas the sparse tensor space V_n is given by $V_n = \bigoplus_{1 \leq \mathbf{k}, |\mathbf{k}| \leq n+d-1} W_{\mathbf{k}}$ (only the spaces $W_{\mathbf{k}}$ corresponding to the entries above the diagonal are used to construct V_n).

Consider the discretization of the Dirichlet problem in Ω : the discretization error of the Galerkin method with the approximation space V_n (respectively $V_{n\mathbf{1}}$) is of the same order as the best fit error when approximating the solution of the continuous problem by a function of V_n (respectively $V_{n\mathbf{1}}$). Let us assume that u is smooth. We know that $\inf_{v \in V_{n\mathbf{1}}} \|v - u\|_{H^1(\Omega)} \leq C 2^{-n} |u|_{W^{2,2}(\Omega)}$, where $|u|_{W^{2,2}(\Omega)}^2 = \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(\Omega)}^2$. Since V_n is much smaller than $V_{n\mathbf{1}}$, a similar estimate is not true for $\inf_{v \in V_n} \|v - u\|_{H^1(\Omega)}$. GRIEBEL, SCHNEIDER and ZENGER [1992] have proved the following theorem.

THEOREM 5.1. *If $u \in H_0^1(\Omega) \cap X^{2,2}(\Omega)$ and if $u_n \in V_n$ is the component of the expansion of u in the hierarchical representation,*

$$\|u - u_n\|_{L^2(\Omega)} \leq \left(\frac{2^{-2n+1}}{12^d} \sum_{k=0}^{d-1} \binom{n+d-1}{k} \right) |u|_{2,2} = O(2^{-2n} n^{d-1}) |u|_{2,2}, \quad (5.40)$$

$$|u - u_n|_{H^1(\Omega)} \leq \left(\frac{2^{-n} d}{\sqrt{3} 6^{d-1}} \right) |u|_{2,2} = O(2^{-n}) |u|_{2,2}. \quad (5.41)$$

Theorem 5.1 says that under the assumption that $u \in H_0^1(\Omega) \cap X^{2,2}(\Omega)$ (which is a rather strong regularity assumption, much stronger than the assumption $u \in H_0^1(\Omega) \cap W^{2,2}(\Omega)$ required when the full tensor product space is used), using the sparse approximation space V_n instead of the full tensor space $V_{n\mathbf{1}}$ does not deteriorate the accuracy, at least with respect to the H^1 seminorm. There is a moderate deterioration for the L^2 norm of the error.

In our presentation, we have focused on sparse methods based on tensorizing one-dimensional hierarchical bases made of hat functions. This technique can be generalized to other classes of basis functions, for example, higher order piecewise polynomial functions or wavelets as in Fig. 5.5.

5.3.3. Sparse grids

Before defining finite-difference methods on sparse grids, we need to introduce new notations and concepts.

Consider the one variable shape functions: $\phi_{\ell,i}(x) = \phi(2^\ell x - i)$, $\ell \geq 1$, $1 \leq i \leq 2^\ell - 1$, and call V_ℓ the space spanned by $(\phi_{\ell,i})_{1 \leq i \leq 2^\ell - 1}$. Call W_ℓ the subspace of V_ℓ spanned by $(\phi_{\ell,2i-1})_{1 \leq i \leq 2^{\ell-1}}$. We have $V_\ell = W_\ell \oplus V_{\ell-1}$. We have already seen that $V_1 \subset \dots \subset V_\ell \subset V_{\ell+1} \subset \dots$ is a multiresolution analysis of $H_0^1((0, 1))$. For a function $u \in \mathcal{C}^0([0, 1])$ s.t. $u(0) = u(1) = 0$, we have

$$u = \sum_{\ell=1}^{\infty} \sum_{i=1}^{2^{\ell-1}} u_{\ell,i} \phi_{\ell,2i-1},$$

and the projection of u on V_ℓ is

$$\sum_{i=1}^{2^{\ell-1}} u_{\ell,i} \phi_{\ell,i} = \sum_{k=1}^{\ell} \sum_{i=1}^{2^{k-1}} u_{k,i} \phi_{k,2i-1}.$$

The change of coordinates $(u^{\ell,i})_{i=1,\dots,2^{\ell-1}} \mapsto (u_{k,i})_{k=1,\dots,\ell, i=1,\dots,2^{k-1}}$ is called T_ℓ . We call U_ℓ and U^ℓ the column vectors: $U^\ell = (u^{\ell,1}, \dots, u^{\ell,2^{\ell-1}}) \in \mathbb{R}^{2^{\ell-1}}$ and $U_\ell = (u_{\ell,1}, \dots, u_{\ell,2^{\ell-1}}) \in \mathbb{R}^{2^{\ell-1}}$. We have

$$T_\ell U^\ell = \begin{pmatrix} U_1 \\ \vdots \\ U_\ell \end{pmatrix}.$$

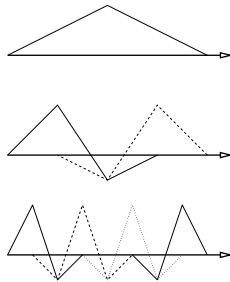


FIG. 5.5 An example of wavelets.

We denote by P^ℓ the restriction operator

$$P^\ell : \mathcal{C}^0([0, 1]) \rightarrow \mathbb{R}^{2^\ell-1}, \quad P^\ell u = U^\ell. \quad (5.42)$$

Note that T_ℓ^{-1} is the representation of the operator P^ℓ in the wavelet basis, that is,

$$P^\ell \left(\sum_{k \leq \ell} \sum_{i=1}^{2^{k-1}} u_{k,i} \phi_{k,2i+1} \right) = T_\ell^{-1} \begin{pmatrix} U_1 \\ \vdots \\ U_\ell \end{pmatrix}.$$

We introduce the interpolation operator I^ℓ :

$$I^\ell : \mathbb{R}^{2^\ell-1} \rightarrow \mathcal{C}^0([0, 1]), \quad I^\ell U = \sum_{i=1}^{2^\ell-1} u^i \phi_{\ell,i}. \quad (5.43)$$

We also denote by D^ℓ the finite-difference operator for the discretization of $\frac{d^2}{dx^2}$:

$$D^\ell : \mathbb{R}^{2^\ell-1} \rightarrow \mathbb{R}^{2^\ell-1}, \quad \forall U, V \in \mathbb{R}^{2^\ell-1} \quad (D^\ell U, V) = 2^\ell \int_0^1 (I^\ell U)' (I^\ell V)'. \quad (5.44)$$

We consider the uniform grids of $(0, 1)$: $\omega^\ell = 2^{-\ell}\{1, \dots, 2^\ell - 1\}$. For $\ell \in \mathbb{N}^d$, $1 \leq \ell$, we introduce the Cartesian grid of Ω : $\Omega^\ell = \prod_{i=1}^d \omega^{\ell_i}$. A grid function on Ω^ℓ is a mapping from Ω^ℓ to $\mathbb{R}$. The space of the grid functions on Ω^ℓ is exactly $\prod_{i=1}^d \mathbb{R}^{2^{\ell_i}-1}$. The mapping $(u^i)_{1 \leq i \leq 2^\ell-1} \mapsto u = \sum_{1 \leq i \leq 2^\ell-1} u^i \phi_{\ell,i}$ is an isomorphism from the space of the grid functions on Ω^ℓ onto V_ℓ defined in (5.33). Moreover, the function u can be written on the wavelet basis $u = \sum_{1 \leq k \leq \ell} \sum_{i \in I_k} u_{k,i} \phi_{k,i}$. Calling U_k the vector $(u_{k,i})_{i \in I_k}$, the grid function will be represented by the family $(U_k)_{1 \leq k \leq \ell}$.

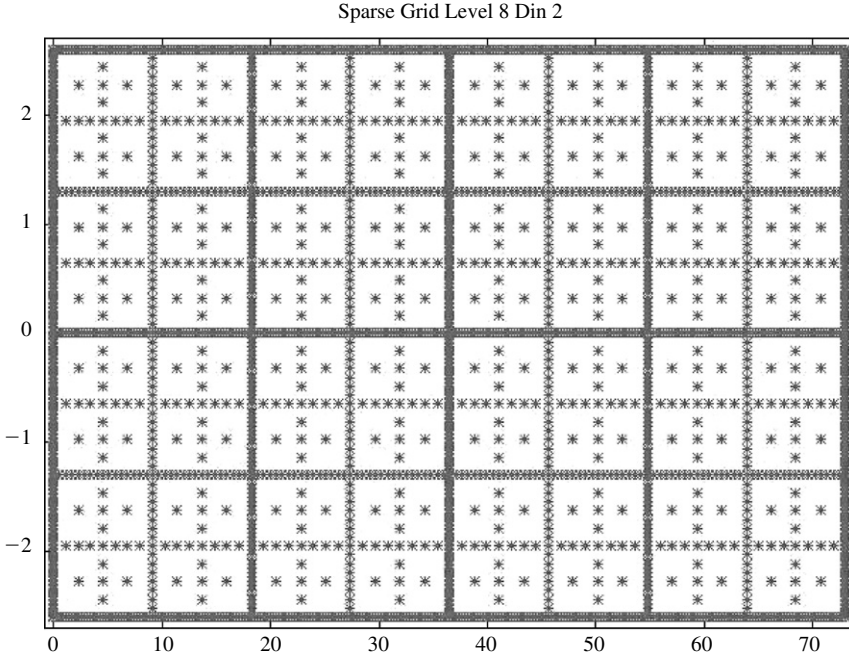
For a positive integer n , we define the sparse grid Ω^n as follows:

$$\Omega^n = \cup_{1 \leq \ell, |\ell| \leq n+d-1} \Omega^\ell \subset \Omega^{n1}. \quad (5.45)$$

An example of a sparse grid in dimension $d = 2$ is presented in Fig. 5.6.

A grid function on Ω^n is a mapping from Ω^n to $\mathbb{R}$. The space of the grid functions on Ω^n is isomorphic to V_n defined in (5.38). As for the full tensor grid, a grid function on Ω^n can be represented on the wavelet basis by $\sum_{1 \leq k, |k| \leq n} \sum_{i \in I_k} u_{k,i} \phi_{k,i}$. Calling U_k the vector $(u_{k,i})_{i \in I_k}$, the sparse grid function will be represented by the family $(U_k)_{1 \leq k, |k| \leq n+d-1}$.

We now define the sparse finite-difference discretization of $\frac{\partial^2}{\partial x_1^2}$: given the vectors $\check{k} = (k_2, \dots, k_d) \in \mathbb{N}^{d-1}$, $\check{i} \in I_{\check{k}}$ and a sparse grid function represented by $(U_k)_{1 \leq k, |k| \leq n+d-1}$, let $\tilde{k}$ be the positive integer $\tilde{k} = n + d - 1 - |\check{k}|$; we introduce $U_{\check{k}}$ by

FIG. 5.6 An example of a sparse grid for $d = 2, n + 1 = 8$.

$$U_{\check{k}, \check{i}} = \begin{pmatrix} U_{(1, \check{k}, \check{i})} \\ \vdots \\ U_{(\tilde{k}, \check{k}, \check{i})} \end{pmatrix}, \quad \text{where } U_{(j, \check{k}, \check{i})} = \left(u_{(j, \check{k}), (m, \check{i})} \right)_{\{m \text{ odd}, 1 \leq m \leq 2^j - 1\}}^T.$$

The sparse grid discretization of the operator $\frac{\partial^2}{\partial x_1^2}$ is

$$(U_k)_{1 \leq k, |k| \leq n+d-1} \mapsto (V_k)_{1 \leq k, |k| \leq n+d-1} \quad (5.46)$$

$$\text{such that } V_{\check{k}, \check{i}} = T_{\check{k}} D_{\check{k}} T_{\check{k}}^{-1} U_{\check{k}, \check{i}}, \quad \forall \check{k}, \check{i} \in I_k.$$

The sparse grid discretization of the operators $\frac{\partial^2}{\partial x_j^2}, \frac{\partial}{\partial x_j}, j = 1, \dots, d$, can be done in a similar way.

It is natural to define the restriction operator $P^\ell : u \mapsto u|_{\Omega^\ell}$ and the interpolation operator $I^\ell = I^{\ell_1} \otimes \dots \otimes I^{\ell_d} : \prod_{i=1}^d \mathbb{R}^{2^{\ell_i}-1} \rightarrow C^0(\Omega)$. The finite-difference approximation of $\partial_{x_1}^2 u$ on the grid Ω^ℓ is $(I^\ell \circ (D^\ell \otimes Id) \circ P^\ell)(u)$. It has been proved by Koster (see KOSTER [2000]) that the sparse grid approximation of $\partial_{x_1}^2 u$ can be written in terms of these finite-difference operators.

THEOREM 5.2. *For a function $u \in C^0(\Omega)$ s.t. $u = 0$ on $\partial\Omega$, we note $D_n(u)$ the function of V_n whose expansion in the wavelet basis is given by $(V_k)_{1 \leq k, |k| \leq n+d-1}$ in (5.46), where $(U_k)_{1 \leq k, |k| \leq n+d-1}$ is the expansion on the wavelet basis of the projection of u on V_n . Then,*

$$D_n(u) = \left(\sum_{1 \leq k, |k| \leq n+d-1} f(k) I^k \circ (D^{k_1} \otimes Id) \circ P^k \right) (u), \quad (5.47)$$

where $f(k)$ is recursively defined by

$$\begin{aligned} f(k) &= 0, & \text{if } |k| > n+d-1 \text{ or } k < \mathbf{1}, \\ f(k) &= 1 - \sum_{\ell: k < \ell} f(\ell), & \text{if } |k| \leq n+d-1 \text{ and } k \geq \mathbf{1}. \end{aligned} \quad (5.48)$$

Before stating a consistency estimate, let us introduce some Hölder spaces: let α belong to $\mathbb{R}_+^d$. Call $[\alpha]$ the vector of $\mathbb{N}^d$ whose i th component is the integer part of α_i . Call $\{\alpha\} = \alpha - [\alpha]$. We note $C^\alpha(\bar{\Omega})$ the space of continuous functions u such that for all $\beta \leq [\alpha]$, $D^\beta u$ is continuous and

$$\sup \left\{ \frac{|D^{[\alpha]} u(\mathbf{x} + \mathbf{h}) - D^{[\alpha]} u(\mathbf{x})|}{|h_1|^{[\alpha_1]} \dots |h_d|^{[\alpha_d]}}, \mathbf{x}, \mathbf{x} + \mathbf{h} \in \Omega, |h_i| > 0, i = 1, \dots, d \right\} < +\infty.$$

The last quantity corresponds to a seminorm on $C^\alpha(\bar{\Omega})$, which we call $|u|_{C^\alpha(\bar{\Omega})}$.

Theorem 5.2 is the key to the following consistency estimate, obtained by KOSTER [2000]:

THEOREM 5.3. *Assume that $u \in C^\alpha(\bar{\Omega})$, where $\alpha_1 > 2$, $\alpha_i > 0$, $i = 2, \dots, d$, and $u = 0$ on $\partial\Omega$. Let P^n be the restriction operator on the sparse grid Ω_n : $P^n(u) = u(\Omega_n)$. We have the consistency error estimate*

$$\| P^n \left(\frac{\partial^2 u}{\partial x_1^2} \right) - P^n \circ D_n(u) \|_\infty \leq C n^{d-1} 2^{-n \min(\alpha_1-2, \alpha_2, \dots, \alpha_d, 2)} |u|_{C^\alpha(\bar{\Omega})}. \quad (5.49)$$

Similarly, for the sparse discretization of the Laplace operator, the consistency error may be bounded by $C n^{d-1} 2^{-n \min(\alpha_1-2, \alpha_2-2, \dots, \alpha_d-2, 2)} |u|_{C^\alpha(\bar{\Omega})}$ if $u \in C^\alpha(\bar{\Omega})$, with $\alpha_i > 2$, $i = 1, \dots, d$.

We find that the sparse grid discretization of Δ is consistent and the consistency error is almost of the same order (up to the factor n^{d-1}) as the consistency error obtained with a full tensor grid.

We are left with studying the stability of the sparse grid discretization. As far as we know, there is, unfortunately, no theoretical stability estimates. There is even no proof that the matrix D arising in the discrete problem is invertible. Indeed, D does not fall into the well-studied classes of matrices: in particular, D is neither a symmetric nor a M matrix. No discrete maximum principle is available. Nevertheless, numerical tests were

done by SCHIEKOFER [1998], indicating that the stability constant, that is, $\|D^{-1}\|_\infty$ is bounded by Cn^{d-1} .

If such a stability estimate is true, we see that the sparse grid discretization of the Poisson problem is convergent, with an error of the order of

$$n^{2d-2} 2^{-n \min(\alpha_1-2, \alpha_2-2, \dots, \alpha_d-2, 2)},$$

if $u \in \mathcal{C}^\alpha(\bar{\Omega})$ with $\alpha_i > 2, i = 1, \dots, d$.

5.3.4. The combination technique

For a linear PDE in Ω with Dirichlet conditions, there is an alternative technique that consists of separately computing the approximations of the solution with standard finite-difference schemes on all the Cartesian grids Ω^ℓ , $1 \leq \ell$, $|\ell| \leq n + d - 1$, and suitably combining these solutions (see GRIEBEL, SCHNEIDER and ZENGER [1992], REISINGER, REISINGER and WITTUM [2007], ZENGER [1991]): the discrete solution is

$$\sum_{1 \leq \ell, |\ell| \leq n+d-1} f(\ell) u_\ell = \sum_{\ell=n}^{n+d-1} a_{\ell-n} \sum_{\ell, |\ell|=\ell} u_\ell,$$

where u_ℓ is the discrete solution computed with the standard finite-difference scheme on Ω^ℓ , $f(\ell)$ is defined in (5.48) and

$$a_j = (-1)^{d-1-j} \binom{d-1}{j}, \quad 0 \leq j \leq d-1.$$

The choice of the coefficients a_j comes from

- performing a multidimensional Taylor expansion of the error between the solution to the continuous problem and its approximation by a linear finite-difference scheme on a Cartesian grid of steps $(h_1, \dots, h_d)$ with respect to $h_1, \dots, h_d$.
- combining the discrete solutions on the Cartesian grids Ω^ℓ , $1 \leq \ell$, $|\ell| \leq n + d - 1$, in order to cancel the larger terms in the above-mentioned Taylor expansions.

Doing so, there is an approximation error (and not only a consistency estimate) (see REISINGER, REISINGER and WITTUM [2007]); for a second-order scheme and a sufficiently smooth u , the error in maximum norm is bounded by

$$C_d(n + 2(d-1))^{d-1} 2^{-2n}.$$

Applications to option pricing Sparse methods have been applied for pricing derivatives by several authors, in particular REISINGER [2004], VON PETERSDOFF and SCHWAB [2004] with wavelets.

For option pricing, one of the main difficulty is that the payoff function is generally not smooth and, furthermore, the locus of its singularity has no relation with the directions of the sparse grid (or sparse tensor product); therefore, the error of the sparse approximation will increase (blow up) near maturity.

For basket options with a payoff depending on the weighted sum $\sum_{i=1}^d \alpha_i S_i$, the change of variable (4.17) proposed by REISINGER [2004] may be used; for a Cartesian

grid in the new variables $(y_i)_{1 \leq i \leq d}$ or for a sparse grid obtained by removing nodes from the last Cartesian grid, the locus of the singularity is an hyperplane perpendicular to one of the grid's directions. This enables grid refinement in the last direction, which decreases the error while keeping the size of the discrete problem reasonable. The resulting grid is sparse in the directions parallel to the last hyperplane and nonuniformly refined in the remaining direction. The price to pay is a more complicated PDE. Of course, this trick is not possible with other options such as best-of options; more involved refinement strategies have then to be used (see GRIEBEL [1998] and the examples below).

To compensate the loss of regularity at maturity, VON PETERSDOFF and SCHWAB [2004] have proposed to use a time stepping with a very nonuniform time grid suitably refined near maturity.

An even more difficult case is that of American options (see Section 6) because the pricing function exhibits a singularity at the exercise boundary, which is an unknown and cannot be related to the grid's directions.

As an illustration, we plot the pricing function for a put on an average of two assets computed with a sparse grid (see Fig. 5.7). Here, the singularity of the payoff is not aligned with the grid. In Fig. 5.8, we show an adapted sparse grid for the same baskets on two assets. The sparse grid is computed by progressively enriching an initial coarse grid. The mesh is refined near a node if the corresponding coefficient in the multilevel expansion of the discrete solution is larger than a threshold.

We will see later that sparse methods prove more useful for option pricing on a single asset but with a multifactor stochastic volatility (see Section 7.4). In this case, the payoff function depends on the price variable only. Hence, the singularity is located on an hyperplane in the price/volatilities space, and sparse grids can be used in an easy way.

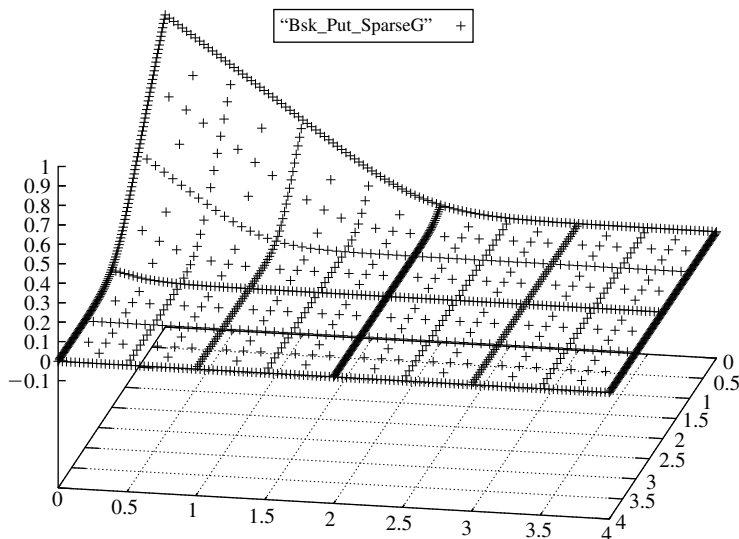


FIG. 5.7 The pricing function of a European option on a basket of two assets, computed on a sparse grid (many thanks to David Pommier who wrote the software and gave us this figure).

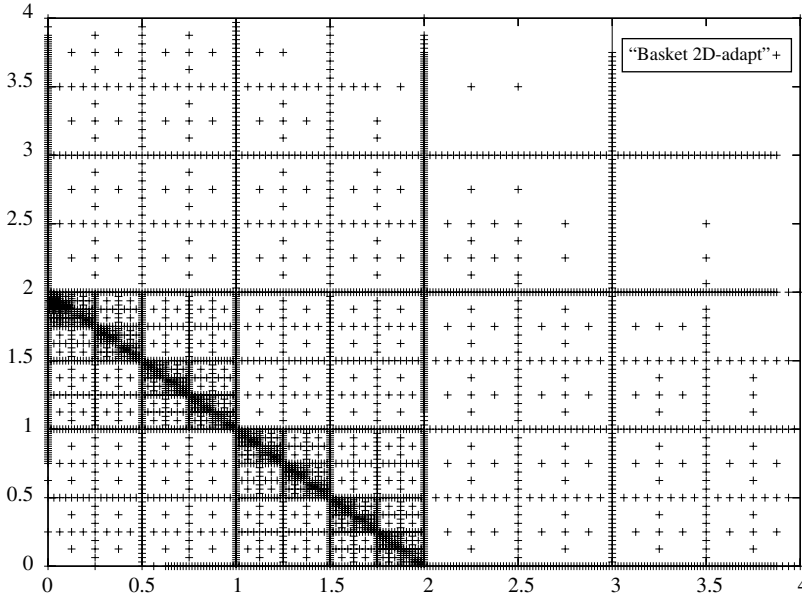


FIG. 5.8 Adapted sparse grid for an option on a basket of two assets (thanks to David Pommier).

6. American basket options

We consider an American option on the d risky assets whose prices $S_{i,t}$ are the processes described at the beginning of Section 4. The maturity of the option is T and its payoff function is $P_o : \mathbb{R}_+^d \mapsto \mathbb{R}_+$.

The Black–Scholes model leads to the following formula for the price of the American option at time t : under the risk-neutral probability,

$$P_t = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^* \left(e^{-\int_t^\tau r(s) ds} P_o(S_{1,\tau}, \dots, S_{d,\tau}) \middle| F_t \right), \quad (6.1)$$

where $\mathcal{T}_{t,T}$ denotes the set of stopping times in $[t, T]$ (see JAILLET, LAMBERTON and LAPEYRE [1990], KARATZAS [1988]). It is clear that $P_t \geq P_o(S_{1,t}, \dots, S_{d,t})$.

Under suitable assumptions on the payoff P_o and on the volatilities, it can be proven that P_t is a function of $S_{1,t}, \dots, S_{d,t}$ and t , that is, $P_t = P(S_{1,t}, \dots, S_{d,t}, t)$, and the pricing function P is the solution to a variational inequality (see (6.6), which is the variational form of the following linear complementarity problem:

$$\begin{aligned} \frac{\partial P}{\partial t} + LP - rP &\leq 0, & \text{in } \mathbb{R}_+^d \times [0, T), \\ P &\geq P_o, & \text{in } \mathbb{R}_+^d \times [0, T), \\ \left(\frac{\partial P}{\partial t} + LP - rP \right) (P - P_o) &= 0, & \text{in } \mathbb{R}_+^d \times [0, T), \\ P|_{t=T} &= P_o, & \text{in } \mathbb{R}_+^d, \end{aligned} \quad (6.2)$$

where L is given by (4.3). The proof of this result goes beyond the scope of this paragraph. It can be found in BENSOUSSAN and LIONS [1984] or in JAILLET, LAMBERTON and LAPEYRE [1990].

6.1. The variational inequality

Here, we assume that $P_o \in L^2(\mathbb{R}_+^d)$. It is possible to deal with other payoff functions by using Sobolev spaces with different weights at infinity (see BENSOUSSAN and LIONS [1984] or JAILLET, LAMBERTON and LAPEYRE [1990]). Calling t the time to maturity, (6.2) becomes

$$\begin{aligned} \frac{\partial P}{\partial t} - LP + rP &\geq 0, & \text{in } \mathbb{R}_+^d \times (0, T], \\ P &\geq P_o, & \text{in } \mathbb{R}_+^d \times (0, T], \\ \left(\frac{\partial P}{\partial t} - LP + rP\right)(P - P_o) &= 0, & \text{in } \mathbb{R}_+^d \times (0, T], \\ P|_{t=0} &= P_o, & \text{in } \mathbb{R}_+^d, \end{aligned} \quad (6.3)$$

where L is given by (4.3). To write the variational formulation of (6.3), we use the same Sobolev space V as for the European option (see (4.31)). We call $\mathbb{K}$ the closed and convex subset of V ,

$$\mathbb{K} = \{v \in V, v \geq P_o \text{ a.e. in } \mathbb{R}_+^d\}, \quad (6.4)$$

and we introduce $\mathcal{K}$,

$$\mathcal{K} = \{v \in L^2(0, T; V), \text{ s.t. } v(t) \in \mathbb{K} \text{ for a.a. } t \in (0, T)\}. \quad (6.5)$$

Following LIONS [1969], a variational formulation of (6.3) is to

$$\begin{aligned} &\text{find } P \in \mathcal{K} \text{ such that } \frac{dP}{dt} \in L^2(0, T; V') \text{ and } P(t=0) = P_o \text{ and satisfying} \\ \forall v \in \mathcal{K}, \quad &\int_0^T \left\langle \frac{dP}{dt}(t), v(t) - P(t) \right\rangle dt + \int_0^T a_t(P(t), v(t) - P(t)) dt \geq 0, \end{aligned} \quad (6.6)$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between V' and V and a_t is the bilinear form introduced in (4.30).

By adapting the results in LIONS [1969] (see also KINDERLEHRER and STAMPACCHIA [1980]), one can prove the following theorem.

THEOREM 6.1. *Under the assumptions (4.33) (4.34) (4.35) and if r is a bounded function defined on $(0, T)$, the variational inequality (6.6) has a unique solution P . Furthermore, $P \in C^0([0, T]; L^2(\mathbb{R}_+^d))$, $\frac{\partial P}{\partial t} \in L^2(\mathbb{R}_+^d \times (0, T))$, $S_i S_j \frac{\partial^2 P}{\partial S_i \partial S_j} \in L^2(\mathbb{R}_+^d \times (0, T))$, $i, j = 1, \dots, d$, and P satisfies the linear complementarity problem (6.3).*

More properties can be proved under stronger assumptions on the payoff function and the coefficients, for example, the volatilities are constant and the payoff function is piecewise linear and continuous with compact support, then P is continuous on $\mathbb{R}_+^d \times$

$[0, T]$. If the coefficients are constant and under some special assumptions on the payoff, it can be proved that $P(S, t)$ is nondecreasing with respect to t (t is the time to maturity).

6.2. The exercise region

The exercise region at time t is the set

$$\{S \in \mathbb{R}_+^d, \text{ s.t. } P(S, t) = P_\circ(S)\}.$$

The theoretical results concerning the exercise region for American options on baskets strongly depend on the payoff. VILLENEUVE [2004] has proved that if the coefficients are constant with $r > 0$ and $P_\circ$ is bounded and continuous, then the exercise region is nonempty. The shape of the exercise region and its behavior near maturity is studied as well by VILLENEUVE [2004] for a particular class of payoff functions.

Examples of exercise regions for American best-of options computed by FEM will be given below (see Fig. 6.1). It will be seen that the exercise boundary, that is, the boundary of the exercise region may exhibit rather strong singularities.

6.3. Finite element methods

Assuming that $P_\circ$ has compact support, we truncate the domain as for the European option, that is, we consider $\Omega = (0, \bar{S})^d$ (where $\bar{S}$ is large enough so that the support of $P_\circ$ is strictly contained in Ω) and Γ_0 given in (5.1). We choose to impose homogeneous Dirichlet artificial boundary conditions on Γ_0 . The Sobolev space V to work with is given in (5.4), and the new definition of $\mathbb{K}$ is $\mathbb{K} = \{v \in V, v \geq P_\circ \text{ a.e in } \Omega\}$. Changing $\mathcal{K}$ accordingly (see (6.5)), the new variational inequality is (6.6), where a_t is given by (5.5). We are now ready to propose a finite-element discretization.

We introduce a partition of the interval $[0, T]$ into subintervals $[t_{m-1}, t_m]$, $1 \leq m \leq M$, with $\delta t_i = t_i - t_{i-1}$, $\delta t = \max_i \delta t_i$. We choose a triangulation $\mathcal{T}_h$ of Ω , and we define V_h by the following:

$$V_h = \{v_h \in V, \forall \omega \in \mathcal{T}_h, v_h|_\omega \in \mathcal{P}_1\}, \quad (6.7)$$

where $\mathcal{P}_1$ is the space of affine functions. For simplicity, we assume that $P_\circ \in V_h$. We define the closed and convex subset $\mathbb{K}_h$ of V_h by

$$\mathbb{K}_h = \{v \in V_h, v_h \geq P_\circ \text{ in } \bar{\Omega}\}. \quad (6.8)$$

The discrete problem arising from the implicit Euler scheme is

$$\begin{aligned} &\text{find } (P^m)_{0 \leq m \leq M} \in \mathbb{K}_h \text{ satisfying} \\ &P^0 = P_\circ, \end{aligned} \quad (6.9)$$

and for all m , $1 \leq m \leq M$,

$$\forall v \in \mathbb{K}_h, \quad \left(P^m - P^{m-1}, v - P^m \right) + \delta t_m a_{t_m}(P^m, v - P^m) \geq 0. \quad (6.10)$$

Expressing P^m , $0 \leq m \leq M$, and v in the nodal basis of V_h , (6.10) is equivalent to the finite-dimensional linear complementary system

$$\begin{aligned} \mathbf{M}(\mathbf{U}^m - \mathbf{U}^{m-1}) + \delta t_m \mathbf{A}^m \mathbf{U}^m &\geq 0, \\ \mathbf{U}^m &\geq \mathbf{U}^0, \\ (\mathbf{U}^m - \mathbf{U}^0)^T (\mathbf{M}(\mathbf{U}^m - \mathbf{U}^{m-1}) + \delta t_m \mathbf{A}^m \mathbf{U}^m) &= 0, \end{aligned} \quad (6.11)$$

where for two vectors $\mathbf{U}$ and $\mathbf{V}$, $\mathbf{U} \geq \mathbf{V}$ means that all the coordinates of $\mathbf{U} - \mathbf{V}$ are nonnegative.

Assuming that the coefficients satisfy the assumptions (4.33) (4.34) (4.35), it can be proved that for δt small enough (6.11) has a unique solution for all $m = 1, \dots, M$. Stability and convergence in the natural energy norm can be proved.

Mesh adaptivity based on a posteriori error estimates is possible. The description of the error estimators for parabolic variational inequalities goes beyond the scope of this chapter. We refer to CHEN and NOCHETTO [2000], NOCHETTO, SIEBERT and VEESER [2003, 2005], VEESER [2001] for a posteriori error estimates and mesh refinement strategies for elliptic variational inequalities. In the parabolic case, a strategy similar to the one by BERGAM, BERNARDI and MGHAZLI [2005] is studied by ACHDOU, HECHT and POMMIER.

6.4. Algorithms

6.4.1. The projected SOR algorithm

Let us write (6.11) in the simpler form

$$\begin{aligned} \mathbf{B}\mathbf{u} &\geq \mathbf{f}, \\ \mathbf{u} &\geq \mathbf{g}, \\ (\mathbf{u} - \mathbf{g})^T (\mathbf{B}\mathbf{u} - \mathbf{f}) &= 0. \end{aligned} \quad (6.12)$$

The projected SOR algorithm is an iterative method for solving (6.12). Let ω be a positive real number. The idea is to approximate $\mathbf{u}$ by using a one-step recursion formula $\mathbf{u}^{(k+1)} = \psi(\mathbf{u}^{(k)})$ (starting from an initial guess $\mathbf{u}^{(0)}$), where ψ is the nonlinear mapping in $\mathbb{R}^N$:

$$\begin{aligned} \psi : \mathbf{v} &\mapsto \mathbf{w} = \psi(\mathbf{v}) : \\ \forall i = 1, \dots, N, \quad w_i &= \max(g_i, y_i), \quad \text{and } y_i \text{ is given by} \\ \frac{1}{\omega} \mathbf{B}_{ii} y_i + \sum_{j < i} \mathbf{B}_{ij} w_j &= f_i + \left(\frac{1}{\omega} - 1\right) \mathbf{B}_{ii} v_i - \sum_{j > i} \mathbf{B}_{ij} v_j. \end{aligned} \quad (6.13)$$

This construction is a modification of the so-called SOR method used for systems of linear equations (see AXELSSON [1994], GOLUB and VAN LOAN [1989], SAAD [1996] for iterative methods). For solving approximately the system $\mathbf{B}\mathbf{v} = \mathbf{f}$, the SOR algorithm constructs the sequence $(\mathbf{v}^{(k)})_k$ (starting from an initial guess $\mathbf{v}^{(0)}$) by the recursion:

$$\forall i = 1, \dots, N, \quad \frac{1}{\omega} \mathbf{B}_{ii} v_i^{(k+1)} + \sum_{j < i} \mathbf{B}_{ij} v_j^{(k+1)} = f_i + \left(\frac{1}{\omega} - 1\right) \mathbf{B}_{ii} v_i^{(k)} - \sum_{j > i} \mathbf{B}_{ij} v_j^{(k)}.$$

PROPOSITION 6.1. *If $\mathbf{B}$ is a diagonal dominant matrix and if $0 < \omega \leq 1$, then the mapping ψ defined in (6.13) is a contraction in $\mathbb{R}^N$ for the norm $\|\cdot\|_\infty$ ($\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq N} |v_i|$). The fixed point of ψ is $\mathbf{u}$.*

Under the assumptions of Proposition 6.1, the sequence constructed by the PSOR (projected SOR) algorithm converges to $\mathbf{u}$. The speed of convergence depends on the matrix $\mathbf{B}$ and on the relaxation parameter ω . The convergence is generally slow for ill-conditioned matrices.

6.4.2. Primal-dual methods

Following ITO and KUNISCH [2003], we first go back to the semidiscrete problem: find $P^m \in \mathbb{K}$ such that

$$\forall v \in \mathbb{K}, \quad (P^m - P^{m-1}, v - P^m) + \delta t_m a_{t_m}(P^m, v - P^m) \geq 0.$$

For any positive constant c , this is equivalent to finding $P^m \in V$ and a Lagrange multiplier $\mu \in V'$ such that

$$\begin{aligned} \forall v \in V, \quad & \frac{1}{\delta t_m} (P^m - P^{m-1}, v) + a_{t_m}(P^m, v) - \langle \mu, v \rangle = 0, \\ & \mu = \max(0, \mu - c(P^m - P^0)). \end{aligned} \quad (6.14)$$

When using an iterative method for solving (6.14), that is, when constructing a sequence $(P^{m,j}, \mu^j)$ for approximating (P^m, μ) , the Lagrange multiplier μ^j may not be a function if the gradient of $P^{m,j}$ jumps, whereas μ may be a function. Therefore, a dual method (i.e., an iterative method for computing μ) may be difficult to use. As a remedy, ITO and KUNISCH [2003] considered a one-parameter family of regularized problems based on smoothing the equation for μ as follows:

$$\mu = \alpha \max(0, \mu - c(P^m - P^0)), \quad (6.15)$$

for $0 < \alpha < 1$, which is equivalent to

$$\mu = \max(0, -\chi(P^m - P^0)), \quad (6.16)$$

for $\chi = c\alpha/(1 - \alpha) \in (0, +\infty)$. We may consider a generalized version of (6.16)

$$\mu = \max(0, \bar{\mu} - \chi(P^m - P^0)) \quad (6.17)$$

where $\bar{\mu}$ is a fixed function. This turns out to be useful when the complementarity condition is not strict.

It is now possible to study the fully regularized problem

$$\begin{aligned} \forall v \in V, \quad & \frac{1}{\delta t_m} (P^m - P^{m-1}, v) + a_{t_m}(P^m, v) - \langle \mu, v \rangle = 0, \\ & \mu = \max(0, \bar{\mu} - \chi(P^m - P^0)) \end{aligned} \quad (6.18)$$

and prove that it has a unique solution, with μ a square-integrable function. A primal-dual active set algorithm for solving (6.18) is as follows:

6.4.3. Primal-dual active set algorithm

- Choose $P^{m,0}$ and set $k = 0$
- Loop

1. Set

$$\begin{aligned}\mathcal{A}^{-,k+1} &= \{S : \bar{\mu}^k(S) - \chi(P^{m,k}(S) - P^0(S)) > 0\}, \\ \mathcal{A}^{+,k+1} &= (0, \bar{S}) \setminus \mathcal{A}^{-,k+1}.\end{aligned}$$

2. Solve for $P^{m,k+1} \in V: \forall v \in V$,

$$\begin{aligned}0 &= \frac{1}{\delta t_m} (P^{m,k+1} - P^{m-1}, v) + a_{t_m}(P^{m,k+1}, v) \\ &\quad - (\bar{\mu} - \chi(P^{m,k+1} - P^0), 1_{\mathcal{A}^{-,k+1}} v).\end{aligned}\tag{6.19}$$

3. Set

$$\mu^{k+1} = \begin{cases} 0 & \text{on } \mathcal{A}^{+,k+1}, \\ \bar{\mu} - \chi(P^{m,k+1} - P^0) & \text{on } \mathcal{A}^{-,k+1} \end{cases}\tag{6.20}$$

4. Set $k = k + 1$.

Calling A_m the operator from V to V' : $\langle A_m v, w \rangle = \frac{1}{\delta t_m} (v, w) + a_{t_m}(v, w)$ and $F : V \times L^2(\mathbb{R}_+^d) \rightarrow V' \times L^2(\mathbb{R}_+^d)$

$$F(v, \mu) = \begin{pmatrix} A_m v + \mu - \frac{P^{m-1}}{\delta t_m} \\ \mu - \max(0, \bar{\mu} - \chi(v - P^0)) \end{pmatrix},$$

it is proved by ITO and KUNISCH [2003] that $G(v, \mu) : V \times L^2(\mathbb{R}_+^d) \rightarrow V' \times L^2(\mathbb{R}_+^d)$,

$$G(v, \mu)h = \begin{pmatrix} A_m h_1 + h_2 \\ h_2 - \chi 1_{\{\bar{\mu} - \chi(v - P^0) > 0\}} h_1 \end{pmatrix}$$

is a generalized derivative of F in the sense that

$$\lim_{\|h\| \rightarrow 0} \frac{\|F(v + h_1, \mu + h_2) - F(v, \mu) - G(v + h_1, \mu + h_2)h\|}{\|h\|} = 0.$$

Note that

$$G(P^{m,k}, \mu^k)h = \begin{pmatrix} A_m h_1 + h_2 \\ h_2 - \chi 1_{\mathcal{A}^{-,k+1}} h_1 \end{pmatrix}.$$

Thus, the above primal-dual active set algorithm can be seen as a semismooth Newton method applied to F , that is,

$$(P^{m,k+1}, \mu^{k+1}) = (P^{m,k}, \mu^k) - G^{-1}(P^{m,k}, \mu^k)F(P^{m,k}, \mu^k).\tag{6.21}$$

Indeed, calling $(\delta P^m, \delta\mu) = (P^{m,k+1} - P^{m,k}, \mu^{k+1} - \mu^k)$, it is straightforward to see that in the primal-dual active set algorithm, we have

$$\begin{aligned} A_m \delta P^m + \delta\mu &= -A_m P^{m,k} - \mu^k + \frac{P^{m-1}}{\delta t_m}, \\ \delta\mu &= -\mu^k \text{ on } \mathcal{A}^{+,k+1}, \\ \delta\mu - \chi \delta P^m &= -\mu^k + \bar{\mu} - \chi(P^{m,k} - P^0) \text{ on } \mathcal{A}^{-,k+1}, \end{aligned}$$

which is precisely (6.21).

ITO and KUNISCH [2003], by using the results proved by HINTERMÜLLER, ITO and KUNISCH [2002], established that the primal-dual active set algorithm converges from any initial guess and if the initial guess is sufficiently close to the solution of (6.18), then the convergence is superlinear.

To solve (6.14), it is possible to successively compute the solutions $(P^m(\chi_\ell), \mu(\chi_\ell))$ to (6.18) for a sequence of parameters (χ_ℓ) converging to $+\infty$ using $(P^m(\chi_\ell), \mu(\chi_\ell))$ as an initial guess for the primal-dual active set algorithm for $(P^m(\chi_{\ell+1}), \mu(\chi_{\ell+1}))$.

Of course, it is possible to use the same algorithm for the fully discrete problem. Convergence results hold in the discrete case if there is a discrete maximum principle. The algorithm amounts to solving a sequence of systems of linear equations and the matrix of the system varies at each iteration.

Examples The discrete method discussed above has been applied to compute the pricing function of an American best-of put option on a two assets basket, $P_o(S_1, S_2) = (K - \max(S_1, S_2))^+$. The artificial boundary Γ_0 is $\{\max(S_1, S_2) = \bar{S} = 200\}$. Homogeneous Dirichlet conditions are imposed on Γ_0 . We have chosen two examples:

1. In the first example, the parameters are

$$\sigma_1 = 0.2, \quad \sigma_2 = 0.1, \quad r = 0.05, \quad \rho = -0.6, \quad \text{and } K = 100.$$

2. In the second example, the parameters are

$$\sigma_1 = \sigma_2 = 0.2, \quad r = 0.05, \quad \rho = 0, \quad \text{and } K = 50.$$

The implicit Euler scheme has been used with a uniform time step of 1/250 year. For solving the linear complementarity problems (6.11), we have used the regularized active set method with the regularization parameters $\chi = 10^7$ and $\bar{\mu} = 0$ (see (6.17)). Mesh adaption in the (S_1, S_2) variable has been performed every 1/10 year. The adaptive strategy is close to the one used in FreeFem, and the mesh is refined in the contact set. This may be unnecessary if the obstacle belongs to the finite-element space. We refer to ACHDOU, HECHT and POMMIER for a better adaptive strategy based on local error indicators where the mesh is not refined in the so-called strong contact region (see also CHEN and NOCHETTO [2000], NOCHETTO, SIEBERT and VEESER [2003, 2005], VEESER [2001]) for elliptic contact problems.

In Fig. 6.2, we have plotted the adapted mesh (left) and the contours of the pricing function (right) 1 year to maturity for the first example. Note that the contours exhibit right angles in the exercise region. In Fig. 6.1, we plot the exercise region 1 year to maturity for the first example (top) and for the second example (bottom). One sees that

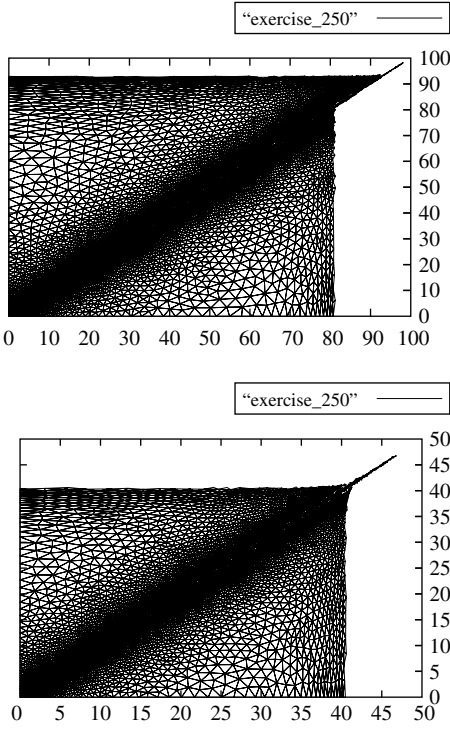


FIG. 6.1 The exercise region 1 year to maturity. Top: $k = 100$, $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.6$. Bottom: $K = 50$, $\sigma_1 = \sigma_2 = 0.2$, and $\rho = 0$.

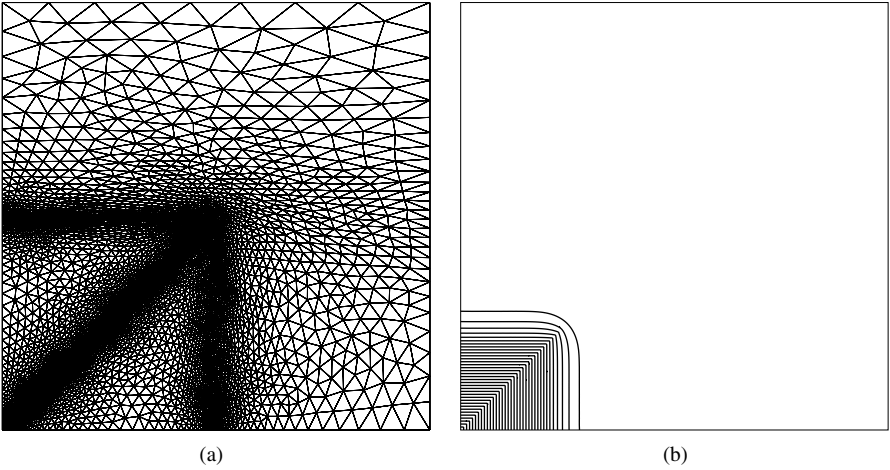


FIG. 6.2 The adapted mesh and the contours of P 1 year to maturity. $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.6$.

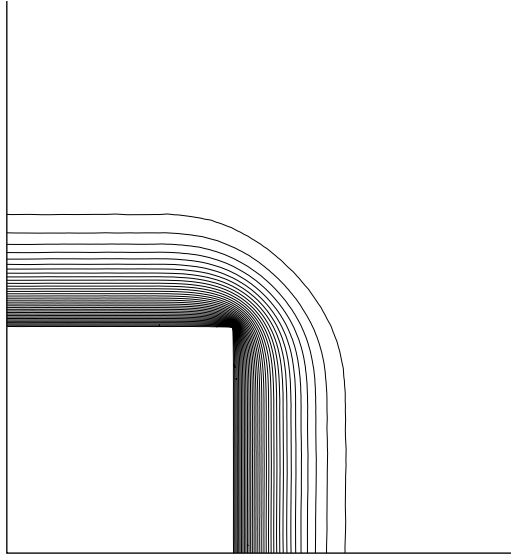


FIG. 6.3 The contours of P for an American binary option 1 year to maturity. $P_o(S_1, S_2) = 50.1_{\{\max(S_1, S_2) < 50\}}$, $\sigma_1 = \sigma_2 = 0.2$, and $\rho = 0$.

the exercise boundary has singularities. It is also visible that the mesh has been adapted near the exercise boundary.

Finally, we consider an American binary option whose payoff is

$$P_o(S_1, S_2) = K 1_{\{\max(S_1, S_2) < K\}},$$

with the parameters of the second example. The exercise region does not evolve. The contours of the pricing function 1 year to maturity are plotted in Fig. 6.3.

6.4.4. Penalty methods

Assume that P_o is bounded and $P_o \in V$, where V is given in (4.31). A penalized version of (6.3) is to look for Q_ϵ such that

$$\begin{aligned} \frac{\partial Q_\epsilon}{\partial t} - LQ_\epsilon + rQ_\epsilon - \mathcal{V}_\epsilon(Q_\epsilon) &= LP_o - rP_o, & \text{in } \mathbb{R}_+^d \times (0, T], \\ Q_\epsilon|_{t=0} &= 0 & \text{in } \mathbb{R}_+^d, \end{aligned} \quad (6.22)$$

where L is given by (4.3), ϵ is a small positive parameter, and $\mathcal{V}_\epsilon : \mathbb{R} \rightarrow \mathbb{R}_+$ is a sequence of C^2 nonincreasing functions such that

- $\mathcal{V}_\epsilon(z) = 0$ if $z > 0$.
- $\mathcal{V}_\epsilon(z) = -\frac{z}{\epsilon}$ if $z < -\epsilon$.
- $\mathcal{V}'_\epsilon$ and $\mathcal{V}''_\epsilon$ are bounded by $\frac{C}{\epsilon}$.

It is reasonable to think that Q_ϵ is a good approximation of $Q = P - P_o$ as ϵ tends to zero.

The theory of the weak solutions of parabolic partial differential with monotone operators (see LIONS [1969]) can be used because $\mathcal{V}_\epsilon$ is nonincreasing. One can prove that (6.22) has a unique weak solution and Q_ϵ tends to Q in the natural norms associated with the variational problem (6.6). In some cases, error estimates can be found.

The initial value problem (6.22) can, in turn, be approximated by the FEM, for example, using, an implicit Euler scheme and piecewise affine functions on a triangulation of Ω , each time step is of the form

$$\mathbf{M}(\mathbf{Q}^m - \mathbf{Q}^{m-1}) + \delta t_m (\mathbf{A}^m \mathbf{Q}^m + \mathcal{V}_\epsilon(\mathbf{Q}^m)) = -\delta t_m \mathbf{A}^m \mathbf{U}^0, \quad (6.23)$$

where $\mathbf{M}$ and $\mathbf{A}$ are the matrices in (5.16) and $\mathcal{V}_\epsilon : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a nonlinear operator such that

$$(\mathcal{V}_\epsilon(\mathbf{Q}^m))_i \sim \int_{\Omega} w_i \mathcal{V}_\epsilon \left(\sum_{j=1}^N q_j^m w_j \right).$$

Here, the sign $\sim$ means that a suitable elementary quadrature rule has been used to approximate $\int_K \mathcal{V}_\epsilon \left(\sum_{j=1}^N q_j^m w_j \right) w_i$, $K \in \mathcal{T}_h$. Newton iterations can be used for solving (6.23), and $\mathbf{Q}^{m-1}$ can be used as an initial guess.

A simpler nonlinear problem is

$$\mathbf{M}(\mathbf{Q}^m - \mathbf{Q}^{m-1}) + \delta t_m (\mathbf{A}^m \mathbf{Q}^m + \mathcal{V}_\epsilon(\mathbf{Q}^m)) = -\delta t_m \mathbf{A}^m \mathbf{U}^0, \quad (6.24)$$

where $\mathcal{V}_\epsilon(\mathbf{Q}^m)$ is the vector defined by $(\mathcal{V}_\epsilon(\mathbf{Q}^m))_i = \mathcal{V}_\epsilon(q_i^m)$, $i = 1, \dots, N$. Convergence may be proved for (6.23) or (6.24).

A similar method has been carefully tested for American options in the slightly different context of a stochastic volatility model (see ZVAN, FORSYTH and VETZAL [1998]).

The computing times of the penalty methods and the primal-dual active set methods are comparable.

6.4.5. Projection Schemes

Let us go back to the sequence of fully discrete linear complementarity problems (6.11) arising from the implicit Euler scheme. It is possible to write them in an equivalent manner using a Lagrange multiplier $\Lambda^m \in \mathbb{R}^N$,

$$\begin{aligned} \mathbf{M}(\mathbf{U}^m - \mathbf{U}^{m-1}) + \delta t_m \mathbf{A}^m \mathbf{U}^m &= \delta t_m \Lambda^m, \\ \mathbf{U}^m &\geq \mathbf{U}^0, \\ \Lambda^m &\geq 0, \\ (\mathbf{U}^m - \mathbf{U}^0)^T \Lambda^m &= 0. \end{aligned} \quad (6.25)$$

Following LIONS and MERCIER [1979] (see also GLOWINSKI, LIONS, and TRÉMOLIÈRES [1981] and IKONEN and TOIVANEN [2004]) we modify (6.25) in order to obtain a new operator splitting time scheme where each time step is made of two substeps: the first substep is that of the implicit Euler scheme for the corresponding parabolic equation (European option) and the second step is a projection step. Operator splitting

schemes are very popular in computational fluid dynamics for incompressible fluids, and they are often called projection schemes (see ACHDOU and GUERMOND [2000], GUERMOND [1999], GUERMOND and QUARTAPPELLE [1998]). More precisely, the first substep reads

$$\mathbf{M}(\tilde{\mathbf{U}}^m - \mathbf{U}^{m-1}) + \delta t_m \mathbf{A}^m \tilde{\mathbf{U}}^m = \delta t_m \Lambda^{m-1}, \quad (6.26)$$

where the unknown is $\tilde{\mathbf{U}}^m$ [(6.26) is a system of linear equations], whereas the second substep is

$$\begin{aligned} \mathbf{M}(\mathbf{U}^m - \tilde{\mathbf{U}}^m) &= \delta t_m (\Lambda^m - \Lambda^{m-1}), \\ \mathbf{U}^m &\geq \mathbf{U}^0, \\ \Lambda^m &\geq 0, \\ (\mathbf{U}^m - \mathbf{U}^0)^T \Lambda^m &= 0. \end{aligned} \quad (6.27)$$

The unknowns of the linear complementarity problem (6.27) are $\mathbf{U}^m$ and Λ^m , and the matrix of the problem is $\mathbf{M}$ instead of $\mathbf{M} + \delta t_m \mathbf{A}^m$ in (6.25). Since

1. $\mathbf{M}$ is symmetric and positive definite,
2. generally, $\mathbf{M}$ has a much smaller condition number than $\mathbf{M} + \delta t_m \mathbf{A}^m$,

the linear complementarity problem (6.27) is easier to solve than (6.25). For (6.27), either the PSOR algorithm or projected gradient algorithms are efficient. Furthermore, these algorithms can be easily modified in order to use diagonal preconditioning for $\mathbf{M}$ when the finite-element mesh is highly nonuniform.

If the operator splitting scheme is used either with mass lumping (see Remark 5.5) or with finite-difference schemes instead of finite elements as in IKONEN and TOIVANEN [2004], then (6.27) becomes easier because $\mathbf{M}$ is replaced with a diagonal matrix and the projection substep consists of solving as many decoupled one-dimensional linear complementarity problems as there are nodes in the grid.

Other schemes can be modified to produce operator splitting schemes: for example, the Crank–Nicolson scheme

$$\begin{aligned} \mathbf{M}(\mathbf{U}^m - \mathbf{U}^{m-1}) + \frac{1}{2} \delta t_m (\mathbf{A}^m \mathbf{U}^m + \mathbf{A}^{m-1} \mathbf{U}^{m-1}) &= \delta t_m \Lambda^m, \\ \mathbf{U}^m &\geq \mathbf{U}^0, \\ \Lambda^m &\geq 0, \\ (\mathbf{U}^m - \mathbf{U}^0)^T \Lambda^m &= 0, \end{aligned} \quad (6.28)$$

becomes

- linear substep

$$\mathbf{M}(\tilde{\mathbf{U}}^m - \mathbf{U}^{m-1}) + \frac{1}{2} \delta t_m (\mathbf{A}^m \tilde{\mathbf{U}}^m + \mathbf{A}^{m-1} \mathbf{U}^{m-1}) = \delta t_m \Lambda^{m-1}, \quad (6.29)$$

- Projection substep (6.27).

The second-order accurate backward Euler scheme

$$\begin{aligned} M(U^m - \frac{4}{3}U^{m-1} + \frac{1}{3}U^{m-2}) + \frac{2}{3}\delta t_m A^m U^m &= \frac{2}{3}\delta t_m \Lambda^m, \\ U^m &\geq U^0, \\ \Lambda^m &\geq 0, \\ (U^m - U^0)^T \Lambda^m &= 0, \end{aligned} \tag{6.30}$$

becomes

- linear substep

$$M(\tilde{U}^m - \frac{4}{3}U^{m-1} + \frac{1}{3}U^{m-2}) + \frac{2}{3}\delta t_m A^m \tilde{U}^m = \frac{2}{3}\delta t_m \Lambda^{m-1}, \tag{6.31}$$

- projection substep

$$\begin{aligned} M(U^m - \tilde{U}^m) &= \frac{2}{3}\delta t_m (\Lambda^m - \Lambda^{m-1}), \\ U^m &\geq U^0, \\ \Lambda^m &\geq 0, \\ (U^m - U^0)^T \Lambda^m &= 0. \end{aligned} \tag{6.32}$$

It is possible to estimate the error produced by using the projection schemes above. This was done by IKONEN and TOIVANEN [2004] in the context of finite differences: for the Crank–Nicolson scheme with constant time steps and time-invariant coefficients, it was shown that the error between the solution to (6.28) and (6.29, 6.27) is of the order of δt^2 if some stability assumptions are satisfied (A is diagonally dominant with positive diagonal entries). Therefore, both the Crank–Nicolson and the projected Crank–Nicolson schemes are second order in time. It is also possible to obtain error estimates in Sobolev norms for the three schemes above by using the techniques in GUERMOND [1999], GUERMOND and QUARTAPELLE [1998].

In terms of computing time, we found that the projection scheme is clearly faster than the primal-dual algorithm above for a comparable accuracy. In our experience, the only difficulty posed by the projection schemes is that they are not fully compatible with a strategy of dynamic mesh adaption (i.e., when the mesh is adapted between two time steps).

REMARK 6.1. Operator splitting can also be applied to the penalized problems (6.23) or (6.24).

Example We use the projected Euler scheme (6.26, 6.27) for computing the pricing in the first example in Section 6.4.2. In Fig. 6.4, we plot the exercise region 1 year to maturity. This figure must be compared with Fig. 6.1(top). The difference is not visible.

6.4.6. Multigrid methods

Multigrid methods can also be used for linear complementarity problems: one possibility is to modify the primal-dual algorithm described above, recall that each iteration of

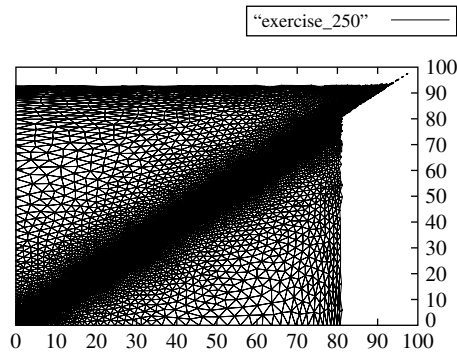


FIG. 6.4 The exercise region 1 year to maturity. $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\rho = -0.6$. This has to be compared with Fig. 6.1

such algorithms requires the solution to a linear boundary value problem in a varying subdomain. The idea is to use a multigrid method for these linear substeps. The main difficulty lies in the fact that one cannot obtain a hierarchy of meshes of the subdomains by simply taking subsets of the hierarchical meshes for the whole domain. We refer to HOPPE [1987] for such algorithms for finite-difference schemes. In the same context, but for finite elements, additive multilevel preconditioners were proposed by HOPPE and KORNHUBER [1994].

Another possible way is to design a multigrid cycle for the full nonlinear problem. This has been done by IKONEN and TOIVANEN [2006], OOSTERLEE [2003], REISINGER and WITTUM [2004].

6.4.7. Conclusion

We have presented different algorithms relevant in the context of American options pricing. The PSOR and primal-dual methods are algorithms for solving the linear complementarity problem at each time step. Multigrid methods can be used as a component of the primal-dual methods or for directly solving the linear complementarity problem. The penalty method replaces the linear complementarity problem by a penalized nonlinear problem. The projection schemes consist of dividing each time step into two substeps, the first one for diffusion and the other one to enforce the constraints.

In our experience, the projection schemes are fast (faster than the primal-dual methods not combined with multigrid), and their implementation is very simple and general. Multigrid methods may be faster, but their implementation certainly needs more care.

7. Stochastic volatility

7.1. Volatility models with one stochastic process

We consider a financial asset whose price is given by the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t, \quad (7.1)$$

where $\mu S_t dt$ is a drift term, (W_t) is a Brownian motion, and (σ_t) is the volatility. As we have seen before, the simplest models take a constant deterministic volatility, but these models are generally too coarse to match real market prices. Here, we assume that (σ_t) is a stochastic process taking nonnegative values, satisfying a stochastic differential equation driven by a second Brownian motion $\hat{Z}_t$, not perfectly correlated to W_t . More precisely, using the notations of FOUQUE, PAPANICOLAOU and SIRCAR [2000], we assume that $\sigma_t = f(Y_t)$, where f is some positive function and (Y_t) is the driving process; the most common choices for (Y_t) are

- lognormal process

$$dY_t = c_1 Y_t dt + c_2 Y_t d\hat{Z}_t, \quad (7.2)$$

where c_1 and c_2 are positive constants.

- mean reverting Orstein-Uhlenbeck (OU) process

$$dY_t = \alpha(m - Y_t)dt + \beta d\hat{Z}_t, \quad (7.3)$$

where α and β are positive constants.

- Cox–Ingersoll–Ross (CIR) process

$$dY_t = \kappa(m - Y_t)dt + \lambda\sqrt{Y_t}d\hat{Z}_t, \quad (7.4)$$

where κ , m , and λ are positive constants.

One of the important feature of the second and third processes is their mean-reversion: the drift term in the stochastic differential equation for Y_t pulls the process back to the long-run mean level m . For example, if (Y_t) is a mean reverting OU process satisfying (7.3), then the law of Y_t knowing Y_0 is

$$\mathcal{N}\left(m + (Y_0 - m)e^{-\alpha t}, \frac{\beta^2}{2\alpha}(1 - e^{-2\alpha t})\right).$$

Therefore, m is the limit of the mean value of Y_t as $t \rightarrow +\infty$, and $\frac{1}{\alpha}$ is the characteristic time of mean reversion. The parameter α is called the rate of mean reversion. The ratio $\frac{\beta^2}{2\alpha}$ is the limit of the variance of Y_t as $t \rightarrow +\infty$. The long-run distribution of the OU process is $\mathcal{N}\left(m, \frac{\beta^2}{2\alpha}\right)$.

The Brownian motion $\hat{Z}_t$ may be correlated to W_t , and it can be written as a linear combination of W_t and an independent Brownian motion Z_t ,

$$\hat{Z}_t = \rho W_t + \sqrt{1 - \rho^2} Z_t, \quad (7.5)$$

where the correlation factor ρ lies in $[-1, 1]$.

Table 7.1 summarizes the most popular choices of stochastic volatility models with one stochastic process.

TABLE 7.1
Frequently used models of stochastic volatilities

Authors	ρ	$f(y)$	Y_t process
HULL and WHITE [1987]	$\rho = 0$	$f(y) = \sqrt{y}$	lognormal
CHESNEY and SCOTT [1989]	$\rho = 0$	$f(y) = e^y$	mean reverting OU
STEIN [1991]	$\rho = 0$	$f(y) = y $	mean reverting OU
HESTON [1993]	$\rho \neq 0$	$f(y) = \sqrt{y}$	CIR

7.2. European options with stochastic volatility

Consider a European option on the previously mentioned asset, with expiration date T and payoff function $P_o(S_T)$. Its price at time $t < T$ will depend on t , the price of the underlying asset S_t , and Y_t . We denote by $P(S_t, Y_t, t)$ the price of the option and by $\tilde{r}(t)$ the interest rate.

The option is priced using a no-arbitrage principle and the two-dimensional Itô's formula. However, since there are two risk factors, it is not possible to construct a hedged portfolio containing simply one option and shares of the underlying asset. One says that the market is incomplete. Instead, one can try to build a hedged portfolio containing options with two different maturities and shares of the underlying assets; only for simplicity, we restrict the discussion to the case when (Y_t) is an OU process satisfying (7.3), with $(\tilde{Z}_t)$ given by (7.5), but what follows can be generalized to any Markovian Itô driving process:

$$dY_t = \mu_Y(t, Y_t)dt + \sigma_Y(t, Y_t)d\tilde{Z}_t.$$

Let us try to build a self-financing hedged portfolio containing a_t shares of the underlying asset, one option with expiration date T_1 whose price is

$$P_t^{(1)} = P^{(1)}(S_t, Y_t, t)$$

and b_t options with a larger expiration date $T_2 > T_1$, whose price is

$$P_t^{(2)} = P^{(2)}(S_t, Y_t, t).$$

The value of the portfolio is c_t . The no-arbitrage principle yields that for $t < T_1$,

$$dc_t = a_t dS_t + dP_t^{(1)} + b_t dP_t^{(2)} = \tilde{r}_t c_t dt = \tilde{r}_t (a_t S_t + P_t^{(1)} + b_t P_t^{(2)}) dt. \quad (7.6)$$

The two-dimensional Itô formula permits $dP_t^{(1)}$ and $dP_t^{(2)}$ to be expressed as combinations of dt , dW_t , and dZ_t . The right-hand side of (7.6) does not contain dZ_t , thus

$$b_t = -\frac{\frac{\partial P^{(2)}}{\partial y}}{\frac{\partial P^{(1)}}{\partial y}}.$$

From the last equation and since the right-hand side of (7.6) does not contain dW_t , we deduce

$$a_t + \frac{\partial P^{(1)}}{\partial S} + b_t \frac{\partial P^{(2)}}{\partial S} = 0.$$

Comparing the dt terms in (7.6) and substituting the values of a_t and b_t , we obtain

$$\begin{aligned} & \frac{1}{\frac{\partial P^{(1)}}{\partial y}} \left(\frac{\partial P^{(1)}}{\partial t} + \frac{1}{2} f(y)^2 S^2 \frac{\partial^2 P^{(1)}}{\partial S^2} + \rho \beta S f(y) \frac{\partial^2 P^{(1)}}{\partial S \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2 P^{(1)}}{\partial y^2} \right. \\ & \quad \left. + \tilde{r}(t) \left(S \frac{\partial P^{(1)}}{\partial S} - P^{(1)} \right) \right) = \\ & \frac{1}{\frac{\partial P^{(2)}}{\partial y}} \left(\frac{\partial P^{(2)}}{\partial t} + \frac{1}{2} f(y)^2 S^2 \frac{\partial^2 P^{(2)}}{\partial S^2} + \rho \beta S f(y) \frac{\partial^2 P^{(2)}}{\partial S \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2 P^{(2)}}{\partial y^2} \right. \\ & \quad \left. + \tilde{r}(t) \left(S \frac{\partial P^{(2)}}{\partial S} - P^{(2)} \right) \right). \end{aligned}$$

In the above equation, the left-hand side does not depend on T_2 and the right-hand side does not depend on T_1 . Therefore, there exists a function $g(S, y, t)$ such that

$$\begin{aligned} & \frac{1}{\frac{\partial P}{\partial y}} \left(\frac{\partial P}{\partial t} + \frac{1}{2} f(y)^2 S^2 \frac{\partial^2 P}{\partial S^2} + \rho \beta S f(y) \frac{\partial^2 P}{\partial S \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2 P}{\partial y^2} + \tilde{r}(t) \left(S \frac{\partial P}{\partial S} - P \right) \right) \\ & = g(S, y, t). \end{aligned}$$

Writing $g(S, y, t) = \alpha(y - m) + \beta \tilde{\Lambda}(S, y, t)$ makes the infinitesimal generator of the OU process explicit in the last equation. We obtain

$$\begin{aligned} & \frac{\partial P}{\partial t} + \frac{1}{2} f(y)^2 S^2 \frac{\partial^2 P}{\partial S^2} + \rho \beta S f(y) \frac{\partial^2 P}{\partial S \partial y} + \frac{1}{2} \beta^2 \frac{\partial^2 P}{\partial y^2} \\ & + \tilde{r}(t) \left(S \frac{\partial P}{\partial S} - P \right) + \left(\alpha(m - y) - \beta \tilde{\Lambda}(S, y, t) \right) \frac{\partial P}{\partial y} = 0, \quad \begin{cases} 0 \leq t < T, \\ S > 0, y \in \mathbb{R}, \end{cases} \end{aligned} \quad (7.7)$$

where

$$\tilde{\Lambda}(S, y, t) = \rho \frac{\mu - \tilde{r}(t)}{f(y)} + \sqrt{1 - \rho^2} \tilde{\gamma}(S, y, t), \quad (7.8)$$

with the terminal condition $P(S, y, T) = P_\circ(S)$.

The function $\tilde{\gamma}(S, y, t)$ (return on the volatility risk) can be chosen arbitrarily.

As explained by FOUQUE, PAPANICOLAOU and SIRCAR [2000], we can group the differential operator in (7.7) as follows:

$$\begin{aligned} & \underbrace{\frac{\partial P}{\partial t} + \frac{1}{2}f(y)^2S^2\frac{\partial^2 P}{\partial S^2} + \tilde{r}(t)\left(S\frac{\partial P}{\partial S} - P\right)}_{BS_{f(y)}} + \underbrace{\rho\beta Sf(y)\frac{\partial^2 P}{\partial S\partial y}}_{\text{correlation}} \\ & + \underbrace{\frac{1}{2}\beta^2\frac{\partial^2 P}{\partial y^2} + \alpha(m-y)\frac{\partial P}{\partial y}}_{\text{Orstein Uhlenbeck}} - \underbrace{\beta\tilde{\Lambda}(S, y, t)\frac{\partial P}{\partial y}}_{\text{premium}}. \end{aligned} \quad (7.9)$$

The term $\beta\tilde{\Lambda}(S, y, t)\frac{\partial P}{\partial y}$ is a premium on the volatility risk: the reason to decompose $\tilde{\Lambda}$ as in (7.8) is that in the perfectly correlated case ($|\rho| = 1$, complete market), it is possible to find the equation satisfied by P by a simpler no-arbitrage argument with a hedged portfolio containing only the option and shares of the underlying assets. In this case, the equation found for P is

$$\begin{aligned} & \frac{\partial P}{\partial t} + \frac{1}{2}f(y)^2S^2\frac{\partial^2 P}{\partial S^2} + \rho\beta Sf(y)\frac{\partial^2 P}{\partial S\partial y} + \frac{1}{2}\beta^2\frac{\partial^2 P}{\partial y^2} \\ & + \tilde{r}(t)\left(S\frac{\partial P}{\partial S} - P\right) + \left(\alpha(m-y) - \beta\rho\frac{\mu - \tilde{r}(t)}{f(y)}\right)\frac{\partial P}{\partial y} = 0, \quad \begin{cases} 0 \leq t < T, \\ S > 0, y \in \mathbb{R}. \end{cases} \end{aligned} \quad (7.10)$$

The term $\frac{\mu - \tilde{r}(t)}{f(y)}$ is called the excess return to risk ratio.

Finally, with (7.7), the Itô formula and (7.8)

$$dP(S_t, Y_t, t) = (Sf(Y_t)\frac{\partial P}{\partial S} + \beta\rho\frac{\partial P}{\partial y})(\frac{\mu - \tilde{r}}{f(Y_t)}dt + dW_t) + \beta\sqrt{1 - \rho^2}\frac{\partial P}{\partial y}(\tilde{\gamma}dt + dZ_t)$$

from which we see that the function $\tilde{\gamma}$ is the contribution of the second source of randomness dZ_t to the risk premium. The function $\tilde{\gamma}$ is called the market price of the volatility risk or the *risk premium factor*.

Similarly, assuming that (Y_t) is a CIR process satisfying (7.4), one obtains

$$\begin{aligned} & \underbrace{\frac{\partial P}{\partial t} + \frac{1}{2}f(y)^2S^2\frac{\partial^2 P}{\partial S^2} + \tilde{r}(t)\left(S\frac{\partial P}{\partial S} - P\right)}_{BS_{f(y)}} + \underbrace{\rho\lambda S\sqrt{y}f(y)\frac{\partial^2 P}{\partial S\partial y}}_{\text{correlation}} \\ & + \underbrace{\frac{1}{2}\lambda^2y\frac{\partial^2 P}{\partial y^2} + \kappa(m-y)\frac{\partial P}{\partial y}}_{\text{CIR}} - \underbrace{\lambda\sqrt{y}\tilde{\Lambda}(S, y, t)\frac{\partial P}{\partial y}}_{\text{premium}} = 0, \end{aligned} \quad (7.11)$$

where $\tilde{\Lambda}(S, y, t)$ is given by (7.8).

REMARK 7.1. It is possible to obtain (7.7) and (7.11) by using a more mathematically sound risk-neutral theory and the market price of the volatility risk appears from Girsanov's theorem (see FOUQUE, PAPANICOLAOU and SIRCAR [2000]), §2.5.

REMARK 7.2. For the Heston model, a closed-form solution in terms of integrals is available (see HESTON [1993]).

The initial value problem for Stein–Stein’s model We discuss the mathematical analysis of the initial value problem with (7.7) in the case when $\rho = 0$ and $f(y) = |y|$ (Stein–Stein’s model). The goal is to study variational formulations of (7.7), and obtain global energy estimates. These estimates are useful for studying discrete approximations by, for example, FEM. Variational formulations are also particularly useful for the linear complementarity problems obtained when pricing American options. This paragraph summarizes the results contained in ACHDOU and TCHOU [2002] and ACHDOU, FRANCHI, and TCHOU [2005].

For simplicity, we assume that the market price of risk $\tilde{\gamma}$ is bounded independently of t , S , and y . The variance of the invariant distribution of the OU process, that is, $v^2 = \frac{\beta^2}{2\alpha}$, will play an important role in what follows.

In order to obtain a forward parabolic equation, we work with the time to maturity, that is, $T - t \rightarrow t$. With the aim of deriving a variational formulation, we make the change of unknown

$$u(S, y, t) = P(S, y, T - t)e^{-(1-\eta)\frac{(y-m)^2}{2v^2}}, \quad (7.12)$$

where η is a parameter such that $0 < \eta < 1$; we are going to impose that u tends to 0 as y tends to ∞ . Indeed, if $\tilde{\Lambda} = 0$, then one can find a solution to (7.7) of the form $g(t)e^{\frac{(y-m)^2}{2v^2}}$; imposing that $u(S, y, t)$ tends to 0 as y tends to ∞ prevents such a behavior for large values of y . The parameter η will not be important for practical computations because in any case, we have to truncate the domain and suppress large values of y .

With the notations $r(t) = \tilde{r}(T - t)$, $\gamma(t) = \tilde{\gamma}(T - t)$, and $\Lambda(t) = \tilde{\Lambda}(T - t)$, the new unknown u satisfies the degenerate parabolic PDE

$$\begin{aligned} & \frac{\partial u}{\partial t} - \frac{1}{2}y^2S^2\frac{\partial^2 u}{\partial S^2} - r(t)\left(S\frac{\partial u}{\partial S} - u\right) - \frac{1}{2}\beta^2\frac{\partial^2 u}{\partial y^2} \\ & + (-\alpha(y - m) + \beta\Lambda(S, y, t))\frac{\partial u}{\partial y} + \left(2\frac{\alpha}{\beta}\Lambda(S, y, t)(y - m) - \alpha\right)u \\ & + \eta\left(2\alpha(y - m)\frac{\partial u}{\partial y} + 2\frac{\alpha^2}{\beta^2}(1 - \eta)(y - m)^2u - 2\frac{\alpha}{\beta}\Lambda(y - m)u + \alpha u\right) = 0. \end{aligned} \quad (7.13)$$

The equation is degenerate near the axis $y = 0$ because the coefficient in front of $S^2\frac{\partial^2 P}{\partial S^2}$ vanishes on $y = 0$.

Expanding Λ and denoting by $\mathcal{L}_t$ the linear partial differential operator

$$\begin{aligned} & \mathcal{L}_t v \\ & = -\frac{1}{2}y^2S^2\frac{\partial^2 v}{\partial S^2} - \frac{1}{2}\beta^2\frac{\partial^2 v}{\partial y^2} - r(t)S\frac{\partial v}{\partial S} + (-(1 - 2\eta)\alpha(y - m) + \beta\gamma(S, y, t))\frac{\partial v}{\partial y} \\ & + \left(r(t) + 2\frac{\alpha^2}{\beta^2}\eta(1 - \eta)(y - m)^2 + 2(1 - \eta)\frac{\alpha}{\beta}(y - m)(\gamma(S, y, t) - \alpha(1 - \eta))\right)v, \end{aligned} \quad (7.14)$$

we obtain

$$\frac{\partial u}{\partial t} + \mathcal{L}_t u = 0. \quad (7.15)$$

We denote by Q the open half-plane $Q = \mathbb{R}_+ \times \mathbb{R}$. Let us consider the weighted Sobolev space V

$$V = \left\{ v : \left(\sqrt{1+y^2}v, \frac{\partial v}{\partial y}, S|y|\frac{\partial v}{\partial S} \right) \in (L^2(Q))^3 \right\}. \quad (7.16)$$

This space with the norm

$$|||v|||_V = \left(\int_Q (1+y^2)v^2 + \left(\frac{\partial v}{\partial y}\right)^2 + S^2 y^2 \left(\frac{\partial v}{\partial S}\right)^2 \right)^{\frac{1}{2}} \quad (7.17)$$

is a Hilbert space, and it has the following properties:

1. V is separable.
2. Calling $\mathcal{D}(Q)$ the space of smooth functions with compact support in Q , $\mathcal{D}(Q) \subset V$ and $\mathcal{D}(Q)$ is dense in V .
3. V is dense in $L^2(Q)$.

The crucial point is point 7.2, which can be proved by an argument due to Friedrichs (theorem 4.2 in FRIEDRICHS [1944]). We also have the following:

LEMMA 7.1. *For any function v in V ,*

$$\int_Q y^2 v^2 \leq 4 \int_Q y^2 S^2 \left(\frac{\partial v}{\partial S}\right)^2. \quad (7.18)$$

The seminorm

$$\|v\|_V = \left(\int_Q v^2 + \left(\frac{\partial v}{\partial y}\right)^2 + S^2 y^2 \left(\frac{\partial v}{\partial S}\right)^2 \right)^{\frac{1}{2}} \quad (7.19)$$

is, in fact, a norm in V , equivalent to $|||\cdot|||_V$.

We call V' the dual of V . In order to use the general theory of LIONS and MAGENES [1968] on parabolic equations, we first need to prove the following lemma.

LEMMA 7.2. *The operator $v \rightarrow S \frac{\partial v}{\partial S}$ is continuous from V into V' .*

PROOF. Call X and Y the differential operators

$$X(v) = Sy \frac{\partial v}{\partial S} + \beta \frac{\partial v}{\partial y}, \quad Y(v) = Sy \frac{\partial v}{\partial S} - \beta \frac{\partial v}{\partial y}. \quad (7.20)$$

The operators X and Y are continuous operators from V into $L^2(Q)$ and their adjoints are

$$\begin{aligned} X^T(v) &= -S y \frac{\partial v}{\partial S} - \beta \frac{\partial v}{\partial y} - yv = -X(v) - yv, \\ Y^T v &= -S y \frac{\partial v}{\partial S} + \beta \frac{\partial v}{\partial y} - yv = -Y(v) - yv. \end{aligned} \quad (7.21)$$

Consider the commutator $[X, Y] = XY - YX$, it can be checked that

$$[X, Y](v) = 2\beta S \frac{\partial v}{\partial S}. \quad (7.22)$$

Therefore, for $v \in V$ and $w \in \mathcal{D}(Q)$,

$$(2\beta S \frac{\partial v}{\partial S}, w) = - \int_Q Y(v)(X(w) + yw) + \int_Q X(v)(Y(w) + yw), \quad (7.23)$$

and from (7.18), there exists a constant C such that

$$|(2\beta S \frac{\partial v}{\partial S}, w)| \leq C \|v\|_V \|w\|_V.$$

To conclude, we use the density of $\mathcal{D}(Q)$ into V . □

Lemma 7.2 implies that the operator $\mathcal{L}_t$ is continuous from V to its dual V' . Calling a_t the bilinear form defined on $V \times V$ by $a_t(u, v) = \langle \mathcal{L}_t u, v \rangle$, we have

$$\begin{aligned} a_t(u, v) &= \frac{1}{2} \int_Q y^2 S^2 \frac{\partial u}{\partial S} \frac{\partial v}{\partial S} + \int_Q y^2 S \frac{\partial u}{\partial S} v + \frac{\beta^2}{2} \int_Q \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} \\ &\quad + \frac{r(t)}{2\beta} \left(\int_Q Y(u)(X(v) + yv) - \int_Q X(u)(Y(v) + yv) \right) \\ &\quad + \int_Q ((2\eta - 1)\alpha(y - m) + \beta\gamma(S, y, t)) \frac{\partial u}{\partial y} v \\ &\quad + \int_Q \left(r(t) + (1 - \eta) \left(2 \frac{\alpha^2}{\beta^2} \eta(y - m)^2 + 2 \frac{\alpha}{\beta} (y - m)\gamma(S, y, t) - \alpha \right) \right) uv. \end{aligned} \quad (7.24)$$

PROPOSITION 7.1. *Assume that r is a bounded function of time and γ is bounded by a constant. The bilinear form a_t is continuous on $V \times V$, with a continuity constant independent of t .*

We also need a Gårding inequality.

PROPOSITION 7.2. *Assume that r is a bounded function of time and γ is bounded by a constant Γ . If $\alpha > \beta$, then there exist two positive constants C and c independent of t*

and two constants, $0 < \eta_1 < \eta_2 < 1$, such that, for $\eta_1 < \eta < \eta_2$ and for any $v \in V$,

$$a_t(v, v) \geq C \|v\|_V^2 - c \|v\|_{L^2(Q)}^2. \quad (7.25)$$

From Propositions 7.1 and 7.2, we can prove the existence and uniqueness of weak solutions to the initial value problem with (7.15).

THEOREM 7.1. *Assume that $\alpha > \beta$ and η has been chosen as in Proposition 7.2. Then, for any $u_\circ \in L^2(Q)$, there exists a unique u in $L^2(0, T; V) \cap C^0([0, T]; L^2(Q))$, with $\frac{\partial u}{\partial t} \in L^2(0, T; V')$ such that for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in V$,*

$$-\int_0^T \phi'(t) \left(\int_Q u(t)v \right) dt + \int_0^T \phi(t) a_t(u, v) dt = 0 \quad (7.26)$$

and

$$u(t=0) = u_\circ. \quad (7.27)$$

The mapping $u_\circ \mapsto u$ is continuous from $L^2(Q)$ to $L^2(0, T; V) \cap C^0([0, T]; L^2(Q))$.

REMARK 7.3. The ratio $\frac{2\alpha^2}{\beta^2}$ is exactly the ratio between the rate of mean reversion and the asymptotic variance of the volatility. The assumption in Theorem 7.1 says that the rate of mean reversion should not be too small compared with the asymptotic variance of the volatility. This condition is usually satisfied in practice since α is often much larger than the asymptotic variance $\frac{\beta^2}{2\alpha}$.

It is possible to prove a maximum principle (see ACHDOU and TCHOU [2002]): as a consequence, in the case of a vanilla put, we see that the weak solution given by Theorem 7.1 has a financially correct behavior.

PROPOSITION 7.3. *Assume that the coefficients are smooth and bounded and that $\alpha > \beta$. If $P_\circ(S, y) = (K - S)^+$, then the function*

$$P(t, S, y) = u(T - t, S, y) e^{(1-\eta) \frac{(y-m)^2}{2v^2}},$$

where u is the solution to (7.26), (7.27) with $u_\circ = e^{(1-\eta) \frac{(y-m)^2}{2v^2}} P_\circ$, satisfies

$$(S - Ke^{-r(T-t)})^- \leq P(t, S, y) \leq Ke^{-r(T-t)}, \quad (7.28)$$

and we have the put-call parity $C(t, S, y) - P(t, S, y) = S - Ke^{-r(T-t)}$ if C is the pricing function of the corresponding call option.

Consider now $\mathcal{L}_t$ as an unbounded operator defined on $L^2(Q)$ and call D_t the domain of $\mathcal{L}_t$, that is, $\{v \in V \text{ s.t. } \mathcal{L}_t v \in L^2(Q)\}$. In ACHDOU, FRANCHI, and TCHOU [2005], have shown that D_t does not depend on t .

THEOREM 7.2. *If for all t , $r(t) > 0$, then D_t does not depend on t : $D_t = D$.*

Moreover, if there exists a constant $r_0 > 0$ such that $r(t) > r_0$ a.e., and $\frac{\alpha^2}{\beta^2} > 2$, then for well-chosen values of η (in particular such that $2\frac{\alpha^2}{\beta^2}\eta(1 - \eta) > 1$),

$$D = \left\{ v \in V; y^2 S^2 \frac{\partial^2 v}{\partial S^2}, \frac{\partial^2 v}{\partial y^2}, yS \frac{\partial^2 v}{\partial S \partial y}, S \frac{\partial v}{\partial S}, y \frac{\partial v}{\partial y}, y^2 v \in L^2(Q) \right\}. \quad (7.29)$$

Then, from general results of Kato (see PAZY [1983] theorem 5.6.8.), regularity results on the solution to (7.26) (7.27) can be obtained.

THEOREM 7.3. *Assume that there exists ζ , $0 < \zeta \leq 1$, such that γ belongs to $C^\zeta([0, T], L^\infty(Q))$ and r is a Hölder function of time with exponent ζ . Assume also that $r(t) > r_0$ for a positive constant r_0 and $\frac{\alpha^2}{\beta^2} > 2$. Then, for η chosen as in Proposition 7.2 and Theorem 7.2 if $u_\circ$ belongs to D defined by (7.29), then the solution u to (7.26) (7.27) belongs to $C^1((0, T); L^2(Q)) \cap C^0([0, T]; D)$, and the functional equation in $L^2(Q)$*

$$u'(t) + \mathcal{L}_t u(t) = 0 \quad (7.30)$$

is satisfied pointwise in $[0, T]$.

Furthermore, for $u_\circ \in L^2(Q)$, the solution to (7.26) (7.27) also belongs to $C^1((\tau, T); L^2(Q)) \cap C^0([\tau, T]; D)$, for all $\tau > 0$, and satisfies

$$\|u'(t)\|_{L^2(Q)} + \|\mathcal{L}_t u(t)\|_{L^2(Q)} \leq \frac{C}{t}, \quad \text{for } t > 0.$$

REMARK 7.4. The same kind of analysis is possible for an extended Stein–Stein’s model with a nonzero correlation factor, but in this case (still assuming that $\tilde{\gamma}$ is bounded), one has to cope with the term $\rho \frac{\mu - \tilde{r}(t)}{|y|} \frac{\partial P}{\partial y}$, which becomes singular on the axis $y = 0$; therefore, one may need to impose a Dirichlet condition on the axis $y = 0$ of the form

$$P(S, 0, t) = g(S, t), \quad 0 \leq t < T, \quad S > 0, \quad (7.31)$$

where

$$\begin{aligned} \frac{\partial g}{\partial t} + \tilde{r}(t) \left(S \frac{\partial g}{\partial S} - g \right) &= 0 \quad 0 \leq t < T, \quad S > 0, \\ g(S, t = T) &= P_\circ(S, 0). \end{aligned} \quad (7.32)$$

7.2.1. The initial value problem for Heston’s model

We aim at making an analysis for Heston’s model in the same spirit as the one proposed above for the Stein–Stein’s model. We consider the PDE (7.11), in the simple case when $\rho = 0$. We also assume that $\tilde{\gamma}$ is bounded independently of t , S and y , and we define $\gamma(S, y, t) = \tilde{\gamma}(S, y, T - t)$. We need to set the variational formulation in a suitable weighted Sobolev space compatible with the operator degeneracy on the axis $y = 0$. For that, we introduce a smooth positive function ψ defined on $\mathbb{R}_+$ such that $\psi(y) = \sqrt{y}$

on $(0, m)$ and $\psi(y) = e^{-dy}$ on $(2m, +\infty)$ for some positive parameter d . We will fix d later.

We introduce the new unknown function

$$u(S, y, t) = \psi(y)P(S, y, T - t). \quad (7.33)$$

Denoting by $\mathcal{L}_t$ the linear partial differential operator

$$\begin{aligned} \mathcal{L}_t v = & -\frac{1}{2}yS^2 \frac{\partial^2 v}{\partial S^2} - r(t) \left(S \frac{\partial v}{\partial S} - v \right) \\ & - \frac{\lambda^2}{2}y \left(\frac{\partial^2 v}{\partial y^2} - 2 \frac{\psi'}{\psi} \frac{\partial v}{\partial y} - \psi \left(\frac{\psi'}{\psi^2} \right)' v \right) - \kappa(m - y) \left(\frac{\partial v}{\partial y} - \frac{\psi'}{\psi} v \right) \\ & + \lambda \sqrt{y} \gamma(S, y, t) \left(\frac{\partial v}{\partial y} - \frac{\psi'}{\psi} v \right), \end{aligned} \quad (7.34)$$

we obtain

$$\frac{\partial u}{\partial t} + \mathcal{L}_t u = 0. \quad (7.35)$$

The equation clearly becomes degenerate on the axis $y = 0$ because the coefficients in front of the two operators $\frac{\partial^2 u}{\partial y^2}$ and $S^2 \frac{\partial^2 u}{\partial S^2}$ vanish. We denote by Q the open sector $Q = \mathbb{R}_+ \times \mathbb{R}_+$. Let us consider the weighted Sobolev space V :

$$V = \left\{ v : \left(v, \frac{v}{\sqrt{y}}, \sqrt{y} \frac{\partial v}{\partial y}, \sqrt{y} S \frac{\partial v}{\partial S} \right) \in (L^2(Q))^4 \right\}. \quad (7.36)$$

This space with the norm

$$\|v\|_V = \left(\int_Q \left(1 + \frac{1}{y} \right) v^2 + y \left(\frac{\partial v}{\partial y} \right)^2 + y S^2 \left(\frac{\partial v}{\partial S} \right)^2 \right)^{\frac{1}{2}} \quad (7.37)$$

is a Hilbert space, and it has the following properties:

1. V is separable.
2. Calling $\mathcal{D}(Q)$ the space of smooth functions with compact support in Q , $\mathcal{D}(Q) \subset V$ and $\mathcal{D}(Q)$ are dense in V
3. V is dense in $L^2(Q)$.
4. For any function v in V ,

$$\int_Q y v^2 \leq 4 \int_Q y S^2 \left(\frac{\partial v}{\partial S} \right)^2. \quad (7.38)$$

REMARK 7.5. The reason for imposing that $\frac{v}{\sqrt{y}}$ be square integrable will appear in Lemma 7.3. Note that the functions $v(y) = \log^\sigma(y)$, with $0 < \sigma < \frac{1}{2}$, are such that v and $\sqrt{y}v'$ are square integrable near $y = 0$, but $\frac{v}{\sqrt{y}}$ is not square integrable.

LEMMA 7.3. *The operator $v \rightarrow S \frac{\partial v}{\partial S}$ is continuous from V into V' .*

PROOF. Call X and Y the differential operators

$$X(v) = \sqrt{y} S \frac{\partial v}{\partial S}, \quad Y(v) = \sqrt{y} \frac{\partial v}{\partial y}, \quad (7.39)$$

The operators X and Y are continuous operators from V into $L^2(Q)$, and their adjoints are

$$X^T(v) = -X(v) - \sqrt{y}v, \quad Y^T v = -Y(v) - \frac{1}{2\sqrt{y}}v. \quad (7.40)$$

It can be checked that

$$[X, Y](v) = -\frac{1}{2} S \frac{\partial v}{\partial S}. \quad (7.41)$$

Therefore, for $v \in V$ and $w \in \mathcal{D}(Q)$,

$$\left(S \frac{\partial v}{\partial S}, w \right) = 2 \int_Q Y(v)(X(w) + \sqrt{y}w) - 2 \int_Q X(v) \left(Y(w) + \frac{1}{2\sqrt{y}}w \right), \quad (7.42)$$

and from (7.38), there exists a constant C such that

$$\left| \left(S \frac{\partial v}{\partial S}, w \right) \right| \leq C \|v\|_V \|w\|_V.$$

To conclude, we use the density of $\mathcal{D}(Q)$ into V . □

Lemma 7.3 and the assumption made on ψ imply that the operator $\mathcal{L}_t$ is continuous from V to its dual V' because, in particular, the functions $\frac{\psi'}{\psi}$, $\frac{1}{y} \frac{d}{dy} (y \frac{\psi'}{\psi})$, and $\psi \frac{d}{dy} (\frac{\psi'}{\psi^2})$ are bounded on $[2m, +\infty)$.

Calling a_t the bilinear form defined on $V \times V$ by $a_t(u, v) = \langle \mathcal{L}_t u, v \rangle$, we have

$$\begin{aligned} a_t(u, v) = & \frac{1}{2} \int_Q y S^2 \frac{\partial u}{\partial S} \frac{\partial v}{\partial S} + \int_Q y S \frac{\partial u}{\partial S} v \\ & + \frac{\lambda^2}{2} \left(\int_Q y \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} + \int_Q \frac{\partial u}{\partial y} v + 2 \int_Q y \frac{\psi'}{\psi} \frac{\partial u}{\partial y} v + \int_Q y \psi \left(\frac{\psi'}{\psi^2} \right)' uv \right) \\ & - 2r(t) \left(\int_Q Y(u)(X(v) + \sqrt{y}v) - \int_Q X(u)(Y(v) + \frac{1}{2\sqrt{y}}v) \right) + r(t) \int_Q uv \\ & - \kappa \int_Q (m - y) \left(\frac{\partial u}{\partial y} - \frac{\psi'}{\psi} u \right) v + \lambda \int_Q \sqrt{y} \gamma(t, S, y) \left(\frac{\partial u}{\partial y} - \frac{\psi'}{\psi} u \right) v. \end{aligned} \quad (7.43)$$

Assume that r is a bounded function of time and γ is bounded by a constant. The bilinear form a_t is continuous on $V \times V$, with a continuity constant independent of t . We also need a Gårding type inequality.

PROPOSITION 7.4. *Assume that r is a bounded function of time and γ is bounded by a constant Γ . If*

$$\kappa \min(m, \kappa) > \frac{3}{4}\lambda^2, \quad (7.44)$$

one can choose d (see the definition of ψ) such that there exist two positive constants C and c independent of t ,

$$a_t(v, v) \geq C\|v\|_V^2 - c\|v\|_{L^2(Q)}^2, \quad \forall v \in V. \quad (7.45)$$

PROOF. It is enough to prove (7.45) for $v \in \mathcal{D}(Q)$. Several integrations by part lead to

$$\begin{aligned} a_t(v, v) = & \frac{1}{2} \int_Q y S^2 \left(\frac{\partial v}{\partial S} \right)^2 - \frac{1}{2} \int_Q y v^2 + \frac{3}{2} r(t) \int_Q v^2 + \frac{\lambda^2}{2} \int_Q y \left(\frac{\partial v}{\partial y} \right)^2 \\ & - \int_Q \left(\frac{\lambda^2}{2} \left(y \frac{\psi'}{\psi} \right)' + \kappa y \frac{\psi'}{\psi} \right) v^2 + \lambda \int_Q \sqrt{y} \gamma(t, S, y) \left(\frac{\partial v}{\partial y} - \frac{\psi'}{\psi} v \right) v \\ & + \int_Q \left(\kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) - \frac{\kappa}{2} \right) v^2. \end{aligned}$$

For brevity, we skip many details, and we only focus on the main two steps of the proof.

First Step

Note that if $y < m$, then $\kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) = \frac{1}{2y} (\kappa m - \frac{3}{4}\lambda^2)$. From this and (7.44), we see that the quantity $\int_{Q \cap \{y < m\}} \frac{v^2}{y}$ is bounded by a positive factor times $\int_{Q \cap \{y < m\}} \left(\kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) \right) v^2$. On the other hand, the quantity

$$\begin{aligned} & \int_{Q \cap \{y < 2m\}} \left| -\frac{1}{2}y - \left(\frac{\lambda^2}{2} \left(y \frac{\psi'}{\psi} \right)' + \kappa y \frac{\psi'}{\psi} \right) \right| v^2 \\ & + \int_{Q \cap \{m < y < 2m\}} \left| \kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) \right| v^2 \end{aligned}$$

is bounded by a constant times $\int_{Q \cap \{y < 2m\}} v^2$.

Second Step

Calling

$$I = \int_{Q \cap \{y > 2m\}} \left(-\frac{1}{2}y - \left(\frac{\lambda^2}{2} \left(y \frac{\psi'}{\psi} \right)' + \kappa y \frac{\psi'}{\psi} \right) + \left(\kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) \right) \right) v^2,$$

we choose ψ such that there is an estimate of the form

$$I \geq -\alpha \int_Q y S^2 \left(\frac{\partial v}{\partial S} \right)^2 - \beta \int_Q v^2,$$

for some $\alpha < \frac{1}{2}$ and $\beta \geq 0$. In view of (7.38), this will be a consequence of the estimate

$$I \geq -\frac{\alpha}{4} \int_Q y v^2 - \beta \int_Q v^2.$$

Therefore, let us look for ψ such that

$$\begin{aligned} \int_{Q \cap \{y > 2m\}} & \left(\left(\frac{\alpha}{4} - \frac{1}{2} \right) y - \left(\frac{\lambda^2}{2} \left(y \frac{\psi'}{\psi} \right)' + \kappa y \frac{\psi'}{\psi} \right) \right. \\ & \left. + \left(\kappa m \frac{\psi'}{\psi} + \frac{\lambda^2}{2} \left(y \psi \left(\frac{\psi'}{\psi^2} \right)' \right) \right) \right) v^2 \geq -\beta \int_Q v^2. \end{aligned}$$

Since $\psi(y) = e^{-dy}$, $y > 2m$, the integrand above is

$$\left(\frac{\alpha}{4} - \frac{1}{2} \right) y - \frac{1}{2} \lambda^2 (d^2 y - d) + \kappa(y - m)d,$$

so the estimate will be true if one can choose d such that $-\frac{1}{2} \lambda^2 d^2 + \kappa d - \frac{3}{8} > 0$. This is the case if $\kappa^2 > \frac{3}{4} \lambda^2$. $\square$

The continuity of a_t and the Gårding's inequality yield the existence and uniqueness of a weak solution to the initial value problem with equation (7.35).

THEOREM 7.4. *Under the assumptions of Proposition 7.4, for any $u_o \in L^2(Q)$, there exists a unique u in $L^2(0, T; V) \cap C^0([0, T]; L^2(Q))$, with $\frac{\partial u}{\partial t} \in L^2(0, T; V')$ such that for any smooth function $\phi \in \mathcal{D}(0, T)$, for any $v \in V$,*

$$-\int_0^T \phi'(t) \left(\int_Q u(t)v \right) dt + \int_0^T \phi(t) a_t(u, v) dt = 0 \quad (7.46)$$

and

$$u(t=0) = u_o. \quad (7.47)$$

The mapping $u_o \mapsto u$ is continuous from $L^2(Q)$ to $L^2(0, T; V) \cap C^0([0, T]; L^2(Q))$.

REMARK 7.6. The assumption (7.44) plays the same role as the assumption $\alpha > \beta$ in the discussion of the Stein–Stein's model.

REMARK 7.7. Note that no boundary condition has been imposed on ∂Q .

For example, if $P_o(S) = (K - S)^+$, then $u_o(S, y) = \psi(y)(K - S)^+$ is a square-integrable function, and Theorem 7.4 can be applied to the case of a European put.

A numerical method for pricing options with Stein–Stein’s model We consider the PDE (7.9) with $f(y) = |y|$. We assume that the interest rate r is constant, and we take $\rho = 0$, and $\tilde{\gamma} = 0$. The partial differential equation is rewritten in divergence form, with the new variable $T - t \rightarrow t$:

$$\begin{aligned} 0 = & \frac{\partial P}{\partial t} - \frac{1}{2} \left(\frac{\partial}{\partial S} \left(y^2 S^2 \frac{\partial P}{\partial S} \right) + \frac{\partial}{\partial y} \left(\beta^2 \frac{\partial P}{\partial y} \right) \right) \\ & + (y^2 S - rS) \frac{\partial P}{\partial S} + \alpha(y - m) \frac{\partial P}{\partial y} + rP. \end{aligned} \quad (7.48)$$

A first-order implicit Euler scheme is used for time discretization:

$$\begin{aligned} 0 = & \frac{P^m - P^{m-1}}{\delta t} - \frac{1}{2} \left(\frac{\partial}{\partial S} \left(y^2 S^2 \frac{\partial P^m}{\partial S} \right) + \frac{\partial}{\partial y} \left(\beta^2 \frac{\partial P^m}{\partial y} \right) \right) \\ & + (y^2 S - rS) \frac{\partial P^m}{\partial S} + \alpha(y - m) \frac{\partial P^m}{\partial y} + rP^m. \end{aligned} \quad (7.49)$$

Aiming at obtaining a discrete version of this equation, we first truncate the domain, that is, we introduce the rectangle $\Omega = (0, \bar{S}) \times (-\bar{y}, \bar{y})$, with $\bar{S}$ and $\bar{y}$ large enough. We are looking for a numerical approximation of P in Ω . For that, we first need to supply reasonable artificial boundary conditions on boundaries of $\partial\Omega$, in agreement with the payoff function. Let us discuss the artificial boundary condition for a vanilla put option, that is, $P_o(S) = (K - S)^+$.

- On $\partial\Omega \cap \{S = \bar{S}\}$, we impose $\frac{\partial P}{\partial S}(\bar{S}, y, t) = 0$. Such a condition is reasonable if $\bar{S}$ is large enough compared with K .
- On $\partial\Omega \cap \{y = \pm\bar{y}\}$, finding an accurate artificial boundary condition is not easy. However, if $\bar{y}$ is chosen such that $\alpha(\bar{y} \pm m) \gg \beta^2$, then for $y \sim \pm\bar{y}$, the coefficient of the advection term in the y direction, $\alpha(y - m)$, is much larger in absolute value than the coefficient of the diffusion in the y direction, $\frac{\beta^2}{2}$. Furthermore, near $y = \pm\bar{y}$, the vertical velocity $\alpha(y - m)$ is directed outward Ω . Therefore, the error caused by an artificial condition on $y = \pm\bar{y}$ will be damped away from the boundaries and localized in boundary layers whose width is of the order of $\frac{\beta^2}{\alpha\bar{y}}$. Therefore, Dirichlet boundary conditions, $P(S, \pm\bar{y}, t) = 0$, will not cause a large error for $|y|$ small enough compared with $\bar{y}$, for example $|y| \leq \frac{\bar{y}}{2}$, even though these conditions are not satisfied at all by the exact solution.
- No boundary condition is needed on $S = 0$ because of the degeneracy of the equation.

Let V_h be the space of continuous piecewise linear functions on a triangulation of Ω , which are equal to zero on $y = \pm \bar{y}$. We consider the following finite-element discretization: $\forall v_h \in V_h$,

$$\begin{aligned} \int_{\Omega} \left(\frac{1}{\delta t} + r \right) P^m v + \frac{1}{2} \int_{\Omega} y^2 S^2 \frac{\partial P^m}{\partial S} \frac{\partial v}{\partial S} + \frac{\beta^2}{2} \int_{\Omega} \frac{\partial P^m}{\partial y} \frac{\partial v}{\partial y} \\ + \int_{\Omega} (y^2 S - rS) \frac{\partial P^m}{\partial S} v + \alpha \int_{\Omega} (y - m) \frac{\partial P^m}{\partial y} v = \frac{1}{\delta t} \int_{\Omega} P^{m-1} v \end{aligned} \quad (7.50)$$

To illustrate this, let us take the parameters

$$r = 0.05, \quad \alpha = 1, \quad \nu = 0.5, \quad m = 0.2, \quad \text{and} \quad K = 100. \quad (7.51)$$

The goal is to approximate P in the domain $(0, \bar{S}) \times (-1.5, 1.5)$ for t smaller than 1. We choose $\bar{S} = 800$. For computing the solution in $(0, \bar{S}) \times (-1.5, 1.5)$, we choose the larger domain $\Omega = (0, \bar{S}) \times (-3, 3)$, $\bar{y} = 3$.

We compute the pricing function of the put option 1 year to maturity. The time step has been set to 6 days. The artificial boundary conditions $u(S, \pm \bar{y}, t) = K e^{-rt}$ have been used. In Fig. 7.1, we display the contours of the solution in $\Omega = (0, 800) \times (-3, 3)$. In this figure, it is very well seen that the artificial boundary conditions on $y = \pm 3$ do not affect the solution in the region $|y| < 1.5$.

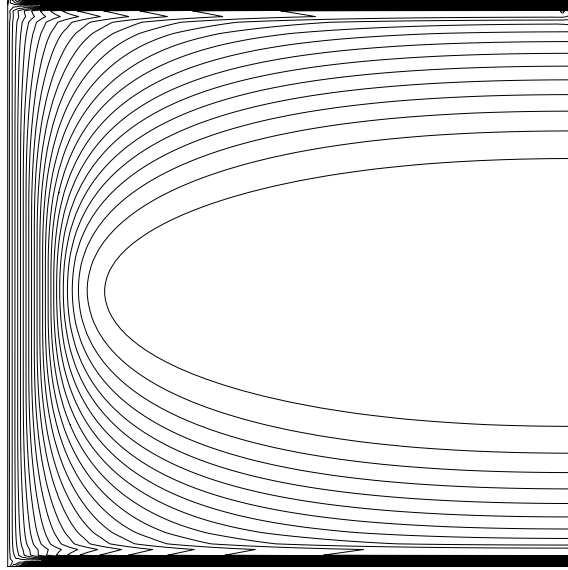


FIG. 7.1 The contours of the price computed in $\Omega = (0, 800) \times (-3, 3)$: note the boundary layers due to artificial boundary conditions on $y = \pm 3$.

REMARK 7.8. The choice of $\alpha = 1$ is not quite realistic from a financial viewpoint, if the asset is linked to stocks, because the mean reversion rate is generally larger. When the asset corresponds to interest rates, such values of α are reasonable. When the mean reversion rate is large, it is possible to carry out an asymptotic expansion of the solution as in FOUQUE, PAPANICOLAOU and SIRCAR [2000], and we believe that the variational setting introduced above justifies these expansions.

The analysis of convergence of the FEM for the elliptic part of operator appearing in (7.48) was performed in ACHDOU, FRANCHI, and TCHOU [2005]. In this article, it was shown that the finite-element error analysis theory as in the studies by CIARLET [1978, 1991] could be performed as soon as the family of meshes under consideration satisfy a regularity assumption with respect to an intrinsic metric associated with the degenerate operator; more precisely, defining X and Y the smooth vector fields in $\mathbb{R}^2$,

$$X = Sy\partial_S, \quad Y = \partial_y, \quad (7.52)$$

we say that an absolutely continuous curve $\gamma : [0, T] \longrightarrow \mathbb{R}^2$ is a subunit curve with respect to X if for any $\xi \in \mathbb{R}^2$

$$\langle \dot{\gamma}(t), \xi \rangle^2 \leq \langle X(\gamma(t)), \xi \rangle^2 + \langle Y(\gamma(t)), \xi \rangle^2$$

for a.e. $t \in [0, T]$. Following NAGEL, STEIN and WAINGER [1985]), we define the intrinsic or Carnot–Carathéodory metric $d: \forall P_1, P_2 \in \mathbb{R}^2$,

$$d(P_1, P_2) = \inf \{T > 0 : \text{there exists a subunit curve } \gamma, \\ \gamma : [0, T] \longrightarrow \mathbb{R}^2, \quad \gamma(0) = P_1, \quad \gamma(T) = P_2\}.$$

If the above set of curves is empty, we take $d(P_1, P_2) = \infty$.

A study by ACHDOU, FRANCHI, and TCHOU [2005], it was essentially required that the family of meshes be regular with respect to the Carnot–Carathéodory metric: there exists a parameter σ such that, for any triangle T , one can find two Carnot–Carathéodory balls, the first one containing T and the other one contained in T , such that calling r_1 and r_2 their Carnot–Carathéodory radii ($r_1 > r_2$) we have $\frac{r_1}{r_2} \leq \sigma$. Under such an assumption, it is possible to construct a local regularization operator similar to that of CLÉMENT [1975], and then to obtain optimal error estimates. A study by ACHDOU, FRANCHI, and TCHOU [2005], examples of meshes satisfying the regularity assumption were given.

A numerical method for pricing options with Heston's model With Heston's model, in contrast to the last example, the advection in the y variable does not dominate the diffusion as $y \rightarrow \infty$. Therefore, inexact boundary conditions on an artificial boundary $y = \bar{y}$ may produce large errors. For this reason, another strategy has been chosen: instead of truncating the domain in the variable y , we have used a suitable change of variable in order to map the y -domain, that is, $\mathbb{R}_+$ onto the interval $(0, 1)$.

We want to approximate P given by (7.11). The idea is to make the change of variables $z = y/(y + 1)$, which maps $\mathbb{R}_+$ onto $(0, 1)$. The inverse map is $y = z/(1 - z)$.

In the variables $(T - t, S, z)$, the PDE becomes

$$\frac{\partial \check{P}}{\partial t} + \check{\mathcal{L}}_t \check{P} = 0 \quad (7.53)$$

for $t \in (0, T]$, $S \in \mathbb{R}_+$, and $z \in (0, 1)$, where

$$\begin{aligned} \check{\mathcal{L}}_t v = & -\frac{1}{2} \frac{z}{1-z} S^2 \frac{\partial^2 v}{\partial S^2} - r(t) \left(S \frac{\partial v}{\partial S} - v \right) - \rho \lambda S(1-z) z \frac{\partial^2 v}{\partial S \partial z} \\ & - \frac{\lambda^2}{2} z(1-z)^2 \left((1-z) \frac{\partial^2 v}{\partial z^2} - 2 \frac{\partial v}{\partial z} \right) - \kappa \left(m - \frac{z}{1-z} \right) (1-z)^2 \frac{\partial v}{\partial z}, \end{aligned} \quad (7.54)$$

assuming that $\tilde{\gamma} = 0$.

No boundary condition is needed on the axis $z = 0$ because the partial differential operator becomes degenerate there: indeed, all the coefficients of the second derivatives vanish; moreover, the first-order derivatives correspond to an advection with an outgoing velocity, that is, the coefficient in front of $\frac{\partial v}{\partial z}$ is negative near $z = 0$ (its value is $-\kappa m$). Similarly, no boundary condition is needed on the axis $z = 1$ because $\check{\mathcal{L}}$ becomes degenerate near $z = 1$.

On the other hand, one can truncate the domain in the S variable. The new problem is posed in the rectangle $(0, \bar{S}) \times (0, 1)$, which allows for the use of a FEM. We have refined the mesh near the strike.

We have made the following choice of parameters:

$$r = 0, \quad \rho = -0.5, \quad \kappa = 2.5, \quad \lambda = 0.5, \quad m = 0.06, \quad K = 1.$$

We have taken $\bar{S} = 4$. An approximation of P solution to (7.11) is obtained from the finite-element approximation of $\check{P}$ (or possibly $e^{-\gamma y} \check{P}$) by performing the inverse change of variable $z \mapsto y = z/(1-z)$. The pricing functions of the put and the call half-year to maturity are displayed in Fig. 7.2. The computed prices are in good agreement with the closed form obtained by HESTON [1993].

7.3. American options with stochastic volatility

In this paragraph, we discuss the pricing of an American put option with the Stein–Stein stochastic volatility model. The parameters of the model are given by (7.51). The domain truncation is also the same as in the example of the European option. Piecewise linear finite elements are used for the discretization. We have chosen to use the first-order operator splitting or projection scheme described in Section 6.4. A similar method has been studied by IKONEN and TOIVANEN [2004] for Heston’s model. In order to capture the exercise boundary, we have adapted the mesh in the variables S and y . In Fig. 7.3, we have plotted the contours of the pricing function 1 year to maturity: the exercise zone clearly appears, indeed, it corresponds to the zone where the pricing function matches the function $S \mapsto K - S$, that is, where the contours are vertical straight lines. Figure 7.3 has to be compared with Fig. 7.1 for the European option. In Fig. 7.4, we have plotted

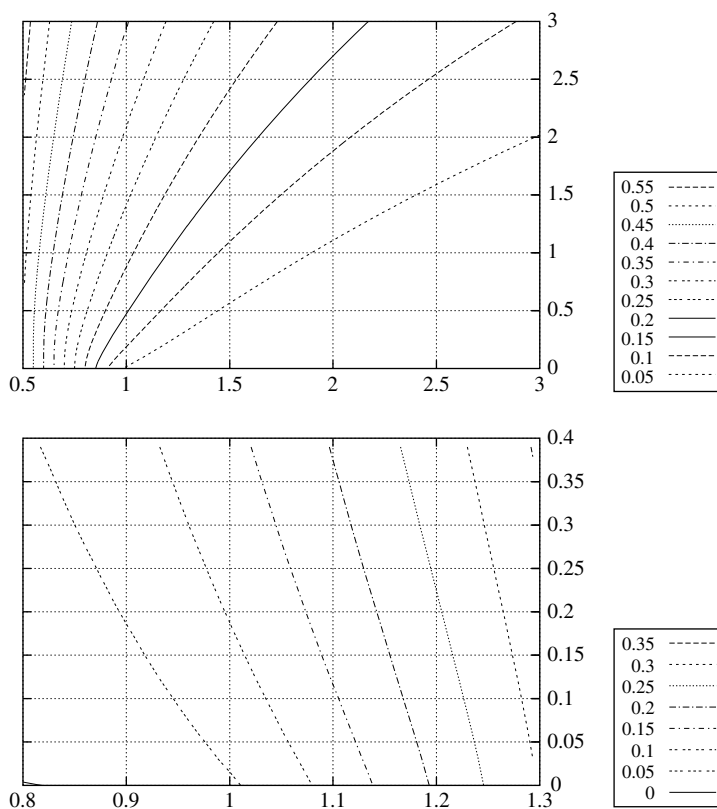


FIG. 7.2 The contours of the pricing function of a put(top)/call(bottom) option with Heston model half-year to maturity.

the exercise region 1 year to maturity. The mesh is visible. It is refined near the exercise boundary.

IKONEN and TOIVANEN [2006] have proposed specific finite-difference methods and solution procedures for American options with Heston's model. Special discretization and grid are designed in such a way that the resulting matrix be an M matrix. The scheme is a seven-point scheme, and upwinding is used when necessary. A specific alternating direction splitting scheme is proposed. Each substep consists of solving a one-dimensional linear complementarity problem by the Brennan–Schwartz algorithm (see BRENNAN and SCHWARTZ [1977]). Three directions are used: the two axes and the first diagonal. For these algorithms to work, one needs that the tridiagonal matrices used in the substeps are M matrices. This is not true, in general, but the scheme and the grid have been designed so that this condition holds. It is also necessary that the exercise boundary intersects the directions once at most. This condition is not proved, but no counterexample has been found in the computations. IKONEN and TOIVANEN [2006] have compared this solution procedure with four other methods, in particular,

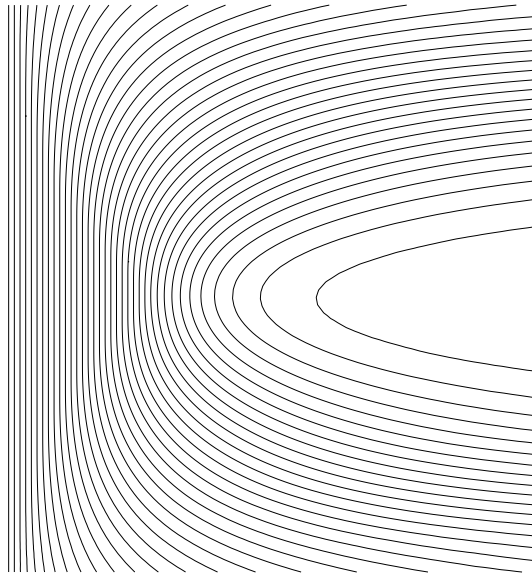


FIG. 7.3 The pricing function of the American option 1 year to maturity. The exercise zone is clearly visible.

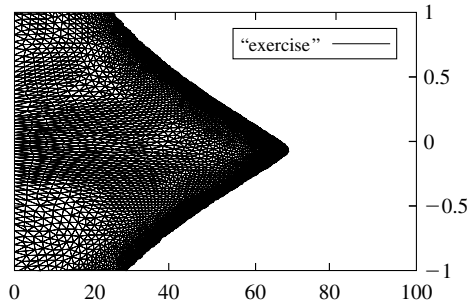


FIG. 7.4 The exercise zone 1 year to maturity. One clearly sees that the mesh has been refined near the exercise boundary.

the previously described projection scheme and a multigrid algorithm; they have shown that the alternating direction method performs best. On the other hand, the last solution procedure has been tailored for this problem, and its robustness has to be assessed.

7.4. Volatility models with several stochastic variables

We give the example of a generalized multifactor Scott model: we consider d fully correlated OU processes

$$dY_t^{(i)} = -\lambda_i Y_t^{(i)} dt + \beta_i dB_t, \quad i = 1, \dots, d, \quad (7.55)$$

where B_t is one-dimensional Brownian motion. We consider an asset whose price is a lognormal process:

$$dS_t = r(t)S_t dt + \sigma_t S_t dW_t,$$

where W_t is a Brownian motion independent of B_t and the volatility σ_t is of the form

$$\sigma_t = \sigma \left(\sum_{i=1}^d Y_t^{(i)}, t \right).$$

The price of the option is $P(S_t, Y_t^{(i)}, \dots, Y_t^{(d)}, t)$, where P satisfies the $d + 2$ dimensional PDE:

$$\begin{aligned} \frac{\partial P}{\partial t} + \frac{1}{2} \sigma^2(z_1, t) S^2 \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} \\ + \frac{1}{2} \sum_{i,j} \beta_i \beta_j \frac{\partial^2 P}{\partial y_i \partial y_j} + \rho \beta_i \sigma(z_1, t) S \frac{\partial^2 P}{\partial S \partial y_i} - \sum_{i=1}^d \lambda_i y_i \frac{\partial P}{\partial y_i} - rP = 0, \end{aligned}$$

where $z_1 = \sum_{i=1}^d y_i$. One can make the change of variables $z = Qy$, where

$$Q = \begin{pmatrix} 1 & 1 & \dots & \dots & 1 \\ -\beta_2/\beta_1 & 1 & 0 & \dots & 0 \\ -\beta_3/\beta_1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\beta_d/\beta_1 & 0 & \dots & 0 & 1 \end{pmatrix}$$

and get the PDE

$$\begin{aligned} \frac{\partial P}{\partial t} + \frac{1}{2} \sigma^2(z_1, t) S^2 \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} + \frac{1}{2} \beta^2 \frac{\partial^2 P}{\partial z_1^2} \\ + \rho \beta \sigma(z_1, t) S \frac{\partial^2 P}{\partial S \partial z_1} - z^T L^T \nabla_z P - rP = 0, \end{aligned}$$

where $\beta = \sum_{i=1}^d \beta_i$ and $L = Q \text{Diag}(\lambda_1, \dots, \lambda_d) Q^{-1}$. This linear PDE is parabolic with respect to the variables S and z_1 and hyperbolic with respect to z_i , $1 < i \leq d$.

Sparse grid methods can be used for approximating the five-variable function P . We give an example provided: the parameters of the three factors models are

- interest rate: $r = 5\%$,
- spot price–volatility correlation: $\rho = -0.5$,
- mean value of the volatility: $\sigma = 0.2$.
- parameters of the OU processes: $\lambda \approx (29.27, 2.45, 0.108)$, $\beta = (1.26, 0.42, 0.42)$.

With these parameters, one may truncate the domain of computation because the velocity in the advection terms in the PDE is directed outward near the artificial boundaries, and

TABLE 7.2
Price of a European call 1 year to maturity. These results have
been obtained by D. Pommier.

Spot	Level = 7	Level = 8	Level = 9	Monte Carlo
0.80	1.51	1.48	1.45	1.41
0.85	2.86	2.81	2.78	2.75
0.90	4.78	4.72	4.71	4.67
0.95	7.29	7.23	7.24	7.22
1.00	10.32	10.31	10.31	10.32
1.05	13.82	13.84	13.85	13.88
1.10	17.72	17.75	17.77	17.81
1.15	21.92	21.94	21.96	22.01
1.20	26.31	26.35	26.38	26.43

the errors produced by the inexact boundary conditions are small sufficiently far from the artificial boundaries.

In Table 7.2, we compute the price of a European call option 1 year to maturity for several spot values, the strike is 1, for $y_i = 0, i = 1, 2, 3$. The payoff function depends on the variable S only. Hence, the singularity is located on an hyperplane in the spot/volatilities space, and sparse grids are well adapted. A Crank–Nicolson scheme with a time step of 0.01 year has been used.

We compare the results of the sparse grid method with refinement levels of 7, 8, and 9 with a Monte Carlo simulation. Note that the sparse grid approximation is sharper for spots larger than or equal to the strike because the sparse grid is relatively coarser for small spots.

The tests have been done on a 2.66-GHz Intel Xeon processors with 1.5-Gb RAM. The computing time is approximately 4 minutes for $n = 7$, 15 minutes for $n = 8$, and 1 hour for $n = 9$.

Sensitivity and Calibration

8. Sensitivity

It is important to compute the sensitivity of options' prices to parameters such as the spot price or the volatility. The partial derivatives with respect to the relevant parameters are called the *Greeks*, let C be the price of a vanilla European call:

- the δ (delta) is its derivative with respect to the stock price S : $\frac{\partial C}{\partial S}$.
- the Θ or time decay is its derivative with respect to time: $\frac{\partial C}{\partial t}$.
- the vega κ is its derivative with respect to the volatility σ ,
- the rho ρ is its derivative with respect to the interest rate, $\frac{\partial C}{\partial r}$,
- η is its derivative with respect to the strike K
- finally, the gamma is the rate of change of its delta : $\frac{\partial^2 C}{\partial S^2}$.

Equations can be derived for these by directly differentiating the PDE and the boundary conditions that define C . Automatic differentiation (AD) of a computer code for option pricing provides a way to do that efficiently and automatically.

8.1. Automatic differentiation

Automatic differentiation of computer programs is based on the idea that every line of a program can be differentiated analytically. Consider, for instance, the C program that computes $J = (u^2 - u_d^2)$ for some values of u and u_d

```
int main()
{
    double J, aux, u=2.5, u_d=1;
    aux = u-u_d;
    J=aux*(u+u_d);
    cout<<"J="<<J<<endl;
}
```

It can be made to compute the derivative of J with respect to u by adding above each instruction its derivative with respect to u :

```
int main()
{
    double dJdu,J, dauxdu,aux, u=2, u_d=0.1;
    dauxdu=1;
    aux = u-u_d;
```

```

dJdu=dauxdu*(u+u_d)+aux;
J=aux*(u+u_d);
cout<<"J="<<J<<" dJdu="<<dJdu<<endl;
}

```

However, it is more systematic to compute differentials, and consider that every `double` or `float` variable has an infinitesimal variation, potentially and initially set to zero.

Let $J(u) = (u - u_d)(u + u_d)$, then its differential is

$$\delta J = (u - u_d)(\delta u + \delta u_d) + (u + u_d)(\delta u - \delta u_d). \quad (8.1)$$

Obviously, the derivative of J with respect to u is obtained by setting $\delta u = 1$ and $\delta u_d = 0$. Now, suppose that J is programmed in C/C++ by

```

double J(double u, double u_d){
    double z = u-u_d;
    z=z*(u+u_d);
    return z;
}
int main(){double u=2, u_d=0.1;
    cout << J(u,u_d) << endl;
}

```

A program that computes J and its differential can be obtained by writing above each differentiable line its differentiated form:

```

double JandDJ(double u, double u_d, double du, double du_d,
    double *pdz){double dz = du - du_d, z = u-u_d;
    double dJ = dz*(u+u_d) + z*(du + du_d);
    z = z*(u+u_d);
    *pdz = dz;
    return z;
}
int main(){
    double dJ, u=2,u_d = 0.1;
    cout << J(u,u_d,1,0,&dJ) << endl;
}

```

Except for the embarrassing problem of returning both z, dz instead of z , the procedure can be automatized by introducing a structured type holding the value of the variable and of its derivative:

```

struct {double val[2];} dreal;

dreal JandDJ(dreal u, dreal u_d) {
    dreal z;
    z.val[1]=u.val[1]-u_d.val[1];
    z.val[0]=u.val[0]-u_d.val[0];
    z.val[1]=z.val[1]*(u.val[0]+u_d.val[0])+z.val[0]*(u.val[1]+u_d.val[1]);
    z.val[0]=z.val[0]*(u.val[0]+u_d.val[0]);
    return z;
}

```

```
int main() {
    dreal u, dJ;
    u.val[0]=2;u_d.val[0]=0.1;u.val[1]=1;u_d.val[1]=0;
    cout <<J(u,u_d).val[0]<<J(u,u_d,1,0).val[1];
}
```

8.1.1. The class *ddouble*

In C++, the program can be simplified further by redefining the operators $=$, $-$, $+$, and $*$. Then, a class has to be used instead of a struct!

```
class ddouble {
public: double val[2];
    ddouble(double a, double b=0){ v[0] = a; v[1]=b;}
    ddouble operator=(const ddouble& a)
    { val[1] = a.val[1]; val[0]=a.val[0];
      return *this;
    }
}

friend ddouble operator - (const ddouble& a, const ddouble& b){
    ddouble c;
    c.v[1] = a.v[1] - b.v[1];    // (a-b)'=a'-b'
    c.v[0] = a.v[0] - b.v[0];
    return c;
}

friend ddouble operator + (const ddouble& a, const ddouble& b){
    ddouble c;
    c.v[1] = a.v[1] + b.v[1];    // (a-b)'=a'-b'
    c.v[0] = a.v[0] + b.v[0];
    return c;
}

friend ddouble operator * (const ddouble& a, const ddouble& b){
    ddouble c;
    c.v[1] = a.v[1]*b.v[0] + a.v[0]* b.v[1];
    c.v[0] = a.v[0] * b.v[0];
    return c;
}
};
```

For more complex programs, all the operators and all the functions of algebra have to be redefined in the class *ddouble*. For instance,

```
ddouble ddouble::sqrt(const ddouble& x){
    ddouble c;
    c.v[0] = sqrt(x.v[0]);
    c.v[1] = 0.5*x.v[1]/sqrt(x.v[0]);
    return c;
}
```

Obviously, it creates a problem at $x = 0$, but there the function is nondifferentiable.

The library *ddouble.hpp* implements correctly all these; it can be downloaded from <http://www.ann.jussieu.fr/pironeau>.

8.2. Sensitivity by automatic differentiation

The most obvious application of AD is for the computation of the sensitivity with respect to the parameters of the models (the *Greeks*). For this, there is very little to change in the computer program; for example, consider a local volatility function of S and t :

$$\sigma = \sigma_0(1 + a\sqrt{|S - K|})(1 + bt). \quad (8.2)$$

The following changes are made to compute the sensitivity with respect to b at $b = 1$ when $a = 0.01$ and $\sigma_0 = 0.3$.

1. In the file `optionhb.hpp`, we make now an explicit use of `ddouble.hpp`

```
#include "ddouble.hpp" // for automatic differentiation
// typedef double ddouble; // for compatibility with AD.
```

2. In the main function, the definition of b has value 1 in the second argument to indicate that derivatives are with respect to this variable.

```
int main(){
    VarMesh m(50,100,0.5,300.); // uniform nb x, nb t, T, xmax
    const ddouble b(1.,1.); // for derivatives with respect to b at b=1
    const double K=100, a=0.01;
    Option p(1,&m,K,0.05,0.3); // B&S,mesh,strike,r,sigma
    for(int i = 1; i < m.nT; i++)
        for(int j = 1; j < m.nX[i]; j++) // local volatility
            p.sigma[i][j] *= (1 + a*sqr(fabs(m.x[i][j]-K)))*(1+b*m.t[i]);
    p.calc();
}
```

Results are in $p.u[i][j].val[0]$ for the price of the put and in $p.u[i][j].val[1]$ for its derivative with respect to b . Both are shown in Fig. 8.1.

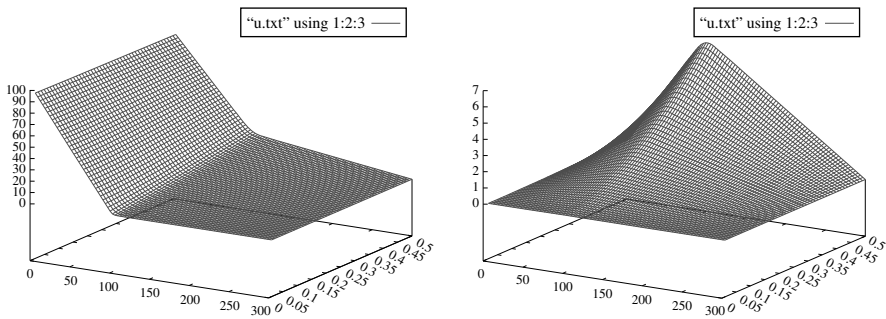


FIG. 8.1 The pricing function of the put option (left) and its derivative with respect to the parameter b when σ is given by (8.2).

9. Calibration

9.1. Limitation of the Black–Scholes model: the need for calibration

Consider a European option on a given stock with a maturity T and a payoff function P_0 and assume that this option is on the market. Call p its present price. Also, assume the risk-free interest rate is the constant r . One may associate with p the so-called implied volatility, the volatility σ_{imp} , such that the price given by formula (1.6) at time $t = 0$ with $\sigma = \sigma_{\text{imp}}$ coincides with p . If the Black–Scholes model was sharp, then the implied volatility would not depend on the payoff function P_0 . Unfortunately, for example, vanilla European puts or calls, it is often observed that the implied volatility is far from constant. Rather, it is often a convex function of the strike price. This phenomenon is known as the volatility smile. A possible explanation for the volatility smile is that the deeply out-of-the-money options are less liquid, thus relatively more expensive than the options in the money.

This shows that the critical parameter in the Black–Scholes model is the volatility σ . Assuming σ to be constant and using (1.3) often lead to poor predictions of the options prices. The volatility smile is the price paid for the too great simplicity of Black–Scholes assumptions. It is, thus, necessary to use more involved models that must be calibrated.

Let us first explain what the term *calibration* means: consider an arbitrage-free market described by a probability measure $\mathbb{P}$ on a scenario space $(\Omega, \mathcal{A})$. There is a risk-free asset whose price at time τ is $e^{r\tau}$, $r \geq 0$ and a risky asset whose price at time τ is S_τ . Specifying an arbitrage-free option pricing model necessitates the choice of a risk-neutral measure, that is, a probability $\mathbb{P}^*$ equivalent to $\mathbb{P}$ such that the discounted price $(e^{-r\tau} S_\tau)_{\tau \in [0, T]}$ is a martingale under $\mathbb{P}^*$. Such a probability measure $\mathbb{P}^*$ allows for the pricing of European options; consider a European option with payoff P_0 at maturity $t \leq T$: its price at time $\tau \leq t$ is $P_\tau = e^{-r(t-\tau)} \mathbb{E}^{\mathbb{P}^*}(P_0(S_t) | \mathcal{F}_\tau)$, where $(\mathcal{F}_\tau)_{\tau \in [0, T]}$ is the natural filtration.

The pricing model $\mathbb{P}^*$ must be compatible with the prices of the options observed on the market, whose number may be large. Model calibration consists of finding $\mathbb{P}^*$ such that the discounted price $(e^{-r\tau} S_\tau)_{\tau \in [0, T]}$ is a martingale and such that the option prices computed with the model coincide with the observed option prices. This is an inverse problem.

Popular extensions to the Black–Scholes model are

- local volatility models: the volatility is a function of time and the spot price, that is, $\sigma_t = \sigma(S_t, t)$. With suitable assumptions on the regularity and the behavior at infinity of the function σ , (1.6) holds and $P_t = p(S_t, t)$, where p satisfies the final value problem (1.3), in which σ varies with t and S . Calibration of local volatility has been much studied (see ACHDOU and PIRONNEAU [2002], ANDERSEN and BROTHERTON-RATCLIFFE [1998], DUPIRE [1997], JACKSON, SÜLI and HOWISON [1998] for volatility calibration with European options and ACHDOU [2005], ACHDOU and PIRONNEAU [2005] with American options).
- stochastic volatility models: one assumes that $\sigma_t = f(y_t)$, where y_t is a continuous time stochastic process, correlated or not to the process driving S_t (see Section 7).

Stochastic volatility calibration has been performed by NAYAK and PAPANICOLAOU [2006].

- Lévy driven spot price: one may generalize the Black–Scholes model by assuming that the spot price is driven by a more general stochastic process, a Lévy process CONT and TANKOV [2003], MERTON [1976]. Lévy processes are processes with stationary and independent increments that are continuous in probability. For a Lévy process X_τ on a filtered probability space with probability $\mathbb{P}$, the Lévy–Khintchine formula says that there exists a function $\chi : \mathbb{R} \rightarrow \mathbb{C}$ such that

$$\mathbb{E}(e^{iuX_\tau}) = e^{\tau\chi(u)},$$

$$\chi(u) = -\frac{\sigma^2 u^2}{2} + i\beta u + \int_{|z|<1} (e^{iuz} - 1 - iuz)v(dz) + \int_{|z|>1} (e^{iuz} - 1)v(dz),$$

for $\sigma \geq 0$, $\beta \in \mathbb{R}$ and a positive measure ν on $\mathbb{R} \setminus \{0\}$ such that $\int_{\mathbb{R}} \min(1, z^2)\nu(dz) < +\infty$. The measure ν is called the Lévy measure of X . We focus on Lévy measure with a density, $\nu(dz) = k(z)dz$. Assume that the discounted price of the risky asset is a square-integrable martingale under $\mathbb{P}$ and that it is represented as the exponential of a Lévy process:

$$e^{-r\tau} S_\tau = S_0 e^{X_\tau}.$$

The martingale property is that $\mathbb{E}(e^{X_\tau}) = 1$, that is,

$$\int_{|z|>1} e^z \nu(dz) < \infty \quad \text{and} \quad \beta = -\frac{\sigma^2}{2} - \int_{\mathbb{R}} (e^z - 1 - z1_{|z|\leq 1})k(z)(dz),$$

and the square integrability comes from the condition $\int_{|z|>1} e^{2z}k(z)dz < \infty$.

With such models, the pricing function for a European option is obtained by solving a PIDE, with a nonlocal term (see CONT and TANKOV [2003], PHAM [1998] for the analysis of this equation and ACHDOU and PIRONNEAU [2005], CONT and VOLTCHKOVA [2003], MATACHE, NITSCHKE and SCHWAB [2003], MATACHE, VON PETERSDOFF and SCHWAB [2004] for numerical methods based on the PIDE). Calibration of Lévy models with European options has been discussed by CONT and TANKOV [2004, 2006].

In this paragraph, we assume that the model is characterized by parameters θ in a suitable class Θ .

The last two classes of models (stochastic volatility and Lévy driven assets) describe incomplete markets (see CONT and TANKOV [2004]): the knowledge of the historical price process alone does not allow to compute the option prices in a unique manner. When the option prices do not determine the model completely, additional information may be introduced by specifying a prior model. If the historical price process has been estimated statistically from the time series of the underlying asset, this knowledge has to be applied in the inverse problem; calling $\mathbb{P}_0$ the prior probability measure obtained as an estimation of $\mathbb{P}$, the inverse problem may be cast in a least-square formulation of

the type: find $\theta \in \Theta$ that minimizes

$$\sum_{i \in I} \omega_i (P^\theta(0, S_0, t_i, x_i) - \bar{p}_i)^2 + \rho J_2(\mathbb{P}^\theta, \mathbb{P}_0), \quad (9.1)$$

where

- ω_i are suitable positive weights,
- S_0 is the price of the underlying asset today,
- $P^\theta(0, S_0, t_i, x_i)$ is the price of the option with maturity t_i strike x_i , computed with the pricing model associated with θ ,
- $\rho J_2(\mathbb{P}^\theta, \mathbb{P}_0)$ is a regularization term that measures the closeness of the model $\mathbb{P}^\theta$ to the prior. The number $\rho > 0$ is called the regularization parameter. This functional has two roles: (1) it stabilizes the inverse problem, and for that, ρ should be large enough and J_2 should be convex or at least convex in a large enough region; (2) it guarantees that $\mathbb{P}^\theta$ remains close to $\mathbb{P}_0$ in some sense. The choice of J_2 is very important: $J_2(\mathbb{P}^\theta, \mathbb{P}_0)$ is often chosen as the relative entropy of the pricing measure $\mathbb{P}^\theta$ with respect to the prior model $\mathbb{P}_0$ (see AVELLANEDA [1998]), because the relative entropy becomes infinite if $\mathbb{P}^\theta$ is not equivalent to $\mathbb{P}_0$. Some authors have argued that such a choice may be too conservative in some cases, for two reasons: (a) the historical data that determine the prior may be missing or partially available and (b) in the context of volatility calibration, once the volatility is specified under $\mathbb{P}_0$, then the volatility under $\mathbb{P}^\theta$ must be the same for the relative entropy to be finite. A different approach was considered that allowed for volatility calibration (see AVELLANEDA, FRIEDMAN, HOLMES and SAMPERI [1997]).

Note that local volatility models describe complete markets; however, an additional regularization cost functional is necessary, as explained in the paragraph below.

9.2. The calibration of local volatility

Consider a family of I call options on the same asset with maturities $(T_i)_{1 \leq i \leq I}$ and strike prices $(K_i)_{1 \leq i \leq I}$. Assume that the options are available on the market, so one can observe their prices today. Call $(\bar{C}_i)_{1 \leq i \leq I}$ their prices. Assume that the spot price today is S_0 . Calibrating the local volatility amounts to finding a local volatility surface $S, t \rightarrow \sigma(S, t)$ such that if $C_i(S, t, K_i, T_i)$ is computed by solving the boundary value problem

$$\frac{\partial C_i}{\partial t} + \frac{\sigma(S, t)^2}{2} S^2 \frac{\partial^2 C_i}{\partial S^2} + rS \frac{\partial C_i}{\partial S} + rC_i = 0, \quad C_i(S, T_i) = (S - K_i)^+, \quad (9.2)$$

then $C_i(S_0, 0)$ coincides with the observed price $\bar{C}_i$, for $1 \leq i \leq I$.

A natural idea for this is to use least squares, that is, to minimize a functional

$$J_{LS} : \sigma \mapsto \sum_{i \in I} \omega_i |\bar{C}_i - C_i(S_0, 0)|^2$$

for σ in a suitable function set Σ , where ω_i are positive weights and C_i is computed by solving (9.2). The evaluation of J_{LS} requires the solution of I initial value problems.

The set Σ where the volatility is to be found must be chosen in order to ensure that from a minimizing sequence one can extract at least a subsequence that converges in Σ and that its limit is, indeed, a solution to the least square problem. For example, Σ may be a compact subset of a Hilbert space W such that the mapping J_{LS} is continuous in W (lower semicontinuous would be enough). In practice, W has a finite dimension and is compactly embedded in the space of bounded and continuous functions σ such that $S\partial_S\sigma$ is bounded. Thus, the existence of a solution to the minimization problem is most often guaranteed. What is more difficult to guarantee is uniqueness and stability: is there a unique solution to the least square problem? If yes, is the solution insensitive to small variations of the data? The answer to these questions is no, and we say that the problem is *ill-posed*.

As a possible cure to ill-posedness, one usually modifies the problem by minimizing the functional

$$J : \sigma \mapsto J_{LS}(\sigma) + J_R(\sigma)$$

instead of J_{LS} , where J_R is a sufficiently large strongly convex functional defined on W and containing some financially relevant information. For example, one may choose

$$J_R(\sigma) = \omega \|\sigma - \bar{\sigma}\|^2,$$

where ω is some positive weight, $\|\cdot\|$ is a norm in W , and $\bar{\sigma}$ is a prior local volatility, which may come from a historical knowledge. The difficulty is that ω must not be too large not to perturb the inverse problem too much but not too small to guarantee some stability. The art of the practitioner lies in the choice of J_R .

9.3. Gradient methods

Once the least square problem is chosen, we are left with proposing a strategy for the construction of minimizing sequences. If J_{LS} and J_R are C^1 functional, then gradient methods may be used. The principle of gradient methods is as follows: let $\delta\sigma$ be an admissible variation and assume that J is Frechet differentiable with respect to σ , then by definition,

$$\delta J = J(\sigma + \delta\sigma) - J(\sigma) = \text{grad}J(\sigma) \cdot \delta\sigma + o(|\delta\sigma|). \quad (9.3)$$

So, $\delta\sigma = -\rho \text{grad}J(\sigma)$ causes a decrease in J at least when ρ is small.

9.3.1. Gradient algorithm with Armijo rule

1. Choose an initial σ^0 and 2 numbers $0 < \alpha < 1$, $\rho_0 > 0$.
2. Set $H^m = -\text{grad}J(\sigma^m)$.
3. Compute by dichotomy a signed integer k such that, with $\rho = \rho_0 2^k$,

$$\begin{aligned} J(\sigma^m + \rho H^m) - J(\sigma^m) &\leq \rho \alpha \text{grad}J(\sigma^m) \cdot H^m, \\ J(\sigma^m + 2\rho H^m) - J(\sigma^m) &\geq 2\rho \alpha \text{grad}J(\sigma^m) \cdot H^m. \end{aligned} \quad (9.4)$$

4. Set $\sigma^{m+1} = \sigma^m + \rho H^m$ and proceed to the next iteration.

9.3.2. Conjugate gradient method

In the Polak-Ribiere (POLAK [1997]) version, one changes the definition of H^m , $m > 1$, to

$$H^m = -\text{grad}J(\sigma^m) + \gamma H^{m-1} \text{ with } \gamma = \frac{\text{grad}J(\sigma^m) \cdot (\text{grad}J(\sigma^m) - \text{grad}J(\sigma^{m-1}))}{\|\text{grad}J(\sigma^{m-1})\|^2}.$$

Under strong hypothesis such as local convexity and twice differentiability of J , this algorithm converges to a local minimum superlinearly, that is, faster than any geometric progression.

The drawbacks and advantages of gradient methods are well known. On the one hand, they do not guarantee convergence to the global minimum if the functional is not convex because the iterates can be trapped near a local minimum. On the other hand, they are fast and accurate when the initial guess is close enough to the minimum. For these reasons, gradient methods are often combined with techniques that permit to localize the global minimum but that are slow, like simulated annealing or evolutionary algorithms.

9.4. A finite dimensional example

To reduce the size of the problem, we parametrize the volatility surface by a linear-quadratic spline. The idea is that we want the local volatility to be close to linear in the money and to have a convex shape out of the money; this is desired at each time t , but the bounds may depend on t linearly. Hence, the surface is given by

$$\sigma(S, t) = \begin{cases} a + (2\frac{\sigma_1 - a}{S_1} - \frac{\sigma_2 - \sigma_1}{S_2 - S_1})S + (\frac{\sigma_2 - \sigma_1}{S_2 - S_1} - \frac{\sigma_1 - a}{S_1})\frac{S^2}{S_1}, & \text{if } S < S_1, \\ \sigma_2 \frac{S - S_1}{S_2 - S_1} + \sigma_1 \frac{S_2 - S}{S_2 - S_1}, & \text{if } S_1 \leq S \leq S_2, \\ \sigma_2 + (S - S_2)\frac{\sigma_2 - \sigma_1}{S_2 - S_1} + \left((S - S_2)\frac{\sigma_2 - \sigma_1}{S_2 - S_1}\right)^2, & \text{if } S > S_2, \end{cases} \quad (9.5)$$

where S_i , σ_i , and $i = 1, 2$ are linear with respect to t :

$$S_i = S_{i1} \left(1 - \frac{t}{T}\right) + S_{i2} \frac{t}{T}, \quad \sigma_i = \sigma_{i1} \left(1 - \frac{t}{T}\right) + \sigma_{i2} \frac{t}{T}. \quad (9.6)$$

The local volatility is, thus, $\mathcal{C}^1$ regular and linear with respect to S for $S_1 \leq S \leq S_2$; it takes the values σ_1 at $S = S_1$, σ_2 at $S = S_2$, and a at $S = 0$. For a, T given, the local volatility depends on eight parameters: S_{ij} , σ_{ij} . However, the representation (9.5) (9.6) does not make sense unless $0 < S_1(t) < S_2(t)$, so it is better to set

$$z_0 = \sqrt{S_{11}}, \quad z_1 = \sqrt{S_{21}}, \quad z_2 = \sqrt{S_{21} - S_{11}}, \quad z_3 = \sqrt{S_{22} - S_{12}},$$

$$z_{2*i+j+1} = \frac{\sqrt{\sigma_{ij}}}{1 + \sqrt{\sigma_{ij}}}, \quad i, j = 1, 2, \quad (9.7)$$

so as to work with the set of parameters $\{z_i\}_0^7$; the last formula gives $\sigma_{ij} = z_{2*i+j+1}^2 / (1 + z_{2*i+j+1}^2) \in (0, 1)$ for all possible values of $z_{2*i+j+1}$. This is a simple way to force σ in $(0, 1)$ for all values of the parameters z .

9.4.1. A test case

Table 9.1 contains the observed prices of several European calls on the same asset. The parameter r is constant, $r = 0.03$. The spot price is 1418.3.

The least square problem is

$$\min_{\sigma \in \Sigma} \sum_{i=1}^I |C_i(S_0, 0) - C_i|^2, \quad \text{subject to (9.2).} \quad (9.8)$$

The conjugate gradient algorithm was used to solve (9.8) with the initial guess for the parameter vector

$$z_0 = \sqrt{1500}, \quad z_1 = \sqrt{1700}, \quad z_2 = z_3 = \sqrt{150}, \quad z_{4+i} = \sqrt{\frac{0.4}{1+i-0.4}} \quad i = 0..3.$$

The cost function J has $I = 57$ terms. The other parameters are $a = 4$ and $T = 3$ (years). The results are shown in Fig. 9.1. The strange shape of the volatility surface is due to the parametrization. If in (9.5), we change the definition for $S > S_2$ to be

$$\sigma_2 + (S - S_2) \frac{\sigma_2 - \sigma_1}{S_2 - S_1} + \left((S - S_2) \frac{\sigma_2(0) - \sigma_1(0)}{S_2(0) - S_1(0)} \right)^2, \quad \text{if } S > S_2, \quad (9.9)$$

then, the shape is more natural (Fig. 9.2), and the same precision is obtained.

9.5. Dupire's equation

Solving (9.8) by a conjugate gradient method is time consuming because the Black–Scholes PDE has to be solved I times at each iteration and for each ρ trial in the dichotomy (9.4). DUPIRE [1994] noticed that the function $(K, \tau) \mapsto C(S, t, K, \tau)$ (now t and S are fixed) satisfies a forward parabolic PDE. One can obtain this either by reasoning directly on (1.6) or by PDE arguments using adjoint operators. We present the second approach.

PROPOSITION 9.1. (*Dupire*) *Let v be solution in $\mathbb{R}_+ \times (0, T)$ to*

$$\frac{\partial v}{\partial t} - \frac{1}{2} \sigma^2(S, t) S^2 \frac{\partial^2 v}{\partial S^2} + rS \frac{\partial v}{\partial S} = 0, \quad v(S, 0) = (S_0 - S)^+, \quad (9.10)$$

then

$$C(S_0, 0, K, \tau) = v(K, \tau), \quad (9.11)$$

and $p \equiv \frac{\partial^2 v}{\partial S^2}$ satisfies the boundary value problem

TABLE 9.1
The prices of a family of calls on the same asset

Strike	1 month	2 months	6 months	12 months	24 months	36 months
700						733
800						650.6
900						569.8
1000					467.8	
1100					385.3	
1150					345.4	
1175			265.2			
1200			242	266.1	306.6	
1215				253.4		
1225			219	245		
1250			196.6	224.2	269.2	
1275			174.5	203.9	251	
1300			152.9	184.1	233.2	
1325			131.9	164.9	215.8	
1350			111.7	146.3	198.9	
1365			100			
1375	50.6	60	92.5	139	182.6	
1380	46.1	55.8				
1385	41.8	51.8				
1390	37.5	47.9				
1395	33.4	44				
1400	29.4	40.3	74.5	128.4	166.7	215.9
1405	25.6	36.7				
1410	21.9	33.2				
1415	18.7	29.8				
1420	15.4	26.6				
1425	12.7	23.8	58	111.4	151.5	
1430	10	20.7				
1435	8	18.2				
1440	6.3	15.7				
1445	4.4	13.4				
1450	3.1	11.3	43.3	95.2	136.9	187.5
1455	2.05	9.6				
1460	1.45	7.9				
1475			30.6	80.2		
1500			20.3	54	109.6	160.8
1525			12.6	42.7		
1550			7.5	33		
1575				24.7		
1600			1.95	18.2	64.5	113.9
1700					32.7	75.7
1800					15.5	
1900					5.2	

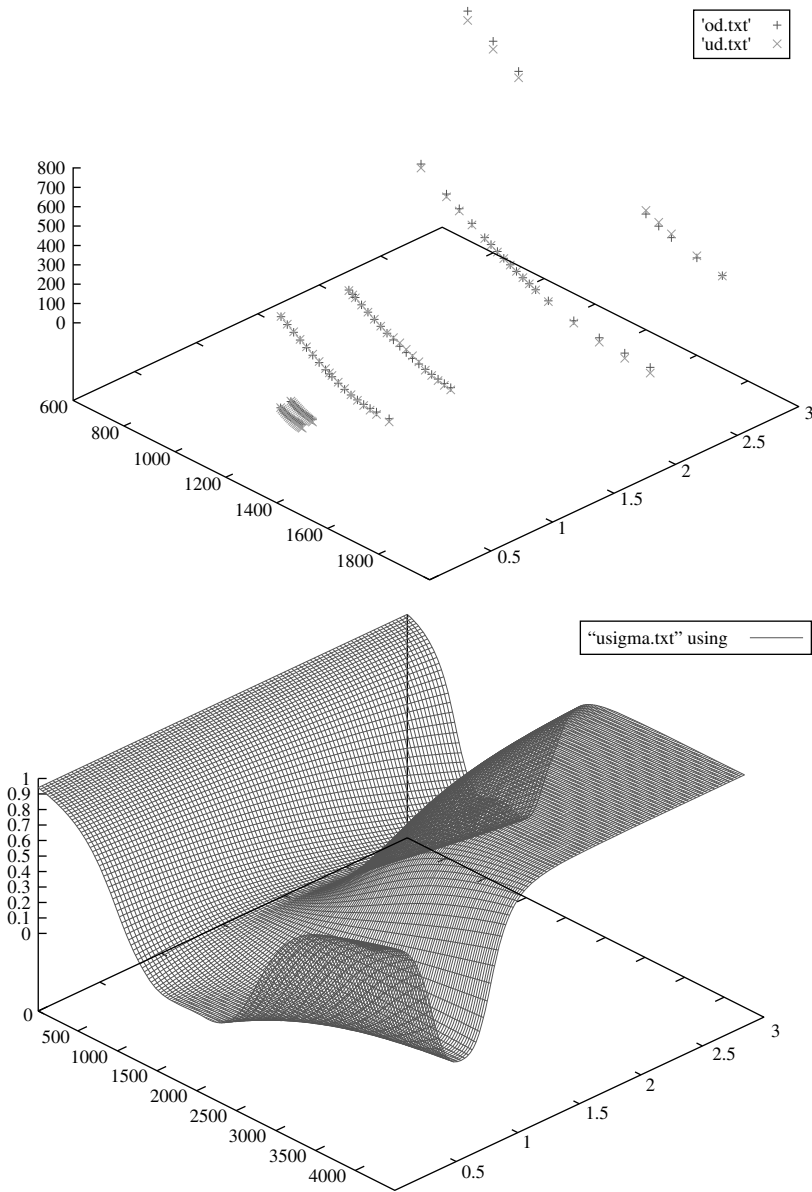


FIG. 9.1 Difference between observations and model predictions (top) obtained with the local volatility (bottom) produced by 100 iterations of conjugate gradient on (9.8). The cost function is reduced from 500 000 to 9000 and the L^2 -norm of the gradient from 10^{11} to 10^7 .

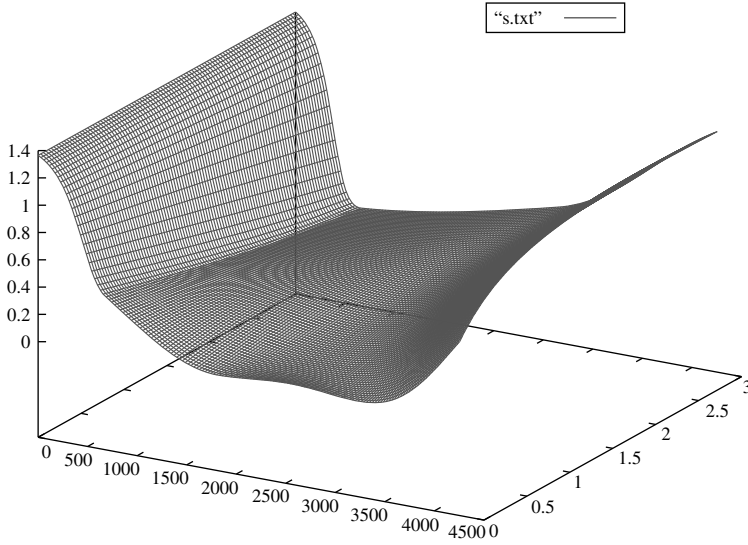


FIG. 9.2 Same as Fig. 9.1 but with (9.9). Convergence is also not perfect. After 100 iterations, J is reduced from 500 000 to 12 000, giving an average error at each observed point larger than 1%.

$$\frac{\partial p}{\partial t} - \frac{\partial^2}{\partial S^2} \left(\frac{\sigma^2(S, t) S^2}{2} p \right) + \frac{\partial}{\partial S} (rSp) + rp = 0, \quad p(S, 0) = \delta_{S_0}(S), \quad (9.12)$$

where δ_{S_0} is the Dirac mass at S_0 .

PROOF. Call Q the rectangular domain $Q = (S_m, S_M) \times (0, \tau)$, with $0 \leq S_m < S_M$ and consider the boundary value problem in Q

$$\frac{\partial u}{\partial t} + \eta(S, t) \frac{\partial^2 u}{\partial S^2} + \mu \frac{\partial u}{\partial S} - ru = 0, \quad u(S, \tau) = u_\tau(S), \quad (9.13)$$

where $\eta(S, t) = \frac{1}{2} \sigma^2(S, t) S^2$ and $\mu = rS$. Multiplying (9.13) by the solution p to (9.12), integrating on Q , and using an integration by part in time and Green's formula in the variable S yield

$$u(S_0, 0) = \int_{S_m}^{S_M} u_\tau(S) p(S, \tau) dS + \int_0^\tau \left[p \eta \frac{\partial u}{\partial S} - u \frac{\partial}{\partial S} (\eta p) + p \mu u \right]_{S_m}^{S_M} dS. \quad (9.14)$$

We take $u(S, t) = C(S, t, K, \tau)$, $S_m = 0$, and $S_M = +\infty$. The second integral in (9.14) vanishes because $u \sim S$ and p tends to 0 faster than S^{-1} as S tends to infinity. Let v be a double primitive of p , that is, a function such that $\frac{\partial^2 v}{\partial S^2} = p$, then, integrating (9.12) twice yields

$$\frac{\partial v}{\partial t} - \frac{1}{2} \sigma^2(S, t) S^2 \frac{\partial^2 v}{\partial S^2} + rS \frac{\partial v}{\partial S} = aS + b, \quad v(S, 0) = c + dS + (S - S_0)^+, \quad (9.15)$$

where a, b, c , and d are integration constants. But $u_\tau = (S - K)^+$, so $\frac{\partial^2 u_\tau}{\partial S^2}(S) = \delta_K(S)$, where δ_K is the Dirac mass at K . A double integration by part applied to (9.14) yields

$$u(S_0, 0) = \int_0^\infty u_\tau(S) \frac{\partial^2 v}{\partial S^2}(S)(S, \tau) dS = \int_0^\infty v(S, \tau) \frac{\partial^2 u_\tau}{\partial S^2}(S) dS + \left[u_\tau \frac{\partial v}{\partial S} - v \frac{\partial u_\tau}{\partial S} \right]_0^\infty. \quad (9.16)$$

Therefore, if v vanishes at ∞ , we obtain

$$u(S_0, 0) = v(K, \tau)$$

because $u_\tau(0) = \partial_S u_\tau(0) = 0$.

By choosing $a = b = 0$ and $c = S_0$, $d = -1$, we obtain (9.11) because $(S - S_0)^+ - (S - S_0) = (S - S_0)^- = (S_0 - S)^+$, so v vanishes at infinity, and the initial condition in (9.10) is the desired one. $\square$

REMARK 9.1. If the underlying asset yields a distributed dividend, $qS_t dt$, then the pricing function satisfies

$$\begin{aligned} \frac{\partial C}{\partial t} + \frac{\sigma^2(S, t) S^2}{2} \frac{\partial^2 C}{\partial S^2} + (r - q)S \frac{\partial C}{\partial S} - rC &= 0 & \text{in } \mathbb{R}_+ \times [0, \tau), \\ C(S, \tau) &= (S - K)^+ & \text{in } \mathbb{R}_+, \end{aligned} \quad (9.17)$$

and the related Dupire's equation is

$$\frac{\partial v}{\partial t} - \frac{1}{2} \sigma^2(S, t) S^2 \frac{\partial^2 v}{\partial S^2} + (r - q)S \frac{\partial v}{\partial S} + qv = 0, \quad (9.18)$$

for $\tau \geq t > 0$ and $S \in \mathbb{R}_+$.

Numerical results A finite-difference method implicit in time of order one (Euler's scheme) is used for (9.2) and (9.10); the parameters are $K = 100$, $r = 0.06$, $\sigma = 0.4$, and 200 time steps and 250 mesh points for S . We have compared the numerical results for (9.2) with (9.11) by solving (9.10) for all the values of S_0 used to display $S_0 \rightarrow C(S_0, 0)$. Very good accuracy is found (see Fig. 9.3).

Extension To more complex option is possible as long as the PDE is linear. The previous argument does not work with American options (modeled by a variational inequality). It is not always possible to find a double primitive of the adjoint equation, but the discrete equivalent of (9.14) can be found when a variational numerical method such as the FEM is used.

For simplicity, assume that the coefficients in (9.13) do not depend on time, and consider a numerical discretization of (9.13) with an Euler implicit scheme in time with time step δt and piecewise linear finite elements in the variable S . The scheme can be written in matrix form

$$(B + A)u^n - Bu^{n+1} = 0, \quad (9.19)$$

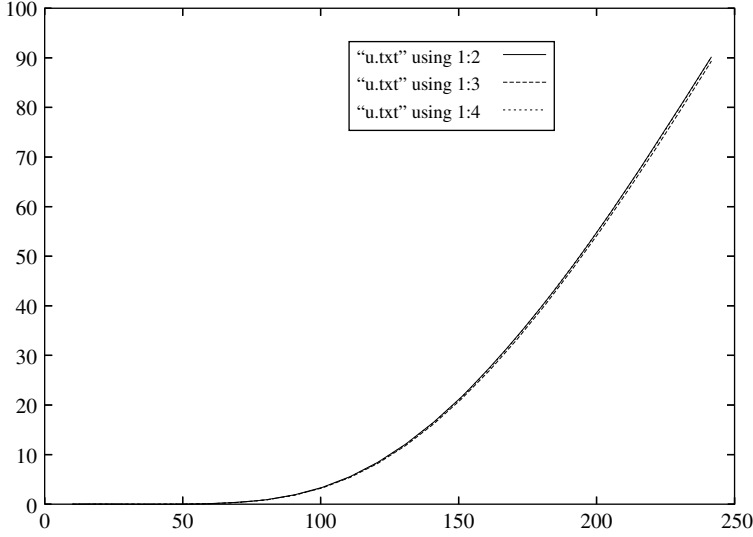


FIG. 9.3 The pricing function of European call option 1 year to maturity computed by three different methods: (a) by solving (9.2), (b) by the Dupire formula (9.10, 9.11), (c) by the Black–Scholes analytic formula.

where $\mathbf{u}^n$ is the vector of the nodal values of the piecewise linear function u_h^n , which approximates $u(\cdot, n\delta t)$. Given a vector $\mathbf{p}^0$, introduce the sequence of vectors $\mathbf{p}^n$ obtained by iterating

$$(\mathbf{A} + \mathbf{B})^T \mathbf{p}^{n+1} - \mathbf{B}^T \mathbf{p}^n = 0. \quad (9.20)$$

Now, notice that (9.19) multiplied by $(\mathbf{p}^{n+1})^T$ gives

$$0 = (\mathbf{p}^{n+1})^T (\mathbf{A} + \mathbf{B}) \mathbf{u}^n - (\mathbf{p}^{n+1})^T \mathbf{B} \mathbf{u}^{n+1} = (\mathbf{p}^n)^T \mathbf{B} \mathbf{u}^n - (\mathbf{p}^{n+1})^T \mathbf{B} \mathbf{u}^{n+1}, \quad (9.21)$$

where the last equality has used (9.20). Summing up over all n gives

$$(\mathbf{p}^0)^T \mathbf{B} \mathbf{u}^0 = (\mathbf{p}^N)^T \mathbf{B} \mathbf{u}^N. \quad (9.22)$$

Choosing $\mathbf{p}_j^0 = \delta_{ij}$, $j = 0, \dots, M$ gives the discrete equivalent of (9.14)

$$\left(\sum_{j=0}^M \mathbf{B}_{ij} \right) \mathbf{u}_i^0 \approx (\mathbf{B} \mathbf{u}^0)_i = (\mathbf{p}^N)^T \mathbf{B} \mathbf{u}^N. \quad (9.23)$$

9.6. A new least square problem

A natural idea is to somehow interpolate the observed prices by a sufficiently smooth function $\tilde{v} : \mathbb{R}_+ \times [0, \max_{i \in I} T_i] \rightarrow \mathbb{R}_+$, then use (9.18) with $v = \tilde{v}$ in order to obtain an

approximation σ . For example, bicubic splines may be used. This approach has several serious drawbacks:

- it is difficult to design an interpolation process such that $\frac{\partial^2 \tilde{v}}{\partial S^2}$ does not take the value 0 and the obtained approximation of the squared volatility is nonnegative.
- there is an infinity of possible interpolations of $\tilde{C}_i$ at (K_i, T_i) , $1 \leq i \leq I$, and for two possible choices, the volatility obtained by (9.18) may differ considerably.

We see that financially relevant additional information have to be added to the interpolation process.

Another natural idea is to consider a new least square problem where the state function satisfies Dupire's equation:

$$\min_{\sigma \in \Sigma} J(\sigma) = \sum_{i=1}^I \omega_i |v(K_i, T_i) - \tilde{C}_i|^2 + J_R(\sigma) \quad \text{subject to (9.18)}. \quad (9.24)$$

The evaluation of J requires solving only one boundary value problem instead of I in the first least square problem.

9.6.1. Numerical results

The same test problem as in Section 9.4 was solved: the local volatility is represented by a spline, and a gradient method with AD is used. Naturally, the computing time is more than 50 times faster: on a 2-GHz dual core Intel, 50 iterations on a 100×50 grid takes 0.5 seconds. The results are shown in Figs. 9.4–9.6.

9.6.2. Adjoint equation

Gradient methods require the differentiation of J with respect to σ . The gradient of J can be computed with AD of computer programs if the chosen representation of σ is done with a small number of parameters. Alternatively, when σ has a large number of degrees of freedom, AD becomes too expensive, and an analytic procedure is needed.

Since J_R explicitly depends on σ , its gradient is easily computed. The gradient of $J_{LS} = \sum_{i=1}^I |v(K_i, T_i) - \tilde{C}_i|^2$ is more difficult to evaluate because the prices $v(K_i, T_i)$ depend on σ in an indirect way: one finds to evaluate the variations of $v(K_i, T_i)$ caused by a small variation of σ ; calling $\delta\sigma$ the variation of σ and δv the induced variation of v , one finds by differentiating (9.18) that $\delta v(S, t = 0) = 0$ and

$$\partial_t \delta v - \frac{\sigma^2(S, t) S^2}{2} \partial_{SS}^2 \delta v + (r - q) S \partial_S \delta v + q \delta v = \sigma \delta \sigma S^2 \partial_{SS}^2 v. \quad (9.25)$$

To express δJ in terms of $\delta\sigma$, an adjoint state function P is introduced, as the solution to the adjoint problem: find the function P such that $P(\bar{T}, \cdot) = 0$ and for $t < \bar{T}$,

$$\partial_t P + \partial_{SS}^2 \left(\frac{\sigma^2 S^2}{2} P \right) - \partial_S (P(r - q)S) - qP = 2 \sum_{i \in I} \omega_i (v(K_i, T_i) - \tilde{C}_i) \delta_{K_i, T_i}, \quad (9.26)$$

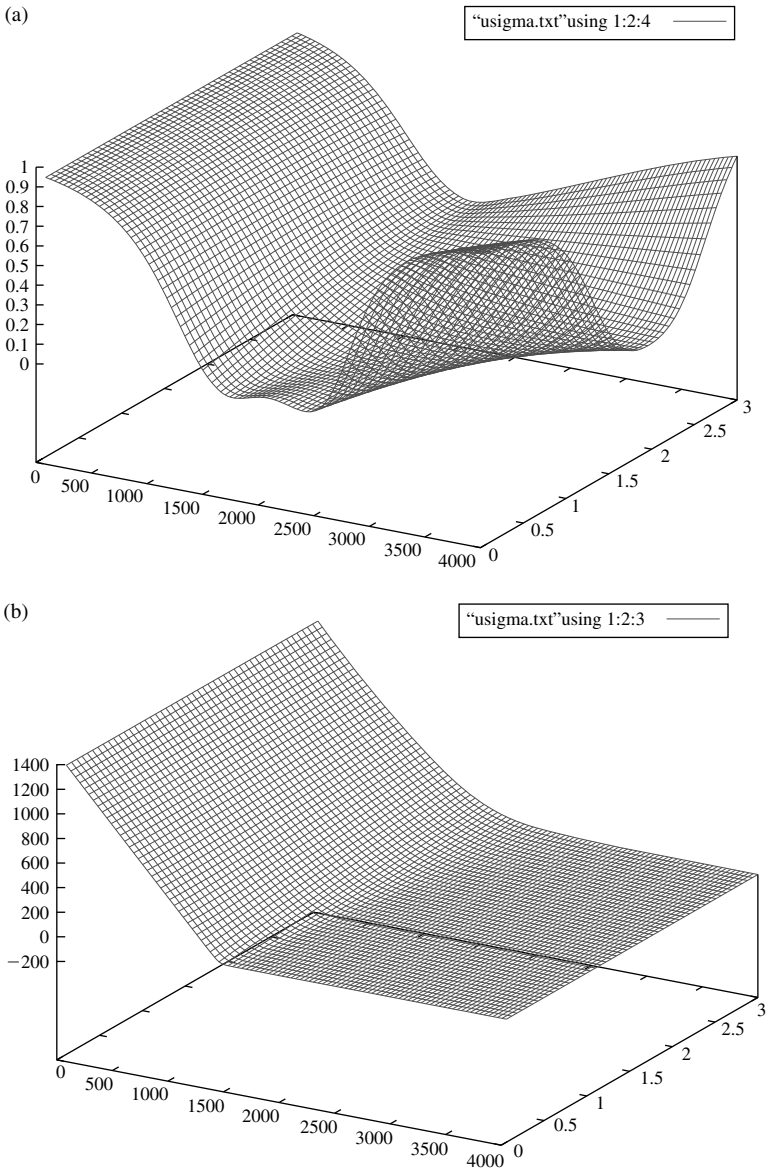


FIG. 9.4 Local volatility surface generated by the optimization algorithm from the parametrization (9.5) and using Dupire's equation. The price surface as a function of strike and maturity is shown (bottom).

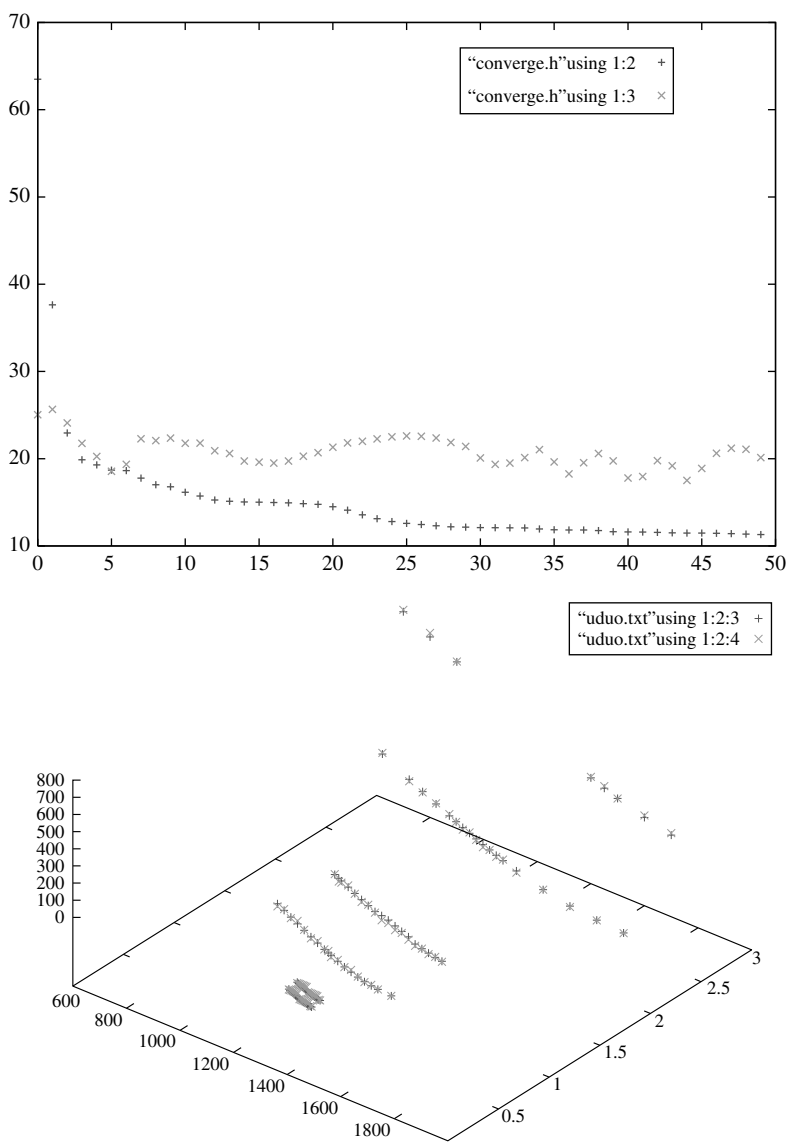


FIG. 9.5 Convergence of $J/1000$ to its local minimum and of the logarithm of the squared norm of the gradient to $-\infty$ as a function of iteration number. Comparison between all observed prices (target) and the prices generated by the model is shown (bottom).

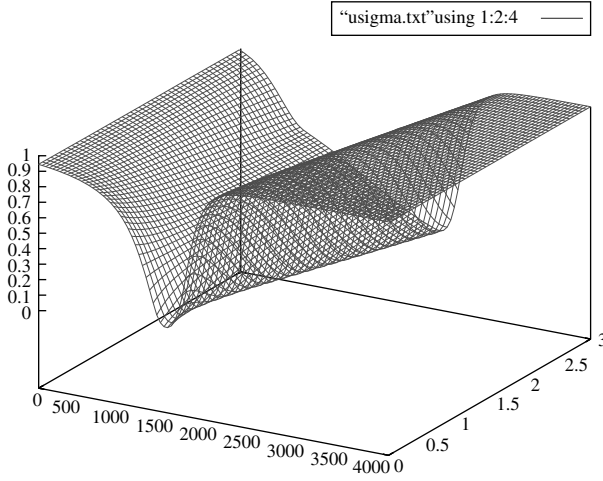


FIG. 9.6 Same as Fig. 9.4 but with (9.9). A total of 100 iterations of conjugate gradient decrease J from 500 000 to 1200, giving an average error smaller than 1%.

where $\bar{T}$ is an arbitrary time greater than $\max_{i \in I} T_i$, and in the right-hand side, δ_{K_i, T_i} denote Dirac functions at (K_i, T_i) . The meaning of (9.26) is the following:

$$\begin{aligned}
 & - \int_Q \left(\partial_t w - \frac{\sigma^2 S^2}{2} \partial_{SS}^2 w + (r - q) S \partial_S w + q w \right) \\
 P &= 2 \sum_{i \in I} \omega_i (v(K_i, T_i) - \bar{C}_i) w(K_i, T_i),
 \end{aligned} \tag{9.27}$$

where $Q = \mathbb{R}_+ \times (0, \bar{T})$, and w is any function such that $w \in L^2((0, \bar{T}), V)$, with $\partial_t w \in L^2(Q)$ and $S^2 \partial_{SS}^2 w \in L^2(Q)$. Taking $v = \delta v$ in (9.27) and using (9.25), one finds

$$\begin{aligned}
 2 \sum_{i \in I} \omega_i (v(K_i, T_i) - \bar{C}_i) \delta v(K_i, T_i) &= 2 \sum_{i \in I} \omega_i (v(K_i, T_i) - \bar{C}_i) \langle \delta_{K_i, T_i}, \delta v \rangle \\
 &= - \int_Q \left(\partial_t \delta v - \frac{\sigma S^2}{2} \partial_{SS}^2 \delta v + (r - q) S \partial_S \delta v + q \delta v \right) \\
 P &= - \int_Q \sigma \delta \sigma S^2 P \partial_{SS}^2 v.
 \end{aligned}$$

We have worked in a formal way, but all the integrations above can be justified. This leads to the estimate

$$\left| \delta J_{LS} + \int_Q \sigma \delta \sigma S^2 P \partial_{SS}^2 v \right| \leq c \|\delta \sigma\|_{L^\infty(Q)}^2,$$

which implies that J_{LS} is differentiable and its differential at point σ is given by

$$DJ_{LS}(\sigma) : \eta \mapsto - \int_Q \sigma \eta S^2 P(\sigma) \partial_{SS}^2 v(\sigma),$$

where $P(\sigma)$ satisfies (9.26), and $v(\sigma)$ satisfies (9.18). We see that the gradient of J_{LS} can be evaluated. When (9.18) is discretized with finite elements, all what has been done can be repeated (with a discrete adjoint problem), and the gradient of the functional can be evaluated in the same way. Let us stress that the gradient $DJ_{LS}(\sigma)$ is computed exactly, which would not be the case with a finite-difference method.

Local volatility can also be calibrated with American options, but it is not possible to find the analogue of Dupire's equation. Thus, in the context of a least square approach, the evaluation of the cost function requires the solution of I variational inequalities, which is computationally expensive (see ACHDOU [2005], ACHDOU and PIRONNEAU [2005]). Nevertheless, it is also possible to find necessary optimality conditions involving an adjoint state.

References

- ACHDOU, Y. (2005). An inverse problem for a parabolic variational inequality arising in volatility calibration with American options. *SIAM J. Control. Optim.* **43** (5), 1583–1615 (electronic).
- ACHDOU, Y., FRANCHI, B., TCHOU, N. (2005). A partial differential equation connected to option pricing with stochastic volatility: regularity results and discretization. *Math. Comput.* **74** (251), 1291–1322 (electronic).
- ACHDOU, Y., GUERMOND, J.-L. (2000). Convergence analysis of a finite element projection/Lagrange-Galerkin method for the incompressible Navier-Stokes equations. *SIAM J. Numer. Anal.* **37** (3), 799–826 (electronic).
- ACHDOU, Y., HECHT, F., POMMIER, D. Space-time a posteriori error estimates for variational inequalities, *J. of Scientific Computing*, to appear.
- ACHDOU, Y., PIRONNEAU, O. (2002). Volatility smile by multilevel least squares. *Int. J. Theor. Appl. Finance* **5** (6), 619–643.
- ACHDOU, Y., PIRONNEAU, O. (2005). *Computational Methods for Option Pricing. Frontiers in Applied Mathematics, vol. 30* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA).
- ACHDOU, Y., PIRONNEAU, O. (2005). Numerical procedure for calibration of volatility with American options. *Appl. Math. Finance* **12** (3), 201–241.
- ACHDOU, Y., TCHOU, N. (2002). Variational analysis for the Black and Scholes equation with stochastic volatility. *M2AN Math. Model. Numer. Anal.* **36** (3), 373–395.
- ANDERSEN, L.B.G., BROTHERTON-RATCLIFFE, R. (1998). The equity option volatility smile: an implicit finite difference approach. *J. Comput. Finance* **1** (2), 5–32.
- AVELLANEDA, M. (1998). Minimum entropy calibration of asset pricing models. *Int. J. Theor. Appl. Finance* **1**, 5–37.
- AVELLANEDA, M., FRIEDMAN, M., HOLMES, C., SAMPERI, D. (1997). Calibrating volatility surfaces via relative entropy minimization. *Appl. Math. Finance* **4**, 37–64.
- AXELSSON, O. (1994). *Iterative Solution Methods* (Cambridge University Press, New York, NY).
- BACHELIER, L. (1995). *Théorie de la spéculation*. Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. Éditions Jacques Gabay, Sceaux. Théorie mathématique du jeu. [Mathematical theory of games], Reprint of the 1900 original.
- BARLES, G. (1994). *Solutions de viscosité des équations de Hamilton-Jacobi. Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 17* (Springer-Verlag, Paris, France).
- BENSOUSSAN, A., LIONS, J.-L. (1984). *Impulse Control and Quasivariational Inequalities*. μ (Gauthier-Villars, Montrouge, France). Transl. from French by J.M. Cole.
- BERGAM, A., BERNARDI, C., MGHAZLI, Z. (2005). A posteriori analysis of the finite element discretization of some parabolic equations. *Math. Comp.* **74** (251), 1117–1138 (electronic).
- BERNARDI, C., MADAY, Y. (1997). Spectral methods. In: *Handbook of Numerical Analysis, vol. V* (North-Holland, Amsterdam, UK), pp. 209–485.
- BLACK, F., SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Pol. Econ.* **81**, 637–659.
- BRAESS, D. (2001). *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics* second ed. (Cambridge University Press, Cambridge, UK).
- BRAESS, D., HACKBUSCH, W. (1983). A new convergence proof for the multigrid method including the V-cycle. *SIAM J. Numer. Anal.* **20** (5), 967–975.
- BRAMBLE, J.H., PASCIAK, J.E., XU, J. (1990). Parallel multilevel preconditioners. *Math. Comp.* **55** (191), 1–22.
- BRENNAN, M.J., SCHWARTZ, E.S. (1977). The valuation of the American put option. *J. Finance* **32**, 449–462.

- BRENNER, S.C., SCOTT, R. (1994). *The Mathematical Theory of Finite Element Methods. Texts in Applied Mathematics, vol. 15* (Springer-Verlag, New York, NY).
- BRIGGS, W.L. (1987). *A Multigrid Tutorial*. (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA).
- BUNGARTZ, H.J., GRIEBEL, M. (2004). Sparse grids. *Acta Numerica* **13**, 1–123.
- CHEN, Z., NOCHETTO, R.H. (2000). Residual type a posteriori error estimates for elliptic obstacle problems. *Numer. Math.* **84** (4), 527–548.
- CHESNEY, M., SCOTT, L. (1989). Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* JSTOR, **24** (3), 267–284.
- CIARLET, P.G. (1978). *The Finite Element Method for Elliptic Problems* (North-Holland, Amsterdam, UK).
- CIARLET, P.G. (1991). Basic error estimates for elliptic problems. In: *Handbook of Numerical Analysis, vol. II* (North-Holland, Amsterdam, UK), pp. 17–351.
- CLÉMENT, P. (1975). Approximation by finite element functions using local regularization. *RAIRO, Sér. Rouge Anal. Numér.* **9** (R-2), 77–84.
- CONT, R., TANKOV, P. (2003). *Financial Modelling with Jump Processes* (Chapman and Hall).
- CONT, R., TANKOV, P. (2004). Nonparametric calibration of jump-diffusion option pricing models. *J. Comput. Finance* **7** (3), 1–49.
- CONT, R., TANKOV, P. (2006). Retrieving Lévy processes from option prices: regularization of an ill-posed inverse problem. *SIAM J. Control Optim.* **45** (1), 1–25 (electronic).
- CONT, R., VOLTCHKOVA, E. (2003). Finite difference methods for option pricing in jump-diffusion and exponential Lévy models. Rapport Interne 513, CMAP, Ecole Polytechnique.
- COURANT, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.* **49**, 1–23.
- CRANDALL, M.G., ISHII, H., LIONS, P.-L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)* **27** (1), 1–67.
- DUPIRE, B. (1994). Pricing with a smile. *Risk* 18–20.
- DUPIRE, B. (1997). Pricing and hedging with smiles. In: *Mathematics of Derivative Securities (Cambridge, 1995)* (Cambridge University Press, Cambridge, UK), pp. 103–111.
- ERIKSSON, K., ESTEP, D., HANSBO, P., JOHNSON, C. (1995). Introduction to computational methods for differential equations. In: *Theory and Numerics of Ordinary and Partial Differential Equations, IV*, pp. 77–122 (Oxford University Press, New York, NY).
- ERIKSSON, K., JOHNSON, C. (1991). Adaptive finite element methods for parabolic problems. I. A linear model problem. *SIAM J. Numer. Anal.* **28** (1), 43–77.
- ERIKSSON, K., JOHNSON, C. (1995). Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.* **32** (3), 706–740.
- ERN, A., GUERMOND, J.-L. (2004). *Theory and Practice of Finite Elements. Applied Mathematical Sciences, vol. 159* (Springer-Verlag, New York, NY).
- EYMARD, R., GALLOUËT, T., HERBIN, R. (2000). Finite volume methods. In: *Handbook of Numerical Analysis, vol. VII* (North-Holland, Amsterdam, UK), pp. 713–1020.
- FLEMING, W.H., SONER, H.M. (1993). *Controlled Markov Processes and Viscosity Solutions. Applications of Mathematics (New York), vol. 25* (Springer-Verlag, New York, NY).
- FOUQUE, J.P., PAPANICOLAOU, G., SIRCAR, K.R. (2000). *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge University Press, Cambridge, UK).
- FRIEDMAN, A. (1964). *Partial Differential Equations of Parabolic Type* (Prentice-Hall Inc., Englewood Cliffs, NJ).
- FRIEDRICHS, K.O. (1944). The identity of weak and strong extensions of differential operators. *Trans. Amer. Math. Soc.* **55**, 132–151.
- GEORGE, P.L., HECHT, F., SALTEL, E. (1991). Automatic mesh generator with specified boundary. *Comput. Methods Appl. Mech. Engrg.* **92** (3), 269–288.
- GLOWINSKI, R., LIONS, J.-L., TRÉMOLIÈRES, R. (1981). *Numerical Analysis of Variational Inequalities. Studies in Mathematics and its Applications, vol. 8* (North-Holland Publishing Co., Amsterdam, UK). Transl. from French.

- GOLUB, G.H., VAN LOAN, C.F. (1989). *Matrix Computations* second ed. (John Hopkins University Press).
- GREENBAUM, A. (1997). *Iterative Methods for Solving Linear Systems. Frontiers in Applied Mathematics*, vol. 17 (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA).
- GRIEBEL, M. (1998). Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing* **61** (2), 151–179.
- GRIEBEL, M., OSWALD, P. (1995). Tensor-product-type subspace splittings and multilevel iterative methods for anisotropic problems. *Adv. Comput. Math.* **4**, 171–206.
- GRIEBEL, M., SCHNEIDER, M., ZENGER, C. (1992). A combination technique for the solution of sparse grid problems. In *Proceedings of the IMACS International Symposium on Iterative Methods in Linear Algebra* (Elsevier, Amsterdam, Netherlands) pp. 263–281.
- GRIEWANK, A. (2000). *Evaluating derivatives: principles and techniques of algorithmic differentiation*, (SIAM, Philadelphia, PA).
- GUERMOND, J.-L. (1999). Un résultat de convergence d'ordre deux en temps pour l'approximation des équations de Navier-Stokes par une technique de projection incrémentale. *M2AN Math. Model. Numer. Anal.* **33** (1), 169–189.
- GUERMOND, J.-L., QUARTELLA, L. (1998). On the approximation of the unsteady Navier-Stokes equations by finite element projection methods. *Numer. Math.* **80** (2), 207–238.
- HASCOET, L., PASCUAL, V. (2004). Tapenade 2.1 users guide. *Writing* 09.
- HESTON, S. (1993). A closed form solution for options with stochastic volatility with application to bond and currency options. *Rev. Financ. Stud.* 327–343.
- HINTERMÜLLER, M., ITO, K., KUNISCH, K. (2002). The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.* **13** (3), 865–888 (electronic) (2003).
- HOPPE, R.H.W., KORNUBER, R. (1994). Adaptive multilevel methods for obstacle problems. *SIAM J. Numer. Anal.* **31** (2), 301–323.
- HOPPE, R.H.W. (1987). Multigrid algorithms for variational inequalities. *SIAM J. Numer. Anal.* **24** (5), 1046–1065.
- HULL, J.C., WHITE, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281–300.
- IKONEN, S., TOIVANEN, J. (2004). Operator splitting methods for American option pricing. *Appl. Math. Lett.* **17** (7), 809–814.
- IKONEN, S., TOIVANEN, J. (2006). Efficient numerical methods for pricing American options under stochastic volatility. to appear in *Numer. Methods. Partial Differ. Equ.*
- ITO, K., KUNISCH, K. (2003). Semi-smooth Newton methods for variational inequalities of the first kind. *M2AN Math. Model. Numer. Anal.* **37** (1), 41–62.
- JACKSON, N., SÜLI, E., HOWISON, S. (1998). Computation of deterministic volatility surfaces. *Appl. Math. Finance* **2** (2), 5–37.
- JAILLET, P., LAMBERTON, D., LAPEYRE, B. (1990). Variational inequalities and the pricing of American options. *Acta Appl. Math.* **21** (3), 263–289.
- KARATZAS, I. (1988). On the pricing of American options. *Appl. Math. Optim.* **17** (1), 37–60.
- KARATZAS, I., SHREVE, S.E. (1991). *Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics*, vol. 113 second ed. (Springer-Verlag, New York, NY).
- KINDERLEHRER, D., STAMPACCHIA, G. (1980). *An Introduction to Variational Inequalities and their Application* (Academic Press).
- KOSTER, F. (2000). A proof of the consistency of the finite difference technique on sparse grids. *Computing* **65** (3), 247–261.
- KRYLOV, N.V. (1996). *Lectures on Elliptic and Parabolic Equations in Hölder Spaces* (American Mathematical Society, Providence, RI).
- LADYŽENSKAJA, O.A., SOLONNIKOV, V.A., URAL'CEVA, N.N. (1967). *Linear and quasilinear equations of parabolic type*. Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23 (American Mathematical Society, Providence, RI).
- LAMBERTON, D., LAPEYRE, B. (1997). *Introduction au Calcul Stochastique appliqué à la Finance* (Ellipses).
- LIONS, J.L. (1969). *Quelques Méthodes de Résolution des Problèmes aux Limites non Linéaires* (Dunod, Paris, France).

- LIONS, J.L., MAGENES, E. (1968). *Problèmes aux limites non homogènes et applications, vol. I and II* (Dunod, Paris, France).
- LIONS, P.-L., MERCIER, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16** (6), 964–979.
- MATACHE, A.-M., NITSCHKE, P.-A., SCHWAB, C. (2003). Wavelet Galerkin pricing of American options on Lévy driven assets. Research Report SAM 2003–06.
- MATACHE, A.M., VON PETERSDOFF, T., SCHWAB, C. (2004). Fast deterministic pricing of Lévy driven assets. *Math. Model. Numer. Anal.* **38** (1), 37–72.
- MERTON, R.C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.* **3**, 125–144.
- MERTON, R.C. (1973). Theory of rational option pricing. *Bell. J. Econ. Manage. Sci.* **4**, 141–183.
- MEURANT, G. (1999). *Computer Solution of Large Linear Systems. Studies in Mathematics and its Applications*, vol. 28 (North-Holland Publishing Co., Amsterdam, Netherlands).
- NAGEL, A., STEIN, E.M., WAINGER, S. (1985). Balls and metrics defined by vector fields. I. Basic properties. *Acta Math.* **155** (1–2), 103–147.
- NAYAK, S., PAPANICOLAOU, G. (2006). Stochastic volatility surface estimation.
- NOCHETTO, R.H., SIEBERT, K.G., VEESER, A. (2003). Pointwise a posteriori error control for elliptic obstacle problems. *Numer. Math.* **95** (1), 163–195.
- NOCHETTO, R.H., SIEBERT, K.G., VEESER, A. (2005). Fully localized a posteriori error estimators and barrier sets for contact problems. *SIAM J. Numer. Anal.* **42** (5), 2118–2135 (electronic).
- OOSTERLEE, C.W. (2003). On multigrid for linear complementarity problems with application to American-style options. *Electron. Trans. Numer. Anal.* **15**, 165–185 (electronic). In: *Tenth Copper Mountain Conference on Multigrid Methods* (Copper Mountain, CO, 2001).
- PAZY, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences*, vol. 44 (Springer-Verlag, New York, NY).
- PHAM, H. (1998). Optimal stopping of controlled jump-diffusion processes: a viscosity solution approach. *J. Math. Syst* **8** (1), 1–27.
- POLAK, E. (1997). *Optimization, Algorithms and Consistent Approximations. Applied Mathematical Sciences*, vol. 124 (Springer-Verlag, New York, NY).
- PROTTER, M.H., WEINBERGER, H.F. (1984). *Maximum Principles in Differential Equations*. (Springer-Verlag, New York, NY). Corrected reprint of the 1967 original.
- QUARTERONI, A. (1991). An introduction to spectral methods for partial differential equations. In *Advances in Numerical Analysis*, vol. 1 (Lancaster, 1990) (Oxford Sci. Publ., Oxford Univ. Press, New York, NY), pp. 96–146.
- RAVIART, P.-A., THOMAS, J.-M. (1983). *Introduction à L'analyse Numérique des équations aux Dérivées Partielles* (Masson, Paris, France).
- REISINGER, C. Analysis of linear difference schemes in the sparse grid combination technique. preprint, (available at: <http://eprints.maths.ox.ac.uk>).
- REISINGER, C. (2004). Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben. PhD thesis, Universität Heidelberg.
- REISINGER, C., WITTUM, G. (2004). On multigrid for anisotropic equations and variational inequalities: pricing multi-dimensional European and American options. *Comput. Vis. Sci.* **7** (3–4), 189–197.
- REISINGER, C., WITTUM, G. (2007). Efficient hierarchical approximation of high-dimensional option pricing problems. *SIAM J. Sci. Comput.* **29** (1), 440–458 (electronic).
- RICHTMYER, R.D., MORTON, K.W. (1994). *Difference Methods for Initial-Value Problems* second ed. (Robert E. Krieger Publishing Co. Inc., Malabar, FL).
- RUGE, J.W., STÜBEN, K. (1987). Algebraic multigrid. In *Multigrid Methods. Frontiers in Applied Mathematics*, vol. 3 (SIAM, Philadelphia, PA), pp. 73–130.
- SAAD, Y. (1996). *Iterative Methods for Sparse Linear Systems* (PWS Publishing Company).
- SCHIEKOFER, T. (1998). Die methode der finiten differenzen auf dünnen gittern zur lösung elliptischer und parabolischer. PhD thesis, Universität Bonn.
- SCHÖTZAU, D., SCHWAB, C. (2001). hp-discontinuous Galerkin time-stepping for parabolic problems. *C. R. Acad. Sci. Paris Sér. I Math.* **333** (12), 1121–1126.

- STEIN, E., STEIN, J. (1991). Stock price distributions with stochastic volatility : an analytic approach. *Rev. Financ. Stud.* **4** (4), 727–752.
- STRANG, G., FIX, G.J. (1973). *An Analysis of the Finite Element Method* (Prentice-Hall, Englewood Cliffs, N.J.).
- THOMÉE, V. (1997). *Galerkin Finite Element Methods for Parabolic Problems. Springer Series in Computational Mathematics, vol. 25* (Springer-Verlag, Berlin).
- VEESER, A. (2001). Efficient and reliable a posteriori error estimators for elliptic obstacle problems. *SIAM J. Numer. Anal.* **39** (1), 146–167 (electronic).
- VERFURTH, R. (1996). *A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques* (Wiley, Chichester).
- VILLENEUVE, S. (2004). Exercise regions of American options on several assets. *Finance Stoch.* **3** (3), 295–322.
- VON PETERSDOFF, T., SCHWAB, C. (2004). Numerical solutions of parabolic equations in high dimensions. *Math. Model. Numer. Anal.* **38** (1), 93–128.
- WERDER, T., GERDES, K., SCHÖTZAU, D., SCHWAB, C. (2001). *hp*-discontinuous Galerkin time stepping for parabolic problems. *Comput. Methods Appl. Mech. Engrg.* **190** (49–50), 6685–6708.
- YSERENTANT, H. (1993). Old and New Convergence Proofs for Multigrid Methods. *Acta Numerica* (Cambridge Univ. Press, Cambridge), pp. 285–326.
- ZENGER, C. (1991). Sparse grids. In: Hackbusch, W. (ed.), *Parallel Algorithms for Partial Differential Equations. Notes on Numerical Fluid Mechanics, vol. 31* (Braunschweig/Wiesbaden, Vieweg).
- ZIENKIEWICZ, O.C., TAYLOR, R.L. (2000). *The Finite Element Method. vol. 1* fifth ed. (Butterworth-Heinemann, Oxford, UK). The basis.
- ZVAN, R., FORSYTH, P.A., VETZAL, K.R. (1998). Penalty methods for American options with stochastic volatility. *J. Comput. Appl. Math.* **91** (2), 199–218.

Advanced Monte Carlo Methods for Barrier and Related Exotic Options

Emmanuel Gobet

*Laboratoire Jean Kuntzmann, Université de Grenoble and CNRS, BP 53, 38041
Grenoble cedex 9, France.
E-mail address: emmanuel.gobet@imag.fr*

Abstract

In this work, we present advanced Monte Carlo techniques applied to the pricing of barrier options and other related exotic contracts. It covers in particular the Brownian bridge approaches, the barrier shifting techniques (BAST), and their extensions. We leverage the link between discrete and continuous monitoring to design efficient schemes, which can be applied to the Black–Scholes model but also to stochastic volatility or Merton’s jump models. This is supported by theoretical results and numerical experiments.

1. Introduction

In this chapter, we review and extend advanced techniques for the valuation of barrier options (initially introduced by MERTON [1973]) and other financial contracts, whose activation depends on whether an underlying asset has reached a specified level (the barrier) or not. From the Monte Carlo point of view, these types of payoff are difficult to simulate because they are strongly path dependent and they are discontinuous with respect to the path of the monitored process. The option prototype is the Down and Out Call (DOC), which is a European call with strike K and expiration date T , but it is paid if and only if the asset X has not reached the lower level D before expiration. This is a knockout option, whose payoff can be written as follows:

$$\Psi_T = \mathbf{1}_{\forall t < T: X_t > D} (X_T - K)_+.$$

On the other hand, a knock-in option comes into effect only when the barrier is reached. However, a knockout option plus a knock-in option with same parameters readily gives a vanilla option; thus, by absence of arbitrage, one needs to focus on only one type, and the focus will be on the knockout in what follows. We mention that all these options are also known as trigger options. Additionally, a rebate can be paid at the trigger when the option ceases to exist (in the case of a knockout option).

The type of barrier options can be widened with many respects. First, it can be designed for several assets simultaneously, with barriers on each asset. Double barriers (lower and upper) or time-dependent barriers are complementary choices. Also, there may be a discrete-time monitoring instead of a continuous one: for instance, the option is not knocked out if the asset remains above the barrier, daily at a fixing hour. Moreover, it is quite fruitful to connect the discrete and continuous monitoring.

At this level of description, the knowledge of the stochastic model does not matter. It could be given by the solution of a Brownian stochastic differential equations (SDE), and it could include jumps as well. It could be related to Forex markets, where these options are of the most popular types, but related assets can be traded on equity or fixed income markets as well. Usually, they are over-the-counter contracts. Within the structured products, it is also quite common to meet barrier-type options.

The problem to simulate such payoff Ψ_T is that one has to compare the value of the asset and of the barrier at any time before expiration date T to decide whether the option is knocked out or not, whereas one is able to simulate the path only at a fixed number, N , of times (the monitoring dates) $(t_i)_{1 \leq i \leq N}$. If the monitoring frequency is high, one expects that the payoffs Ψ_T and $\Psi_T^N = \mathbf{1}_{\forall t_i < T: X_{t_i} > D}(X_T - K)_+$ are close to each other. This may be false for some simulation scenario but true when one takes the expectation, which is the criterion of interest for valuation. However, the monitoring bias $\mathbb{E}\Psi_T^N - \mathbb{E}\Psi_T$ shrinks to 0 quite slowly (as $N^{-1/2}$, see the references hereafter), and besides this, it is positive because we may not detect that a trigger has occurred between two successive dates. The bad consequence is to systematically overestimate the price (here, we omit the discount factor when referring to the price).

To overcome this difficulty, there are two numerical strategies.

- First, to discretize the state space of the asset X (or its logarithm). Then, simulating the successive hitting times of this space grid is a smart way (see ROGERS and STAPLETON [1998]) to well control the behavior of the path and thus to reduce the simulation bias. Actually to be optimal, one should choose carefully the grid so that the barrier lies in the grid. This is closely related to binomial or multinomial tree methods. We refer to BOYLE and LAU [1994], BOYLE and TIAN [1998], CHEUK and VORST [1996], DERMAN, KANI, ERGENER, and BARDHAN [1995], RITCHKEN [1995], ROGERS and STAPLETON [1998], and RUBINSTEIN and REINER [1991]. This approach becomes difficult to design, especially for several dependant assets and when jumps may occur.
- Second, to take into account that the option may be knocked out between monitoring times. One possibility is to simulate the trigger event conditionally on the known values of X . This is known as Brownian bridge techniques, and this dates back to BALDI [1995]. This and related refinements will be discussed in Section 2. Another

possibility consists in shifting the trigger in order to compensate the overestimation. This idea, initially put in a financial framework by BROADIE, GLASSERMAN, and KOU [1997] for a single asset in the Black–Scholes model, can be extended in many ways (multiassets, local volatility, time-dependent barrier, jumps, etc). Actually, it turns out to be quite flexible. It is discussed in Section 3.

Section 4 brings together several financial examples, where we illustrate how the previous methods perform. We do not present results related to the computations of Greeks (see GOBET [2004] and the references therein).

2. Brownian bridge techniques

2.1. A toy example

Before dealing with advanced techniques designed for sophisticated models, we begin with the simplest financial example of Black–Scholes model with constant coefficients, used to price a DOC option. Thus, under the risk-neutral probability, the dynamics of the underlying asset X is given by

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t,$$

where W is a standard Wiener process. Here, σ is the volatility, and μ is the drift under the risk-neutral probability (generally equal to the interest rate minus the continuous dividend rate if X is an equity or to the difference of interest rates of both economies if X is an exchange rate). To simplify even more, consider that the discount factor equals 1 (zero interest rate). Then, the fair price is given by the expectation

$$\mathbb{E}(\mathbf{1}_{\forall t < T: X_t > D}(X_T - K)_+),$$

which we may evaluate by Monte Carlo methods. Of course, in this toy example, one has a closed formula for this price (see RUBINSTEIN and REINER [1991] and Eq. (3.14)), which is useful for numerics to check the validity of a procedure.

2.1.1. Without the simulation of the trigger.

For the simulation, we may proceed as follows.

- first, simulate X_T by¹

$$X_T \stackrel{d}{=} x_0 \exp((\mu - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z),$$

where Z is Gaussian variable with zero mean and unit variance.

- second, compute (analytically) the conditional trigger probability

$$p(x, y, T, D, |\sigma|) = \mathbb{P}(\exists t < T : X_t \leq D | X_0 = x, X_T = y) \quad (2.1)$$

¹ $U \stackrel{d}{=} V$ means that U and V have the same distribution.

and as an output, take

$$[1 - p(x_0, X_T, T, D, |\sigma|)](X_T - K)_+.$$

Draw many independent simulations of the output, and by averaging them out, one gets an approximation of

$$\begin{aligned} & \mathbb{E}([1 - p(x_0, X_T, T, D, |\sigma|)](X_T - K)_+) \\ &= \mathbb{E}(\mathbb{E}(\mathbf{1}_{\forall t < T: X_t > D} | X_0, X_T)(X_T - K)_+) \\ &= \mathbb{E}(\mathbf{1}_{\forall t < T: X_t > D}(X_T - K)_+). \end{aligned}$$

This leads to an unbiased Monte Carlo procedure provided that we are able to compute the conditional trigger probability. Actually, this quantity is explicit, and it is given by

$$p(x, y, T, D, |\sigma|) = \begin{cases} 1 & \text{if } x \text{ or } y \text{ are below } D, \\ \exp(-2 \frac{\log(x/D) \log(y/D)}{|\sigma|^2 T}) & \text{otherwise.} \end{cases} \quad (2.2)$$

This expression easily follows from the known distribution of the value of a Brownian motion and its running maximum at a given time (see REVUZ and YOR [1994]). Actually, this is equal to the probability that a Brownian bridge (namely, $(\log(X_t))_t$ conditionally to $\log(X_0)$ and $\log(X_T)$) reaches the level $\log(D)$: this gives the label of such approach as Brownian bridge techniques. It is worth noticing that the conditional trigger probability does not depend on the drift μ .

PSEUDOCODE

```
Z=gauss(0,1);
X=X0*exp((mu-0.5*sigma^2)*T+ sigma*sqrt(T)*Z);
p=p(X0,X,T,D,|sigma|);
return (1-p)*max(X-K,0);
```

2.1.2. With the simulation of the trigger.

In this toy example, to generate a scenario, we only need to draw one random variable (i.e., Z) and to weight the call payoff by the nontrigger probability. Note that one could also simulate the trigger by taking as an output

$$\mathbf{1}_{U > p(0, x_0, X_T, T, D, |\sigma|)}(X_T - K)_+,$$

with an extra random variable U , independent of Z and uniformly distributed on $[0, 1]$. The procedure is still unbiased since the expectation remains unchanged. Only the variance is modified, and it is now larger because the variance of the conditional expectation is smaller than the variance. Thus, the confidence interval is wider, and in this example, simulating the trigger is not relevant to the computational efficiency.

2.2. Easy extensions

From the simple principle described above, we can derive several easy extensions, which allow us to handle less-specific payoffs. Whereas in the previous examples, there was

no advantage to generate the underlying asset along a time grid, it is no longer true if the coefficients μ and σ are time dependent, say piece-wise constant on each subinterval² $[t_i, t_{i+1}]$ and equal to μ_i and σ_i . In that case, one has to generate $(X_{t_i})_i$ at those times

$$X_{t_{i+1}} \stackrel{d}{=} X_{t_i} \exp\left((\mu_i - \frac{1}{2}\sigma_i^2)\Delta_i + \sigma_i\sqrt{\Delta_i}Z_i\right)$$

with i.i.d. random variables $Z_i \stackrel{d}{=} Z$. Then, as an output, take

$$(X_T - K)_+ \prod_{i=0}^{N-1} (1 - p(X_{t_i}, X_{t_{i+1}}, \Delta_i, D, |\sigma_i|)), \quad (2.3)$$

where the function p is still defined by (2.2).

PSEUDOCODE

```

X=X0;
prob=1;
for i=0 to N-1
    Z=gauss(0,1);
    Y=X*exp((mu_i-0.5*sigma_i^2)*Delta_i
            + sigma_i*sqrt(Delta_i)*Z);
    prob=prob*(1-p(X,Y,Delta_i,D,|sigma_i|));
    X=Y;
return prob*max(X-K,0);

```

If the barrier D is time dependent and piecewise constant, at each time step, one has to evaluate $p(X_{t_i}, X_{t_{i+1}}, \Delta_i, D_i, |\sigma_i|)$ instead of $p(X_{t_i}, X_{t_{i+1}}, \Delta_i, D, |\sigma_i|)$. Then, the procedure remains unbiased.

If there is an upper barrier U , we proceed very analogously, except that the conditional trigger probability p is now given by

$$p(x, y, \Delta, U, |\sigma|) = \begin{cases} 1 & \text{if } x \text{ or } y \text{ are above } U, \\ \exp\left(-\frac{\log(x/U)\log(y/U)}{|\sigma|^2\Delta}\right) & \text{otherwise.} \end{cases} \quad (2.4)$$

As in the toy example, we could alternatively simulate the trigger event. In the same way, it slightly increases the variance of simulations. But now, the pseudocode may be computationally cheaper since the loop over i can be stopped as soon as the option is knocked out.

For a double-barrier option (with down and up barriers D and U), the expression for p is still explicit:

$$p(x, y, \Delta, D, U, |\sigma|) = \begin{cases} 1 & \text{if } x \text{ or } y \text{ are above } U \text{ or if } x \text{ or } y \text{ are below } D, \\ \sum_{k=-\infty}^{+\infty} \left[\exp\left(-2\frac{k\log(U/D)(k\log(U/D)+\log(y/x))}{|\sigma|^2\Delta}\right) \right. \\ \left. - \exp\left(-2\frac{(k\log(U/D)+\log(x/U))(k\log(U/D)+\log(y/U))}{|\sigma|^2\Delta}\right) \right] & \text{otherwise.} \end{cases} \quad (2.5)$$

²Regarding the notations, we put $t_0 = 0$, $t_N = T$, and $\Delta_i = t_{i+1} - t_i$.

However, from large-deviation arguments (see BALDI [1995]), we know that one has to consider only the barrier that is the closest to x and y (and neglect the other one) and at a distance of order $|\sigma|\sqrt{\Delta_i}$, and then, the evaluation of p as a series boils down to the computation of one term. For a dense time grid (Δ_i small), there is no possible ambiguity for such choice of the closest barrier, but it may be questionable if Δ_i is not small (X_{t_i} may be close to U and $X_{t_{i+1}}$ close to D).

Notice that it is also straightforward to take into account in the simulation procedure that a jump component lies in the dynamics of X . For instance, if

$$\frac{dX_t}{X_{t-}} = \mu dt + \sigma dW_t + Y_t dN_t, \quad (2.6)$$

where N is a Poisson process with parameter λ and Y stands for the random jumps, we proceed as follows. Simulate the $k = N_T$ jump times $(\tau_i)_i$ up to time T ($\tau_0 = 0$). Between two jumps, X behaves like a geometric Brownian motion for which we can apply Brownian bridge techniques. This heuristic is fully justified by the fact that the Brownian part and the jump part are independent. Thus, we average out independent simulations of

$$(X_T - K)_+ (1 - p(X_{\tau_k}, X_T, T - \tau_k, D, U, |\sigma|)) \times \prod_{i=0}^{k-1} (1 - p(X_{\tau_i}, X_{\tau_{i+1}}^-, \tau_{i+1} - \tau_i, D, U, |\sigma|)) \quad (2.7)$$

to get asymptotically and without bias the required expectation.

2.3. Further approximations

Up to now, we have described only unbiased procedures. But too often, models and payoffs are so complex that approximations are necessary to get through the numerics.

First, when the underlying process has a fairly general dynamics such as

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

(with matrix notations) with nonconstant (or nonlinear) coefficients μ and σ , it is likely that one cannot simulate exactly the process at given times. One alternative is to use an Euler scheme (based on a regular grid $(t_i)_{0 \leq i \leq N}$ with time step $\Delta = T/N$), which is defined by $X_0^N = X_0$ and

$$X_{t_{i+1}}^N = X_{t_i}^N + \mu(t_i, X_{t_i}^N)\Delta + \sigma(t_i, X_{t_i}^N)(W_{t_{i+1}} - W_{t_i}).$$

If X is a coordinate-wise positive process, it is better to use the Euler scheme on the log-process $\log(X)$ to keep this positivity property. Regarding the Brownian bridge techniques, they are still applicable because on each interval $[t_i, t_{i+1}]$, X^N is a Brownian motion with drift: $X_t^N = X_{t_i}^N + \mu(t_i, X_{t_i}^N)(t - t_i) + \sigma(t_i, X_{t_i}^N)(W_t - W_{t_i})$. Thus, $(X_t^N)_{t_i \leq t \leq t_{i+1}}$ conditionally to $X_{t_i}^N = x$ and $X_{t_{i+1}}^N = y$ has the same law as $(\bar{x}_t = x + \sigma(t_i, x)(W_t - W_{t_i}))_{t_i \leq t \leq t_{i+1}}$ conditionally to $\bar{x}_{t_{i+1}} = y$ (note that the law of the

Brownian bridge does not depend anymore of the initial drift). Now suppose, for instance, that the trigger is made via a knockout on the first asset X_1 at the lower level D and that the payoff has the general form $f(X_T)$. Thus, the conditional trigger probability of the Euler scheme is given by

$$\mathbb{P}(\exists t \in [t_i, t_{i+1}] : X_{1,t}^N \leq D | X_{t_i}^N, X_{t_{i+1}}^N) = p(X_{t_i}^N, X_{t_{i+1}}^N, \Delta, D, |\sigma_1(t_i, X_{t_i}^N)|)$$

with $\sigma_1(\cdot)$ is the diffusion coefficient of X_1 (equal to the first row of $\sigma(\cdot)$) and

$$p(x, y, \Delta, D, \bar{\sigma}) = \begin{cases} 1 & \text{if } x_1 \text{ or } y_1 \text{ are below } D, \\ \exp(-2 \frac{(x_1 - D)(y_1 - D)}{\bar{\sigma}^2 \Delta}) & \text{otherwise.} \end{cases}$$

Then, as an output take

$$f(X_T^N) \prod_{i=0}^{N-1} (1 - p(X_{t_i}^N, X_{t_{i+1}}^N, \Delta, D, |\sigma_1(t_i, X_{t_i}^N)|))$$

analogously to (2.3).

PSEUDOCODE

```

X=X0;
prob=1;
for i=0 to N-1
    // Gaussian vector for the Brownian increments
    Z=gauss(0, Id);
    // component by component
    Y=X+mu(t_i, X)*Delta +sigma(t_i, X)*sqrt(Delta)*Z;
    prob=prob*(1-p(X, Y, Delta, D, |sigma_1(t_i, X)|));
    X=Y;
return prob*f(X);

```

In this situation, the Monte Carlo procedure is biased because we use the Euler scheme. Indeed, $(X_{t_i}^N)_{0 \leq i \leq N}$ are only approximative simulation values of $(X_{t_i})_{0 \leq i \leq N}$: the accuracy in L_p norm is of order $\Delta^{1/2}$ (i.e., $\|\sup_{i \leq N} |X_{t_i}^N - X_{t_i}|\|_p = O(\Delta^{1/2})$). But one knows that for the evaluation of $\mathbb{E}(f(X_T))$, the accuracy becomes a $O(\Delta)$ (this is the weak error, see BALLY and TALAY [1996] for the mathematical analysis). The extension of this error analysis to barrier options has been carried out in GOBET [2000], and it is shown that the weak error is still of order Δ . Hence, provided that one suitably weights the payoff f by the conditional trigger probability of the Euler scheme, the simulation bias is as small as if there was no barrier. It also means that it is not worth using a Milshtein scheme because the weak error has the same magnitude (in addition, we recall that the Milshtein scheme may be harder to use in general models because of restrictive conditions on σ and its derivatives).

The crucial point in the above arguments is that one can analytically compute the trigger probability of the Euler scheme bridge (or equivalently of a Brownian bridge). For upper barrier and double barriers, this is possible as explained in Section 2.2. Difficulties really arise when barriers are multiple. When the trigger is associated with the first exit time of a given domain $\mathcal{D}$ (i.e., the option payoff is of the form $\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)$),

one may use large-deviation arguments (see BALDI [1995]) to get an accurate approximation of

$$p(x, y, \Delta, \mathcal{D}, \sigma(t_i, x)) = \mathbb{P}(\exists t \in [t_i, t_{i+1}] : X_t^N \notin \mathcal{D} | X_{t_i}^N = x, X_{t_{i+1}}^N = y)$$

as Δ goes to 0. An asymptotic expansion for this probability is available (for smooth domain). But it is computational demanding and thus this is of limited interest for real-time computations. One may alternatively think of locally approximating the domain by a half-space (see GOBET [2001]). This may be formalized as follows. For $X_{t_i}^N = x$ close to the boundary $\partial\mathcal{D}$, denote by $\pi(x)$ its projection on $\partial\mathcal{D}$ and by $n(x)$ the associated inward normal vector at this point. In the computation of $p_i(x, y, \Delta)$, then replace $\mathcal{D}$ by the half-space containing x and delimited by the tangent hyperplane at the point $\pi(x)$. Then, the conditional trigger probability becomes explicit because this boils down to a one-dimensional situation. This leads to

$$p(x, y, \Delta, \mathcal{D}, \sigma(t_i, x)) \approx \begin{cases} 1 & \text{if } x \text{ or } y \text{ is outside } \mathcal{D}, \\ \exp(-2 \frac{d(x, \mathcal{D})d(y, \mathcal{D})}{|\sigma(t_i, x) \cdot n(x)|^2 \Delta}) & \text{otherwise.} \end{cases}$$

One can prove (see GOBET [2001]) that the simulation bias is still of order Δ as before when the conditional trigger probability was exactly computed. But this result is valid if the domain is smooth enough; while in practice for multiple barriers, the domain is of the form

$$\mathcal{D} =]D_1, U_1[\times \cdots \times]D_d, U_d[\quad (2.8)$$

(where d is the number of assets); thus, it exhibits corners. Hence, for each asset, there is a single or double barrier for which the individual conditional trigger probability is easy to evaluate³:

$$\begin{aligned} p_{i,j}(x, y, \Delta) &= \mathbb{P}(\exists t \in [t_i, t_{i+1}] : X_{j,t}^N \notin]D_j, U_j[| X_{t_i}^N = x, X_{t_{i+1}}^N = y) \\ &= \begin{cases} 1 & \text{if } x_j \text{ or } y_j \text{ are above } U_j \text{ or below } D_j, \\ \sum_{k=-\infty}^{+\infty} \left[\exp(-2 \frac{k(U_j - D_j)(k(U_j - D_j) + y_j - x_j)}{|\sigma_j(t_i, x)|^2 \Delta}) \right. \\ \quad \left. - \exp(-2 \frac{(k(U_j - D_j) + x_j - U_j)(k(U_j - D_j) + y_j - U_j)}{|\sigma_j(t_i, x)|^2 \Delta}) \right] & \text{otherwise.} \end{cases} \end{aligned}$$

If the assets are not correlated, the individual trigger events on the interval $[t_i, t_{i+1}]$ are independent and thus

$$1 - p(x, y, \Delta, \mathcal{D}, \sigma(t_i, x)) = \prod_{j=1}^d (1 - p_{i,j}(x, y, \Delta)). \quad (2.9)$$

But in the case of correlated assets, there is no closed formula for p . In that case, SHEVCHENKO [2003] suggests to use the theory of copulas (JOE [1997]) in order to

³If there is only one barrier, the above series reduces to one term.

derive a lower and an upper bounds for p . We recall the Frechet bounds: for any events $(A_j)_{1 \leq j \leq d}$, one has

$$\left(\sum_{j=1}^d \mathbb{P}(A_j) - (d-1) \right)_+ \leq \mathbb{P} \left(\bigcap_{1 \leq j \leq d} A_j \right) \leq \min_{1 \leq j \leq d} \mathbb{P}(A_j). \quad (2.10)$$

In our case, it writes

$$\left(1 - \sum_{j=1}^d p_{i,j}(x, y, \Delta) \right)_+ \leq 1 - p(x, y, \Delta, \mathcal{D}, \sigma(t_i, x)) \leq 1 - \max_{1 \leq j \leq d} p_{i,j}(x, y, \Delta). \quad (2.11)$$

Denote by $1 - p_i^L(x, y, \Delta)$ and $1 - p_i^U(x, y, \Delta)$ the above lower and upper bounds. Then clearly, for positive payoff f , one has

$$f(X_T^N) \prod_{i=0}^{N-1} (1 - p_i^L(X_{t_i}^N, X_{t_{i+1}}^N, \Delta)) \quad (2.12)$$

$$\begin{aligned} &\leq f(X_T^N) \prod_{i=0}^{N-1} (1 - p(X_{t_i}^N, X_{t_{i+1}}^N, \Delta, \mathcal{D}, \sigma(t_i, x))) \\ &\leq f(X_T^N) \prod_{i=0}^{N-1} (1 - p_i^U(X_{t_i}^N, X_{t_{i+1}}^N, \Delta)), \end{aligned} \quad (2.13)$$

which gives two Monte Carlo estimates of $\mathbb{E}(\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T))$. These are lower and upper bounds if the $(X_{t_i})_i$ is simulated without error. In a study by SHEVCHENKO [2003], it is proposed to take the average as an estimator.

Numerical results are brought together with the methods of barrier shifting in Section 4.

3. Shifting the barrier

We have mentioned in the preceding section that for the evaluation of $\mathbb{E}(\Psi_T)$ with $\Psi_T = \mathbf{1}_{\forall t < T: X_t > D} (X_T - K)_+$ (DOC), the simulation of $\Psi_T^N = \mathbf{1}_{\forall t_i < T: X_{t_i} > D} (X_T - K)_+$ yields a systematic overestimation because by monitoring the process only at times $(t_i)_i$, we ignore its possible exit between these times. However, this positive bias shrinks to 0 as the number of monitoring dates N goes to infinity. One natural idea to compensate the bias is to shift the barrier D inside the activation zone of the option (i.e., one increases D to get D_N) and thus to compute

$$\Psi_T^{N, \text{shift}} = \mathbf{1}_{\forall t_i < T: X_{t_i} > D_N} (X_T - K)_+.$$

Of course, the new barrier D_N has to be tuned accurately in order to exactly remove the overestimation bias. This is not a trivial issue, but the way has been paved by BROADIE,

GLASSERMAN, and KOU [1999] for a single asset X in the Black–Scholes model with constant volatility σ . Namely, by setting

$$D_N = D \exp(0.5826\sigma\sqrt{\Delta}), \quad (3.1)$$

one gets

$$\mathbb{E}(\Psi_T^{N,\text{shift}}) - \mathbb{E}(\Psi_T) = o(\Delta^{1/2})$$

instead of $O(\Delta^{1/2})$ without shifting the barrier. The constant 0.5826... is defined later in (3.6).

Recently, GOBET and MENOZZI [2007] have established that there is a universal rule for shifting the barrier. We discuss this later. The purpose of this section is to provide several refinements about the idea of shifting the barrier, for the pricing of continuously monitored barrier option and the pricing of discrete barrier option, in the case of constant and nonconstant volatilities, including jumps or not. Numerical experiments are discussed in Section 4.

3.1. Understanding the influence of the monitoring frequency

Before coming into the details of barrier shifting, it is essential to understand well what the main bias term in $\Psi_T^N - \Psi_T$ is.

During the last decade, a lot of attention has been paid to the study of the associated convergence. Regarding the rate of convergence, it is proved quite generally that it is at most of order $\Delta^{1/2}$. Namely, for a smooth domain $\mathcal{D}$ and a general Itô process X , one has (see GOBET and MENOZZI [2007])

$$\mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) = O(\Delta^{1/2}).$$

In addition, for a positive payoff function f and for a solution X to a nondegenerate SDE, the above error is bounded from below by a positive constant times $\Delta^{1/2}$ (see GOBET and MENOZZI [2004]). These theoretical results are supported by numerical experiments, which illustrate well that a discrete sampling for such exotic contracts yields a positive bias, slowly decreasing. This is analogous when in addition the Euler scheme is used to approximate the underlying process X . Hence, in the following, we neglect this extra error whose contribution turns to be of order Δ (see Section 2.3).

The key point in the error analysis is to make the connection with the asymptotics of the overshoot Y^N . The overshoot is defined by the distance of the process to the barrier (the boundary of $\mathcal{D}$) at the discrete trigger $\tau^N = \inf\{t_i : X_{t_i} \notin \mathcal{D}\}$:

$$Y^N = d(X_{\tau^N}, \partial\mathcal{D}). \quad (3.2)$$

The important role of the overshoot has been underlined for years in the binomial tree methods. When the first node of the tree $(\log(X_0) + i\sigma\sqrt{\Delta})_i$ above the upper barrier $\log(U)$ coincides with the barrier, the overshoot of the binomial tree vanishes. In that case, one observes that the binomial tree method is quite accurate. Otherwise, the overshoot

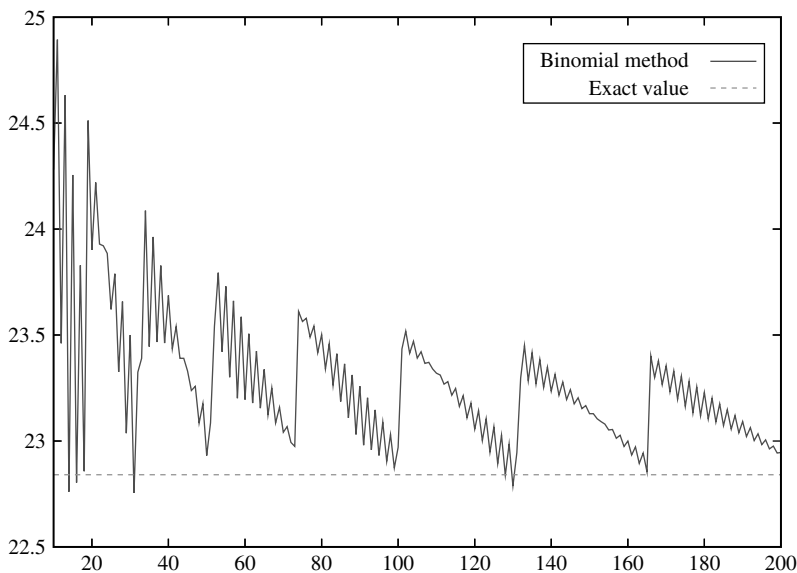


FIG. 3.1 Up and Out Put (nondiscounted) with strike 900 and upper barrier at 1150. Underlying asset: initial value = 1000, $T = 1$ year, interest rate = 0.05, and volatility = 0.2.

is equal to $\sigma\sqrt{\Delta} \left(1 - \text{Frac}\left(\frac{\log(U/x_0)}{\sigma\sqrt{\Delta}}\right)\right)$, where $\text{Frac}(x)$ denotes the fractional part of x (with the convention $\text{Frac}(x) = 1$ if $x \in \mathbb{N}$). Then, one knows that the numerical error is essentially proportional to this overshoot (see GOBET [1999] for a proof) and thus creates nasty oscillations as N varies (see Fig. 3.1).

In the general case, to decompose the error, we use the process $M_t = \mathbb{E}(\mathbf{1}_{T < \tau_t} f(X_T) | \mathcal{F}_t)$, where $\tau_t = \inf\{s \geq t : X_s \notin \mathcal{D}\}$ is the trigger time after t . Clearly, if no trigger has occurred before t , then M_t is the price at time t of the barrier option. We collect below the main properties related to M :

- (i) $X_t \notin \mathcal{D} \implies \tau_t = t$ and $M_t = 0$;
- (ii) for any given time t , $(M_{s \wedge \tau_t})_{t \leq s \leq T}$ is a martingale⁴;
- (iii) under nondegeneracy⁵ on X and mild smoothness assumptions on μ , σ and $\mathcal{D}$, one has

$$M_t = u(t, X_t),$$

where u is a smooth function on $[0, T] \times \mathcal{D}$, vanishing on $[0, T] \times \mathcal{D}^c$ and solving a parabolic partial differential equations (PDE) with Cauchy–Dirichlet boundary conditions: $u'_t + \sum_i \mu_i u'_{x_i} + \frac{1}{2} \sum_{i,j} [\sigma \sigma^*]_{i,j} u''_{x_i, x_j} = 0$ for $t < T, x \in \mathcal{D}$; $u(T, x) = f(x)$ for $x \in \mathcal{D}$; $u(t, x) = 0$ for $x \notin \mathcal{D}$ and $t < T$.

⁴Indeed, one easily checks that $M_{s \wedge \tau_t} = \mathbf{1}_{s \leq \tau_t} M_s = \mathbb{E}(\mathbf{1}_{T < \tau_t} f(X_T) | \mathcal{F}_s)$.

⁵Ellipticity or hypoellipticity condition plus noncharacteristic boundary condition.

Then, it follows

$$\begin{aligned}
& \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\
&= \mathbb{E}(M_{T \wedge \tau^N} - M_0) = \sum_{0 \leq i < N} \mathbb{E}(M_{t_{i+1} \wedge \tau^N} - M_{t_i \wedge \tau^N}) \text{ (properties i-iii)} \\
&= \sum_{0 \leq i < N} \mathbb{E} \mathbf{1}_{t_i < \tau^N} (M_{t_{i+1}} - M_{t_i}) \\
&= \sum_{0 \leq i < N} \mathbb{E} \mathbf{1}_{t_i < \tau^N} (M_{t_{i+1}} - M_{t_{i+1} \wedge \tau_{t_i}}) + \sum_{0 \leq i < N} \mathbb{E} \mathbf{1}_{t_i < \tau^N} (M_{t_{i+1} \wedge \tau_{t_i}} - M_{t_i}) \\
&= \sum_{0 \leq i < N} \mathbb{E} \mathbf{1}_{t_i < \tau^N} \mathbf{1}_{\tau_{t_i} < t_{i+1}} (M_{t_{i+1}} - M_{\tau_{t_i}}) \text{ (properties i-ii) and } \tau_{t_i} \geq t_i).
\end{aligned}$$

Because M vanishes when X is outside $\mathcal{D}$ and u is smooth inside $\mathcal{D}$, one has $M_{t_{i+1}} - M_{\tau_{t_i}} = \mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}} \nabla_x u(\tau_{t_i}, X_{\tau_{t_i}}) \cdot (X_{t_{i+1}} - X_{\tau_{t_i}})$ plus a term whose expectation is of order Δ and which can be neglected (see GOBET and MENOZZI [2007] for details). This gives

$$\begin{aligned}
& \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\
&= \sum_{0 \leq i < N} \mathbb{E} \left(\mathbf{1}_{t_i < \tau^N} \mathbf{1}_{\tau_{t_i} < t_{i+1}} \mathbb{E} \left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}} \nabla_x u(\tau_{t_i}, X_{\tau_{t_i}}) \cdot (X_{t_{i+1}} - X_{\tau_{t_i}}) | \mathcal{F}_{\tau_{t_i}} \right) \right) + O(\Delta).
\end{aligned}$$

Denote by $n(s)$ the inward normal unit vector at $s \in \partial\mathcal{D}$ and set $\partial_n u(t, s) = \nabla_x u(t, s) \cdot n(s)$. Since $u = 0$ on $\partial\mathcal{D}$, $\nabla_x u(t, s)$ lies only in the normal direction. In addition, with a probability exponentially close to 1 with respect to N , on $\mathcal{F}_{\tau_{t_i}}$ $X_{t_{i+1}}$ is close to the boundary and the normal component of $X_{t_{i+1}} - X_{\tau_{t_i}}$ has an amplitude equal to $d(X_{t_{i+1}}, \partial\mathcal{D})$. Bringing together all these remarks leads to

$$\begin{aligned}
& \mathbb{E}(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}} \nabla_x u(\tau_{t_i}, X_{\tau_{t_i}}) \cdot (X_{t_{i+1}} - X_{\tau_{t_i}}) | \mathcal{F}_{\tau_{t_i}}) \\
&= \partial_n u(\tau_{t_i}, X_{\tau_{t_i}}) \mathbb{E}(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}} d(X_{t_{i+1}}, \partial\mathcal{D}) | \mathcal{F}_{\tau_{t_i}}) + o(\Delta) \\
&= \partial_n u(\tau_{t_i}, X_{\tau_{t_i}}) \mathbb{E}(\mathbf{1}_{X_{t_{i+1}} \notin \mathcal{D}} d(X_{t_{i+1}}, \partial\mathcal{D}) | \mathcal{F}_{\tau_{t_i}}) + O(\Delta) \\
&= \partial_n u(\tau_{t_i}, X_{\tau_{t_i}}) \mathbb{E}(Y^N | \mathcal{F}_{\tau_{t_i}}) + O(\Delta).
\end{aligned}$$

Finally, we obtain the following result.

THEOREM 3.1. *Discrete monitoring and continuous monitoring yield a difference of prices that is proportional to the expected weighted overshoot. It writes*

$$\begin{aligned}
& \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\
&= \sum_{0 \leq i < N} \mathbb{E}(\mathbf{1}_{t_{i+1} = \tau^N} \partial_n u(\tau_{t_i}, X_{\tau_{t_i}}) Y^N) + O(\Delta). \tag{3.3}
\end{aligned}$$

Since the increments of X are of order $\Delta^{1/2}$, the same order applies to the overshoot Y^N , and this justifies why the discrete monitoring yields a simulation bias of this order. More interesting is to see that an expansion at the order $\Delta^{1/2}$ is available, provided that the limit of the triplet $(\tau^N, X_{\tau^N}, \Delta^{-1/2}Y^N)$ can be identified. When $(X_t = x_0 + W_t)_t$ is a linear Brownian motion and $\mathcal{D} =]-\infty, U[$, finding the asymptotics of the renormalized overshoot $\sqrt{\frac{N}{T}}Y^N$ can be made as follows. Set $(s_i = \sqrt{\frac{N}{T}}W_{iT/N})_i$, which defines a Gaussian random walk, and put $\tau^u = \inf\{i \geq 0 : s_i > u\}$. Clearly, one has

$$\sqrt{\frac{N}{T}}Y^N := \sqrt{\frac{N}{T}}(X_{\tau^N} - U) = [s_{\tau^u} - u]_{u=u_N} := y(u_N), \quad (3.4)$$

where $u_N = \sqrt{\frac{N}{T}}(U - x_0)$ goes to infinity with N . From the renewal theory (see SIEGMUND [1979]), $s_{\tau^u} - u = y(u)$ weakly converges as u goes to infinity to a random variable Y , whose cumulative function H is given by

$$H(y) := (\mathbb{E}[s_{\tau^0}])^{-1} \int_0^y \mathbb{P}[s_{\tau^0} > z] dz. \quad (3.5)$$

In our shifting approach, the quantity that plays a crucial role [remember (3.1)] is the expected asymptotic renormalized overshoot

$$\bar{y}(\infty) = \mathbb{E}(Y) = \lim_{u \rightarrow \infty} \mathbb{E}(y(u)) = \frac{\mathbb{E}[s_{\tau^0}^2]}{2\mathbb{E}[s_{\tau^0}]} = -\frac{\zeta(1/2)}{\sqrt{2\pi}} = 0.5826... \quad (3.6)$$

(see SIEGMUND [1979]). One also knows from CHANG and PERES [1997] that

$$\bar{y}(0) := \mathbb{E}(y(0)) = \mathbb{E}[s_{\tau^0}] = \frac{1}{\sqrt{2}} = 0.7071... \quad (3.7)$$

Unfortunately, the other values of $\bar{y}(u) = \mathbb{E}(y(u))$ are not known (see Appendix. A for a numerical approximation).

The previous results by Siegmund on the scalar Brownian motion have been recently extended to general diffusion processes (and to the associated Euler scheme).

THEOREM 3.2. (GOBET and MENOZZI [2007]). *The sequence $(\tau^N, X_{\tau^N}, \Delta^{-1/2}Y^N)_N$ weakly converges to $(\tau, X_\tau, |n \sigma(\tau, X_\tau)|Y)$, where Y is a random variable independent of (τ, X_τ) and whose cumulative function is equal to H (given in (3.5)).*

As a consequence and relying on (easy) extra uniform integrability results, one can pass to the limit in Theorem 3.1 to get

$$\begin{aligned} & \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\ &= \bar{y}(\infty) \sqrt{\frac{T}{N}} \mathbb{E}(\mathbf{1}_{\tau < T} \partial_n u(\tau, X_\tau) |n \sigma(\tau, X_\tau)|) + o(N^{-1/2}). \end{aligned} \quad (3.8)$$

As required, this is an overestimation (for positive payoff f) because the inward normal derivative of u is positive.

If we look carefully at the assumptions in the quoted references, the domain $\mathcal{D}$ needs to be a little smooth (of class C^2), and it does not allow us to directly apply the results to domains with corners. Actually, the limitation is essentially due to technical considerations related to good controls of the derivatives of u , solution of the PDE. In a study by MENOZZI [2006], these technicalities are handled in the case of a bidimensional drifted Brownian motion.

As a computational consequence of this result, the Romberg extrapolation techniques can be applied to get a more accurate procedure. Namely, by using two monitoring frequencies N/T and $2N/T$ and by averaging out independent simulations of

$$\Psi_T^{N,\text{Romberg}} = f(X_T) \left(\sqrt{2} \mathbf{1}_{\forall i \leq 2N: X_{iT/(2N)} \in \mathcal{D}} - \mathbf{1}_{\forall i \leq N: X_{iT/N} \in \mathcal{D}} \right) / (\sqrt{2} - 1), \quad (3.9)$$

we get

$$\mathbb{E} \Psi_T^{N,\text{Romberg}} - \mathbb{E} \Psi_T = o(N^{-1/2}). \quad (3.10)$$

Hence, at the first sight, we get a more accurate estimation of the price. However, there are two drawbacks. First, the computational time has essentially been multiplied by a factor 3 (because one simulates with N and $2N$ time steps). Second, the variance is approximately multiplied by a factor $((\sqrt{2})^2 + 1^2)/(\sqrt{2} - 1)^2 \approx 17.5$, which increases the statistical error of the Monte Carlo method.

3.2. Derivation of the sensitivity to the barriers

As mentioned in the introduction of this section, to compensate the systematic overestimation of discrete monitoring, we may think of slightly increasing the trigger zone by pushing the barrier toward the initial value of the underlying process X . The question is how much should we shift the barrier? The answer is strongly related to the sensitivity of the option price to the barrier.

We start with the simplest case of a drifted Brownian motion for $X_t = x + \mu t + \sigma W_t$ and an upper barrier U . In that case, due to explicit conditional trigger probability, one gets

$$\begin{aligned} & \partial_U \mathbb{E}(\mathbf{1}_{\forall t < T: X_t < U} f(X_T)) \\ &= \partial_U \int_{-\infty}^U f(y) \left(1 - \exp\left(-2 \frac{(U-x)(U-y)}{\sigma^2 T}\right) \right) \frac{1}{\sigma \sqrt{2\pi T}} \exp\left(-\frac{(y-x-\mu T)^2}{2\sigma^2 T}\right) dy \\ &= \int_{-\infty}^U f(y) \exp\left(-2 \frac{(U-x)(U-y)}{\sigma^2 T}\right) \frac{2(2U-x-y)}{\sigma^3 \sqrt{2\pi T^3}} \exp\left(-\frac{(y-x-\mu T)^2}{2\sigma^2 T}\right) dy \\ &= \int_{-\infty}^U f(y) \frac{2}{\sigma^2} \exp\left(\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2}{2\sigma^2} T\right) h^0(T, 2U-x-y) dy, \end{aligned}$$

where we set

$$h^\mu(t, b) = \frac{|b|}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(b - \mu t)^2}{2\sigma^2 t}\right) = h^0(t, b) \exp\left(\frac{\mu b}{\sigma^2} - \frac{\mu^2}{2\sigma^2} t\right).$$

The first hitting time τ for X of the level U has a density equal to $h^\mu(t, U - x)$ (see KARATZAS and SHREVE [1991]). The above relation with h^0 can be also obtained using the Girsanov theorem.

For the purpose of the barrier shifting, it is more convenient to relate $\partial_U \mathbb{E}(\mathbf{1}_{\forall t < T: X_t < U} f(X_T))$ to the gradient of the solution u of the PDE at the barrier. For this, use the convolution equality

$$h^0(T, 2U - x - y) = \int_0^T h^0(t, U - x) h^0(T - t, U - y) dt$$

($U > x, U > y$), which results from the definition of h and the independence of hitting times of $(x + \sigma W_t)_t$ of successive levels. Then,

$$\begin{aligned} \partial_U \mathbb{E}(\mathbf{1}_{\forall t < T: X_t < U} f(X_T)) &= \\ \int_0^T h^\mu(t, U - x) \int_{-\infty}^U f(y) \frac{2}{\sigma^2} \exp\left(\frac{\mu(y - U)}{\sigma^2} - \frac{\mu^2}{2\sigma^2}(T - t)\right) h^0(T - t, U - y) dy dt. \end{aligned}$$

In the same way, a direct computation shows that the above integral with respect to y is actually equal to $-\partial_x u(t, U)$, which finally gives

$$\partial_U \mathbb{E}(\mathbf{1}_{\forall t < T: X_t < U} f(X_T)) = -\mathbb{E}(\mathbf{1}_{\tau \leq T} \partial_x u(\tau, X_\tau)). \quad (3.11)$$

This formula is remarkably simple. It is not coinciding with the Brownian setting. It turns out that it can be extended to general diffusion models. More interesting is to notice that this gradient formula looks like the main term appearing in the discrete monitoring, this is discussed a bit later.

In the general case, the above explicit computations are no more explicit, and we have to carry out an alternative way to proceed. We give below another proof using basic stochastic calculus arguments, which has the great advantage to handle a large generality of trigger zones and stochastic processes. Before, we mention that related problems have been considered in the PDE literature. Indeed, the price function is a solution of Cauchy–Dirichlet PDE of the same type than the position of a elastic structure to which external forces are applied. Hence, computing the barrier sensitivity is very analogous to computing the shape sensitivity of elastic structures (see ALLAIRE [2002] and references therein about shape optimization). Sensitivities with respect to the domain are classic issues in the numerical analysis literature and date back to Hadamard in the beginning of the century. Recently in a study by COSTANTINI, GOBET, and EL KAROUI [2006], extensions to time-dependent domains and to general diffusion processes have been obtained by probabilistic techniques. The proof is rather elementary, and we give it in the previous case of an upper barrier U and linear diffusion process X . Denote by $\tau^\epsilon = \inf\{t \geq 0 : X_t \geq U - \epsilon\}$ the first hitting time of $U - \epsilon$ by X ($\epsilon > 0$). Clearly,

$(\tau^\epsilon)_\epsilon$ defines an increasing sequence as $\epsilon \downarrow 0$, bounded by τ , and it is not hard⁶ to show that it also converges to τ . We aim at computing the limit of $[\mathbb{E}(\mathbf{1}_{T < \tau} f(X_T)) - \mathbb{E}(\mathbf{1}_{T < \tau^\epsilon} f(X_T))]/\epsilon$ as $\epsilon \downarrow 0$, which gives the left derivative⁷ of $\mathbb{E}(\mathbf{1}_{\forall t < T: X_t < U} f(X_T))$ with respect to U .

Note that $\mathbb{E}(\mathbf{1}_{T < \tau} f(X_T)) = u(0, x) = \mathbb{E}(u(T \wedge \tau^\epsilon, X_{T \wedge \tau^\epsilon}))$ since $(u(s \wedge \tau, X_{s \wedge \tau}))_{s \leq T}$ is a martingale and $\tau^\epsilon < \tau$. In addition, $\mathbf{1}_{T < \tau^\epsilon} f(X_T) = u(T \wedge \tau^\epsilon, X_{T \wedge \tau^\epsilon}) - \mathbf{1}_{\tau^\epsilon \leq T} u(\tau^\epsilon, X_{\tau^\epsilon})$ because on $T < \tau^\epsilon$, $f(X_T) = u(T, X_T)$ ($X_T < U$). Writing $X_{\tau^\epsilon} = U - \epsilon$ and $u(\tau^\epsilon, U) = 0$, one finally gets

$$\frac{\mathbb{E}(\mathbf{1}_{T < \tau} f(X_T)) - \mathbb{E}(\mathbf{1}_{T < \tau^\epsilon} f(X_T))}{\epsilon} = \mathbb{E} \left(\mathbf{1}_{\tau^\epsilon \leq T} \frac{u(\tau^\epsilon, U - \epsilon) - u(\tau^\epsilon, U)}{\epsilon} \right).$$

Passing to the limit, we clearly obtain the equality (3.11), not only for a drifted Brownian motion but also for a general diffusion model for X .

We now state a more general result for a multidimensional domain $\mathcal{D} \subset \mathbb{R}^d$ and take, for instance, $\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)$ as a general form⁸ of the option payoff for an initial trigger zone $\mathcal{D}$: the price function $u(t, x)$ is solution of a parabolic PDE as mentioned before. We allow the domain $\mathcal{D}$ to be perturbed in various directions with various amplitudes (parameterized by a function Θ and a small parameter ϵ), which may result in a time-dependent domain in that way:

$$\mathcal{D}_t^\epsilon = \{x : x + \epsilon \Theta(t, x) \in \mathcal{D}\}.$$

We focus on the price for the option associated with the perturbed trigger zone: it is equal to $\mathbb{E}(\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}_t^\epsilon} f(X_T))$. It is proved⁹ (under suitable mild smoothness assumptions and ellipticity condition, see COSTANTINI, GOBET, and EL KAROUI [2006] for details) that the latter price as a function of ϵ is continuously differentiable, and its derivative at $\epsilon = 0$ is given by

$$\mathbb{E}[\mathbf{1}_{\tau < T} \partial_n u(\tau, X_\tau) [\Theta \cdot n](\tau, X_\tau)],$$

where $n(x)$ is the inward normal unit vector at $\partial \mathcal{D}$ at the point x and $\tau = \inf\{t > 0 : X_t \notin \mathcal{D}\}$ is the first exit time of $(X_s)_{s \geq 0}$ from $\mathcal{D}$. Equivalently, this differentiability result can be stated as follows.

THEOREM 3.3. *The price of knockedout option is differentiable with respect to trigger zone, and the sensitivity is defined by*

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\forall t < T: X_t + \epsilon \Theta(t, X_t) \in \mathcal{D}} f(X_T)) &= \mathbb{E}(\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\ &+ \epsilon \mathbb{E}[\mathbf{1}_{\tau < T} \partial_n u(\tau, X_\tau) [\Theta \cdot n](\tau, X_\tau)] + o(\epsilon). \end{aligned} \quad (3.12)$$

⁶One needs here to assume that the diffusion coefficient of X does not vanish at the barrier.

⁷The computations for the right derivative are analogous, and we skip them.

⁸Without extra difficulties, a rebate could be paid at the trigger time, and the domain may be time dependent as well (see COSTANTINI, GOBET, and EL KAROUI [2006] for details).

⁹The arguments are similar to those used before for the case of one-dimensional barrier U .

3.3. Connecting discrete/continuous monitoring: the Barrier Shifting Techniques (BAST)

We recall that from (3.8), one gets

$$\begin{aligned} & \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T) - \mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) \\ &= \bar{y}(\infty) \sqrt{\frac{T}{N}} \mathbb{E}(\mathbf{1}_{\tau < T} \partial_n u(\tau, X_\tau) |n \sigma(\tau, X_\tau)|) + o(N^{-1/2}). \end{aligned}$$

Thus, combining this with (3.12) for the choice $\theta(t, x) = \bar{y}(\infty) |n \sigma(t, x)| n(x)$ and $\epsilon = \sqrt{\frac{T}{N}}$, one obtains several simple relations (Theorems 3.4, 3.5, and 3.6) between prices of discrete and continuous monitored barrier options with adjusted trigger zones. In the following, we refer to these results as BAST (Barrier Shifting Techniques).

THEOREM 3.4 (From Discrete to Continuous (D2C)). *The price of a discrete barrier option equals that of a continuous barrier option with shifted barrier:*

$$\begin{aligned} & \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} \in \mathcal{D}} f(X_T)) - \mathbb{E}(\mathbf{1}_{\forall t < T: X_t + \bar{y}(\infty) |n \sigma(t, X_t)| \sqrt{\frac{T}{N}} \in \mathcal{D}} f(X_T)) \\ &= o(N^{-1/2}). \end{aligned} \tag{3.13}$$

In the above expression, $n(x)$ has to be understood as the inward normal to $\partial \mathcal{D}$ at the closest point to x on $\partial \mathcal{D}$ (i.e., its projection on the boundary).

The above Theorem 3.4 is an extension of results by BROADIE, GLASSERMAN, and KOU [1999] to multi-asset and multibarrier options. Such equality is useful to numerically evaluate the price of discrete barrier options only in situations where we are able to compute more efficiently the price of continuous ones.

- This is the case for linear arithmetic/geometric Brownian motion with one single/double barrier for which there is a closed formula, which is instantaneous to compute. This is the idea developed by BROADIE, GLASSERMAN, and KOU [1999]. For instance, for the Down and Out Call in a geometric Brownian motion model with volatility σ and with dividend rate q , one has

$$\begin{aligned} \text{DOC}(X_0, D) &= \mathbb{E}(e^{-rT} \mathbf{1}_{\forall t < T: X_t > D} (X_T - K)_+) \\ &= \text{Call}_{\text{BS}}(X_0) - \left(\frac{X_0}{D}\right)^{1-2\frac{(r-q)}{\sigma^2}} \text{Call}_{\text{BS}}\left(\frac{D^2}{X_0}\right), \end{aligned} \tag{3.14}$$

where $\text{Call}_{\text{BS}}(X_0)$ is the usual Black–Scholes price of the vanilla Call when the initial spot equals X_0 . In that case,

$$X_t + \bar{y}(\infty) |n \sigma(t, X_t)| \sqrt{\frac{T}{N}} \in \mathcal{D} \iff X_t + \bar{y}(\infty) \sigma X_t \sqrt{\frac{T}{N}} > D,$$

which is equivalent (at the order $o(N^{-1/2})$) to

$$X_t > D \exp\left(-\bar{y}(\infty)\sigma\sqrt{\frac{T}{N}}\right).$$

If we denote by $\text{DOC}_N(X_0, D)$ the price of the similar DOC with N monitoring dates, we get the approximation (see BROADIE, GLASSERMAN, and KOU [1999])

$$\text{DOC}_N(X_0, D) = \text{DOC}\left(X_0, D \exp\left(-\bar{y}(\infty)\sigma\sqrt{\frac{T}{N}}\right)\right) + o(N^{-1/2}). \quad (3.15)$$

In the next paragraph, we discuss this approximation, and we give a simple additional improvement when X_0 is close to the barrier or when N is small.

- This may also be applied if we can perform a efficient Monte Carlo method for the continuous barrier option, using, for instance, the Brownian bridge techniques (see Section 2). To illustrate this, consider again a DOC with a Merton's model¹⁰ of type (2.6). Then, the payoff (2.7) (written with a lower barrier D) has to be replaced with

$$\begin{aligned} (X_T - K)_+ & \left(1 - p\left(X_{\tau_k}, X_T, T - \tau_k, D \exp\left(-\bar{y}(\infty)\sigma\sqrt{\frac{T}{N}}\right), |\sigma|\right) \right) \\ & \times \prod_{i=0}^{k-1} \left(1 - p\left(X_{\tau_i}, X_{\tau_{i+1}^-}, \tau_{i+1} - \tau_i, D \exp(-\bar{y}(\infty)\sigma\sqrt{\frac{T}{N}}), |\sigma|\right) \right). \end{aligned} \quad (3.16)$$

In the same way, one can go from the discrete to continuous barrier pricing; we can approximate continuous barrier option using discrete ones with a Barrier shifting procedure.

THEOREM 3.5 (From Continuous to Discrete (C2D)). *The price of a continuous barrier option equals that of a discrete barrier option with shifted barrier:*

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\forall t < T: X_t \in \mathcal{D}} f(X_T)) & - \mathbb{E}(\mathbf{1}_{\forall t_i < T: X_{t_i} - \bar{y}(\infty)|n\sigma|n(t_i, X_{t_i})\sqrt{\frac{T}{N}} \in \mathcal{D}} f(X_T)) \\ & = o(N^{-1/2}). \end{aligned} \quad (3.17)$$

This procedure avoids using Brownian bridge techniques and related variants. This is particularly interesting for multiasset and multibarrier contracts where the computation of joint trigger probabilities is not possible (see the discussion in paragraph 2.3).

Also, for contracts with a large number of monitoring dates $N_1 \gg 1$ (daily for instance), we approximate them by the similar one with a medium number of monitoring

¹⁰In this jump diffusion model, the previous discussion is not supported by theoretical results. Nevertheless, the intuition is clear. Indeed, if the option is triggered when a jump occurs, the monitoring frequency has no influence in these aspects. Hence, the link between discrete and continuous barrier options really relies on the good understanding of the Brownian part, as we have done before.

dates $1 \ll N_2 \ll N_1$, by suitably adjusting the barrier twice. This saves computational time because it is essentially linear with the number of dates.

THEOREM 3.6 (From Discrete to Discrete (D2D)). *Two discrete barrier options with two different monitoring dates $(t_{1,i} = iT/N_1)_i$ and $(t_{2,i} = iT/N_2)_i$ are equal provided that barriers are conveniently shifted:*

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\forall t_{1,i} < T: X_{t_{1,i}} \in \mathcal{D}} f(X_T)) - \mathbb{E}(\mathbf{1}_{\forall t_{2,i} < T: X_{t_{2,i}} + \bar{y}(\infty) | n \sigma | n(t_{2,i}, X_{t_{2,i}})} \left(\sqrt{\frac{T}{N_1}} - \sqrt{\frac{T}{N_2}} \right) \in \mathcal{D}} f(X_T)) \\ = o(N_2^{-1/2}). \end{aligned}$$

(3.18)

3.4. An additional improvement: the adjusted BAST (ABAST)

Before presenting several numerical illustrations of BAST, we provide an additional improvement when the processes are close to the barrier or when the number of dates, N , are relatively small.

To understand the necessity of this improvement, we report in Table 3.1 few numerical values borrowed from BROADIE, GLASSERMAN, and KOU [1999]. It deals with the price of discrete DOC option using the closed formula for the continuous one in a Black–Scholes model [see Formula (3.15)]. One should carefully look at the accuracy of the approximation when the barrier D varies and the number, N , of dates as well. On the one hand, the approximation performs very well for large N or for barriers, D , far from the initial spot X_0 . On the other hand, in the other cases, it can be quite bad and may provide errors of order of 5% and even more.

By a careful look at the proof of the influence of the discrete monitoring (which is summed up in Equality (3.8)), one notices that the crucial quantity for the limit is

TABLE 3.1

D2C approximation for the discrete DOC in a Black–Scholes model with volatility $\sigma = 0.3$, interest rate $r = 0.1$, $K = X_0 = 100$, $T = 0.2$. $N = 50$, $N = 25$ or $N = 5$

N	Barrier D	True value	Continuous monitoring	Continuous monitoring with BAST (and relative error in %)
50	87	6.281	6.244	6.281 (0.0)
	91	5.977	5.808	5.977 (0.0)
	95	4.907	4.398	4.907 (0.0)
	99	2.337	1.171	2.271 (−2.8)
25	87	6.292	6.244	6.293 (0.0)
	91	6.032	5.808	6.033 (0.0)
	95	5.081	4.398	5.084 (0.0)
	99	2.813	1.171	2.673 (−5.0)
5	91	6.187	5.808	6.194 (0.1)
	95	5.671	4.398	5.646 (−0.5)
	97	5.167	3.060	5.028 (−2.7)
	99	4.489	1.171	4.053 (−9.8)

the expected overshoot

$$\mathbb{E}(Y^N) = \sqrt{\frac{T}{N}} \mathbb{E}(y(u_N)) = \sqrt{\frac{T}{N}} \bar{y}(u_N)$$

when the level (directly expressed in the logarithmic variables) is defined by $u_N = (\log(X_0) - \log(D))/(\sigma\sqrt{\frac{T}{N}})$. Since usually u_N is very large (since N is large and D is not close to X_0), we may focus only on the asymptotic value $\bar{y}(\infty) \approx 0.5826$ as in BROADIE, GLASSERMAN, and KOU [1997]. Indeed, in those cases, we have observed a very good performance of the BAST. However, for medium¹¹ values of u_N , one should replace $\bar{y}(\infty)$ with $\bar{y}(u_N)$. In the above case of a discrete DOC, this means that Equality (3.15) becomes

$$\text{DOC}_N(X_0, D) = \text{DOC}\left(X_0, D \exp\left[-\bar{y}(\log(X_0/D)/(\sigma\sqrt{\frac{T}{N}}))\sigma\sqrt{\frac{T}{N}}\right]\right) + o(N^{-1/2}). \quad (3.19)$$

This is what we call in the following Adjusted BAST (ABAST). The additional problem that one has to solve is the computation of the function $u \mapsto \bar{y}(u)$, which is not obvious because only the values in 0 and ∞ are known. A numerical study (reported in Appendix A) shows that one has the approximation

$$\bar{y}(u) \approx 0.5826 + 0.1245 \exp(-2.7u^{1.2}) \quad (3.20)$$

with less than 1% error. We present in Table 3.2 the same results as before, including now the ABAST. We also provide the value of u_N . In view of (3.20), there is no additional

TABLE 3.2
D2C approximation for the discrete DOC in a Black–Scholes model with volatility $\sigma = 0.3$, interest rate $r = 0.1$, $K = X_0 = 100$, $T = 0.2$. $N = 50$, $N = 25$ or $N = 5$

N	Barrier D	True value	BAST (and relative error in %)	u_N	ABAST (and relative error in %)
50	87	6.281	6.281 (0.0)	7.33	6.281 (0.0)
	91	5.977	5.977 (0.0)	4.97	5.977 (0.0)
	95	4.907	4.907 (0.0)	2.70	4.907 (0.0)
	99	2.337	2.271 (−2.8)	0.53	2.332 (−0.2)
25	87	6.292	6.293 (0.0)	5.18	6.281 (0.0)
	91	6.032	6.033 (0.0)	3.51	6.033 (0.0)
	95	5.081	5.084 (0.0)	1.91	5.0841 (0.0)
	99	2.813	2.673 (−5.0)	0.37	2.794 (−0.7)
5	91	6.187	6.194 (0.1)	1.57	6.194 (0.0)
	95	5.671	5.646 (−0.5)	0.85	5.663 (−0.1)
	97	5.167	5.028 (−2.7)	0.51	5.111 (−1.1)
	99	4.489	4.053 (−9.8)	0.17	4.353 (−3.1)

¹¹In other words, the convergence stated in (3.8) is not uniform in X_0 .

improvement in the case $u_N > 1$ (but in this case, the approximation is already very accurate), which is confirmed by the numerical experiments. In the converse case, the improvement is significant and leads to very good results when N is not small ($N = 25$ and $N = 50$); the accuracy is now robust in the distance of the spot from the barrier. The improvement is not so spectacular for $N = 5$, but this is not surprising because all the analyses related to barrier sensitivities and to the influence of the discrete monitoring are based on the asymptotics $T/N \rightarrow 0$.

4. Numerical tests

We conclude this work by presenting different illustrations of the Brownian bridge techniques and of ABAST on practical examples.

4.1. Merton's model

Consider first the case of discrete DOC (with N monitoring dates) for which we apply the D2C approximation described in (3.16) (with modifying $\bar{y}(\infty)$ as indicated before). The advantage is mainly computational because instead of simulating the underlying process at the N monitoring dates and at the N_J jump times,¹² the formula (3.16) requires the knowledge of the underlying process at the N_J jump times plus the maturity T . In the following example, we take $\lambda = 1$ and $T = 0.5$, which requires in average to simulate the process at $0.5 + 1$ dates instead of $N + 2$ dates. This can be much faster. To complete our description, we numerically observe that the confidence interval width is not significantly modified. We report numerical results in Table 4.1 when the barrier and N vary. It turns out that, up to the statistical error due to the Monte Carlo simulations, we get a very good accuracy for high-frequency monitoring (N large), exactly when such procedure may be interesting to save computational time. In Table 4.2, analogous results are given for a larger jump intensity ($\lambda = 4$). This is still satisfactory.

4.2. Up and out put in the Heston stochastic volatility model

In this one asset example, one considers the payoff of a continuous Up and Out Put (UOP) option

$$\mathbf{1}_{\forall t \leq T: X_t < U} (K - X_T)_+,$$

where the underlying asset (under the valuation probability) is defined by a Heston model

$$\begin{aligned} \frac{dX_t}{X_t} &= r dt + \sqrt{V_t} dW_t, \\ dV_t &= \beta(\alpha - V_t)dt + \sigma_V \sqrt{V_t} dB_t, \end{aligned}$$

¹² Which, in any case, is necessary for the simulation procedure.

TABLE 4.1

D2C approximation for the discrete DOC in a Merton model. Volatility $\sigma = 0.3$; lognormal jumps with mean $\mu_J = -0.02$, standard deviation $\sigma_J = 0.2$ and intensity $\lambda = 1$; interest rate $r = 0.05$, $K = X_0 = 100$, $T = 0.5$. $N = 25$ (weekly), or $N = 125$ (daily). 1 000 000 simulations. The true value is obtained using Monte Carlo simulations. Values in parentheses indicate the half-width of the 95% confidence interval

N	Barrier D	Discrete DOC (true value)	Continuous DOC with ABAST	Error (in percent)
25	89	9.65 (0.035)	9.68 (0.035)	0.3
	91	9.02 (0.035)	9.05 (0.034)	0.3
	93	8.16 (0.034)	8.20 (0.033)	0.5
	95	7.04 (0.033)	7.08 (0.032)	0.6
	97	5.69 (0.030)	5.67 (0.029)	-0.4
	99	4.21 (0.027)	3.93 (0.024)	-6.6
125	89	9.32 (0.035)	9.33 (0.035)	0.1
	91	8.55 (0.034)	8.56 (0.034)	0.1
	93	7.52 (0.033)	7.53 (0.033)	0.1
	95	6.22 (0.031)	6.22 (0.031)	0.0
	97	4.55 (0.028)	4.56 (0.027)	0.2
	99	2.60 (0.022)	2.55 (0.020)	-1.9

TABLE 4.2

Similar to Table (4.1) with jump intensity $\lambda = 4$

N	Barrier D	Discrete DOC (true value)	Continuous DOC with ABAST	Error (in %)
25	89	11.84 (0.050)	11.93 (0.050)	0.8
	91	10.99 (0.049)	11.10 (0.049)	1.0
	93	9.90 (0.048)	10.03 (0.047)	1.3
	95	8.52 (0.045)	8.68 (0.044)	1.9
	97	6.89 (0.041)	6.98 (0.040)	1.3
	99	5.05 (0.037)	4.89 (0.033)	-3.2
125	89	11.49 (0.050)	11.51 (0.050)	0.2
	91	10.50 (0.049)	10.52 (0.049)	0.2
	93	9.22 (0.047)	9.26 (0.046)	0.4
	95	7.62 (0.044)	7.65 (0.043)	0.4
	97	5.63 (0.039)	5.67 (0.038)	0.7
	99	3.27 (0.030)	3.22 (0.028)	-1.5

where B and W are two correlated Brownian motions (with correlation $\rho = 0.1$). For the next experiments, we take $X_0 = 40$, $K = 42$, and $r = 0.03$ for the interest rate, $T = 0.5$ for the expiration date, and the following values for the stochastic volatility: $\beta = 4$, $\alpha = 0.04 = V_0$, and $\sigma_V = 0.15$.

We apply here the C2D approximation given in Theorem 3.5. For the simulation of the volatility component, we use the symmetrized Euler scheme with step size h , which writes $V_{(i+1)h} = |V_{ih} + \beta(\alpha - V_{ih})h + \sigma_V \sqrt{V_{ih}}(B_{(i+1)h} - B_{ih})|$. Since we focus on the

monitoring error, we choose a small step size $h = 1/250$ (daily) to ensure that the related discretization error can be neglected. In the pseudocode below, we only give the details related to the trigger once the asset and its volatility are generated.

PSEUDOCODE

```
Delta=T/N;
// Brownian bridge techniques
prob=1;
for i=0 to N-1
    prob=prob*(1-p(X[i],X[i+1],Delta,U,sqrt(V[i])));
payoffBrownianBridge=prob*max(0,K-X[N]);
// ABAST
exitABAST=0;
cste_ABAST=y_bar(log(U/X[0])/sqrt(V[0]*Delta));
for i=0 to N-1
    U_ABAST=U*exp(-cste_ABAST*sqrt(V[i]*Delta));
    if (X[i+1]>U_ABAST) exitABAST=1;
if (exitABAST==1) payoffABAST=0;
else payoffABAST=max(0,K-X[N]);
```

In Table 4.3, we observe that the Brownian bridge techniques still yield very accurate results with respect to the number of monitoring dates N . Indeed, for each given barrier U , the output values are almost constant with respect to N . As usual, the discrete monitoring procedure provides a price overestimation, which becomes larger and larger as the trigger probability (U closer to the initial spot). Regarding ABAST, the accuracy is very good, except for $U = 41$, where the asymptotics is recovered only for large values of N . In this example of one single barrier on one single asset, the procedure using Brownian bridges is more efficient.

4.3. Down out call on three assets

We borrow this example from SHEVCHENKO [2003]. The buyer of the contract receives at T the payoff $(X_{1,T} - K)_+$, that is, a Call on the first asset, with knocked-out lower barriers on the three assets. The assets are modelled by correlated geometric Brownian motions with constant coefficients. Our Table 4.4 (which completes table 4 in SHEVCHENKO [2003]) compares 4 estimators: 1) using the simple discrete monitoring; 2) and 3) using the lower and upper copula bounds to compute the conditional trigger probability (see inequalities (2.12 and 2.13)); 4) using ABAST. In the latter case, according to Theorem 3.5, each barrier is shifted separately: the lower barrier for the asset X_k becomes $D_k \exp(\bar{y}(u_{k,N})\sigma_k\sqrt{\frac{T}{N}})$, where $u_{k,N} = \log(X_{k,0}/D_k)/(\sigma_k\sqrt{\frac{T}{N}})$. Contrary to the Brownian bridge techniques, we do not need to take into account the correlation between assets to adjust the barriers. This theoretical choice is confirmed by numerical experiments in Table 4.4 and 4.5.

It is observed by SHEVCHENKO [2003] that the spread between the prices given by upper and lower copula bounds shrinks to 0 as N goes to infinity. In Table 4.4, indeed, it is small, and it is of order of the statistical error. The ABAST behaves also very well.

TABLE 4.3

Continuous UOP in a Heston model with parameters $X_0 = 40, K = 42, r = 0.03, T = 0.5, \beta = 4, \alpha = 0.04 = V_0, \sigma_V = 0.15, \rho = 0.1$. Upper barriers from $U = 41$ to $U = 46$. Comparison of discrete monitoring, Brownian bridge techniques, ABAST (C2D approximation), as N increases: $N = 13, N = 26$ (weekly), $N = 52, N = 126$ (daily). 10 000 000 simulations: the 95% statistical errors are smaller than 0.002 and are not reported

Barrier U	Discrete UOP / Brownian bridge/ ABAST ($N = 13$)	Discrete UOP / Brownian bridge / ABAST ($N = 26$)	Discrete UOP / Brownian bridge / ABAST ($N = 52$)	Discrete UOP/ / Brownian bridge / ABAST ($N = 126$)
41	1.729	1.516	1.364	1.233
	0.984	0.984	0.983	0.983
	1.133	1.040	0.995	0.982
42	2.244	2.110	2.008	1.913
	1.730	1.728	1.727	1.727
	1.752	1.726	1.725	1.727
43	2.595	2.510	2.442	2.378
	2.253	2.251	2.250	2.250
	2.251	2.250	2.250	2.250
44	2.808	2.756	2.714	2.674
	2.594	2.593	2.592	2.592
	2.591	2.592	2.592	2.592
45	2.926	2.897	2.873	2.850
	2.802	2.801	2.801	2.801
	2.800	2.801	2.801	2.801
46	2.988	2.973	2.960	2.948
	2.920	2.920	2.920	2.920
	2.920	2.920	2.919	2.920

TABLE 4.4

DOC on three assets. $\sigma_1 = \sigma_2 = \sigma_3 = 0.4, \rho_{i,j} = \rho = 0.5$ for $i \neq j, X_{1,0} = X_{2,0} = X_{3,0} = 100, K = 100, r = 0.05$, and $T = 1$. Barrier down: $D_1 = D_2 = D_3 = 80$. Comparison of discrete monitoring, Brownian bridge techniques with lower and upper copula bounds, ABAST (C2D approximation) as N increases. 4 000 000 simulations. Values in parentheses indicate the half-width of the 95% confidence interval

N	Discrete monitoring	Copula lower bound	Copula upper bound	ABAST
8	10.68 (0.03)	7.48 (0.02)	7.66 (0.02)	7.53 (0.02)
16	9.85 (0.03)	7.53 (0.02)	7.59 (0.02)	7.52 (0.02)
32	9.22 (0.03)	7.55 (0.02)	7.57 (0.02)	7.54 (0.02)
64	8.74 (0.03)	7.54 (0.02)	7.55 (0.02)	7.53 (0.02)
128	8.41 (0.03)	7.55 (0.02)	7.55 (0.02)	7.55 (0.02)

TABLE 4.5

DOC on three assets. Parameters analogous to Table 4.4 but for $D_1 = D_2 = D_3 = 90$. Impact of the correlation ρ on the accuracy of the numerical methods. Values in parentheses indicate the half-width of the 95% confidence interval

ρ	N	Discrete monitoring	Copula lower bound	Copula upper bound	ABAST
0	16	1.124 (0.009)	0.347 (0.005)	0.383 (0.005)	0.385 (0.005)
	32	0.855 (0.008)	0.352 (0.005)	0.362 (0.005)	0.357 (0.005)
	64	0.688 (0.007)	0.359 (0.005)	0.361 (0.005)	0.357 (0.005)
	128	0.579 (0.007)	0.359 (0.005)	0.360 (0.005)	0.359 (0.005)
0.25	16	2.73 (0.02)	1.12 (0.01)	1.22 (0.01)	1.24 (0.01)
	32	2.21 (0.01)	1.15 (0.01)	1.18 (0.01)	1.17 (0.01)
	64	1.87 (0.01)	1.15 (0.01)	1.16 (0.01)	1.15 (0.01)
	128	1.66 (0.01)	1.16 (0.01)	1.16 (0.01)	1.16 (0.01)
0.5	16	4.98 (0.02)	2.45 (0.01)	2.67 (0.01)	2.72 (0.02)
	32	4.22 (0.02)	2.52 (0.01)	2.59 (0.01)	2.57 (0.02)
	64	3.71 (0.02)	2.54 (0.01)	2.56 (0.01)	2.54 (0.02)
	128	3.39 (0.02)	2.56 (0.01)	2.56 (0.01)	2.56 (0.02)
0.75	16	7.93 (0.02)	4.45 (0.02)	4.89 (0.02)	5.01 (0.02)
	32	6.99 (0.02)	4.64 (0.02)	4.81 (0.02)	4.79 (0.02)
	64	6.33 (0.02)	4.69 (0.02)	4.76 (0.02)	4.73 (0.02)
	128	5.89 (0.02)	4.72 (0.02)	4.75 (0.02)	4.74 (0.02)
1	16	13.02 (0.03)	7.70 (0.02)	9.71 (0.03)	10.02 (0.03)
	32	12.15 (0.03)	8.39 (0.02)	9.70 (0.03)	9.74 (0.03)
	64	11.49 (0.03)	8.78 (0.02)	9.69 (0.03)	9.68 (0.03)
	128	11.02 (0.03)	9.06 (0.03)	9.71 (0.03)	9.71 (0.03)

In the next experiments listed in Table 4.5, we take the barriers closer to the initial spot. In that case, the upper and lower copula bounds have a more significant impact on the price spread. Note that the larger the correlation is, the larger the spread is. This would be even worse if we took $D_k \approx X_{k,0}$. The accuracy of ABAST does not seem to be affected by the correlation changes.

4.4. Down out call on a basket

In this paragraph, we consider a DOC with two barriers, but the two barriers are applied to a basket of six assets. The payoff is given by $\mathbf{1}_{\forall t < T: B_{1,t} > D_1, B_{2,t} > D_2} (B_{1,T} - K)_+$, where the two baskets are given by

$$B_{1,t} = 0.5X_{1,t} + 0.3X_{2,t} + 0.2X_{3,t} \text{ and } B_{2,t} = (X_{4,t} + X_{5,t} + X_{6,t})/3.$$

In the following, the six assets X_k are modeled by correlated geometric Brownian motions, with constant correlation $\rho = 0.4$ and constant volatilities $\sigma_k = 0.3$. Their initial values are all equal to 100. The option expiration is $T = 1$ year, the interest rate

$r = 0.05$. The barriers $D_1 = D_2 = D$ are taken as 85, 90, and 95. During the ABAST simulation procedure, these barriers are shifted to $D_k \exp(\bar{y}(u_{k,N})\sigma_{k,t_i}\sqrt{\frac{T}{N}})$, where we should take for σ_{k,t_i} the volatility of the basket B_k at time t_i . Since the basket does not follow a geometric Brownian motion dynamics, the determination of σ_{k,t_i} has to be made carefully. For this, consider a general Basket on different assets X_i with weights p_i :

$$B_t = \sum_i p_i X_{i,t}. \quad (4.1)$$

We define $\sigma_{B,t}$, its volatility at time t , by

$$\frac{dB_t}{B_t} = \dots dt + \sigma_{B,t} dW_t,$$

where W is a Brownian motion. If X_i has a dynamics of the form $\frac{dX_{i,t}}{X_{i,t}} = \dots dt + \sigma_{i,t}^X dW_{i,t}$, where the Brownian motions $(W_i)_i$ are correlated, then by equating the quadratic variations of both sides of (4.1), we easily get

$$[B_t \sigma_{B,t}]^2 = \sum_{i,j} p_i p_j S_{i,t} S_{j,t} \sigma_{i,t}^X \sigma_{j,t}^X \rho_{i,j}.$$

This gives the value σ_{B,t_i} , which should be used to shift the barrier at time t_i . In Table 4.6, we present numerical results for different values of the barriers. The aim of these tests is to check the accuracy of ABAST. Indeed, the usual discrete monitoring procedure still yields a slow convergence with respect to N , while for ABAST, it is much quicker. For a weekly monitoring (roughly $N = 50$), this approach gives a bias smaller than the statistical error.

4.5. BLAC down out option

This example is aimed at illustrating the D2D approximation from Theorem 3.6. The Basket Lock Active Coupon (BLAC) Down Out option pays 1 Euro if at most one of five underlying assets has touched a lower barrier L before the expiration. This is a bit

TABLE 4.6
DOC on a basket with six assets and two barriers. Accuracy of ABAST as N increases.
1 000 000 simulations. For all the results, the statistical errors are smaller than 0.033

N	Discrete monitoring	ABAST	Discrete monitoring	ABAST	Discrete monitoring	ABAST
	$D = 85$	$D = 85$	$D = 90$	$D = 90$	$D = 95$	$D = 95$
12	10.15	9.03	8.36	6.51	5.73	3.52
24	9.90	9.06	7.88	6.49	4.95	3.20
50	9.70	9.09	7.53	6.53	4.39	3.10
100	9.51	9.07	7.22	6.50	4.00	3.06
250	9.35	9.06	6.97	6.51	3.66	3.06

TABLE 4.7

BLAC Down Out option on 5 assets, with daily monitoring $N_1 = 250$. D2D ABAST approximation with weekly monitoring $N_2 = 52$. 1 000 000 simulations. Values in parentheses indicate the half-width of the 95% confidence interval

Barrier	True price	ABAST $N_2 = 52$	ABAST $N_2 = 26$	ABAST $N_2 = 12$
70	0.519 (0.001)	0.519 (0.001)	0.517 (0.001)	0.512 (0.001)
75	0.387 (0.001)	0.386 (0.001)	0.385 (0.001)	0.381 (0.001)
80	0.2632 (0.0009)	0.2628 (0.0009)	0.2617 (0.0009)	0.2583 (0.0009)
85	0.1578 (0.0007)	0.1578 (0.0007)	0.1559 (0.0007)	0.1551 (0.0007)
90	0.0773 (0.0005)	0.0773 (0.0005)	0.0764 (0.0005)	0.0790 (0.0005)

more sophisticated than a simple barrier option. Usually, the assets are daily monitored to determine whether the barrier has been touched and if so by how many assets. For contract expirations larger than 1 year, it may be computational demanding for the pricing, and one may be interested in speeding it up by monitoring the assets only weekly or monthly and by adjusting the barrier L (see Theorem 3.6 with $N_1 = 250$ and $N_2 = 52$ or $N_2 = 12$). Note that the barrier may be shifted differently for each asset because their volatilities may differ. In our experiments, the five assets $(X_k)_{k \leq 5}$ are modeled by correlated geometric Brownian motions, with constant correlation $\rho = 0.5$ and constant volatilities $\sigma_1 = 0.3$, $\sigma_2 = 0.32$, $\sigma_3 = 0.35$, $\sigma_4 = 0.38$, and $\sigma_5 = 0.4$. Their initial values are all equal to 100, and the interest rate equals 0. The option expiration is $T = 1$ year. We report in Table 4.7 first the prices computed by Monte Carlo simulations of the assets at N_1 dates and second those with $N_2 = 52, 26, 12$ dates but with shifted barriers. The accuracy is very good, and the computational time has been approximately divided by 5, 10, or 20.

References

- ALLAIRE, G. (2002). *Shape Optimization by the Homogenization Method*, first ed. (Springer Verlag, New York, NY).
- BALDI, P. (1995). Exact asymptotics for the probability of exit from a domain and applications to simulation. *Ann. Probab.* **23** (4), 1644–1670.
- BALLY, V., TALAY, D. (1996). The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Probab. Theory. Rel.* **104-1**, 43–60.
- BOYLE, P., LAU, S. (1994). Bumping up against the barrier with the binomial method. *J. Derivatives* **1**, 6–14.
- BOYLE, P., TIAN, Y. (1998). An explicit finite difference approach to the pricing of barrier options. *Appl. Math. Financ.* **5**, 17–43.
- BROADIE, M., GLASSERMAN, P., KOU, S. (1997). A continuity correction for discrete barrier options. *Math. Financ.* **7**, 325–349.
- BROADIE, M., GLASSERMAN, P., KOU, S. (1999). Connecting discrete and continuous path-dependent options. *Financ. Stoch.* **3**, 55–82.
- CHANG, J., PERES, Y. (1997). Ladder heights, Gaussian random walks and the Riemann zeta function. *Ann. Probab.* **25** (2), 787–802.
- CHEUK, T., VORST, T. (1996). Complex barrier options. *J. Derivatives* **4**, 8–22.
- COSTANTINI, C., GOBET, E., EL KAROUI, N.E. (2006). Boundary sensitivities for diffusion processes in time dependent domains. *Appl. Math. Opt.* **54** (2), 159–187.
- DERMAN, E., KANI, I., ERGENER, D., BARDHAN, I. (1995). Enhanced numerical methods for options with barriers. *Financ. Analysts J.* **51** (6), 65–74.
- GOBET, E. (1999). Analysis of the zigzag convergence for barrier options with binomial trees. Technical Report, (Prépublication 536 du laboratoire PMA Paris 6, France) (Available at: <http://www.proba.jussieu.fr/mathdoc/preprints/>).
- GOBET, E. (2000). Euler schemes for the weak approximation of killed diffusion. *Stoch. Proc. Appl.* **87**, 167–197.
- GOBET, E. (2001). Euler schemes and half-space approximation for the simulation of diffusions in a domain. *ESAIM Probabil. Stat.* **5**, 261–297.
- GOBET, E. (2004). Revisiting the Greeks for European and American options. In: Akahori, J., Ogawa, S., Watanabe, S. (eds.), *Stochastic Processes and Applications to Mathematical Finance* (World Scientific), pp. 53–71.
- GOBET, E., MENOZZI, S. (2004). Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme. *Stoch. Proc. Appl.* **112** (2), 201–223.
- GOBET, E., MENOZZI, S. (2007a). *Discrete Sampling of Functionals of Itô Processes. Séminaire de Probabilités XL - Lecture Notes in Mathematics 1899* (Springer Verlag), 355–374.
- GOBET, E., MENOZZI, S. (2007b). Stopped diffusion processes: Overshoots and boundary correction. Technical Report. (Preprint available on Arxiv <http://fr.arxiv.org/abs/0706.4042>).
- JOE, H. (1997). *Multivariate Models and Dependence Concepts* (Chapman and Hall).
- KARATZAS, I., SHREVE, S. (1991). *Brownian Motion And Stochastic Calculus*, Second ed. (Springer Verlag).
- MENOZZI, S. (2006). Improved simulation for the killed Brownian Motion in a cone. *SIAM J. Numer. Anal.* **44** (6), 2610–2632.
- MERTON, R. (1973). Theory of rational option pricing. *Bell J. Econ. and Manage. Sci.* **4**, 141–183.
- REVUZ, D., YOR, M. (1994). *Continuous Martingales and Brownian Motion*, second ed. In: *Grundlehren der Mathematischen Wissenschaften*. (Springer, Berlin, Germany).

- RITCHKEN, P. (1995). On pricing barrier options. *J. Derivatives* **3**, 19–28.
- ROGERS, L., STAPLETON, E. (1998). Fast accurate binomial pricing. *Financ. Stoch.* **2**, 3–17.
- RUBINSTEIN, M., REINER, E. (1991). Breaking down the barriers. *Risk* **4** (8), 28–35.
- SHEVCHENKO, P. (2003). Addressing the bias in Monte Carlo pricing of multi-asset options with multiple barriers through discrete sampling. *J. Comput. Financ.* **6** (3), 1–20.
- SIEGMUND, D. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. Appl. Probab.* **11** (4), 701–719.

Appendix A. Numerical approximation of the expected overshoot of a Gaussian random walk

The necessity of well approximating the expected overshoot $\bar{y}(u) = \mathbb{E}(s_{\tau^u} - u)$ for arbitrary levels u is explained in Paragraph 3.4. It aims at reducing the simulation bias when the initial spot is close to the barriers.

From the numerical point of view, the computation of $\bar{y}(u)$ using a Monte Carlo method is not straightforward. The objective of this section is to provide a few related facts. On the one hand, it seems that one has only to simulate a Gaussian random walk $(s_i)_{i \geq 0}$, until the hitting time τ^u of the level u and then compute the overshoot $s_{\tau^u} - u$. On the other hand, although τ^u is finite with probability 1, it has an infinite mean. This implies that our simulation procedure with M independent paths will finish after some time, but it may take a huge computational time!!

To practically compute the expected overshoot, one should stop a simulation when it is too long, say when $\tau^u > t$ for an appropriate choice of a large t to ensure a desired accuracy. For this, put

$$\bar{y}(u, t) := \mathbb{E}((s_{\tau^u \wedge t} - u)_+). \quad (\text{A.1})$$

This is the quantity that we can compute by the simulation of the Gaussian random walk until time t (or less). Since $s_i < u$ for $i < \tau^u$, one also has

$$\bar{y}(u, t) = \mathbb{E}(\mathbf{1}_{\tau^u \leq t}(s_{\tau^u} - u)). \quad (\text{A.2})$$

Thus, $\bar{y}(u, t)$ is increasing with respect to t , and by monotone convergence theorem, its limit as $t \rightarrow \infty$ is $\bar{y}(u)$. To estimate the approximation error, it is useful to derive upper bounds related to the overshoot.

LEMMA A.1. *For any $p > 0$ and for any $u > 0$, one has*

$$\mathbb{E}(s_{\tau^u} - u)^p \leq \mathbb{E}|s_1|^p. \quad (\text{A.3})$$

PROOF. We imbed the Gaussian random walk into a standard Brownian motion W : $s_i = W_i$. Put $\tau_i = \inf\{s > i : W_s = u\}$. Then, writing

$$s_{\tau^u} - u = \sum_{i \geq 1} \mathbf{1}_{\tau^u > i-1} (s_i - u)_+ = \sum_{i \geq 1} \mathbf{1}_{\tau^u > i-1} \mathbf{1}_{\tau_{i-1} < i} (W_i - W_{\tau_i})_+$$

and using the strong Markov property, the scaling invariance and the symmetry property of W , it follows that

$$\mathbb{E}(s_{\tau^u} - u)^p = \sum_{i \geq 1} \mathbb{E}(\mathbf{1}_{\tau^u > i-1} \mathbf{1}_{\tau_{i-1} < i} (W_i - W_{\tau_i})_+^p)$$

$$\begin{aligned}
&\leq \sum_{i \geq 1} \mathbb{E}(\mathbf{1}_{\tau^u > i-1} \mathbf{1}_{\tau_{i-1} < i}) \mathbb{E}([s_1]_+^p) \\
&= 2\mathbb{E}([s_1]_+^p) \sum_{i \geq 1} \mathbb{E}(\mathbf{1}_{\tau^u > i-1} \mathbf{1}_{s_i > u}) = 2\mathbb{E}([s_1]_+^p) = \mathbb{E}(|s_1|^p).
\end{aligned}$$

□

For $p = 1$, by a direct computation of $\mathbb{E}|s_1|$, we get the uniform upper bound

$$\bar{y}(u) \leq 2/\sqrt{2\pi} = 0.7979.$$

We recall from CHANG and PERES [1997] that $\bar{y}(0) = 1/\sqrt{2} = 0.7071$ and $\bar{y}(\infty) = 0.5826$, which is not far from the upper bound.

To discuss the accuracy of $\bar{y}(u, t)$ for large t , we need to upper bound

$$\begin{aligned}
\mathbb{P}(\tau^u > t) &= \mathbb{P}(\max_{i \leq t} W_i < u) \\
&\leq \mathbb{P}(\max_{s \leq t} W_s < u + \sqrt{3 \log t}) + \mathbb{P}(\max_{i \leq t} W_i < u; \max_{s \leq t} W_s \geq u + \sqrt{3 \log t}).
\end{aligned}$$

The term $\sqrt{3 \log t}$ is chosen in order to approximately equate each contribution. The first one is equal to

$$\mathbb{P}(|W_t| < u + \sqrt{3 \log t}) \leq \frac{2}{\sqrt{2\pi t}}(u + \sqrt{3 \log t}).$$

The second contribution is upper bounded by

$$\begin{aligned}
\mathbb{P}(\exists i \leq t-1 : \sup_{s \in [i, i+1]} (W_s - W_i) > \sqrt{3 \log t}) &\leq t\mathbb{P}(|W_1| > \sqrt{3 \log t}) \\
&\leq 2t \exp(-(\sqrt{3 \log t})^2/2) = \frac{2}{\sqrt{t}}.
\end{aligned}$$

By the Holder inequality, we deduce that for any $p \geq 1$ and $u \leq 4$

$$\begin{aligned}
0 \leq \bar{y}(u) - \bar{y}(u, t) &= \mathbb{E}((s_{\tau^u} - u)\mathbf{1}_{\tau^u > t}) \leq \|s_{\tau^u} - u\|_{2p}(\mathbb{P}(\tau^u > t))^{(2p-1)/(2p)} \\
&\leq \left(\frac{(2p!)}{2^p p!}\right)^{1/(2p)} \left(\frac{2}{\sqrt{2\pi t}}(4 + \sqrt{2\pi} + \sqrt{3 \log t})\right)^{(2p-1)/(2p)}.
\end{aligned}$$

For instance, an accuracy of 0.0025 is achieved for $t = 3.15 \times 10^8$ (take $p = 7$ in the upper bound). To tune the number of simulations M in the evaluation of $\bar{y}(u, t)$, we note that the variance of $(s_{\tau^u} - u)\mathbf{1}_{\tau^u \leq t}$ is bounded by $\mathbb{E}(s_1^2) = 1$ (owing to Lemma A.1). Thus, $M = 1.96^2/0.0025^2 \approx 615000$ yields a statistical error of order 0.0025 with probability 95% and thus an overall error of 0.005 (i.e., a posteriori less than 1% error).

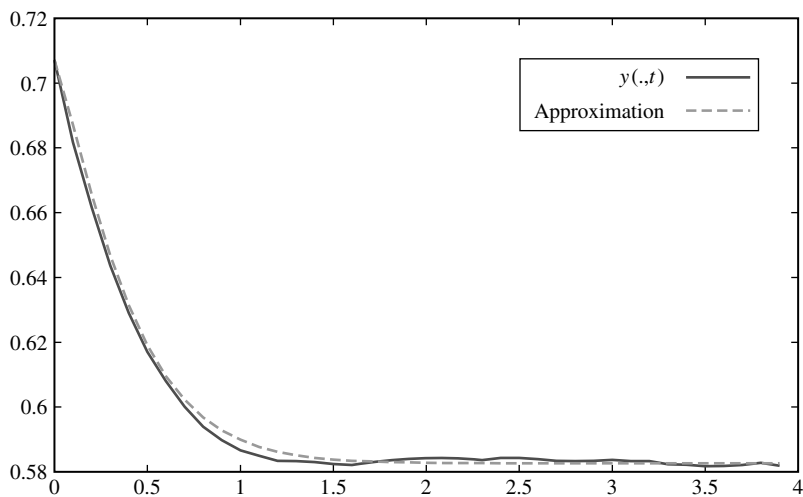


FIG. A.1 The numerical values of $\bar{y}(\cdot, t)$ and its approximation (A.4).

The estimation of $\bar{y}(\cdot, t)$ that we obtain is plotted in Fig. A.1. A good approximation may be achieved using an exponential type function of the form

$$\bar{y}(u) \approx \bar{y}(\infty) + (\bar{y}(0) - \bar{y}(\infty)) \exp(-2.7u^{1.2}) \quad (\text{A.4})$$

with $\bar{y}(0) = 0.7071$ and $\bar{y}(\infty) = 0.5826$. It achieves an accuracy at least of 1%.

Real Options

Alain Bensoussan

*International Center for Decision and Risk Analysis, ICDRiA,
School of Management, SM30, University of Texas at Dallas,
800 W. Campbell Rd, Richardson, TX 75080-0688, USA
E-mail Addresses: Alain.Bensoussan@utdallas.edu*

Abstract

We present here the theory of real options using extensively the techniques of variational inequalities. This powerful technique is the right tool to capture the various possibilities that express the flexibility, the main background of real options.

1. Introduction

A key reference for the theory of real options is the study by DIXIT and PINDYCK [1994], although the title of the book is different. Variational inequalities developed in a different context were initially used for stochastic control in the study by BENSOUSSAN and LIONS [1982]. We show here that this technique is the appropriate one to handle the flexibility, which is the hallmark of real options. Real options theory is an approach to mitigate risks of investment projects that stems from two ideas. The first idea is hedging, borrowed from financial options, when market considerations can be introduced. We refer to situations in which the project risk is correlated to the market risk, and one can use tradable assets to hedge. The second idea is flexibility. The option concept pertains to the range of decisions. In particular, one may scale down or up the project, one may stop it, or one may change orientation. This flexibility allows to react properly when information is obtained on the uncertainties of the evolution. We shall first review what can be obtained from financial theory and adapted to investment decisions in the continuous time and in the discrete time. We will then treat the flexibility in the decision making.

2. Tradable assets

2.1. Complete market model

We assume continuous time. The randomness is characterized by n standard independent Wiener processes $w_j(t)$. We denote

$$\mathcal{F}^t = \sigma(w_j(s), j = 1, \dots, n; s \leq t).$$

There are n basic assets on the market whose prices are denoted by $Y_i(t)$ whose evolution is governed by

$$dY_i(t) = Y_i(t)(\alpha_i(t)dt + \sigma_{ij}(t)dw_j),$$

where $\alpha_i(t)$ and $\sigma_{ij}(t)$ are processes adapted to $\mathcal{F}^t$. The market is complete when the matrix $\sigma(t)$ is invertible. In this case, the information obtained by observing the evolution of the prices of assets is sufficient to recover the underlying source of noise modeled by the Wiener processes. In addition to the random assets, there is a riskless asset whose evolution is characterized by

$$Y_0(t) = \exp rt.$$

Denoting by $\alpha(t)$ the vector with components $\alpha_i(t)$, we consider the process

$$\theta(t) = \sigma^{-1}(t)(\alpha(t) - r\mathbb{I}),$$

whose definition makes direct use of the invertibility of the matrix $\sigma(t)$. We next define $Z(t)$ by the relation

$$dZ(t) = -Z(t)\theta(t) \cdot dw(t), \quad Z(0) = 1.$$

A key property is the following

PROPOSITION 2.1. *The processes*

$$Z(t)Y_i(t) \exp -rt \text{ is a } \mathcal{F}^t \text{ martingale.}$$

From the martingale property, we can write

$$E[Z(T)Y_i(T) \exp -rT | \mathcal{F}^t] = Z(t)Y_i(t) \exp -rt, \quad \forall T > t$$

hence,

$$Y_i(t) = E\left(\frac{Z(T)}{Z(t)} Y_i(T) \exp -r(T-t) | \mathcal{F}^t\right),$$

which can also be written as follows

$$Y_i(t) = E[Y_i(T) | \mathcal{F}^t] \exp -r(T-t) + \text{cov}\left(\frac{Y_i(T)}{\exp r(T-t)}, \frac{Z(T)}{Z(t)} | \mathcal{F}^t\right). \quad (2.1)$$

This formulation is similar to that of the pricing of assets in the Markowitz model. The first term in the right-hand side is the expected value at time t of the value $Y_i(T)$ discounted with the free-risk discount rate r . The second term is a premium linked to the risk. This risk premium is expressed by the covariance between the value at T discounted at rate r and a fixed market indicator, which does not depend on i . This indicator is similar to the Market portfolio and the covariance is similar to the β of the asset.

Let us now take $T = t + \delta$ with δ small. We can write

$$Y_i(t) \exp r\delta = E[Y_i(t + \delta)|\mathcal{F}^t] + \text{cov}\left(Y_i(t + \delta), \frac{Z(T)}{Z(t)}|\mathcal{F}^t\right),$$

and approximating $\exp r\delta$ by $1 + r\delta$, we get the formulation

$$E[\delta Y_i(t)|\mathcal{F}^t] = rY_i(t)\delta - Y_i(t) \sum_j \sigma_{ij}(t) \text{cov}\left(\delta w_j(t), \frac{\delta Z(t)}{Z(t)}|\mathcal{F}^t\right), \quad (2.2)$$

where

$$\delta f(t) = f(t + \delta) - f(t).$$

The risk premium can be expressed in terms of the covariance between the noises affecting the price values $Y_i(t)$ and the market indicator.

We can also express Eq. (2.2) as follows

$$\alpha_i dt = rdt - \sum_j \sigma_{ij}(t) \text{cov}\left(dw_j(t), \frac{dZ(t)}{Z(t)}|\mathcal{F}^t\right), \quad (2.3)$$

which can be seen as a consequence of Ito's calculus and the definition of $\theta(t)$. Note also the equivalence with

$$\alpha_i dt = rdt - \text{cov}\left(\frac{dY_i(t)}{Y_i(t)}, \frac{dZ(t)}{Z(t)}|\mathcal{F}^t\right). \quad (2.4)$$

2.2. Risk premium and CCAPM

We can recover these formulas by the Consumer Capital Asset Pricing Model (CCAPM) approach. In this approach, consider a generic "investor" who invests in the market. He should be representative of the behavior of investors on the market. Suppose that at time t his portfolio is made of an amount $\pi_i(t)$ of asset i and an amount $\pi_f(t)$ of riskless asset. His wealth at time t is then given by

$$X(t) = \sum_i \pi_i(t) Y_i(t) + \pi_f(t) \exp rt.$$

The evolution of $X(t)$ is only governed by the evolution of the portfolio. This implies

$$dX(t) = \sum_i \pi_i(t) dY_i(t) + \pi_f(t) r \exp rt.$$

We can eliminate $\pi_f(t)$ between these two relations. Besides, we use the amounts relative to the wealth, namely,

$$\varpi_i(t) = \frac{\pi_i(t)Y_i(t)}{X(t)}.$$

With these elements, we can derive the evolution of the wealth as follows

$$dX(t) = rX(t)dt + X(t)\varpi(t).\sigma(t)(dw(t) + \theta(t)dt),$$

and thus one observes that $X(t)Z(t) \exp -rt$ is also a $\mathcal{F}^t$ -martingale and more precisely

$$d(X(t)Z(t) \exp -rt) = X(t)Z(t) \exp -rt(\sigma^*(t)\varpi(t) - \theta(t)).dw(t). \quad (2.5)$$

This martingale property has an important consequence. First,

$$E(X(T)Z(T) \exp -rT) = X(0),$$

where the initial wealth is known and not random. To a given portfolio, evolution corresponds a given wealth profile. It is remarkable that one can invert the statement. If one considers an arbitrary wealth at time T , which is a $\mathcal{F}^T$ -measurable random variable satisfying the previous constraint, called the budget equation, then the wealth at any time is determined by the martingale property, and it is realized by a unique portfolio, thanks to the representation of martingales with respect to a filtration generated by Wiener processes.

The generic investor will choose his portfolio according to some utility function U depending on his final wealth. In other words, he seeks to maximize

$$EU(X(T)) \exp -\beta T,$$

where β is a subjective discount linked to the investor, as the utility function is. From this consideration, he can find his optimal final wealth, provided the budget equation constraint is respected. So, his problem reduces to finding $X(T)$ such that

$$\begin{aligned} &\text{maximize } EU(X(T)) \exp -\beta T \\ &\text{under the constraint } E(X(T)Z(T) \exp -rT) = X(0). \end{aligned}$$

We may introduce a Lagrange multiplier λ to deal with the constraint, and the optimal final wealth $\hat{X}(T)$ must satisfy

$$U'(\hat{X}(T)) \exp -\beta T = \lambda Z(T) \exp -rT.$$

Taking the conditional expectation with respect to $\mathcal{F}^t$, we deduce

$$\frac{U'(\hat{X}(T))}{E[U'(\hat{X}(T))|\mathcal{F}^t]} = \frac{Z(T)}{Z(t)}.$$

We plug this value in the pricing relation of assets to get

$$Y_i(t) = E[Y_i(T)|\mathcal{F}^t] \exp -r(T-t) + \text{cov} \left(\frac{Y_i(T)}{\exp r(T-t)}, \frac{U'(\hat{X}(T))}{E[U'(\hat{X}(T))|\mathcal{F}^t]} | \mathcal{F}^t \right), \quad (2.6)$$

which is a CCAPM pricing formula.

2.3. Definition of a tradable asset

A tradable asset is an asset whose value $Q(t)$ evolves according to the relation

$$dQ = Q(a(t)dt + b(t) \cdot dw(t))$$

with a drift such that

$$a(t) = r - E \left[\text{cov} \left(\frac{dQ}{Q}, \frac{dZ}{Z} \right) | \mathcal{F}^t \right].$$

The logic is that one could constitute a portfolio of the basic assets and risk less asset, which is equal to $Q(t)$, carries identical risk as the tradable asset, and consequently provides the same expected return, in view of the absence of arbitrage. This portfolio is equal to

$$X(t) = \sum_{i=1}^n \pi_i(t) Y_i(t) + \pi_f(t) \exp rt = Q(t),$$

and its evolution is

$$dX(t) = \sum_{i=1}^n \pi_i(t) dY_i(t) + r\pi_f(t) \exp rt dt.$$

So,

$$dX(t) = rQ(t)dt + \sum_{i=1}^n \pi_i(t) Y_i(t) (\alpha_i(t) - r)dt + \sum_{i,j=1}^n \pi_i(t) Y_i(t) \sigma_{ij}(t) dw_j.$$

Equating the risk term between dQ and dX yields

$$b_j(t)Q(t) = \sum_{i=1}^n \pi_i(t) Y_i(t) \sigma_{ij}(t),$$

and by equating the expected return, we obtain

$$a(t)Q(t) = rQ(t) + \sum_{i=1}^n \pi_i(t) Y_i(t) (\alpha_i(t) - r).$$

We use Eq. (2.3) to obtain

$$a(t)Q(t)dt = rQ(t)dt - \sum_{i=1}^n \pi_i(t)Y_i(t) \sum_{j=1}^n \sigma_{ij}(t)E \left[\text{cov} \left(dw_j, \frac{dZ}{Z} \right) | \mathcal{F}^t \right],$$

hence,

$$a(t)Q(t)dt = rQ(t)dt - Q(t) \sum_{j=1}^n b_j(t)E \left[\text{cov} \left(dw_j, \frac{dZ}{Z} \right) | \mathcal{F}^t \right].$$

Finally, we obtain

$$a(t)dt = rdt - E \left[\text{cov} \left(\frac{dQ(t)}{Q(t)}, \frac{dZ(t)}{Z(t)} \right) | \mathcal{F}^t \right]. \quad (2.7)$$

It is easy to derive the portfolio, which replicates the tradable asset $Q(t)$. Indeed, we have

$$\pi_i(t)Y_i(t) = Q(t) \sum_j ((\sigma^*)^{-1})_{ij}(t)b_j(t)$$

and thus

$$\pi_f(t) \exp rt = Q(t) - Q(t) \sum_{ij} ((\sigma^*)^{-1})_{ij}(t)b_j(t).$$

An equivalent way to obtain the relation (2.7) is to constitute a portfolio made of the tradable asset $Q(t)$ and a short position on the market assets $Y_i(t)$. Such a portfolio leads to a wealth

$$x(t) = Q(t) - \sum_i \pi_i(t)Y_i(t).$$

We recall that a short position on an asset means that one sells immediately the asset and receives the corresponding cash.

If $Q(t)$ is tradable, one should find the values of $\pi_i(t)$ so that the preceding portfolio is risk free, hence satisfies

$$dx(t) = rx(t)dt.$$

The fact that it is risk free results from the idea that by taking a short position on the market assets we eliminate the risk. The advantage of this approach is also that one does not introduce cash explicitly.

2.4. Case of dividends

Let us assume now that the market assets bear a coupon. More precisely, the asset $Y_i(t)$ yields a coupon $\delta_i(t)$ per unit value and per unit of time, hence $Y_i(t)$ yields $Y_i(t)\delta_i(t)dt$ during the period $t, t + dt$. Consider a tradable asset that carries a dividend denoted by $q(t)$ per unit value and per unit of time. The processes $\delta_i(t)$ and $q(t)$ are adapted to $\mathcal{F}^t$. The formulas change slightly. Indeed, during dt the increase of wealth due to the possession of $Y_i(t)$ is $dY_i(t) + Y_i(t)\delta_i(t)dt$, arising from the increase of price of the asset and the from the payment of the dividend.

If we consider the riskless portfolio

$$x(t) = Q(t) - \sum_i \pi_i(t)Y_i(t),$$

then we can write

$$dx(t) = Q(a + q)dt - \sum_i \pi_i Y_i(\alpha_i + \delta_i)dt + \sum_j \left(Qb_j - \sum_i \pi_i Y_i \sigma_{ij} \right) dw_j.$$

Proceeding as above, we obtain the formula

$$\begin{aligned} (a(t) + q(t))dt = rdt + \sum_{ij} \delta_i(t)((\sigma^*)^{-1})_{ij}(t)b_j(t)dt \\ - E \left[\text{cov} \left(\frac{dQ(t)}{Q(t)}, \frac{dZ(t)}{Z(t)} \right) | \mathcal{F}^t \right]. \end{aligned} \quad (2.8)$$

We can naturally apply formula (2.8) to $Y_k(t)$. We must use the values

$$a = \alpha_k; \quad b_j = \sigma_{kj}; \quad q = \delta_k$$

Plugging these values in (2.8) yields immediately (2.3). Indeed, this formula is not affected directly by the coupons on $Y_i(t)$ since they express uniquely the risk part of the asset.

3. Valuation of contingent claims

3.1. Claims on market assets

A claim is a right on the underlying assets to be exercised in a contract. We will consider first a claim on the market assets $Y_i(t)$ components of a vector $Y(t)$. Since the approach is based on Ito's calculus, we will assume that

$$\begin{aligned} \alpha_i(t) &= \alpha_i(Y(t), t), \\ \sigma_{ij}(t) &= \sigma_{ij}(Y(t), t), \\ \delta_i(t) &= \delta_i(Y(t), t) \end{aligned}$$

in which the functions on the right-hand side are deterministic (the randomness is captured through the process $Y(t)$).

A claim on the market assets has a value at time t , which is a function of the values $Y(t)$, expressed as $F(Y(t), t)$. It is a capital asset and is fully paid. We assume that the possession of this claim generates a profit per unit of time $q(F, t)$, which is equivalent to a dividend

$$q(t) = \frac{q(F, t)}{F}$$

in which the argument F has to be replaced with $F(Y(t), t)$ to obtain the precise value. We consider

$$Q(t) = F(Y(t), t)$$

as a tradable asset, which carries the above dividend.

We can write

$$dQ(t) = Q \left(a dt + \sum_j b_j dw_j \right),$$

where the values of a and b_j can be obtained, thanks to Ito's calculus and the Ito differentials of $Y_i(t)$. We get easily

$$Fa = \frac{\partial F}{\partial t} + \sum_i \alpha_i Y_i \frac{\partial F}{\partial Y_i} + \sum_{ij} \frac{1}{2} \frac{\partial^2 F}{\partial Y_i \partial Y_j} \sigma_{ij} Y_i Y_j$$

$$Fb_j = \sum_i \frac{\partial F}{\partial Y_i} Y_i \sigma_{ij}.$$

We can, then, apply the formula (2.8) and cancel identical terms on both sides.

We can state the following.

THEOREM 3.1. *A tradable contingent claim on the market assets $F(Y, t)$ must be a solution of the partial differential equation (PDE)*

$$\frac{\partial F}{\partial t} + \sum_i (r - \delta_i) Y_i \frac{\partial F}{\partial Y_i} + \sum_{ij} \frac{1}{2} \frac{\partial^2 F}{\partial Y_i \partial Y_j} \sigma_{ij} Y_i Y_j - rF + q = 0. \quad (3.1)$$

The boundary conditions are part of the description of the contract related to the contingent claim. When we will consider real options as contingent claims, we will make precise these boundary conditions.

3.2. Claims on a tradable asset

Suppose now that we have a tradable asset $Q(t)$, which is not one of the market assets and we have a claim on this asset. The tradable asset itself is described by the Ito differential

$$dQ = Q(a(Q, t)dt + \sum_j b_j(Q, t)dw_j(t)).$$

We shall assume that it carries a dividend per unit value and unit of time denoted by $\delta(Q, t)$. We do not use the notation $q(t)$ for the dividend since it will be used for the contingent claim. Since the asset is tradable, we have the relation see Eq. (2.8)

$$(\delta(Q, t) + a(Q, t))dt = r + \sum_{ij} \delta_i(t)((\sigma^*)^{-1})_{ij}(t)b_j(Q, t)dt - E \left[\text{cov} \left(\sum_j b_j(Q, t)dw_j(t), \frac{dZ(t)}{Z(t)} \right) | \mathcal{F}^t \right].$$

The contingent claim is valued by a function $F(Q, t)$, and it brings a profit $q(Q, t)$. Proceeding as in the case of contingent claims on market assets, we derive easily the following equation for the valuation function

$$\frac{\partial F}{\partial t} + (r - \delta(Q, t))Q \frac{\partial F}{\partial Q} + \frac{1}{2} \frac{\partial^2 F}{\partial Q^2} |b(Q, t)|^2 - rF + q(Q, t) = 0. \quad (3.2)$$

3.3. Claims on a nontradable asset

Some limited extensions can be obtained for claims on nontradable assets. We assume that the nontradable asset obeys the stochastic differential

$$dQ = Q(a(Q, t)dt + \sum_j b_j(Q, t)dw_j(t)),$$

but since this asset is nontradable, the drift $a(Q, t)$ does not satisfy the relation (2.8). This asset carries a dividend $\delta(Q, t)$ per unit value and per unit of time. We follow the model of DIXIT and PINDYCK (see [1994]). We assume there are tradable assets, the value of which is denoted by $Q_i(t)$, $i = 1, \dots, m$ verifying

$$dQ_i = Q_i(A_i(Q, t)dt + \sum_j B_{ij}(Q, t)dw_j(t)),$$

which provides a dividend $D_i(Q, t)$ per unit value and per unit of time.

We note that in the above model, the functions $A_i(Q, t)$, $B_{ij}(Q, t)$, and $D_i(Q, t)$ depend on Q and not on Q_i . The main assumption is that the vector b can be written as a linear combination of the vectors B_i , whose components are B_{ij} . Namely,

$$b = \sum_{i=1}^m \gamma_i B_i.$$

This assumption is satisfied when $m = n$ and

the matrix $B_{ij}(Q, t)$ is invertible.

The contingent claim is a function $F(Q, t)$. We consider a portfolio made of the contingent claim and a short position on the assets $Q_i(t)$, whose wealth is given by

$$x(t) = F(Q, t) - \sum_i \pi_i(t) Q_i(t),$$

and we want this portfolio to be risk free. A routine calculation shows that, to ensure risk free, we must have the relations

$$\frac{\partial F}{\partial Q} Q b_j(Q, t) = \sum_i \pi_i(t) Q_i(t) B_{ij}(Q, t).$$

By the assumption, we introduce the vector $\gamma(Q, t)$ of components $\gamma_i(Q, t)$, which in the case of invertibility of the matrix B is given by $(B^*(Q, t))^{-1} b(Q, t)$. The risk-free condition is equivalent to the portfolio

$$\pi_i(t) Q_i(t) = \gamma_i(Q, t) Q \frac{\partial F}{\partial Q}.$$

Expressing that this portfolio satisfies

$$dx = rx(t)dt,$$

we obtain the valuation equation

$$\frac{\partial F}{\partial t} + (a - \sum_i (\gamma_i(A_i + D_i - r)) Q \frac{\partial F}{\partial Q} + \frac{1}{2} \frac{\partial^2 F}{\partial Q^2} |b(Q, t)|^2 - rF + q(Q, t) = 0. \quad (3.3)$$

We can also recover the relation (3.3) by expressing directly that the assets $Q_i(t)$ and the contingent claim $F(Q, t)$ are tradable. We must have

$$A_i + D_i = r + \sum_{kj} ((\sigma^*)^{-1})_{kj} B_{ij} \delta_k - E \left[\text{cov} \left(\sum_j B_{ij} dw_j(t), \frac{dZ(t)}{Z(t)} \right) | \mathcal{F}^t \right]. \quad (3.4)$$

EXERCISE 3.1. Obtain Eq. (3.3) by using (3.4), computing the Ito differential of F and expressing that the claim $F(Q, t)$ is tradable.

3.4. Valuation of futures

A future is a contingent claim, which is not payable at the time of agreement. It is not a capital asset. It is an agreement to perform an exchange at some future date. Therefore, it does not carry dividends or any interest rate. If $F(Y(t), t)$ represents the value of a future based on market assets, then we can form a portfolio with the future and a short position on the market assets, which is risk free. However, setting

$$x(t) = F(Y(t), t) - \sum_i \pi_i(t) Y_i(t),$$

we have

$$dx(t) = -r \sum_i \pi_i(t) Y_i(t) dt$$

since F cannot yield any interest rate.

Developing the above conditions, we obtain

$$\frac{\partial F}{\partial t} + \sum_i (r - \delta_i) Y_i \frac{\partial F}{\partial Y_i} + \sum_{ij} \frac{1}{2} \frac{\partial^2 F}{\partial Y_i \partial Y_j} \sigma_{ij} Y_i Y_j = 0 \quad (3.5)$$

4. Valuation of a project

4.1. Description of the model

A project is characterized by an output called $P(t)$. This is the basic “asset,” which is the source of value. It is measured in dollars. It is a stochastic process governed by the equation

$$dP = \alpha P dt + P \sum_j \sigma_j dw_j, \quad (4.1)$$

where $w(t)$ is a the n -dimensional Wiener process $w_1, \dots, w_n$ modeling the randomness of the economy. So the output carries the same randomness as the economy in general. So we can consider $P(t)$ as an asset, which may be traded or not. To simplify a little bit, we shall assume that α and σ are constant and deterministic, as are the noise correlations σ_{ij} and the coupons δ_i of the basic assets of the economy $Y_i(t)$.

Suppose P carries a dividend δ per unit value and per unit of time. It is constant to simplify. If P is tradable, we have the relation

$$\alpha + \delta = \mu,$$

where μ called the risk-adjusted expected rate of return of P is given by (see Eq. (2.8))

$$\mu = r + \sum_{ij} ((\sigma^*)^{-1})_{ij} \sigma_j \delta_i - \sum_j \text{cov} \left(\sigma_j dw_j, \frac{dZ}{Z} \right).$$

Even when there is no dividend, it is economically meaningful to assume $\mu > \alpha$ since α is the expected capital gain. This assumption implies that the randomness involved in P brings more opportunities than risks. P should generate more wealth than its expected capital gain. The difference $\mu - \alpha$ can be denoted by $\delta > 0$ and acts as a dividend.

If P is not tradable, we shall assume that P is spanned by financial markets, which means that there exist traded assets $Q_i(t)$ perfectly correlated with P as follows

$$dQ_i = Q_i \left(A_i dt + \sum_j B_{ij} dw_j \right).$$

We assume that

$$\sigma_j = \sum_i \gamma_i B_{ij}.$$

In addition, we require that

$$\sum_i \gamma_i = 1.$$

With these assumptions we are in the same situation as above by defining

$$\mu = \sum_i \gamma_i (A_i + D_i)$$

and

$$\delta = \mu - \alpha.$$

4.2. The valuation equation

The project carries a flow of profits given by $\pi(P, t)$ when the output is P at time t . Denote by $V(P, t)$ the value of owning the project (we can also think of a firm instead of a project and speak about the value of the firm). If the output itself is tradable, then we have

$$\alpha = \mu - \delta,$$

where μ is the risk-adjusted rate of return of the output

$$\mu = r + \sum_{ij} ((\sigma^*)^{-1})_{ij} \sigma_j \delta_i - \sum_j \text{cov} \left(\sigma_j dw_j, \frac{dZ}{Z} \right).$$

Considering the value of ownership of the project as a tradable asset and writing its differential, we have

$$dV = \left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial P} P \alpha + \frac{1}{2} \frac{\partial^2 V}{\partial P^2} P^2 \sigma^2 \right) dt + \frac{\partial V}{\partial P} P \sum_j \sigma_j dw_j,$$

where we have used the notation

$$\sigma^2 = \sum_j \sigma_j^2.$$

The risk-adjusted rate of return of the ownership of the project is then

$$\tilde{\mu} = r + \frac{\partial V}{\partial P} \frac{P}{V} \left[\sum_{ij} ((\sigma^*)^{-1})_{ij} \sigma_j \delta_i - \sum_j \text{cov} \left(\sigma_j dw_j, \frac{dZ}{Z} \right) \right].$$

Therefore,

$$\tilde{\mu} - r = \frac{1}{V} \frac{\partial V}{\partial P} P(\mu - r).$$

We then write that the expected capital return on the ownership of the project plus the profit flow per unit value is equal to $\tilde{\mu}$. This means

$$V\tilde{\mu} = \pi + \frac{\partial V}{\partial t} + \frac{\partial V}{\partial P} P\alpha + \frac{1}{2} \frac{\partial^2 V}{\partial P^2} P^2 \sigma^2.$$

Replacing α and eliminating common terms yield

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 V}{\partial P^2} P^2 \sigma^2 - rV + \pi = 0. \quad (4.2)$$

4.3. Solution of the valuation equation

Eq. (4.2) is a partial differential equation whose space variable P lies in $(0, \infty)$. We need boundary conditions. As far as time is concerned, we will, in most cases, look for stationary solutions of Eq. (4.2), namely,

$$\frac{\partial V}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 V}{\partial P^2} P^2 \sigma^2 - rV + \pi = 0, \quad (4.3)$$

which is possible when coefficients are independent of time (as we assumed), the profit flow is also independent of time, and the horizon is infinite. If we cannot consider the horizon as infinite, we may consider the following terminal condition

$$V(P, T) = 0,$$

whose interpretation is clear.

Concerning the variable P we need conditions at 0 and ∞ . For $P = 0$, we see formally on Eq. (4.3) that

$$V(0) = \frac{\pi(0)}{r}.$$

It is natural to assume that $\pi(0) = 0$, so $V(0) = 0$. This assumes implicitly that V does not have a singularity at 0. In fact, the condition $V(0)$, makes perfect economic sense. Concerning the condition at infinity, we need to specify a growth condition. We shall require a growth condition similar to that of $\pi(P)$. Leaving aside a particular solution, which will be valid for P large, the general solution is an exponential

$$V(P) = \exp \beta P$$

with β to be a solution of

$$\frac{1}{2} \sigma^2 \beta(\beta - 1) + (r - \delta)\beta - r = 0.$$

The roots of this quadratic expression are β_1 , and β_2 , with

$$\beta_1 = \frac{1}{2} - \frac{r - \delta}{\sigma^2} + \sqrt{\left[\frac{1}{2} - \frac{r - \delta}{\sigma^2}\right]^2 + \frac{2r}{\sigma^2}},$$

$$\beta_2 = \frac{1}{2} - \frac{r - \delta}{\sigma^2} - \sqrt{\left[\frac{1}{2} - \frac{r - \delta}{\sigma^2}\right]^2 + \frac{2r}{\sigma^2}},$$

and

$$\beta_1 > 1, \beta_2 < 0$$

EXERCISE 4.1. Show that if $\pi(P) = P$, then

$$V(P) = \frac{P}{\delta}.$$

EXERCISE 4.2. Assume

$$\pi(P) = (P - C)^+,$$

then one has

$$V(P) = \begin{cases} V_1 P^{\beta_1} & \text{if } P \leq C \\ V_2 P^{\beta_2} + \frac{P}{\delta} - \frac{C}{r} & \text{if } P \geq C, \end{cases}$$

where

$$V_1 = \frac{C^{1-\beta_1}}{\beta_1 - \beta_2} \left(\frac{\beta_2}{r} - \frac{\beta_2 - 1}{\delta} \right)$$

$$V_2 = \frac{C^{1-\beta_2}}{\beta_1 - \beta_2} \left(\frac{\beta_1}{r} - \frac{\beta_1 - 1}{\delta} \right).$$

Note that $V_1, V_2 > 0$.

The previous profit flow function is interpreted as follows: there is an operating cost C and when $P < C$, then the project can be interrupted, with no cost. Whenever $P \geq C$, the project can be resumed with no cost either.

EXERCISE 4.3. Assume

$$\pi(P) = kP^\gamma, \gamma > 1,$$

then

$$V(P) = \frac{kP^\gamma}{\rho}$$

with

$$\rho = r - \gamma(r - \delta) - \frac{1}{2}\sigma^2\gamma(\gamma - 1).$$

The preceding profit flow is motivated by the following optimization

$$\pi(P) = \max_v [Ph(v) - C(v)],$$

where $h(v)$ represents a production function and $C(v)$ a variable operating cost. The particular case

$$h(v) = v^\theta, 0 < \theta < 1, C(v) = c$$

yields the function chosen above with

$$\gamma = \frac{1}{1 - \theta} > 1.$$

4.4. Probabilistic interpretation

We can interpret $V(P, t)$ as follows. The output evolution is described by

$$dP = P((r - \delta)ds + \sum_j \sigma_j dw_j(s)), s \geq t; \quad P(t) = P.$$

Note that the drift has been changed from α to $r - \delta$. Then, we have

$$V(P, t) = E \int_t^\infty \exp -r(s - t) \pi(P(s), s) ds.$$

Note that the discount applicable is the riskless discount r .

EXERCISE 4.4. Obtain by the probabilistic interpretation $V(P)$ in the cases $\pi(P) = P$, $\pi(P) = (P - C)^+$, and $\pi(P) = kP^\gamma$.

4.5. Possibility of death

Let us suppose that the project can be stopped accidentally by an external event, which arrives at a random time τ distributed with an exponential density with rate λ . Using the probabilistic interpretation, we can write

$$V(P, t) = E \left[\int_t^\tau \exp -r(s - t) \pi(P(s), s) ds \mid \tau > t \right].$$

The conditional density of τ given $\tau > t$ is $\lambda \exp -\lambda(\theta - t)d\theta$ on the interval t, ∞ .

EXERCISE 4.5. Check that

$$V(P, t) = E \left[\int_t^\infty \exp -(r + \lambda)(s - t) \pi(P(s), s) ds \right].$$

5. Valuation of an option to invest

5.1. Motivation

In the previous section, we have discussed the value $V(P, t)$ of a project depending on a random output whose value at time t is P . The value stream of this project is a profit flow $\pi(P, t)$. To simplify the discussion, we will consider the stationary case $V(P)$ corresponding to a profit flow $\pi(P)$. Note that whenever $\pi(P)$ is an increasing function of P (a natural assumption), $V(P)$ is also increasing in P . This can be seen by using the probabilistic formula and recognizing that $P(s)$ being log normal with initial condition P is increasing in P . The problem we face now is that of investing in the project an amount I . In other words, we pay a price I to get a project of value $V(P)$. Under net present value (NPV) approach, we will invest if $P \geq P_0$, where

$$I = V(P_0).$$

5.2. The option approach

Although natural and used constantly, the NPV approach has a serious flaw. It rules out one possibility, that of postponing the decision to invest to wait for more favorable values of P . It does not take into consideration the flexibility, which is key in decision making. The option approach aims at introducing a flexibility in the time of decision. At any time, we can either invest immediately, in which case we get $V(P) - I$, or postpone a little bit of time and consider that we have a contingent claim to be valued according to valuation techniques already discussed. We denote by $F(P)$ the value of the option. What is the problem to be solved to find this function.

Since on the branches of the alternative is to invest immediately, we must have

$$F(P) \geq V(P) - I, \quad \forall P.$$

The other branch is to keep the option. To proceed, we must use valuation concepts. Let us assume, for instance, that the output P is not directly tradable, but there exists a tradable asset Q governed by

$$dQ = Q(Adt + \sum_j \sigma_j dw_j(t)).$$

We can form a portfolio made of the option and a short position in the tradable asset Q .

To achieve a riskless portfolio, we must assume $F(P)$ to be smooth, so we can apply Ito's calculus. A portfolio $F(P(t)) - \pi(t)Q(t)$ will be riskless, if

$$Q(t)\pi(t) = P(t)F'(P(t)).$$

If we keep the option, we get just the increase in capital, since the option by itself does not carry any dividend. The expected increase in capital is

$$F'(P)P\alpha + \frac{1}{2}F''(P)\sigma^2 - F'(P)PA.$$

Since it is risk free, it cannot be larger than $r(F - PF'(P))$. As usual, we set $\delta = A - \alpha > 0$ and we get the inequality

$$F'(P)P(r - \delta) + \frac{1}{2}F''(P)\sigma^2 - rF \leq 0.$$

Since the alternative has only two branches, at any time, one of the inequality must become an equality.

5.3. Variational inequality

We can summarize by stating that $F(P)$ is solution of the following set of differential inequalities and complementarity slackness condition

$$\begin{aligned} F(P) &\geq V(P) - I \\ F'(P)P(r - \delta) + \frac{1}{2}F''(P)\sigma^2 - rF &\leq 0 \\ (F(P) - V(P) + I)(F'(P)P(r - \delta) + \frac{1}{2}F''(P)\sigma^2 - rF) &= 0. \end{aligned} \tag{5.1}$$

This problem is called a variational inequality (VI). It must be completed by boundary conditions and smoothness conditions. We take $F(0) = 0$, and we assume a growth condition similar to that of $V(P)$, hence of $\pi(P)$. In addition, we require F to be continuously differentiable.

We can give a probabilistic interpretation for $F(P)$ as a problem of optimal stopping. Recall that we must consider that $P(t)$ has the differential

$$dP = P((r - \delta)ds + \sum_j \sigma_j dw_j(s)); \quad P(0) = P.$$

Consider now stopping times τ adapted to the σ -field generated by the process $P(t)$, then we have

$$F(P) = \max_{\tau} E[(V(P(\tau)) - I) \exp -r\tau]. \tag{5.2}$$

To solve the VI, we look for a value P^* , such that

$$\begin{aligned} F'(P)P(r - \delta) + \frac{1}{2}F''(P)\sigma^2 - rF &= 0, \quad P < P^* \\ F(P) &= V(P) - I, \quad P \geq P^* \\ F'(P^*) &= V'(P^*). \end{aligned}$$

Define

$$\Gamma(P) = F(P) - V(P) + I.$$

Then, we have

$$\begin{aligned}\Gamma(P) &= 0, P \geq P^* \\ \Gamma'(P)P(r - \delta) + \frac{1}{2}\Gamma''(P)\sigma^2 - r\Gamma &= \pi(P) - rI, P < P^* \\ \Gamma'(P^*) &= 0.\end{aligned}\tag{5.3}$$

THEOREM 5.1. *Assume that $\pi(P)$ increases and that a solution of the system (5.3) exists. Then,*

$$F(P) = \Gamma(P) + V(P) - I$$

is solution of the VI (5.1).

PROOF. We have to check

$$\begin{aligned}\Gamma(P) &> 0, P < P^* \\ \Gamma'(P)P(r - \delta) + \frac{1}{2}\Gamma''(P)\sigma^2 - r\Gamma &\leq \pi(P) - rI, P > P^*.\end{aligned}\tag{5.4}$$

First, we must have

$$\pi(P^*) - rI > 0.$$

Indeed, consider

$$\gamma(P) = \Gamma'(P).$$

It is the solution of

$$\gamma'P(r - \delta + \sigma^2) + \frac{1}{2}\sigma^2P^2\gamma'' - r\gamma = \pi'(P)$$

and $\gamma(P^*) = 0$. Since $\pi'(P) > 0$, it follows that

$$\gamma(P) < 0, P < P^*.$$

This implies also

$$\gamma'(P^*) = \Gamma''(P^*) > 0,$$

hence necessarily from the equation letting $P \uparrow P^*$, we obtain the desired property. Hence, the second part of condition (5.4) is satisfied, since $\pi(P)$ is increasing.

To prove the first part of assertion (5.4), we first notice that $\Gamma(P) > 0$ for P sufficiently close to P^* . Indeed, we may write the formula

$$\Gamma(P) = (P - P^*)^2 \int_0^1 \int_0^1 \lambda \Gamma''(P^* + \lambda\mu(P - P^*)) d\lambda d\mu,$$

and the integrand is positive for P sufficiently close to P^* . This, of course, postulates that $\Gamma''(P)$ is a continuous function. Introduce next P_* defined by

$$\pi(P_*) = rI, \quad P_* < P^*,$$

then, we first prove that

$$\Gamma(P) > 0, \quad \forall P_* < P < P^*.$$

If this assertion is not true, then there exists $P_0 \neq P_*$, with $P_* < P_0 < P^*$ and $\Gamma(P_0) = 0$. Let $\bar{P}$ be this point. One should have

$$\Gamma'(\bar{P}) = 0, \quad \Gamma''(\bar{P}) < 0, \quad \Gamma(\bar{P}) > 0,$$

which is impossible since $\pi(\bar{P}) - rI > 0$. Hence,

$$\Gamma(P) > 0, \quad \forall P_* \leq P < P^*.$$

On $[0, P_*]$, let $\underline{P}$ be the minimum of $\Gamma(P)$. We cannot have $\Gamma(\underline{P}) < 0$. If so, $\underline{P}$ is in the interior of the interval and

$$\Gamma'(\underline{P}) = 0, \quad \Gamma''(\underline{P}) > 0, \quad \Gamma(\underline{P}) < 0.$$

Hence, $\pi(\bar{P}) - rI > 0$, which is impossible. The proof has been completed. $\square$

5.4. Parabolic VI

We consider now the nonstationary case: the problem (5.1) becomes

$$\begin{aligned} F(P, t) &\geq V(P, t) - I \\ \frac{\partial F}{\partial t} + \frac{\partial F}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 F}{\partial P^2} \sigma^2 - rF &\leq 0 \\ (F(P, t) - V(P, t) + I) \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 F}{\partial P^2} \sigma^2 - rF \right) &= 0 \\ F(P, T) &= 0. \end{aligned} \tag{5.5}$$

This problem is called a parabolic VI.

We cannot use the same approach as that used for stationary VI (called also elliptic VI). This is because of the extra term, which is the partial derivative with respect to time. We are interested in the same type of solution, namely, find a free boundary $P^*(t)$ with F satisfying

$$\begin{aligned} \frac{\partial F}{\partial t} + \frac{\partial F}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 F}{\partial P^2} \sigma^2 - rF &= 0, \quad P < P^*(t) \\ F(P, t) &= V(P, t) - I, \quad P \geq P^*(t) \\ \frac{\partial F}{\partial P}(P^*(t), t) &= 0, \quad F(P, T) = 0. \end{aligned} \tag{5.6}$$

THEOREM 5.2. Assume $\frac{\partial \pi}{\partial t} \leq 0$, $\frac{\partial \pi}{\partial P} \geq 0$, and $\pi(0, t) = 0$. Then, there exists a continuous differentiable function F solution of the parabolic VI (5.5), which is of the form (5.6)

PROOF. It is convenient to introduce

$$\Gamma(P, t) = F(P, t) - V(P, t) + I$$

and to write the parabolic VI, as follows:

$$\begin{aligned} \Gamma(P, t) &\geq 0 \\ \frac{\partial \Gamma}{\partial t} + \frac{\partial \Gamma}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial P^2} \sigma^2 - r\Gamma &\leq \pi(P, t) - rI \\ \Gamma(P, t) \left(\frac{\partial \Gamma}{\partial t} + \frac{\partial \Gamma}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial P^2} \sigma^2 - r\Gamma - \pi(P, t) + rI \right) &= 0 \\ \Gamma(P, T) &= I \end{aligned} \tag{5.7}$$

We shall show the existence of a continuously differentiable solution of (5.7), which satisfies

$$\frac{\partial \Gamma}{\partial t} \geq 0, \quad \frac{\partial \Gamma}{\partial P} \leq 0. \tag{5.8}$$

If this is true, then note that $\Gamma(0, t)$ satisfies

$$\begin{aligned} \frac{\partial \Gamma}{\partial t}(0, t) - r\Gamma(0, t) &\leq -rI \\ \Gamma(0, t) \left(\frac{\partial \Gamma}{\partial t}(0, t) - r\Gamma(0, t) + rI \right) &= 0 \\ \Gamma(0, t) &\geq 0, \quad \Gamma(0, T) = I; \end{aligned}$$

therefore, $\Gamma(0, t) = I$. □

We then define

$$P^*(t) = \{\inf P | \Gamma(P, t) = 0.\}$$

Note that $P^*(T) = \infty$. By definition

$$\frac{\partial \Gamma}{\partial t} + \frac{\partial \Gamma}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial P^2} \sigma^2 - r\Gamma = \pi(P, t) - rI, \quad P < P^*(t).$$

Next for $P > P^*(t)$, we have

$$0 \leq \Gamma(P, t) \leq \Gamma(P^*(t), t) = 0.$$

Hence, $\Gamma(P, t) = 0$ for $P > P^*(t)$. Clearly,

$$P^*(t) > P_*(t) \tag{5.9}$$

with

$$\pi(P_*(t), t) = rI.$$

The study of the parabolic VI is done using the penalty approximation

$$\frac{\partial \Gamma^\epsilon}{\partial t} + \frac{\partial \Gamma^\epsilon}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 \Gamma^\epsilon}{\partial P^2} \sigma^2 + \frac{1}{\epsilon} (\Gamma^\epsilon)^- - r\Gamma^\epsilon = \pi(P, t) - rI$$

with the initial condition

$$\Gamma^\epsilon(P, T) = I.$$

By classical methods, the solution of the penalty approximation converges toward a solution of the parabolic VI.

However, if we consider

$$\eta^\epsilon(P, t) = \frac{\partial \Gamma^\epsilon}{\partial t},$$

then it is the solution of

$$\frac{\partial \eta^\epsilon}{\partial t} + \frac{\partial \eta^\epsilon}{\partial P} P(r - \delta) + \frac{1}{2} \frac{\partial^2 \eta^\epsilon}{\partial P^2} \sigma^2 - \frac{\eta^\epsilon}{\epsilon} \mathbb{I}_{\Gamma^\epsilon < 0} - r\eta^\epsilon = \frac{\partial \pi(P, t)}{\partial t}$$

with final condition

$$\eta^\epsilon(P, T) = \pi(P, T)$$

from which (and the assumption) it follows that

$$\eta^\epsilon(P, t) \geq 0.$$

Similarly, consider

$$\gamma^\epsilon(P, t) = \frac{\partial \Gamma^\epsilon}{\partial P},$$

which is the solution of

$$\frac{\partial \gamma^\epsilon}{\partial t} + \frac{\partial \gamma^\epsilon}{\partial P} P(r - \delta + \sigma^2) + \frac{1}{2} \frac{\partial^2 \gamma^\epsilon}{\partial P^2} \sigma^2 - \frac{\gamma^\epsilon}{\epsilon} \mathbb{I}_{\Gamma^\epsilon < 0} - \delta \gamma^\epsilon = \frac{\partial \pi(P, t)}{\partial P},$$

with final condition

$$\gamma^\epsilon(P, T) = 0$$

from which (and the assumption again) we obtain

$$\gamma^\epsilon(P, t) \leq 0.$$

Letting ϵ go to 0, we obtain the properties (5.8).

Note that from the sign conditions (5.8), we deduce

$$\frac{dP^*(t)}{dt} > 0, \quad (5.10)$$

which follows from differentiating the relation

$$\Gamma(P^*(t), t) = 0.$$

5.5. Comparison with NPV

From the preceding characterization of the value of the option to invest in a project $F(P, t)$ it follows that we use the following decision rule.

$$\text{We invest in the project if } P(t) \geq P^*(t), \quad (5.11)$$

where $P(t)$ is the value of the output at time t representing the present time (more precisely the time of decision). On the other hand, the traditional NPV decision rule tells that one invests at time t provided that

$$V(P(t), t) \geq I.$$

A natural question is to compare these two decision rules.

We begin by stating the property

$$v(P, t) = \frac{\partial V(P, t)}{\partial P} > 0. \quad (5.12)$$

This is similar to the reverse property for Γ (see (5.8)). Differentiating Eq. (4.2) in P , we obtain

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial P} P(r - \delta + \sigma^2) + \frac{1}{2} \frac{\partial^2 v}{\partial P^2} \sigma^2 - \delta v + \frac{\partial \pi(P, t)}{\partial P} = 0,$$

with final condition

$$v(P, T) = 0.$$

It follows easily that $v(P, t)$ is positive. We may define $\hat{P}(t)$ such that

$$V(\hat{P}(t), t) = I.$$

The NPV rule is equivalent to the following rule

$$\text{we invest in the project if } P(t) \geq \hat{P}(t) \quad (5.13)$$

We prove the following.

THEOREM 5.3. *Under the assumptions of Theorem 5.2, we have*

$$\hat{P}(t) < P^*(t).$$

So the NPV rule is always wrong. One invests much too early with this rule.

PROOF. The result will follow from the fact that

$$F(P^*(t), t) = V(P^*(t), t) - I > 0.$$

From the definition of $\hat{P}(t)$ and the monotonicity property of $V(P, t)$ with respect to P the result is obtained. We remark that $F(P, t)$ is the solution of the problem

$$\frac{\partial F}{\partial t} + \frac{\partial F}{\partial P}P(r - \delta) + \frac{1}{2} \frac{\partial^2 F}{\partial P^2} \sigma^2 - rF = 0, \quad P < P^*(t)$$

$$F(0, t) = 0 \quad F(P, T) = 0$$

$$\frac{\partial F}{\partial P}(P^*(t), t) = \frac{\partial V}{\partial P}(P^*(t), t) > 0.$$

We claim that

$$F(P, t) \geq 0, \quad \forall P, t \text{ such that } P \leq P^*(t), \quad t \leq T.$$

It is sufficient to consider a minimum P_0, t_0 of the function $F(P, t)$ and to prove that $F(P_0, t_0) \geq 0$. Suppose we have $F(P_0, t_0) < 0$. We cannot have $P_0 = 0$ or $t_0 = T$. We cannot have either $P_0 = P^*(t_0)$. Indeed, since P_0, t_0 is a minimum of $F(P, t)$,

$$F(P_0 - \epsilon, t_0) \geq F(P_0, t_0)$$

from which we get $\frac{\partial F}{\partial P}(P_0, t_0) \leq 0$, which is not possible if $P_0 = P^*(t_0)$. So the point P_0, t_0 is in the interior of the domain, for which the partial differential equation holds. A look at the equation shows that $F(P_0, t_0) > 0$, which is impossible. Hence, the positivity of F . Finally, this positivity combined with the strict positivity of the partial derivative in P at $P^*(t)$ implies the strict positivity of $F(P^*(t), t)$, hence the result. $\square$

The quantity

$$F(P^*(t), t) = V(P^*(t), t) - I > 0$$

represents the value of the option at time t , hence the value of flexibility in the decision of investment at time t .

EXERCISE 5.1. Solve the parabolic VI in the case of the profit flow

$$\pi(P, t) = P \exp -\lambda t.$$

Show that

$$V(P, t) = \frac{P}{\lambda + \delta} \exp -\lambda t (1 - \exp -(\lambda + \delta)(T - t))$$

$$P^*(t) = \frac{\beta_1}{\beta_1 - 1} \frac{I(\lambda + \delta)}{\exp -\lambda t (1 - \exp -(\lambda + \delta)(T - t))}$$

hence,

$$V(P^*(t), t) = \frac{\beta_1}{\beta_1 - 1} I.$$

6. Extensions

The model studied so far has many limitations. There is a single project under consideration (or a single firm). The price for investment is fixed. Within this framework, the investment will be decided. The only question is when there is no possibility of abandonment or temporary mothballing. Also, the investment is external, there is no policy of building capital. The objective in the following sections is to relax some of these limitations and also to consider how the market mechanism will lead to an equilibrium.

7. Uncertainties on investment

7.1. Assumptions

We consider here a model in which not only the value of the output flow is random but also the investment needed can vary with time. More specifically, we have

$$dP = P \left(\alpha_P dt + \sum_j \sigma_{Pj} dw_j \right)$$

and a similar relation for the investment

$$dI = I \left(\alpha_I dt + \sum_j \sigma_{Ij} dw_j \right).$$

To fix the ideas that both P and I are tradable assets, we will consider their risk-adjusted expected returns μ_P, μ_I and assume as usual

$$\mu_P - \alpha_P = \delta_P > 0, \quad \mu_I - \alpha_I = \delta_I > 0,$$

and δ_P, δ_I can be interpreted as dividends associated with the output flow or the investment flow.

We will write

$$\sigma_P^2 = \sum_j \sigma_{Pj}^2, \quad \sigma_I^2 = \sum_j \sigma_{Ij}^2,$$

and

$$\rho \sigma_P \sigma_I = \sum_j \sigma_{Pj} \sigma_{Ij}.$$

We suppose to simplify that the profit flow of the project is given by

$$\pi(P) = P.$$

We will then consider a stationary model. We first note that the value of the project does not depend on the investment; therefore, in view of the profit flow and Exercise 4.1, we have

$$V(P) = \frac{P}{\delta_P}.$$

7.2. Valuation of the option

The value of the option $F(P, I)$ depends on P and I . It did not appear explicitly when I was just a constant, but now its evolution must be taken into consideration. By standard arguments, one obtains that $F(P, I)$ must be the solution of the VI

$$\begin{aligned} \frac{\partial F}{\partial P} P(r - \delta_P) + \frac{\partial F}{\partial I} I(r - \delta_I) \\ + \frac{1}{2} \frac{\partial^2 F}{\partial P^2} P^2 \sigma_P^2 + \frac{1}{2} \frac{\partial^2 F}{\partial I^2} I^2 \sigma_I^2 + \frac{\partial^2 F}{\partial P \partial I} P I \rho \sigma_P \sigma_I - rF \leq 0 \end{aligned} \quad (7.1)$$

$$F - \frac{P}{\delta_P} + I \geq 0$$

$$\text{product} = 0,$$

where “product” means the product of the two quantities on the left-hand side of the inequalities.

It is convenient to introduce

$$\Gamma(P, I) = F(P, I) - \frac{P}{\delta_P} + I,$$

and we get the problem

$$\begin{aligned} \frac{\partial \Gamma}{\partial P} P(r - \delta_P) + \frac{\partial \Gamma}{\partial I} I(r - \delta_I) + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial P^2} P^2 \sigma_P^2 \\ + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial I^2} I^2 \sigma_I^2 + \frac{\partial^2 \Gamma}{\partial P \partial I} P I \rho \sigma_P \sigma_I - r\Gamma \leq P - I\delta_I \end{aligned} \quad (7.2)$$

$$\Gamma \geq 0$$

$$\text{product} = 0.$$

Fortunately, this two-dimensional problem can be reduced to a one-dimensional problem by scaling considerations. In fact, the solution can be expressed by

$$\Gamma(P, I) = I z\left(\frac{P}{I}\right)$$

and $z(x)$ is the solution of the VI.

$$\begin{aligned} \frac{\partial z}{\partial x} x(\delta_I - \delta_P) + \frac{1}{2} \frac{\partial^2 z}{\partial x^2} x^2 (\sigma_P^2 + \sigma_I^2 - 2\rho\sigma_P\sigma_I) - z\delta_I &\leq \\ x - \delta_I z &\geq 0 \\ \text{product} &= 0. \end{aligned} \tag{7.3}$$

This VI can be easily solved. Considering the quadratic form

$$\beta(\delta_I - \delta_P) + \frac{1}{2}\beta(\beta - 1)(\sigma_P^2 + \sigma_I^2 - 2\rho\sigma_P\sigma_I) - \delta_I = 0$$

and its root $\beta_1 > 1$, we deduce a threshold value

$$x^* = \frac{\beta_1}{\beta_1 - 1} \delta_P$$

and the solution is given by

$$z = f_1 x^{\beta_1} - \frac{x}{\delta_P} + 1, \quad x \leq x^*$$

and

$$z = 0, \quad x \geq x^*.$$

The constant f_1 is such that there is continuity for $x = x^*$.

8. Uncertainties due to incentives

8.1. Setting of the model

We suppose in this model that the investment cost I can be reduced by an incentive to invest decided by government. However, the decision of the government is random. The incentive can be introduced or withdrawn according to a birth and death process. More precisely, the investment cost is a stochastic process given by the model

$$I(t) = I(1 - \theta\eta(t)),$$

where $\eta(t)$ is a stochastic process independent of $P(t)$. The process $\eta(t)$ can take two values, 1 or 0. It evolves as a Markov chain with

$$\text{Prob}(\eta(t + dt) = 1 | \eta(t) = 1) = 1 - \lambda_0 dt$$

$$\text{Prob}(\eta(t + dt) = 0 | \eta(t) = 0) = 1 - \lambda_1 dt.$$

The output flow is governed by

$$dP = P \left(\alpha dt + \sum_j \sigma_j dw_j \right)$$

and

$$\alpha = \mu - \delta,$$

where μ is as usual (assuming that P is tradable) the risk-adjusted rate of return. Since the value of the project does not depend on I , we still have

$$V(P) = \frac{P}{\delta}.$$

The value of the option depends on the process $\eta(t)$. So if $\eta(0) = \eta$, it is a function $F(P, \eta)$. Since η can take only two values, it is convenient to write

$$F(P, 0) = F^0(P), \quad F(P, 1) = F^1(P).$$

8.2. System of VI

The functions $F^0(P)$ and $F^1(P)$ are solutions of a system of coupled VI. To establish it, a convenient way is to use the probabilistic interpretation, recalling that the drift α must be replaced with $r - \delta$. We have

$$F(P, \eta) = \max_{\tau} E \left[\left(\frac{P(\tau)}{\delta} - I(1 - \theta\eta(t)) \right) \exp -r\tau \right].$$

To simplify the notation, define the differential operator on smooth functions $\phi(P)$ by

$$A\phi(P) = \phi'(P)P(r - \delta) + \frac{1}{2}\sigma^2 P^2 \phi''(P) - r\phi(P).$$

By the standard dynamic programming arguments, one obtains easily the system

$$\begin{aligned} AF^0(P) + \lambda_1(F^1(P) - F^0(P)) &\leq 0 \\ F^0(P) - \frac{P}{\delta} + I &\geq 0 \end{aligned} \tag{8.1}$$

product = 0

$$\begin{aligned} AF^1(P) + \lambda_0(F^0(P) - F^1(P)) &\leq 0 \\ F^1(P) - \frac{P}{\delta} + I(1 - \theta) &\geq 0 \end{aligned} \tag{8.2}$$

product = 0.

8.3. Solution of the system

The solution of the system is guided by intuition. There is a threshold of the output flow for which it makes sense to invest whether or not the incentive is in place. Similarly, there is a threshold below which one will not invest even when the incentive is in place.

In between, one will invest when the incentive is in place and will postpone decisions when the incentive is not in place. So, we look for two numbers $0 < P^1 < P^0$ such that

$$AF^0 + \lambda_1(F^1 - F^0) = 0$$

$$AF^1 + \lambda_0(F^0 - F^1) = 0$$

$$\forall P < P^1$$

$$AF^0 + \lambda_1(F^1 - F^0) = 0$$

$$F^1(P) = \frac{P}{\delta} - I(1 - \theta)$$

$$\forall P^1 < P < P^0$$

and finally

$$F^0(P) = \frac{P}{\delta} - I$$

$$F^1(P) = \frac{P}{\delta} - I(1 - \theta)$$

$$\forall P^0 < P.$$

One can check that three differential operators play a role in the solution. They are denoted as follows:

$$A_0\phi(P) = A\phi(P)$$

$$A_1\phi(P) = A\phi(P) - \lambda_1\phi(P)$$

$$A_2\phi(P) = A\phi(P) - (\lambda_1 + \lambda_0)\phi(P).$$

The notation is explained by the presence of zero λ term, one λ term, or two λ terms.

We associate to these differential operators second-order algebraic equations

$$\frac{1}{2}\sigma^2\beta(\beta - 1) + (r - \delta)\beta - r = 0$$

$$\frac{1}{2}\sigma^2\beta(\beta - 1) + (r - \delta)\beta - (r + \lambda_1) = 0$$

$$\frac{1}{2}\sigma^2\beta(\beta - 1) + (r - \delta)\beta - (r + \lambda_1 + \lambda_0) = 0,$$

whose roots are denoted as follows:

$$\beta(0)_1, \beta(0)_2; \quad \beta(1)_1, \beta(1)_2; \quad \beta(2)_1, \beta(2)_2,$$

where the second index is 1 for the root larger than 1, and 2 for the negative root.

EXERCISE 8.1. Show that one can write for $P < P^1$

$$F^1 = \frac{\lambda_0 \lambda_1 F_1^a P^{\beta(0)_1} + \lambda_0 F_1^s P^{\beta(2)_1}}{\lambda_0 + \lambda_1}$$

$$F^0 = \frac{\lambda_0 \lambda_1 F_1^a P^{\beta(0)_1} - \lambda_1 F_1^s P^{\beta(2)_1}}{\lambda_0 + \lambda_1}.$$

Next for $P^1 < P < P^0$, we have

$$F^1 = \frac{P}{\delta} - I(1 - \delta)$$

$$F^0 = F_1^0 P^{\beta(1)_1} + F_2^0 P^{\beta(1)_2} + \frac{\lambda_1}{\delta} \frac{P}{\delta + \lambda_1} - \lambda_1 \frac{I(1 - \theta)}{r + \lambda_1}.$$

For $P > P^0$, the values of $F^0(P)$ and $F^1(P)$ are known. We, thus, have six constants entering into the definition of $F^0 F^1$, namely, $F_1^a, F_1^s, F_1^0, F_2^0, P^0$, and P^1 . We can express six matching conditions. At P^1 we can write two conditions for F^0 and two for F^1 . Next, at P^0 we can write two conditions for F^0 . This defines completely the solution. It remains to show that this solution solves indeed the system of VI (8.1) and (8.2).

9. The option of abandonment

9.1. Setting of the model

In the preceding models, once the investment is decided the project is continued to its end and a value is collected. We have also considered the possibility of temporary suspension whenever $P < C$ and resuming the activity when $P > C$ with no penalty in both cases. This resulted in simply taking a profit flow function given by

$$\pi(P) = (P - C)^+.$$

We consider now the possibility of abandonment. This means we may decide to stop the project when it is active. This will entail a fixed cost denoted by E . If such a decision occurs, we are put back in the situation before the decision of investment was made. We may, thus, start again, but from scratch, paying the same investment cost I . So, there is no benefit from previous investment. Let us consider a stationary model. The option part (decision to invest) is the same as before, provided, of course, we have the right value function for the project, namely, we have (recalling the definition of the operator A)

$$\begin{aligned} AF(P) &\leq 0 \\ F(P) - V(P) + I &\geq 0 \\ AF(P)(F(P) - V(P) + I) &= 0. \end{aligned} \tag{9.1}$$

However, now the value $V(P)$ is not defined independently of F . Indeed, if one stops the project, then one goes back to the situation before investing. In addition, when the project runs, it faces a profit flow $P - C$, since the possibility of postponement with no

penalty is no longer present. It follows that $V(P)$ is the solution of

$$\begin{aligned} AV(P) + P - C &\leq 0 \\ V(P) - F(P) + E &\geq 0 \\ (AV(P) + P - C)(V(P) - F(P) + E) &= 0. \end{aligned} \tag{9.2}$$

9.2. Two-sided VI

Note first the compatibility condition

$$I + E \geq 0,$$

which is obvious whenever the two quantities I and E are positive. However, this leaves room for negative E . This possibility is useful whenever there is some cash recovered when the project is dismantled, which is a realistic situation. However, we must have $I + E > 0$ to avoid situations in which one could get cash but with continuous investment and disinvestment. The nice feature is that the function

$$\Gamma(P) = F(P) - V(P) + I$$

is still the solution to a single problem (no coupling). However, it is a two-sided VI and not a one-sided VI.

The problem is expressed as follows:

$$\begin{aligned} 0 &\leq \Gamma(P) \leq I + E \\ \text{if } 0 < \Gamma(P) < I + E, &\text{ then } A\Gamma(P) = P - C - rI \\ \text{if } \Gamma(P) = 0, &\text{ then } A\Gamma(P) \leq P - C - rI \\ \text{if } \Gamma(P) = I + E, &\text{ then } A\Gamma(P) \geq P - C - rI. \end{aligned} \tag{9.3}$$

Since $AC = -rC$, the two last conditions reduce to

$$\begin{aligned} \Gamma(P) = 0 &\Rightarrow P - C - rI > 0 \\ \Gamma(P) = I + E &\Rightarrow P - C + rE < 0. \end{aligned}$$

9.3. Solution of the two-sided VI

The two last conditions guide the intuition. For a sufficiently low output P , one should not only not invest but also cut the investment. For a sufficiently large output one should not only continue the project but also invest if the project has not yet started. In between, one should continue a project already started but one should not invest. Therefore, one looks for two thresholds $0 < P_L < P_H$ such that

$$A\Gamma(P) = P - C - rI, \forall P_L < P < P_H$$

and

$$\Gamma(P) = 0, \forall P > P_H$$

$$\Gamma(P) = I + E, \forall P < P_L.$$

Clearly, we must have

$$P_H \geq C + rI, \quad P_L \leq C - rE.$$

We will have in the interval (P_L, P_H)

$$\Gamma(P) = \Gamma_1 P^{\beta_1} + \Gamma_2 P^{\beta_2} - \frac{P}{\delta} + \frac{C}{r} + I,$$

and we have four constants to obtain, Γ_1 , Γ_2 and P_H , P_L . There are also four conditions to express the continuity of $\Gamma(P)$ and its derivative at points P_L and P_H .

We obtain the following system

$$\Gamma_1 P_L^{\beta_1} + \Gamma_2 P_L^{\beta_2} - \frac{P_L}{\delta} + \frac{C}{r} = E$$

$$\Gamma_1 P_H^{\beta_1} + \Gamma_2 P_H^{\beta_2} - \frac{P_H}{\delta} + \frac{C}{r} + I = 0$$

$$\Gamma_1 \beta_1 P_L^{\beta_1-1} + \Gamma_2 \beta_2 P_L^{\beta_2-1} - \frac{1}{\delta} = 0$$

$$\Gamma_1 \beta_1 P_H^{\beta_1-1} + \Gamma_2 \beta_2 P_H^{\beta_2-1} - \frac{1}{\delta} = 0.$$

We derive the following system for P_H and P_L

$$\frac{\frac{1}{\delta}(1 - \beta_2) + \frac{\beta_2}{P_L} \left(\frac{C}{r} - E \right)}{P_L^{\beta_1-1}} = \frac{\frac{1}{\delta}(1 - \beta_2) + \frac{\beta_2}{P_H} \left(\frac{C}{r} + I \right)}{P_H^{\beta_1-1}}$$

$$\frac{\frac{1}{\delta}(\beta_1 - 1) - \frac{\beta_1}{P_L} \left(\frac{C}{r} - E \right)}{P_L^{\beta_2-1}} = \frac{\frac{1}{\delta}(\beta_1 - 1) + \frac{\beta_2}{P_H} \left(\frac{C}{r} + I \right)}{P_H^{\beta_1-1}}$$

It can be shown that this system has a unique solution with $P_L < P_H$ (see DIXIT [1989]).

10. The option of mothballing

10.1. Description of the model

We introduce a new possibility for an active project, that of mothballing instead of abandoning. From a situation of mothballing, a project can be reactivated or abandoned. In a situation of mothballing, the profit flow is lost and in addition a maintenance cost M must be paid. It is smaller than the operating cost C . To put a project in a situation of

mothballing incurs a fixed cost E_M . To reactive from mothballing implies a fixed cost E_R . Finally, to abandon a project from mothballing represents a cost E_S . The quantity

$$E = E_S + E_M$$

represents the fixed cost of abandoning an active project. In writing the VI for V , we will need a new function $H(P)$, which represents the value of the option of mothballing. So, in fact, we will have a coupled system for three functions F , V , and H .

10.2. Variational inequalities

We consider the stationary case. The value of the option to invest is governed by the VI.

$$\begin{aligned} AF(P) &\leq 0 \\ F(P) - V(P) + I &\geq 0 \\ AF(P)(F(P) - V(P) + I) &= 0. \end{aligned} \tag{10.1}$$

The problem of which $V(P)$ is a solution is now given by

$$\begin{aligned} AV(P) + P - C &\leq 0 \\ V(P) &\geq H(P) - E_M \\ (AV(P) + P - C)(V(P) - H(P) + E_M) &= 0. \end{aligned} \tag{10.2}$$

In (10.2), the function $H(P)$ represents the value of mothballing. It is itself governed by the following VI

$$\begin{aligned} AH(P) - M &\leq 0 \\ H(P) &\geq \max(V(P) - E_R, F - E_S) \\ (AH(P) - M)[H(P) - \max(V(P) - E_R, F - E_S)] &= 0. \end{aligned} \tag{10.3}$$

The second inequality expresses the fact that if mothballing is stopped, it is to go back to a state of active project or abandon, in which case one is back in the situation preceding investing.

10.3. Solution of the VI

To define the solution, we need four thresholds

$$0 < P_S < P_M < P_R < P_H,$$

which trigger the following situations

$$\text{for } P > P_H, F(P) = V(P) - I$$

$$\text{for } P > P_R, H(P) = V(P) - E_R$$

$$\text{for } P < P_M, V(P) = H(P) - E_M$$

$$\text{for } P < P_S, H(P) = F(P) - E_S.$$

As a consequence, we have

$$\text{for } P < P_S, V(P) = F(P) - E.$$

So below P_S , an active project abandons without going to the mothballing step. So P_S and P_H correspond to the thresholds P_L and P_H defined when there was no option of mothballing.

We can next define completely the three functions by solving the differential equations in the respective intervals. We can state

$$F(P) = F_1 P^{\beta_1} \text{ for } P < P_H$$

$$F(P) = V(P) - I \text{ for } P > P_H$$

$$V(P) = V_2 P^{\beta_2} + \frac{P}{\delta} - \frac{C}{r} \text{ for } P > P_M$$

$$V(P) = H(P) - E_M \text{ for } P < P_M$$

$$H(P) = H_1 P^{\beta_1} + H_2 P^{\beta_2} - \frac{M}{r} \text{ for } P_S < P < P_R$$

$$H(P) = F(P) - E_S \text{ for } P < P_S$$

$$H(P) = V(P) - E_R \text{ for } P > P_R.$$

The solution depends on eight constants, the four thresholds and the values of constants F_1 , G_2 , H_1 , and H_2 . We write eight matching conditions (two per threshold). There is some decoupling. The values $H_1, G_2 - H_2, P_R, P_M$ are solutions of the system

$$-H_1 P_R^{\beta_1} + (G_2 - H_2) P_R^{\beta_2} + \frac{P_R}{\delta} - \frac{C - M}{r} - E_R = 0$$

$$-H_1 \beta_1 P_R^{\beta_1 - 1} + (G_2 - H_2) \beta_2 P_R^{\beta_2 - 1} + \frac{1}{\delta} = 0$$

$$-H_1 P_M^{\beta_1} + (G_2 - H_2) P_M^{\beta_2} + \frac{P_M}{\delta} - \frac{C - M}{r} + E_M = 0$$

$$-H_1 \beta_1 P_M^{\beta_1 - 1} + (G_2 - H_2) \beta_2 P_M^{\beta_2 - 1} + \frac{1}{\delta} = 0.$$

Similarly, we define $F_1 - H_1, H_2, P_S, P_H$ by the system

$$(F_1 - H_1)P_S^{\beta_1} - H_2P_S^{\beta_2} + \frac{M}{r} - E_S = 0$$

$$(F_1 - H_1)\beta_1 P_S^{\beta_1-1} - H_2\beta_2 P_S^{\beta_2-1} = 0$$

$$F_1P_H^{\beta_1} - G_2P_H^{\beta_2} - \frac{P_H}{\delta} + \frac{C}{r} + I = 0$$

$$F_1\beta_1 P_H^{\beta_1-1} - G_2\beta_2 P_H^{\beta_2-1} - \frac{1}{\delta} = 0.$$

11. Reflecting barriers

11.1. Barrier at high output

We want to prevent the output $P(t)$ from going beyond a barrier $\bar{P}$. To justify this model, we refer to the fact that at high values of the output, the project is submitted to competing projects from other firms. At the barrier, we assume that sufficiently many firms will compete preventing the output value to go beyond the barrier. The stochastic process that models the output is the following

$$dP = P(\alpha dt + \sum_j \sigma_j dw_j(t)) - d\xi(t).$$

We have, in fact, a pair $P(t)$ and $\xi(t)$ uniquely defined by the following properties: $\xi(t)$ is a continuous process adapted to the filtration generated by the Wiener process, and $P(t)$ is continuous adapted and has the Ito differential written above. Moreover, $\xi(t)$ is nondecreasing and does not increase when $P(t) < \bar{P}$.

We assume that the profit flow is the output value itself P . We want to define the value $V(P)$ of an active project, for $P \leq \bar{P}$. For $P < \bar{P}$, the evolution of $P(t)$ is the same as the nonreflected process, some by standard arguments we shall have

$$AV(P) + P = 0, \forall P < \bar{P}.$$

We need a boundary condition at $\bar{P}$. However, at $\bar{P}$ the process is instantaneously reflected so

$$V(\bar{P}) \simeq V(\bar{P} - \epsilon).$$

Since ϵ is arbitrarily small, we must have

$$V'(\bar{P}).$$

EXERCISE 11.1. *Show that the solution of*

$$AV(P) + P = 0, \forall P < \bar{P}$$

$$V(0) = 0, \quad V'(\bar{P}) = 0$$

is given by

$$V(P) = \frac{P}{\delta} - \frac{1}{\delta\beta_1} P^{\beta_1} \bar{P}^{1-\beta_1}. \quad (11.1)$$

We see, in particular, that

$$V(\bar{P}) = \frac{\bar{P}}{\delta} \frac{\beta_1 - 1}{\beta_1}. \quad (11.2)$$

We now turn to the value $F(P)$ of the option to invest in the project. At $\bar{P}$, it makes no sense to wait since the output will decline certainly in the near future. So, we must have

$$F(\bar{P}) = V(\bar{P}) - I.$$

For $P < \bar{P}$ we are in the same situation as without barrier. Therefore, $F(P)$ is the solution to the problem

$$\begin{aligned} AF(P) &\leq 0, \quad F(P) - V(P) + I \geq 0 \\ AF(P)(F(P) - V(P) + I) &= 0, \quad \forall P < \bar{P} \\ F(\bar{P}) &= V(\bar{P}) - I. \end{aligned} \quad (11.3)$$

EXERCISE 11.2. Assume that

$$P^* = \frac{\beta_1}{\beta_1 - 1} \delta I < \bar{P},$$

then show that

$$\begin{aligned} F(P) &= F_1 P^{\beta_1}, \quad P < P^* \\ F(P) &= V(P) - I, \quad P^* \leq P \leq \bar{P} \\ F_1 &= \frac{1}{\delta\beta_1} (P^{*1-\beta_1} - \bar{P}^{1-\beta_1}). \end{aligned} \quad (11.4)$$

We see that $F_1 > 0$, hence the value of the option is positive for $P \leq P^*$.

Can we reach an equilibrium at $\bar{P}$? We need to have indifference between the interest to invest or to stay out. This means that

$$V(\bar{P}) = I.$$

Recalling the formula for $V(\bar{P})$, we check easily that

$$\bar{P} = P^*.$$

This implies $F_1 = 0$, and the value of the option is always 0.

11.2. Case of two barriers

The model with one barrier can be generalized to two barriers, $\underline{P}$ and $\bar{P}$, with a process $P(t)$ governed by

$$dP = P(\alpha dt + \sum_j \sigma_j dw_j(t)) - d\xi(t) + d\eta(t),$$

where $P(t)$, $\xi(t)$, and $\eta(t)$ are adapted continuous processes, $\xi(t)$ and $\eta(t)$ are nondecreasing, and

$$\begin{aligned} d\xi(t) &= 0, \text{ if } P(t) < \bar{P} \\ d\eta(t) &= 0, \text{ if } P(t) > \underline{P}. \end{aligned}$$

These conditions determine a unique triple $P(t)$, $\xi(t)$, and $\eta(t)$. The logic of the lower barrier is the same as the upper barrier. The output $P(t)$ is sufficiently low and competing projects will be naturally withdrawn.

We adapt the language for the upper barrier to the lower one. We now have a system $F(P)$, $V(P)$ solution of

$$\begin{aligned} AF(P) &\leq 0, \quad F(P) - V(P) + I \geq 0 \\ AF(P)(F(P) - V(P) + I) &= 0, \quad \forall P < \bar{P} \\ F'(\underline{P}) &= 0, \quad F(\bar{P}) = V(\bar{P}) - I \\ AV(P) + P - C &\leq 0 \\ V(P) - F(P) + E &\geq 0 \\ (AV(P) + P - C)(V(P) - F(P) + E) &= 0 \\ V(\underline{P}) &= F(\underline{P}) - E, \quad V'(\bar{P}) = 0, \end{aligned} \tag{11.5}$$

where E represents the fixed cost of abandon.

We need two thresholds with

$$\underline{P} < P_L < P_H < \bar{P},$$

and we have the formulas

$$\begin{aligned} V(P) &= V_1 P^{\beta_1} + V_2 P^{\beta_2} + \frac{P}{\delta} - \frac{C}{r}, \quad \text{for } P_L < P < \bar{P} \\ F(P) &= F_1 P^{\beta_1} + F_2 P^{\beta_2}, \quad \text{for } \underline{P} < P < P_H. \end{aligned}$$

So, we have six constants

$$P_L, P_H, F_1, F_2, V_1, V_2$$

to define. We can write six conditions

$$F'(\underline{P}) = 0, \quad F(P_H) = V(P_H) - I, \quad F'(P_H) = V'(P_H)$$

$$V'(\bar{P}) = 0, \quad V(P_L) = F(P_L) - E, \quad V'(P_L) = F'(P_L).$$

EXERCISE 11.3. Show that $P_H, P_L, V_1 - F_1, V_2 - F_2$ can be defined by the system

$$\begin{aligned} (V_1 - F_1)P_H^{\beta_1} + (V_2 - F_2)P_H^{\beta_2} + \frac{P_H}{\delta} - \frac{C}{r} &= I \\ (V_1 - F_1)\beta_1 P_H^{\beta_1-1} + (V_2 - F_2)\beta_2 P_H^{\beta_2-1} + \frac{1}{\delta} &= 0 \\ (V_1 - F_1)P_L^{\beta_1} + (V_2 - F_2)P_L^{\beta_2} + \frac{P_L}{\delta} - \frac{C}{r} &= -E \\ (V_1 - F_1)\beta_1 P_L^{\beta_1-1} + (V_2 - F_2)\beta_2 P_L^{\beta_2-1} + \frac{1}{\delta} &= 0. \end{aligned} \tag{11.6}$$

We complete with the relations

$$\begin{aligned} F_1\beta_1 \underline{P}^{\beta_1-1} + F_2\beta_2 \underline{P}^{\beta_2-1} &= 0 \\ V_1\beta_1 \bar{P}^{\beta_1-1} + V_2\beta_2 \bar{P}^{\beta_2-1} &= 0. \end{aligned} \tag{11.7}$$

Note that P_H and P_L do not depend on $\underline{P}$ and $\bar{P}$, provided the values are in the interval $(\underline{P}, \bar{P})$

We will define equilibrium in this setup by the condition

$$P_L = \underline{P}, \quad P_H = \bar{P},$$

arguing that at equilibrium the general thresholds and those for an individual project coincide. We, then, have

$$F'(\underline{P}) = F'(\bar{P}) = V'(\bar{P}) = 0,$$

from which one obtains

$$F_1 = F_2 = 0;$$

therefore,

$$F(P) = 0, \quad \forall P.$$

There is no value in the option of abandonment. Moreover,

$$F_1 = F_2 = 0;$$

therefore,

$$V(\underline{P}) = -E, \quad V(\bar{P}) = I.$$

12. Equilibrium model

12.1. General description

From a better economic interpretation, we replace projects by firms. Instead of abandoning a project, we will speak of firms becoming idle, and instead of investing in a project, we will speak of firms becoming active. We consider an economy in which N firms per unit of time enter in the market at any time. In an equilibrium, there will be permanently Q firms that are active. However, in this model, when a firm enters in the market, it does not become immediately active. It can wait sometime before becoming active, for which it will have to pay the usual fee I . To explain this fact, we assume that the output value is given by

$$P(t) = X(t)D(Q).$$

Here, $D(Q)$ represents the demand, which depends on the number of active firms. It is a given function. The process $X(t)$ describes the random component of the output value.

The random component is modeled by

$$dX = X(\alpha dt + \sigma dw(t)).$$

There is, in addition, an initial value ξ , which is also random and independent of $w(t)$. We assume that it has a probability density $g(x)$. To enter into the market means that a firm will learn about the initial value of its random component. It has to pay a price R for this information. After it has observed this initial value, it may decide to become active or to wait till a better value of the output. So, in permanence, there will be Q active firms and M waiting firms. The numbers N , Q , M are the unknowns of a possible equilibrium. However, if firms live forever, there cannot be any equilibrium. So, we assume that there is an exogenous death process. The lifetime is, thus, a random variable T , whose distribution is exponential with rate λ . Moreover, T is assumed to be independent of ξ and $w(t)$. In the following analysis, we will not introduce risk-adjusted expected return to simplify the presentation and firms are risk neutral.

12.2. Value of options

A firm entering in the market pays the fee R to observe the initial value ξ of its random component. During its lifetime, it starts by waiting for some time before it becomes active, for which an extra I has to be paid. The external death process applies in the same way to idle and active firms. The value of an active firm, knowing that its initial $\xi = x$, is given by

$$V(x, Q) = E \left[\int_0^T X(t)D(Q) \exp -rt dt | X(0) = x \right],$$

where we have emphasized the dependence in Q . It is easy to check that

$$V(x, Q) = \frac{x D(Q)}{r + \lambda - \alpha}.$$

The value of an option to become active is given by

$$F(x, Q) = \max_{\tau} E[V(X(\tau)) \mathbb{I}_{\tau < T} \exp -r\tau | X(0) = x],$$

where τ is any stopping time. The probabilistic interpretation is the easiest way to F . However, for computations, we refer to the VI formulation.

Let us define the differential operator

$$A\phi(x) = x\alpha\phi'(x) + \frac{1}{2}x^2\sigma^2\phi''(x) - (r + \lambda)\phi(x).$$

Then, $F(x, Q)$ is the solution of

$$AF(x, Q) \leq 0$$

$$F(x, Q) - V(x, Q) + I \geq 0$$

$$AF(x, Q)(F(x, Q) - V(x, Q) + I) = 0.$$

There will be a threshold $x^*(Q)$ given explicitly by

$$x^* = \frac{\beta_1 I(r + \lambda - \alpha)}{(\beta_1 - 1)D(Q)}$$

and

$$F(x, Q) = F_1(Q)x^{\beta_1}, \quad \text{for } x \leq x^*$$

$$F(x, Q) = V(x, Q) - I, \quad \text{for } x \geq x^*.$$

12.3. Choice of Q

We complete the definition of $F(x, Q)$ by noting that β_1 is the root larger than 1 of the second-order equation

$$\beta\alpha + \frac{\sigma^2}{2}\beta(\beta - 1) - (r + \alpha) = 0$$

and

$$F_1(Q) = \left(\frac{D(Q)}{(r + \lambda - \alpha)\beta_1} \right)^{\beta_1} \left(\frac{\beta_1 - 1}{I} \right)^{\beta_1 - 1}.$$

Having defined the function $F(x, Q)$, the value of Q is given by the equation

$$EF(\xi, Q) = \int_0^\infty g(x)F(x, Q)dx.$$

12.4. Distribution of firms

For the sequel, it is convenient to work with the stochastic process

$$Y(t) = \log X(t),$$

which is the solution of

$$dY = vdt + \sigma dw,$$

with

$$v = \alpha - \frac{1}{2}\sigma^2.$$

We set

$$y^* = \log x^*$$

and

$$Y(0) = \eta = \log \xi,$$

where the probability distribution of η is $h(x)$.

Let us consider again that the process has a lifetime T with probability density an exponential with rate λ . Define also

$$\tau = \{\inf t | Y(t) \geq y^*\}$$

We are interested in the following probability distribution

$$\psi(y, t)dy = \text{Prob} \{y < Y(t) < y + dy \text{ and } t < T \wedge \tau\}, \text{ for } y < y^*.$$

This function is the solution of the Kolmogorov equation

$$\frac{\partial \psi}{\partial t} + v \frac{\partial \psi}{\partial y} - \frac{1}{2} \frac{\partial^2 \psi}{\partial y^2} + \lambda \psi = 0$$

$$\psi(y^*, t) = 0, \quad \psi(-\infty, t) = 0$$

$$\psi(y, 0) = h(y).$$

The boundary conditions are clear. To derive the PDE, one uses the following relation for $0 < y < y^*$, $t > 0$

$$\psi(y, t) \simeq E\psi(y - vdt - \sigma dw, t - dt)(1 - \lambda dt)$$

and expands the right-hand side. Now, the number of firms arrived between s and $s + ds$ for which $y < Y(t) < y + dy$ is, thus, $N\psi(y, t - s)ds$. It follows that

$$\text{number of firms for which } y < Y(t) < y + dt = N\phi_t(y) = N \int_0^t \psi(y, s)ds.$$

At equilibrium, firms are in fixed positions. For those waiting, we can state that

$$\text{number of firms waiting at equilibrium in position } y = N\phi(y) = N \int_0^\infty \psi(y, s) ds.$$

We obtain the differential equation

$$\begin{aligned} \nu \frac{\partial \phi}{\partial y} - \frac{1}{2} \frac{\partial^2 \phi}{\partial y^2} + \lambda \phi &= h(y) \\ \phi(y^*) &= \phi(-\infty) = 0. \end{aligned} \quad (12.1)$$

12.5. Number of new entrants at equilibrium

We want to obtain the value of N , which sustains an equilibrium. Let us first check that the rate of activation, which is defined by

$$\frac{\text{Number of firms becoming active in an interval } dt}{dt} = -\frac{N}{2} \phi'(y^*) \sigma^2.$$

An easy way to obtain this result is to consider the binomial approximation of the Wiener process. From a position y , the particle moves during the interval dt to $y + \sigma\sqrt{dt}$ with probability $\frac{1}{2}(1 + \frac{\nu}{\sigma}\sqrt{dt})$ and to position $y - \sigma\sqrt{dt}$ with probability $\frac{1}{2}(1 - \frac{\nu}{\sigma}\sqrt{dt})$.

It follows that the number of new active firms during the interval dt is approximately

$$N\phi(y^* - \sigma\sqrt{dt})(1 - \lambda dt) \frac{1}{2}(1 + \frac{\nu}{\sigma}\sqrt{dt}),$$

which is approximately

$$-\frac{N}{2} \phi'(y^*) \sigma^2 dt,$$

which implies the result. We deduce that the number of new active firms in the interval dt is

$$-\frac{N}{2} \phi'(y^*) \sigma^2 dt + N(1 - H(y^*))dt,$$

where $H(y)$ is the cumulative distribution function corresponding to the density $h(y)$. The additional term above corresponds to the number of firms, which are immediately active once they enter in the market.

In order to keep the number of active firms fixed equal to Q , the number of new activated firms during dt must coincide with the number that disappear during dt , which is $Q\lambda dt$. Therefore, we obtain the relation

$$Q\lambda = -\frac{N}{2} \phi'(y^*) \sigma^2 + N(1 - H(y^*)),$$

which defines the number N of new entrants per unit of time.

References

- BENSOUSSAN, A., LIONS, J.L. (1982). *Applications of Variational Inequalities in Stochastic Control* (North Holland, Amsterdam, The Netherlands).
- DIXIT, A.K. (1989). Entry and exit decisions under uncertainty. *J. Polit. Econ.* **97**, 620–638.
- DIXIT, A.K., PINDYCK, R.S. (1994). *Investment Under Uncertainty* (Princeton University Press, Princeton, NJ).

Anticipative Stochastic Control for Lévy Processes With Application to Insider Trading

Agnès Sulem

INRIA, Domaine de Voluceau, Rocquencourt, B.P.105, F-78153 Le Chesnay Cedex, France.

E-mail address: agnes.sulem@inria.fr

Arturo Kohatsu-Higa

INRIA, Domaine de Voluceau, Rocquencourt, B.P.105, F-78153 Le Chesnay Cedex, France.

E-mail address: arturo.kohatsu@inria.fr

Bernt Øksendal

Centre of Mathematics for Applications (CMA) and Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway.

E-mail address: giulian@math.uio.no

Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway.

E-mail address: oksendal@math.uio.no

Frank Proske

Centre of Mathematics for Applications (CMA) and Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway.

E-mail address: proske@math.uio.no

Giulia Di Nunno

Centre of Mathematics for Applications (CMA) and Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway.

E-mail address: giulian@math.uio.no

Thilo Meyer-Brandis

*Centre of Mathematics for Applications (CMA) and Department of Mathematics,
University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway.
E-mail address: meyerbr@math.uio.no*

Abstract

An insider is an agent who has access to larger information than the one given by the development of the market events and who takes advantage of this in optimizing his position in the market. In this chapter, we consider the optimization problem of an insider who is so influential in the market to affect the price dynamics: in this sense, he is called a “large” insider. The optimal portfolio problem for a general utility function is studied for a financial market driven by a Lévy process in the framework of forward anticipating calculus.

1. Introduction

The modeling of insider trading is a challenge that recently has been taken up by many scientists with the aim of understanding the behavior and quantifying the gain of a trader who takes advantage of some extra information, that is, not deducible from the market behavior itself, that he may happen to have at his disposal.

Thus, in a market model on the probability space $(\Omega, \mathcal{F}, P)$ with two investment possibilities such as

- a bond with price $S_0(t)$, $t \in [0, T]$,
- a stock with price $S_1(t)$, $t \in [0, T]$,

an “honest” agent is taking decisions relying only on the flow of information

$$\mathbb{F} := \{\mathcal{F}_t \subset \mathcal{F}, 0 \leq t \leq T\}$$

given by the development of the market events, while an “insider” would rely on the flow of information

$$\mathbb{H} := \{\mathcal{H}_t \subset \mathcal{F}, 0 \leq t \leq T\} : \quad \mathcal{H}_t \supset \mathcal{F}_t.$$

Therefore, the insider’s portfolios are in general stochastic processes adapted to $\mathbb{H}$.

Different aspects of the insider trading have been considered and with different approaches. It is rather hard to mention all past and recent achievements, so we will restrict ourselves to the papers that have mostly inspired the present work.

The subject we are dealing with is the optimization problem

$$\max_{\pi \in \mathcal{A}} E[U(X_\pi(T))] \tag{1.1}$$

of an insider who wants to maximize the expected utility of his final wealth $X_\pi(T)$ given by the dynamics

$$dX_\pi(t) = (1 - \pi(t))X_\pi(t)dS_0(t) + \pi(t)X_\pi(t)dS_1(t), \quad X_\pi(0) > 0,$$

over all admissible choices of portfolios $\pi \in \mathcal{A}$ (see Section 3).

Optimization problems of this kind have been studied widely. Here, we mention the pivotal work of KARATZAS and PIKOVSKY [1996]. They were considering the problem (1.1) for a market driven by a Brownian motion and a logarithmic utility function in the framework of classical enlargement of filtrations. This framework applies under the a priori assumption that the $\mathbb{F}$ -adapted Brownian motion driving the market is a semimartingale with respect to $\mathbb{H}$. This assumption is often difficult, if not impossible, to be verified since it depends on the kind of information $\mathbb{H}$ available to the insider.

In a study by BIAGINI and ØKSENDAL [2005], a general approach is suggested to the modeling of insider trading that overcomes the need of the above assumption in the framework of forward anticipating calculus. In this setting, the authors give a solution to problem (1.1) for a general utility function. However, they restrict themselves to the case of markets driven by Brownian motion only.

REMARK. The reasons for taking this approach into account can be summarized in the following points:

1. The forward integral provides the natural interpretation of the gains from the trade process. Indeed, suppose that a trader buys one stock at a random time τ_1 and keeps it until the random time $\tau_2 > \tau_1$. When he sells it, the gain obtained is $S_1(\tau_2) - S_1(\tau_1) = \int 1_{(\tau_1, \tau_2]}(t) d^- S_1(t)$, where the integral is a forward stochastic integral.
2. If the integrand is càglàd (i.e., left continuous and with right-sided limits), the forward integral may be regarded as the limit of Riemann sums (see BIAGINI and ØKSENDAL [2005] and KOHATSU-HIGA and SULEM [2006]).
3. If the stochastic process driving the market happens to be a semimartingale with respect to the insider filtration $\mathbb{H}$, then the corresponding stochastic integral coincides with the forward integral.

In a study by ØKSENDAL and SULEM [2004], the forward integral and anticipative calculus are used to study the optimal portfolio problem with logarithmic utility for a trader with partial information in a (Lévy–Brownian type) anticipative market (e.g., a market influenced by insiders).

The study of ØKSENDAL and SULEM [2004] is extended by KOHATSU-HIGA and SULEM [2006] to cover the case when there are no a priori assumptions about the relation between the information available to the trader and the information generated by the possibly anticipative market. Here, the market is assumed to be driven by Brownian motion and the utility function is logarithmic.

DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005] and [2006] extended the forward integration to the case of compensated Poisson random measures and thus to more general Lévy processes and solve problem (1.1) in the case of a logarithmic utility function. This extension of framework to Lévy processes is motivated by the ongoing discussion on the better fitting of these models to real financial markets than the ones driven only by Brownian motion. Here, we can refer, for example, to the studies by BARNDORFF-NIELSEN [1998], CONT and TANKOV [2004], EBERLEIN and RAIBLE [1999], and SCHOUTENS [2003].

In the same line of BIAGINI and ØKSENDAL [2005] and relying on the achievements DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005] and DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2006], we now solve problem (1.1) for a general utility function and for a general Lévy process. This represents the major contribution of this chapter.

Besides, there is also another element of novelty. In fact, inspired by the studies of CUOCO and CVITANIC [1998] and KOHATSU-HIGA and SULEM [2006], we consider the problem (1.1) from the point of view of a trader so influential in the market that his decisions affect the price process dynamics. In this sense, our dealer is called “large” trader. In a study by CUOCO and CVITANIC [1998], the effect of the trader’s positions on the prices is exogenously specified. In this chapter, we chose to use a similar approach (see Eqs. (3.1)–(3.2)). This visible effect of a large trader on the price dynamics may arise because of the volumes traded or also because the other market investors may suppose, though without certainty, that the large trader is an insider. Note that actually in the study by CUOCO and CVITANIC [1998], the large trader is not an insider. On the other hand, the study by KOHATSU-HIGA and SULEM [2006] considers a similar model for prices but extends the analysis to the cases in which the large trader is truly an insider. The analysis by KOHATSU-HIGA and SULEM [2006], is however, restricted to the case of logarithmic utility and Brownian-motion-driven dynamics.

In this chapter, as said, the major concern is the solution to an optimal portfolio problem from the point of view of a “large insider,” and we do not attempt to discuss the price formation here. This would require a study of equilibria under asymmetric information. For this, we can refer to the seminal articles by KYLE [1985], BACK [1992], and the recent literature in this line.

This chapter is organized as follows. In Section 2, we recall the basic tools of forward calculus for Lévy processes and in particular the Itô formula (see Theorem 2.1), which are then applied in Section 3 where criteria for the existence of the solution of the “large” insider’s portfolio optimization problem (1.1) are given. In Section 4, some examples are considered.

For related works in the context of insider modeling and portfolio optimization see ELLIOTT and JEANBLANC [1998], ELLIOTT, GEMAN and KORKIE [1997], KOHATSU-HIGA and SULEM [2006], KOHATSU-HIGA and YAMAZATO [2004], KUNITA [2004], and ØKSENDAL [2006].

We emphasize that our study is mainly intended as a survey. Thus, we have left out some technical details in some of the proofs.

2. Framework: forward anticipating calculus

In this section, we briefly recall some properties of the forward integral. We can refer to the studies by BIAGINI and ØKSENDAL [2005], NUALART and PARDOUX [1988], RUSSO and VALLOIS [1993], RUSSO and VALLOIS [1995], and RUSSO and VALLOIS [2000] for information on the forward integration with respect to the Brownian motion and to the study by DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005] for the integration with respect to the compensated Poisson random measure.

As mentioned in the introduction, we are interested in a Lévy process

$$\eta(t) = \sigma B(t) + \int_0^t \int_{\mathbb{R}_0} z \tilde{N}(ds, dz), \quad t \in [0, T], \quad (2.1)$$

on the complete filtered probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{F} = \{\mathcal{F}_t \subset \mathcal{F}, 0 \leq t \leq T\}$ ($\mathcal{F}_0$ trivial up to events of measure 0) with a finite time horizon $T > 0$. In the Itô representation (2.1) (Itô [1978]) of the Lévy process, we can distinguish the standard Brownian motion $B(t)$, $t \in [0, T]$ ($B(0) = 0$), the constant $\sigma \in \mathbb{R}$, and the compensated Poisson random measure

$$\tilde{N}(dt, dz) = N(dt, dz) - \nu(dz)dt.$$

Here, $\nu(dz)$, $z \in \mathbb{R}_0$, is a σ -finite Borel measure on $\mathbb{R}_0 = (-\infty, 0) \cup (0, \infty)$, such that

$$\int_{\mathbb{R}_0} z^2 \nu(dz) < \infty.$$

Then, $E[\eta^2(t)] = t(\sigma^2 + \int_{\mathbb{R}_0} z^2 \nu(dz)) < \infty$ for all $t \in [0, T]$. For more information on Lévy processes, we can refer to the studies by APPLEBAUM [2004], BERTOIN [1996], PROTTER [2003] and SATO [1999].

The following definition is by RUSSO and VALLOIS [1995].

DEFINITION 2.1. We say that the (measurable) stochastic process $\varphi = \varphi(t)$, $t \in [0, T]$, is *forward integrable* over the interval $[0, T]$ with respect to the Brownian motion if there exists a process $I(t)$, $t \in [0, T]$, such that

$$\sup_{t \in [0, T]} \left| \int_0^t \varphi(s) \frac{B(s + \varepsilon) - B(s)}{\varepsilon} ds - I(t) \right| \longrightarrow 0, \quad \varepsilon \rightarrow 0, \quad (2.2)$$

in probability. Then, for any $t \in [0, T]$,

$$I(t) = \int_0^t \varphi(s) d^- B(s)$$

is called the *forward integral* of φ with respect to the Brownian motion on $[0, t]$.

The corresponding definition of forward integral with respect to the compensated Poisson random measure is given by DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005]. Here, a modified version of what is suggested by DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005] is actually given to be in the line with the definition suggested by RUSSO and VALLOIS [1995]. Note that these definitions are such that the Itô formulae for forward integrals with respect to the Brownian motion and the compensated Poisson random measure hold true (see RUSSO and VALLOIS [1995] and DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005]).

DEFINITION 2.2. We say that the (measurable) random field $\psi = \psi(t, z)$, $t \in [0, T]$, $z \in \mathbb{R}_0$, is *forward integrable* over $[0, T]$ with respect to the compensated Poisson random measure if there exists a process $J(t)$, $t \in [0, T]$, such that

$$\sup_{t \in [0, T]} \left| \int_0^t \int_{\mathbb{R}_0} \psi(s, z) 1_{U_n}(z) \tilde{N}(ds, dz) - J(t) \right| \rightarrow 0, \quad n \rightarrow \infty, \quad (2.3)$$

in probability. Here, U_n , $n = 1, 2, \dots$, is an increasing sequence of compact sets $U_n \subseteq \mathbb{R}_0$ with $\nu(U_n) < \infty$, such that $\bigcup_n U_n = \mathbb{R}_0$. Then, for any $t \in [0, T]$,

$$J(t) = \int_0^t \int_{\mathbb{R}_0} \psi(s, z) \tilde{N}(d^-s, dz)$$

is called the *forward integral* of ψ with respect to the compensated Poisson random measure on $[0, t]$.

REMARK 2.1. (1) If the integrands in the above definitions are adapted to the filtration $\mathbb{F}$, then the limits (2.1) and (2.2) coincide with the Itô integral. In particular, if we consider the stronger convergence in $L_2(P)$ in the above definitions, we obtain an extension of the classical Itô integral. This is useful for the forthcoming applications and is the case we take into account in the sequel.

(2) If G is a random variable, then

$$\begin{aligned} G \cdot \left[\int_0^T \varphi(t) d^-B(t) + \int_0^T \int_{\mathbb{R}_0} \psi(t, z) \tilde{N}(d^-t, dz) \right] \\ = \int_0^T G \varphi(t) d^-B(t) + \int_0^T \int_{\mathbb{R}_0} G \psi(t, z) \tilde{N}(d^-t, dz). \end{aligned} \quad (2.4)$$

Note that this property does not hold in general for the Itô integral.

DEFINITION 2.3. A *forward process* is a measurable stochastic function $X(t) = X(\omega, t)$, $\omega \in \Omega$, $t \in [0, T]$, that admits the representation

$$X(t) = x + \int_0^t \alpha(s) ds + \int_0^t \varphi(s) d^-B(s) + \int_0^t \int_{\mathbb{R}_0} \psi(s, z) \tilde{N}(d^-s, dz), \quad (2.5)$$

where $x = X(0)$ is a constant. A shorthand notation for (2.5) is

$$d^-X(t) = \alpha(t) dt + \varphi(t) d^-B(t) + \int_{\mathbb{R}_0} \psi(t, z) \tilde{N}(d^-t, dz), \quad X(0) = x. \quad (2.6)$$

We call $d^-X(t)$ the *forward differential* of $X(t)$, $t \in [0, T]$.

REMARK 2.2. There is a relation between the forward integral and the Skorohod integral (see DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005], lemma 4.1). Using

this, we can see that under mild conditions, there is a càdlàg version of the process $X(t)$, $t \in [0, T]$. From now on, we will consider and use this càdlàg version.

We can now state the *Itô formula for forward integrals* (see RUSSO and VALLOIS [1995] and RUSSO and VALLOIS [2000], for the Brownian motion case, and DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2005], for the compensated Poisson random measure case).

THEOREM 2.1. *Let $X(t)$, $t \in [0, T]$, be a forward process of the form (2.5) and assume that $\psi(\omega, t, z)$ is continuous in z around zero for (ω, t) -a.e. and $\int_0^T \int_{\mathbb{R}_0} \psi(t, z)^2 \nu(dz) dt < \infty$ ω -a.s. Let $f \in C^2(\mathbb{R})$. Then, the forward differential of $Y(t) = f(X(t))$, $t \in [0, T]$, is given by the following formula:*

$$\begin{aligned} d^- Y(t) = & \left[f'(X(t))\alpha(t) + \frac{1}{2} f''(X(t))\varphi^2(t) \right. \\ & + \int_{\mathbb{R}_0} \left(f(X(t^-) + \psi(t, z)) - f(X(t^-)) - f'(X(t^-))\psi(t, z) \right) \nu(dz) \Big] dt \\ & + f'(X(t))\varphi(t) d^- B(t) + \int_{\mathbb{R}_0} \left(f(X(t^-) + \psi(t, z)) \right. \\ & \left. \left. - f(X(t^-)) \right) \tilde{N}(d^- t, dz), \end{aligned} \quad (2.7)$$

where $f'(x) = \frac{d}{dx} f(x)$ and $f''(x) = \frac{d^2}{dx^2} f(x)$, $x \in \mathbb{R}$.

3. Optimal portfolio problem for a “large” insider

In this section, we study the existence of an optimal portfolio for the problem (1.1).

Let us consider the following market model with a finite time horizon $T > 0$ and two investment possibilities:

- a bond with price dynamics

$$\begin{cases} dS_0(t) = r(t)S_0(t)dt, & t \in (0, T], \\ S_0(0) = 1 \end{cases} \quad (3.1)$$

- a stock with price dynamics

$$\begin{cases} dS_1(t) = S_1(t^-) [\mu(t, \pi(t))dt + \sigma(t)d^- B(t) + \int_{\mathbb{R}_0} \theta(t, z) \tilde{N}(d^- t, dz)], & t \in (0, T], \\ S_1(0) > 0 \end{cases} \quad (3.2)$$

on the complete probability space $(\Omega, \mathcal{F}, P)$. The stochastic coefficients $r(t)$, $\mu(t, \pi)$, $\sigma(t)$ and $\theta(t, z)$, $t \in [0, T]$, $z \in \mathbb{R}_0$, are measurable, càglàd processes with respect to the

parameter t , adapted to some given filtration $\mathbb{G}$, for each constant value of π . Here, $\mathbb{G} := \{\mathcal{G}_t \subset \mathcal{F}, t \in [0, T]\}$ is a filtration with

$$\mathcal{G}_t \supset \mathcal{F}_t, \quad t \in [0, T].$$

We also assume that $\theta(t, z) > -1$, $dt \times \nu(dz)$ -a.e. and that

$$E \int_0^T \left\{ |r(t)| + |\mu(t)| + \sigma^2(t) + \int_{\mathbb{R}_0} \theta^2(t, z) \nu(dz) \right\} dt < \infty.$$

We recall that $\mathbb{F} := \{\mathcal{F}_t \subset \mathcal{F}, t \in [0, T]\}$ is the filtration generated by the development of the noise events, that is, the driving processes $B(t)$ and $\tilde{N}(t, z)$, $t \in [0, T]$, $z \in \mathbb{R}_0$.

In this model, the coefficient $\mu(t)$, $t \in [0, T]$, depends on the portfolio choice $\pi(t)$, $t \in [0, T]$, of an insider who has access to the information represented by the filtration $\mathbb{H} := \{\mathcal{H}_t \subset \mathcal{F}, t \in [0, T]\}$ with

$$\mathcal{H}_t \supset \mathcal{G}_t \supset \mathcal{F}_t, \quad t \in [0, T].$$

Accordingly, the insider's portfolio $\pi = \pi(t)$, $t \in [0, T]$, is a stochastic process adapted to $\mathbb{H}$. With the above conditions on μ , we intend to model a possible situation in which an insider is so influential in the market to affect the prices with his choices. In this sense, we discuss about a large insider.

This exogenous model for the price dynamics (3.1)–(3.2) is in line with CUOCO and CVITANIC [1998]. In the study by CUOCO and CVITANIC [1998], a dependence of the coefficient r on the portfolio π is also considered. In this chapter, this can also be mathematically carried through without substantial change; however, the assumption that the return of the bond depends on the agent's portfolio could be considered unrealistic.

We consider the insider's wealth process to be given by

$$\begin{aligned} dX_\pi(t) = X_\pi(t^-) & \left\{ [r(t) + (\mu(t, \pi(t)) - r(t))\pi(t)] dt \right. \\ & \left. + \pi(t)\sigma(t)d^-B(t) + \pi(t) \int_{\mathbb{R}_0} \theta(t, z)\tilde{N}(d^-t, dz) \right\}, \end{aligned} \quad (3.3)$$

with initial capital $X_\pi(0) = x > 0$. In the sequel, we put $x = 1$ for simplicity in notation. By the Itô formula for forward integrals, see Theorem 2.1, the final wealth of the admissible portfolio π is the unique solution of Eq. (3.3):

$$\begin{aligned} X_\pi(t) = \exp & \left\{ \int_0^t \left[r(s) + (\mu(s, \pi(s)) - r(s))\pi(s) \right. \right. \\ & \left. \left. - \frac{1}{2}\sigma^2(s)\pi^2(s) \right] ds - \int_0^t \int_{\mathbb{R}_0} \left[\pi(s)\theta(s, z) - \ln(1 + \pi(s)\theta(s, z)) \right] \nu(dz) ds \right. \\ & \left. + \int_0^t \pi(s)\sigma(s)d^-B(s) + \int_0^t \int_{\mathbb{R}_0} \ln(1 + \pi(s)\theta(s, z))\tilde{N}(d^-s, dz) \right\}. \end{aligned} \quad (3.4)$$

Taking the point of view of an insider, with the only purpose of understanding his opportunities in the market, we are interested in solving the optimization problem

$$\Phi := \sup_{\pi \in \mathcal{A}} E[U(X_\pi(T))] = E[U(X_{\pi^*}(T))], \quad (3.5)$$

for the given utility function

$$U : [0, \infty) \longrightarrow [-\infty, \infty),$$

which is a nondecreasing, concave, and lower semicontinuous, which we assume to be continuously differentiable on $(0, \infty)$. Here, the controls belong to the set $\mathcal{A}$ of admissible portfolios characterized as follows.

DEFINITION 3.1. The set $\mathcal{A}$ of *admissible portfolios* consists of all processes $\pi = \pi(t)$, $t \in [0, T]$, such that

- π is càglàd and adapted to the filtration $\mathbb{H}$; (3.6)

- $\pi(t)\sigma(t)$, $t \in [0, T]$ is forward integrable with respect to $d^-B(t)$; (3.7)

- $\pi(t)\theta(t, z)$, $t \in [0, T]$, $z \in \mathbb{R}_0$ is forward integrable with respect to $\tilde{N}(d^-t, dz)$; (3.8)

- $\pi(t)\theta(t, z) > -1 + \epsilon_\pi$ for a.a. (t, z) with respect to $dt \times \nu(dz)$, for some $\epsilon_\pi \in (0, 1)$ depending on π ; (3.9)

- $E \int_0^T \left\{ |\mu(s, \pi(s)) - r(s)| |\pi(s)| + (1 + \sigma^2(s)) \pi^2(s) + \int_{\mathbb{R}_0} \pi^2(s) \theta^2(s, z) \nu(dz) \right\} ds < \infty$ (3.10)

- $\ln(1 + \pi(t)\theta(t, z))$ is forward integrable with respect to $\tilde{N}(d^-t, dz)$; (3.11)

$$E[U(X_\pi(T))] < \infty \text{ and } 0 < E[U'(X_\pi(T))X_\pi(T)] < \infty,$$

where $U'(w) = \frac{d}{dw} U(w)$, $w \geq 0$. (3.12)

- For all $\pi, \beta \in \mathcal{A}$, with β bounded, there exists a $\zeta > 0$ such that the family

$$\{U'(X_{\pi+\delta\beta}(T))X_{\pi+\delta\beta}(T) | M_{\pi+\delta\beta}(T)\}_{\delta \in (-\zeta, \zeta)} \quad (3.13)$$

is uniformly integrable. Note that, for $\pi \in \mathcal{A}$ and $\beta \in \mathcal{A}$ bounded, $\pi + \delta\beta \in \mathcal{A}$ for any $\delta \in (-\zeta, \zeta)$ with ζ small enough. Here, the stochastic process $M_\pi(t)$, $t \in [0, T]$, is

defined as

$$\begin{aligned}
 M_\pi(t) := & \int_0^t \left\{ \mu(s, \pi(s)) - r(s) + \mu'(s, \pi(s))\pi(s) \right. \\
 & \left. - \sigma^2(s)\pi(s) - \int_{\mathbb{R}_0} \frac{\pi(s)\theta^2(s, z)}{1 + \pi(s)\theta(s, z)} \nu(dz) \right\} ds \\
 & + \int_0^t \sigma(s) d^-B(s) + \int_0^t \int_{\mathbb{R}_0} \frac{\theta(s, z)}{1 + \pi(s)\theta(s, z)} \tilde{N}(d^-s, dz), \tag{3.14}
 \end{aligned}$$

where $\mu'(s, \pi) = \frac{\partial}{\partial \pi} \mu(s, \pi)$.

REMARK 3.1. Condition (3.13) may be difficult to verify. Here, we give some examples of conditions under which it holds.

First, consider $M(\delta) := M_{\pi+\delta\beta}(T)$. The uniform integrability of $\{M(\delta)\}_{\delta \in (-\zeta, \zeta)}$ is assured by

$$\sup_{\delta \in (-\zeta, \zeta)} E[|M|^p(\delta)] < \infty \quad \text{for some } p > 1.$$

Observe that, since $\pi, \beta \in \mathcal{A}$ (see (3.9)), we have that $1 + (\pi(s) + \delta\beta(s))\theta(s, z) \geq \epsilon_\pi - \zeta \, d\tau \times \nu(dz)$ -a.e. for some $\zeta \in (0, \epsilon_\pi)$. Moreover, for $\epsilon > 0$,

$$\begin{aligned}
 & \int_0^T \int_{|z| \geq \epsilon} \frac{\theta(s, z)}{1 + (\pi(s) + \delta\beta(s))\theta(s, z)} \tilde{N}(d^-s, dz) \\
 &= \int_0^T \int_{|z| \geq \epsilon} \frac{\theta(s, z)}{1 + (\pi(s) + \delta\beta(s))\theta(s, z)} \tilde{N}(ds, dz).
 \end{aligned}$$

Thus, we have that

$$\begin{aligned}
 & E \left[\left(\int_0^T \int_{|z| \geq \epsilon} \frac{\theta(s, z)}{1 + (\pi(s) + \delta\beta(s))\theta(s, z)} \tilde{N}(d^-s, dz) \right)^2 \right] \\
 & \leq \frac{1}{(\epsilon_\pi - \zeta)^2} E \left[\int_0^T \int_{|z| \geq \epsilon} \theta^2(s, z) \nu(dz) ds \right] < \infty.
 \end{aligned}$$

So, if

$$E \left[\left(\int_0^T \sigma(s) d^-B(s) \right)^2 \right] < \infty \text{ and } E \left[\left(\int_0^T \int_{|z| < \epsilon} |\theta(s, z)| \tilde{N}(d^-s, dz) \right)^2 \right] < \infty$$

(see Remark 2.1 (1)), we have that $E[M^2(\delta)] < \infty$ uniformly in $\delta \in (-\zeta, \zeta)$ if, for example, the coefficients μ, μ', r , and σ are bounded. This shows that (3.13) holds if $U'(x)x$ is uniformly bounded for $x \in (0, \infty)$. This is the case, for example, of $U(x) = \ln x$ and $U(x) = -\exp\{-\lambda x\}$ ($\lambda > 0$).

Similarly, in the case of power utility function

$$U(x) = \frac{1}{\gamma} x^\gamma, \quad x > 0 \quad \text{for some } \gamma \in (0, 1),$$

we see that $U'(X_{\pi+\delta\beta}(T))X_{\pi+\delta\beta}(T)|M(\delta)| = X_{\pi+\delta\beta}^\gamma(T)|M(\delta)|$ and condition (3.13) would be satisfied if

$$\sup_{\delta \in (-\zeta, \zeta)} E[(X_{\pi+\delta\beta}^\gamma(T)|M(\delta)|)^p] < \infty \quad \text{for some } p > 1.$$

Note that we can write

$$X_{\pi+\delta\beta}(T) = X_\pi(T)N(\delta),$$

where

$$\begin{aligned} N(\delta) := \exp \bigg\{ & \int_0^T [(\mu(s, \pi(s) + \delta\beta(s)) - r(s))\delta\beta(s) + (\mu(s, \pi(s) + \delta\beta(s)) \\ & - \mu(s, \pi(s))\pi(s) - \sigma^2(s)\delta\beta(s)\pi(s) - \frac{1}{2}\sigma^2(s)\delta^2\beta^2(s)]ds \\ & + \int_0^T \delta\sigma(s)\beta(s)d^-B(s) + \int_0^T \int_{\mathbb{R}_0} [\ln(1 + (\pi(s) + \delta\beta(s))\theta(s, z)) \\ & - \ln(1 + \pi(s)\theta(s, z)) - \delta\beta(s)\theta(s, z)]\nu(dz)ds \\ & + \int_0^T \int_{\mathbb{R}_0} [\ln(1 + (\pi(s) + \delta\beta(s))\theta(s, z)) - \ln(1 + \pi(s)\theta(s, z))] \tilde{N}(d^-s, dz) \bigg\}. \end{aligned}$$

From the iterated application of the Hölder inequality, we have

$$\begin{aligned} & E[(X_{\pi+\delta\beta}^\gamma(T)|M(\delta)|)^p] \\ & \leq (E[(X_\pi(T))^{\gamma p a_1 b_1}])^{\frac{1}{a_1 b_1}} (E[(N(\delta))^{\gamma p a_1 b_2}])^{\frac{1}{a_1 b_2}} (E[(|M(\delta)|)^{p a_2}])^{\frac{1}{a_2}}, \end{aligned}$$

where a_1, a_2 : $\frac{1}{a_1} + \frac{1}{a_2} = 1$ and b_1, b_2 : $\frac{1}{b_1} + \frac{1}{b_2} = 1$. Then, we can choose $a_1 = \frac{2}{2-p}$, $a_2 = \frac{2}{p}$ and also $b_1 = \frac{2-p}{\gamma p}$, $b_2 = \frac{2-p}{2-p-\gamma p}$ for some $p \in (1, \frac{2}{\gamma+1})$. Hence,

$$\begin{aligned} & E[(X_{\pi+\delta\beta}^\gamma(T)|M(\delta)|)^p] \\ & \leq (E[(X_\pi(T))^2])^{\frac{\gamma p}{2}} (E[(N(\delta))^{\frac{2\gamma p}{2-p-\gamma p}}])^{\frac{2-p-\gamma p}{2}} (E[(|M(\delta)|)^2])^{\frac{p}{2}}. \end{aligned}$$

If the value $X_\pi(T)$ in (3.4) satisfies

$$E[(X_\pi(T))^2] < \infty, \tag{3.15}$$

then the condition (3.13) holds if

$$\sup_{\delta \in (-\zeta, \zeta)} E[(N(\delta))^{\frac{2\gamma p}{2-p-\gamma p}}] < \infty.$$

Since (3.10) holds, it is enough, for example, that μ, μ', r , and σ are bounded to have $E[(N(\delta))^{\frac{2\gamma p}{2-p-\gamma p}}] < \infty$ uniformly in $\delta \in (-\zeta, \zeta)$. Note that condition (3.15) is verified, for example, if for all $K > 0$

$$E\left[\exp\left\{K\left(\int_0^T |\pi(s)|ds + \left|\int_0^T \pi(s)\sigma(s)d^-B(s)\right| + \left|\int_0^T \int_{\mathbb{R}_0} \ln(1 + \pi(s)\theta(s, z))\tilde{N}(d^-s, dz)\right|\right)\right\}\right] < \infty.$$

By similar arguments, we can also treat the case of a utility function such with $U'(x)$ uniformly bounded for $x \in (0, \infty)$. We omit the details. $\square$

The forward stochastic calculus gives an adequate mathematical framework in which we can proceed to solve the optimization problem (3.5). Define

$$J(\pi) := E[U(X_\pi(T))], \quad \pi \in \mathcal{A}.$$

First, let us suppose that π is locally optimal for the insider, in the same that $J(\pi) \geq J(\pi + \delta\beta)$ for all $\beta \in \mathcal{A}$ bounded, and for all δ small enough. Since the function $J(\pi + \delta\beta)$ is maximal at π , by Eqs. (3.13) and (2.4), we have that

$$\begin{aligned} 0 &= \frac{d}{d\delta} J(\pi + \delta\beta)|_{\delta=0} \\ &= E\left[U'(X_\pi(T))X_\pi(T)\left\{\int_0^T \beta(s)\left[\mu(s, \pi(s)) - r(s) + \mu'(s, \pi(s))\pi(s) \right. \right. \right. \\ &\quad \left. \left. - \sigma^2(s)\pi(s) - \int_{\mathbb{R}_0} \left\{\theta(s, z) - \frac{\theta(s, z)}{1 + \pi(s)\theta(s, z)}\right\}\nu(dz)\right]ds \right. \\ &\quad \left. + \int_0^T \beta(s)\sigma(s)d^-B(s) + \int_0^T \int_{\mathbb{R}_0} \frac{\beta(s)\theta(s, z)}{1 + \pi(s)\theta(s, z)}\tilde{N}(d^-s, dz)\right\}\right]. \end{aligned} \quad (3.16)$$

Now, let us fix $t \in [0, T)$ and $h > 0$ such that $t + h \leq T$. We can choose $\beta \in \mathcal{A}$ of the form

$$\beta(s) = \alpha\chi_{(t, t+h]}(s), \quad 0 \leq s \leq T,$$

where α is an arbitrary bounded $\mathcal{H}_t$ -measurable random variable. Then, (3.16) gives

$$\begin{aligned} 0 &= E\left[U'(X_\pi(T))X_\pi(T)\left\{\int_t^{t+h} [\mu(s, \pi(s)) - r(s) \right. \right. \\ &\quad \left. \left. + \mu'(s, \pi(s))\pi(s) - \sigma^2(s)\pi(s) - \int_{\mathbb{R}_0} \frac{\pi(s)\theta^2(s, z)}{1 + \pi(s)\theta(s, z)}\nu(dz)]ds \right. \right. \\ &\quad \left. \left. + \int_t^{t+h} \sigma(s)d^-B(s) + \int_t^{t+h} \int_{\mathbb{R}_0} \frac{\theta(s, z)}{1 + \pi(s)\theta(s, z)}\tilde{N}(d^-s, dz)\right\} \cdot \alpha\right]. \end{aligned} \quad (3.17)$$

Since this holds for all such α , we can conclude that

$$E[F_\pi(T)(M_\pi(t+h) - M_\pi(t))|\mathcal{H}_t] = 0, \quad (3.18)$$

where

$$F_\pi(T) := \frac{U'(X_\pi(T))X_\pi(T)}{E[U'(X_\pi(T))X_\pi(T)]} \quad (3.19)$$

and

$$\begin{aligned} M_\pi(t) := & \int_0^t \left\{ \mu(s, \pi(s)) - r(s) + \mu'(s, \pi(s))\pi(s) \right. \\ & \left. - \sigma^2(s)\pi(s) - \int_{\mathbb{R}_0} \frac{\pi(s)\theta^2(s, z)}{1 + \pi(s)\theta(s, z)} \nu(dz) \right\} ds \\ & + \int_0^t \sigma(s) d^- B(s) + \int_0^t \int_{\mathbb{R}_0} \frac{\theta(s, z)}{1 + \pi(s)\theta(s, z)} \tilde{N}(d^-s, dz), \quad t \in [0, T] \end{aligned} \quad (3.20)$$

- cf. (3.14). Define the probability measure Q_π on $(\Omega, \mathcal{H}_T)$ by

$$Q_\pi(d\omega) := F_\pi(T)P(d\omega) \quad (3.21)$$

and denote the expectation with respect to the measure Q_π by E_{Q_π} . Then, by (3.19), we have

$$E_{Q_\pi}[M_\pi(t+h) - M_\pi(t)|\mathcal{H}_t] = \frac{E[F_\pi(T)(M_\pi(t+h) - M_\pi(t))|\mathcal{H}_t]}{E[F_\pi(T)|\mathcal{H}_t]} = 0.$$

Hence, the process $M_\pi(t)$, $t \in [0, T]$ is a $(\mathbb{H}, Q_\pi)$ -martingale (i.e., a martingale with respect to the filtration $\mathbb{H}$ and under the probability measure Q_π).

On the other hand, the argument can be reversed as follows. If $M_\pi(t)$, $t \in [0, T]$, is a $(\mathbb{H}, Q_\pi)$ -martingale, then

$$E[F_\pi(T)(M_\pi(t+h) - M_\pi(t))|\mathcal{H}_t] = 0,$$

for all $h > 0$ such that $0 \leq t < t+h \leq T$, which is (3.18). Or equivalently,

$$E[\alpha F_\pi(T)(M_\pi(t+h) - M_\pi(t))] = 0$$

for all bounded $\mathcal{H}_t$ -measurable $\alpha \in \mathcal{A}$. Hence, (3.17) holds for all such α . Taking linear combinations, we see that (3.16) holds for all càglàd step processes $\beta \in \mathcal{A}$. By our assumptions (3.7) and (3.8) on $\mathcal{A}$ we get, by an approximation argument, that (3.16) holds for all $\beta \in \mathcal{A}$. If the function $g(\delta) := E[U(X_{\pi+\delta\beta}(T))]$, $\delta \in (-\zeta, \zeta)$ is concave for each $\beta \in \mathcal{A}$, we conclude that its maximum is achieved at $\delta = 0$. Hence, we have proved the following result.

THEOREM 3.1. (1) If the stochastic process $\pi \in \mathcal{A}$ is locally optimal for the problem (3.5), then the stochastic process $M_\pi(t)$, $t \in [0, T]$, is an $(\mathbb{H}, Q_\pi)$ martingale. (2) Conversely, if the function $g(\delta) := E[U(X_{\pi+\delta\beta}(T))]$, $\delta \in (-\zeta, \zeta)$, is concave for each $\beta \in \mathcal{A}$ and $M_\pi(t)$, $t \in [0, T]$, is an $(\mathbb{H}, Q_\pi)$ martingale, then $\pi \in \mathcal{A}$ is locally optimal for the problem (3.5).

REMARK. Since the composition of a concave increasing function with a concave function is concave, we can see that a sufficient condition for the function $g(\delta)$, $\delta \in (-\zeta, \zeta)$, to be concave is that the function

$$\Lambda(s) : \pi \longrightarrow r(s) + (\mu(s, \pi) - r(s))\pi - \frac{1}{2}\sigma^2(s)\pi^2 \quad (3.22)$$

is concave for all $s \in [0, T]$. For this, it is sufficient that $\mu(s, \cdot)$ are C^2 for all s and that

$$\mu''(s, \pi)\pi + 2\mu'(s, \pi) - \sigma^2 \leq 0 \quad (3.23)$$

for all s, π . Here, we have set $\mu' = \frac{\partial \mu}{\partial \pi}$ and $\mu'' = \frac{\partial^2 \mu}{\partial \pi^2}$.

Moreover, we also obtain the following result

THEOREM 3.2. (1) A stochastic process $\pi \in \mathcal{A}$ is optimal for the problem (3.5) only if the process

$$\hat{M}_\pi(t) := M_\pi(t) - \int_0^t \frac{d[M_\pi, Z_\pi](s)}{Z_\pi(s)}, \quad t \in [0, T], \quad (3.24)$$

is an $(\mathbb{H}, P)$ martingale (i.e., a martingale with respect to the filtration $\mathbb{H}$ and under the probability measure P). Here,

$$Z_\pi(t) := E_{Q_\pi} \left[\frac{dP}{dQ_\pi} | \mathcal{H}_t \right] = (E[F_\pi(T) | \mathcal{H}_t])^{-1}, \quad t \in [0, T]. \quad (3.25)$$

(2) Conversely, if $g(\delta) := E[U(X_{\pi+\delta\beta}(T))]$, $\delta \in (-\zeta, \zeta)$, is concave and (3.24) is an $(\mathbb{H}, P)$ martingale, then $\pi \in \mathcal{A}$ is optimal for the problem (3.5).

PROOF. If $\pi \in \mathcal{A}$ is an optimal portfolio for an insider, then by Theorem 3.1 we know that $M_\pi(t)$, $t \in [0, T]$ is an $(\mathbb{H}, Q_\pi)$ martingale. Applying the Girsanov theorem (see PROTTER [2003] theorem 3.35), we obtain that

$$\hat{M}_\pi(t) := M_\pi(t) - \int_0^t \frac{d[M_\pi, Z_\pi](s)}{Z_\pi(s)}, \quad t \in [0, T],$$

is an $(\mathbb{H}, P)$ martingale with

$$Z_\pi(t) = E_{Q_\pi} \left[\frac{dP}{dQ_\pi} | \mathcal{H}_t \right] = E \left[(F_\pi(T))^{-1} \frac{F_\pi(T)}{E[F_\pi(T) | \mathcal{H}_t]} | \mathcal{H}_t \right] = (E[F_\pi(T) | \mathcal{H}_t])^{-1}.$$

Conversely, if $\hat{M}_\pi(t)$, $t \in [0, T]$, is $(\mathbb{H}, P)$ martingale, then $M_\pi(t)$, $t \in [0, T]$, is an $(\mathbb{H}, Q_\pi)$ martingale. Hence, π is optimal by Theorem 3.1. $\square$

4. Examples

In this section, we give some examples to illustrate the contents of the main results in Section 3.

EXAMPLE 4.1. Suppose that

$$\sigma(t) \neq 0, \theta = 0 \text{ and } \mathcal{H}_t = \mathcal{F}_t \vee \sigma(B(T_0)), \text{ for all } t \in [0, T] \text{ (for some } T_0 > T), \quad (4.1)$$

that is, we consider a market driven by the Brownian motion only and where the insider's filtration is a classical example of enlargement of the filtration $\mathbb{F}$ by the knowledge derived from the value of the Brownian motion at some future time $T_0 > T$. Then, we obtain the following result.

THEOREM 4.1. *Suppose that the function Λ in (3.22) is concave for all $s \in [0, T]$. A portfolio $\pi \in \mathcal{A}$ is optimal for the problem (3.5) if and only if $d[M_\pi, Z_\pi](t)$ is absolutely continuous with respect to the Lebesgue measure dt and*

$$\begin{aligned} & \mu'(t, \pi(t))\pi(t) + \mu(t, \pi(t)) - r(t) \\ & - \sigma^2(t)\pi(t) + \sigma(t) \left[\frac{B(T_0) - B(t)}{T_0 - t} - \frac{1}{Z_\pi(t)} \frac{d}{dt} [B, Z_\pi](t) \right] = 0. \end{aligned} \quad (4.2)$$

PROOF. By Theorem 3.2, the portfolio $\pi \in \mathcal{A}$ is optimal for the problem (3.5) if and only if the process

$$\begin{aligned} \hat{M}_\pi(t) = & \int_0^t \{ \mu'(s, \pi(s))\pi(s) + \mu(s, \pi(s)) - r(s) \\ & - \sigma^2(s)\pi(s) \} ds + \int_0^t \sigma(s) d^- B(s) - \int_0^t \frac{d[M_\pi, Z_\pi](s)}{Z_\pi(s)} \end{aligned} \quad (4.3)$$

is an $(\mathbb{H}, P)$ martingale. Since $\hat{M}_\pi(t)$ is continuous and has quadratic variation

$$[\hat{M}_\pi, \hat{M}_\pi](t) = \int_0^t \sigma^2(s) ds,$$

we conclude that $\hat{M}_\pi(t)$ can be written as

$$\hat{M}_\pi(t) = \int_0^t \sigma(s) d\hat{B}(s) \quad (4.4)$$

for some $(\mathbb{H}, P)$ Brownian motion $\hat{B}$.

On the other hand, from the result of Itô [1978], we know that $B(t)$ is a semimartingale with respect to $(\mathbb{H}, P)$ with decomposition

$$B(t) = \tilde{B}(t) + \int_0^t \frac{B(T_0) - B(s)}{T_0 - s} ds, \quad 0 \leq t \leq T, \quad (4.5)$$

for some $(\mathbb{H}, P)$ Brownian motion $\tilde{B}(t)$. Combining (4.3)–(4.5), we get

$$\begin{aligned} \sigma(t)d\hat{B}(t) &= d\hat{M}_\pi(t) = \{\mu'(t, \pi(t))\pi(t) \\ &\quad + \mu(t, \pi(t)) - r(t) - \sigma^2(t)\pi(t)\}dt + \sigma(t)d\tilde{B}(t) \\ &\quad + \sigma(t)\frac{B(T_0) - B(t)}{T_0 - t}dt - \frac{d[M_\pi, Z_\pi](t)}{Z_\pi(t)}. \end{aligned} \quad (4.6)$$

By uniqueness of the semimartingale decomposition of $\hat{M}_\pi(t)$ with respect to $(\mathbb{H}, P)$, we conclude that $\hat{B}(t) = \tilde{B}(t)$ and

$$\begin{aligned} &\{\mu'(t, \pi(t))\pi(t) + \mu(t, \pi(t)) - r(t) - \sigma^2(t)\pi(t) \\ &\quad \sigma(t)\frac{B(T_0) - B(t)}{T_0 - t}\}dt - \frac{d[M_\pi, Z_\pi](t)}{Z_\pi(t)} = 0. \end{aligned} \quad (4.7)$$

From this, we deduce that $d[M_\pi, Z_\pi](t) = \sigma(t)d[B, Z_\pi](t)$ is absolutely continuous with respect to dt and (4.2) follows. $\square$

COROLLARY 4.1. *Assume that (4.1) holds and, in addition, that*

$$\mu(t, \pi) = \mu_0(t) + a(t)\pi \quad (4.8)$$

for some $\mathbb{F}$ -adapted processes μ_0 and a with $0 \leq a(t) \leq \frac{1}{2}\sigma^2(t)$, $t \in [0, T]$, which do not depend on π . Then, $\pi \in \mathcal{A}$ is optimal if and only if $d[M_\pi, Z_\pi](t)$ is absolutely continuous with respect to dt and

$$(\sigma^2(t) - 2a(t))\pi(t) = \mu_0(t) - r(t) + \sigma(t)\left[\frac{B(T_0) - B(t)}{T_0 - t} - \frac{1}{Z_\pi(t)}\frac{d[B, Z_\pi](t)}{dt}\right]. \quad (4.9)$$

PROOF. In this case, we have that $\mu'(t, \pi(t)) = a(t)$. Therefore, the function Λ defined in (3.22) is concave (by (3.23)), and the result follows from Theorem 4.1. $\square$

Next, we give an example for a pure-jump financial market.

EXAMPLE 4.2. Suppose that

$$\sigma(t) = 0 \quad \text{and} \quad \theta(t, z) = \beta z, \quad (4.10)$$

where $\beta z > -1$ $\nu(dz)$ -a.e. ($\beta > 0$) and that

$$\mathcal{H}_t = \mathcal{F}_t \vee \sigma(\eta(T_0)) \quad \text{for some } T_0 > T, \quad (4.11)$$

where

$$\eta(t) = \int_0^t \int_{\mathbb{R}_0} z \tilde{N}(ds, dz),$$

(i.e., the insider's filtration is the enlargement of $\mathbb{F}$ by the knowledge derived from some future value $\eta(T_0)$ of the market driving process). Then, by the result of Itô, as extended by Kurtz (see PROTTER [2003] pp 256), the process

$$\hat{\eta}(t) := \eta(t) - \int_0^t \frac{\eta(T_0) - \eta(s)}{T_0 - s} ds \quad (4.12)$$

is an $(\mathbb{H}, P)$ martingale. By proposition 5.2 in DI NUNNO, MEYER-BRANDIS, ØKSENDAL and PROSKE [2006], the $\mathbb{H}$ -compensating measure $\nu_{\mathbb{H}}$ of the jump measure N is given by

$$\begin{aligned} \nu_{\mathbb{H}}(ds, dz) &= \nu_{\mathbb{F}}(dz)ds + E\left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) | \mathcal{H}_s\right] ds \\ &= E\left[\frac{1}{T_0 - s} \int_s^{T_0} N(dr, dz) | \mathcal{H}_s\right] ds, \end{aligned} \quad (4.13)$$

where $\nu_{\mathbb{F}} = \nu$. This implies that the $\mathbb{H}$ -compensated random measure $\tilde{N}_{\mathbb{H}}$ is related to $\tilde{N}_{\mathbb{F}} = \tilde{N}$ by

$$\begin{aligned} \tilde{N}_{\mathbb{H}}(ds, dz) &= N(ds, dz) - \nu_{\mathbb{H}}(ds, dz) \\ &= \tilde{N}(ds, dz) - E\left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) | \mathcal{H}_s\right] ds. \end{aligned} \quad (4.14)$$

Hence, directly from the definition of the forward integral, we have

$$\begin{aligned} \int_0^t \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} \tilde{N}(d^-s, dz) &= \int_0^t \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} \tilde{N}_{\mathbb{H}}(ds, dz) \\ &\quad + \int_0^t \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} \\ &\quad \times E\left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) | \mathcal{H}_s\right] ds. \end{aligned} \quad (4.15)$$

By Theorem 3.2, a portfolio $\pi \in \mathcal{A}$ is optimal if and only if the process

$$\begin{aligned} \hat{M}_{\pi}(t) &= \int_0^t \left\{ \mu(s, \pi(s)) - r(s) + \mu'(s, \pi(s))\pi(s) \right. \\ &\quad \left. - \int_{\mathbb{R}_0} \frac{\beta^2 z^2 \pi(s)}{1 + \pi(s)\beta z} \nu(dz) \right\} ds \\ &\quad + \int_0^t \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} \tilde{N}(d^-s, dz) - \int_0^t \frac{d[M_{\pi}, Z_{\pi}](s)}{Z_{\pi}(s)} \end{aligned} \quad (4.16)$$

is an $(\mathbb{H}, P)$ martingale. Therefore, if we put

$$\begin{aligned} G_\pi(s) &:= \mu(s, \pi(s)) - r(s) + \mu'(s, \pi(s))\pi(s) \\ &\quad - \int_{\mathbb{R}_0} \frac{\beta^2 z^2 \pi(s)}{1 + \pi(s)\beta z} \nu(dz) \\ &\quad + \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} E\left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) | \mathcal{H}_s\right], \end{aligned} \quad (4.17)$$

and combine (4.15) and (4.16), we obtain that the process

$$\hat{M}_\pi(t) = \int_0^t G_\pi(s) ds - \int_0^t \frac{d[M_\pi, Z_\pi](s)}{Z_\pi(s)} + \int_0^t \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} \tilde{N}_{\mathbb{H}}(ds, dz)$$

is an $(\mathbb{H}, P)$ martingale. This is possible if and only if

$$\int_0^t G_\pi(s) ds - \int_0^t \frac{d[M_\pi, Z_\pi](s)}{Z_\pi(s)} = 0, \quad \text{for all } t \in [0, T].$$

This implies that $d[M_\pi, Z_\pi](t)$ is absolutely continuous with respect to the Lebesgue measure dt . We have, thus, proved the following statement.

THEOREM 4.2. *Assume that (4.10) and (4.11) hold. Then, $\pi \in \mathcal{A}$ is optimal if and only if $d[M_\pi, Z_\pi](t)$ is absolutely continuous with respect to the Lebesgue measure dt and*

$$G_\pi(t) = \frac{1}{Z_\pi(t)} \frac{d}{dt}[M_\pi, Z_\pi](t) \quad \text{for almost all } t \in [0, T], \quad (4.18)$$

where G_π is given by (4.17).

In analogy with Corollary 4.1, we get the following result in the special case when the influence of the trader on the market is given by (4.8).

COROLLARY 4.2. *Assume that (4.10) and (4.11) hold and, in addition, that also (4.8) holds. Then, $\pi \in \mathcal{A}$ is optimal if and only if $d[M_\pi, Z_\pi](t)$ is absolutely continuous with respect to the Lebesgue measure dt and*

$$\begin{aligned} \pi(s) \int_{\mathbb{R}_0} \frac{\beta^2 z^2}{1 + \pi(s)\beta z} \nu(dz) - \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} E\left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) | \mathcal{H}_s\right] \\ - 2a(s)\pi(s) = \mu_0(s) - r(s) - \frac{1}{Z_\pi(s)} \frac{d}{ds}[M_\pi, Z_\pi](s). \end{aligned} \quad (4.19)$$

COROLLARY 4.3. *Suppose that (4.8), (4.10), and (4.11) hold and that*

$$U(x) = \ln x, \quad x \geq 0. \quad (4.20)$$

Then, $\pi \in \mathcal{A}$ is optimal if and only if

$$\begin{aligned} \pi(s) \int_{\mathbb{R}_0} \frac{\beta^2 z^2}{1 + \pi(s)\beta z} \nu(dz) - \int_{\mathbb{R}_0} \frac{\beta z}{1 + \pi(s)\beta z} E \left[\frac{1}{T_0 - s} \int_s^{T_0} \tilde{N}(dr, dz) \middle| \mathcal{H}_s \right] \\ - 2a(s)\pi(s) = \mu_0(s) - r(s). \end{aligned}$$

PROOF. If $U(x) = \ln x$, then $F_\pi(T) = 1 = Z_\pi(t)$, $t \in [0, T]$. Hence, $[M_\pi, Z_\pi] = 0$. $\square$

5. Acknowledgment

The authors would like to thank Terje Bjuland for his useful comments.

References

- APPLEBAUM, D. (2004). *Lévy Processes and Stochastic Calculus* (Cambridge University Press).
- BACK, K. (1992). Insider trading in continuous time. *Rev. Financ. Stud.* **5**, 387–409.
- BARNDORFF-NIELSEN, O. (1998). Processes of normal inverse Gaussian type. *Financ. Stoch.* **1**, 41–68.
- BERTOIN, J. (1996). *Lévy Processes* (Cambridge University Press).
- BIAGINI, F., ØKSENDAL, B. (2005). A general stochastic calculus approach to insider trading. *Appl. Math. Optim.* **52**, 167–181.
- CUOCO, D., CVITANIC, J. (1998). Optimal consumption choices for a “large” investor. *J. Econ. Dynam. Control* **22**, 401–436.
- CONT, R., TANKOV, P. (2004). *Financial Modelling with Jump Processes* (Chapman and Hall).
- DI NUNNO, G., MEYER-BRANDIS, T., ØKSENDAL, B., PROSKE, F. (2005). Malliavin Calculus for Lévy processes. *Infin. Dimens. Anal. Quantum Probab. Relat. Fields.* **8**, 235–258.
- DI NUNNO, G., MEYER-BRANDIS, T., ØKSENDAL, B., PROSKE, F. (2006). Optimal portfolio for an insider in a market driven by Lévy processes. *Quant. Financ.* **6**, 83–94.
- ELLIOTT, R., JEANBLANC, M. (1998). Incomplete markets with jumps and informed agents. *Math. Method Oper. Res.* **50**, 475–492.
- ELLIOTT, R., GEMAN, H., KORKIE, R. (1997). Portfolio optimization and contingent claim pricing with differential information. *Stoch. Stoch. Rep.* **60**, 185–203.
- EBERLEIN, E., RAIBLE, S. (1999). Term structure models driven by Lévy processes. *Math. Finance.* **9**, 31–53.
- ITÔ, K. (1978). Extension of stochastic integrals. In Proceedings of International symposium on stochastic Differential of Equations, Wiley 1978, pp. 95–109.
- KARATZAS, I., PIKOVSKY, I. (1996). Anticipating Portfolio Optimization. *Adv. Appl. Prob.* **28**, 1095–1122.
- KYLE, A. (1985). Continuous auctions and insider trading. *Econometrica* **53**, 1315–1335.
- KOHATSU-HIGA, A., SULEM, A. (2006). Utility maximization in an insider influenced market. *Math. Finance.* **16**, 153–179.
- KOHATSU-HIGA, A., SULEM, A. (2006). A Large Trader-Insider Model, Proceedings of the Ritsumeikan International Symposium, Japan, March 2005, In: Akahori, J., Ogawa, S., Watanabe, S (eds.), *Stochastic Processes and Applications to Mathematical Finance*, (World Scientific), pp. 101–124.
- KOHATSU-HIGA, A., YAMAZATO, M. (2004). Enlargement of filtrations with random times for processes with jumps. Preprint.
- KUNITA, H. (2004). Variational equality and portfolio optimization for price processes with jumps. In: Akahori, J., Ogawa, S., Watanabe, S., (eds.), *Processes and Applications to Mathematical Finance, Proceedings of the Ritsumeikan International Symposium Kusatsu, Shiga, Japan, March* (Ritsumeikan University, Japan).
- NUALART, D., PARDOUX, E. (1988). Stochastic calculus with anticipating integrands, *Probab. Theory Rel.* **78**, 535–581.
- NUALART, D., SCHOUTENS, W. (2000). Chaotic and predictable representations for Lévy processes. *Stochastic Process. Appl.* **90**, 109–122.
- ØKSENDAL, B. (2006). A universal optimal consumption rate for an insider. *Math. Finance.* **16**, 119–129.
- ØKSENDAL, B., SULEM, A. (2004). Partial observation control in an anticipating environment. *Russian Math. Surveys.* **50**, 355–375.
- PROTTER, P. (2003). *Stochastic Integration and Differential Equations*, Second ed. (Springer-Verlag).
- RUSSO, F., VALLOIS, P. (1993). Forward, backward and symmetric stochastic integration. *Prob. Theory Rel. Fields.* **97**, 403–421.

- RUSO, F., VALLOIS, P. (1995). The generalized covariation process and Itô formula. *Stoch. Proc. Appl.* **59**, 81–104.
- RUSO, F., VALLOIS, P. (2000). Stochastic calculus with respect to continuous finite quadratic variation processes. *Stoch. Stoch. Rep.* **70**, 1–40.
- SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Studies in Advanced Mathematics, vol. 68 (Cambridge University Press).
- SCHOUTENS, W. (2003). *Lévy Processes in Finance* (Wiley).

Optimal Quantization for Finance: From Random Vectors to Stochastic Processes

Gilles Pagès

*Laboratoire de Probabilités et Modèles aléatoires, UMR 7599,
Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France.
E-mail address: gpa@ccr.jussieu.fr*

Jacques Printems

*LAMA, Université Paris 12,
E-mail address: printems@univ-paris12.fr*

Abstract

In this chapter, we present an overview of the recent developments of vector quantization and functional quantization and their applications as a numerical method in finance, with an emphasis on the quadratic case. Quantization is a way to approximate a random vector or a stochastic process, viewed as a Hilbert-valued random variable, using a nearest neighbor projection on a finite codebook. We make a review of cubature formulas to approximate expectation, an conditional expectation, including the introduction of a quantization-based Richardson–Romberg extrapolation method. The optimal quadratic quantization of the Brownian motion is presented in full detail. A special emphasis is made on the computational aspects and the numerical applications, in particular, the pricing of different kinds of options in various fields (swing options on gas and options in a Heston stochastic volatility model).

1. Introduction

Quantization is a way to discretize the path space of a random phenomenon: a random vector in finite dimension and a stochastic process in infinite dimension. Optimal vector quantization theory (finite dimensional) of random vectors finds its origin in the early 1950s in order to discretize some emitted signal (see GERSHO and GRAY [1992] or GRAF and LUSCHGY [2000]). It was further developed by specialists in signal processing and in information theory.

The infinite-dimensional case started to be extensively investigated in the early 2000s by several authors (see PAGÈS [2000], LUSCHGY and PAGÈS [2002, 2004, 2006], DEREICH and SCHEUTZOW [2003, 2006], WILBERTZ [2005], GRAF, LUSCHGY and PAGÈS [2007]).

Let us consider a Hilbertian setting. One considers a random vector X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking its values in a separable Hilbert space $(H, (\cdot|\cdot)_H)$ (equipped with its natural Borel σ -algebra) and satisfying $\mathbb{E}|X|^2 < +\infty$. When H is an Euclidean space $(\mathbb{R}^d)$, one speaks about vector quantization. When H is an infinite-dimensional space like $L^2_T := L^2([0, T], dt)$ (endowed with the usual Hilbertian norm $\|f\|_{L^2_T} := (\int_0^T f^2(t)dt)^{\frac{1}{2}}$), one speaks of functional quantization (denoted by L^2_T from now on). A (bimeasurable) stochastic process $(X_t)_{t \in [0, T]}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying $\|X(\omega)\|_{L^2_T} < +\infty$ $\mathbb{P}(d\omega)$ -a.s. can always be seen, once possibly modified on a $\mathbb{P}$ -negligible set, as an L^2_T -valued random variable. Although we will focus on the Hilbertian framework, other choices are possible for H , in particular, some more general Banach settings like $L^p([0, T], dt)$ or $\mathcal{C}([0, T], \mathbb{R})$ spaces.

This chapter is organized as follows: in Section 2 and in its subsections we introduce quadratic quantization in a Hilbertian setting. In Section 3, we focus on optimal quantization, including some extensions to nonquadratic quantization. Section 4 is devoted to some quantized cubature formulae. Section 5.1 provides some classical background on the quantization rate in finite dimension. Section 6 deals with functional quantizations of Gaussian processes, like the Brownian motion, with a special emphasis on the numerical aspects. We present here what is, to our guess, the first large-scale numerical optimization of the quadratic quantization of the Brownian motion. We compare it to the optimal product quantization, formerly investigated in a study by PAGÈS and PRINTEMS [2005]. In Section 7, we propose a constructive approach to the functional quantization of scalar or multidimensional diffusions (in the Stratonovich sense). In Section 8, we show how to use functional quantization to price path-dependent options like Asian options (in a Heston stochastic volatility model). We conclude by some recent results showing how to derive universal (often optimal) functional quantization rate from time regularity of a process described in Section 9 and by a few examples in Section 10 about the specific methods that produce some lower bounds (this important subject as many others like the connections with small deviation theory is not treated in this numerically oriented overview). As concerns statistical applications of functional quantization, we refer to the studies by TARPEY and KINATEDER [2003], TARPEY, PETKOVA, and OGDEN [2003].

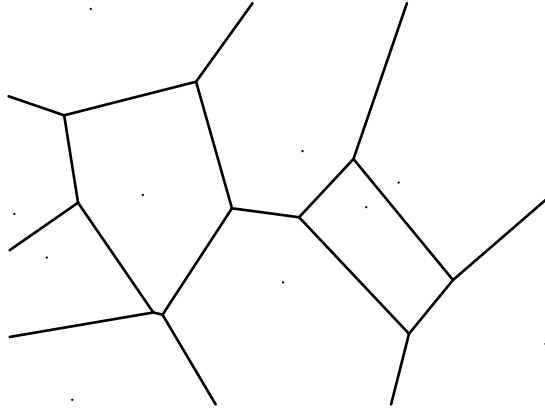


FIG. 1.1 A two-dimensional 10-quantizer $\Gamma = \{x_1, \dots, x_{10}\}$ and its Voronoi diagram.

NOTATIONS.

- $a_n \approx b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n \sim b_n$ means $a_n = b_n + o(a_n)$.
- If $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (H, |\cdot|_H)$ (Hilbert space), then $\|X\|_2 = (\mathbb{E}|X|_H^2)^{\frac{1}{2}}$.
- $\lfloor x \rfloor$ denotes the integral part of the real x .

2. What is quadratic quantization?

Let $(H, (\cdot|\cdot)_H)$ denote a separable Hilbert space. Let $X \in L_H^2(\mathbb{P})$, that is, a random vector $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto H$ (H is endowed with its Borel σ -algebra) such that $\mathbb{E}|X|_H^2 < +\infty$. An N -quantizer (or N -codebook) is defined as a subset

$$\Gamma := \{x_1, \dots, x_N\} \subset H$$

with $\text{card } \Gamma = N$. In numerical applications, Γ is also called *grid*. Then, one can quantize (or simply discretize) X by $q(X)$, where $q : H \mapsto \Gamma$ is a Borel function. It is straightforward that

$$\forall \omega \in \Omega, \quad |X(\omega) - q(X(\omega))|_H \geq d(X(\omega), \Gamma) = \min_{1 \leq i \leq N} |X(\omega) - x_i|_H$$

so that the best pointwise approximation of X is provided by considering for q a nearest neighbor projection on Γ , denoted by Proj_Γ . Such a projection is in one-to-one correspondence with the Voronoi partitions (or diagrams) of H induced by Γ , that is, the Borel partitions of H satisfying

$$C_i(\Gamma) \subset \left\{ \xi \in H : |\xi - x_i|_H = \min_{1 \leq j \leq N} |\xi - x_j|_H \right\} = \overline{C}_i(\Gamma), \quad i = 1, \dots, N,$$

where $\overline{C_i}(\Gamma)$ denotes the closure of $C_i(\Gamma)$ in H (this heavily uses the Hilbert structure). Then,

$$\text{Proj}_\Gamma(\xi) := \sum_{i=1}^N x_i \mathbf{1}_{C_i(\Gamma)}(\xi)$$

is a nearest neighbor projection on Γ . These projections only differ on the boundaries of the Voronoi cells $C_i(\Gamma)$, $i = 1, \dots, N$. All Voronoi partitions have the same boundary contained in the union of the median hyperplanes defined by the pairs (x_i, x_j) , $i \neq j$. Fig. 1.1 represents the Voronoi diagram defined by a (random) 10-tuple in $\mathbb{R}^2$. Then, one defines a Voronoi N -quantization of X by setting for every $\omega \in \Omega$,

$$\hat{X}^\Gamma(\omega) := \text{Proj}_\Gamma(X(\omega)) = \sum_{i=1}^N x_i \mathbf{1}_{C_i(\Gamma)}(X(\omega)).$$

One clearly has, still for every $\omega \in \Omega$, that

$$|X(\omega) - \hat{X}^\Gamma(\omega)|_H = \text{dist}_H(X(\omega), \Gamma) = \min_{1 \leq i \leq N} |X(\omega) - x_i|_H.$$

The mean (quadratic) quantization error is then defined by

$$e(\Gamma, X, H) = \|X - \hat{X}^\Gamma\|_2 = \sqrt{\mathbb{E} \left(\min_{1 \leq i \leq N} |X - x_i|_H^2 \right)}. \quad (2.1)$$

The distribution of $\hat{X}^\Gamma$ as a random vector is given by the N -tuple $(\mathbb{P}(X \in C_i(\Gamma)))_{1 \leq i \leq N}$ of the Voronoi cells. This distribution clearly depends on the choice of the Voronoi

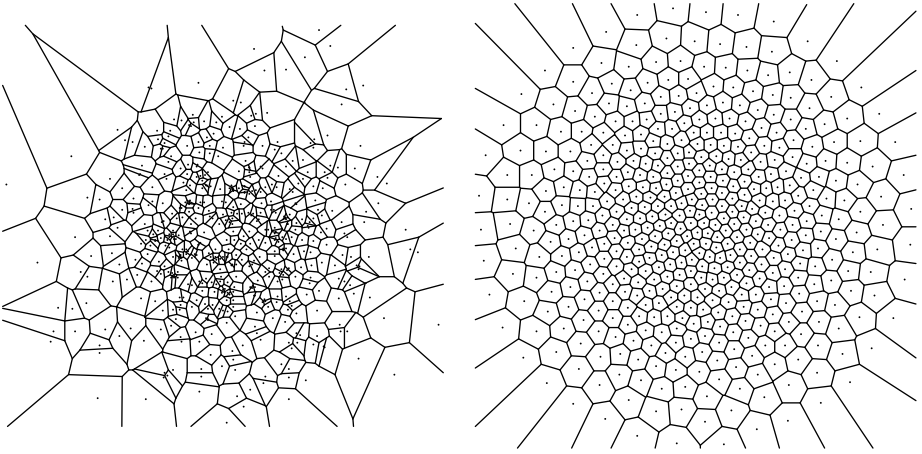


FIG. 2.1 Two N -quantizers (and their Voronoi diagram) related to bi-variate normal distribution $\mathcal{N}(0; I_2)$ ($N = 500$); which one is the best?

partition as emphasized by the following elementary situation: if $H = \mathbb{R}$, the distribution of X is given by $\mathbb{P}_X = \frac{1}{3}(\delta_0 + \delta_{1/2} + \delta_1)$, $N = 2$ and $\Gamma = \{0, 1\}$ since $1/2 \in \partial C_0(\Gamma) \cap \partial C_1(\Gamma)$. However, if $\mathbb{P}_X$ weights no hyperplane, the distribution of $\hat{X}^\Gamma$ depends only on Γ .

As concerns terminology, vector quantization is concerned with the finite-dimensional case, when $\dim H < +\infty$, and is a rather old story, going back to the early 1950s when it was designed in the field of signal processing and then mainly developed in the community of information theory. The term *functional quantization*, probably introduced by LUSCHGY and PAGÈS [2002], PAGÈS [2000], deals with the infinite-dimensional case including the more general Banach-valued setting. The term *functional* comes from the fact that a typical infinite-dimensional Hilbert space is the function space $H = L_T^2$. Then, any (bimeasurable) process $X: ([0, T] \times \Omega, \text{Bor}([0, T]) \otimes \mathcal{A}) \rightarrow (\mathbb{R}, \text{Bor}(\mathbb{R}))$ can be seen as a random vector taking values in the set of Borel functions on $[0, T]$. Furthermore, $((t, \omega) \mapsto X_t(\omega)) \in L^2(dt \otimes d\mathbb{P})$ if and only if $(\omega \mapsto X_\cdot(\omega)) \in L_H^2(\mathbb{P})$ since

$$\int_{[0, T] \times \Omega} X_t^2(\omega) dt \mathbb{P}(d\omega) = \int_{\Omega} \mathbb{P}(d\omega) \int_0^T X_t^2(\omega) dt = \mathbb{E} \|X_\cdot\|_{L_T^2}^2.$$

3. Optimal (quadratic) quantization

At this stage, we are lead to wonder whether it is possible to design some optimally fitted grids to a given distribution $\mathbb{P}_X$, that is, which induce the lowest possible mean quantization error among all grids of size at most N (see e.g. Fig. 2.1). This amounts to the following optimization problem

$$e_N(X, H) := \inf_{\Gamma \subset H, \text{card}(\Gamma) \leq N} e(\Gamma, X, H). \quad (3.1)$$

It is convenient at this stage to make a correspondence between quantizers of size at most N and N -tuples of H^N : to any N -tuple $x := (x_1, \dots, x_N)$ corresponds a quantizer $\Gamma := \Gamma(x) = \{x_i, i = 1, \dots, N\}$ (of size at most N). One introduces the quadratic distortion, denoted by D_N^X , defined by H^N as a (symmetric) function by

$$\begin{aligned} D_N^X &: H^N \longrightarrow \mathbb{R}_+ \\ (x_1, \dots, x_N) &\longmapsto \mathbb{E} \left(\min_{1 \leq i \leq N} |X - x_i|_H^2 \right). \end{aligned}$$

Note that combining (2.1) and the definition of the distortion show that

$$D_N^X(x_1, \dots, x_N) = \mathbb{E} \left(\min_{1 \leq i \leq N} |X - x_i|_H^2 \right) = \mathbb{E} \left(d(X, \Gamma(x))^2 \right) = \|X - \hat{X}^{\Gamma(x)}\|_2^2$$

so that

$$e_N(X, H) = \inf_{(x_1, \dots, x_N) \in H^N} \sqrt{D_N^X(x_1, \dots, x_N)}.$$

The following proposition shows the existence of an optimal N -tuple $x^{(N,*)} \in H^N$ such that $e_N(X, H) = \sqrt{D_N^X(x^{(N,*)})}$. The corresponding optimal quantizer at level N is denoted by $\Gamma^{(N,*)} := \Gamma(x^{(N,*)})$. In finite dimensions, we refer to POLLARD [1982] and in infinite-dimensional settings to CUESTA-ALBERTOS and MATRÁN [1988] and PÁRNA [1990]; one may also refer to PAGÈS [1993], GRAF and LUSCHGY [2000], and LUSCHGY and PAGÈS [2002]. For recent developments on existence and pathwise regularity of optimal quantizer, see GRAF ET AL. [2007].

PROPOSITION 3.1.

- (a) The function D_N^X is lower semicontinuous for the product weak topology on H^N .
- (b) The function D_N^X reaches a minimum at a N -tuple $x^{(N,*)}$ (so that $\Gamma^{(N,*)}$ is an optimal quantizer at level N).
 - If $\text{card}(\text{supp}(\mathbb{P}_X)) \geq N$, the quantizer has full size N (i.e., $\text{card}(\Gamma^{(N,*)}) = N$) and $e_N(X, H) < e_{N-1}(X, H)$.
 - If $\text{card}(\text{supp}(\mathbb{P}_X)) \leq N$, $e_N(X, H) = 0$.

Furthermore, $\lim_N e_N(X, H) = 0$.

- (c) Any optimal (Voronoi) quantization at level N , $\hat{X}^{\Gamma^{(N,*)}}$ satisfies

$$\hat{X}^{\Gamma^{(N,*)}} = \mathbb{E}(X \mid \sigma(\hat{X}^{\Gamma^{(N,*)}})), \quad (3.2)$$

where $\sigma(\hat{X}^{\Gamma^{(N,*)}})$ denotes the σ -algebra generated by $\hat{X}^{\Gamma^{(N,*)}}$.

- (d) Any optimal (quadratic) quantization at level N is a best least square (i.e., $L^2(\mathbb{P})$) approximation of X among all H -valued random variables taking at most N values:

$$e_N(X, H) = \|X - \hat{X}^{\Gamma^{(N,*)}}\|_2 = \min\{\|X - Y\|_2, Y : (\Omega, \mathcal{A}) \rightarrow H, \text{card}(Y(\Omega)) \leq N\}.$$

PROOF. (sketch of):

- (a) The claim follows from the l.s.c. of $\xi \mapsto |\xi|_H$ for the weak topology and Fatou's lemma.
- (b) One proceeds by induction on N . If $N = 1$, the optimal one-quantizer is $x^{(N,*)} = \{\mathbb{E} X\}$ and $e_2(X, H) = \|X - \mathbb{E} X\|_2$.

□

Assume now that an optimal quantizer $x^{(N,*)} = (x_1^{(N,*)}, \dots, x_N^{(N,*)})$ does exist at level N .

- If $\text{card}(\text{supp}(\mathbb{P})) \leq N$, then the $N + 1$ -tuple $(x^{(N,*)}, x_N^{(N,*)})$ (among other possibilities) is also optimal at level $N + 1$ and $e_{N+1}(X, H) = e_N(X, H) = 0$.
- Otherwise, $\text{card}(\text{supp}(\mathbb{P})) \geq N + 1$, hence $x^{(N,*)}$ has pairwise distinct components and there exists $\xi_{N+1} \in \text{supp}(\mathbb{P}_X) \setminus \{x_i^{(N,*)}, i = 1, \dots, N\} \neq \emptyset$.

Then, with obvious notations,

$$D_{N+1}^X((x^{(N,*)}, \xi_{N+1})) < D_N^X(x^{(N,*)}).$$

Then, the set $F_{N+1} := \{x \in H^{N+1} \mid D_{N+1}^X(x) \leq D_{N+1}^X((x^{(N,*)}, \xi_{N+1}))\}$ is nonempty, weakly closed since D_{N+1}^X is l.s.c.. Furthermore, it is bounded in H^{N+1} . Otherwise, there would exist a sequence $x_{(m)} \in H^{N+1}$ such that $|x_{(m), i_m}|_H = \max_i |x_{(m), i}|_H \rightarrow +\infty$ as $m \rightarrow \infty$. Then, by Fatou's lemma, one checks that

$$\liminf_{m \rightarrow \infty} D_{N+1}^X(x_{(m)}) \geq D_N^X(x^{(N,*)}) > D_{N+1}^X((x^{(N,*)}, \xi_{N+1})).$$

Consequently, F_{N+1} is weakly compact and the minimum of D_{N+1}^X on F_{N+1} is clearly its minimum over the whole space H^{N+1} . In particular,

$$e_{N+1}(X, H) \leq D_{N+1}^X((x^{(N,*)}, \xi_{N+1})) < e_N(X, H).$$

If $\text{card}(\text{supp}(\mathbb{P})) = N + 1$, set $x^{(N+1,*)} = \text{supp}(\mathbb{P})$ (as sets) so that $t X = \hat{X}^{\Gamma^{(N+1,*)}}$, which implies $e_{N+1}(X, H) = 0$.

To establish that $e_N(X, H)$ goes to 0, one considers an everywhere dense sequence $(z_k)_{k \geq 1}$ in the separable space H . Then, $d(\{z_1, \dots, z_N\}, X(\omega))$ goes to 0 as $N \rightarrow \infty$ for every $\omega \in \Omega$. Furthermore, $d(\{z_1, \dots, z_N\}, X(\omega))^2 \leq |X(\omega) - z_1|_H^2 \in L^1(\mathbb{P})$. One concludes by the Lebesgue dominated convergence theorem that $D_N^X(z_1, \dots, z_N)$ goes to 0 as $N \rightarrow \infty$.

(c) and (d) temporarily set $\hat{X}^* := \hat{X}^{\Gamma^{(N,*)}}$ for convenience. Let $Y : (\Omega, \mathcal{A}) \rightarrow H$ be a random vector taking at most N values. Set $\Gamma := Y(\Omega)$. Since $\hat{X}^\Gamma$ is a Voronoi quantization of X induced by Γ ,

$$|X - \hat{X}^\Gamma|_H = d(X, \Gamma) \leq |X - Y|_H$$

so that

$$\|X - \hat{X}^\Gamma\|_2 \leq \|X - Y\|_2.$$

On the other hand, the optimality of $\Gamma^{(N,*)}$ implies

$$\|X - \hat{X}^*\|_2 \leq \|X - \hat{X}^\Gamma\|_2.$$

Consequently,

$$\|X - \hat{X}^*\|_2 \leq \min \{ \|X - Y\|_2, Y : (\Omega, \mathcal{A}) \rightarrow H, \text{card}(Y(\Omega)) \leq N \}.$$

The inequality holds as an equality since $\hat{X}^*$ takes at most N values. Furthermore, considering random vectors of the form $Y = g(\hat{X})$ (which take at most as many values as the size of $\Gamma^{(N,*)}$) shows, going back to the very definition of conditional expectation, that $\hat{X}^* = \mathbb{E}(X \mid \hat{X}^*)$ $\mathbb{P}$ -a.s. $\diamond$

Item (c) introduces a very important notion in (quadratic) quantization.

DEFINITION 3.1. A quantizer $\Gamma \subset H$ is *stationary* (or *self-consistent*) if (there is a nearest-neighbor projection such that $\hat{X}^\Gamma = \text{Proj}_\Gamma(X)$ satisfying)

$$\hat{X}^\Gamma = \mathbb{E}(X | \hat{X}^\Gamma). \quad (3.3)$$

Note, in particular, that any stationary quantization satisfies $\mathbb{E}X = \mathbb{E}\hat{X}^\Gamma$.

As shown by Proposition 3.1(c) any quadratic optimal quantizer at level N is stationary. Usually, at least when $d \geq 2$, there are other stationary quantizers: indeed, the distortion function D_N^X is $|\cdot|_H$ -differentiable at N -quantizers $x \in H^N$ with pairwise distinct components and

$$\nabla D_N^X(x) = 2 \left(\int_{C_i(x)} (x_i - \xi) \mathbb{P}_X(d\xi) \right)_{1 \leq i \leq N} = 2 \left(\mathbb{E}(\hat{X}^{\Gamma(x)} - X) \mathbf{1}_{\{\hat{X}^{\Gamma(x)} = x_i\}} \right)_{1 \leq i \leq N}.$$

Hence, any critical point of D_N^X is a stationary quantizer.

REMARKS and COMMENTS.

- In fact (see GRAF and LUSCHGY [2000], theorem 4.2, pp 38), the Voronoi partitions of $\Gamma^{(N,*)}$ always have a $\mathbb{P}_X$ -negligible boundary so that (3.3) holds for any Voronoi diagram induced by Γ .
- The problem of the uniqueness of optimal quantizer (viewed as a set) is not mentioned in the above proposition. In higher dimension, this essentially never occurs. In one dimension, uniqueness of the optimal N -quantizer was first established by FLEISCHER [1964] with strictly log-concave density function. This was successively extended by KIEFFER [1983] and TRUSHKIN [1982] and lead to the following criterion (for more general “loss” functions than the square function):

If the distribution of X is absolutely continuous with a log-concave density function, then, for every $N \geq 1$, there exists only one stationary quantizer of size N , which turns out to be the optimal quantizer at level N .

More recently, a more geometric approach to uniqueness based on the Mountain Pass lemma first developed by LAMBERTON and PAGÈS [1996] and then generalized by COHORT [1998] provided a slight extension of the above criterion (in terms of loss functions).

This log-concavity assumption is satisfied by many families of probability distributions like the uniform distribution on compact intervals, the normal distributions, and the gamma distributions. There are examples of distributions with a non-log-concave density function having a unique optimal quantizer for every $N \geq 1$ (see the Pareto distribution in FORT and PAGÈS [2004]). On the other hand, simple examples of scalar distributions having multiple optimal quantizers at a given level can be found in the study by GRAF and LUSCHGY [2000].

- A stationary quantizer can be suboptimal. This will be emphasized in Section 6 for the Brownian motion (but it is also true for finite-dimensional Gaussian random

vectors), where some families of suboptimal quantizers—the product quantizers designed from the Karhunen–Loève (K-L) basis—are stationary quantizers.

- For the uniform distribution over an interval $[a, b]$, there is a closed form for the optimal quantizer at level N given by $\Gamma^{(N,*)} = \{a + (2k - 1)\frac{b-a}{N}, k = 1, \dots, N\}$. This N -quantizer is optimal not only in the quadratic case but also for any L^r -quantization (see a definition further on). In general, there is no such closed form, either in one or in higher dimension. However, a study FORT and PAGÈS [2004] obtained some semiclosed forms for several families of (scalar) distributions including the exponential and the Pareto distributions: all the optimal quantizers can be expressed using a single underlying sequence $(a_k)_{k \geq 1}$ defined by an induction $a_{k+1} = F(a_k)$.
- In one dimension, as soon as the optimal quantizer at level N is unique (as a set or as an N -tuple with increasing components), it is generally possible to compute it as the solution of the stationarity Eq. (3.2) either by a zero search (Newton–Raphson gradient descent) or by a fixed point (like the specific Lloyd I procedure, see KIEFFER [1982]) procedure.
- In higher dimension, deterministic optimization methods become intractable, and one uses stochastic procedures to compute optimal quantizers. We decided to postpone the short overview on these aspects to Section 6, devoted to the optimal functional quantization of the Brownian motion, where the case of Gaussian vectors with (diagonal) covariance matrix is considered. All stochastic optimization approaches rely on some repeated nearest-neighbor searches: our procedures include some fast (exact) algorithms for that purpose (like K - d -tree, see FRIEDMAN, BENTLEY and FINKEL [1977]). So far, the most efficient methods are also based on the so-called splitting method, which increases progressively the quantization level N (this must be understood when looking for a systematic quantization of a distribution). This method is directly inspired by the induction developed in the proof of claim (b) of Proposition 3.1 since one designs the starting value of the optimization procedure at size $N + 1$ by “merging” the optimized N -quantizer obtained at level N with one further point of $\mathbb{R}^d$, usually randomly sampled with respect to an appropriate distribution (see PAGÈS and PRINTEMPS [2003] for a discussion). For normal distributions $\mathcal{N}(0; I_d)$, alternative starting values living on a sphere with an appropriate radius seem to yield the same accuracy for a given size N without splitting (see PAGÈS and SAGNA [2007]).
- As concerns functional quantization, for example, $H = L^2_t$, there is a close connection between the regularity of optimal (or even stationary) quantizers and that of $t \mapsto X_t$ from $[0, T]$ into $L^2(\mathbb{P})$. Furthermore, as concerns optimal quantizers of Gaussian processes, one shows (see LUSCHGY and PAGÈS [2002]) that they belong to the reproducing space of their covariance operator, for example, to the Cameron–Martin space $H^1 = \{\int_0^\cdot \dot{h}_s ds, \dot{h} \in L^2_t\}$ when $X = W$. Other properties of optimal quantization of Gaussian processes are established by LUSCHGY and PAGÈS [2002].

Extensions to the $L^r(\mathbb{P})$ -quantization of random variables. In this chapter, we focus on the purely quadratic framework (L^2_t and $L^2(\mathbb{P})$ -norms), essentially because it is a natural (and somewhat easier) framework for the computation of optimized grids for

the Brownian motion and for some first applications (like the pricing of path-dependent options, see Section 8). But a more general and natural framework is to consider the functional quantization of random vectors taking values in a separable Banach space $(E, |\cdot|_E)$. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, |\cdot|_E)$ such that $\mathbb{E}|X|_E^r < +\infty$ for some $r \geq 1$ (the case $0 < r < 1$ can also be taken into consideration).

The N -level $(L^r(\mathbb{P}), |\cdot|_E)$ -quantization problem for $X \in L^r_E(\mathbb{P})$ reads

$$e_{N,r}(X, E) := \inf \{ \|X - \hat{X}^\Gamma\|_r, \Gamma \subset E, \text{card}(\Gamma) \leq N \}.$$

The main examples for $(E, |\cdot|_E)$ are the non-Euclidean norms on $\mathbb{R}^d$, the functional spaces $L^p_T(\mu) := L^p([0, T], \mu(dt))$, $1 \leq p \leq \infty$, equipped with its usual norm, $(E, |\cdot|_E) = (\mathcal{C}([0, T]), \|\cdot\|_{\sup})$, etc. As concerns the existence of an optimal quantizer, it holds true for reflexive Banach spaces (see PÄRMA [1990]) and $E = L^1_T$, but otherwise it may fail even when $N = 1$ (see GRAF, LUSCHGY and PAGÈS [2007]). In finite dimension, the Euclidean feature is not crucial (see GRAF and LUSCHGY [2000]). In the functional setting, many results originally obtained in a Hilbert setting have been extended to the Banach setting either for existence or for regularity results (see GRAF, LUSCHGY and PAGÈS [2007]) or for rates (see DEREICH [2005a], DEREICH and SCHEUTZOW [2006], LUSCHGY and PAGÈS [2004], LUSCHGY and PAGÈS [2007]).

4. Cubature formulae: conditional expectation and numerical integration

Let $F : H \rightarrow \mathbb{R}$ be a continuous functional (with respect to the norm $|\cdot|_H$) and let $\Gamma \subset H$ be an N -quantizer. It is natural to approximate $\mathbb{E}(F(X))$ by $\mathbb{E}(F(\hat{X}^\Gamma))$. This quantity $\mathbb{E}(F(\hat{X}^\Gamma))$ is simply the finite-weighted sum

$$\mathbb{E}(F(\hat{X}^\Gamma)) = \sum_{i=1}^N F(x_i) \mathbb{P}(\hat{X}^\Gamma = x_i). \quad (4.1)$$

Numerical computation of $\mathbb{E}(F(\hat{X}^\Gamma))$ is possible as soon as $F(\xi)$ can be computed at any $\xi \in H$ and the distribution $(\mathbb{P}(\hat{X}^\Gamma = x_i))_{1 \leq i \leq N}$ of $\hat{X}^\Gamma$ is known. The induced quantization error $\|X - \hat{X}^\Gamma\|_2$ is used to control the error (see below). These quantities related to the quantizer Γ are also called *companion parameters*.

Likewise, one can consider a priori the $\sigma(\hat{X}^\Gamma)$ -measurable random variable $F(\hat{X}^\Gamma)$ as a good approximation of the conditional expectation $\mathbb{E}(F(X) | \hat{X}^\Gamma)$.

4.1. Lipschitz functionals

Assume that the functional F is Lipschitz continuous on H . Then,

$$\left| \mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma) \right| \leq [F]_{\text{Lip}} \mathbb{E}(|X - \hat{X}^\Gamma| | \hat{X}^\Gamma)$$

so that, for every real exponent $r \geq 1$,

$$\|\mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma)\|_r \leq [F]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_r$$

(where we applied conditional Jensen inequality to the convex function $u \mapsto u^r$). In particular, using $\mathbb{E} F(X) = \mathbb{E}(\mathbb{E}(F(X) | \hat{X}^\Gamma))$, one derives (with $r = 1$) that

$$\begin{aligned} \left| \mathbb{E} F(X) - \mathbb{E} F(\hat{X}^\Gamma) \right| &\leq \|\mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma)\|_1 \\ &\leq [F]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_1. \end{aligned}$$

Finally, using the monotony of the $L^r(\mathbb{P})$ -norms as a function of r yields

$$\left| \mathbb{E} F(X) - \mathbb{E} F(\hat{X}^\Gamma) \right| \leq [F]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_1 \leq [F]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_2. \quad (4.2)$$

In fact, considering the Lipschitz functional $F(\xi) := d(\xi, \Gamma)$ shows that

$$\|X - \hat{X}^\Gamma\|_1 = \sup_{[F]_{\text{Lip}} \leq 1} \left| \mathbb{E} F(X) - \mathbb{E} F(\hat{X}^\Gamma) \right|. \quad (4.3)$$

By the Lipschitz functionals making up a characterizing family for the weak convergence of probability measures on H , one derives that, for any sequence of N -quantizers Γ^N satisfying $\|X - \hat{X}^{\Gamma^N}\|_1 \rightarrow 0$ as $N \rightarrow \infty$,

$$\sum_{1 \leq i \leq N} \mathbb{P}(\hat{X}^{\Gamma^N} = x_i^N) \delta_{x_i^N} \xrightarrow{(H)} \mathbb{P}_X,$$

where $\xrightarrow{(H)}$ denotes the weak convergence of probability measures on $(H, |\cdot|_H)$.

4.2. Convex functionals

If $F : H \rightarrow \mathbb{R}$ is a convex functional and $\hat{X}$ is a stationary quantization of X , a straightforward application of Jensen inequality yields

$$\mathbb{E}(F(X) | \hat{X}) \geq F(\hat{X})$$

so that $\mathbb{E}(F(\hat{X})) \leq \mathbb{E}(F(X))$.

4.3. Differentiable functionals with Lipschitz differentials

Assume now that F is differentiable on H , with a Lipschitz continuous differential DF , and that the quantizer Γ is stationary (see Eq. (3.3)). A Taylor expansion yields

$$\left| F(X) - F(\hat{X}^\Gamma) - DF(\hat{X}^\Gamma).(X - \hat{X}^\Gamma) \right| \leq [DF]_{\text{Lip}} |X - \hat{X}^\Gamma|^2.$$

Taking conditional expectation, given $\hat{X}^\Gamma$, yields

$$\left| \mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma) - \mathbb{E}(DF(\hat{X}^\Gamma).(X - \hat{X}^\Gamma) | \hat{X}^\Gamma) \right| \leq [DF]_{\text{Lip}} \mathbb{E}(|X - \hat{X}^\Gamma|^2 | \hat{X}^\Gamma).$$

Now, using that the random variable $DF(\hat{X}^\Gamma)$ is $\sigma(\hat{X}^\Gamma)$ -measurable, one has

$$\mathbb{E}(DF(\hat{X}^\Gamma).(X - \hat{X}^\Gamma)) = \mathbb{E}(DF(\hat{X}^\Gamma).\mathbb{E}(X - \hat{X}^\Gamma | \hat{X}^\Gamma)) = 0$$

so that

$$\left| \mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma) \right| \leq [DF]_{\text{Lip}} \mathbb{E}(|X - \hat{X}^\Gamma|^2 | \hat{X}^\Gamma).$$

Then, for every real exponent $r \geq 1$,

$$\left\| \mathbb{E}(F(X) | \hat{X}^\Gamma) - F(\hat{X}^\Gamma) \right\|_r \leq [DF]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_{2r}^2. \quad (4.4)$$

In particular, when $r = 1$, one derives like in the former setting

$$\left| \mathbb{E}F(X) - \mathbb{E}F(\hat{X}^\Gamma) \right| \leq [DF]_{\text{Lip}} \|X - \hat{X}^\Gamma\|_2^2. \quad (4.5)$$

In fact, the above inequality holds provided F is $\mathcal{C}^1$ with Lipschitz differential on every Voronoi cell $C_i(\Gamma)$. A characterization similar to (4.3) based on these functionals could be established.

Some variant of these cubature formulae can be found in PAGÈS and PRINTEMS [2003] or GRAF ET AL. [2006] for functions or functionals F having only some local Lipschitz regularity.

4.4. Quantized approximation of $\mathbb{E}(F(X) | Y)$

Let X and Y be two H -valued random vectors defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $F : H \rightarrow \mathbb{R}$ be a Borel functional. The natural idea is to approximate $\mathbb{E}(F(X) | Y)$ by the quantized conditional expectation $\mathbb{E}(F(\hat{X}) | \hat{Y})$, where $\hat{X}$ and $\hat{Y}$ are quantizations of X and Y , respectively.

Let $\varphi_F : H \rightarrow \mathbb{R}$ be a (Borel) version of the conditional expectation, that is, satisfying

$$\mathbb{E}(F(X) | Y) = \varphi_F(Y).$$

Usually, no closed form is available for the function φ_F but some regularity property can be established, especially in a (Feller) Markovian framework. Thus, assume that both F and φ_F are Lipschitz continuous with Lipschitz coefficients $[F]_{\text{Lip}}$ and $[\varphi_F]_{\text{Lip}}$. Then,

$$\mathbb{E}(F(X) | Y) - \mathbb{E}(F(\hat{X}) | \hat{Y}) = \mathbb{E}(F(X) | Y) - \mathbb{E}(F(X) | \hat{Y}) + \mathbb{E}(F(X) - F(\hat{X}) | \hat{Y}).$$

Hence, assuming that $\hat{Y}$ is $\sigma(Y)$ -measurable and that conditional expectation is an L^2 -contraction,

$$\begin{aligned} \|\mathbb{E}(F(X) | Y) - \mathbb{E}(F(X) | \hat{Y})\|_2 &= \|\mathbb{E}(F(X) | Y) - \mathbb{E}(\mathbb{E}(F(\hat{X}) | \hat{Y}) | Y)\|_2 \\ &\leq \|\varphi_F(Y) - \mathbb{E}(F(X) | \hat{Y})\|_2 \\ &= \|\varphi_F(Y) - \mathbb{E}(\varphi_F(Y) | \hat{Y})\|_2 \\ &\leq \|\varphi_F(Y) - \varphi_F(\hat{Y})\|_2. \end{aligned}$$

The last inequality follows from the definition of conditional expectation, given $\hat{Y}$ as the best quadratic approximation among $\sigma(\hat{Y})$ -measurable random variables. On the other hand, still assuming that $\mathbb{E}(\cdot | \sigma(\hat{Y}))$ is an L^2 -contraction and this time that F is Lipschitz continuous yields

$$\|\mathbb{E}(F(X) - F(\hat{X}) | \hat{Y})\|_2 \leq \|F(X) - F(\hat{X})\|_2 \leq [F]_{\text{Lip}} \|X - \hat{X}\|_2.$$

Finally,

$$\|\mathbb{E}(F(X) | Y) - \mathbb{E}(F(\hat{X}) | \hat{Y})\|_2 \leq [F]_{\text{Lip}} \|X - \hat{X}\|_2 + [\varphi_F]_{\text{Lip}} \|Y - \hat{Y}\|_2.$$

In the nonquadratic case, the above inequality remains valid provided $[\varphi_F]_{\text{Lip}}$ is replaced by $2[\varphi_F]_{\text{Lip}}$.

5. Vector quantization

5.1. Vector quantization rate ($H = \mathbb{R}^d$)

The fact that $e_N(X, \mathbb{R}^d)$ is a nonincreasing sequence that goes to 0 as N goes to ∞ is a rather simple result established in Proposition 3.1. Its (sharp) rate of convergence to 0 is a much more challenging problem. An answer is provided by the so-called Zador theorem stated below.

This theorem was first stated and established for distributions with compact supports by Zador (see ZADOR [1963, 1982]). Then, a first extension to general probability distributions on $\mathbb{R}^d$ is developed by BUCKLEW and WISE [1982]. The first mathematically rigorous proof can be found in a study by GRAF and LUSCHGY [2000], and relies on a random quantization argument (called upon in a step of the proof sometimes called Pierce lemma). We also provide a nonasymptotic error bound that can be seen as simple reformulation of this Pierce lemma. It turns out to be very useful for applications.

THEOREM 5.1 (a) SHARP RATE (see GRAF and LUSCHGY [2000]). *Let $r > 0$ and $X \in L^{r+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi) d\xi + \nu(d\xi)$ be the canonical decomposition of the distribution of X (ν and the Lebesgue measure are singular). Then (if $\varphi \not\equiv 0$),*

$$e_{N,r}(X, \mathbb{R}^d) \sim \tilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+r}}(u) du \right)^{\frac{1}{d} + \frac{1}{r}} \times N^{-\frac{1}{d}} \quad \text{as } N \rightarrow +\infty, \quad (5.1)$$

where $\tilde{J}_{r,d} \in (0, \infty)$.

(b) NONASYMPTOTIC UPPER BOUND (see LUSCHGY and PAGÈS [2007]). *Let $d \geq 1$. There exists $C_{d,r,\eta} \in (0, \infty)$ such that, for every $\mathbb{R}^d$ -valued random vector X ,*

$$\forall N \geq 1, \quad e_{N,r}(X, \mathbb{R}^d) \leq C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}.$$

REMARKS.

- The real constant $\tilde{J}_{r,d}$ clearly corresponds to the case of the uniform distribution over the unit hypercube $[0, 1]^d$ for which the slightly more precise statement holds

$$\lim_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \inf_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = \tilde{J}_{r,d}.$$

The proof is based on a self-similarity argument. The value of $\tilde{J}_{r,d}$ depends on the reference norm on $\mathbb{R}^d$. When $d = 1$, elementary computations show that $\tilde{J}_{r,1} = (r+1)^{-\frac{1}{r}}/2$. When $d = 2$, with the canonical Euclidean norm, one shows (see NEWMAN [1982] for a proof (see also GRAF and LUSCHGY [2000]) that $\tilde{J}_{2,d} = \sqrt{\frac{5}{18\sqrt{3}}}$. Its exact value is unknown for $d \geq 3$ but, still for the canonical Euclidean norm, one has (see GRAF and LUSCHGY [2000]) using some random quantization arguments,

$$\tilde{J}_{2,d} \sim \sqrt{\frac{d}{2\pi e}} \approx \sqrt{\frac{d}{17,08}} \quad \text{as } d \rightarrow +\infty.$$

- When $\varphi \equiv 0$, the distribution of X is purely singular. The rate (5.1) still holds in the sense that $\lim_N N^{\frac{1}{d}} e_{N,r}(X, \mathbb{R}^d) = 0$. Consequently, this is not the right asymptotics. The quantization problem for singular measures (like uniform distribution on fractal compact sets) has been extensively investigated by several authors, leading to the definition of a quantization dimension in connection with the rate of convergence of the quantization error on these sets. For more details, we refer to GRAF and LUSCHGY [2000], GRAF and LUSCHGY [2005] and the references therein.
- A more naive way to quantize the uniform distribution on the unit hypercube is to proceed by product quantization, that is, by quantizing the marginals of the uniform distribution. If $N = m^d$, $m \geq 1$, one easily proves that the best quadratic product quantizer (for the canonical Euclidean norm on $\mathbb{R}^d$) is the “midpoint square grid”

$$\Gamma^{sq,N} = \left(\frac{2i_1 - 1}{2m}, \dots, \frac{2i_d - 1}{2m} \right)_{1 \leq i_1, \dots, i_d \leq m},$$

which induces a quadratic quantization error equal to

$$\sqrt{\frac{d}{12}} \times N^{-\frac{1}{d}}.$$

Consequently, product quantizers are still rate optimal in every dimension d . Moreover, note that the ratio of these two rates remains bounded as $d \uparrow \infty$.

- For a brief discussion and comparison with quasi-Monte Carlo methods, we refer to PAGÈS [2007] and the references therein. Let us simply recall that sequences (or sets) with low discrepancy are uniformly distributed sequence over the unit d -dimensional hypercube $[0, 1]^d$. When used instead of (pseudo-)random numbers to integrate a function f with bounded variations, the rate of convergence is

theoretically “almost dimension free”: it is the product of the variation of f by the discrepancy, which behaves like $O\left(\frac{\log(N)^d}{N}\right)$ (for sequences). However, such functions become less and less “standard” in higher dimension. When implemented with Lipschitz continuous functions, the quasi-Monte Carlo (QMC) method does face the curse of dimensionality with theoretical performances which seem to be worse than optimal quantizers of the uniform distribution over $[0, 1]^d$, namely $O\left(\frac{\log(N)}{N^{\frac{1}{d}}}\right)$ (still for sequences) owing to Proinov’s theorem (PROINOV [1988]).

- The nonasymptotic Zador theorem stated and established by LUSCHGY and PAGÈS [2007] is essentially a variant of the so-called Pierce lemma (see GRAF and LUSCHGY [2000]). Many developments and heuristics about the rate of convergence of some quantization-based algorithms for American option pricing, stochastic control, or nonlinear filtering (see PAGÈS, PHAM and PRINTEMS [2003]) can be significantly simplified or established rigorously by calling upon this result. This is emphasized by the example below devoted to swing options.

5.2. Examples of application of optimal vector quantization

5.2.1. Numerical integration (II): Richardson–Romberg extrapolation versus curse of dimensionality

Combining the above cubature formula (4.1) and the rate of convergence of the (optimal) quantization error, the theoretical critical dimension to use quantization-based cubature formulae seems to be $d = 4$ when compared to Monte Carlo simulation (at least for continuously differentiable functions). Several numerical tests have been carried out and reported by PAGÈS, PHAM and PRINTEMS [2003] and PAGÈS and PRINTEMS [2003] to evaluate more precisely the effect of the so-called curse of dimensionality. The benchmark was made of several European payoffs on a geometric index made of d independent assets in a Black–Scholes model: *vanilla put and put spread options and their smoothed versions*. No control variate was used. The absence of correlation is not a realistic assumption in finance but is clearly more challenging as a benchmark for numerical integration. Once the dimension d and the quantizer size N have been chosen, we compared the resulting integration error to a symmetric confidence interval with total length equal to two standard deviations of a Monte Carlo (MC) estimator based on N simulated data $\left(\frac{\sigma_{\text{payoff}}}{\sqrt{N}}\right)$. Furthermore, σ_{payoff} has been computed by a Monte Carlo simulation on 10^4 simulated data of the payoff.

The results turned out to be more favorable to quantization than predicted by theoretical bounds, mainly because we carried out our tests with rather small values of N , whereas curse of dimensionality is an asymptotic bound. Until the dimension 4, the larger N is, the more quantization outperforms MC simulation. When the dimension $d \geq 5$, quantization always outperforms MC (in the above sense) until a critical size $N_c(d)$, which decreases as d increases.

RICHARDSON–ROMBERG (R-R) EXTRAPOLATION. In this section, we provide a method to push ahead these critical sizes, at least for *smooth enough functionals*. Let $F : \mathbb{R}^d \rightarrow \mathbb{R}$ be a twice differentiable functional with Lipschitz–Hessian $D^2 F$. Let $(\hat{X}^{(N)})_{N \geq 1}$ be a

sequence of optimal quadratic quantizations. Then,

$$\mathbb{E}(F(X)) = \mathbb{E}(F(\hat{X}^{(N)})) + \frac{1}{2}\mathbb{E}\left(D^2 F(\hat{X}^{(N)}).(X - \hat{X}^{(N)})^{\otimes 2}\right) + O\left(\mathbb{E}|X - \hat{X}|^3\right) \quad (5.2)$$

Under some assumptions that are satisfied by most usual distributions (including the normal one), it is proved by GRAF, LUSCHGY and PAGÈS [2006] as a special case of a general theorem about the asymptotic behavior of L^s of sequences of optimal L^r -quantizers for $s \in (r, r + d)$ that

$$\mathbb{E}|X - \hat{X}|^3 = O(N^{-\frac{3}{d}}) \quad \text{if } d \geq 2,$$

or $\mathbb{E}|X - \hat{X}|^3 = O(N^{-\frac{3-\varepsilon}{d}})$, $\varepsilon > 0$, if $d = 2$. Furthermore, if we make the conjecture that

$$\mathbb{E}(D^2 F(\hat{X}^{(N)}).(X - \hat{X}^{(N)})^{\otimes 2}) = c_{F,X} N^{-\frac{2}{d}} + O(N^{-\frac{3}{d}}), \quad (5.3)$$

it becomes possible to implement an R-R extrapolation to compute $\mathbb{E}(F(X))$. Namely, one considers two sizes N_1 and N_2 (in practice, one often sets $N_1 = N/2$ and $N_2 = N$). Then, combining (5.2) with N_1 and N_2 ,

$$\mathbb{E}(F(X)) = \frac{N_2^{\frac{2}{d}} \mathbb{E}(F(\hat{X}^{(N_2)})) - N_1^{\frac{2}{d}} \mathbb{E}(F(\hat{X}^{(N_1)}))}{N_2^{\frac{2}{d}} - N_1^{\frac{2}{d}}} + O\left(\frac{1}{(N_1 \wedge N_2)^{\frac{1}{d}} (N_2^{\frac{2}{d}} - N_1^{\frac{2}{d}})}\right).$$

In Section 8.1, a similar procedure is tested in an infinite-dimensional setting: $\mathbb{R}^d$ is replaced by the Hilbert space $H = L^2([0, T], dt)$ viewed as a state of paths for a stochastic process X (namely, the Brownian motion).

NUMERICAL ILLUSTRATION: In order to evaluate the effect of the R-R technique described above, numerical computations have been carried out in the case of the regularized versions of some put spread options on geometric indices in dimension $d = 4, 6, 8, 10$. By “regularized,” we mean that the payoff at maturity $T = 1$ has been replaced by its price function at time $T' < T$. Numerical integration was performed using the Gaussian optimal grids of size $N = 2^k$, $k = 2, \dots, 12$ (available at the Web site www.quantize.maths-fi.com).

We consider again one of the test functions implemented by PAGÈS and PRINTEMS [2003] (pp 152). These test functions were borrowed from classical option pricing in mathematical finance: one considers d independent traded assets $S^1, \dots, S^d$ following a d -dimensional Black–Scholes dynamics (under its risk-neutral probability)

$$S_t^i = s_0^i \exp\left((r - \frac{\sigma^2}{2})t + \sigma\sqrt{t}Z^{i,t}\right), \quad i = 1, \dots, d,$$

where $Z^{i,t} = W_t^i/\sqrt{t}$ and $W = (W^1, \dots, W^d)$ is a d -dimensional standard Brownian motion. We also assume that $S_0^i = s_0 > 0$, $i = 1, \dots, d$ and that the d assets share the

same volatility $\sigma^i = \sigma > 0$. One considers the geometric index $I_t = (S_t^1 \dots S_t^d)^{\frac{1}{d}}$. One shows that $e^{-\frac{\sigma^2}{2}(\frac{1}{d}-1)t_{J_t}}$ has itself a risk-neutral Black–Scholes dynamics. We want to test the regularized put spread option on this geometric index with strikes $K_1 < K_2$ (at time $T/2$). Let $\psi(s_0, K_1, K_2, r, \sigma, T)$ the premium at time 0 of a put spread on any of the assets S^i .

$$\psi(x, K_1, K_2, r, \sigma, T) = \pi(x, K_2, r, \sigma, T) - \pi(x, K_1, r, \sigma, T)$$

$$\pi(x, K, r, \sigma, T) = Ke^{-rT} \operatorname{erf}(-d_2) - x \operatorname{erf}(-d_1),$$

$$d_1 = \frac{\log(x/K) + (r + \frac{\sigma^2}{2d})T}{\sigma\sqrt{T/d}}, \quad d_2 = d_1 - \sigma\sqrt{T/d}.$$

Using the martingale property of the discounted value of the premium of a European option yields that the premium $e^{-rT} \mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+)$ of the put spread option on I satisfies, on the one hand,

$$e^{-rT} \mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+) = \psi(s_0 e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)T}, K_1, K_2, r, \sigma/\sqrt{d}, T)$$

and, on the other hand,

$$e^{-rT} \mathbb{E}((K_1 - I_T)_+ - (K_2 - I_T)_+) = \mathbb{E} g(Z),$$

where

$$g(Z) = e^{-rT/2} \psi(s_0 e^{\frac{\sigma^2}{2}(\frac{1}{d}-1)\frac{T}{2}} I_{\frac{T}{2}}, K_1, K_2, r, \sigma, T/2)$$

and $Z = (Z^1, \dots, Z^d)^{\frac{T}{2}} \stackrel{d}{=} \mathcal{N}(0; I_d)$. The numerical specifications of the function g are as follows:

$$s_0 = 100, \quad K_1 = 98, \quad K_2 = 102, \quad r = 5\%, \quad \sigma = 20\%, \quad T = 2.$$

The results are shown below (see Fig. 5.1) in a log-log scale for the dimensions $d = 4, 6, 8, 10$.

First, we recover the theoretical rates (namely, $-2/d$) of convergence for the error bounds. Indeed, some slopes $\beta(d)$ can be derived (using a regression) for the quantization errors and we found $\beta(4) = -0.48$, $\beta(6) = -0.33$, $\beta(8) = -0.25$, and $\beta(10) = -0.23$ for $d = 10$ (see Fig. 5.1). These rates plead for the implementation of R-R extrapolation. Also note that, as already reported by PAGÈS and PRINTEMPS [2003], when $d \geq 5$, quantization still outperforms MC simulations (in the above sense) up to a critical number $N_c(d)$ of points ($N_c(6) \sim 5000$, $N_c(7) \sim 1000$, $N_c(8) \sim 500$, etc).

As concerns the R-R extrapolation method itself, note first that it always gives better results than crude quantization. As regards, the comparison with Monte Carlo simulation, no critical number of points $N_{\text{Romb}}(d)$ comes out beyond which MC simulation outperforms R-R extrapolation. This means that $N_{\text{Romb}}(d)$ is greater than the range of use of quantization-based cubature formulas in our benchmark, namely 5000.

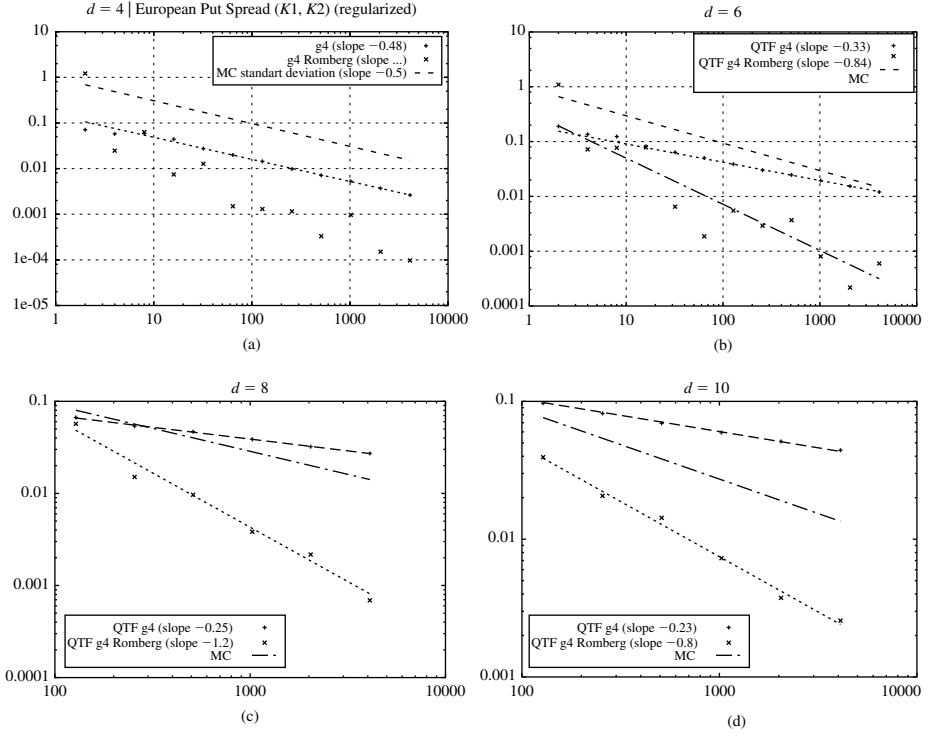


FIG. 5.1 Errors and standard deviations as functions of the number of points N in a log-log scale. The quantization error is shown by the symbol $+$ and the R - R extrapolation error by the symbol $\times$. The dashed line without crosses denotes the standard deviation of the Monte Carlo estimator. (a) $d = 4$, (b) $d = 6$, (c) $d = 8$, and (d) $d = 10$.

The R - R extrapolation techniques are commonly known to be unstable, and indeed, it has not been always possible to estimate satisfactorily its rate of convergence on our benchmark. But when a significant slope (in a log-log scale) can be estimated from the R - R errors (like for $d = 8$ and $d = 10$ in Fig. 5.1 (c), (d)), its absolute value is larger than $1/2$, and so, these extrapolations always outperform the MC method even for large values of N . As a by-product, our results plead in favor of the conjecture (5.3) and lead to think that R - R extrapolation is a powerful tool to accelerate numerical integration by optimal quantization, even in higher dimension.

5.3. An application to the pricing of swing options

Optimal-quantization-based algorithms have been already devised to solve several multi-dimensional nonlinear problems, from multiasset American style options (BALLY, PAGÈS and PRINTEMS [2001, 2003, 2005], BALLY and PAGÈS [2003a,b] to nonlinear filtering and portfolio management (see PAGÈS, PHAM and PRINTEMS [2003]). Here, we present a new application developed by Gaz de France (French gas company) to price swing options contracts. For a detailed version of this section, we refer to the original works

by BARDOU, BOUTHEMY and PAGÈS [2007a,b]. The holder of such a contract daily purchases a quantity of gas, say q_{t_k} at time t_k , $k = 0, \dots, n-1$, at a price K_{t_k} , which can be deterministic or random (e.g., an oil-based index). These quantities are subject to two kinds of constraints, some local daily constraints $q_{\min} \leq q_{t_k} \leq q_{\max}$ and some global constraint about the total amount of purchased gas $Q_{\min} \leq q_0 + \dots + q_{t_{n-1}} \leq Q_{\max}$ (with $n q_{\min} \leq Q_{\min} \leq Q_{\max} \leq n q_{\max}$). The spot price S_{t_k} at time t_k of gas is usually not quoted on gas markets but is usually approximated by the day ahead price. The dynamics of the gas price itself is usually multifactorial, which makes it non-Markovian. However, it depends on a multidimensional underlying Markov structure process, which can be Gaussian or not. For the sake of simplicity, we assume here that (S_{t_k}) has a Markov dynamics and that the exercise prices K_{t_k} are deterministic (and there is no interest rate). Then, given $q_{\min}, q_{\max}, Q_{\min}, Q_{\max}$ and if $\bar{q}_{t_k} := q_0 + q_{t_1} + \dots + q_{t_{k-1}}$ denotes the purchased quantity prior to time t_k , the price of this contract at time t_k is given by

$$P(t_k, \bar{q}_{t_k}, S_{t_k}) := \inf \left\{ \mathbb{E} \left(\sum_{\ell=k}^{n-1} q_{t_\ell} (S_{t_\ell} - K_{t_\ell}) \mid S_{t_k} \right), q_{t_\ell} \in \mathcal{F}_{t_{\ell-1}}^S, \right. \\ \left. q_{\min} \leq q_{t_\ell} \leq q_{\max}, Q_{\min} \leq \bar{q}_{t_\ell} \leq Q_{\max} \right\}$$

where $\mathcal{F}_{t_\ell}^S = \sigma(S_0, S_{t_1}, \dots, S_{t_\ell})$ and $\in$ stands for measurability with respect to a σ -field. This formula shows that this pricing problem, is a stochastic control problem, where the purchased quantity process appears as the control variable. This price satisfies the following dynamic programming principle

$$P(t_k, \bar{q}, S_{t_k}) = \sup_{q \in [q_{\min}, q_{\max}]} \left(q(S_{t_k} - K_{t_k}) + \mathbb{E} \left(P(t_{k+1}, \bar{q} + q, S_{t_{k+1}}) \mid S_{t_k} \right) \right).$$

It is shown by BARDOU, BOUTHEMY and PAGÈS [2007b] that the optimal control does exist under mild integrability assumptions but is not always *bang-bang* in general (owing to prediction errors). However, if $\frac{Q_{\max} - n q_{\min}}{q_{\max} - q_{\min}}$ and $\frac{Q_{\min} - n q_{\min}}{q_{\max} - q_{\min}}$ are integers, then the optimal control $q_{t_k}^*$ is always $\{q_{\min}, q_{\max}\}$ -valued. Then, one defines the quantized dynamic programming formula by setting

$$\hat{P}(t_k, Q, \hat{S}_{t_k}) := \max_{q \in [q_{\min}, q_{\max}]} \left(q(S_{t_k} - K_{t_k}) + \mathbb{E} \left(\hat{P}(t_{k+1}, Q + q, \hat{S}_{t_{k+1}}) \mid \hat{S}_{t_k} \right) \right),$$

where $\hat{S}_{t_k}$ is an N_k -quantization of S_{t_k} obtained by a nearest-neighbor projection on an optimal (quadratic) N_k -quantizer $\Gamma_k = \{x_1^k, \dots, x_{N_k}^k\}$. This quantization approach amounts to approximating the transitions $\mathcal{L}(S_{t_{k+1}} \mid S_{t_k})$ by the finitely valued transition $\mathcal{L}(\hat{S}_{t_{k+1}} \mid \hat{S}_{t_k})$. All these grids and transition weights make up a so-called quantization tree. In a Gaussian framework, some grids can be obtained from precomputed normalized optimal grids (available at the Web site www.quantize.maths-fi.com). Otherwise, this quantization optimization step can be performed by some stochastic optimization procedures (like randomized Lloyd's I and *CLVQ* procedures, see Section 6.3 for an example in a Gaussian framework). The next step is the computation of the quantized

transitions $\hat{\pi}_{ij}^k := \mathbb{P}(\hat{S}_{t_{k+1}} = x_j^{k+1} | \hat{S}_{t_k} = x_i^k)$ by a Monte Carlo simulation. Both steps are mainly based on repeated nearest-neighbor searches. They can be carried out off-line since they do not depend on the payoff characteristics (the K_{t_k} s). However, using some fast nearest-neighbor procedures, typically the K - d -tree algorithm introduced by FRIEDMAN, BENTLEY and FINKEL [1977] or some improved versions like (Principal Axis Tree (see McNAMES [2001])) drastically reduces the complexity of this phase (hence its duration) in higher dimension. Now with modern computing devices it becomes possible in many applications to include both phases in the online computations, which makes it as flexible as Monte Carlo-based methods like regression methods.

Assume that $S_0 = s_0 \in (0, \infty)$ and that $\hat{S}_0 = s_0$. It is proved BARDOU, BOUTHEMY and PAGÈS [2007b] that

$$|\hat{P}(0, 0, s_0) - P(0, 0, s_0)| \leq C \sum_{k=0}^{n-1} \|S_{t_k} - \hat{S}_{t_k}\|_2.$$

If all the quantizations $\hat{S}_{t_k}$ of S_{t_k} are optimal (with size $N_k = \bar{N}$), this provides a $O(\frac{n}{\bar{N}^{\frac{1}{d}}})$ rate of convergence. In fact, numerical evidences show that the observed rate is usually

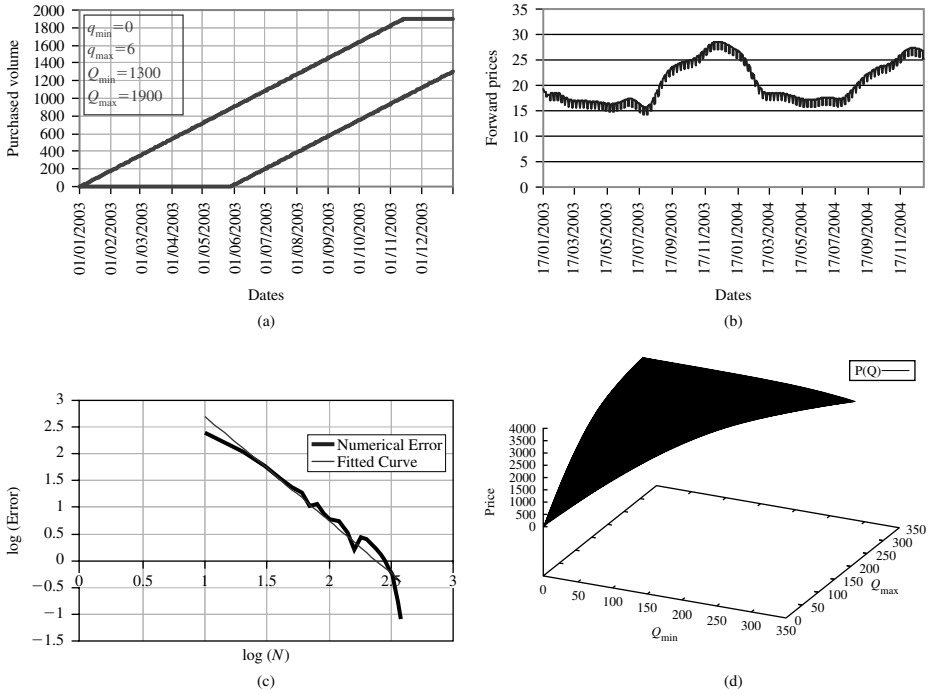


FIG. 5.2 The parameters are those given in the numerical illustration. (a) Constraint set, (b) daily forward curve, (c) numerical convergence as a function of the optimal grid size (log-log scale) and (d) the graph of the price as a function of the global constraints.

$O(\frac{n}{N^{\frac{1}{d}}})$, that is, somewhat similar to the rates obtained in the cubature formula for differentiable functions. The choice of the N_k s can be refined like for American options following the lines of BALLY and PAGÈS [2003a].

A comparison carried out by BARDOU, BOUTHEMY and PAGÈS [2007b] suggests that the quantization approach (including the transition computation) is significantly faster than the least squares regression (LSR) methods à la Longstaff–Schwartz.

NUMERICAL ILLUSTRATION: We consider the one-factor Toy model given by

$$S_t = F_{0,t} \exp\left(\sigma \int_0^t e^{-\alpha(t-s)} dW_s - \frac{1}{2} \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t})\right),$$

where $\sigma = 70\%$, $\alpha = 4$, and $t_k = k/n$. The future prices are real data (January 17, 2003) corresponding to the first part of the curve in Fig. 5.2(b). The contract parameters are $q_{\min} = 0$, $q_{\max} = 6$, $Q_{\min} = 1300$, $Q_{\max} = 1900$, $K_{t_k} = K$, and $n = 365$ (1 year).

Note the slope in Fig. 5.2(d) is $1.96 \approx 2$. The main asset of quantization is to directly approximate the underlying Markov dynamics (in particular, when dealing with multifactor models): it is a model-driven method, which has its ability to “capture” automatically the correlation structure of the asset, which becomes quickly impossible with multinomial trees as the number of factors increases.

6. Optimal quadratic functional quantization of Gaussian processes

Optimal quadratic functional quantization of Gaussian processes is closely related to their so-called K-L expansion, which can be seen in some sense as some infinite-dimensional principal component analysis of a (Gaussian) process. Before stating a general result for Gaussian processes, we start by the standard Brownian motion: it is the most important example in view of (numerical) applications and for this process, everything can be made explicit.

6.1. Brownian motion

One considers the Hilbert space $H = L_T^2 := L^2([0, T], dt)$, $(f|g)_2 = \int_0^T f(t)g(t)dt$, $\|f\|_{L_T^2} = \sqrt{(f|f)_2}$. The covariance operator C_W of the Brownian motion $W = (W_t)_{t \in [0, T]}$ is defined on L_T^2 by

$$C_W(f) := \mathbb{E}((f, W)_2 W) = \left(t \mapsto \int_0^T (s \wedge t) f(s) ds\right).$$

It is a symmetric positive trace class operator, which can be diagonalized in the so-called K-L orthonormal basis $(e_n^W)_{n \geq 1}$ of L_T^2 , with eigenvalues $(\lambda_n)_{n \geq 1}$, given by

$$e_n^W(t) = \sqrt{\frac{2}{T}} \sin\left(\pi\left(n - \frac{1}{2}\right)\frac{t}{T}\right), \quad \lambda_n = \left(\frac{T}{\pi\left(n - \frac{1}{2}\right)}\right)^2, \quad n \geq 1.$$

This classical result can be established as a simple exercise by solving the functional equation $C_W(f) = \lambda f$. In particular, one can expand W itself on this basis so that

$$W \stackrel{L_T^2}{=} \sum_{n \geq 1} (W|e_n^W)_2 e_n^W.$$

Now, the orthonormality of the (K-L) basis implies, using Fubini's theorem,

$$\mathbb{E}((W|e_k^W)_2 (W|e_\ell^W)_2) = (e_k^W | C_W(e_\ell^W))_2 = \lambda_\ell \delta_{k\ell},$$

where $\delta_{k\ell}$ denotes the Kronecker symbol. Hence, the Gaussian sequence $((W|e_n^W)_2)_{n \geq 1}$ is pairwise noncorrelated, which implies that these random variables are independent. The above identity also implies that $\text{Var}((W|e_n^W)_2) = \lambda_n$. Finally, this shows that

$$W \stackrel{L_T^2}{=} \sum_{n \geq 1} \sqrt{\lambda_n} \xi_n e_n^W, \quad (6.1)$$

where $\xi_n := (W|e_n^W)_2 / \sqrt{\lambda_n}$, $n \geq 1$, is an i.i.d. sequence of $\mathcal{N}(0; 1)$ -distributed random variables. Furthermore, this K-L expansion converges in a much stronger sense since $\sup_{t \in [0, T]} |W_t - \sum_{k=1}^n \sqrt{\lambda_k} \xi_k e_k^W(t)| \rightarrow 0$ $\mathbb{P}$ -a.s. and

$$\| \sup_{[0, T]} |W_t - \sum_{1 \leq k \leq n} \sqrt{\lambda_k} \xi_k e_k^W(t)| \|_2 = O\left(\sqrt{\log n/n}\right)$$

(see LUSCHGY and PAGÈS [2007]). Similar results (with various rates) hold true for a wide class of Gaussian processes expanded on “admissible” basis (see LUSCHGY and PAGÈS [2007]).

THEOREM 6.1 (LUSCHGY and PAGÈS [2002], LUSCHGY and PAGÈS [2004], and LUSCHGY, PAGÈS and WILBERTZ [2007]). *Let Γ^N , $N \geq 1$, be a sequence of optimal N -quantizers for W .*

(a) *For every $N \geq 1$, $\text{span}(\Gamma^N) = \text{span}\{e_1^W, \dots, e_{d(N)}^W\}$ with $d(N) \gtrsim \frac{1}{2} \log N$.*

(b) $e_N(W, L_T^2) = \|W - \hat{W}^{\Gamma^N}\|_2 \sim \frac{T\sqrt{2}}{\pi} \frac{1}{\sqrt{\log N}}$ as $N \rightarrow \infty$.

REMARKS.

- The fact, confirmed by numerical experiments (see Section 6.3, Fig. 6.4), that $d(N) \sim \log N$ holds as a conjecture.

- Denoting Π_d the orthogonal projection on $\text{span}\{e_1^W, \dots, e_d^W\}$, one derives from (a) that $\hat{W}^{\Gamma^N} = \widehat{\Pi_{d(N)}(W)}^{\Gamma^N}$ (optimal quantization at level N) and

$$\begin{aligned} \|W - \hat{W}^{\Gamma^N}\|_2^2 &= \|\Pi_{d(N)}(W) - \widehat{\Pi_{d(N)}(W)}^{\Gamma^N}\|_2^2 + \|W - \Pi_{d(N)}(W)\|_2^2 \\ &= e_N \left(Z_{d(N)}, \mathbb{R}^{d(N)} \right)^2 + \sum_{n \geq d(N)+1} \lambda_n, \end{aligned}$$

$$\text{where } Z_{d(N)} \stackrel{\mathcal{L}}{\sim} \Pi_{d(N)}(W) \sim \bigotimes_{k=1}^{d(N)} \mathcal{N}(0; \lambda_k).$$

6.2. Centered Gaussian processes

Theorem 6.1 devoted to the standard Brownian motion is a particular case of a more general theorem, which holds for a wide class of Gaussian processes.

THEOREM 6.2 (LUSCHGY and PAGÈS [2002], LUSCHGY and PAGÈS [2004]). *Let $X = (X_t)_{t \in [0, T]}$ be a Gaussian process with K -L eigensystem $(\lambda_n^X, e_n^X)_{n \geq 1}$ (with $\lambda_1 \geq \lambda_2 \geq \dots$ is nonincreasing). Let Γ^N , $N \geq 1$, be a sequence of quadratic optimal N -quantizers for X . Assume*

$$\lambda_n^X \sim \frac{\kappa}{n^b} \quad \text{as } n \rightarrow \infty \quad (b > 1).$$

- (a) $\text{span}(\Gamma^N) = \text{span}\{e_1^X, \dots, e_{d^X(N)}^X\}$ and $d^X(N) \gtrsim \frac{1}{b^{1/(b-1)}} \frac{2}{b} \log N$.
 (b) $e_N(X, L_T^2) = \|X - \hat{X}^{\Gamma^N}\|_2 \sim \sqrt{\kappa} \sqrt{b^b(b-1)^{-1}} (2 \log N)^{-\frac{b-1}{2}}$.

REMARKS.

- The above result admits an extension to the case $\lambda_n^X \sim \varphi(n)$ as $n \rightarrow \infty$ with φ regularly varying, index $-b \leq -1$ (see LUSCHGY and PAGÈS [2004]). In LUSCHGY and PAGÈS [2002], upper or lower bounds are also established when

$$(\lambda_n^X \leq \varphi(n), \quad n \geq 1) \quad \text{or} \quad (\lambda_n^X \geq \varphi(n), \quad n \geq 1).$$

- The sharp asymptotics $d^X(N) \sim \frac{2}{b} \log N$ holds as a conjecture.

Applications to classical (centered) Gaussian processes.

- Brownian bridge: $X_t := W_t - \frac{t}{T} W_T$, $t \in [0, T]$ and $e_n^X(t) = \sqrt{2/T} \sin(\pi n \frac{t}{T})$, $\lambda_n = (\frac{T}{\pi n})^2$ so that $e_N(X, L_T^2) \sim T \frac{\sqrt{2}}{\pi} (\log N)^{-\frac{1}{2}}$.
- Fractional Brownian motion with Hurst constant $H \in (0, 1)$

$$e_N(W^H, L_T^2) \sim T^{H+\frac{1}{2}} c(H) (\log N)^{-H},$$

where $c(H) = \left(\frac{\Gamma(2H) \sin(\pi H)(1+2H)}{\pi} \right)^{\frac{1}{2}} \left(\frac{1+2H}{2\pi} \right)^H$ and $\Gamma(t)$ denotes the Gamma function at $t > 0$.

- Some further explicit sharp rates can be derived from the above theorem for other classes of Gaussian stochastic processes (see LUSCHGY and PAGÈS [2004]) like the fractional Ornstein–Uhlenbeck processes, the Gaussian diffusions, a wide class Gaussian stationary processes (the quantization rate is derived from the high-frequency asymptotics of its spectral density, assumed to be square integrable on the real line) for the m -folded integrated Brownian motion, the fractional Brownian sheet, etc.
- Of course, some upper bounds can be derived for some even wider classes of processes, based on the first remark (see LUSCHGY and PAGÈS [2002]).

Extensions to $r, p \neq 2$ When the processes have some self-similarity properties, it is possible to obtain some sharp rates in the nonpurely quadratic case: this has been done for fractional Brownian motion DEREICH and SCHEUTZOW [2006] using some quite different techniques in which self-similarity properties play crucial role. It leads to the following sharp rates, for $p \in [1, +\infty]$ and $r \in (0, \infty)$

$$e_{N,r}(W^H, L_T^p) \sim T^{H+\frac{1}{2}} c(r, H) (\log N)^{-H}, \quad c(r, H) \in (0, +\infty).$$

6.3. Numerical optimization of quadratic functional quantization

Thanks to the scaling property of Brownian motion, one may focus on the normalized case $T = 1$. The numerical approach to optimal quantization of the Brownian motion is essentially based on Theorem 6.1 and the remark that follows: indeed, these results show that quadratic optimal functional quantization of a centered Gaussian process reduces to a finite-dimensional optimal quantization problem for a Gaussian distribution with a diagonal covariance structure. Namely, the optimization problem at level N reads

$$(\mathcal{O}_N) \equiv \begin{cases} e_N(W, L_T^2)^2 & := e_N(Z_{\mathbf{d}(N)}, \mathbb{R}^{\mathbf{d}(N)})^2 + \sum_{k \geq \mathbf{d}(N)+1} \lambda_k \\ \text{where } Z_{\mathbf{d}(N)} & \overset{\mathcal{L}}{\sim} \bigotimes_{k=1}^{\mathbf{d}(N)} \mathcal{N}(0, \lambda_k). \end{cases}$$

Moreover, if $\beta^N := \{\beta_1^N, \dots, \beta_N^N\}$ denotes an optimal N -quantizer of $Z_{\mathbf{d}(N)}$, then the optimal N -quantizer Γ^N of W reads $\Gamma^N = \{x_1^N, \dots, x_N^N\}$ with

$$x_i^N(t) = \sum_{1 \leq \ell \leq \mathbf{d}(N)} (\beta_i^N)^\ell e_\ell^W(t), \quad i = 1, \dots, N. \quad (6.2)$$

The good news is that $(\mathcal{O}_N)$ is in fact a finite-dimensional quantization optimization problem for each $N \geq 1$. The bad news is that the problem is somewhat ill-conditioned since the decrease of the eigenvalues of W is very steep for small values of n : $\lambda_1 =$

$0.40528 \dots, \lambda_2 = 0.04503 \dots \approx \lambda_1/10$. This is probably one reason for which former attempts to produce good quantizations of the Brownian motion first focused on other kinds of quantizers like scalar product quantizers (see PAGÈS and PRINTEM [2005] and Section 6.4) or d -dimensional block product quantizations (see WILBERTZ [2005], LUSCHGY, PAGÈS and WILBERTZ [2007]).

Optimization of the (quadratic) quantization of $\mathbb{R}^d$ -valued random vectors has been extensively investigated since the early 1950s, first in one-dimension, then in higher dimension when the cost of numerical Monte Carlo simulation was drastically cut down (see GERSHO and GRAY [1992]). Recent application of optimal vector quantization to numerics turned out to be much more demanding in terms of accuracy. In that direction, one may cite PAGÈS and PRINTEM [2003], MRAD and BEN HAMIDA [2006] (mainly focused on numerical optimization of the quadratic quantization of normal distributions). To apply these methods, it is more convenient to rewrite our optimization problem with respect to the standard d -dimensional distribution $\mathcal{N}(0; I_d)$ by simply considering the Euclidean norm derived from the covariance matrix $\text{Diag}(\lambda_1, \dots, \lambda_{d(N)})$, that is,

$$(\mathcal{O}_N) \Leftrightarrow \begin{cases} N\text{-optimal quantization of } \bigotimes_{k=1}^{d(N)} \mathcal{N}(0, 1) \\ \text{for the covariance norm } |(z_1, \dots, z_{d(N)})|^2 = \sum_{k=1}^{d(N)} \lambda_k z_k^2. \end{cases}$$

The main point is, of course, that the dimension $d(N)$ is unknown. However (see Fig. 6.4), one clearly verifies on small values of N that the conjecture $d(N) \sim \log N$ is most likely true. Then, for higher values of N , one relies on it to shift from one dimension to another following the rule $d(N) = d, N \in \{e^d, \dots, e^{d+1} - 1\}$.

6.3.1. A toolbox for quantization optimization: a short overview

Here is a short overview of stochastic optimization methods to compute optimal or at least locally optimal quantizers in finite dimension. For more detail, we refer to PAGÈS and PRINTEM [2003] and the references therein. Let $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0; I_d)$.

Competitive learning vector quantization (CLVQ). This procedure is a recursive stochastic approximation gradient descent based on the integral representation of the gradient $\nabla D_N^Z(x)$, $x \in H^n$ (temporarily coming back to N -tuple notation), of the distortion as the expectation of a local gradient, that is,

$$\forall x^N \in H^N, \quad \nabla D_N^Z(x^N) = \mathbb{E}(\nabla D_N^Z(x^N, \zeta)), \quad \zeta_k \text{ i.i.d.}, \quad \zeta_1 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, I_d)$$

so that, starting from $x^N(0) \in (\mathbb{R}^d)^N$, one sets

$$\forall k \geq 0, \quad x^N(k+1) = x^N(k) - \frac{c}{k+1} \nabla D_N^Z(x^N(k), \zeta_{k+1}),$$

where $c \in (0, 1]$ is a real constant to be tuned. As set, this looks quite formal but the operating CLVQ procedure consists of two, phases at each iteration.

- (i) Competitive Phase: Search of the nearest-neighbor $x^N(k)_{i^*(k+1)}$ of ζ_{k+1} among the components of $x^N(k)_i$, $i = 1, \dots, N$ (using a “winning convention” in case of conflict on the boundary of the Voronoi cells).
- (ii) Cooperative Phase: One moves the winning component toward ζ_{k+1} using a dilatation, that is, $x^N(k+1)_{i^*(k+1)} = \text{dilatation}_{\zeta_{k+1}, 1 - \frac{c}{k+1}}(x^N(k)_{i^*(k+1)})$.

This procedure is useful for small or medium values of N . For an extensive study of this procedure, which turns out to be singular in the world of recursive stochastic approximation algorithms, we refer to PAGÈS [1998]. For general background on stochastic approximation, we refer to BENVENISTE, MÉTIVIER and PRIOURET [1990], KUSHNER and YIN [2003].

The randomized “Lloyd I procedure.” This is the randomization of the stationarity-based fixed-point procedure since any optimal quantizer satisfies (3.3):

$$\hat{Z}^{x^N(k+1)} = \mathbb{E}(Z \mid \hat{Z}^{x^N(k)}), \quad x^N(0) \subset \mathbb{R}^d.$$

At every iteration, the conditional expectation $\mathbb{E}(Z \mid \hat{Z}^{x^N(k)})$ is computed using a Monte Carlo simulation. For more details about practical aspects of Lloyd I procedure, we refer to PAGÈS and PRINTEMS [2003]. In MRAD and BEN HAMIDA [2006], an approach based on genetic evolutionary algorithms is developed.

For both procedures, one may substitute a sequence of quasi-random numbers to the usual pseudorandom sequence. This often speeds up the rate of convergence of the method, although this can only be proved (see BENVENISTE, MÉTIVIER and PRIOURET [1990]) for a very specific class of stochastic algorithms (to which *CLVQ* does not belong).

The most important step to preserve the accuracy of the quantization as N (and $d(N)$) increase is to use the so-called splitting method, which finds its origin in the proof of the existence of an optimal N -quantizer: once the optimization of a quantization grid of size N is achieved, one specifies the starting grid for the size $N+1$ or more generally $N+v$, $v \geq 1$, by merging the optimized grid of size N resulting from the former procedure with v points sampled independently from the normal distribution with probability density proportional to $\varphi^{\frac{d}{d+2}}$, where φ denotes the p.d.f. of $\mathcal{N}(0; I_d)$. This rather unexpected choice is motivated by the fact that this distribution provides the lowest in average random quantization error (see COHORT [1998]).

As a result, to be downloaded on the Web site PAGÈS and PRINTEMS [2005]: www.quantize.maths-fi.com

- Optimized stationary codebooks for W : in practice, the N -quantizers β^N of the distribution $\otimes_{k=1}^{d(N)} \mathcal{N}(0; \lambda_k)$, $N=1$ up to 10,000 ($d(N)$ runs from 1 to 9).
- Companion parameters:
 - distribution of $\hat{W}^{\Gamma^N}$: $\mathbb{P}(\hat{W}^{\Gamma^N} = x_i^N) = \mathbb{P}(\hat{Z}_{d(N)}^{\beta^N} = \beta_i^N)$ ($\leftarrow$ in $\mathbb{R}^{d(N)}$).
 - The quadratic quantization error: $\|W - \hat{W}^{\Gamma^N}\|_2$.

See Figs. 6.1, 6.2 and 6.3 for some examples of optimal quantizers Γ^N (and their counter parts β^N in $\mathbb{R}^{d(N)}$).

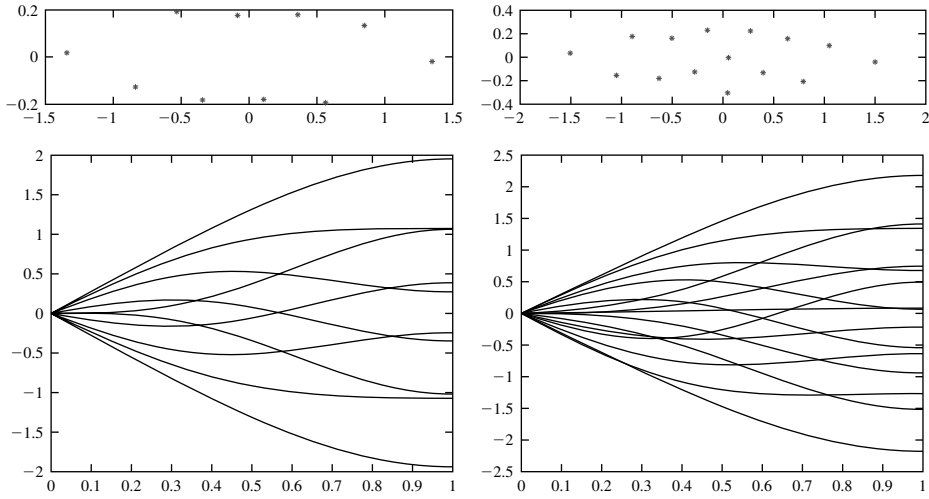


FIG. 6.1 Optimized functional quantization of the Brownian motion W for $N = 10, 15$ ($d(N) = 2$). Top: β^N depicted in $\mathbb{R}^2$. Bottom: the optimized N -quantizer Γ^N .

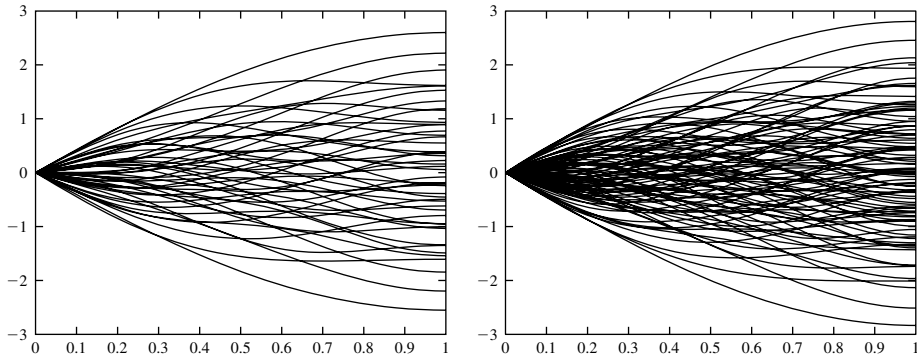


FIG. 6.2 Optimized functional quantization of the Brownian motion W . The N -quantizers Γ^N . Left: $N = 48$ ($d(N) = 3$). Right: $N = 96$, $d(96) = 4$.

REMARKS.

- Both stochastic optimization procedures that we described above can, of course, be implemented to produce optimal (or optimized) grids of any multidimensional probability distribution on $\mathbb{R}^d$, having in mind that as the dimension d increases the second one becomes the most efficient.
- These procedures are based on a nearest-neighbor search among N points. A naive implementation of such a procedure has a linear complexity in N and becomes very demanding in higher dimension. So, to drastically reduce this optimization phase as well as that devoted to the weight estimation of the resulting optimal quantizer, one can call upon some fast nearest-neighbor procedure like that originally developed

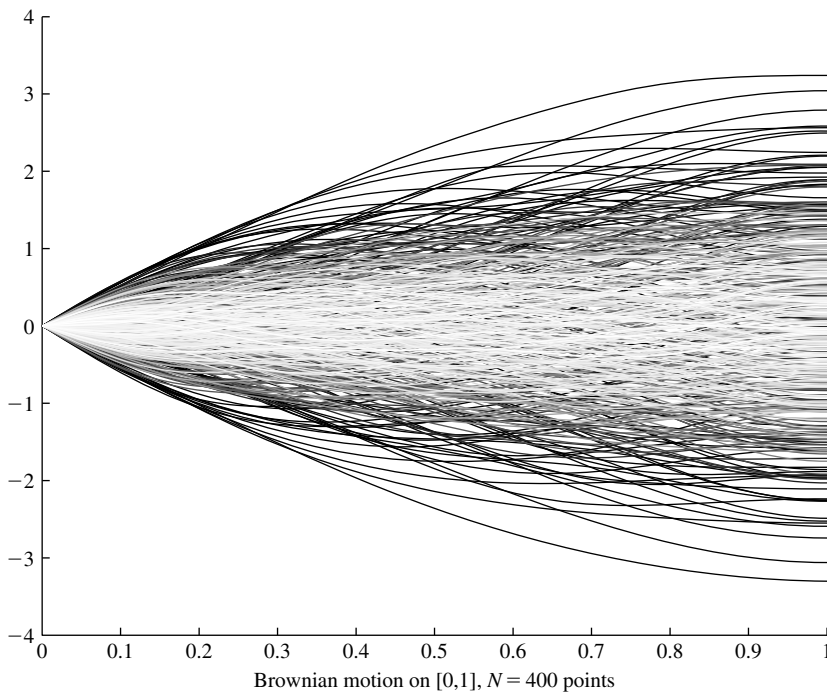


FIG. 6.3 Optimized N -quantizer Γ^N of the Brownian motion W with $N = 400$. The grey level of the paths codes their weights.

by Bentley and analyzed in a seminal paper FRIEDMAN, BENTLEY and FINKEL [1977], which is based on the notion of k - d -tree introduced for that purpose by the authors. It reduces the complexity of the search down to $O(\log N)$. The (relative) efficiency of the method increases as the dimension of the state space increases.

6.4. An alternative: product functional quantization

Scalar product functional quantization is a quantization method that produces rate optimal suboptimal quantizers. They were used by LUSCHGY and PAGÈS [2002] to provide exact rate (although not sharp) for a very large class of processes. The first attempt to use functional quantization for numerical computation with the Brownian motion was achieved with these quantizers (see PAGÈS and PRINTEMS [2005]). We will see further on their assets. What follows is presented for the Brownian motion but would work for a large class of centered Gaussian processes.

Let us consider again the expansion of W in its K - L basis:

$$W \stackrel{L_T^2}{=} \sum_{n \geq 1} \sqrt{\lambda_n} \xi_n e_n^W,$$

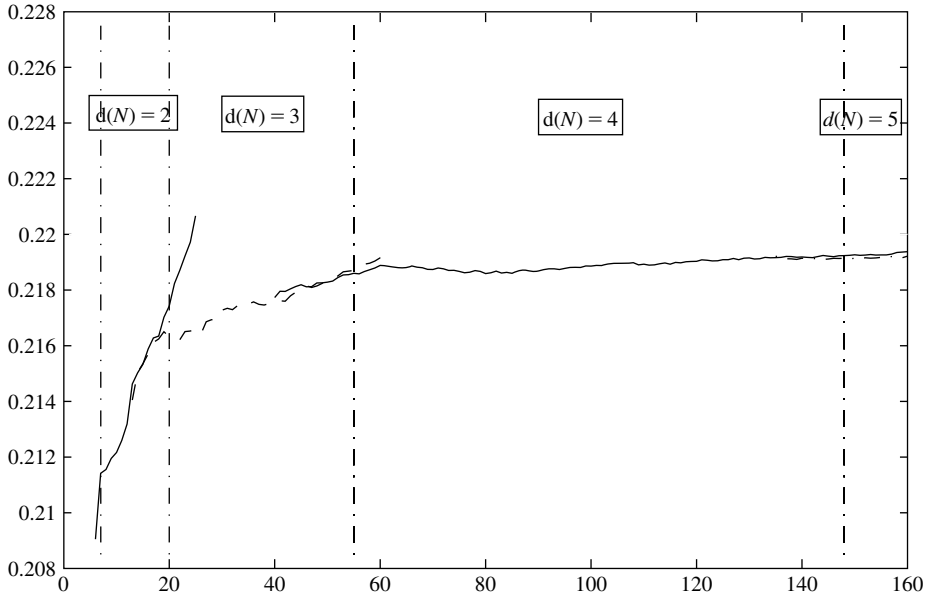


FIG. 6.4 Optimal functional quantization of the Brownian motion. $N \mapsto \log N (e_N(W, L_T^2))^2$, $N \in \{6, \dots, 160\}$. Vertical dashed lines: critical dimensions for $d(N)$, $e^2 \approx 7$, $e^3 \approx 20$, $e^4 \approx 55$, $e^5 \approx 148$.

where $(\xi_n)_{n \geq 1}$ is an i.i.d. sequence $\mathcal{N}(0; 1)$ -distributed random variables (keep in mind this convergence also holds a.s. uniformly in $t \in [0, T]$). The idea is simply to quantize these (normalized) random coordinates ξ_n : for every $n \geq 1$, one considers an optimal N_n -quantization of ξ_n , denoted by $\hat{\xi}_n^{(N_n)}$ ($N_n \geq 1$). For $n > m$, set $N_n = 1$ and $\hat{\xi}_n^{(N_n)} = 0$ (which is the optimal one-quantization). The integer m is called the *length* of the product quantization. Then, one sets

$$\hat{W}_t^{(N_1, \dots, N_m, \text{prod})} := \sum_{n=1}^m \sqrt{\lambda_n} \hat{\xi}_n^{(N_n)} e_n^W(t) = \sum_{n=1}^m \sqrt{\lambda_n} \hat{\xi}_n^{(N_n)} e_n^W(t).$$

Such a quantizer takes $\prod_{n=1}^m N_n \leq N$ values.

If one denotes by $\alpha^M = \{\alpha_1^M, \dots, \alpha_M^M\}$ the (unique) optimal quadratic M -quantizer of the $\mathcal{N}(0; 1)$ -distribution, the underlying quantizer of the above quantization $\hat{W}^{(N_1, \dots, N_m, \text{prod})}$ can be expressed as follows (if one introduces the appropriate multiindexation): for every multiindex $\underline{i} := (i_1, \dots, i_m) \in \prod_{n=1}^m \{1, \dots, N_n\}$, set

$$x_{\underline{i}}^{(N)}(t) := \sum_{n=1}^m \sqrt{\lambda_n} \alpha_{i_n}^{(N_n)} e_n^W(t) \text{ and } \Gamma^{N_1, \dots, N_m, \text{prod}} := \left\{ x_{\underline{i}}^{(N)}, \underline{i} \in \prod_{n=1}^m \{1, \dots, N_n\} \right\}.$$

Then, the product quantization $\hat{W}^{(N_1, \dots, N_m, \text{prod})}$ can be rewritten as

$$\hat{W}_t^{(N_1, \dots, N_m, \text{prod})} = \sum_{\underline{i}} \mathbf{1}_{\{W \in C_{\underline{i}}(\Gamma^{N_1, \dots, N_m, \text{prod}})\}} x_{\underline{i}}^{(N)}(t),$$

where the Voronoi cell of $x_{\underline{i}}^{(N)}$ is given by

$$C_{\underline{i}}(\Gamma^{N_1, \dots, N_m, \text{prod}}) = \prod_{n=1}^m (\alpha_{i_n - \frac{1}{2}}^{(N_n)}, \alpha_{i_n + \frac{1}{2}}^{(N_n)}), \quad \alpha_{i \pm \frac{1}{2}}^{(M)} := \frac{\alpha_i^{(M)} + \alpha_{i \pm 1}^{(M)}}{2},$$

$$\alpha_0 = -\infty, \alpha_{M+1} = +\infty.$$

6.4.1. Quantization rate by product quantizers

It is clear that the optimal product quantizer is the solution to the optimal integral bit allocation

$$\min \left\{ \|W - \hat{W}^{(N_1, \dots, N_m, \text{prod})}\|_2, N_1, \dots, N_m \geq 1, N_1 \times \dots \times N_m \leq N, m \geq 1 \right\}. \quad (6.3)$$

Expanding $\|W - \hat{W}^{(N_1, \dots, N_m, \text{prod})}\|_2^2 = \|W - \hat{W}^{(N_1, \dots, N_m, \text{prod})}\|_{L_T^2}^2$ yields

$$\|W - \hat{W}^{(N_1, \dots, N_m, \text{prod})}\|_2^2 = \sum_{n \geq 1} \lambda_n \|\hat{\xi}_n^{(N_n)} - \xi_n\|_2^2 \quad (6.4)$$

$$= \sum_{n=1}^m \lambda_n (e_{N_n}^2(\mathcal{N}(0; 1), \mathbb{R}) - 1) + \frac{T^2}{2} \quad (6.5)$$

since

$$\sum_{n \geq 1} \lambda_n = \mathbb{E} \sum_{n \geq 1} (W | e_n^W)_2^2 = \mathbb{E} \int_0^T W_t^2 dt = \int_0^T t dt = \frac{T^2}{2}.$$

THEOREM 6.3 (see LUSCHGY and PAGÈS [2002]). *For every $N \geq 1$, there exists an optimal scalar product quantizer of size at most N (or at level N), denoted by $\hat{W}^{(N, \text{prod})}$, of the Brownian motion defined as the solution to the minimization problem (6.3). Furthermore, these optimal product quantizers make up a rate optimal sequence: there exists a real constant $c_W > 0$ such that*

$$\|W - \hat{W}^{(N, \text{prod})}\|_2 \leq \frac{c_W T}{(\log N)^{\frac{1}{2}}}.$$

Proof (sketch of). By scaling, one may assume without loss of generality that $T = 1$. Combining (6.4) and Zador's theorem shows

$$\begin{aligned} \|W - \hat{W}^{(N_1, \dots, N_m, \text{prod})}\|_2^2 &\leq C \left(\sum_{n=1}^m \frac{1}{n^2 N_n^2} \right) + \sum_{n \geq m+1} \lambda_n \\ &\leq C' \left(\sum_{n=1}^m \frac{1}{n^2 N_n^2} + \frac{1}{m} \right), \end{aligned}$$

with $\prod_n N_n \leq N$. Setting $m := m(N) = \lfloor \log N \rfloor$ and $N_k = \lfloor \frac{(m!N)^{\frac{1}{m}}}{k} \rfloor \geq 1$, $k = 1, \dots, m$, yields the announced upper bound. $\diamond$

REMARKS.

- One can show that the length $m(N)$ of the optimal quadratic product quantizer satisfies

$$m(N) \sim \log N \quad \text{as} \quad N \rightarrow +\infty.$$

- The most striking fact is that very few ingredients are necessary to make the proof work as far as the quantization rate is concerned. We only need the basis of L_T^2 on which W is expanded to be orthonormal or the random coordinates to be orthogonal in $L^2(\mathbb{P})$. This robustness of the proof has been used to obtain some upper bounds for very wide classes of Gaussian processes by considering alternative orthonormal basis of L_T^2 like the Haar basis for processes having self-similarity properties (see LUSCHGY and PAGÈS [2002]), or trigonometric basis for stationary processes (see LUSCHGY and PAGÈS [2002]). More recently, combined with the nonasymptotic Zador's theorem, it was used to provide some connections between mean regularity of stochastic processes and quantization rate (see Section 9 and LUSCHGY and PAGÈS [2007]).
- Block quantizers combined with large deviation estimates were used to provide the sharp rate obtained in Theorem 6.1 LUSCHGY and PAGÈS [2004].
- d -dimensional block quantization is also possible, possibly with varying block size, providing a constructive approach to sharp rate see WILBERTZ [2005] and LUSCHGY, PAGÈS and WILBERTZ [2007].
- A similar approach can also provide some $L^r(\mathbb{P})$ -rates for product quantization with respect to the sup-norm over $[0, T]$, see LUSCHGY and PAGÈS [2007].

6.4.2. How to use product quantizers for numerical computations?

For numerics, one can assume by a scaling argument that $T = 1$. To use product quantizers for numerics, we need to have access to the quantizers (or grid) at a given level N , their weights (and the quantization error). All these quantities are available with product quantizers. In fact, the first attempts to use functional quantization for numerics (path-dependent option pricing) were carried out with product quantizers (see PAGÈS and PRINTEMPS [2005]).

TABLE 6.1
Optimal product quantization of the Brownian motion: optimal allocations for $N = 10^k$, $k = 0, \dots, 5$.

N	N_{rec}	Quantization error	Optimal allocation
1	1	0.7071	1
10	10	0.3138	5-2
100	96	0.2264	12-4-2
1 000	966	0.1881	23-7-3-2
10 000	9 984	0.1626	26-8-4-3-2-2
100 000	97 920	0.1461	34 - 10 - 6 - 4 - 3 - 2 - 2

- The optimal product quantizers (denoted by $\Gamma^{(N, \text{prod})}$) at level N are explicit, given the optimal quantizers of the scalar normal distribution $\mathcal{N}(0; 1)$. In fact, the optimal allocation of the size N_i of each marginal has been already achieved up to very high values of N . Some typical optimal allocation (and the resulting quadratic quantization error) is reported in Table 6.1 below. The integer N_{rec} denotes the effective size of the optimal product quantizer.
- The weights $\mathbb{P}(\hat{W}^{(N, \text{prod})} = x_i)$ are explicit too: the normalized coordinates ξ_n of W in its K - L basis are independent, consequently,

$$\begin{aligned} \mathbb{P}(\hat{W}^{(N, \text{prod})} = x_i) &= \mathbb{P}(\hat{\xi}_n^{(N_n)} = \alpha_{i_n}^{(N_n)}, n = 1, \dots, m(N)) \\ &= \prod_{n=1}^{m(N)} \underbrace{\mathbb{P}(\hat{\xi}_n^{(N_n)} = \alpha_{i_n}^{(N_n)})}_{1D \text{ (tabulated) weights}}. \end{aligned}$$

- Eq. (6.5) shows that the (squared) quantization error of a product quantizer can be straightforwardly computed as soon as one knows the eigenvalues and the (squared) quantization error of the normal distributions for the N_i s.

The optimal allocations up to $N = 12\,000$ can be downloaded on the Web site (PAGÈS and PRINTEMS [2005]) as well as the necessary one-dimensional optimal quantizers (including the weights and the quantization error) of the scalar normal distribution (up to a size of 500, which is enough for this purpose). Some examples of optimal product quantizers are displayed in Figs. 6.5 and 6.6.

For numerical purpose, we are also interested in the stationarity property since such quantizers produce lower (weak) errors in cubature formulas.

PROBLEM 6.1 (see PAGÈS and PRINTEMS [2005]). The product quantizers obtained from the K - L basis are stationary quantizers (although suboptimal).

PROOF. First, note that

$$\hat{W}^{N, \text{prod}} = \sum_{n \geq 1} \sqrt{\lambda_n} \hat{\xi}_n^{(N_n)} \mathbf{e}_n(t)$$

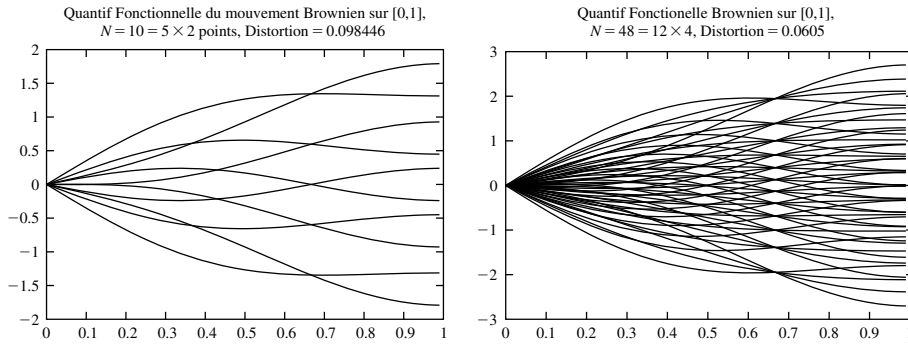


FIG. 6.5 Product quantization of the Brownian motion: the N_{rec} -quantizer $\Gamma^{(N, \text{prod})}$. $N = 10$: $N_{\text{rec}} = 10$ and $N = 50$: $N_{\text{rec}} = 12 \times 4 = 48$.

so that $\sigma(\hat{W}^{N, \text{prod}}) = \sigma(\hat{\xi}_k^{(N_k)}, k \geq 1)$. Consequently,

$$\begin{aligned} \mathbb{E}(W | \hat{W}^{N, \text{prod}}) &= \mathbb{E}(W | \sigma(\hat{\xi}_k^{(N_k)}, k \geq 1)) \\ \mathbb{E}(W | \hat{W}^{N, \text{prod}}) &= \sum_{n \geq 1} \sqrt{\lambda_n} \mathbb{E}(\xi_n | \sigma(\hat{\xi}_k^{(N_k)}, k \geq 1)) e_n^W \\ &\stackrel{i.i.d.}{=} \sum_{n \geq 1} \sqrt{\lambda_n} \mathbb{E}(\xi_n | \hat{\xi}_n^{(N_n)}) e_n^W \\ &= \sum_{n \geq 1} \sqrt{\lambda_n} \hat{\xi}_n^{(N_n)} e_n^W = \hat{W}. \end{aligned}$$

□

REMARKS.

- This result is no longer true for product quantizers based on other orthonormal basis.
- This shows the existence of nonoptimal stationary quantizers.

6.5. Optimal versus product quadratic functional quantization ($T = 1$)

(NUMERICAL) OPTIMIZED QUANTIZATION: By scaling, we can assume without loss of generality that $T = 1$. We carried out a huge optimization task in order to produce some optimized quantization grids for the Brownian motion by solving numerically $(\mathcal{O}_N)$ for $N = 1$ up to $N = 10\,000$.

$$e_N(W, L_T^2)^2 \approx \frac{0.2195}{\log N}, \quad N = 1, \dots, 10\,000.$$

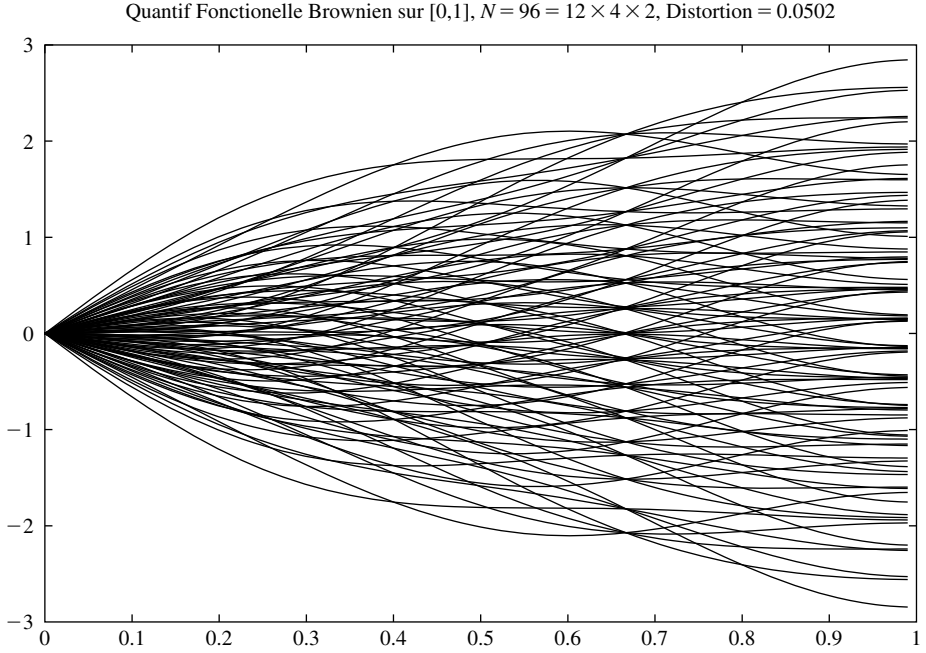


FIG. 6.6 Product quantization of the Brownian motion: the N_{rec} -quantizer $\Gamma^{(N, \text{prod})}$. $N = 100$: $N_{\text{rec}} = 12 \times 4 \times 2 = 96$.

This value (see Fig. 6.7(left)) is significantly greater than the theoretical (asymptotic) bound given by Theorem 6.1, which is

$$\lim_N \log Ne_N(W, L_T^2)^2 = \frac{2}{\pi^2} = 0.2026 \dots$$

Our guess, supported by our numerical experiments, is that in fact $N \mapsto \log Ne_N(W, L_T^2)^2$ is possibly not monotone but unimodal.

OPTIMAL PRODUCT QUANTIZATION: as shown in Fig. 6.7 (bottom), one has approximately

$$\begin{aligned} \min \left\{ \|W - \hat{W}\|_{L_T^2}^2, 1 \leq N_1 \dots N_m \leq N, m \geq 1 \right\} \\ = \|W - \hat{W}^{(N, \text{prod})}\|_2^2 \approx \frac{0.245}{\log N}. \end{aligned}$$

OPTIMAL d -DIMENSIONAL BLOCK PRODUCT QUANTIZATION: let us briefly mention this approach developed by WILBERTZ [2005] in which product quantization is achieved by quantizing some marginal blocks of size 1, 2, or 3. By this approach, the corresponding constant is approximately 0.23, that is, roughly in between scalar product quantization and optimized numeric quantization.

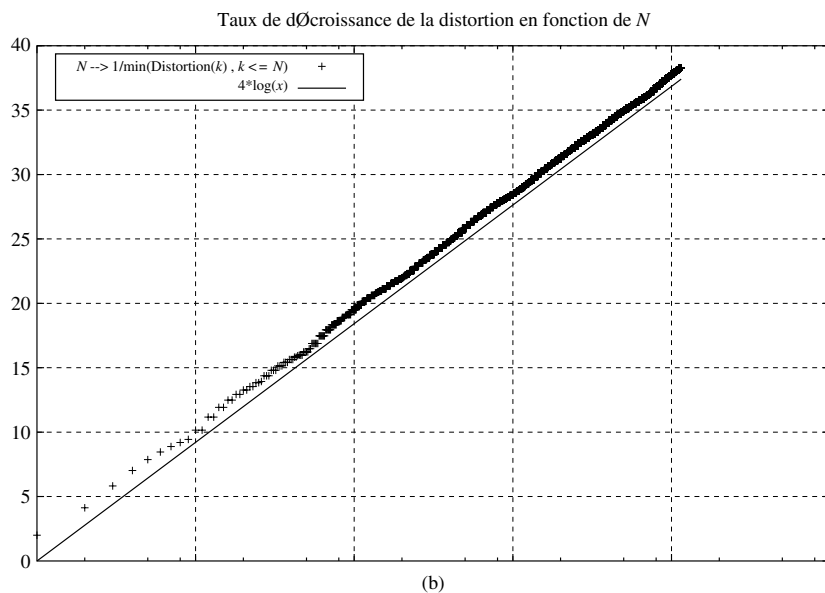
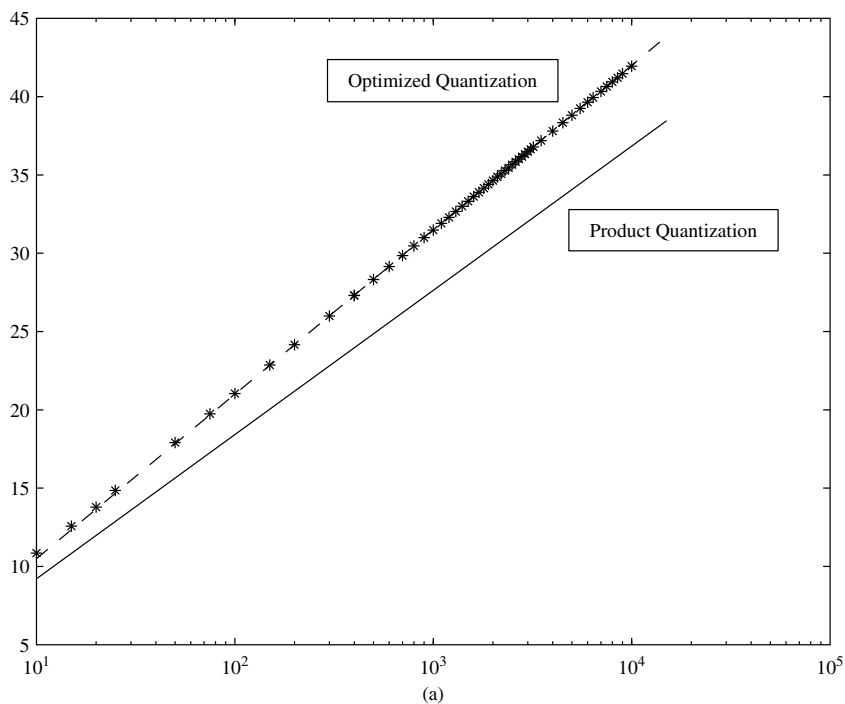


FIG. 6.7 Numerical quantization rates. Top (optimal quantization). Line+++ : $\log N \mapsto (\|W - \hat{W}^N\|_2)^{-2}$. Dashed line: $\log N \mapsto \log N/0.2194$. Solid line: $\log N \mapsto \log N/0.25$. Bottom (product quantization). Line+++ : $\log N \mapsto (\min_{1 \leq k \leq N} \|W - \hat{W}^{k, \text{prod}}\|_2^2)^{-1}$. Solid line: $\log N \mapsto \log N/0.25$.

The conclusion, confirmed by our numerical experiments on option pricing (see Section 8), is that

- Optimal quantization is significantly more accurate on numerical experiments but is much more demanding since it needs to keep off-line or at least to handle large files (say 1 GB for $N = 10\,000$).
- Both approaches are included in the option pricer PREMIA (MATHFI Project, Inria). An online benchmark is available on the Web site (PAGÈS and PRINTEMS [2005]).

7. Constructive functional quantization of diffusions

7.1. Rate optimality for scalar Brownian diffusions

One considers on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ an homogenous Brownian diffusion process:

$$dX_t = b(X_t)dt + \vartheta(X_t) dW_t, \quad X_0 = x_0 \in \mathbb{R},$$

where b and ϑ are continuous on $\mathbb{R}$ with at most linear growth (*i.e.*, $|b(x)| + |\sigma(x)| \leq C(1 + |x|)$) so that at least a weak solution to the equation exists.

To devise a constructive way to quantize the diffusion X , it seems natural to start from a rate optimal quantization of the Brownian motion and to obtain some “good” (but how good?) quantizers for the diffusion by solving an appropriate Ordinary Differential Equation (ODE). So let $\Gamma^N = (w_1^N, \dots, w_N^N)$, $N \geq 1$, be a sequence of stationary rate optimal N -quantizers of W . One considers the following (noncoupled) integral equations:

$$\begin{aligned} dx_i^{(N)}(t) = & \left(b(x_i^{(N)}(t)) - \frac{1}{2} \vartheta \theta'(x_i^{(N)}(t)) \right) dt \\ & + \vartheta(t, x_i^{(N)}(t)) dw_i^N(t), \quad i = 1, \dots, N. \end{aligned} \quad (7.1)$$

Set

$$\tilde{X}_t^{x^{(N)}} = \sum_{i=1}^N x_i^{(N)}(t) \mathbf{1}_{\{\hat{W}^{\Gamma^N} = w_i^N\}}.$$

The process $\tilde{X}^{x^{(N)}}$ is a non-Voronoi quantizer (since it is defined using the Voronoi diagram of W). What is interesting is that it is a computable quantizer (once the above integral equations have been solved) since the weights $\mathbb{P}(\hat{W}^{\Gamma^N} = w_i^N)$ are known. The Voronoi quantization defined by $x^{(N)}$ induces a lower quantization error, but we have no access to its weights for numerics. The good news is that $\tilde{X}^{x^{(N)}}$ is already rate optimal.

THEOREM 7.1 (LUSCHGY and PAGÈS [2006]). *Assume that b is differentiable, ϑ is positive twice differentiable and that $b' - b \frac{\vartheta'}{\vartheta} - \frac{1}{2} \vartheta \vartheta''$ is bounded. Then,*

$$e_N(X, L_T^2) \leq \|X - \tilde{X}^{x^{(N)}}\|_2 = O((\log N)^{-\frac{1}{2}}).$$

If, furthermore, $\vartheta \geq \varepsilon_0 > 0$, then $e_N(X, L_T^2) \approx (\log N)^{-\frac{1}{2}}$.

REMARKS.

- For some results in the nonhomogenous case, we refer to LUSCHGY and PAGÈS [2006]. Furthermore, the above estimates still hold true for the $(L^r(\mathbb{P}), L_T^p)$ -quantization, $1 < r, p < +\infty$ provided $\|W - \hat{W}^{\Gamma^N}\|_{L_T^p} = O((\log N)^{-\frac{1}{2}})$.
- This result is closely connected to the Doss–Sussman approach (see Doss [1977]), and in fact, the results can be extended to some classes multidimensional diffusions (whose diffusion coefficient is the inverse of the gradient of a diffeomorphism), which include several standard multidimensional financial models (including the Black–Scholes model).
- A sharp quantization rate $e_{N,r}(X, L_T^p) \sim c(\log N)^{-\frac{1}{2}}$ for scalar elliptic diffusions is established by DEREICH [2005a,b] using a nonconstructive approach, $1 \leq p \leq \infty$.

EXAMPLE 7.1. Rate optimal product quantization of the Ornstein–Uhlenbeck process.

$$dX_t = -kX_t dt + \vartheta dW_t, \quad X_0 = x_0.$$

One solves the noncoupled integral (linear) system

$$x_i(t) = x_0 - k \int_0^t x_i(s) ds + \vartheta w_i^N(t),$$

where $\Gamma^N := \{w_1^N, \dots, w_N^N\}$, $N \geq 1$ is a rate optimal sequence of quantizers

$$w_i^N(t) = \sqrt{\frac{2}{T}} \sum_{\ell \geq 1} \varpi_{i,\ell} \frac{T}{\pi(\ell - 1/2)} \sin\left(\pi(\ell - 1/2) \frac{t}{T}\right), \quad i \in I_N.$$

If Γ^N is optimal for W , then $\varpi_{i,\ell} := (\beta_i^N)^\ell$, $i = 1, \dots, N$, $1 \leq \ell \leq d(N)$ with the notations introduced in (6.2). If Γ^N is an optimal product quantizer (and $N_1, \dots, N_\ell, \dots$ denote the optimal size allocation), then $\varpi_{i,\ell} = \alpha_{i_\ell}^{(N_\ell)}$, where $i := (i_1, \dots, i_\ell, \dots) \in \prod_{\ell \geq 1} \{1, \dots, N_\ell\}$. Elementary computations show that

$$x_i^N(t) = e^{-kt} x_0 + \vartheta \sum_{\ell \geq 1} \chi_{i_\ell}^{(N_\ell)} \tilde{c}_\ell \varphi_\ell(t),$$

$$\text{with} \quad \tilde{c}_\ell = \frac{T^2}{(\pi(\ell - 1/2))^2 + (kT)^2}$$

$$\text{and} \quad \varphi_\ell(t) := \sqrt{\frac{2}{T}} \left(\frac{\pi}{T} (\ell - 1/2) \sin\left(\pi(\ell - 1/2) \frac{t}{T}\right) + k \left(\cos\left(\pi(\ell - 1/2) \frac{t}{T}\right) - e^{-kt} \right) \right).$$

7.2. Multidimensional diffusions for Stratanovich Stochastic Differential Equations (SDE)

The correcting term $-\frac{1}{2}\vartheta\vartheta'$ coming up in the integral equations suggests to consider directly some diffusion in the Stratanovich sense

$$dX_t = b(t, X_t) dt + \vartheta(t, X_t) \circ dW_t \quad X_0 = x_0 \in \mathbb{R}^d, \quad t \in [0, T].$$

(see REVUZ and YOR [1999] for an introduction), where $W = (W^1, \dots, W^d)$, is a d -dimensional standard Brownian motion.

In that framework, we need to introduce the notion of p -variation: a continuous function $x : [0, T] \rightarrow \mathbb{R}^d$ has finite p -variations if

$$\text{Var}_{p,[0,T]}(x) := \sup \left\{ \left(\sum_{i=0}^{k-1} |x(t_i) - x(t_{i+1})|^p \right)^{\frac{1}{p}}, \right. \\ \left. 0 \leq t_0 \leq t_1 \leq \dots \leq t_k \leq T, k \geq 1 \right\} < +\infty.$$

Then, $d_p(x, x') = |x(0) - x'(0)| + \text{Var}_{p,[0,T]}(x - x')$ defines a distance on the set of functions with finite p -variations. It is classical background that $\text{Var}_{p,[0,T]}(W(\omega)) < +\infty$ $\mathbb{P}(d\omega)$ -a.s. for every $p > 2$.

One way to quantize W at level (at most) N is to quantize each component W^i at level $\lfloor \sqrt[d]{N} \rfloor$. One shows (see LUSCHGY and PAGÈS [2004]) that $\|W - (\hat{W}^{1, \lfloor \sqrt[d]{N} \rfloor}, \dots, \hat{W}^{d, \lfloor \sqrt[d]{N} \rfloor})\|_2 = O((\log N)^{-\frac{1}{2}})$.

Let $\mathcal{C}_b^r([0, T] \times \mathbb{R}^d)$, $r > 0$, denote the set of $\lfloor r \rfloor$ -times differentiable bounded functions $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ with bounded partial derivatives up to order $\lfloor r \rfloor$ and whose partial derivatives of order $\lfloor r \rfloor$ are $(r - \lfloor r \rfloor)$ -Hölder.

THEOREM 7.2 (see PAGÈS and SELLAMI [2007]). *Let $b, \vartheta \in \mathcal{C}_b^{2+\alpha}([0, T] \times \mathbb{R}^d)$ ($\alpha > 0$) and let $\Gamma^N = \{w_1^N, \dots, w_N^N\}$, $N \geq 1$, be a sequence of N -quantizers of the standard d -dimensional Brownian motion W such that $\|W - \hat{W}^{\Gamma^N}\|_2 \rightarrow 0$ as $N \rightarrow \infty$. Let*

$$\tilde{X}_t^{x^{(N)}} := \sum_{i=1}^N x_i^{(N)}(t) \mathbf{1}_{\{\hat{W}=w_i^N\}},$$

where for every $i \in \{1, \dots, N\}$, $x_i^{(N)}$ is solution to

$$\text{ODE}_i \equiv dx_i^{(N)}(t) = b(t, x_i^{(N)}(t))dt + \vartheta(t, x_i^{(N)}(t))dw_i^N(t), \quad x_i^{(N)}(0) = x.$$

Then, for every $p \in (2, \infty)$,

$$\text{Var}_{p,[0,T]}(\tilde{X}^{x^{(N)}} - X) \xrightarrow{\mathbb{P}} 0 \quad \text{as } N \rightarrow \infty.$$

REMARKS.

- The keys of this results are the Kolmogorov criterion, stationarity (in a slightly extended sense), and the connection with rough paths theory (see LEJAY [2003] for an introduction to rough paths theory, convergence in p -variation, etc).
- In is general setting, we have no convergence rate, although we conjecture that $\tilde{X}^{x^{(N)}}$ remains rate optimal if $\hat{W}^{\Gamma^N}$ is.
- There are also some results about the convergence of stochastic integrals of the form $\int_0^t g(\hat{W}_s^N) d\hat{B}_s^N \rightarrow \int_0^t g(W_s) \circ dB_s$ with some rates of convergence when $W = B$ or W and B are independent (depending on the regularity of the function g , see PAGÈS and SELLAMI [2007]).

7.3. About the quantization of multidimensional Brownian motion

We assume $\mathbb{R}^d$ is equipped with the canonical Euclidean norm $|(x^1, \dots, x^d)|^2 = (x^1)^2 + \dots + (x^d)^2$. Let $W = (W^1, \dots, W^d)$ be a d -dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

The most elementary way to quantize W is to quantize each marginal component W^i at a N^i -level so that $N^1 \dots N^d \leq N$. This appears as a spatial product quantization. This is a simple, somewhat flexible approach for applications since N^i can be chosen different from $\lfloor N^{\frac{1}{d}} \rfloor$. However, it is clearly not optimal. Furthermore, it suffers like any product like quantization from the instability of the product $\prod_{1 \leq i \leq d} N^i$. One easily shows that if $N^i = \lfloor N^{\frac{1}{d}} \rfloor$, $i = 1, \dots, d$, then $\|W - (\hat{W}^{(N^i)})_{1 \leq i \leq d}\|_2 = O((\log N)^{-\frac{1}{2}})$ (see LUSCHGY and PAGÈS [2006]).

An alternative way is to quantize the K - L expansion of W given by

$$W^i = \sum_{n \geq 1} \sqrt{\lambda_n^W} \xi_n^i e_n^W, \quad i = 1, \dots, d, \quad \xi_n = (\xi_n^1, \dots, \xi_n^d) \sim \mathcal{N}(0; I_d),$$

by setting

$$(\hat{W})^i = \sum_{n \geq 1} \sqrt{\lambda_n^W} \pi^i(\hat{\xi}_n^{(N_n)}) e_n^W,$$

where $\pi^i(x^1, \dots, x^d) = x^i$ and $\hat{\xi}_n^{(N_n)}$ is an optimal N_n -quantization of ξ_n .

8. Applications to path-dependent option pricing

The typical functionals F defined on $(L_T^2, |\cdot|_{L_T^2})$ for which $\mathbb{E}(F(W))$ can be approximated by the cubature formulae (4.2) and (4.5) are of the form $F(\omega) := \varphi\left(\int_0^T f(t, \omega(t)) dt\right) \mathbf{1}_{\{\omega \in \mathcal{C}([0, T], \mathbb{R})\}}$, where $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is locally Lipschitz continuous in the second variable, namely,

$$\forall t \in [0, T], \forall u, v \in \mathbb{R}, |f(t, u) - f(t, v)| \leq C_f |u - v| (1 + g(|u|) + g(|v|))$$

(with $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ increasing, convex and $g(\sup_{t \in [0, T]} |W_t|) \in L^2(\mathbb{P})$) and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous. One could consider for ω some càdlàg functions as well. A classical example is the Asian payoff in a Black–Scholes model

$$F(\omega) = \exp(-rT) \left(\frac{1}{T} \int_0^T s_0 \exp(\sigma \omega(t) + (r - \sigma^2/2)t) dt - K \right)_+.$$

8.1. Numerical integration (III): Richardson–Romberg log-extrapolation

Let $F : L_T^2 \rightarrow \mathbb{R}$ be a three times $|\cdot|_{L_T^2}$ -differentiable functional with bounded differentials. Assume $\hat{W}^{(N)}$, $N \geq 1$, is a sequence of a rate-optimal stationary quantizations of the standard Brownian motion W . Assume, furthermore, that

$$\mathbb{E}(D^2 F(\hat{W}^{(N)}).(W - \hat{W}^{(N)})^{\otimes 2}) \sim \frac{c}{\log N} \quad \text{as } N \rightarrow \infty \quad (8.1)$$

and

$$\mathbb{E}|W - \hat{W}^{(N)}|_{L_T^2}^3 = O\left((\log N)^{-\frac{3}{2}}\right). \quad (8.2)$$

Then, a higher order Taylor expansion yields

$$\begin{aligned} F(W) &= F(\hat{W}^{(N)}) + DF(\hat{W}^{(N)}).(W - \hat{W}^{(N)}) + \frac{1}{2} D^2 F(\hat{W}^{(N)}).(W - \hat{W}^{(N)})^{\otimes 2} \\ &\quad + \frac{1}{6} D^3 F(\zeta).(W - \hat{W}^{(N)})^{\otimes 3}, \quad \zeta \in (\hat{W}^{(N)}, W), \\ \mathbb{E} F(W) &= \mathbb{E} F(\hat{W}^{(N)}) + \frac{c}{2 \log N} + o\left((\log N)^{-\frac{3}{2} + \varepsilon}\right). \end{aligned}$$

Then, one can design a log R - R extrapolation by considering N , N' , $N < N'$ (e.g., $N' \approx 4N$), so that

$$\mathbb{E}(F(W)) = \frac{\log N' \times \mathbb{E}(F(\hat{W}^{(N')})) - \log N \times \mathbb{E}(F(\hat{W}^{(N)}))}{\log N' - \log N} + o\left((\log N)^{-\frac{3}{2} + \varepsilon}\right).$$

For practical implementation, it is suggested by WILBERTZ [2005] to replace $\log N$ by the more consistent “estimator” $\|W - \hat{W}^{(N)}\|_2^{-2}$.

In fact, Assumption (8.1) holds true for optimal product quantization when F is polynomial function F , $d^0 F = 2$. Assumption (8.2) holds true in that case (see GRAF, LUSCHGY and PAGÈS [2006]). As concerns optimal quantization, these statements, are still conjectures. However, given that $\hat{W}$ and $W - \hat{W}$ are independent (see LUSCHGY and PAGÈS [2002]), (8.1) is equivalent to the simple case where $D^2 F(\hat{W}^{(N)})$ is constant.

Note that the above extrapolation or some variants can be implemented with other stochastic processes in accordance with the rate of convergence of the quantization error.

8.2. Asian option pricing in a Heston stochastic volatility model

In this section, we will price an Asian call option in a Heston stochastic volatility model using some optimal (or at least optimized) functional quantizations of the two Brownian motions that drive the diffusion. This model has already been considered by PAGÈS and PRINTEMPS [2005], who investigated numerical aspects of functional quantization for the first time. The implementation was made with some optimal product quantizations of the Brownian motions. The Heston stochastic volatility model was introduced by HESTON [1993] to model stock price dynamics. Its popularity partly comes from the existence of semiclosed forms for vanilla European options, based on inverse Fourier transform and from its ability to reproduce some skewness shape of the implied volatility surface. We consider it under its risk-neutral probability measure, namely,

$$\begin{aligned} dS_t &= S_t(r dt + \sqrt{v_t} dW_t^1), & S_0 &= s_0 > 0, & (\text{risky asset}) \\ dv_t &= k(a - v_t)dt + \vartheta \sqrt{v_t} dW_t^2, & v_0 &> 0 \text{ with } d\langle W^1, W^2 \rangle_t = \rho dt, & \rho \in [-1, 1], \end{aligned}$$

where ϑ, k, a such that $\vartheta^2/(4ak) < 1$. We consider the Asian call payoff with maturity T and strike K . No closed form is available for its premium

$$\text{AsCall}^{\text{Hest}} = e^{-rT} \mathbb{E} \left(\frac{1}{T} \int_0^T S_s ds - K \right)_+.$$

We briefly recall how to proceed (see PAGÈS and PRINTEMPS [2005] for details): first, one projects W^1 on W^2 so that $W^1 = \rho W^2 + \sqrt{1 - \rho^2} \tilde{W}^1$ and decompose S_t as

$$\begin{aligned} S_t &= s_0 \exp \left(\left(r - \frac{1}{2} \bar{v}_t \right) t + \rho \int_0^t \sqrt{v_s} dW_s^2 \right) \exp \left(\sqrt{1 - \rho^2} \int_0^t \sqrt{v_s} d\tilde{W}_s^1 \right) \\ &= s_0 \exp \left(t \left(\left(r - \frac{\rho a k}{\vartheta} \right) + \bar{v}_t \left(\frac{\rho k}{\vartheta} - \frac{1}{2} \right) \right) + \frac{\rho}{\vartheta} (v_t - v_0) \right) \\ &\quad \exp \left(\sqrt{1 - \rho^2} \int_0^t \sqrt{v_s} d\tilde{W}_s^1 \right), \end{aligned}$$

where we have used the notation $\bar{v}_t = \frac{1}{t} \int_0^t v_s ds$ and where we have used the dynamics of (v_t) in the second line. The chaining rule for conditional expectations yields

$$\text{AsCall}^{\text{Hest}}(s_0, K) = e^{-rT} \mathbb{E} \left(\mathbb{E} \left(\left(\frac{1}{T} \int_0^T S_s ds - K \right)_+ \middle| \sigma(W_t^2, 0 \leq t \leq T) \right) \right).$$

Combining these two expressions and assuming $\tilde{W}^1$ and W^2 are independent imply that $\text{AsCall}^{\text{Hest}}(s_0, K)$ is a functional of $(\tilde{W}_t^1, v_t)$ (as concerns the squared volatility process v , only v_T and $\bar{v}_T$ are involved).

Let $\Gamma^{N_k} = \{w_1^{N_k}, \dots, w_{N_k}^{N_k}\}$ be an N_k -quantizer of the Brownian motion for $k = 1, 2$ and set the (non-Voronoi) (N_1, N_2) -quantization of (v_t, S_t) by

$$\begin{aligned}\hat{v}_t^{N_1} &= \sum_{i=1}^{N_1} y_i^{N_1}(t) \mathbf{1}_{C_i(\Gamma^{N_1})}(W^2) \\ \hat{S}_t^{N_1, N_2} &= \sum_{1 \leq i \leq N_1} \sum_{1 \leq j \leq N_2} s_{i,j}^{N_1, N_2}(t) \mathbf{1}_{C_i(\Gamma^{N_1})}(W^2) \mathbf{1}_{C_j(\Gamma^{N_2})}(\tilde{W}^1),\end{aligned}$$

where $y_i = y_i^{N_1}$ and $s_{i,j} = s_{i,j}^{N_1, N_2}$, for $i = 1, \dots, N_1$ and $j = 1, \dots, N_2$, are the solutions of the following ordinary differential equations:

$$\begin{aligned}\frac{dy_i}{dt}(t) &= k \left(a - y_i(t) - \frac{\vartheta^2}{4k} \right) + \vartheta \sqrt{y_i(t)} \frac{dw_i^{N_1}}{dt}(t), \quad y_i(0) = v_0, \\ \frac{ds_{i,j}}{dt}(t) &= s_{i,j}(t) \left(r - \frac{1}{2} y_i(t) - \frac{\rho \vartheta}{4} + \right. \\ &\quad \left. \sqrt{y_i(t)} \left(\rho \frac{dw_i^{N_1}}{dt}(t) + \sqrt{1 - \rho^2} \frac{dw_j^{N_2}}{dt}(t) \right) \right), \quad s_{i,j}(0) = s_0.\end{aligned}\tag{8.3}$$

These ODEs are solved using, for example, a Runge-Kutta scheme. Note that these formulæ require the computation of the $N_1 \times N_2$ quantized stochastic integrals $\int_0^t \sqrt{y_i^{N_1}}(s) dw_j^{N_2}(s)$ (which corresponds to the independent case). Let us point out that it is well known that (8.3) can be solved more or less in an explicit way for special sets of parameter of the model (when $\theta^2 = 4ak$) as emphasized by ROGERS [1995]; in this case, there is no more time integration error and the computations can be made significantly faster (see PAGÈS and PRINTEMPS [2005]). It is not the case of the selected parameters in the present chapter (see further on). The weights of the product cells $\{\tilde{W}^1 \in C_j(\Gamma^{N_2}), W^2 \in C_i(\Gamma^{N_1})\}$ are given by

$$\mathbb{P}(\tilde{W}^1 \in C_j(w^{N_2}), W^2 \in C_i(w^{N_1})) = \mathbb{P}(\tilde{W}^1 \in C_j(\Gamma^{N_2})) \mathbb{P}(W^2 \in C_i(\Gamma^{N_1}))$$

owing to the independence of $\tilde{W}^1$ and W^e . Eventually, the premium is approximated by

$$\begin{aligned}\widehat{\text{AsCall}}_{N_1, N_2}^{\text{Hest}}(s_0, K) &= e^{-rT} \mathbb{E} \left(\frac{1}{T} \int_0^T \hat{S}_t^{N_1, N_2} dt - K \right)_+ \\ &= e^{-rT} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\frac{1}{T} \int_0^T s_{i,j}^{N_1, N_2}(t) dt - K \right)_+ \\ &\quad \times \mathbb{P}(\tilde{W}^1 \in C_j(\Gamma^{N_2})) \mathbb{P}(W^2 \in C_i(\Gamma^{N_1})).\end{aligned}$$

We follow the guidelines of the methodology introduced by PAGÈS and PRINTEMPS [2005]: we compute the crude quantized premium for two sizes (N_1, N_2) and (N'_1, N'_2) ,

then proceed a space R - R – log extrapolation, where $\log(N)$ (respectively, $\log(N')$) is replaced by $\log(N_1 N_2)$ (respectively, by $\log(N'_1 N'_2)$). Finally, we make a K -linear interpolation based on the (Asian) forward moneyness $s_0 e^{rT} \left(\frac{1 - e^{-rT}}{rT} \right) \approx s_0 e^{rT}$ and the Asian call put parity formula

$$\text{AsianCall}^{\text{Hest}}(s_0, K) = \text{AsianPut}^{\text{Hest}}(s_0, K) + s_0 \left(\frac{1 - e^{-rT}}{rT} \right) - K e^{-rT}.$$

To be precise, we set with obvious notations

$$(i) \widehat{\text{AsCallParity}}^{\text{Hest}}_{N_1, N_2}(s_0, K) = \widehat{\text{AsPut}}^{\text{Hest}}_{N_1, N_2}(s_0, K) + s_0 \left(\frac{1 - e^{-rT}}{rT} \right) - K e^{-rT},$$

$$(ii) \widehat{\text{AsCallInterp}}^{\text{Hest}}_{N_1, N_2} = \frac{(K_{\max} - K) \widehat{\text{AsCallParity}}^{\text{Hest}}_{N_1, N_2} + (K - K_{\min}) \widehat{\text{AsCall}}^{\text{Hest}}_{N_1, N_2}}{K_{\max} - K_{\min}}.$$

The anchor strikes $K_{\min}$ and $K_{\max}$ of the interpolation are chosen symmetric with respect to the forward moneyness. At $K_{\max}$, the call is deep out of the money: we use the R - R extrapolated functional quantization computation; at $K_{\min}$, the call is deep in the money: one computes the call by parity. Between $K_{\min}$ and $K_{\max}$, we proceed a linear interpolation in K . This interpolation yields the best results within the range of variance of our model, compared with other extrapolations like the regression approach. Further comments on this step are made at the end of the section.

- *Parameters of the Heston model:* $s_0 = 100$, $k = 2$, $a = 0.01$, $\rho = 0.5$, $v_0 = 10\%$, $\vartheta = 20\%$.
- *Parameters of the option portfolio:* $T = 1$, $K = 99, \dots, 111$ (13 strikes).
- Fig. 8.1(b) depicts a N -quantizer of the Heston volatility process with $N = 400$ for this set of parameters obtained from an optimal N -quantizer of the Brownian motion (see Fig. 8.1(a)) by solving (8.3).
- The reference price has been computed by a Monte Carlo simulation of size $M_{\text{MC}} = 10^8$ (including a time R - R extrapolation of the Euler scheme with $2n = 256$).
- The differential equations (8.3) are solved with the parameters of the quantization cubature formulae $\Delta t = 1/32$, with couples of quantization levels $(N_1, N_2, N'_1, N'_2) = (1000, 1000, 100, 100)$, $(3200, 3200, 400, 400)$. See Fig. 8.3 for an illustration of the convergence rate in time.

Functional quantization can compute a whole vector (more than 10) option premia for the Asian option in the Heston model with 1 cent accuracy in less than 1 second (implementation in C on a 2.5-GHz processor) see Fig. 8.2.

Further numerical tests carried out or in progress with the Black-Scholes or *SABR* models (vanilla, Asian European options) confirmed that the efficiency of the model is not model dependent.

Let us now give some insights about the choice of the couple (N_1, N_2) that has to be used in (8.3). Other numerical experiments suggest that it depends on both the standard

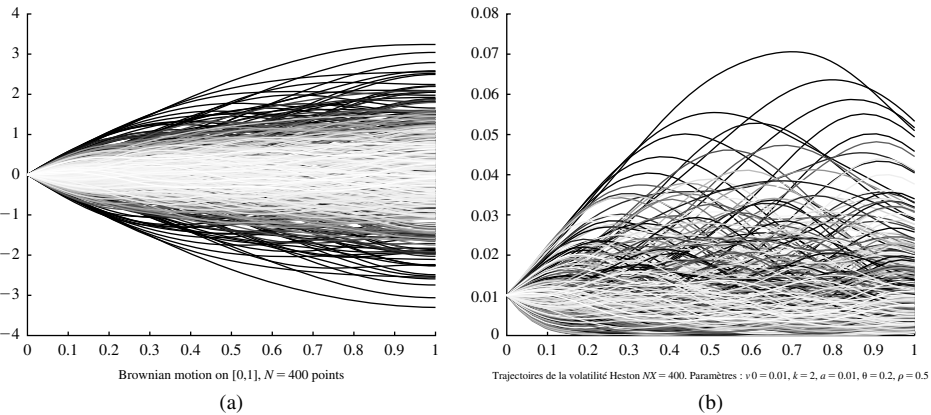


FIG. 8.1 N -quantizer of the Heston squared volatility process (v_t) ($N = 400$) resulting from an (optimized) N -quantizer of W .

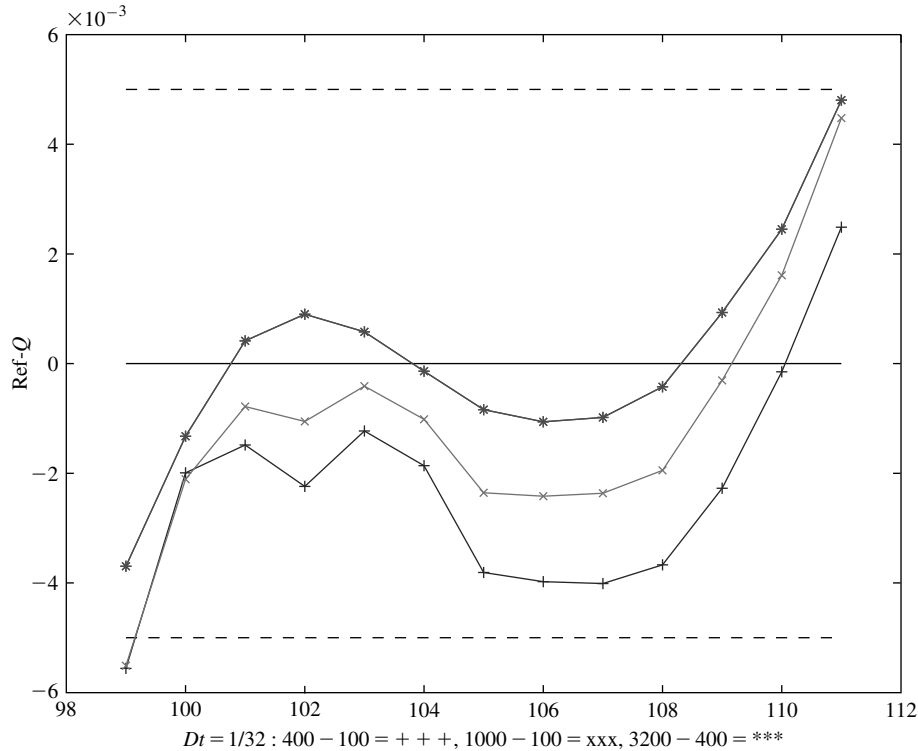


FIG. 8.2 Quantized diffusions based on optimal functional quantization: Pricing by K -interpolated, log R - R -extrapolated functional quantization prices as a function of K : absolute error with $(N_1, N_2, N'_1, N'_2) = (400, 400, 100, 100), (1000, 1000, 100, 100)$ and $(3200, 3200, 400, 400)$. $T = 1, s_0 = 100, K \in \{99, \dots, 111\}; k = 2, a = 0.01, \rho = 0.5, \theta = 0.2$.

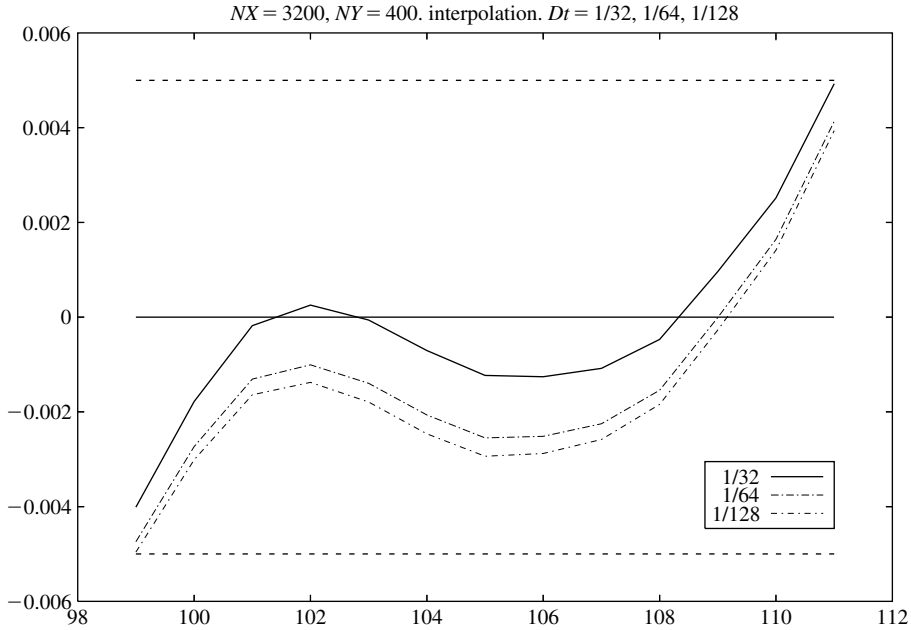


FIG. 8.3 Quantized diffusions based on optimal functional quantization: Pricing by K interpolated, log- R -extrapolated functional quantization price as a function of K : convergence as $\Delta t \in \{1/32, 1/64, 1/128\}$ with $(N_1, N_2, N'_1, N'_2) = (3200, 3200, 400, 400)$ (absolute error). $T = 1$, $s_0 = 100$, $K \in \{99, \dots, 111\}$; $k = 2$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.2$.

deviation $\text{StDev}(S_T)$ of S_T and the sign of ρ . For low values of $\text{StDev}(S_T)$ (let us say 10% of s_0), square quantizers (i.e. $N_1 = N_2$) are relevant. It means that, for a given size $N_1 \times N_2$, a good precision can be obtained using a square quantizer. However, when $\text{StDev}(S_T)$ increases, further numerical simulations not reproduced here show that this choice of a square quantizer is not optimal for a given complexity. To be precise, we have done numerical simulations with various 4-tuples (N_1, N_2, N'_1, N'_2) for two sets of model parameters such that the variance is about 200 for $\tilde{S}_T$ in the two cases ($s_0 = 100$, $r = 0.1$, $k = 2$, $(a, \rho) \in \{(0.06; -0.6), (0.047; 0.6)\}$). The primes denote the second couple used in the R - R extrapolation.

Then, we have selected the 4-tuples (N_1, N_2, N'_1, N'_2) giving “good” call prices (the reference prices are still computed by a Monte Carlo simulation of size $M_{\text{MC}} = 10^8$, and “good” prices means that they stand within one standard deviation interval) around the forward moneyness ($K \in \{95, 100, 105, 110\}$) and having the smallest quantization errors as regards $\text{AsCall} - \text{AsCallParity}$, that is,

$$\mathbb{E} \int_0^T (S_t - \hat{S}_t^{N_1, N_2}) dt.$$

This second criterion has been chosen since it is more “parameter free” than the option prices themselves.

TABLE 8.1

Asian Heston prices with the parameters : $s_0 = 100$, $r = 0.1$, $k = 2$, $v_0 = a = 0.06$, $\theta = 0.5$ and $\rho = -0.6$ and where K here denotes the strikes prices. Here $(N_1, N_2, N'_1, N'_2) = (1700, 800, 2000, 900)$, i.e. $\alpha \sim \alpha' \sim 0.36$. The final prices have been computing using formula (8.4).

K	95	100	105	110
Reference price	11.200	7.948	5.244	3.170
$LSR\widehat{AsCall}^{Hest}$	11.204 (0.036%)	7.947 (0.012%)	5.246 (0.038%)	3.172 (0.063%)

In order to discuss the results, we set $\alpha := (N_1 - N_2)/(N_1 + N_2)$ and $\alpha' := (N'_1 - N'_2)/(N'_1 + N'_2)$ and we showed in a plane (α, α') the 4-tuples selected above. Two conclusions can, then, be drawn. The first conclusion is that the “good” prices lie in the (α, α') along the diagonal $\alpha' = \alpha$. Not surprisingly, it means that during the R - R extrapolation, the couples (N_1, N_2, N'_1, N'_2) have to be chosen so that the relative parts of the quantizations of the volatility and the asset remain the same. The second information is that, in the case $\rho < 0$, most 4-tuples lie along the positive part of the diagonal $\alpha \sim \alpha' > 0$, whereas when $\rho > 0$, most of them lies in the negative part $\alpha \sim \alpha' < 0$. It means that a good choice is $N_1 > N_2$ when $\rho < 0$ and $N_1 < N_2$ when $\rho > 0$.

Let us also remark that the choice of the K -interpolation in (ii) should be tempered when the variance of $\bar{S}_T$ increases. For high variances, a better choice is to use a LSR -based variance reduction technique inherited from Monte Carlo method. Namely, we interpolate $AsCall$ and $AsCallParity$ (as given by (ii)) by computing (using quantization)

$$LSR\widehat{AsCall} := \widehat{AsCall} - \frac{\widehat{Cov}(X, X - Y)}{\widehat{Var}(X - Y)} \times (\widehat{AsCall} - \widehat{AsCallParity}) \quad (8.4)$$

given that $X = e^{-rT} \left(\frac{1}{T} \int_0^T \hat{S}_s ds - K \right)_t$ and $X - Y = e^{-rT} \frac{1}{T} \int_0^T \hat{S}_t dt - s_0 \frac{1 - e^{-rT}}{rT}$ (^{Hest} and the size of the grid are dropped for notational convenience). This approach needs further computations (covariances and variances). When the variance of S_T is not too high, the extremal values of λ at anchor strikes are nearly 0 or 1, respectively, and the LSR interpolation coincides with the (slightly faster) K -linear interpolation. An example of numerical results is reported in Table 8.1 (with a negative correlation).

8.3. Comparison: optimized quantization versus (optimal) product quantization

The comparison is balanced and probably needs some further in situ experiments since it may depend on the modes of the computation. However, it seems that product quantizers (as those implemented by PAGÈS and PRINTEMS [2005]) are from two to four times less efficient than optimal quantizers within our range of application (small values of the N_i 's

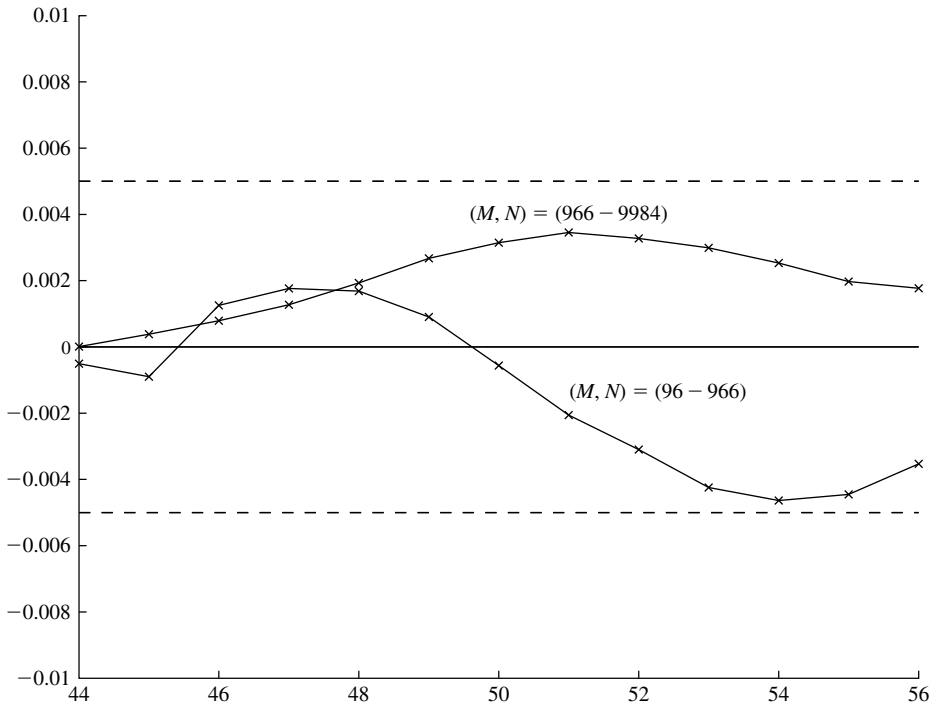


FIG. 8.4 Quantized diffusions based on optimal product quantization: Pricing by K -linear interpolation of R - R log-extrapolations as a function of K (absolute error) with $N_1 = N_2 = N$, $N'_1 = N'_2 = n$, $(n, N) = (96, 966)$, $(n, N) = (966, 9984)$, $T = 1$, $s_0 = 50$, $K \in \{44, \dots, 56\}$; $k = 2$, $a = 0.01$, $\rho = 0.5$, $\vartheta = 0.1$.

and N'_i 's). See Fig. 8.4 for a numerical test (with $N_1 = N_2$, $N'_1 = N'_2$). On the other hand, the design of product quantizer from one-dimensional scalar quantizers is easy and can be made from some light elementary “bricks” (the scalar quantizer up to $N = 35$ and the optimal allocation rules). Thus, the whole set of data needed to design all optimal product quantizers up to $N = 10\,000$ is approximately 500 KB, whereas one optimal quantizer with size 10 000 $\approx$ 1 MB...

9. Universal quantization rate and mean regularity

The following theorem points out the connection between functional quantization rate and mean regularity of $t \mapsto X_t$ from $[0, T]$ to $L^r(\mathbb{P})$.

THEOREM 9.1 (LUSCHGY and PAGÈS [2007]). *Let $X = (X_t)_{t \in [0, T]}$ be a stochastic process. If there is $r^* \in (0, \infty)$ and $a \in (0, 1]$ such that*

$$X_0 \in L^{r^*}(\mathbb{P}), \quad \|X_t - X_s\|_{L^{r^*}(\mathbb{P})} \leq C_X |t - s|^a,$$

for some positive real constant $C_X > 0$, then

$$\forall p, r \in (0, r^*), \quad e_{N,r}(X, L_T^p) = O((\log N)^{-a}).$$

The proof is based on a constructive approach, which involves the Haar basis (instead of K-L basis), the nonasymptotic version Zador theorem, and product functional quantization. Roughly speaking, we use the unconditionality of the Haar basis in every L_T^p (when $1 < p < \infty$) and its wavelet feature, that is, its ability to “code” the path regularity of a function on the decay rate of its coordinates.

EXAMPLES (see LUSCHGY and PAGÈS [2007]):

- d -dimensional Itô processes (includes d -dim diffusions with sublinear coefficients) with $a = 1/2$.
- General Lévy process X with Lévy measure ν with square-integrable big jumps. If X has a Brownian component, then $a = 1/2$; otherwise, if $\beta(X) > 0$, where $\beta(X) := \inf \{ \theta : \int |y|^\theta \nu(dy) < +\infty \} \in (0, 2)$ (Blumenthal–Gettoor index of X), then $a = 1/\beta(X)$. This rate is the exact rate, that is,

$$e_{N,r}(X, L_T^p) \approx (\log N)^{-a}$$

for many classes of Lévy processes like symmetric stable processes, Lévy processes having a Brownian component, etc (see LUSCHGY and PAGÈS [2007] for further examples).

- When X is a compound Poisson process, then $\beta(X) = 0$ and 1 shows, still with constructive methods, that

$$e_N(X) = O(e^{-(\log N)^\vartheta}), \quad \vartheta \in (0, 1),$$

which is in-between the finite- and infinite-dimensional settings.

10. About lower bounds for functional quantization

In this overview, we gave no clue toward lower bounds for functional quantization although most of the rates we mentioned are either exact ($\approx$) or sharp ($\sim$) (we tried to emphasize the numerical aspects). Several approaches can be developed to get some lower bounds. Historically, the first one was to rely on subadditivity property of the quantization error derived from self-similarity of the distribution: this works with the uniform distribution over $[0, 1]^d$ but also in an infinite-dimensional framework (see DEREICH and SCHEUTZOW [2006] for the fractional Brownian motion).

A second approach consists in pointing out the connection with the Shannon–Kolmogorov entropy (see LUSCHGY and PAGÈS [2002]) using that the entropy of a random variable taking at most N values is at most $\log N$.

A third connection can be made with small deviation theory (see DEREICH and SCHEUTZOW [2003], GRAF, LUSCHGY and PAGÈS [2003], LUSCHGY and PAGÈS [2007]). Thus, in the study by GRAF, LUSCHGY and PAGÈS [2003], a connection is established between (functional) quantization and small ball deviation for Gaussian processes. In

particular, this approach provides a method to derive a lower bound for the quantization rate from some upper bound for the small deviation problem. A careful reading of the proof of theorem 1.2 in GRAF, LUSCHGY and PAGÈS [2003] shows that this small deviation lower bound holds for any *unimodal* (with respect to 0) nonzero process. To be precise, assume that $\mathbb{P}_X$ is L_T^p -unimodal, that is, there exists a real $\varepsilon_0 > 0$ such that

$$\forall x \in L_T^p, \forall \varepsilon \in (0, \varepsilon_0], \quad \mathbb{P}(|X - x|_{L_T^p} \leq \varepsilon) \leq \mathbb{P}(|X|_{L_T^p} \leq \varepsilon).$$

For centered Gaussian processes (or processes “subordinated” to Gaussian processes), this follows from the Anderson inequality (when $p \geq 1$). If

$$G(-\log(\mathbb{P}(|X|_{L_T^p} \leq \varepsilon))) = \Omega(1/\varepsilon) \quad \text{as } \varepsilon \rightarrow 0$$

for some increasing unbounded function $G : (0, \infty) \rightarrow (0, \infty)$, then

$$\forall c > 1, \quad \liminf_N G(\log(cN)) e_{N,r}(X, L_T^p) > 0, \quad r \in (0, \infty). \quad (10.1)$$

This approach is efficient in the nonquadratic case as emphasized by LUSCHGY and PAGÈS [2007], where several universal bounds are shown to be optimal using this approach.

11. Toward new applications: a guided Monte Carlo method

This section provides some preliminary (and theoretical) elements about a quantization-based stratification method to reduce the variance of a Monte Carlo simulation. It can be seen as a guided Monte Carlo method or a hybrid quantization/Monte Carlo method. This method has been introduced by PAGÈS and PRINTEMS [2005] for Lipschitz functionals of the Brownian motion. Here, we will mainly focus on the finite-dimensional case and consider some more regular functions. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ be square-integrable random vector. For more details and some simulation results, we refer to PAGÈS and PRINTEMS [2007].

▷ *Lipschitz functions.* Let $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a Lipschitz function. In order to compute $\mathbb{E}(F(X))$, one writes

$$\begin{aligned} \mathbb{E}(F(X)) &= \mathbb{E}(F(\hat{X}^N)) + \mathbb{E}\left(F(X) - F(\hat{X}^N)\right) \\ &= \underbrace{\mathbb{E}(F(\hat{X}^N))}_{(a)} + \underbrace{\frac{1}{M} \sum_{m=1}^M F(X^{(m)}) - F(\widehat{X^{(m)}}^N)}_{(b)} + R_{N,M}, \end{aligned} \quad (11.1)$$

where $X^{(m)}$, $m = 1, \dots, M$ are M independent copies of the standard Brownian motion, $\widehat{\cdot}^N$ denotes the nearest-neighbor projection on a fixed N -quantizer, and $R_{N,M}$ is a remainder term defined by (11.1). Term (a) can be computed by quantization, and term (b) can

computed by a Monte Carlo simulation. Then,

$$\|R_{N,M}\|_2 = \frac{\sigma(F(X) - F(\hat{X}^N))}{\sqrt{M}} \leq \frac{\|F(X) - F(\hat{X}^N)\|_2}{\sqrt{M}} \quad \text{and,} \\ \sqrt{M} R_{N,M} \xrightarrow{L} \mathcal{N}(0; \text{Var}(F(X) - F(\hat{X}^N)))$$

as $M \rightarrow +\infty$, where $\sigma(Y)$ denotes the standard deviation of a random variable Y . Consequently, if F is simply a Lipschitz function and if $(\hat{X}^N)_{N \geq 1}$ is a rate optimal sequence quantization of X , then

$$\|F(X) - F(\hat{X}^N)\|_2 \leq \frac{[F]_{\text{Lip}} C_X}{N^{\frac{1}{d}}} \quad \text{and} \quad \|R_{N,M}\|_{L^2_T} \leq \frac{[F]_{\text{Lip}} C_X}{M^{\frac{1}{2}} N^{\frac{1}{d}}}.$$

The main gap to implement this variance reduction method is the nearest-neighbor search needed at each step to simulate $\widehat{X}^{(m)N}$ from $X^{(m)}$ (assuming that the quantizer is known as well as its weights). If one uses a product quantization, this search reduces to the marginals and the complexity of a nearest number search on the real line based on dichotomy is approximately $\frac{\log N}{\log 2}$ (once the N points of interest are sorted). As concerns nonproduct quantizations, fast nearest-neighbor search procedures can also be implemented (if N is not too small, see FRIEDMAN, BENTLEY and FINKEL [1977]).

▷ *Quantization based stratification.* In many natural situations (e.g. $d = 1$ or $d \geq 2$ with a product quantizer X^N , one has an explicit expression for the conditional distribution of X given $\hat{X}^N = \text{Proj}_{\Gamma_N}(X)$ that is, we can write

$$X = \varphi(\hat{X}^N, U)$$

where U is independent of $\hat{X}^N$ with a distribution $\mu := \mathbb{P}_U$ on $\mathbb{R}^q$ and $\varphi : \Gamma_N \times \mathbb{R}^q \rightarrow \mathbb{R}^d$ a Borel function. The probability distribution μ is supposed to be easy to simulate and φ easy to compute. In fact, from a theoretical viewpoint, one may always assume that U is uniformly distributed on a unit hypercube $[0, 1]^q$ (or even the unit interval $[0, 1]$). Then, one derives that for every Borel function $F : \mathbb{R}^d \rightarrow \mathbb{R}$ in $L^1(\mathbb{P}_X)$,

$$\begin{aligned} \mathbb{E} F(X) &= \mathbb{E}(\mathbb{E}(F(\varphi(\hat{X}^N, U)) | U)) \\ &= \mathbb{E}(\Phi_F(U)) \end{aligned} \tag{11.2}$$

with

$$\Phi_F(u) := \mathbb{E}(F(X) | U = u) = \sum_{i=1}^N \mathbb{P}(\hat{X}^N = x_i) F(\varphi(x_i, u)), \quad u \in [0, 1]^q$$

since $\hat{X}^N$ and U are independent. Then, one derives that

$$\begin{aligned} \text{Var}(\Phi_F(U)) &= \text{Var}(\Phi_F(U) - \mathbb{E} F(\hat{X}^N)) \\ &= \text{Var}(\mathbb{E}(F(\varphi(\hat{X}^N, U)) - F(\hat{X}^N) | U)) \\ &\leq \text{Var}(F(X) - F(\hat{X}^N)) \end{aligned} \tag{11.3}$$

where we used that conditional expectation is an $L^2(\mathbb{P})$ -contraction, preserves expectation and again that $\widehat{X}^N$ and U are independent. Furthermore the *above last inequality is strict* (except if $F(X) = F(\widehat{X}^N) + G(U)$). As expected this stratification produces an additional variance reduction with respect to the above regular hybrid Monte Carlo-quantization method.

12. Acknowledgment

We thank S. Graf (University of Passau), H. Luschgy (University of Trier), and B. Wilbertz (University of Trier) for all the fruitful discussions and collaborations we have about quantization. S. Bouthemy (Gaz de France) made the simulations of the chapter on swing options.

References

- ABAYA, E.F., WISE, G.L. (1982). On the existence of optimal quantizers. *IEEE Trans. Inform. Theory* **28**, 937–940.
- ABAYA, E.F., WISE, G.L. (1984). Some remarks on the existence of optimal quantizers. *Stat. Probab. Lett.* **2**, 349–351.
- BALLY, V., PAGÈS, G. (2003a). A quantization algorithm for solving discrete time multidimensional optimal stopping problems. *Bernoulli* **9** (6), 1003–1049.
- BALLY, V., PAGÈS, G. (2003b). Error analysis of the quantization algorithm for obstacle problems, *Stoch. Proc. Appl.* **106** (1), 1–40.
- BALLY, V., PAGÈS, G., PRINTEMS, J. (2001). A Stochastic quantization method for nonlinear problems, *Monte Carlo Methods Appl.* **7** (1), 21–34.
- BALLY, V., PAGÈS, G., PRINTEMS, J. (2003). First order schemes in the numerical quantization method. *Math. Financ.* **13** (1), 1–16.
- BALLY, V., PAGÈS, G., PRINTEMS, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Math. Financ.* **15** (1), 119–168.
- BARDOU, O., BOUTHEMY, S., PAGÈS, G. (2007a). Pricing swing options using optimal quantization, pre-print LPMA-1146, To appear in Journal of Applied Finance.
- BARDOU, O., BOUTHEMY, S., PAGÈS, G. (2007b). When are swing option bang-bang and how to use it, pre-print LPMA-1141, submitted.
- BENVENISTE, A., MÉTIVIER, M., PRIURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*, Translated from the French by Stephen S. Wilson. Applications of Mathematics **22** (Springer-Verlag, Berlin, Germany), pp. 365.
- BOULEAU, N., LÉPINGLE, D. (1994). Numerical Methods For Stochastic Processes. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication* (John Wiley and Sons Inc., New York, NY), pp. 359. ISBN 0-471-54641-0.
- BUCKLEW, J.A., WISE, G.L. (1982). Multidimensional asymptotic quantization theory with r^{th} power distortion. *IEEE Trans. Inform. Theory* **28** (2), 239–247.
- COHORT, P. (1998). A geometric method for uniqueness of locally optimal quantizer. Preprint LPMA-464 and Ph.D. thesis, *Sur quelques problèmes de quantification*, 2000, Univ. Paris 6.
- CUESTA-ALBERTOS, J.A., MATRÁN, C. (1988). The strong law of large numbers for k -means and best possible nets of Banach valued random variables. *Probab. Theory. Rel.* **78**, 523–534.
- DELATTRE, S., FORT, J.-C., PAGÈS, G. (2004). Local distortion and μ -mass of the cells of one dimensional asymptotically optimal quantizers. *Commun. Stat.* **33** (5), 1087–1118.
- DEREICH, S. (2008a). The coding complexity of diffusion processes under $L^p[0, 1]$ -norm distortion, pre-print, *Stoch. proc. Appl.* **118** (6), 938–951.
- DEREICH, S. (2008b). The coding complexity of diffusion processes under supremum norm distortion, pre-print, *Stoch. proc. Appl.* **118** (6), 917–937.
- DEREICH, S., FEHRINGER, F., MATOUSSI, A., SCHEUTZOW, M. (2003). On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theor. Probab.* **16**, 249–265.
- DEREICH, S., SCHEUTZOW, M. (2006). High resolution quantization and entropy coding for fractional Brownian motions, *Electron. J. Probab.* **11**, 700–722.

- DOSS, H. (1977). Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Probab. Statist.* **13** (2), 99–125.
- FLEISCHER, P.E. (1964). Sufficient conditions for achieving minimum distortion in a quantizer. *IEEE Int. Conv. Rec.* part I, 104–111.
- FORT, J.-C., PAGÈS, G. (2004). Asymptotics of optimal quantizers for some scalar distributions. *J. Comput. Appl. Math.* **146**, 253–275.
- FRIEDMAN, J.H., BENTLEY, J.L., FINKEL, R.A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM. T. Math. Software* **3** (3), 209–226.
- GERSHO, A., GRAY, R.M. (1992). *Vector Quantization and Signal Compression* (Kluwer, Boston, MA).
- GRAF, S., LUSCHGY, H. (2000). *Foundations of Quantization for Probability Distributions. Lect. Notes in Math. 1730* (Springer, Berlin, Germany), pp. 230.
- GRAF, S., LUSCHGY, H. (2005). The point density measure in the quantization of self-similar probabilities. *Math. Proc. Cambridge Phil. Soc.* **138**, 513–531.
- GRAF, S., LUSCHGY, H., PAGÈS, G. (2003). Functional quantization and small ball probabilities for Gaussian processes. *J. Theoret. Probab.* **16** (4), 1047–1062.
- GRAF, S., LUSCHGY, H., PAGÈS, G. (2007). Optimal quantizers for Radon random vectors in a Banach space. *J. Approx. Theory.* **144**, 27–53.
- GRAF, S., LUSCHGY, H., PAGÈS, G. (2008). Distortion mismatch in the quantization of probability measures, 18 *ESAIM Probab. Stat.* **12**, 127–153.
- HESTON, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.* **6** (2), 327–343.
- KIEFFER, J.C. (1982). Exponential rate of convergence for Lloyd's Method I. *IEEE Trans. Inform. Theory* **28** (2), 205–210.
- KIEFFER, J.C. (1983). Uniqueness of locally optimal quantizer for log-concave density and convex error weighting functions. *IEEE Trans. Inform. Theory* **29**, 42–47.
- KUSHNER, H.J., YIN, G.G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Second edition. Applications of Mathematics **35**. Stochastic Modelling and Applied Probability. (Springer-Verlag, New York, NY), pp. 474.
- LAMBERTON, D., PAGÈS, G. (1996). On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm. In: Verleysen, M. (ed.), *Proceedings of the ESANN'96*, (Editions D Facto, Bruxelles, Belgium), pp. 97–106.
- LAPEYRE, B., SAB, K., PAGÈS, G. (1990). Sequences with low discrepancy. Generalization and application to Robbins-Monro algorithm. *Stat* **21** (2), 251–272.
- LEJAY, A. (2003). An introduction to rough paths. *Séminaire de Probabilités XXXVII*, Lecture Notes in Mathematics 1832, (Springer, Berlin, Germany), pp. 1–59.
- LUSCHGY, H., PAGÈS, G. (2002). Functional quantization of Gaussian processes. *J. Funct. Anal.* **196** (2), 486–531.
- LUSCHGY, H., PAGÈS, G. (2004). Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.* **32** (2), 1574–1599.
- LUSCHGY, H., PAGÈS, G. (2006). Functional quantization of a class of Brownian diffusions: A constructive approach. *Stoch. proc. Appl.* **116**, 310–336.
- LUSCHGY, H., PAGÈS, G., WILBERTZ, B. (2007). Asymptotically optimal quantization schemes for Gaussian processes. To appear in *ESAIM Probab. Stat.*
- LUSCHGY, H., PAGÈS, G. (2007a). Expansion of Gaussian processes and Hilbert frames. Submitted.
- LUSCHGY, H., PAGÈS, G. (2007b). High-resolution product quantization for Gaussian processes under sup-norm distortion. *Bernoulli* **13** (3), 653–671.
- LUSCHGY, H., PAGÈS, G. (2008). Functional quantization rate and mean regularity of processes with an application to Lévy processes. *Ann. Appl. Probab.* **18** (2), 427–469.
- MCMANES, J. (2001). A fast nearest-neighbor algorithm based on a principal axis search tree. *IEEE T. Pattern. Anal.* **23** (9), 964–976.
- MRAD, M., BEN HAMIDA, S. (2006). Optimal quantization: evolutionary algorithm vs stochastic gradient. *Proceedings of the 9th Joint Conference on Information Sciences*.

- NEWMAN, D.J. (1982). The hexagon theorem. *IEEE Trans. Inform. Theory* **28**, 137–138.
- PÄRNA, K. (1990). On the existence and weak convergence of k -centers in Banach spaces. *Tartu Ülikooli Toimetised* **893**, 17–287.
- PAGÈS, G. (1993). Voronoi tessellation, space quantization algorithm and numerical integration. In: Verleysen, M. (ed.), *Proceedings of the ESANN'93*, (Editions D Facto, Bruxelles, Belgium), pp. 221–228.
- PAGÈS, G. (1998). A space vector quantization method for numerical integration. *J. Comput. Appl. Math.* **89**, 1–38.
- PAGÈS, G. (2000). Functional quantization: a first approach. Preprint CMP12-04-00 (Univ., Paris 12 France).
- PAGÈS, G. (2007). Quadratic optimal functional quantization methods and numerical applications. In: *Proceedings of MCQMC Ulm'06* (Springer, Berlin, Germany) 101–142.
- PAGÈS, G., PHAM, H., PRINTEMS, J. (2003). Optimal quantization methods and applications to numerical methods in finance. In: Rachev, S.T. (ed.), *Handbook of Computational and Numerical Methods in Finance* (Birkhäuser, Boston, MA), pp. 429.
- PAGÈS, G., PRINTEMS, J. (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Appl.* **9** (2), 135–165.
- PAGÈS, G., PRINTEMS, J. (2005). Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Appl.* **11** (4), 407–446.
- PAGÈS, G., PRINTEMS, J. (2005). Website devoted to vector and functional optimal quantization. <http://www.quantize.maths-fi.com>.
- PAGÈS, G., PRINTEMS, J. (2007). A hybrid Monte Carlo quantization method, in progress.
- PAGÈS, G., SAGNA, A. (2007). Asymptotics of the radius of an L^r -optimal sequence of quantizers, pre-pub. LPMA-1224, Univ. Paris 6 (France), submitted.
- PAGÈS, G., SELLAMI, A. (2007). Convergence of multi-dimensional quantized SDE's. pre-pub. LPMP 1196.
- POLLARD, D. (1982). Quantization and the method of k -means. *IEEE Trans. Inform. Theory* **28** (2), 199–205.
- PROINOV, P.D. (1988). Discrepancy and integration of continuous functions. *J. Approx. Theory* **52**, 121–131.
- REVUZ, D., YOR, M. (1999). *Continuous martingales and Brownian motion*, third ed. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **293**, (Springer-Verlag, Berlin, Germany), pp. 602.
- ROGERS, L.C.G. (1995). Which model for the term structure of interest rates should one use? *Math. Financ. IMA* **65**, 93–116.
- SAGNA, A. (2007). Universal L^s -rate-optimality of L^r -optimal quantizers by dilatation-contraction, pre-pub. LPMA 1164, Univ. Paris 6 (France), To appear in *ESAIM Probab. Stat.*
- TARPEY, T., KINATEDER, K.K.J. (2003a). Clustering functional data. *J. Classif.* **20**, 93–114.
- TARPEY, T., PETKOVA, E., OGDEN, R.T. (2003b). Profiling placebo responders by self-consistent partitioning of functional data. *J. Amer. Statist. Assoc.* **98**, 850–858.
- TRUSHKIN, A.V. (1982). Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions. *IEEE Trans. Inform. Theory* **28** (2), 187–198.
- WILBERTZ, B. (2005). Computational aspects of functional quantization for Gaussian measures and applications, diploma thesis, Univ. Trier, Germany.
- ZADOR, P.L. (1963). Development and evaluation of procedures for quantizing multivariate distributions, Ph.D. dissertation, Stanford Univ., Stanford, CA.
- ZADOR, P.L. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory* **28** (2), 139–149.

Stochastic Clock and Financial Markets

Hélyette Geman

*Birkbeck, University of London & ESSEC Business School,
Birkbeck, University of London–Malet Street, Bloomsbury–London WC1E 7HX
E-mail address: h.geman@bbk.ac.uk*

Abstract

Brownian motion played a central role throughout the twentieth century in probability theory. The same statement is even truer in finance, with the introduction in 1900 by the French mathematician Louis Bachelier of an arithmetic Brownian motion (or a version of it) to represent stock price dynamics. This process was pragmatically transformed by Samuelson in 1965 into a geometric Brownian motion ensuring the positivity of stock prices.

More recently, the elegant martingale property under an equivalent probability measure derived from the no-arbitrage assumption combined with Monroe's theorem on the representation of semimartingales has led to write asset prices as time-changed Brownian motion. Independently, CLARK [1973] had the original idea of writing cotton future prices as subordinated processes, with Brownian motion as the driving process. Over the last few years, time changes have been used to account for different speeds in market activity in relation to news arrival as the stochastic clock goes faster during periods of intense trading. They have also allowed us to uncover new classes of processes in asset price modeling.

1. Introduction

The twentieth century started with the pioneer dissertation of Louis BACHELIER [1900] and the introduction of Brownian motion for stock price modeling. It also ended with Brownian motion as a central element in the representation of the dynamics of assets such as bonds, commodities, or stocks. The reasonable assumption of the nonexistence of arbitrage opportunities in financial markets led to the first fundamental theorem of asset pricing. From the representation of discounted asset prices as martingales under an

equivalent martingale measure, the semimartingale property for asset prices under the real probability measure was then derived and in turn, the expression of these prices as time-changed Brownian motion.

Since the seminal papers by BLACK and SCHOLES [1973] and MERTON [1973] on the pricing of options, the theory of no-arbitrage has played a central role in finance. It is, in fact, amazing how much can be deduced from the simple economic assumption that it is not possible in a financial market to make profits with zero investment and without bearing any risk. Unsurprisingly, practitioners in various sectors of the economy are prepared to subscribe to this assumption, hence the harvest the flurry of results derived from it. Pioneering work on the relation between no-arbitrage arguments and martingale theory was conducted in the late seventies and early eighties by Harrison–Kreps–Pliska: HARRISON and KREPS [1979] introduced in a discrete-time setting the notion of equivalent martingale measure. HARRISON and PLISKA [1981] examined the particular case of complete markets and established the unicity of the equivalent martingale measure. A vast amount of research grew out of these remarkable results: DALANG, MORTON and WILLINGER [1990] extended the discrete-time results to a general probability space Ω . DELBAEN and SCHACHERMAYER [1994] chose general semimartingales for the price processes of primitive assets and established the following result.

1.1. First Fundamental Theorem of Asset Pricing

The market model is arbitrage free if and only if there exists a probability measure Q equivalent to P (and often called equivalent martingale measure) with respect to which the discounted prices of primitive securities are martingales.

We consider the classical setting of a filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{F}, P)$ whose filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ represents the flow of information accruing to the agents in the economy, that is, right continuous and $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$; we are essentially considering in this chapter a finite horizon T . The security market consists of $(n + 1)$ primitive securities: $(S_i(t))_{0 \leq t \leq T}$, $i = 1 \dots n$ denotes the price process of the n stocks and S_0 is the money-market account that grows at a rate r supposed to be constant in the Black–Scholes–Merton and Harrison–Kreps–Pliska setting. Initially, the process S is only assumed to be locally bounded, a fairly general assumption that covers in particular the case of continuous price processes. We assume that the process S is a semimartingale, adapted to the filtration $\mathcal{F}_t$ and satisfying the condition of being right continuous and limited to the left. The semimartingale property has to prevail for the process S in an arbitrage-free market: by the first fundamental theorem of asset pricing mentioned above, the discounted stock price process is a martingale under an equivalent martingale measure; hence, the stock price process has to be a semimartingale under the initial probability measure P .

A self-financing portfolio Π is defined as a pair (x, H) , where the constant x is the initial value of the portfolio and $H = (H^i)_{0 \leq i \leq n}$ is a predictable S -integrable process specifying the amount of each asset in the portfolio.

The value process of such a portfolio Π at time t is given by

$$V(t) = x_0 + \int_0^t H_u \cdot dS_u \quad 0 \leq t \leq T.$$

In order to rule out strategies generating arbitrage opportunities by going through times where the portfolio value is very negative, HARRISON and PLISKA [1981] defined a predictable, S integrable process H as admissible if there exists a positive constant C such that

$$(H \cdot S)_t = \int_0^t H_u \cdot dS_u \geq -C \quad \text{for } 0 \leq t \leq T.$$

This condition has been required in all the subsequent literature; it also has the merit to be consistent with the reality of financial markets since margin calls imply that losses are bounded.

In the particular case of a discrete-time representation, each R^d -valued process $(H_t)_{t=1}^T$ that is predictable (i.e., each H_t is $F_{(t-1)}$ measurable) is S integrable, and the stochastic integral $(H \cdot S)$ reduces to a sum

$$(H \cdot S) = \int_0^T H_u \cdot dS_u = \sum_{u=1}^T H_u \cdot (\Delta S_u) = \sum_{u=1}^T H_u \cdot (S_u - S_{u-1}),$$

where $H_u \cdot \Delta S_u$ denotes that the inner product of the vectors H_u and $\Delta S_u = S_u - S_{u-1}$ belong to R^d . Of course, such a trading strategy H is admissible if the underlying probability space Ω is finite.

We define a contingent claim as an element of $L^\infty(\Omega, \mathcal{F}, P)$. A contingent claim C is said to be attainable if there exists a self-financing trading strategy H whose terminal value at date T is equal to C . Assuming momentarily zero interest rates for simplicity, we denote by A_0 the subspace of $L^\infty(\Omega, \mathcal{F}, P)$ formed by the random variables $(H \cdot S)_T$, representing the value at time T of attainable contingent claims, and denote by J the linear space spanned by A_0 and the constant 1. The no-arbitrage assumption implies that the set J and the positive orthant with the origin deleted, denoted as K , have an empty intersection. Hence, from the Hahn–Banach theorem, there exists (at least) a hyperplane G containing J and such that $G \cap K = \emptyset$.

We can then define the linear functional χ by $\chi/G = 0$ and $\chi(1) = 1$. This linear functional χ may be identified with a probability measure Q on F by $\chi = \frac{dQ}{dP}$, and χ is strictly positive if and only if the probability measure Q is equivalent to P . In addition, χ vanishes on A_0 if and only if S is a martingale under Q , and this provides a brief proof of the first fundamental theorem of asset pricing (the other implication being simple to demonstrate). The proof is extended to nonzero (constant) interest rates in a nonelementary manner (see ARTZNER and DELBAEN [1989]) and stochastic interest rates in a study by GEMAN [1989].

2. Time changes in mathematics

2.1. Time changes: the origins

The presence of time changes in probability theory can be traced back to the French mathematician Doeblin who, in 1940, studied real-valued diffusions and wrote the

“Fundamental martingales” attached to a diffusion as time-changed Brownian motion. VOLKONSKI [1958] used time changes to study Markov processes. A considerable contribution to the subject was brought by ITÔ and McKEAN [1965] (see also FELLER [1964]) who showed that time changes allow to replace the study of diffusions by the study of Brownian motion. McKEAN [2001], in his beautifully entitled paper *Scale and Clock*, revisited space and time transformations and how they allow to reduce the study of complex semimartingales to that of more familiar processes. In the framework of finance and in continuity with our previous section on the martingale representation of discounted stock prices, we need to mention the theorem by DUBINS and SCHWARZ [1965] and DAMBIS [1965]:

“Any continuous martingale is a time-changed Brownian motion.”

The time change that transforms the process S to this new scale is the quadratic variation, which is the limit of the sum of squared increments when the time step goes to zero. For standard Brownian motion, the quadratic variation over any interval is equal to the length of the interval. If the correct limiting procedure is applied, then the sum of squared increments of a continuous martingale converges to the quadratic variation. This quadratic variation has recently become in finance the subject of great attention with the development of instruments such as variance swaps and options on volatility index (see CARR, GEMAN, MADAN and YOR [2005]).

Continuing the review of the major mathematical results on time changes, we need to mention the theorem by LAMPERTI [1972], which establishes that the exponential of a Lévy process is a time-changed self-similar Markov process. Obviously, a Brownian motion with drift is a particular case of a Lévy process. WILLIAMS [1974] showed that the exponential of Brownian motion with drift is a time-changed Bessel process. It is on this representation of a geometric Brownian motion as a time-changed squared Bessel process that GEMAN and YOR [1993] built their exact valuation of Asian options in the Black–Scholes setting; in contrast to geometric Brownian motion, the class of squared Bessel processes is stable by additivity, and this property is obviously quite valuable for the pricing of contingent claims related to the arithmetic average of a stock or commodity price.

Thirteen years after the Dubins–Schwarz theorem, MONROE [1978] extended the result to semimartingales and established that

“Any semi-martingale can be written as a time-changed Brownian motion.”

More formally, it is required that there exists a filtration $\mathcal{G}_u$ with respect to which the Brownian motion $W(u)$ is adapted and that $T(t)$ is an increasing process of stopping times adapted to this filtration $(\mathcal{G}_u)$.

2.2. Market activity and transaction clock

The first idea to use such a clock for analyzing financial processes was introduced by CLARK [1973]. Clark was analyzing cotton futures price data, and in order to address the nonnormality of observed returns, he wrote the price process as a subordinated process

$$S(t) = W(X(t)),$$

TABLE 2.1
Descriptive statistics of S&P 500 future prices at various time resolutions over
the period 1993–1997

	Mean	Variance	Skewness	Kurtosis
1 minute	1,77857E-6	8,21128E-08	1,109329038	58,59028989
15 minutes	2,50594E-05	1,1629E-06	−0,443443082	13,85515387
30 minutes	4,75712E-05	1,95136E-06	−0,092978342	6,653238051
Hour by hour	8,76093E-05	3,92529E-06	−0,135290879	5,97606376

where he conjectured that the process W had to be Brownian motion, and the economic interpretation of the subordinator X was the cumulative volume of traded contracts.

Note that subordination was introduced in harmonic analysis (and not in probability theory) by BOCHNER [1955] and that subordinators are restrictive time changes as they are increasing processes with independent increments.

ANÉ and GEMAN [2000] analyzed a high-frequency database of S&P future contracts over the period 1993–1997. They exhibited the increasing deviations from normality of realized returns over shorter time intervals.

Revisiting Clark's brilliant conjecture,¹ they demonstrated that the structure of semi-martingales necessarily prevails for stock prices S by bringing together the no-arbitrage assumption and Monroe's theorem to establish that any stock price return may be written as

$$\ln S(t) = W(T(t)),$$

where W is Brownian motion, and T is an almost surely increasing process.

They showed in a general nonparametric setting that in order to recover a quasiperfect normality of returns, the transaction clock is better represented by the number of trades than by the volume. JONES, KAUL, and LIPTON [1994] had shown that conditional on the number of trades, the volume was hardly an explanatory factor for the volatility.

Moreover, ANÉ and GEMAN [2000] showed that, under the assumption of independence of W and T , the above expression of $\ln S(t)$ leads to the representation of the return as a mixture of normal distributions, in line with the empirical evidence exhibited by HARRIS [1986].

Conducting the analysis of varying market activity in its most obvious form, a vast body of academic literature has examined trading/nontrading time effects on the risk and return characteristics of various securities; nontrading time refers to the periods in which the principal markets where the security is traded are closed (and the transaction clock goes at a very low pace). Trading time refers to the period in which a security is openly traded in a central market [e.g., New York stock Exchange (NYSE), Chicago Board of Trade (CBOT)] or an active over-the-counter market. These studies (DYL and MABERLY [1986]) first focused on differing returns/variances between weekdays and weekends. Subsequent studies (GEMAN and SCHNEEWEIS [1991]) also tested for intertemporal changes in asset risk as measured by return variance of overnight and daytime periods as

¹I am grateful to Joe Horowitz for bringing Monroe's paper to my attention during Summer 1997.

well as intraday time intervals. Results in both the cash (FRENCH and ROLL [1986]) and the futures markets (CORNELL [1983]) indicated greater return variance during trading time than during nontrading time. GEMAN and SCHNEEWEIS [1991] argued that “the non-stationarity in asset return variance should be discussed in the context of calendar time and transaction time hypotheses.” FRENCH and ROLL [1986] conducted an empirical analysis of the impact of information on the difference between trading and nontrading time stock return variance. They concluded that information accumulates more slowly when the NYSE and AMEX are closed, resulting in lower volatility in these markets during weekends and holidays. FRENCH, SCHWERT and STAMBAUGH [1987] and SAMUELSON [2001] showed that expected returns are higher for more volatile stocks since investors are rewarded for taking more risk. Hence, the validity of the semimartingale model discussed in the previous section for stock prices: the sum of a martingale and a trend process, which is unknown but assumed to be fairly smooth, continuous and locally of finite variation.

2.3. Stochastic volatility and information arrival

Financial markets go through hectic and calm periods. In hectic markets, the fluctuations in prices are large. In calm markets, price fluctuations tend to be moderate. The simplest representation of the size of fluctuations is volatility, the central quantity in financial markets. Financial time series exhibit, among other features, the property that periods of high volatility, or large fluctuations of the stock or commodity price, tend to cluster as shown in Figs. 2.1 and 2.2.

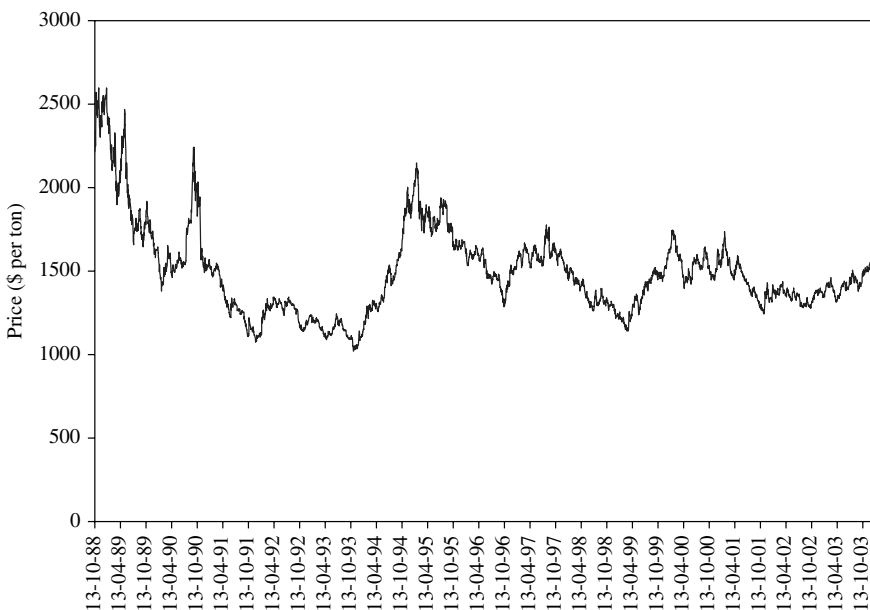


FIG. 2.1 Aluminum nearby future prices on the London Metal Exchange from Oct 1988–Dec 2003.

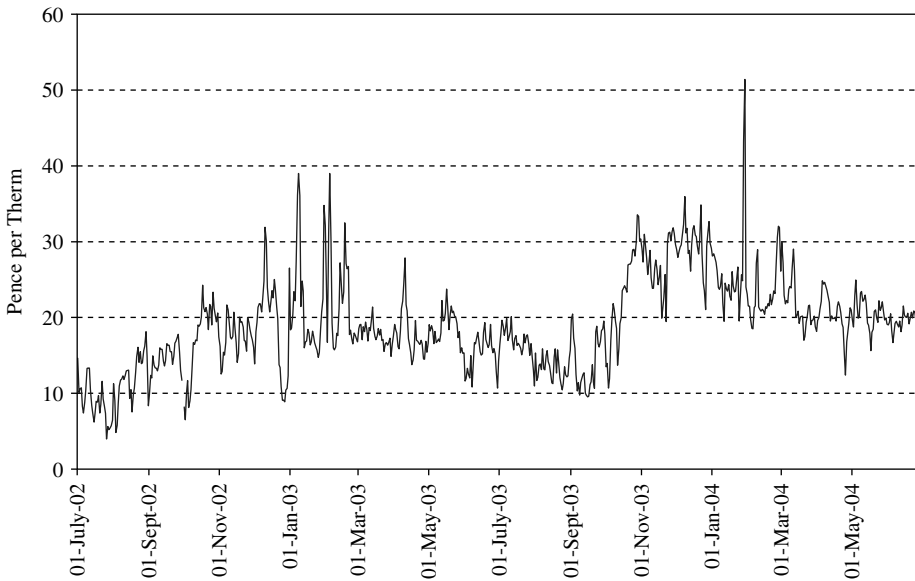


FIG. 2.2 UK National Balancing Point gas price.

MANDELBROT and TAYLOR [1967], CLARK [1973], KARPOFF [1987], SCHWERT, and FRENCH and STAMBAUGH [1987], RICHARDSON and SMITH [1994] have suggested to link asset return volatility to the flow of information arrival. This flow is not uniform through time and not always directly observable. Its most obvious components include quote arrivals, dividend announcements, macroeconomic data release, or markets closures. Fig. 2.3 shows the dramatic effect on the share price of the oil company Royal Dutch Shell due to the announcement in January 2004 of a large downward adjustment in the estimation of oil reserves owned by the company. At the same time, oil prices were sharply increasing under the combined effect of growth in demand and production uncertainties in major oil producing countries.

GEMAN and YOR [1993] proposed to model a nonconstant volatility by introducing a clock that measures financial time: the clock runs fast if trading is brisk and runs slowly if trading is dull. We can observe the property in a deterministic setting: by self-similarity of Brownian motion, an increase in the scale parameter σ may be interpreted as an increase in speed

$$(\sigma W(t), t \geq 0) \stackrel{\text{law}}{=} W(\sigma^2 t) t \geq 0, \quad \text{for any } \sigma > 0.$$

Hence, volatility appears as closely related to time change; doubling the volatility σ will speed up the financial time by a factor four. BICK [1995] revisited portfolio insurance strategies in the context of stochastic volatility. Instead of facing an unwanted outcome of the portfolio insurance strategy at the horizon H , he suggested to roll the strategy at

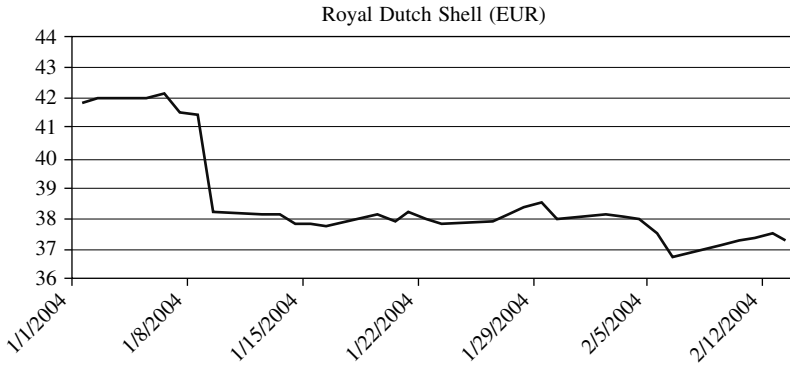


FIG. 2.3 Royal Dutch Shell price over the period January, 1–February 12, 2004.

the first time τ_b such that

$$\int_0^{\tau_b} \sigma^2(s) ds = b^2, \quad (2.1)$$

where $b(b > 0)$ is the volatility parameter chosen at inception of the portfolio strategy at date 0. Assuming that $\sigma(t)$ is continuous and $\mathcal{F}_t$ adapted, it is easy to show that the stopping time

$$\tau_b = \inf \left\{ u \geq 0 : \int_0^u \sigma^2(s) ds = b^2 \right\} \quad (2.2)$$

is the first instant at which the option replication is correct and hence the portfolio value is equal to the desired target.

GEMAN and YOR [1993] look at the distribution of the variable τ_b in the HULL and WHITE [1987] model of stochastic volatility where both the stock price and variance are driven by a geometric Brownian motion

$$\begin{cases} \frac{dS(t)}{S(t)} = \mu_1 dt + \sigma(t) dW^1(t), \\ \frac{dy(t)}{y(t)} = \mu_2 dt + \eta(t) dW^2(t), \end{cases}$$

where $d < W^1$, $W^2 >_t = \rho dt$, and $y(t) = [\sigma(t)]^2$.

The squared volatility following a geometric Brownian motion can be written as a squared Bessel process through a theorem LAMPERTI [1972].

Hence, the volatility itself may be written as

$$\sigma(t) = R_{\sigma_0}^{(v)} \left(\frac{4t}{\int_0^{y^2} \sigma^2(u) du} \right),$$

where $(R_{\sigma_0}^{(v)}(t))_{t \geq 0}$ is a Bessel process starting at σ_0 , with index

$$v = \frac{2\mu_2}{\eta^2} - 1.$$

In order to identify the stopping time τ_b , we need to invert Eq. (2.1), and this leads to

$$\tau_b = \int_0^b \frac{du}{[R_{\sigma_0}^{(v)}(u)]^2}.$$

The probability density f_b of τ_b does not have a simple expression, but its Laplace transform has an explicit expression (see YOR [1980]):

$$\int_0^{+\infty} f_b(x) e^{-\lambda x} dx = \frac{1}{\Gamma(k)} \int_0^1 \exp\left(\frac{-2u\sigma_0^2}{b\eta^2}\right) \left(\frac{2u\sigma_0^2}{b\eta^2}\right)^k (1-\mu)^{\mu-k} du,$$

where $\mu = \left(\frac{8\lambda}{\eta^2} + v^2\right)^{\frac{1}{2}}$ and $k = \frac{\mu-v}{2}$.

EYDELAND and GEMAN [1995] proposed an inversion of this Laplace transform. By linearity, the same method can be applied to obtain the expectation of τ_b , that is, the average time at which replication will be achieved and the roll of the portfolio insurance strategy takes place. Note that the same type of time-change technique can be applied to other models of stochastic volatility and the option trader can compare the different answers obtained for the distributions of the stopping time of exact replication.

2.4. Stochastic time and jump processes

GEMAN, MADAN and YOR [2001] (GMY) argued that asset price processes arising from market-clearing conditions should be modeled as pure jump processes, with no continuous martingale component. Since continuity and normality can always be obtained after a time change, they studied various examples of time changes and showed that in all cases, the time changes are related to measures of economic activity. For the most general class of processes, the time change is a size-weighted sum of order arrivals.

The possibility of discontinuities or jumps in asset prices has a long history in the economics literature. MERTON [1976] considered the addition of a jump component to the classical geometric Brownian motion model for the pricing of options on stocks. BATES [1996] and BAKSHI, CAO and CHEN [1997] proposed models that contain a diffusion component in addition to a low or finite-activity jump part. The diffusion component accounts for high activity in price fluctuations, while the jump component is used to account for rare and extreme movements. By contrast, GMY accounts for the small and high activities and rare large moves of the price process in a unified and connected manner: all motions occur via jumps. High activity is accounted for by a large (in fact infinite) number of small jumps. The various jump sizes are analytically connected by the requirement that smaller jumps occur at a higher rate than larger jumps. In this family of Lévy processes, the property of an infinite number of small moves is shared with the diffusion-based models, with the additional attractive feature that the sum of absolute

changes in price is finite, while for diffusions, this quantity is infinite (for diffusions, the price changes must be squared before they sum to a finite value). This makes possible the design and pricing of contracts such as droptions based on the instantaneous upward, downward, or total variability (positive, negative, or absolute price jump size) of underlying asset prices, in addition to the more traditional contracts with payoffs that are functionally related to the level of the underlying price. These processes include the α -stable processes (for $\alpha < 1$) that were studied by MANDELBROT [1963].

The empirical literature that has related price changes to measure of activity (see KARPOFF [1987], GALLANT, ROSSI and TAUCHEN [1992], JONES, KAUL, and LIPTON [1994]) has considered either the number of trades or the volume as relevant measures of activity. GEMAN, MADAN and YOR [2001] argued that time changes must be the processes with jumps that are locally uncertain, since they are related to demand and supply shocks in the market.

Writing $S(t) = W(T(t))$, we see that the continuity of $(S(t))$ is equivalent to the continuity of $(T(t))$. If the time change is continuous, it must essentially be a stochastic integral with respect to another Brownian motion (see REVUZ and YOR [1994]): denoting $T(t)$ as the time change, it must satisfy an equation of the type

$$dT(t) = a(t)dt + b(t)dB(t),$$

where $(B(t))$ is a Brownian motion. For the time change to be an almost surely increasing process, the diffusion component $b(t)$ must be zero. Then, the time change would be locally deterministic, which is in contradiction with its fundamental interpretation in terms of supply and demand order flow.

The equation

$$S(t) = W(T(t))$$

implies that the study of price processes for market economies may be reduced to the study of time changes for Brownian motion. We can note that this is a powerful reduced-form representation of a complex phenomenon involving multidimensional considerations, those of modeling supply, demand and their interaction through market clearing, to a single entity: the correct unit of time for the economy with respect to which we have a Brownian motion.

Hence, the investigations may focus on theoretically identifying and interpreting $T(t)$ from a knowledge of the process $S(t)$ through historical data. GMY defines a process as exhibiting a high level of activity if there is no interval of time in which prices are constant throughout the time interval.

An important structural property of Lévy densities attached to stock prices is that of monotonicity. One expects that jumps of larger sizes have lower arrival rates than jumps of smaller sizes. This property amounts to asserting for differentiable densities that the derivative is negative for positive jump sizes and positive for negative jump sizes.

We want to go further in that direction and introduce the property of complete monotonicity for the density. If we focus our attention on the density corresponding to positive jumps (this does not mean that we assume symmetry of the Lévy density), a completely monotone Lévy density on R^+ will be decreasing and convex, its derivative will be

increasing and concave and so on. Structural restrictions of this sort are useful in limiting the modeling set, given the wide class of choices that are otherwise available to model the Lévy density, which is basically any positive function that integrates the minimum of x^2 and 1. Complete monotonicity has the interesting property of linking analytically the arrival rate of large jumps to that of small ones by requiring the latter to be large than the former. The presence of such a feature makes it possible to learn about larger jumps from observing smaller ones. In this regard, we note that the by jump diffusion model MERTON [1976] is not completely monotone as the normal density shifts from being a concave function near zero to a convex function near infinity. On the other hand, the exponentially distributed jump size is the foundation for all completely monotone Lévy densities (accordingly, they have been largely used in insurance to model losses attached to weather events).

2.5. Stable processes as time-changed Brownian motion

For an increasing stable process of index $\alpha < 1$, the Lévy measure is

$$v(dx) = \frac{1}{x^{\alpha+1}} dx \quad \text{for } x > 0.$$

The difference $X(t)$ of two independent copies of such a process is the symmetric stable process of index α with characteristic function

$$E[\exp(iuX(t))] = \exp(-tc|u|^\alpha)$$

for a positive constant c .

If we compute the characteristic function of an independent Brownian motion evaluated at an independent increasing stable process of index α , we obtain

$$E[\exp(iuW(T(t)))] = E[\exp(-u^2T(t)/2)] = \exp(-t(c/2)|u|^{2\alpha})$$

or a symmetric stable process of index 2α .

It follows from the observation that the difference of two increasing stable α processes for $\alpha < 1$ is Brownian motion evaluated at an increasing stable $\alpha/2$ process.

2.6. The normal inverse gaussian process

Barndorff and Nielsen (1998) proposed the normal inverse Gaussian (NIG) distribution as a possible model for the stock price. This process may also be represented as a time-changed Brownian motion, where the time change $T(t)$ is the first passage time of another independent Brownian motion with drift to the level t . The time change is, therefore, an inverse Gaussian process, and when one evaluates a Brownian motion at this time, this suggests the nomenclature of a normal inverse Gaussian process.

We note that the inverse Gaussian process is a homogenous Lévy process that is, in fact, a stable process of index $\alpha = \frac{1}{2}$. We observed that if $2\alpha < 1$, time-changing Brownian motion with such a process leads to the symmetric stable process of index $\alpha < 1$. For $\alpha = \frac{1}{2}$, we show below that the process is of infinite variation. In general, for

$W(T(t))$ to be a process of bounded variation, we must have that $\int (1 \wedge |x|) \tilde{\nu}(dx) < \infty$, where $\tilde{\nu}$ is the Lévy measure of the time-changed Brownian motion.

Returning to the expression of the NIG process, it is defined as

$$X_{\text{NIG}}(t; \sigma, v, \theta) = \theta T_t^v + \sigma W(T_t^v),$$

where for any positive t , T_t^v is the first time a Brownian motion with drift v reaches the positive level t . The density of T_t^v is inverse Gaussian, and its Laplace transform has the simple expression

$$E[\exp(-\lambda T_t^v)] = \exp\left[-t(\sqrt{2\lambda + v^2} - v)\right].$$

This leads, in turn, to a fairly simple expression of the characteristic function of the NIG process in terms of the three parameters θ , v , and σ

$$E[e^{iuX_{\text{NIG}}(t)}] = E\left[\exp(iu\theta T_t^v - \sigma^2 \frac{u^2}{2} T_t^v)\right] = \exp\left[-t\sqrt{v^2 - 2iu\theta + \sigma^2 u^2} - v\right]$$

The NIG belongs to the family of hyperbolic distributions introduced by BARNDORFF-NIELSEN and HALGREEN [1977] who showed, in particular, that the hyperbolic distribution can be represented as a mixture of normal densities, where the mixing distribution is a generalized inverse Gaussian density. GEMAN [2002] emphasized that one of the merits of the expression of the stock price return as a time-changed Brownian motion

$$S(t) = W(T(t))$$

resides in the fact that it easily leads to the representation of the return as a mixture of normal distributions, where the mixing factor is conveyed by the time change, that is, by the market activity. Loosely stated, it means that one needs to mix enough normal distributions to account for the skewness, kurtosis, and other deviations from normality exhibited by stock returns, with a mixing process that is not necessarily continuous. The mixture of normal distribution hypothesis (MDH) has often been offered in the finance literature. RICHARDSON and SMITH [1994], outside any time change discussion, proposed to test it by measuring the daily flow of information, the information that precisely drives market activity and the stochastic clock!

2.7. The CGMY process with stochastic volatility

CARR, GEMAN, MADAN and YOR [2002] introduced a pure jump Lévy process to model stock prices, defined by its Lévy density

$$k_{\text{CGMY}}(x) = \begin{cases} \frac{C e^{-Mx}}{x^{1+Y}}, & x > 0 \\ \frac{C e^{-G|x|}}{|x|^{1+Y}}, & x < 0 \end{cases}$$

and showed that the parameter γ characterizes the activity intensity of the market to which the process is calibrated.

As any Lévy process, the CGMY process has independent increments that do not allow to capture effects such as volatility clustering that have been well documented in the finance literatures. In order to better calibrate the volatility surface, CARR, GEMAN, MADAN and YOR [2003] proposed to introduce in the CGMY model stochastic volatility in the form of a time change, leading to a return process

$$R(t) = X_{\text{CGMY}}(T(t)), \quad (2.3)$$

where the time change is meant to create autocorrelations of returns and clustering of volatility. Since the time change has to be increased, they chose for $T(t)$ the integral of a mean-reverting positive process, namely, the square-root process:

$$T(t) = \int_0^t y(u) du,$$

where

$$dy(t) = k(\eta - y)dt + \lambda\sqrt{y}dB(t).$$

The process described in (2.3) performs much better when calibrating S&P option prices through strikes and maturities (see CARR, GEMAN, MADAN and YOR [2003]).

To conclude this section, we should observe that the representation

$$R(t) = X(T(t)), \quad (2.4)$$

where X is not necessarily a Brownian motion

T , the time change, chosen to translate the desired properties of stochastic volatility may be quite powerful if the two processes X and T are fully known, in particular, in terms of trajectories. In order to price exotic options, one can build Monte Carlo simulations of the stock process and avoid the hurdles created by the unobservable nature of volatility in stochastic volatility models.

We have shown that by changing the probability measure (and the numéraire in the economy) or changing the clock, asset price processes can be expressed as martingales or even Brownian motion. The martingale representation is immediately extended to contingent claims in the case of complete markets where there is a unique martingale measure for each chosen numéraire. In the case of incomplete markets, we are facing many martingale measures; moreover, self-financing portfolios are not in general numéraire invariant, or in turn the pricing and hedging of contingent claims (see GOURIÉROUX and LAURENT [1998] for the case of the minimal variance measure). These elements, among others, illustrate the numerous difficulties attached to incomplete markets. Given the importance of the numéraire-invariance property, for instance, when managing a book of options involving several currencies, this feature may be a constraint one wishes to incorporate when choosing among the different answers to market incompleteness.

References

- ANÉ T. and GEMAN H. (2000) Order Flow, transaction clock and normality of asset returns, *Journal of Finance* **55**, 2259–2285.
- ARTZNER, P., DELBAEN, F. (1989). Term structure of interest rates: the martingale approach. *Adv. Appl. Math.* **10**, 95–129.
- BACHELIER, L. (1900). Théorie de la spéculation. *Ann. Sci. Ec. Norm. Supér.* **17**, 21–86.
- BAKSHI, G.S., CAO, C., CHEN, Z.W. (1997). Empirical performance of alternative option pricing models. *J. Finance* **52**, 2003–2049.
- BARNDORFF-NIELSEN, O.E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2**, 41–68.
- BARNDORFF-NIELSEN, O.E., HALGREEN, O. (1977). Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Z. für Wahrscheinlichkeitstheorie verw. Geb* **38**, 309–312.
- BATES, D. (1996). Jumps and stochastic volatility: exchange rate processes in deutschemark options. *Rev. Financ. Studies* **9**, 69–108.
- BICK, A. (1995). Quadratic-variation-based dynamic strategies. *Manage. Sci.* **41** (4), 722–732.
- BLACK, F., SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Econ.* **81**, 637–659.
- BOCHNER, S. (1955). *Harmonic Analysis and the Theory of Probability* (University of California Press, Berkeley, CA).
- CARR, P., GEMAN, H., MADAN, D., YOR, M. (2002). The fine structure of asset returns: an empirical investigation. *J. Bus.* **75**, 305–332.
- CARR, P., GEMAN, H., MADAN, D., YOR, M. (2003). Stochastic volatility for Lévy processes. *Math. Finance* **13**, 345–382.
- CARR, P., GEMAN, H., MADAN, D., YOR, M. (2005). Pricing options on realized variance. *Finance Stoch.* **4** (3), 453–478.
- CLARK, P. (1973). A subordinated stochastic process with finite variance for speculative prices. *Econometrica* **41**, 135–156.
- CORNELL, B. (1983). Money supply announcements and interest rates: another view. *J. Bus.* **56** (1), 1–23.
- DALANG, R.C., MORTON, A., WILLINGER, W. (1990). Equivalent martingale measures and no-arbitrage in stochastic securities market models. *Stochastics* **29** (2), 185–201.
- DAMBS, K.E. (1965). On the decomposition of continuous martingales. *Theor. Prob. Appl.* **10**, 401–410.
- DELBAEN, F., SCHACHERMAYER, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300**, 465–520.
- DUBINS, L., SCHWARZ, G. (1965). On continuous martingales. *Proc. Nat. Acad. Sci. USA* **53**, 913–916.
- DYL, E., MABERLY, E. (1986). The weekly pattern in stock index futures: a further note. *Journal of Finance* **1149**–1152.
- EYDELAND, A., GEMAN, H. (1995). Asian options revisited: inverting the laplace transform. *RISK*.
- FELLER, W. (1964). *An Introduction to Probability Theory and its Applications*, vol. 2 (Wiley, New York, NY).
- FRENCH, K., ROLL, R. (1986). Stock return variance: the arrival of information and the reaction of traders. *J. Financ. Econ.* **17**, 5–26.
- FRENCH, K., SCHWERT, G., STAMBAUGH, R. (1987). Expected stock returns and volatility. *J. Financ. Econ.* **19**, 3–29.
- GALLANT, A.R., ROSSI, P.E., TAUCHEN, G. (1992). Stock prices and volume. *Rev. Financ. Studies* **5**, 199–242.
- GEMAN, H. (1989). The Importance of the Forward Neutral Probability Measure for Stochastic Interest Rates. ESSEC Working Paper.

- GEMAN, H. (2002). Pure jump Lévy processes for asset price modelling. *J. Bank. Finance* 1297–1316.
- GEMAN, H., MADAN, D., YOR, M. (2001). Time changes for Lévy processes. *Math. Finance* 11, 79–96.
- GEMAN, H., SCHNEEWEIS, T. (1991). Trading time-non trading time effects on French futures markets. In: *Accounting and Financial Globalization* (Quorum, New York, NY).
- GEMAN, H., YOR, M. (1993). Bessel processes, Asian options and perpetuities. *Math. Finance* 4 (3), 349–375.
- GOURIÉROUX, C., LAURENT, J.P. (1998). Mean-variance hedging and numéraire. *Math. Finance* 8 (3), 179–200.
- HARRIS, L. (1986). Cross-security tests of the mixture of distributions hypothesis. *J. Finance Quant. Analysis* 21, 39–46.
- HARRISON, J.M., KREPS, D. (1979). Martingales and arbitrage in multiperiod securities market. *J. Econ. Theory* 20, 381–408.
- HARRISON, J.M., PLISKA, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stoch. Processes Appl.* 11, 381–408.
- HULL, J., WHITE, J. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* 42 (2), 281–300.
- ITÔ, K., MCKEAN, H.P. (1965). *Diffusion Processes and Their Sample Paths* (Springer-Verlag, Berlin, Germany).
- JONES, C., KAUL, G., LIPTON, M.L. (1994). Transactions, volume and volatility. *Rev. Financ. Studies* 7, 631–651.
- KARPOFF, J. (1987). The relation between price changes and trading volume: a survey. *J. Financ. Quant. Analysis* 22, 109–126.
- LAMPERTI, J. (1972). Semi-stable Markov processes. *Z. für Wahrscheinlichkeitstheorie Verw. Geb.* 205–255.
- MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Bus.* 36, 394–419.
- MANDELBROT, B., TAYLOR, H. (1967). On the distribution of stock prices differences. *Oper. Res.* 15, 1057–1062.
- MCKEAN, H. (2001). Scale and clock. In: Geman, H., Madan, D., Pliska, S., Vorst, T. (eds.), In: *The First World Bachelier Congress* (Springer-Finance).
- MERTON, R. (1973). The theory of rational option pricing. *Bell J. Econ. Manage. Sci.* 4, 141–183.
- MERTON, R. (1976). Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.* 3, 125–144.
- MONROE, I. (1978). Processes that can be embedded in Brownian motion. *Ann. Probabil.* 6, 42–56.
- REVUZ, D., YOR, M. (1994). *Continuous Martingales and Brownian Motion* (Springer-Verlag, Berlin, Germany).
- RICHARDSON, M., SMITH, T. (1994). A direct test in the mixture of distributions hypothesis: measuring the daily flow of information. *J. Financ. Quant. Analysis* 29, 101–116.
- SAMUELSON, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Ind. Manage. Rev.* 6 (Spring), 41–49.
- SAMUELSON, P. (1965). Rational theory of warrant pricing. *Ind. Manage. Rev.* 6 (Spring), 13–31.
- SAMUELSON, P. (2001). Finance theory in a lifetime. In: Geman, H., Madan, D., Pliska, S., Vorst, T. (eds.), In: *The First Bachelier World Congress* (Springer).
- SCHWERT, G.W., FRENCH, K.R., STAMBAUGH, R.F. (1987). Expected stock returns and volatility. *J. Financ. Econ.* 19, 3–29.
- VOLKONSKI, V.A. (1958). Random, substitution of time in strong Markov processes. *Teor. Veroyatnost* 3, 332–350.
- WILLIAMS, D. (1974). Path decomposition and continuous of local time for the dimensional diffusions. *Proc. Lond. Math. Soc.* 3 (28), 738–768.
- YOR, M. (1980). Loi de l'indice du lacet Brownien et distribution de Hartmann Watson. *Z. für Wahrscheinlichkeitstheorie Verw. Gebiete*, 53, 71–95.

Analytical Approximate Solutions to American Barrier and Lookback Option Values

Qiang Zhang

Department of Economics and Finance, City University of Hong Kong, 83 Tat chee avenue, Kowloon, Hong Kong

Tanya Taksar

Risk and Analytic Department, Open Link Financial, Uniondale, NY, 11553, USA

Abstract

In this chapter, we present analytical approximate solutions to the values of American barrier options and American lookback strike options. In barrier options, one specifies a barrier. Once the value of the underlying asset reaches the barrier, the “out” barrier option becomes worthless and the “in” barrier option becomes alive. Lookback options are path-dependent options whose payoff depends on the maximum or the minimum realized value of the underlying asset over the life of the option. Our theoretical predictions for the values of these American-style exotic options are in excellent agreement with the results obtained from direct numerical computations.

1. Introduction

Due to the fact that the trading of options written on stocks, currency exchanges, and commodities and commodity futures has grown dramatically over the last two decades in a variety of financial markets, option pricing has become a very active research area in finance. Most of the options traded in the United States and Canada are American style that can be exercised at any time up to and including the expiration date of the option. In contrast to American-style options, European-style options can only be exercised on their

expiration date. So far, analytical solutions to American options have not been found for vanilla options, the simplest form of American options. In the case of exotic options, additional complications are present. In this chapter, we present analytical approximate solutions to the values of American barrier options and American lookback options. The analytical approximation solution derived in this chapter can be used to price the values of American barrier calls and puts and American lookback strike options on stocks, commodities and commodity futures, and currency exchanges.

Barrier options are options with the feature that once the price of the underlying asset reaches a specified value known as the barrier, the options become worthless (an “out” barrier) or just become alive (an “in” barrier). Lookback options are the options for which the payoff at the date of exercise or maturity depends on the maximum value or the minimum value of the underlying asset over the life of the option. Both the barrier options and the lookback options are path-dependent options, which have more complicated structures than vanilla options. So far, analytical solutions to American barrier options and American lookback options have not been found.

The purpose of this chapter is to develop analytical approximate solutions to American barrier options and American lookback options. We show that the analytical approximate solutions derived in this chapter provide quite accurate predictions for the values of these types of American exotic options. Excellent agreements between the theoretical predictions and the numerical solutions obtained from a finite-difference method are demonstrated.

We follow the model developed by MERTON [1973], BLACK and SCHOLES [1973] and BLACK [1976] in which the value of an option is governed by the equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - q)S \frac{\partial V}{\partial S} - rV = 0, \quad (1.1)$$

with the payoff function given by

$$V(T, S) = \max(S - K, 0) \quad \text{and} \quad V(T, S) = \max(K - S, 0) \quad (1.2)$$

for a vanilla call option and a vanilla put option, respectively. Here S is the price of the underlying asset, V is the price of the option on the underlying asset, r is the short-term risk-free interest rate, and S and σ are the spot price and volatility of the underlying asset, respectively. K is the strike price, T is the expiration date of the option, and $T - t$ is the time left before the expiration of the option. For options on a dividend paying stock, q is the continuous dividend paying rate. For options on foreign currency exchange, q is the risk-free foreign interest rate. For options on a commodity, $(r - q)S$ is the cost of carrying the commodity. For options on commodity futures, $q = r$. Eq. (1.1) was studied by MERTON [1973] for options on commodities, by BLACK and SCHOLES [1973] for options on a nondividend paying stock (i.e., $q = 0$), and by BLACK [1976] for options on commodity futures (i.e., $q = r$). The underlying concepts of this approach can be found in these articles. General background on option pricing can be found in studies by HULL [1993] and WILMOTT, DEWYNNE and HOWISON [1993].

For American options, there may exist a critical price. Once the price of the underlying asset crosses that critical price, it is optimal to exercise the option rather than continue

holding it. The following conditions hold at the critical price:

$$V(t, S) = F(S) \quad \text{and} \quad \frac{\partial V(t, S)}{\partial S} = \frac{\partial F(S)}{\partial S} \quad \text{at} \quad S = S_c. \quad (1.3)$$

Here, $F(S)$ is the payoff function. For vanilla call and put options, $F(S)$ is given by (1.2). Since the critical price cannot be determined a priori, the condition (1.3) and the Black–Scholes Eq. (1.1), together with the final condition at the expiration date, form a free-boundary problem. A detailed explanation to the conditions for the early exercise of call and put options can be found in STOLL and WHALEY [1986].

MACMILLAN [1986] derived an approximate solution to American put options on a nondividend paying stock. BARONE-ADESI and WHALEY [1987] derived an approximate solution to American options on commodities and commodity futures. In both these studies, only American vanilla options were considered. In this chapter, we apply the method of quadratic approximation to American barrier and lookback options. We show that the analytical approximate solutions developed by BARONE-ADESI and WHALEY [1987] and MACMILLAN [1986] are the special cases of the analytical approximate solutions presented in this chapter. Other approaches for analytical approximations to American put options can be found in GESKE and JOHNSON [1984] and JOHNSON [1983].

In Section 2, we present the analytical approximate solutions to American barrier options for a general payoff function. In Section 3, we present the analytical approximate solutions to American lookback options. We also show that, as the maturity tend to infinity, our approximate solutions tend to the exact solutions for perpetual American type barrier and lookback options. In Section 4, we validate the approximate solutions by comparing their predictions with the numerical solutions to these options obtained from a finite-difference method.

2. Barrier options

Barrier options are exotic options. A barrier option has less value than the corresponding option without a barrier since it gives less rights to the option holder. There are two types of barrier options: “in” and “out”. An option with an “out” barrier at X becomes worthless once the value of the underlying asset reaches the “out” barrier X . An option with an “in” barrier at X cannot be exercised if the value of the underlying asset never reaches the barrier X . Once the price of the underlying asset reaches the “in” barrier X , the “in” option has the same value as an option without a barrier. It is obvious that a European “out” barrier option plus a European in barrier option (when the both have the same the barrier value X) is equivalent to a European option without a barrier:

$$\text{“out” barrier option} + \text{“in” barrier option} = \text{option without barrier.}$$

This relation holds for both European call and put options and for arbitrary payoff functions. We comment that this relation may not hold for American style barrier options since the American “out” and “in” barrier options may have different critical prices. See HUDSON [1991] and RUBINSTEIN and REINER [1991] for barrier options.

2.1. European “out” barrier options

Let $v(t, S)$ be a European option for a general payoff without barrier. It is easy to show that the corresponding European barrier option is given by

$$v_{\text{out}}(t, S) = v(t, S) - \left(\frac{S}{X}\right)^{1-\frac{2(r-q)}{\sigma^2}} v\left(t, \frac{X^2}{S}\right). \quad (2.1)$$

From the relationship between the European “in” and “out” barrier options, the second term on the right-hand side of (2.1) is the value of the European “in” barrier option, namely,

$$v_{\text{in}}(t, S) = \left(\frac{S}{X}\right)^{1-\frac{2(r-q)}{\sigma^2}} v\left(t, \frac{X^2}{S}\right).$$

For example, for a European barrier call option with $X < K$, $v(t, S)$ is the price of European vanilla call options:

$$v(t, S) = c(t, S) = S e^{-q(T-t)} N(d_1) - K e^{-r(T-t)} N(d_2). \quad (2.2)$$

For a European barrier put option with $X > K$, $v(t, S)$ is the price of European vanilla call options:

$$v(t, S) = p(t, S) = K e^{-r(T-t)} N(-d_2) - S e^{-q(T-t)} N(-d_1). \quad (2.3)$$

$N(\cdot)$ is the cumulative univariate normal distribution, d_1 and d_2 are given by

$$d_1 = \frac{\log(S/K) + (r - q + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T - t}. \quad (2.4)$$

2.2. American barrier “out” options

In this subsection, we derive the analytical approximate solutions to the values of American “out” barrier options. The value of American “out” barrier options are determined by the Black–Scholes Eq. (1.1), a payoff function $F(S)$, free boundary conditions (1.3) at the critical price, and the boundary condition at the “out” barrier X

$$V_{\text{out}}(t, X) = 0. \quad (2.5)$$

We assume that the payoff function $F(S)$ is a continuous function of S and that the payoff function vanishes at the “out” barrier X , that is, $F(X) = 0$.

We express the price of the American “out” barrier option $V_{\text{out}}(t, S)$ as a sum of the price of the corresponding European “out” barrier option $v_{\text{out}}(t, S)$ and an additional premium $a(t, S)$ for the privilege of early exercise:

$$V_{\text{out}}(t, S) = v_{\text{out}}(t, S) + a(t, S). \quad (2.6)$$

Then, $a(t, S)$ also satisfies the Black–Scholes equation:

$$\frac{\partial a}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 a}{\partial S^2} + (r - q)S \frac{\partial a}{\partial S} - ra = 0. \quad (2.7)$$

Let us define

$$\alpha = \frac{2r}{\sigma^2}, \quad h = 1 - e^{-r(T-t)}, \quad \beta = \frac{2(r-q)}{\sigma^2}, \quad a(t, S) = hg(h, S). \quad (2.8)$$

Then, (2.7) can be written as

$$hS^2 \frac{\partial^2 g}{\partial S^2} + h\beta S \frac{\partial g}{\partial S} - \alpha g - h(1-h)\alpha \frac{\partial g}{\partial h} = 0. \quad (2.9)$$

The terminology *quadratic approximation* refers to the approximation that sets the last term $h(1-h)\alpha \frac{\partial g}{\partial h}$ in (2.9) to zero BARONE-ADESI and WHALEY [1987], MACMILLAN [1986]. This approximation is motivated by the facts that when $T - t \rightarrow \infty$, the factor $1 - h$ tends to zero (assuming r is nonzero) and when $T \rightarrow t$, the factor h tends to zero. Therefore, under the quadratic approximation, one only needs to solve the ordinary differential equation

$$hS^2 \frac{\partial^2 g}{\partial S^2} + h\beta S \frac{\partial g}{\partial S} - \alpha g = 0. \quad (2.10)$$

Eq. (2.10) admits two independent solutions: S^{γ_1} and S^{γ_2} , where

$$\begin{aligned} \gamma_1 &= -\frac{\beta-1}{2} - \frac{1}{2} \left[(\beta-1)^2 + \frac{4\alpha}{h} \right]^{1/2} \quad \text{and} \\ \gamma_2 &= -\frac{\beta-1}{2} + \frac{1}{2} \left[(\beta-1)^2 + \frac{4\alpha}{h} \right]^{1/2}. \end{aligned} \quad (2.11)$$

Therefore, under the quadratic approximation, the American “out” barrier options can be expressed as

$$V_{\text{out}}(t, S) = v_{\text{out}}(t, S) + C_1 \left(\frac{S}{S_c} \right)^{\gamma_1} + C_2 \left(\frac{S}{S_c} \right)^{\gamma_2}. \quad (2.12)$$

Here, C_1 and C_2 are time-dependent constants and S_c is the critical price. From the boundary condition (2.5) at the “out” barrier $S = X$, it follows that

$$C_1 \left(\frac{X}{S_c} \right)^{\gamma_1} + C_2 \left(\frac{X}{S_c} \right)^{\gamma_2} = 0. \quad (2.13)$$

Here, we have used the fact that the European “out” barrier option also vanishes at the “out” barrier, that is, $v_{\text{out}}(t, X) = 0$. Furthermore, from the boundary conditions at the

free boundary $S = S_c$ given by (1.3), we have

$$\begin{aligned} v_{\text{out}}(t, S_c) + C_1 + C_2 &= F(S_c) \quad \text{and} \quad \frac{\partial}{\partial S} v_{\text{out}}(t, S) \Big|_{S=S_c} + \frac{C_1 \gamma_1}{S_c} + \frac{C_2 \gamma_2}{S_c} \\ &= \frac{\partial F(S)}{\partial S} \Big|_{S=S_c}. \end{aligned} \quad (2.14)$$

Solving C_1 and C_2 from (2.13) and the first equation in (2.14), we obtain

$$C_1 = \frac{F(S_c) - v_{\text{out}}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \left(\frac{S_c}{X}\right)^{\gamma_1} \quad \text{and} \quad C_2 = \frac{F(S_c) - v_{\text{out}}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_2} - \left(\frac{S_c}{X}\right)^{\gamma_1}} \left(\frac{S_c}{X}\right)^{\gamma_2}. \quad (2.15)$$

Finally, after substituting C_1 and C_2 from (2.15) into (2.12), we have an analytical expression for an approximate solution to American “out” barrier options (before crossing the barrier or reaching the critical price),

$$V_{\text{out}}(t, S) = v_{\text{out}}(t, S) + \frac{F(S_c) - v_{\text{out}}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \left[\left(\frac{S}{X}\right)^{\gamma_1} - \left(\frac{S}{X}\right)^{\gamma_2} \right]. \quad (2.16)$$

The critical price S_c in (2.16) is determined by the following algebraic equation resulting from substituting C_1 and C_2 from (2.15) into the second equation in (2.14):

$$\begin{aligned} \frac{\partial F(S)}{\partial S} \Big|_{S=S_c} &= \frac{\partial v_{\text{out}}(t, S)}{\partial S} \Big|_{S=S_c} \\ &+ \frac{F(S_c) - v_{\text{out}}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \frac{1}{S_c} \left[\gamma_1 \left(\frac{S_c}{X}\right)^{\gamma_1} - \gamma_2 \left(\frac{S_c}{X}\right)^{\gamma_2} \right]. \end{aligned} \quad (2.17)$$

With the formula given by (2.16), we can determine the values of the American “out” barrier options. The procedure is to solve the critical price from (2.17) first, and then substitute the result into (2.16) to determine the value of American barrier “out” options.

A down-and-out call option is a barrier call option with $F(S) = \max(0, S - K)$ and $X < K$ and a up-and-out put option is a barrier put option with $F(S) = \max(0, K - S)$ and $K < X$. From (2.1), the European down-and-out call before crossing the barrier is given by

$$c_{\text{d.o.}}(t, S) = c(t, S) - \left(\frac{S}{X}\right)^{1 - \frac{2(r-q)}{\sigma^2}} c\left(t, \frac{X^2}{S}\right), \quad X < K, \quad (2.18)$$

and the European up-and-out put options before crossing the barrier is given by

$$p_{\text{u.o.}}(t, S) = p(t, S) - \left(\frac{S}{X}\right)^{1 - \frac{2(r-q)}{\sigma^2}} p\left(t, \frac{X^2}{S}\right), \quad K < X. \quad (2.19)$$

Then, from (2.16) and (2.17), the value of an American down-and-out call with $X < K$ is given by

$$C_{d.o.}(t, S) = \begin{cases} 0 & 0 \leq S \leq S_c \\ c_{d.o.}(t, S) + \frac{S_c - K - c_{out}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \left[\left(\frac{S}{X}\right)^{\gamma_1} - \left(\frac{S}{X}\right)^{\gamma_2} \right] & X \leq S < S_c \\ S - K & S_c \leq S, \end{cases} \quad (2.20)$$

with the critical price S_c determined from the algebraic equation

$$1 = \frac{\partial c_{d.o.}(t, S)}{\partial S} \Big|_{S=S_c} + \frac{S_c - K - c_{d.o.}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \frac{1}{S_c} \left[\gamma_1 \left(\frac{S_c}{X}\right)^{\gamma_1} - \gamma_2 \left(\frac{S_c}{X}\right)^{\gamma_2} \right]. \quad (2.21)$$

The value of an American up-and-out put with $K < X$ is given by

$$P_{u.o.}(t, S) = \begin{cases} K - S & 0 \leq S \leq S_c \\ p_{u.o.}(t, S) + \frac{K - S_c - p_{u.o.}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \left[\left(\frac{S}{X}\right)^{\gamma_1} - \left(\frac{S}{X}\right)^{\gamma_2} \right] & S_c < S < X \\ 0 & X \leq S, \end{cases} \quad (2.22)$$

with the critical price determined from the algebraic equation

$$-1 = \frac{\partial p_{u.o.}(t, S)}{\partial S} \Big|_{S=S_c} + \frac{K - S_c - p_{u.o.}(t, S_c)}{\left(\frac{S_c}{X}\right)^{\gamma_1} - \left(\frac{S_c}{X}\right)^{\gamma_2}} \frac{1}{S_c} \left[\gamma_1 \left(\frac{S_c}{X}\right)^{\gamma_1} - \gamma_2 \left(\frac{S_c}{X}\right)^{\gamma_2} \right]. \quad (2.23)$$

2.3. American “in” barrier options

In this subsection, we derive the analytical approximate solutions for the values of the American “in” barrier options. The value of the American “in” barrier options is determined by the Black–Scholes Eq. (1.1), and the boundary conditions at the barrier X is

$$V_{in}(t, X) = V(t, X). \quad (2.24)$$

Here, $V(t, X)$ is the solution to the corresponding American option without a barrier. The payoff on maturity is zero ($V(S, T) = 0$) since without crossing the barrier the option can never come alive.

For American “in” barrier options, once the price of the underlying asset reaches the barrier, the barrier no longer exists, and the value of the option will be the same as the value of an American option without the barrier. Therefore, we only need to consider the value of American “in” barrier options before the price of the underlying asset reaches

the in barrier. However, before S reaches the “in” barrier, the option has not yet become “alive” in the sense that one cannot exercise the option. Therefore, American “in” barrier option does not have an explicit free boundary. The free boundary problem only occurs implicitly as a boundary condition at the barrier $S = X$, that is, $V_{\text{in}}(t, X) = V(t, X)$. This due to the fact that in order to determine $V(t, X)$, one may need to solve a free boundary value problem in $V(t, S)$.

Although American “in” barrier options do not have an explicit free boundary, the value of the American “in” barrier option is at least as high as the value of the European in barrier option since the boundary value at the “in” barrier of an American option is at least as high as the boundary value at the “in” barrier of the European option.

The solutions under the quadratic approximation to American vanilla options without barrier are determined by BARONE-ADESI and WHALEY [1987] and MACMILLAN [1986]. The results for American vanilla call options, BARONE-ADESI and WHALEY [1987], are given by

$$C(t, S) = \begin{cases} c(t, S) + \frac{1}{\gamma_2} \left[1 - \frac{\partial}{\partial S} c(t, S_c) \right] S_c (S/S_c)^{\gamma_2} & S < S_c \\ S - K & S \geq S_c, \end{cases} \quad (2.25)$$

with the critical price determined from

$$c(t, S_c) + \frac{1}{\gamma_2} \left[1 - \frac{\partial}{\partial S} c(t, S_c) \right] S_c = S_c - K, \quad (2.26)$$

and the results for American vanilla put options are given by BARONE-ADESI and WHALEY [1987], MACMILLAN [1986].

$$P(t, S) = \begin{cases} p(t, S) - \frac{1}{\gamma_1} \left[1 + \frac{\partial}{\partial S} p(t, S_c) \right] S_c (S/S_c)^{\gamma_1} & S > S_c \\ S - K & S \leq S_c, \end{cases} \quad (2.27)$$

with the critical price determined from

$$p(t, S_c) - \frac{1}{\gamma_1} \left[1 + \frac{\partial}{\partial S} p(t, S_c) \right] S_c = K - S_c. \quad (2.28)$$

Here, $c(t, \cdot)$ and $p(t, \cdot)$ are given by (2.2) and (2.3), respectively.

Of course, in the case of call options, we assume $q \neq 0$. Otherwise, the value of an American call option is the same as the value of a European call option.

We now show that the analytical approximate solution given in this chapter for American barrier options is in a class larger than that derived by BARONE-ADESI and WHALEY [1987] and MACMILLAN [1986] for American vanilla options. In other words, the approximate solutions derived by BARONE-ADESI and WHALEY [1987] and MACMILLAN [1986] are special cases of the approximate solutions given in this chapter. We examine the call options first. It is easy to see that in the limit $X \rightarrow 0$, the European down-and-out call option (2.18) approaches the European vanilla call option (2.2). From the condition $\gamma_1 \leq 0 \leq \gamma_2$ [see (2.11)] and the condition $X < S$ for down-and-out call options, it is easy to check that in the limit $X \rightarrow 0$, (2.20) approaches (2.25) and (2.21) approaches (2.26).

For American “out” barrier put options, in the limit $X \rightarrow \infty$, (2.22) approaches (2.27) and (2.23) approaches (2.28). Therefore, the approximate solution of American up-and-out put options to the approximate solution of American vanilla put options in the limit $X \rightarrow \infty$.

3. Lookback options

A path-dependent option is an option for which the payoff at the date of exercise or maturity depends on the history of the price of the underlying asset. Path-dependent options have more complicated payoff structure than vanilla options. Lookback options are one of the most common types of path-dependent options. We assume that the lookback options are based on continuous sampling of the price of the underlying asset.

3.1. European lookback options

The analytical expression for the value of European lookback strike call options with a payoff $c_{lb}(T, S, S_{\min}) = S - S_{\min}$ is given by

$$\begin{aligned} c_{lb}(S, S_{\min}, t) = & -S_{\min}e^{-r(T-t)} \left[N(a_2) - \frac{1}{\beta}(S/S_{\min})^{1-\beta} N(-a_3) \right] \\ & + Se^{-q(T-t)} \left[N(a_1) - \frac{1}{\beta} N(-a_1) \right], \end{aligned} \quad (3.1)$$

where $a_1 = [\log(\frac{S}{S_{\min}}) + (r - q + \frac{\sigma^2}{2})(T - t)]/(\sigma\sqrt{T - t})$, $a_2 = a_1 - \sigma\sqrt{T - t}$, $a_3 = a_1 - 2(r - q)/(\sigma\sqrt{T - t})$, and $\beta = \frac{2(r-q)}{\sigma^2}$. The analytical expression for the value of European lookback strike put options with payoff $p_{lb}(T, S, S_{\max}) = S_{\max} - S$ is given by

$$\begin{aligned} p_{lb}(S, S_{\max}, t) = & S_{\max}e^{-r(T-t)} \left[N(b_1) - \frac{1}{\beta}(S_{\max}/S)^{\beta-1} N(-b_3) \right] \\ & - Se^{-q(T-t)} \left[N(b_2) - \frac{1}{\beta} N(-b_2) \right], \end{aligned} \quad (3.2)$$

where $b_1 = [\log(\frac{S_{\max}}{S}) + (r - q + \frac{\sigma^2}{2})(T - t)]/(\sigma\sqrt{T - t})$, $b_2 = b_1 - \sigma\sqrt{T - t}$, and $b_3 = b_1 - 2(r - q)/(\sigma\sqrt{T - t})$. These analytical solutions to European lookback options can be found in the studies by BABBS [1986, 1992], GARMAN [1989], GOLDMAN, SOSIN and GATTO [1979], HULL [1993], WILMOTT, DEWYNNE and HOWISON [1993].

When $r = q$, as in the case of lookback options on commodity futures, one needs to determine the limit $\beta \rightarrow 0$ of the expressions (3.1) and (3.2). It is easy to check that in

the limit $\beta \rightarrow 0$, that is, $r \rightarrow q$, the European lookback call option (3.1) becomes

$$c_{lb}(S, S_{\min}, t) = Se^{-r(T-t)} \left[N(a_1) + \frac{\sigma^2}{2}(T-t)N(-a_1) + \log(S_{\min}/S)N(-a_3) \right. \\ \left. + \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}e^{-a_1^2/2} \right] - S_{\min}e^{-r(T-t)}N(a_2), \quad (3.3)$$

and the European lookback put option (3.2) becomes

$$p_{lb}(S, S_{\max}, t) = -Se^{-r(T-t)} \left[N(b_2) + \log(S_{\max}/S)N(-b_3) - \frac{\sigma^2}{2}(T-t)N(-b_2) \right. \\ \left. - \frac{\sigma\sqrt{T-t}}{\sqrt{2\pi}}e^{-b_3^2/2} \right] + S_{\max}e^{-r(T-t)}N(b_1). \quad (3.4)$$

The expressions for a_1, a_2, a_3, b_1, b_2 , and b_3 here are the same as before but with $r = q$.

In the next subsection, we will derive analytical approximate solutions to American lookback strike call and put options.

3.2. American lookback options

We consider American lookback strike call options first. The value of an American lookback call option based on continuous sampling is determined by the Black–Scholes Eq. (1.1), the payoff function

$$C_{lb}(T, S, S_{\min}) = S - S_{\min}, \quad (3.5)$$

the free-boundary conditions at the critical price S_c

$$C_{lb}(t, S_c, S_{\min}) = S_c - S_{\min}, \quad \frac{\partial}{\partial S} C_{lb}(t, S_c, S_{\min}) = 1, \quad (3.6)$$

$$\frac{\partial}{\partial S_{\min}} C_{lb}(t, S_c, S_{\min}) = -1, \quad (3.7)$$

and the boundary condition at $S_{\min}$ due to continuous sampling

$$\frac{\partial}{\partial S_{\min}} C_{lb}(T, S, S_{\min}) = 0 \quad \text{at} \quad S = S_{\min}. \quad (3.8)$$

Eq. (3.8) comes from the fact that the value of the lookback option for continuous sampling is insensitive to a small change in the current extreme value $S_{\min}$ when it reaches a new low, namely, $S - S_{\min}$. This is because the probability of the current new low $S_{\min}$ still remaining as the historical low on the expiration date is zero. It is easy to see from the payoff function given by (3.5) that the American lookback strike call option should have the functional form

$$C_{lb}(t, S, S_{\min}) = S_{\min} F_{\text{call}} \left(t, \frac{S}{S_{\min}} \right). \quad (3.9)$$

By direct evaluation of the differential operators in (3.6) and (3.7) with the C_{lb} given by (3.9), it is straightforward to show that once (3.6) is satisfied, (3.7) is automatically satisfied. Therefore, we can drop (3.7) from our system.

Applying the quadratic approximation, we obtain the following general form for the approximate solution of American lookback rate call option:

$$C_{lb}(t, S, S_{\min}) = c_{lb}(t, S, S_{\min}) + A_1 S_{\min} \left(\frac{S}{S_{\min}} \right)^{\gamma_1} + A_2 S_{\min} \left(\frac{S}{S_{\min}} \right)^{\gamma_2}, \quad (3.10)$$

where γ_1 and γ_2 are given by (2.11). Note that (3.10) satisfies the functional form for the American lookback call option given by (3.9). Here, A_1 and A_2 depend only on t and the ratio $S_c/S_{\min}$. We comment that since S_c is proportional to $S_{\min}$, A_1 and A_2 do not depend on $S_{\min}$ alone. We now apply boundary conditions to determine A_1 , A_2 , and S_c .

From the condition (3.8), it follows that

$$(1 - \gamma_1)A_1 + (1 - \gamma_2)A_2 = 0. \quad (3.11)$$

Here, we have used the fact that the derivative of the European lookback strike call with respect to $S_{\min}$ vanishes at $S = S_{\min}$. Furthermore, from the conditions at the free boundary $S = S_c$ given by (3.6), we have

$$c_{lb}(t, S_c, S_{\min}) + A_1 S_{\min} \left(\frac{S_c}{S_{\min}} \right)^{\gamma_1} + A_2 S_{\min} \left(\frac{S_c}{S_{\min}} \right)^{\gamma_2} = S_c - S_{\min}, \quad (3.12)$$

$$\frac{\partial}{\partial S} c_{lb}(t, S_c, S_{\min}) + A_1 \gamma_1 \left(\frac{S_c}{S_{\min}} \right)^{\gamma_1-1} + A_2 \gamma_2 \left(\frac{S_c}{S_{\min}} \right)^{\gamma_2-1} = 1. \quad (3.13)$$

By solving A_1 and A_2 from (3.11) and (3.12), we obtain

$$A_1 = \frac{S_c - S_{\min} - c_{lb}(t, S_c, S_{\min})}{S_{\min} \left[\left(\frac{S_c}{S_{\min}} \right)^{\gamma_1-\gamma_2} - \frac{\gamma_1-1}{\gamma_2-1} \right]} \left(\frac{S_c}{S_{\min}} \right)^{-\gamma_2} \quad \text{and} \quad A_2 = -\frac{1-\gamma_2}{1-\gamma_1} A_1. \quad (3.14)$$

Finally, after substituting A_1 and A_2 from (3.14) into (3.10), we have an analytical expression for an approximate solution to the American lookback strike call option,

$$C_{lb}(t, S, S_{\min}) = \begin{cases} c_{lb}(t, S, S_{\min}) + \frac{S_c - S_{\min} - c_{lb}(t, S_c, S_{\min})}{\left(\frac{S_c}{S_{\min}} \right)^{\gamma_1-\gamma_2} - \frac{\gamma_1-1}{\gamma_2-1}} \left[\left(\frac{S}{S_{\min}} \right)^{\gamma_1-\gamma_2} - \frac{\gamma_1-1}{\gamma_2-1} \right] \left(\frac{S}{S_c} \right)^{\gamma_2} & S < S_c \\ S - S_{\min} & S \geq S_c. \end{cases} \quad (3.15)$$

The critical price in (3.15) is determined by the following algebraic equation resulting from substituting A_1 and A_2 from (3.14) into (3.13):

$$\frac{\partial}{\partial S} c_{lb}(t, S_c, S_{\min}) + \frac{S_c - S_{\min} - c_{lb}(t, S_c, S_{\min})}{S_c \left[\left(\frac{S_c}{S_{\min}} \right)^{\gamma_1-\gamma_2} - \frac{\gamma_1-1}{\gamma_2-1} \right]} \left[\gamma_1 \left(\frac{S_c}{S_{\min}} \right)^{\gamma_1-\gamma_2} - \frac{\gamma_1-1}{\gamma_2-1} \gamma_2 \right] = 1. \quad (3.16)$$

Now, we consider American lookback strike put options. The value of an American lookback strike put option is determined by the Black–Scholes Eq. (1.1); the payoff function

$$P_{lb}(T, S, S_{\max}) = S_{\max} - S, \quad (3.17)$$

the conditions at free boundary S_c ,

$$P_{lb}(t, S_c, S_{\max}) = S_{\max} - S_c, \quad \frac{\partial}{\partial S} P_{lb}(t, S_c, S_{\max}) = -1, \quad (3.18)$$

$$\frac{\partial}{\partial S_{\max}} P_{lb}(t, S_c, S_{\max}) = 1, \quad (3.19)$$

and the condition at $S_{\max}$,

$$\frac{\partial}{\partial S_{\max}} P_{lb}(T, S, S_{\max}) = 0 \quad \text{at} \quad S = S_{\max}. \quad (3.20)$$

The condition (3.20) comes from the property that, due to continuous sampling, P_{lb} is insensitive to a small change in the current extreme value $S_{\max}$. For the payoff function given by (3.17), the American lookback strike put option should have the functional form

$$P_{lb}(t, S, S_{\max}) = S_{\max} F_{\text{put}}\left(t, \frac{S}{S_{\max}}\right). \quad (3.21)$$

The derivation for the approximate solution to the American lookback strike put option is similar to the one given for American lookback strike call option. Following the same procedure, we have the following analytical expression for an approximate solution to American lookback strike put option:

$$P_{lb}(t, S, S_{\max}) = \begin{cases} p_{lb}(t, S, S_{\max}) - \frac{S_c - S_{\max} + p_{lb}(t, S_c, S_{\max})}{\left(\frac{S_c}{S_{\max}}\right)^{\gamma_1 - \gamma_2} - \frac{\gamma_1 - 1}{\gamma_2 - 1}} \left[\left(\frac{S}{S_{\max}}\right)^{\gamma_1 - \gamma_2} - \frac{\gamma_1 - 1}{\gamma_2 - 1} \right] \left(\frac{S}{S_c}\right)^{\gamma_2} & S < S_c \\ S_{\max} - S & S \geq S_c. \end{cases} \quad (3.22)$$

The critical price in (3.22) is determined by the following algebraic equation:

$$\begin{aligned} \frac{\partial}{\partial S} P_{lb}(t, S_c, S_{\max}) - \frac{S_c - S_{\max} + p_{lb}(t, S_c, S_{\max})}{S_c \left[\left(\frac{S_c}{S_{\max}}\right)^{\gamma_1 - \gamma_2} - \frac{\gamma_1 - 1}{\gamma_2 - 1} \right]} \\ \times \left[\gamma_1 \left(\frac{S_c}{S_{\max}}\right)^{\gamma_1 - \gamma_2} - \frac{\gamma_1 - 1}{\gamma_2 - 1} \gamma_2 \right] = -1. \end{aligned} \quad (3.23)$$

Perpetual options have no expiration date, that is, $T \rightarrow \infty$. In this case, the term that we neglected under the quadratic approximation is identically zero for all time. Therefore, by

taking the limit $T \rightarrow \infty$ in our approximate analytical solutions, derived under quadratic approximation, to American barrier and lookback options over finite time horizon, we obtain the exact analytical solutions to perpetual barrier and lookback options. The procedure of taking limit is straightforward, and resulting formulae for the exact solutions to perpetual exotic options are similar to ones for the finite-time approximate solutions, but the expressions become much simpler.

4. Validation of analytical approximate solutions

In this section, we present the results of the validation study by comparing the predictions of the analytical approximate solutions derived in Sections 2 and 3 with the numerical solutions obtained by a finite-difference method.

The numerical solutions for the values of the American barrier and lookback options are determined in the following way.

For the American barrier options, we take the Black–Scholes Eq. (1.1), the boundary condition at the barrier (2.5), and the free-boundary conditions at the critical price (1.3). For American barrier call options, we take the payoff function $\max\{0, S - K\}$, while for American barrier put options, we take the payoff function $\max\{0, K - S\}$. The numerical solutions to the American lookback options are determined by solving (1.1), together with (3.5)–(3.7) for call options and (3.17)–(3.19) for put options.

Now, we outline the procedure for solving these equations. We apply the transform $S = Je^x$, $V = JU$, and $\tau = T - t$ to obtain a parabolic equation with constant coefficients. Here, $J = K$ for barrier options and $J = S_{\max}$ or $S_{\min}$ for lookback options. To the resulting constant coefficient parabolic equation, we apply the forward-time, central-space, finite-difference approximation for the temporal variable τ and the spatial variable x . The barrier appears naturally as a boundary condition in the finite-difference method. It occupies one of the end grid points at every time level. In the case of lookback options, we apply the one-side finite-difference scheme to the term containing the first-order derivative with respect to x in boundary conditions at minimum or maximum stock price.

The values of the American options are initialized by the payoff function at $\tau = 0$, that is, at the expiration date $t = T$ and propagate forward in τ , that is, backward in t , using the following three steps.

- Step 1: update the value of the American option by the solution to the discretized Black–Scholes equations based on the backward-time, central-space, finite-difference scheme.
- Step 2: determine the critical price that is defined as the intersection point of the payoff function and the piecewise linear interpolation function of U at adjacent grid points of x .
- Step 3: further, update the value of u by the value of the payoff function at the grid points where the value of U is lower than the payoff function.

We carry out these three steps for each time step. Finally, we express the results in terms of original variables.

The backward-time, central-space method is one of the most commonly used finite-difference schemes for solving partial differential equations. The implementation of this scheme is standard STRICKWERDA [1989]. It is well known that this scheme is implicit and unconditionally stable. It has a local truncation error of order $(\Delta x)^3$ and a global error of order $(\Delta x)^2$. Here, Δx is the spatial grid spacing. The temporal grid spacing $\Delta \tau$ is proportional to $(\Delta x)^2$. These theoretical properties of the scheme can be found in a study by STRICKWERDA [1989]. For barrier options, we used the spatial grid spacing $\Delta x = 5 \times 10^{-3}$ and the temporal grid spacing $\Delta \tau = 2.5 \times 10^{-5}$. For lookback options, we used the spatial grid spacing $\Delta x = 1.25 \times 10^{-3}$ and the temporal grid spacing $\Delta \tau = 6.25 \times 10^{-6}$. Therefore, the error in numerical solutions is in the order of 10^{-5} . This is sufficient for the purpose of our validation study. We have also run the computation with finer grids, and the results remain the same. Before computing American-style exotic options, we used the method to compute the corresponding European-style exotic options and checked that the results indeed agreed with the analytical solutions for the European-style exotic options.

Binomial and trinomial tree methods are alternative approaches in numerical computations. Studies based on these approaches can be found in BOYLE and LAU [1994] and RITCHEN [1995] for barrier options and in BABBS [1986, 1992], CHEUK and VORST [1997] and KAT [1995] for lookback options.

We have compared the predictions of our analytical approximate solutions for the values of the American barrier options and American lookback strike options with the results from numerical computations. The tests have been performed for options on commodities ($r - q \neq 0$), options on commodity futures ($r - q = 0$), and put options on nondividend paying stocks ($q = 0$). We found that the theoretical predictions are in excellent agreement with the results from the computations.

We now show the representative results of these validation studies. In Tables 4.1–4.4, we compare the theoretical predictions with the results from numerical computations for the values of American “out” barrier call options, American “out” barrier put options, American lookback strike call options, and American lookback strike put options, respectively. The predictions of the corresponding European options are also presented in each table. The values of r , q , σ , and $T - t$ are listed in first column of each table. As shown in these tables, our theoretical predictions are in excellent agreement with the results from numerical computations. We comment that the evaluations of our theoretical predictions take no time because we only need to find the root of an algebraic equation.

In summary, we have developed analytical approximate solutions for American barrier options and American lookback options. Both the types of options are path-dependent options and have complicated structures. We have provided the theoretical predictions for the values of these types of American options and shown that the predictions are in excellent agreement with the results from numerical computations. The results presented in this chapter will be important for both theoretical considerations and practical applications.

The quadratical approximation is a theoretical approach for solving problems involving free boundaries. Previously, this approach had only been applied to American vanilla options. Under quadratical approximation, Black–Scholes equation has two fundamental

TABLE 4.1

Comparison between the predictions of our analytical approximate solution (labeled as Quad.) and the results from numerical computations (labeled as Num.) for American “out” call options. Here, the strike price is $K = 100$ and the “out” barrier is set at $X = 70$. The predictions of the European “out” barrier call options (labeled as Euro.) are also shown

$r - q = 0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	0.052	0.849	4.441	11.662	20.898	30.698	40.588
$\sigma = 0.2$	Num.	0.000	0.052	0.850	4.442	11.663	20.903	30.701	40.590
$T - t = 0.25$	Euro.	0.000	0.052	0.849	4.441	11.662	20.898	30.698	40.588
$r - q = 0.04$	S	70	80	90	100	110	120	130	140
$r = 0.12$	Quad.	0.000	0.052	0.841	4.397	11.547	20.694	30.404	40.218
$\sigma = 0.2$	Num.	0.000	0.052	0.841	4.397	11.546	20.694	30.402	40.217
$T - t = 0.25$	Euro.	0.000	0.052	0.841	4.396	11.546	20.690	30.392	40.184
$r - q = 0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	1.255	3.819	8.350	14.797	22.716	31.584	40.987
$\sigma = 0.4$	Num.	0.000	1.255	3.819	8.351	14.798	22.717	31.589	40.987
$T - t = 0.25$	Euro.	0.000	1.255	3.819	8.349	14.796	22.714	31.579	40.979
$r - q = 0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	0.414	2.180	6.496	13.425	22.060	31.483	41.185
$\sigma = 0.2$	Num.	0.000	0.414	2.181	6.501	13.429	22.063	31.484	41.182
$T - t = 0.50$	Euro.	0.000	0.414	2.180	6.496	13.424	22.059	31.481	41.179
$r - q = -0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	0.032	0.590	3.525	10.315	20.000	30.000	40.000
$\sigma = 0.2$	Num.	0.000	0.031	0.580	3.523	10.355	20.000	30.000	40.000
$T - t = 0.25$	Euro.	0.000	0.029	0.570	3.421	9.847	18.618	28.159	37.884
$r - q = -0.04$	S	70	80	90	100	110	120	130	140
$r = 0.12$	Quad.	0.000	0.032	0.587	3.507	10.289	20.000	30.000	40.000
$\sigma = 0.2$	Num.	0.000	0.031	0.575	3.500	10.325	20.000	30.000	40.000
$T - t = 0.25$	Euro.	0.000	0.029	0.564	3.387	9.749	18.433	27.879	37.468
$r - q = -0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	1.035	3.279	7.410	13.502	21.233	30.182	40.000
$\sigma = 0.4$	Num.	0.000	1.032	3.265	7.405	13.525	21.293	30.247	40.000
$T - t = 0.25$	Euro.	0.000	1.018	3.229	7.291	13.248	20.728	29.232	38.337
$r - q = -0.04$	S	70	80	90	100	110	120	130	140
$r = 0.08$	Quad.	0.000	0.227	1.387	4.724	10.995	20.000	30.000	40.000
$\sigma = 0.2$	Num.	0.000	0.220	1.359	4.709	10.997	20.000	30.000	40.000
$T - t = 0.50$	Euro.	0.000	0.209	1.312	4.465	10.163	17.851	26.621	35.838

TABLE 4.2

Comparison between the predictions of our analytical approximate solution (labeled as Quad.) and the results from numerical computations (labeled as Num.) for American “out” put options. Here, the strike price is $K = 100$ and the “out” barrier is set at $X = 130$. The predictions of the European “out” barrier call options (labeled as Euro.) are also shown

$r - q = 0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.000	30.000	20.000	10.183	3.544	0.798	0.117	0.000
$\sigma = 0.2$	Num.	40.000	30.000	20.000	10.211	3.542	0.791	0.115	0.000
$T - t = 0.25$	Euro.	38.617	28.717	18.868	9.765	3.455	0.777	0.112	0.000
$r - q = 0.04$	S	60	70	80	90	100	110	120	130
$r = 0.12$	Quad.	40.000	30.000	20.000	10.160	3.525	0.794	0.117	0.000
$\sigma = 0.2$	Num.	40.000	30.000	20.000	10.197	3.523	0.780	0.116	0.000
$T - t = 0.25$	Euro.	38.233	28.431	18.680	9.667	3.421	0.769	0.110	0.000
$r - q = 0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.000	30.000	20.527	12.921	7.428	3.843	1.594	0.000
$\sigma = 0.4$	Num.	40.000	30.000	20.586	12.953	7.435	3.843	1.589	0.000
$T - t = 0.25$	Euro.	38.647	28.993	20.105	12.734	7.338	3.800	1.576	0.000
$r - q = 0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.000	30.000	20.000	10.705	4.770	1.754	0.510	0.000
$\sigma = 0.2$	Num.	40.000	30.000	20.000	10.755	4.766	1.728	0.499	0.000
$T - t = 0.50$	Euro.	37.268	27.498	18.077	10.041	4.555	1.677	0.485	0.000
$r - q = -0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.000	30.120	20.149	11.251	4.397	1.118	0.183	0.000
$\sigma = 0.2$	Num.	40.000	30.116	20.417	11.251	4.397	1.118	0.183	0.000
$T - t = 0.25$	Euro.	39.793	30.083	20.413	11.250	4.396	1.118	0.183	0.000
$r - q = -0.04$	S	60	70	80	90	100	110	120	130
$r = 0.12$	Quad.	40.000	30.000	20.248	11.146	4.355	1.107	0.182	0.000
$\sigma = 0.2$	Num.	40.000	30.000	20.239	11.142	4.354	1.107	0.182	0.000
$T - t = 0.25$	Euro.	39.397	29.790	20.210	11.138	4.353	1.107	0.182	0.000
$r - q = -0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.021	30.379	21.463	13.923	8.247	4.400	1.886	0.000
$\sigma = 0.4$	Num.	40.014	30.357	21.442	13.919	8.244	4.404	1.909	0.000
$T - t = 0.25$	Euro.	39.815	30.302	21.430	13.908	8.239	4.397	1.884	0.000
$r - q = -0.04$	S	60	70	80	90	100	110	120	130
$r = 0.08$	Quad.	40.000	30.280	20.982	12.645	6.372	2.645	0.870	0.000
$\sigma = 0.2$	Num.	40.000	30.261	20.974	12.640	6.370	2.645	0.871	0.000
$T - t = 0.50$	Euro.	39.573	30.169	20.947	12.632	6.367	2.643	0.869	0.000

TABLE 4.3

Comparison between the predictions of our analytical approximate solution (labeled as Quad.) and the results from numerical computations (labeled as Num.) for American lookback strike call options. The predictions of the European lookback strike call options (labeled as Euro.) are also shown. The values of the lookback strike call options shown are scaled by $S_{\min}$, namely, $C_{lb}/S_{\min}$ and $c_{lb}/S_{\min}$ for American- and European-style options, respectively

$r - q = -0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.1437	0.1736	0.2314	0.3093	0.4007	0.5000	0.6000	0.7000
$\sigma = 0.4$	Num.	0.1432	0.1736	0.2320	0.3103	0.4014	0.5000	0.6000	0.7000
$T - t = 0.25$	Euro.	0.1413	0.1709	0.2272	0.3020	0.3878	0.4796	0.5742	0.6703
$r - q = -0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.0982	0.1297	0.2018	0.3000	0.4000	0.5000	0.6000	0.7000
$\sigma = 0.2$	Num.	0.0974	0.1301	0.2026	0.3000	0.4000	0.5000	0.6000	0.7000
$T - t = 0.50$	Euro.	0.0935	0.1231	0.1862	0.2685	0.3590	0.4522	0.5461	0.6402
$r - q = -0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.0724	0.1117	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000
$\sigma = 0.2$	Num.	0.0723	0.1123	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000
$T - t = 0.25$	Euro.	0.0707	0.1082	0.1878	0.2818	0.3785	0.4755	0.5725	0.6696
$r - q = -0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.12$	Quad.	0.0720	0.1113	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000
$\sigma = 0.2$	Num.	0.0718	0.1118	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000
$T - t = 0.25$	Euro.	0.0700	0.1071	0.1860	0.2790	0.3747	0.4707	0.5668	0.6629
$r - q = 0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.0812	0.1248	0.2101	0.3071	0.4059	0.5049	0.6039	0.7029
$\sigma = 0.2$	Num.	0.0812	0.1248	0.2101	0.3071	0.4059	0.5049	0.6039	0.7029
$T - t = 0.25$	Euro.	0.0812	0.1248	0.2101	0.3071	0.4059	0.5049	0.6039	0.7029
$r - q = 0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.12$	Quad.	0.0804	0.1235	0.2081	0.3042	0.4022	0.5008	0.6000	0.7000
$\sigma = 0.2$	Num.	0.0804	0.1235	0.2081	0.3041	0.4021	0.5006	0.6000	0.7000
$T - t = 0.25$	Euro.	0.0804	0.1235	0.2080	0.3040	0.4018	0.4998	0.5979	0.6959
$r - q = 0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.1526	0.1850	0.2456	0.3244	0.4137	0.5087	0.6054	0.7037
$\sigma = 0.4$	Num.	0.1526	0.1850	0.2456	0.3244	0.4137	0.5082	0.6054	0.7037
$T - t = 0.25$	Euro.	0.1526	0.1850	0.2455	0.3243	0.4136	0.5081	0.6052	0.7034
$r - q = 0.04$	$S/S_{\min}$	1.0000	1.1000	1.2000	1.3000	1.4000	1.5000	1.6000	1.7000
$r = 0.08$	Quad.	0.1148	0.1523	0.2261	0.3162	0.4122	0.5097	0.6078	0.7059
$\sigma = 0.2$	Num.	0.1148	0.1523	0.2261	0.3162	0.4122	0.5097	0.6077	0.7059
$T - t = 0.50$	Euro.	0.1148	0.1523	0.2261	0.3162	0.4121	0.5096	0.6075	0.7055

TABLE 4.4

Comparisons between the predictions of our analytical approximate solution (labeled as Quad.) and the results from numerical computations (labeled as Num.) for American lookback strike put options. The predictions of the European lookback strike put options (labeled as Euro.) are also shown. The values of the lookback strike put options shown here are scaled by $S_{\max}$, namely, $P_{lb}/S_{\max}$ and $p_{lb}/S_{\max}$ for American- and European-style options, respectively

$r - q = -0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3012	0.2045	0.1190	0.0853
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3011	0.2045	0.1190	0.0853
$T - t = 0.25$	Euro.	0.6891	0.5920	0.4950	0.3979	0.3009	0.2044	0.1190	0.0853
$r - q = -0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.12$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2028	0.1179	0.0845
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2028	0.1179	0.0845
$T - t = 0.25$	Euro.	0.6822	0.5861	0.4900	0.3940	0.2979	0.2024	0.1178	0.0844
$r - q = -0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4004	0.3061	0.2265	0.1771	0.1707
$\sigma = 0.4$	Num.	0.7000	0.6000	0.5000	0.4003	0.3059	0.2263	0.1770	0.1707
$T - t = 0.25$	Euro.	0.6891	0.5920	0.4950	0.3984	0.3054	0.2262	0.1769	0.1707
$r - q = -0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3030	0.2123	0.1426	0.1221
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3026	0.2120	0.1426	0.1221
$T - t = 0.50$	Euro.	0.6783	0.5841	0.4849	0.3957	0.3018	0.2119	0.1425	0.1220
$r - q = 0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1088	0.0775
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1088	0.0777
$T - t = 0.25$	Euro.	0.6832	0.5842	0.4852	0.3862	0.2872	0.1892	0.1058	0.0763
$r - q = 0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.12$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1083	0.0770
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1083	0.0771
$T - t = 0.25$	Euro.	0.6764	0.5784	0.4803	0.3822	0.2843	0.1873	0.1047	0.0755
$r - q = 0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3004	0.2176	0.1692	0.1636
$\sigma = 0.4$	Num.	0.7000	0.6000	0.5000	0.4000	0.3005	0.2177	0.1693	0.1640
$T - t = 0.25$	Euro.	0.6832	0.5842	0.4852	0.3868	0.2928	0.2145	0.1676	0.1625
$r - q = 0.04$	$S/S_{\max}$	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
$r = 0.08$	Quad.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1258	0.1079
$\sigma = 0.2$	Num.	0.7000	0.6000	0.5000	0.4000	0.3000	0.2000	0.1259	0.1085
$T - t = 0.50$	Euro.	0.6667	0.5687	0.4707	0.3727	0.2753	0.1847	0.1208	0.1051

solutions S^{γ_1} and S^{γ_2} . For American vanilla options, one can only use one fundamental solution, the other solution diverges in certain limits ($S \rightarrow 0$ for call and $S \rightarrow \infty$ for put). In this chapter, we have demonstrated that one can use both the fundamental solutions to study more complicated problems.

5. Acknowledgments

The work of Q. Zhang was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, Project CityU 103807. This work was supported by City University of Hong Kong, grant No. 7001752.

References

- BABBS (1986). Fx hindsight options. Working paper.
- BABBS (1992). Binomial valuations of lookback options. Working paper.
- BARONE-ADESI, G., WHALEY, R.E. (1987). Efficient analytical approximation of American option values. *J. Financ.* **2**, 301–320.
- BLACK, F. (1976). The pricing of commodity contracts. *J. Financ. Econ.* **3**, 167–179.
- BLACK, F., SCHOLES, M.S. (1973). The pricing of options and corporate liabilities. *J. Financ. Econ.* **81**, 637–654.
- BOYLE, P.P., LAU, S.H. (1994). Bumping up against the barrier with the binomial method. *J. Derivatives*, 6–14.
- CHEUK, T.H.F., VORST, T.C.F. (1997). Currency lookback options and observation frequency: a binomial approach. *J. Int. Money Financ.* **16**, 8–22.
- GARMAN, M. (1989). Recollection in tranquility. *Risk*.
- GESKE, R., JOHNSON, H.E. (1984). The American put valued analytically. *J. Financ.* **39**, 1511–1524.
- GOLDMAN, B., SOSIN, H., GATTO, M.A. (1979). Path-dependent options: buy at the low, sell at the high. *J. Financ.* **34**, 1111–1127.
- HUDSON, M. (1991). The value of going out. *RISK* **34**.
- HULL, J.C. (1993). *Options, Futures, and Other Derivatives* (Prentice Hall, Englewood Cliffs, NJ).
- JOHNSON, H.E. (1983). An analytic approximation for the American put price. *J. Financ. Quant. Anal.* **18**, 141–148.
- KAT, H. (1995). Pricing lookback options using binomial trees: an evaluation. *J. Financ. Eng.* **4**, 375–397.
- MACMILLAN, L.W. (1986). Analytic approximation for the American put options. *Advances in Futures and Options Research* **1**, 119–139.
- MERTON, R.C. (1973). The theory of rational option pricing. *Bell J. Econ. Manage. Sci.* **4**, 141–183.
- RITCHEN, P. (1995). On pricing barrier options. *J. Derivatives* 19–28.
- RUBINSTEIN, M., REINER, E. (1991). Breaking down the barrier. *Risk* 19–28.
- STOLL, H.R., WHALEY, R.E. (1986). The new option instruments: arbitrageable linkages and valuation. *Advances in Futures and Options Research* **1**, 25–62.
- STRICKWERDA, J.C. (1989). *Finite Difference Schemes and Partial Differential Equations* (Wadsworth & Brook/Cole, Pacific Grove, CA).
- WILMOTT, P., DEWYNNE, J., HOWISON, S. (1993). *Option Pricing, Mathematical Models and Computation* (Oxford Financial Press, Oxford, UK).

Asset Prices With Regime-Switching Variance Gamma Dynamics

Andrew J. Royal

Haskayne School of Business, University of Calgary, Calgary, CANADA T2N 1N4
Email address: ajroyal@ucalgary.ca

Robert J. Elliott

Haskayne School of Business, University of Calgary, Calgary, CANADA T2N 1N4
Email address: relliott@ucalgary.ca

Abstract

Recently, Elliott and Osakwe have discussed option pricing when the price process has dynamics described by a regime-switching Lévy process. The regime switching is determined by an observable Markov chain. In this chapter, a related framework is considered, but the regime-switching chain is not observed directly. Its state and dynamics can only be estimated using some new filters. The results are tested empirically for option prices using S&P data.

1. Introduction

Empirical work has suggested that financial modelling should move away from the standard log-normal dynamics of the Black–Scholes framework. ELLIOTT and OSAKWE [2006] discussed option pricing when the price of the underlying asset has dynamics given by a regime-switching Lévy process. The regime-switching is intended to model the state of the market or economy, and it is described mathematically by a finite-state Markov chain. In a study by ELLIOTT and OSAKWE [2006], this chain is taken to be fully observed.

The work presented in this chapter considers a related framework where the underlying asset price has variance gamma (VG) dynamics, which switch according to a Markov chain. However, the Markov chain is not supposed to be observed directly. Information about the chain must be estimated from observed processes, such as the return of the underlying asset. New filters are obtained for the chain in this context. The results are tested empirically using S&P data.

2. The model for asset price returns

Consider a filtered probability space $\{\Omega, \mathcal{F}, \mathbf{G}, \bar{P}\}$. We suppose there are two independent stochastic processes defined on this space denoted by X and Y . The two processes generate the complete right continuous filtration $\mathbf{G} \equiv \{\mathcal{G}_t\}_{t \geq 0}$, where $\mathcal{G}_t$ is the right continuous complete filtration generated from $\mathcal{G}_t^0 = \sigma\{X_s, Y_s; s \leq t\}$. $\mathbf{G}$ will be referred to as the global filtration. The global filtration is distinct from two other natural filtrations on the probability space, namely $\mathcal{F}_t^0 = \sigma\{X_s; 0 \leq s \leq t\}$ and $\mathcal{Y}_t^0 = \sigma\{Y_s; 0 \leq s \leq t\}$. Then,

$$\mathbf{F} \equiv \{\mathcal{F}_t\}_{t \geq 0},$$

$$\mathbf{Y} \equiv \{\mathcal{Y}_t\}_{t \geq 0},$$

where $\{\mathcal{F}_t\}$ (resp. $\{\mathcal{Y}_t\}$) is the right continuous, complete filtration generated by $\{\mathcal{F}_t^0\}$ (resp. $\{\mathcal{Y}_t^0\}$).

The state of the economy or market will be modelled by a process, X , which is a continuous-time, finite-state Markov chain, taking values in the set of N canonical basis vectors. More specifically, the stochastic process $X(\omega, t) : \Omega \times [0, \infty) \rightarrow \mathcal{L}$, where the set $\mathcal{L}$ is defined as

$$\mathcal{L} \equiv \{\mathbf{e}_1, \dots, \mathbf{e}_N\} \subset \mathfrak{R}^N,$$

and $\mathbf{e}_i \in \mathfrak{R}^N$ has 1 in the i^{th} position and zero elsewhere. As usual, we shall write $X_t(\omega) = X(\omega, t)$, depending on context. Write $p_t^i = \bar{P}(X_t = \mathbf{e}_i | X_0)$ and $p_t = (p_t^1, p_t^2, \dots, p_t^N)'$. The evolution of the chain is usually described in terms of its rate matrix (or Q matrix) A so that

$$\frac{dp_t}{dt} = Ap_t.$$

Then, X has semimartingale decomposition (see ELLIOTT, AGGOUN and MOORE [1995])

$$X_t(\omega) = X_0(\omega) + \int_0^t A X_s(\omega) ds + V_t.$$

Here, V_t is a $(\bar{P}, \mathbf{G})$ martingale with values in $\mathfrak{R}^N$, which by definition is independent of Y .

The observation process will be a map $Y(\omega, t) : \Omega \times [0, \infty) \rightarrow \mathfrak{R}$, which we shall suppose is a VG process. This can be represented in a number of equivalent ways.

2.1. First representation—time-changed brownian motion

The first representation of the VG process is as a time-changed Brownian motion with drift. Write $Z_t(\omega) = \theta t + \sigma B_t(\omega)$, where B_t is a Brownian motion. Here, the drift is θt and the instantaneous variance σ^2 . We shall write $G_t^\nu(\omega)$ for a gamma process independent of the Brownian motion. That is,

$$G_{t+h}^\nu - G_t^\nu \stackrel{\text{law}}{\equiv} \gamma\left(\frac{h}{\nu}, \frac{1}{\nu}\right),$$

where $\gamma(\frac{h}{\nu}, \frac{1}{\nu})$ is a gamma variable with mean h and variance νh . We write Y as a Brownian motion subordinated by a gamma process, that is,

$$Y_t \stackrel{\text{law}}{=} Z_{G_t^\nu}.$$

This representation has a natural economic interpretation: “packets” of information affecting security prices arrive at the market randomly according to a gamma process. At those times, the size of the returns will be given by the drifted Brownian motion, Z evaluated at the new time. The gamma process is discontinuous and consequently the composition of the VG process is too.

2.2. Second representation—difference of two gamma processes

The first representation makes it easy to write down the characteristic function by using a conditional expectation argument. We have

$$\begin{aligned} E[e^{iuY_t}] &= E[E[\exp(iuZ_{G_t^\nu})|G_t^\nu]] \\ &= E\left[\exp\left(\left(iu\theta - \frac{\sigma^2 u^2}{2}\right)G_t^\nu\right)\right] \\ &= \left(1 - iu\theta\nu + \frac{\sigma^2 u^2 \nu}{2}\right)^{-\frac{t}{\nu}}. \end{aligned} \quad (2.1)$$

A second representation of a VG process is via a decomposition of the characteristic function in (2.1).

Write

$$\begin{aligned} \frac{1}{C} &= \nu, \\ \frac{1}{G} &= \sqrt{\frac{\theta^2 \nu^2}{4} + \frac{\sigma^2 \nu}{2}} - \frac{\theta \nu}{2}, \\ \frac{1}{M} &= \sqrt{\frac{\theta^2 \nu^2}{4} + \frac{\sigma^2 \nu}{2}} + \frac{\theta \nu}{2}. \end{aligned} \quad (2.2)$$

For ease of computation, we note that

$$\frac{1}{GM} = \frac{\sigma^2 v}{2}$$

and $\frac{1}{M} - \frac{1}{G} = \theta v$.

Consequently, we may factor Eq. (2.1) as

$$\left(1 - iu\theta v + \frac{\sigma^2 u^2 v}{2}\right)^{-\frac{t}{v}} = \left(1 - \frac{i u}{M}\right)^{-Ct} \left(1 + \frac{i u}{G}\right)^{-Ct}. \quad (2.3)$$

The importance of this factorization is that we can interpret the VG process as the difference of two independent gamma processes. Specifically, if $G \sim \gamma(a, b)$, then its characteristic function is given by

$$E[e^{iuG}] = \left(1 - \frac{i u}{b}\right)^{-a}.$$

Consequently, the characteristic function of the difference of two independent gamma processes $G_i \sim \gamma(a_i, b_i)$ is given by

$$\begin{aligned} E[e^{iu(G_1 - G_2)}] &= E[e^{iuG_1}]E[e^{-iuG_2}] \\ &= \left(1 - \frac{i u}{b_1}\right)^{-a_1} \left(1 + \frac{i u}{b_2}\right)^{-a_2}, \end{aligned}$$

which has the same functional form as Eq. (2.3).

2.3. Third representation—Lévy measure

In the context of Lévy processes, the characteristic function takes on extra special significance through the following theorem.

THEOREM 2.1 (Lévy–Khintchine formula). *If Y is a square-integrable Lévy process, then its characteristic function may be written in the following way*

$$E[e^{iuY_t}] = \exp(t\phi(u)),$$

where the unit log-characteristic function is written as

$$\phi(u) = i u b - \frac{1}{2} \sigma^2 u^2 + \int_{\mathfrak{N}} (e^{i u x} - 1 - i u x I_{\{|x| \leq 1\}}) v(dx).$$

Here, $v(dx)$ is a sigma-finite measure, which satisfies $v(\{0\}) = 0$ and the integrability condition

$$\int_{\mathfrak{N}} (1 \wedge x^2) v(dx) < \infty.$$

We also have $\sigma > 0$ and b are real numbers.

DEFINITION 2.1 (Lévy triple). The triple (b, σ, ν) described in the Lévy–Khintchine theorem is referred to as the Lévy triple and uniquely defines a Lévy process. We can calculate the Lévy triple by using the following identity (see SATO [2004], example 8.10 or SHIRYAEV [1999], p. 206.):

$$\begin{aligned}
 (1 + \beta v)^{-1} &= \exp(-\log(1 + \beta v)) \\
 &= \exp\left(-\int_0^v \frac{\beta ds}{1 + \beta s}\right) \\
 &= \exp\left(-\int_0^v \int_0^\infty e^{-x(\frac{1}{\beta} + s)} dx ds\right) \\
 &= \exp\left(-\int_0^\infty \int_0^v e^{-xs} ds e^{-x/\beta} dx\right) \\
 &= \exp\left(\int_0^\infty (e^{-xv} - 1) \frac{e^{-x/\beta}}{x} dx\right).
 \end{aligned}$$

By analytic continuation (see section 52, example 1 in BROWN and CHURCHILL [1996]), we may extend this result to the half-plane $v \in \{a + bi \in \mathbb{C} | a \geq 0\}$ for all values of $\beta \in \Re$.

The importance of this last result is that it allows one to determine the Lévy measure for the VG process. When the last equation is evaluated at the points $v = \pm iu$ and $\beta = \frac{1}{M}, \frac{1}{G}$, (both of which are in the domain of the analytic function $(1 + \beta u)^{-1}$), combined with Eq. (2.3), we have the following representation for the VG process

$$\begin{aligned}
 &\left(1 - \frac{iu}{M}\right)^{-Ct} \left(1 + \frac{iu}{G}\right)^{-Ct} \\
 &= \exp\left(Ct \int_0^\infty (e^{iux} - 1) \frac{e^{-Mx}}{x} dx\right) \exp\left(Ct \int_0^\infty (e^{-iux} - 1) \frac{e^{-Gx}}{x} dx\right) \\
 &= \exp\left(Ct \int_0^\infty (e^{iux} - 1) \frac{e^{-Mx}}{x} dx\right) \exp\left(Ct \int_{-\infty}^0 (e^{iux} - 1) \frac{e^{-G|x|}}{|x|} dx\right) \\
 &= \exp\left(t \int_{\Re} (e^{iux} - 1) \nu(dx)\right) \\
 &= \exp\left(t i u b + t \int_{\Re} (e^{iux} - 1 - iux I_{\{|x| \leq 1\}}) \nu(dx)\right),
 \end{aligned}$$

where we have written $\nu(dy) = k(y)dy$ as the Lévy measure, where k is given by

$$k(y) = \frac{C \exp(-My)}{y} I_{\{y > 0\}} + \frac{C \exp(-G|y|)}{|y|} I_{\{y < 0\}}. \quad (2.4)$$

and the drift, b , is given by

$$\begin{aligned} b &= -C \int_{-1}^0 e^{-G|x|} dx + C \int_0^1 e^{-Mx} dx \\ &= \frac{C}{M} (1 - e^{-M}) - \frac{C}{G} (1 - e^{-G}). \end{aligned}$$

Thus, the Lévy triple is given by $(b, 0, \nu)$.

REMARK 2.1 (the drift, b). Usually, the b is ignored as the Lévy–Khintchine formula has many different representations. Specifically, if we exchange the function $xI_{\{|x| \leq 1\}}$ for a more general truncation function h , which has compact support (i.e., is zero outside of a closed and bounded set in $\mathfrak{R}$, and behaves like $h(x) = x$ near the origin), we find that the drift will become dependent on this truncation function with the Lévy measure remaining invariant.

REMARK 2.2 (the volatility, σ). The Lévy–Kintchine formula decomposes a general Lévy process into the sum of two independent random processes: the first part is a continuous Brownian motion, with volatility σ and the second part is a pure jump process with Lévy measure ν . Given that the VG process has $\sigma = 0$, we can say that this process has no continuous Brownian motion component: the VG process is a pure jump process.

One might well wonder what use the Lévy triple representation has, but, in fact, it has a fairly simple explanation found in the following.

2.4. Fourth representation—predictable compensator

The last representation of Y is via its predictable compensator, which is related to its Lévy measure in a fundamental way.

DEFINITION 2.2 (Random measure). A random measure on $\mathfrak{R} \times \mathfrak{R}_+$ is a family $\mu = \{\mu(dy, dt; \omega); \omega \in \Omega\}$ of nonnegative measures on the measure space $(\mathfrak{R} \times \mathfrak{R}_+, \mathbf{B}(\mathfrak{R}) \otimes \mathbf{B}(\mathfrak{R}_+))$, which satisfy $\mu(\mathfrak{R} \times \{0\}; \omega) = 0$. First (dropping the ω subscript), the random (jump) measure induced by Y is defined as follows:

$$\begin{aligned} \mu(A, (0, t]) &\triangleq \sum_{0 < s \leq t} I_{\{\Delta Y_s \in A, \Delta Y_s \neq 0\}} \\ &= \sum_{s \in (0, t]} I_{\{\Delta Y_s \in A\}} \delta_{(s, \Delta Y_s)}((0, t] \times A). \end{aligned}$$

Here, $A \subset \mathfrak{R}$ is any Borel set that does not contain an open interval surrounding the origin, thus ensuring that μ is finite. As usual, δ_a is the Dirac measure for the point $a \in \mathfrak{R}_+ \times \mathfrak{R}$.

REMARK 2.3. The quantity $\mu(A, (0, t])$ is integer valued and can be intuitively thought of as the number of jumps of “size” A that the process Y makes in the time interval

$[0, t]$. The integral of a suitable function $f : \mathfrak{R} \times \mathfrak{R}_+ \times \Omega \rightarrow \mathfrak{R}$ with respect to this jump measure is defined by

$$\begin{aligned}(f * \mu)_t &= \int_{(0,t]} \int_{\mathfrak{R}} f(y, s-) \mu(dy, ds) \\ &= \sum_{0 < s \leq t} f(\Delta Y_s, s I_{\{\Delta Y_s \neq 0\}}).\end{aligned}$$

Here, we have suppressed the ω notation for convenience. Next, we shall define the predictable compensator, ν^P , of this quantity as a random measure such that $E[(f * \nu^P)_t] = E[(f * \mu)_t]$ for all $t \geq 0$ and all f such that f is a left continuous process (i.e., f is predictable) (see theorem 2.1.8 of JACOD and SHIRYAEV [1987] for the existence of ν^P). Then, we may calculate the Lévy measure in terms of the predictable compensator by

$$\begin{aligned}e^{iuX_t} &= 1 + \sum_{0 < s \leq t} e^{iuX_s} - e^{iuX_{s-}} \\ &= 1 + \sum_{0 < s \leq t} e^{iuX_{s-}} (e^{iu\Delta X_s} - 1) \\ &= 1 + \int_0^t e^{iuX_{s-}} \int_{\mathfrak{R}} (e^{iux} - 1) \mu(dx; ds) \\ &= 1 + \int_0^t e^{iuX_{s-}} \int_{\mathfrak{R}} (e^{iux} - 1) (\mu(dx; ds) - \nu^P(dx; ds)) \\ &\quad + \int_0^t e^{iuX_{s-}} \int_{\mathfrak{R}} (e^{iux} - 1) \nu^P(dx; ds) \\ &= 1 + \text{martingale} + \int_0^t e^{iuX_{s-}} \int_{\mathfrak{R}} (e^{iux} - 1) \nu^P(dx; ds).\end{aligned}$$

Taking expectations and solving, we observe that

$$\begin{aligned}E_P[e^{iuX_t}] &= 1 + \int_0^t E_P[e^{iuX_{s-}}] \int_{\mathfrak{R}} (e^{iux} - 1) \nu(dx, ds) \\ &= \exp\left(\int_0^t \int_{\mathfrak{R}} (e^{iux} - 1) \nu(dx; ds)\right).\end{aligned}$$

Equating this with the Lévy–Kintchine formula, we see that the compensator measure $\nu(dx; ds)$ is the same as the Lévy measure. That is, the predictable compensator of the VG process Y is given by

$$\nu^P(dy; ds) = k(y) dy ds = \left(\frac{C \exp(-My)}{y} I_{\{y>0\}} + \frac{C \exp(-G|y|)}{|y|} I_{\{y<0\}} \right) dy ds \quad (2.5)$$

The compensator measure is very useful as it allows the use of the tools of stochastic calculus. This is implicit in the next section.

3. VG processes under absolutely continuous changes of measures

One of the interesting tools in the filtering techniques developed in the following sections is the concept of a reference measure. Under the reference probability measure, the process Y acts and behaves as if it were a VG process, while under the historical measure, it acts like a VG process with regime-switching parameters.

However, we want the reference probability to be equivalent to the historical measure up to a change of measure. For this to work properly, we need the historical measure to be absolutely continuous with respect to the reference measure.

At this point, it is useful to introduce a truncated version of the VG process such that we can ensure that the change of measure is well defined.

Instead of considering Y , we shall consider the tails of this process. For any $\epsilon > 0$, write

$$Y_t^\epsilon = Y_0 + \sum_{0 < s \leq t} \Delta Y_s I_{\{|\Delta Y_s| > \epsilon\}}.$$

For $A \in \mathcal{B}(\mathfrak{R})$, the integer-valued random measure is then defined by $\mu^\epsilon : \mathfrak{R} \times \mathfrak{R}_+ \times \Omega \rightarrow \mathbb{N}$ by

$$\mu^\epsilon(A \times (0, t])(\omega) = \sum_{s \in (0, t]} I_{\{\Delta Y_s^\epsilon(\omega) \in A\}} \delta_{(s, \Delta Y_s^\epsilon)}((0, t] \times A).$$

The truncated VG process makes this sum almost surely finite because infinite jump activity of size less than ϵ is ignored. For convenience, we have omitted the ω from this notation.

Then, we may define the jump times $T_0^\epsilon = 0$ and $T_{n+1}^\epsilon = \inf\{s > T_n^\epsilon; \{Y_t^\epsilon - Y_{T_n^\epsilon}^\epsilon\} > \epsilon\}$. The integral of a suitable function $f : \mathfrak{R} \times \mathfrak{R}_+ \times \Omega \rightarrow \mathfrak{R}$ is defined by

$$\begin{aligned} (f * \mu^\epsilon)_t &= \int_{(0, t]} \int_{\mathfrak{R}} f(y, s-) \mu^\epsilon(dy, ds) \\ &= \sum_{0 < s \leq t} f(\Delta Y_s^\epsilon, s) I_{\{\Delta Y_s^\epsilon \neq 0\}} \\ &= \sum_{n \geq 0} f(\Delta Y_{T_n^\epsilon}^\epsilon, T_n^\epsilon) I_{\{T_n^\epsilon \leq t\}}. \end{aligned}$$

If Y is a VG process with predictable compensator $k(y)dyds$, then Y^ϵ is our truncated VG process with predictable compensator $v^\epsilon(dy; ds) = k(y)I_{\{|y| \geq \epsilon\}}dyds = k^\epsilon(y)dyds$. Thus, the integral $y * (\mu^\epsilon - v^\epsilon)$ is, by definition, a martingale.

In fact, if we set

$$b^\epsilon = \frac{C}{M} (e^{-\epsilon} - e^{-M}) - \frac{C}{G} (e^{-\epsilon} - e^{-G}).$$

then the Lévy triple of Y^ϵ will be $(b^\epsilon, 0, v^\epsilon)$.

DEFINITION 3.1 (Local absolutely continuous measures). P is locally absolutely continuous with respect to P' (written $P \ll_{\text{loc}} P'$), if $P|_{\mathcal{F}_t} \ll P'|_{\mathcal{F}_t}$ for all $t \geq 0$.

The importance of this truncation is that the integrability conditions of the following theorem hold.

THEOREM 3.1. If, under P_i , the Lévy triple is given by $(b_i, 0, \nu_i)$ for $i = 1, 2$, then $P_1 \ll_{\text{loc}} P_2$ provided that

1. $\int_{-1}^1 \{xk_1(x) - xk_2(x)\} dx < \infty$,
2. $b_1 = b_2 + \int_{-1}^1 x(k_1(x) - k_2(x)) dx$, and
3. $\int_{\mathbb{R}} (\sqrt{k(x)} - \sqrt{k'(x)})^2 dx < \infty$.

PROOF. See JACOD and SHIRYAEV [1987], theorem 4.4.39(c). Note, we have simplified the conditions as there is no Brownian motion component in the Lévy process that we are analyzing. Also, in JACOD and SHIRYAEV [1987], the theorem is for general processes of independent increments, not just Lévy processes.

Note that the first and second conditions in Theorem 3.1 hold without the need for truncation; however, the third condition can fail as the following example shows. $\square$

EXAMPLE 3.1 (Two gamma processes that are not absolutely continuous). Suppose we have two Lévy densities given by $k_i(y) = \frac{C_i e^{-M_i y}}{y} I_{\{y>0\}}$ for $i = 1, 2$ with $C_1 \neq C_2$ but $M_1 = M_2$. Then, Condition 3 is calculated by

$$\begin{aligned} \int_0^\infty (\sqrt{k_1(y)} - \sqrt{k_2(y)})^2 dy &= \int_0^\infty \frac{e^{-M_1 y}}{y} (\sqrt{C_1} - \sqrt{C_2})^2 dy \\ &> (\sqrt{C_1} - \sqrt{C_2})^2 \left(\int_0^1 \frac{dy}{y} + \int_1^\infty \frac{e^{-M_1 y}}{y} dy \right) \\ &= \infty. \end{aligned}$$

Consequently, this theorem shows that it do *not* always have $P \ll_{\text{loc}} P'$ for arbitrary VG processes. We can ensure that this condition holds if we truncate the process near the origin. This is a critical point when it comes to the estimation of parameters based on the Expectation Maximization (EM) algorithm estimation scheme.

REMARK 3.1. The conditions in the theorem are symmetric in P_1 and P_2 when truncation is used. This shows us that we actually have $P_1 \sim_{\text{loc}} P_2$. There is one other way to ensure that condition 3 holds and is of interest for derivative pricing (see ELLIOTT and OSAKWE [2006]).

EXAMPLE 3.2 (Absolutely continuous subfamilies of gamma processes). Suppose we have two Lévy densities given by $k_i(y) = \frac{C_i e^{-M_i y}}{y} I_{\{y>0\}}$ for $i = 1, 2$ with $C_1 = C_2$.

Then, we shall show that without any truncation $P_1 \stackrel{\text{loc}}{\sim} P_2$. Without loss of generality, suppose that $M_2 > M_1$. Then,

$$\int_0^\infty \left(\sqrt{k_1(y)} - \sqrt{k_2(y)} \right)^2 dy = \int_0^\infty C_1 e^{-M_1 y} \left(\frac{1 - e^{-\frac{(M_2 - M_1)}{2} y}}{\sqrt{y}} \right)^2 dy.$$

But $f(y) = e^{-y}$ is a convex function that must satisfy the defining condition $f'(0)(y - 0) \leq f(y) - f(0)$ or $-y \leq e^{-y} - 1$ for all $y \in \Re$. That is,

$$\frac{1 - e^{-\frac{(M_2 - M_1)}{2} y}}{\sqrt{y}} \leq \frac{M_2 - M_1}{2} \sqrt{y}.$$

Condition 3 of Theorem 3.1 now holds because

$$\int_0^\infty C_1 e^{-M_1 y} \left(\frac{1 - e^{-\frac{(M_2 - M_1)}{2} y}}{\sqrt{y}} \right)^2 dy \leq C_1 \left(\frac{M_2 - M_1}{2} \right)^2 \int_0^\infty y e^{-M_1 y} dy < \infty.$$

REMARK 3.2. Unfortunately, for the EM algorithm, we need to be able to update all parameters, not just G and the M . Thus, the EM algorithm would become much simpler if we knew with certainty what C was.

4. Estimating model parameters

4.1. The reference measure

For the rest of the chapter, we shall write Y (resp. μ , k) for the truncated VG process, Y^ϵ (resp. for the jump measure μ^ϵ , for the compensator measure k^ϵ), to avoid the more cumbersome notation.

Suppose under $\bar{P}$, Y is a truncated VG process with parameters $C = 1$, $G = M = \sqrt{2}$. For $j = 1, \dots, m$, $C_j, M_j, G_j \in \Re_+$, consider a likelihood function $L_j^\epsilon(y)$ defined by

$$\begin{aligned} L_j^\epsilon(y) &= C_j \exp(- (M_j - \sqrt{2})y) I_{\{y > \epsilon\}} \\ &\quad + C_j \exp(- (G_j - \sqrt{2})|y|) I_{\{y < -\epsilon\}} + I_{\{|y| \leq \epsilon\}} \\ &= \frac{k_j(y)}{k(y)} I_{\{|y| > \epsilon\}} + I_{\{|y| \leq \epsilon\}}, \end{aligned}$$

where $k(y)$ (resp. $k_j(y) I_{\{|y| > \epsilon\}}$) is the Lévy measure of a VG process with parameters $C = 1$, $G = M = \sqrt{2}$ (resp. a VG process with parameters C_j, M_j, G_j). Consider two

processes on $(\Omega, \mathcal{F}, \mathbf{G}, \bar{P})$ defined by

$$\begin{aligned}\bar{U}_t &= \int_{(0,t]} \sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle > \int_{\mathfrak{R}} (L_j^\epsilon(y) - 1) \{ \mu(dy; ds) - k(y) dy ds \}, \\ \bar{\Lambda}_t &= 1 + \int_{(0,t]} \bar{\Lambda}_{s-} d\bar{U}_s.\end{aligned}$$

Then, $\bar{\Lambda}$ is given by the Doléans–Dade exponential (see JACOD and SHIRYAEV [1987]),

$$\begin{aligned}\bar{\Lambda}_t &= E_t(\bar{U}) \\ &= e^{\bar{U}_t} \prod_{0 < s \leq t} (1 + \Delta \bar{U}_s) e^{-\Delta \bar{U}_s} \\ &= \exp \left(\int_{(0,t]} \sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle > \int_{\mathfrak{R}} \log(L_j^\epsilon(y)) \mu(dy; ds) \right. \\ &\quad \left. - \int_{(0,t]} \sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle > \int_{\mathfrak{R}} (L_j^\epsilon(y) - 1) k(y) dy ds \right).\end{aligned}$$

We may now consider the change of measure

$$\left. \frac{dP}{d\bar{P}} \right|_{\mathcal{G}_t} \equiv \bar{\Lambda}_t, \quad t \geq 0.$$

We require that $P \sim \bar{P}$, which forced the truncation of our Lévy measure from $k(y)$ to $k(y)I_{\{|y|>\epsilon\}}$. (For further information on this point, see JACOD and SHIRYAEV [1987] and the chapter on Hellinger integrals.) Under the measure P (which will be referred to as the historical measure), it will be shown that

$$Y_t^\epsilon - \sum_{j=1}^N \int_{(0,t]} \langle X_{s-}, \mathbf{e}_j \rangle > \int_{\mathfrak{R}} y k_j(y) I_{\{|y|>\epsilon\}} dy ds$$

is a martingale. In other words, under P , the process Y^ϵ has predictable compensator defined by the measure

$$\sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle > k_j(y) I_{\{|y|>\epsilon\}} dy ds.$$

For this, we need the following lemma.

LEMMA 4.1. Z_t is a P martingale if and only if $Z_t \bar{\Lambda}_t$ is a $\bar{P}$ martingale.

PROOF. See JACOD and SHIRYAEV [1987], proposition 3.3.8. □

Set $Z_t = Y_t^\epsilon - \sum_{j=1}^N \int_{(0,t]} < X_{s-}, \mathbf{e}_j > \int_{\mathfrak{N}} y k_j(y) I_{\{|y|>\epsilon\}} dy ds$. Then, by the definition of the square bracket process (see JACOD and SHIRYAEV [1987], p 51, definition 1.4.45), we have

$$\begin{aligned}
 & Z_t \bar{\Lambda}_t \\
 &= Y_0 + \int_{(0,t]} Z_{s-} d\bar{\Lambda}_s + \int_{(0,t]} \bar{\Lambda}_{s-} dZ_s + [Z, \bar{\Lambda}]_t \\
 &= Y_0 + \int_{(0,t]} Z_{s-} d\bar{\Lambda}_s + \int_{(0,t]} \bar{\Lambda}_{s-} dY_s^\epsilon \\
 &\quad - \sum_{j=1}^N \int_{(0,t]} \bar{\Lambda}_{s-} < X_{s-}, \mathbf{e}_j > \int_{\mathfrak{N}} y k_j(y) I_{\{|y|>\epsilon\}} dy ds \\
 &\quad + \sum_{s \in (0,t]} \Delta Y_s^\epsilon \Delta \bar{\Lambda}_s \\
 &= Y_0 + \int_{(0,t]} Z_{s-} d\bar{\Lambda}_s \\
 &\quad + \sum_{j=1}^N \int_{(0,t]} \bar{\Lambda}_{s-} < X_{s-}, \mathbf{e}_j > \int_{\mathfrak{N}} y I_{\{|y|>\epsilon\}} (\mu(dy; ds) - k_j(y) dy ds) \\
 &\quad + \sum_{j=1}^N \int_{(0,t]} \bar{\Lambda}_{s-} < X_{s-}, \mathbf{e}_j > \int_{\mathfrak{N}} y (L_j^\epsilon(y) - 1) \mu(dy; ds) \\
 &= Y_0 + \int_{(0,t]} Z_{s-} d\bar{\Lambda}_s \\
 &\quad + \sum_{j=1}^N \int_{(0,t]} \bar{\Lambda}_{s-} < X_{s-}, \mathbf{e}_j > \int_{\mathfrak{N}} y L_j^\epsilon(y) I_{\{|y|>\epsilon\}} (\mu(dy; ds) - k(y) dy ds).
 \end{aligned}$$

The last equation shows us that $Z_t \bar{\Lambda}_t$ is a $(\bar{P}, \mathbf{G})$ martingale, and by Lemma 4.1, it follows that Z_t is a $\mathcal{P}, \mathbf{G}$ -martingale.

4.2. Estimation

This section's objective is to calculate a linear Zakai equation, which is the basis for much of the rest of the chapter. That is, we calculate quantities such as $E[H_t X_t | \mathcal{Y}_t]$, where H_t is a $\mathcal{G}_t$ -measurable process. More specifically, we want to find estimates of processes of the form

$$H_t = H_0 + \int_0^t \alpha_s ds + \int_0^t \beta'_s dV_s + \int_0^t \int_{\mathfrak{N}} \xi_s \mu(dy; ds).$$

Here α , ξ , and β are predictable square-integrable processes, with $\beta \in \mathfrak{R}^N$. In a sense, H_t spans the space of $\mathcal{G}_t$ -adapted processes. Results of ELLIOTT, AGGOUN and MOORE [1995] illustrate how a recursion of unnormalized estimates of H leads to a linear Zakai equation. This is done using the abstract Bayes' rule (see ELLIOTT, AGGOUN and MOORE [1995]), which we now illustrate. Write $E[\cdot]$ and $(\bar{E}[\cdot])$, for expectations taken with respect to P , and $(\bar{P})$.

LEMMA 4.2 (Conditional Bayes' formula).

$$E[H_t X_t | \mathcal{Y}_t] = \frac{\bar{E}[\bar{\Lambda}_t H_t X_t | \mathcal{Y}_t]}{\bar{E}[\bar{\Lambda}_t | \mathcal{Y}_t]}.$$

PROOF. See ELLIOTT, AGGOUN and MOORE [1995], theorem 3.2, p 23. $\square$

Write $q_t(HX) = \bar{E}[\bar{\Lambda}_t H_t X_t | \mathcal{Y}_t]$. We use Lemma 4.2 to calculate the linear Zakai equation described in the following theorem.

THEOREM 4.1 (Linear Zakai equation).

$$\begin{aligned} q_t(HX) = & q_0(HX) \\ & + \int_0^t \left\{ q_s(HAX + \alpha X) + \sum_{i,j=1}^N a_{ji} < q_s((\beta^j - \beta^i)X), \mathbf{e}_i > (\mathbf{e}_j - \mathbf{e}_i) \right\} ds \\ & + \int_0^t \int_{\mathfrak{R}} \sum_{j=1}^N < q_{s-}(\xi_s(y)X), \mathbf{e}_j > L_j^\epsilon(y) \mu(dy; ds) \mathbf{e}_j \\ & + \int_0^t \sum_{j=1}^N < q_{s-}(HX), \mathbf{e}_j > \int_{\mathfrak{R}} (L_j^\epsilon(y) - 1) \{\mu(dy; ds) - k(y) dy ds\} \mathbf{e}_j. \end{aligned} \quad (4.1)$$

PROOF. Using the Ito rule, we have

$$\begin{aligned} \bar{\Lambda}_t H_t X_t = & \bar{\Lambda}_0 H_0 X_0 + \int_0^t \bar{\Lambda}_s \left\{ H_s A X_s + \alpha_s X_s \right. \\ & + \sum_{i,j=1}^N a_{ji} < (\beta_s^j - \beta_s^i) X_s, \mathbf{e}_i > (\mathbf{e}_j - \mathbf{e}_i) \Big\} ds \\ & + \int_0^t \bar{\Lambda}_{s-} H_{s-} dV_s + \int_0^t \bar{\Lambda}_{s-} X_{s-} \beta'_s dV_s \\ & + \int_0^t \bar{\Lambda}_{s-} \sum_{i,j=1}^N a_{ji} < (\beta_s^j - \beta_s^i) X_{s-}, \mathbf{e}_i > < \mathbf{e}_i, dV_s > (\mathbf{e}_j - \mathbf{e}_i) \end{aligned}$$

$$\begin{aligned}
& + \int_0^t \int_{\mathfrak{R}} \bar{\Lambda}_{s-} \sum_{j=1}^N \langle \xi_s(y) X_{s-}, \mathbf{e}_j \rangle \mu(dy; ds) \mathbf{e}_j \\
& + \int_0^t \bar{\Lambda}_{s-} H_{s-} \sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle \int_{\mathfrak{R}} (L_j^\xi(y) - 1) \{\mu(dy; ds) - k(y) dy ds\} \mathbf{e}_j \\
& + \sum_{0 < s \leq t} \Delta \bar{\Lambda}_s \Delta (HX)_s.
\end{aligned}$$

We may simplify the very last term noting $\Delta X_s \Delta \bar{\Lambda}_s = 0$ a.s. and

$$\sum_{0 < s \leq t} \Delta \bar{\Lambda}_s \Delta (HX)_s = \int_{(0,t]} \int_{\mathfrak{R}} \bar{\Lambda}_{s-} \sum_{j=1}^N \langle \xi_s(y) X_{s-}, \mathbf{e}_j \rangle (L_j^\xi(y) - 1) \mu(dy; ds) \mathbf{e}_j.$$

Note that under $\bar{P}$, $\bar{\Lambda}_s H_s X_s$ is independent of the sigma field $\sigma\{Y_u; s \leq u \leq t\}$ for all $t \geq s$. Secondly, by definition, the measure $\mu(dy; ds)$ is $\mathcal{Y}_t$ measurable for $s \leq t$. Thirdly, the expectation of the $(\bar{P}, \mathbf{G})$ -martingale V_t is zero. Consequently, the result follows. $\square$

4.3. Parameter estimation

Here, we derive an EM algorithm for parameter updating. The parameter space for our model is given by

$$\begin{aligned}
\Theta = & \left\{ (a_{ij})_{i,j=1}^{N,N}, (C_j)_{j=1}^N, (G_j)_{j=1}^N, (M_j)_{j=1}^N; \right. \\
& \left. C_j, G_j, M_j, a_{ij} \geq 0, i \neq j, i, j = 1, \dots, N, \text{ and } \sum_{j=1}^N a_{ij} = 0, i = 1, \dots, N \right\}.
\end{aligned}$$

Consider a fixed $\theta \in \Theta$. Associated with this parameter, we have a probability P_θ under which the process X has rate matrix A , and Y is a VG process with parameters C_j, G_j , and M_j when X is in state j . We wish to estimate a better $\hat{\theta} \in \Theta$. For this, we use the EM algorithm: given observations $\{y_t\}_{0 \leq t \leq T}$ choose

$$\hat{\theta} \in \operatorname{argmax}_{\psi \in \Theta} E_\theta \left[\ln \frac{dP_\psi}{dP_\theta} \mid \mathcal{Y}_T \right].$$

We then define $P_{\hat{\theta}}$ so that our model has parameter $\hat{\theta}$. Firstly, we endeavor to find the Radon–Nikodym derivative that changes the drift of the state process from $\int_0^t A X_s ds$ to $\int_0^t \hat{A} X_s ds$. Consider the counting process, J_t^{ij} , $i \neq j$, which counts the number of jumps from state i to j in the interval $(0, t]$. This has representation

$$\begin{aligned}
J_t^{ij} &= \int_0^t \langle X_{s-}, \mathbf{e}_i \rangle \langle dX_s, \mathbf{e}_j \rangle \\
&= \int_0^t \langle X_{s-}, \mathbf{e}_i \rangle \langle AX_s ds, \mathbf{e}_j \rangle + \int_0^t \langle X_{s-}, \mathbf{e}_i \rangle \langle dV_s, \mathbf{e}_j \rangle \\
&= \int_0^t a_{ji} \langle X_s, \mathbf{e}_i \rangle ds + V_t^{ij} \\
&= a_{ji} O_t^i + V_t^{ij}.
\end{aligned} \tag{4.2}$$

Here, O_t^i is the occupation time of the state process in state i up to time t ; also, the third equality holds because the set $\{s \in [0, t]; X_{s-} \neq X_s\}$ is a.s. finite and has Lebesgue measure zero.

Write

$$\begin{aligned}
\hat{\Lambda}_t^{ij} &= E_t \left(\left(\frac{\hat{a}_{ji}}{a_{ji}} - 1 \right) V^{ij} \right) \\
&= \left(\frac{\hat{a}_{ji}}{a_{ji}} \right)^{J_t^{ij}} \exp(-(\hat{a}_{ji} - a_{ji}) O_t^i).
\end{aligned}$$

Then, the product $\prod_{i \neq j} \hat{\Lambda}_t^{ij}$ defines the required Radon–Nikodym derivative. Write k (resp. $\hat{k}$) for the Lévy measure with parameters $\{C_j, G_j, M_j\}$ (resp. $\{\hat{C}_j, \hat{G}_j, \hat{M}_j\}$)

$$\hat{L}_j^\epsilon(y) = \frac{\hat{k}(y)}{k(y)} I_{\{|y| > \epsilon\}} + I_{\{|y| \leq \epsilon\}}$$

$$\begin{aligned}
\text{and } \hat{U}_t &= \int_{(0,t]} \sum_{j=1}^N \langle X_{s-}, \mathbf{e}_j \rangle \int_{\mathfrak{N}} (\hat{L}_j^\epsilon(y) - 1) \{\mu(dy; ds) - k(y) dy ds\}, \\
\hat{\Lambda}_t &= E_t(\hat{U}) \prod_{i \neq j} \hat{\Lambda}_t^{ij}.
\end{aligned}$$

We now define the measure $\hat{P}$ by

$$\frac{d\hat{P}}{dP} \Big|_{\mathcal{G}} = \hat{\Lambda}_t.$$

Then, under $\hat{P}$, the observation process has parameter $\hat{\theta}$. Now

$$\begin{aligned}
\ln \hat{\Lambda}_T &= \ln E_T(\hat{U}) + \sum_{i \neq j} \ln \hat{\Lambda}_T^{ij} \\
&= - \left((\hat{L}^\epsilon - 1) * k \right)_T + \left((\ln \hat{L}^\epsilon) * \mu \right)_T \\
&\quad + \sum_{i \neq j} \left(J_T^{ij} \ln \left(\frac{\hat{a}_{ij}}{a_{ij}} \right) - (\hat{a}_{ij} - a_{ij}) O_T^i \right).
\end{aligned}$$

Thus, the conditional log-likelihood function looks like

$$\begin{aligned}\ell(\hat{\theta}, \theta) &= E_{\theta}[\ln \hat{\Lambda}_T \mid \mathcal{Y}_T] + E_{\theta}\left[\sum_{i \neq j} \left(J_T^{ij} \ln \left(\frac{\hat{a}_{ij}}{a_{ij}}\right) - (\hat{a}_{ij} - a_{ij})\mathcal{O}_T^i\right) \mid \mathcal{Y}_T\right] \\ &= E_{\theta}\left[-((\hat{L}^{\epsilon} - 1) * k)_T + ((\ln \hat{L}^{\epsilon}) * \mu)_T \mid \mathcal{Y}_T\right] \\ &\quad + \sum_{i \neq j} \left(J_T^{ij} \ln \left(\frac{\hat{a}_{ij}}{a_{ij}}\right) - (\hat{a}_{ij} - a_{ij})\mathcal{O}_T^i\right).\end{aligned}$$

Here, $\mathcal{J}_T^{ij} = E_{\theta}[J_T^{ij} \mid \mathcal{Y}_T]$ and $\mathcal{O}_T^i = E_{\theta}[\mathcal{O}_T^i \mid \mathcal{Y}_T]$. We also introduce the statistics

$$P_T^j = \int_0^T \langle X_{s-}, \mathbf{e}_j \rangle \int_{\epsilon}^{\infty} y \mu(dy; ds) \quad (4.3)$$

$$Q_T^j = \int_0^T \langle X_{s-}, \mathbf{e}_j \rangle \int_{-\infty}^{-\epsilon} |y| \mu(dy; ds), \quad (4.4)$$

$$R_T^j = \int_0^T \langle X_{s-}, \mathbf{e}_j \rangle \mu(|y| > \epsilon; ds) \quad (4.5)$$

and the corresponding estimates $\mathcal{P}_T^j = E[P_T^j \mid \mathcal{Y}_T]$, $\mathcal{Q}_T^j = E[Q_T^j \mid \mathcal{Y}_T]$, and $\mathcal{R}_T^j = E[R_T^j \mid \mathcal{Y}_T]$. Taking derivatives with respect to the parameter $\hat{\theta}$, we get the following result.

THEOREM 4.2 (Parameter updates). *The parameter updates, given the observations, $\mathcal{Y}_T$, are given by*

$$\hat{a}_{ji} = \frac{\mathcal{J}_T^{ij}}{\mathcal{O}_T^i}, \quad \forall i \neq j \quad (4.6)$$

$$\hat{C}_j = \frac{\mathcal{R}_T^j}{(E_1(\hat{G}_j \epsilon) + E_1(\hat{M}_j \epsilon))\mathcal{O}_T^j}, \quad (4.7)$$

$$\hat{M}_j = \frac{\hat{C}_j \mathcal{O}_T^j}{\mathcal{P}_T^j}, \quad (4.8)$$

$$\text{and } \hat{G}_j = \frac{\hat{C}_j \mathcal{O}_T^j}{\mathcal{Q}_T^j}. \quad (4.9)$$

Here, $E_1(x) = \int_x^{\infty} \frac{\exp(-s)}{s} ds$ is the exponential integral function described in ABRAMOWITZ and STEGUN [1965, p 228]. Also note that due to the constraint on the rate matrix making all columns sum to zero,

$$\hat{a}_{ii} = - \sum_{j=1, j \neq i}^N \hat{a}_{ji}. \quad (4.10)$$

PROOF. This follows from the preceeding discussion. $\square$

5. Robust statistics

5.1. Clark's gauge transformation

The linear Zakai Eq. (4.1) can be transformed into a simpler stochastic differential than the one derived, one that has significant numerical benefits. For $\ell = 1, \dots, N$, write

$$\begin{aligned} q_t(HX)^\ell &= \langle q_t(HX), \mathbf{e}_\ell \rangle, \\ U_t^\ell &= \int_{(0,t]} \int_{\mathfrak{R}} \left(\frac{1}{L_\ell^\epsilon(y)} - 1 \right) \{ \mu(dy; ds) - k_\ell(y) dy ds \}, \\ \lambda_t^\ell &= 1 + \int_{(0,t]} \lambda_{s-}^\ell dU_s^\ell \\ &= E_t(U^\ell). \end{aligned}$$

Write $\Gamma_t = \text{diag}(\lambda_t^\ell)$ and $\bar{q}_t = \Gamma_t q_t$. Then, we have the following result.

THEOREM 5.1 (Robust filters).

$$\begin{aligned} \bar{q}_t(HX) &= \bar{q}_0(HX) + \int_0^t \{ \bar{q}_s(HAX + \alpha X) \\ &\quad + \sum_{i,j=1}^N \Gamma_s a_{ji} (\mathbf{e}_j - \mathbf{e}_i) \mathbf{e}_i' \Gamma_s^{-1} \bar{q}_s((\beta^j - \beta^i)X) \} ds \\ &\quad + \int_0^t \int_{\mathfrak{R}} \bar{q}_{s-}(\xi(y)X) \mu(dy; ds). \end{aligned}$$

PROOF. According to Eq. (4.1), we have

$$\begin{aligned} \bar{q}_t(HX)^\ell &\equiv \lambda_s^\ell q_t(HX)^\ell \\ &= \bar{q}_0(HX)^\ell + \int_0^t \left\{ \bar{q}_s(HAX + \alpha X)^\ell \right. \\ &\quad + \sum_{i,j=1}^N \lambda_s^\ell a_{ji} q_s((\beta^j - \beta^i)X)^i (\mathbf{e}_j - \mathbf{e}_i)' \mathbf{e}_\ell \left. \right\} ds \\ &\quad + \int_0^t \int_{\mathfrak{R}} \bar{q}_{s-}(\xi(y)X)^\ell L_\ell^\epsilon(y) \mu(dy; ds) \end{aligned}$$

$$\begin{aligned}
& + \int_0^t \bar{q}_{s-} (HX)^\ell \int_{\mathfrak{R}} (L_\ell^\epsilon(y) - 1) \{\mu(dy; ds) - k(y) dy ds\} \\
& + \int_0^t \bar{q}_{s-} (HX)^\ell dU_s^\ell \\
& + \sum_{0 < s \leq t} \left(\bar{q}_{s-} (\xi(\Delta Y) X)^\ell L_\ell^\epsilon(\Delta Y_s) + \bar{q}_{s-} (HX)^\ell (L_\ell^\epsilon(\Delta Y_s) - 1) \right) \left(\frac{1}{L_\ell^\epsilon(\Delta Y_s)} - 1 \right) \\
& = \bar{q}_0 (HX)^\ell + \int_0^t \left\{ \bar{q}_s (HAX + \alpha X)^\ell + \sum_{i,j=1}^N \lambda_s^\ell a_{ji} q_s ((\beta^j - \beta^i) X)^i (\mathbf{e}_j - \mathbf{e}_i)' \mathbf{e}_\ell \right\} ds \\
& + \int_0^t \int_{\mathfrak{R}} \bar{q}_{s-} (\xi(y) X)^\ell \mu(dy; ds).
\end{aligned}$$

The result follows. $\square$

REMARK 5.1. The mapping $q \mapsto \bar{q}$ is often referred to as a gauge transformation and was first used to calculate robust filters by CLARK [1978].

5.2. Parameter estimation

This section calculates the recursive equations for the parameter updates. As we are using VG dynamics, some of the filters are new and have a different form.

5.2.1. State estimates

Set $\alpha = \xi = 0$, $\beta = 0 \in \mathfrak{R}^N$ and $H_t = H_0 = 1$. Then, we obtain the following ordinary differential equation:

$$\begin{aligned}
\bar{q}_t(X)^\ell &= \bar{q}_0(X)^\ell + \int_0^t \bar{q}_s(AX)^\ell ds \\
&= \bar{q}_0(X)^\ell + \int_0^t \mathbf{e}_\ell' \Gamma_s A \Gamma_s^{-1} \bar{q}_s(X) ds.
\end{aligned}$$

Consequently,

$$\bar{q}_t(X) = \bar{q}_0(X) + \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(X) ds. \quad (5.1)$$

To calculate the normalized probability, we need only calculate

$$E_\theta[X \mid \mathcal{Y}_T] = \frac{\bar{q}_t(X)}{\bar{q}_t(X)' \mathbf{1}_N}.$$

5.2.2. Occupation time of the state process

Write $H_t = O_t^j$ for the occupation time of the hidden Markov chain in state j up to time t . Write $\mathcal{O}_t^j = E_\theta[\mathcal{O}_t^j | \mathcal{Y}_t]$ for the corresponding estimate. Also, set $H_0 = 0$, $\alpha_s = \langle X_{s-}, \mathbf{e}_j \rangle$, $\beta = 0 \in \mathbb{R}^N$, and $\xi = 0$. Then,

$$\begin{aligned} \bar{q}_t(O^j X)^\ell &= \int_0^t \bar{q}_s(O^j A X + \langle X_{s-}, \mathbf{e}_j \rangle X)^\ell ds \\ &= \int_0^t \mathbf{e}_\ell' \Gamma_s A \Gamma_s^{-1} \bar{q}_s(O^j X) + \mathbf{e}_\ell' \mathbf{e}_j \bar{q}_s(O^j X) ds. \end{aligned}$$

Consequently,

$$\bar{q}_t(O^j X) = \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(O^j X) + \mathbf{e}_j \bar{q}_s(X) ds. \quad (5.2)$$

5.2.3. Number of jumps from state i to state j

Write $H_t = J_t^{ij}$ for the number of jumps made by the state process up to time t . Write $\mathcal{J}_t^{ij} = E_\theta[J_t^{ij} | \mathcal{Y}_t]$ for the corresponding estimate. According to Eq. (4.3), set $H_0 = 0$, $\alpha_s = a_{ji} \langle X_{s-}, \mathbf{e}_i \rangle$, $\beta = \langle X_{s-}, \mathbf{e}_i \rangle \mathbf{e}_j$, and $\xi(y) = 0$. Then,

$$\begin{aligned} \bar{q}_t(J^{ij} X) &= \int_0^t \bar{q}_s(J^{ij} A X + a_{ji} \langle X_{s-}, \mathbf{e}_i \rangle X) ds \\ &\quad + \int_0^t \Gamma_s a_{ji} (\mathbf{e}_j - \mathbf{e}_i) \mathbf{e}_i' \bar{q}_s(X) ds. \end{aligned}$$

Consequently,

$$\bar{q}_t(J^{ij} X) = \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(J^{ij} X) ds + \int_0^t \Gamma_s a_{ji} \mathbf{e}_j \mathbf{e}_i' \Gamma_s^{-1} \bar{q}_s(X) ds. \quad (5.3)$$

5.2.4. Cumulative size of positive jumps in the observation process

Write $H_t = P_t^j$ for the accumulated size of positive jumps made by the observations of size greater than ϵ when the state of the hidden Markov chain is in state j up to time t . Write $\mathcal{P}_t^j = E_\theta[P_t^j | \mathcal{Y}_t]$ for the corresponding estimate. According to Eq. (4.3), set $H_0 = 0$, $\alpha_s = 0$, $\beta = 0 \in \mathbb{R}^N$, and $\xi(y) = \langle X_{s-}, \mathbf{e}_j \rangle y I_{\{y \geq \epsilon\}}$. Then,

$$\begin{aligned} \bar{q}_t(P^j X)^\ell &= \int_0^t \bar{q}_s(P^j A X)^\ell ds \\ &\quad + \int_0^t \bar{q}_{s-}(\langle X_{s-}, \mathbf{e}_j \rangle X)^\ell \int_\epsilon^\infty y \mu(dy; ds). \end{aligned}$$

Consequently,

$$\bar{q}_t(P^j X) = \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(P^j X) ds + \int_0^t \bar{q}_{s-}(X) \int_{\epsilon}^{\infty} y \mu(dy; ds) \mathbf{e}_j. \quad (5.4)$$

5.2.5. Cumulative size of negative jumps in the observation process

Write $H_t = Q_t^j$ for the accumulated size of negative jumps made by the observations of size greater than ϵ when the state of the hidden Markov chain is in state j up to time t . Write $\mathcal{Q}_t^j = E_{\theta}[Q_t^j | \mathcal{Y}_t]$ for the corresponding estimate. According to Eq. (4.4), set $H_0 = 0$, $\alpha_s = 0$, $\beta = 0 \in \mathfrak{R}^N$, and $\xi(y) = \langle X_{s-}, \mathbf{e}_j \rangle y I_{\{y \leq -\epsilon\}}$. Then,

$$\begin{aligned} \bar{q}_t(Q^j X)^{\ell} &= \int_0^t \bar{q}_s(Q^j A X)^{\ell} ds \\ &\quad + \int_0^t \bar{q}_{s-}(\langle X_{s-}, \mathbf{e}_j \rangle X)^{\ell} \int_{-\infty}^{-\epsilon} y \mu(dy; ds). \end{aligned}$$

Consequently,

$$\bar{q}_t(Q^j X) = \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(Q^j X) ds + \int_0^t \bar{q}_{s-}(X) \int_{-\infty}^{-\epsilon} y \mu(dy; ds) \mathbf{e}_j. \quad (5.5)$$

5.2.6. Number of jumps in the observation process

Write $H_t = R_t^j$ for the number of jumps made by the observations of size greater than ϵ when the state of the hidden Markov chain is in state j up to time t . Write $\mathcal{R}_t^j = E_{\theta}[R_t^j | \mathcal{Y}_t]$ for the corresponding estimate. According to Eq. (4.5), set $H_0 = 0$, $\alpha_s = 0$, $\beta = 0 \in \mathfrak{R}^N$, and $\xi(y) = \langle X_{s-}, \mathbf{e}_j \rangle I_{\{|y| \geq \epsilon\}}$. Then,

$$\begin{aligned} \bar{q}_t(R^j X)^{\ell} &= \int_0^t \bar{q}_s(R^j A X)^{\ell} ds \\ &\quad + \int_0^t \bar{q}_{s-}(\langle X_{s-}, \mathbf{e}_j \rangle X)^{\ell} \mu(|y| \geq \epsilon; ds). \end{aligned}$$

Consequently,

$$\bar{q}_t(R^j X) = \int_0^t \Gamma_s A \Gamma_s^{-1} \bar{q}_s(R^j X) ds + \int_0^t \bar{q}_{s-}(X) \mu(|y| \geq \epsilon; ds) \mathbf{e}_j. \quad (5.6)$$

5.3. Discretization

Now that we have closed-form solutions for the parameter estimates and the recursive formulae to calculate them, we wish to test how well they perform. We begin with observations recorded on the interval $[0, T]$. For some integer N , we consider the partition $\mathcal{P}_N = \{0 = t_0, t_1, \dots, t_N = T\}$, where t_k is an increasing sequence of times and $\tau_k =$

$t_k - t_{k-1}$. The simple Euler scheme is used to approximate the recursion of state estimates in Eq. (5.1) by

$$\begin{aligned} q_0(X) &= \pi \in \mathfrak{R}^N, \\ \Gamma_{t_{k+1}} q_{t_{k+1}}(X) &= \Gamma_{t_k} q_{t_k}(X) + \int_{t_k}^{t_{k+1}} \Gamma_s A q_s(X) ds, \\ &\approx \Gamma_{t_k} q_{t_k}(X) + \tau_{k+1} \Gamma_{t_k} A q_{t_k}(X), \\ \text{or } q_{t_{k+1}}(X) &= \Gamma_{t_{k+1}}^{-1} \Gamma_{t_k} (I_N + \tau A) q_{t_k}(X). \end{aligned}$$

Here, we have written I_N as the $N \times N$ identity matrix. For small values of τ , the quantity $(I_N + \tau A)$ approximates a transition matrix for a discrete-time Markov process, an object we shall call $\tilde{A}$. We shall denote the diagonal matrix $\Gamma_{t_{k+1}}^{-1} \Gamma_{t_k}$ by $\tilde{B}_k$. $\tilde{B}_k$ has the following simplification:

$$\begin{aligned} \tilde{B}_k^{j,j} &\equiv (\Gamma_{t_{k+1}}^{-1} \Gamma_{t_k})^{j,j} \\ &= \exp \left(\int_{t_k}^{t_{k+1}} \int_{\mathfrak{N}} \left(\frac{1}{L_j^\epsilon(y)} - 1 \right) k_j(y) dy ds \right. \\ &\quad \left. + \int_{(t_k, t_{k+1}]} \int_{\mathfrak{N}} \log(L_j^\epsilon(y)) \mu(dy; ds) \right). \end{aligned}$$

We note that

$$\begin{aligned} &\int_{t_k}^{t_{k+1}} \int_{\mathfrak{N}} \left(\frac{1}{L_j^\epsilon(y)} - 1 \right) k_j(y) dy ds \\ &= \tau_{k+1} \left(\int_{\epsilon}^{\infty} + \int_{-\infty}^{-\epsilon} \right) k(y) - k_j(y) dy \\ &= \tau_{k+1} (E_1(\sqrt{2}\epsilon) - C_j E_1(M_j \epsilon) + E_1(\sqrt{2}\epsilon) - C_j E_1(G_j \epsilon)) \\ &= \tau_{k+1} (2E_1(\sqrt{2}\epsilon) - C_j (E_1(M_j \epsilon) + E_1(G_j \epsilon))). \end{aligned}$$

We can approximate this quantity to arbitrary accuracy using a truncation of the series expansion of the exponential integral (which was defined in Theorem 4.2)

$$E_1(x) = -\gamma - \ln(x) - \sum_{n=1}^{\infty} \frac{(-x)^n}{nn!}.$$

Here, we have used Euler's constant $\gamma = 0.5772156649 \dots$. Also,

$$\begin{aligned}
 & \int_{t_k}^{t_{k+1}} \int_{\mathfrak{N}} \log(L_j^\epsilon(y)) \mu(dy; ds) \\
 &= - \int_{t_k}^{t_{k+1}} \int_{\epsilon}^{\infty} (M_j - \sqrt{2}) y \mu(dy; ds) - \int_{t_k}^{t_{k+1}} \int_{-\infty}^{-\epsilon} (G_j - \sqrt{2}) |y| \mu(dy; ds) \\
 & \quad + \ln C_j \mu(|\Delta Y_s| > \epsilon, s \in (t_k, t_{k+1}]) \\
 &= -(M_j - \sqrt{2}) \sum_{s \in (t_k, t_{k+1}]} \Delta Y_s I_{\{\Delta Y_s > \epsilon\}} - (G_j - \sqrt{2}) \sum_{s \in (t_k, t_{k+1}]} \Delta Y_s I_{\{\Delta Y_s < -\epsilon\}} \\
 & \quad + \ln C_j \mu(|\Delta Y_s| > \epsilon, s \in (t_k, t_{k+1}]).
 \end{aligned}$$

We obtain

$$\begin{aligned}
 \tilde{B}_k^{j,j} &\equiv (\Gamma_{t_{k+1}}^{-1} \Gamma_{t_k})^{j,j} \\
 &= \exp\left(\tau_{k+1} \left(E_1(\sqrt{2}\epsilon) - C_j E_1(M_j \epsilon) + E_1(\sqrt{2}\epsilon) - C_j E_1(G_j \epsilon)\right)\right) \\
 & \quad \times C_j^{\sum_{t_k < s \leq t_{k+1}} I_{\{|\Delta Y_s| > \epsilon\}}} \\
 & \quad \times \exp\left(- (M_j - \sqrt{2}) \sum_{t_k < s \leq t_{k+1}} \Delta Y_s I_{\{\Delta Y_s > \epsilon\}}\right) \\
 & \quad \times \exp\left(- (G_j - \sqrt{2}) \sum_{t_k < s \leq t_{k+1}} \Delta Y_s I_{\{\Delta Y_s < -\epsilon\}}\right).
 \end{aligned}$$

Similarly, we may approximate the recursive densities in Section 5 by the following formulae.

State estimates

$$\begin{aligned}
 q_0(X) &= \pi \in \mathfrak{N}^N, \\
 q_{t_{k+1}}(X) &\approx \tilde{B}_k \tilde{A} q_{t_k}(X).
 \end{aligned}$$

Occupation time

$$\begin{aligned}
 q_0(O^j X) &= \mathbf{0} \in \mathfrak{N}^N, \\
 q_{t_{k+1}}(O^j X) &\approx \tilde{B}_k \tilde{A} q_{t_k}(O^j X) + \tau_{k+1} \tilde{B}_k \mathbf{e}_j \mathbf{e}_j' q_{t_k}(X).
 \end{aligned}$$

Number of jumps from state i to state j , $i \neq j$

$$\begin{aligned}
 q_0(J^{ij} X) &= \mathbf{0} \in \mathfrak{N}^N, \\
 q_{t_{k+1}}(J^{ij} X) &\approx \tilde{B}_k \tilde{A} q_{t_k}(J^{ij} X) + \tau_{k+1} a_{ji} \tilde{B}_k \mathbf{e}_j \mathbf{e}_i' q_{t_k}(X).
 \end{aligned}$$

Cumulative size of positive jumps in the observation process

$$q_0(P^j X) = \mathbf{0} \in \mathbb{R}^N,$$

$$q_{t_{k+1}}(P^j X) \approx \tilde{B}_k \tilde{A} q_{t_k}(P^j X) + \tilde{B}_k \mathbf{e}_j \mathbf{e}_j' q_{t_k}(X) \times \sum_{t_k < s \leq t_{k+1}} \Delta Y_s I_{\{\Delta Y_s > \epsilon\}}.$$

Cumulative size of negative jumps in the observation process

$$q_0(Q^j X) = \mathbf{0} \in \mathbb{R}^N,$$

$$q_{t_{k+1}}(Q^j X) \approx \tilde{B}_k \tilde{A} q_{t_k}(Q^j X) + \tilde{B}_k \mathbf{e}_j \mathbf{e}_j' q_{t_k}(X) \times \sum_{t_k < s \leq t_{k+1}} \Delta Y_s I_{\{\Delta Y_s < -\epsilon\}}.$$

Number of jumps in the observation process

$$q_0(R^j X) = \mathbf{0} \in \mathbb{R}^N,$$

$$q_{t_{k+1}}(R^j X) \approx \tilde{B}_k \tilde{A} q_{t_k}(R^j X) + \tilde{B}_k \mathbf{e}_j \mathbf{e}_j' q_{t_k}(X) \times \sum_{t_k < s \leq t_{k+1}} I_{\{|\Delta Y_s| \geq \epsilon\}}.$$

6. Simulation

The simulation we use the discretization where the $t_k = T_k^\epsilon$ are, in fact, the jump times of the observation process, Y . In this case, we have $\sum_{t_k < s \leq t_{k+1}} \Delta Y_s I_{\{\Delta Y_s > \epsilon\}} = \Delta Y_{t_{k+1}} I_{\{\Delta Y_{t_{k+1}} > \epsilon\}}$ and $\sum_{t_k < s \leq t_{k+1}} I_{\{|\Delta Y_s| \geq \epsilon\}} = I_{\{|\Delta Y_{t_{k+1}}| \geq \epsilon\}}$. We can then calculate the recursive equations given above, whereupon we substitute these “structural equations” into the new parameter estimates calculated in Eqs. (4.6).

We draw the observations from a simulated process with parameters given by

$$A = \begin{bmatrix} -0.025 & 0.015 \\ 0.025 & -0.015 \end{bmatrix}, \quad \theta = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0.05 \\ 0.15 \end{bmatrix}, \quad \text{and } v = \begin{bmatrix} 5.0 \\ 3.0 \end{bmatrix}.$$

We then use an initial guess of the EM algorithm given by

$$A^{(0)} = \begin{bmatrix} -0.2 & 0.1 \\ 0.2 & -0.1 \end{bmatrix}, \quad \theta^{(0)} = \begin{bmatrix} -0.05 \\ 0.25 \end{bmatrix}, \quad \sigma^{(0)} = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}, \quad \text{and } v^{(0)} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}.$$

The results for the state estimation are shown in Fig. 6.1, while the convergence of the parameters is shown in Fig. 6.2.

The filter works reasonably well. Future research might include the following. Firstly, one would like to know how the value of ϵ affects the parameter estimation. Ideally, parameter estimation should be independent of this number. As ϵ increases, fewer jumps will be recorded, and the times between jumps will increase. The time discretization then (either fixed step length or random) becomes critical. Secondly, numerical maximum likelihood estimates would help show if the bias among estimates shown in Fig. 6.2 is a result of the maximum likelihood scheme presented in this chapter. Moreover, the accuracy of the estimates produced would be quantified by the estimates of the information matrix and the resulting standard errors.

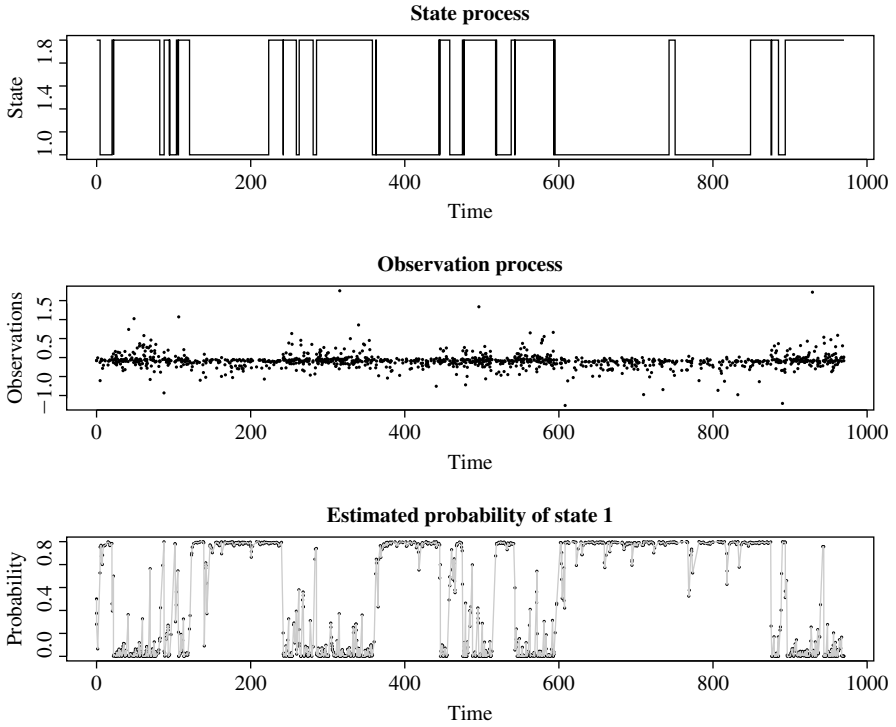


FIG. 6.1 State estimates from filtering the VG process. The upper pane shows the (hidden) state process, $\{X_t\}$. The middle pane indicates observations, $\{Y_t\}$, while the dotted gray line in the bottom pane indicates the estimated state process, $P^{(50)}(X_t = \mathbf{e}_1 | \mathcal{Y}_t) = \frac{\langle q_t(\mathcal{X}_t), \mathbf{e}_1 \rangle}{\langle q_t(\mathcal{X}_t), \mathbf{1}_2 \rangle}$. Note that the estimates of the probability are using the estimated parameters after the 50th pass of the EM algorithm.

7. Statistical results for the S&P500

In this section, some empirical results are presented. Using the estimation scheme presented in this chapter, we find the estimated point estimates for a two-state model given by $(A^{(50)}, \theta^{(50)}, \sigma^{(50)}, \text{ and } \nu^{(50)})$, where

$$A^{(50)} = \begin{bmatrix} -0.2582563 & 0.3234474 \\ 0.2582563 & -0.3234474 \end{bmatrix},$$

$$\theta^{(50)} = \begin{bmatrix} 0.14348039 \\ 0.06117406 \end{bmatrix},$$

$$\nu^{(50)} = \begin{bmatrix} 0.04702351 \\ 0.05296345 \end{bmatrix},$$

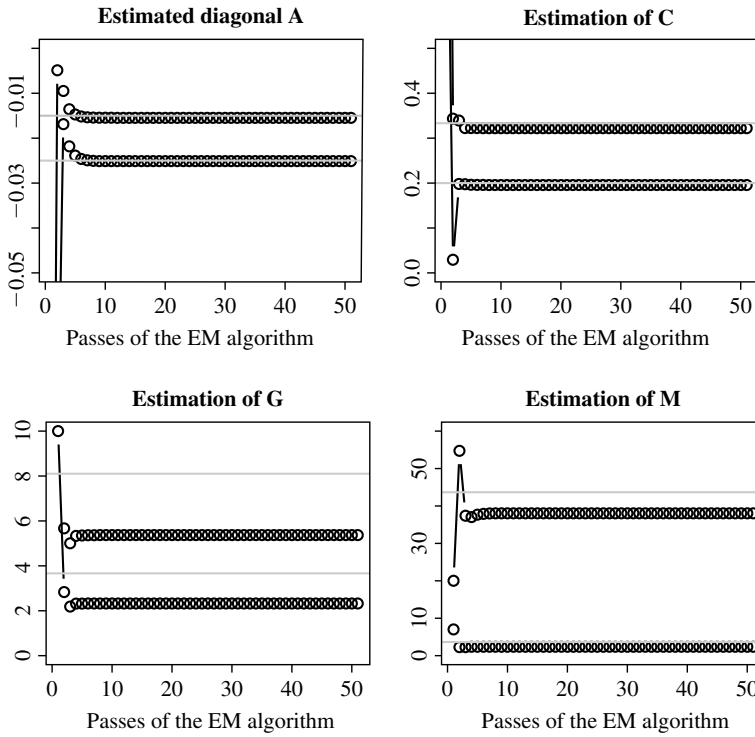


FIG. 6.2 Estimates of the parameters of the VG process. The gray lines in each picture indicate the true values of the parameters.

$$\text{and } \sigma^{(50)} = \begin{bmatrix} 0.1992016 \\ 0.3913205 \end{bmatrix}.$$

A diagram of the working filter for X is shown in Fig. (6.3). The filter clearly shows the switching occurring between two main regimes. State 1 appears to be prominent during the mid-1990s and in the last year, while state 2 is much more in evidence during times of high volatility and market crashes. Note that our parameter estimates correspond to a fairly reasonable (approximate) conditional yearly standard deviation of

$$\sqrt{(\theta^{(50)})^2 \nu^{(50)} + (\sigma^{(50)})^2} = \begin{bmatrix} 0.2016168 \\ 0.3915737 \end{bmatrix}.$$

This shows that we have managed to decompose the time series into two states: one with high volatility (the second state) and one with low volatility. This is only a guide for yearly volatility as observations were not taken at regular intervals.

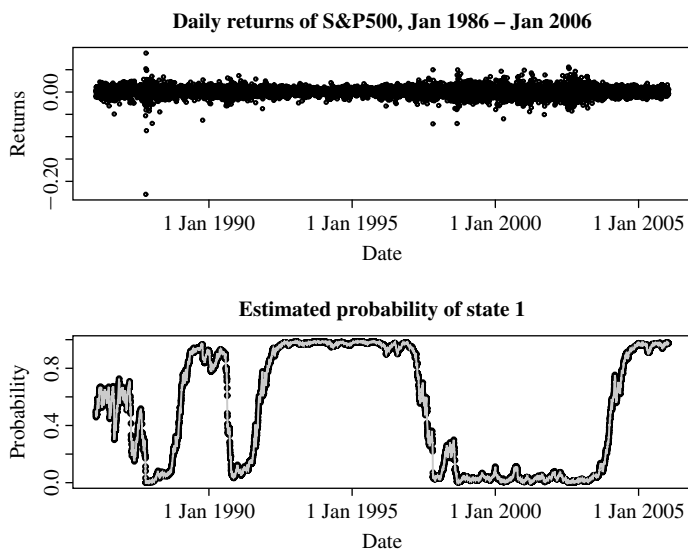


FIG. 6.3 Daily observations of the S&P500 (top plot) with filtered state estimates (bottom plot).

8. Conclusion

Pricing of options using this filtering methodology now can use these estimates shown in this model combined with the approach shown by ELLIOTT and OSAKWE [2006]. That is, we used the estimates given and substituted them into the Fourier transform and then inverted to give the density. However, this is outside the scope of this chapter and is left to future study. See CARR and MADAN [1999].

References

- ABRAMOWITZ, M., STEGUN I.E. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Table* (Dover).
- BROWN, J.W., CHURCHILL, R.V. (1996). *Complex Variables and Applications*, sixth ed. (McGraw-Hill, New York, NY).
- CARR, P., MADAN, D. (1999). Option valuation using the fast Fourier transform. *J. of Comput Financ.* **2**, 61–73.
- CLARK, J.M.C. (1978). The design of robust approximation to the stochastic differential equations of non-linear filtering. In: SKWIRZYNSKI, J.K. (ed.), *Communication Systems and Random Process Theory: Number 25 in Applied Science* (Sijthoff and Noordhoff, Netherlands), pp. 721–734.
- ELLIOTT, R.J., AGGOUN, L., MOORE, J.B. (1995). *Hidden Markov Models: Estimation and Control* (Springer, New York, NY).
- ELLIOTT, R.J., OSAKWE, C-J.U. (2006). Option pricing for pure jump processes with Markov switching compensators. *Financ. Stoch.* **10**, 250–275.
- JACOD, J., SHIRYAEV, A.N. (1987) *Limit Theorems for Stochastic Process* (Springer-Verlag, Berlin, Germany).
- SATO, KEN-ITI (2004). *Lévy Process and Infinitely Divisible Distributions* (Cambridge University Press, Cambridge).
- SHIRYAEV, A.N. (1999). *Essentials of Stochastic Finance: Facts, Models, Theory* (World Scientific, New Jersey, NJ).

Index

- abandonment option, 559–561
- ABAST, 515–523
- abstract markets, 149–164
 - classes of, 149
 - definition, 149
- ADD. *See* average-percentage drawdown
- adjoint equation, 486–490
- adjusted BAST (ABAST), 515–523
- admissible portfolios, 581
- algebraic multigrid methods, 427
- American barrier options, 666
 - numerical solutions for, 677
- American basket options, 437–449
 - algorithms, 440–449
 - exercise region, 439
 - variational inequality, 438–439
- American “in” barrier options, 671–673
- American lookback options, 666, 674
- American option price, 289–294, 374
 - algorithm, 290–292
 - approximations, 170, 183, 185
 - dual formulations of, 171–175
 - high-biased estimate of, 172, 173, 177, 180
 - low-biased estimate of, 172–174, 177
 - price errors in, 173–175
 - primal formulations of, 171–175
 - variance errors in, 173
- American “out” barrier options, 668–671
- American put option price, 175, 182, 183
 - parameters of one-dimensional, 177, 184
- analytical approximation solution, 666, 677–683
- approximate replication property, 317
- arbitrage
 - relative to market, 139–141, 154
 - short-horizon, 122–125
 - sufficient intrinsic volatility leading to, 137–142
- arbitrary payoff functions, 667
- arbitrary time horizons, relative arbitrage on, 125
- Archimedean axiom, 37–39
- Asian call payoff, 635
- Asian option pricing, 635–640
- asset price, 12
 - dynamics, 57
 - first fundamental theorem of, 650–651
 - model for, 686
 - process, 77
- assets, 8, 20
 - contingent claims on
 - market, 537–538
 - nontradable, 539–540
 - DOC on, 519–521
 - growth rates, 105
 - primary, 76
 - risky, 21, 24
 - tradable. *See* tradable assets
 - traded financial, 79
 - of volatility process, 11
- asymptotic expansion, projected hedging
 - martingales from, 182–184
- Atlas model, 157, 158, 162
- automatic differentiation (AD)
 - application of, 474
 - of computer programs, 471–473
 - sensitivity by, 474
- AVaR. *See* average value at risk
- Avellaneda–Friedman–Holmes–Samperi’s approach, 11
- average-percentage drawdown (ADD), 191
 - definition, 198
 - dominant strategy under, 215–217
 - long-term asymptotic behavior, 214, 215
 - for MMV strategy, 214–217
 - portfolio frontier, 215
 - for PUM strategy, 214–217
 - for SPM strategy, 214–217

- average value at risk (AVaR), 77
 - coherent risk measure, 36
- axioms, 38
 - Archimedean, 37–39
 - of full and weak certainty independence, 41
 - independence, 37, 39, 41
 - of uncertainty aversion, formulating, 42
- backward stochastic differential equations (BSDE)
 - techniques, 58, 66–68
- barrier options, 666, 667
 - American “in”, 671–673
 - American “out”, 668–671
 - European “in”, 672
 - European “out”, 668
- barriers
 - case of two, 566–567
 - derivation of sensitivity to, 510–512
 - discrete, price of, 513
 - at high output, 564–565
 - reflecting, 564–567
- barrier shifting, 505–517
 - derivation of sensitivity to barriers, 510–512
 - monitoring frequency, influence of, 506–510
- barrier shifting techniques (BAST), 513–515
- barrier-style option price, 172
- Basel II framework, 68
- Basket Lock Active Coupon (BLAC) down out option, 522–523
- BAST, 513–515
- Bayes’ formula, 330, 697
- Bellman equation
 - of dynamic programming, 312, 321–322
 - framework, 319
 - functional iteration, 319
- below-mean semi variance (SV), 191, 198, 199
 - definition, 196
 - dominant strategy under, 205
 - long-term asymptotic behavior, 200, 201
 - for MMV strategy, 200–206
 - portfolio frontier, 202
 - for PUM strategy, 200–206
 - for SPM strategy, 200–206
- Bessel process, 150–155
 - Lamperti representation, 151
 - time-changed and scaled squared, 133
- biased-price estimator, 170
- BICG-stab (stabilized biconjugate gradient method), 421
- binomial distribution, 195, 204
- binomial model, 301, 303, 306, 310
- binomial tree methods, 506
 - Birkhoff ergodic theorem, 151
 - Bismut–Elworthy formula, delta of European options using, 285–289
 - Bismut weight computation, 288, 289
 - BLAC down out option, 522–523
 - Black–Scholes assumptions, 377
 - Black–Scholes dynamics, 610, 611
 - Black–Scholes equation, 364, 668, 669, 671, 674, 678
 - weak formulation of, 383–384
 - Black–Scholes formula, 129, 131
 - Black–Scholes framework, 685
 - Black–Scholes market
 - multidimensional, 77, 78
 - risky asset in, 74, 75
 - Black–Scholes model, 182, 183, 373, 377, 378, 403, 437, 499, 506, 515, 516
 - limitation of, 475–477
 - martingale properties for, 476
 - Black–Scholes type, 276, 280
 - bond price process, 22
 - Borel function, 597, 599
 - Borel probability measures, 36
 - boundary conditions, 538
 - Dirichlet, 463
 - Brennan–Schwartz algorithm, 467
 - Brownian bridge techniques, 499–505, 514
 - approximations, 502–505
 - easy extensions, 500–502
 - with simulation of trigger, 500
 - toy example, 499–500
 - without simulation of trigger, 499–500
 - Brownian diffusion, 630
 - for SDE, 632–633
 - Brownian dimension, of stochastic model, explicative, 8–10
 - Brownian filtration, 115, 116, 131
 - Brownian motion, 71, 77, 114, 192, 194, 229, 361, 362, 366, 575, 587, 595, 602, 615–617, 649, 652, 658
 - arithmetic, 513
 - drifted, 510, 512
 - functional quantization of
 - optimized, 621
 - product, 626
 - geometric, 74, 79–81, 126, 502, 519, 521–523
 - independent scalar, 158
 - multidimensional, 633
 - one-dimensional, 133, 150, 151
 - rate optimal quantization of, 630
 - scalar, 509
 - time-changed, 687
 - stable processes as, 659

- Brownian motion model, 8, 513
- BSDE techniques, 58, 66–68
- budget equation, 534
- càglàd processes, 579
- calibration, 475–490
 - of local volatility, 477–478
 - methods, in finance, 11, 12
- canonical Euclidean norm, 608
- capacity theory, least favorable measures and
 - relation to, 49–53
- capital asset pricing model
 - in continuous time framework, 320–322
 - in discrete time, 299
- capital at risk (CaR), 197
- capital constraints, 68, 78
- capital distribution curve, 156
 - estimation of parameters from, 163
 - stability of, 163, 164
 - steady-state, 161–164
- capital gains, 121, 137
- capital losses, 121, 137
- capital regulation, 68
- capital requirement, 32
 - impact of, 69, 75
- capitalization
 - of market, 100, 122, 150, 151, 156
 - ranked, 159, 160
 - relative, 101
 - stock, 157
- cash invariance, 32
- Cauchy–Dirichlet boundary conditions, 507
- Cauchy–Dirichlet PDE, 511
- Cauchy–Schwarz inequality, 124, 358, 359
- CCAPM. *See* consumer capital asset pricing model
- C2D (continuous to discrete) approximation, 514, 518
- center for research in securities prices (CRSP), 137, 139
- CGMY process with stochastic volatility, 660–661
- Chicago Board of Trade (CBOT), 653
- chi-square distribution, 164
- CIR. *See* Cox–Ingersoll–Ross
- claims, contingent. *See* contingent claims
- Clark’s gauge transformation, 701–702
- CLVQ. *See* competitive learning vector quantization procedure
- coherent monetary utility functional, 32
- coherent risk measure. *See* convex risk measure
- combination technique, 435–437
- commodity market, 80
- comonotonic independence, 42
- comonotonicity, 50
- companion parameters, 604
- competitive learning vector quantization (CLVQ)
 - procedure, 619
 - competitive phase, 620
 - cooperative phase, 620
- complete market, projection techniques for
 - coherent utility functionals in, 45–49
- computer programs
 - automatic differentiation (AD) of, 471–473
 - class double, 473
- concave monetary utility functional, 34, 39, 40, 59, 60
 - law-invariant, 36
 - minimal penalty function, 32
- conditional probability, 301
- conditional value at risk (CVaR), 191
 - definition, 197, 211
 - dominant strategy under, 212–214
 - long-term asymptotic behavior of, 211
 - for MMV strategy, 209–214
 - portfolio frontier under, 212
 - for PUM strategy, 209–214
 - for SPM strategy, 209–214
 - versus VaR, 212–214
- conjugate gradient algorithm, 479, 480
- constant relative risk aversion (CRRA) utility, 77, 80
 - functions, 56, 61
- constructive functional quantization, 630
- consumer capital asset pricing model (CCAPM), 533
 - risk premium and, 533–535
- consumption rate process, 79
- contingent claims
 - American pricing, 131
 - futures valuation, 540–541
 - hedging price of, 130
 - on market assets, 537–538
 - on nontradable asset, 539–540
 - seller of, 127
 - on tradable assets, 538–539
 - valuation of, 537, 540–541
- continuous up and out put (UOP) option, 517
- continuous-time portfolio strategies. *See also specific types*
 - downside ratios comparison, 204
- control problems
 - approximation of, 334
 - error analysis, 334
 - filtering, 329–331
 - partial observation discrete time, 326, 328, 329
 - quantization of, numerical approximation by, 331–334
- control process, 326, 328
- convex conjugate function, 54

- convex functional, 605
- convex risk measure, 31, 32
- cooperative phase, 620
- cost function, 19, 328, 330, 334
- covariance matrix, 277
 - relative, 102
- covariance process, 104
 - properties of, relative, 102–106
- Cox-Ingersoll-Ross (CIR)
 - equation, 365
 - process, 450
 - stochastic volatility factor, 182
- Cramer–Rao lower bounds, 5, 6
- Crank–Nicolson scheme, 447, 448
- CRRA. *See* constant relative risk aversion utility
- CRSP, 137, 139
- Cubature formulae, 604
 - conditional expectation, 604
 - convex functional, 605
 - Lipschitz functionals, 604–605
- cumulative size
 - of negative jumps, 704, 707
 - of positive jumps, 703–704, 707
- curse of dimensionality, 427
 - versus* R-R extrapolation, 609–612
- CVaR. *See* conditional value at risk

- D2D (discrete to discrete) approximation, 52, 515
- delta approximation
 - errors in, 179
 - for European option prices, 179
- descriptive theory. *See* stochastic portfolio theory (SPT)
- deterministic equation, 264–267
- differentiable functionals, with lipschitz
 - differentials, 605–606, 644–645
- differential operators, 258–261, 269, 277
- diffusions, 5
 - approximation of quantiles of, 14–18
 - multidimensional Markov, 15
 - one-dimensional, 364–366
 - recombining binomial tree approximations for, 361–367
 - two-dimensional, 366, 367
- Dirac distribution, 283
- Dirichlet conditions, 416
 - homogeneous, 443
- discretization, 418–419, 467
 - of Dirichlet problem, 430
 - finite-difference operator for, 432
 - finite-element, 439, 464
 - Galerkin, 425
 - sparse grid, 433
 - time, 463
- discretization error, 16, 341, 358, 398, 430
- distorsion. *See* quantization error
- distribution of firms, 570
- diverse market model, 126, 127
- diversity weighting, 153, 154
- diversity-weighted index of large stocks
 - leakage in, 145, 146
- dividend rate processes, 80
 - exogenous, 79
- DOC. *See* down and out call
- Doleans–Dade exponential, 63
- Doob–Meyer decomposition, 171, 173, 176, 180
- Doob’s maximal inequality, 174
- Doss–Sussman approach, 631
- down and out call (DOC), 497, 514–518, 670, 671
 - on basket, 521–522
 - discrete, 515, 516–518
 - in geometric Brownian motion model, 513
 - on three assets, 519–521
- downside ratio, 202–204
- downside risk measures, 189, 198, 199. *See also*
 - below-mean semi variance; conditional value at risk; value at risk
- drawdown risk measures, 189. *See also*
 - average-percentage drawdown; maximum-percentage drawdown
- drift process, 136
 - cumulative, 137, 138
- dual formulations, of American option prices, 171–175
- dual value function, 56, 57
 - of robust utility maximization problem, 62, 67
- duality formula, 287
- duality techniques
 - advantage of, 53
 - in incomplete markets, 53–57
- duality theory. *See* duality techniques
- Dubins–Schwarz theorem, 652
- Dunford–Pettis theorem, 34
- Dupire’s equation, 388, 480–485
 - numerical results, 484
- dynamic programming algorithm, 338
 - for filtered control problem, 331
- dynamic programming approach
 - derivative of value function, 312–314
 - setting, 311
- dynamic programming equation, 312
- dynamic risk measurement, 76

- economic agent, 36, 37
 - optimal investment problem for, 42
 - regulated, 80
 - to solve maximization problem, 74
 - unregulated, 79

- El Karoui–Hounkpatin’s method, 11
- EM algorithm, 693, 694, 698, 709
- EMM. *See* equivalent martingale measure
- entropic monetary utility functional, 34, 35
 - maximization of, 45
- entropic penalty function, 59
- entropy function, 137
- equilibrium model, 568–571
- equilibrium price, 79
- equity markets, 93, 122
 - cumulative excess growth of U.S., 139, 140
 - SPT for, 93
- equivalent martingale measure (EMM)
 - completeness without, 128–130
 - hedging and optimization without, 127, 128
 - non existence of, 115
 - utility maximization in absence of, 131–133
- erroneous stopping rules, 25
- error indicators
 - for fully discrete problem, 397–398
 - for semidiscrete problem in time, 395–397
 - upper bounds for, 399–400
- estimators
 - defining scale invariant, 10
 - maximum likelihood, 5
 - price gap and variance of biased, 170, 178
- Euler scheme, 439, 502, 503, 518
 - defined, 15
 - implicit, 395, 418, 424, 439, 443
 - first-order, 424, 463
 - second-order, 448
 - for time discretization, 463
- European barrier option, 668
 - numerical solutions for, 677
- European basket options, 403–413
 - change of variables, 405–408
 - consequences of maximum principle, 410–411
 - localization, 414–415
 - error, 415–416
 - numerical methods for, 414–437
 - variational formulation, 411–413
- European call option, 130, 670
 - hedging price of, 130, 131
 - quadratic hedging of, 345–347
 - shortfall hedging of, 347–348
 - on stock, 130
- European in barrier option, 672
- European lookback options, 673–674
- European option hedging
 - in partially observed stochastic volatility model, 335–337
 - quadratic criterion
 - call option, 345–347
 - put option, 337–343
 - shortfall risk criterion
 - call option, 347–348
 - put option, 343–345
- European options, delta of
 - call option with payoff, 280
 - digital option with payoff, 280
 - sensitivity computations of, 273, 285
 - Bismut–Elworthy formula for, 285–289
 - integration by parts formula for, 273–285
 - using geometrical model
 - finite-difference method, 281
 - Malliavin calculus, 281
 - using Vasicek model
 - finite-difference method, 280, 283, 284
 - Malliavin calculus, 280, 283, 284
- European options, with stochastic volatility, 451–466
- European payoffs, 609
- European put option
 - price
 - counterpart, 176–178
 - under partial and complete observation, 341, 342
 - quadratic hedging of, 337–343
 - shortfall hedging of, 343–345
- European up-and-out put options, 670
- European vanilla call option, 377, 380, 386
- exogenous model, for price dynamics, 580
- expectation maximization (EM) algorithm, 693, 694, 698, 709
- expected return rate, 193, 198, 199
- expected shortfall, 36, 77
- exponential law, 268
- exponential likelihood ratio process, 23
- fast-diffusion equation, 234
- Fatou property, 33, 44, 45
- Fatou’s lemma, 109
- FEM. *See* finite element method
- filter process, 330, 337
- finance, calibration methods in, 11, 12
- financial market, 68, 76, 80, 95–100
 - Basel II framework on, 68
 - diversity for, 111–113
 - modelling, 30, 69
 - stability of, 68
- financial market model, 192
- finite-difference method, 274, 280
 - delta of European options using, 281, 282
 - on sparse grids, 431
 - versus Malliavin estimator, 284
- finite-difference operator, for discretization, 432
- finite dimensional example, 479–480
- finite-element discretization, 439, 464

- finite element method (FEM), 374, 388–394, 417, 439–440
 - C++ implementation, 390–394
 - discrete problem in matrix form, 419–424
 - discretization, 418–419
 - numerical implementation of, 421
 - time semidiscrete problem, 418
- first stochastic dominance (FSD) order, 192
- forward anticipating calculus, 576–579
- forward integrals, 576
 - definition of, 577
 - Itô formulae for, 577, 579
- forward performance process, 228
 - CARA and CRRA, 245–252
 - construction of, 236
 - definition of, 230
- forward process, 578
- forward stochastic calculus, 584
- fractional Brownian motion, 617
- Frechet bounds, 505
- FSD order, 192
- Fubini's theorem, 616
- full certainty independence, 41, 42
- functional quantization, 599
 - lower bounds for, 642–643
 - optimized, 621
 - product, 626
- functionally generated portfolios, 135–147
- fundamental theorem, asset pricing of, 650–651
- Galerkin discretizations, 425
- Galerkin methods, 425, 429–431
 - sparse, 429–431
- Galerkin projection, 426
- gamma processes, difference of two, 687–688
- Gårding's inequality, 395
- Gaussian function, 380
- Gaussian process
 - centered, 617–618
 - optimal quadratic functional quantization of, 615
- Gaussian random walk, numerical approximation of, 526–528
- Gaussian white noise, 336
- g-divergence, 35
 - minimizing, 46
- geometric multigrid method, 424
- Girsanov theorem, 114, 511, 586
- GMRES (generalized minimum residual method), 421
- gradient methods, 421, 478
 - with Armijo rule, 478
 - conjugate, 479
 - principle of, 478
- growth optimality, 154
- growth-optimal portfolio, 108, 109
 - numéraire property of, 109
- Halmos–Savage theorem, 46, 55
- Hamilton–Jacobi–Bellman–Issacs equation, 19, 20
- HARA utility function. *See* hyperbolic absolute risk aversion utility function
- hedging error, maximizing, 176
- hedging martingales, 170–172, 175–179, 182–183
 - from asymptotic expansion, 182–184
- hedging supermartingales, 179–181
- Heston stochastic volatility model, 635–640
 - up and out put in, 517–519
- Heston's model, 182
 - initial value problem for, 458–461
 - numerical method for pricing options with, 465–466
- high-biased estimates, of American option price, 172, 173, 177, 180, 184
- high-biased estimator, Monte Carlo estimator for, 176
- Hilbert space, 412, 596
 - properties of, 455, 459
- Hilbert-valued random variable, 595
- HJB equation, 63
- homogeneous Dirichlet conditions, 443
- honest agent, 574
- horizon-unbiased utilities, 230
- horse race lotteries, 38
- Huber–Strassen theorem, 49, 50
- hybrid model, 158
- hybrid quantization, 643
- hyperbolic absolute risk aversion (HARA) utility function, 57, 61, 74, 77
- IMA. *See* internal model approach
- incentives, uncertainties due to, 556–559
- independence axiom, 37
- insider trading, aspects of, 574
- integer-valued random measure, 692
- integration by parts formula, 261–263
 - for computing conditional expectations, 270
 - for computing Delta of European options, 273–285
- for pure jump diffusion processes, 263–273
 - using
 - jump amplitudes, 267–272
 - jump times, 272, 273
- interest-rate process, 130
 - for money market, 96
- internal model approach (IMA), 78
- intrinsic volatility, leading to arbitrage, 137–142
- invest option, valuation of an, 546

- investment
 - assumptions, 554–555
 - cost, 556
 - uncertainties on, 554
 - valuation of option, 555–556
- investment performance measurement
 - benchmark and market view processes, 231
 - under local risk tolerance, 228–252
- investor, 77
 - preferences, 36
 - robust utility functional of, 44
 - VaR, 75
- iterative methods, 421, 426
 - convergence of, 423
 - for solving linear systems, 423
- Itô formulae, for forward integrals, 577, 579
- Itô integral, 578
- Itô processes, 8, 14, 77, 96, 228
 - d -dimensional, 76
- Ito's lemma, 172, 182, 186
- Itô's rule, 108, 116, 119, 142
- Jensen's inequality, 130, 605
- jump amplitudes
 - integration by parts formula using, 267–272
 - Malliavin estimators computation using, 277
- jump times
 - integration by parts formula using, 272, 273
 - Malliavin estimators computation using, 278
- K-L expansion, 615
- Knightian uncertainty, 37
- knock-in option, 498
- knockout option, 497, 498
 - price of, 512
- Kohonen algorithm, 333
- Kolmogorov equation, 570
- Lagrange finite elements, 418–419
- Lagrange multiplier, 309, 312, 318, 320
- Laplace operator, sparse discretization of, 434
- large insider, optimal portfolio problem for, 579–586
- large trader, visible effect of, 576
- law invariant, 34
 - concave monetary utility functional, 36
- least favorable measure, 45, 46, 49–53
- least square problem, 485
 - numerical results, 486
- least squares method, 175, 177, 178, 184
- least squares regression (LSR), 615
- Lebesgue dominated convergence theorem, 601
- Lebesgue measure, 257, 268, 272
- Lévy density, 658–660
- Lévy driven spot price, 476
- Lévy process, 476, 575, 577, 652, 661, 688–690
- Lévy–Khinchine formula, 476, 688, 690
- linear systems, iterative methods for solving, 423
- linear Zakai equation, 696, 697
- Lipschitz differentials, differentiable functionals
 - with, 605–606, 644–645
- Lipschitz functionals, 604–605
- Lloyd I procedure, 620
- local risk tolerance, 232–234
 - asymptotically linear, 234–240
 - and differential input surfaces, 237–239
 - investment performance measurement under, 228–252
 - with power, logarithmic, and exponential utilities, 235
- local times, estimation of, 145
- localization function, 274, 276
- log-concave density function, 602
- log-concavity assumption, 602
- lognormal process, 450, 469
- Lookback options, 666, 673
 - American, 674–677
 - European, 673–674
- lotteries, 36, 42
 - monetary character of, 37
- low-biased estimates
 - of American option price, 172, 173, 177, 184
 - variance of, 175
- low-biased estimator, 170, 171
- LSR, 615
- Malliavin calculus
 - delta of European options using, 281, 282
 - differential operators, 258–261
 - framework, 257, 258
 - integration by parts formula. *See* integration by parts formula
- Malliavin covariance matrix, 17, 259, 273
- Malliavin derivative, 259, 274
- Malliavin estimators
 - computation, 277–279
 - localized and nonlocalized, 280
 - using jump amplitudes, 277, 284, 285
 - using jump times, 278, 284, 285
 - variance of, 283, 284
 - versus* finite-difference method, 284
- Malliavin Monte Carlo estimators, 280
- Malliavin weight, 279
- market, 95–100
 - activity, 652–654
 - assets, contingent claims on, 537–538
 - beating, 116
 - capitalization of, 100, 122

- market (*continued*)
 - coherent, 106
 - completeness, 302
 - duality techniques in incomplete, 53
 - equity, 93
 - financial, 68, 80, 95–100
 - intrinsic volatility of, 104
 - portfolio, 100, 101, 125
 - shortest time to beat, 116
 - uniformly weakly diverse on, 111, 112, 120, 129
 - volatility, as drivers of growth, 105, 106
- market diversity, 111–113
 - cumulative change, 121
 - leading to arbitrage, 118–122
 - measure of, 119, 120
 - on time horizon, 111, 126
- market equilibrium models, 78–81
- market models, 5, 43, 66
 - coherent, 101
 - complete, 31, 43, 45, 49
 - diverse, 126
 - financial, 58
 - incomplete, 57
 - multidimensional, 14
 - volatility-stabilized, 149–155
- market portfolio, 100, 101, 119
 - excess growth rate of, 104, 138, 154
 - shorting, 155
- market weight, 135
 - equal, 141, 154
 - process, 101
 - rank for, 142–144
 - smooth functions of, 135
- market's intrinsic volatility, 104
- Markov chain, 290, 326, 328, 333, 336, 685, 703
- Markov inequality, 359
- Markov models, 13
- Markov–Feller process, 13
- Markovian framework
 - martingale properties, 316
 - setting, 314–316
- Markovian Itô driving process, 451
- Markovian markets, model risk P&Ls for
 - misspecified hedging strategies in, 13, 14
- Markovian process, 182
- Markowitz model, 533
- martingale
 - from asymptotic expansion, 182–184
 - controls, effects of, 179
 - properties, 318
 - for Black–Scholes model, 476
 - Markovian framework, 316
 - optimal portfolio and consumption, 308
 - of stochastic integrals, 177
 - theory, 100, 650
- mass lumping, 424
- master equation, 146, 147
- master formula, 135, 136
- maximum-percentage drawdown (MDD), 191
 - definition, 198
 - dominant strategy under, 218
 - long-term asymptotic behavior of, 217, 218
 - for MMV strategy, 217, 218
 - portfolio frontier under, 218
 - for PUM strategy, 217, 218
 - for SPM strategy, 217, 218
- MC estimator, 609
- MC simulation, 609, 614, 643
- MDD. *See* maximum-percentage drawdown
- MDH, 660
- mean regularity, 641–642
- mean-risk approach, 191
- Merton's model, 517
- mesh adaptivity, 394–401, 424
 - computation of bounds, 400–401
 - error indicators for
 - fully discrete problem, 397–398
 - semidiscrete problem in time, 395–397
 - upper bounds, 399–400
 - multiple meshes and integration, 394–395
- minimal penalty function, 55
 - of concave monetary utility functional, 32–35, 55
- Minkowski's inequality, 358, 359
- mirror portfolios, 122–125
- mixture of normal distribution hypothesis (MDH), 660
- MMV. *See* modified mean-variance
- model ambiguity, 37
- model risk, 20–25
 - control of, 20, 25, 26
- P&Ls
 - function, 14
 - for misspecified hedging strategies in Markovian markets, 13
- stochastic game to face, 19, 20
- value function, 19
- model uncertainty, 37, 39, 66, 71
 - reduction of, 39
- modeling issues, 3
- modified mean-variance (MMV), 192, 193–194
 - analytical formula, 200, 206, 210
 - average-percentage drawdown for, 214–217
 - below-mean semi variance for, 200–206
 - conditional value at risk for, 209–214
 - correlation among risk measures for, 220
 - downside ratio, 202–204
 - final wealth distributions for, 203

- in long-term investment, 200, 201, 207, 208, 211
- maximum-percentage drawdown for, 217, 218
- value at risk for, 206–209
- monetary risk measures, 31–42
- monetary utility functionals, 31–34
 - coherent, 36, 45
 - concave, 34, 40
 - with minimal penalty function, 32
 - entropic, 34, 35
 - examples of, 34–36
- money market, 80, 95
 - account, 76, 79
 - interest-rate process for, 96
 - investment in, 99
- monotonicity, 32, 37, 39, 40, 659
- Monroe's theorem, 653
- Monte Carlo (MC) algorithm, 274
- Monte Carlo (MC) error, 17
- Monte Carlo (MC) estimator, 609
 - for high-biased, 176
 - for low-biased, 176
- Monte Carlo (MC) simulation, 12, 15, 25, 170, 178, 182, 214, 219, 330, 333, 609, 614, 643
 - of VaR of model risk P&Ls, 13–18
- mothballing option, 561–563
- multidimensional Brownian motion, 633
- multidimensional market models, 14
- multigrid method, 448
 - algebraic, 427
 - geometric, 424
- multiscale stochastic volatility models, 179, 181, 182
- net present value (NPV) approach, 546, 552–554
- Neumann conditions, 416
- New York stock Exchange (NYSE), 653
- NIG process, 659–660
- nodal basis, 420, 425
- noise, testing of, 6–8
- non-log-concave density function, 602
- nonsatiable investor, optimal strategy for, 192
- nonsatiable risk-averse investor, optimal strategy for, 192
- nontradable asset, contingent claims on, 539–540
- nontrading time, 653
- normal inverse Gaussian (NIG) process, 659–660
- Novikov condition, 362
- Novikov's theorem, 60
- NPV approach, 546, 552–554
- N -quantizers, 598, 603
- numéraire-invariance property, 103, 104, 119
- numéraire property, 132
- numerical error, 507
- numerical integration (II), (III), 609–612, 634
- numerical methods
 - classes of, 374
 - for European basket options, 414–437
- numerical optimization, of quadratic functional quantization, 618–619
- numerical representation, robust preferences and their, 36–42
- numerical tests, 517–523
- NYSE, 653
- observation process
 - negative jumps in, 704, 707
 - number of jumps, 704, 706, 707
 - positive jumps in, 703–704, 707
- occupation time, 703, 706
- ODE, 630
- operator splitting scheme, 447
- optimal asset allocations, 242
- optimal benchmarked portfolio, 233
- optimal consumption process, 321
- optimal investment problem, 42
 - formulation of, 43–45
- optimal investment strategy, 232, 233
- optimal portfolio allocation strategy, 23, 25
- optimal portfolio and consumption
 - martingale properties, 308
 - optimality conditions, 309
 - setting of problem, 307
- optimal portfolio problem, 575, 576
 - for large insider, 579–586
- optimal product quantization, 628
- optimal quadratic functional quantization, of
 - Gaussian processes, 615
- optimal quadratic quantization, 595, 599
 - uniqueness of, 602
- optimal stopping rule, 175
- optimal stopping time, 175, 176, 180
- optimal trading strategy, 77, 78
- optimal vector quantization, application of, 609
- optimal wealth, 240, 243, 318, 321
- optimization problem
 - dynamic robust, reduction of, 72
 - static, 72
- optimized functional quantization, of Brownian motion, 621
- optimized quantization, 627
 - versus* (optimal) product quantization, 640–641
- option pricing, 373, 435, 665
- optional sampling theorem, 171, 172
- ordinary differential equation (ODE), 630
- Orstein–Uhlenbeck (OU) operator, 260
- Orstein–Uhlenbeck (OU) process, 182, 365, 450, 631

- overshoot, 507, 509, 516
 - definition of, 506
 - role of, 506
- parabolic VI, 549–552
- parameter estimation, 698–701, 707
 - from capital distribution curve, 163
- Pareto distributions, 603
- partial differential equation (PDE), 18, 20, 374, 377–388, 404, 507, 511, 538, 543
 - changes of variables, 378–379
 - classical solutions, 381
 - convexity, 386–388
 - quasilinear, 66
 - variational framework, 382–388
 - weak solutions
 - maximum principle for, 385–386
 - regularity of, 384–385
- partial equilibrium models, 69
- partial integrodifferential equations (PIDE), 374
- path-dependent options, 666, 673
- payoff, 42–44, 497, 521
- payoff functions, 377, 403, 404, 410, 411, 435, 506, 668, 677
 - arbitrary, 667
 - integrable, 182
- PDE. *See* partial differential equation
- penalty approximation, 551
- penalty function, 43
 - minimal, 55
 - of concave monetary utility functional, 32–35
- penalty methods, 445–446
- PIDE, 374
- Pierce lemma, 607, 609
- P&Ls. *See* profit and losses
- Poisson point measure, 263
- Poisson process, 263
- Poisson random measures, 575–577
 - compensated, 576, 577
- portfolio, 95–100
 - defining, 124
 - diversification, 105, 106
 - diversity-weighted, 119, 145
 - simulation of, 137
 - entropy-weighted, 137, 138
 - equally weighted, 122, 154, 155
 - excess growth rate, 98, 103, 141
 - functionally generated, 135–147
 - to generated trading strategy, 100
 - growth rate, 98, 105, 106, 108, 141
 - insurance, general equilibrium models of, 81
 - large and small stock, 144
 - long-only, 97, 105–107, 125
 - creating, 124
 - market, 100, 101, 119
 - mirror, 122–125
 - optimization, 106–109
 - relative performance, 143, 144, 146, 155
 - relative return of arbitrary, 103
 - seed, 124
 - simulation of diversity-weighted, 137
 - variance, minimization of, 106
 - wealth processes of, 135
 - weight of, 120, 141
- portfolio choice, 29
 - problem of, 29, 30, 49, 71
 - robust, 42–68
 - under robust constraints, 68–81
- portfolio frontier
 - average-percentage drawdown, 215
 - under below-mean semi variance, 202
 - under conditional value at risk, 212
 - under maximum-percentage drawdown, 218
 - under value at risk, 208, 209
- portfolio-generating functions, 135–147
- posteriori error, 418
- power utility maximization (PUM), 192, 196
 - analytical formula, 200, 207, 211
 - average-percentage drawdown, 214–217
 - below-mean semi variance for, 200–206
 - conditional value at risk for, 209–214
 - correlation among risk measures for, 220
 - downside ratio, 202–204
 - final wealth distributions for, 203
 - in long-term investment, 200, 201, 207, 208, 211
 - maximum-percentage drawdown for, 217, 218
 - value at risk for, 206–209
- predictable compensator, 690–692
- preference order, 37
- price
 - equilibrium, 79
 - error, in American option price estimation, 174
 - regulation on, 78
 - of swing options, 612
 - upper hedging, 128–131
- price dynamics, 8, 579
 - asset, 57
 - exogenous model for, 580
- price process, 6, 79, 476
 - asset, 77
 - continuous and discontinuous, 7, 8
 - determination, in equilibrium, 80
 - exogenous, 69
 - modelling of discounted of assets, 43
 - primary security, 69, 76
 - of stock, 70
- prices of states of nature, 302, 303

- pricing function, 378, 476
- primal formulations, of American option prices, 171
- primal value function, 62
- primal-dual active set algorithm, 442–445
- primal-dual methods, 441
- primary assets, 13, 76
- probability distribution, 300, 301
- probability measures, 30, 42, 43, 49, 50, 114
 - Borel, 36
 - subjective, 70, 72–74
- probability transition matrix, 333, 336
- product functional quantization, 622
 - of Brownian motion, 626
 - quantization rate, 624–625
- product quantization, optimized quantization *versus*, 640–641
- product quantizer, 624–625
 - for numerical computations, 625–627
- profit and losses (P&Ls), 31, 32
 - function, 14
 - Monte Carlo approximations of VaR model risk, 13
- Proinov's theorem, 609
- project valuation, 541
 - death possibility, 545
 - probabilistic interpretation, 545
 - valuation equation, 542–545
- projected SOR (PSOR) algorithm, 440–441
- projection schemes, 446–448
- projection techniques, for coherent utility
 - functionals in complete market, 45–49
- proportionality factor, 322
- pseudoreplicating portfolio, self-financed, 14
- PSOR algorithm, 440–441
- PUM. *See* power utility maximization
- pure jump diffusion processes, integration by parts
 - formula for, 263–273
- put-call parity, 129, 386
- QMC method, 609
- quadratic approximation, 606–607, 669
- quadratic criterion, 106
- quadratic estimation error, 6
- quadratic functional quantization
 - numerical optimization of, 618–619
 - optimal *versus* product, 627–630
 - toolbox for, 619–622
- quadratic hedging
 - of European call option, 345–347
 - of European put option, 337–343
 - versus* shortfall hedging, 345, 348
- quadratic quantization, 597, 599
- quantization, 337, 595
 - of control problems, 331–334
 - optimal quadratic, 595
 - optimal vector, 609
 - optimized, 640–641
 - of pair filter observation, 332
 - product, 640–641
 - of random variables, 603–694
 - quantization error, 327, 332, 334, 598, 608
 - quantization rate, 624–625
 - and mean regularity, 641–642
 - universal, 641
 - quantization tree, 613
 - quantization-based algorithms, 609
 - quasi-Monte Carlo (QMC) method, 609
- Radon–Nikodym densities, 72
- Radon–Nikodym derivative, 50, 304, 317, 698, 699
- ramifications, 130, 131
- random variables, quantization of, 603–604
- Rank, for market weights, 142–144
- rank-based models, 155–164. *See also* volatility-stabilized market models
- rate optimal quantization, 630
- rate process
 - consumption, 79
 - exogenous dividend, 79
- real market, 20
- real options theory, 531
- real-valued function, 381
- recombining binomial tree model, approximations
 - for diffusions, 361–367
- reference probability measure, 692
- reflecting barriers, 564–567
- regime-switching, 685
- regularity assumptions, 407
- regulatory rules, for financial institutions, 68
- relative arbitrage
 - on arbitrary time horizons, 125
 - and consequences, 113–118
- relative return process, of stock, 102
- renewal theory, 509
- replication portfolio, 304–307
- Richardson–Romberg (R-R) extrapolation, 609–610, 612, 634
 - versus* curse of dimensionality, 609–612
- risk aversion. *See* preference order
- risk constraints
 - capital at, 77
 - downside, 74, 78
 - economic impact of, 71
 - semidynamic, 75–77, 81
 - shortfall, 71

- risk constraints (*continued*)
 - static, 69–75, 78–81
 - VaR, 74, 75
- risk measure, 53, 68–69, 76, 77, 189. *See also specific types*
 - convex, 31, 32
 - definition of, 32
 - correlations between different, 219, 220
 - defined, 70
 - downside, 71
 - dynamic, 76
 - minimizing, 81
 - monetary, 31–34
 - static, 76
 - transfer of, 81
 - VaR, 80
 - variance, 190
- risk premium, 533
 - and CCAPM, 533–535
- risk tolerance process, 232, 244
- risk-averse investor, optimal strategy for, 191
- risk-aversion parameter, 63
- risk-free asset, 475
- riskless asset, 373
- riskless asset price, 229, 336
- risk-neutral probability, 303, 304, 317
- risk-sensitive control, 45
- risky asset, 21, 24, 373, 476
 - in Black-Scholes market, 74–75
 - defined, 58
- risky asset price, 192, 193, 228, 229, 328, 329, 336
- risky securities, 228
- robust constraints, portfolio choice under, 68–81
- robust preferences, and numerical representation, 36–42
- robust problem, 47, 54, 56
- robust statistical test theory, 45
- Robust statistics, 701–707
- robust utility functional, 57
 - of investor, 44
- robust utility maximization problem, 47, 49, 57
 - dual-value function of, 62, 67
 - reduction of, 46
 - value function of, 61, 65
- robust value function, 55, 56
- Rogers' dual formulation, 170
- Romberg extrapolation techniques, 15
- Rothschild–Stiglitz (R-S) stochastic dominance order, 191
- R-R extrapolation. *See* Richardson–Romberg extrapolation
- Runge–Kutta scheme, 636
- scale invariant estimator, 10
- scale-invariance property, diversity function of, 146
- Scott model, 468
- SDE. *See* stochastic differential equation
- second stochastic dominance (SSD) order, 192
- semimartingale decomposition, 109
- sensitivity, 471–474
 - by automatic differentiation, 474
- shifting barrier. *See* barrier shifting
- shortfall hedging
 - of European call option, 347–348
 - of European put option, 343–345
 - versus* quadratic hedging, 345, 348
- shortfall probability minimization (SPM), 192, 195
 - analytical formula, 200, 206, 210
 - average-percentage drawdown for, 214–217
 - below-mean semi variance for, 200–206
 - conditional value at risk for, 209–214
 - correlation among risk measures for, 220
 - downside ratio, 202–204
 - final wealth distributions for, 203
 - in long-term investment, 200, 201, 207, 208, 211
 - maximum-percentage drawdown for, 217, 218
 - value at risk for, 206–209
- short-horizon arbitrage, 122–125
- signal-to-noise ratio, 107
- single-agent models. *See* partial equilibrium models
- Skorohod integral, 258, 259, 578
- Skorohod problem, 160
- Snell's envelop process, 173
- Sobolev norms, 382–383
- SOR. *See* successive over relaxation
- S&P500
 - daily returns of, 710
 - statistical results, 708–710
- space discretization, 333
- sparse discretization, of Laplace operator, 434
- sparse finite-difference discretization, 432
- sparse methods, 427
 - applications to option pricing, 435
 - combination technique for, 435–437
 - discretization, 433–435
 - Galerkin method, 429–431
 - grid-based method, 431–435, 469
 - notations and preliminary results, 427–429
 - principle of, 427
- SPM. *See* shortfall probability minimization
- SPT. *See* stochastic portfolio theory
- square-integrability property, 127
- SSD order, 192

- standard differential calculus, 261
- standard errors
 - of low-biased estimates, for American option prices, 177
 - reduction of, 178
- standard utility maximization problem, 46, 48
- stationary quantization, 602
- stationary VI, 549
- statistical arbitrage, 122
- statistical models, explicative brownian dimension of, 8–10
- statistical procedures, limitations of, 5–10
- Stein–Stein’s model, 458
 - initial value problem for, 454
 - numerical method for pricing options with, 463
- stochastic control techniques, 66, 67
 - solution with, 57–66
- stochastic differential, 320
- stochastic differential equation (SDE), 15–17, 58, 79, 233, 240, 361, 377, 498
 - weak solution of, 362, 363
- stochastic equation, 96
- stochastic equity market models, simple. *See* abstract markets
- stochastic game, to face model risk, 19, 20
- stochastic integrals, martingale property of, 177
- stochastic models, 5
- stochastic optimization, 619
- stochastic portfolio theory (SPT), 89
 - for equity markets, 93
 - for financial assets, 93
- stochastic process, 307, 556, 686
- stochastic time and jump processes, 657–659
- stochastic variables, volatility models with several, 468–470
- stochastic variational calculus. *See* Malliavin calculus
- stochastic volatility, 181–184, 449–470
 - American options with, 466–468
 - CGMY process with, 660–661
 - European options with, 451–466
 - and information arrival, 654–657
- stochastic volatility model, 183, 374, 451, 475–476
 - parameters used in the two-factor, 184
- stock(s), 22, 95, 96, 128, 137
 - asymptotic growth-rate of, 158
 - buying-low-and-selling-high of, 122
 - capitalization of, 120, 130, 157
 - European call option on, 130
 - fluctuation of individual, 150
 - growth rate of, 96, 112, 152
 - price process of, 70
 - relative return process, 102
 - variances of, 103, 104
 - volatility, reduction of, 80
- stock index, cap-weighted large, 137
- stock market, 80
 - weights of individual, 104
- stock option, 127
 - upper hedging price of, 130
- stock price, 24, 100, 131
 - deflated, 115, 116
 - process, logarithmic representation, 96
 - risky, 175
- Stratanovich Stochastic Differential Equations (SDE), 632–633
- strict local martingales, 113–118
 - exponential, computing, 153
- strict local martingales galore, 115
- strict local supermartingales, 115, 118
- successive over relaxation (SOR), 424
 - algorithm, 440–441
- supermartingale
 - characterization, 180
 - control, 182
 - hedging, 179–181
 - properties, 129, 131, 186
- superreplication principle, 410
- Tanaka-type formula, 159
- technical analysis techniques, 20–25
- time-changed Brownian motion, 687
 - stable processes as, 659
- time changes, 651–661
 - origins of, 651–652
- time discretization, Euler scheme for, 463
- time horizon, 100, 117
 - arbitrary, 125
 - increase of, 130, 131
 - long, 141, 142
 - market diversity on, 111, 126
 - relative arbitrage on arbitrary, 125
 - weakly diverse over, 111, 118, 119
- time-rescaling process, 232, 233
- total consumption rate, 79
- tradable assets, 532–537
 - contingent claims on, 538–539
 - definition of, 535–536
 - dividends, 537
 - evolution of prices, 532
 - market model, 532–533
 - Markowitz model, 533
- trading strategy, 99
 - admissible, for initial capital, 100
 - generated by portfolio, 100
 - log-optimal, 132, 153
 - optimal, 77, 78

- trading time, 653
- transaction clock, 652–654
- translation invariance. *See* cash invariance
- translation property, 40, 74
- two-sided VI, 560
 - solution of, 560–561
- UBSR, 71
- uncertainty aversion, 38
- underlying asset, 377, 499, 517
 - price of, 451
- uniform boundedness condition, 98, 105, 114
- universal quantization rate, 641
- up-and-out put option, 670
- upper hedging price, 128–131
- utility-based shortfall risk (UBSR), 71
- utility functions. *See* constant relative risk aversion (CRRRA) utility
- valuation equation, 542–543
 - solutions, 543–545
- value at risk (VaR), 69, 77, 191, 198
 - conditional, 36
 - definition, 197, 207
 - dominant strategy under, 209, 210
 - investor, 75
 - long-term asymptotic behavior, 207, 208
 - for MMV strategy, 206–209
 - of model risk P&Ls, 13–18
 - portfolio frontier under, 208, 209
 - for PUM strategy, 206–209
 - risk constraint, 80, 81
 - risk management, 80
 - for SPM strategy, 206–209
 - versus* CVaR, 212–214
- value function, 19, 20, 22, 23, 56
 - of robust problem, 54
 - of robust utility maximization problem, 61, 64
- vanilla call option, 377, 386, 471, 475, 666
- vanilla put option, 383, 387, 463, 475, 666
- VaR. *See* value at risk
- variance, 190
 - of biased estimators, 170, 178
 - of Malliavin estimators, 283
 - for MMV strategy, 200
 - for PUM strategy, 200
 - for SPM strategy, 200
- variance-covariance process, 76, 78
- variance gamma (VG) dynamics, 686
- variational inequality (VI), 547, 562
 - parabolic, 549–552
 - solution, 562–564
 - stationary, 549
 - system of, 557
 - two-sided, 560
- Vasicek model, 276, 277, 280, 288, 366
 - delta of European options using, 280, 281, 288
- vector quantization, 599, 607
 - application of optimal, 609
 - rate, 607–609
- VG. *See* variance gamma (VG) dynamics
- VG process
 - absolutely continuous changes of measures, 692
 - estimates of parameters of, 709
 - Lévy processes, 688–690
 - with parameters, 694, 698
 - predictable compensator, 690–692
 - representation of, 687–692
 - time-changed Brownian motion, 687
- VI. *See* variational inequality
- volatility, 377, 378
 - indicators, 20
 - matrix, 116
 - calibration of, 8
- volatility models
 - local, 58, 475
 - with one stochastic process, 449–451
 - with several stochastic variables, 468–470
- volatility smile, 475
- volatility-stabilized market models, 149–155, 164
- von Neumann–Morgenstern theory, 36–39
- von Neumann–Morgenstern utility, 74, 77
- Voronoi cells, 598
- Voronoi tessellations, 332
- weak certainty independence, 38–40
- weak maximum principle, 385
- wealth
 - calculation of, 77
 - generated by portfolio, 125
 - investment, in asset, 76
 - robust expected utility of, 70
- wealth process, 70, 93, 132, 326, 329, 337
 - deflated, 115, 117
 - optimal, 132
 - discounted, 114
 - log-optimal, 132
 - portfolio, 70, 135
 - SDE of, 76
- Weierstrass intermediate value theorem, 208
- weighted Sobolev norms, 382–383
- Wiener processes, 314, 532
- Zador theorem, 335, 607, 609, 625, 642
- Zakai equation, linear, 696, 697
- zero growth rate, 157