MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 221–257

OPTIMAL INVESTMENT UNDER RELATIVE PERFORMANCE CONCERNS

GILLES-EDOUARD ESPINOSA AND NIZAR TOUZI

Centre de Mathématiques Appliquées, Ecole Polytechnique Paris

We consider the problem of optimal investment when agents take into account their relative performance by comparison to their peers. Given N interacting agents, we consider the following optimization problem for agent i, $1 \leq i \leq N$:

$$\sup_{\pi^i \in \mathcal{A}_i} \mathbb{E} U_i \left((1-\lambda_i) X_T^{\pi^i} + \lambda_i \left(X_T^{\pi^i} - \bar{X}_T^{i,\pi} \right) \right),$$

where U_i is the utility function of agent i, π^i his portfolio, X^{π^i} his wealth, $\bar{X}^{i,\pi}$ the average wealth of his peers, and λ_i is the parameter of relative interest for agent i. Together with some mild technical conditions, we assume that the portfolio of each agent i is restricted in some subset A_i. We show existence and uniqueness of a Nash equilibrium in the following situations:

- unconstrained agents,
- constrained agents with exponential utilities and Black–Scholes financial market.

We also investigate the limit when the number of agents N goes to infinity. Finally, when the constraints sets are vector spaces, we study the impact of the λ_is on the risk of the market.

KEY WORDS: portfolio optimization, relative concerns, Nash equilibrium, differential game, backward stochastic differential equations.

1. INTRODUCTION

The seminal papers of Merton (1969, 1971) generated a huge literature extending the optimal investment problem in various directions and using different techniques. We refer to Pliska (1986), Cox and Huang (1989), or Karatzas, Lehoczky, and Shreve (1987) for the complete market situation, to Cvitanic and Karatzas (1992) or Zariphopoulou (1994) for constrained portfolios; to Constantinides and Magill (1976), Davis and Norman (1990), Shreve and Soner (1994), Duffie and Sun (1990), or Akian, Menaldi, and Sulem (1995) for transactions costs; to Constantinides (1983), Jouini, Koehl, and Touzi (1997, 1999), Damon, Spatt, and Zhang (2001), or Ben Tahar, Soner, and Touzi (2008a, 2008b) for

The research is supported by the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, the Chair *Derivatives of the Future* sponsored by the Fédération Bancaire Française, and the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

We are grateful to Jérôme Lebuchoux for initiating this study as well as for his helpful comments and advice. We also thank Jean-Michel Lasry for his support and for many interesting discussions and remarks, Christoph Frei and Gonçalo dos Reis for many helpful discussions as well as pointing out some mistakes, and also, Ramon van Handel and Ronnie Sircar for some interesting discussions. Finally, we are grateful to two anonymous referees who helped to improve a previous version of the paper.

Manuscript received September 2011; final revision received December 2012.

Address correspondence to Gilles-Edouard Espinosa, Ecole Polytechnique-CMAP, Route de Saclay, Palaiseau 91128, France; e-mail: gilles-edouard.espinosa@polytechnique.org.

DOI: 10.1111/mafi.12034

taxes; and to He and Pearson (1991a,b), Karatzas, Lehoczky, Shreve, and Xu (1991), Kramkov and Schachermayer (1999, 2003), or Kramkov and Sirbu (2006) for general incomplete markets.

However, in all of these works, no interaction between agents is taken into account. The most natural framework to model such interaction would be a general equilibrium model where the behavior of the investors is coupled through the market equilibrium conditions. But this typically leads to untractable calculations. Instead, we shall model the interactions based on some simplified context of comparison of the performance to that of the competitors or to some benchmark. A return of 5% during a crisis is not equivalent to the same return during a financial bubble. Moreover, human beings tend to compare themselves to their peers. In fact, economic and sociological studies have emphasized the importance of relative concerns in human behaviors, see Veblen (1899) for the sociological part, and Abel (1990), Gali (1994), Gomez, Priestley, and Zapatero (2007), or DeMarzo, Kaniel, and Kremer (2008) for economic works, considering simple models in discrete time frameworks.

In this paper, we study the optimal investment problem under relative performance concerns, in a continuous-time framework. More precisely, there are N particular investors that compare themselves to each other. Agents are heterogeneous (different utility functions and different constraints sets) and instead of considering only his absolute wealth, each agent takes into account a convex combination of his wealth (with weight $1-\lambda$, $\lambda \in [0, 1]$) and the difference between his wealth and the average wealth of the other investors (with weight λ). This creates interactions between agents and therefore leads to a differential game with N players. We also consider that each agent's portfolio must stay in a set of constraints.

In the context of a complete market situation where all agents have access to the entire financial market, we prove existence and uniqueness of a Nash equilibrium for general utility functions. The optimal performances at equilibrium are explicit, and therefore allow for many interesting qualitative results.

We next turn to the case where the agents have different access to the financial market, i.e., their portfolio constraints sets are different. Our solution approach requires to restrict the utility functions to the exponential framework. Then, assuming mainly that the agents positions are constrained to lie in closed convex subsets, and that the drift and volatility of the log prices are deterministic, we show the existence and uniqueness of a Nash equilibrium, using the backward stochastic differential equation (BSDE) techniques introduced by El Karoui and Rouge (2000) and further developed by Hu, Imkeller, and Müller (2005). The Nash equilibrium optimal positions are more explicit in the case of constraints defined by linear subspaces. In this setting, we analyze the limit when the number of players N goes to infinity where the situation considerably simplifies in the spirit of mean field games, see Lasry and Lions (2007). Notice that our problem does not fit in the framework of Lasry and Lions (2007) for the two following reasons. First, in Lasry and Lions (2007), the authors consider similar agents, which is not the case in the present paper, as the utility functions, the parameters λ_is, and the sets of constraint can be specific. More importantly, in Lasry and Lions (2007), the sources of randomness of two different agents are independent.

We finally investigate the impact of the interaction coefficient λ. Under some additional assumptions, which are satisfied in many examples, we show that the local volatility of the wealth of each agent is nondecreasing with respect to λ. In other words, the more investors are concerned about each other (λ large), the more risky is the (equilibrium) portfolio of each investor. However, in general, this can fail to hold. But in the limit,

N goes to infinity, the same phenomenon holds for the average portfolio of the market, without any additional assumption. Roughly speaking, this means that the global risk of the market increases with λ, although it can fail for the portfolio of some specific agent.

This paper is organized as follows. Section 2 introduces the problem. In Section 3, we solve the complete market situation, for general utility functions. In Section 4, we deal with the general case with exponential utility functions and portfolios that are constrained to remain inside closed convex sets. In Section 5, we restrict the sets of constraints to linear spaces that allow us in particular to derive some interesting economic implications.

REMARK 1.1. This paper is the submitted version from the content of the PhD thesis of the first author Espinosa (2010). The delayed date of submission is due to our wish to provide more relevant examples. Based on the content of Espinosa (2010), Frei and dos Reis developed an interesting article, Frei and dos Reis (2011), during their Post-Doc at Ecole Polytechnique. Their article, in particular, highlights the difficulty in the existence and uniqueness of the quadratic multidimensional backward SDE of the present paper, and establishes the existence of a *sequentially delayed* Nash equilibrium in the general case. Frei and dos Reis (2011) also provide some extensions of our results, and some shorter proofs. But they crucially make use of many results of this paper.

Notations. $\mathbb{H}^2(\mathbb{R}^m)$ denotes the space of all predictable processes φ, with values in $\mathbb{R}^m$, and satisfying $\mathbb{E}\int_0^T |\varphi_t|^2 dt < \infty$. The corresponding localized space is denoted by $\mathbb{H}^2_{loc}(\mathbb{R}^m)$. When there is no risk of confusion, we simply write $\mathbb{H}^2$ and $\mathbb{H}^2_{loc}$.

2. PROBLEM FORMULATION

Let W be a d-dimensional Brownian motion on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote by $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ the corresponding completed canonical filtration. We assume that $\mathcal{F}$ is generated by W. Let $T > 0$ be the investment horizon so that $t \in [0, T]$. Given two $\mathbb{F}$-predictable processes θ taking values in $\mathbb{R}^d$ and σ taking values in $\mathbb{R}^{d\times d}$, satisfying:

$$\sigma \text{ symmetric, uniformly definite positive, } \int_0^T |\sigma_t|^2 dt < +\infty \text{ a.s.,} \tag{2.1}$$

$$\text{and } \theta \text{ is bounded, } dt \otimes d\mathbb{P}\text{−a.e,} \tag{2.2}$$

we consider a market with a nonrisky asset with interest rate $r = 0$ and a d-dimensional risky asset $S = (S^1, \ldots, S^d)$ given by the following dynamics:

$$dS_t = \operatorname{diag}(S_t)\sigma_t(\theta_t dt + dW_t), \tag{2.3}$$

where for $x \in \mathbb{R}^d$, $\operatorname{diag}(x)$ is the diagonal matrix with ith diagonal term equal to x^i.

A portfolio is an $\mathbb{F}$-predictable process $\{\pi_t, t \in [0, T]\}$ taking values in $\mathbb{R}^d$. Here, π_t^j is the amount invested in the jth risky asset at time t. Under the self-financing condition, the associated wealth process X_t^π is defined by

$$X_t^\pi = X_0 + \int_0^t \pi_r \cdot \operatorname{diag}(S_r)^{-1} dS_r,\ t \in [0, T].$$

Given an integer $N \geq 2$, we consider N portfolio managers whose preferences are characterized by a utility function $U_i : \mathbb{R} \to \mathbb{R}$, for each $i = 1, \ldots, N$. We assume that U_i is

C^1, increasing, strictly concave and satisfies Inada conditions:

$$U_i'(-\infty) = +\infty,\ U_i'(+\infty) = 0. \tag{2.4}$$

In addition, we assume that each investor is concerned about the average performance of his peers. Given the portfolio strategies π^i, $i = 1, \dots, N$, of the managers, we introduce the average performance viewed by agent i as

$$\bar{X}^{i,(\pi^j)_{j\neq i}} := \frac{1}{N-1}\sum_{j\neq i} X^{\pi^j}. \tag{2.5}$$

The portfolio optimization problem of the ith agent is then defined by

$$\begin{aligned} V_0^i\left((\pi^j)_{j\neq i}\right) := V_0^i &:= \sup_{\pi^i\in\mathcal{A}^i} \mathbb{E}\big[U_i\big((1-\lambda_i)X_T^{\pi^i} + \lambda_i\big(X_T^{\pi^i} - \bar{X}_T^{i,(\pi^j)_{j\neq i}}\big)\big)\big] \\ &= \sup_{\pi^i\in\mathcal{A}^i} \mathbb{E}\big[U_i\big(X_T^{\pi^i} - \lambda_i \bar{X}_T^{i,(\pi^j)_{j\neq i}}\big)\big],\ 1 \le i \le N, \end{aligned} \tag{2.6}$$

where $\lambda_i \in [0, 1]$ measures the sensitivity of agent i to the performance of his peers, and the set of admissible portfolios $\mathcal{A}^i$ will be defined later. Roughly speaking, we impose integrability conditions as well as the constraints π^i take values in A_i, a given closed convex subset of $\mathbb{R}^d$.

Our main interest is to find a Nash equilibrium in the context where each agent is "small" in the sense that his actions do not impact the market prices S.

DEFINITION 2.1. A Nash equilibrium for the N portfolio managers is an N-tuple $(\hat{\pi}^1, \dots, \hat{\pi}^N) \in \mathcal{A}^1 \times \dots \mathcal{A}^N$ such that, for every $i = 1, \dots, N$, given $(\hat{\pi}^j)_{j\neq i}$, the portfolio strategy $\hat{\pi}^i$ is a solution of the portfolio optimization problem $V_0^i((\hat{\pi}^j)_{j\neq i})$.

If in addition, for each $i = 1, \dots, N$, $\hat{\pi}^i$ is a deterministic and continuous function of $t \in [0, T]$, we say that $(\hat{\pi}^1, \dots, \hat{\pi}^N)$ is a deterministic Nash equilibrium.

Our main results are the following:

MAIN THEOREM 2.2. *Assume that for each $i = 1, \dots, N$, $U_i(x) = -e^{-\frac{x}{\eta_i}}$ for some constant $\eta_i > 0$, the portfolio constraints sets A_i are closed convex, $\prod_{i=1}^N \lambda_i < 1$, and that there exists a Nash equilibrium $(\tilde{\pi}^1, \dots, \tilde{\pi}^N)$. Then, there exists a solution to the N-dimensional quadratic BSDE* (4.20) *below, and $(\tilde{\pi}^1, \dots, \tilde{\pi}^N)$ can be expressed in terms of this solution.*

Unfortunately, the well-posedness of BSDE (4.20) below is an open problem in the present literature, thus preventing Main Theorem 2.2 from providing a characterization of Nash equilibria. We refer in particular to Frei and dos Reis (2011) who further explored this question following a previous version of this paper based on Espinosa (2010). In particular, Frei and dos Reis (2011) highlight some examples of nonexistence by allowing the agents to have some final reward.

In the context of deterministic coefficients, we obtain a complete characterization.

MAIN THEOREM 2.3. *Assume that θ and σ are deterministic and continuous functions of $t \in [0, T]$, and that for each $i = 1, \dots, N$, $U_i(x) = -e^{-\frac{x}{\eta_i}}$ for some constant $\eta_i > 0$, the portfolio constraints sets A_i are closed convex, and $\prod_{i=1}^N \lambda_i < 1$. Then, there exists a unique deterministic Nash equilibrium.*

We also observe that the unique Nash equilibrium of Main Theorem 2.3 can be computed explicitly in various interesting situations that will be described throughout the paper.

In order to simplify notations, from now on, we will write

$$X_t^i := X_t^{\pi^i} \quad \text{and} \quad \bar{X}_t^i := \bar{X}_t^{i,(\pi^j)_{j\neq i}}, t \in [0, T].$$

In Section 3, we shall consider the complete market situation in which the portfolios will be free of constraints (in other words, $A_i = \mathbb{R}^d$ for each i). This will be solved for general utility functions. In the next sections, we will derive results for more general types of constraints, but we will focus on the case of exponential utility functions: $U_i(x) = -e^{-\frac{x}{\eta_i}}$. We will first consider the general case in Section 4, and then in Section 5, we will focus on the case of linear constraints, where the A_is are (vector) subspaces of $\mathbb{R}^d$.

3. THE COMPLETE MARKET SITUATION

In this section, we consider the case where there are no constraints on the portfolios:

$$A_i = \mathbb{R}^d, \quad \text{for all } i = 1, \dots, N.$$

In the present situation, the density of the unique equivalent martingale measure is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\int_0^T \theta(u)\cdot dW_u - \frac{1}{2}\int_0^T |\theta(u)|^2 du}. \tag{3.1}$$

We shall denote by $\mathbb{E}^{\mathbb{Q}}$ the expectation under $\mathbb{Q}$.

In contrast with the general results in the subsequent sections, the complete market situation can be solved for general utility functions. In this case, the set of admissible strategies $\mathcal{A} = \mathcal{A}_i$ is the set of predictable processes π such that

$$\begin{aligned} &\sigma\pi \in \mathbb{H}^2_{loc}(\mathbb{R}^d),\ X^\pi \text{ is a} \mathbb{Q}\text{-martingale}, \\ &\text{and } U_j\big(-2^k\big(X_T^\pi\big)^+\big), U_j\big(-2^k\big(X_T^\pi\big)^-\big) \in \mathbb{L}^1(\mathbb{P}) \quad \text{for all } j \le N \quad \text{and } k \in \mathbb{N}. \end{aligned} \tag{3.2}$$

To simplify the notations and presentation in this introductory example, we also assume that all agents have the same relative performance coefficient λ:

$$\lambda_i = \lambda \in [0, 1), \quad \text{for all } i = 1, \dots, N, \tag{3.3}$$

see, however, Remark 3.4. Nevertheless, we allow the investors to have different utility functions U_i and different initial endowments $x^i \in \mathbb{R}$. We denote

$$\bar{x}^i := \frac{1}{N-1}\sum_{j\neq i} x^j, i = 1, \dots, N.$$

3.1. Single-Agent Optimization

We first check that the single-agent optimization problem is well-posed under the following additional conditions:

$$\text{for all } y > 0, \mathbb{E}\left|U_i \circ I_i\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right| < \infty, \mathbb{E}^{\mathbb{Q}}\left|I_i\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right| < \infty, \text{ and} \tag{3.4}$$

$$U_j\left(-2^k I_i\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^+\right), \quad U_j\left(-2^k I_i\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^-\right) \in \mathbb{L}^1(\mathbb{P}) \quad \text{for all } j \le N \quad \text{and } k \in \mathbb{N}.$$

LEMMA 3.1. *Under condition (3.4), we have* $U_i(X_T^i - \lambda \bar{X}_T^i) \in \mathbb{L}^1(\mathbb{P})$ *for all* $(\pi^1, \dots, \pi^N) \in \mathcal{A}_1 \times \dots \times \mathcal{A}_N$ *and* $i = 1, \dots, N$.

Proof. Using the convex dual of $-U_i(-x)$, we have for any $y > 0$:

$$U_i(X_T^i - \lambda \bar{X}_T^i) \le \left| U_i \circ I_i \left(y \frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right| + y \frac{d\mathbb{Q}}{d\mathbb{P}} \left(|X_T^i - \lambda \bar{X}_T^i| + \left| I_i \left(y \frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right| \right).$$

The right-hand side is integrable under $\mathbb{P}$ by the admissibility conditions (3.2) and the integrability assumptions (3.4). Then, $U_i(X_T^i - \lambda \bar{X}_T^i)^+$ is integrable. On the other hand, by the increase of U_i, we have $U_i(x - y) \ge U_i(-x^- - y^+) \ge -|U_i(-2x^-)| - |U_i(-2y^+)|$. Then, it follows from the concavity of U_i that

$$\begin{aligned} U_i(X_T^i - \lambda \bar{X}_T^i) &\ge (1-\lambda) U_i(X_T^i) + \frac{\lambda}{N-1} \sum_{j \ne i} U_i(X_T^i - X_T^j) \\ &\ge (1-\lambda) U_i(-2(X_T^i)^-) - \frac{\lambda}{N-1} \sum_{j \ne i} \{ |U_i(-2(X_T^i)^-)| + |U_i(-2(X_T^j)^+)| \} \\ &\ge -|U_i(-2(X_T^i)^-)| - \frac{\lambda}{N-1} \sum_{j \ne i} |U_i(-2(X_T^j)^+)|. \end{aligned}$$

Hence, $U_i(X_T^i - \lambda \bar{X}_T^i) \in \mathbb{L}^1(\mathbb{P})$ by our definition of $\mathcal{A}_i$. □

We now characterize the optimal portfolio and wealth of each agent, given the strategies of his peers. In other words, we try to find the best response of agent i to the strategies of his peers. As in the classical case of optimal investment in complete market, we will use the convex dual of $-U_i(-x)$. Since U_i is strictly concave and C^1, we can define $I_i := (U_i')^{-1}$, which is a bijection from $\mathbb{R}_+^*$ onto $\mathbb{R}$ because of (2.4). The main result of this section requires the following integrability conditions:

LEMMA 3.2. *For any* $i = 1, \dots, N$, *let the strategies* $\pi^j \in \mathcal{A}$ *for* $j \ne i$ *be given. Then, under (3.4), there exists a unique optimal portfolio for the optimization problem (2.6) of agent* i *with optimal final wealth:*

$$X_T^{i*} = I_i \left(y^i \frac{d\mathbb{Q}}{d\mathbb{P}} \right) + \lambda \bar{X}_T^i, \quad \text{where } y^i \text{ is defined by } \mathbb{E}^{\mathbb{Q}} I_i \left(y^i \frac{d\mathbb{Q}}{d\mathbb{P}} \right) = x^i - \lambda \bar{x}^i. \tag{3.5}$$

Proof. Since the market is complete, and under conditions (3.2) and (3.4), there exist portfolio strategies π_i^* such that $X_T^{i*} = X_T^{\pi_i^*}$ is a $\mathbb{Q}$-martingale. We only verify that $\pi_i^* \in \mathcal{A}_i$ for all i, the rest of the proof is omitted as it follows the classical martingale approach in the simple complete market framework. Writing $I_i := I_i(y^i \frac{d\mathbb{Q}}{d\mathbb{P}})$, we compute that

$$\begin{aligned} U_j(0) \ge U_j(-2^k (X_T^{i*})^\pm) &\ge U_j(-2^k I_i^\pm - 2^k (\bar{X}_T^i)^\pm) \\ &\ge \frac{1}{N-1} \sum_{\ell \ne i} U_j(-2^k I_i^\pm - 2^k (X_T^\ell)^\pm) \\ &\ge \frac{-1}{N-1} \sum_{\ell \ne i} \{ |U_j(-2^{k+1} I_i^\pm)| + |U_j(2^{k+1} (X_T^\ell)^\pm)| \}. \end{aligned}$$

Then, the required integrability follows from (3.4) and the definition (3.2) of the sets of admissible portfolios $\mathcal{A}_i$, $i = 1, \dots, N$. □

3.2. Partial Nash Equilibrium

The second step is to search for a Nash equilibrium between the N agents. Let $\mathbf{X_N} := (X_T^{i*})_{1\le i\le N}$ be the vector of terminal wealth of the investors associated to $(\pi^1, \dots, \pi^N)$. From Lemma 3.2, $(\pi^1, \dots, \pi^N)$ is a Nash equilibrium if and only if we have

$$A_N \mathbf{X_N} = J_N,$$

$$\text{where } A_N = \begin{pmatrix} 1 & & -\frac{\lambda}{N-1} \\ & \ddots & \\ -\frac{\lambda}{N-1} & & 1 \end{pmatrix} \in M_N(\mathbb{R}); \quad J_N = \left(I_i\left(y^i \frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right)_{1\le i\le N}.$$

Under the condition $\lambda \neq 1$ in (3.3), it follows that A_N is invertible and we can compute explicitly that

$$A_N^{-1} = \begin{pmatrix} 1 + \frac{\lambda^2}{(1-\lambda)(N+\lambda-1)} & & \frac{\lambda}{(1-\lambda)(N+\lambda-1)} \\ & \ddots & \\ \frac{\lambda}{(1-\lambda)(N+\lambda-1)} & & 1 + \frac{\lambda^2}{(1-\lambda)(N+\lambda-1)} \end{pmatrix},$$

thus providing the existence of a unique Nash equilibrium:

THEOREM 3.3. *There exists a unique Nash equilibrium, and the equilibrium terminal wealth for each $i = 1, \dots, N$ is given by*

$$\hat{X}_T^i = \left(1 + \frac{\lambda^2}{(1-\lambda)(N+\lambda-1)}\right) I_i\left(y^i \frac{d\mathbb{Q}}{d\mathbb{P}}\right) + \frac{\lambda}{(1-\lambda)(N+\lambda-1)} \sum_{j\neq i} I_j\left(y^j \frac{d\mathbb{Q}}{d\mathbb{P}}\right).$$

REMARK 3.4. In the case of specific λ_is, the previous arguments can be adapted. In the expression of A_N, λ_i appears on the ith line instead of λ, A_N is invertible if and only if $\prod_{i=1}^N \lambda_i < 1$ (for more details, see the proof of Lemma 4.6 below), and then its inverse is given by

$$\left(A_N^{-1}\right)_{ii} = 1 + \frac{\lambda_i^N \sum_{k\neq i} \frac{\lambda_k^N}{1+\lambda_k^N}}{1 - \sum_{k\neq i} \frac{\lambda_k^N(1+\lambda_i^N)}{1+\lambda_k^N}}, \quad \text{and} \quad \left(A_N^{-1}\right)_{ij} = \frac{\frac{\lambda_i^N}{1+\lambda_j^N}}{1 - \sum_{k\neq i} \frac{\lambda_k^N(1+\lambda_i^N)}{1+\lambda_k^N}} \quad \text{for } i \neq j,$$

where we denoted $\lambda_i^N := \lambda_i/(N-1)$. The equilibrium performances are given by

$$\hat{X}_T^i = \sum_{j=1}^N \left(A_N^{-1}\right)_{ij} I_j\left(y^j \frac{d\mathbb{Q}}{d\mathbb{P}}\right), \quad i = 1, \dots, N.$$

REMARK 3.5. In the case $\lambda = 1$, it turns out that there exist either an infinity of Nash equilibria or no Nash equilibrium. Indeed, in this case, A_N is of rank $N-1$. Therefore, if J_N belongs to the image of A_N, then there is an affine space of dimension 1 of Nash equilibria, while if J_N is not in the image of A_N, then there is no Nash equilibrium.

In particular, in the exponential utility context (further developed below), we directly compute that $J_N = A_N x + \eta \int_0^T \theta(t)\cdot(\theta(t)dt + dW_t)$, where x is the vector of initial data

x^i and η is the vector of risk tolerances η_i of each agent. Therefore, J_N belongs to the image of A_N if and only if η belongs to it.

3.3. The Exponential Utility Case

In order to push further the analysis of the complete market situation, we now consider the exponential utility case:

$$U_i(x) = -e^{-\frac{x}{\eta_i}},\ x \in \mathbb{R}, \tag{3.6}$$

where $\eta_i > 0$ is the risk tolerance parameter for agent i, i.e., the inverse of his absolute risk aversion coefficient. We denote the average risk tolerance by

$$\bar{\eta}_N := \frac{1}{N}\sum_{j=1}^{N}\eta_j. \tag{3.7}$$

In the present context, $I_i(y) = -\eta_i \ln(\eta_i y)$, so that the equilibrium wealth process is

$$\hat{X}^i_T = x^i - \frac{\eta_i}{1-\lambda}\left[\left(1 - \frac{\lambda N}{N+\lambda-1}\right) + \frac{\lambda N}{N+\lambda-1}\frac{\bar{\eta}_N}{\eta_i}\right]\left(\ln\frac{d\mathbb{Q}}{d\mathbb{P}} - \mathbb{E}^{\mathbb{Q}}\ln\frac{d\mathbb{Q}}{d\mathbb{P}}\right).$$

We denote by $\hat{\pi}^{i,N,\lambda}$ the corresponding equilibrium portfolio strategy of agent i, where we emphasize its dependence on the parameters N and λ.

In order to have explicit formulas, we assume that the risk premium θ is a (deterministic) continuous function of t. Then, it is well known that the classical portfolio optimization problem with no interaction between managers leads to the optimal portfolios

$$\hat{\pi}^{0,i}_t := \eta_i\sigma_t^{-1}\theta(t), \quad t \in [0, T].$$

PROPOSITION 3.6. *In the above setting, the equilibrium portfolio for agent i is given by*

$$\hat{\pi}^{i,N,\lambda}_t = k^{i,N}_\lambda \hat{\pi}^{0,i}_t, \quad \textit{where } k^{i,N}_\lambda := \frac{1}{1-\lambda}\left[\left(1 - \frac{\lambda N}{N+\lambda-1}\right) + \frac{\lambda N}{N+\lambda-1}\frac{\bar{\eta}_N}{\eta_i}\right].$$

REMARK 3.7. Assume further that $\bar{\eta}_N \longrightarrow \eta > 0$ as $N \to \infty$. Then, $k^{i,N}_\lambda \longrightarrow 1 + \frac{\lambda}{1-\lambda}\frac{\eta}{\eta_i}$. In particular, if all agents have the same risk aversion coefficient $\eta_i = \eta > 0$, then

$$\hat{\pi}^{i,N,\lambda} = \hat{\pi}^\lambda := \frac{1}{1-\lambda}\hat{\pi}^0_t. \quad \text{for all } i.$$

REMARK 3.8. In the case of similar agents, i.e., for any $i = 1, \dots, N$, $\eta_i = \eta$ and $\lambda_i = \lambda$, we can find the equilibrium portfolio very easily. Indeed, by symmetry considerations, all the X^is must be equal, $X^i = \bar{X}^i$, and the optimization problem reduces to

$$\sup_\pi -\mathbb{E}e^{-\frac{1-\lambda}{\eta}X^i_T}.$$

This is the classical case with η replaced by $\frac{\eta}{1-\lambda}$ so that the optimal portfolio is given by $\hat{\pi}_t = \eta\sigma_t^{-1}\theta(t)/(1-\lambda)$, in agreement with our results.

In the general case of the following sections, we will not always be able to conclude anything on the behavior of every agent; therefore, we introduce the following definition:

DEFINITION 3.9. The market index and the corresponding market portfolio are defined by

$$\bar{X}_t := \frac{1}{N}\sum_{i=1}^{N} X_t^i \quad \text{and} \quad \bar{\pi}_t := \frac{1}{N}\sum_{i=1}^{N} \pi_t^i,\ t \in [0, T].$$

We recall the definition of the Sharpe ratio SR and introduce the variance risk ratio VRR:

$$\text{SR} = \frac{\text{expected excess return}}{\text{volatility}}, \quad \text{VRR} := \frac{\text{expected excess return}}{\text{variance}}. \tag{3.8}$$

For practical purposes, the VRR is a better criterion for the two following reasons:

- VRR is robust to the investment duration, while SR is not: for a time period L and a scalar $k > 0$, we have $\text{SR}(kL) = k\,\text{SR}(L)$, while $\text{VRR}(kL) = \text{VRR}(L)$.
- VRR accounts for the illiquidity risk related to the size of the position, while SR does not: for a portfolio X and a scalar $k > 0$, we have $\text{SR}(kX) = \text{SR}(X)$ and $\text{VRR}(kX) = \text{VRR}(X)/k$.

We have the following results for the impact of λ:

PROPOSITION 3.10.

(i) For any linear form φ, $|\varphi(\hat{\pi}_t^{i,N,\lambda})|$ is increasing w.r.t. λ.
(ii) The dynamics of the market index and the corresponding market portfolio are given by

$$d\bar{X}_t = \frac{\bar{\eta}_N}{1-\lambda}\theta(t)\cdot[\theta(t)dt + dW_t] \quad \text{and} \quad \bar{\pi}_t = \frac{\bar{\eta}_N}{1-\lambda}\sigma_t^{-1}\theta(t).$$

In particular, for any linear form φ, $|\varphi(\bar{\pi}_t)|$ is increasing w.r.t. λ.

Proof. (ii) is immediate, so we only prove (i). By Proposition 3.6, $\hat{\pi}^{i,N,\lambda} = k_\lambda^{i,N}\hat{\pi}_t^{0,i}$, and we directly compute that

$$\frac{\partial k_\lambda^{i,N}}{\partial \lambda} = \frac{1}{(1-\lambda)^2(N+\lambda-1)^2}\left[N+\lambda-1+N\big(\lambda+(N-1)(1-\lambda)\big)\left(\frac{\bar{\eta}_N}{\eta_i}-1\right)\right].$$

By definition of $\bar{\eta}_N$ in (3.7) and the fact that $\eta_j \geq 0$ for all j, we have $\frac{\bar{\eta}_N}{\eta_i} - 1 \geq \frac{1-N}{N}$. Therefore,

$$\begin{aligned}(1-\lambda)^2(N+\lambda-1)^2\frac{\partial k_\lambda}{\partial \lambda} &\geq N(N+\lambda-1) - \lambda(N-1) - (N-1)^2(1-\lambda)\\ &\geq (N-1)(1-\lambda) + \lambda(N^2-N+1) > 0.\end{aligned}$$

□

In words, Proposition 3.10 states that the more investors are concerned about each other, the more risk they will undertake. In each investment direction, the global position of agents, described by $|\varphi(\bar{\pi}_t)|$, will increase with λ and in the limit $\lambda \to 1$, we even have a limit of infinite positions $|\varphi(\bar{\pi}_t)| \to \infty$ a.s. Furthermore, the drift and volatility of the market index are both increasing w.r.t. λ. The corresponding Sharpe ratio is $\text{SR} = |\theta(t)|$, independent of λ, while the variance risk ratio is $\text{VRR} = \frac{1-\lambda}{\bar{\eta}_N}$, a decreasing function of λ. This is a perverse aspect of the present financial markets that may provide an explanation of the emergence of financial bubbles, when managers use the Sharpe ratio as a reliable indicator.

3.4. General Equilibrium

In the previous sections, the price process S was given exogeneously. We now analyze the effect of the relative performance coefficient λ when the price process S is determined at the equilibrium.

For each fixed price process S, defined as in Section 2, there exists a unique Nash equilibrium in the sense of Definition 2.1. Similar to Karatzas and Shreve (1998), our objective is to search for a market equilibrium price S that is consistent with market equilibrium conditions:

$$\sum_{i=1}^{N} \pi_t^{i,j} = K^j S_t^j \quad \text{for all } j = 1, \dots, d \quad \text{and } t \in [0, T], \tag{3.9}$$

$$\sum_{i=1}^{N} x^i = \sum_{j=1}^{d} K^j S_0^j, \tag{3.10}$$

where K^j is a constant such that $K^j S_t^j$ is the market capitalization of the jth firm. Equation (3.9) says that the total amount invested in the stocks of the jth firm is equal to the market capitalization of this firm. Equation (3.10) says that the initial endowment of the investors equals the initial market capitalizations. With $\mathbf{1} := (1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^d$, we observe that (3.9) and (3.10) imply that

$$\begin{aligned}
\sum_{i=1}^{N} X_t^i &= \sum_{i=1}^{N} \left(x^i + \int_0^t \pi_u^i \cdot \operatorname{diag}(S_u)^{-1} dS_u \right) \\
&= \sum_{i=1}^{N} x^i + \sum_{j=1}^{d} \int_0^t K^j dS_u^j \\
&= \sum_{i=1}^{N} x^i + \sum_{j=1}^{d} K^j \left(S_t^j - S_0^j\right) = \sum_{j=1}^{d} K^j S_t^j = \sum_{i=1}^{N} \pi_t^i \cdot \mathbf{1},
\end{aligned}$$

i.e., the total amount invested in the nonrisky asset is zero at any time $t \in [0, T]$.

DEFINITION 3.11. We say that a process S is an equilibrium market if there exists a Nash equilibrium $\hat{\pi} = (\hat{\pi}^1, \dots, \hat{\pi}^N)$ associated with the price dynamics S, in the sense of Definition 2.1, such that S and $\hat{\pi}$ satisfy (3.9) and (3.10).

In order to simplify notations, we set $K^j = k^j N$ and $k := (k^1, \dots, k^d)$.

PROPOSITION 3.12. *Let θ be a deterministic and continuous function of $t \in [0, T]$. Then, there exists an equilibrium market whose risk premium is θ. Moreover, in this equilibrium market, the market index is given by*

$$\bar{X}_t = \bar{x} + \frac{\bar{\eta}_N}{1-\lambda} \int_0^t \theta(u) \cdot (\theta(u)\, du + dW_u),\ t \in [0, T].$$

Proof. By Proposition 3.10 (ii), it follows that

$$S_t = \frac{\bar{\eta}_N}{1-\lambda}\operatorname{diag}(k)\sigma(t)^{-1}\theta(t).$$

Notice that the previous equation does not define σ uniquely for $d > 1$.

Conversely, let θ be some given continuous function. Then, we can choose a diagonal matrix $\sigma_t = \sigma(t, S_t)$, with diagonal elements

$$\sigma^{ii}(t, S_t) = \frac{\bar{\eta}_N}{(1-\lambda)k^i S_t^i}\theta^i(t).$$

Notice that σ satisfies the conditions for S to be a strong solution of (2.3). Then, it follows from Proposition 3.6 that:

$$d\bar{X}_t = \frac{1}{N}\sum_{i=1}^{N} dX_t^i = \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_t^i \cdot \operatorname{diag}(S_t)^{-1}dS_t = \frac{\bar{\eta}_N}{1-\lambda}\theta(t)\cdot(\theta(t)dt + dW_t).$$

□

We next analyze the impact of λ on the drift and the volatility of the market index. Despite the multiplicity of market equilibria, they all lead to similar conclusions. Let us, for example, assume that the risk premium is independent of λ. Then, the drift of the market index is $\eta|\theta(t)|^2/(1-\lambda)$ and the volatility is $\eta|\theta(t)|/(1-\lambda)$; thus, both are increasing w.r.t. λ and with the same order. We may interpret this equilibrium as a financial bubble, where the return and the volatility are both increased by the agents' interactions. An alternative interpretation for a fund manager is that for the same given return, the agents' interaction coefficient increases the volatility of the optimal portfolio.

Notice that in the present setting, the variance risk ratio VRR $= (1-\lambda)/\eta$ is decreasing in λ and tends to zero as $\lambda \to 1$. This indicates that according to this criterion, the agents' interactions lead to market inefficiency.

4. GENERAL CONSTRAINTS WITH EXPONENTIAL UTILITY

In the rest of this paper, we consider a general case with constrained portfolios. We assume

(4.1) $\qquad A_i$ is a closed convex set of $\mathbb{R}^d$, $\quad$ for all $i = 1, \ldots, N$.

We denote by P_t^i the orthogonal projection on $\sigma_t A_i$, which is well defined by (4.1). For $x \in \mathbb{R}^d$, we denote $\operatorname{dist}(x, \sigma_t A_i) := |x - P_t^i x|$ the Euclidean distance from x to the closed convex subset $\sigma_t A_i$.

REMARK 4.1. Recall that for a closed convex set A in a Euclidean space, the orthogonal projection on A, denoted by P, is well defined, is a contraction, and satisfies for any $x, y \in \mathbb{R}^d$:

$$|P(x) - P(y)|^2 \le (x-y)\cdot(P(x) - P(y)) \le |x-y|^2.$$

Moreover, $P(x)$ is the only point satisfying $(x - P(x))\cdot(a - P(x)) \le 0$ for all $a \in A$.

For technical reasons, we restrict our analysis to exponential utility functions (3.6).

DEFINITION 4.2. The set of admissible strategies $\mathcal{A}_i$ is the collection of all predictable processes π with values in A_i, $dt \otimes d\mathbb{P}-$a.e., such that $\sigma\pi \in \mathbb{H}^2_{loc}(\mathbb{R}^d)$ and such that the family

$$\{e^{\pm(X^\pi_\tau - X^\pi_\nu)};\ \nu, \tau \text{ stopping times on } [0, T] \text{ with } \nu \le \tau \text{ a.s.}\} \tag{4.2}$$

is uniformly bounded in $\mathbb{L}^p(\mathbb{P})$ for all $p > 0$.

In comparison with the admissibility conditions of Section 3, the previous definition requires the uniform boundedness condition of the above family, which is needed in order to prove a dynamic programming principle similar to Lim and Quenez (2009).

REMARK 4.3. The set of admissible strategies introduced in Definition 4.2 is strictly smaller than the corresponding set in the one-agent framework of Hu, Imkeller, and Müller (2005). This deficiency was further overcome by Frei and dos Reis (2011) who used the results of the present paper (which are contained in Espinosa 2010).

4.1. A Formal Argument

In this section, we provide a formal argument that helps to understand the construction of Nash equilibrium of the subsequent section. For fixed $i = 1, \dots, N$, we rewrite (2.6) as

$$V^i_0 := \sup_{\pi^i \in \mathcal{A}_i} \mathbb{E}\big[U_i\big(X^{\pi^i}_T - \tilde{\xi}^i\big)\big], \quad \text{where } \tilde{\xi}^i := \lambda_i \bar{X}^i_T =: \lambda_i \bar{x}^i + \tilde{\xi}^i_0. \tag{4.3}$$

Then, following El Karoui and Rouge (2000) or Hu, Imkeller, and Müller (2005), we expect that the value function V^i_0 and the corresponding optimal solution be given by

$$V^i_0 = -e^{-(x^i - \lambda_i \bar{x}^i - \tilde{Y}^i_0)/\eta_i}, \quad \sigma_t \hat{\pi}^i_t = P^i_t(\tilde{\zeta}^i_t + \eta_i \theta_t) \quad \text{for all } t \in [0, T],$$

and $(\tilde{Y}^i, \tilde{\zeta}^i)$ is the solution of the quadratic BSDE:

$$\tilde{Y}^i_t = \tilde{\xi}^i_0 + \int_t^T \Big(-\tilde{\zeta}^i_u \cdot \theta_u - \frac{\eta_i}{2}|\theta_u|^2 + \tilde{f}^i_u\big(\tilde{\zeta}^i_u + \eta_i\theta_u\big)\Big)\, du - \int_t^T \tilde{\zeta}^i_u \cdot dW_u,\ t \le T, \tag{4.4}$$

where the generator $\tilde{f}^i$ is given by

$$\tilde{f}^i_t(z^i) := \frac{1}{2\eta_i} \operatorname{dist}(z^i, \sigma_t A_i)^2, \quad z^i \in \mathbb{R}^d. \tag{4.5}$$

This suggests that one can search for a Nash equilibrium by solving the BSDEs (4.4) for all $i = 1, \dots, N$. However, this raises the following difficulties:

- the final data $\tilde{\xi}^i_0$ do not have to be bounded as it is defined in (4.3) through the performance of the other portfolio managers;
- the situation is even worse because the final data $\tilde{\xi}^i_0$ induce a coupling of the BSDEs (4.4) for $i = 1, \dots, N$. To express this coupling in a more transparent way, we substitute the expressions of $\tilde{\xi}^i_0$ and rewrite (4.4) for $t = 0$ into:

$$\tilde{Y}^i_0 = \eta_i \xi + \int_0^T \tilde{f}^i_u(\zeta^i_u)\, du - \int_0^T \left(\zeta^i_u - \lambda^N_i \sum_{j \ne i} P^j_u\big(\zeta^j_u\big)\right) \cdot dB_u,$$

where $B := W + \int_0^{\cdot} \theta_r dr$ is the Brownian motion under the equivalent martingale measure

$$\lambda_i^N := \frac{\lambda_i}{N-1}, \quad \zeta_t^i := \tilde{\zeta}_t^i + \eta_i \theta_t, \quad t \in [0, T], \tag{4.6}$$

and the final data are expressed in terms of the unbounded r.v.

$$\xi := \int_0^T \theta_u \cdot dB_u - \frac{1}{2}\int_0^T |\theta_u|^2 du. \tag{4.7}$$

Then, $\tilde{Y}_0 = Y_0$, where (Y, ζ) is defined by the BSDE

$$Y_t^i = \eta_i \xi + \int_t^T \tilde{f}_u^i(\zeta_u^i)\, du - \int_t^T \left(\zeta_u^i - \lambda_i^N \sum_{j \neq i} P_u^j(\zeta_u^j)\right) \cdot dB_u. \tag{4.8}$$

In order to sketch (4.8) into the BSDEs framework, we further introduce the mapping $\varphi_t : \mathbb{R}^{Nd} \longrightarrow \mathbb{R}^{Nd}$ defined by the components:

$$\varphi_t^i(\zeta^1, \ldots, \zeta^N) := \zeta^i - \lambda_i^N \sum_{j \neq i} P_t^j(\zeta^j) \quad \text{for all } \zeta^1, \ldots, \zeta^N \in \mathbb{R}^d. \tag{4.9}$$

It turns out that the mapping φ_t is invertible under fairly general conditions. We shall prove this result in Lemma 4.6 for general convex constraints and in Lemma 5.1 in the case of linear constraints. We denote $\psi_t := \varphi_t^{-1}$ and $\psi_t^i(z)$ the ith block component of size d of $\varphi_t^{-1}(z)$. Then, one can rewrite (4.8) as

$$Y_t^i = \eta_i \xi + \int_t^T f_u^i(Z_u)\, du - \int_t^T Z_u^i \cdot dB_u, \tag{4.10}$$

where the generator f^i is now given by

$$f_t^i(z) := \tilde{f}_t^i\big(\psi_t^i(z)\big) \quad \text{for all } z = (z^1, \ldots, z^N) \in \mathbb{R}^{Nd}. \tag{4.11}$$

A Nash equilibrium should then satisfy for each i:

$$\hat{\pi}_t^i = \sigma_t^{-1} P_t^i\big(\psi_t^i(Z_t)\big), \quad i = 1, \ldots, N. \tag{4.12}$$

4.2. Auxiliary Results

Our first objective is to verify that the map φ introduced in (4.9) is invertible. The crucial condition for the rest of this section is

$$\prod_{1 \le i \le N} \lambda_i < 1. \tag{4.13}$$

Recall the notation λ_i^N from (4.6).

LEMMA 4.4. *Under (4.1) and (4.13), for any $t \in [0, T]$, the map $I + \lambda_j^N P_t^j$ is a bijection on $\mathbb{R}^d$ and its inverse is a contraction, for all $j = 1, \ldots, N$.*

Proof. Let $t \in [0, T]$ be fixed, for ease of notation, we omit all t subscripts. Since $\sigma_t A_j$ is a closed convex set, from Remark 4.3, $(x - y) \cdot (P^j(x) - P^j(y)) \ge |P^j(x) - P^j(y)|^2 \ge 0$, for any $x, y \in \mathbb{R}^d$. Notice that $I + \lambda_j^N P^j$ is a bijection if and only if, for all $y \in \mathbb{R}^d$, the map

$$f_y(x) := y - \lambda_j^N P^j(x) \tag{4.14}$$

has a unique fixed point. Since P^j is a contraction, we compute, for any x, x' in $\mathbb{R}^d$:

$$|f_y(x) - f_y(x')| = \lambda_j^N |P^j(x) - P^j(x')| \leq \lambda_j^N |x - x'| = \frac{\lambda_j}{N-1}|x - x'|.$$

Case 1: If $N \geq 3$ or $\lambda_j < 1$, then f_y is a strict contraction of $\mathbb{R}^d$. We prove now that the inverse of $I + \lambda_j^N P^j$ is a contraction. Indeed, if $x \neq y$, we have

$$\begin{aligned} \left|x - y + \lambda_j^N (P^j(x) - P^j(y))\right|^2 &= |x - y|^2 + \left(\lambda_j^N\right)^2 |P^j(x) - P^j(y)|^2 \\ &\quad + 2\lambda_j^N (x - y) \cdot (P^j(x) - P^j(y)) \\ &\geq |x - y|^2 > 0, \end{aligned} \tag{4.15}$$

where we used the fact that $(x - y) \cdot (P^j(x) - P^j(y)) \geq 0$, see Remark 4.15.

Case 2: If $N = 2$ and $\lambda_j = 1$, f_y fails to be a strict contraction. However, (4.15) still holds and implies that $I + P^j$ is one-to-one. Using Lemma 4.5 below, we get the bijection property of $I + P^j$ and the contraction property of the inverse function follows from (4.15). □

LEMMA 4.5. *Let A be a closed convex set of $\mathbb{R}^d$. Then $(I + P_A)(\mathbb{R}^d) = \mathbb{R}^d$.*

Proof. Let $B := 2A = \{y \in \mathbb{R}^d; \exists x \in A, y = 2x\}$, and let us prove that

$$P_A\left(y - \frac{1}{2} P_B(y)\right) = \frac{1}{2} P_B(y) \quad \text{for all } y \in \mathbb{R}^d. \tag{4.16}$$

This implies that $y = (I + P_A)(y - \frac{1}{2} P_B(y)) \in (I + P_A)(\mathbb{R}^d)$ for all $y \in \mathbb{R}^d$, which gives the required result.

To prove (4.16), define $x := \frac{1}{2} P_B(y)$ and $z := y - 2x$. By Remark 4.3, $P_B(y)$ is the only point in B satisfying $(y - P_B(y)) \cdot (b - P_B(y)) \leq 0$ for all $b \in B$. In other words, we have for any $b \in B$, $z \cdot (b - 2x) \leq 0$, or by definition of B, for any $a \in A$, $z \cdot (2a - 2x) \leq 0$; hence,

$$(x + z - x) \cdot (a - x) \leq 0 \quad \text{for all } a \in A,$$

which means that $x = P_A(x + z)$ and therefore $(I + P_A)(x + z) = x + z + x = y$. □

Recall the definition of φ in (4.9).

LEMMA 4.6. *Under (4.1) and (4.13), we have for $t \in [0, T]$:*

(i) φ_t is a bijection of $\mathbb{R}^{Nd}$, and we write $\psi_t := \varphi_t^{-1}$.
(ii) ψ_t is Lipschitz continuous with a constant depending only on N and the λ_is.

Proof. For ease of notation, we omit all t subscripts. For arbitrary $z = (z^1, \ldots, z^N)$ in $\mathbb{R}^{Nd}$, we want to find a solution $\zeta \in \mathbb{R}^{Nd}$ to the following system:

$$\varphi^i(\zeta) = \zeta^i - \lambda_i^N \sum_{j \neq i} P^j(\zeta^j) = z^i, \quad 1 \leq i \leq N. \tag{4.17}$$

Subtracting λ_j times equation i from λ_i times equation j in (4.17), we see that

$$\lambda_i \left(I + \lambda_j^N P^j\right)(\zeta^j) = \lambda_j \left(I + \lambda_i^N P^i\right)(\zeta^i) + \lambda_i z^j - \lambda_j z^i, \quad i, j = 1, \dots, N. \tag{4.18}$$

1. From Lemma 4.4, we know that $I + \lambda_j^N P^j$ is a bijection; thus, from (4.18), we compute

$$\sum_{j \neq i} P^j(\zeta^j) = \sum_{j \neq i} P^j \circ \left(I + \lambda_j^N P^j\right)^{-1} \left(\frac{\lambda_j}{\lambda_i} (I + \lambda_i^N P^i)(\zeta^i) + z^j - \frac{\lambda_j}{\lambda_i} z^i \right),$$

so that from (4.17)

$$\zeta^i = z^i + \lambda_i^N \sum_{j \neq i} P^j \circ \left(I + \lambda_j^N P^j\right)^{-1} \left(\frac{\lambda_j}{\lambda_i} (I + \lambda_i^N P^i)(\zeta^i) + z^j - \frac{\lambda_j}{\lambda_i} z^i \right) =: g^{i,z}(\zeta^i). \tag{4.19}$$

2. We next show that under Condition (4.13), $g^{i,z}$ has a unique fixed point. We have

$$\begin{aligned} \left|\left(I + \lambda_j^N P^j\right)(x) - \left(I + \lambda_j^N P^j\right)(y)\right|^2 &= |x - y|^2 + 2\lambda_j^N (x - y) \cdot (P^j(x) - P^j(y)) \\ &\quad + \left(\lambda_j^N\right)^2 |P^j(x) - P^j(y)|^2 \\ &\geq \left(1 + 2\lambda_j^N + \left(\lambda_j^N\right)^2\right) |P^j(x) - P^j(y)|^2 \\ &\geq \left(1 + \lambda_j^N\right)^2 |P^j(x) - P^j(y)|^2. \end{aligned}$$

Therefore, $P^j \circ (I + \lambda_j^N P^j)^{-1}$ is $\frac{1}{1+\lambda_j^N}$−Lipschitz. Then, since $(I + \lambda_i^N P^i)$ is $1 + \lambda_i^N$−Lipschitz:

$$|g^{i,z}(x) - g^{i,z}(y)| \leq \frac{1}{N-1} \sum_{j \neq i} \frac{\lambda_j}{1 + \lambda_j^N} \left(1 + \lambda_i^N\right) |x - y|.$$

Notice that $\frac{\lambda_j}{1+\lambda_j^N}(1 + \lambda_i^N) \leq \max(\lambda_i, \lambda_j)$, with equality if and only if $\lambda_i = \lambda_j$. Therefore, condition (4.13) implies that $K^i := \frac{1}{N-1} \sum_{j \neq i} \frac{\lambda_j}{1+\lambda_j^N}(1 + \lambda_i^N) < 1$, where K^i depends only on N and the λ_js. Then, $g^{i,z}$ is a strict contraction and admits a unique fixed point that we write $\{\psi(z)\}^i$. It is then immediate that $\zeta = \psi(z)$ is the unique solution of (4.17).
3. Finally, we prove that ψ is Lipschitz with a constant depending only on N and the λ_js. Let $z_1, z_2 \in \mathbb{R}^{Nd}$, from (4.19), we compute

$$|\psi(z_1)^i - \psi(z_2)^i| \leq |z_1^i - z_2^i| + K^i |\psi(z_1)^i - \psi(z_2)^i| + 2 \sup_{1 \leq j \leq N} |z_1^j - z_2^j|.$$

Since $K := \sup_{1 \leq j \leq N} K^j < 1$, we get $\sup_{1 \leq j \leq N} |\psi(z_1)^j - \psi(z_2)^j| \leq \frac{3}{1-K} \sup_{1 \leq j \leq N} |z_1^j - z_2^j|$, which completes the proof since K depends only on N and the λ_js. □

4.3. The Main Results

Similar to the classical literature on portfolio optimization with exponential utility (El Karoui and Rouge 2000; Hu, Imkeller, and Müller 2005; Mania and Schweizer 2005), we first establish a connection between Nash equilibria and a quadratic multidimensional BSDE.

THEOREM 4.7. *Under (4.1) and (4.13), assume that* $(\tilde{\pi}^1, \ldots, \tilde{\pi}^N)$ *is a Nash equilibrium. Then, there exists a solution* $(Y, Z) \in \mathbb{H}^2(\mathbb{R}^N) \times \mathbb{H}^2_{loc}(\mathbb{R}^{Nd})$ *of the following N-dimensional BSDE:*

$$Y_t^i = \eta_i \xi + \frac{1}{2\eta_i} \int_t^T \left|\left(I - P_u^i\right) \circ \psi_u^i(Z_u)\right|^2 du - \int_t^T Z_u^i \cdot dB_u, \tag{4.20}$$

where ξ *is defined by (4.7), and we have*

$$\tilde{\pi}_t^i = \sigma_t^{-1} P_t^i\left(\psi_t^i(Z_t)\right) \quad \textit{and} \quad V_i = -e^{-\frac{1}{\eta_i}(x^i - \lambda_i \bar{x}^i - Y_0^i)}.$$

Proof. See Section 4.5. □

Our second main result focuses on the multidimensional Black–Scholes financial market, where we can guess an explicit solution to the BSDE (4.20). Although no uniqueness result is available for the BSDE (4.20) in this context, the following complete characterization is obtained by means of a PDE verification argument.

In view of Lemma 4.6, under Condition (4.13), the maps

$$\bar{\psi}_t^i(x) := \psi_t^i(\eta_1 x, \ldots, \eta_N x) \quad \text{for all } x \in \mathbb{R}^d, i = 1, \ldots, N, t \in [0, T], \tag{4.21}$$

are well defined and Lipschitz continuous on $\mathbb{R}^d$.

THEOREM 4.8. *Under (4.1) and (4.13), assume that* σ *and* θ *are deterministic continuous functions. Then, there exists a unique deterministic Nash equilibrium:*

$$\hat{\pi}_t^i = \sigma(t)^{-1} P_t^i \circ \bar{\psi}_t^i(\theta(t)) \quad \textit{for all } t \in [0, T]. \tag{4.22}$$

Moreover, the value function for agent i at equilibrium is given by

$$V_i = -e^{-\frac{1}{\eta_i}(x^i - \lambda_i \bar{x}^i - Y_0^i)}, \quad Y_0^i = -\frac{\eta_i}{2} \int_0^T |\theta(t)|^2 dt + \frac{1}{2\eta_i} \int_0^T \left|\left(I - P_t^i\right) \circ \bar{\psi}_t^i(\theta(t))\right|^2 dt.$$

Proof. See Section 4.6. □

We conclude this section by two simple examples. More interesting situations will be obtained later under the additional condition that the constraints sets are linear.

EXAMPLE 4.9 (Common investment). Let $\sigma = I_d, \lambda_i = \lambda, \eta_i = \eta$, and $A_i = \bar{B}(x, r)$ for some $x \in \mathbb{R}^d$ and $r > 0$, $i = 1, \ldots, N$. Here, $\bar{B}(x, r)$ is the closed ball centered at x with radius $r > 0$ for the canonical Euclidean norm of $\mathbb{R}^d$. Using Theorem 4.8, we compute the following equilibrium portfolio:

$$\hat{\pi}_t^i = P\left(\frac{\eta\theta(t)}{1-\lambda}\right) = \begin{cases} \dfrac{\eta\theta(t)}{1-\lambda} & \text{if } \dfrac{\eta\theta(t)}{1-\lambda} \in \bar{B}(x, r) \\ x + \dfrac{r}{\left|\frac{\eta\theta(t)}{1-\lambda} - x\right|} \left(\dfrac{\eta\theta(t)}{1-\lambda} - x\right) & \text{otherwise.} \end{cases}$$

Notice in particular that as one could expect, $\hat{\pi}^i_t - x$ is collinear to $\frac{\eta\theta(t)}{1-\lambda} - x$ and that $\hat{\pi}^i_t$ is in the boundary of $\bar{B}(x, r)$ whenever $\frac{\eta\theta(t)}{1-\lambda} \notin \bar{B}(x, r)$. One can prove that $|\hat{\pi}|$ is nondecreasing w.r.t. λ and η. Notice also that this expression is independent of N.

EXAMPLE 4.10 (Specific independent investments). Let $\sigma = I_d$, $\lambda_i = \lambda$, $\eta_i = \eta$, and $A_i = [a_i, b_i]e_i$, for some $a_i \le b_i$, $i = 1, \dots, N$. Here, $(e_j, 1 \le j \le d)$ is the canonical basis of $\mathbb{R}^d$. Using Theorem 4.8, we compute the following equilibrium portfolio for agent i:

$$\hat{\pi}^i_t = P^i\big(\eta\theta(t)\big) = a_i \vee \big(\eta\theta(t)\big) \wedge b_i.$$

This is exactly the same expression as in the classical case with no interaction between managers. Hence, the equilibrium portfolio is not affected by λ and N.

REMARK 4.11. Suppose that the portfolio constraints sets A_i are not convex. Then, we have to face two major problems. First, the projection operators A_i are not well defined. Second, and more importantly, the map φ may fail to be one-to-one or surjective onto $\mathbb{R}^{Nd}$. The failure of the one-to-one property means that there could exist more than one Nash equilibrium. However, the failure of the surjectivity onto $\mathbb{R}^{Nd}$, as illustrated by Examples A.1 and A.2 in the Appendix section, would lead to a constrained (N-dimensional) BSDE with no additional nondecreasing penalization process. Such BSDEs do not have solutions even in the case of Lipschitz generators, meaning that there is no Nash equilibrium in this context.

4.4. Infinite Managers Asymptotics

In the spirit of the theory of mean-field games, see Lasry and Lions (2007), we examine the situation when the number of mangers N increases to infinity with the hope of getting some more explicit qualitative results with behavioral implications. In this section, we assume that the number of assets d is not affected by the increase of the number of managers, see, however, the examples of Section 5.3. We also specialize the discussion to the case where the agents have similar preferences and only differ by their specific access to market.

The following result is similar to Proposition 5.1 in Espinosa (2010). Therefore, the proof is omitted.

PROPOSITION 4.12. *Let* $\lambda_j = \lambda \in [0, 1)$ *and* $\eta_j = \eta > 0$ *for all* $j \ge 1$. *Assume* $\frac{1}{N}\sum_{i=1}^N P^i_t \longrightarrow U^1_t$ *uniformly on any compact subsets, for all* $t \in [0, T]$ *(respectively, uniformly on* $[0, T] \times K$, *for any compact subset* K *of* $\mathbb{R}^d$*). Then,*

$$\hat{\pi}^{i,N}_t \longrightarrow \hat{\pi}^{i,\infty}_t := \sigma(t)^{-1} \circ P^i_t \circ \big(I - U^1_t \circ (\lambda I)\big)^{-1}\big(\eta_i\theta(t) + U^1_t(0)\big)$$

for all $t \in [0, T]$ *(respectively, uniformly in* $t \in [0, T]$*).*

4.5. Proof of Theorem 4.7

Assume that $(\tilde{\pi}^1, \dots, \tilde{\pi}^N)$ is a Nash equilibrium for our problem. First, by Hölder's inequality, the admissibility conditions for all $i = 1, \dots, N$ imply that $e^{-\frac{1}{\eta_i}(X^i_T - \lambda_i \bar{X}^i_T)}$ belongs to $\mathbb{L}^p$, for any $p > 0$. Let $\mathcal{T}$ be the set of all stopping times with values in $[0, T]$, we define the following family of random variables:

$$J^{i,\pi}(\tau) := \mathbb{E}\big[-e^{-\frac{1}{\eta_i}(\int_\tau^T \sigma_u\pi_u\cdot dB_u-\lambda_i(\bar{X}^i_T-\bar{x}^i))}|\mathcal{F}_\tau\big], \tag{4.23}$$

$$\mathcal{V}^i(\tau) := \operatorname*{ess\,sup}_{\pi\in\mathcal{A}_i} J^{i,\pi}(\tau) \quad \text{for all } \tau\in\mathcal{T} \quad \text{so that } \mathcal{V}^i(0)=e^{\frac{1}{\eta_i}(x^i-\lambda_i\bar{x}^i)}V_0^i. \tag{4.24}$$

1. By Lemma 4.13 below, the family $\{\mathcal{V}^i(\tau);\tau\in\mathcal{T}\}$ satisfies a supermartingale property. Indeed, let $\beta_t^{i,\pi} := e^{-\frac{1}{\eta_i}\int_0^t\sigma(u)\pi_u\cdot dB_u}$ for all $\pi\in\mathcal{A}_i$, we have

$$\beta_\tau^{i,\pi}\mathcal{V}_\tau^i \ge \mathbb{E}(\beta_\theta^{i,\pi}\mathcal{V}_\theta^i|\mathcal{F}_\tau) \quad \text{for all stopping times } \tau\le\theta.$$

Then, we can extract a process $(\mathcal{V}_t^i)$ that is càdlàg and consistent with the family defined previously in the sense that $\mathcal{V}_\tau^i=\mathcal{V}^i(\tau)$ a.s. (see Karatzas and Shreve 1991, proposition I.3.14 p. 16, for more details). Moreover, this process also satisfies the dynamic programming principle stated in Lemma 4.13 so that for any $\pi\in\mathcal{A}_i$, the process $\beta^{i,\pi}\mathcal{V}^i$ is a $\mathbb{P}$-supermartingale.

The definition of a Nash equilibrium implies that $\tilde{\pi}^i$ is optimal for agent i, i.e.,

$$\begin{aligned}\mathcal{V}_0^i &= \sup_{\pi\in\mathcal{A}_i}\mathbb{E}-e^{-\frac{1}{\eta_i}(X_T^\pi-x^i-\lambda_i(\bar{X}^i_T-\bar{x}^i))}\\ &= \mathbb{E}-e^{-\frac{1}{\eta_i}(X_T^{\tilde{\pi}^i}-x^i-\lambda_i(\bar{X}^i_T-\bar{x}^i))},\end{aligned} \tag{4.25}$$

which implies that the process $\beta^{i,\tilde{\pi}^i}\mathcal{V}^i$ is a square integrable martingale, as the conditional expectation of a r.v. in $\mathbb{L}^2$.

2. We now show that the adapted and continuous process:

$$\gamma_t^i := X_t^{\tilde{\pi}^i}-x^i+\eta_i\ln\big(-\beta_t^{i,\tilde{\pi}^i}\mathcal{V}_t^i\big),\ t\in[0,T], \tag{4.26}$$

solves the required BSDE.

(a) First, by Jensen's inequality, and the fact that $\ln x\le x$ for any $x>0$, we have

$$\begin{aligned}-\frac{1}{\eta_i}\mathbb{E}\big[X_T^{\tilde{\pi}^i}-x^i-\lambda_i\big(\bar{X}^i_T-\bar{x}^i\big)|\mathcal{F}_t\big] &\le \ln\big(-\beta_t^{i,\tilde{\pi}^i}\mathcal{V}_t^i\big)\\ &\le \mathbb{E}\big[-e^{-\frac{1}{\eta_i}(X_T^{\tilde{\pi}^i}-x^i-\lambda_i(\bar{X}^i_T-\bar{x}^i))}|\mathcal{F}_t\big].\end{aligned} \tag{4.27}$$

By the admissibility conditions, both sides of (4.27) belong to $\mathbb{H}^2$, as conditional expectations of random variables in $\mathbb{L}^2$. Since $X^{\tilde{\pi}^i}$ is also in $\mathbb{H}^2$, we see that γ^i is in $\mathbb{H}^2$. Then, for all $\pi\in\mathcal{A}_i$, we have that

$$M_t^{i,\pi} := -e^{-\frac{1}{\eta_i}(X_t^\pi-x^i-\gamma_t^i)} = \tilde{M}_t^i e^{-\frac{1}{\eta_i}(X_t^\pi-X_t^{\tilde{\pi}^i})}, \quad t\in[0,T],$$

where $\tilde{M}^i=\beta^{i,\tilde{\pi}^i}\mathcal{V}^i$ is a square integrable martingale. By Hölder's inequality, it follows that $e^{-\frac{1}{\eta_i}(X_t^\pi-X_t^{\tilde{\pi}^i})}\in\mathbb{L}^p$ for all $p>0$. Then, $M^{i,\pi}$ is integrable.

(b) In this step, we prove that $M^{i,\pi}$ is a supermartingale for all $\pi\in\mathcal{A}_i$. Assume, to the contrary, that there exists $\pi\in\mathcal{A}_i$, $t\ge s$ and $A\in\mathbb{F}_s$, with $\mathbb{P}(A)>0$ and such that

$$\mathbb{E}\big(-e^{-\frac{1}{\eta_i}(X_t^\pi-x^i-\gamma_t^i)}|\mathcal{F}_s\big) > -e^{-\frac{1}{\eta_i}(X_s^\pi-x^i-\gamma_s^i)} \text{ on } A,$$

and let us work toward a contradiction. Define:

$$\hat{\pi}_u(\omega) := \pi_u(\omega)\mathbf{1}_{\{[s,t]\times A\}}(u,\omega) + \tilde{\pi}_u(\omega)\mathbf{1}_{\{([s,t]\times A)^c\}}(u,\omega).$$

Since $A \in \mathcal{F}_s$, using Hölder's inequality, we see that $\hat{\pi} \in \mathcal{A}_i$ and we have

$$\begin{aligned}\mathcal{V}_0^i &\geq \mathbb{E} - e^{-\frac{1}{\eta_i}(X_T^{\hat{\pi}} - x^i - \gamma_T^i)} \\ &= \mathbb{E}\big[\mathbb{E}\{- e^{-\frac{1}{\eta_i}(X_T^{\hat{\pi}} - x^i - \gamma_T^i)} | \mathcal{F}_t\}\big] = \mathbb{E} - e^{-\frac{1}{\eta_i}(X_t^{\hat{\pi}} - x^i - \gamma_t^i)}\end{aligned}$$

by the fact that $\hat{\pi} = \tilde{\pi}$ on $[t, T]$ and $\tilde{M}^i$ is a martingale. Since $\mathbb{P}[A] > 0$, $A \in \mathcal{F}_s$, and recalling the definition of $\hat{\pi}$, this implies that

$$\begin{aligned}\mathcal{V}_0^i &\geq \mathbb{E}\big[\mathbb{E}\{ - e^{-\frac{1}{\eta_i}(X_t^{\hat{\pi}} - x^i - \gamma_t^i)} | \mathcal{F}_s\}\big] \\ &= \mathbb{E}\big[\mathbb{E}\{ - e^{-\frac{1}{\eta_i}(X_t^{\pi} - x^i - \gamma_t^i)} | \mathcal{F}_s\} 1_A + \mathbb{E}\{ - e^{-\frac{1}{\eta_i}(X_t^{\tilde{\pi}} - x^i - \gamma_t^i)} | \mathcal{F}_s\} 1_{A^c})\big] \\ &> \mathbb{E} - e^{-\frac{1}{\eta_i}(X_s^{\tilde{\pi}} - x^i - \gamma_s^i)} = -e^{\frac{1}{\eta_i}\gamma_0^i} = \mathcal{V}_0^i,\end{aligned}$$

which provides the required contradiction.

(c) Since $\tilde{M}^i = \beta^{i,\tilde{\pi}^i}\mathcal{V}^i$ is a martingale, it follows from the martingale representation theorem that $\tilde{M}^i$ is an Itô process. Therefore, (4.26) implies that γ^i is also an Itô process defined by some coefficients b^i and ζ^i:

$$d\gamma_t^i = -b_t^i dt + \zeta_t^i \cdot dW_t \quad \text{with } (\gamma^i, \zeta^i) \in \mathbb{H}^2(\mathbb{R}) \times \mathbb{H}_{loc}^2(\mathbb{R}^d). \tag{4.28}$$

Moreover, by Jensen's inequality, $\ln(-M^{i,\tilde{\pi}^i})$ is a supermartingale, and by (4.27), it is bounded in $\mathbb{L}^2$. Therefore, it admits a Doob–Meyer decomposition $\ln(-M^{i,\tilde{\pi}^i}) = N + A$, where N is a (uniformly integrable) martingale and A a decreasing process. The martingale representation theorem then implies that there exists a process $\delta \in \mathbb{H}_{loc}^2(\mathbb{R}^d)$ such that $N_t = \int_0^t \delta_u \cdot dW_u$. Using (4.27) and (4.28), we get $\zeta_t^i = \sigma_t \tilde{\pi}_t^i + \eta_i \delta_t$.

(d) We next compute the drift of $M^{i,\pi}$. From the previous supermartingale and martingale properties of $M^{i,\pi}$ and $\tilde{M}^i$, respectively, together with (4.28), we get

$$\begin{aligned} b_t^i &\leq \frac{1}{2\eta_i}\left|\sigma_t\pi_t - \left(\zeta_t^i + \eta_i\theta_t\right)\right|^2 - \frac{\eta_i}{2}|\theta_t|^2 - \zeta_t^i \cdot \theta_t \quad \text{for all } \pi \in \mathcal{A}_i, \\ &\text{and } \mathrm{b}_t^i = \frac{1}{2\eta_i}\left|\sigma_t\tilde{\pi}_t^i - \left(\zeta_t^i + \eta_i\theta_t\right)\right|^2 - \frac{\eta_i}{2}|\theta_t|^2 - \zeta_t^i \cdot \theta_t. \end{aligned}$$

This implies that

$$\begin{aligned} \tilde{\pi}_t^i &= \sigma_t^{-1} P_t^i \left(\zeta_t^i + \eta_i\theta_t\right), \\ b_t^i &= f^i(t, \zeta_t^i) = \frac{1}{2\eta_i} d\left(\zeta_t^i + \eta_i\theta_t, \sigma_t A_i\right)^2 - \frac{\eta_i}{2}|\theta_t|^2 - \zeta_t^i \cdot \theta_t, \end{aligned} \tag{4.29}$$

and therefore $(\gamma^i, \zeta^i) \in \mathbb{H}^2(\mathbb{R}) \times \mathbb{H}_{loc}^2(\mathbb{R}^d)$ is a solution of the BSDE:

$$\begin{aligned} d\gamma_t^i &= \left(\zeta_t^i \cdot \theta_t + \frac{\eta_i|\theta_t|^2}{2} - \frac{1}{2\eta_i}\left|\left(I - P_t^i\right)\left(\zeta_t^i + \eta_i\theta_t\right)\right|^2\right) dt + \zeta_t^i \cdot dW_t, \\ \gamma_T^i &= \lambda_i\left(\bar{X}_T^i - \bar{x}_i\right) = \lambda_i^N \sum_{j \neq i} \int_0^T \tilde{\pi}_u^j \cdot \sigma_u(dW_u + \theta_u du). \end{aligned}$$

Recalling that $dB_t = dW_t + \theta_t dt$, we can write it:

$$(4.30)\qquad d\gamma_t^i = \left(\frac{\eta_i|\theta_t|^2}{2} - \frac{\eta_i}{2}\big|\big(I - P_t^i\big)\big(\zeta_t^i + \eta_i\theta_t\big)\big|^2\right)dt + \zeta_t^i \cdot dB_t,$$
$$\gamma_T^i = \lambda_i\big(\bar{X}_T^i - \bar{x}_i\big) = \lambda_i^N \sum_{j\neq i}\int_0^T \tilde{\pi}_u^j \cdot \sigma_u dB_u.$$

3. We finally put together the N BSDEs obtained in step 2. Since $(\tilde{\pi}^1, \dots, \tilde{\pi}^N)$ is a Nash equilibrium, equation (4.30) holds for each $i = 1, \dots, N$. Replacing the value of $\tilde{\pi}^j$ by (4.29) in the expression of γ^i and writing $\Gamma^i := \zeta^i + \eta_i\theta$, we see that (γ^i, Γ^i) must satisfy for each $t \in [0, T]$:

$$\gamma_t^i = \lambda_i^N \sum_{j\neq i}\int_0^T P_u^j\big(\Gamma_u^j\big)\cdot dB_u - \frac{\eta_i}{2}\int_t^T |\theta_u|^2 du + \frac{1}{2\eta_i}\int_t^T \big|\big(I - P_u^i\big)\big(\Gamma_u^i\big)\big|^2 du$$
$$- \int_t^T \big(\Gamma_u^i - \eta_i\theta_u\big)\cdot dB_u,$$

so that the adapted process $Y_t^i := \gamma_t^i - \frac{\eta_i}{2}\int_0^t |\theta_u|^2 du + \frac{\eta_i}{2}\int_0^t \theta_u \cdot dB_u - \lambda_i^N \sum_{j\neq i}\int_0^t P_u^j(\Gamma_u^j)\cdot dB_u$, $t \in [0, T]$, satisfies:

$$Y_t^i = \eta_i\xi + \frac{1}{2\eta_i}\int_t^T \big|\big(I - P_u^i\big)\big(\Gamma_u^i\big)\big|^2 du - \int_t^T \left(\Gamma_u^i - \lambda_i^N \sum_{j\neq i} P_u^j\big(\Gamma_u^j\big)\right)\cdot dB_u,$$

with $(Y^i, \Gamma^i) \in \mathbb{H}^2(\mathbb{R}) \times \mathbb{H}^2_{loc}(\mathbb{R}^d)$. We finally define

$$Z_t^i := \varphi_t^i(\Gamma_t) = \Gamma_t^i - \lambda_i^N \sum_{j\neq i} P_t^j\big(\Gamma_t^j\big).$$

Under (4.13), using Lemma 4.6, we know that φ_t is invertible. As a consequence, $(Y, Z) \in \mathbb{H}^2(\mathbb{R}^N) \times \mathbb{H}^2_{loc}(\mathbb{R}^{Nd})$ is a solution of the following system of BSDEs:

$$Y_0^i = \eta_i\xi + \frac{1}{2\eta_i}\int_0^T \big|\big(I - P_t^i\big)\big(\psi_t^i(Z_t)\big)\big|^2 dt - \int_0^T Z_t^i \cdot dB_t.$$

Moreover, for each i, the equilibrium portfolio is given by

$$\sigma(t)\tilde{\pi}_t^i = P_t^i[\psi_t(Z_t)^i], \quad t \in [0, T].$$

The following dynamic programming principle was used in step 1 of the previous proof.

LEMMA 4.13 (Dynamic Programming). *For any stopping times $\tau \leq \nu$ in $\mathcal{T}$, we have*

$$\mathcal{V}^i(\tau) = ess\sup_{\pi\in\mathcal{A}_i} \mathbb{E}\big[e^{-\frac{1}{\eta_i}\int_\tau^\nu \sigma_u\pi_u\cdot, dB_u}\mathcal{V}^i(\nu)|\mathcal{F}_\tau\big], \quad i = 1, \dots, N.$$

Proof. Let $\tau \leq \nu \leq T$ a.s. We first obtain by the tower property that

$$\begin{aligned}\mathcal{V}^i(\tau) &= \operatorname*{ess\,sup}_{\pi\in\mathcal{A}_i} \mathbb{E}\big[\mathbb{E}\big[-e^{-\frac{1}{\eta_i}(\int_\nu^T \sigma_u\pi_u\cdot dB_u - \lambda_i(\bar{X}^i_T - \bar{x}^i))}|\mathcal{F}_\nu\big]e^{-\frac{1}{\eta_i}\int_\tau^\nu \sigma(u)\pi_u\cdot dB_u}|\mathcal{F}_\tau\big]\\ &\leq \operatorname*{ess\,sup}_{\pi\in\mathcal{A}_i} \mathbb{E}\big[e^{-\frac{1}{\eta_i}\int_\tau^\nu \sigma_u\pi_u\cdot dB_u}\mathcal{V}^i(\nu)|\mathcal{F}_\tau\big].\end{aligned}$$

To prove the converse inequality, we fix $\pi^0 \in \mathcal{A}_i$ and we observe that $J^{i,\pi}(\nu)$ defined by (4.23) depends on π only through its values on $[\nu, T]$. Therefore, we have the identity:

$$\mathcal{V}^i(\nu) = \operatorname*{ess\,sup}_{\pi\in\mathcal{A}_i(\nu)} J^{i,\pi}(\nu), \quad \text{where } \mathcal{A}_i(\nu) := \{\pi \in \mathcal{A}_i; \pi = \pi^0 \quad \text{on } [0,\nu], dt\otimes d\mathbb{P}\text{-a.e}\}.$$

We next observe that the family $\{J^{i,\pi}(\nu), \pi \in \mathcal{A}_i(\nu)\}$ is closed under pairwise maximization. Indeed, let π_1, π_2 in $\mathcal{A}_i(\nu)$, $A := \{\omega \in \Omega; J^{i,\pi_1}(\nu)(\omega) \geq J^{i,\pi_2}(\nu)(\omega)\}$ and define the process $\pi := 1_A\pi_1 + 1_{\Omega\setminus A}\pi_2$. Since $\pi^1 = \pi^2 = \pi^0$ on $[\![0,\nu]\!]$, and $A \in \mathcal{F}_\nu$, it is immediate that π is predictable. We compute $\mathbb{E}e^{\pm\frac{p}{\eta_i}(X^\pi_\tau - X^\pi_\nu)} = \mathbb{E}e^{\pm\frac{p}{\eta_i}\int_\vartheta^\tau \sigma_t\pi^1_t\cdot dB_t}1_A + \mathbb{E}e^{\pm\frac{p}{\eta_i}\int_\vartheta^\tau \sigma_t\pi^2_t\cdot dB_t}1_{\Omega\setminus A}$ so that since $\pi^1, \pi^2 \in \mathcal{A}_i(\nu)$, the family $\{e^{\pm\frac{1}{\eta_i}(X^\pi_\tau - X^\pi_\vartheta)}; \vartheta \leq \tau \in \mathcal{T}\}$ is uniformly bounded in any $\mathbb{L}^p$, $p > 1$. Therefore, $\pi \in \mathcal{A}_i(\nu)$ and it is immediate that $J^{i,\pi}(\nu) = \max(J^{i,\pi_1}(\nu), J^{i,\pi_2}(\nu))$. Then, it follows from Theorem A.3 (p. 324 in Karatzas and Shreve 1998) that there exists a sequence $(\hat{\pi}_n)$ satisfying:

- $\forall n, \hat{\pi}_n = \pi^0$ on $[\![0,\nu]\!]$
- $(J^{i,\hat{\pi}_n}(\nu))$ is nondecreasing and converges to $\mathcal{V}^i(\nu)$.

Then, we have

$$J^{i,\hat{\pi}_n}(\tau) = \mathbb{E}[J^{i,\hat{\pi}_n}(\nu)e^{-\frac{1}{\eta_i}\int_\tau^\nu \sigma_u\pi^0_u\cdot dB_u}|\mathcal{F}_\tau].$$

Since $J^{i,\hat{\pi}_n}(\nu)$ is nondecreasing and converges to $\mathcal{V}^i(\nu)$, it follows from the monotone convergence theorem that

$$\mathcal{V}^i(\tau) \geq \mathbb{E}\big[e^{-\frac{1}{\eta_i}\int_\tau^\nu \sigma_u\pi^0_u\cdot dB_u}\mathcal{V}^i(\nu)|\mathcal{F}_\tau\big],$$

and the required inequality follows from the arbitrariness of π^0. □

4.6. Proof of Theorem 4.8

1. We first prove that the portfolio (4.22) is indeed a Nash equilibrium. The idea is to show that we can make the formal computations of Section 4.1 in the reverse sense.

(a) Let

$$\xi_t := \int_0^t \theta(u).dB_u - \frac{1}{2}\int_0^T |\theta(u)|^2 du, \quad t \in [0,T], \tag{4.31}$$

Since θ and σ are deterministic and continuous functions, the functions P^is are also deterministic and continuous w.r.t. $(t,z) \in [0,T]\times\mathbb{R}^d$. Let us prove that the same holds for ψ, and therefore for that $\hat{\pi}^i_t := \sigma(t)^{-1}P^i_t \circ \psi^i_t(Z_t)$ is deterministic and continuous w.r.t. $t \in [0,T]$. Indeed, it is immediate that φ is deterministic and continuous w.r.t. (t,ζ) so that ψ is a deterministic function of (t,z). Then, from Lemma 4.6, under condition (4.13), ψ_t is Lipschitz in z, uniformly in t, so that there exists a constant $K > 0$ such that for all $t \in [0,T]$, and all $z, z' \in \mathbb{R}^d$, $|\psi_t(z) - \psi_t(z')| \leq |z - z'|$. Let $t_n \to t$, $z \in \mathbb{R}^{Nd}$, and $\zeta := \varphi_t(z)$.

We define $z_n := \varphi_{t_n}(\zeta)$ for each n. Since φ is continuous w.r.t. t, $z_n \to z$, and we have, for all n, $\psi_{t_n}(z_n) = \zeta$, so that $|\psi_{t_n}(z) - \zeta| = |\psi_{t_n}(z) - \psi_{t_n}(z_n)| \le K|z - z_n| \to 0$. Therefore, ψ is continuous w.r.t. t. Then, if $z_n \to z$ and $t_n \to t$, we compute $|\psi_{t_n}(z_n) - \psi_t(z)| \le |\psi_{t_n}(z_n) - \psi_{t_n}(z)| + |\psi_{t_n}(z) - \psi_t(z)| \to 0$, since ψ is continuous w.r.t. t and Lipschitz in z uniformly in t. As a consequence, we can define the following adapted and continuous processes:

$$Z_t^i := \eta_i \theta(t) \quad \text{and } Y_t^i := \eta_i \xi_t + \frac{1}{2\eta_i}\int_t^T \left|\left(I - P_u^i\right) \circ \psi_u^i(Z_u)\right|^2 du, \quad t \in [0, T].$$

Then, we directly verify that (Y, Z) satisfies the following N-dimensional BSDE:

$$Y_t^i = \eta_i \xi + \frac{1}{2\eta_i}\int_t^T \left|\left(I - P_u^i\right)\left(\psi_u^i(Z_u)\right)\right|^2 du - \int_t^T Z_u^i \cdot dB_u.$$

Set

$$\begin{aligned}
\gamma_t^i &= Y_t^i + \frac{\eta_i}{2}\int_0^t |\theta(u)|^2 du - \eta_i \int_0^t \theta(u) \cdot dB_u + \lambda_i^N \sum_{j \neq i} \int_0^t P_u^j\left(\psi_u^j(Z_u)\right) \cdot dB_u, \\
\zeta_t^i &= \psi_t(Z_t)^i - \eta_i \theta(t) = \left(\bar{\psi}_t^i - \eta_i I\right)(\theta(t)).
\end{aligned}$$

By the same computations as in Section 4.1, we see that for all $i = 1, \dots, N$, (γ^i, ζ^i) is a solution of the one-dimensional BSDE:

$$\begin{aligned}
d\gamma_t^i &= \left(\zeta_t^i \cdot \theta(t) + \frac{\eta_i |\theta(t)|^2}{2} - \frac{1}{2\eta_i}\left|\left(I - P_t^i\right)\left(\zeta_t^i + \eta_i \theta(t)\right)\right|^2\right) dt + \zeta_t^i \cdot dW_t, \\
\gamma_T^i &= \lambda_i^N \sum_{j \neq i} \int_0^T \hat{\pi}_u^j \cdot \sigma(u)(dW_u + \theta(u)\, du).
\end{aligned}$$

Then, using the definition of φ and ψ, we can rewrite γ^i as

$$\begin{aligned}
\gamma_t^i &= -\frac{\eta_i}{2}\int_0^t |\theta(u)|^2 du + \frac{1}{2\eta_i}\int_t^T \left|\left(I - P_u^i\right) \circ \bar{\psi}_u^i(\theta(u))\right|^2 du + \lambda_i^N \sum_{j \neq i} \int_0^t P_u^j[\psi_u(Z_u)]^j \cdot dB_u \\
&= -\frac{\eta_i}{2}\int_0^t |\theta(u)|^2 du + \frac{1}{2\eta_i}\int_t^T \left|\left(I - P_u^i\right) \circ \bar{\psi}_u^i(\theta(u))\right|^2 du + \int_0^t \zeta_u^i \cdot dB_u \\
&= -\frac{\eta_i}{2}\int_0^t |\theta(u)|^2 du + \frac{1}{2\eta_i}\int_t^T \left|\left(I - P_u^i\right) \circ \bar{\psi}_u^i(\theta(u))\right|^2 du + \int_0^t \left(\bar{\psi}_u^i - \eta_i I\right)(\theta(u)) \cdot dB_u.
\end{aligned}$$

(b) Throughout this step, we fix an integer $i \in \{1, \dots, N\}$, and we define

$$M_t^\pi := -e^{-\frac{1}{\eta_i}\left(X_t^{i,\pi} - x^i - \gamma_t^i\right)} \quad \text{for all } \pi \in \mathcal{A}_i.$$

By Itô's formula, it follows that M^π is a local supermartingale for each $\pi \in \mathcal{A}_i$, and $M^{\hat{\pi}^i}$ is a local martingale. Then, there exist increasing sequences of stopping times (τ_n^π) in $\mathcal{T}$ such that for each π, $\tau_n^\pi \to T$ a.s. and for each n and any $s \le t$:

$$\mathbb{E}\left[M_{t \wedge \tau_n^\pi}^\pi \middle| \mathcal{F}_s\right] \le M_{s \wedge \tau_n^\pi}^\pi \quad \text{for all} \quad \pi \in \mathcal{A}_i \quad \text{and} \quad \mathbb{E}\left[M_{t \wedge \tau_n^{\hat{\pi}}}^{\hat{\pi}^i} \middle| \mathcal{F}_s\right] = M_{s \wedge \tau_n^{\hat{\pi}}}^{\hat{\pi}^i}. \tag{4.32}$$

We next introduce the measure $\mathbb{Q}^i$, equivalent to $\mathbb{P}$, defined by its Radon–Nikodym density:

$$(4.33)\qquad L_t^i = \frac{d\mathbb{Q}^i}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = e^{\int_0^t (\frac{1}{\eta_i}\bar{\psi}_u^i - I)(\theta(u))\cdot dW_u - \frac{1}{2}\int_0^t |(\frac{1}{\eta_i}\bar{\psi}_u^i - I)(\theta(u))|^2 du}.$$

We denote by $\mathbb{E}^i$ the expectation operator under $\mathbb{Q}^i$. Since θ is a deterministic and continuous function on $[0, T]$, $-\frac{\eta_i}{2}\int_0^t |\theta(u)|^2 du + \frac{1}{2\eta_i}\int_t^T |(I - P_u^i) \circ \bar{\psi}_u^i(\theta(u))|^2 du$ is bounded. Then, for any $\pi \in \mathcal{A}_i$:

$$(4.34)\qquad \begin{aligned}\mathbb{E} M_{t\wedge\tau_n}^\pi = {}& \frac{1}{L_s^i}\mathbb{E}^i\Big[- e^{-\frac{1}{\eta_i}(X_{t\wedge\tau_n}^\pi - x^i) - \frac{1}{2}\int_0^{t\wedge\tau_n} |\theta(u)|^2 du + \frac{1}{2\eta_i^2}\int_{t\wedge\tau_n}^T |(I-P_u^i)\circ f_u^i(\theta(u))|^2 du} \\ & \times e^{\int_0^{t\wedge\tau_n}(\frac{1}{\eta_i}\bar{\psi}_u^i - I)(\theta(u))\cdot\theta(u)du + \frac{1}{2}\int_0^{t\wedge\tau_n} |(\frac{1}{\eta_i}\bar{\psi}_u^i - I)(\theta(u))|^2 du}\Big],\end{aligned}$$

where we simply denoted $\tau_n := \tau_n^\pi$. In (4.34), all the terms inside the expectation other than $e^{-\frac{1}{\eta_i}X_{t\wedge\tau_n}^\pi}$ are bounded. We shall prove in step 1(c) below that the family $\{e^{-\frac{1}{\eta_i}X_\tau^\pi}; \tau \in \mathcal{T}\}$ is uniformly integrable under $\mathbb{Q}^i$. Hence, the sequence of processes inside the expectation in (4.34) is uniformly integrable under $\mathbb{Q}^i$, and we may apply the dominated convergence theorem to pass to the limit $n \to \infty$, and we obtain $\lim_{n\to\infty} \mathbb{E} M_{t\wedge\tau_n}^\pi = \mathbb{E} M_t^\pi$. Together with (4.32), this implies that

$$\begin{aligned}&\mathbb{E} - e^{-\frac{1}{\eta_i}(X_t^\pi - x^i - \gamma_t^i)} \le -e^{\frac{1}{\eta_i}\gamma_0^i} \quad \text{for all } \pi \in \mathcal{A}_i \text{ and}\\ &\mathbb{E} - e^{-\frac{1}{\eta_i}(X_t^{\hat{\pi}^i} - x^i - \gamma_t^i)} = -e^{\frac{1}{\eta_i}\gamma_0^i}.\end{aligned}$$

Multiplying by $e^{-\frac{1}{\eta_i}(x^i - \lambda_i \bar{x}^i)}$, we finally get $V_i = -e^{-\frac{1}{\eta_i}(x^i - \lambda_i\bar{x}^i - Y_0^i)}$, since $Y_0^i = \gamma_0^i$, and $\hat{\pi}^i$ is optimal for agent i. Hence, $(\hat{\pi}^1, \dots, \hat{\pi}^N)$ is a Nash equilibrium.

(c) In this step, we prove that the family $\{Y_\tau := e^{-\frac{1}{\eta_i}X_\tau^{i,\pi}} : \tau \in \mathcal{T}\}$ is $\mathbb{Q}^i$ uniformly integrable for all $\pi \in \mathcal{A}_i$. Fix some $p > 1$. Then, by the admissibility condition, the family $\{Y_\tau : \tau \in \mathcal{T}\}$ is uniformly bounded in $\mathbb{L}^p(\mathbb{P})$. With $r := (1 + p)/2$, it follows that the family $\{Y_\tau^r : \tau \in \mathcal{T}\}$ is uniformly integrable. Then, for all $c > 0$ and $\tau \in \mathcal{T}$, it follows from Hölder's inequality:

$$\mathbb{E}^{\mathbb{Q}^i}[Y_\tau 1_{Y_\tau \ge c}] = \mathbb{E}\big[L_T^i Y_\tau 1_{Y_\tau\ge c}\big] \le \big\|L_T^i\big\|_{\mathbb{L}^q(\mathbb{P})}\|Y_\tau^r 1_{Y_\tau\ge c}\|_{\mathbb{L}^r(\mathbb{P})},$$

where q is defined by $(1/q) + (1/r) = 1$. Since $\{Y^r : \tau \in \mathcal{T}\}$ is uniformly integrable, the last term uniformly goes to 0 as $c \to \infty$.

REMARK 4.14. We mention again that based on a first version of this paper contained in the PhD thesis Espinosa (2010), Frei and dos Reis developed many interesting extensions in Frei and dos Reis (2011). In particular, their argument for the existence part of Theorem 4.8 is shorter than our previous step 1.

2. We now prove uniqueness by using a verification argument.

(a) Let $(\pi^1, \dots, \pi^N)$ be a deterministic Nash equilibrium, and define for all $i = 1, \dots, N$:

$$(4.35)\qquad u^i(t, x, y) := -e^{-\frac{1}{\eta_i}(x - \lambda_i y) - \frac{1}{2}\int_t^T |\theta(u)|^2 du + \frac{1}{2\eta_i^2}\int_t^T |(I - P_u^i)(\eta_i\theta(u) + \lambda_i\sigma(u)\bar{\pi}_N^i(u))|^2 du},$$

where

$$\bar{\pi}^i_N(u) := \frac{1}{N-1}\sum_{j\neq i}\pi^j(u).$$

Since π^j is a continuous function for all $j = 1, \dots, N$, the functions u^i are C^1 in the t variable. Direct calculation reveals that u^i is a classical solution of the equation:

$$-\partial_t u^i - \sup_{p\in A_i} L^p u^i = 0 \quad \text{and} \quad u^i(T, x, y) = -e^{-(x-\lambda_i y)/\eta_i},$$

where for all $p \in A_i$, L^p is the linear second-order differential operator:

$$\begin{aligned} L^p &:= \sigma(t)\bar{\pi}^i_N(t)\cdot\theta(t)\partial_y + \frac{1}{2}\big|\sigma(t)\bar{\pi}^i_N(t)\big|^2\partial^2_{yy} \\ &\quad + \sigma(t)p.\theta(t)\partial_x + \sigma(t)p\cdot\sigma(t)\bar{\pi}^i_N(t)\partial^2_{xy} + \frac{1}{2}|\sigma(t)p|^2\partial^2_{xx}, \end{aligned}$$

and the supremum is attained at a unique point

$$\pi^*_t := \sigma(t)^{-1}P^i_t\big(\eta_i\theta(t) + \lambda_i\sigma(t)\bar{\pi}^i_N(t)\big). \tag{4.36}$$

(b) In this step, we prove that $u^i(0, X^i_0, \bar{X}^i_0) = V^i_0$. First, by Itô's formula, we have for all $\pi \in \mathcal{A}_i$:

(4.37)
$$u^i(t, x, y) = u^i\big(\tau_n, X^\pi_{\tau_n}, \bar{X}^i_{\tau_n}\big) - \int_t^{\tau_n} L^\pi u^i\big(r, X^\pi_r, \bar{X}^i_r\big)\,dr - \int_t^{\tau_n}\big(\pi - \bar{\pi}^i_N\big)(r)\cdot\sigma(r)\,dW_r,$$

where $\tau_n := \inf\{r \ge t, |X^\pi_r - x| \ge n \text{ or } |\bar{X}^i_r - \bar{x}^i| \ge n\}$. Taking conditional expectations in (4.37), and using the fact that $L^{\pi_t}u^i \le 0$ for any $\pi \in \mathcal{A}_i$, we get

$$u^i(t, x, y) \ge \mathbb{E}_{t,x,y}u^i(\tau_n, X^\pi_{\tau_n}, \bar{X}^i_{\tau_n}) \quad \text{for all } \pi \in \mathcal{A}_i. \tag{4.38}$$

Since the π_js, σ, and θ are continuous deterministic functions and $\pi^i \in \mathcal{A}_i$, it follows from Hölder's inequality that $\{e^{-\frac{1}{\eta_i}(X^i_\tau - \lambda_i\bar{X}^i_\tau)}, \tau \in \mathcal{T}\}$ is uniformly bounded in any $\mathbb{L}^p$. By the definition of U^i, this property is immediately inherited by the family $\{u^i(\tau, X^\pi_\tau, \bar{X}^i_\tau), \tau \in \mathcal{T}\}$. Therefore, taking the limit $n \to \infty$ in (4.38), we get $u^i(t, x, y) \ge \mathbb{E}_{t,x,y}e^{-\frac{1}{\eta_i}(X^\pi_T - \lambda_i\bar{X}^i_T)}$. By the arbitrariness of $\pi \in \mathcal{A}_i$, this implies that $u^i(0, X^i_0, \bar{X}^i_0) \ge V^i_0$.

We next observe that $\pi^* \in \mathcal{A}_i$ and the inequality in (4.38) is turned into an equality if π^* is substituted to π. By the dominated convergence theorem, this provides

$$u^i(t, x, y) = \mathbb{E}_{t,x,y}e^{-\frac{1}{\eta_i}(X^{\pi^*}_T - \lambda_i\bar{X}^i_T)},$$

which, in view of (4.38), shows that $u^i(0, X^i_0, \bar{X}^i_0) = V^i_0$.

(c) To see that the continuous deterministic Nash equilibrium is unique, consider another continuous deterministic Nash equilibrium $(\hat{\pi}^1, \dots, \hat{\pi}^N)$, and denote by $\hat{u}^i$ the corresponding value functions as in (4.35). It suffices to observe that $L^\pi\hat{u}^i < 0$ on any nonempty open subset B of $[0, T]$ such that $\pi \neq \pi^*$ on B, and the inequality (4.38) is strict. Therefore, any Nash equilibrium must satisfy (4.36) for every $i = 1, \dots, N$.

Set $\hat{\gamma}_t^i := \sigma(t)\hat{\pi}_t^i$, and let $\hat{\gamma}$ be the matrix whose ith line is $\hat{\gamma}^i$. From the previous argument, $(\hat{\pi}^1, \ldots, \hat{\pi}^N)$ is a Nash equilibrium if and only:

$$\Gamma_t^i(\hat{\gamma}_t) := P_t^i\left(\eta_i\theta(t) + \lambda_i^N \sum_{j\neq i} \hat{\gamma}_t^j\right) = \hat{\gamma}_t^i, \qquad i = 1, \ldots, N, t \in [0, T], \tag{4.39}$$

i.e., $\hat{\gamma}_t$ is a fixed point of Γ_t for all $t \in [0, T]$. Using Lemma 4.15 below, we have the uniqueness of a Nash equilibrium. Finally, the expression for V^i at equilibrium follows from the last statement of Lemma 4.15 together with (4.35).

Recall the function $\bar{\psi}^i$ defined in (4.21).

LEMMA 4.15. *Under (4.13), the function Γ_t defined in (4.39) has a unique fixed point $\hat{\gamma}_t$ for all $t \in [0, T]$, given by*

$$\hat{\gamma}_t^i = P_t^i \circ \bar{\psi}_t^i(\theta) \quad \textit{and satisfying } \bar{\psi}_t^i(\theta) = \eta_i\theta + \lambda_i^N \sum_{j\neq i} \hat{\gamma}_t^j.$$

Proof.

1. Since P_t^i is a contraction, we compute

$$|\Gamma_t(x_1) - \Gamma_t(x_2)|_1 := \sum_{i=1}^N \left|\Gamma_t^i(x_1) - \Gamma_t^i(x_2)\right| \le \sum_{i=1}^N \frac{1}{N-1} \sum_{j\neq i} \left|x_1^j - x_2^j\right| = |x_1 - x_2|_1,$$

proving that Γ_t is a contraction.

2. We next show that $(\Gamma_t)^2 := \Gamma_t \circ \Gamma_t$ is a strict contraction. Indeed, under (4.13), we may assume without loss of generality that $\lambda_1 < 1$. Then,

$$\left|\Gamma_t^i \circ \Gamma_t(x_1) - \Gamma_t^i \circ \Gamma_t(x_2)\right| \le \lambda_i^N \sum_{j\neq i} \left|\Gamma_t^j(x_1) - \Gamma_t^j(x_2)\right| \le \lambda_i^N \sum_{j\neq i} \lambda_j^N \sum_{k\neq j} \left|x_1^k - x_2^k\right|,$$

so that

$$\begin{aligned} |(\Gamma_t)^2(x_1) - (\Gamma_t)^2(x_2)|_1 &\le \sum_{i=1}^N \sum_{j\neq i} \sum_{k\neq j} \lambda_i^N \lambda_j^N \left|x_1^k - x_2^k\right| \\ &\le \frac{\lambda_1}{(N-1)^2}\left(\sum_{k=1}^N (N-2)\left|x_1^k - x_2^k\right| + \sum_{k\neq 1} \left|x_1^k - x_2^k\right|\right) \\ &\quad + \frac{N-2}{N-1}\left|x_1^1 - x_2^1\right| + \frac{N-1+(N-2)^2}{(N-1)^2} \sum_{k\neq 1} \left|x_1^k - x_2^k\right| \\ &\le \left(\frac{\lambda_1(N-2)}{(N-1)^2} + \frac{(N-2)^2 + N - 1}{(N-1)^2}\right) |x_1 - x_2|_1. \end{aligned}$$

Observe that $N - 2 + N - 1 + (N-2)^2 = (N-1)^2$. Then, $\lambda_1 < 1$ implies that $(\Gamma_t)^2$ is a strict contraction.

3. Therefore, $(\Gamma_t)^n$ is a strict contraction as well for any $n \ge 2$. As a consequence, $(\Gamma_t)^2$, $(\Gamma_t)^3$, and $(\Gamma_t)^6$, respectively, admit a unique fixed point x_2, x_3, and x_6, respectively. It is immediate that x_2 and x_3 are also fixed points for $(\Gamma_t)^6$; therefore, $x_2 = x_3 = x_6$, and finally, $x_2 = (\Gamma_t)^3(x_2) = \Gamma_t \circ (\Gamma_t)^2(x_2) = \Gamma_t(x_2)$ so that x_2 is a

fixed point of Γ_t. The uniqueness is immediate since a fixed point of Γ_t is also a fixed point of $(\Gamma_t)^2$.

4. Let $\Theta \in \mathbb{R}^{Nd}$ be defined by $\Theta^i = \eta_i \theta$. By definition of ψ_t in Lemma 4.6, $\varphi_t^i \circ \psi_t(\Theta) = \eta_i \theta$ for all $i = 1, \dots, N$. Using the definition of φ_t in (4.9), this implies that

$$\psi_t^i(\Theta) = \eta_i \theta + \lambda_i^N \sum_{j \neq i} P_t^j \circ \psi_t^j(\Theta). \tag{4.40}$$

Applying P_t^i and setting $\hat{\gamma}_t^i = P_t^i \circ \psi_t^i(\Theta)$, this provides $\hat{\gamma}_t^i = \Gamma_t^i(\hat{\gamma}_t)$, for each $i = 1, \dots, N$. By the definition of $\bar{\psi}^i$ together with the expression of ψ, we have $\bar{\psi}_t^i(\theta) = \psi_t^i(\Theta)$ so that $\hat{\gamma}_t^i = P_t^i \circ \bar{\psi}_t^i(\theta)$. Plugging it into (4.40) provides the last statement of the lemma. □

5. LINEAR PORTFOLIO CONSTRAINTS

We now focus on the case where the sets of constraints are such that

$$A_i \text{ is a vector subspace of } \mathbb{R}^d, \quad \text{for all } i = 1, \dots, N. \tag{5.1}$$

Our main objective in this section is to exploit the linearity of the projection operators P^i in order to derive more explicit results.

5.1. Nash Equilibrium Under Linear Portfolio Constraints

In the present context, we show that condition (4.13) in Theorem 4.8 can be weakened to

$$\prod_{i=1}^N \lambda_i < 1 \quad \text{or} \quad \bigcap_{i=1}^N A_i = \{0\}. \tag{5.2}$$

In view of Lemma 4.4 (which is obvious in the present linear case), the map

$$R_t^i := \frac{1}{N-1} \sum_{j \neq i} \lambda_j P_t^j \left(I + \lambda_j^N P_t^j\right)^{-1} \left(I + \lambda_i^N P_t^i\right) \tag{5.3}$$

is well defined. Moreover, for any $j = 1, \dots, N$, since P_t^j is a projection, we compute that $(I + \lambda_j^N P_t^j)^{-1} = I - \frac{\lambda_j^N}{1+\lambda_j^N} P_t^j$ so that

$$R_t^i = \sum_{j \neq i} \frac{\lambda_j^N}{1 + \lambda_j^N} P_t^j \left(I + \lambda_i^N P_t^i\right).$$

The following statement is more precise than Lemma 4.6.

LEMMA 5.1. *Let $(A_i)_{1 \leq i \leq N}$ be vector subspaces of $\mathbb{R}^d$. Then, for all $t \in [0, T]$:*

(i) the linear operator φ_t is invertible if and only if (5.2) is satisfied,
(ii) this condition is equivalent to the invertibility of the linear operators $I - R_t^i$, $i = 1, \dots, N$,

(iii) under (5.2), the ith component of $\psi_t = \varphi_t^{-1}$ is given by

$$\psi_t^i(z) = (I - R_t^i)^{-1} \left(z^i + \sum_{j \neq i} \frac{1}{1 + \lambda_j^N} P_t^j \left(\lambda_i^N z^j - \lambda_j^N z^i \right) \right).$$

The proof of this lemma is reported in Section 5.4. We now proceed to the characterization of Nash equilibria in the context of the multivariate Black–Scholes financial market. From Lemma 5.1, if condition (5.2) is satisfied, $\bar{\psi}^i$ defined by (4.21) is well defined, is a linear operator, and has the following expression:

(5.4)

$$\bar{\psi}_t^i = M_t^i := \left(I - \sum_{j \neq i} \frac{\lambda_j^N}{1 + \lambda_j^N} P_t^j \left(I + \lambda_i^N P_t^i \right) \right)^{-1} \left(\eta_i I + \sum_{j \neq i} \frac{\lambda_i^N \eta_j - \lambda_j^N \eta_i}{1 + \lambda_j^N} P_t^j \right).$$

THEOREM 5.2. *Assume that σ and θ are deterministic, and (5.2) is satisfied. Then, there exists a unique deterministic Nash equilibrium given by*

$$\hat{\pi}_t^i = \sigma(t)^{-1} P_t^i M_t^i \theta(t) \quad \text{for } i = 1, \ldots, N, t \in [0, T].$$

Moreover, the value function for agent i at equilibrium is given by

$$V_i = -e^{-\frac{1}{\eta_i}(x^i - \lambda_i \bar{x}^i - Y_0^i)}$$

where

$$Y_0^i = -\frac{\eta_i}{2} \int_0^T |\theta(t)|^2 dt + \frac{1}{2\eta_i} \int_0^T \left| \left(I - P_t^i \right) M_t^i \theta(t) \right|^2 dt.$$

Proof. Follow the lines of the proof of Theorem 4.8, replacing Lemma 4.6 by Lemma 5.1 and Lemma 4.15 by the following Lemma 5.3. □

LEMMA 5.3. *Let $\theta \in \mathbb{R}^d$ be arbitrary and $\Gamma : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$ be defined for any $\gamma \in \mathbb{R}^{Nd}$ by*

$$\Gamma^i(\gamma) = P^i \left(\eta_i \theta + \lambda_i^N \sum_{j \neq i} \gamma^j \right).$$

Then, under (5.2), Γ admits a unique fixed point $\hat{\gamma}$ given by $\hat{\gamma}^i = P^i \bar{\psi}^i(\theta)$.

The proof of this lemma is reported in Section 5.4. We illustrate the previous Nash equilibrium in the context of symmetric managers with different access to the financial market.

EXAMPLE 5.4 (Similar agents with different investment constraints). Assume that σ and θ are deterministic, and let $\lambda_j = \lambda \in [0, 1)$ and $\eta_j = \eta > 0$, $j = 1, \ldots, N$. Then, there exists a unique deterministic Nash equilibrium given by

$$\hat{\pi}_t^i = \eta \sigma(t)^{-1} P_t^i \left(I - \frac{\lambda^N}{1 + \lambda^N} \sum_{j \neq i} P_t^j \left(I + \lambda^N P_t^i \right) \right)^{-1} \theta(t), \quad i = 1, \ldots, N.$$

We conclude this section with the following qualitative result that shows, in particular, that the managers' interactions induce an overinvestment on the risky assets, and imply that the market portfolio $\bar{\pi}$ of Definition 3.9 is nondecreasing in the interaction coefficients λ_i, in agreement with Proposition 3.10. This result requires a quite restrictive condition that, however, covers many examples, see also Remark 5.6 below.

PROPOSITION 5.5. *Assume that the projection operators P^i commute, i.e., $P^i P^j = P^j P^i$ for all $i, j = 1, \dots, N$. Then, under the conditions of Theorem 5.2, Agent i's equilibrium portfolio is such that $|\sigma(t)\hat{\pi}_t^i|$ is nondecreasing w.r.t. λ_j and η_j, for all $i, j = 1, \dots, N$ and $t \in [0, T]$.*

Proof. We fix an agent $i = 1, \dots, N$, and omit all t-dependence. The assumption that the P^i's commute is equivalent to the existence of an orthonormal basis $\{u_k, k = 1, \dots, d\}$ such that for all i, u_k is an eigenvector of P_i for all k. We write $P^i u_k = \varepsilon_{i,k} u_k$, and we observe that $\varepsilon_{i,k} \in \{0, 1\}$ by the fact that P^i is a projection. Then, by the explicit expression of $\hat{\pi}^i$ in Theorem 5.2, writing $\theta = \sum_{k=1}^d \theta^k u_k$, we directly compute that $|\sigma\hat{\pi}^i|^2 = \sum_{k=1}^d (\theta^k)^2 (\ell_{i,k})^2$, where

$$\ell_{i,k} = \varepsilon_{i,k} \left(1 - \sum_{m \neq i} \frac{\lambda_m^N \varepsilon_{m,k}}{1 + \lambda_m^N} \left(1 + \lambda_i^N \varepsilon_{i,k}\right)\right)^{-1} \left(\eta_i + \sum_{m \neq i} \frac{\lambda_i^N \eta_m - \lambda_m^N \eta_i}{1 + \lambda_m^N} \varepsilon_{m,k}\right). \tag{5.5}$$

We now verify that $\ell_{i,k}$ is nondecreasing w.r.t. λ_j and η_j, for all $j = 1, \dots, N$ and $k = 1, \dots, d$, which implies the required result by the orthogonality of the basis $\{u_k, k = 1, \dots, d\}$.

- That $\ell_{i,k}$ is nondecreasing in η_j is obvious from (5.5).
- That $\ell_{i,k}$ is nondecreasing in λ_i is also obvious from (5.5).
- Finally, for $j \neq i$, we directly differentiate (5.5), and see that the sign of $\partial \ell_{i,k} / \partial \lambda_j^N$ is given by the sign of

$$\begin{aligned} &\varepsilon_{i,k}\varepsilon_{j,k} \left(\left(1 + \lambda_i^N \varepsilon_{i,k}\right) \left(\eta_i + \sum_{m \neq i} \frac{\lambda_i^N \eta_m - \lambda_m^N \eta_i}{1 + \lambda_m^N} \varepsilon_{m,k}\right) \right. \\ &\qquad \left. - \eta_i \left(1 - \sum_{m \neq i} \frac{\lambda_m^N \varepsilon_{m,k}}{1 + \lambda_m^N} \left(1 + \lambda_i^N \varepsilon_{i,k}\right)\right)\right) \\ &\quad = \varepsilon_{i,k}\varepsilon_{j,k} \left(\lambda_i^N \eta_i + \lambda_i^N \left(1 + \lambda_i^N \sum_{m \neq i} \frac{\eta_m}{1 + \lambda_m^N} \varepsilon_{m,k}\right)\right) \geq 0. \end{aligned}$$

□

REMARK 5.6. The statement of Proposition 5.5 is not valid for general portfolio constraints, as illustrated by the following example. Let $N = d = 2$, $A_1 = \mathbb{R}e_1$, $A_2 = \mathbb{R}(e_1 + e_2)$, and $\sigma = I$. Then, the projection operators P^1 and P^2 are defined by the following matrices in the basis (e_1, e_2):

$$P^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad P^2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix},$$

respectively. By direct calculation, $|\hat{\pi}^1| = \frac{1}{2-\lambda_1\lambda_2}|(2\eta_1 + \lambda_1\eta_2)\theta_1 + \lambda_1\eta_2\theta_2|$, which can be increasing or decreasing in η_i and λ_i, $i = 1, 2$ for appropriate choices of the risk premium θ.

5.2. Infinite Managers Asymptotics

We now investigate the limiting behavior when the number of agents N goes to infinity with fixed number of assets d.

Recall that $|.|$ denotes the canonical Euclidean norm on $\mathbb{R}^d$, and $\mathcal{L}(\mathbb{R}^d)$ is the space of linear mappings on $\mathbb{R}^d$ endowed with operator norm $\|U\| = \sup_{|x|=1} |U(x)|$ for all $U \in \mathcal{L}(\mathbb{R}^d)$.

PROPOSITION 5.7. *Let d be fixed and the sequence $(\eta_i)_{i\in\mathbb{N}}$ bounded in $\mathbb{R}$. Assume that*

$$\frac{1}{N}\sum_{i=1}^{N} \lambda_i P_t^i \longrightarrow U_t^\lambda in\mathcal{L}(\mathbb{R}^d) \quad \textit{and} \quad \frac{1}{N}\sum_{i=1}^{N} \eta_i P_t^i \to U_t^\eta \quad \textit{in} \quad \mathcal{L}(\mathbb{R}^d), \tag{5.6}$$

for all (respectively, uniformly in) $t \in [0, T]$. Assume further that $\|U_t^\lambda\| < 1$, $t \in [0, 1]$. Then,

$$\hat{\pi}_t^{i,N} \longrightarrow \hat{\pi}_t^{i,\infty} := \sigma(t)^{-1} P_t^i (I - U_t^\lambda)^{-1}\big(\eta_i\big(I - U_t^\lambda\big) + \lambda_i U_t^\eta\big)\theta(t)$$

for all (respectively, uniformly in) $t \in [0, T]$.

Proof. By Theorem 5.2, we have $\hat{\pi}_t^{i,N} = \sigma(t)^{-1} P_t^i A_t^i B_t^i \theta(t)$, where

$$A_t^i := \left(I - \sum_{j\neq i} \frac{\lambda_j^N}{1+\lambda_j^N} P_t^j (I + \lambda_i^N P_t^i)\right)^{-1} \quad \text{and} \quad B_t^i := \eta_i I + \sum_{j\neq i} \frac{\lambda_i^N \eta_j - \lambda_j^N \eta_i}{1+\lambda_j^N} P_t^j.$$

Since $\| P_t^j \| \leq 1$, we have

$$\begin{aligned}
&\left\| \frac{1}{N-1}\sum_{j\neq i} \frac{\lambda_j}{1+\lambda_j^N} P_t^j - \frac{1}{N}\sum_{j=1}^{N} \lambda_j P_t^j \right\| \\
&\quad \leq \frac{1}{N-1}\left\| \sum_{j\neq i} \frac{\lambda_j}{1+\lambda_j^N} P_t^j - \lambda_j P_t^j \right\| + \left\| \frac{1}{N-1}\sum_{j\neq i} \lambda_j P_t^j - \frac{1}{N}\sum_{j=1}^{N} \lambda_j P_t^j \right\| \\
&\quad \leq \left\| \frac{1}{(N-1)^2}\sum_{j\neq i} \frac{\lambda_j^2}{1+\lambda_j^N} P_t^j \right\| + \left\| \frac{1}{N}\lambda_i P_t^i + \frac{1}{N(N-1)}\sum_{j\neq i} \lambda_j P_t^j \right\| \leq \frac{3}{N}.
\end{aligned}$$

Similarly, by the boundedness of the sequence $(\eta_i)_{i\geq 1}$:

$$\left\| \frac{1}{N-1}\sum_{j\neq i} \frac{\eta_j}{1+\lambda_j^N} P_t^j - \frac{1}{N}\sum_{j=1}^{N} \eta_j P_t^j \right\| \leq \frac{3|\eta|_\infty}{N}.$$

Then, as $N \to \infty$, we have in $\mathcal{L}(\mathbb{R}^d)$

$$I + \lambda_i^N P_t^i \to I, \quad \frac{1}{N-1}\sum_{j \neq i} \frac{\lambda_j}{1+\lambda_j^N} P_t^j \longrightarrow U_t^\lambda, \quad \frac{1}{N-1}\sum_{j \neq i} \frac{\eta_j}{1+\lambda_j^N} P_t^j \longrightarrow U_t^\eta,$$

and $A_t^i \to (I - U_t^\lambda)^{-1}$, $B_t^i \to \eta_i I + \lambda_i U_t^\eta - \eta_i U_t^\lambda$. Under the condition $\|U_t^\lambda\| < 1$, the limit is finite. Moreover, the convergence is uniform in t whenever the convergence (5.6) holds uniformly in t. □

EXAMPLE 5.8 (Symmetric agents with different access to the financial market). Let $\lambda_i = \lambda \in [0, 1)$ and $\eta_i = \eta > 0$, $i \geq 1$. Then, the limiting Nash equilibrium portfolio reduces to

$$\hat{\pi}_t^{i,\infty} = \eta\sigma(t)^{-1} P_t^i \big(I - \lambda U_t^1\big)^{-1}\theta(t), \quad t \in [0, T], i \geq 1.$$

EXAMPLE 5.9 (Symmetric agents with finite market access possibilities). In the context of the previous example, suppose further that $\{A_i, i \geq 1\} = \{A_j, j = 1, \dots, p\}$ for some integer $p > 1$. We denote by k_j^N the number of agents with portfolio constraint A_j, and we assume that $k_j^N/N \longrightarrow \kappa_j \in [0, 1]$ for all $j = 1, \dots, p$. Then, an immediate application of Proposition 5.7 provides the limit Nash equilibrium portfolio:

$$\hat{\pi}_t^{i,\infty} = \eta\sigma(t)^{-1} P_t^i \left(I - \lambda \sum_{j=1}^{p} \kappa_j P_t^j\right)^{-1} \theta(t).$$

REMARK 5.10. We may also adopt the following probabilistic point of view to reformulate Proposition 5.7. Assume that there is a continuum of independent players modeled through a probability space $(\Delta, \mathcal{D}, \mu)$ independent from the space $(\Omega, \mathcal{F}, \mathbb{P})$ describing the financial market uncertainty. In such a setting, the market interactions, the risk tolerance, and the projection operators are defined by the random variables λ and η and the process $P = \{P_t, t \in [0, T]\}$ taking values, respectively, in $[0, 1]$, $(0, +\infty)$ and $\mathcal{L}(\mathbb{R}^d)$. The limiting Nash equilibrium portfolio is then given by

$$\hat{\pi}_t^{i,\infty} := \sigma(t)^{-1} P_t(I - \mu(\lambda P_t))^{-1}(\eta(I - \mu(\lambda P_t)) + \lambda\mu(\eta P_t))\theta(t),$$

provided that $\mu(\lambda\|P_t\|) + \mu(\eta\|P_t\|) < \infty$ and $\|\mu(\lambda P_t)\| < 1$.

Our next comment concerns the asymptotics of the market index $\bar{X}^N$ and the market portfolio $\bar{\pi}^N$ of Definition 3.9.

REMARK 5.11. In the context of Remark 5.10, we further assume that the random variables λ, η, and P_t are independent, and we denote $\bar{P}_t := \mu(P_t)$, $\bar{\lambda} := \mu(\lambda)$, $\bar{\eta} := \mu(\eta)$. Then, under the condition $\bar{\lambda}\bar{P}_t < 1$, the limit market portfolio and market index are given by

$$\bar{\pi}_t^\infty = \sigma(t)^{-1}\bar{v}_t^\infty, \; \bar{X}_t^\infty = \bar{x} + \int_0^t \bar{v}_t^\infty \cdot \big(dW_t + \theta(t)dt\big), \text{ where } v_t^\infty := \bar{\eta}\bar{U}_t\big(I - \bar{\lambda}\bar{U}_t\big)^{-1}\theta(t).$$

In particular, we have the following observations that are consistent with Proposition 3.10:

- the drift of the market index is nonnegative,
- the drift and the volatility of the market index are nondecreasing in $\bar{\eta}$ and $\bar{\lambda}$,

- the VRR index of the market portfolio is given by

$$\overline{\text{VRR}}_t^\infty = \frac{\theta(t)\cdot \bar{P}_t(I-\bar{\lambda}\bar{P}_t)^{-1}\theta(t)}{\theta(t)\cdot(\bar{P}_t(I-\bar{\lambda}\bar{P}_t))^2\theta(t)},$$

and is nonincreasing in $\bar{\eta}$ and $\bar{\lambda}$.

5.3. Examples with Linear Constraints

For simplicity, except for Example 5.16, we assume that the agents are symmetric $\lambda_i = \lambda$ and $\eta_i = \eta$ for $i = 1, \ldots, N$ and only differ by their access to the financial market.

Except for the last Example 5.17, we shall consider a diagonal multidimensional Black–Scholes model with volatility matrix $\sigma = I_d$, i.e., the risky assets price processes are independent.

Under the conditions of Theorem 5.2, the optimal Nash equilibrium is given by

$$(5.7)\quad \hat{\pi}_t^i = \eta P^i \left(I - \frac{\frac{\lambda}{N-1}}{1+\frac{\lambda}{N-1}} \sum_{j\neq i} P^j \left(I + \frac{\lambda}{N-1} P^i \right)\right)^{-1} \theta(t) \quad \text{for } i = 1, \ldots, N,$$

see Example 5.4. Let $(e_1, \ldots, e_d)$ be the canonical basis of $\mathbb{R}^d$.

EXAMPLE 5.12. Let $d = N$ and $A_i = \mathbb{R}e_i$, $i = 1, \ldots, N$. Notice that $\cap_{i=1}^n A_i = \{0\}$. Then, Theorem 5.2 applies for all $\lambda \in [0, 1]$. The projection matrices P^i are all diagonal with unique nonzero diagonal entry $P^i_{i,i} = 1$. The calculation of the Nash equilibrium is then easy and provides

$$\hat{\pi}_t^i = \eta\sigma(t)^{-1}\theta_i(t)e_i, \quad i = 1, \ldots, N.$$

Hence, in agreement with the economic intuition, the interaction has no impact in this example, and the optimal Nash equilibrium portfolio coincides with the classical case with no interactions ($\lambda = 0$).

EXAMPLE 5.13. Let $d = 3$, $N = 2$, and $A_1 = \mathbb{R}e_1 + \mathbb{R}e_2$, $A_2 = \mathbb{R}e_2 + \mathbb{R}e_3$. Since $A_1 \cap A_2 \neq \{0\}$, Theorem 5.2 requires that $\lambda \in [0, 1)$. In the present context, the projection matrices are diagonal with $P^1_{1,1} = P^1_{2,2} = 1$, $P^1_{3,3} = 0$, and $P^2_{1,1} = 0$, $P^2_{2,2} = P^2_{3,3} = 1$. An easy calculation provides the optimal Nash equilibrium:

$$\hat{\pi}_t^1 = \eta\theta^1(t)e_1 + \frac{\eta}{1-\lambda}\theta^2(t)e_2 \quad\text{and}\quad \hat{\pi}_t^2 = \frac{\eta}{1-\lambda}\theta^2(t)e_2 + \eta\theta^3(t)e_3.$$

Notice that the optimal investment in the first and the third stock for agent 1 and agent 2, respectively, is the same as in the classical case ($\lambda = 0$). However, the investment in stock 2, which both agents can trade, is dilated by the factor $(1-\lambda)^{-1} \in [1, +\infty)$.

EXAMPLE 5.14. Let $d = N = 3$ and $A_1 = \mathbb{R}e_1 + \mathbb{R}e_2$, $A_2 = \mathbb{R}e_2 + \mathbb{R}e_3$, $A_3 = \mathbb{R}e_3$. Since $A_1 \cap A_2 \cap A_3 = \{0\}$, Theorem 5.2 applies for $\lambda \in [0, 1]$. The projection matrices P^1 and P^2 are the same as in the previous example, and we similarly see that P^3 is diagonal with $P^3_{1,1} = P^3_{2,2} = 0$, $P^3_{3,3} = 1$. Direct calculation provides the optimal Nash equilibrium:

$$\hat{\pi}_t^1 = \eta\theta^1(t)e_1 + \frac{\eta}{1-\frac{\lambda}{2}}\theta^2(t)e_2, \quad \hat{\pi}_t^2 = \frac{\eta}{1-\frac{\lambda}{2}}\theta^2(t)e_2 + \frac{\eta}{1-\frac{\lambda}{2}}\theta^3(t)e_3,$$
$$\hat{\pi}_t^3 = \frac{\eta}{1-\frac{\lambda}{2}}\theta^3(t)e_3.$$

Similar to the previous example, we see that the optimal investment in the first stock for agent 1 and agent 2, respectively, is the same as in the classical case ($\lambda = 0$), while the investment in stocks 2 and 3, which can both be traded by two agents, is dilated by the factor $(1 - \frac{\lambda}{2})^{-1} \in [1, +\infty)$. Notice that the dilation factor in the present example is smaller than that of the previous one.

EXAMPLE 5.15 (Investment with respect to hyperplanes). Let $d = N$ and $A_i = (\mathbb{R}e_i)^\perp$. In words, each manager has access to the whole market except for its own stock or those of the firms for which some private information is available to the manager. Direct calculation from the expression of Theorem 5.2 provides the following unique Nash equilibrium:

$$\hat{\pi}^{i,N} = \frac{\eta}{1 - \lambda + \frac{\lambda}{N-1}} \sum_{j \neq i} \theta_j e_j, \quad i = 1, \dots, N.$$

EXAMPLE 5.16 (Groups of managers investing in independent sectors). We assume that there are d groups of managers. The jth group consists of k_j symmetric agents with risk tolerance coefficient η_j, interaction coefficient λ_j, and market access defined by the constraints set $A_j = \mathbb{R}e_j$. The total number of managers is $N = \sum_{j=1}^d k_j$. Then, it follows from Theorem 5.2 that the Nash equilibrium portfolio for an agent of the jth group is

$$\begin{aligned}
\hat{\pi}^j &= P^j \left(I - \sum_{m \neq j} k_m \frac{\lambda_m^N}{1 + \lambda_m^N} P^m \big(I + \lambda_j^N P^j\big) - (k_j - 1) \frac{\lambda_j^N}{1 + \lambda_j^N} P^j \big(I + \lambda_j^N P^j\big) \right)^{-1} \\
&\quad \times \left(\eta_j I + \sum_{m \neq j} \frac{\lambda_j^N \eta_m - \lambda_m^N \eta_j}{1 + \lambda_m^N} P^m \right) \theta \\
&= P^j \left(I - \sum_{m \neq j} k_m \frac{\lambda_m^N}{1 + \lambda_m^N} P^m - (k_j - 1) \lambda_j^N P^j \right)^{-1} \left(\eta_j I + \sum_{m \neq j} \frac{\lambda_j^N \eta_m - \lambda_m^N \eta_j}{1 + \lambda_m^N} P^m \right) \theta,
\end{aligned}$$

where we used the fact that $P^j P^m = 0$ for $m \neq j$. The inverse matrix in the previous expression can be computed explicitly, and we get

$$\hat{\pi}^j = P^j \left(\frac{1}{1 - (k_j - 1)\lambda_j^N} P^j + \sum_{m \neq j} \frac{1}{1 - k_m \frac{\lambda_m^N}{1 + \lambda_m^N}} P^m \right) \left(\eta_j I + \sum_{m \neq j} \frac{\lambda_j^N \eta_m - \lambda_m^N \eta_j}{1 + \lambda_m^N} P^m \right) \theta.$$

Using again the fact that $P^j P^m = 0$ for $m \neq j$, we see that

$$\hat{\pi}^j = \frac{\eta_j}{1 - \frac{k_j - 1}{N} \lambda_j} \theta_j e_j \quad \text{for each agent of group } j, \quad j = 1, \dots, d.$$

EXAMPLE 5.17 (Correlated investments). Let $d = N$, $A_i = \mathbb{R}e_i$, $i = 1, \dots, N$, and

$$\theta = \theta_N \sigma \sum_{i=1}^d e_i, \qquad \sigma^2 = \sigma_N^2 \begin{pmatrix} 1 & & \rho^2 \\ & \ddots & \\ \rho^2 & & 1 \end{pmatrix}, \tag{5.8}$$

for some $\theta_N \in \mathbb{R}$, $\rho \in (-1, 1)$ and $\sigma_N > 0$.

Since σ is invertible, $(u_i := \sigma e_i)_{1\le i\le d}$ forms a basis of $\mathbb{R}^d$. We directly verify that for $j \neq i$ and $x = \sum_{i=1}^d x_i u_i$,

$$P^j(x) = \left(x_j + \rho^2 \sum_{k\neq j} x_k \right) u_j, \quad P^j P^i(x) = \rho^2 \left(x_i + \rho^2 \sum_{k\neq i} x_k \right) u_j. \tag{5.9}$$

By (5.7), the Nash equilibrium portfolio for the ith manager is given by $\hat{\pi}^i_t = \eta P^i x$, where x satisfies

$$\left(I - \frac{\frac{\lambda}{N-1}}{1+\frac{\lambda}{N-1}} \sum_{j\neq i} P^j \left(I + \frac{\lambda}{N-1} P^i \right) \right) x = \theta \quad \text{for } i = 1, \dots, N.$$

Given the particular structure of the risk premium in (5.8), we search for a solution of this linear system of the form $x = x_i u_i + x_0 \sum_{k\neq i} u_k$. By (5.9), this reduces the previous linear system to

$$\theta = x_i u_i + \sum_{j\neq i} \left(x_0 - \frac{\lambda}{N-1} \rho^2 x_i - \frac{\frac{\lambda}{N-1}}{1+\frac{\lambda}{N-1}} (1 + (N-2)\rho^2 + \lambda\rho^4) x_0 \right) u_j,$$

and provides the solution of the system

$$x_i = \theta_N \quad \text{and} \quad x_0 = \frac{(1 + \frac{\lambda}{N-1}\rho^2)\theta_N}{1 - \frac{\frac{\lambda}{N-1}}{1+\frac{\lambda}{N-1}}(1 + (N-2)\rho^2 + \lambda\rho^4)},$$

and therefore, using again (5.9), the Nash equilibrium $\hat{\pi}^i = \eta P^i x$ is given by

$$\hat{\pi}^i = \eta\theta_N \left(1 + \frac{(N-1)\rho^2(1 + \frac{\lambda}{N-1}\rho^2)}{1 - \frac{\frac{\lambda}{N-1}}{1+\frac{\lambda}{N-1}}(1 + (N-2)\rho^2 + \lambda\rho^4)} \right) u_i, \quad i = 1, \dots, N.$$

We finally observe that $\hat{\pi}^i \sim \eta\theta_N \frac{1+(N-1-\lambda)\rho^2}{1-\lambda\rho^2}$ as $N \to \infty$. Then,

$$\hat{\pi}^i \sim \frac{\eta\theta_0\rho^2}{1-\lambda\rho^2} u_i \quad \text{whenever } \theta_N \equiv \frac{\theta_0}{N} \text{ as } N \to \infty.$$

This shows that the Nash equilibrium portfolio consists again of a dilation of the no-interaction optimal portfolio. However, in the present context, in addition to the dilation due to the interaction coefficient λ, there is an additional dilation caused by the correlation coefficient ρ. The dilation factor is increasing both in λ and ρ.

5.4. Proof of Technical Lemmas

Proof of Lemma 5.1. We omit all t subscripts. For arbitrary $z^1, \dots, z^N$ in $\mathbb{R}^d$, we want to find a unique solution to the system:

$$z^i - \lambda_i^N \sum_{j\neq i} P^j(z^j) = \zeta^i, \quad 1 \le i \le N. \tag{5.10}$$

1. We reduce (5.10) to a simpler form. Subtracting λ_i times equation j from λ_j times equation i in (5.10), we get for any i, j

$$\lambda_i \left(I + \lambda_j^N P^j\right) z^j = \lambda_j \left(I + \lambda_i^N P^i\right) z^i + \lambda_i \zeta^j - \lambda_j \zeta^i.$$

Since $(I + \lambda_j^N P^j)^{-1} = I - \frac{\lambda_j^N}{1+\lambda_j^N} P^j$, we have

$$\lambda_i P^j z^j = \frac{1}{1+\lambda_j^N} P^j \big(\lambda_j \left(I + \lambda_i^N P^i\right) z^i + \lambda_i \zeta^j - \lambda_j \zeta^i\big).$$

Thus, using (5.10) it follows that

$$\zeta^i = z^i - \frac{1}{N-1} \sum_{j \neq i} \frac{1}{1+\lambda_j^N} P^j \left[\lambda_j \left(I + \lambda_i^N P^i\right) z^i + \lambda_i \zeta^j - \lambda_j \zeta^i\right],$$

and we can rewrite (5.10) equivalently as

(5.11)
$$\left(I - \sum_{j \neq i} \frac{\lambda_j^N}{1+\lambda_j^N} P^j \left(I + \lambda_i^N P^i\right)\right) z^i = \zeta^i + \frac{1}{N-1} \sum_{j \neq i} \frac{1}{1+\lambda_j^N} P^j \left(\lambda_i \zeta^j - \lambda_j \zeta^i\right),$$

so that the invertibility of φ is equivalent to the invertibility of the linear operators $I - R^i$, for $i = 1, \dots, N$, where the R^is are introduced in the statement of the lemma.

2. We prove that the $I - R^i$s are all invertible iff (5.2) holds true.

(a) First, assume that $\lambda_j = 1$ for all j and that $x \in \bigcap_{j=1}^N A_j \neq \{0\}$ satisfies $x \neq 0$. Then, we have for any j, $P^j x = x$ and so:

$$R^i x = \frac{1}{N-1} \sum_{j \neq i} \frac{1}{1 + \frac{1}{N-1}} P^j \left(I + \frac{1}{N-1} P^i\right) x = x.$$

Therefore, $I - R^i$ is not invertible.

(b) Conversely, assume that (5.2) holds true. We consider two separate cases.

- If $\lambda_{i_0} < 1$, for some $i_0 \in \{1, \dots, N\}$, then we estimate that

$$\frac{\lambda_{i_0}^N}{1+\lambda_{i_0}^N} < \frac{\frac{1}{N-1}}{1 + \frac{1}{N-1}} \quad \text{and} \quad \frac{\lambda_j^N}{1+\lambda_j^N} \leq \frac{\frac{1}{N-1}}{1 + \frac{1}{N-1}} \quad \text{for any } j \neq i_0.$$

  Then, since $1 + \lambda_{i_0}^N < 1 + \frac{1}{N-1}$, for any i and any $x \neq 0$, $|R^i x| < |x|$, proving that $I - R^i$ is invertible.
- If $\lambda_i = 1$, for all $i = 1, \dots, N$ and $\bigcap_{i=1}^N A_i = \{0\}$. Let $x \in Ker(I - R^i)$ for some i, using the fact that the P^js are contractions, we have

$$|x| = |R^i x| = \left| \frac{1}{N-1} \sum_{j \neq i} \frac{1}{1 + \frac{1}{N-1}} P^j \left(I + \frac{1}{N-1} P^i\right) x \right|$$
$$\leq \frac{1}{N-1} \sum_{j \neq i} |x| = |x|,$$

so that equality holds in the above inequality, which can only happen if $P^j x = x$ for all $j = 1, \ldots, N$, which implies that $x \in \bigcap_{j=1}^N A_j$, and therefore $x = 0$, which completes the proof. □

Proof of Lemma 5.3. We want to show that the system $\eta_i P^i\theta + \lambda_i^N \sum_{j\neq i} P^i\gamma^j = \gamma^i$, for all $i = 1, \ldots, N$, has a unique solution, or equivalently that $\lambda_i^N \sum_{j\neq i} P^i\gamma^j - \gamma^i = 0$ is satisfied for all $i = 1, \ldots, N$ if and only if $\gamma = 0$. Writing this linear system $A\gamma = 0$, we have

$$\left|\lambda_i^N \sum_{j\neq i} P^i\gamma^j - \gamma^i\right| \geq |\gamma^i| - \lambda_i^N \sum_{j\neq i} |P^i\gamma^j|,$$

so that $\gamma \in \operatorname{Ker} A$ implies that $|\gamma^i| = |\gamma^j|$ for any i, j. Having equality for i implies that $P^i\gamma^j = \gamma^j$ for all j, the γ^js are all collinear (i included) and $\lambda_i = 1$. Therefore, if $\prod_i \lambda_i < 1$ or $\cap_i A_i = \{0\}$, the previous inequality becomes strict if $\gamma \in \operatorname{Ker} A \neq 0$.

Then, as in the proof of Lemma 4.15, we have $\hat{\gamma}^i = \Gamma^i(\hat{\gamma})$, for each $i = 1, \ldots, N$. □

APPENDIX

EXAMPLE A.1. Let $N = 2$, $\sigma = I_d$, $\lambda_i = \lambda$, and $A_i = A := \{x \in \mathbb{R}^d; |x_1| \geq 1\}$, $i = 1, 2$. The projection is uniquely determined for $x_1 \neq 0$, and we can take, for example, the following:

$$P(x) = \begin{cases} x, & \text{if } x \in A \\ (1, x_2, \ldots, x_d)^t, & \text{if } x_1 \in [0, 1) \\ (-1, x_2, \ldots, x_d)^t, & \text{if } x_1 \in (-1, 0). \end{cases}$$

If φ was surjective onto $\mathbb{R}^{2d}$, then subtracting the expressions of φ^1 and φ^2, we see that $I + \lambda P$ would be surjective onto $\mathbb{R}^d$. Let $y \in \mathbb{R}^d$, we want to find x such that $x + \lambda P(x) = y$.

- If $x_1 \geq 1$, then $(1 + \lambda)x_1 = y_1$ so that $y_1 \geq 1 + \lambda$;
- if $x_1 \in [0, 1)$, then $x_1 + \lambda = y_1$ so that $y_1 \in [\lambda, 1 + \lambda)$;
- if $x_1 \in (-1, 0)$, then $x_1 - \lambda = y_1$ so that $y_1 \in (-1 - \lambda, -\lambda)$;
- if $x_1 \leq -1$, then $(1 + \lambda)x_1 = y_1$ so that $y_1 \leq -1 - \lambda$.

Therefore, $\{x \in \mathbb{R}^d; x_1 \in [-\lambda, \lambda)\}$ is not attained by $I + \lambda P$ so that as soon as $\lambda > 0$, φ is not surjective. Moreover, the interior of the complementary of its image is nonempty.

EXAMPLE A.2. Let $A_i = B := \{x \in \mathbb{R}^d; |x| \geq 1\}$, the complement of the unit (open) ball. The projection is uniquely determined for $x \neq 0$, and we can for example take:

$$P(x) = \begin{cases} x, & \text{if } x \in B, \\ \dfrac{1}{|x|}x, & \text{if } |x| \in (0, 1), \\ 1_d, & \text{if } x = 0. \end{cases}$$

Similar to the previous example, in order to have φ surjective, we need $I + \lambda P$ surjective onto $\mathbb{R}^d$. If $y \in \mathbb{R}^d$, and $x + \lambda P(x) = y$, we compute

- if $|x| \geq 1$, then $(1+\lambda)x = y$ so that $|y| \geq 1+\lambda$;
- if $|x| \in (0,1)$, then $(1+\frac{\lambda}{|x|})x = y$ so that $|y| \in (\lambda, 1+\lambda)$;
- if $x = 0$, then $y = \lambda 1_d$.

Therefore, $\{x \in \mathbb{R}^d; |x| < \lambda\}$ is not attained by $I + \lambda P$, so again as soon as $\lambda > 0$, φ is not surjective. Moreover, the interior of the complementary of its image is nonempty.

REFERENCES

ABEL, A. B. (1990): Asset Prices Under Habit Formation and Catching Up with the Joneses, *Am. Econ. Rev.* 80, 38–42.

AKIAN, M., J.-L. MENALDI, and A. SULEM (1995): Multi-Asset Portfolio Selection Problem with Transaction Costs. Probabilités Numériques (Paris, 1992), *Math. Comput. Simul.* 38, 163–172.

BEN TAHAR, I., H. M. SONER, and N. TOUZI (2008a): Modelling Continuous-Time Financial Markets with Capital Gains Taxes, Preprint.

BEN TAHAR, I., H. M. SONER, and N. TOUZI (2008b): The Dynamic Programming Equation for the Problem of Optimal Investment Under Capital Gains Taxes, *SIAM J. Control Optim.* 46, 1779–1801.

CONSTANTINIDES, G. M. (1983): Capital Market Equilibrium with Personal Taxes, *Econometrica* 51, 611–636.

CONSTANTINIDES, G. M., and M. J. P. MAGILL (1976): Portfolio Selection with Transaction Costs, *J. Econ. Theory* 13, 245–263.

COX, J., and C. F. HUANG (1989): Optimal Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process, *J. Econ. Theory* 49, 33–83.

CVITANIC, J., and I. KARATZAS (1992): Convex Duality in Constrained Portfolio Optimization, *Ann. Appl. Probab.* 2, 767–818.

DAMON, R. M., C. S. SPATT, and H. H. ZHANG (2001): Optimal Consumption and Investment with Capital Gains Taxes, *Rev. Financ. Stud.* 14, 583–616.

DAVIS, M. H. A., and A. R. NORMAN (1990): Portfolio Selection with Transaction Costs, *Math. Oper. Res.* 15(4), 676–713.

DEMARZO, P., R. KANIEL, and I. KREMER (2008): Relative Wealth Concerns and Financial Bubbles, *Rev. Financ. Stud.* 21(1), 19–50.

DUFFIE, D., and T. SUN (1990): Transaction Costs and Portfolio Choice in a Discrete-Continuous-Time Setting, *J. Econ. Dyn. Control* 14, 35–51.

ESPINOSA, G.-E. (2010): Stochastic Control Methods for Optimal Portfolio Investment, PhD Manuscript, Ecole Polytechnique.

ESPINOSA, G.-E. (2010): Optimal Investment Under Relative Performance Concerns with Local Risk Penalization, Preprint.

FREI, C., and G. DOS REIS (2011): A Financial Market with Interacting Investors: Does an Equilibrium Exist? *Math. Financ. Econ.* 4, 161–182.

GALI, J. (1994). Keeping Up with the Joneses: Consumption Externalities, Portfolio Choice and Asset Prices, *J. Money Credit Banking* 26, 1–8.

GOMEZ, J. P., R. PRIESTLEY, and F. ZAPATERO (2007): The Impact of Keeping Up with the Joneses Behavior on Asset Prices and Portfolio Choice, *Finance Res. Lett.* 4(2), 95–103.

HE, H., and N. D. PEARSON (1991a): Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Finite-Dimensional Case, *Math. Finance* 1, 1–10.

HE, H., and N. D. PEARSON (1991b): Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Infinite-Dimensional Case, *J. Econ. Theory* 54, 259–304.

HU, Y., P. IMKELLER, and M. MÜLLER (2005): Utility Maximization in Incomplete Markets, *Ann. Appl. Probab.* 15(3), 1691–1712.

JOUINI, E., P.-F. KOEHL, and N. TOUZI (1997): Optimal Investment with Taxes: An Optimal Control Problem with Endogeneous Delay, *Nonlinear Anal. Theory Methods Appl.* 37, 31–56.

JOUINI, E., P.-F. KOEHL, and N. TOUZI (1999): Optimal Investment with Taxes: An Existence Result, *J. Math. Econ.* 33, 373–388.

KARATZAS, I., J. P. LEHOCZKY, and S. E. SHREVE (1987): Optimal Portfolio and Consumption Decisions for a "Small Investor" on a Finite Horizon, *SIAM J. Control Optim.* 25, 1557–1586.

KARATZAS, I., J. P. LEHOCZKY, S. E. SHREVE, and G. XU (1991): Martingale and Duality Methods for Utility Maximization in an Incomplete Market, *SIAM J. Control Optim.* 29, 702–730.

KAROUI, N. EL, and R. ROUGE (2000): Pricing via Utility Maximization and Entropy, *Math. Finance* 10, 259–276.

KARATZAS, I., and S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.

KARATZAS, I., and S. E. SHREVE (1998): *Methods of Mathematical Finance*, New York: Springer-Verlag.

KRAMKOV, D., and W. SCHACHERMAYER (1999): The Condition on the Asymptotic Elasticity of Utility Functions and Optimal Investment in Incomplete Markets, *Ann. Appl. Probab.* 9, 904–950.

KRAMKOV, D., and W. SCHACHERMAYER (2003): Necessary and Sufficient Conditions in the Problem of Optimal Investment in Incomplete Markets, *Ann. Appl. Probab.* 13(4), 1504–1516.

KRAMKOV, D., and M. SIRBU (2006): On the Two-Times Differentiability of the Value Functions in the Problem of Optimal Investment in Incomplete Markets, *Ann. Appl. Probab.* 16(3), 1352–1384.

LASRY, J.-M., and P.-L. LIONS (2007): Mean Field Games, *Jpn. J. Math.* 2, 229–260.

LIM, T., and M.-C. QUENEZ (2009): Utility Maximization in Incomplete Markets with Default, Preprint.

MANIA, M., and M. SCHWEIZER (2005): Dynamic Exponential Utility Indifference Valuation, *Ann. Appl. Probab.* 15(3), 2113–2143.

MERTON, R. C. (1969): Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Model, *Rev. Econ. Statist.* 51, 247–257.

MERTON, R. C. (1971): Optimum Consumption and Portfolio Rules in a Continuous-Time Model, *J. Econ. Theory* 3, 373–413.

PLISKA, S. R. (1986): A Stochastic Calculus Model of Continuous Trading: Optimal Portfolios, *Math. Oper. Res.* 11, 371–382.

SHREVE, S. E., and H. M. SONER (1994): Optimal Investment and Consumption with Transaction Costs, *Ann. Appl. Probab.* 4, 609–692.

VEBLEN, T. B. (1899): *The Theory of the Leisure Class: An Economic Study of Institutions*, New York: Penguin.

ZARIPHOPOULOU, T. (1994): Consumption-Investment Models with Constraints, *SIAM J. Control Optim.* 32(1), 59–85.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 258–287

ROBUST UTILITY MAXIMIZATION IN NONDOMINATED MODELS WITH 2BSDE: THE UNCERTAIN VOLATILITY MODEL

ANIS MATOUSSI

Université du Maine and CMAP, Ecole Polytechnique

DYLAN POSSAMAÏ

Université Paris-Dauphine

CHAO ZHOU

CMAP, Ecole Polytechnique

The problem of robust utility maximization in an incomplete market with volatility uncertainty is considered, in the sense that the volatility of the market is only assumed to lie between two given bounds. The set of all possible models (probability measures) considered here is nondominated. We propose studying this problem in the framework of second-order backward stochastic differential equations (2BSDEs for short) with quadratic growth generators. We show for exponential, power, and logarithmic utilities that the value function of the problem can be written as the initial value of a particular 2BSDE and prove existence of an optimal strategy. Finally, several examples which shed more light on the problem and its links with the classical utility maximization one are provided. In particular, we show that in some cases, the upper bound of the volatility interval plays a central role, exactly as in the option pricing problem with uncertain volatility models.

KEY WORDS: second-order backward stochastic differential equation, quadratic growth, robust utility maximization, volatility uncertainty.

1. INTRODUCTION

One of the most prominent questions of mathematical finance literature is the so-called problem of utility maximization. It is a problem of optimal investment faced by an economic agent who has the opportunity to invest in a financial market consisting of a riskless asset and (for simplicity) one risky asset. Given a fixed investment horizon T, the aim of the agent is to find an optimal allocation between the two assets, so as to maximize his “welfare” at time T. Following the seminal work of Von Neumann and Morgenstern (1944), where they assumed that the preference of the agent could be represented by a

The authors would like to thank two anonymous referees and the associate editor for their helpful remarks and comments. Research supported by the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, the Chair *Derivatives of the Future* sponsored by the Fédération Bancaire Française, and the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

Manuscript received January 2012; final revision received November 2012.

Address correspondence to Anis Matoussi, Université du Maine—Mathématiques, avenue Olivier Messiaen Lemans, Sarthe 72085, France; e-mail: anis.matoussi@univ-lemans.fr.

DOI: 10.1111/mafi.12031

utility function U and a given probability measure $\mathbb{P}$ reflecting his views, the now classical formulation of the problem consists in solving the optimization problem

$$V(x) := \sup_{\pi \in \mathcal{A}} \mathbb{E}^{\mathbb{P}}\left[U(X_T^{x,\pi} - \xi)\right],$$

where $\mathcal{A}$ is the set of admissible strategies π for the agent, $X_T^{x,\pi}$ is his wealth process at time T computed with initial capital x, trading strategy π, and terminal liability ξ.

In the 1960s, Merton (1969) was the first to study and solve this problem in the particular case where the risky asset follows a Black–Scholes model, there are no restrictions on the admissible strategies (that is to say in a complete market), the utility function is of power type, and the liability is equal to 0. The proof relies on classical techniques of stochastic control theory, since he manages to solve the Hamilton–Jacobi–Bellman partial differential equation (PDE) associated with the problem explicitly, and then uses a verification argument. The problem in complete markets but with general utility functions was only solved in the 1980s by Pliska (1986), using techniques from convex duality. Following these papers, many have tried to weaken their assumptions, notably the completeness assumption on the market, which was too restrictive and unrealistic from the point of view of applications. One possible direction of generalization is to impose constraints on the strategies of the investor. Following the first works of Cvitanić and Karatzas (1992) and Zariphopoulou (1994), where once more convex duality techniques were used, the beginning of the 21st century saw the emergence of a link between this optimal investment problem and the theory of backward stochastic differential equations (BSDEs for short). These objects were first introduced by Bismut (1973) in the linear case, then generalized by Pardoux and Peng (1990) to Lipschitz generators. On a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\leq t\leq T}, \mathbb{P})$ generated by an $\mathbb{R}^d$-valued Brownian motion B, a solution to a BSDE consists of a pair of progressively measurable processes (Y, Z) such that

$$Y_t = \xi + \int_t^T f_s(Y_s, Z_s)ds - \int_t^T Z_s dB_s, \ t \in [0, T], \ \mathbb{P} - \text{a.s.}$$

where f (also called the driver) is a progressively measurable function and ξ is a $\mathcal{F}_T$-measurable random variable.

El Karoui and Rouge (2000) considered the problem of indifference pricing with an exponential utility function (which is linked to the optimal investment problem) in the case where the strategies are constrained to stay in a given closed and convex set. They proved that the value function of the problem was related to the initial value of a BSDE with a driver of quadratic growth in the Z part. Building upon these results, Hu, Imkeller, and Müller (2005) generalized the approach to the case of logarithmic and power utilities with strategies constrained in a closed set.

Another direction of generalization of the original Merton problem is related to the question of model uncertainty. Indeed, in all the above formulations, a probability measure $\mathbb{P}$ is fixed. It means that the investor knows the "historical" probability $\mathbb{P}$ that describes the dynamics of the state process. In reality, the investor may have some uncertainty on this probability, which means that there can be several objective probability measures to consider. The problem then becomes a robust utility maximization and can be written as follows:

$$V^{\xi}(x) := \sup_{\pi \in \mathcal{A}} \inf_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}}[U(X_T^{x,\pi} - \xi)],$$

where $\mathcal{P}$ is the set of all considered possible probability measures.

In this case, the properties of the set $\mathcal{P}$ become crucial in order to solve the problem. Preliminary results in the literature were limited to dominated sets. A set $\mathcal{P}$ is said to be dominated if every probability measure $\mathbb{P} \in \mathcal{P}$ is absolutely continuous with respect to some reference probability measure in $\mathcal{P}$. For instance, this covers the case of drift uncertainty. In this framework, the problem was introduced by Gilboa and Schmeidler (1989). Anderson, Hansen, and Sargent (2003) and Hansen et al. (2006) then introduced and discussed the basic problem of robust utility maximization, penalized by a relative entropy term of the model uncertainty $\mathbb{Q} \in \mathcal{P}$ with respect to a given reference probability measure. Inspired by these latter works, Bordigoni, Matoussi, and Schweizer (2007) solved the robust problem (the minimization part) in a more general semimartingale framework by using stochastic control techniques and proved that the solution was related to a quadratic semimartingale BSDE. Among others, results in the robust maximization problem were also obtained by Gundel (2005), Schied and Wu (2005), or Skiadas (2003) in the case of continuous filtrations. The overall approach relies essentially on convex duality ideas.

The situation becomes more intricate when the set $\mathcal{P}$ is no longer dominated, which happens when introducing volatility uncertainty, in the sense that the volatility process is only assumed to lie between two given bounds. Although the problem of option pricing under volatility uncertainty has been solved for a long time (see Avellaneda, Levy, and Paras 1995 and Lyons 1995 for instance), the problem of utility maximization was not addressed until recently by Denis and Kervarec (2007) (see also Talay and Zheng 2002, where it is analyzed under the framework of stochastic games and Hamilton–Jacobi–Bellman–Isaacs equations). They first establish a duality theory for robust utility maximization and then show that there is a least favorable probability measure and an optimal strategy. However, their utility function U is supposed to be bounded and to satisfy Inada conditions. Recently, Epstein and Ji (2011) formulated a model of utility in a continuous-time framework that captures the decision maker's concern with ambiguity or model uncertainty, even though they did not study the maximization problem of robust utility per se. More recently, Tevzadze, Toronjadze, and Uzunashvili (2012) studied a related robust utility maximization problem for exponential and power utility functions (and also for mean-square error criteria). We will compare our results and theirs in Section 7 (see Remark 7.2).

The intuition at the core of our work is that, exactly as the problem of utility maximization under constraints was linked to BSDEs with quadratic growth, the problem of robust utility maximization under volatility uncertainty should be linked to some kind of backward equations. In fact, the right objects to consider in this case are the so-called second-order BSDEs (2BSDEs for short) which were introduced for the first time by Cheridito et al. (2007). However, they were not able to provide a complete theory of existence and uniqueness. Hence, a reformulation was proposed by Soner, Touzi, and Zhang (2012), who provided a well-posedness theory for 2BSDEs under uniform Lipschitz conditions similar to those of Pardoux and Peng. Their key idea was to reinforce the condition that the 2BSDE must hold $\mathbb{P}-$a.s. for every probability measure $\mathbb{P}$ in a nondominated class of mutually singular measures (see Section 2 for precise definitions). The theory being very recent, the literature remains rather limited. However, we refer the interested reader to Possamaï (2010) and Possamaï and Zhou (2012) who, respectively, extended these well-posedness results to generators with linear and quadratic growth. In our main result, we show that in incomplete markets with volatility uncertainty, the solution of the robust utility maximization problem (for exponential, logarithmic, and power utilities) is related to the initial value of a particular 2BSDE with quadratic growth, as suggested by intuition. We also emphasize a specificity in our approach when it comes to the sets of admissible strategies considered. Usually, when dealing with this type of problems (see, for instance, El Karoui and Rouge 2000; Hu et al. 2005), an exponential uniform integrability assumption is made on the trading strategies. Our approach relies instead on integrability assumptions of BMO type on the trading strategies. The mathematical justifications are detailed in Remarks 4.2 and 4.5. However, there is also a financial interpretation. Indeed, as explained in Frei, Mocha, and Westray (2012) which also adopts the BMO framework, this assumption corresponds to a situation where the market price of risk is assumed to be BMO. Exactly as in the case of a bounded market price of risk, it implies that the minimum martingale measure is a true probability measure, and therefore that the market is without arbitrage, in the sense of No Free Lunch with Vanishing Risk.

The paper is organized as follows: in Section 2, we recall the 2BSDEs framework and some useful results. Inspired by El Karoui and Rouge (2000) and Hu et al. (2005), in Sections 3, 4, 5, and 6 we solve the problem for robust exponential utility, robust power utility, and robust logarithmic utility, which, unlike in Denis and Kervarec (2007), are not bounded. Finally, in Section 7, we give some examples where we can explicitly solve the robust utility maximization problems by finding the solution of the associated 2BSDEs, and we provide some insights and comparisons with the classical dynamic programming approach adopted in the seminal work of Merton (1969).

2. PRELIMINARIES

We will start by recalling some notations and notions related to the theory of 2BSDEs, which are the main tool in our approach to the robust utility maximization problem.

2.1. Probability Spaces

Let $\Omega := \{\omega \in C([0, T])^d,\ \omega(0) = 0\}$ be the canonical space, B the canonical process, and $\mathbb{F} = (\mathcal{F}_t)_{0\leq t\leq T}$ the filtration generated by B. We will also make use of the right-limit $\mathbb{F}^+ = (\mathcal{F}_{t^+})_{0\leq t\leq T}$ of $\mathbb{F}$. Let $\mathbb{P}_0$ be the Wiener measure. As in Soner et al. (2012), we can construct the quadratic variation of B and its density $\widehat{a}$ pathwise.

Let $\overline{\mathcal{P}}_W$ denote the set of all local martingale measures $\mathbb{P}$ such that $\mathbb{P}-$a.s. for $t \in [0, T]$

$$\text{(2.1)} \qquad \langle B\rangle_t \text{ is absolutely continuous with respect to } t \text{ and } \widehat{a} \text{ takes values in } \mathbb{S}_d^{>0},$$

where $\mathbb{S}_d^{>0}$ denotes the space of all $d \times d$ real-valued positive definite matrices. As in Soner et al. (2012), we concentrate on the subclass $\overline{\mathcal{P}}_S \subset \overline{\mathcal{P}}_W$ consisting of all probability measures

$$\text{(2.2)} \qquad \mathbb{P}^\alpha := \mathbb{P}_0 \circ (X^\alpha)^{-1} \text{ where } X_t^\alpha := \int_0^t \alpha_s^{1/2} dB_s,\ t \in [0, T],\ \mathbb{P}_0 - \text{a.s.}$$

for some $\mathbb{F}$-progressively measurable process α in $\mathbb{S}_d^{>0}$ with $\int_0^T |\alpha_t| dt < +\infty$, $\mathbb{P}_0 - \text{a.s.}$ Notice that $\alpha_s^{1/2}$ is just the square-root of the positive definite matrix α_s.

Finally, we fix $\underline{a}, \overline{a} \in \mathbb{S}_d^{>0}$ such that $\underline{a} \leq \overline{a}$ (for the usual order on positive definite matrices, i.e., $(\overline{a} - \underline{a}) \in \mathbb{S}_d^{>0}$) and we define the class

$$\mathcal{P}_H := \left\{\mathbb{P} \in \overline{\mathcal{P}}_S \text{ s.t. } \underline{a} \leq \widehat{a}_t \leq \overline{a},\ dt \times \mathbb{P} - \text{a.e.}\right\},$$

which is a particular case of definition 2.6 in Soner et al. (2012), the main differences here being that the two bounds on $\widehat{a}$ are independent of the probability measures and that $\widehat{F}^0$ (introduced and defined below in Section 2.3) is bounded. Throughout the paper, it is assumed that $\mathcal{P}_H$ is not empty.

For every $(t, \mathbb{P}) \in [0, T] \times \mathcal{P}_H$, we also define the class of probability measures which coincide with $\mathbb{P}$ up to t^+

$$\mathcal{P}_H\big(t^+, \mathbb{P}\big) := \big\{\mathbb{P}' \in \mathcal{P}_H,\ \mathbb{P}' = \mathbb{P} \text{ on } \mathcal{F}_{t^+}\big\}.$$

DEFINITION 2.1. We say that a property holds $\mathcal{P}_H$-quasi-surely ($\mathcal{P}_H$-q.s. for short) if it holds $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}_H$.

REMARK 2.2. The filtration $\mathbb{F}^+$ is right-continuous but not complete under each $\mathbb{P} \in \mathcal{P}_H$. Moreover, it is not possible to complete the filtration for each $\mathbb{P}$ since the measures are singular. It is of course a major drawback since many results of the general theory of processes rely on the fact that the underlying filtrations satisfy the usual hypotheses of right-continuity and completeness. However, this problem was solved in lemma 2.4 of Soner, Touzi, and Zhang (2011), which implies that for every $\mathbb{P} \in \mathcal{P}_H$, we can always consider a version of our processes which is progressively measurable for the completion of $\mathbb{F}^+$ under $\mathbb{P}$.

2.2. Spaces and Norms

We now recall from Possamaï and Zhou (2012) the spaces and norms which will be needed for the formulation of the quadratic 2BSDEs.

$\mathbb{L}_H^\infty$ is the space of random variables which are bounded quasi-surely endowed with the norm

$$\|\xi\|_{\mathbb{L}_H^\infty} := \sup_{\mathbb{P}\in\mathcal{P}_H} \|\xi\|_{L^\infty(\mathbb{P})}.$$

For $p \geq 1$, L_H^p is the space of random variables with

$$\|\xi\|_{L_H^p}^p := \sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[|\xi|^p\right] < +\infty.$$

For $p \geq 1$, $\mathbb{H}_H^p$ denotes the space of all $\mathbb{F}^+$-progressively measurable $\mathbb{R}^d$-valued processes Z with

$$\|Z\|_{\mathbb{H}_H^p}^p := \sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[\left(\int_0^T |\widehat{a}_t^{1/2} Z_t|^2 dt\right)^{\frac{p}{2}}\right] < +\infty.$$

$\mathbb{BMO}(\mathcal{P}_\mathrm{H})$ denotes the space of all $\mathbb{F}^+$-progressively measurable $\mathbb{R}^d$-valued processes Z with

$$\|Z\|_{\mathbb{BMO}(\mathcal{P}_\mathrm{H})} := \sup_{\mathbb{P}\in\mathcal{P}_H} \left\|\int_0^\cdot Z_s dB_s\right\|_{\mathrm{BMO}(\mathbb{P})},$$

where $\|\cdot\|_{\mathrm{BMO}(\mathbb{P})}$ is the usual BMO($\mathbb{P}$) norm under $\mathbb{P}$, that is to say

$$\left\|\int_0^\cdot Z_s dB_s\right\|_{\mathrm{BMO}(\mathbb{P})}^2 := \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{0,T}}^{\mathbb{P}} \left\|\mathbb{E}_\tau^{\mathbb{P}}\left[\int_\tau^T |\widehat{a}_t Z_t|^2\, dt\right]\right\|_{L^\infty(\mathbb{P})},$$

where $\mathcal{T}_{0,T}$ denotes the stopping times with value in $[0, T]$. We abuse notation and say that $\int_0^\cdot Z_s dB_s$ is a $\mathbb{BMO}(\mathcal{P}_\mathrm{H})$ martingale if $Z \in \mathbb{BMO}(\mathcal{P}_\mathrm{H})$.

$\mathbb{D}_H^\infty$ denotes the space of all $\mathbb{F}^+$-progressively measurable $\mathbb{R}$-valued processes Y with

$$\mathcal{P}_H - q.s. \text{ càdàg paths, and } \|Y\|_{\mathbb{D}_H^\infty} := \sup_{0\leq t\leq T} \|Y_t\|_{L_H^\infty} < +\infty.$$

We also recall the following space which is important in the formulation of Lipschitz 2BSDEs in Soner et al. (2012). For any $\kappa \in (1, 2]$, $\mathbb{L}_H^{2,\kappa}$ is the space of random variables ξ such that

$$\|\xi\|^2_{\mathbb{L}_H^{2,\kappa}} := \sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^\mathbb{P}\left[\operatorname*{ess\,sup}^\mathbb{P}_{0\leq t\leq T} \operatorname*{ess\,sup}^\mathbb{P}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})} \left(\mathbb{E}_t^{\mathbb{P}'}[|\xi|^\kappa]\right)^{\frac{2}{\kappa}}\right] < +\infty.$$

Finally, we denote by $\mathrm{UC}_b(\Omega)$ the collection of all bounded and uniformly continuous maps $\xi : \Omega \to \mathbb{R}$ with respect to the $\|\cdot\|_\infty$-norm, and we let

$$\mathcal{L}_H^\infty := \text{the closure of } \mathrm{UC}_b(\Omega) \text{ under the norm } \|\cdot\|_{\mathbb{L}_H^\infty},$$

and

$$\mathcal{L}_H^{2,\kappa} := \text{the closure of } \mathrm{UC}_b(\Omega) \text{ under the norm } \|\cdot\|_{\mathbb{L}_H^{2,\kappa}}.$$

2.3. The Quadratic Generator

We consider a map $H_t(\omega, z, \gamma) : [0, T] \times \Omega \times \mathbb{R}^d \times D_H \to \mathbb{R}$, where $D_H \subset \mathbb{R}^{d\times d}$ is a given subset containing 0.

Define the corresponding conjugate of H with respect to γ by

$$F_t(\omega, z, a) := \sup_{\gamma\in D_H}\left\{\frac{1}{2}\mathrm{Tr}(a\gamma) - H_t(\omega, z, \gamma)\right\} \text{ for } a \in \mathbb{S}_d^{>0},$$
$$\widehat{F}_t(z) := F_t(\omega, z, \widehat{a}_t) \text{ and } \widehat{F}_t^0 := \widehat{F}_t(0).$$

We denote by $D_{F_t(z)}$ the domain of F in a for a fixed (t, ω, z). As in Possamaï and Zhou (2012), the generator F is supposed to verify either

ASSUMPTION 2.3.

(i) The domain $D_{F_t(z)} = D_{F_t}$ is independent of (ω, z).
(ii) For fixed (z, a), F is $\mathbb{F}$-progressively measurable.
(iii) F is uniformly continuous in ω for the $||\cdot||_\infty$ norm.
(iv) F is continuous in z and has the following growth property. There exists $(\alpha, \gamma) \in \mathbb{R}_+ \times \mathbb{R}_+/\{0\}$ such that

$$|F_t(\omega, z, a)| \leq \alpha + \frac{\gamma}{2}\left|a^{1/2}z\right|^2, \textit{ for all } (t, z, \omega, a).$$

(v) F is C^2 in z, and there are constants r and θ such that for all (t, ω, z, a),

$$|D_z F_t(\omega, z, a)| \leq r + \theta\left|a^{1/2}z\right|, \ |D^2_{zz} F_t(\omega, y, z, a)| \leq \theta.$$

or

ASSUMPTION 2.4. *Let points* (i) *through* (iv) *of Assumption 2.3 hold, and*

(vi) $\exists \mu > 0$ *and a progressively measurable process* $\phi \in \mathbb{BMO}(\mathcal{P}_{\mathrm{H}})$ *such that for all* (t, z, z', ω, a),

$$\left|F_t(\omega, z, a) - F_t(\omega, z', a) - \phi_t.a^{1/2}(z - z')\right| \leq \mu a^{1/2}|z - z'| \left(\left|a^{1/2}z\right| + \left|a^{1/2}z'\right|\right).$$

REMARK 2.5. Notice that Assumption 2.3(iv) implies that $\underset{0\leq t\leq T}{\operatorname{ess\,sup}} \left|\widehat{F}_t^0\right| \in \mathbb{L}_H^\infty$.

2.4. Quadratic 2BSDE

In the sequel, we will have to deal with the following type of 2BSDEs

$$Y_t = \xi - \int_t^T \widehat{F}_s(Z_s)\,ds - \int_t^T Z_s dB_s + K_T - K_t,\ 0 \leq t \leq T,\ \mathcal{P}_H - q.s. \tag{2.3}$$

DEFINITION 2.6. Given $\xi \in \mathcal{L}_H^\infty$, we say $(Y, Z) \in \mathbb{D}_H^\infty \times \mathbb{H}_H^2$ is a solution to the 2BSDE (2.3) if

- $Y_T = \xi,\ \mathcal{P}_H - q.s.$
- For each $\mathbb{P} \in \mathcal{P}_H$, the process $K^{\mathbb{P}}$ defined below has nondecreasing paths, $\mathbb{P}$-a.s.

$$K_t^{\mathbb{P}} := Y_0 - Y_t + \int_0^t \widehat{F}_s(Z_s)\,ds + \int_0^t Z_s dB_s,\ 0 \leq t \leq T,\ \mathbb{P} - a.s. \tag{2.4}$$

- The family of processes $\left\{K^{\mathbb{P}}, \mathbb{P} \in \mathcal{P}_H\right\}$ defined in (2.4) satisfies the following minimum condition:

$$K_t^{\mathbb{P}} = \underset{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} E_t^{\mathbb{P}'}[K_T^{\mathbb{P}'}],\ \mathbb{P} - \text{a.s. for all } \mathbb{P} \in \mathcal{P}_H, t \in [0, T]. \tag{2.5}$$

Moreover, if the family $\{K^{\mathbb{P}}, \mathbb{P} \in \mathcal{P}_H\}$ can be aggregated into a universal process K, that is to say that for all $t \in [0, T]$

$$K_t = K_t^{\mathbb{P}},\ \mathbb{P} - a.s.,\ \forall \mathbb{P} \in \mathcal{P}_H,$$

then we call (Y, Z, K) a solution of the 2BSDE (2.3).

Here, one of the results proved in Possamaï and Zhou (2012) is recalled (see theorems 3.3 and 4.1).

THEOREM 2.7. *Let* $\xi \in \mathcal{L}_H^\infty$. *Under Assumption 2.3 or 2.4 with the addition that the norm of* ξ *and the* $\mathbb{L}_H^\infty$*-norm of* $\underset{0\leq t\leq T}{\operatorname{ess\,sup}}\ |\widehat{F}_t^0|$ *are small enough, there is a unique solution* $(Y, Z) \in \mathbb{D}_H^\infty \times \mathbb{H}_H^2$ *of the* $2BSDE$ *(2.3). Moreover, for all* $t \in [0, T]$ *and every* $\mathbb{P} \in \mathcal{P}_H$

$$Y_t = \underset{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}{\operatorname{ess\,sup}^{\mathbb{P}}} y_t^{\mathbb{P}'},\ \mathbb{P} - a.s., \tag{2.6}$$

where $y^{\mathbb{P}'}$ *is the solution under* $\mathbb{P}'$ *of the BSDE with generator* $\widehat{F}$ *and terminal condition* ξ.

REMARK 2.8. Assumption 2.4 is weaker than Assumption 2.3, but is sufficient for the existence of a solution to the quadratic 2BSDE defined above only if the norms of the terminal condition ξ and $\widehat{F}^0$ are small enough. Notice that since, for power and logarithmic utilities, the terminal condition will be equal to 0, we only have a restriction on the norm of $\widehat{F}^0$ in these cases. We also emphasize that these restrictions are in no way necessary to obtain existence, but are artifacts of the type of proofs used in Possamaï and Zhou (2012) to obtain existence of a 2BSDE with quadratic growth. Indeed, the proof relies at some point on the fact that solutions of standard BSDEs can be obtained through Picard iterations. Notice that this property was already needed in Soner et al. (2012). However, with a generator of quadratic growth, such a property was shown by Tevzadze (2008) only if Assumption 2.3(v) holds (see proposition 2), or if the terminal condition and $\widehat{F}^0$ are small enough and if Assumption 2.4(v) holds (see proposition 1). We conjecture that existence of solutions of 2BSDEs with quadratic growth should hold under less restrictive assumptions similar to those in Barrieu and El Karoui (2011) (for instance ξ would not need to be bounded and the generator would only need to be of quadratic growth), but this is left for future research.

REMARK 2.9. The representation (2.6) gives some insight into 2BSDEs. Since Y can be written as a supremum of solution of BSDEs, we can interpret the increasing processes $K^{\mathbb{P}}$ as the instruments allowing Y to remain above the corresponding $y^{\mathbb{P}}$. It is similar to reflected BSDEs with a lower obstacle. Moreover, the minimum condition (2.5) tells us that this is done in a minimal way, making it the counterpart of the Skorokhod condition in our context.

3. ROBUST UTILITY MAXIMIZATION

We will now present the main problem of the paper and introduce a financial market with volatility uncertainty. The financial market consists of one bond with zero interest rate and d stocks. The price process is given by

$$dS_t = \operatorname{diag}[S_t](b_t dt + dB_t),\ \mathcal{P}_H - q.s.,$$

where b is an $\mathbb{R}^d$-valued uniformly bounded stochastic process which is uniformly continuous in ω for the $||\cdot||_\infty$ norm.

REMARK 3.1. The volatility is implicitly embedded in the model. Indeed, under each $\mathbb{P} \in \mathcal{P}_H$, we have $dB_s \equiv \widehat{a}_t^{1/2} dW_t^{\mathbb{P}}$ where $W^{\mathbb{P}}$ is a Brownian motion under $\mathbb{P}$. Therefore, $\widehat{a}^{1/2}$ plays the role of volatility under each $\mathbb{P}$ and thus makes it possible to model the volatility uncertainty. We also note that we make the uniform continuity assumption for b to ensure that the generators of the 2BSDEs obtained later satisfy Assumption 2.3 or 2.4.

We then denote $\pi = (\pi_t)_{0\leq t\leq T}$ a trading strategy, which is a d-dimensional $\mathbb{F}^+$-progressively measurable process, supposed to take its value in some closed set A. We refer to Definitions 4.1, 5.1, and 6.1 in the following sections for the precise definitions of the set of admissible strategies $\mathcal{A}$ for the three utility functions studied.

The process π_t^i describes the amount of money invested in stock i at time t, with $1 \leq i \leq d$. The number of shares is $\frac{\pi_t^i}{S_t^i}$. So the liquidation value of a trading strategy π with positive initial capital x is given by the following wealth process:

$$X_t^\pi = x + \int_0^t \pi_s(dB_s + b_s ds),\ 0 \leq t \leq T,\ \mathcal{P}_H - q.s.$$

Since zero interest rate was assumed, the amount of money in the bank π^0 does not appear in the wealth process X.

The problem of the investor in this financial market is to maximize the expected utility under model uncertainty of his terminal wealth $X_T^\pi - \xi$, where ξ is a liability, that is to say an $\mathcal{F}_T$-measurable random variable. This liability could represent the value of any option or contract maturing at time T. It will always be assumed that $\xi \in \mathcal{L}_H^\infty$.

Denote by U the utility function of the investor. The value function V of the maximization problem therefore becomes

$$V^\xi(x) := \sup_{\pi \in \mathcal{A}} \inf_{\mathbb{Q} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{Q}}[U(X_T^\pi - \xi)]. \tag{3.1}$$

In the case where $\mathcal{P}_H$ contains only one probability measure, the problem reduces to the classical utility maximization problem.

REMARK 3.2. Due to the construction of 2BSDEs, we must have $\xi \in \mathcal{L}_H^\infty$. It is easy to see that ξ can be constant, deterministic, or in the form of $g(B_T)$ where g is a Lipschitz bounded function, such as a Put or a Call spread payoff function. However, it can be noted that vanilla options payoffs with underlying S may not be in $\mathcal{L}_H^\infty$. Indeed, we have in the one-dimensional framework

$$S_T = S_0 \exp\left(\int_0^T b_t dt - \frac{1}{2}\langle B\rangle_T + B_T\right),\ \mathcal{P}_H - q.s.$$

Since the quadratic variation of the canonical process can be written as follows:

$$\overline{\lim_{n\to+\infty}} \sum_{i \leq 2^n t} \left(B_{\frac{i+1}{2^n}}(\omega) - B_{\frac{i}{2^n}}(\omega)\right)^2,$$

it is not too difficult to see that S can be approximated by a sequence of random variables in $\mathrm{UC_b}(\Omega)$. Besides, this sequence converges in $\mathcal{L}_H^2$. However, we cannot be sure that it also converges in $\mathcal{L}_H^\infty$, which is the space of interest here.

Of course, in the uncertainty framework, it seems to be a major drawback. Nevertheless, to deal with these options, it is sufficient to redo the whole 2BSDE construction from scratch but taking the exponential of the Brownian motion under the Wiener measure as the canonical process instead of the Brownian motion itself. It would amount to restrict ourselves to the subset $\mathcal{P}_H^+$ of $\mathcal{P}_H$, containing only those $\mathbb{P} \in \mathcal{P}_H$ such that the canonical process is a positive continuous local martingale under $\mathbb{P}$.

To find the value function V^ξ and an optimal trading strategy π^*, we follow the ideas of the general *martingale optimality principle* approach in El Karoui and Rouge (2000) and Hu et al. (2005), but adapt it here to a nonlinear framework. Note that $\mathcal{A}$ is the admissibility set of the strategies π.

Let $\{R^\pi\}_{\pi\in\mathcal{A}}$ be a family of processes which satisfy the following properties.

PROPERTIES 3.3.

(i) $R_T^\pi = U(X_T^\pi - \xi)$ *for all* $\pi \in \mathcal{A}$.
(ii) $R_0^\pi = R_0$ *is constant for all* $\pi \in \mathcal{A}$.
(iii) We have

$$R_t^\pi \geq \operatorname*{ess\,inf}^{\mathbb{P}}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})} \mathbb{E}_t^{\mathbb{P}'}[R_T^\pi],\ \forall \pi \in \mathcal{A}$$

$$R_t^{\pi^*} = \operatorname*{ess\,inf}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(t^+,\mathbb{P})} \mathbb{E}_t^{\mathbb{P}'}[R_T^{\pi^*}] \textit{ for some } \pi^* \in \mathcal{A},\ \mathbb{P}-\text{a.s.} \textit{ for all } \mathbb{P} \in \mathcal{P}_H.$$

Then it follows that

$$\inf_{\mathbb{P} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{P}}[U(X_T^\pi - \xi)] \leq R_0 = \inf_{\mathbb{P} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{P}}[U(X_T^{\pi^*} - \xi)] = V^\xi(x). \tag{3.2}$$

In the following sections, we will follow the ideas of Hu et al. (2005) to construct such a family for our three utility functions U.

4. ROBUST EXPONENTIAL UTILITY

In this section, the exponential utility function which is defined as

$$U(x) = -\exp(-\beta x),\ \ x \in \mathbb{R} \text{ for } \beta > 0,$$

will be considered. In this context, the set of admissible trading strategies is defined as follows:

DEFINITION 4.1. Let A be a closed set in $\mathbb{R}^d$. The set of admissible trading strategies $\mathcal{A}$ consists of all d-dimensional progressively measurable processes, $\pi = (\pi_t)_{0 \leq t \leq T}$ satisfying

$$\pi \in \mathbb{B}\text{MO}(\mathcal{P}_H) \text{ and } \pi_t \in A,\ \ dt \otimes \mathcal{P}_H - \text{a.e.}$$

REMARK 4.2. Many authors have shed light on the natural link between BMO class, exponential uniformly integrable class and BSDEs with quadratic growth. See Barrieu and El Karoui (2011), Briand and Hu (2006), and Hu et al. (2005) among others. In the standard utility maximization problem studied in Hu et al. (2005), their trading strategies satisfy a uniform integrability assumption on the family $(\exp(X_\tau^\pi))_\tau$. Since the optimal strategy is a BMO martingale, it is easy to see that the utility maximization problem can also be solved if the uniform integrability assumption is replaced by a BMO assumption. However, at the end of the day, those two assumptions are deeply linked, as shown in the context of quadratic semimartingales in Barrieu and El Karoui (2011). Nonetheless, in our framework, as explained below in Remark 4.5, it is necessary to generalize the BMO martingale assumption instead of the uniform integrability assumption. Moreover, as recalled in the introduction, from a financial point of view these admissibility sets are related to absence of arbitrage in the market considered.

4.1. Characterization of the Value Function and Existence of Optimal Strategies

The investor wants to solve the maximization problem

$$V^\xi(x) := \sup_{\pi \in \mathcal{A}} \inf_{\mathbb{Q} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{Q}}\left[-\exp\left(X_T^\pi - \xi\right)\right]. \tag{4.1}$$

In order to construct a process R^π which satisfies the Properties 3.3, we set

$$R_t^\pi = -\exp(-\beta(X_t^\pi - Y_t)),\ \ t \in [0, T],\ \ \pi \in \mathcal{A},$$

where $(Y, Z) \in \mathbb{D}_H^\infty \times \mathbb{H}_H^2$ is the unique solution of a 2BSDE with a generator $\widehat{F}$ to be determined

$$Y_t = \xi - \int_t^T Z_s dB_s - \int_t^T \widehat{F}(s, Z_s)\, ds + K_T^{\mathbb{P}} - K_t^{\mathbb{P}},\ \mathbb{P} - \text{a.s.},\ \forall \mathbb{P} \in \mathcal{P}_H.$$

REMARK 4.3. From theorem 3.3 of Possamaï and Zhou (2012), we have the following representation:

$$Y_t = \operatorname*{ess\,sup}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})} y_t^{\mathbb{P}'}.$$

Therefore, in general Y_0 is only $\mathcal{F}_{0^+}$-measurable and therefore not a constant. But by proposition 4.2 of Possamaï and Zhou (2012), we know that the process Y is actually $\mathbb{F}$-measurable (it is true when the terminal condition is in $\mathrm{UC_b}(\Omega)$ and by passing to the limit when the terminal condition is in $\mathcal{L}_H^\infty$). This and the above representation easily imply that

$$Y_0 = \operatorname*{ess\,sup}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(0^+, \mathbb{P})} y_0^{\mathbb{P}'} = \sup_{\mathbb{P}' \in \mathcal{P}_H} y_0^{\mathbb{P}'}.$$

The Blumenthal Zero-One law then ensures that Y_0 is a constant.

Let us now define for all $a \in \mathbb{S}_d^{>0}$ such that $\underline{a} \leq a \leq \overline{a}$ the set A_a by

$$A_a := a^{1/2} A = \left\{ a^{1/2} b,\ b \in A \right\}.$$

For any $a \in [\underline{a}, \overline{a}]$, the set A_a is still closed. Moreover, since $A \neq \varnothing$ we have

$$\min\left\{ |r|\,,\ r \in A_a \right\} \leq k \tag{4.2}$$

for some constant k independent of a. We can now state the main result of this section

THEOREM 4.4. *Assume that $\xi \in \mathcal{L}_H^\infty$ and either that $\|\xi\|_{\mathbb{L}_H^\infty} + \sup_{0 \leq t \leq T} \|b_t\|_{\mathbb{L}_H^\infty}$ is small and that $0 \in A$, or that the set A is C^2 (in the sense that its border is a C^2 Jordan arc). Then, the value function of the optimization problem (4.1) is given by*

$$V^\xi(x) = -\exp\left(-\beta\,(x - Y_0)\right),$$

where Y_0 is defined as the initial value of the unique solution $(Y, Z) \in \mathbb{D}_H^\infty \times \mathbb{H}_H^2$ to the following 2BSDE:

$$Y_t = \xi - \int_t^T Z_s dB_s - \int_t^T \widehat{F}_s(Z_s)\, ds + K_T^{\mathbb{P}} - K_t^{\mathbb{P}},\ \mathbb{P} - \text{a.s.},\ \forall \mathbb{P} \in \mathcal{P}_H. \tag{4.3}$$

The generator is defined as follows:

$$\widehat{F}_t(\omega, z) := F_t(\omega, z, \widehat{a}_t), \tag{4.4}$$

where for all $t \in [0, T]$, $z \in \mathbb{R}^d$ and $a \in \mathbb{S}_d^{>0}$

$$F_t(\omega, z, a) = -\frac{\beta}{2} \operatorname{dist}^2\left(a^{1/2} z + \frac{1}{\beta} \theta_t(\omega),\ A_a \right) + z' a^{1/2} \theta_t(\omega) + \frac{1}{2\beta} \left|\theta_t(\omega)\right|^2,$$

where $\theta_t(\omega) := a^{-\frac{1}{2}} b_t(\omega)$ *and for all* $x \in \mathbb{R}^d$ *and* $E \subset \mathbb{R}^d$, $\mathrm{dist}(x, E)$ *is the distance from* x *to* E.

Moreover, there is an optimal trading strategy π^* *satisfying*

$$\widehat{a}_t^{1/2}\pi_t^* \in \Pi_{A_{\widehat{a}_t}}\left(\widehat{a}_t^{1/2} Z_t + \frac{1}{\beta}\widehat{\theta}_t\right), \quad t \in [0, T], \ \mathcal{P}_H - q.s. \tag{4.5}$$

with $\widehat{\theta}_t := \widehat{a}_t^{-1/2} b_t$.

Proof. The proof is divided into five steps. First, it is shown that the 2BSDE with the generator defined in (4.4) has indeed a unique solution. Then, we prove a multiplicative decomposition for the process R^π and some BMO integrability results on the process Z and the optimal strategy π^*. Using these results, we are then able to show that (iii) of Properties 3.3 holds.

Step 1: We show first that the 2BSDE (4.3) has a unique solution. We need to verify that the generator $\widehat{F}$ satisfies the conditions of Assumption 2.4 or 2.3.
First of all, F defined above is a convex function of a, and thus for any $t \in [0, T]$, F can be written as the Fenchel transform of a function

$$H_t(\omega, z, \gamma) := \sup_{a \in D_F}\left\{\frac{1}{2}\mathrm{Tr}(a\gamma) - F_t(\omega, z, a)\right\} \text{ for } \gamma \in \mathbb{R}^{d\times d}.$$

It is clear that F satisfies the first two conditions of either Assumption 2.4 or 2.3. For Assumptions 2.4(iii) and 2.3(iii), the assumption of boundedness and uniform continuity in ω on b implies that b^2 is uniformly continuous in ω. Since b and b^2 are the only nondeterministic terms in F, then F is also uniformly continuous in ω.
As we consider the distance function to a closed set, we know that it is attained for some element of $\mathbb{R}^d$. Besides, as recalled earlier in (4.2), there is a constant $k \geq 0$ such that

$$\min\left\{|d|,\ d \in A_{\widehat{a}_t}\right\} \leq k,\ dt \otimes \mathbb{P} - \text{a.e., for all } \mathbb{P} \in \mathcal{P}_H.$$

Then we get, for all $z \in \mathbb{R}^d$, $t \in [0, T]$,

$$\mathrm{dist}^2\left(\widehat{a}_t^{1/2} z + \frac{1}{\beta}\widehat{\theta}_t, A_{\widehat{a}_t}\right) \leq 2\left|\widehat{a}_t^{1/2} z\right|^2 + 2\left(\frac{1}{\beta}\left|\widehat{\theta}_t\right| + k\right)^2.$$

Thus, we obtain from the boundedness of $\widehat{\theta}$

$$\left|\widehat{F}_t(z)\right| \leq c_0 + c_1\left|\widehat{a}_t^{1/2} z\right|^2,$$

that is to say that Assumptions 2.4(iv) and 2.3(iv) are satisfied.
Finally, Assumption 2.4(v) is clear from the Lipschitz property of the distance function, and Assumption 2.3(v) is also clear from the regularity assumption on the border of A.
The terminal condition ξ is in $\mathcal{L}_H^\infty$ and we have proved that the generator $\widehat{F}$ satisfies Assumption 2.4 or 2.3. Moreover, by definition of F, it is clear that if b has a small $\mathbb{L}_H^\infty$-norm and if $0 \in A$, then $\widehat{F}^0$ also has a small $\mathbb{L}_H^\infty$-norm. Indeed, we have

$$\widehat{F}_t^0 = -\frac{\beta}{2}\mathrm{dist}\left(\frac{\theta_t}{\beta}, A_{\widehat{a}_t}\right) + \frac{1}{2\beta}|\theta_t|^2,$$

which tends to 0 as b_t and thus θ_t goes to 0 (it is clear for the second term on the right-hand side, and for the first, continuity of the distance function and the fact $0 \in A$ ensure the result).

Therefore, Theorem 2.7 states that the 2BSDE (4.3) has a unique solution in $\mathbb{D}^\infty_H \times \mathbb{H}^2_H$.
Step 2: We first decompose R^π as the product of a process M^π and a nonincreasing process N^π which is constant for some $\pi^* \in \mathcal{A}$.
Define for all $\mathbb{P} \in \mathcal{P}_H$ any for any $t \in [0, T]$

$$M^\pi_t = e^{-\beta(x-Y_0)}\exp\left(-\int_0^t \beta(\pi_s - Z_s)\,dB_s - \frac{1}{2}\int_0^t \beta^2\left|\widehat{a}_s^{1/2}(\pi_s - Z_s)\right|^2 ds - \beta K^{\mathbb{P}}_t\right),\quad \mathbb{P}-\text{a.s.}$$

We can then write for all $t \in [0, T]$

$$R^\pi_t = M^\pi_t N^\pi_t,$$

with

$$N^\pi_t = -\exp\left(\int_0^t v(s,\pi_s,Z_s)\,ds\right),$$

and

$$v(t,\pi,z) = -\beta\pi b_t + \beta\widehat{F}_t(z) + \frac{1}{2}\beta^2\left|\widehat{a}_t^{1/2}(\pi - z)\right|^2.$$

Clearly, for every $t \in [0, T]$, $v(t,\pi_t,Z_t)$ can be rewritten in the following form:

$$\begin{aligned}\frac{1}{\beta}v(t,\pi_t,Z_t) &= \frac{\beta}{2}\left|\widehat{a}_t^{1/2}\pi_t\right|^2 - \beta\pi_t'\widehat{a}_t^{1/2}\left(\widehat{a}_t^{1/2}Z_t + \frac{1}{\beta}\widehat{\theta}_t\right) + \frac{\beta}{2}\left|\widehat{a}_t^{1/2}Z_t\right|^2 + \widehat{F}_t(Z_t)\\ &= \frac{\beta}{2}\left|\widehat{a}_t^{1/2}\pi_t - \left(\widehat{a}_t^{1/2}Z_t + \frac{1}{\beta}\widehat{\theta}_t\right)\right|^2 - Z_t'\widehat{a}_t^{1/2}\widehat{\theta}_t - \frac{1}{2\beta}\left|\widehat{\theta}_t\right|^2 + \widehat{F}_t(Z_t).\end{aligned}$$

By a classical measurable selection theorem (see El Karoui 1981 or lemma 3.1 in El Karoui, Peng, and Quenez 1994), we can define a progressively measurable process π^* satisfying (4.5). Then, it follows from the definition of $\widehat{F}$ that $\mathcal{P}_H - q.s.$

- $v(t,\pi_t,Z_t) \geq 0$ for all $\pi \in \mathcal{A}$, $t \in [0, t]$.
- $v(t,\pi^*_t,Z_t) = 0$, $t \in [0, T]$,

which implies that the process N^π is always nonincreasing for all π and is equal to -1 for π^*.
Step 3: In this step, we show that the processes

$$\int_0^\cdot Z_s\,dB_s,\quad \int_0^\cdot \pi^*_s\,dB_s$$

are $\mathbb{B}\mathrm{MO}(\mathcal{P}_\mathrm{H})$ martingales.
First of all, by lemma 2.1 in Possamaï and Zhou (2012), we know that $\int_0^\cdot Z_s\,dB_s$ is a $\mathbb{B}\mathrm{MO}(\mathcal{P}_\mathrm{H})$ martingale. By the triangle inequality and the definition of π^* together with (4.2), we have for all $t \in [0, T]$

$$\begin{aligned}\left|\widehat{a}_t^{1/2}\pi^*_t\right| &\leq \left|\widehat{a}_t^{1/2}Z_t + \frac{1}{\beta}\widehat{\theta}_t\right| + \left|\widehat{a}_t^{1/2}\pi^*_t - \left(\widehat{a}_t^{1/2}Z_t + \frac{1}{\beta}\widehat{\theta}_t\right)\right|\\ &\leq 2\left|\widehat{a}_t^{1/2}Z_t\right| + \frac{2}{\beta}\left|\widehat{\theta}_t\right| + k \leq 2\left|\widehat{a}_t^{1/2}Z_t\right| + k_1,\end{aligned}$$

where k_1 is a bound on $\widehat{\theta}$. Then, for every probability $\mathbb{P} \in \mathcal{P}_H$ and every stopping time $\tau \leq T$,

$$\mathbb{E}_\tau^{\mathbb{P}}\left[\int_\tau^T \left|\widehat{a}_t^{1/2}\pi_t^*\right|^2 dt\right] \leq \mathbb{E}_\tau^{\mathbb{P}}\left[\int_\tau^T 8\left|\widehat{a}_t^{1/2}Z_t\right|^2 dt + 2Tk_1^2\right],$$

and therefore

$$\|\pi^*\|^2_{\mathbb{BMO}(\mathcal{P}_\mathrm{H})} \leq 8\|Z\|^2_{\mathbb{BMO}(\mathcal{P}_\mathrm{H})} + 2Tk_1^2,$$

which implies the $\mathbb{BMO}(\mathcal{P}_\mathrm{H})$ martingale property of $\int_0^\cdot \pi_s^* dB_s$ as desired.
Step 4: We prove that $\pi^* \in \mathcal{A}$ and $R^{\pi^*} \equiv -M^{\pi^*}$ satisfies (iii) of Properties 3.3, that is to say for all $t \in [0, T]$

$$\operatorname*{ess\,sup}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})} \mathbb{E}_t^{\mathbb{P}'}\left[M_T^{\pi^*}\right] = M_t^{\pi^*}, \ \ \mathbb{P}-\text{a.s.}, \ \forall \mathbb{P} \in \mathcal{P}_H.$$

For a fixed $\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})$, we denote

$$L_t := \int_0^t \beta(\pi_s^* - Z_s)\, dB_s + \frac{1}{2}\int_0^t \beta^2 \left|\widehat{a}_s^{1/2}(\pi_s^* - Z_s)\right|^2 ds + \beta K_t^{\mathbb{P}'}, \ 0 \leq t \leq T,$$

then with Itô's formula and thanks to the $\mathbb{BMO}(\mathcal{P}_\mathrm{H})$ property proved in Step 3, we obtain for every $t \in [0, T]$,

$$\begin{aligned} \text{(4.6)} \qquad \mathbb{E}_t^{\mathbb{P}'}\left[M_T^{\pi^*}\right] - M_t^{\pi^*} &= -\beta\mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s-}^{\pi^*} dK_s^{\mathbb{P}'}\right] \\ &\quad + \mathbb{E}_t^{\mathbb{P}'}\left[\sum_{t \leq s \leq T} e^{-L_s} - e^{-L_{s-}} + e^{-L_{s-}}(L_s - L_{s-})\right]. \end{aligned}$$

First, we prove

$$\operatorname*{ess\,inf}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})} \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s-}^{\pi^*} dK_s^{\mathbb{P}'}\right] = 0, \ \ t \in [0, T], \ \ \mathbb{P}-\text{a.s.}$$

For every t and every $\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})$, we have

$$0 \leq \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s-}^{\pi^*} dK_s^{\mathbb{P}'}\right] \leq \mathbb{E}_t^{\mathbb{P}'}\left[\left(\sup_{0 \leq s \leq T} M_s^{\pi^*}\right)\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)\right].$$

Besides, since $K^{\mathbb{P}'}$ is nondecreasing, we obtain for all $s \geq t$

$$M_s^{\pi^*} \leq e^{-\beta(x - Y_0)}\mathcal{E}\left(\beta\int_0^s \left(Z_u - \pi_u^*\right) dB_u\right).$$

Then, again thanks to Step 3, we know that

$$\left(Z_s - \pi_s^*\right) \in \mathbb{BMO}(\mathcal{P}_\mathrm{H}),$$

and thus the exponential martingale above is a uniformly integrable martingale for all $\mathbb{P}$ and is in L_H^r for some $r > 1$ (see lemma 2.2 in Possamaï and Zhou 2012). Thus, by

Hölder inequality, we have for all $t \in [0, T]$

$$\mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s^-}^{\pi^*} dK_s^{\mathbb{P}'}\right] \leq e^{\beta(Y_0 - x)} \mathbb{E}_t^{\mathbb{P}'}\left[\sup_{0 \leq s \leq T} \mathcal{E}^r\left(\beta \int_0^s \left(Z_u - \pi_u^*\right) dB_u\right)\right]^{\frac{1}{r}}$$
$$\times \mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)^q\right]^{\frac{1}{q}}.$$

With Doob's maximal inequality, we have for every $t \in [0, T]$

$$\mathbb{E}_t^{\mathbb{P}'}\left[\sup_{0 \leq s \leq T} \mathcal{E}^r\left(\beta \int_0^s \left(Z_u - \pi_u^*\right) dB_u\right)\right]^{1/r} \leq$$
$$C \mathbb{E}_t^{\mathbb{P}'}\left[\mathcal{E}^r\left(\beta \int_0^T \left(Z_u - \pi_u^*\right) dB_u\right)\right]^{1/r} < +\infty,$$

where C is an universal constant which can change value from line to line.

From Cauchy–Schwarz inequality, we see for $0 \leq t \leq T$

$$\mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)^q\right]^{1/q} \leq C\left(\mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)\right] \mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)^{2q-1}\right]\right)^{\frac{1}{2q}}$$
$$\leq C\left(\underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,sup}^{\mathbb{P}}} \mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)^{2q-1}\right]\right)^{\frac{1}{2q}}$$
$$\times \left(\mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)\right]\right)^{\frac{1}{2q}}.$$

Arguing as in the proof of theorem 3.1 in Possamaï and Zhou (2012) we know that for $t \in [0, T]$

$$\left(\underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,sup}^{\mathbb{P}}} \mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)^{2q-1}\right]\right)^{\frac{1}{2q}} < +\infty, \ \mathbb{P} - \text{a.s.}$$

Hence, we obtain for $0 \leq t \leq T$

$$0 \leq \underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s^-}^{\pi^*} dK_s^{\mathbb{P}'}\right] \leq C \underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \left(\mathbb{E}_t^{\mathbb{P}'}\left[\left(K_T^{\mathbb{P}'} - K_t^{\mathbb{P}'}\right)\right]\right)^{\frac{1}{2q}} = 0, \ \mathbb{P} - \text{a.s.},$$

which means

$$\underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s^-}^{\pi^*} dK_s^{\mathbb{P}'}\right] = 0, \ 0 \leq t \leq T, \ \mathbb{P} - \text{a.s.}$$

Finally, we have for every $t \in [0, T]$

$$\underset{\mathbb{P}' \in \mathcal{P}_H(t^+, \mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s^-}^{\pi^*} dK_s^{\mathbb{P}'} - \sum_{t \leq s \leq T} \exp(-\beta L_s) - \exp(-\beta L_{s^-})\right.$$
$$\left. + \beta \exp(-\beta L_{s^-})(L_s - L_{s^-})\right]$$

$$
\begin{aligned}
&\leq \operatorname*{ess\,inf}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_t^{\mathbb{P}'}\left[\int_t^T M_{s^-}^{\pi^*} dK_s^{\mathbb{P}'}\right] \\
&\quad - \operatorname*{ess\,inf}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_t^{\mathbb{P}'}\left[\sum_{t\leq s\leq T} \exp(-\beta L_s) - \exp(-\beta L_{s^-}) + \beta \exp(-\beta L_{s^-})(L_s - L_{s^-})\right] \\
&\leq 0,\ \mathbb{P}-\text{a.s.},
\end{aligned}
$$

because the function $x \to \exp(-x)$ is convex and the jumps of L are positive. Hence, using (4.6), we have for every $t \in [0, T]$

$$
\operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_t^{\mathbb{P}'}\left[M_T^{\pi^*} - M_t^{\pi^*}\right] \geq 0,\ \mathbb{P}-\text{a.s.}
$$

However, by definition M^{π^*} is the product of a martingale and a positive nonincreasing process and is therefore a supermartingale. It implies that for every $t \in [0, T]$

$$
\operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_t^{\mathbb{P}'}\left[M_T^{\pi^*} - M_t^{\pi^*}\right] = 0,\ \mathbb{P}-\text{a.s.}
$$

Finally, π^* is an admissible strategy, R^{π^*} satisfies (iii) of Properties 3.3, and

$$
R_0^{\pi^*} = \inf_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[-\exp\left(-\beta\left(x + \int_0^T \pi_s^* (dB_s + \theta_s ds) - \xi\right)\right)\right] = -\exp\left(-\beta\left(x - Y_0\right)\right).
$$

Step 5: Next we will show that for all $\pi \in \mathcal{A}$, R^π satisfies (iii) of Properties 3.3, that is, for every $t \in [0, T]$

$$
\operatorname*{ess\,inf}_{\mathbb{P}'\in\mathcal{P}_H(t^+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_t^{\mathbb{P}'}[-\exp(-\beta(X_T^\pi - \xi))] \leq R_t^\pi,\ \mathbb{P}-\text{a.s.}
$$

Since $\pi \in \mathcal{A}$, the process

$$
\int_0^\cdot (Z_s - \pi_s)\, dB_s
$$

is a $\mathbb{B}$MO($\mathcal{P}_\mathrm{H}$) martingale. Then, the process

$$
G^\pi = \exp\left(-\beta(x - Y_0)\right) \mathcal{E}\left(-\beta \int_0^\cdot (\pi_s - Z_s)\, dB_s\right)
$$

is a uniformly integrable martingale under each $\mathbb{P} \in \mathcal{P}_H$.
As in the previous steps, we write R^π as $R^\pi = M^\pi N^\pi$, where N^π is a negative nonincreasing process. We then have for $0 \leq s \leq t \leq T$

$$
\operatorname*{ess\,inf}_{\mathbb{P}'\in\mathcal{P}_H(s+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_s^{\mathbb{P}'}\left[M_t^\pi N_t^\pi\right] \leq \operatorname*{ess\,inf}_{\mathbb{P}'\in\mathcal{P}_H(s+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_s^{\mathbb{P}'}\left[M_t^\pi N_s^\pi\right] = \operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}_H(s+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_s^{\mathbb{P}'}\left[M_t^\pi\right] N_s^\pi,\ \mathbb{P}-\text{a.s.},
$$

because N^π is negative. By the same arguments as in Step 3 for M^{π^*}, we have for $0 \leq s \leq t \leq T$

$$
\operatorname*{ess\,sup}_{\mathbb{P}'\in\mathcal{P}_H(s+,\mathbb{P})}^{\mathbb{P}} \mathbb{E}_s^{\mathbb{P}'}\left[M_t^\pi\right] = M_s^\pi,\ \mathbb{P}-\text{a.s.}
$$

Therefore, the following inequality holds for $0 \leq s \leq t \leq T$

$$\underset{\mathbb{P}' \in \mathcal{P}_H(s+,\mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \mathbb{E}_s^{\mathbb{P}'} \left[R_t^{\pi} \right] \leq R_s^{\pi}, \ \mathbb{P} - \text{a.s.},$$

which ends the proof. □

REMARK 4.5. Here, it can be seen why it is essential in this context to have strong integrability assumptions on the trading strategies. Indeed, in the proof of the above property for M^{π^*}, the fact that the stochastic integral

$$\int_0^{\cdot} \pi_s^* dB_s$$

is a $\mathbb{B}\mathrm{MO}(\mathcal{P}_\mathrm{H})$ martingale allowed us to control the moments of its stochastic exponential, which in turn allowed us to deduce from the minimal property for $K^{\mathbb{P}}$ a similar minimal property for

$$\int_0^{\cdot} M_s^{\pi^*} dK_s^{\mathbb{P}}.$$

This term is new when compared with the context of Hu et al. (2005). To deal with it, the $\mathbb{B}\mathrm{MO}(\mathcal{P}_\mathrm{H})$ property has to be imposed. Note however that since the optimal strategy already has that property, we do not lose much by restricting the strategies.

REMARK 4.6. We note that the approach still works when there are no constraints on trading strategies. In this case, the 2BSDE related to the maximization problem has a uniformly Lipschitz generator, and we are in the context of complete markets. Then, the theory developed in Soner et al. (2012) for Lipschitz 2BSDEs can also be used.

4.7. A Min-Max Property

By comparing the value function of our robust utility maximization problem and the one presented in Hu et al. (2005) for standard utility maximization problem, we are able to obtain a min-max property similar to that obtained by Denis and Kervarec (2007). We observe that we were only able to prove this property after having solved the initial problem, unlike in the approach of Denis and Kervarec (2007).

THEOREM 4.7. *Under the previous assumptions on the probability measures set $\mathcal{P}_H$ and the admissible strategies set $\mathcal{A}$, the following min-max property holds:*

$$\sup_{\pi \in \mathcal{A}} \inf_{\mathbb{P} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right] = \inf_{\mathbb{P} \in \mathcal{P}_H} \sup_{\pi \in \mathcal{A}} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right] = \inf_{\mathbb{P} \in \mathcal{P}_H} \sup_{\pi \in \mathcal{A}^{\mathbb{P}}} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right],$$

where $\mathcal{A}^{\mathbb{P}}$ is the set consisting of trading strategies π which are in A, $\mathbb{P} -$ a.s.*, and such that the process $\int_0^{\cdot} \pi_s dB_s$ is a $BMO(\mathbb{P})$ martingale.*

Proof. First note that we have

$$D := \sup_{\pi \in \mathcal{A}} \inf_{\mathbb{P} \in \mathcal{P}_H} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right] \leq \inf_{\mathbb{P} \in \mathcal{P}_H} \sup_{\pi \in \mathcal{A}} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right] \leq \inf_{\mathbb{P} \in \mathcal{P}_H} \sup_{\pi \in \mathcal{A}^{\mathbb{P}}} \mathbb{E}^{\mathbb{P}} \left[R_T^{\pi} \right] =: C.$$

Indeed, the first inequality is obvious and the second one follows from the fact that for all $\mathbb{P}$, $\mathcal{A} \subset \mathcal{A}^{\mathbb{P}}$.

That $C \leq D$ remains to be proved. By the previous sections, we know that

$$D = -\exp\left(-\beta\left(x - Y_0\right)\right).$$

Moreover, we know from Possamaï and Zhou (2012) that we have a representation for Y_0,

$$Y_0 = \sup_{\mathbb{P}\in\mathcal{P}_H} y_0^{\mathbb{P}},$$

where $y_0^{\mathbb{P}}$ is the solution of the standard BSDE with the same generator $\widehat{F}$. On the other hand, it can be observed from Hu et al. (2005) that

$$C = \inf_{\mathbb{P}\in\mathcal{P}_H} \left[-\exp\left(-\beta\left(x - y_0^{\mathbb{P}}\right)\right)\right]$$

implying that $C = D$. □

4.3. Indifference Pricing via Robust Utility Maximization

It has been shown in El Karoui and Rouge (2000) that in a market model with constraints on the portfolios, if the indifference price for a claim Φ is defined as the smallest number p such that

$$\sup_{\pi} \mathbb{E}\left[-\exp\left(-\beta\left(X^{x+p,\pi} - \Phi\right)\right)\right] \geq \sup_{\pi} \mathbb{E}\left[-\exp\left(-\beta X^{x,\pi}\right)\right],$$

where $X^{x,\pi}$ is the wealth associated with the portfolio π and initial value x, then this problem turns into the resolution of a BSDE with quadratic growth generator.

In this framework of uncertain volatility, the problem of indifference pricing of a contingent claim Φ boils down to solving the following equation in p:

$$V^0(x) = V^{\Phi}(x + p).$$

Thanks to our results, we know that if $\Phi \in \mathcal{L}_H^{\infty}$ then the two sides of the above equality can be computed by solving 2BSDEs. The indifference price p can therefore be calculated as soon as the 2BSDEs can be solved (explicitly or numerically). Two examples are provided in Section 7.

5. ROBUST POWER UTILITY

In this section, we will consider the power utility function

$$U(x) = -\frac{1}{\gamma} x^{-\gamma}, \; x > 0, \; \gamma > 0.$$

Here, a different notion of trading strategy will be used: $\rho = (\rho^i)_{i=1,\dots,d}$ denotes the proportion of wealth invested in stock i. The number of shares of stock i is given by $\frac{\rho_t^i X_t}{S_t^i}$.

Then, the wealth process is defined as

$$X_t^{\rho} = x + \int_0^t \sum_{i=1}^{d} \frac{X_s^{\rho} \rho_s^i}{S_s^i} dS_s^i = x + \int_0^t X_s^{\rho} \rho_s \left(dB_s + b_s ds\right), \; \mathcal{P}_H - q.s. \tag{5.1}$$

and the initial capital x is positive.

In the present setting, the set of admissible strategies is defined as follows:

DEFINITION 5.1. Let A be a closed set in $\mathbb{R}^d$. The set of admissible trading strategies $\mathcal{A}$ consists of all $\mathbb{R}^d$-valued progressively measurable processes $\rho = (\rho_t)_{0\leq t\leq T}$ satisfying

$$\rho \in \mathbb{BMO}(\mathcal{P}_H) \text{ and } \rho \in A,\ dt \otimes \mathcal{P}_H - \text{a.e.}$$

The wealth process X^ρ can be written as

$$X_t^\rho = x\mathcal{E}\left(\int_0^t \rho_s(dB_s + b_s ds)\right),\ t \in [0, T],\ \mathcal{P}_H - q.s.$$

Then for every $\rho \in \mathcal{A}$, the wealth process X^ρ is a local $\mathbb{P}$-martingale bounded from below, hence, a $\mathbb{P}$-supermartingale, for all $\mathbb{P} \in \mathcal{P}_H$.

We suppose that there is no liability ($\xi = 0$). Then, the investor faces the maximization problem

$$V(x) = \sup_{\rho\in\mathcal{A}} \inf_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[U(X_T^\rho)\right]. \tag{5.2}$$

In order to find the value function and an optimal strategy, we follow the method outlined in the previous section for the exponential utility. Therefore, we have to construct a stochastic process R^ρ with terminal value

$$R_T^\rho = U\left(x + \int_0^T X_s^\rho \rho_s \frac{dS_s}{S_s}\right),$$

and which satisfies Properties 3.3. Then the value function will be given by $V(x) = R_0$.

Applying the utility function to the wealth process yields

$$-\frac{1}{\gamma}\left(X_t^\rho\right)^{-\gamma} = -\frac{1}{\gamma}x^{-\gamma}\exp\left(-\int_0^t \gamma\rho_s dB_s - \int_0^t \gamma\rho_s b_s ds + \frac{1}{2}\int_0^t \gamma\left|\widehat{a}_s^{1/2}\rho_s\right|^2 ds\right). \tag{5.3}$$

This equation suggests the following choice:

$$R_t^\rho = -\frac{1}{\gamma}x^{-\gamma}\exp\left(-\int_0^t \gamma\rho_s dB_s - \int_0^t \gamma\rho_s b_s ds + \frac{1}{2}\int_0^t \gamma\left|\widehat{a}_s^{1/2}\rho_s\right|^2 ds + Y_t\right),$$

where $(Y, Z) \in \mathbb{D}_H^\infty \times \mathbb{H}_H^2$ is the unique solution of the following 2BSDE:

$$Y_t = 0 - \int_t^T Z_s\, dBs - \int_t^T \widehat{F}_s(Z_s)\, ds + K_T - K_t,\ t \in [0, T],\ \mathcal{P}_H - q.s. \tag{5.4}$$

In order to get (iii) of Properties 3.3 for R^ρ, we have to construct $\widehat{F}_t(z)$ such that, for $t \in [0, T]$

$$\gamma\rho_t b_t - \frac{1}{2}\gamma\left|\widehat{a}_t^{1/2}\rho_t\right|^2 - \widehat{F}_t(Z_t) \leq -\frac{1}{2}\left|\widehat{a}_t^{1/2}(\gamma\rho_t - Z_t)\right|^2 \text{ for all } \rho \in \mathcal{A} \tag{5.5}$$

with equality for some $\rho^* \in \mathcal{A}$. It is equivalent to

$$\widehat{F}_t(Z_t) \geq\ -\frac{1}{2}\gamma\,(1+\gamma)\left|\widehat{a}_t^{1/2}\rho_t - \frac{1}{1+\gamma}\left(-\widehat{a}_t^{1/2}Z_t + \widehat{\theta}_t\right)\right|^2$$

$$-\frac{1}{2}\frac{\gamma\left|-\widehat{a}_t^{1/2}Z_t+\widehat{\theta}_t\right|^2}{1+\gamma}+\frac{1}{2}\left|\widehat{a}_t^{1/2}Z_t\right|^2$$

with $\widehat{\theta}_t := \widehat{a}_t^{-1/2}b_t$.

Hence, the appropriate choice for $\widehat{F}$ is

$$\widehat{F}_t(z) = -\frac{\gamma(1+\gamma)}{2}\mathrm{dist}^2\left(\frac{-\widehat{a}_t^{1/2}z+\widehat{\theta}_t}{1+\gamma}, A_{\widehat{a}_t}\right)+\frac{\gamma\left|-\widehat{a}_t^{1/2}z+\widehat{\theta}_t\right|^2}{2(1+\gamma)}+\frac{1}{2}\left|\widehat{a}_t^{1/2}z\right|^2, \tag{5.6}$$

and a candidate for the optimal strategy must satisfy

$$\widehat{a}_t^{1/2}\rho_t^* \in \Pi_{A_{\widehat{a}_t}}\left(\frac{1}{1+\gamma}\left(-\widehat{a}_t^{1/2}Z_t+\widehat{\theta}_t\right)\right),\ t\in[0,T].$$

The above results are summarized in the following theorem.

THEOREM 5.2. *Assume either that the drift b verifies that* $\sup_{0\le t\le T}\|b_t\|_{\mathbb{L}_H^\infty}$ *is small and that the set A contains* 0, *or that the set A is C^2 (in the sense that its border is a C^2 Jordan arc). Then, the value function of the optimization problem (5.2) is given by*

$$V(x) = -\frac{1}{\gamma}x^{-\gamma}\exp(Y_0)\ \text{ for } x>0,$$

where Y_0 is defined as the initial value of the unique solution $(Y,Z)\in\mathbb{D}_H^\infty\times\mathbb{H}_H^2$ of the quadratic 2BSDE

$$Y_t = 0-\int_t^T Z_s\,dBs-\int_t^T \widehat{F}_s(Z_s)\,ds+K_T-K_t,\ t\in[0,T],\ \mathcal{P}_H-q.s., \tag{5.7}$$

where $\widehat{F}$ is given by (5.6).

Moreover, there is an optimal trading strategy $\rho^*\in\mathcal{A}$ with the property

$$\widehat{a}_t^{1/2}\rho_t^* \in \Pi_{A_{\widehat{a}_t}}\left(\frac{1}{1+\gamma}\left(-\widehat{a}_t^{1/2}Z_t+\widehat{\theta}_t\right)\right),\ t\in[0,T], \tag{5.8}$$

with $\widehat{\theta}_t := \widehat{a}_t^{-1/2}b_t$.

Proof. The proof is very similar to the case of robust exponential utility. First, it can be shown with the same arguments that the generator $\widehat{F}$ satisfies the conditions of Assumption 2.3 or 2.4. Hence, there exists a unique solution to the 2BSDE (5.7).

Let ρ^* denote the progressively measurable process, constructed with a measurable selection theorem, which realizes the distance in the definition of $\widehat{F}$. The same arguments as in the case of robust exponential utility show that $\rho^*\in\mathcal{A}$.

Then with the choice made for $\widehat{F}$, we have the following multiplicative decomposition:

$$R_t^\rho = -\frac{1}{\gamma}x^{-\gamma}\mathcal{E}\left(-\int_0^t(\gamma\rho_s-Z_s)\,dB_s\right)e^{-\gamma K_t^{\mathbb{P}}}\exp\left(-\int_0^t v_s\,ds\right),$$

where

$$v_t = \gamma\rho_t b_t-\frac{1}{2}\gamma\left|\widehat{a}_t^{1/2}\rho_t\right|^2-\widehat{F}_t(Z_t)+\frac{1}{2}\left|\widehat{a}_t^{1/2}(\gamma\rho_t-Z_t)\right|^2\le 0,\ dt\otimes\mathbb{P}\text{−a.e.}$$

Since the stochastic integral $\int_0^t (\rho_s - Z_s)\, dB_s$ is a $\mathbb{B}\mathrm{MO}(\mathcal{P}_{\mathrm{H}})$ martingale, the stochastic exponential above is a uniformly integrable martingale. By exactly the same arguments as before, we have

$$\underset{\mathbb{P}'\in\mathcal{P}_H(s+,\mathbb{P})}{\operatorname{ess\,inf}^{\mathbb{P}}} \mathbb{E}_s^{\mathbb{P}'}\left[R_t^\rho\right] \le R_s^\rho, \ \ s \le t, \ \mathbb{P}-\text{a.s.},$$

with equality for ρ^*.

Hence, the terminal value R_T^ρ is the utility of the terminal wealth of the trading strategy ρ. Consequently,

$$\inf_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[U\left(X_T^\rho\right)\right] \le R_0 = -\frac{1}{\gamma}x^{-\gamma}\exp(Y_0) \ \text{ for all } \rho \in \mathcal{A}. \qquad \square$$

REMARK 5.3. Of course, the min-max property of Theorem 4.7 still holds.

6. ROBUST LOGARITHMIC UTILITY

In this section, we consider the logarithmic utility function

$$U(x) = \log(x), \ \ x > 0.$$

Here, we use the same notion of trading strategies as in the power utility case, $\rho = (\rho^i)_{i=1,\dots,d}$ denotes the part of the wealth invested in stock i. The number of shares of stock i is given by $\frac{\rho_t^i X_t}{S_t^i}$. Then, the wealth process is defined as

$$X_t^\rho = x + \int_0^t \sum_{i=1}^d \frac{X_s^\rho \rho_s^i}{S_s^i} dS_s^i = x + \int_0^t X_s^\rho \rho_s\left(dB_s + b_s ds\right), \ \mathcal{P}_H - q.s. \tag{6.1}$$

and the initial capital x is positive.

The wealth process X^ρ can be written as

$$X_t^\rho = x\mathcal{E}\left(\int_0^t \rho_s(dB_s + b_s ds)\right), \ t \in [0, T], \ \mathcal{P}_H - q.s.$$

In this case, the set of admissible strategies is defined as follows.

DEFINITION 6.1. Let A be a closed set in $\mathbb{R}^d$. The set of admissible trading strategies $\mathcal{A}$ consists of all $\mathbb{R}^d$-valued progressively measurable processes ρ satisfying

$$\sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[\int_0^T |\widehat{a}_t^{1/2}\rho_t|^2 dt\right] < \infty$$

and $\rho \in A, \ dt \otimes d\mathbb{P} - \text{a.s.}, \ \forall \mathbb{P} \in \mathcal{P}_H$.

For the logarithmic utility, we assume the agent has no liability at time T ($\xi = 0$). Then the optimization problem is given by

$$\begin{aligned} V(x) &= \sup_{\rho\in\mathcal{A}} \inf_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}[\log(X_T^\rho)] \\ &= \log(x) + \sup_{\rho\in\mathcal{A}} \inf_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[\int_0^T \rho_s dB_s + \int_0^T (\rho_s b_s - \frac{1}{2}|\widehat{a}_s^{1/2}\rho_s|^2)ds\right]. \end{aligned} \tag{6.2}$$

We have the following theorem.

THEOREM 6.2. *Assume either that the drift b verifies that $\sup_{0\leq t\leq T}\|b_t\|_{\mathbb{L}^\infty_H}$ is small and that the set A contains 0, or that the set A is C^2 (in the sense that its border is a C^2 Jordan arc). Then, the value function of the optimization problem (6.2) is given by*

$$V(x) = \log(x) - Y_0 \text{ for } x > 0,$$

where Y_0 is defined as the initial value of the unique solution $(Y, Z) \in \mathbb{D}^\infty_H \times \mathbb{H}^2_H$ of the quadratic 2BSDE

$$Y_t = 0 - \int_t^T Z_s\, dB_s - \int_t^T \widehat{F}_s ds + K^{\mathbb{P}}_T - K^{\mathbb{P}}_t,\ t \in [0, T],\ \mathbb{P} - \text{a.s.},\ \forall \mathbb{P} \in \mathcal{P}_H. \tag{6.3}$$

The generator is defined by

$$\widehat{F}_s = -\frac{1}{2}\operatorname{dist}^2(\widehat{\theta}_s, A_{\widehat{a}}) + \frac{1}{2}|\widehat{\theta}_s|^2,$$

where $\widehat{\theta}_t := \widehat{a}_t^{-1/2} b_t$.

Moreover, there exists an optimal trading strategy $\rho^* \in \mathcal{A}$ with the property

$$\widehat{a}_t^{1/2}\rho^*_t \in \Pi_{A_{\widehat{a}_t}}\left(\widehat{\theta}_t\right). \tag{6.4}$$

Proof. The proof is very similar to the case of exponential and power utility. First, we show that there is an unique solution to the 2BSDE (6.3). We then write, for $t \in [0, T]$

$$R^\rho_t = M^\rho_t + N^\rho_t,$$

where

$$M^\rho_t = \log(x) - Y_0 + \int_0^t (\rho_s - Z_s)\, dB_s + K^{\mathbb{P}}_t,$$
$$N^\rho_t = \int_0^t \left(-\frac{1}{2}\left|\widehat{a}_s^{1/2}\rho_s - \widehat{\theta}_s\right|^2 + \frac{1}{2}\left|\widehat{\theta}_s\right|^2 - \widehat{F}_s \right) ds.$$

Similarly, we prove that ρ^*, which can be constructed by means of a classical measurable selection argument, is in $\mathcal{A}$. Note in particular that ρ^* only depends on $\widehat{\theta}$, $\widehat{a}^{1/2}$ and the closed set A describing the constraints on the trading strategies.

Next, due to Definition 6.1, the stochastic integral in R^ρ is a martingale under each $\mathbb{P}$ for all $\rho \in \mathcal{A}$. Moreover, $\widehat{F}$ is chosen to make the process N^ρ nonincreasing for all ρ and a constant for ρ^*. Thus, the minimum condition of $K^{\mathbb{P}}$ implies that R^ρ satisfies (iii) of Properties 3.3.

Furthermore, the initial value Y_0 of the simple 2BSDE (6.3) satisfies

$$Y_0 = -\sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[\int_0^T \widehat{F}_s ds\right].$$

Hence,

$$V(x) = R^{\rho^*}_0(x) = \log(x) + \sup_{\mathbb{P}\in\mathcal{P}_H} \mathbb{E}^{\mathbb{P}}\left[\int_0^T \widehat{F}_s ds\right]. \qquad \square$$

REMARK 6.3. Of course, the min-max property of Theorem 4.7 still holds. Moreover, it is an easy exercise to show that the 2BSDE has a unique solution in this case given by

$$Y_t = \operatorname*{ess\,sup}^{\mathbb{P}}_{\mathbb{P}' \in \mathcal{P}_H(t^+,\mathbb{P})} \mathbb{E}^{\mathbb{P}'}\left[\int_t^T \frac{1}{2}\left(\operatorname{dist}^2(\theta_s, A_{\widehat{a}_s}) - |\theta_s|^2\right) ds\right], \ \mathbb{P}-\text{a.s.}, \ \textit{for all}\ \mathbb{P} \in \mathcal{P}_H.$$

7. EXAMPLES

In general, it is difficult to solve BSDEs and 2BSDEs explicitly. In this section, some examples with an explicit solution will be given. In particular, we show how the optimal probability measure is chosen. In all our examples, we will work in dimension one, $d = 1$.

First, robust exponential utility is dealt with. We consider the case where there are no constraints on trading strategies, that is $A = \mathbb{R}$. Then, the associated 2BSDE has a generator which is linear in z. In the first example, we consider a deterministic terminal liability ξ and show that our result can be compared with the one obtained by solving the Hamilton–Jacobi–Bellman (HJB) equation in the standard Merton's approach, working with the probability measure associated with the constant process $\overline{a}$. In the second example, we show that with a random payoff $\xi = -B_T^2$, where B is the canonical process, we end up with an optimal probability measure which is not of Bang–Bang type (Bang–Bang type means that, under this probability measure, the density of the quadratic variation $\widehat{a}$ takes only the two extreme values, $\underline{a}$ and $\overline{a}$). We emphasize that this example does not have real financial significance, but nonetheless shows that one cannot expect the optimal probability measure to depend only on the two bounds for the volatility unlike with option pricing in the uncertain volatility model.

7.1. Example 1: Deterministic Payoff

In this example, we suppose that b is a constant in $\mathbb{R}$. From Theorem 4.4, we know that the value function of the robust maximization problem is given by

$$V^{\xi}(x) = -\exp\left(-\beta\left(x - Y_0\right)\right),$$

where Y is the solution of a 2BSDE with quadratic generator. When there are no constraints, the 2BSDE can be written as follows:

$$Y_t = \xi - \int_t^T Z_s\, dB_s - \int_t^T \widehat{F}_s(Z_s)\, ds + K_T^{\mathbb{P}} - K_t^{\mathbb{P}}, \ \mathbb{P}-\text{a.s.}, \ \forall \mathbb{P} \in \mathcal{P}_H$$

and the generator is given by

$$\widehat{F}_t(z) := F_t(\omega, z, \widehat{a}) = bz + \frac{b^2}{2\beta\widehat{a}}.$$

Then, the corresponding BSDEs can be solved explicitly with the same generator under each $\mathbb{P}$. By setting

$$M_t = e^{-\int_0^t \frac{1}{2} b^2 \widehat{a}_s^{-1} ds - \int_0^t b\widehat{a}_s^{-1} dB_s},$$

and applying Itô's formula to $y_t^{\mathbb{P}} M_t$, we have

$$y_0^{\mathbb{P}} = \mathbb{E}^{\mathbb{P}}\left[\xi M_T - \frac{b^2}{2\beta}\int_0^T \widehat{a}_s^{-1} M_s ds\right].$$

Since $\underline{a} \leq \widehat{a} \leq \overline{a}$, we derive that

$$y_0^{\mathbb{P}} \leq \xi - \frac{1}{2\beta}\frac{b^2}{\overline{a}}T.$$

Therefore, by the representation of Y, we have

$$Y_0 \leq \xi - \frac{1}{2\beta}\frac{b^2}{\overline{a}}T.$$

Moreover, under the specific probability measure $\mathbb{P}^{\overline{a}} \in \mathcal{P}_H$, we have

$$y_0^{\mathbb{P}^{\overline{a}}} = \xi - \frac{1}{2\beta}\frac{b^2}{\overline{a}}T.$$

It implies that $Y_0 = y_0^{\mathbb{P}^{\overline{a}}}$, which means that the robust utility maximization problem is degenerated and is equivalent to a standard utility maximization problem under the probability measure $\mathbb{P}^{\overline{a}}$. This result is discussed in more details in Example 7.3 below.

7.2. Example 2: Nondeterministic Payoff

In this subsection, we consider a nondeterministic payoff $\xi = -B_T^2$. As in the first example, there are no constraints on trading strategies. Then, the 2BSDE has a linear generator. We can verify that $-B_T^2$ can be written as the limit under the norm $\|\cdot\|_{\mathbb{L}_H^{2,\kappa}}$ of a sequence which is in $\mathrm{UC_b}(\Omega)$, and thus is in $\mathcal{L}_H^{2,\kappa}$, which is the terminal condition set for 2BSDEs with Lipschitz generator (these sets are defined in Section 2.2). Here, we suppose that b is a deterministic continuous function of time t.

By the same method as in the previous example, let

$$M_t = e^{-\int_0^t \frac{1}{2} b_s^2 \widehat{a}_s^{-1} ds - \int_0^t b_s \widehat{a}_s^{-1} dB_s},$$

then we obtain

$$y_0^{\mathbb{P}} = \mathbb{E}^{\mathbb{P}}\left[-M_T B_T^2 - \int_0^T \frac{b_s^2}{2\beta}\widehat{a}_s^{-1} M_s ds\right].$$

By applying Itô's formula to $M_t B_t$, we have

$$d M_t B_t = M_t dB_t + B_t d M_t - b_t M_t dt.$$

Since b is deterministic, by taking expectation under $\mathbb{P}$ and localizing if necessary, we obtain

$$\mathbb{E}^{\mathbb{P}}[M_T B_T] = \mathbb{E}^{\mathbb{P}}\left[-\int_0^T b_t M_t dt\right] = -\int_0^T b_t dt.$$

Again, by applying Itô's formula to $-M_t B_t^2$, we have

$$-dM_t B_t^2 = -2M_t B_t dB_t - B_t^2 dM_t - \widehat{a}_t M_t dt + 2b_t M_t B_t dt.$$

Therefore, $y_0^{\mathbb{P}}$ can be rewritten as

$$y_0^{\mathbb{P}} = \mathbb{E}^{\mathbb{P}}\left[\int_0^T -M_t\left(\widehat{a}_t + \frac{b_t^2}{2\beta\widehat{a}_t}\right)dt\right] - \int_0^T 2b_t\left(\int_0^t b_s ds\right)dt.$$

By analyzing the map $g : x \in \mathbb{R}^+ \longmapsto x - \frac{b_t^2}{2\beta x}$, we know that $g'(x) = 1 - \frac{b_t^2}{2\beta x^2}$, implying that g is nondecreasing when $x^2 \geq \frac{b_t^2}{2\beta}$.

Let it now be assumed that b is a deterministic positive continuous and nondecreasing function of time t such that

$$\frac{b_0^2}{2\beta} \leq \underline{a}^2 \leq \overline{a}^2 \leq \frac{b_T^2}{2\beta}.$$

Let $\underline{t}$ be such that $\frac{b_{\underline{t}}^2}{2\beta} = \underline{a}$ and $\overline{t}$ be such that $\frac{b_{\overline{t}}^2}{2\beta} = \overline{a}$, and define

$$a_t^* := \underline{a}\mathbf{1}_{0\leq t<\underline{t}} + \frac{b_t}{\sqrt{2\beta}}\mathbf{1}_{\underline{t}\leq t<\overline{t}} + \overline{a}\mathbf{1}_{\overline{t}\leq t\leq T},\ 0 \leq t \leq T,$$

then as in Example 7.1, we can show that $\mathbb{P}^{a^*}$ is an optimal probability measure, which is not of Bang–Bang type.

7.3. Example 3: Merton's Approach for Robust Power Utility

Here, we deal with robust power utility. As in Example 7.1, we suppose that b is a constant in $\mathbb{R}$ and $\xi = 0$. First, we consider the case where $A = \mathbb{R}$. From Theorem 5.2, $\widehat{F}_t(z)$ can be rewritten as

$$\widehat{F}_t(z) = \frac{\gamma\left|-\widehat{a}_t^{1/2}z + b\widehat{a}_t^{-1/2}\right|^2}{2(1+\gamma)} + \frac{1}{2}\left|\widehat{a}_t^{1/2}z\right|^2,$$

which is quadratic and linear in z.

Then the corresponding BSDEs can be solved explicitly under each probability measure $\mathbb{P}$. We use an exponential transformation and let

$$\alpha := 1 + \frac{\gamma}{1+\gamma},\ y'\mathbb{P} := e^{-\alpha y^{\mathbb{P}}},\ z'\mathbb{P} := e^{-\alpha y^{\mathbb{P}}}z^{\mathbb{P}}.$$

By applying Itô's formula, we know that $(y'\mathbb{P}, z'\mathbb{P})$ is the solution of the following linear BSDE:

$$dy'\mathbb{P}_t = -\alpha y'\mathbb{P}_t\left[\frac{\gamma}{2(1+\gamma)}\left(b^2\widehat{a}_t^{-1} - 2bz_t^{\mathbb{P}}\right)dt + z'\mathbb{P}_t dB_t\right],$$

with the terminal condition $y'\mathbb{P}_T = 1$.

For $t \in [0, T]$, let

$$\lambda_t := \frac{\alpha\gamma}{2(1+\gamma)} b^2 \widehat{a}_t^{-1}, \ \eta_t := -\frac{\gamma}{2(1+\gamma)} 2b\widehat{a}_t^{-1/2}, \text{ and } M_t := e^{\int_0^t \lambda_s - \frac{\eta_s^2}{2} ds + \int_0^t \widehat{a}_s^{-1/2} \eta_s dB_s}.$$

By applying Itô's formula to $y'\mathbb{P}_t M_t$, we obtain

$$y'\mathbb{P}_t = \mathbb{E}_t^{\mathbb{P}}[M_T / M_t], \text{ so } y_0^{\mathbb{P}} = -\frac{1}{\alpha} \ln\left(\mathbb{E}^{\mathbb{P}}[M_T]\right).$$

Since $\underline{a} \leq \widehat{a} \leq \overline{a}$, we derive that

$$y_0^{\mathbb{P}} \leq -\frac{\gamma}{2(1+\gamma)} \frac{b^2}{\overline{a}} T.$$

Thus by the representation of Y, we have

$$Y_0 \leq -\frac{\gamma}{2(1+\gamma)} \frac{b^2}{\overline{a}} T.$$

Moreover, under the specific probability measure $\mathbb{P}^{\overline{a}} \in \mathcal{P}_H$, we have

$$y_0^{\mathbb{P}^{\overline{a}}} = -\frac{\gamma}{2(1+\gamma)} \frac{b^2}{\overline{a}} T.$$

It implies that $Y_0 = y_0^{\mathbb{P}^{\overline{a}}}$. Thus, the value of the robust power utility maximization problem is

$$V(x) = -\frac{1}{\gamma} x^{-\gamma} \exp(Y_0).$$

As in Example 7.1, the robust utility maximization problem degenerates, and becomes a standard utility maximization problem under the probability measure $\mathbb{P}^{\overline{a}}$. In order to shed more light on this somehow surprising result, we first recall the HJB equation obtained by Merton (1969) in the standard utility maximization problem

$$-\frac{\partial v}{\partial t} - \sup_{\delta \in A} [\mathcal{L}^{\delta,\alpha} v(t,x)] = 0$$

together with the terminal condition

$$v(T,x) = U(x) := -\frac{x^{-\gamma}}{\gamma}, \ x \in \mathbb{R}_+, \ \gamma > 0,$$

where

$$\mathcal{L}^{\delta,\alpha} v(t,x) = x\delta b \frac{\partial v}{\partial x} + \frac{1}{2} x^2 \delta^2 \alpha \frac{\partial^2 v}{\partial x^2},$$

with a constant volatility $\alpha^{1/2}$.

It turns out that, when $A = \mathbb{R}$, the value function is given by

$$v(t,x) = \exp\left(\frac{b^2}{2\alpha} \frac{-\gamma}{(1+\gamma)} (T-t)\right) U(x), \ (t,x) \in [0,T] \times \mathbb{R}_+.$$

Let $\alpha = \overline{a}$, we have $v(0, x) = V(x)$, which is the result given by our 2BSDE method. Intuitively and formally speaking (in the case of controls taking values in compact sets, it has actually been proved under other technical conditions in Talay and Zheng 2002 that the solution to the stochastic game we consider is indeed a viscosity solution of the equation below, see also Remark 7.2), the HJB equation for the robust maximization problem should then be

$$-\frac{\partial v}{\partial t} - \sup_{\delta \in A} \ \alpha \in [\underline{a}, \overline{a}] \inf [\mathcal{L}^{\delta,\alpha} v(t, x)] = 0$$

together with the terminal condition $v(T, x) = U(x),\ x \in \mathbb{R}_+$.

Note that the value function obtained from our 2BSDE approach solves the above PDE, confirming the intuition that it is the correct PDE to consider in this context. Now assume that $A = \mathbb{R}$. If the second derivative of v is positive, then the term

$$\sup_{\delta \in A} \inf_{\alpha \in [\underline{a}, \overline{a}]} [\mathcal{L}^{\delta,\alpha} v(t, x)]$$

becomes infinite, so the above PDE has no meaning. It implies that v should be concave. Then $\overline{a}$ is the minimizer. It explains why the robust utility maximization problem degenerates in the case $A = \mathbb{R}$. From a financial point of view, this is the same type of result as in the problem of superreplication of an option with convex payoff under volatility uncertainty. Then, similarly to the so-called robustness of the Black–Scholes formula, this leads to the fact that the probability measure with the highest volatility corresponds to the worst case for the investor. However, it is clear that when, for instance, we impose no short-sale and no large sales constraints (that is to say A is a segment), the problem should not degenerate and the optimal probability measure switches between the two bounds $\underline{a}$ and $\overline{a}$.

Finally, notice that using the language of G-expectation introduced by Peng (2010), if we let

$$G(\Gamma) = \frac{1}{2} \sup_{\underline{a} \le \alpha \le \overline{a}} \alpha \Gamma = \frac{1}{2} \left(\overline{a} \, (\Gamma)^+ - \underline{a} \, (\Gamma)^- \right),$$

then the above PDE can be rewritten as follows:

$$-\frac{\partial v}{\partial t} + \inf_{\delta \in A} [\mathcal{L}^{\delta,\underline{a},\overline{a}} v(t, x)] = 0, \tag{7.1}$$

where

$$\mathcal{L}^{\delta,\underline{a},\overline{a}} v(t, x) = x^2 \delta^2 G \left(-\frac{\partial^2 v}{\partial x^2} \right).$$

Then, our PDE plays the same role for Merton's PDE as the Black–Scholes–Barenblatt PDE plays for the usual Black–Scholes PDE, by replacing the second-order derivative terms by their nonlinear versions.

REMARK 7.1. It could be interesting to consider more general constraints for the volatility process. For instance, we may hope to consider cases where $\underline{a}$ can become 0 and $\overline{a}$ can become $+\infty$. From the point of view of existence and uniqueness of the 2BSDEs with quadratic growth considered here, all the results still hold, since there is no uniform bound on $\widehat{a}$ for the set of probability measures considered in Possamaï and Zhou (2012, see definition 2.2). However, the boundedness assumption is crucial to retain the BMO integrability of the optimal strategy and thus also crucial for our proofs. We think that without it, the problem could still be solved but by now using the dynamic programming and PDE approach that we mentioned. However, delicate problems would arise in the sense that on the one hand, if $\underline{a} = 0$, then the PDE will become degenerate and one should then have to consider solutions in the viscosity sense, and on the other hand, if $\overline{a} = +\infty$, the PDE will have to be understood in the sense of boundary layers.

Another possible generalization would be to consider time-dependent or stochastic uncertainty sets for the volatility. It would be possible if we were able to weaken Assumption 2.3(i), which was already crucial in the proofs of existence and uniqueness in Soner et al. (2012). One first step in this direction has been taken by Nutz (2013) where he defines a notion of G-expectation (which roughly corresponds to a 2BSDE with a generator equal to 0) with a stochastic domain of volatility uncertainty.

REMARK 7.2. In Tevzadze et al. (2012), a similar problem of robust utility maximization is considered. They consider a financial market consisting of a riskless asset, a risky asset with unknown drift and volatility, and a nontradable asset with known coefficients. Their aim is to solve the robust utility maximization problem without terminal liability and without constraints for exponential and power utilities, by means of the dynamic programming approach already used in Talay and Zheng (2002). They managed to show that the value function of their problem solves a PDE similar to (7.1), and also that (see proposition 2.2) the optimal probability measure was of Bang–Bang type, thus confirming our intuition in their particular framework. Besides, they give some semi-explicit characterization of the optimal strategies and of the optimal probability measures. From a technical point of view, the main difference between our two approaches, beyond the methodology used, is that their set of generalized controls (that is to say their set of probability measures) is compact for the weak topology, because it corresponds to the larger set $\overline{\mathcal{P}}_W$ defined in Section 2. It is also the framework adopted in Denis and Kervarec (2007). However, as shown in Denis and Martini (2006) for instance, our smaller set $\mathcal{P}_H$ is only relatively compact for the weak topology. Nonetheless, working with this smaller set has no effect from the point of view of applications, and more importantly makes it possible to obtain results which are not attainable by their PDE methods, for instance with non-Markovian terminal liability ξ and also when the set of trading strategies is constrained in an arbitrary closed set.

REFERENCES

ANDERSON, E., L. P. SARGENT, and T. HANSEN (2003): A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection, *J. Eur. Econ. Assoc.* 1, 68–123.

AVELLANEDA, M., A. LEVY, and A. PARAS (1995): Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities, *Appl. Math. Finance* 2, 73–88.

BARRIEU, P., and N. EL KAROUI (2011): Monotone Stability of Quadratic Semimartingales with Applications to Unbounded General Quadratic BSDEs, Preprint arXiv:1101.5282.

BISMUT, J. M. (1973): Conjugate Convex Functions in Optimal Stochastic Control, *J. Math. Anal. Appl.* 44, 384–404.

BORDIGONI, G., A. MATOUSSI, and M. SCHWEIZER (2007): A Stochastic Control Approach to a Robust Utility Maximization Problem, in *Stochastic Analysis and Applications, Abel Symposium* 2, pp. 125–151.

BRIAND, P., and Y. HU (2006): BSDE with Quadratic Growth and Unbounded Terminal Value, *Probab. Theory Relat. Fields* 136, 604–618.

CHERIDITO, P., H. M. SONER, N. TOUZI, and N. VICTOIR (2007): Second Order Backward Stochastic Differential Equations and Fully Non-Linear Parabolic PDEs, *Commun. Pure Appl. Math.* 60(7), 1081–1110.

CVITANIĆ, J., and I. KARATZAS (1992): Convex Duality in Constrained Portfolio Optimization, *Ann. Appl. Probab*. 2, 767–818.

DENIS, L., and M. KERVAREC (2007): Utility Functions and Optimal Investment in Non-Dominated Models, Preprint, hal:00371215.

DENIS, L., and C. MARTINI (2006): A Theoretical Framework for the Pricing of Contingent Claims in the Presence of Model Uncertainty, *Ann. Appl. Probab*. 16(2), 827–852.

EL KAROUI, N. (1981): Les aspects probabilistes du contrôle stochastique, in *Ecole d'Eté de Probabilités de Saint-Flour IX-1979*, Vol. 876 of Lecture Notes in Mathematics, Berlin: Springer, pp. 73–238.

EL KAROUI, N., S. PENG, and M. C. QUENEZ (1994): Backward Stochastic Differential Equations in Finance, *Math. Finance* 7(1), 1–71.

EL KAROUI, N., and R. ROUGE (2000): Pricing via Utility Maximization and Entropy, *Math. Finance* 10, 259–276.

EPSTEIN, L. G., and S. JI (2011): Ambiguous Volatility, Possibility and Utility in Continuous Time, Preprint arxiv:1103.1652.

FREI, C., M. MOCHA, and N. WESTRAY (2012): BSDEs in Utility Maximization with BMO Market Price of Risk, *Stoch. Process. Appl.* 122, 2486–2519.

GILBOA, I., and D. SCHMEIDLER (1989): Maximin Expected Utility with a Non-Unique Prior, *J. Math. Econ*. 18, 141–153.

GUNDEL, A. (2005): Robust Utility Maximization for Complete and Incomplete Market Models, *Finance Stoch*. 9, 151–176.

HANSEN, L. P., T. J. SARGENT, G. A. TURMUHAMBETOVA, and N. WILLIAMS (2006): Robust Control and Model Misspecification, *J. Econ. Theory* 128, 45–90.

HU, Y., P. IMKELLER, and M. MÜLLER (2005): Utility Maximization in Incomplete Markets, *Ann. Appl. Probab*. 15(3), 1691–1712.

LYONS, F. (1995): Uncertain Volatility and the Risk-Free Synthesis of Derivatives, *J. Appl. Finance* 2:117–133.

MERTON, R. (1969): Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case, *Rev. Econ. Stat*. 51, 239–265.

NUTZ, M. (2013): Random G-Expectations, *Ann. Appl. Probab.* 23, 1755–1777.

PARDOUX, E., and S. PENG (1990): Adapted Solution of a Backward Stochastic Differential Equation, *Syst. Control Lett.*, 14, 55–61.

PENG, S. (2010): Nonlinear Expectations and Stochastic Calculus under Uncertainty, Preprint, arXiv:1002.4546.

POSSAMAÏ, D. (2010): Second Order Backward Stochastic Differential Equations with Continuous Coefficient, Preprint, arXiv:1201.1049.

POSSAMAÏ, D., and C. ZHOU (2012): Second Order Backward Stochastic Differential Equations with Quadratic Growth, Preprint arXiv:1201.1050.

PLISKA, S. R. (1986): A Stochastic Calculus Model of Continuous Trading: Optimal Portfolios, *Math. Oper. Res.* 11, 371–382.

SCHIED, A., and C.-T. WU (2005): Duality Theory for Optimal Investments under Model Uncertainty, *Stat. Decis*. 23:199–217.

SKIADAS, C. (2003): Robust Control and Recursive Utility, *Finance Stoch.* 7, 475–489.

SONER, H. M., N. TOUZI, and J. ZHANG (2011): Quasi-Sure Stochastic Analysis through Aggregation, *Elect. J. Probab*. 16, 1844–1879.

SONER, H. M., N. TOUZI, and J. ZHANG (2012): Wellposedness of Second Order BSDE's, *Probab. Theory Relat. Fields* 153(1–2), 149–190.

TALAY, D. and Z. ZHENG (2002): Worst Case Model Risk Management, *Finance Stoch.* 6, 517–537.

TEVZADZE, R. (2008): Solvability of Backward Stochastic Differential Equations with Quadratic Growth, *Stoch. Process. Appl.* 118, 503–515.

TEVZADZE, R., T. TORONJADZE, and T. UZUNASHVILI (2012): Robust Utility Maximization for Diffusion Market Model with Misspecified Coefficients, Forthcoming in *Finance Stoch*.

VON NEUMANN, J., and O. MORGENSTERN (1944): Theory of Games and Economic Behavior, Princeton, New Jersey: Princeton University Press.

ZARIPHOPOULOU, T. (1994): Consumption-Investment Models with Constraints, *SIAM J. Control Optim*. 32(1), 59–85.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 288–310

AN ONLINE PORTFOLIO SELECTION ALGORITHM WITH REGRET LOGARITHMIC IN PRICE VARIATION

ELAD HAZAN

Technion—Israel Institute of Technology

SATYEN KALE

IBM T.J. Watson Research Center

We present a novel efficient algorithm for portfolio selection which theoretically attains two desirable properties:

1. Worst-case guarantee: the algorithm is universal in the sense that it asymptotically performs almost as well as the best constant rebalanced portfolio determined in hindsight from the realized market prices. Furthermore, it attains the tightest known bounds on the regret, or the log-wealth difference relative to the best constant rebalanced portfolio. We prove that the regret of the algorithm is bounded by $O(\log Q)$, where Q is the quadratic variation of the stock prices. This is the first improvement upon Cover's (1991) seminal work that attains a regret bound of $O(\log T)$, where T is the number of trading iterations.
2. Average-case guarantee: in the Geometric Brownian Motion (GBM) model of stock prices, our algorithm attains tighter regret bounds, which are provably impossible in the worst-case. Hence, when the GBM model is a good approximation of the behavior of market, the new algorithm has an advantage over previous ones, albeit retaining worst-case guarantees.

We derive this algorithm as a special case of a novel and more general method for online convex optimization with exp-concave loss functions.[1]

KEY WORDS: regret minimization, portfolio selection, online convex optimization.

1. INTRODUCTION

A widely used model in mathematical finance for stock prices is the Geometric Brownian Motion (GBM). This model has been successfully applied to study and price financial instruments, and is the basis of much investing in practice. However, while it is often an acceptable approximation, the GBM model is not always valid empirically. This motivates a worst-case approach to investing, called universal portfolio management, where the objective is to maximize wealth relative to the wealth earned by the best fixed portfolio in hindsight. In this paper, we tie the two approaches, and design an investment

Elad Hazan is supported by ISF grant 810/11. Part of this work was done when Satyen Kale was at Microsoft Research and Yahoo! Research.

Manuscript received June 2011; final revision received April 2012.

Address correspondence to Elad Hazan, Technion—Israel Institute of Technology, Haifa 32000, Israel; e-mail: ehazan@ie.technion.ac.il.

[1] An extended abstract of this work appeared in Hazan and Kale (2009).

DOI: 10.1111/mafi.12006

strategy which is universal in the worst case, and yet capable of significantly improved performance when the market is closely approximated by the GBM model.

"Average-case" Investing. Much of mathematical finance theory is devoted to the modeling of stock prices and devising investment strategies that maximize wealth gain (while controlling risk). Typically, such investment strategies involve estimating fitting a parametric model of stock prices to observed data. Such strategies are geared to the *average-case market* (in the formal computer science sense), and are naturally susceptible to drastic deviations from the model, as witnessed in the recent stock market crash.

Even so, empirically the GBM (Osborne 1959; Bachelier 1900) has enjoyed great predictive success and every year significant investments are made assuming this model. One illustration of its wide acceptability is that Black and Scholes (1973) used this same model in their Nobel prize-winning work on pricing options on stocks.

"Worst-case" Investing. The fragility of average-case models in the face of rare but dramatic deviations led Cover (1991) to take a worst-case approach to investing in stocks. The performance of an online investment algorithm for *arbitrary* sequences of stock price returns is measured with respect to the best constant rebalanced portfolio (CRP; see Cover 1991) in hindsight. A universal portfolio selection algorithm is one that obtains sublinear (in the number of trading periods T) *regret*, which is the difference in the logarithms of the final wealths obtained by the two.

Cover (1991) gave the first universal portfolio selection algorithm with regret bounded by $O(\log T)$. There has been much follow-up work after Cover's seminal work, such as Helmbold et al. (1998), Merhav and Feder (1993), Kalai and Vempala (2003), Blum and Kalai (1900), and Hazan, Agarwal, and Kale (2007), which focused on either obtaining alternate universal algorithms or improving the efficiency of Cover's algorithm. However, the best regret bound is still $O(\log T)$.

This dependence of the regret on the number of trading periods is not entirely satisfactory for two main reasons. First, *a priori* it is not clear why the online algorithm should have high regret (growing with the number of iterations) in an unchanging environment. As an extreme example, consider a setting with two stocks where one has an "upward drift" of 1% daily, whereas the second stock remains at the same price. One would expect to "figure out" this pattern quickly and focus on the first stock, thus attaining a constant fraction of the wealth of the best CRP in the long run, i.e., constant regret, unlike the worst-case bound of $O(\log T)$.

The second problem arises from trading frequency. Suppose we need to invest over a fixed period of time, say a year. Trading more frequently potentially leads to higher wealth gain, by capitalizing on short-term stock movements. However, increasing trading frequency increases T, and thus one may expect more regret. The problem is actually even worse: since we measure regret as a difference of logarithms of the final wealths, a regret bound of $O(\log T)$ implies a polynomial in T factor ratio between the final wealths. In reality, however, experiments by Agarwal et al. (2006) show that some known online algorithms actually *improve* with increasing trading frequency.

Bridging Worst-case and Average-case Investing. Both these issues are resolved if one can show that the regret of a "good" online algorithm depends on total *variation* in the sequence of stock returns, rather than purely on the number of iterations. If the stock return sequence has low variation, we expect our algorithm to be able to perform better. If we trade more frequently, then the per iteration variation should go down correspondingly, so the total variation stays the same.

We analyze a portfolio selection algorithm and prove that its regret is bounded by $O(\log Q)$, where Q (formally defined in Section 1.2) is the sum of squared deviations of the returns from their mean. Since $Q \leq T$ (after appropriate normalization), we improve over previous regret bounds and retain the worst-case robustness. Furthermore, in an average-case model such as GBM, the variation can be tied very nicely to the volatility parameter, which explains the experimental observation the regret does not increase with increasing trading frequency. Our algorithm is efficient, and its implementation requires constant time per iteration (independent of the number of game iterations).

1.1. New Techniques and Comparison to Related Work

Cesa-Bianchi, Mansour, and Stoltz (2007) initiated work on relating worst-case regret to the variation in the data for the related learning problem of prediction from expert advice, and conjectured that the optimal regret bounds should depend on the observed variation of the cost sequence. Recently, this conjecture was proved and regret bounds of $\tilde{O}(\sqrt{Q})$ were obtained in the full information and bandit linear optimization settings (Hazan and Kale 2010, 2011), where Q is the variation in the cost sequence. In this paper, we give an exponential improvement in regret, viz. $O(\log Q)$, for the case of online *exp-concave* optimization, which includes portfolio selection as a special case.

Another approach to connecting worst-case to average-case investing was taken by Jamshidian (1992) and Cross and Barron (2003). They considered a model of "continuous trading," where there are T "trading intervals," and in each the online investor chooses a *fixed* portfolio which is rebalanced k times with $k \to \infty$. They prove familiar regret bounds of $O(\log T)$ (independent of k) in this model w.r.t. the best fixed portfolio which is rebalanced $T \times k$ times. In this model our algorithm attains the tighter regret bounds of $O(\log Q)$, although our algorithm has more flexibility. Furthermore their algorithms, being extensions of Cover's algorithm, may require exponential time in general.[2]

Our bounds of $O(\log Q)$ regret require completely different techniques compared to the $\tilde{O}(\sqrt{Q})$ regret bounds of Hazan and Kale (2010, 2011). These previous bounds are based on first-order gradient descent methods which are too weak to obtain $O(\log Q)$ regret. Instead we have to use the second-order Newton step ideas based on Hazan et al. (2007) (in particular, the Hessian of the cost functions).

The second-order techniques of Hazan et al. (2007) are, however, not sensitive enough to obtain $O(\log Q)$ bounds. This is because progress was measured in terms of the distance between successive portfolios in the usual Euclidean norm, which is insensitive to variation in the cost sequence. In this paper, we introduce a different analysis technique, based on analyzing the distance between successive predictions using norms that keep changing from iteration to iteration and are actually sensitive to the variation.

A key technical step in the analysis is a lemma (Lemma 2.3) which bounds the sum of differences of successive Cesaro means of a sequence of vectors by the logarithm of its variation. This lemma, which may be useful in other contexts when variation bounds on the regret are desired, is proved using the Kahn–Karush–Tucker conditions, and also improves the regret bounds in previous papers.

[2] Cross and Barron give an efficient implementation for some interesting special cases, under assumptions on the variation in returns and bounds on the magnitude of the returns, and assuming $k \to \infty$. A truly efficient implementation of their algorithm can probably be obtained using the techniques of Kalai and Vempala.

1.2. The Model and Statement of Results

Portfolio Management. In the universal portfolio management model Cover (1991), an online investor iteratively distributes her wealth over n assets before observing the change in asset price. In each iteration $t = 1, 2, \ldots$ the investor commits to an n-dimensional distribution of her wealth, $x_t \in \Delta_n = \{\sum_i x_i = 1, x \geq 0\}$. She then observes a *price relatives vector* $r_t \in \mathbb{R}^n_+$, where $r_t(i)$ is the ratio between the closing price of the ith asset on trading period t and the opening price. In the tth trading period, the wealth of the investor changes by a factor of $(r_t \cdot x_t)$. The overall change in wealth is thus $\prod_t (r_t \cdot x_t)$. Since in a typical market wealth grows at an exponential rate, we measure performance by the exponential growth rate, which is $\log \prod_t (r_t \cdot x_t) = \sum_t \log(r_t \cdot x_t)$. A *constant rebalanced portfolio* (CRP) is an investment strategy which rebalances the wealth in every iteration to keep a fixed distribution. Thus, for a CRP $x \in \Delta_n$, the change in wealth is $\prod_t (r_t \cdot x)$.

The *regret* of the investor is defined to be the difference between the exponential growth rate of her investment strategy and that of the best CRP strategy in hindsight, i.e.,

$$\text{Regret} := \max_{x^* \in \Delta_n} \sum_t \log(r_t \cdot x^*) - \sum_t \log(r_t \cdot x_t).$$

Note that the regret does not change if we scale all the returns in any particular period by the same amount. So we assume w.l.o.g. that in all periods t, $\max_i r_t(i) = 1$. We assume that there is known parameter $r > 0$, such that for all periods t, $\min_{t,i} r_t(i) \geq r$. We call r the *market variability* parameter. This is the only restriction we put on the stock price returns; they could be chosen adversarially as long as they respect the market variability bound.

Online Convex Optimization. In the online convex optimization problem Zinkevich (2003), which generalizes universal portfolio management, the decision space is a closed, bounded, convex set $K \in \mathbb{R}^n$, and we are sequentially given a series of convex cost[3] functions $f_t : K \to \mathbb{R}$ for $t = 1, 2, \ldots$. The algorithm iteratively produces a point $x_t \in K$ in every round t, without knowledge of f_t (but using the past sequence of cost functions), and incurs the cost $f_t(x_t)$. The regret at time T is defined to be

$$\text{Regret} := \sum_{t=1}^{T} f_t(x_t) - \min_{x \in K} \sum_{t=1}^{T} f_t(x).$$

In this paper, we restrict our attention to convex cost functions which can be written as $f_t(x) = g(v_t \cdot x)$ for some univariate convex function g and a parameter vector $v_t \in \mathbb{R}^n$ (for example, in the portfolio management problem, $K = \Delta_n, f_t(x) = -\log(r_t \cdot x)$, $g = -\log$, and $v_t = r_t$).

Thus, the cost functions are *parameterized* by the vectors $v_1, v_2, \ldots, v_T$. Our bounds will be expressed as a function of the *quadratic variability* of the parameter vectors v_1, $v_2, \ldots, v_T$, defined as

$$Q(v_1, \ldots, v_T) := \min_{\mu} \sum_{t=1}^{T} \|v_t - \mu\|^2.$$

[3] Note the difference from the portfolio selection problem: here we have convex *cost* functions, rather than concave payoff functions. The portfolio selection problem is obtained by using $-\log$ as the cost function.

This expression is minimized at $\mu = \frac{1}{T}\sum_{t=1}^{T} v_t$, and thus the quadratic variation is just $T-1$ times the sample variance of the sequence of vectors $\{v_1, \dots, v_t\}$. Note however that the sequence can be generated adversarially rather than by some stochastic process. We shall refer to this as simply Q if the vectors are clear from the context.

Main Theorem. In the setup of the online convex optimization problem above, we have the following algorithmic result:

THEOREM 1.1. *Let the cost functions be of the form* $f_t(x) = g(v_t \cdot x)$. *Assume that there are parameters* $R, D, a, b > 0$ *such that the following conditions hold:*

1. *for all* t, $\|v_t\| \le R$,
2. *for all* $x \in K$, *we have* $\|x\| \le D$,
3. *for all* $x \in K$, *and for all* t, *either* $g'(v_t \cdot x) \in [0, a]$ or $g'(v_t \cdot x) \in [-a, 0]$, *and*
4. *for all* $x \in K$, *and for all* t, $g''(v_t \cdot x) \ge b$.

Then there is an algorithm that guarantees the following regret bound:

$$\text{Regret} = O((a^2 n/b)\log(1 + bQ + bR^2) + aRD\log(1 + Q/R^2) + D^2).$$

Now we apply Theorem 1.1 to the portfolio selection problem. First, we estimate the relevant parameters. We have $\|r_t\| \le \sqrt{n}$ since all $r_t(i) \le 1$, thus $R = \sqrt{n}$. For any $x \in \Delta_n$, $\|x\| \le 1$, so $D = 1$. $g'(v_t \cdot x) = -\frac{1}{(v_t \cdot x)}$, and thus $g'(v_t \cdot x) \in [-\frac{1}{r}, 0]$, so $a = \frac{1}{r}$. Finally, $g''(v_t \cdot x) = \frac{1}{(v_t \cdot x)^2} \ge 1$, so $b = 1$. Applying Theorem 1.1 we get the following corollary:

COROLLARY 1.2. *For the portfolio selection problem over n assets, there is an algorithm that attains the following regret bound:*

$$\text{Regret} = O\left(\frac{n}{r^2}\log(Q + n)\right).$$

2. BOUNDING THE REGRET BY THE OBSERVED VARIATION IN RETURNS

2.1. Preliminaries

All matrices are assumed to be real symmetric matrices in $\mathbb{R}^{n \times n}$, where n is the number of stocks. We use the notation $A \succeq B$ to say that $A - B$ is positive semidefinite. We require the notion of a norm of a vector x induced by a positive definite matrix M, defined as $\|x\|_M = \sqrt{x^\top Mx}$. The following simple generalization of the Cauchy–Schwartz inequality is used in the analysis:

$$\forall x, y \in \mathbb{R}^n : x \cdot y \le \|x\|_M \|y\|_{M^{-1}}.$$

We denote by $|A|$ the determinant of a matrix A, and by $A \bullet B = \mathbf{Tr}(AB) = \sum_{ij} A_{ij} B_{ij}$. As we are concerned with logarithmic regret bounds, potential functions which behave like harmonic series come into play. A generalization of harmonic series to high dimensions is the *vector-harmonic series*, which is a series of quadratic forms that can be expressed as (here $A \succ 0$ is a positive definite matrix, and $v_1, v_2, \ldots$ are vectors in $\mathbb{R}^n$):

$$v_1^\top \left(A + v_1 v_1^\top\right)^{-1} v_1,\ v_2^\top \left(A + v_1 v_1^\top + v_2 v_2^\top\right)^{-1} v_2, \ldots, v_t^\top \left(A + \sum_{\tau=1}^{t} v_\tau v_\tau^\top\right)^{-1} v_t, \ldots$$

The following lemma is from Hazan et al. (2007) (and proven in Appendix A for completeness):

LEMMA 2.1. *For a vector harmonic series given by an initial matrix A and vectors v_1, $v_2, \ldots, v_T$, we have*

$$\sum_{t=1}^{T} v_t^\top \left(A + \sum_{\tau=1}^{t} v_\tau v_\tau^\top\right)^{-1} v_t \le \log \left[\frac{\left| A + \sum_{\tau=1}^{T} v_\tau v_\tau^\top \right|}{|A|} \right].$$

The reader can note that in one dimension, if all vectors $v_t = 1$ and $A = 1$, then the series above reduces exactly to the regular harmonic series whose sum is bounded, of course, by $\log(T+1)$.

Henceforth, we will denote by $\sum_t$ the summation $\sum_{t=1}^{T}$.

2.2. Algorithm and Analysis

We analyze the following algorithm and prove that it attains logarithmic regret with respect to the observed variation (rather than number of iterations). The algorithm follows the generic algorithmic scheme of "Follow-The-Regularized-Leader" (FTRL) with squared Euclidean regularization.

Algorithm Exp-Concave-FTL. In iteration t, use the point x_t defined as:

$$x_t \triangleq \arg\min_{x \in \Delta_n} \left(\sum_{\tau=1}^{t-1} f_\tau(x) + \frac{1}{2}\|x\|^2 \right). \tag{2.1}$$

Note the mathematical program which the algorithm solves is convex, and can be solved in time polynomial in the dimension and number of iterations. The running time, however, for solving this convex program can be quite high. In Section 4, for the specific problem of portfolio selection, where $f_t(x) = -\log(r_t \cdot x)$, we give a faster implementation whose per iteration running time is independent of the number of iterations.

We now proceed to prove Theorem 1.1.

Proof [Theorem 1.1]. First, we note that the algorithm is running a "Follow-the-leader" procedure on the cost functions $f_0, f_1, f_2, \ldots$ where $f_0(x) = \frac{1}{2}\|x\|^2$ is a fictitious period 0 cost function. In other words, in each iteration, it chooses the point that would have minimized the total cost under all the observed functions so far (and, additionally, a fictitious initial cost function f_0). This point is referred to as the leader in that round.

The first step in analyzing such an algorithm is to use a stability lemma from Kalai and Vempala (2005), which bounds the regret of any Follow-the-leader algorithm by the difference in costs (under f_t) of the current prediction x_t and the next one x_{t+1}, plus an additional error term which comes from the regularization. Thus, we have (recall the notation $\sum_t \equiv \sum_{t=1}^{T}$)

$$\begin{aligned}
\text{Regret} &\leq \sum_t f_t(x_t) - f_t(x_{t+1}) + \frac{1}{2}(\|x^*\|^2 - \|x_0\|^2) \\
&\leq \sum_t \nabla f_t(x_t) \cdot (x_t - x_{t+1}) + \frac{1}{2} D^2 \\
&= \sum_t g'(v_t \cdot x_t)[v_t \cdot (x_t - x_{t+1})] + \frac{1}{2} D^2
\end{aligned} \tag{2.2}$$

The second inequality is because f_t is convex. The last equality follows because $\nabla f_t(x_t) = g'(x_t \cdot v_t)v_t$. Now, we need a handle on $x_t - x_{t+1}$. For this, define $F_t = \sum_{\tau=0}^{t-1} f_\tau$, and note that x_t minimizes f_t over K. Consider the difference in the gradients of F_{t+1} evaluated at x_{t+1} and x_t:

$$\begin{aligned}
\nabla F_{t+1}(x_{t+1}) - \nabla F_{t+1}(x_t) &= \sum_{\tau=0}^{t} \nabla f_\tau(x_{t+1}) - \nabla f_\tau(x_t) && (2.3)\\
&= \sum_{\tau=1}^{t} \left[g'(v_\tau \cdot x_{t+1}) - g'(v_\tau \cdot x_t)\right] v_\tau + (x_{t+1} - x_t) \\
&= \sum_{\tau=1}^{t} \left[\nabla g'\left(v_\tau \cdot \zeta_\tau^t\right) \cdot (x_{t+1} - x_t)\right] v_\tau + (x_{t+1} - x_t) \\
&= \sum_{\tau=1}^{t} g''\left(v_\tau \cdot \zeta_\tau^t\right) v_\tau v_\tau^\top (x_{t+1} - x_t) + (x_{t+1} - x_t). && (2.4)
\end{aligned}$$

Equation (2.3) follows by applying the Taylor expansion of the (multi-variate) function $g'(v_\tau \cdot x)$ at point x_t, for some point ζ_τ^t on the line segment joining x_t and x_{t+1}. The equation (2.4) follows from the observation that $\nabla g'(v_\tau \cdot x) = g''(v_\tau \cdot x)v_\tau$.

Define $A_t = \sum_{\tau=1}^{t} g''(v_\tau \cdot \zeta_\tau^t) v_\tau v_\tau^\top + I$, where I is the identity matrix, and $\Delta x_t = x_{t+1} - x_t$. Then equation (2.4) can be rewritten as:

$$\nabla F_{t+1}(x_{t+1}) - \nabla F_t(x_t) - g'(v_t \cdot x_t)v_t = A_t \Delta x_t. \tag{2.5}$$

Now, since x_t minimizes the convex function f_t over the convex set K, a standard inequality of convex optimization (see Boyd and Vandenberghe 2004) states that for any point $y \in K$, we have $\nabla f_t(x_t) \cdot (y - x_t) \geq 0$. Thus, for $y = x_{t+1}$, we get that $\nabla f_t(x_t) \cdot (x_{t+1} - x_t) \geq 0$. Similarly, we get that $\nabla F_{t+1}(x_{t+1}) \cdot (x_t - x_{t+1}) \geq 0$. Putting these two inequalities together, we get that

$$(\nabla F_{t+1}(x_{t+1}) - \nabla F_t(x_t)) \cdot \Delta x_t \leq 0. \tag{2.6}$$

Thus, using the expression for $A_t \Delta x_t$ from (2.5) we have

$$\begin{aligned}
\|\Delta x_t\|_{A_t}^2 &= A_t \Delta x_t \cdot \Delta x_t \\
&= (\nabla F_{t+1}(x_{t+1}) - \nabla F_t(x_t) - g'(v_t \cdot x_t)v_t) \cdot \Delta x_t \\
&\leq g'(v_t \cdot x_t)[v_t \cdot (x_t - x_{t+1})]
\end{aligned} \tag{2.7}$$

(from (2.6)).

Assume that $g'(v_t \cdot x) \in [-a, 0]$ for all $x \in K$ and all t. The other case is handled similarly. Inequality (2.7) implies that $g'(v_t \cdot x_t)$ and $v_t \cdot (x_t - x_{t+1})$ have the same sign. Thus, we can upper bound

$$g'(v_t \cdot x_t)[v_t \cdot (x_t - x_{t+1})] \le a(v_t \cdot \Delta x_t). \tag{2.8}$$

Define $\tilde{v}_t = v_t - \mu_t$, $\mu_t = \frac{1}{t+1} \sum_{\tau=1}^{t} v_\tau$. Then, we have

$$\sum_t v_t \cdot \Delta x_t = \sum_t \tilde{v}_t \cdot \Delta x_t + \sum_{t=2}^{T} x_t(\mu_{t-1} - \mu_t) - x_1\mu_1 + x_{T+1}\mu_T, \tag{2.9}$$

Now, define $\rho = \rho(v_1, \dots, v_T) = \sum_{t=1}^{T-1} \|\mu_{t+1} - \mu_t\|$. Then we bound

$$\begin{aligned} &\sum_{t=2}^{T} x_t(\mu_{t-1} - \mu_t) - x_1\mu_1 + x_{T+1}\mu_T \\ &\quad \le \sum_{t=2}^{T} \|x_t\|\|\mu_{t-1} - \mu_t\| + \|x_1\|\|\mu_1\| + \|x_{T+1}\|\|\mu_T\| \\ &\quad \le D\rho + 2DR. \end{aligned} \tag{2.10}$$

We will bound ρ momentarily. For now, we turn to bounding the first term of (2.9) using the Cauchy–Schwartz generalization as follows:

$$\tilde{v}_t \cdot \Delta x_t \le \|\tilde{v}_t\|_{A_t^{-1}} \|\Delta x_t\|_{A_t}. \tag{2.11}$$

By the usual Cauchy–Schwartz inequality,

$$\begin{aligned} \sum_t \|\tilde{v}_t\|_{A_t^{-1}} \|\Delta x_t\|_{A_t} &\le \sqrt{\sum_t \|\tilde{v}_t\|_{A_t^{-1}}^2} \cdot \sqrt{\sum_t \|\Delta x_t\|_{A_t}^2} \\ &\le \sqrt{\sum_t \|\tilde{v}_t\|_{A_t^{-1}}^2} \cdot \sqrt{\sum_t a(v_t \cdot \Delta x_t)} \end{aligned}$$

from (2.7) and (2.8). We conclude, using (2.9), (2.10), and (2.11), that

$$\sum_t a(v_t \cdot \Delta x_t) \le a\sqrt{\sum_t \|\tilde{v}_t\|_{A_t^{-1}}^2} \cdot \sqrt{\sum_t a(v_t \cdot \Delta x_t)} + aD\rho + 2aDR.$$

This implies (using the AM-GM inequality applied to the first term on the RHS) that

$$\sum_t a(v_t \cdot \Delta x_t) \le a^2 \sum_t \|\tilde{v}_t\|_{A_t^{-1}}^2 + 2aD\rho + 4aDR.$$

Plugging this into the regret bound (2.2) we obtain, via (2.8),

$$\text{Regret} \le a^2 \sum_t \|\tilde{v}_t\|_{A_t^{-1}}^2 + 2aD\rho + 4aDR + \frac{1}{2}D^2.$$

The proof is completed by the following two lemmas (Lemmas 2.2 and 2.3) which bound the RHS. The first term is a vector harmonic series, and the second term can be bounded by a (regular) harmonic series. $\square$

LEMMA 2.2. $\sum_t \|\tilde v_t\|^2_{A_t^{-1}} \le \frac{5n}{b}\log\left[1+bQ+bR^2\right]$.

Proof. We have $A_t = \sum_{\tau=1}^t g''(v_\tau\cdot\zeta_\tau^t)v_\tau v_\tau^\top + I$. Since $g''(v_t\cdot\zeta_\tau^t)\ge b$, we have $A_t \succeq I + b\sum_{\tau=1}^t v_\tau v_\tau^\top$. Using the fact that $\tilde v_t = v_t - \mu_t$ and $\mu_t = \frac{1}{t+1}\sum_{\tau\le t} v_\tau$ we get that

$$\begin{aligned}\sum_{\tau=1}^t \tilde v_\tau \tilde v_\tau^\top &= \sum_{s=1}^t \left(v_s - \frac{1}{s+1}\sum_{\tau\le s} v_\tau\right)\left(v_s - \frac{1}{s+1}\sum_{\tau\le s} v_\tau\right)^\top \\ &= \sum_{s=1}^t\left(1+\sum_{\tau=s}^t \frac{1}{(\tau+1)^2} - \frac{2}{s+1}\right)v_s v_s^\top \\ &\quad + \sum_{s=1}^t\sum_{r<s}\left(-\frac{1}{s+1}+\sum_{\tau=s}^t\frac{1}{(\tau+1)^2}\right)\left[v_r v_s^\top + v_s v_r^\top\right].\end{aligned}$$

Since

$$\frac{1}{s+1}-\frac{1}{t+2} = \int_{s+1}^{t+2}\frac{1}{x^2}dx \le \sum_{\tau=s}^t \frac{1}{(\tau+1)^2} \le \int_s^{t+1}\frac{1}{x^2}dx = \frac{1}{s}-\frac{1}{t+1}, \tag{2.12}$$

we get that

$$\left|-\frac{1}{s+1}+\sum_{\tau=s}^t\frac{1}{(\tau+1)^2}\right| \le \frac{1}{s^2}+\frac{1}{t}. \tag{2.13}$$

Since $(v_r \pm v_s)(v_r\pm v_s)^\top \succeq 0$, we have $(v_r v_r^\top + v_s v_s^\top) \succeq \pm(v_r v_s^\top + v_s v_r^\top)$, and so

$$\begin{aligned}\left(-\frac{1}{s+1}+\sum_{\tau=s}^t\frac{1}{(\tau+1)^2}\right)\left[v_r v_s^\top + v_s v_r^\top\right] &\preceq \left|-\frac{1}{s+1}+\sum_{\tau=s}^t\frac{1}{(\tau+1)^2}\right|\left[v_r v_r^\top + v_s v_s^\top\right] \\ &\preceq \left(\frac{1}{s^2}+\frac{1}{t}\right)\left[v_r v_r^\top + v_s v_s^\top\right],\end{aligned}$$

by (2.13). Also by (2.12), we have $(1+\sum_{\tau=s}^t \frac{1}{(\tau+1)^2} - \frac{2}{s+1}) \le 1 + \frac{1}{s} - \frac{1}{t+1} - \frac{2}{s+1} \le 1$, so we have

$$\begin{aligned}\sum_{\tau=1}^t \tilde v_\tau\tilde v_\tau^\top &\preceq \sum_{s=1}^t v_s v_s^\top + \sum_{s=1}^t\sum_{r<s}\left(\frac{1}{s^2}+\frac{1}{t}\right)\left[v_r v_r^\top + v_s v_s^\top\right] \\ &\preceq 3\sum_{s=1}^t v_s v_s^\top + \sum_{s=1}^t\left(\frac{1}{s^2}+\frac{1}{t}\right)\sum_{r<s} v_r v_r^\top \\ &\preceq 3\sum_{s=1}^t v_s v_s^\top + \sum_{s=1}^t v_s v_s^\top \sum_{r=s+1}^t\left(\frac{1}{r^2}+\frac{1}{t}\right) \\ &\preceq 3\sum_{s=1}^t v_s v_s^\top + \sum_{s=1}^t\left(1+\frac{1}{s}\right)v_s v_s^\top \\ &\preceq 5\sum_{s=1}^t v_s v_s^\top.\end{aligned}$$

Let $\tilde{A}_t = I + b\sum_\tau \tilde{v}_\tau \tilde{v}_\tau^\top$. Note that the inequality above shows that $\tilde{A}_t \preceq 5A_t$. Thus, using Lemma 2.1, we get

$$(2.14)\qquad \sum_t \|\tilde{v}_t\|^2_{A_t^{-1}} = \sum_t \tilde{v}_t A_t^{-1} \tilde{v}_t \le \frac{5}{b} \sum_t [\sqrt{b}\tilde{v}_t]^\top \tilde{A}_t^{-1} [\sqrt{b}\tilde{v}_t] \le \frac{5}{b} \log\left[\frac{|\tilde{A}_T|}{|\tilde{A}_0|}\right].$$

To bound the latter quantity note that $|\tilde{A}_0| = |I| = 1$, and that

$$|\tilde{A}_T| = |I + b\sum_t \tilde{v}_t \tilde{v}_t^\top| \le \left(1 + b\sum_t \|\tilde{v}_t\|_2^2\right)^n = (1 + b\tilde{Q})^n$$

where $\tilde{Q} = \sum_t \|\tilde{v}_t\|^2 = \sum_t \|v_t - \mu_t\|^2$. Lemma 2.4 shows that $\tilde{Q} \le Q + R^2$. This implies that $|\tilde{A}_T| \le (1 + bQ + bR^2)^n$ and the proof is completed by substituting this bound into (2.14). □

LEMMA 2.3. $\rho(v_1, \ldots, v_T) \le 4R[\log(1 + Q/R^2) + 2]$.

Proof. Define, for $\tau = 0, 1, 2, \ldots, T$, the vector $u_\tau = v_\tau - \mu$, where $\mu = \frac{1}{T}\sum_{t=1}^T v_t$. Let $v_0 = 0$ and $u_0 = -\mu$. We have

$$\sum_{t=1}^T \|u_t\|^2 = \sum_{t=1}^T \|v_t - \mu\|^2 = Q.$$

We have

$$\begin{aligned}
\|\mu_{t+1} - \mu_t\| &= \left\| \frac{1}{t+2}\sum_{\tau=0}^{t+1} v_\tau - \frac{1}{t+1}\sum_{\tau=0}^{t} v_\tau \right\| \\
&= \left\| \frac{1}{t+2}\sum_{\tau=0}^{t+1} u_\tau - \frac{1}{t+1}\sum_{\tau=0}^{t} u_\tau \right\| \\
&\le \frac{1}{(t+1)^2}\sum_{\tau=0}^{t} \|u_\tau\| + \frac{1}{t+1}\|u_{t+1}\|.
\end{aligned}$$

Summing up over all iterations,

$$\begin{aligned}
\rho = \sum_{t=1}^{T-1} \|\mu_{t+1} - \mu_t\| &\le \sum_{t=1}^{T-1} \left(\frac{1}{(t+1)^2}\sum_{\tau=0}^{t} \|u_\tau\| + \frac{1}{t+1}\|u_{t+1}\| \right) \\
&\le \|u_0\| \cdot \sum_{t=1}^{T-1} \frac{1}{(t+1)^2} + \sum_{t=1}^{T} \|u_t\| \cdot \left(\frac{1}{t} + \sum_{\tau=t}^{T-1} \frac{1}{(\tau+1)^2} \right) \\
&\le \|u_0\| + \sum_{t=1}^{T} \frac{2}{t}\|u_t\| \\
&\le 4R\,[\log(1 + Q/R^2) + 2].
\end{aligned}$$

The second inequality follows because $\sum_{\tau=t}^{\infty} \frac{1}{(\tau+1)^2} \le \int_{x=t}^{\infty} \frac{1}{x^2}dx = \frac{1}{t}$. The last inequality uses the following facts:

1. Since $v_0 = 0$, $u_0 = -\mu$, and hence $\|u_0\| = \|\mu\| \le R$ since for all t, $\|v_t\| \le R$.
2. Note $\|u_t\| = \|v_t - \mu\| \le \|v_t\| + \|\mu\| \le 2R$. Applying Lemma 2.5 below with $x_t = \|u_t\|/2R$ for $t = 1, 2, \ldots, T$, and using the fact that $\sum_{t=1}^T x_t^2 = \sum_{t=1}^T \|u_t\|^2/4R^2 \le Q/R^2$, we get that

$$\sum_{t=1}^{T} \frac{2}{t}\|u_t\| \le 4R(\log(1 + Q/R^2) + 1). \qquad \square$$

LEMMA 2.4. $\tilde{Q} \le Q + R^2$.

Proof. Consider the Be-The-Leader (BTL) algorithm played on the sequence of cost functions $c_t(x) = \|v_t - x\|^2$, for $t = 0, 1, 2, \ldots, T$, with $v_0 = 0$, when the convex domain is the ball of radius R. The BTL algorithm is as follows. On round t, this algorithm chooses the point that minimizes $\sum_{\tau=0}^{t} c_t(x)$ over the domain. It is easy to see that this point is exactly $\frac{1}{t+1}\sum_{\tau=0}^{t} v_t = \mu_t$. Thus, the cost of the algorithm is $\sum_{\tau=0}^{t} \|v_t - \mu_t\|^2 = \tilde{Q}$, since $\mu_0 = v_0 = 0$, and the first period cost is thus 0. The best fixed point in hindsight is μ_T. Thus, the cost of the best fixed point in hindsight, μ_T, is $\sum_{\tau=0}^{t} \|v_t - \mu_T\|^2 = Q + \|\mu_T\|^2$. Kalai and Vempala (2005) prove that the BTL algorithm incurs 0 regret, i.e., $\tilde{Q} \le Q + \|\mu_T\|^2 \le Q + R^2$. $\square$

LEMMA 2.5. *Suppose that* $0 \le x_t \le 1$ *and* $\sum_t x_t^2 \le Q$. *Then*

$$\sum_{t=1}^{T} \frac{x_t}{t} \le \log(1 + Q) + 1.$$

Proof. By Lemma 2.6, the values of x_t that maximize $\sum_{t=1}^T x_t/t$ must have the following structure: there is a k such that for all $t \le k$, we have $x_t = 1$, and for any index $t > k$, we have $x_{k+1}/x_t \ge (1/k)/(1/t)$, which implies that $x_t \le k/t$. We first note that $k \le Q$, since $Q \ge \sum_{t=1}^k x_t^2 = k$. Now, we can bound the value as follows:

$$\begin{aligned}\sum_{t=1}^{T} \frac{x_t}{t} &\le \sum_{t=1}^{k} \frac{1}{t} + \sum_{t=k+1}^{T} \frac{k}{t^2} \\ &\le \log(k+1) + k \cdot \frac{1}{k} \\ &= \log(1 + Q) + 1. \end{aligned} \qquad \square$$

LEMMA 2.6. *Let* $a_1 \ge a_2 \ge \ldots a_n > 0$. *Then the optimal solution of*

$$\max\left\{\sum_i a_i x_i \,:\, 0 \le x_i \le 1 \text{ and } \sum_i x_i^2 \le Q\right\}$$

has the following properties: $x_1 \ge x_2 \ge \ldots x_n$, *and for any pair of indices* i, j, *with* $i < j$, *either* $x_i = 1$, $x_i = 0$ or $x_i/x_j \ge a_i/a_j$.

Proof. The fact that in the optimal solution $x_1 \ge x_2 \ge \ldots x_n$ is obvious, since otherwise we could permute the x_i's to be in decreasing order and increase the value.

The second fact follows by the Karush–Kuhn–Tucker (KKT) optimality conditions, which imply the existence of constants μ, $\lambda_1, \ldots, \lambda_n$, $\rho_1, \ldots, \rho_n$ for which the optimal vector x satisfies (here e_i is the ith standard basis vector, and $a = (a_1, a_2, \ldots, a_n)$):

$$-a + 2\mu x + \sum_i (\lambda_i + \rho_i) e_i = 0.$$

By KKT theory, complementary slackness implies that the constants λ_i, ρ_i are equal to zero for all indices of the solution which satisfy $x_i \notin \{0, 1\}$. For these coordinates, the KKT equation is

$$a_i - 2\mu x_i = 0,$$

which implies the lemma. □

3. IMPLICATIONS IN THE GEOMETRIC BROWNIAN MOTION MODEL

We begin with a brief description of the model. The model assumes that stocks can be traded continuously, and that at any time, the fractional change in the stock price within an infinitesimal time interval is normally distributed, with mean and variance proportional to the length of the interval. The randomness is due to many infinitesimal trades that jar the price, much like particles in a physical medium are jarred about by other particles, leading to the classical Brownian motion.

Formally, the model is parameterized by two quantities, the *drift* μ, which is the long-term trend of the stock prices, and *volatility* σ, which characterizes deviations from the long-term trend. The parameter σ is typically specified as *annualized* volatility, i.e., the standard deviation of the stock's logarithmic returns in 1 year. Thus, a trading interval of [0, 1] specifies 1 year. The model postulates that the stock price at time t, S_t, follows a geometric Brownian motion with drift μ and volatility σ

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

where W_t is a continuous-time stochastic process known as the Wiener process or simply Brownian motion. The Wiener process is characterized by three facts:

1. $W_0 = 0$,
2. W_t is almost surely continuous, and
3. for any two disjoint time intervals $[s_1, t_1]$ and $[s_2, t_2]$, the random variables $W_{t_1} - W_{s_1}$ and $W_{t_2} - W_{s_2}$ are independent zero mean Gaussian random variables with variance $t_1 - s_1$ and $t_2 - s_2$, respectively.

Using Itō's lemma (see, for example, Karatzas and Shreve 2004), it can be shown that the stock price at time t is given by

$$S_t = S_0 \exp((\mu - \sigma^2/2)t + \sigma W_t). \tag{3.1}$$

Now, we consider a situation where we have n stocks in the GBM model. Let $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ be the vector of drifts, and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be the vector of (annualized) volatilities. Suppose we trade for 1 year. We now study the effect of trading frequency on the quadratic variation of the stock price returns. For this, assume that the year-long trading interval is sub-divided into T equally sized intervals of length $1/T$, and we trade at the end of each such interval. Let $r_t = (r_t(1), r_t(2), \ldots, r_t(n))$ be the vector of stock returns in the tth trading period. We assume that T is "large enough," which is taken to mean that it is larger than $\mu(i), \sigma(i), (\frac{\mu(i)}{\sigma(i)})^2$ for any i.

Then using the facts of the Wiener process stated above, we can prove the following lemma, which shows that the expected quadratic variation is essentially the same regardless of trading frequency, while its variance decreases with trading frequency.

LEMMA 3.1. *In the setup of trading n stocks in the GBM model over 1 year with T trading periods, where $T \gg \mu(i), \sigma(i), \forall i$, there is a vector v such that*

$$\mathbf{E}\left[\sum_{t=1}^{T} \|r_t - v\|^2\right] \leq \|\sigma\|^2 \left(1 + O\left(\frac{1}{T}\right)\right)$$

and

$$\mathrm{VAR}\left[\sum_{t=1}^{T} \|r_t - v\|^2\right] \leq \frac{3\|\sigma\|^4}{T} \left(1 + O\left(\frac{1}{T}\right)\right),$$

regardless of how the stocks are correlated.

Proof. For every stock i, it follows from the GBM equation for the stock prices ($S_t = S_0 \exp((\mu - \sigma^2/2)t + \sigma W_t)$) that its return $r_t(i)$ in period t is given by

$$r_t(i) = \exp\left(\left(\mu(i) - \frac{\sigma(i)^2}{2}\right)\frac{1}{T} + \sigma(i) X_t(i)\right),$$

where $X_t(i) \sim \mathcal{N}(0, \frac{1}{T})$. Thus, for any given stock i, the returns $r_1(i), r_2(i), \ldots, r_T(i)$ are i.i.d. log-normal random variables, with parameters:

$$r_t(i) \sim \ln\mathcal{N}\left(\frac{\mu(i)}{T} - \frac{\sigma(i)^2}{2T}, \frac{\sigma(i)^2}{T}\right)$$

Recall that for a log-normal random variable $X \sim \ln\mathcal{N}(\mu, \sigma^2)$ the mean and variance are given by $\mathbf{E}[X] = e^{\mu+\sigma^2/2}$ and $\mathrm{VAR}[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.

Define the vector v as $v(i) = \mathbf{E}[r_t(i)] = e^{\mu(i)/T}$. Then, assuming that $T > \mu(i), \sigma(i)$, we have using the exponential approximation $e^x \leq 1 + x + x^2$ for $-1 \leq x \leq 1$:

$$\begin{aligned}
\mathbf{E}[(r_t(i) - v(i))^2] = \mathrm{VAR}(r_t(i)) &= (e^{\sigma(i)^2/T} - 1)e^{2\mu(i)/T} \\
&\leq \left(\frac{\sigma(i)^2}{T} + O\left(\frac{1}{T^2}\right)\right) \cdot e^{2\mu(i)/T} \\
&\leq \frac{\sigma(i)^2}{T}\left(1 + O\left(\frac{1}{T}\right)\right),
\end{aligned}$$

where the last equality uses the Taylor approximation of the exponential and the fact that $\frac{2\mu(i)}{T} \ll 1$.

Summing up over all stocks i and all periods t, and using linearity of expectation, we get the first part of the lemma:

$$\mathbf{E}\left[\sum_{t=1}^{T} \|r_t - v\|^2\right] = \sum_{t=1}^{T}\sum_{i=1}^{n} \mathbf{E}[(r_t(i) - v(i))^2] \leq \|\sigma\|^2\left(1 + O\left(\frac{1}{T}\right)\right).$$

As for the bound on the variance, we first bound the quantity $\mathbf{E}[(r_t(i)-v(i))^4]$, using the fact that if $X \sim \ln\mathcal{N}(\mu,\sigma^2)$, then $X^a \sim \ln\mathcal{N}(a\mu, a^2\sigma^2)$. We denote $\tilde{\mu} = \frac{\mu(i)}{T} - \frac{\sigma(i)^2}{2T}$, $\tilde{\sigma}^2 = \frac{\sigma(i)^2}{T}$.

$$\begin{aligned}
\mathbf{E}[(r_t(i)-v(i))^4] &= \mathbf{E}[r_t(i)^4 - 4r_t(i)^3 v(i) + 6r_t(i)^2 v(i)^2 - 4r_t(i)v(i)^3 + v(i)^4] \\
&= e^{4\tilde{\mu}+8\tilde{\sigma}^2} - 4e^{3\tilde{\mu}+\frac{9}{2}\tilde{\sigma}^2} e^{\tilde{\mu}+\tilde{\sigma}^2/2} + 6e^{2\tilde{\mu}+2\tilde{\sigma}^2} e^{2\tilde{\mu}+\tilde{\sigma}^2} \\
&\quad - 4e^{\tilde{\mu}+\tilde{\sigma}^2/2} e^{3\tilde{\mu}+\frac{3}{2}\tilde{\sigma}^2} + e^{4\tilde{\mu}+2\tilde{\sigma}^2} \\
&= e^{4\tilde{\mu}+8\tilde{\sigma}^2} - 4e^{4\tilde{\mu}+5\tilde{\sigma}^2} + 6e^{4\tilde{\mu}+3\tilde{\sigma}^2} - 4e^{4\tilde{\mu}+2\tilde{\sigma}^2} + e^{4\tilde{\mu}+2\tilde{\sigma}^2} \\
&= e^{4\tilde{\mu}}(e^{8\tilde{\sigma}^2} - 4e^{5\tilde{\sigma}^2} + 6e^{3\tilde{\sigma}^2} - 3e^{2\tilde{\sigma}^2}) \\
&= (e^{\tilde{\sigma}^2}-1)^2 e^{4\tilde{\mu}}(e^{6\tilde{\sigma}^2} + 2e^{5\tilde{\sigma}^2} + 3e^{4\tilde{\sigma}^2} - 3e^{2\tilde{\sigma}^2}) \\
&\le \left(\frac{\sigma(i)^4}{T^2} + O\left(\frac{1}{T^3}\right)\right) \cdot e^{4\tilde{\mu}}(e^{6\tilde{\sigma}^2} + 2e^{5\tilde{\sigma}^2} + 3e^{4\tilde{\sigma}^2} - 3e^{2\tilde{\sigma}^2}) \\
&= \frac{\sigma(i)^4}{T^2}\left(3 + O\left(\frac{1}{T}\right)\right) = \frac{3\sigma(i)^4}{T^2}\left(1 + O\left(\frac{1}{T}\right)\right).
\end{aligned}$$

Above the fifth equality follows from the polynomial factorization:

$$x^8 - 4x^5 + 6x^3 - 3x^2 = (x-1)^2(x^6 + 2x^5 + 3x^4 - 3x^2).$$

Thus for all stocks i and time periods t, we have:

$$\mathrm{VAR}[(r_t(i)-v(i))^2] \le \mathbf{E}[(r_t(i)-v(i))^4] \le \frac{3\sigma(i)^4}{T^2}\left(1 + O\left(\frac{1}{T}\right)\right). \tag{3.2}$$

Now, we use the following inequality which follows from the Cauchy–Schwarz inequality: for any random variables $X_1, X_2, \dots, X_n$:

$$\mathrm{VAR}\left[\sum_{i=1}^n X_i\right] \le \left(\sum_{i=1}^n \sqrt{\mathrm{VAR}[X_i]}\right)^2.$$

Hence, using inequality (3.2), we get:

$$\mathrm{VAR}[\|r_t - v\|^2] \le \left(\sum_{i=1}^n \sqrt{\mathrm{VAR}[(r_t(i)-v(i))^2]}\right)^2 \le \frac{3\|\sigma\|^4}{T^2}\left(1 + O\left(\frac{1}{T}\right)\right).$$

Finally, using the fact that in the GBM model, the returns are independent between periods we get

$$\mathrm{VAR}\left[\sum_{t=1}^T \|r_t - v\|^2\right] \le \frac{3\|\sigma\|^4}{T}\left(1 + O\left(\frac{1}{T}\right)\right). \qquad \square$$

Applying this bound in our algorithm, we obtain the following regret bound from Corollary 1.2.

THEOREM 3.2. *In the setup of Lemma 3.1, for any $\delta > 0$, with probability at least $1 - \delta$, we have*

$$\text{Regret} = O\left(n \log\left(\left(1 + \frac{3}{\sqrt{\delta T}}\right)\|\sigma\|^2 + n\right)\right).$$

Proof. First, we show that the market variability parameter, r, is $\Omega(1)$ with high probability. Fix any stock i. The return $r_t(i)$ for stock i at time period t is a log-normal random variable: $r_t(i) = \exp(N_t(i))$ where $N_t(i) \sim \mathcal{N}((\mu(i) - \frac{\sigma(i)^2}{2})\frac{1}{T}, \frac{\sigma(i)^2}{T})$. Using standard tail bounds for normal variables (which is given in Lemma B.1 in the Appendix for completeness), we have

$$\Pr\left[\left|N_t(i) - \left(\mu(i) - \frac{\sigma(i)^2}{2}\right)\frac{1}{T}\right| > \frac{\sigma(i)}{\sqrt{T}}\sqrt{2\log(2nT/\delta)}\right] < e^{-\log(2nT/\delta)} = \frac{\delta}{2nT}.$$

Thus, assuming $T \gg \mu(i), \sigma(i)$, with probability at least $1 - \frac{\delta}{2nT}$, we have

$$|N_t(i)| = O\left(\frac{\sigma(i)}{\sqrt{T}}\sqrt{\log(nT/\delta)}\right).$$

Since $r_t(i) = \exp(N_t(i))$, we conclude that with probability at least $1 - \frac{\delta}{2nT}$, we have

$$|r_t(i) - 1| = O\left(\frac{\sigma(i)}{\sqrt{T}}\sqrt{\log(nT/\delta)}\right).$$

Applying a union bound over all stocks and all periods, we conclude that with probability at least $1 - \frac{\delta}{2}$, the market variability parameter r is at least $1 - O(\frac{\sigma}{\sqrt{T}}\sqrt{\log(nT/\delta)}) > 0.5$, if $T > \Omega(\sigma^2 \log(nT/\delta))$.

Now we bound the total variation. Applying Chebyshev's inequality to the random variable $\sum_{t=1}^{T} \|r_t - v\|^2$ and using Lemma 3.1, we have

$$\Pr\left[\sum_{t=1}^{T} \|r_t - v\|^2 > \left(1 + \frac{3}{\sqrt{\delta T}}\right)\|\sigma\|^2\right] < \frac{\frac{3\|\sigma\|^4}{T}\left(1 + O\left(\frac{1}{T}\right)\right)}{\frac{9\|\sigma\|^4}{\delta T}} < \frac{\delta}{2}.$$

The result now follows by a union bound applied to our regret bound of Corollary 1.2. $\square$

Theorem 3.2 shows that one expects to achieve constant regret independent of the trading frequency, as long as the total trading period is fixed. This result is only useful if increasing trading frequency improves the performance of the best constant rebalanced portfolio. Indeed, this has been observed empirically (see, e.g., Agarwal et al. 2006 and Section 5).

To obtain a theoretical justification for increasing trading frequency, we consider an example where we have two stocks that follow independent Black–Scholes models with the same drifts, but different volatilities σ_1, σ_2. The same drift assumption is necessary because in the long run, the best CRP is the one that puts all its wealth on the stock with the greater drift. We normalize the drifts to be equal to 0, this doesn't change the performance in any qualitative manner.

Since the drift is 0, the expected return of either stock in any trading period is 1; and since the returns in each period are independent, the expected final change in wealth, which is the product of the returns, is also 1. Thus, in expectation, any CRP (indeed, any portfolio selection strategy) has overall return 1. We therefore turn to a different criterion for selecting a CRP. The risk of an investment strategy is measured by the variance of its payoff; thus, if different investment strategies have the same expected payoff, then the one to choose is the one with minimum variance. We therefore choose the CRP with the least variance.

LEMMA 3.3. *In the setup where we trade two stocks with zero drift and volatilities σ_1, σ_2, the variance of the minimum variance CRP decreases as the trading frequency increases.*

Proof. To compute the minimum variance CRP, we first compute the variance of the CRP $(p, 1-p)$:

$$\begin{aligned}\mathrm{VAR}\left[\prod_t (pr_t(1) + (1-p)r_t(2))\right] &= \mathbf{E}\left[\prod_t (pr_t(1) + (1-p)r_t(2))^2\right] - 1 \\ &= \prod_t \mathbf{E}[(pr_t(1) + (1-p)r_t(2))^2] - 1.\end{aligned}$$

The second equality above follows from the independence of the randomness in each period. Thus, the minimum variance CRP is obtained by minimizing $\mathbf{E}[(pr_t(1) + (1-p)r_t(2))^2]$ over the range $p \in [0, 1]$. We now note that in the Black–Scholes model, $r_t(1) = e^{X_1}$ where $X_1 \sim \mathcal{N}(-\frac{\sigma_1^2}{2T}, \frac{\sigma_1^2}{T})$, and $r_t(2) = e^{X_2}$ where $X_2 \sim \mathcal{N}(-\frac{\sigma_2^2}{2T}, \frac{\sigma_2^2}{T})$, and thus

$$\begin{aligned}&\mathbf{E}[(pr_t(1) + (1-p)r_t(2))^2] \\ &\quad = \mathbf{E}[p^2 e^{2X_1} + 2p(1-p)e^{X_1+X_2} + (1-p)^2 e^{2X_2}] \\ &\quad = p^2 e^{\frac{\sigma_1^2}{T}} + 2p(1-p) + (1-p)^2 e^{\frac{\sigma_2^2}{T}}.\end{aligned}$$

This is minimized at $p = \frac{\exp(\frac{\sigma_2^2}{T})-1}{\exp(\frac{\sigma_1^2}{T})+\exp(\frac{\sigma_2^2}{T})-2}$, and at this point, its value is $\frac{\exp(\frac{\sigma_1^2+\sigma_2^2}{T})-1}{\exp(\frac{\sigma_1^2}{T})+\exp(\frac{\sigma_2^2}{T})-2}$. Thus, the variance of the final payoff becomes

$$\left[\frac{\exp\left(\frac{\sigma_1^2+\sigma_2^2}{T}\right) - 1}{\exp\left(\frac{\sigma_1^2}{T}\right) + \exp\left(\frac{\sigma_2^2}{T}\right) - 2}\right]^T - 1.$$

This is a decreasing function of T. To prove this, by direct computation we have

$$\begin{aligned}
&\frac{d}{dT}\left[\left[\frac{\exp\left(\frac{\sigma_1^2+\sigma_2^2}{T}\right)-1}{\exp\left(\frac{\sigma_1^2}{T}\right)+\exp\left(\frac{\sigma_2^2}{T}\right)-2}\right]^T-1\right]\\
&=\left[\frac{\exp\left(\frac{\sigma_1^2+\sigma_2^2}{T}\right)-1}{\exp\left(\frac{\sigma_1^2}{T}\right)+\exp\left(\frac{\sigma_2^2}{T}\right)-2}\right]^T\cdot\log\left[\frac{\exp\left(\frac{\sigma_1^2+\sigma_2^2}{T}\right)-1}{\exp\left(\frac{\sigma_1^2}{T}\right)+\exp\left(\frac{\sigma_2^2}{T}\right)-2}\right]\\
&\quad\times\frac{-\left(\exp\left(\frac{\sigma_1^2}{T}\right)-1\right)^2\exp\left(\frac{\sigma_2^2}{T}\right)\frac{\sigma_2^2}{T^2}-\left(\exp(\frac{\sigma_2^2}{T})-1\right)^2\exp\left(\frac{\sigma_1^2}{T}\right)\frac{\sigma_1^2}{T^2}}{\left(\exp\left(\frac{\sigma_1^2}{T}\right)+\exp\left(\frac{\sigma_2^2}{T}\right)-2\right)^2}\\
&<0.
\end{aligned}$$

The inequality above uses the fact that $\exp(\frac{\sigma_1^2+\sigma_2^2}{T})-1>\exp(\frac{\sigma_1^2}{T})+\exp(\frac{\sigma_2^2}{T})-2$ which can be verified by using the power series expansion of $\exp(x)$. □

Thus, increasing the trading frequency decreases the variance of the minimum variance CRP, which implies that it gets less risky to trade more frequently; in other words, the more frequently we trade, the more likely the payoff will be close to the expected value. On the other hand, as we show in Theorem 3.2, the regret does not change even if we trade more often; thus, one expects to see improving performance of our algorithm as the trading frequency increases.

4. FASTER IMPLEMENTATION

In this section we describe a more efficient algorithm compared to the one from the main body of the paper. The regret bound deteriorates slightly, though it is still logarithmic in the total quadratic variation. The algorithm is based on the online Newton method, introduced in Hazan et al. (2007), and is described in the following figure. For simplicity, we focus on the portfolio management problem, although it is likely that similar ideas can work for general loss functions of the type we consider in this paper.

Algorithm 1 Faster quadratic-variation universal algorithm

for $t=1$ to T **do**
  Use $x_t \triangleq \arg\min_{x\in\Delta_n}\left(\sum_{\tau=1}^{t-1}\tilde{f}_\tau(x)+\frac{1}{2}\|x\|^2\right)$.
  Receive return vector r_t.
  Let $\tilde{f}_t(x)=-\log(r_t\cdot x_t)-\frac{r_t\cdot(x-x_t)}{(r_t\cdot x_t)}+\frac{r(r_t\cdot(x-x_t))^2}{8(r_t\cdot x_t)^2}$.
end for

The basic idea is to bound the cost functions by a paraboloid approximation in the FTRL algorithm. The paraboloid approximation only increases the regret, but since it is a simple quadratic function, the running time of the the FTRL algorithm is improved greatly. All we need to do is optimize the *sum* of quadratic cost functions, which has a compact representation (unlike the sum of log functions), over the n-dimensional simplex. This optimization can be carried out in time $O(n^{3.5})$ using interior point methods (assuming real number operations can be carried out in $O(1)$ time). Using observations made in Hazan et al. (2007) it is possible to further speed up the algorithm and attain a running time proportional to $O(n^3)$.

We have the following regret bound for the algorithm:

THEOREM 4.1. *For the portfolio selection problem, the regret of algorithm 1 is bounded by*

$$\text{Regret} = O\left(\frac{n}{r^3}\log(Q+n)\right).$$

Proof. We first describe the paraboloid approximation to the cost functions that we use in the algorithm instead of actual cost functions. This approximation, based on a more general lemma from Hazan et al. (2007), has the following property, for all x and y in the simplex and any return vector v with coordinates in $[r, 1]$:

$$-\log(v\cdot x) \geq -\log(v\cdot y) - \frac{v\cdot(x-y)}{(v\cdot y)} + \frac{r(v\cdot(x-y))^2}{8(v\cdot y)^2}.$$

Thus for any t, $\tilde{f}_t(x_t) = -\log(r_t\cdot x_t)$, and for any $x\in\Delta_n$, $\tilde{f}_t(x)\leq -\log(r_t\cdot x)$. Thus, if x^* is the best CRP in hindsight, we have the following bound on the regret of algorithm 1:

$$\begin{aligned}\text{Regret} &= \sum_t -\log(r_t\cdot x_t) - \sum_t -\log(r_t\cdot x^*)\\ &\leq \sum_t \tilde{f}_t(x_t) - \sum_t \tilde{f}_t(x^*).\end{aligned}$$

The RHS above is bounded by the regret of algorithm 1 assuming that the cost functions are $\tilde{f}_t$. We therefore proceed to bound this regret of the algorithm with cost functions $\tilde{f}_t$.

The cost functions $\tilde{f}_t$ can be written in terms of the univariate functions

$$g_t(y) = \frac{r}{8(r_t\cdot x_t)^2}\cdot y^2 - \frac{1}{r_t\cdot x_t}\left(1+\frac{r}{4}\right)\cdot y + \frac{r}{8} + 1 - \log(r_t\cdot x_t)$$

as $\tilde{f}_t(x) = g_t(r_t\cdot x_t)$. Now we note that even though the statement of the main theorem assumes that the cost functions can be written in terms of a *single* univariate function g for all t, the proof of the theorem is flexible enough to handle different functions g_t for different t, as long as conditions 3. and 4. in the main theorem on the first and second derivatives of g hold uniformly with the same constants a and b for all functions g_t.

Furthermore, the proof only requires the bound a on the magnitude of the first derivatives at the points x_t which the algorithm produces. Thus, we can now estimate the a and b parameters for the g_t functions as follows: $g_t'(r_t\cdot x_t) = -\frac{1}{(r_t\cdot x_t)} \in [-\frac{1}{r}, 0]$, so we choose

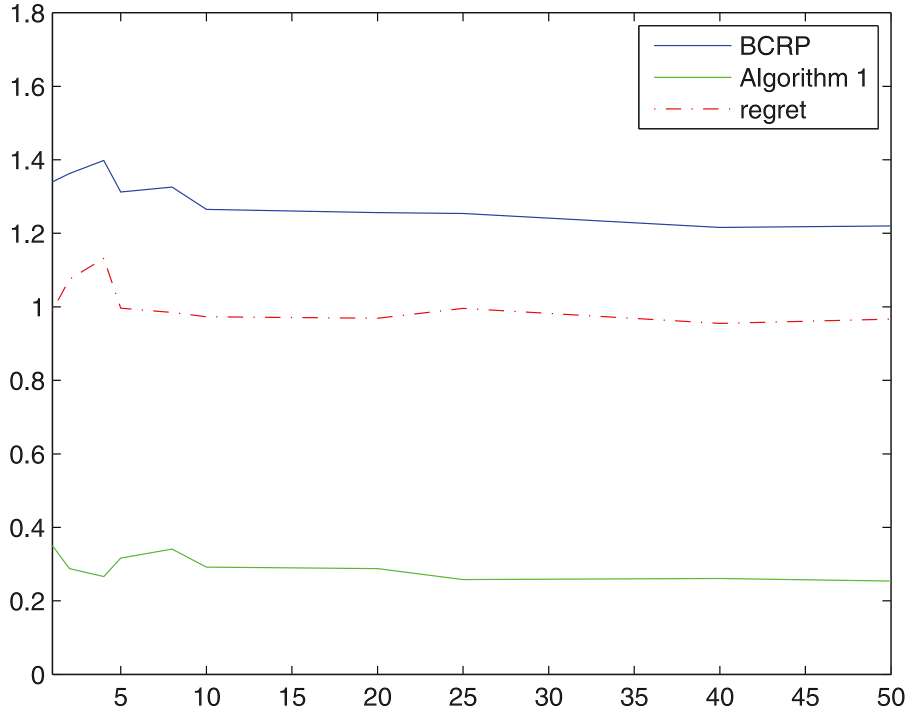


FIGURE 5.1. Best CRP and algorithm exp-concave-FTL vs. trading period. The x-axis denotes the trading period in days, and the y-axis is the log-wealth factor (i.e., a real number).

$a = \frac{1}{r}$. For any portfolio $x \in \Delta_n$, and $g_t''(r_t \cdot x) = \frac{r}{4(r_t \cdot x)^2} \geq \frac{r}{4}$, thus we choose $b = \frac{r}{4}$. The regret bound is now obtained via the bound of the main theorem. □

5. EXPERIMENTS

We tested the performance of the best CRP and Algorithm Exp-Concave-FTL as well as its regret on stock market data. The following graph in Figure 5.1 was generated using real NYSE quotes for the 1,000 trading days from 2001 and 2005 obtained form Yahoo! finance. We randomly chose twenty S&P 500 stocks, and computed the performance of the best CRP and Algorithm Exp-Concave-FTL on the relevant trading period, varying the trading frequency from daily to every 50 trading days (only integer periods which divide 1,000 were tested). The regret, i.e., difference between the log wealth of both methods is also depicted.

For various choices of stocks, the fact that the regret remains pretty much constant was consistent, as predicted by the theoretical arguments in the paper.

The change in performance of the best CRP with respect to trading frequency was not conclusive, at times agreeing with theory and at times not. In one sense, however, this change does agree with theory: in the previous work of Agarwal et al. (2006), Figure 6 depicts the performance of several online portfolio management as varies with the trading period. These more comprehensive experiments measure the average annual percentage yield (APY) in an experiment of sampling a set of stocks from the S&P 500 (the average is taken over the different samples of random stocks from S&P 500). These experiments clearly show that on an average, the performance of many online algorithms, as well as that of the best CRP, improves with trading frequency.

6. CONCLUSIONS

We have presented an efficient algorithm for regret minimization with exp-concave loss functions whose regret strictly improves upon the state of the art. For the problem of portfolio selection, the regret is bounded in terms of the observed variation in stock returns rather than the number of iterations. This is the first theoretical improvement in regret bounds for universal portfolio selection algorithms since the work of Cover (1991).

We show how this fact implies that in the standard GBM model for stock prices the regret does not increase with trading frequency, hence giving the first universal portfolio selection algorithm whose performance improves when the underlying assets are close to GBM. This serves as a bridge between universal portfolio theory and stochastic portfolio theory.

Open Questions. It remains an intriguing open question to improve the dependence of the regret in portfolio selection in terms of other important parameters aside from the quadratic variability: our dependence on the number of stocks in the portfolio (previously denoted n) is linear. In contrast Helmbold et al. (1998) obtain a logarithmic dependence in this parameter. Our regret bounds also depend on the market variability parameter (denoted r), whereas Cover's original algorithm does not have this dependence at all. Is it possible to obtain a $O(\log Q)$ regert bound, via an efficient algorithm, that behaves better with respect to n, r?

APPENDIX A: PROOF OF THE VECTOR HARMONIC SERIES LEMMA

The proof is based on the following fact, whose one-dimensional analogue is an easy consequence of the Taylor expansion of the logarithm.

LEMMA A.1. *Let $A \succeq B \succ 0$ be positive definite matrices. Then*

$$A^{-1} \bullet (A - B) \leq \log \frac{|A|}{|B|},$$

where $|A|$ denotes the determinant of matrix A.

Proof. For any positive definite matrix C, denote by $\lambda_1(C)$, $\lambda_2(C)$, ..., $\lambda_n(C)$ its (positive) eigenvalues. Denote by $\mathbf{Tr}(C)$ the trace of the matrix, which is equal to the sum of the diagonal entries of C, and also to the sum of its eigenvalues.

Note that for the matrix product $A \bullet B = \sum_{i,j=1}^{n} A_{ij} B_{ij}$ defined earlier, we have $A \bullet B = \mathbf{Tr}(AB)$ (where AB is the standard matrix multiplication), since the trace is equal to the sum of the diagonal entries. Therefore,

$$\begin{aligned}
A^{-1} \bullet (A - B) &= \mathbf{Tr}(A^{-1}(A - B)) \\
&= \mathbf{Tr}(A^{-1/2}(A - B)A^{-1/2}) \\
&= \mathbf{Tr}(I - A^{-1/2}BA^{-1/2}) \\
&= \sum_{i=1}^{n}[1 - \lambda_i(A^{-1/2}BA^{-1/2})] \\
&\qquad \left(\because \mathbf{Tr}(C) = \sum_{i=1}^{n} \lambda_i(C)\right) \\
&\leq -\sum_{i=1}^{n} \log[\lambda_i(A^{-1/2}BA^{-1/2})] \\
&\qquad (\because 1 - x \leq -\log(x)) \\
&= -\log\left[\prod_{i=1}^{n} \lambda_i(A^{-1/2}BA^{-1/2})\right] \\
&= -\log|A^{-1/2}BA^{-1/2}| \\
&= \log\left[\frac{|A|}{|B|}\right].
\end{aligned}$$

In the last equality we use the following facts about the determinant of matrices: $|A| = \prod_{i=1}^{n} \lambda_i(A)$, $|AB| = |A||B|$ and $|A^{-1}| = \frac{1}{|A|}$. □

We can now prove the vector harmonic series Lemma 3:

LEMMA A.2.

$$\sum_{t=1}^{T} v_t^{\top}\left(A + \sum_{\tau=1}^{t} v_\tau v_\tau^{\top}\right)^{-1} v_t \leq \log\left[\frac{\left|A + \sum_{\tau=1}^{T} v_\tau v_\tau^{\top}\right|}{|A|}\right].$$

Proof. Let $A_t := A + \sum_{\tau=1}^{t} v_\tau v_\tau^{\top}$ and denote $A_0 = A$. Now by the lemma above

$$\begin{aligned}
\sum_{t=1}^{T} v_t A_t^{-1} v_t &= \sum_t A_t^{-1} \bullet v_t v_t^{\top} \\
&= \sum_t A_t^{-1} \bullet (A_t - A_{t-1}) \\
&\leq \sum_i \log \frac{|A_t|}{|A_{t-1}|} \\
&= \log\left[\frac{|A_T|}{|A_0|}\right].
\end{aligned}$$

□

APPENDIX B: TAIL BOUNDS FOR NORMAL RANDOM VARIABLES

The following standard bound on the tail of the Normal distribution is given here for completeness.

LEMMA B.1. *Let $X \sim N(\mu, \sigma^2)$, then*

$$\Pr[|X - \mu| > x] \leq \frac{\sigma}{x} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Proof. First, consider the case $\mu = 0$, $\sigma = 1$. Then by definition

$$\Pr[X > x] = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\, dt.$$

Since the exponent is a convex function, we can upper bound it by the first derivative at the point $t = x$ and obtain:

$$\int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\, dt \leq \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-x^2/2 - x(t - x))\, dt$$

where the new exponent is the linearization of $-t^2/2$ at $t = x$. Then pull out factors which do not depend on t to get

$$\frac{\exp(x^2/2)}{\sqrt{2\pi}} \int_x^\infty \exp(-xt)\, dt$$

and doing that last integral gives the bound:

$$\Pr[X > x] \leq \frac{1}{\sqrt{2\pi}\, x} \exp(-x^2/2).$$

Hence, by symmetry of the distribution, we get for $X \sim N(0, 1)$:

$$\Pr[|X| > x] \leq \frac{\sqrt{2}}{\sqrt{\pi}\, x} \exp(-x^2/2) \leq \frac{1}{x} \exp(-x^2/2).$$

Next, consider $Y \sim N(\mu, \sigma^2) = \mu + \sigma N(0, 1)$. Hence,

$$\Pr[|Y - \mu| > x] = \Pr\left[|X| > \frac{x}{\sigma}\right] \leq \frac{\sigma}{x} \exp\left(-\frac{x^2}{2\sigma^2}\right). \qquad \square$$

REFERENCES

AGARWAL, A., E. HAZAN, S. KALE, and R. E. SCHAPIRE (2006): Algorithms for Portfolio Management Based on the Newton Method, in *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, New York, NY, USA: ACM, pp. 9–16.

BACHELIER, L. (1900): Théorie de la spéculation, *Annales Scientifiques de l'École Normale Supérieure* 3(17), 21–86.

BLACK, F., and M. SCHOLES (1973): The Pricing of Options and Corporate Liabilities, *J. Polit. Econ.* 81(3), 637–654.

BLUM, A., and A. KALAI (1999): Universal Portfolios with and without Transaction Costs, *Mach. Learn.* 35(3), 193–205.

BOYD, S., and L. VANDENBERGHE (2004): *Convex Optimization*, New York: Cambridge University Press.

CESA-BIANCHI, N., Y. MANSOUR, and G. STOLTZ (2007): Improved Second-Order Bounds for Prediction with Expert Advice, *Machine Learning* 66(2-3), 321–352.

COVER, T. (1991): Universal Portfolios, *Math. Finance* 1, 1–19.

CROSS, J. E., and A. R. BARRON (2003): Efficient Universal Portfolios for Past Dependent Target Classes, *Math. Finance* 13(2), 245–276.

HAZAN, E., A. AGARWAL, and S. KALE (2007): Logarithmic Regret Algorithms for Online Convex Optimization, *Mach. Learn.* 69(2-3), 169–192.

HAZAN, E., and S. KALE (2009): On Stochastic and Worst-Case Models for Investing, in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, eds., Cambridge: MIT Press, pp. 709–717.

HAZAN, E., and S. KALE (2010): Extracting Certainty from Uncertainty: Regret Bounded by Variation in Costs, *Mach. Learn.* 80, 165–188.

HAZAN, E., and S. KALE (2011): Better Algorithms for Benign Bandits, *J. Mach. Learn. Res.* 12, 1287–1311.

HELMBOLD, D. P., R. E. SCHAPIRE, Y. SINGER, and M. K. WARMUTH (1998): On-line Portfolio Selection Using Multiplicative Updates, *Math. Finance* 8(4), 325–347.

JAMSHIDIAN, F. (1992): Asymptotically Optimal Portfolios, *Math. Finance* 2, 131–150.

KALAI, A., and S. VEMPALA (2003): Efficient Algorithms for Universal Portfolios, *J. Mach. Learn. Res.* 3, 423–440.

KALAI, A., and S. VEMPALA (2005): Efficient Algorithms for Online Decision Problems, *J. Comput. Syst. Sci.* 71(3), 291–307.

KARATZAS, I., and S. E. SHREVE (2004): *Brownian Motion and Stochastic Calculus*, New York: Springer Verlag.

MERHAV, N., and M. FEDER (1993): Universal Schemes for Sequential Decision from Individual Data Sequences, *IEEE Trans. Inform. Theory* 39, 1280–1292.

OSBORNE, M. F. M. (1959): Brownian Motion in the Stock Market, *Operat. Res.* 2, 145–173.

ZINKEVICH, M. (2003): Online Convex Programming and Generalized Infinitesimal Gradient Ascent, in T. Fawcett and N. Mishra, eds, *ICML*, AAAI Press, pp. 928–936.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 311–338

THE EFFECT OF TRADING FUTURES ON SHORT SALE CONSTRAINTS

ROBERT JARROW

Cornell University

PHILIP PROTTER

Columbia University

SERGIO PULIDO

Carnegie Mellon University

It is commonly believed that the trading of futures on a commodity enables the market to overcome short selling constraints on the spot commodity itself. This belief is embedded in the notion that trading strategies involving futures contracts enable traders to replicate the payoffs as if they were short the spot commodity. The purpose of this paper is to investigate this common belief in a general arbitrage-free semimartingale financial model with trading in futures and a short selling prohibition on the spot commodity. We show via various examples that, in general, this common belief is incorrect. Furthermore, we provide a set of sufficient conditions, albeit very restrictive, under which the common belief is true.

KEY WORDS: short sale constraints, futures contracts, futures prices, complete markets, martingale representation, supermartingale measures, overpricing hypothesis, bubbles.

1. INTRODUCTION

In the recent financial crisis, short sales bans and restrictions were used by regulators to stabilize declining prices (see Beber and Pagano 2013; Boehmer, Jones, and Zhang 2011). For the same reasons, in most third world emerging markets, short selling is either not allowed or is not possible due to inadequate financial infrastructures (see Bris, Goetzmann, and Zhu 2007; Charoenrook and Daouk 2005). Even in well-developed markets, for many commodities, short selling is impossible because the underlying asset cannot be borrowed, e.g., residential homes. When an asset cannot be shorted, it is often believed that the market may be "overpriced" because the market value does not reflect negative sentiments. This is called the *overpricing hypothesis* (see for instance Miller 1977; chapter 7 of Fabozzi 2004; Boulton and Braga Alves 2010; Brown and Rogers 2012). Such an overpricing can lead to an inefficient allocation of investments within the economy. Consequently, the inability to short an asset is often used as one justification for introducing the trading of futures contracts on that asset (see for instance

We would like to thank Professor Walter Schachermayer for providing the main ideas that led to Example 6.9. We would also like to thank the anonymous referees for their careful reviews of the manuscript. P. Protter is supported in part by NSF grant DMS-0906995.

Manuscript received June 2011; final revision received July 2012.

Address correspondence to Robert Jarrow, Johnson School of Management, 451 Sage Hall, Cornell University, Ithaca, NY 14853, USA; e-mail: raj15@cornell.edu.

DOI: 10.1111/mafi.12013

chapter 3 in Duffie 1989). The common belief is that trading futures contracts on an asset enables the market effectively to short the spot asset, thereby completing the market. The purpose of this paper is to investigate the validity of this belief in a general arbitrage-free semimartingale financial model with trading in futures and short sale prohibitions on the spot asset.

We show via various examples that, in general, this common belief is incorrect: one cannot use futures to replicate the payoffs of being short an asset. Furthermore, we provide a set of sufficient conditions, albeit very restrictive, where this common belief is true. These results are based on the version of the Fundamental Theorem of Asset Pricing proved in Pulido (2011). In addition, when futures contracts do overcome short selling constraints, we also investigate whether trading in only a single index of futures on the spot commodities (instead of all the futures on the spot commodities) is sufficient to complete the market. In general, the answer is again negative. However, hedging strategies involving all the futures on the spot commodities can be replaced by trading strategies using an index futures and only long trades on the futures contracts themselves. Here the long positions on the individual spot asset futures can be approximated by long positions on the spot commodities themselves, yielding an approximation to the desired result. The error term in this approximation depends on the magnitude of the market overpricing. This insight is particularly important for commodities that cannot be sold short and where futures on the individual asset are impractical, but where futures on an index of the commodities (a portfolio) can be traded. Two prime examples satisfying these conditions are: (i) residential homes and (ii) individual equities when short sale constraints are imposed.

The paper is organized as follows. In Section 2, we present two examples that motivate the discussion in the rest of the paper. In Section 3, we extend the Fundamental Theorem of Asset Pricing (FTAP) in markets with short sale prohibitions, as presented in Pulido (2011), to markets with positive interest rates and assets with cash flows. In Section 4, we study the implications that the FTAP imposes on futures prices. In Section 5, under various hypotheses, we characterize arbitrage-free futures prices when interest rates are identically zero. We show that futures prices exhibit more randomness than is present in the spot asset's price process, even under zero interest rates. We also give in Section 6 sufficient conditions under which futures contracts can be used to short the underlying asset, along with mathematical examples where this is not the case and where the spot asset cannot be hedged with traded futures. These examples complement those presented in Section 2. In Section 7, we extend the previous results to positive interest rate economies. Finally, in Section 8 we consider futures contracts on an index and show when trading an index futures is sufficient to overcome the short sale restrictions on the individual stocks themselves.

2. PRELIMINARY EXAMPLES

In this section, we present two examples to show that under short sale restrictions, the common belief that the completeness property of the underlying asset price process (see Definition 6.1) is inherited by the futures price process is incorrect. Other examples will be presented in Section 6. These additional examples will better illustrate the results of Sections 5 and 6 (see Theorems 5.2 and 6.5). As we will see in Section 4, in a market with short sale prohibitions the set of *risk-neutral measures* for the futures market corresponds to the set of equivalent probability measures under which the discounted underlying asset price process (plus discounted cash flows) is a **supermartingale** and the futures price process is a **local martingale.** Furthermore, in the examples provided the futures price process is actually a martingale, which implies that futures prices do not have type 3 bubbles (see Jarrow, Protter, and Shimbo 2010). For simplicity these examples also assume that the default free spot rate of interest is zero and the spot commodities have no cash flows.

The first example is in discrete time. It can be extended to a continuous time example with jumps as in Example 5.5.

EXAMPLE 2.1. Suppose that for $t = 1, \dots, T$ the underlying asset price process is given by

$$S_t = S_0 \prod_{k=1}^{t} (1 + R_k),$$

where S_0 is a positive constant, under P, $R_1, \dots, R_T$ are i.i.d random variables with $P(R_1 = r) = \frac{1}{2}$, $P(R_1 = -r) = \frac{1}{2}$, and $0 < r < 1$. It is known that this financial market, where S trades and the filtration considered is the minimal filtration generated by S, is complete under P. Fix $0 < t^* < t$. Define a measure Q^* such that for $n > t^*$, on the set $\{S_{t^*} = x\}$,

$$Q^*(R_n = -r \mid \mathcal{F}_{n-1}) = p_x \in (1/2, 1),$$

and $Q^*(R_n = r \mid \mathcal{F}_{n-1}) = 1 - p_x$. Assume additionally that Q^* coincides with P over $\sigma(R_1, \dots, R_{t^*})$. Q^* is a probability measure equivalent to P such that S is a Q^*-supermartingale. As we will prove in the next section, if we define a futures price process $F_{t,T}$ (see Definition 4.1) by the formula $F_{t,T} = E^{Q^*}[S_T \mid \mathcal{F}_t]$ (where $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ is the minimal filtration generated by S), the extended market where both S and futures contracts on S (with maturity T) trade, with short sale prohibitions on S, is arbitrage-free (see Proposition 4.2). Observe that in this example, on the set $\{S_{t^*} = x\}$,

$$\begin{aligned}
E^{Q^*}[S_T \mid \mathcal{F}_{t^*}] &= x \sum_{k=0}^{T-t^*} \binom{T-t^*}{k} (1-r)^k (1+r)^{T-t^*-k} p_x^k (1-p_x)^{T-t^*-k} \\
&= x((1-r)p_x + (1+r)(1-p_x))^{T-t^*} \\
&= x((1+r) - 2rp_x)^{T-t^*}.
\end{aligned}$$

Suppose that a fix positive constant c satisfies $0 < c < S_{t^*}$ and

$$r > 1 - \left(\frac{c}{S_{t^*}}\right)^{\frac{1}{T-t^*}}.$$

If we let

$$p_x := \frac{(1+r) - \left(\frac{c}{x}\right)^{\frac{1}{T-t^*}}}{2r} \in (1/2, 1),$$

we have that $F_{t,T} = E^{Q^*}[S_T \mid \mathcal{F}_t]$ is identically equal to the constant c for $t \leq t^*$ and the futures market is not complete under Q^*. Trivially, the futures cannot be used to hedge the underlying's price risk, either long or short. In other words, the futures market does not inherit the completeness property of the underlying market (see Figure 2.1). In other

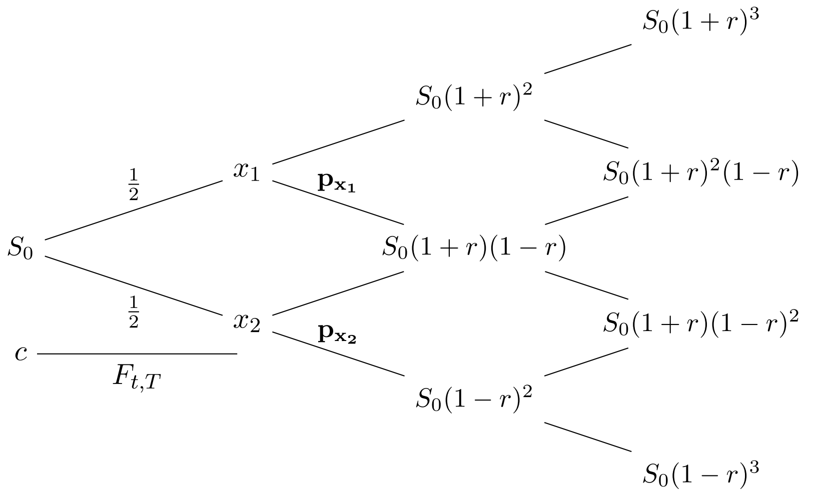


FIGURE 2.1. This is a discrete time representation of Example 2.1. In this figure $t^* = 1$. The probability of going down after time 1 depends on whether the path goes through x_1 or x_2 at time 1. The futures price is constant and equal to c between times 0 and 1.

cases, for instance if the constants p_x are all equal to a constant $p \in (1/2, 1)$, the futures market is complete under Q^*.

For continuous processes we have the following example considered in Cox and Hobson (2005) when studying price bubbles.

EXAMPLE 2.2 (An Asset Price Process with a Bubble). Suppose that for $t < T$ the underlying asset price process S is given by

$$S = 1 + \mathcal{E}\left(\int_0^\cdot \frac{dB_s}{\sqrt{T-s}}\right),$$

where B is a P-Brownian motion. We have that if

$$X := \int_0^\cdot \frac{dB_s}{\sqrt{T-s}},$$

then

$$[X]_t = \ln\left(\frac{T}{T-t}\right) \quad \text{and} \quad \lim_{t\to T}[X]_t = \infty,\ P\text{-a.s.}$$

Since

$$S_t = 1 + \exp\left(X_t - \frac{1}{2}[X]_t\right) = 1 + \exp\left([X]_t\left(\frac{X_t}{[X]_t} - \frac{1}{2}\right)\right),$$

and $\lim_{t\to T}\frac{X_t}{[X]_t} = 0$ (see problem 2.9.3 and theorem 3.4.6 in Karatzas and Shreve (1998a)), if we define $S_T \equiv 1$, then S is a continuous strict local martingale on $[0, T]$, i.e., S is a P-local martingale on $[0, T]$ that is not a P-martingale.

The market consisting of this risky asset price alone is complete. In this case $F_{t,T} := E^P[S_T \mid \mathcal{F}_t]$ (where $\mathcal{F} = (\mathcal{F}_t)_t$ is the minimal filtration generated by S) is constant and equal to 1. Again trivially, although the futures market is arbitrage-free, the futures cannot be used to hedge the underlying's price risk, and the futures market is not complete under P.

For markets with short sales restrictions, we would like to know when futures can be used to hedge the underlying asset's price risk. Before we answer this question in Section 6, we first investigate the dynamics of futures prices consistent with No Free Lunch with Vanishing Risk (NFLVR-S) when short sale prohibitions exist in the underlying market.

3. THE FTAP WITH SHORT SALE PROHIBITIONS

This section extends the results in Pulido (2011) to the case where the money market account's value is stochastic and risky assets have cash flows. We consider a finite trading horizon $[0, T]$ and assume that there are N risky assets trading. We suppose, as in the seminal work of Delbaen and Schachermayer (1994), that the price processes of the N risky assets are nonnegative locally bounded P-semimartingales on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, P)$, where $\mathbb{F} := (\mathcal{F}_t)_{0\leq t\leq T}$ satisfies the usual hypotheses. The probability measure P denotes the statistical probability measure. We further assume that $\mathcal{F}_0$ is P-trivial and $\mathcal{F}_T = \mathcal{F}$. Hence, all random variables measurable with respect to $\mathcal{F}_0$ are P-almost surely constant and there is no additional source of randomness on the probability space. Given two probability measures Q and P on $\mathcal{F}$, we write $Q \sim P$ when P and Q are equivalent.

We denote by $S := (S^i)_{1\leq i\leq N}$ the $\mathbb{R}^N$-valued stochastic price processes for the risky assets. We assume that S^0, the value of a money market account, is a positive and predictable P-semimartingale bounded away from 0. This is almost without loss of generality. In typical models, the money market account's value grows by continuously compounding at the default free spot rate of interest, i.e., the money market account's value equals a constant times the exponential of a Lebesque integral of the spot rate of interest across time. As such, the money market account's value is continuous a.e., even if the spot rate of interest follows a discontinuous sample path process. We denote by $\tilde{S}^i = (S^0)^{-1}S^i$ the *discounted* price process of asset i for $1 \leq i \leq N$ and by $\tilde{S} := (\tilde{S}^1, \ldots, \tilde{S}^N)$ the vector of discounted price processes.

We assume that for each asset i, $1 \leq i \leq N$, there exists a cumulative cash flow process D^i, which is an $\mathbb{F}$-adapted P-semimartingale. We let $D := (D^1, \ldots, D^N)$ be the vector of cumulative cash flow processes.

Also for $1 \leq i \leq N$ we define $\tilde{D}^i := \frac{1}{S^0} \cdot D^i$ (throughout this paper the $\cdot$ symbol represents stochastic integration, see Protter 2005). This integral is well defined since S^0 is nonnegative and bounded away from 0. Finally, let $\tilde{D} := (\tilde{D}^1, \ldots, \tilde{D}^N)$ be the vector of discounted cumulative cash flows.

Last, to capture short sale prohibitions on some of the traded assets, we assume that the first $d \leq N$ assets can be sold short in an admissible way (see Definition 3.1), but the last $N - d$ assets cannot, i.e., we define the admissible strategies as follows.

DEFINITION 3.1. A vector valued process $H = (H^1, \ldots, H^N)$ is called an admissible trading strategy if

(i) $H \in L(\tilde{S} + \tilde{D})$, i.e., H is integrable in the stochastic sense with respect to $\tilde{S} + \tilde{D}$ (see Protter 2005).

(ii) $H_0 = 0$.

(iii) $(H \cdot (\tilde{S} + \tilde{D})) \geq -\alpha$ for some $\alpha > 0$.
(iv) $H^i \geq 0$ for all $i > d$.

We let $\mathcal{A}$ be the set of admissible trading strategies.

We define as in Pulido (2011) the sets

$$\mathcal{K} := \{(H \cdot (\tilde{S} + \tilde{D}))_T : H \in \mathcal{A}\}, \tag{3.1}$$

$$\mathcal{C} := \left(\mathcal{K} - L^0_+(P)\right) \cap L^\infty(P), \tag{3.2}$$

where $L^0_+(P)$ is the space of equivalence classes of nonnegative finite random variables, and $L^\infty(P)$ is the space of P-essentially bounded random variables. With these sets, the arbitrage-free conditions of No Arbitrage under short sales prohibition (NA-S) and No Free Lunch with Vanishing Risk under short sales prohibition (NFLVR-S) are defined as follows. (NA-S) holds if $\mathcal{C} \cap L^0_+(P) = \{0\}$. (NFLVR-S) holds if $\overline{\mathcal{C}} \cap L^0_+(P) = \{0\}$, where $\overline{\mathcal{C}}$ is the closure of $\mathcal{C}$ with respect with the $\|\cdot\|_\infty$ norm in $L^\infty(P)$. Because these definitions are standard we provide no additional explanation. Throughout this paper whenever we write (NFLVR-S) we mean No Free Lunch with Vanishing Risk under short sales prohibition.

The following theorem is an immediate consequence of theorem 3.10 in Pulido (2011) and the fact that S^0 is a positive semimartingale bounded away from 0.

THEOREM 3.2. *Under the additional assumption that for all i, $\tilde{D}^i$ is a locally bounded semimartingale under P, the condition of (NFLVR-S) is equivalent to*

$$\mathcal{M}_{sup}(\tilde{S} + \tilde{D}) \neq \emptyset, \tag{3.3}$$

where $\mathcal{M}_{sup}(\tilde{S} + \tilde{D})$ is the set of probability measures $Q \sim P$ such that $\tilde{S}^i + \tilde{D}^i$ is a Q-local martingale for $1 \leq i \leq d$ and $\tilde{S}^i + \tilde{D}^i$ is a Q-local supermartingale for $d < i \leq N$. If in addition $\tilde{D}^i$ is bounded from below, then $\tilde{S}^i + \tilde{D}^i$ is a Q-supermartingale for Q in $\mathcal{M}_{sup}(\tilde{S} + \tilde{D})$.

Proof. This corresponds to theorem 3.10 in Pulido (2011) after replacing S by $\tilde{S} + \tilde{D}$. The only difference is that $\tilde{S} + \tilde{D}$ is not always nonnegative. However, a careful inspection of the proof of Proposition 3.2 in Pulido (2011) shows that the conclusion of this theorem holds. □

Although the previous theorem seems to be folklore knowledge (see for instance Cvitanić and Karatzas 1993; Jouini and Kallal 1995; Schürger 1996; Frittelli 1997; Pham and Touzi 1999; Napp 2003), we could not find a proof at this level of generality for semimartingales with jumps. Our proof relies on a modification of a known proposition by Ansel and Stricker (1994, see their proposition 3.3), which is proved in Pulido (2011). This theorem will be used to study the conditions under which futures contracts can be used to overcome short sale prohibitions on the underlying commodities.

4. ARBITRAGE-FREE FUTURES PRICES

This section explores the implications that the Fundamental Theorem of Asset Pricing (FTAP) under short sale prohibitions (Theorem 3.2) has on futures prices and their hedging strategies. Initially, to simplify our notation, we assume that there is only one risky asset trading and that this asset cannot be sold short (i.e., we assume that $d = 0$, $N = 1$). We denote by S the price of the risky asset and call it *the underlying asset price process* or *spot price process*. We further assume that there are no cash flows associated to S. As we did in the previous section, we denote by S^0 the price of the money market account, and assume that it is a positive predictable P-semimartingale bounded away from zero. $\tilde{S} := (S^0)^{-1} S$ therefore corresponds to the *discounted price process* of the risky asset.

The previous section shows that (NFLVR-S) guarantees the existence of a probability measure in $\mathcal{M}_{sup}(\tilde{S})$ (see Theorem 3.2). This set of probability measures contains the set of measures $\mathcal{M}_{loc}(\tilde{S})$ defined by

$$\mathcal{M}_{loc}(\tilde{S}) := \{Q \sim P : \tilde{S} \text{ is a } Q\text{-local martingale}\}. \tag{4.1}$$

Furthermore, the Fundamental Theorem of Asset Pricing (FTAP) presented in Delbaen and Schachermayer (1994), shows that this later set $\mathcal{M}_{loc}(\tilde{S}) \neq \emptyset$ if and only if the condition of (NFLVR-S) holds for admissible trading strategies, without restriction (iv) in Definition 3.1. Hence, with the short sale prohibitions on S, the set of *risk neutral* measures is enlarged from $\mathcal{M}_{loc}(\tilde{S})$ to $\mathcal{M}_{sup}(\tilde{S})$. This makes intuitive sense. If the set of trading strategies is restricted by short sale constraints, then the set of price processes consistent with no-arbitrage should increase, and it does.

Suppose that for some $Q^* \in \mathcal{M}_{sup}(\tilde{S})$, $\tilde{S}_t^* := E^{Q^*}[\tilde{S}_T \mid \mathcal{F}_t]$ corresponds to *the discounted fundamental price of S at time t*. It has been argued that in markets with short sale prohibitions, under certain hypotheses on agents' beliefs, the measure Q^* is a strict supermartingale measure, in the sense that $\tilde{S}_0 > \tilde{S}_0^*$. This is known as the *overpricing hypothesis* (see for instance Miller 1977; chapter 7 of Fabozzi 2004; Boulton and Braga Alves 2010; Brown and Rogers 2012). In the case when Q^* is a *strict local martingale measure*, the asset S has a price bubble (see Jarrow, Protter, and Shimbo 2007; Jarrow et al. 2010). This case has been studied extensively (see for instance Delbaen and Schachermayer 1995; Cox and Hobson 2005; Madan and Yor 2006; Jarrow et al. 2007, 2010; Ekström and Tysk 2009; Pal and Protter 2010; Mijatović and Urusov 2012; Ruf 2012).

For hedging, it is often argued that by trading in futures markets, short sale prohibitions can be overcome (see chapter 7 of Duffie 1989; chapter 3 of Fabozzi 2004). The previous examples (see Section 2) show that this is not always true. In the subsequent sections we study both the consequences of short sale prohibitions and the overpricing hypothesis on the ability of hedging strategies using futures to generate a short position in the underlying asset price process S. Our analysis differs from that used in the bubbles literature because we consider supermartingale measures, rather than local martingale measures. Consequently, our conclusions and examples apply to a larger variety of models, including both the Black–Scholes and discrete time models.

We define futures contracts as in Karatzas and Shreve (1998b) and Jarrow and Protter (2009).

DEFINITION 4.1. *A futures contract on a risky asset with price process S and maturity time T is a financial instrument with cash flows $dF_{t,T}$ such that*

(i) $F_{t,T}$ is a nonnegative $\mathbb{F}$-adapted P-semimartingale with $F_{T,T} = S_T$, and
(ii) the market price of the cash flows $(dF_{t,T})_t$ is zero at all times.

$F_{t,T}$ is called the futures price process.

Condition (ii) in this definition makes the futures price process dependent on the market price. Note that the futures price process is different from the market price of the futures contract which is always zero. Using the notation of Section 3, if we denote by S^2 a futures contract on S with maturity T, the discounted market price of S^2 is equal to 0 (i.e., $\tilde{S^2} = 0$) and the cash flow associated with S^2 is $dF_{\cdot,T}$ (i.e., $dD^2 = dF_{\cdot,T}$). Investors are allowed to take both long and short positions in a futures contract.

Intuitively a long futures contract on S at time t obligates the holder to purchase the risky asset at time T at a fixed price $F_{t,T}$, specified at time t. The payment is arranged in different installments determined by the fluctuations of the futures prices over the contract's life. In practice this is accomplished by using a margin account. When futures prices increase, the increment is deposited in the margin account. And when futures prices decrease, the negative increment is withdrawn from the margin account. The futures margin account is said to be *marked-to-market*. The cumulative changes in the futures price between time t and the contract's maturity sum to the spot price of the asset at time T less the original fixed price $F_{t,T}$.

Unlike shorting the underlying asset, taking a short position in a futures contract is unrestricted and it is equivalent to selling the futures contract. A short position in a futures contract is entitled to a stream of cash flows exactly opposite to a long position. In practice, a futures transaction is made through a clearing house where for each seller the clearing house finds a buyer on the other side of the contract. Acting as a middle man between the buyer and seller, the clearing house reduces counterparty risk, and it also facilitates reduced transaction costs and increased market liquidity.

Finally, there is no cash flow paid when entering a long or short position in a futures contract (see (ii) in Definition 4.1), other than the necessity of setting up a margin account (for a detailed exposition we refer the reader to Duffie 1989). However, under our assumptions such margin requirements impose no additional restrictions on the set of trading strategies. This is because a margin account is equivalent to holding the money market account, and borrowing and lending is unrestricted in our model.

4.1. NFLVR-S IN THE EXTENDED MARKET

This section presents some necessary and sufficient conditions on futures price processes under which the underlying asset price process S and the futures contract on S with maturity T satisfy the no arbitrage condition of (NFLVR-S).

PROPOSITION 4.2. *If the futures price process $F_{t,T}$ is a Q-local martingale for some $Q \in \mathcal{M}_{sup}(\tilde{S})$ then the extended market where both the underlying risky asset S and the futures contract on S trade satisfies (NFLVR-S). Conversely, if furthermore S^0 is locally bounded from above, $\tilde{D} := (S^0)^{-1} \cdot F_{\cdot,T}$ is locally bounded, and the extended market satisfies (NFLVR-S), then there exists a $Q \in \mathcal{M}_{sup}(\tilde{S})$ such that $F_{t,T}$ is a Q-local martingale.*

Proof. ($\Rightarrow$) We have that $\tilde{D} = (S^0)^{-1} \cdot F_{\cdot,T}$ is a Q-local martingale since $(F_{t,T})_t$ is a Q-local martingale and $(S^0)^{-1}$ is bounded (S^0 is bounded away from 0). The conclusion follows from Theorem 3.2.

($\Leftarrow$) Assume that S^0 is locally bounded from above and $\tilde{D} = (S^0)^{-1} \cdot F_{\cdot,T}$ is locally bounded. By Theorem 3.2 there exists $Q \in \mathcal{M}_{sup}(\tilde{S})$ such that $\tilde{D}$ is a Q-local martingale, and since $F_{t,T} = F_{0,T} + (S^0 \cdot \tilde{D})_t$ and S^0 is locally bounded, $F_{\cdot,T}$ is a Q-local martingale as well. $\square$

REMARK 4.3. In the above proof the process $\tilde{D}$ corresponds to the discounted cash flows from a futures contract on S with maturity T. Notice that in order for the extended market to satisfy (NFLVR-S) it suffices to ensure the existence of a measure $Q \in \mathcal{M}_{sup}(\tilde{S})$ such that $(F_{t,T})_t$ is a Q-local martingale and not necessarily a Q-martingale (see Jarrow and Protter 2009). Also, since usually $|\mathcal{M}_{sup}(\tilde{S})| > 1$ this proposition shows that the futures price process is not uniquely determined by the underlying asset price process and (NFLVR-S). Indeed, for arbitrary $Q \in \mathcal{M}_{sup}(\tilde{S})$

$$F_{t,T} := E^Q[S_T \mid \mathcal{F}_t],$$

defines the futures price of a futures contract in such a way that the extended market satisfies (NFLVR-S). In principle, these futures price processes could differ across different risk neutral measures (see Example 4.5).

Taking into account the remarks just made we impose the following assumption on the futures contract on S.

ASSUMPTION 4.4. *The futures price process $F_{\cdot,T}$ is a Q^*-martingale, for at least one $Q^* \in \mathcal{M}_{sup}(\tilde{S})$.*

This implies that the extended market including both S and F satisfies (NFLVR-S) and the futures contract on S does not have a bubble with respect to $Q^* \in \mathcal{M}_{sup}(\tilde{S})$ (see Jarrow et al. 2007, 2010).

The purpose of the next example is to illustrate, under Assumption 4.4, the relationship between the arbitrage-free futures price and the underlying asset's price processes. This example shows that the futures price process inherits the dynamics of the underlying asset price process if additional assumptions are made on the measure Q^* of Assumption 4.4 (see equations (4.2) and (4.3)). Furthermore, under these additional assumptions, an explicit formula characterizing the difference between the underlying asset price process and the futures price can be derived (see equation (4.4)). The example motivates Theorem 5.2. It is important to point out that Assumption 4.4 is not a necessary condition for the futures price process to inherit the structure of the asset's price process. For instance, in Example 2.2, $F_{\cdot,T} = S$ defines a futures price process consistent with (NFLVR-S) which has the same dynamics as the spot price, but Assumption 4.4 does not hold in this case.

EXAMPLE 4.5 (The Black–Scholes Model with Short Sale Prohibitions). Assume that

$$dS_t = S_t(\mu dt + \sigma dB_t),$$

where B_t is an $\mathbb{F}$-Brownian motion under P, and μ and σ are positive constants, i.e.,

$$S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right).$$

Assume that the default free spot rate of interest is identically equal to 0 so that $\tilde{S} = S$. Furthermore, suppose that $\mathbb{F}$ is the filtration generated by B, and that $Q^* \in \mathcal{M}_{sup}(\tilde{S})$. By the martingale representation theorem for Brownian motion (see theorem IV-43 in Protter 2005) we have that there exists a predictable process η such that for all $t \in [0, T]$,

$$Z_t := E^P\left[\frac{dQ^*}{dP}\bigg|\mathcal{F}_t\right] = \mathcal{E}(\eta \cdot B)_t.$$

Girsanov's Theorem (theorem III-40 in Protter 2005) implies that under Q^*, S has the semimartingale decomposition

$$\begin{aligned} dS_t &= \sigma S_t\,(dB_t - \eta_t\,dt) + S_t(\sigma\eta_t + \mu)\,dt \\ &= \sigma S_t dB_t^* + S_t(\sigma\eta_t + \mu)\,dt, \end{aligned}$$

where B^* is an $\mathbb{F}$-Brownian motion under Q^*. Since $Q^* \in \mathcal{M}_{sup}(\tilde{S})$, we have that S is a Q^*-supermartingale, and we conclude that the finite variation process

$$\int_0^{\cdot} S_s(\sigma\eta_s + \mu)\,ds$$

is indistinguishable from a nonincreasing process and $\eta \leq -\frac{\mu}{\sigma}$ $P\otimes\lambda$-almost surely, where λ denotes Lebesgue measure on $[0, T]$. Recall that $\frac{\mu}{\sigma}$ is commonly known as the market price of risk.

Now, according to Assumption 4.4 the futures price process of a futures contract on S with maturity T is given by

$$\begin{aligned} F_{t,T} &= E^{Q^*}[S_T \mid \mathcal{F}_t] \\ &= E^{Q^*}\left[S_0 \exp\left(\sigma B_T^* + \int_0^T \left(\sigma\eta_s + \mu - \frac{\sigma^2}{2}\right) ds\right) \middle|\, \mathcal{F}_t\right]. \end{aligned}$$

To obtain a more explicit formula for the futures price process at time t we could impose the additional assumption that

$$\int_t^T \sigma\eta_s\,ds \text{ is } \mathcal{F}_t\text{-measurable.} \tag{4.2}$$

In essence, this additional assumption implies that the futures price process has no additional randomness over the contract's entire maturity, other than that present in the price process itself. In this case we can conclude that

$$F_{t,T} = S_0 \exp\left(\int_0^T (\sigma\eta_s + \mu)\,ds\right)\mathcal{E}(\sigma B^*)_t. \tag{4.3}$$

Under no short sale prohibitions and zero interest rates with $\eta \equiv -\frac{\mu}{\sigma}$, the dynamics of the futures price process is given by

$$S_0\mathcal{E}(\sigma B^*)_t.$$

Hence, in a market with short sale prohibitions, if the overpricing hypothesis and (4.2) hold, then at time t the futures price process has the additional discounting factor

$$\exp\left(\int_0^T (\sigma\eta_s + \mu)\,ds\right).$$

It is interesting to note that in this case

$$\begin{aligned} \frac{F_{t,T}}{S_t} &= \frac{\exp\left(\int_0^T (\sigma\eta_s + \mu)\,ds\right)\mathcal{E}(\sigma B^*)_t}{\exp(\mu t)\mathcal{E}(\sigma B)_t} \\ &= \exp\left(\int_t^T (\sigma\eta_s + \mu)\,ds\right). \end{aligned} \tag{4.4}$$

Since $F_{\cdot,T}$ and S are observable, one can estimate $\eta + \frac{\mu}{\sigma}$ from market price observations. Of course, in order for (4.4) to hold, we are not taking into consideration interest rates (see Section 7) and we are adding an additional assumption on the process η, namely (4.2) (see Theorem 5.2). Equation (4.4) agrees with the empirical evidence on the effect of short sales restrictions on futures prices and stock returns (see for instance Fung and Draper 1999; Boulton and Braga Alves 2010).

REMARK 4.6. Condition (4.2) is a strong assumption. It effectively assumes that the futures price process exhibits no additional randomness over the contract's entire maturity, other than that reflected in the underlying asset's price process itself. There is no reason to believe this must be true a priori. Equivalently, condition (4.2) essentially assumes that the future evolution of the process $\sigma\eta_u$ for $t \leq u \leq T$ is observable at time t. Again, under the filtration generated by the Brownian filtration, there is no reason to believe that this must be true a priori.

This condition might hold if one enlarges the original filtration to include knowledge of $(\eta_u)_{t\leq u\leq T}$ by appealing to the theory of *enlargement of filtrations*. A simple way to do this would be to include the entire process $(\eta_u)_{0\leq u\leq T}$ in $\mathcal{F}_0$ via what is known as an initial enlargement, using Jacod's Criterion (see e.g., chapter VI of Protter 2005). But this approach seems inappropriate for processes such as η which arise through martingale representation. An alternative solution might be to include gradually the process η in the filtration via a dynamic enlargement, but this theory is still in its infancy, see Kchia and Protter (2011).

5. ZERO-INTEREST RATES

This section generalizes Example 4.5 and characterizes the relation between the futures and asset price processes. As in the previous section, we continue to assume that the spot rate of interest is identically equal to 0 so $S = \tilde{S}$. We can describe the dynamics of the futures prices under Assumption 4.4. Before we do so let's recall the following lemma.

LEMMA 5.1. *Suppose that the supermartingale $S > 0$. Then S has one and only one multiplicative decomposition of the form $S = LD$ where L is a positive Q^*-local martingale and D is a positive, predictable, and nonincreasing process with $D_0 = S_0$. Furthermore, if $S = S_0 + N + V$ is the Doob–Meyer decomposition of the Q^*-supermartingale S, with N a Q^*-local martingale, and V a predictable and nonincreasing process with $N_0 = V_0 = 0$ then*

$$L = \mathcal{E}\left(\left(\frac{1}{S_- + \Delta V}\right) \cdot N\right), \tag{5.1}$$

$$D = S_0\left(\mathcal{E}\left(-\left(\frac{1}{S_- + \Delta V}\right) \cdot V\right)\right)^{-1}. \tag{5.2}$$

Proof. The proof of this result can be found in section VI-2-a of Jacod (1979). See also section II-8b of Jacod and Shiryaev (2003). □

By using this lemma and the ideas contained in Example 4.5 we have the following result.

THEOREM 5.2. *Suppose that $S > 0$ and Assumption 4.4 holds. With the notation of Lemma 5.1 we have the following.*

1. *If N is a Q^*-martingale then $F_{t,T} = S_0 + N_t + E^{Q^*}[V_T \mid \mathcal{F}_t]$. Furthermore, if $V_T \in \mathcal{F}_{t_0}$ for some $t_0 < T$, then $F_{t,T} = S_0 + N_t + V_T$ for all $t \in [t_0, T]$.*
2. *If L is a Q^*-martingale and $D_T \in \mathcal{F}_{t_0}$ for some $t_0 < T$, then $F_{t,T} = L_t D_T$ for all $t \in [t_0, T]$.*

Proof. This theorem follows simply from the expression $F_{t,T} = E^{Q^*}[S_T \mid \mathcal{F}_t]$. □

REMARK 5.3. Since this result does not take into consideration interest rates, the futures price coincides with the forward price. Recall that the forward price at time t is the value of K such that the price of a forward contract with delivery price K and maturity T has market price 0 at time t. Assuming that there is no bubble for the forward contract with respect to Q^* this holds if and only if $E^{Q^*}[S_T - K \mid \mathcal{F}_t] = 0$ and $K = E^{Q^*}[S_T \mid \mathcal{F}_t]$.

REMARK 5.4 (Black–Scholes continued). If we assume that the hypotheses of the previous theorem hold, then in the previous Black–Scholes example, the futures price process differs from the asset price process for $t \in [t_0, T]$ by the factor

$$\frac{F_{t,T}}{S_t} = \frac{D_T}{D_t}.$$

When using futures contracts to hedge the asset price process under short sale prohibitions it is important to recognize the existence of this additional factor. Observe that in Example 4.5, $D_t = \exp(\int_0^t (\sigma \eta_s + \mu)\, ds)$ and the previous theorem agrees with (4.3).

Although the conclusion of the previous theorem might seem trivial at first, Theorem 5.2 motivates the discussion in the next section. As the following example shows, the dynamics of the futures price process on the interval $[0, t_0)$ might be very different from the underlying asset's price process and consequently not inherit properties such as completeness (see Definition 6.1) on this interval.

EXAMPLE 5.5. Suppose that $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ and $T = 3$. Suppose that the price process of an asset that cannot be sold short is given by S, where $S_t \equiv \frac{21}{16}$ for $0 \le t < 1$; $S_t(\omega_1) = S_t(\omega_2) = \frac{11}{8}$, $S_t(\omega_3) = S_t(\omega_4) = \frac{10}{8}$ for $1 \le t < 2$; $S_t(\omega_1) = \frac{7}{4}$, $S_t(\omega_2) = \frac{1}{4}$, $S_t(\omega_3) = \frac{3}{2}$, $S_t(\omega_4) = \frac{1}{2}$ for $2 \le t \le 3$. Let $\mathbb{F}$ be the minimal filtration generated by S.

We have in this case that P given by $P(\omega_1) = P(\omega_3) = \frac{3}{8}$, $P(\omega_2) = P(\omega_4) = \frac{1}{8}$ is a martingale-measure for S. Furthermore, P is the only measure that makes S a martingale.

Next, suppose that Q^* is given by $Q^*(\omega_1) = Q^*(\omega_2) = Q^*(\omega_3) = Q^*(\omega_4) = \frac{1}{4}$. Then $Q^* \in \mathcal{M}_{sup}(S)$. We have that $E^{Q^*}[S_t \mid \mathcal{F}_1] \equiv 1$ for $2 \le t \le 3$. It is easy to see that the canonical decomposition of the special semimartingale S relative to Q^* is $S = S_0 + N + V$, where $V_t \equiv 0$ for $0 \le t < 2$ and $V_t = 1 - S_1$ for $2 \le t \le 3$, $N_t = 0$ for $0 \le t < 1$, $N_t = S_1 - S_0$ for $1 \le t < 2$ and $N_t = S_2 + (S_1 - 1) - S_0$ for $2 \le t \le 3$. We have that

$$\begin{aligned}
\left(\frac{1}{S_- + \Delta V} \cdot N\right)_t &= \frac{S_1 - S_0}{S_0 + 0} 1_{[1,3]}(t) + \frac{S_2 - 1}{S_1 + (1 - S_1)} 1_{[2,3]}(t) \\
&= \frac{S_1 - S_0}{S_0} 1_{[1,3]}(t) + (S_2 - 1) 1_{[2,3]}(t), \\
\left(\frac{1}{S_- + \Delta V} \cdot V\right)_t &= \frac{1 - S_1}{S_1 + (1 - S_1)} 1_{[2,3]}(t) \\
&= (1 - S_1) 1_{[2,3]}(t).
\end{aligned}$$

The conditions of Theorem 5.2 are satisfied in this case with $t_0 = 1$.

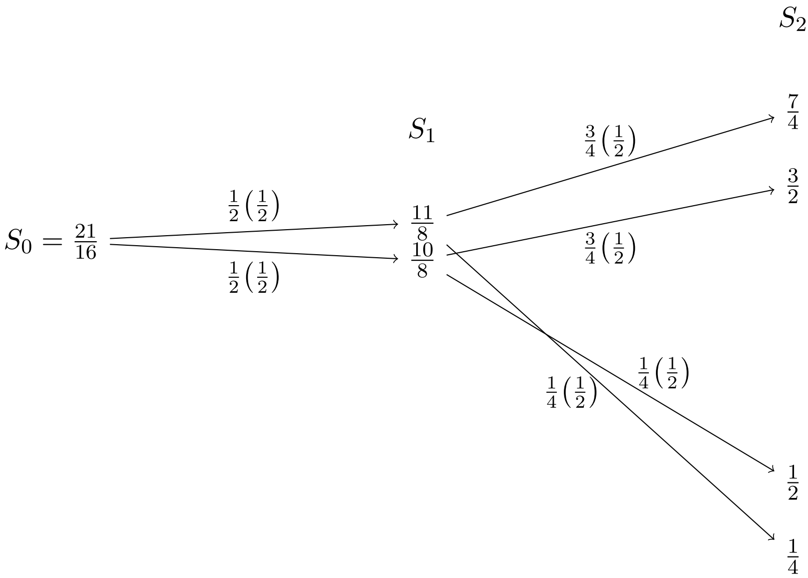


FIGURE 5.1. This is a discrete time representation of Example 5.5. We think of the $\omega_i' s$ as the possible paths and the probabilities of going through each branch are written above the branch (the Q^*-probabilities being those in parentheses).

Observe that in this case $F_{t,T} \equiv 1$ for $0 \leq t < 2$. The multiplicative decomposition of S (see Lemma 5.1) is given by

$$D = \begin{cases} S_0 & \text{if } \ 0 \leq t < 2 \\ \dfrac{S_0}{S_1} & \text{for } \ 2 \leq t \leq 3, \end{cases}$$

$$L = \begin{cases} 1 & \text{for } \ 0 \leq t < 1 \\ \dfrac{S_1}{S_0} & \text{for } \ 1 \leq t < 2 \\ \dfrac{S_1 S_2}{S_0} & \text{for } \ 2 \leq t \leq 3, \end{cases}$$

and the conclusion of the theorem does not hold for $t \in [0, 1)$. A discrete time representation of this example can be found in Figure 5.1.

In this example the futures contract cannot be used to hedge positions on S before time 2. In the next section we prove that when the hypothesis of Theorem 5.2 holds with $t_0 = 0$, market completeness in the underlying asset price process is inherited by the futures market. We will also provide a striking continuous time example (Example 6.9) when this fails, which complements the examples provided in Section 2 and Example 5.5.

6. THE REPRESENTATION PROPERTY

Suppose that the financial market is complete without short sale prohibitions. It is commonly believed that, without taking into account interest rates, the introduction of trading in futures contracts completes the market. In this section we provide sufficient conditions under which this belief is true, and we provide counterexamples to show that it is not true in general.

The concept of a *complete market* is usually defined in terms of a predictable representation property for the assets' price processes (see for instance the seminal works of Harrison and Pliska 1981, 1983). In this section we discuss some conditions under which this property on the underlying asset price process is inherited by the futures price process. To keep the notation simple, as before we assume that there is only one underlying asset price process S trading in the market and that the spot interest rate and associated cash flows are identically zero. We furthermore assume that there exists a $P^* \in \mathcal{M}_{loc}(S)$. In other words, we assume that S satisfies (NFLVR-S) for admissible integrands without short sale prohibitions.

DEFINITION 6.1 *(Market Completeness). We say that the financial model is complete under P^*, or that S has the completeness property with respect to P^*, if every P^*-local martingale M can be expressed as $M = M_0 + (H \cdot S)$ for some $H \in L(S)$.*

Given the results in the previous section it is interesting to study whether the local martingale part N, of the Doob–Meyer decomposition of S under Q^*, inherits the completeness property of the process S. This, in turn, will enable us to derive sufficient conditions under which the futures market inherits the completeness property from the underlying asset market (see Theorem 6.5). To prove this result in general, we first have to prove it for continuous processes.

THEOREM 6.2. *Assume that S is a continuous process that has the completeness property with respect to P^*, and N is the local martingale part in the canonical decomposition of S under Q^*. Then N has the completeness property with respect to Q^*.*

Proof. The proof of this result when S is a P^*-Brownian motion can be found in proposition 5.8.1.2. of Chesney, Jeanblanc, and Yor (2009). The proof can be mimicked step by step for any continuous process and the conclusion follows from Girsanov's Theorem (theorem III-40 in Protter 2005). □

The following result generalizes the previous theorem. It shows that in general the local martingale part of the canonical decomposition of S under Q^* inherits the completeness property of the underlying asset price process S.

THEOREM 6.3. *Assume that S has the completeness property with respect to P^*, and N is the local martingale part in the canonical decomposition of S under Q^*. Then N has the completeness property with respect to Q^*.*

Proof. We let $(0, c, K, A)$ be a good version of the semimartingale characteristics as in (A.1) of Appendix A, with P replaced by P^*. By localization, it is enough to show that any Q^*-martingale X in $\mathcal{H}_1(Q^*)$ (see Protter 2005, p. 195) with $X_0 = 0$ can be written as $X = H \cdot N$ for some $H \in L(N)$. For a semimartingale Z with good version of the semimartingale characteristics (b^Z, c^Z, K^Z, A^Z) we denote by $L_{c,Z}(\omega, t)$ the eigenspace of $c^Z(\omega, t)$ corresponding to the eigenvalue 1 and by $L_{d,Z}(\omega, t)$ the support of the kernel $K^Z(\omega, t)$. By theorem 4.82 in Jacod (1979), to show the above-mentioned representation for N it is enough to show that:

(i) Any continuous Q^*-martingale Y in $\mathcal{H}_1(Q^*)$ can be written as $K \cdot N^c$ for some predictable process K, where N^c is the continuous Q^*-local martingale part of N.
(ii) $L_{c,N}(\omega, t) \cap L_{d,N}(\omega, t) = \{0\}$ $M^{Q^*}_{A^N}-$ a.s. in (ω, t) (see Jacod 1979).

By the same theorem, we have the same properties for S with respect to P^*. (i) follows by Theorem 6.2 and the fact that N^c is the Q^*-local martingale part of S^c. To see (ii), observe that $L_{c,N}$ coincides with $L_{c,S}$, that $L_{d,N}$ is a subset of $L_{d,S}$ by equation (B.1) of Appendix B, and that we can take $A := A^S = A^N$. Hence, $L_{c,N}(\omega, t) \cap L_{d,N}(\omega, t) = \{0\}$ $M_A^{P^*}$−a.s. in (ω, t). Since $Q^* \sim P^*$, this implies that $L_{c,N}(\omega, t) \cap L_{d,N}(\omega, t) = \{0\}$ $M_A^{Q^*}$−a.s. in (ω, t). □

REMARK 6.4. This theorem shows that the completeness property of the local martingale part in the Q^*-canonical decomposition of S is inherited from the completeness property of S with respect to P^*. Theorem 11.29 in Jacod (1979) shows that this condition is necessary for the supermartingale measure Q^* to be an extreme point among all the measures in $\mathcal{M}_{sup}(S)$, but in general the two conditions are not equivalent. Hence, even for measures that are not extreme in $\mathcal{M}_{sup}(S)$, the completeness property of the local martingale is inherited from the underlying process S.

Theorem 6.3 allows us to provide a set of sufficient conditions that guarantees the futures market is complete, even when asset price bubbles exist.

THEOREM 6.5. *Assume the hypotheses of Theorem 5.2 (2) hold with $t_0 = 0$ and that the financial market where S trades is complete under P^*. Then, the futures market where a futures contract with price process $F_{t,T} = E^{Q^*}[S_T \mid \mathcal{F}_t]$ trades is complete under Q^*, i.e., all Q^*-local martingales are of the form $x + (H \cdot F_{\cdot,T})$ for some constant x and $H \in L(F_{\cdot,T})$.*

Proof. If the hypotheses of Theorem 5.2 (2) hold with $t_0 = 0$ then the futures price process dynamics are given by

$$dF_{t,T} = S_0 D_T dL_t = \frac{S_0 D_T L_{t-}}{S_{t-} + (\Delta V)_t} dN_t, \tag{6.1}$$

where $S = S_0 + N + V$ is the Doob–Meyer decomposition of the Q^*-supermartingale S. It follows easily from equation (6.1) and Theorem 6.3 that in this case the futures market is complete with respect to Q^*. □

Theorem 6.5 states that if the discount factor D_T is $\mathcal{F}_0$ measurable, the interest rates are zero, the spot price process is strictly positive, and the futures price process can be written as a conditional expectation, then trading in the futures price process completes the market when the underlying spot price process has restricted short sales. This theorem is perhaps the source of the common belief mentioned earlier because in the classical literature on futures contracts, these three assumptions are common. Indeed, stochastic interest rates were usually not considered so no distinction was usually made between forward and futures contracts, and futures prices were defined as the conditional expectation in Assumption 4.4.

When we relax the hypotheses of this theorem, the completeness property of the underlying asset price process need not be inherited by the futures price process. As we showed in the examples of Section 2, the common belief that trading in futures completes the market when short selling in the spot asset is prohibited is false in general.

EXAMPLE 6.6. In Example 2.1 the conclusion of Theorem 6.5 does not apply. In this case the discount factor D_T is not $\mathcal{F}_0$-measurable.

EXAMPLE 6.7. It is easy to check that the model presented in Example 5.5 is complete under $P = P^*$. However, since $F_{\cdot,T}$ is constant on the interval [0, 2) the futures market is not complete under Q^*. As we already pointed out, in this example the hypothesis of Theorem 5.2 does not hold with $t_0 = 0$, because the discount factor D_T is not $\mathcal{F}_0$-measurable. This example complements Example 2.1, because in this case the probability measure Q^* at time 1 is independent of the level of the underlying asset price, while in Example 2.1 the probability measure Q^* depends on the level of the underlying asset price at time t^*.

EXAMPLE 6.8. In Example 2.2 the conclusion of Theorem 6.5 fails. In this case even though the discount factor D is constant and is identically equal to S_0, the local martingale factor $L(=S/S_0)$ is not a martingale with respect to the measure $Q^*(=P)$.

The following example is even more striking, and it is motivated by Example 2.1. It shows that the futures price process can lack the completeness property even if the underlying asset price process is a continuous martingale under P^* satisfying the completeness property with respect to its minimal filtration. This example is the generalization to continuous time of the idea behind Example 2.1, where after a certain time the measure Q^* is dependent on the level of the underlying asset price process at a fixed time t^*.

EXAMPLE 6.9. In this example, we first construct a price process for a risky asset on a complete and filtered probability space $(\Omega, \mathcal{G}, \mathbb{G}, P^*)$. We denote this price process as S and we show it has the martingale representation property with respect to its own filtration $\mathbb{F}$. We then construct a futures process F for S and show that, under $\mathbb{F}$, the futures process does not have martingale representation. This shows that martingale representation (and hence market completeness) is not—in general—inherited by a futures process, even if the risky asset price process has the property.

We begin with the construction of S, which is slightly complicated. We let W be a standard Brownian motion on $(\Omega, \mathcal{G}, \mathbb{G} := (\mathcal{G}_t), P^*)$. Given $x \in \mathbb{R}$, on the compact time interval [0, 2] we can change to an equivalent probability measure P^x so that under P^x the process $X_t = W_t + xt$ becomes a standard Brownian motion with no drift, on [0, 1]. By Girsanov's Theorem we have that P^x given by

$$\frac{dP^x}{dP^*} = \mathcal{E}(-xW)_1,$$

satisfies the desired property. We next define a $\mathbb{G}$ stopping time τ by

$$\tau = \inf\{t : |W_t| \geq 1\} \wedge 1,$$

and

$$f(x) = E^{P^*}[W_\tau \mathcal{E}(-xW_\tau)] = E^{P^*}\left[W_\tau \exp\left\{-xW_\tau - \frac{x^2}{2}\tau\right\}\right] = E^{P^x}[W_\tau]. \tag{6.2}$$

Since X is a standard Brownian motion on [0, 1] under P^x and τ is a stopping time bounded by 1, $E^{P^x}[X_\tau] = 0$; but

$$0 = E^{P^x}[X_\tau] = E^{P^x}[W_\tau + x\tau] = E^{P^x}[W_\tau] + xE^{P^x}[\tau] = f(x) + xE^{P^x}[\tau]$$

whence $f(x) = -xE^{P^x}[\tau]$. Since $|W_\tau|$ is bounded by 1 and τ strictly positive a.s. we know $-1 < f(x) < 0$ for all $x > 0$. By Lebesgue's dominated convergence theorem we have that f is continuous on $\mathbb{R}_+$. Also, $f(0) = 0$. Therefore $f([0, 1]) = [-R, 0]$ for some constant $R \in (0, 1)$.

Recall we are working under the filtration $\mathbb{G}$, relative to which W is a standard Brownian motion under P^*. We define two stopping times ρ and $\tilde{\tau}$ as follows.

$$\rho = \inf\left\{t : |W_t| \geq \frac{R}{2}\right\} \wedge 1,$$

and

$$\tilde{\tau} = \inf\{t \geq 1 : |W_t - W_1| \geq 1\} \wedge 2.$$

Define

$$S_t = 2 + W_{t\wedge\rho} + (W_{t\wedge\tilde{\tau}} - W_1)1_{[1,2]}(t) > 0.$$

Finally, we let $\mathbb{F} := (\mathcal{F}_t)$ denote the filtration generated by S. We are only interested in $(\mathcal{F}_t)_{0\leq t\leq 2}$.

Note that $\rho < \tau$ and $\rho < \tilde{\tau}$ as well. If we let $\mathcal{H}_t = \sigma\{W_s - W_1; 1 \leq s \leq t\}$ then the filtration $\mathbb{H} = (\mathcal{H}_t)_{t\geq 1}$ is independent of $\mathcal{F}_1$ by the independent increments of W under P^*. Moreover if $\tilde{W}_t = W_t - W_1$ for $t \geq 1$, then $(\tilde{W}_t, \mathbb{H})_{t\geq 1}$ is a Brownian motion independent from $(W_{t\wedge 1})$.

We can also write S as a stochastic integral: $S_t - 2 = \int_0^t H_s dW_s$ where $H_t = 1_{[0,\rho]}(t) + 1_{(1,\tilde{\tau}]}(t)$; then S is at least a P^*-local martingale, but also S is a bounded process and hence it is a P^*-martingale. Note further that S has the same intervals of constancy as $t \mapsto [S, S]_t$ and hence is constant on the random intervals $[\rho, 1$ and $\tilde{\tau}, 2]$.

We are now ready to show that $(S, \mathbb{F})$ has martingale representation for $[0, 2]$. We first need to show that S is martingale for its own filtration $\mathbb{F}$. However since martingales for an arbitrary filtration to which they are adapted remain martingales for their own filtration, we only need to show S is a martingale for a larger filtration, and this we have already done. To show S has martingale representation, we can invoke the Second Fundamental Theorem of Asset Pricing. This theorem states that $(S, \mathbb{F})$ has the martingale representation property on $[0, 2]$ if for any other measure Q equivalent to P^* on $\mathcal{F}_2$ such that $(S, \mathbb{F})$ is a Q-local martingale, we have that $Q = P^*$ on $\mathcal{F}_2$. This result essentially corresponds to corollary 11.4 in Jacod (1979) (see also Harrison and Pliska 1983 and section 9.5 of Chesney et al. 2009). Let Q be a probability measure equivalent to P on $\mathcal{F}_2$ such that $(S, \mathbb{F})$ is a Q-local martingale. Then since the quadratic variation of S is $[S, S]_t = \int_0^t (1_{[0,\rho]}(s) + 1_{(1,\tilde{\tau}]}(s))\, ds$, by Lévy's theorem, $(S_t - S_0)$ is a Brownian motion on the stochastic interval $[0, \rho]$ with respect to Q and $(\mathcal{F}_t)_{0\leq t\leq 1}$, and $(S_t - S_1)$ is also a Brownian on the interval $[1, \tilde{\tau}]$ with respect to Q and $(\mathcal{H}_t)_{1\leq t\leq 2}$ which is independent of $\mathcal{F}_1$. Additionally, S is constant on the intervals $[\rho, 1]$ and $[\tilde{\tau}, 2]$. Therefore under Q the process S has exactly the same distribution as it does for P^* unless ρ and/or $\tilde{\tau}$ have different distributions for Q and P^*. Now, ρ can be defined in terms of the paths of S before ρ, so it will have the same distribution under both P^* and Q, since S is a Brownian motion for each time before ρ due to Lévy's theorem. And $\tilde{\tau}$ can be defined in terms of the paths of S after 1 up to $\tilde{\tau}$, and since again $(S_t - S_1)_{t\geq 1}$ behaves like a standard Brownian motion for both Q and P^*, the distribution of $\tilde{\tau}$ is the same under both P^* and Q. Therefore $Q = P^*$ on $\mathcal{F}_2$ and since Q was arbitrarily chosen by the Second Fundamental Theorem of Asset Pricing we have that $(S, \mathbb{F})$ has martingale representation on $[0, 2]$.

We now turn our attention to the futures process $F = (F_t)_{t\geq 0}$. First, for $y \in [-R, 0]$ define

$$\mu(y) = \inf\{x \geq 0 : f(x) = y\},$$

and Q^* on $\mathcal{F}_2$ by

$$\frac{dQ^*}{dP^*|_{\mathcal{F}_2}} = \mathcal{E}\left(-\int_1^2 \mu(-R/2 - S_1 + 2)\, dS_s\right).$$

Notice that Q^* is well defined because the function μ is bounded on $[-R, 0]$ by 1. Intuitively, the measure Q^* gives drift to the paths of S according to their level at time 1. We have by Bayes' formula, the tower property for conditional expectations, the independence lemma for conditional expectations, and the Markov property of the Brownian motion W that

$$\begin{aligned}
E^{Q^*}[S_2 - S_1|\mathcal{F}_1] &= E^{Q^*}[W_{\tilde{\tau}} - W_1|\mathcal{F}_1] \\
&= E^{P^*}\left[E^{P^*}\left[(W_{\tilde{\tau}} - W_1)\exp\left\{-\mu(-R/2 - S_1 + 2)(W_{\tilde{\tau}} - W_1)\right.\right.\right. \\
&\qquad \left.\left.\left. - \frac{(\mu(-R/2 - S_1 + 2))^2}{2}(\tilde{\tau} - 1)\right\}\middle|\mathcal{G}_1\right]\middle|\mathcal{F}_1\right] \\
&= f(\mu(-R/2 - S_1 + 2)) \\
&= -R/2 - S_1 + 2.
\end{aligned}$$

Therefore,

$$E^{Q^*}[S_2 \mid \mathcal{F}_t] = E^{Q^*}[S_2 - S_1 \mid \mathcal{F}_t] + S_1 \equiv 2 - R/2 \quad \text{for} \quad t \leq 1,$$

S is a $(Q^*, \mathbb{F})$-supermartingale ($\mu(-R/2 - S_1 + 2) \geq 0$) and

$$F_t := E^{Q^*}[S_2 \mid \mathcal{F}_t].$$

Note that F is a Q^* martingale, being the projection of S_2 onto a filtration $\mathbb{F}$, but the filtration $\mathbb{F}$ is not constant on $[0, 1]$ while F is constant there. Since any stochastic integral with respect to F must also be constant where F is constant, $(F, \mathbb{F})$ does not have martingale representation for Q^*.

In the Brownian framework we can obtain *inter alia* necessary and sufficient conditions for market completeness to be inherited by the futures price process, albeit within the confines of a heavily hypothesized paradigm.

EXAMPLE 6.10. Suppose that $dS_t = S_t(\mu dt + \sigma dB_t)$ with B a P-Brownian motion and let $Q^* \in \mathcal{M}_{sup}(S)$ be as in Example 4.5. By following the steps of Example 4.5 we can prove that

$$F_{t,T} = S_0 E^{Q^*}\left[\exp\left(\int_0^T (\mu + \sigma\eta_s)\, ds\right)\mathcal{E}(\sigma B^*)_T \middle| \mathcal{F}_t\right],$$

where B^* is a Q^*-martingale with respect to $\mathbb{F}$ and η is a predictable process such that $(\mu + \sigma\eta) \leq 0$ $P \otimes \lambda$-almost surely, where λ is Lebesgue measure on $[0, T]$.

Assume that $\mathbb{F}$ is the minimal filtration generated by B^* (this is a rather delicate assumption, and an interesting discussion on this subject can be found in chapter V of Revuz and Yor 1999 and section 5.7.1 of Chesney et al. 2009). Assume that $\mu + \sigma\eta - \frac{\sigma^2}{2} = -k(B_t^*)$ for some continuous and nonnegative function k (this hypothesis is also quite delicate since we are imposing a particular functional form dependence of η_t on B_t and η_s for $s < t$).

We can then write

$$\begin{aligned} F_{t,T} &= S_0 E^{Q^*}\left[\exp\left(-\int_0^T k(B_s^*)\,ds\right)\exp\left(\sigma B_T^*\right)\middle|\mathcal{F}_t\right] \\ &= S_0\exp\left(-\int_0^t k(B_s^*)\,ds\right)E^{Q^*}\left[\exp\left(-\int_t^T k\left(B_s^*\right)dt\right)\exp\left(\sigma B_T^*\right)\middle|\mathcal{F}_t\right]. \end{aligned}$$

Using the Markovian property of B^* with respect to $\mathcal{F}_t$ and assuming that the hypotheses of the theorem of Feynman–Kac (theorem 4.4.2 in Karatzas and Shreve 1998a) hold then we can write

$$F_{t,T} = S_0\exp\left(-\int_0^t k\left(B_s^*\right)ds\right)v\left(t,\,B_t^*\right),$$

where

$$v(t,x) = E^x\left[\exp\left(-\int_0^{T-t} k\left(B_s^*\right)ds\right)\exp\left(\sigma B_{T-t}^*\right)\right]$$

solves the Cauchy-problem

$$\begin{aligned} -\frac{\partial v}{\partial t} + kv &= \frac{1}{2}\Delta v; \quad \text{on } [0,T)\times\mathbb{R} \\ v(T,x) &= \exp(\sigma x); \quad x\in\mathbb{R}. \end{aligned}$$

By Itô's formula

$$dF_{t,T} = S_0\exp\left(-\int_0^t k\left(B_s^*\right)ds\right)v_x\left(t,\,B_t^*\right)dB_t^*,$$

where v_x denotes the first partial derivative of v is the x argument.

We would have in this case that the futures market is complete under P^* (for the original filtration $\mathbb{F}$) if and only if the process $(v_x(t,\,B_t^*))_{0\le t\le T}$ is nonzero $P\otimes\lambda$-almost surely, where λ is Lebesgue measure on $[0,T]$.

An alternative way to describe the dynamics of the futures price process when the process η is not deterministic uses Malliavin Calculus. Under technical assumptions on η, by the Generalized Clark-Ocone formula (see Karatzas and Ocone 1991) we have that

$$dF_{t,T} = \left(E^{Q^*}[D_tF|\mathcal{F}_t] + E^{Q^*}\left[F\int_t^T D_t\eta_u\,dB_u^*\middle|\mathcal{F}_t\right]\right)dB_t^*,$$

where $F = S_T = S_0\exp(\mu T + \sigma B_T - \frac{\sigma^2}{2}T)$ and D_t denotes the Malliavin derivative. By using properties of the Malliavin derivative it is straightforward to see that

$$D_tF = \sigma F.$$

This implies that

$$dF_{t,T} = \left(\sigma F_{t,T} + E^{Q^*}\left[S_T\int_t^T D_t\eta_u\,dB_u^*\middle|\mathcal{F}_t\right]\right)dB_t^*.$$

When η is deterministic this yields

$$dF_{t,T} = \sigma F_{t,T}\,dB_t^*,$$

and $F_{t,T} = F_{0,T}\mathcal{E}(\sigma B^*)_t = S_0 \exp(\int_0^T (\mu + \sigma\eta_s)\,ds)\mathcal{E}(\sigma B^*)_t$, as already observed in Example 4.5.

However, when η is random, the volatility of the process $F_{t,T}$ is not necessarily equal to σ and has an additional term equal to

$$E^{Q^*}\left[S_T \int_t^T D_t\eta_u\,dB_u^*\middle|\mathcal{F}_t\right].$$

This difference in volatility in markets with short sales prohibition was experimentally confirmed in Charoenrook and Daouk (2005). Also observe that the futures market is complete if and only if

$$\left(\sigma F_{t,T} + E^{Q^*}\left[S_T \int_t^T D_t\eta_u\,dB_u^*\middle|\mathcal{F}_t\right]\right) \neq 0 \tag{6.3}$$

Lebesgue-almost everywhere $t \in [0, T]$, P-almost surely.

Under the hypothesis of this example, expression (6.3) provides necessary and sufficient conditions for the futures market to inherit the completeness property from the underlying asset price process. Essentially, the volatility of the futures price process must not be zero for any open time interval over the futures contract's maturity. If the volatility is zero, then the futures price does not reflect the randomness generating the underlying asset price process. And of course, the futures market would not be complete.

7. STOCHASTIC INTEREST RATES

In the previous sections the spot rate of interest was equal to zero, and futures and forward prices coincided (see Remark 5.3). It is well known that without short sales restrictions or dividend payments, futures prices differ from forward prices by a factor depending on interest rates (see for instance Duffie 1989; Jarrow and Protter 2009). We have showed that with short sale prohibitions, when the overpricing hypothesis holds, and when the spot rate of interest is zero, there is a difference between the asset price process and the futures price process originating from the multiplicative decomposition of the underlying asset price process with respect to the market's pricing measure Q^* (see Theorem 5.2). Hence, in a market with: (1) short sale prohibitions, (2) stochastic interest rates, (3) when the overpricing hypothesis holds, and (4) when the spot price process has a bubble, then the difference between futures and the spot asset prices can be expressed as a combination of two factors, the one relating to interest rates and the aforementioned "short sales prohibition discount factor." In this section, we exhibit explicit formulas for this difference. We have the following extension of Theorem 5.2.

THEOREM 7.1. *Suppose that $S > 0$ and Assumption 4.4 holds. Suppose that the canonical Doob–Meyer decomposition of the Q^*-supermartingale $\tilde{S}$ is $\tilde{S} = \tilde{S}_0 + \tilde{N} + \tilde{V}$ and that the multiplicative decomposition of $\tilde{S}$ under Q^* is $\tilde{S} = \tilde{L}\tilde{D}$. Where*

$$\tilde{L} = \mathcal{E}\left(\left(\frac{1}{\tilde{S}_- + \Delta\tilde{V}}\right)\cdot\tilde{N}\right),$$

$$\tilde{D} = \tilde{S}_0\left(\mathcal{E}\left(-\left(\frac{1}{\tilde{S}_- + \Delta\tilde{V}}\right)\cdot\tilde{V}\right)\right)^{-1}.$$

Then the following holds.

1. *If* $\tilde{N}$ *is a* Q^**-martingale and* $\tilde{V}_T, S_T^0 \in \mathcal{F}_{t_0}$ *for some* $t_0 < T$ *then* $F_{t,T} = S_0^T(\tilde{S}_0 + \tilde{N}_t + \tilde{V}_T)$ *for all* $t \in [t_0, T]$.
2. *If* $\tilde{L}$ *is a* Q^**-martingale and* $\tilde{D}_T, S_T^0 \in \mathcal{F}_{t_0}$ *for some* $t_0 < T$, *then* $F_{t,T} = L_t S_T^0 D_T$ *for all* $t \in [t_0, T]$.

Proof. This theorem follows directly from

$$F_{t,T} = E^{Q^*}[S_T \mid \mathcal{F}_t] = S_T^0 E^{Q^*}[\tilde{S}_T \mid \mathcal{F}_t] \quad \text{for } t \geq t_0. \qquad \square$$

When the assumptions on S in the previous theorem do not hold, then there is an alternative way to represent the futures price process using additional assumptions on the stochastic evolution of the term structure of interest rates. Define

$$p(t, T) := E^{Q^*}\left[\left(S_T^0\right)^{-1} \mid \mathcal{F}_t\right], \tag{7.1}$$

to be the (discounted) price at time t of a zero-coupon bond with maturity T. We have the following alternative characterization of the futures price process.

THEOREM 7.2. *With the notation of Theorem 7.1 and under Assumption 4.4, let*

$$X := \left(\frac{1}{\tilde{S}_- + \Delta \tilde{V}}\right) \cdot \tilde{N},$$
$$R := -\left(\frac{1}{\tilde{S}_- + \Delta \tilde{V}}\right) \cdot \tilde{V},$$

and suppose that

$$p(\cdot, T) = p(0, T)\mathcal{E}(Y),$$

for a Q^**-local martingale* Y *with* $\Delta Y > -1$. *Then, the futures price process is given by*

$$F_{t,T} = \frac{\tilde{S}_0}{p(0, T)} E^{Q^*}\left[\frac{\mathcal{E}(Z)_T}{\mathcal{E}(R)_T}\middle| \mathcal{F}_t\right], \tag{7.2}$$

where

$$Z := X - Y - [X^c - Y^c, Y^c] - \sum_{s \leq \cdot} \left(\Delta(X - Y)_s \frac{\Delta Y_s}{1 + \Delta Y_s}\right),$$

and X^c, Y^c *are the continuous parts of the* Q^**-local martingales* X *and* Y, *respectively.*

Proof. Observe that by Lemma 5.1 and Theorem 7.1

$$\begin{aligned}
F_{t,T} &= E^{Q^*}[S_T \mid \mathcal{F}_t] \\
&= E^{Q^*}\left[\tilde{S}_T\left(S_T^0\right) \middle| \mathcal{F}_t\right] \\
&= E^{Q^*}\left[\frac{\tilde{S}_T}{p(T, T)}\middle| \mathcal{F}_t\right] \\
&= \frac{\tilde{S}_0}{p(0, T)} E^{Q^*}\left[\frac{\mathcal{E}(X)_T}{\mathcal{E}(Y)_T \mathcal{E}(R)_T}\middle| \mathcal{F}_t\right].
\end{aligned}$$

The theorem follows from lemma 3.4 in Karatzas and Kardaras (2007), which shows that

$$\frac{\mathcal{E}(X)}{\mathcal{E}(Y)} = \mathcal{E}(Z). \qquad \square$$

REMARK 7.3. The ideas presented in this theorem extend those in Amin and Jarrow (1992). In their work, there is no short sales prohibition (so $R \equiv 0$), the processes X and Y are continuous, $[X - Y, Y]$ is assumed to be deterministic, and $\mathcal{E}(X - Y)$ is a Q^*-martingale. In this case we can rewrite formula (7.2) as

$$\begin{aligned} F_{t,T} &= \frac{\tilde{S}_0 \exp(-[X - Y, Y]_T)}{p(0, T)} \mathcal{E}(X - Y)_t \\ &= \frac{\tilde{S}_0 \exp(-[X - Y, Y]_T + [X - Y, Y]_t)}{p(0, T)} \mathcal{E}(Z)_t \\ &= \frac{\tilde{S}_t}{p(t, T)} \exp(-[X - Y, Y]_T + [X - Y, Y]_t), \end{aligned}$$

which corresponds to equation (3.26) in Amin and Jarrow (1992).

These results exhibit explicitly the fact that the discount factors of futures prices have two sources of randomness, X and Y, one coming from the underlying asset price process and another from stochastic interest rates Y. Hence, in order to use futures contracts to hedge the asset's price randomness, it is not only important to adjust for the "short sales prohibition discount factor" $\mathcal{E}(R)^{-1}$, but also for the interest rate factor as well. Traded zero-coupon bonds are needed to hedge this later randomness.

8. THE MULTIDIMENSIONAL CASE AND FUTURES ON AN INDEX

This section presents the natural extension of the previous results to multidimensional markets. We also study the behavior of futures on an index, and the effect of short sale prohibitions on hedging strategies using these instruments. These results are of particular importance for assets that cannot be short sold, where futures on the individual asset are impractical, but where futures on an index of the assets (a portfolio) can be traded. The two prime examples satisfying these conditions are residential homes, and individual equities with short sale constraints. Another example would be for electricity and electricity futures, where the underlying commodity—electricity—cannot be stored in significant quantities.

We again adopt the notation used in Section 2. We assume that there are N securities trading in the market, from which the first d assets can be sold short in an admissible way (see Definition 3.1) and the last $N - d$ cannot. Let $S = (S^1, \ldots, S^N)$ represent the price processes of these assets. In this section, to simplify our notation we will assume again that the spot rate of interest is identically equal to zero, however we note that the results can be extended to the case of stochastic interest rates using the ideas from the previous section.

Throughout this section we will assume that $S^i > 0$ for all $i \leq N$. For each i, the Doob–Meyer decomposition of the Q^*-supermartingale S^i is $S^i = S^i_0 + N^i + V^i$. For each i, we denote by L^i and D^i the local martingale and predictable part, respectively, in the multiplicative decomposition $S^i = L^i D^i$. For each $i \leq N$ we suppose that Assumption 4.4 holds for the futures price of a futures contract on S^i with maturity T. In other words, we assume that if $F^i_{\cdot,T}$ is the futures price of a futures contract on S^i with maturity T then

ASSUMPTION 8.1.

$$F^i_{\cdot,T} \text{ is a } Q^*\text{-martingale.}$$

We denote by $F_\cdot = (F^1_{\cdot,T}, \ldots, F^N_{\cdot,T})$ the vector of futures price processes. For fixed *deterministic* positive weights ω_i we define the index value by

$$I = \sum_{i=1}^{N} \omega_i S^i. \tag{8.1}$$

The following observation immediately follows.

PROPOSITION 8.2. *If I is a Q^*-martingale, then for all $i \leq N$, S^i is a Q^*-martingale.*

Proof. Since S^i is a Q^*-supermartingale for all $i \leq N$ we have that if $E^{Q^*}[S^i_T|\mathcal{F}_i] < S^i_t$ for some $i \leq N$ and $t < T$ then

$$I_t = E^{Q^*}[I_T \mid \mathcal{F}_t] = \sum_{i \leq N} \omega_i E^{Q^*}\left[S^i_T \middle| \mathcal{F}_t\right] < \sum_{i \leq N} \omega_i S^i_t = I_t,$$

which is a contradiction. Then, it must be that S^i is a Q^*-martingale for all $i \leq N$.

This proposition shows that if the index I is traded, sold short, and has no bubble then none of the underlying asset price processes have bubbles either, using the definiton of bubbles in Jarrow et al. (2007, 2010).

It is also of interest to study futures contracts on indexes. In this regard we make the following assumption.

ASSUMPTION 8.3. *The futures price of a futures contract on I, $F^I_{\cdot,T}$, is a Q^*-martingale.*

The following theorem describes the behavior of futures on an index and a related hedging result.

THEOREM 8.4. *Assume that for all $i \leq N$ D^i is deterministic and L^i is a true Q^*-martingale and that Assumptions 8.1 and 8.3 hold. We have that if for some probability measure $P^* \sim P$*

$$\mathcal{M}_{loc}(S) = \{Q \sim P : \ S^i \textit{ is a } Q\textit{-local martingale for all } i \leq N\} = \{P^*\}$$

then,

$$\mathcal{M}_{loc}(F) := \{Q \sim P : \ F^i_{\cdot,T} \textit{ is a } Q\textit{-local martingale for all } i \leq N\} = \{Q^*\}.$$

Additionally, any Q^*-local martingale M can be written as

$$M = M_0 + (H \cdot F),$$

for some predictable process $H \in L(F)$, where $L(F)$ is the space of predictable processes integrable with respect to F.

If additionally $H^i \in L(F^i_{\cdot,T})$ for all $i \leq N$ then

$$M = M_0 + (K \cdot Y),$$

where $Y = (F^1_{\cdot,T}, \ldots, F^N_{\cdot,T}, F^I)$ *and* K *is a predictable process in* $L(Y)$ *such that* $K^i \geq 0$ *for all* $i \leq N$.

Proof. That $\mathcal{M}_{loc}(F) = \{Q^*\}$ and that any Q^*-local martingale M can be written as

$$M = M_0 + (H \cdot F),$$

for some predictable process $H \in L(F)$, follows from the Second Fundamental Theorem of Asset pricing (see corollary 11.4 in Jacod 1979; Harrison and Pliska 1983; section 9.5 of Chesney et al. 2009) and Theorem 6.3 (the same proofs apply to this framework). It remains to prove the representation property with respect to the futures contract on I. Let M be an arbitrary Q^*-martingale and $H \in L(F)$ such that

$$M = M_0 + (H \cdot M).$$

We have that for each $i \leq N$

$$-dF^i_{t,T} = \frac{1}{\omega_i}\left(-dF^I_t + \sum_{j \leq N, j \neq i} \omega_j dF^j_{t,T}\right). \tag{8.2}$$

If additionally $H^i \in L(F^i_{\cdot,T})$ for all $i \leq N$, we can write (see Jacod 1980)

$$\begin{aligned}
H_t dF_t &= \sum_{i \leq N} H^i_t dF^i_{t,T} \\
&= \sum_{i \leq N} \left(H^i_t 1_{\{H^i_t \geq 0\}} dF^i_{t,T} + H^i_t 1_{\{H^i_t < 0\}} dF^i_{t,T}\right) \\
&= \sum_{i \leq N} \left(H^i_t 1_{\{H^i_t \geq 0\}} dF^i_{t,T} + (-H^i_t) 1_{\{H^i_t < 0\}} (-dF^i_{t,T})\right) \\
&= \sum_{i \leq N} \left(H^i_t 1_{\{H^i_t \geq 0\}} dF^i_{t,T} + (-H^i_t) 1_{\{H^i_t < 0\}} \frac{1}{\omega_i}\left(-dF^I_{t,T} + \sum_{j \leq N, j \neq i} \omega_j dF^j_{t,T}\right)\right).
\end{aligned}$$

Observe that the last equation can be rewritten as

$$H_t dF_t = \sum_{i \leq N} K^i_t dF^i_{t,T} + K^{N+1}_t dF^I_{t,T},$$

with $K^i \geq 0$ for all $i \leq N$. □

REMARK 8.5. This theorem states that under certain technical assumptions, hedging strategies involving all the futures contracts on the asset price processes can be replaced by a trading strategy using a futures on an index and trades that are only *long* on these futures contracts.

It is also important to note that in order to hedge claims using the futures contract on an index, one needs the futures contracts on the individual asset price processes, rather than the asset price processes themselves. A short position on $F^I_{\cdot,T}$ combined with a long position on the individual asset price processes (modulus some constant coefficients) is not necessarily equivalent to a short position on the asset price process due to the factors D^i by which asset and futures prices differ (see equation (8.2)). More precisely, we have that

$$dF^i_{t,T} = D_T dL^i_t,$$

while

$$dS^i_t = D^i_{t-} dL^i_t + L^i_{t-} dD^i_t + d\left[D^i_t, L^i_t\right].$$

Nonetheless, we see how the difference between the dynamics of the futures and asset price processes depends on the "overpricing discount factor" D.

APPENDIX A: SEMIMARTINGALE CHARACTERISTICS UNDER THE REFERENCE MEASURE

For a more detailed discussion of the results presented below we refer the reader to chapter III of Jacod and Shiryaev (2003). To simplify our notation, in this appendix we will assume that $S^0 \equiv 1$ and there are no cash flows. However under mild hypotheses, by the results presented in Section 3 (see Theorem 3.2), the analysis below can be extended to markets with stochastic interest rates and assets with cash flows. To obtain such an extension, one replaces the underlying asset price process by the discounted process plus its discounted cash flows.

Let S be an $\mathbb{R}^N$-valued process representing the prices of the risky assets in the market. Since we have assumed that S is a P-semimartingale, it has a canonical representation given by

$$S = S_0 + S^c + (x1_{\{|x|\leq 1\}}) * (\mu^S - \nu) + (x1_{\{|x|>1\}}) * \mu^S + B,$$

where $*$ denotes integration with respect to a random measure (see section II-1a of Jacod and Shiryaev 2003) and

(i) S^c is a continuous P-local martingale starting at 0, known as the continuous martingale part of S,

(ii) μ^S is the random measure associated to the jumps of S defined by

$$\mu^S([0, t] \times A) = \sum_{s\leq t} 1_{A\setminus\{0\}}(\Delta S_s),$$

for $0 \leq t \leq T$ and $A \subset \mathbb{R}^N$,

(iii) ν is the compensator of the random measure μ^S (see theorem II-1.8 in Jacod and Shiryaev 2003),

(iv) B is a predictable $\mathbb{R}^N$-valued process with components of finite variation.

If we define $C_{i,j} = [(S^c)^i, (S^c)^j]$ then (B, C, ν) are known as the **semimartingale characteristics of S under P** with respect to the canonical truncation function $h(x) = x1_{\{|x|\leq 1\}}$. According to proposition II-2..9 in Jacod and Shiryaev (2003) one can find a version of the characteristics (B, C, ν) of S of the form

$$B = b \cdot A,\quad C = c \cdot A,\quad \nu(\omega, dt, dx) = dA_t(\omega) K_{\omega,t}(dx), \tag{A.1}$$

where A is a predictable locally integrable nondecreasing process; b and c are predictable processes, with b taking values in $\mathbb{R}^N$ and c taking values in the set of symmetric nonnegative $N \times N$ matrices; and $K_{\omega,t}(dx)$ is a transition kernel from $(\Omega \times [0, T], \mathcal{P})$ into $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ which satisfies

$$\begin{aligned}
&K_{\omega,t}(\{0\}) = 0, \qquad \int K_{\omega,t}(dx)\left(|x|^2 \wedge 1\right) \leq 1,\\
&\Delta A_t(\omega) > 0 \Rightarrow b_t(\omega) = \int K_{\omega,t}(dx) x 1_{\{|x \leq 1|\}},\\
&\Delta A_t(\omega) K_{\omega,t}(\mathbb{R}^N) \leq 1.
\end{aligned}$$

APPENDIX B: SEMIMARTINGALE CHARACTERISTICS UNDER THE PRICING MEASURE

Now, given the measure Q^* (see Assumption 4.4), Girsanov's Theorem for semimartingales (theorem III-3.24 in Jacod and Shiryaev 2003) implies that there exists a nonnegative $\tilde{\mathcal{P}} := \mathcal{P} \otimes \mathcal{B}(\mathbb{R}^N)$-measurable function Y^* (where $\mathcal{B}(\mathbb{R}^N)$ is the Borel sigma-algebra on $\mathbb{R}^N$ and $\otimes$ denotes the product sigma-algebra) and a predictable process β^* satisfying

$$\begin{aligned}
&|x 1_{\{|x| \leq 1\}}(Y^* - 1)| * \nu_t < \infty \ Q^*\text{-almost surely for all } t \in [0, T],\\
&\left|\sum_{j \leq N} c^{ij} \beta^{*,j}\right| \cdot A_t < \infty \quad \text{and} \quad \left(\sum_{j,k \leq N} \beta^{*,j} c^{jk} \beta^{*,k}\right) \cdot\\
&\quad A_t < \infty \ Q\text{-almost surely for all } i \text{ and } t \in [0, T],\\
&\nu(\omega; \{t\} \times E) = 1 \Rightarrow \int Y^*(\omega, t, x) \nu(\omega; \{t\} \times dx) = 1,
\end{aligned}$$

and such that the characteristics of S relative to Q^* are

$$\begin{aligned}
\text{(B.1)} \qquad B^* &= B + c\beta^* + x 1_{\{|x| \leq 1\}}(Y^* - 1) * \nu,\\
C^* &= C,\\
\nu^* &= Y^* \cdot \nu, \text{ where } Y^* \cdot \nu(\omega; dt, dx) = \nu(\omega; dt, dx) Y^*(\omega, t, x).
\end{aligned}$$

Furthermore, according to lemma III-5.17 in Jacod and Shiryaev (2003), the density process Z^* of Q^* relative to P has the form

$$Z^* = 1 + (Z^*_- \beta^*) \cdot S^c + Z^*_- \left(Y^* - 1 + \frac{\hat{Y}^* - a}{1 - a} 1_{\{a < 1\}}\right) * (\mu^S - \nu) + Z',$$

where

(i) Z' is a P-local martingale with $Z'_0 = 0$ and $[(Z'^c, (S^i)^c] = 0$ for all $i \leq N$ and $M^P_{\mu^S}[\Delta Z' \mid \tilde{\mathcal{P}}] = 0$ (see III-3.15 in Jacod and Shiryaev 2003),

(ii)

$$a_t(\omega) = \nu(\omega; \{t\} \times \mathbb{R}^N),$$

(iii)

$$\hat{Y}^*{}_t(\omega) = \begin{cases} \displaystyle\int \nu(\omega; \{t\} \times dx) Y^*(\omega, t, x) & \text{if this integral converges,} \\ \infty & \text{otherwise.} \end{cases}$$

REFERENCES

AMIN, K. I., and R. JARROW (1992): Pricing Options on Risky Assets in a Stochastic Interest Rate Economy, *Math. Finance* 2(4), 217–237.

ANSEL, J. P., and C. STRICKER (1994): Couverture des Actifs Contingents et Prix Maximum, *Ann. de l'Institut Henri Poincaré (B) Probab. Stat.* 30(2), 303–315.

BEBER, A., and M. PAGANO (2013): Short-Selling Bans around the World: Evidence from the 2007-09 Crisis, *J. Finance* 68, 343–381.

BOEHMER, E., C. M. JONES, and X. ZHANG (2011): Shackling Short Sellers: The 2008 Shorting Ban, Working Paper, Available at SSRN: http://ssrn.com/abstract=1412844.

BOULTON, T. J., and M. V. BRAGA ALVES (2010): The Skinny on the 2008 Naked Short-Sale Restrictions, *J. Finan. Markets* 13(4), 397–421.

BRIS, A., W. N. GOETZMANN, and N. ZHU (2007). Efficiency and the Bear: Short Sales and Markets around the World, *J. Finance* 62(3), 1029–1079.

BROWN, A. A., and L. C. G. ROGERS (2012): Diverse Beliefs, *Stoch. Int. J. Probab. Stoch. Process*. Forthcoming.

CHAROENROOK, A., and H. DAOUK (2005): A Study of Market-wide Short-selling Restrictions, Working Paper, Available at: SSRN: http://ssrn.com/abstract=687562.

CHESNEY, M., M. JEANBLANC, and M. YOR (2009): *Mathematical Methods for Financial Markets*, London: Springer-Verlag.

COX, A., and D. HOBSON (2005): Local Martingales, Bubbles and Option Prices, *Finance Stoch.* 9(4), 477–492.

CVITANIĆ, J., and I. KARATZAS (1993): Hedging Contingent Claims with Constrained Portfolios, *Ann. Appl. Probab.* 3(3), 652–681.

DELBAEN, F., and W. SCHACHERMAYER (1994): A General Version of the Fundamental Theorem of Asset Pricing, *Math. Ann.* 300(1), 463–520.

DELBAEN, F., and W. SCHACHERMAYER (1995): Arbitrage Possibilities in Bessel Processes and Their Relations to Local Martingales, *Probab. Theory Relat. Fields* 102(3), 357–366.

DUFFIE, D. (1989): *Futures Markets*, Englewood Cliffs, NJ: Prentice Hall.

EKSTRÖM, E., and J. TYSK (2009): Bubbles, Convexity and the Black-Scholes Equation, *Ann. Appl. Probab.* 19(4), 1369–1384.

FABOZZI, F. J. (2004): *Short Selling: Strategies, Risks, and Rewards*, The Frank J. Fabozzi Series, Hoboken, NJ: John Wiley & Sons, Inc.

FRITTELLI, M. (1997): Semimartingales and Asset Pricing under Constraints, in *Mathematics of Derivative Securities (Publications of the Newton Institute)*, M. A. H. Dempster and S. R. Pliska, eds., Cambridge, UK: Cambridge University Press, pp. 265–277.

FUNG, J. K. W., and P. DRAPER (1999): Mispricing of Index Futures Contracts and Short Sales Constraints, *J. Futures Markets* 19(6), 695–715.

HARRISON, J. M., and S. R. PLISKA (1981): Martingales and Stochastic Integrals in the Theory of Continuous Trading, *Stoch. Process. Appl.* 11(3), 215–260.

HARRISON, J. M., and S. R. PLISKA (1983): A Stochastic Calculus Model of Continuous Trading: Complete Markets, *Stoch. Process. Appl.* 15(3), 313–316.

JACOD, J. (1979): *Calcul stochastique et problèmes de martingales*, Vol. 714 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer.

JACOD, J. (1980): Intégrales Stochastiques par Rapport à une Semimartingale Vectorielle et Changements de Filtration, *Séminaire de Probabilités de Strasbourg* 14, 161–172.

JACOD, J., and A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes*, 2nd ed., Berlin, Heidelberg: Springer-Verlag.

JARROW, R., and P. PROTTER (2009): Forward and Futures Prices with Bubbles, *Int. J. Theor. Appl. Finance* 12(7), 901–924.

JARROW, R., P. PROTTER, and K. SHIMBO (2007): Asset Price Bubbles in a Complete Market, in *Advances in Mathematical Finance*, *Applied and Numerical Harmonic Analysis*, M. C. Fu, R. A. Jarrow, J-Y. J. Yen, and R. J. Elliott, eds., Boston: Birkhäuser, pp. 97–121.

JARROW, R., P. PROTTER, and K. SHIMBO (2010): Asset Price Bubbles in Incomplete Markets, *Math. Finance* 20(2), 145–185.

JOUINI, E., and H. KALLAL (1995): Arbitrage in Securities Markets with Short-Sales Constraints, *Math. Finance* 5(3), 197–232.

KARATZAS, I., and C. KARDARAS (2007): The Numéraire Portfolio in Semimartingale Financial Models, *Finance Stoch.* 11(4), 447–493.

KARATZAS, I., and D. OCONE (1991): A Generalized Clark Representation Formula, with Applications to Optimal Portfolios, *Stochastics* 37, 187–220.

KARATZAS, I., and S. E. SHREVE (1998a): *Brownian Motion and Stochastic Calculus*, 2nd ed., New York: Springer.

KARATZAS, I., and S. E. SHREVE (1998b): *Methods of Mathematical Finance*, New York: Springer-Verlag.

KCHIA, Y., and P. PROTTER (2011): On Progressive Filtration Expansion with a Process, Preprint, arXiv:1105.1662v1.

MADAN, D. B., and M. YOR (2006): Itô's Integrated Formula for Strict Local Martingales. in *Memoriam Paul-André Meyer*, Vol. 1874 of *Lecture Notes in Mathematics*, M. Emery and M. Yor, eds., Berlin, Heidelberg: Springer, pp. 157–170.

MIJATOVIĆ, A., and M. URUSOV (2012): On the Martingale Property of Certain Local Martingales, *Probab. Theory Relat. Fields* 152(1–2), 1–30.

MILLER, E. M. (1977): Risk, Uncertainty and Divergence of Opinion, *J. Finance* 32(4), 1151–1168.

NAPP, C. (2003): The Dalang–Morton–Willinger Theorem under Cone Constraints, *J. Math. Econ.* 39(1-2), 111–126.

PAL, S., and P. PROTTER (2010): Analysis of Strict Local Martingales via h-Transforms, *Stoch. Process. Appl.* 120(8), 1424–1443.

PHAM, H., and N. TOUZI (1999): The Fundamental Theorem of Asset Pricing with Cone Constraints, *J. Math. Econ.* 31(2), 265–279.

PROTTER, P. (2005): *Stochastic Integration and Differential Equations*, 2nd ed., Corrected 3rd printing, Berlin, Heidelberg: Springer.

PULIDO, S. (2011): The Fundamental Theorem of Asset Pricing, the Hedging Problem and Maximal Claims in Financial Markets with Short Sales Prohibitions, Preprint, arXiv:1012.3102v4.

REVUZ, D., and M. YOR (1999): *Continuous Martingales and Brownian Motion*, 3rd ed., Berlin, Heidelberg: Springer.

RUF, J. (2012): Hedging under Arbitrage, *Math. Finance*, doi: 10.1111/j.1467-9965.2011.00502.x.

SCHÜRGER, K. (1996): On the Existence of Equivalent τ-Measures in Finite Discrete Time, *Stoch. Process. Appl.* 61(1), 109–128.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 339–370

DUAL REPRESENTATIONS FOR GENERAL MULTIPLE STOPPING PROBLEMS

CHRISTIAN BENDER

Universität des Saarlandes

JOHN SCHOENMAKERS

Weierstrass Institute for Applied Analysis and Stochastics

JIANING ZHANG

Weierstrass Institute for Applied Analysis and Stochastics

In this paper, we study the dual representation for generalized multiple stopping problems, hence the pricing problem of general multiple exercise options. We derive a dual representation which allows for cash flows which are subject to volume constraints modeled by integer-valued adapted processes and refraction periods modeled by stopping times. As such, this extends the works by Schoenmakers (2012), Bender (2011a), Bender (2011b), Aleksandrov and Hambly (2010), and Meinshausen and Hambly (2004) on multiple exercise options, which either take into consideration a refraction period or volume constraints, but not both simultaneously. We also allow more flexible cash flow structures than the additive structure in the above references. For example, some exponential utility problems are covered by our setting. We supplement the theoretical results with an explicit Monte Carlo algorithm for constructing confidence intervals for prices of multiple exercise options and illustrate it with a numerical study on the pricing of a swing option in an electricity market.

KEY WORDS: general multiple stopping, dual representations, multiple exercise options, volume constraints, refraction period.

1. INTRODUCTION

The last decades have seen ground breaking developments of Monte Carlo methods for American options based on multidimensional underlying price processes. In the late 1990s the regression-based methods by Carriere (1996), Longstaff and Schwartz (2001), and Tsitsiklis and Van Roy (2001) may be considered as main breakthroughs. In general, these methods provide lower bounds on the option price by constructing an approximation to the optimal exercise (stopping) time via regression on a set of basis functions. As such these approaches are termed "primal." At the beginning of this century Rogers (2002) and independently Haugh and Kogan (2004) provided the next breakthrough by presenting a "dual" representation for the optimal stopping problem corresponding to the American option pricing problem. In this representation the option price is expressed as infimum of an expectation over a set of martingales. (The key behind this dual representation can

Manuscript received December 2011; final revision received October 2012.
Address correspondence to John Schoenmakers, Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany; e-mail: schoenma@wias-berlin.de.

DOI: 10.1111/mafi.12030

already be found in Davis and Karatzas 1994, in fact.) Although in the primal methods the central problem is to find a "good" stopping time, in the dual problem one needs to find a "good" martingale, which leads to an upper bound for the price of an American option. As one of the standard numerical approaches to compute dual upper bounds for American options by Monte Carlo we refer to Andersen and Broadie (2004).

During the same time, in the emerging electricity markets products with multiple exercise opportunities, such as "swing options," became popular. Naturally, the pricing of such a product leads to a multiple stopping problem, which requires a numerical method for resolution. In this respect, generalization of the existing primal regression methods for standard optimal stopping was just a matter of routine. Further, Bender and Schoenmakers (2006) developed a kind of policy iteration for multiple stopping. However, regarding the dual approach the situation was less clear. Meinshausen and Hambly (2004) proposed a dual representation for the multiple stopping problem by expressing the excess value due to each additional exercise right by an infimum of an expectation over a set of martingales and a set of stopping times. This line of research was carried out further by Aleksandrov and Hambly (2010) and Bender (2011b) in the context of dual pricing of multiexercise options under volume constraints. Recently, Schoenmakers (2012) introduced a dual representation for the price of a multiple exercise option instead of the dual representation for the excess value of an additional right. This new dual representation involves an infimum over martingales only and can thus be considered as a more natural extension of the dual representation for single exercise options. This approach was generalized by Bender (2011a) to a continuous time setting involving (constant) refraction periods.

In the meantime Kobylanski, Quenez, and Rouy-Mironescu (2011) introduced and studied multiple stopping problems in the primal sense in a far more general context, where the payoff is considered to be some abstract functional of an (ordered) sequence of stopping times. The goal of this paper is to find (pure) martingale dual representations for such generalized multiple stopping problems in a discrete time setting. As we will show, such representations can be constructed even in a most general setting. However, for practical implementation these general representations unfold their full strength only if applied to some more specifically structured cash flows. In this respect, we study a generic payoff structure with both multiplicative and additive structure that incorporates (integer-valued) volume constraints and refraction periods given by stopping times. We furthermore provide an explicit Monte Carlo-based algorithm and carry out a detailed numerical study illustrating the pricing of swing options. Comparing to existing works, the numerical experiments reveal that, by and large, the dual algorithms due to our new representations applied to the problem type considered in Aleksandrov and Hambly (2010) and Bender (2011b) produce tighter upper bounds on the option price, in particular when the number of exercise rights is large. We moreover present a numerical example which involves swing options subject to both volume constraints and refraction periods and give tight confidence intervals for the respective option prices. We underline that the latter example is not covered by the numerical dual methods presented in the literature so far.

It should be mentioned that Brown et al. (2010) developed a dual approach for general optimal control problems via the concept of information relaxation and that Rogers (2007) presented a path-wise dual approach for a specific class of Markovian control problems. A closer investigation reveals that, in fact, the dual representation of Brown et al. (2010) is essentially equivalent to Theorem 2.4 in this paper. However, Theorem 2.4 is derived in a different way and in a different language, namely by directly generalizing the standard dual martingale approach of Rogers (2002), and Haugh and Kogan (2004) (cf. Schoenmakers 2012). Furthermore, to end up with feasible algorithms, the complexity has to be reduced. In this respect, we present a systematic treatment for a class of problems which generalizes those studied previously in the literature. Finally, we note that (virtually simultaneously with the present work) Gyurko, Hambly, and Witte (2011) developed a dual formulation for a specific class of control problems via an approach in the spirit of Rogers (2007).

The structure of the paper is as follows: In Section 2 we derive a dual representation for general multiple stopping problems in terms of a family of martingales. As this family of martingales is typically too large for practical purposes in general, we specialize to a generic cash flow with additive and multiplicative structure which incorporates volume constraints and refraction periods in Section 3. In this section we then prove two dual representations. One is in terms of the Doob decomposition of the Snell envelopes of an auxiliary family of stopping problems, the other one only requires approximations of these Snell envelopes. In Section 4 we explain how to build a Monte Carlo algorithm for computing confidence intervals on the value of the multiple stopping problems based on the results of Section 3 and perform some numerical experiments in the context of swing option pricing.

2. GENERAL MULTIPLE STOPPING PROBLEM

In this section we consider a multiple stopping problem in discrete time $i = 0, \dots, T$, where $T \in \mathbb{N}$ is a fixed and finite time horizon. We further introduce a "cemetery time" $\partial := T + 1$, where all rights will be exercised, which are not exercised up to time T. For a given filtration $(\mathcal{F}_i)_{0 \le i \le \partial}$ and a number L of exercise dates we next consider a cash flow X as a map $X : \{0, \dots, T, \partial\}^L \times \Omega \to \mathbb{R}$ which satisfies for all $0 \le i_1 \le \cdots \le i_L \le \partial$,

$$X_{i_1,\dots,i_L} \text{ is } \mathcal{F}_{i_L}\text{-measurable},$$
$$\mathbb{E}|X_{i_1,\dots,i_L}| < \infty.$$

Now consider the stopping problem (sup:=ess.sup, $\mathbb{E}_i := \mathbb{E}_{\mathcal{F}_i}$),

$$Y_i^{*L} = \sup_{i \le \tau^1 \le \cdots \le \tau^L} \mathbb{E}_i \, X_{\tau^1,\dots,\tau^L}, \tag{2.1}$$

where the supremum runs over the families of ordered stopping times τ^k, $1 \le k \le L$.

In mathematical finance this type of multiple stopping problems has the following two generic applications:

- Suppose that X is the revenue or, more generally, the utility of an investor's revenue depending on a set of exercise decisions. An investor, who aims at maximizing his expected utility by using his exercise rights in an optimal way, is faced with the optimal multiple stopping problem (2.1), where the expectation is taken under the physical measure. Examples 3.2 and 3.3 below on exponential utility and on portfolio liquidation are of this type.
- Suppose that X is the discounted payoff of an option which allows for multiple exercises. Then, similarly to the American option case, an arbitrage-free discounted price for the multiple exercise option is given by the multiple stopping problem (2.1), if the expectation is taken under a pricing measure, i.e., a probability measure under which the discounted prices of tradable and storable basic securities in the underlying

market are (local) martingales. This price depends, of course, on the choice of the pricing measure in the situation of an incomplete market, which can be selected by calibration to liquidly traded options in practice for instance. In Example 3.1 below and in the numerical section we consider the pricing of swing options.

To reduce problem (2.1) to a sequence of single stopping problems let us define for $k = 2, \dots, L$ and $0 \le j_1 \dots \le j_{k-1} \le r \le \partial$,

$$Y_r^{*L-k+1, j_1, \dots, j_{k-1}} := \sup_{r \le \tau^k \le \dots \le \tau^L} \mathbb{E}_r\, X_{j_1, \dots, j_{k-1}, \tau^k, \dots, \tau^L}, \tag{2.2}$$

with the convention that for $k = 1$, we put $Y_r^{*L, \varnothing} := Y_r^{*L}$, and for $k = L + 1$, we put $Y_r^{*0, j_1, \dots, j_L} = X_{j_1, \dots, j_L}$.

PROPOSITION 2.1. *We have the following reduction principle*

$$Y_r^{*L-k+1, j_1, \dots, j_{k-1}} = \sup_{\tau \ge r} \mathbb{E}_r\, Y_\tau^{*L-k, j_1, \dots, j_{k-1}, \tau}, \quad r \ge j_{k-1}. \tag{2.3}$$

Proof. This principle can be straightforwardly proved in an inductive manner, but it can also be considered as a discrete time version of a related result in a continuous time setting from Kobylanski et al. (2011). □

REMARK 2.2. Thanks to the previous proposition and standard results on single stopping problems, we observe that a family of optimal stopping times for (2.1) is given recursively by

$$\begin{aligned} \tau_i^{*1} &= \inf\left\{r \ge i;\ Y_r^{*L-1, r} \ge \mathbb{E}_r\, Y_{r+1}^{*L}\right\}, \\ \tau_i^{*k} &= \inf\left\{r \ge \tau_i^{*k-1};\ Y_r^{*L-k, \tau_i^{*1}, \dots, \tau_i^{*k-1}, r} \ge \mathbb{E}_r\, Y_{r+1}^{*L-k+1, \tau_i^{*1}, \dots, \tau_i^{*k-1}}\right\}, \quad k = 2, \dots, L. \end{aligned}$$

In what follows the following remark turns out to be useful.

REMARK 2.3. We say that a martingale $(M_r)_{r \ge p}$ is a Doob martingale of $(Y_r)_{r \ge p}$, whenever there exists a predictable process $(A_r)_{r \ge p}$, such that $Y_r - M_r + A_r$ is $\mathcal{F}_p$-measurable for any $r \ge p$. In particular, for any two Doob martingales $(M_r)_{r \ge p}$ and $(\widetilde{M}_r)_{r \ge p}$ of $(Y_r)_{r \ge p}$, it holds

$$M_r - M_{r'} = \widetilde{M}_r - \widetilde{M}_{r'} = \sum_{k=r'}^{r-1} (Y_{k+1} - \mathbb{E}_k\, Y_{k+1})$$

for any $r \ge r' \ge p$.

We can now state and prove a dual representation for the general multiple stopping problem in terms of martingales.

THEOREM 2.4 [Dual representation]. *In the setting described above, we have that:*

(i) For any $0 \le i \le \partial$ *and any set of martingales* $(M_r^{L-k+1, j_1, \dots, j_{k-1}})_{r \ge j_{k-1}}$, *where* $1 \le k \le L$, *and* $i =: j_0 \le j_1 \le \dots \le j_{k-1}$, *it holds*

$$Y_i^{*L} \le \mathbb{E}_i \max_{i \le j_1 \le \dots \le j_L \le \partial} \left(X_{j_1, \dots, j_L} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{L-k+1, j_1, \dots, j_{k-1}} - M_{j_k}^{L-k+1, j_1, \dots, j_{k-1}} \right) \right). \tag{2.4}$$

(ii) It holds for $i \geq 0$

$$Y_i^{*L} = \max_{i \leq j_1 \leq \cdots \leq j_L \leq \partial} \left(X_{j_1,\dots,j_L} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{*L-k+1, j_1,\dots,j_{k-1}} - M_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \right),$$

where for $1 \leq k \leq L$, and $i =: j_0 \leq j_1 \leq \cdots \leq j_{k-1}$,$(M_r^{*L-k+1, j_1,\dots,j_{k-1}})_{r \geq j_{k-1}}$ is a Doob martingale of $(Y_r^{*L-k+1, j_1,\dots,j_{k-1}})_{r \geq j_{k-1}}$.

Proof.

(i) For the martingale family as stated we have for any chain of stopping times $0 \leq \tau^1 \leq \cdots \leq \tau^L \leq \partial$,

$$\begin{aligned} &\mathbb{E}_i \sum_{k=1}^{L} \left(M_{\tau^{k-1}}^{L-k+1, \tau^1,\dots,\tau^{k-1}} - M_{\tau^k}^{L-k+1, \tau^1,\dots,\tau^{k-1}} \right) \\ &\quad = \sum_{k=1}^{L} \mathbb{E}_i \mathbb{E}_{\tau^{k-1}} \left(M_{\tau^{k-1}}^{L-k+1, \tau^1,\dots,\tau^{k-1}} - M_{\tau^k}^{L-k+1, \tau^1,\dots,\tau^{k-1}} \right) = 0, \end{aligned}$$

hence,

$$Y_i^{*L} = \sup_{i \leq \tau^1 \leq \cdots \leq \tau^L} \mathbb{E}_i \left(X_{\tau^1,\dots,\tau^L} + \sum_{k=1}^{L} \left(M_{\tau^{k-1}}^{L-k+1, \tau^1,\dots,\tau^{k-1}} - M_{\tau^k}^{L-k+1, \tau^1,\dots,\tau^{k-1}} \right) \right)$$

from which (i) follows directly.

(ii) For any chain $i \leq j_1 \leq \cdots \leq j_L \leq \partial$ we may write (recalling $j_0 := i$)

$$\begin{aligned} (2.5) \qquad & X_{j_1,\dots,j_L} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{*L-k+1, j_1,\dots,j_{k-1}} - M_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \\ &= X_{j_1,\dots,j_L} + \sum_{k=1}^{L} \left(Y_{j_{k-1}}^{*L-k+1, j_1,\dots,j_{k-1}} - Y_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \\ &\quad + \sum_{k=1}^{L} \sum_{l=j_{k-1}}^{j_k - 1} \left(\mathbb{E}_l Y_{l+1}^{*L-k+1, j_1,\dots,j_{k-1}} - Y_l^{*L-k+1, j_1,\dots,j_{k-1}} \right) \\ &= Y_i^{*L} + \sum_{k=1}^{L} \left(Y_{j_k}^{*L-k, j_1,\dots,j_k} - Y_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \\ &\quad + \sum_{k=1}^{L} \sum_{l=j_{k-1}}^{j_k - 1} \left(\mathbb{E}_l Y_{l+1}^{*L-k+1, j_1,\dots,j_{k-1}} - Y_l^{*L-k+1, j_1,\dots,j_{k-1}} \right). \end{aligned}$$

By the reduction principle (2.3) it follows that $(Y_r^{*L-k+1, j_1,\dots,j_{k-1}})_{r \geq j_{k-1}}$ is a supermartingale which dominates the (virtual) cash flow $Y_r^{*L-k, j_1,\dots,j_{k-1},r}$ for $k = 1, \dots, L$. Hence, expression (2.5) is less than or equal to Y_i^{*L}. It thus follows that

$$\max_{i \leq j_1 \leq \cdots \leq j_L \leq \partial} \left(X_{j_1,\dots,j_L} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{*L-k+1, j_1,\dots,j_{k-1}} - M_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \right) \leq Y_i^{*L},$$

and then, an application of (i) finishes the proof. □

A straightforward consequence of Theorem 2.4 is the following dual representation in terms of approximate Snell envelopes.

COROLLARY 2.5. *For any set* $(Y_r^{L-k+1,j_1,\dots,j_{k-1}})_{r\geq j_{k-1}}$ *of approximations to the Snell envelopes* $\big(Y_r^{*L-k+1,j_1,\dots,j_{k-1}}\big)_{r\geq j_{k-1}}$ *with* $Y_{j_L}^{0,j_1,\dots,j_L} := X_{j_1,\dots,j_L}$, *it holds for* $i\geq 0$:

$$
\begin{aligned}
(2.6)\qquad Y_i^{*L} \leq Y_i^L + \mathbb{E}_i \max_{i\leq j_1\leq\cdots\leq j_L\leq\partial} \sum_{k=1}^{L}\Bigg(& Y_{j_k}^{L-k,j_1,\dots,j_k} - Y_{j_k}^{L-k+1,j_1,\dots,j_{k-1}} + \sum_{l=j_{k-1}}^{j_k-1} \\
& \big(\mathbb{E}_l Y_{l+1}^{L-k+1,j_1,\dots,j_{k-1}} - Y_l^{L-k+1,j_1,\dots,j_{k-1}}\big)\Bigg).
\end{aligned}
$$

Equality holds when the Snell envelopes are plugged in.

Proof. Given $(Y_r^{L-k+1,j_1,\dots,j_{k-1}})_{r\geq j_{k-1}}$, we denote a corresponding family of Doob martingales by $(M_r^{L-k+1,j_1,\dots,j_{k-1}})_{r\geq j_{k-1}}$. Following the same manipulations as in (2.5) and recalling that by definition $Y_{j_L}^{0,j_1,\dots,j_L} = X_{j_1,\dots,j_L}$, we get

$$
\begin{aligned}
Y_i^L + \sum_{k=1}^{L}\Bigg(& \big(Y_{j_k}^{L-k,j_1,\dots,j_k} - Y_{j_k}^{L-k+1,j_1,\dots,j_{k-1}}\big) \\
& + \sum_{l=j_{k-1}}^{j_k-1}\big(\mathbb{E}_l Y_{l+1}^{L-k+1,j_1,\dots,j_{k-1}} - Y_l^{L-k+1,j_1,\dots,j_{k-1}}\big)\Bigg) \\
= X_{j_1,\dots,j_L} + & \sum_{k=1}^{L}\big(M_{j_{k-1}}^{L-k+1,j_1,\dots,j_{k-1}} - M_{j_k}^{L-k+1,j_1,\dots,j_{k-1}}\big).
\end{aligned}
$$

Hence, the assertion is a mere reformulation of Theorem 2.4. □

At this point, we stress that the dual representation from Theorem 2.4 relies on families of martingales $(M_r^{L-k+1,j_1,\dots,j_{k-1}})_{r\geq j_{k-1}}$ whose size is parameterized via the $(k-1)$-tuples $(j_1,\dots,j_{k-1})$, $k=1,\dots,L$. Hence, depending on the time horizon T and the number of exercise rights L a huge number of martingales $M^{L-k+1,j_1,\dots,j_{k-1}}$, $k=1,\dots,L$, $0\leq j_1\leq \dots\leq j_{k-1}\leq\partial = T+1$ is required to compute an upper price bound by the above dual formulation. It is thus of great importance to single out situations, in which a family of optimal martingales can be constructed from a much smaller family of auxiliary processes. This will be the topic of Section 3. A motivating example in this respect is the standard multiple stopping problem.

EXAMPLE 2.6 [Standard multiple stopping]. Let Z be a nonnegative adapted process with $Z_j = 0$ for $j=\partial$, i.e., no penalty is imposed for unexercised rights. The standard multiple stopping problem is to maximize $\mathbb{E}[\sum_{k=1}^{L} Z_{\tau^k}]$ over the set of ordered stopping times $\tau^1\leq\cdots\leq\tau^L$ such that $\tau^k<\tau^{k+1}$ or $\tau^k=\tau^{k+1}=\partial$. This means that at most one right can be exercised per day, but an arbitrary number of rights can be left unexercised, i.e., is exercised at time ∂. This problem can be put into our general setting by considering the cash flow

$$X_{i_1,\dots,i_L} = \begin{cases} \sum_{k=1}^{L} Z_{i_k}, & \text{if } i_{j+1} = i_j \Rightarrow i_j = \partial, \\ -N, & \text{else,} \end{cases}$$

for $N \in \mathbb{N}$. Note that the Snell envelope Y_i^{*L} does not depend on the choice of N, because it is never optimal to exercise X in a way which gives a negative payment. Hence, letting N tend to ∞, Theorem 2.4 yields,

$$Y_i^{*L} = \max_{\substack{i \le j_1 \le \cdots \le j_L \\ j_k = j_{k+1} \Rightarrow j_k = \partial}} \left(\sum_{k=1}^{L} Z_{j_k} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{*L-k+1, j_1,\dots,j_{k-1}} - M_{j_k}^{*L-k+1, j_1,\dots,j_{k-1}} \right) \right),$$

where for $1 \le k \le L$, $i \le j_1 < \cdots < j_{k-1}$,$(M_r^{*L-k+1, j_1,\dots,j_{k-1}})_{r \ge j_{k-1}}$ is a Doob martingale of

$$Y_r^{*L-k+1, j_1,\dots,j_{k-1}} = \sum_{p=1}^{k-1} Z_{j_p} + \sup_{\substack{r \le \tau^k \le \cdots \le \tau^L \\ \tau^p = \tau^{p+1} \Rightarrow \tau^p = \partial \text{ and } \tau^k = j_{k-1} \Rightarrow j_{k-1} = \partial}} \mathbb{E}_r \sum_{p=k}^{L} Z_{\tau^p}$$

for $r \ge j_{k-1}$. Define, for $r \ge 0$,

$$Y_r^{*L-k+1} := \sup_{\substack{r \le \tau^k \cdots \le \tau^L \\ \tau^p = \tau^{p+1} \Rightarrow \tau^p = \partial}} \mathbb{E}_r \left(\sum_{p=k}^{L} Z_{\tau^p} \right)$$

and denote the Doob martingale of $(Y_r^{*L-k+1})_{r \ge 0}$ by $(M_r^{*L-k+1})_{r \ge 0}$. As

$$Y_r^{*L-k+1} - Y_r^{*L-k+1, j_1,\dots,j_{k-1}}$$

is $\mathcal{F}_{j_{k-1}}$-measurable for $r \ge j_{k-1}$, we can conclude by Remark 2.3 that $(M_r^{*L-k+1})_{r \ge j_{k-1}}$ is a Doob martingale of $(Y_r^{*L-k+1, j_1,\dots,j_{k-1}})_{r \ge j_{k-1}}$. Hence, we end up with the dual representation

$$Y_i^{*L} = \max_{\substack{i \le j_1 \le \cdots \le j_L \\ j_k = j_{k+1} \Rightarrow j_k = \partial}} \left(\sum_{k=1}^{L} Z_{j_k} + \sum_{k=1}^{L} \left(M_{j_{k-1}}^{*L-k+1} - M_{j_k}^{*L-k+1} \right) \right)$$

of Schoenmakers (2012). Here, the potentially large family of optimal martingales $(M^{*L-k+1, j_1,\dots,j_{k-1}})$, $k = 1, \dots, L$, $0 \le j_1, \dots \le j_{k-1} \le \partial$, collapses, in fact, to a family of L martingales, namely the Doob martingales of Y^{*k}.

3. GENERIC CASH FLOW WITH ADDITIVE AND MULTIPLICATIVE STRUCTURE

We now introduce a generic cash flow structure for which the dual representation simplifies in a similar way than for the standard multiple stopping problem in Example 2.6. To this end, let us consider for each $k = 1, \dots, L$ and $l = 1, \dots, L-1$ two adapted processes

U^k and V^l. We define a "precash flow"

$$\widetilde{X}_{j_1,\ldots,j_L} = \sum_{k=1}^{L} U^k_{j_k} \prod_{l=1}^{k-1} V^l_{j_l},$$

which is assumed to satisfy $\widetilde{X}_{j_1,\ldots,j_L} > -N$ for some (possibly large) $N \in \mathbb{N}$. Concerning the processes U^k and V^l, we suppose that U^k_i is integrable for every $k = 1, \ldots, L$ and $i = 0, \ldots, \partial$, and that V^l_i is strictly positive and bounded from above for every $l = 1, \ldots, L-1$ and $i = 0, \ldots, \partial$. The multiple stopping problem which we have in mind is to optimally exercise this precash flow under some constraints on the set of admissible stopping times, which we now formulate. We first define an adapted volume constraint process v with values in $\{1, \ldots, L\}$ such that v_t is the maximum number of rights one may exercise at t, and such that $v_\partial = L$. To formalize this constraint, we introduce for $p \geq 1$ the mapping $\mathcal{E}_p$ which acts on a nondecreasing p-tuple $(j_1, \ldots, j_p)$ by

$$\mathcal{E}_p(j_1, \ldots, j_p) := \#\{r : 1 \leq r \leq p,\ j_r = j_p\}.$$

Hence, $\mathcal{E}_p$ denotes the number of rights exercised at j_p in the nondecreasing chain $0 \leq j_1 \leq \cdots \leq j_p \leq \partial$. Obviously, an ordered chain of stopping times $\tau^1 \leq \cdots \leq \tau^L$ satisfies the volume constraint if and only if $\mathcal{E}_p(\tau^1, \ldots, \tau^p) \leq v_{\tau^p}$ for every $p = 1, \ldots, L$. The second constraint, which we want to impose, is a refraction period which specifies the minimal waiting time between two exercises at different times. We admit random refraction periods, i.e., at each time i, $0 \leq i < \partial$, we fix a stopping time ρ^i taking values in $\{i + 1, \ldots, \partial\}$. If at least one right is exercised at time i, then the refraction period constraint imposes that the next right must either be exercised at the same time (if consistent with the volume constraint) or otherwise no earlier than ρ^i. A standard case is $\rho^i = (i + \delta) \wedge \partial$, where $1 \leq \delta \leq T$ is deterministic. Both constraints can be summarized by the binary $\mathcal{F}_{j_p}$-measurable random variable

$$\mathcal{C}_p(j_1, \ldots, j_p) := \begin{cases} 1, & \forall_{1 \leq l \leq p} : \mathcal{E}_l(j_1, \ldots, j_l) \leq v_{j_l} \text{ and } \forall_{1 \leq l \leq p} : j_l > j_{l-1} \Longrightarrow j_l \geq \rho^{j_{l-1}}, \\ 0, & \text{else,} \end{cases}$$

which is equal to 1, if and only if the constraints are satisfied when exercising at the p times $j_1 \leq \cdots \leq j_p$.

The dynamic multiple stopping problem which we now study is

$$Y^{*L}_i = \sup_{\substack{i \leq \tau^1 \leq \cdots \leq \tau^L \leq \partial \\ \mathcal{C}_L(\tau^1, \ldots, \tau^L) = 1}} \mathbb{E}_i \left[\sum_{k=1}^{L} U^k_{\tau^k} \prod_{l=1}^{k-1} V^l_{\tau^l}, \right], \tag{3.1}$$

i.e., the supremum is taken over all stopping times with values in $\{i, \ldots, T, \partial\}$ which satisfy the volume constraint and the refraction period constraint. This problem fits in our general (unconstrained) setting by considering the cash flow

$$X_{j_1,\ldots,j_L} = \begin{cases} \widetilde{X}_{j_1,\ldots,j_L}, & \text{if } \mathcal{C}_L(j_1, \ldots, j_L) = 1. \\ -N, & \text{else.} \end{cases} \tag{3.2}$$

To illustrate our motivation for studying the previous cash flow, let us have a look at the following examples.

EXAMPLE 3.1 [Swing options]. We extend the situation in Example 2.6 by imposing volume constraints and refraction periods as described above. Hence, we have

$$V_j^l := 1, l = 1, \dots, L-1, j = 0, \dots, \partial,$$

$$U_j^p := Z_j p = 1, \dots, L, j = 0, \dots, \partial,$$

where we recall that Z is a nonnegative adapted process with $Z_\partial = 0$. The multiple stopping problem then becomes

$$\sup_{\substack{\tau^1 \le \cdots \le \tau^L \\ \mathcal{C}_L(\tau^1, \dots, \tau^L) = 1}} \mathbb{E}\left[\sum_{k=1}^{L} Z_{\tau^k}\right],$$

leading to

$$X_{j_1,\dots,j_L} = \begin{cases} \sum_{k=1}^{L} Z_{j_k}, & \text{if } \mathcal{C}_L(j_1, , \dots, j_L) = 1, \\ -N, & \text{else.} \end{cases}$$

Here any $N \in \mathbb{N}$ can be chosen because Z is nonnegative. A dual approach for this multiple stopping problem was studied by Bender (2011b) and Aleksandrov and Hambly (2010) under volume constraints, but with unit refraction period, i.e., $\rho^i = i + 1$. The case with nontrivial constant refraction period is treated in Bender (2011a), but only under unit volume constraint, i.e., $v_i = 1$. A typical problem in the context of electricity markets which leads to this type of multiple stopping problem is the pricing of Swing option contracts, in which volume constraints and refraction periods are often imposed. This option pricing problem will be explained in more detail in our numerical study in Section 4.

EXAMPLE 3.2 [Exponential utility]. Under the assumptions of the previous example we can also maximize the exponential utility of exercising the cash flow Z_i L-times while obeying the constraints. Given the risk aversion parameter $\alpha \in (0, \infty)$, the corresponding multiple stopping problem becomes

$$\sup_{\substack{\tau^1 \le \cdots \le \tau^L \\ \mathcal{C}_L(\tau^1, \dots, \tau^L) = 1}} \mathbb{E}\left[-e^{-\alpha \sum_{k=1}^{L} Z_{\tau^k}}\right].$$

This problem fits in our setting by considering

$$X_{j_1,\dots,j_L} = \begin{cases} \sum_{k=1}^{L} U_{j_k}^k \prod_{l=1}^{k-1} V_{j_l}^l, & \text{if } \mathcal{C}_L(j_1, , \dots, j_L) = 1, \\ -N, & \text{else} \end{cases}$$

with

$$V_j^l := e^{-\alpha Z_j} > 0 \quad \text{and} \quad U_j^k := \begin{cases} 0, & \text{if } k = 1, \cdots, L-1, \\ -e^{-\alpha Z_j}, & \text{if } k = L, \end{cases}$$

for $j = 0, 1, \ldots, \partial$ and $N \geq 2$.

EXAMPLE 3.3 [Portfolio liquidation]. Suppose a (large) investor on a illiquid market wants to sell out (liquidate) L shares of a stock during the period $\{0, \ldots, T\}$. As such the investor is faced with a portfolio liquidation problem, see for example Almgrem and Chriss (2001), Gatheral and Schied (2011), Schied and Slynko (2011), and the references therein. Let us assume that $\widetilde{S}_j > 0, j = 0, \ldots, T$, is the virtual stock price process reflecting the stock price evolution in the absence of the large investor's trading. In the spirit of Schied and Slynko (2011), section 3.1, we model the price impact of the large investor by a resilience function G which we here apply to the log-price. Hence, the log-stock price $\ln S_{j_k}^{j_1,\ldots,j_{k-1}}$ at time j_k of the sale of the kth share, where $k-1$ shares were already sold at dates $0 \leq j_1 \leq \cdots \leq j_{k-1}$, is given by

$$\ln S_{j_k}^{j_1,\ldots,j_{k-1}} = \ln \widetilde{S}_{j_k} - \sum_{l=1}^{k-1} G(j_k - j_l).$$

We here choose the capped linear resilience function $G(t) = b(1 - at)_+$ for constants a, $b > 0$. Assuming a short time horizon $T \leq 1/a$, the investor is thus faced with a multiple stopping problem

$$\sup_{\tau^1 \leq \cdots \leq \tau^L} \mathbb{E}\left[\sum_{k=1}^{L} S_{\tau^k}^{\tau^1,\ldots,\tau^{k-1}}\right],$$

which fits in our framework by applying, for $0 \leq j_1 \leq \cdots \leq j_L \leq T$, the cash flow

$$\begin{aligned} X_{j_1,\ldots,j_L} &:= \sum_{k=1}^{L} S_{j_k}^{j_1,\ldots,j_{k-1}} = \sum_{k=1}^{L} \widetilde{S}_{j_k} \exp\left(-\sum_{l=1}^{k-1} b(1 - a(j_k - j_l))\right) \\ &= \sum_{k=1}^{L} \widetilde{S}_{j_k} \exp[b(a\, j_k - 1)(k-1)] \prod_{l=1}^{k-1} \exp(-abj_l) \\ &= \sum_{k=1}^{L} U_{j_k}^k \prod_{l=1}^{k-1} V_{j_l}^l \end{aligned}$$

with

$$U_j^k := \widetilde{S}_j \exp[b(aj - 1)(k-1)] \quad \text{and} \quad V_j^l := \exp(-abj).$$

(Note that the cemetery time ∂ is irrelevant in this setting and we can, e.g., set $U_\partial^k = 0$, $V_\partial^l = 1$ to make sure that it is never optimal to exercise at this time.)

Similarly to the situation in Example 2.6, we now introduce a family of auxiliary multiple stopping problems Y_r^{*L-k+1}, which are not parameterized by the times $j_1, \ldots,$ j_{k-1}, at which the first rights were exercised. We will then show that a family of optimal martingales for the original multiple stopping problem (3.1) can be constructed via the Doob decomposition of the auxiliary problems. This then leads to a simplified dual representation for (3.1), which can be implemented in practice even when the maturity T and the number of rights L are large.

Define

$$Y_r^{*L-k+1} := \sup_{\substack{\tau^k,\ldots,\tau^L \\ r\le\tau^k\le\cdots\le\tau^L \text{ and } \mathcal{C}_{L-k+1}(\tau^k,\cdots,\tau^L)=1}} \mathbb{E}_r\left(\sum_{p=k}^{L} U_{\tau^p}^p \prod_{l=k}^{p-1} V_{\tau^l}^l\right) \tag{3.3}$$

with the convention $Y_r^{*0} := 0$. The following proposition states the Bellman principle for this multiple stopping problem.

PROPOSITION 3.4 [Dynamic program]. *For $r \ge 0$ and $1 \le k \le L$ we have,*

$$Y_r^{*L-k+1} = \max\left(\mathbb{E}_r\, Y_{r+1}^{*L-k+1}, \max_{1\le n\le v_r\wedge(L-k+1)}\left(\sum_{p=k}^{k+n-1} U_r^p \prod_{l=k}^{p-1} V_r^l + \prod_{l=k}^{k+n-1} V_r^l \mathbb{E}_r\, Y_{\rho^r}^{*L-k-n+1}\right)\right).$$

Proof. From (3.3) we derive straightforwardly,

$$\begin{aligned}
Y_r^{*L-k+1} &= \max_{0\le n\le v_r\wedge(L-k+1)} \sup_{\substack{r<\tau^{k+n}\le\cdots\le\tau^L \\ \mathcal{C}_{L-k+1}(r,\ldots,r,\tau^{k+n},\ldots,\tau^L)=1}} \\
&\quad\times \mathbb{E}_r\left(\sum_{p=k}^{k+n-1} U_r^p \prod_{l=k}^{p-1} V_r^l + \sum_{p=k+n}^{L} U_{\tau^p}^p \prod_{l=k}^{k+n-1} V_r^l \prod_{l=k+n}^{p-1} V_{\tau^l}^l\right) \\
&= \max\left(\mathbb{E}_r\, Y_{r+1}^{*L-k+1}, \max_{1\le n\le v_r\wedge(L-k+1)}\left(\sum_{p=k}^{k+n-1} U_r^p \prod_{l=k}^{p-1} V_r^l\right.\right. \\
&\quad\left.\left. + \prod_{l=k}^{k+n-1} V_r^l \sup_{\substack{\rho^r\le\tau^{k+n}\le\cdots\le\tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n},\ldots,\tau^L)=1}} \mathbb{E}_r\left(\sum_{p=k+n}^{L} U_{\tau^p}^p \prod_{l=k+n}^{p-1} V_{\tau^l}^l\right)\right)\right).
\end{aligned}$$

A standard argument shows that

$$\begin{aligned}
&\sup_{\substack{\rho^r\le\tau^{k+n}\le\cdots\le\tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n},\ldots,\tau^L)=1}} \mathbb{E}_r\left(\sum_{p=k+n}^{L} U_{\tau^p}^p \prod_{l=k+n}^{p-1} V_{\tau^l}^l\right) \\
&= \mathbb{E}_r \sup_{\substack{\rho^r\le\tau^{k+n}\le\cdots\le\tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n},\ldots,\tau^L)=1}} \mathbb{E}_{\rho^r}\left(\sum_{p=k+n}^{L} U_{\tau^p}^p \prod_{l=k+n}^{p-1} V_{\tau^l}^l\right) = \mathbb{E}_r\, Y_{\rho^r}^{*L-k-n+1},
\end{aligned} \tag{3.4}$$

which concludes the proof. □

We now establish a crucial relationship between the Snell envelopes $Y^{*L-k+1,j_1,\ldots,j_{k-1}}$ and Y^{*L-k+1} defined in (3.3). The following proposition shows that $Y^{*L-k+1,j_1,\ldots,j_{k-1}}$, parameterized by the j_k's, can be represented in terms of Y^{*L-k+1} which avoids the j_k's. Note that, for $k = 1$, both Snell envelopes coincide by definition.

PROPOSITION 3.5. *Suppose* $1 < k \le L + 1$. *Under the condition* $\mathcal{C}_{k-1}(j_1, \dots, j_{k-1}) = 1$, *we have*

(i) for $r > j_{k-1}$ *it holds*

$$Y_r^{*L-k+1, j_1, \dots, j_{k-1}} = \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \mathbb{E}_r\, Y^{*L-k+1}_{\rho^{j_{k-1}} \vee r} \prod_{l=1}^{k-1} V_{j_l}^l. \tag{3.5}$$

(ii) Further it holds

$$\begin{aligned} Y_{j_{k-1}}^{*L-k+1, j_1, \dots, j_{k-1}} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \max_{n \in N(j_1, \dots, j_{k-1})} \\ &\quad \times \left\{ \sum_{p=k}^{k-1+n} U_{j_{k-1}}^p \prod_{l=k}^{p-1} V_{j_{k-1}}^l + \mathbb{E}_{j_{k-1}}\, Y^{*L-k+1-n}_{\rho^{j_{k-1}}} \prod_{l=k}^{k-1+n} V_{j_{k-1}}^l \right\}, \end{aligned} \tag{3.6}$$

where the maximum runs over the $\mathcal{F}_{j_{k-1}}$*-measurable set*

$$N(j_1, \dots, j_{k-1}) := \{n;\ 0 \le n \le (v_{j_{k-1}} - \mathcal{E}_{k-1}(j_1, \dots, j_{k-1})) \wedge (L - k + 1)\}.$$

Proof. For $k = L + 1$ both assertions are implied by the conventions $Y_r^{*0, j_1, \dots, j_L} = X_{j_1, \dots, j_L}$ and $Y_r^{*0} = 0$. Hence, we assume for the remainder of the proof that $1 < k \le L$.

(i) Under $\mathcal{C}_{k-1}(j_1, \dots, j_{k-1}) = 1$, we have for $r > j_{k-1}$

$$\begin{aligned} Y_r^{*L-k+1, j_1, \dots, j_{k-1}} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \\ &\quad \times \sup_{\substack{r \le \tau^k \le \dots \le \tau^L \\ \mathcal{C}_L(j_1, \dots, j_{k-1}, \tau^k, \dots, \tau^L) = 1}} \mathbb{E}_r \left(\sum_{p=k}^{L} U_{\tau^p}^p \prod_{l=k}^{p-1} V_{\tau^l}^l \right). \end{aligned} \tag{3.7}$$

As $r > j_{k-1}$, we obtain, thanks to (3.4),

$$\begin{aligned} &\sup_{\substack{r \le \tau^k \le \dots \le \tau^L \\ \mathcal{C}_L(j_1, \dots, j_{k-1}, \tau^k, \dots, \tau^L) = 1}} \mathbb{E}_r \left(\sum_{p=k}^{L} U_{\tau^p}^p \prod_{l=k}^{p-1} V_{\tau^l}^l \right) \\ &= \mathbb{1}_{\{r < \rho^{j_{k-1}}\}} \sup_{\substack{\rho^{j_{k-1}} \le \tau^k \le \dots \le \tau^L \\ \mathcal{C}_{L-k+1}(\tau^k, \dots, \tau^L) = 1}} \mathbb{E}_r \left(\sum_{p=k}^{L} U_{\tau^p}^p \prod_{l=k}^{p-1} V_{\tau^l}^l \right) \\ &\quad + \mathbb{1}_{\{r \ge \rho^{j_{k-1}}\}} \sup_{\substack{r \le \tau^k \le \dots \le \tau^L \\ \mathcal{C}_{L-k+1}(\tau^k, \dots, \tau^L) = 1}} \mathbb{E}_r \left(\sum_{p=k}^{L} U_{\tau^p}^p \prod_{l=k}^{p-1} V_{\tau^l}^l \right) \\ &= \mathbb{1}_{\{r < \rho^{j_{k-1}}\}} \mathbb{E}_r\, Y^{*L-k+1}_{\rho^{j_{k-1}}} + \mathbb{1}_{\{r \ge \rho^{j_{k-1}}\}} Y_r^{*L-k+1}. \end{aligned} \tag{3.8}$$

Hence, by combining (3.7) and (3.8) we get (i).

(ii) Given that the first $(k-1)$ rights have been exercised at times $j_1 \leq \cdots \leq j_{k-1}$ the number of the remaining $(L-k+1)$ rights which are also exercised at time j_{k-1} must be chosen from the $\mathcal{F}_{j_{k-1}}$-measurable set $N(j_1, \ldots, j_{k-1}) = \{n;\ 0 \leq n \leq (v_{j_{k-1}} - \mathcal{E}_{k-1}(j_1, \ldots, j_{k-1})) \wedge (L-k+1)\}$. These are the only choices which obey the volume constraint at time j_{k-1}. Hence,

$$Y^{*L-k+1, j_1, \ldots, j_{k-1}}_{j_{k-1}} = \max_{n \in N(j_1, \ldots, j_{k-1})} \sup_{\substack{\rho^{j_{k-1}} \leq \tau^{n+k} \leq \cdots \leq \tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n}, \ldots, \tau^L) = 1}} \mathbb{E} X_{j_1, \ldots, j_{k-1}, \cdots, j_{k-1}, \tau^{k+n}, \ldots, \tau^L},$$

where the time index j_{k-1} appears $(n+1)$ times in the nth term. It then follows, for fixed $n \in N(j_1, \ldots, j_{k-1})$,

$$\begin{aligned}
&\sup_{\substack{\rho^{j_{k-1}} \leq \tau^{n+k} \leq \cdots \leq \tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n}, \ldots, \tau^L) = 1}} \mathbb{E} X_{j_1, \ldots, j_{k-1}, \cdots, j_{k-1}, \tau^{k+n}, \ldots, \tau^L} \\
&= \sum_{p=1}^{k-1} U^p_{j_p} \prod_{l=1}^{p-1} V^l_{j_l} + \prod_{l=1}^{k-1} V^l_{j_l} \sum_{p=k}^{k-1+n} U^p_{j_{k-1}} \prod_{l=k}^{p-1} V^l_{j_{k-1}} \\
&\quad + \prod_{l=1}^{k-1} V^l_{j_l} \prod_{l=k}^{k-1+n} V^l_{j_{k-1}} \sup_{\substack{\rho^{j_{k-1}} \leq \tau^{n+k} \leq \cdots \leq \tau^L \\ \mathcal{C}_{L-k-n+1}(\tau^{k+n}, \ldots, \tau^L) = 1}} \mathbb{E}_{j_{k-1}} \sum_{p=k+n}^{L} U^p_{\tau^p} \prod_{l=k+n}^{p-1} V^l_{\tau^p} \\
&= \sum_{p=1}^{k-1} U^p_{j_p} \prod_{l=1}^{p-1} V^l_{j_l} + \prod_{l=1}^{k-1} V^l_{j_l} \sum_{p=k}^{k-1+n} U^p_{j_{k-1}} \prod_{l=k}^{p-1} V^l_{j_{k-1}} + \prod_{l=1}^{k-1} V^l_{j_l} \prod_{l=k}^{k-1+n} V^l_{j_{k-1}} \mathbb{E}_{j_{k-1}} Y^{*L-k-n+1}_{\rho^{j_{k-1}}},
\end{aligned}$$

making again use of (3.4). This implies (3.6). □

3.1. Dual Representation Based on Doob Decompositions

The goal of this subsection is to prove and discuss the following simplified version of the dual representation from Theorem 2.4 for multiple stopping problems of the form (3.1).

THEOREM 3.6. *Suppose Y^{*L}_i is given by (3.1). Then:*

(i) *For any set of martingales $(M^{L-k+1}_r)_{r \geq 0}$, $k = 1, \ldots, L$, and any set of integrable adapted processes $(A^{L-k+1}_r)_{r \geq 0}$, $k = 1, \ldots, L$, it holds for $i \geq 0$ and with $j_0 := i$*

$$Y^{*L}_i \leq \mathbb{E}_i \max_{\substack{i \leq j_1 \leq \cdots \leq j_L \leq \partial \\ \mathcal{C}_L(j_1, \ldots, j_L) = 1}} \Bigg(\sum_{k=1}^{L} U^k_{j_k} \prod_{l=1}^{k-1} V^l_{j_l} + \sum_{k=1}^{L} \prod_{l=1}^{k-1} V^l_{j_l} \Big(M^{L-k+1}_{j_{k-1}} - M^{L-k+1}_{j_k} + \mathbb{E}_{j_k} A^{L-k+1}_{\rho^{j_{k-1}}} - \mathbb{E}_{j_{k-1}} A^{L-k+1}_{\rho^{j_{k-1}}} \Big) \Bigg).$$

(ii) *For every $i \geq 0$ it holds with $j_0 := i$*

$$Y_i^{*L} = \max_{\substack{i \leq j_1 \leq \cdots \leq j_L \leq \partial \\ \mathcal{C}_L(j_1,\ldots,j_L)=1}} \Bigg(\sum_{k=1}^{L} U_{j_k}^k \prod_{l=1}^{k-1} V_{j_l}^l + \sum_{k=1}^{L} \prod_{l=1}^{k-1} V_{j_l}^l \\ \big(M_{j_{k-1}}^{*L-k+1} - M_{j_k}^{*L-k+1} + \mathbb{E}_{j_k} A_{\rho^{j_{k-1}}}^{*L-k+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{*L-k+1} \big) \Bigg),$$

where M^{*L-k+1} , A^{*L-k+1} are the martingale part and the predictable part of the Doob decomposition of the auxiliary Snell envelopes Y^{*L-k+1} in (3.3), *respectively.*

We here recall that the Doob decomposition of Y^{*L-k+1} is the unique decomposition of the form

$$Y_r^{*L-k+1} = Y_0^{*L-k+1} + M_r^{*L-k+1} - A_r^{*L-k+1},$$

where the martingale M_r^{*L-k+1} and the predictable process A_r^{*L-k+1} start in zero at time zero. To prove Theorem 3.6 we need the following auxiliary result.

PROPOSITION 3.7. *Under the assumption of Theorem 3.6, a Doob martingale of $\mathbb{E}_r Y_{\rho^{j_{k-1}} \vee r}^{*L-k+1}$, say $\overline{M}_r^{*L-k+1}$, is determined for $r \geq j_{k-1}$ by*

$$\overline{M}_r^{*L-k+1} - \overline{M}_{j_{k-1}}^{*L-k+1} = M_r^{*L-k+1} - M_{j_{k-1}}^{*L-k+1} + \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{*L-k+1} - \mathbb{E}_r A_{\rho^{j_{k-1}}}^{*L-k+1}.$$

Proof. Using the Doob decomposition we may write

$$\begin{aligned} &\mathbb{E}_r Y_{\rho^{j_{k-1}} \vee r}^{*L-k+1} \\ &= \mathbb{1}_{\{r < \rho^{j_{k-1}}\}} \mathbb{E}_r Y_{\rho^{j_{k-1}}}^{*L-k+1} + \mathbb{1}_{\{r \geq \rho^{j_{k-1}}\}} Y_r^{*L-k+1} \\ &= \mathbb{1}_{\{j_{k-1} \leq r < \rho^{j_{k-1}}\}} \big(Y_{j_{k-1}}^{*L-k+1} + M_r^{*L-k+1} - M_{j_{k-1}}^{*L-k+1} - \mathbb{E}_r A_{\rho^{j_{k-1}}}^{*L-k+1} + A_{j_{k-1}}^{*L-k+1} \big) \\ &\quad + \mathbb{1}_{\{r \geq \rho^{j_{k-1}}\}} \big(Y_{j_{k-1}}^{*L-k+1} + M_r^{*L-k+1} - M_{j_{k-1}}^{*L-k+1} - A_r^{*L-k+1} + A_{j_{k-1}}^{*L-k+1} \big) \\ &= Y_{j_{k-1}}^{*L-k+1} + M_r^{*L-k+1} - M_{j_{k-1}}^{*L-k+1} \\ &\quad - \mathbb{1}_{\{j_{k-1} \leq r < \rho^{j_{k-1}}\}} \mathbb{E}_r A_{\rho^{j_{k-1}}}^{*L-k+1} - \mathbb{1}_{\{r \geq \rho^{j_{k-1}}\}} A_r^{*L-k+1} + A_{j_{k-1}}^{*L-k+1} \\ &= Y_{j_{k-1}}^{*L-k+1} + A_{j_{k-1}}^{*L-k+1} - \mathbb{1}_{\{r \geq \rho^{j_{k-1}}\}} \big(A_r^{*L-k+1} - A_{\rho^{j_{k-1}}}^{*L-k+1} \big) \end{aligned} \tag{3.9}$$

$$+ M_r^{*L-k+1} - M_{j_{k-1}}^{*L-k+1} - \mathbb{E}_r A_{\rho^{j_{k-1}}}^{*L-k+1}. \tag{3.10}$$

Since line (3.9) is the sum of a $\mathcal{F}_{k-1}$-measurable random variable and a predictable process and line (3.10) is a martingale, the proposition follows. □

We now can prove the dual representation.

Proof of Theorem 3.6.

(i) Suppose that, for $k = 1, \ldots, L$, M^{L-k+1} is a martingale and A^{L-k+1} is an adapted and integrable process. Then, the process $M_r^{L-k+1, j_1, \ldots, j_{k-1}}$ defined for $r \geq j_{k-1}$ via

$$M_r^{L-k+1, j_1, \ldots, j_{k-1}} := \prod_{l=1}^{k-1} V_{j_l}^l \big(M_r^{L-k+1} - M_{j_{k-1}}^{L-k+1} + \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k+1} - \mathbb{E}_r A_{\rho^{j_{k-1}}}^{L-k+1} \big)$$

is a martingale due to the boundedness of the V^l's. By Theorem 2.4-*(i)*, we have

$$\begin{aligned} Y_i^{*L} &\leq \mathbb{E}_i \max_{i \leq j_1 \leq \cdots \leq j_L \leq \partial} \left(X_{j_1, \ldots, j_L} + \sum_{k=1}^{L} \big(M_{j_{k-1}}^{L-k+1, j_1, \ldots, j_{k-1}} - M_{j_k}^{L-k+1, j_1, \ldots, j_{k-1}} \big) \right) \\ &= \mathbb{E}_i \max_{i \leq j_1 \leq \cdots \leq j_L \leq \partial} \Bigg(X_{j_1, \ldots, j_L} + \sum_{k=1}^{L} \prod_{l=1}^{k-1} V_{j_l}^l \\ &\qquad \big(M_{j_{k-1}}^{L-k+1} - M_{j_k}^{L-k+1} + \mathbb{E}_{j_k} A_{\rho^{j_{k-1}}}^{L-k+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k+1} \big) \Bigg) \end{aligned}$$

with X as defined in (3.2) for sufficiently large $N \in \mathbb{N}$. Letting N tend to infinity, we observe that maximization only takes place over those $j_1 \leq \cdots \leq j_L$ which satisfy $\mathcal{C}_L(j_1, \ldots, j_L) = 1$. Plugging in the definition of X for those $j_1 \leq \cdots \leq j_L$ yields the assertion.

(ii) We now apply Theorem 2.4-*(ii)* for X as defined in (3.2) with sufficiently large $N \in \mathbb{N}$. Letting N tend to infinity again and substituting the definition of X, we obtain,

$$\begin{aligned} Y_i^{*L} = &\max_{\substack{i \leq j_1 \leq \cdots \leq j_L \leq \partial \\ \mathcal{C}_L(j_1, \ldots, j_L) = 1}} \\ &\times \left(\sum_{k=1}^{L} U_{j_k}^k \prod_{l=1}^{k-1} V_{j_l}^l + \sum_{k=1}^{L} \big(M_{j_{k-1}}^{*L-k+1, j_1, \ldots, j_{k-1}} - M_{j_k}^{*L-k+1, j_1, \ldots, j_{k-1}} \big) \right), \end{aligned}$$

whenever $(M_r^{*L-k+1, j_1, \ldots, j_{k-1}})_{r \geq j_{k-1}}$ are Doob martingales of $(Y_r^{*L-k+1, j_1, \ldots, j_{k-1}})_{r \geq j_{k-1}}$. By Proposition 3.5-*(i)* and Proposition 3.7 we can take

$$\begin{aligned} &M_{j_{k-1}}^{*L-k+1, j_1, \ldots, j_{k-1}} - M_{j_k}^{*L-k+1, j_1, \ldots, j_{k-1}} \\ &= \prod_{l=1}^{k-1} V_{j_l}^l \big(M_{j_{k-1}}^{*L-k+1} - M_{j_k}^{*L-k+1} + \mathbb{E}_{j_k} A_{\rho^{j_{k-1}}}^{*L-k+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{*L-k+1} \big). \end{aligned}$$

□

Theorem 3.6 gives a straightforward generic way to calculate upper bounds for multiple stopping problems of the form (3.1) at time $i = 0$ via Monte Carlo by performing the following steps in a Markovian setting:

1. Solve the dynamic program in Proposition 3.4 for the auxiliary problems Y^{*L-k+1} approximately, and let $\hat{Y}^{L-k+1}$, $k = 1, \ldots, L$, denote the respective approximations.
2. Perform the Doob decomposition of $\hat{Y}^{L-k+1}$, $k = 1, \ldots, L$, numerically, e.g., by one layer of nested Monte Carlo as suggested by Andersen and Broadie (2004) in the context of options with a single early exercise right.

3. Plug the processes which stem from the numerical Doob decomposition into the formula of Theorem 3.6-*(i)* and replace the outer expectation by the sample mean.

This program will be carried out in more detail in Section 4 in the context of Swing options.

Note that for a large maturity and a large number of exercise rights, the path-wise maximum in the dual representation of Theorem 3.6 runs over a huge set. We will now show that, due to the special structure of the payoff in (3.1), this maximum can be computed efficiently by a recursion over the time steps and exercise levels.

Given any L-tuple of martingales $M = (M^1, \dots, M^L)$ and any L-tuple of adapted processes $A = (A^1, \dots, A^L)$, define, for $n = 0, \dots, L$ and $i = 0, \dots, \partial$,

$$\theta_i^{n,L}(M, A) := \max_{\substack{j_0 = i \le j_1 \le \cdots \le j_{L-n} \\ \mathcal{C}_{L-n}(j_1, \dots, j_{L-n}) = 1}} \sum_{k=1}^{L-n} \left(\prod_{l=1}^{k-1} V_{j_l}^{l+n} \right) \times \Big(U_{j_k}^{n+k} - \big(M_{j_k}^{L-n-k+1} - M_{j_{k-1}}^{L-n-k+1} \big) + \mathbb{1}_{\{k>1 \,\wedge\, j_k > j_{k-1}\}} \big(A_{\rho^{j_{k-1}}}^{L-k-n+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k-n+1} \big) \Big).$$

By Theorem 3.6,

$$Y_0^{*L} \le \mathbb{E}[\theta_0^{0,L}(M, A)]$$

for any pair of L-tuples (M, A), and

$$Y_0^{*L} = \theta_0^{0,L}(M^*, A^*)$$

for an optimal pair of L-tuples (M^*, A^*). Generalizing a related formula in Balder, Mahayni, and Schoenmakers (2011) in the context of flexible (or chooser) caps, the expression $\theta_0^{0,L}(M, A)$ can be recursively calculated by the following proposition.

PROPOSITION 3.8. *For every L-tuple of martingales $M = (M^1, \dots, M^L)$ and L-tuple of adapted processes $A = (A^1, \dots, A^L)$ it holds for $i = 0, \dots, T$ and $n = 0, \dots, L$,*

$$\begin{aligned}\theta_i^{n,L}(M, A) = \max \Bigg\{ &\theta_{i+1}^{n,L}(M, A) - \big(M_{i+1}^{L-n} - M_i^{L-n} \big), \max_{\nu=1,\dots,\nu_i \wedge (L-n)} \sum_{k=1}^{\nu} \left(\prod_{l=1}^{k-1} V_i^{l+n} \right) U_i^{n+k} \\ &+ \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n} \right) \Big(\theta_{\rho^i}^{n+\nu,L}(M, A) - \big(M_{\rho^i}^{L-n-\nu} - M_i^{L-n-\nu} \big) \\ &+ A_{\rho^i}^{L-n-\nu} - \mathbb{E}_i A_{\rho^i}^{L-n-\nu} \Big) \Bigg\},\end{aligned}$$

with

$$\theta_{\partial}^{n,L}(M, A) = \sum_{k=1}^{L-n} \left(\prod_{l=1}^{\nu} V_{\partial}^{l+n} \right) U_{\partial}^{n+k}.$$

Proof. The formula for $\theta_\partial^{n,L}(M, A)$ is obvious by definition. To prove the recursive formula, we denote

$$F^{n,L}(j_0,\dots,j_{L-n}) = \sum_{k=1}^{L-n}\left(\prod_{l=1}^{k-1} V_{j_l}^{l+n}\right)\Big(U_{j_k}^{n+k} - \big(M_{j_k}^{L-n-k+1} - M_{j_{k-1}}^{L-n-k+1}\big) \\ + \mathbb{1}_{\{k>1\ \wedge\ j_k>j_{k-1}\}}\big(A_{\rho^{j_{k-1}}}^{L-k-n+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k-n+1}\big)\Big).$$

Then,

$$\theta_i^{n,L}(M, A) = \max_{\nu=0,\dots,\nu_i\wedge(L-n)}\left\{\max_{\substack{j_0=\cdots=j_\nu=i<j_{\nu+1}\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n}(j_1,\dots,j_{L-n})=1}} F^{n,L}(j_0,\dots,j_{L-n})\right\}. \tag{3.11}$$

For $\nu = 0$ we get

$$\begin{aligned}\max_{\substack{j_0=i<j_1\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n}(j_1,\dots,j_{L-n})=1}} & F^{n,L}(j_0,\dots,j_{L-n}) \\ &= \max_{\substack{j_0=i+1\le j_1\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n}(j_1,\dots,j_{L-n})=1}} F^{n,L}(j_0,\dots,j_{L-n}) - \big(M_{i+1}^{L-n} - M_i^{L-n}\big) \\ &= \theta_{i+1}^{n,L}\big(M, A) - (M_{i+1}^{L-n} - M_i^{L-n}\big).\end{aligned} \tag{3.12}$$

For $\nu > 0$ we obtain

$$\begin{aligned}&\max_{\substack{j_0=\cdots=j_\nu=i<j_{\nu+1}\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n}(j_1,\dots,j_{L-n})=1}} F^{n,L}(j_0,\dots,j_{L-n}) \\ &= \sum_{k=1}^{\nu}\left(\prod_{l=1}^{k-1} V_i^{l+n}\right) U_i^{n+k} + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n}\right) \max_{\substack{j_\nu=i,\ \rho^i\le j_{\nu+1}\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n-\nu}(j_{\nu+1},\dots,j_{L-n})=1}} \\ &\quad\times \sum_{k=\nu+1}^{L-n}\left(\prod_{l=\nu+1}^{k-1} V_{j_l}^{l+n}\right)\big(U_{j_k}^{n+k} - M_{j_k}^{L-n-k+1} \\ &\qquad + M_{j_{k-1}}^{L-n-k+1} + \mathbb{1}_{\{j_k>j_{k-1}\}}\big(A_{\rho^{j_{k-1}}}^{L-k-n+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k-n+1}\big)\big) \\ &= \sum_{k=1}^{\nu}\left(\prod_{l=1}^{k-1} V_i^{l+n}\right) U_i^{n+k} + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n}\right)\big(\big(A_{\rho^i}^{L-\nu-n} - \mathbb{E}_i A_{\rho^i}^{L-\nu-n}\big) - \big(M_{\rho^i}^{L-\nu-n} - M_i^{L-\nu-n}\big)\big) \\ &\quad + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n}\right) \max_{\substack{j_\nu=\rho^i\le j_{\nu+1}\le\cdots\le j_{L-n}\\ \mathcal{C}_{L-n-\nu}(j_{\nu+1},\dots,j_{L-n})=1}} \sum_{k=\nu+1}^{L-n}\left(\prod_{l=\nu+1}^{k-1} V_{j_l}^{l+n}\right) \\ &\quad \big(U_{j_k}^{n+k} - M_{j_k}^{L-n-k+1} + M_{j_{k-1}}^{L-n-k+1} + \mathbb{1}_{\{k>\nu+1\ \wedge\ j_k>j_{k-1}\}}\big(A_{\rho^{j_{k-1}}}^{L-k-n+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k-n+1}\big)\big)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{\nu} \left(\prod_{l=1}^{k-1} V_i^{l+n} \right) U_i^{n+k} \\
&\quad + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n} \right) \left(\left(A_{\rho^i}^{L-\nu-n} - \mathbb{E}_i A_{\rho^i}^{L-\nu-n} \right) - \left(M_{\rho^i}^{L-\nu-n} - M_i^{L-\nu-n} \right) \right) \\
&\quad + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n} \right) \max_{\substack{j_0=\rho^i \le j_1 \le \cdots \le j_{L-n-\nu} \\ \mathcal{C}_{L-n-\nu}(j_1,\ldots,j_{L-n-\nu})=1}} \\
&\quad \times \sum_{k=1}^{L-n-\nu} \left(\prod_{l=1}^{k-1} V_{j_l}^{l+n+\nu} \right) \Big(U_{j_k}^{n+\nu+k} - M_{j_k}^{L-n-\nu-k+1} \\
&\quad + M_{j_{k-1}}^{L-n-\nu-k+1} + \mathbb{1}_{\{k>1 \,\wedge\, j_k > j_{k-1}\}} \left(A_{\rho^{j_{k-1}}}^{L-k-n-\nu+1} - \mathbb{E}_{j_{k-1}} A_{\rho^{j_{k-1}}}^{L-k-n-\nu+1} \right) \Big) \\
&= \sum_{k=1}^{\nu} \left(\prod_{l=1}^{k-1} V_i^{l+n} \right) U_i^{n+k} \\
&\quad + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n} \right) \left(\left(A_{\rho^i}^{L-\nu-n} - \mathbb{E}_i A_{\rho^i}^{L-\nu-n} \right) - \left(M_{\rho^i}^{L-\nu-n} - M_i^{L-\nu-n} \right) \right) \\
&\quad + \left(\prod_{\lambda=1}^{\nu} V_i^{\lambda+n} \right) \theta_{\rho^i}^{n+\nu,L}(M, A).
\end{aligned}
$$

Plugging this identity and (3.12) into (3.11) yields the assertion. □

3.2. Dual Representation Based on Snell Envelopes

In this subsection we present a simplified version of the dual representation in Corollary 2.5 in terms of approximate Snell envelopes for the multiple stopping problem of the form (3.1). It reads as follows:

THEOREM 3.9. *Suppose Y_i^{*L} is given by (3.1) for some fixed $0 \le i \le \partial$. Let $(Y^k)_{1\le k\le L}$ be any set of integrable approximations to $(Y^{*k})_{1\le k\le L}$ defined in (3.3). We then have, with the conventions $j_0 := -1$, $\rho^{j_0} := i$, and $Y^0 = 0$,*

$$
\begin{aligned}
Y_i^{*L} - Y_i^L \le \mathbb{E}_i \max_{\substack{i \le j_1 \le \cdots \le j_L \le \partial, \\ \mathcal{C}_L(j_1,\ldots,j_L)=1}} \quad & \sum_{k=1}^{L} \Bigg\{ \sum_{r=\rho^{j_{k-1}}}^{j_k-1} \prod_{l=1}^{k-1} V_{j_l}^l \left(\mathbb{E}_r Y_{r+1}^{L-k+1} - Y_r^{L-k+1} \right) \\
& + \mathbb{1}_{\{j_k > j_{k-1}\}} \prod_{l=1}^{k-1} V_{j_l}^l \Bigg(\max_{1 \le n \le \nu_{j_k} \wedge (L-k+1)} \Bigg\{ \sum_{p=k}^{k+n-1} U_{j_k}^p \prod_{l=k}^{p-1} V_{j_k}^l \\
& \qquad + \prod_{l=k}^{k+n-1} V_{j_k}^l \mathbb{E}_{j_k} Y_{\rho^{j_k}}^{L-k-n+1} \Bigg\} - Y_{j_k}^{L-k+1} \Bigg) \Bigg\}.
\end{aligned}
$$

Moreover, the right-hand side becomes zero if $Y^k = Y^{*k}$, for $k = 1, \ldots, L$.

Proof. Suppose $0 \le i \le \partial$ is fixed and assume that integrable and adapted processes Y^{L-k+1}, $k = 1, \dots, L$, are given which we consider as approximations of the Snell envelopes of the auxiliary multiple stopping problems Y^{*L-k+1}. Following the relationships for the Snell envelopes Y^{*L-k+1} and $Y^{*L-k+1, j_1, \dots, j_{k-1}}$ in Proposition 3.5, we define for $k > 1$ approximations to $Y^{*L-k+1, j_1, \dots, j_{k-1}}$ via

$$Y_r^{L-k+1, j_1, \dots, j_{k-1}} := \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \mathbb{E}_r Y_{\rho^{j_{k-1}} \vee r}^{L-k+1} \prod_{l=1}^{k-1} V_{j_l}^l, \quad r > j_{k-1}, \tag{3.13}$$

and (for $r = j_{k-1}$)

$$\begin{aligned} Y_{j_{k-1}}^{L-k+1, j_1, \dots, j_{k-1}} &:= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l \\ &\quad + \prod_{l=1}^{k-1} V_{j_l}^l \max_{n \in N(j_1, \dots, j_{k-1})} \left\{ \sum_{p=k}^{k-1+n} U_{j_{k-1}}^p \prod_{l=k}^{p-1} V_{j_{k-1}}^l + \prod_{l=k}^{k-1+n} V_{j_{k-1}}^l \mathbb{E}_{j_{k-1}} Y_{\rho^{j_{k-1}}}^{L-k+1-n} \right\}. \end{aligned} \tag{3.14}$$

Clearly, we define, for $k = 1$, $Y^{L, \emptyset} = Y^L$.

Applying Corollary 2.5 for X as defined in (3.1) and the above approximations we obtain,

$$\begin{aligned} Y_i^{*L} \le Y_i^L + \mathbb{E}_i \max_{\substack{i \le j_1 \le \dots \le j_L \le \partial, \\ \mathcal{C}_L(j_1, \dots, j_L) = 1}} \sum_{k=1}^{L} \Bigg(& Y_{j_k}^{L-k, j_1, \dots, j_k} - Y_{j_k}^{L-k+1, j_1, \dots, j_{k-1}} \\ & + \sum_{l=j_{k-1}}^{j_k - 1} \left(\mathbb{E}_l Y_{l+1}^{L-k+1, j_1, \dots, j_{k-1}} - Y_l^{L-k+1, j_1, \dots, j_{k-1}} \right) \Bigg), \end{aligned} \tag{3.15}$$

where we again observe that the path-wise maximum is attained on the set $\mathcal{C}_L(j_1, \dots, j_L) = 1$ by letting N (in the definition of X) tend to infinity.

To prove the upper bound, it is, in view of (3.15), sufficient to show that, for $i \le j_1 \le \dots \le j_L \le \partial$ with $\mathcal{C}_L(j_1, \dots, j_L) = 1$ the following assertions are true:

(i) If $k = 2, \dots, L$ and $j_k > j_{k-1}$ or if $k = 1$, then

$$\begin{aligned} & Y_{j_k}^{L-k, j_1, \dots, j_k} - Y_{j_k}^{L-k+1, j_1, \dots, j_{k-1}} \\ & \quad = \prod_{l=1}^{k-1} V_{j_l}^l \left(\max_{0 \le n \le (v_{j_k} - 1) \wedge (L-k)} \left\{ \sum_{p=k}^{k+n} U_{j_k}^p \prod_{l=k}^{p-1} V_{j_k}^l + \prod_{l=k}^{k+n} V_{j_k}^l \mathbb{E}_{j_k} Y_{\rho^{j_k}}^{L-k-n} \right\} - Y_{j_k}^{L-k+1} \right). \end{aligned}$$

(ii) If $k = 2, \dots, L$ and $j_k = j_{k-1}$, then

$$Y_{j_k}^{L-k, j_1, \dots, j_k} - Y_{j_k}^{L-k+1, j_1, \dots, j_{k-1}} \le 0.$$

(iii) If $k = 1$ and $i \le r \le j_1 - 1$, or if $k = 2, \dots, L$ and $\rho^{j_{k-1}} \le r \le j_k - 1$, then

$$\mathbb{E}_r Y_{r+1}^{L-k+1, j_1, \dots, j_{k-1}} - Y_r^{L-k+1, j_1, \dots, j_{k-1}} = \prod_{l=1}^{k-1} V_{j_l}^l \left(\mathbb{E}_r Y_{r+1}^{L-k+1} - Y_r^{L-k+1} \right).$$

(iv) For $k = 2, \ldots, L$ and $j_{k-1} \le r < \rho^{j_{k-1}}$
$$\mathbb{E}_r\, Y_{r+1}^{L-k+1, j_1,\ldots,j_{k-1}} - Y_r^{L-k+1, j_1,\ldots,j_{k-1}} \le 0.$$

We first show (i). To this end suppose that $k \ge 2$ and $j_k > j_{k-1}$. Then, $\mathcal{E}_k(j_1, \ldots, j_k) = 1$, which implies $N(j_1, \ldots, j_k) = \{n; 0 \le n \le (v_k - 1)\wedge(L-k)\}$. Hence, by (3.14),

$$Y_{j_k}^{L-k, j_1,\ldots,j_k} = \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \max_{0\le n\le (v_{j_k}-1)\wedge(L-k)} \left\{ \sum_{p=k}^{k+n} U_{j_k}^p \prod_{l=k}^{p-1} V_{j_k}^l + \prod_{l=k}^{k+n} V_{j_k}^l \mathbb{E}_{j_k} Y_{\rho^{j_k}}^{L-k-n} \right\}.$$

Subtracting the defining equation (3.13) for $Y_{j_k}^{L-k+1, j_1,\ldots,j_{k-1}}$ from the above expression, we obtain (i), because $j_k \ge \rho^{j_{k-1}}$. For $k = 1$, we again get $\mathcal{E}_1(j_1) = 1$ and (i) follows in the same way, taking the definition $Y_{j_1}^{L,\emptyset} = Y_{j_1}^L$ into account.

To derive (ii), we note that $\mathcal{E}_k(j_1, \ldots, j_k) = \mathcal{E}_{k-1}(j_1, \ldots, j_{k-1}) + 1$ for $j_k = j_{k-1}$. Thus,

$$\begin{aligned}
Y_{j_k}^{L-k, j_1,\ldots,j_k} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \\
&\quad \times \max_{0\le n\le (v_{j_k}-\mathcal{E}_{k-1}(j_1,\ldots,j_{k-1})-1)\wedge(L-k)} \left\{ \sum_{p=k}^{k+n} U_{j_k}^p \prod_{l=k}^{p-1} V_{j_k}^l + \prod_{l=k}^{k+n} V_{j_k}^l \mathbb{E}_{j_k} Y_{\rho^{j_k}}^{L-k-n} \right\} \\
&\le \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \\
&\quad \times \max_{0\le n\le (v_{j_{k-1}}-\mathcal{E}_{k-1}(j_1,\ldots,j_{k-1}))\wedge(L-k+1)} \left\{ \sum_{p=k}^{k+n} U_{j_{k-1}}^p \prod_{l=k}^{p-1} V_{j_{k-1}}^l + \prod_{l=k}^{k+n} V_{j_{k-1}}^l \mathbb{E}_{j_{k-1}} Y_{\rho^{j_{k-1}}}^{L-k-n} \right\} \\
&= Y_{j_{k-1}}^{L-k+1, j_1,\ldots,j_{k-1}} = Y_{j_k}^{L-k+1, j_1,\ldots,j_{k-1}}.
\end{aligned}$$

We next prove (iii). The case $k = 1$ is trivial in view of the definition of $Y^{L,\emptyset}$. Hence, we assume that $k \ge 2$ and $\rho^{j_{k-1}} \le r \le j_k - 1$. Then, $r + 1 > r \ge \rho^{j_{k-1}} > j_{k-1}$ and, thus, by (3.13)

$$\begin{aligned}
\mathbb{E}_r\, Y_{r+1}^{L-k+1, j_1,\ldots,j_{k-1}} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \mathbb{E}_r\, Y_{r+1}^{L-k+1} \prod_{l=1}^{k-1} V_{j_l}^l, \\
Y_r^{L-k+1, j_1,\ldots,j_{k-1}} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + Y_r^{L-k+1} \prod_{l=1}^{k-1} V_{j_l}^l.
\end{aligned}$$

Taking the difference of both equations yields (iii).

It remains to show (iv). For $k \ge 2$ and $j_{k-1} < r < \rho^{j_{k-1}}$, (3.13) implies

$$\mathbb{E}_r\, Y_{r+1}^{L-k+1, j_1,\ldots,j_{k-1}} = \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \mathbb{E}_r\, Y_{\rho^{j_{k-1}}}^{L-k+1} \prod_{l=1}^{k-1} V_{j_l}^l = Y_r^{L-k+1, j_1,\ldots,j_{k-1}}.$$

Finally, for $k \geq 2$ and $r = j_{k-1}$, by (3.13) and (3.14),

$$\begin{aligned}
\mathbb{E}_{j_{k-1}} Y_{j_{k-1}+1}^{L-k+1, j_1, \dots, j_{k-1}} &= \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \mathbb{E}_{j_{k-1}} Y_{\rho^{j_{k-1}}}^{L-k+1} \prod_{l=1}^{k-1} V_{j_l}^l \\
&\leq \sum_{p=1}^{k-1} U_{j_p}^p \prod_{l=1}^{p-1} V_{j_l}^l + \prod_{l=1}^{k-1} V_{j_l}^l \\
&\quad \times \max_{n \in N(j_1, \dots, j_{k-1})} \left\{ \sum_{p=k}^{k-1+n} U_{j_{k-1}}^p \prod_{l=k}^{p-1} V_{j_{k-1}}^l + \prod_{l=k}^{k-1+n} V_{j_{k-1}}^l \mathbb{E}_{j_{k-1}} Y_{\rho^{j_{k-1}}}^{L-k+1-n} \right\} \\
&= Y_{j_{k-1}}^{L-k+1, j_1, \dots, j_{k-1}}.
\end{aligned}$$

Hence, the asserted upper bound for $Y_i^{*L} - Y_i^L$ is shown. This upper bound is zero, if $Y^k = Y^{*k}$, for $k = 1, \dots, L$, because, by Proposition 3.4, Y^{*L-k+1} is a supermartingale and $Y_{j_k}^{*L-k+1}$ dominates

$$\max_{1 \leq n \leq v_{j_k} \wedge (L-k+1)} \left\{ \sum_{p=k}^{k+n-1} U_{j_k}^p \prod_{l=k}^{p-1} V_{j_k}^l + \prod_{l=k}^{k+n-1} V_{j_k}^l \mathbb{E}_{j_k} Y_{\rho^{j_k}}^{*L-k-n+1} \right\}.$$

□

As a spin-off result from Theorem 3.9, we may write the following upper bound for $Y_i^{*,L}$ which avoids the computation of the recursive maximum from Proposition 3.8 (cf. Schoenmakers 2012, remark 3.3, for a related result in the context of the standard multiple stopping problem).

COROLLARY 3.10. *Suppose all assumptions and all conventions of Theorem 3.9 are in force. Then,*

$$\begin{aligned}
Y_i^{*L} - Y_i^L \leq \mathbb{E}_i \Bigg\{ &\sum_{r=i}^{T-1} \max_{0 \leq k < L} \left(\mathcal{V}_{\max}^k \left(\mathbb{E}_r Y_{r+1}^{L-k} - Y_r^{L-k} \right)^+ \right) + \sum_{k=1}^{L} \mathcal{V}_{\max}^{k-1} \\
&\times \max_{i \leq j \leq \partial} \left(\max_{1 \leq n \leq v_{j_k} \wedge (L-k+1)} \left(\sum_{p=k}^{k+n-1} U_j^p \prod_{l=k}^{p-1} V_j^l \prod_{l=k}^{k+n-1} V_j^l \mathbb{E}_j Y_{\rho^j}^{L-k-n+1} \right) - Y_j^{L-k+1} \right)^+ \Bigg\},
\end{aligned}$$

where

$$\mathcal{V}_{\max}^k := \prod_{l=1}^{k} \max_{j \geq i} V_j^l.$$

Moreover, the right-hand side becomes zero if $Y^k = Y^{*k}$, for $k = 1, \dots, L$.

Proof. It is straightforward to check that the upper bound in this corollary is actually an upper bound to the right-hand side of the estimate in Theorem 3.9. That the bound is still tight, i.e., that the right-hand side becomes zero, if $Y^k = Y^{*k}$, for $k = 1, \dots, L$, follows from the same argument as at the end of the proof of Theorem 3.9. □

4. A NUMERICAL EXAMPLE

We provide a numerical example for the dual representation of multiple stopping problems in the context of swing option pricing. Throughout this section, we assume $i = 0$, i.e., we provide confidence bounds for the swing option price at time 0. Precisely, we consider a stylized swing option, similar to those considered in Meinshausen and Hambly (2004) and Bender (2011a). In our setting, the holder of a swing option has the right to buy a certain quantity of electricity in the period from $j = 0, \dots, T$, for a fixed strike price $K > 0$, subject to the restriction that the option allows up to $L \geq 1$ exercise opportunities under the volume constraints v_j, and where a refraction period has to be taken into account. In the numerical implementation we choose $T = 50$ or $T = 300$ and recall that $\partial := T + 1$. The price of electricity, $(S_t)_{t=0,\dots,T}$, is modeled by the following discretized exponential Gaussian Ornstein–Uhlenbeck process

$$\log(S_j) = (1 - k)(\log(S_{j-1}) - \mu) + \mu + \sigma\epsilon_j,\ S_0 = s_0 > 0, \tag{4.1}$$

where $(\epsilon_j)_{j=1,\dots,T}$ is a family of independent standard normal random variables and the parameters are specified by

$$\sigma = 0.5, k = 0.9, \mu = 0, s_0 = 1.$$

We set $S_\partial = 0$, which means that no penalty is imposed, if the holder of the option does not exercise all rights. This model is a discrete version of the exponential Gaussian Ornstein–Uhlenbeck model suggested by Lucia and Schwartz (2002). We stress however that our algorithm can be generically applied to any Markovian model such as the more realistic non-Gaussian exponential Ornstein–Uhlenbeck model studied by Benth, Kallsen, and Meyer-Brandis (2007) and Hambly, Howison, and Kluge (2009). The payoff of the swing option is then given by X in (3.1) with

$$\begin{aligned} V_j^l &:= 1, l = 1, \dots, L-1, j = 0, \dots, \partial, \\ U_j^p &:= Z_j := Z(S_j) := (S_j - K)^+, j = 0, \dots, \partial, p = 1, \dots, L. \end{aligned}$$

In our numerical study we assume that the strike price is $K = 1$. As volume constraints, we consider the situation of a unit volume constraint $v_i = 1$ for $i = 0, \dots, T$ and the situation of a swing option with $v_i = 1$ on weekdays and $v_i = 2$ on Saturdays and Sundays. We recall that swing options are traded as base swing options, peak swing options, and off-peak swing options, see, e.g., Haarbrücker and Kuhn (2009). In an off-peak swing option rights can only be exercised in off-peak periods. As there are twice as many off-peak periods at Saturdays and Sundays than at weekdays, our second volume constraint corresponds to the situation of an off-peak swing option. The refraction period which we impose is a constant refraction period, i.e., $\rho^i = (i + \delta)\wedge\partial$ for various choices of the constant $\delta \in \mathbb{N}$.

In this Markovian framework, we produce confidence intervals for the price of the swing option at time $i = 0$ by applying the following steps. The procedure below can easily be generalized to the generic cash flow structure of Section 3, provided the problem has a Markovian structure. (For notational convenience we only spell out the algorithm for the swing option case.)

4.1. Implementation

Step 1: Precompute an approximation of the continuation values. We employ least squares Monte Carlo regression to obtain an approximation to the continuation values

$$C_j^{*1,l}(S_j) := \mathbb{E}\left[Y_{j+1}^{*l}\middle|\mathcal{F}_j\right] = \mathbb{E}\left[Y_{j+1}^{*l}\middle|S_j\right],\ C_T^{*1,l}(S_T) = 0,$$
$$C_j^{*\delta,l}(S_j) := \mathbb{E}\left[Y_{j+\delta}^{*l}\middle|\mathcal{F}_j\right] = \mathbb{E}\left[Y_{j+\delta}^{*l}\middle|S_j\right],\ C_T^{*\delta,l}(S_T) = 0,$$

with $l = 1, \ldots, L$, where here and in the following $j + 1$ and $j + \delta$ are to be understood as $j + 1 \wedge \partial$ and $j + \delta \wedge \partial$. Recall that $(Y_j^{*l})_{j=0,\ldots,T}$ is given by the dynamic program from Proposition 3.4. We simulate N_1 independent paths $(S_j^m)_{j=0,\ldots,T}^{m=1,\ldots,N_1}$. Choosing as basis functions

$$\psi_1(x) := x, \quad \psi_2(x) := (x - K)^+,$$

we use $(S_j^m)_{j=0,\ldots,T}^{m=1,\ldots,N_1}$ in a straightforward least squares regression procedure to solve the dynamic program approximately, replacing the conditional expectations by the least squares Monte Carlo estimator. This yields approximations to $C_j^{*1,l}(\cdot)$ and $C_j^{*\delta,l}(\cdot)$, denoted by $C_j^{1,l}(\cdot)$ and $C_j^{\delta,l}(\cdot)$, which are linear combinations of the basis functions. We note that alternative numerical methods can be utilized to construct approximate continuation values. For example, Jaillet, Ronn, and Tompaidis (2004) suggest to apply a trinomial forest approximation of the electricity price and Bardou, Bouthemy, and Pagès (2009) study a quantization approach.

Step 2: Compute lower bounds. Given the functions $C_j^{1,l}(\cdot)$ and $C_j^{\delta,l}(\cdot)$, we define a (suboptimal) stopping rule $\left(\tau_j^{p,l}\right)_{1\le p\le l}^{1\le l\le L}$ for $0 \le j \le T$ along a given trajectory $(S_j)_{j=0\ldots,T}$ (which we suppress in the notation below) using the following iteration. Here, $\tau_j^{p,l}$ is interpreted as the time at which the investor exercises the pth right, if l rights are left at time j.

$$
\begin{aligned}
&\tau_j^{0,l} := j - \delta;\\
&p := k := 0;\\
&\textit{while}\ (p < l) \quad \textit{do}\\
&\qquad \tau_j^{p+1,l} := \inf\left\{(\tau_j^{p,l} + \delta) \wedge \partial \le r \le \partial : \max_{1\le n\le v_r\wedge(l-p)} \left(nZ_r + C_r^{\delta,l-p-n}\right) \ge C_r^{1,l-p}\right\},\\
(4.2)&\qquad s := \tau_j^{p+1,l},\\
&\qquad k := \operatorname{argmax}_{1\le n\le v_s\wedge(l-p)}\left(nZ_s + C_s^{\delta,l-p-n}\right),\\
&\qquad \tau_j^{p+1,l} := \tau_j^{p+2,l} := \ldots := \tau_j^{p+k,l} := s,\\
&\qquad p := p + k,\\
&\textit{end}.
\end{aligned}
$$

When $C_j^{1,l}(\cdot)$ and $C_j^{\delta,l}(\cdot)$ are replaced by $C_j^{*1,l}(\cdot)$ and $C_j^{*\delta,l}(\cdot)$, then this family of stopping times is optimal. Hence, $\left(\tau_j^{p,l}\right)_{1\le p\le l}^{1\le l\le L}$ is a good family of stopping times, if the approximations of the continuation values in Step 1 are reasonably close to the true continuation values.

REMARK 4.1.

(i) In the situation of unit volume constraint (i.e., $v \equiv 1$), the stopping rule (4.2) simplifies to $\tau_j^{0,l} = j - \delta$ and

$$\tau_j^{p,l} = \inf\{(\tau_j^{p-1,l} + \delta) \wedge \partial \le r \le \partial : Z_r + C_r^{\delta,l-p} \ge C_r^{1,l-p+1}\},\ 1 \le l \le L, 1 \le p \le l,$$

compare with equation (3.7) in Bender (2011a).

(ii) In the situation of a trivial refraction period (i.e., $\delta = 1$), the above construction of approximate stopping rules is also used in Aleksandrov and Hambly (2010).

Setting

$$\underline{Y}_0^l := \mathbb{E}_0 \sum_{p=1}^{l} Z_{\tau_0^{p,l}},\ \underline{Y}_1^l := \mathbb{E}_1 \sum_{p=1}^{l} Z_{\tau_1^{p,l}},\ \underline{Y}_\delta^l := \mathbb{E}_\delta \sum_{p=1}^{l} Z_{\tau_\delta^{p,l}},$$

we have that $\underline{Y}_0^l$ is a lower bound for Y_0^{*l}. By the tower property of the conditional expectation, we also have

$$\mathbb{E}_0 \underline{Y}_1^l = \mathbb{E}_0 \sum_{p=1}^{l} Z_{\tau_1^{p,l}}, \quad \mathbb{E}_0 \underline{Y}_\delta^l = \mathbb{E}_0 \sum_{p=1}^{l} Z_{\tau_\delta^{p,l}}.$$

As for simulations, we generate a new set of N_2 independent paths of the underlying price process, which we again denote, in abuse of notation, by $(S_j^m)_{j=0,\dots,T}^{m=1,\dots,N_2}$. Along theses N_2 trajectories we compute $\tau_0^{p,l}$ and apply the notation

$$\tau_0^{p,l,m}, \quad 1 \le p \le l, 1 \le m \le N_2.$$

Now the lower biased estimate $\widehat{\underline{Y}}_0^l$ for Y_0^{*l} is calculated by averaging over the N_2 realizations of $\sum_{p=1}^{l} Z_{\tau_0^{p,l}}$, i.e.,

$$\widehat{\underline{Y}}_0^l = \frac{1}{N_2} \sum_{m=1}^{N_2} \sum_{p=1}^{l} Z\big(S^m_{\tau_0^{p,l,m}}\big), \quad 1 \le l \le L. \tag{4.3}$$

Similarly, we also construct approximations

$$\hat{\mathbb{E}}_0 \underline{Y}_1^l = \frac{1}{N_2} \sum_{m=1}^{N_2} \sum_{p=1}^{l} Z\big(S_{\tau_1^{p,l,m}}\big), \quad \hat{\mathbb{E}}_0 \underline{Y}_\delta^l = \frac{1}{N_2} \sum_{m=1}^{N_2} \sum_{p=1}^{l} Z\big(S_{\tau_\delta^{p,l,m}}\big)$$

of $\mathbb{E}_0 \underline{Y}_1^l$ and $\mathbb{E}_0 \underline{Y}_\delta^l$, which we store for later use.

For constructing confidence intervals, we also save the empirical standard deviation $\mathrm{stddev}(\widehat{\underline{Y}}_0^l)$.

Step 3: Compute approximations to the Snell envelopes. Using the stopping rule (4.2), we consider a family of random variables

$$\underline{Y}_j^l := \mathbb{E}_j \sum_{p=1}^{l} Z_{\tau_j^{p,l}},\ 1 \le l \le L, 0 \le j \le T, \tag{4.4}$$

which is an approximation to the Snell envelope $(Y_j^{*l})_{0\le j\le T}^{1\le l\le L}$. We apply the following procedure to simulate $\underline{Y}_j^l$:

We simulate a new set of N_3 paths of the underlying $(S_j^m)_{0\le j\le T}^{1\le m\le N_3}$ (abusing the notation, again). We refer to these paths as the outer paths. We now fix a pair (m, j) and compute approximations of $\underline{Y}_j^l$, $\mathbb{E}_j \underline{Y}_{j+1}^l$, and $\mathbb{E}_j \underline{Y}_{j+\delta}^l$ along the mth outer path which are denoted by $\widehat{\underline{Y}}_j^{l,m}$, $\widehat{\mathbb{E}}_j^m \underline{Y}_{j+1}^l$, and $\widehat{\mathbb{E}}_j^m \underline{Y}_{j+\delta}^l$, respectively. In these approximations, the conditional expectations are replaced by the sample mean over a set of inner simulations. Hence, for the fixed path S^m and the fixed time point j, we generate N_4 independent sample paths of $(S_r)_{r=j,\dots,T}$ under the conditional law given that $S_j = S_j^m$. These inner paths are denoted by $(\bar{S}_r^\nu)_{r=j,\dots,T}^{\nu=1,\dots,N_4}$, suppressing here and in the following the dependence on (m, j). Along the inner paths $\bar{S}^\nu$ we compute the stopping times $\tau_i^{p,l}$ for $i = j, j+1, j+\delta$ in (4.2) and apply the notation

$$\tau_i^{p,l,\nu}, 1 \le p \le l, 1 \le l \le L, \nu = 1, \dots, N_4.$$

We now define

$$\widehat{\underline{Y}}_j^{l,m} := \widehat{\mathbb{E}}_j^m \sum_{p=1}^{l} Z_{\tau_j^{p,l}} := \frac{1}{N_4} \sum_{\nu=1}^{N_4} \sum_{p=1}^{l} Z(\bar{S}^\nu_{\tau_j^{p,l,\nu}}).$$

Similarly, we approximate $\mathbb{E}_j \underline{Y}_{j+1}^l$ for the fixed j along the fixed mth outer path by

$$\widehat{\mathbb{E}}_j^m \underline{Y}_{j+1}^l := \frac{1}{N_4} \sum_{\nu=1}^{N_4} \sum_{p=1}^{l} Z(\bar{S}^\nu_{\tau_{j+1}^{p,l,\nu}}),$$

taking the tower property of the conditional expectation into account. The approximation $\widehat{\mathbb{E}}_j^m \underline{Y}_{j+\delta}^l$ is obtained analogously.

REMARK 4.2. Note that, for $j = 0$ approximations $\widehat{\underline{Y}}_0^l$, $\widehat{\mathbb{E}}_0 \underline{Y}_1^l$, and $\widehat{\mathbb{E}}_0 \underline{Y}_\delta^l$ of $\underline{Y}_0^l$, $\mathbb{E}_0 \underline{Y}_1^l$, and $\mathbb{E}_0 \underline{Y}_\delta^l$ were already obtained based on the N_2-samples in Step 2. As typically $N_2 > N_4$ these approximations are more accurate. Hence, one can perform Step 3 for $j \ge 1$ only and set

$$\widehat{\underline{Y}}_0^{l,m} := \widehat{\underline{Y}}_0^l, \quad \widehat{\mathbb{E}}_0^m \underline{Y}_1^l := \widehat{\mathbb{E}}_0 \underline{Y}_1^l, \quad \widehat{\mathbb{E}}_0^m \underline{Y}_\delta^l := \widehat{\mathbb{E}}_0 \underline{Y}_\delta^l, \quad m = 1, \dots, M.$$

This trick of applying the more accurate nonnested Monte Carlo simulation of Step 2 at time 0 leads to a significant decrease of the variance in the simulation of the upper bound. This is in the same spirit as the computation of low variance upper bounds for the standard stopping problem from Andersen and Broadie (2004).

Step 4: Compute the upper bounds. The Doob decomposition of $\underline{Y}_j^l$ yields the pair $(\underline{M}_j^l, \underline{A}_j^l)$. Note that, due to

$$\underline{M}_{i+1}^l - \underline{M}_i^l = \underline{Y}_{i+1}^l - \mathbb{E}_i \underline{Y}_{i+1}^l,$$

and

$$-(\underline{M}_{i+\delta}^l - \underline{M}_i^l) + \underline{A}_{i+\delta}^l - \mathbb{E}_i \underline{A}_{i+\delta}^l = \mathbb{E}_i \underline{Y}_{i+\delta}^l - \underline{Y}_{i+\delta}^l,$$

we can rewrite the recursion formula in Proposition 3.8 as

$$\theta_i^{n,L} = \max\left\{\theta_{i+1}^{n,L} + \mathbb{E}_i \underline{Y}_{i+1}^{L-n} - \underline{Y}_{i+1}^{L-n}, \max_{1\le\nu\le v_i\wedge(L-n)} \left(\nu Z_i + \theta_{i+\delta}^{n+\nu,L} + \mathbb{E}_i \underline{Y}_{i+\delta}^{L-n-\nu} - \underline{Y}_{i+\delta}^{L-n-\nu}\right)\right\}.$$

We now introduce approximations $\theta_i^{n,L,m}$ of $\theta_i^{n,L}$ along the mth outer path of Step 3 by replacing $\underline{Y}_j^l$, $\mathbb{E}_j \underline{Y}_{j+1}^l$, and $\mathbb{E}_j \underline{Y}_{j+\delta}^l$ with their simulated counterparts $\widehat{\underline{Y}}_j^{l,m}$, $\widehat{\mathbb{E}}_j^m \underline{Y}_{j+1}^l$, and $\widehat{\mathbb{E}}_j^m \underline{Y}_{j+\delta}^l$ constructed in Step 3.

As simulation-based estimate for the upper bound, we use

$$Y_0^{up,L} := \frac{1}{N_3}\sum_{m=1}^{N_3} \theta_0^{0,L,m}.$$

Replacing the conditional expectations by the sample mean in $\theta_0^{0,L,m}$ introduces an additional bias up thanks to Jensen's inequality and the convexity of the maximum. Hence, the estimator $Y_0^{up,L}$ is biased up by Theorem 3.6 and Proposition 3.8.

Finally, a 95% confidence interval on the price of the swing option is given by

$$\left[\widehat{\underline{Y}}_0^L - 1.96 \times \mathrm{stddev}\left(\widehat{\underline{Y}}_0^L\right),\ Y_0^{up,L} + 1.96 \times \mathrm{stddev}\left(Y_0^{up,L}\right)\right].$$

4.2. Numerical Results: Swing Options with Unit Volume Constraints

We now present some numerical results which the above algorithm produces for the swing option contract as specified at the beginning of this section. Let us first consider the situation of a unit volume constraint, i.e., $v_j := 1$ for $j = 0, \dots, T$ with $T = 50$. We recall that $\delta \in \mathbb{N}$ denotes a constant refraction period. In this setting the dual representation of Theorem 3.6 reduces to the one derived in Bender (2011a). In the latter paper the same swing option example is treated numerically but for up to three exercise rights only. Thanks to the recursion formula in Proposition 3.8 we can now efficiently treat the case of a large number of exercise rights. Moreover, the upper bound algorithm in Bender (2011a) differs slightly from the one we propose here. In Bender (2011a) the upper bound is calculated based on the numerical Doob decomposition of

$$Y_j^l = \max\left\{Z_j + C_j^{\delta,l-1}, C_j^{1,l}\right\},$$

although we here utilize the numerical Doob decomposition of $\underline{Y}_j^l := \mathbb{E}_j \sum_{p=1}^l Z_{\tau_j^{p,l}}$.

The choice of simulation parameters in our study is as follows: In Step 1, we choose $N_1 = 1{,}000$ paths for the least squares Monte Carlo regression to approximate the continuation function. In Step 2, the lower bound is simulated using $N_2 = 300{,}000$ paths and in Step 3, we employ $N_3 = 2{,}000$ outer and $N_4 = 100$ inner paths for the computation of the upper bound. Moreover, we use the variance reduction method from Remark 4.2.

Table 4.1 depicts the numerical results for the case of two and three exercise rights for a refraction period ranging from 1 to 20. We observe that the relative length of the 95% confidence intervals is less than 1% in all cases. A comparison with the numerical results in Bender (2011a) shows that the differences in the upper price estimator based on Y^l and $\underline{Y}^l$ are negligible, but the variance reduction method of Remark 4.2 shrinks the confidence interval significantly.

The numerical results for the case of a larger number of exercise rights ($L = 4$, 6, 8, 10) are presented in Table 4.2. Due to the time horizon of 50 days, it may happen that,

TABLE 4.1
Unit Volume Constraints ($v_j \equiv 1$)

δ	L	$\underline{\widehat{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval	L	$\underline{\widehat{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval
1	2	3.3116	3.3211	[3.30738, 3.32229]	3	4.53627	4.54806	[4.53118, 4.54938]
2	2	3.27513	3.28469	[3.27094, 3.28587]	3	4.43753	4.45154	[4.43252, 4.45295]
4	2	3.2313	3.24083	[3.22716, 3.242]	3	4.29996	4.31656	[4.29502, 4.31813]
6	2	3.18613	3.19809	[3.18197, 3.19948]	3	4.15557	4.17514	[4.15063, 4.17697]
8	2	3.13625	3.14984	[3.13213, 3.15143]	3	3.99773	4.01954	[3.99289, 4.02158]
10	2	3.09022	3.10332	[3.08613, 3.1048]	3	3.83377	3.8528	[3.82898, 3.85464]
20	2	2.81521	2.83005	[2.81123, 2.83173]	3	2.91951	2.93649	[2.91536, 2.9383]

Note: Numerical results based on the approximation $\underline{\widehat{Y}}_j^l$ to the Snell envelope via the stopping rule (4.2) for two and three exercise rights.

TABLE 4.2
Unit Volume Constraints ($v_j \equiv 1$)

δ	L	$\underline{\widehat{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval	L	$\underline{\widehat{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval
1	4	5.60136	5.614	[5.59554, 5.61527]	6	7.38677	7.40107	[7.37977, 7.4023]
2	4	5.41347	5.43091	[5.4078, 5.43249]	6	6.94554	6.97364	[6.93882, 6.97562]
4	4	5.13119	5.15543	[5.12562, 5.15741]	6	6.2151	6.25165	[6.2086, 6.2541]
6	4	4.82479	4.85239	[4.81928, 4.85454]	6	5.4079	5.44248	[5.40174, 5.44491]
8	4	4.49057	4.51822	[4.48525, 4.52041]	6	4.68469	4.71574	[4.67908, 4.71808]
10	4	4.13658	4.16231	[4.13141, 4.16444]	6	4.1662	4.19164	[4.16098, 4.19373]
1	8	8.83286	8.84907	[8.82488, 8.85034]	10	10.0219	10.0391	[10.0131, 10.0404]
2	8	8.04508	8.08421	[8.03754, 8.08651]	10	8.80264	8.85093	[8.79443, 8.85353]
4	8	6.66726	6.71129	[6.66021, 6.71389]	10	6.74649	6.78596	[6.73929, 6.78847]
6	8	5.45188	5.48518	[5.44563, 5.48748]	10	5.45187	5.48518	[5.44563, 5.48748]

Note: Numerical results based on the approximation $\underline{\widehat{Y}}_j^l$ to the Snell envelope via the stopping rule (4.2) for a higher number of exercise rights.

for a large number of rights and a large refraction period, some exercise rights cannot be used by the investor. This explains why, e.g., the price bounds for the swing option with refraction period $\delta = 6$ are the same for $L = 8$ and $L = 10$ rights. Concerning the accuracy of our numerical procedure we emphasize that the relative difference between lower and upper bound is still less than 1% even in the case of 10 exercise rights.

4.3. Numerical Results: Off-Peak Swing Option

We now consider an off-peak swing option which allows for buying at most one package of electricity on weekdays and two packages on Saturdays and Sundays. Hence,

TABLE 4.3
Off-Peak Volume Constraints

δ	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval
2	4	5.73736	5.76003	[5.73078, 5.76192]	6	7.55394	7.58547	[7.54595, 7.58773]
4	4	5.51763	5.5468	[5.51105, 5.54915]	6	7.01995	7.06275	[7.01198, 7.06577]
6	4	5.27976	5.30967	[5.27316, 5.31227]	6	6.45197	6.49774	[6.44401, 6.50102]
8	4	5.06733	5.10055	[5.06078, 5.10335]	6	5.9401	5.98546	[5.93238, 5.989]
10	4	4.85039	4.88637	[4.84386, 4.88953]	6	5.46672	5.50782	[5.45916, 5.51116]
2	8	8.97188	9.01806	[8.96279, 9.02078]	10	10.0822	10.1411	[10.0721, 10.1443]
4	8	8.00789	8.06203	[7.99887, 8.06551]	10	8.58082	8.63832	[8.57102, 8.64178]
6	8	7.06562	7.11754	[7.05669, 7.12102]	10	7.33533	7.38481	[7.32577, 7.38835]
8	8	6.18418	6.23051	[6.17596, 6.23403]	10	6.20274	6.24781	[6.19445, 6.2513]

Note: Numerical results for the off-peak swing option for various exercise rights and refraction periods. The simulations are based on the approximation $\widehat{\underline{Y}}_j^l$ to the Snell envelope via the stopping rule (4.2).

we have for $j = 0, \dots, 50$ the volume constraints

$$v_j := \begin{cases} 1, & \text{if } j \text{ is a week day,} \\ 2, & \text{if } j \text{ is a weekend day,} \end{cases} \tag{4.5}$$

where we start in $j = 0$ on a Monday.

We run the above algorithm with $N_1 = 10{,}000$, $N_2 = 300{,}000$, $N_3 = 2{,}000$, and $N_4 = 100$ sample paths. The numerical results for this off-peak swing option are presented in Table 4.3 for various choices of the number L of exercise rights and the length δ of the refraction period. Note that the dual representations yet available in the literature do not cover the case of a nontrivial refraction period ($\delta \neq 1$) in combination with nontrivial volume constraints ($v \neq 1$).

Table 4.3 exhibits the numerical results computed with the algorithm which was described in detail in Section 4.1, although Table 4.4 provides the numerical results for a slight modification of this algorithm which replaces stopping time-based approximations to the Snell envelope by

$$Y_j^l = \max\left\{ \max_{1 \le n \le v_j \wedge l} \left(n Z_j + C_j^{\delta, l-n}\right), C_j^{1,l} \right\},$$

similar to the computation of upper bounds in Bender (2011a). We note that this variant of the algorithm is somewhat cheaper to implement, because the inner paths in Step 3 of Section 4.1 need only be simulated δ steps ahead. A comparison between Tables 4.3 and 4.4 show that the quality of the upper bounds produced by the original algorithm and its variant is almost the same. We observe that the relative length of the 95% confidence interval is less than 1.3% in all cases which demonstrates that the algorithm performs equally well in the presence of volume constraints than in the unit constraint case.

Due to the feature of allowing for exercising twice on weekends in this example, the swing option prices are now higher than in the previous example with unit volume constraint. Moreover, additional rights can now become beneficial in situations in which

TABLE 4.4
Off-Peak Volume with Differently Computed Upper Bounds

δ	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval
2	4	5.73631	5.79467	[5.72494, 5.80299]	6	7.55461	7.62608	[7.54078, 7.63619]
4	4	5.51148	5.54104	[5.50001, 5.54941]	6	7.01482	7.04521	[7.00098, 7.05536]
6	4	5.28808	5.3262	[5.27667, 5.33452]	6	6.458	6.50509	[6.44423, 6.51519]
8	4	5.07143	5.0982	[5.05997, 5.10654]	6	5.93962	5.97165	[5.92618, 5.98171]
10	4	4.85361	4.89511	[4.84228, 4.90326]	6	5.46925	5.51407	[5.45619, 5.52365]
2	8	8.98204	9.06441	[8.96626, 9.07597]	10	10.0889	10.1741	[10.0715, 10.1869]
4	8	8.00261	8.03391	[7.98694, 8.04548]	10	8.57358	8.60182	[8.55655, 8.61451]
6	8	7.07971	7.12341	[7.06422, 7.13479]	10	7.34633	7.38241	[7.32977, 7.39474]
8	8	6.18142	6.2135	[6.16713, 6.2243]	10	6.20036	6.23156	[6.18596, 6.24246]

Note: Numerical results for the off-peak swing option for various exercise rights and refraction periods. The simulations are based on the approximation $Y_j^l = \max\{\max_{1\le n\le v_j\wedge l}(nZ_j + C_j^{\delta,l-n}), C_j^{1,l}\}$.

TABLE 4.5
Off-Peak Volume Constraints for $T = 300$

δ	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval	δ	L	$\widehat{\underline{Y}}_0^L$	$Y_0^{up,L}$	95% confidence interval
	10	20.415	20.594	[20.378, 20.623]		10	19.582	19.766	[19.545, 19.796]
	15	27.081	27.330	[27.036, 27.365]		15	25.120	25.384	[25.075, 25.420]
	20	32.486	32.810	[32.435, 32.850]		20	28.810	29.162	[28.760, 29.202]
5	25	36.854	37.261	[36.798, 37.304]	10	25	30.874	31.286	[30.819, 31.328]
	30	40.965	41.446	[40.903, 41.494]		30	31.636	32.051	[31.578, 32.095]
	35	43.080	43.605	[43.014, 43.655]		35	31.745	32.148	[31.686, 32.192]
	40	45.104	45.674	[45.034, 45.727]		40	31.748	32.150	[31.689, 32.194]

Note: Numerical results for $T = 300$ and reduced numbers of simulation paths.

they could not be exercised under the unit volume constraint (e.g., the additional 10th right when the refraction period is $\delta = 6$).

To test the accuracy of the algorithm for a larger number of rights, we price the same off-peak swing option with $T = 300$ days. To balance the ratio of computational costs and accuracy, we choose $N_1 = 10{,}000$, $N_2 = 30{,}000$, $N_3 = 1{,}000$, and $N_4 = 50$. Moreover, the computation of the upper bounds are based on the Snell envelope approximation

$$Y_j^l = \max\left\{\max_{1\le n\le v_j\wedge l}\left(nZ_j + C_j^{\delta,l-n}\right), C_j^{1,l}\right\}$$

instead of $\underline{Y}_j^l = \mathbb{E}_j \sum_{p=1}^l Z_{\tau_j^{p,l}}$. Table 4.5 depicts the numerical results for $T = 300$. The choice of less simulation paths does not significantly reduce the accuracy because the relative length of the 95% confidence intervals are in a very satisfactory range range of $1.2\% - 1.6\%$ even for up to 40 exercise rights.

TABLE 4.6
Off-Peak Volume Constraints

δ	L	$Y_0^{up,L}$	Upper bound using Bender (2011b)
1	1	1.86485 (0.0019)	1.8638 (0.0019)
1	2	3.40832 (0.003)	3.4078 (0.003)
1	3	4.73509 (0.0037)	4.7368 (0.0038)
1	4	5.90956 (0.0043)	5.9170 (0.0045)
1	5	6.96665 (0.00047)	6.98 (0.0052)
1	6	7.92669 (0.005)	7.9470 (0.0058)
1	7	8.80743 (0.0055)	8.8327 (0.0062)
1	8	9.61643 (0.0058)	9.6493 (0.0069)
1	9	10.3642 (0.0061)	10.4040 (0.0074)
1	10	11.0553 (0.00064)	11.1035 (0.0079)

Note: A comparison between our upper bounds and the upper bounds obtained via the algorithm from Bender (2011b) for the case of unit refraction period. Standard deviations are displayed in parentheses.

In the case of unit refraction period $\delta = 1$, upper price bounds for the off-peak swing option can also be computed by the dual representation of Bender (2011b) for the marginal price of a multiple exercise option. This approach generalizes the ideas of Meinshausen and Hambly (2004): An upper biased estimate for the marginal price of having an additional lth right is computed in terms of one martingale and $(l-1)$ stopping times. By summing up these upper bounds for the marginal prices, one finally ends with an upper biased estimate for the option price. This approach is based on the fact that, roughly speaking, under the assumption of a trivial refraction period ($\delta = 1$) optimal exercise times for the problem with $(l-1)$ rights are also optimal for the problem with l rights, if one adds one additional exercise time in a clever way. This is clearly not possible, in general, in the presence of a nontrivial refraction period. So, it seems that this alternative approach cannot be easily generalized to include refraction periods.

Table 4.6 compares the upper bounds obtained using our method and the method from Bender (2011b) for the unit refraction case $\delta = 1$ for the case $T = 50$. We mention that in Table 4.6, the variance reduction method from Remark 4.2 is not applied for both algorithms. As both methods are run with the same number of sample paths and the nested maximum in our method can be efficiently calculated by the recursion formula in Proposition 3.8, the computational effort is roughly the same for both algorithms. We observe that, as the number of exercise rights increases, our method of directly tackling the Snell envelope produces upper bounds that become lower than the algorithm tackling the marginal values from Bender (2011b). Whereas the differences for $L = 1, \ldots, 4$ are numerically not significant yet, they however become notable starting from $L = 5$ and are striking, e.g., $L = 10$. We also note that the larger L, the better our method performs concerning the variance of the upper bounds. At large, we conclude that if one is mainly interested in the price (and not the marginal price) of the swing option, our new method performs better than the algorithm from Bender (2011b). Moreover, it is applicable to a larger class of problems.

REFERENCES

ALEKSANDROV, N., and B. M. HAMBLY (2010): A Dual Approach to Multiple Exercise Option Problems under Constraints, *Math. Methods Oper. Res.* 71(3), 503–533.

ALMGREM, R., and N. CHRISS (2001): Optimal Execution of Portfolio Transactions, *J. Risk* 3, 5–39.

ANDERSEN, L., and M. BROADIE (2004): A Primal-Dual Simulation Algorithm for Pricing Multi-Dimensional American Options, *Manage. Sci.* 50(9), 1222–1234.

BALDER, S., A. MAHAYNI, and J. SCHOENMAKERS (2011): Primal Dual Linear Monte Carlo Algorithm for Multiple Stopping—An Application to Flexible Caps. WIAS Preprint 1666.

BARDOU, O., S. BOUTHEMY, and G. PAGÈS (2009): Optimal Quantization for the Pricing of Swing Options, *Appl. Math. Finance* 16(1-2), 183–217.

BENDER, C. (2011a): Primal and Dual Pricing of Multiple Exercise Options in Continuous Time, *SIAM J. Financ. Math.* 2, 562–586.

BENDER, C. (2011b): Dual Pricing of Multi-Exercise Options under Volume Constraints, *Finance Stoch.* 15(1), 1–26.

BENDER, C., and J. SCHOENMAKERS (2006): An Iterative Method for Multiple Stopping: Convergence and Stability, *Adv. Appl. Probab.* 38(3), 729–749.

BENTH, F., J. KALLSEN, and T. MEYER-BRANDIS (2007): A Non-Gaussian Ornstein–Uhlenbeck Process for Electricity Spot Price Modeling and Derivatives Pricing, *Appl. Math. Finance* 14, 153–169.

BROWN, D. B., J. E. SMITH, and P. SUN (2010): Information Relaxations and Duality in Stochastic Dynamic Programs, *Oper. Res.* 58(4, part 1), 785–801.

CARRIERE, J. (1996): Valuation of the Early-Exercise Price for Options Using Simulations and Nonparametric Regression, *Insur.: Math. Econ.* 19(1), 19–30.

DAVIS, M., and I. KARATZAS (1994): A Deterministic Approach to Optimal Stopping, in *Probability, Statistics and Optimisation. A Tribute to Peter Whittle*, F. P. Kelly, ed., Chichester: Wiley. Wiley Series in Probability and Mathematical Statistics. Probability and Mathematical Statistics, pp. 455–466.

GATHERAL, J., and A. SCHIED (2011): Optimal Trade Execution under Geometric Brownian Motion in the Almgren and Chriss Framework, *Int. J. Theor. Appl. Finance* 14(3), 353–368.

GYURKO, L., B. HAMBLY, and J. WITTE (2011): Monte Carlo Methods via a Dual Approach for Some Discrete Time Stochastic Control Problems. Working Paper, University of Oxford.

HAARBRÜCKER, G., and D. KUHN (2009): Valuation of Electricity Swing Options by Multistage Stochastic Programming, *Automatica* 45, 889–899.

HAMBLY, B., S. HOWISON, and T. KLUGE (2009): Modeling Spikes and Pricing Swing Options in Electricity Markets, *Quant. Finance* 9, 937–949.

HAUGH, M., and L. KOGAN (2004): Pricing American Options: A Duality Approach, *Oper. Res.* 52(2), 258–270.

JAILLET, P., E. RONN, and S. TOMPAIDIS (2004): Valuation of Commodity-Based Swing Options, *Manag. Sci.* 50, 909–921.

KOBYLANSKI, M., M.-C. QUENEZ, and E. ROUY-MIRONESCU (2011): Optimal Multiple Stopping Time Problem, *Ann. Appl. Probab.* 21(4), 1365–1399.

LONGSTAFF, F., and E. SCHWARTZ (2001): Valuing American Options by Simulation: A Simple Least-Squares Approach, *Rev. Finan. Stud.* 14(1), 113–147.

LUCIA, J., and E. SCHWARTZ (2002): Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange, *Rev. Deriv. Res.* 5, 5–50.

MEINSHAUSEN, N., and B. HAMBLY (2004): Monte Carlo Methods for the Valuation of Multiple-Exercise Options, *Math. Finance* 14(4), 557–583.

ROGERS, L. C. G. (2002): Monte Carlo Valuation of American Options, *Math. Finance* 12(3), 271–286.

ROGERS, L. C. G. (2007): Pathwise Stochastic Optimal Control, *SIAM J. Control Optim.* 46(3), 1116–1132.

SCHIED, A., and A. SLYNKO (2011): Some Mathematical Aspects of Market Impact Modeling, in *Surveys in Stochastic Processes*, J. Blath, P. Imkeller, and S. Roelly, eds., Proceedings of the 33rd SPA, Berlin. EMS Series of Congress Reports, pp. 153–179.

SCHOENMAKERS, J. (2012): A Pure Martingale Dual for Multiple Stopping, *Finance Stoch.* 16, 319–334.

TSITSIKLIS, J., and B. VAN ROY (2001): Regression Methods for Pricing Complex American Style Options, *IEEE Trans. Neural. Net.* 12(14), 694–703.

MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 371–399

ON THE CONSISTENCY OF REGRESSION-BASED MONTE CARLO METHODS FOR PRICING BERMUDAN OPTIONS IN CASE OF ESTIMATED FINANCIAL MODELS

ANDREAS FROMKORTH AND MICHAEL KOHLER

Technische Universität Darmstadt

In many applications of regression-based Monte Carlo methods for pricing, American options in discrete time parameters of the underlying financial model have to be estimated from observed data. In this paper suitably defined nonparametric regression-based Monte Carlo methods are applied to paths of financial models where the parameters converge toward true values of the parameters. For various Black–Scholes, GARCH, and Levy models it is shown that in this case the price estimated from the approximate model converges to the true price.

KEY WORDS: American options, consistency, least squares estimates, nonparametric regression, robustness, regression-based Monte Carlo methods.

1. INTRODUCTION

In this paper we study the problem of numerical evaluation of an American option in discrete time (also called Bermudan option). The holder of such an option has the right to buy or sell the underlying asset for a given strike price at one of the time points $0, 1, \ldots, T$, where T is the so-called majority of the option. It is well known that in complete and arbitrage-free markets the price V_0 of such an option is given by a solution of the optimal stopping problem

$$V_0 = \sup_{\tau \in \mathcal{T}(0,\ldots,T)} \mathbf{E}\{f_\tau(X_\tau)\} \tag{1.1}$$

(cf., e.g., Karatzas and Shreve 1998). Here, f_t is the discounted payoff function, the underlying stochastic process is given by $X_0, X_1, \ldots, X_T$, and $\mathcal{T}(0, \ldots, T)$ is the class of all $\{0, \ldots, T\}$-valued stopping times, i.e., $\tau \in \mathcal{T}(0, \ldots, T)$ is a measurable function of $X_0, \ldots, X_T$ satisfying

$$\{\tau = \alpha\} \in \mathcal{F}(X_0, \ldots, X_\alpha) \quad \text{for all} \quad \alpha \in \{0, \ldots, T\}.$$

Throughout this paper, we assume $X_0 = x_0$ *a.s.* for some $x_0 \in \mathbb{R}^d$, i.e., we start at time zero with some fixed value. Furthermore, we assume that $X_0, X_1, \ldots, X_T$ is a $\mathbb{R}^d$-valued Markov process recording all necessary information about financial variables including prices of the underlying assets as well as additional risk factors driving stochastic volatility or stochastic interest rates. Neither the Markov property nor the form of the payoff as

The authors thank Zizhen Huang and anonymous referees for various suggestions substantially improving the presentation of the results.

Manuscript received Febuary 2012; final revision received October 2012.

Address correspondence to Andreas Fromkorth, Stegmann Systems GmbH, Raiffeisenstr 2 // C1, C2, Rodgau 63110, Germany; e-mail: andreas.fromkorth@stegmannsystems.com.

DOI: 10.1111/mafi.12025

a function of the state X_t are very restrictive and can often be achieved by including supplementary variables.

One way to compute (1.1) is to determine an optimal stopping rule $\tau^* \in \mathcal{T}(0, \dots, T)$ satisfying

$$V_0 = \mathbf{E}\{f_{\tau^*}(X_{\tau^*})\}. \tag{1.2}$$

Let

$$q_t(x) = \sup_{\tau \in \mathcal{T}(t+1,\dots,T)} \mathbf{E}\{ f_\tau(X_\tau) \mid X_t = x\} \tag{1.3}$$

be the so-called continuation value describing the value of the option at time t given $X_t = x$ in case of holding the option rather than exercising it. It follows from the general theory of optimal stopping (cf., e.g., Chow, Robbins, and Siegmund 1971 or Shiryayev 1978) that an optimal stopping rule can be defined by

$$\tau^* = \inf\{s \geq 0 : q_s(X_s) \leq f_s(X_s)\} \tag{1.4}$$

(cf., e.g., section 8.1 in Glasserman 2004 or theorem 1 in Kohler 2010).

The Markov property implies the dynamic programming equations

$$\begin{aligned} &q_T(x) = 0, \\ &q_t(x) = \mathbf{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \mid X_t = x\} \qquad (t = 0, 1, \dots, T-1) \end{aligned} \tag{1.5}$$

(cf., e.g., section 8.1 in Glasserman 2004 or theorem 2 in Kohler 2010). In general, these conditional expectations cannot be computed in applications. The basic idea of regression-based Monte Carlo methods for pricing American options is to apply recursively regression estimates to artificially created samples of

$$(X_t, \max\{f_{t+1}(X_{t+1}), \hat{q}_{n,t+1}(X_{t+1})\}) \tag{1.6}$$

to construct estimates $\hat{q}_{n,t}$ of q_t. This kind of recursive estimation scheme was first proposed by Carrier (1996) for the estimation of so-called value functions. In Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001) it was used in connection with parametric regression to construct estimates of continuation values. Various nonparametric regression estimates have been applied for the estimation of continuation values in Egloff (2005), Egloff, Kohler, and Todorovic (2007), Kohler (2008), Belomestny (2011a), Kohler and Krzyżak (2009), and Kohler, Krzyżak, and Todorovic (2010). There results concerning consistency and rate of convergence of the resulting estimates of the price of the option have been derived. An alternative simulation-based optimization approach for pricing American was proposed and studied in Belomestny (2011b).

From the theoretical point of view there is still one important problem. In applications the distribution of the stock values is unknown. Usually in practice one considers a stochastic model for the stock values (e.g., a Black–Scholes model), estimates the model parameters (in this case the volatility of the underlying asset), and generates sample paths with this estimated distribution. Here it is assumed that the real model is known, but the real model parameters are unknown. So instead of (1.6), artificially generated samples of

$$(\bar{X}_t, \bar{Y}_t) = (\bar{X}_t, \max\{f_{t+1}(\bar{X}_{t+1}), \quad \hat{q}_{n,t+1}(\bar{X}_{t+1})\})$$

are given, where it is assumed that the distribution of $\bar{X}_t$ is close to the distribution of X_t in the sense that it is generated with the same model but slightly different values of the parameters.

In the sequel we investigate how the estimated price of the option behaves in case that the parameters of a given model converge to the true parameter values. Our main result will be that for suitably defined least squares estimates of the continuation values the estimated prices in case of a Black–Scholes model, of a GARCH-model, or of a Levy model converge to the true price in this situation.

Model calibration won't be considered in this paper. Usually, the parameters of the underlying price model are calibrated to option data of plain vanilla options. The corresponding calibration problem in the case of Levy models was rigorously studied in Belomestny and Reiß (2010).

The outline of this paper is as follows: The definition of the nonparametric regression-based Monte Carlo methods which we analyze is given in Section 2. The main result concerning consistency of the estimate in case of an application with estimated parameters of a Black–Scholes, of a GARCH, and of a Levy model is described in Section 3. In Section 4 the results are illustrated by simulated data. The proofs are given in Section 5.

2. DEFINITION OF THE ESTIMATES

In the sequel, we consider a $\mathbb{R}^d$-valued stochastic process

$$(X_t)_{t=0,\dots,T}$$

containing the log prices of the underlyings and at least all informations needed to compute the payoff for arbitrary t. We denote the payoff function with respect to $(X_t)_{t=0,\dots,T}$ at time t by f_t and the corresponding continuation value by q_t.

The consideration of log prices instead of ordinary prices simplifies the integrability condition of the sample paths, which we will need in our theoretical results.

Instead of $(X_t)_{t=0,\dots,T}$ we have only given artificially generated samples of an estimate of $(X_t)_{t=0,\dots,T}$ denoted by

$$\left(\bar{X}_{t,i}^{(n)}\right)_{t=0,\dots,T}, \quad i = 1, \dots, N(n).$$

We will use these sample paths to estimate the continuation values q_t $(t = 0, \dots, T)$. Here, n is the sample size used to calibrate the distribution of the process $(X_t)_{t=0,\dots,T}$ and $N(n)$ denotes the Monte Carlo sample size (so it is possible to choose the Monte Carlo sample size depending on the sample size that has been used to calibrate the model). In situations where the dependency on n is not so relevant, we will sometimes abbreviate $N(n)$ by simply writing N.

We start with

$$\hat{q}_{N,T} = 0. \tag{2.1}$$

Given the estimate $\hat{q}_{N,t+1}$ of q_{t+1} for some $t \in \{0, 1, \dots, T-1\}$, we estimate the conditional expectation in (1.5) by applying the principle of least squares to the data

$$\left\{\left(\bar{X}_{t,i}^{(n)}, \max\left\{f_{t+1}\left(\bar{X}_{t+1,i}^{(n)}\right), \hat{q}_{n,t+1}\left(\bar{X}_{t+1,i}^{(n)}\right)\right\}\right) \quad : \quad i = 1, \dots, N(n)\right\}.$$

To do this, we choose a set $\mathcal{F}_N$ of functions $f : \mathbb{R}^d \to \mathbb{R}$ and define

$$\hat{q}_{N,t}(\cdot) = \arg\min_{g\in\mathcal{F}_N} \frac{1}{N}\sum_{i=1}^{N} \left|g\left(\bar{X}_{t,i}^{(n)}\right) - \max\left\{f_{t+1}\left(\bar{X}_{t+1,i}^{(n)}\right), \hat{q}_{N,t+1}\left(\bar{X}_{t+1,i}^{(n)}\right)\right\}\right|^2,$$

where $x_0 = \arg\min_{x\in\mathcal{D}} h(x)$ means $x_0 \in \mathcal{D}$ and $h(x_0) = \min_{x\in\mathcal{D}} h(x)$ for a function $h : \mathcal{D} \to \mathbb{R}$. Here we assume, for simplicity, that the minimum exists, but we do not require its uniqueness.

To compute the least squares estimate above, we have to specify suitable function sets $\mathcal{F}_N$. In the sequel we will use sets of polynomial spline functions.

Choose $M \in \mathbb{N}_0$, $K_N \in \mathbb{N}$, $A, B \in \mathbb{R}$ with $A < B$ and set $u_k = A + k \cdot (B - A)/K_N$ for $k \in \mathbb{Z}$. Let B_{j,M,K_N}, $j = 1, \ldots, K_N + M$ be the B-spline with support $[u_j, u_{j+M+1}]$ with respect to the knot sequence $(u_k)_{k\in\mathbb{Z}}$ (see, e.g., de Boor 1978, chapter IX or Györfi et al. 2002, section 14.1). The spline spaces which we will use for our estimates in case $d = 1$ will be defined as subspaces of

$$S_{K_N,M}([A, B]) = \left\{ \sum_{j\in\mathbb{Z}:\operatorname{supp}(B_{j,M,K_N})\cap[A,B]\neq\emptyset} a_j \cdot B_{j,M,K_N} : j \in \mathbb{Z}, a_j \in \mathbb{R} \right\}.$$

Restricted on $[A, B]$ the space $S_{K_N,M}([A, B])$ consists of all functions f that are $(M - 1)$-times continuously differentiable on $[A, B]$ and that are on each interval $[u_j, u_{j+1})$ equal to a polynomial of degree M (or less). For our function space we restrict the coefficients in $S_{K_N,M}([A, B])$ such that the functions are bounded and Lipschitz continuous. More precisely, we set

$$\text{(2.2)} \quad S_{K_N,M,\beta_N,\gamma_N}([A, B]) = \Bigg\{ \sum_{j\in\mathbb{Z}} a_j B_{j,M,K_N} : |a_j| \le \beta_N, |a_j - a_{j-1}| \le \gamma_N/K_N,$$
$$a_j = 0 \text{ if } \operatorname{supp}(B_{j,M,K_N}) \cap [A, B] = \emptyset \quad (j \in \mathbb{Z}) \Bigg\}$$

for some $\beta_N, \gamma_N > 0$. By standard results on B-splines and its derivatives (cf., e.g., lemmas 14.4 and 14.6 in Györfi et al. 2002) it can be shown that each function in $S_{K_N,M,\beta_N,\gamma_N}([A, B])$ is bounded in absolute value by β_N and Lipschitz continuous with Lipschitz constant γ_N, which we will need later in the proofs.

In case of higher dimensions we will use tensor product B-splines. Let e_i be the ith unit vector ($i = 1, \ldots, d$). For a multi-index $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, we define the multivariate B-spline $B_k : \mathcal{R}^d \to \mathbb{R}$ of degree $\mathbf{M} = (M_1, \ldots, M_d) \in \mathbb{N}^d$ by

$$B_{k,\mathbf{M},\mathbf{K}_N}(x^{(1)}, \ldots, x^{(d)}) = \prod_{i=1}^{d} B_{k_i,M_i,K_{i,N}}(x^{(i)}) \qquad (x^{(1)}, \ldots, x^{(d)} \in \mathbb{R}),$$

where $\mathbf{K}_N = (K_{1,N}, \ldots, K_{d,N}) \in \mathbb{N}^d$.

Accordingly, we define

$$
\begin{aligned}
(2.3)\qquad & S_{K_N,M,\beta_N,\gamma_N}\left(\bigtimes_{i=1}^{d}[A_i, B_i]\right)\\
&= \Bigg\{ \sum_{j\in\mathbb{Z}^d} a_j B_{j,M,K_N} : |a_j| \le \beta_n, |a_j - a_{j-e_i}| \le \frac{\gamma_n}{\sqrt{d}K_{N,i}} \ (i = 1, \dots, d)\\
&\quad a_j = 0 \ \text{ if } \operatorname{supp}(B_{j,M,K_N}) \cap \bigtimes_{i=1}^{d}[A_i, B_i] = \emptyset \quad (j \in \mathbb{Z}^d) \Bigg\}.
\end{aligned}
$$

The definition of the B-splines implies that $S_{\mathbf{K}_N,\mathbf{M},\beta_N,\gamma_N}([\bigtimes_{i=1}^{d}[A_i, B_i])$ is a subset of a linear vector space of dimension $\prod_{i=1}^{d}(K_i + M_i + 1)$. Furthermore, it follows as above that the functions in $S_{\mathbf{K}_N,\mathbf{M},\beta_N,\gamma_N}(\bigtimes_{i=1}^{d}[A_i, B_i])$ are bounded in absolute value by β_n and are Lipschitz continuous with Lipschitz constant γ_N.

To compute $\hat{q}_{N,t}$ for the function spaces given by (2.2), respectively, (2.3) numerically, one has to solve a quadratic programming problem under linear constraints. For literature about this topic, we refer to Gill and Murray (1983) and Goldfarb and Idnani (1983). In our simulation in Section 4 we speed up the computation of the estimate by choosing β_N and γ_N rather large. In this case, the least squares estimate computed without imposing restrictions on the coefficients, which can be quickly computed by solving a linear equation system (cf., e.g., chapter 10 in Györfi et al. 2002), is contained in the space (2.2) respectively (2.3) and hence coincides with the estimate in the above theorems.

Given the above estimates of the continuation values, we can estimate the price of the option by

$$\hat{V}_{0,N} = \max\{f_0(x_0), \hat{q}_{N,0}(x_0)\}.$$

Because

$$
\begin{aligned}
|\hat{V}_{0,N} - V_0| &= |\max\{f_0(x_0), \hat{q}_{N,0}(x_0)\} - \max\{f_0(x_0), q_0(x_0)\}| \le |\hat{q}_{N,0}(x_0) - q_0(x_0)|\\
(2.4)\qquad &= \left(\int \left|\hat{q}_{N,0}(u) - q_0(u)\right|^2 \mathbf{P}_{X_0}(du)\right)^{1/2}
\end{aligned}
$$

(where the last equality follows from $X_0 = x_0$ *a.s.*), the error of this estimate tends to zero whenever the so-called L_2 error of our estimate of q_0 tends to zero.

Alternatively, we can estimate the price of the option by so-called lower estimates defined by Monte Carlo estimates of the expected payoff of a plug-in version of the stopping rule (1.4), cf., e.g., subsection 8.6.1 in Glasserman (2004). It follows from proposition 1.3 in Belomestny (2011a) that in this case the error of our estimated price of the option tends also to zero if the L_2 errors of the above estimates of the continuation values tend to zero.

3. MAIN RESULTS

In this section we consider three different models for the stock values and present for each model a consistency result of our estimation procedure applied to paths of versions of the model where the parameter values converge to the true values.

3.1. A Black–Scholes Model

In this subsection the stock values are modeled via Black–Scholes theory, and the log-prices are given by $X_{k,t} = \log Z_{k,t}$, where

$$Z_{k,t} = z_{k,0} \cdot e^{r \cdot t} \cdot e^{\sum_{j=1}^{d} \left(\sigma_{k,j} \cdot W_j(t) - \frac{1}{2}\sigma_{k,j}^2 t\right)} \quad (k = 1, \dots, d). \tag{3.1}$$

Here, $r > 0$ is the riskless interest rate, $\sigma_k = (\sigma_{k,1}, \dots, \sigma_{k,d})^T$ is the volatility of the kth stock, $z_{k,0}$ is the initial stock price of the kth stock, and $\{W_j(t) : t \in \mathbb{R}_+\}$ $(j = 1, \dots, d)$ are independent Wiener processes. Because $W_j(1), W_j(2) - W_j(1), \dots, W_j(T) - W_j(T - 1)$ are independent standard normally distributed random variables, we can define $Z_{k,t}$ $(k = 1, \dots, m, t = 0, \dots, T)$ also by

$$Z_{k,0} = z_{k,0} \quad (k = 1, \dots, d)$$

and by

$$Z_{k,t+1} = Z_{k,t} \cdot e^{r} \cdot e^{\sum_{j=1}^{d} \left(\sigma_{k,j} \cdot \epsilon_{t+1,j} - \frac{1}{2}\sigma_{k,j}^2\right)} \quad (k = 1, \dots, d, t = 0, \dots, T - 1),$$

where $(\epsilon_{t,j})_{t \in \{1,\dots,T\}, j \in \{1,\dots,d\}}$ are independent standard normally distributed random variables.

In the sequel we assume that instead of sample paths from

$$(X_t)_{t=0,\dots,T} = ((X_{1,t}, \dots, X_{d,t}))_{t=0,\dots,T},$$

we observe

$$\left(\bar{X}_{t,i}^{(n)}\right)_{t=0,\dots,T} \quad (i = 1, \dots, N(n)),$$

where

$$\bar{X}_{t,i}^{(n)} = \left(\bar{X}_{1,t,i}^{(n)}, \dots, \bar{X}_{d,t,i}^{(n)}\right)^T = \left(\log \bar{Z}_{1,t,i}^{(n)}, \dots, \log \bar{Z}_{d,t,i}^{(n)}\right)^T$$

is given by

$$\bar{Z}_{k,0,i}^{(n)} = z_{k,0} \quad (k = 1, \dots, d)$$

and by

$$\bar{Z}_{k,t+1,i}^{(n)} = \bar{Z}_{k,t,i}^{(n)} \cdot e^{r} \cdot e^{\sum_{j=1}^{d} \left(\hat{\sigma}_{k,j}^{(n)} \cdot \epsilon_{t+1,j,i} - \frac{1}{2}\left(\hat{\sigma}_{k,j}^{(n)}\right)^2\right)} \quad (k = 1, \dots, d, t = 0, \dots, T - 1)$$

for some independent standard normally distributed random variables $\epsilon_{t+1,j,i}$.

THEOREM 3.1. *Assume that the discounted payoff function f_t with respect to the above defined log price process $(X_t)_{t=0,\dots,T}$ is bounded and Lipschitz continuous. Let V_0 and q_t be the corresponding price of the option and continuation values. Let $M \in \mathbb{N}$, $\mathbf{M} = \sum_{i=1}^{d} e_i M$, $K_{N(n)} \in \mathbb{N}$, $\mathbf{K}_{N(n)} = \sum_{i=1}^{d} e_i K_{N(n)}$, $\beta_{N(n)} > 0$, $\gamma_{N(n)} > 0$ and $A_n > 0$ and let the estimate be defined as in Section 2 with*

$$\mathcal{F}_{N(n)} = S_{\mathbf{K}_{N(n)}, \mathbf{M}, \beta_{N(n)}, \gamma_{N(n)}}\left([-A_{N(n)}, A_{N(n)}]^d\right).$$

Assume that the parameters of the function spaces satisfy

$$A_{N(n)} \to \infty, \beta_{N(n)} \to \infty, \gamma_{N(n)} \to \infty, \frac{A_{N(n)}}{K_{N(n)}} \to 0, \frac{\beta_{N(n)}^5 \cdot A_{N(n)}^d K_{N(n)}^d}{n} \to 0 \qquad (n \to \infty).$$

If the parameters of the estimated model converge to the true parameter values in the sense that

$$\gamma_{N(n)} \cdot \left(\hat{\sigma}_{k,j}^{(n)} - \sigma_{k,j}\right) \to 0 \quad (n \to \infty)$$

for all $k, j \in \{1, \dots, d\}$ *then we have for* $t = 0, \dots, T$

$$\int \left|\hat{q}_{N(n),t}(x) - q_t(x)\right|^2 \mathbf{P}_{X_t}(dx) \to 0 \qquad (n \to \infty) \textit{ in probability}$$

and, in addition,

$$\hat{V}_{0,N(n)} \to V_0 \qquad (n \to \infty) \textit{ in probability}.$$

3.2. A GARCH Model

Next we present results for a price process, where the volatility is modeled by a GARCH time series, which was introduced by Bollerslev (1986). We consider this in the form proposed in Duan (1995), where the GARCH process is modified in such a way that the discounted price process is a martingale.

Here the price process $\{S_t\}_{t=0,1,\dots}$ of a stock is modeled by

$$S_t = x_0 \cdot \exp\left(r \cdot t - \frac{1}{2}\sum_{j=1}^{t} h_j + \sum_{j=1}^{t} \sqrt{h_j} \cdot \epsilon_j\right),$$

where $x_0 \in \mathbb{R}_+$ is the value of the stock at time zero, $r > 0$ is the riskless interest rate, $(\epsilon_j)_{j \in \mathbb{Z}}$ are independent standard normally distributed random variables, and where the (random) volatility h_t of the process satisfies

$$h_t = 0 \quad \text{for} \quad t \le 0$$

and

$$h_t = a_0 + \sum_{j=1}^{q} a_j \cdot h_{t-j}\left(\epsilon_{t-j} - \lambda\right)^2 + \sum_{j=1}^{p} b_j \cdot h_{t-j} \quad (t \in \mathbb{N})$$

for some $p, q \in \mathbb{N}_0$ and parameters $\lambda > 0$, $a_j > 0$ $(j = 0, \dots, p)$, and $b_j > 0$ $(j = 1, \dots, p)$.

In applications the parameters λ, $a_0, \dots, a_q$, $b_1, \dots, b_p$ are unknown. Therefore, it is only possible to generate Monte Carlo samples where these parameters are estimated. Given sequences $(\hat{a}_{i,n})_{n \in \mathbb{N}}$ $(i = 0, \dots, q)$, $(\hat{b}_{i,n})_{n \in \mathbb{N}} (i = 1, \dots, p)$, $(\hat{\lambda}_n)_{n \in \mathbb{N}}$ of nonnegative real numbers, where $\hat{a}_{0,n} > 0$ for all $n \in \mathbb{N}$, and independent standard normally distributed random variables $\epsilon_{j,i}$, the samples of the error-behaved logarithmic returns are given by

$$\bar{Z}^{(n)}_{t,i} = \log(x_0) + r \cdot t - \frac{1}{2}\sum_{j=1}^{t} \bar{h}^{(n)}_{j,i} + \sum_{j=1}^{t} \sqrt{\bar{h}^{(n)}_{j,i}} \cdot \epsilon_{j,i}, \tag{3.2}$$

$$\bar{h}^{(n)}_{t,i} = \hat{a}_{0,n} + \sum_{j=1}^{q} \hat{a}_{j,n} \cdot \bar{h}^{(n)}_{t-j,i}(\epsilon_{t-j,i} - \hat{\lambda}_n)^2 + \sum_{j=1}^{p} \hat{b}_{j,n} \cdot \bar{h}^{(n)}_{t-j,i} \tag{3.3}$$

$(i = 1, \ldots, N(n))$, where we set again $\bar{h}^{(n)}_{t,j} = 0$ for $t \leq 0$.

To ensure that the price process is a Markov process, we have to extend the state space. So instead of only $\bar{Z}^{(n)}_{t,i}$, we consider

$$\bar{X}^{(n)}_{t,i} = \left(\bar{Z}^{(n)}_{t,i}, \epsilon_{t,i}, \ldots, \epsilon_{t+1-q,i}, \bar{h}^{(n)}_{t,i}, \ldots, \bar{h}^{(n)}_{t+1-\max\{p,q\},i}\right)^T. \tag{3.4}$$

THEOREM 3.2. *Assume that the payoff function f_t is bounded and Lipschitz continuous. Let V_0 and q_t be the price of the option and the continuation values corresponding to the above defined log price process $\log(S_t)$. Let $\bar{X}^{(n)}_{t,i}$ be defined by (3.4). Let $M \in \mathbb{N}, \mathbf{M} = \sum_{i=1}^{1+q+\max\{p,q\}} e_i M$, $K_{N(n)} \in \mathbb{N}$, $\mathbf{K}_{N(n)} = \sum_{i=1}^{1+q+\max\{p,q\}} e_i K_{N(n)}$, $\beta_{N(n)} > 0$, $\gamma_{N(n)} > 0$ and $A_{N(n)} > 0$. Let the estimates $\hat{q}_{N(n),t}$ be defined as in Section 2, where the function space is given by*

$$\mathcal{F}_{N(n)} = S_{\mathbf{K}_{N(n)},\mathbf{M},\beta_{N(n)},\gamma_{N(n)}}\left(\bigtimes_{i=1}^{1+q+\max\{p,q\}} [A_{i,N(n)}, B_{i,N(n)}]\right).$$

Assume

$$A_{i,N(n)} \to -\infty,\ B_{i,N(n)} \to \infty,\ \frac{\gamma_{N(n)}}{B_{i,N(n)} - A_{i,N(n)}} \to \infty, \quad \gamma_{N(n)}\frac{B_{i,N(n)} - A_{i,N(n)}}{K_{N(n)}} \to 0,$$

for $i = 1, \ldots, 1 + q + \max\{p, q\}$,

$$\beta_{N(n)} \to \infty,$$

and

$$\frac{\beta^5_{N(n)} \prod_{i=1}^{1+q+\max\{p,q\}} K_{i,N(n)}}{N(n)} \to 0 \quad (n \to \infty).$$

Then

$$\gamma_{N(n)}(\hat{a}_{i,n} - a_i) \to 0 \quad (n \to \infty), \quad \gamma_{N(n)}(\hat{b}_{j,n} - b_j) \to 0 \quad (n \to \infty)$$

for all $i \in \{0, \ldots, q\}, j \in \{1, \ldots, p\}$ and

$$\gamma_{N(n)}(\lambda - \hat{\lambda}_n) \to 0 \quad (n \to \infty)$$

imply

$$\int |\hat{q}_{N(n),t}(x) - q_t(x)|^2 \mathbf{P}_{X_{t,1}}(dx) \to 0 \quad \textit{in probability}$$

for all $t = 0, 1 \ldots, T$ *and, in addition,*

$$\hat{V}_{0,N(n)} \to V_0 \quad \textit{in probability}.$$

3.3. A Levy Model

Finally, we present a result for a Lévy processes. Here we consider the following Merton Model (cf. Merton 1976):

$$S_t = x_0 \cdot \exp\left(\mu \cdot t + \sigma \cdot W_t + \sum_{j=1}^{\pi_t} Y_j\right),$$

where $W = (W_t)_{t \in \mathbb{R}_+}$ is a Wiener process, $\pi = (\pi_t)_{t \in \mathbb{R}_+}$ is a Poisson process with parameter λ independent from W, and $Y_1, Y_2, \ldots$ are independent normally distributed random variables with mean m and variance δ^2 independent from W and π. By defining

$$\mu = r - \frac{\sigma^2}{2} - \lambda\left(\exp\left(m + \frac{\delta^2}{2}\right) - 1\right), \tag{3.5}$$

the price process with respect to the martingale measure in the sense of Merton (1976) can be written as

$$S_t = s_0 \exp\left(\mu t + \sum_{s=1}^{t} \sigma \epsilon_s + \sum_{i=1}^{\pi_t} (m + \delta \xi_i)\right), \tag{3.6}$$

where ϵ_s, ξ_i with $i, s \in \mathbb{N}$ are independent standard normal distributed and π_t is Poisson distributed with parameter λt. Again we consider the corresponding log price process

$$X_t = \log s_0 + \mu t + \sum_{s=1}^{t} \sigma \epsilon_s + \sum_{i=1}^{\pi_t} (m + \delta \xi_i).$$

As in the Black–Scholes and the GARCH case we we estimate the parameters σ, λ, m, and δ by $\hat{\sigma}_n$, $\hat{\lambda}_n$, $\hat{m}_n$, and $\hat{\delta}_n$, and consider the logarithm of the returns. According to Merton (1976) we define $\hat{\mu}_n$ by

$$\hat{\mu}_n = r - \frac{\hat{\sigma}_n^2}{2} - \hat{\lambda}_n\left(\exp\left(\hat{m}_n + \frac{\hat{\delta}_n^2}{2}\right) - 1\right)$$

and generate data

$$\bar{X}_{t,i}^{(n)} = \log(x_0) + \hat{\mu}_n t + \sum_{s=1}^{t} \hat{\sigma}_n \epsilon_{s,i} + \sum_{i=1}^{\pi_{t,i}^{(n)}} (\hat{m}_n + \hat{\delta}_n \xi_i) \tag{3.7}$$

$$(t = 0, \ldots, T, i = 1, \ldots, N(n)),$$

where $\epsilon_{i,j}$, $\xi_{i,j}$ are independent standard normal distributed and $\pi_{t,i}^{(n)}$ is a Poisson distributed random variable with parameter $\hat{\lambda}_n t$ independent from $\epsilon_{i,j}$, $\xi_{i,j}$ for all $i, j \in \mathbb{N}$. Note that the above definition of μ_n ensures the martingale property of the discounted price process.

THEOREM 3.3. *Assume that the payoff function f_t is bounded and Lipschitz continuous. Let V_0 and q_t be the price of the option and the continuation values corresponding to the above defined log price process X_t. Let $\bar{X}_{t,i}^{(n)}$ be defined by (3.7). Let $M \in \mathbb{N}$, $K_{N(n)} \in \mathbb{N}$, $\beta_{N(n)} > 0$, $\gamma_{N(n)} > 0$ and $A_{N(n)} > 0$. Letthe estimates $\hat{q}_{N(n),t}$ be defined as in Section 2, where the function space $S_{K_{N(n)},M,\beta_{N(n)},\gamma_{N(n)}}([-A_{N(n)}, A_{N(n)}])$ is defined by (2.2). Assume that*

$$A_{N(n)} \to \infty, \beta_{N(n)} \to \infty, \quad \gamma_{N(n)} \to \infty, \frac{A_{N(n)}}{K_{N(n)}} \to 0, \quad \frac{\beta_{N(n)}^5 \cdot A_{N(n)}^d K_{N(n)}^d}{N(n)} \to 0 \quad (n \to \infty).$$

Then

$$\gamma_{N(n)} \cdot (\hat{\sigma}_n - \sigma) \to 0, \quad \gamma_{N(n)} \cdot (m_n - m) \to 0, \quad \gamma_{N(n)} \cdot (\hat{\delta}_n - \delta) \to 0,$$

and

$$\gamma_{N(n)}^2 \cdot (\hat{\lambda}_n - \lambda) \to 0 \quad (n \to \infty)$$

imply

$$\int \left|\hat{q}_{N(n),t}(x) - q_t(x)\right|^2 \mathbf{P}_{X_{t,1}}(dx) \to 0 \quad \textit{in probability}$$

for all $t \in \{0, \dots, T\}$ and, in addition,

$$\hat{V}_{0,N(n)} \to V_0 \quad \textit{in probability}.$$

REMARK 3.4. It is an open problem whether one can derive rate of convergence results in the above theorems in case of estimation of smooth continuation values. In principle, it should be possible to use known results for the rate of convergency for model calibration algorithms in connection with Theorem 5.2 (which is formulated in the proof section) to derive results for the rate of convergence in the setting presented in this paper.

REMARK 3.5. It is an open problem whether one can derive similar results as in the above theorems without assuming that the payoff functions is bounded. In the proofs of the above theorems we need this boundedness assumption because we apply Theorem 1 of Fromkorth and Kohler (2010), where the boundedness of the regression function is needed.

REMARK 3.6. In practise, the parameters of the spline space can be chosen by splitting of the sample (cf. remark 4 in Fromkorth and Kohler 2010).

4. APPLICATION TO SIMULATED DATA

To demonstrate the finite sample performance of our estimate we first apply it in case of a Duan–GARCH model. To reduce the computational effort, we choose here and in the sequel the parameters β_N and γ_N so large that the least squares estimates which does not impose restrictions on the range of the parameters coincides with the least squares estimate corresponding to (2.3). Basically, this means that we skip the constraints in the optimization problem corresponding to (2.3).

With respect to the martingale measure of Duan (cf. Duan 1995) the logarithm of the price process has the form

$$\log(S_t) = \log(s_0) + rt - \frac{1}{2}\sum_{s=}^{t} h_s + \sum_{s=1}^{t} \sqrt{h_s}\epsilon_s,$$

with

$$h_t = a0 + a_1|h_{t-1}|(\epsilon_{t-1} - \lambda)^2 + b_1 h_{t,1},$$

with independent standard normal distributed random variables $\epsilon_0, \ldots, \epsilon_t$. As parameters of the real price process we take

$$a_0 = 0.0000166, \quad a_1 = 0.144, \quad b_1 = 0.776, \text{ and } \lambda = 0.7138.$$

To ensure the Markov property of the considered process we take

$$X_t = \begin{pmatrix} S_t \\ \exp(h_t) \\ \exp(\epsilon_t) \end{pmatrix}.$$

The riskless rate is assumed to be known as

$$r = 0.05/T \approx 0.001389.$$

As start vector of this process we choose

$$X_0 = \begin{pmatrix} 100 \\ 1 \\ 1 \end{pmatrix}.$$

In applications the parameters a_0, a_1, b_1, and λ are unknown and have to be estimated from historical data. To demonstrate the influence of such an estimate, we consider different error levels for each parameter.

We price a Bermudan capped-straddle option with exercise prices 70, 100, and 130. The payoff function is shown in Figure 4.1. As maturity we choose $T = 36$, so exercising is possible at the time points $t = 0, 1, \ldots, 36$.

As parameters of the spline space we take $A = (0, 0, 0)^T$, $B = (200, 2, 120)^T$, $M = (1, 1, 1)^T$, and $\beta_N = \gamma_N = 5{,}000$. The parameter $K_1 \in \{4, 10\}$ is chosen by splitting of the sample, K_2 and K_3 are set to 2. To estimate the continuation values we use $n_l = 1{,}000$ paths as learning data and $n_t = 1{,}000$ testing data. The so-computed estimates for continuation values are taken as plug-in estimates to evaluate on $n_a = 10{,}000$ newly generated paths. The arithmetic mean of these n_a paths is the estimate of one option price. This procedure is repeated 50 times for each parameter constellation. So we get for every chosen parameter set 50 option prices.

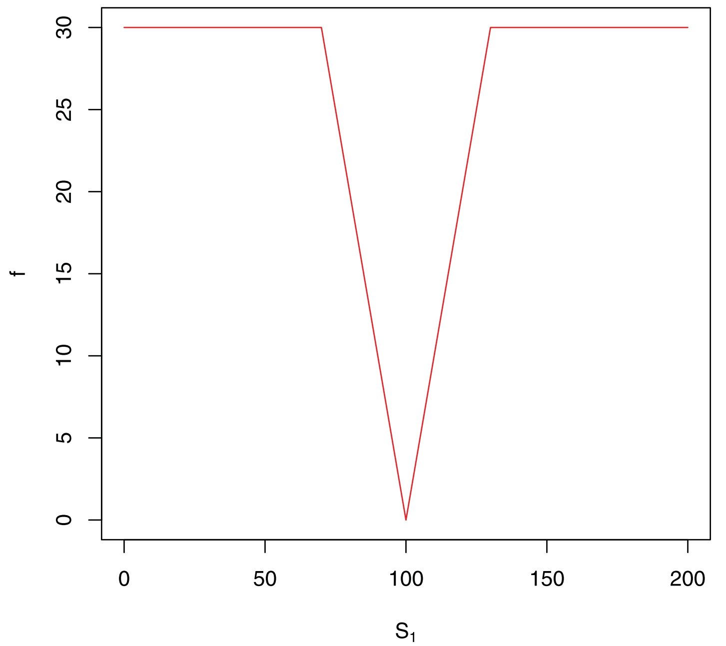


FIGURE 4.1. Payoff function of a capped-straddle with strike prices 70, 100, and 130.

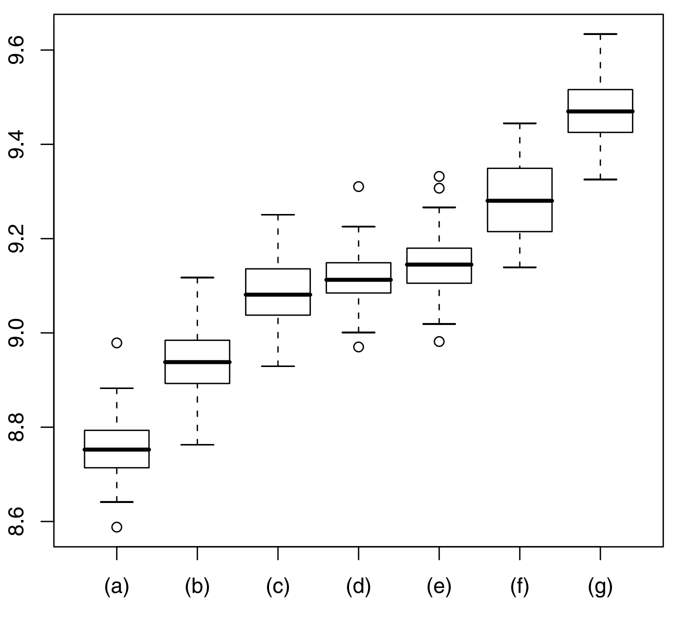


FIGURE 4.2. Simulated option prices in case of a Duan–GARCH model with (a) $a_0 = 0.0001494$, (b) $a_0 = 0.0001577$, (c) $a_0 = 0.00016434$, (d) $a_0 = 0.000166$, (e) $a_0 = 0.00016766$, (f) $a_0 = 0.0001743$, (g) $a_0 = 0.0001826$.

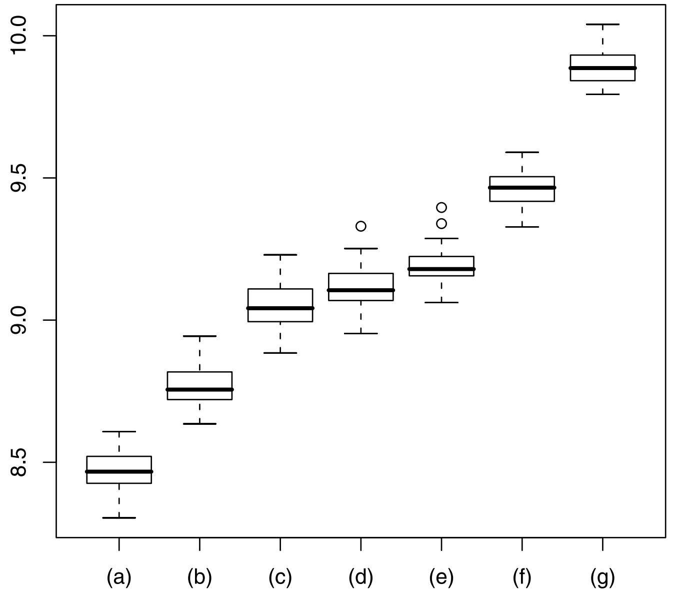


FIGURE 4.3. Simulated option prices in case of a Duan–GARCH model with (a) $a_1 = 0.1296$, (b) $a_1 = 0.1368$, (c) $a_1 = 0.14256$, (d) $a_1 = 0.144$, (e) $a_1 = 0.14544$, (f) $a_1 = 0.1512$, (g) $a_1 = 0.1584$.

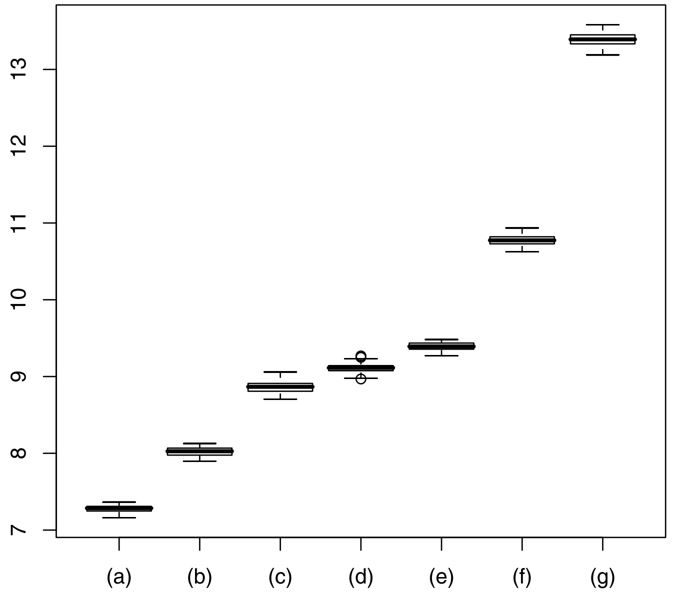


FIGURE 4.4. Simulated option prices in case of a Duan–GARCH model with (a) $b_1 = 0.69840$, (b) $b_1 = 0.7372$, (c) $b_1 = 0.76824$, (d) $b_1 = 0.776$, (e) $b_1 = 0.78376$, (f) $b_1 = 0.8148$, (g) $b_1 = 0.85360$.

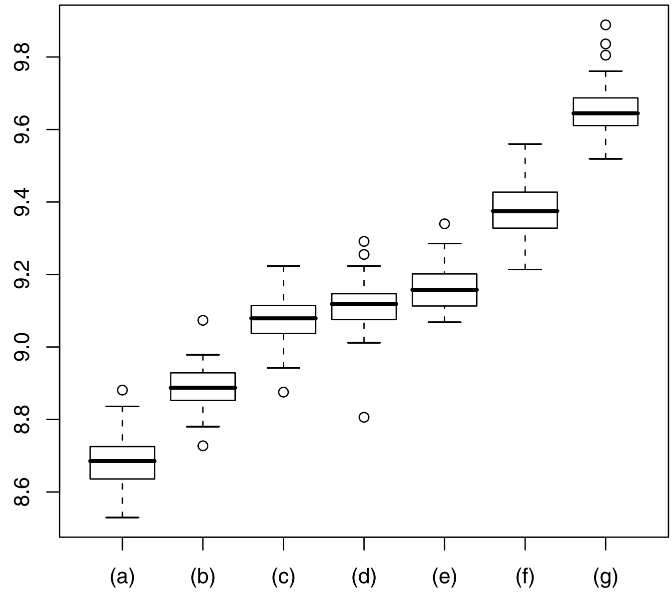


FIGURE 4.5. Simulated option prices in case of a Duan–GARCH model with (a) $\lambda = 0.642420$, (b) $\lambda = 0.678110$, (c) $\lambda = 0.706662$, (d) $\lambda = 0.713800$, (e) $\lambda = 0.720938$, (f) $\lambda = 0.749490$, (g) $\lambda = 0.785180$.

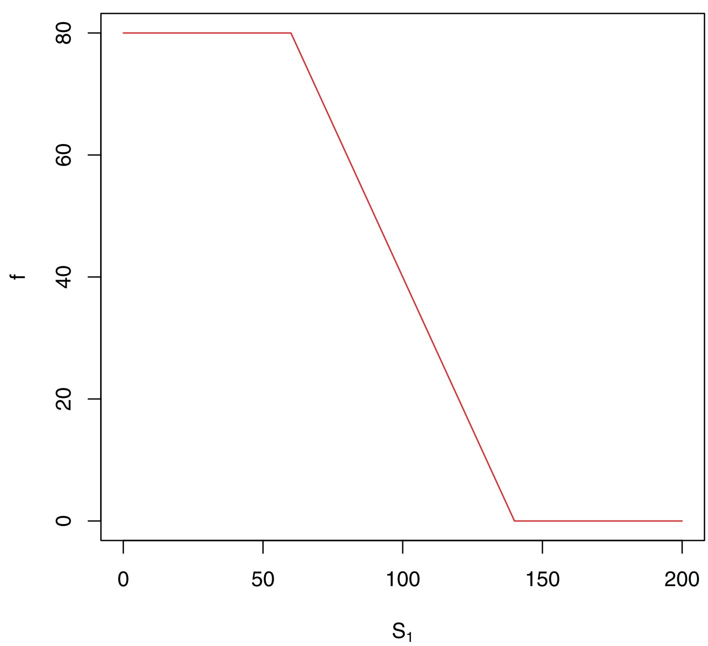


FIGURE 4.6. Payoff function of a bear spread with strikes 60 and 140.

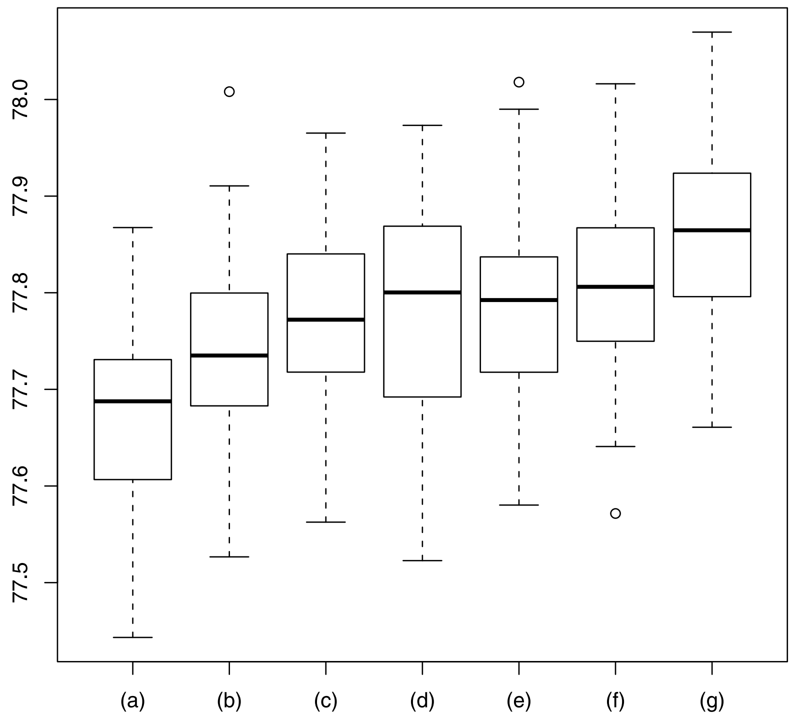


FIGURE 4.7. Simulated option prices in case of a Bermudan bear spread option in a Merton model with (a) $m = 0.18$, (b) $m = 0.19$, (c) $m = 0.198$, (d) $m = 0.2$, (e) $m = 0.202$, (f) $m = 0.21$, (g) $m = 0.22$.

Figures 4.2, 4.3, 4.4, and 4.5 show the simulation results in the form of boxplots. Each of these simulations considers different parameter levels for one of the parameters although all other parameters are set to the value of the real price process defined above.

The next simulation series considers the pricing of a Bermudan bear spread option with strikes 60 and 140. The corresponding payoff function is illustrated as Figure 4.6. The maturity is set to $T = 36$, the riskless rate is $r = 0.05/T$.

This time we use a Merton model for simulating the asset prices. We start with the log price process (3.6). Here we assume, that the real parameters are $\sigma = 0.02/\sqrt{T}$, $m = 0.2$, $\delta = 0.5$, $\lambda = 0.2$. The parameter μ is computed by (3.5).

The parameters of the spline space are $K_n \in \{4, 10\}$, $A = 0$, $B = 250$, $M_n \in \{1, 3\}$, and $\beta_N = \gamma_N = 2,000$. The number of learning data for choosing the spline space parameters is $n_l = 1,000$. The corresponding testing data is $n_t = 1,000$. The so-computed estimates for continuation values are taken as plug-in estimates to evaluate on $n_a = 10,000$ new generated paths. The arithmetic mean of these n_a paths is the estimate of one option price. This procedure is repeated 50 times for each parameter constellation. So we get for every chosen parameter set 50 option prices. Start value of the price process is 100.

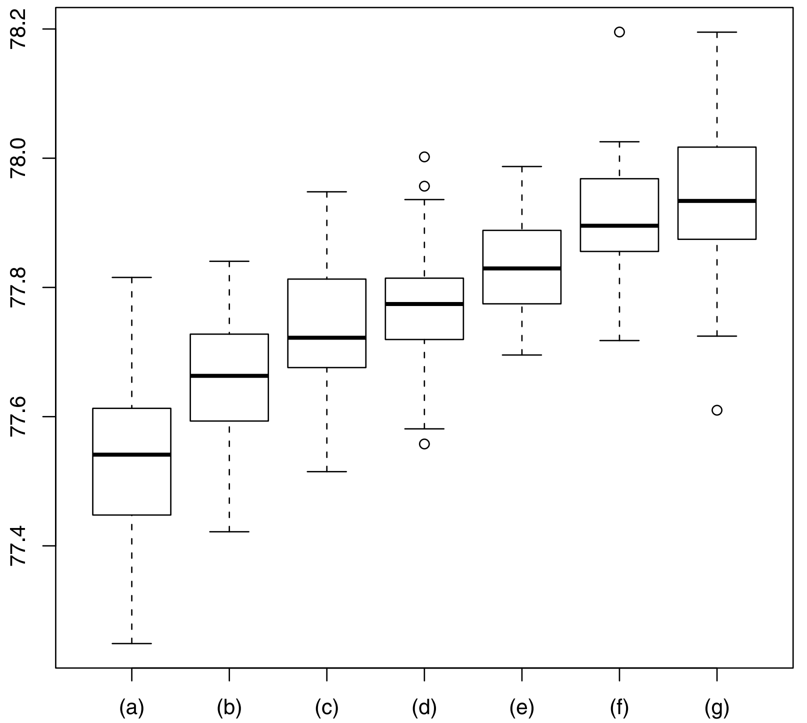


FIGURE 4.8. Simulated option prices in case of a Bermudan bear spread option in a Merton model with (a) $\delta = 0.45$, (b) $\delta = 0.475$, (c) $\delta = 0.495$, (d) $\delta = 0.5$, (e) $\delta = 0.505$, (f) $\delta = 0.525$, (g) $\delta = 0.55$.

Figure 4.7 shows the results, if we choose m as

$$0.18, 0.19, 0.198, 0.2, 0.202, 0.21, 0.22$$

and keep all other parameters fixed.

Figure 4.8 shows the results, if we choose δ as

$$0.45, 0.475, 0.495, 0.5, 0.505, 0.525, 0.55$$

and keep all other parameters fixed.

Figure 4.9 shows the results, if we choose λ as

$$0.18, 0.19, 0.198, 0.2, 0.202, 0.21, 0.22$$

and keep all other parameters fixed.

We do not consider an error in the remaining parameters, because of the similarity with the case of the Black–Scholes model.

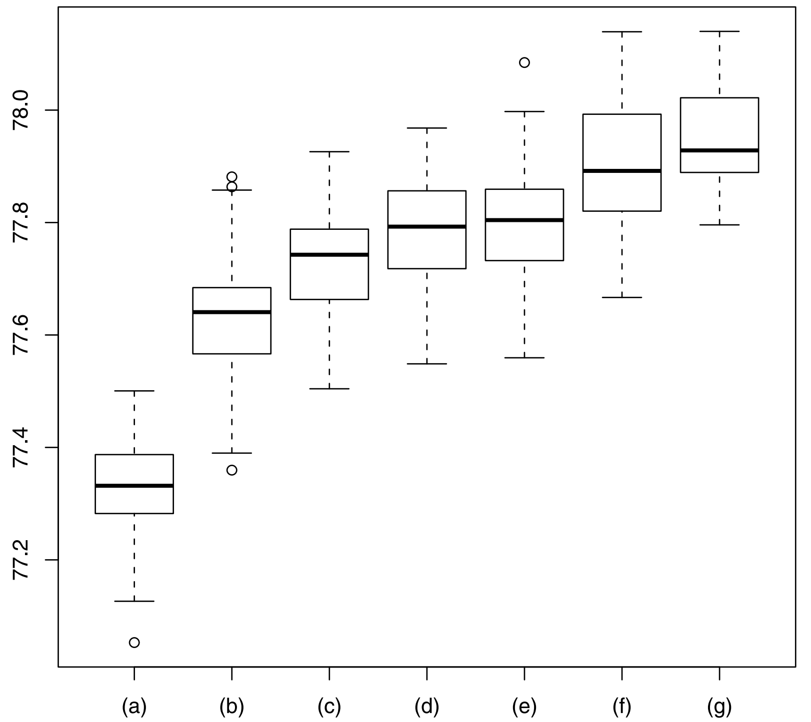


FIGURE 4.9. Simulated option prices in case of a Bermudan bear spread option in a Merton model with (a) $\lambda = 0.18$, (b) $\lambda = 0.19$, (c) $\lambda = 0.198$, (d) $\lambda = 0.2$, (e) $\lambda = 0.202$, (f) $\lambda = 0.21$, (g) $\lambda = 0.22$.

5. PROOFS

5.1. An Auxiliary Lemma

In the proofs we will need the following lemma.

LEMMA 5.1. *Let* $f_t, q_t, \bar{q}_t : \mathcal{R}^d \to \mathbb{R}$ *be functions* ($t = 0, \ldots, T$). *For given* $\mathcal{R}^d$*-valued stochastic processes* $(X_t)_{t=0,\ldots,T}$, $(\bar{X}_t)_{t=0,\ldots,T}$, *and* $t \in \{0, \ldots, T-1\}$ *define*

$$Y_t = \max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \quad \text{and}\, \bar{Y}_t = \max\{f_{t+1}(\bar{X}_{t+1}), \bar{q}_{t+1}(\bar{X}_{t+1})\}.$$

Let $s \in \{0, \ldots, T-1\}$ *and assume that* f_{s+1} *is Lipschitz continuous with Lipschitz constant L. Then*

$$|Y_s - \bar{Y}_s|^2 \leq 2L^2 \| X_{s+1} - \bar{X}_{s+1} \|^2 + 2|q_{s+1}(X_{s+1}) - \bar{q}_{s+1}(\bar{X}_{s+1})|^2.$$

Proof. Using $(a+b)^2 \le 2a^2 + 2b^2$ and $|\max\{a,b\} - \max\{a,c\}| \le |b-c|$ for $a, b, c \in \mathbb{R}$ we get

$$\begin{aligned}
|Y_s - \bar{Y}_s|^2 &\le 2|\max\{f_{s+1}(X_{s+1}), q_{s+1}(X_{s+1})\} - \max\{f_{s+1}(X_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\}|^2 \\
&\quad + 2|\max\{f_{s+1}(X_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\} - \max\{f_{s+1}(\bar{X}_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\}|^2 \\
&\le 2|q_{s+1}(X_{s+1}) - \bar{q}_{s+1}(X_{s+1})|^2 + 2|f_{s+1}(X_{s+1}) - f_{s+1}(\bar{X}_{s+1})|^2.
\end{aligned}$$

By using the Lipschitz property of f_{s+1} we get the desired result. □

5.2. A General Consistency Result

In this subsection we formulate and proof a general consistency result. Here we assume that instead of independent copies

$$(X_{t,i})_{t=0,\dots,T} \quad (i = 1, \dots, N)$$

of the underlying $\mathbb{R}^d$-valued Markov process $(X_t)_{t=0,\dots,T}$, we have given paths

$$(\bar{X}_{t,i}^{(n)})_{t=0,\dots,T} \quad (i = 1, \dots, N(n))$$

such that

$$\frac{1}{N(n)} \sum_{i=1}^{N(n)} \| \bar{X}_{t,i}^{(n)} - X_{t,i} \|^2$$

is small for all t. We define our estimates $\hat{q}_{N,t}$ and $\hat{V}_{0,N}$ as in Section 2 with a general function space $\mathcal{F}_N$. Then the following result holds.

THEOREM 5.2. *Let the discounted payoff function f_t be bounded in absolute value by some $M > 1$ and Lipschitz continuous with Lipschitz constant $L > 0$. Let $\mathcal{F}_{N(n)}$ be a subspace of a linear vector space of dimension $D_{N(n)}$consisting of functions which are bounded in absolute value by some $\beta_{N(n)} > 0$ and which are Lipschitz continuous with respect to some Lipschitz constant $\gamma_{N(n)}$. Assume that*

$$X_{1,t+1}, \dots, X_{N(n),t+1} \text{ and } \bar{X}_{1,t}^{(n)}, \dots, \bar{X}_{N(n),t}^{(n)} \text{ are independent given } X_{1,t}, \dots, X_{N(n),t} \tag{5.1}$$

for all $t = 0, \dots, T-1$. Then

$$\frac{D_{N(n)} \beta_{N(n)}^5}{N(n)} \to 0, \quad D_{N(n)} \beta_{N(n)}^3 \to \infty,$$

$$\inf_{f \in \mathcal{F}_{N(n)}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \to 0 \quad (n \to \infty) \tag{5.2}$$

for all $t = 0, \dots, T-1$ and

$$\frac{\gamma_{N(n)}^2}{n} \sum_{i=1}^{n} \| \bar{X}_{t,i}^{(n)} - X_{t,i} \|^2 \to 0 \quad \textit{in probability} \quad (n \to \infty)$$

for all $t = 0, \dots, T$ *imply*

$$\int |\hat{q}_{N(n),t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \to 0 \quad \text{in probability } (n \to \infty) \tag{5.3}$$

for $t = 0, \dots, T$ *and, in addition*

$$\hat{V}_{0,N(n)} \to V_0 \quad \text{in probability.} \tag{5.4}$$

In the proof we will apply theorem 1 of Fromkorth and Kohler (2011). In case of bounded Y, the sub-Gaussian condition there is trivially fulfilled. Using lemma 9.3 in Györfi et al. (2002), we can conclude from the result there the following lemma:

LEMMA 5.3. *Let* $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ *be independent and identically distributed* $\mathcal{R}^d \times \mathbb{R}$*-valued random vectors with* $|Y| \le \beta$ *a.s. for some* $\beta > 0$*. For each n let* $\mathcal{F}_N$ *be a subset of a linear vector space of dimension* D_n *consisting of functions* $f : \mathbb{R}^d \to \mathbb{R}$ *which are bounded in absolute value by* $\beta_{N(n)}$ *and which are Lipschitz continuous with Lipschitz constant* $\gamma_{N(n)}$*. Given an arbitrary data set*

$$\bar{\mathcal{D}}_{N(n)} = \{(\bar{X}_{1,n}, \bar{Y}_{1,n}), \dots, (\bar{X}_{N(n),n}, \bar{Y}_{N(n),n})\}$$

with the property, that $Y_1, \dots, Y_{N(n)}$ *and* $\bar{X}_{1,n}, \dots, \bar{X}_{N(n),n}$ *are independent given* $X_1, \dots, X_{N(n)}$*, define the estimate* $\bar{m}_{N(n)}$ *by*

$$\bar{m}_{N(n)}(\cdot) = \arg\min_{f \in \mathcal{F}_{N(n)}} \frac{1}{N(n)} \sum_{i=1}^{N(n)} |f(\bar{X}_{i,n}) - \bar{Y}_{i,n}|^2.$$

If

$$\frac{D_{N(n)}\beta^5_{N(n)}}{N(n)} \to 0, \quad D_{N(n)}\beta^3_{N(n)} \to \infty \quad (n \to \infty)$$

then it holds for c sufficiently large that we have for any $n \in \mathbb{N}$

$$\mathbf{P}\left\{\int |\bar{m}_{N(n)}(x) - m(x)|^2 \mu(dx) > c \cdot Z_n\right\} \le c \cdot \exp\left(-c \cdot D_{N(n)} \cdot \beta^3_{N(n)}\right)$$

and

$$\mathbf{P}\left\{\frac{1}{N(n)} \sum_{i=1}^{N(n)} |\bar{m}_{N(n)}(\bar{X}_{i,n}) - m(X_i)|^2 > c \cdot Z_n\right\} \le c \cdot \exp\left(-c \cdot D_{N(n)} \cdot \beta^3_{N(n)}\right),$$

where

$$Z_n = \frac{1}{N(n)} \sum_{i=1}^{N(n)} |Y_i - \bar{Y}_{i,n}|^2 + \gamma^2_{N(n)} \cdot \frac{1}{N(n)} \sum_{i=1}^{N(n)} \|X_i - \bar{X}_{i,n}\|^2 + \frac{D_{N(n)}\beta^5_{N(n)}}{N(n)}$$
$$+ \inf_{f \in \mathcal{F}_{N(n)}} \int |f(x) - m(x)|^2 \mu(dx).$$

Proof. The result follows directly from theorem 1 in Fromkorth and Kohler (2011) and Lemma 9.3 in Györfi et al. (2002). □

Proof of Theorem 5.2. We prove the theorem by backward induction. We start with $t = T$, in which case we have $\hat{q}_{N,T}(x) = q_T(x) = 0$, which implies

$$\int |\hat{q}_{N,s}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \to 0 \quad \text{in probability} \tag{5.5}$$

and

$$\frac{1}{N}\sum_{i=1}^{N} \left|\hat{q}_{N,s}(X_{s,i}) - q_s(\bar{X}_{s,i}^{(n)})\right|^2 \to 0 \quad \text{in probability} \tag{5.6}$$

for $s = T$.

Let $t \in \{0, \dots, T-1\}$ be arbitrary and assume that (5.5) and (5.6) hold for $s = t+1$. In the sequel we show (5.5) and (5.6) for $s = t$. To do this we apply Lemma 5.3 with

$$X_i = X_{i,t}, \quad \bar{X}_{i,n} = \bar{X}_{i,t}^{(n)}, \quad Y_i = \max\{f_{t+1}(X_{i,t+1}), q_{t+1}(X_{i,t+1})\},$$

and

$$\bar{Y}_{i,n} = \max\left\{f_{t+1}\left(\bar{X}_{i,t+1}^{(n)}\right), \quad \hat{q}_{N,t+1}\left(\bar{X}_{i,t+1}^{(n)}\right)\right\}.$$

Here (5.1) implies that $Y_1, \dots, Y_N$ and $\bar{X}_{1,n}, \dots, \bar{X}_{N,n}$ are independent given $X_1, \dots, X_N$, so we can conclude

$$\mathbf{P}\left\{\int |\hat{q}_{N,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) > c_{11} \cdot Z_n\right\} \to 0 \quad (n \to \infty)$$

and

$$\mathbf{P}\left\{\frac{1}{N}\sum_{i=1}^{N} |\hat{q}_{N,t}(\bar{X}_{t,i}) - q_t(X_{t,i})|^2 > c_{11} \cdot Z_n\right\} \to 0 \quad (n \to \infty),$$

where

$$Z_n = \frac{1}{N(n)}\sum_{i=1}^{N(n)} \left|\max\{f_{t+1}(X_{i,t+1}), q_{t+1}(X_{i,t+1})\} - \max\left\{f_{t+1}\left(\bar{X}_{i,t+1}^{(n)}\right), \hat{q}_{N(n),t+1}\left(\bar{X}_{i,t+1}^{(n)}\right)\right\}\right|^2$$
$$+ \gamma_{N(n)}^2 \cdot \frac{1}{N(n)}\sum_{i=1}^{N(n)} \left\| X_{i,t} - \bar{X}_{i,t}^{(n)}\right\|^2 + \frac{D_{N(n)} \cdot \beta_{N(n)}^5}{N(n)} + \inf_{f \in \mathcal{F}_{N(n)}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx).$$

Lemma 5.1 implies that for n sufficiently large (i.e., in case $\gamma_N \geq L$), we have

$$Z_n \leq 2\frac{1}{N(n)}\sum_{i=1}^{N(n)} |\hat{q}_{N(n),t+1}\left(\bar{X}_{i,t+1}^{(n)}\right) - q_{t+1}(X_{i,t+1})|^2 + 3 \cdot \gamma_{N(n)}^2 \cdot \frac{1}{N(n)}\sum_{i=1}^{N(n)} \left\| X_{i,t+1} - \bar{X}_{i,t+1}^{(n)}\right\|^2$$
$$+ \frac{D_{N(n)} \cdot \beta_{N(n)}^5}{N(n)} + \inf_{f \in \mathcal{F}_{N(n)}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx).$$

By the induction hypothesis and the assumptions of the theorem, we have

$$Z_n \to 0 \quad \text{in probability.}$$

The proof of (5.3) is complete, and using (2.4) we also get (5.4). □

We reformulate Theorem 5.2 in case of choosing the function space as a spline space.

COROLLARY 5.4. *Let the discounted payoff function f_t be bounded in absolute value by some $\beta > 1$ and Lipschitz continuous with Lipschitz constant $L > 0$. For $n \in \mathbb{N}$ let A_n, β_n, $\gamma_n > 0$, $K_n \in \mathbb{N}$, $\mathbf{K}_n = (K_n, \dots, K_n)$, $\mathbf{M} \in \mathbb{N}_0^d$ and set*

$$\mathcal{F}_n = S_{\mathbf{K}_n, \mathbf{M}, \beta_n, \gamma_n}([-A_{N(n)}, A_{N(n)}]^d).$$

Assume that

$$(5.7) \quad X_{1,t+1}, \dots, X_{n,t+1} \text{ and } \bar{X}_{1,t}^{(n)}, \dots, \bar{X}_{n,t}^{(n)} \text{ are independent given } X_{1,t}, \dots, X_{N(n),t}$$

for all $t = 1, \dots, T-1$. Then

$$A_{N(n)} \to \infty, \quad \beta_{N(n)} \to \infty, \quad \gamma_{N(n)} \to \infty, \quad \frac{A_{N(n)}}{K_{N(n)}} \to 0, \quad \frac{\beta_{N(n)}^5 \cdot A_{N(n)}^d K_{N(n)}^d}{N(n)} \to 0 \quad (n \to \infty)$$

and

$$\frac{\gamma_{N(n)}^2}{N(n)} \sum_{i=1}^{N(n)} \left\| \bar{X}_{t,i}^{(n)} - X_{t,i} \right\|^2 \to 0 \quad \text{in probability}$$

for all $t = 0, \dots, T$ imply

$$\int |\hat{q}_{N(n),t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \to 0 \quad \text{in probability}$$

for $t = 0, \dots, T$ and, in addition

$$\hat{V}_{0,n} \to V_0 \quad \text{in probability.}$$

Proof. Corollary 5.4 follows directly from Theorem 5.2 if we observe that (5.2) follows from $q_t \in Ł_2(\mathbf{P}_{X_t}$ (which is implied by the boundaries of the payoff function) and approximation properties of spline spaces (cf., e.g., proof of corollary 2 in Fromkorth and Kohler 2011). □

5.3. Proof of Theorem 3.1

Set $X_{t,i} = (X_{1,t,i}, \dots, X_{d,t,i})'$, where

$$X_{k,t,i} = \log(z_{k,0}) + r \cdot t + \sum_{s=1}^{t} \sum_{j=1}^{d} \left(\sigma_{k,j} \cdot \epsilon_{s,j,i} - \frac{1}{2} \cdot \sigma_{k,j}^2 \right).$$

This can be interpreted as an artificial sample of the logarithm of asset values with the real (but unknown) distribution.

For all $t = 0, \dots, T$ it holds

$$
\begin{aligned}
&\| X_{t,1} - \bar{X}_{t,1}^{(n)} \|^2 \\
&= \sum_{k=1}^{d} | X_{k,t,1} - \bar{X}_{k,t,1}^{(n)} |^2 \\
&= \sum_{k=1}^{d} \left| \frac{t}{2} \cdot \sum_{j=1}^{d} \left(\left(\hat{\sigma}_{k,j}^{(n)}\right)^2 - \sigma_{k,j}^2 \right) + \sum_{s=1}^{t} \sum_{j=1}^{d} \left(\left(\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)}\right) \cdot \epsilon_{s,j,1} \right) \right|^2 \\
&\leq \sum_{k=1}^{d} d \cdot (t+1) \cdot \left(\frac{t^2}{4} \cdot \sum_{j=1}^{d} \left(\left(\hat{\sigma}_{k,j}^{(n)}\right)^2 - \sigma_{k,j}^2 \right)^2 + \sum_{s=1}^{t} \sum_{j=1}^{d} \left(\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)}\right)^2 \cdot \epsilon_{s,j,1}^2 \right),
\end{aligned}
$$

where we have used the inequality of Jensen. From this we get

$$
\begin{aligned}
\mathbf{E}\{\gamma_{N(n)}^2 \| X_{t,1} - \bar{X}_{t,1}^{(n)} \|\} &\leq \sum_{k=1}^{d} t \cdot (d+1) \left(\frac{t^2}{4} \cdot \sum_{j=1}^{d} \gamma_{N(n)}^2 \left(\left(\hat{\sigma}_{k,j}^{(n)}\right)^2 - \sigma_{k,j}^2 \right)^2 \right. \\
&\quad \left. + \sum_{s=1}^{t} \sum_{j=1}^{d} \gamma_n^2 \left(\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)}\right)^2 \cdot \mathbf{E}\left(\epsilon_{s,j,1}^2\right) \right) \\
&\leq \sum_{k=1}^{d} t \cdot (d+1) \left(\frac{t^2}{4} \cdot \sum_{j=1}^{d} \gamma_{N(n)}^2 \left(\hat{\sigma}_{k,j}^{(n)} - \sigma_{k,j}\right)^2 \cdot \left(\hat{\sigma}_{k,j}^{(n)} + \sigma_{k,j}\right)^2 \right. \\
&\quad \left. + \; t \cdot \sum_{j=1}^{d} \gamma_{N(n)}^2 \left(\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)}\right)^2 \right) \to 0 \quad (n \to \infty),
\end{aligned}
$$

where the last step follows from the assumptions of the theorem. For arbitrary $\epsilon > 0$, the Markov inequality implies now

$$
\begin{aligned}
\mathbf{P}\left\{ \gamma_{N(n)}^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \| X_{t,i} - \bar{X}_{t,i}^{(n)} \|^2 > \epsilon \right\} &\leq \frac{\mathbf{E}\left\{ \gamma_{N(n)}^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \| X_{t,i} - \bar{X}_{t,i}^{(n)} \|^2 \right\}}{\epsilon} \\
&= \frac{\mathbf{E}\{\gamma_{N(n)}^2 \| X_{t,1} - \bar{X}_{t,1}^{(n)} \|\}}{\epsilon} \to 0 \quad (n \to \infty),
\end{aligned}
$$

this means

$$
\gamma_{N(n)}^2 \cdot \frac{1}{N(n)} \sum_{i=1}^{N(n)} \| X_{t,i} - \bar{X}_{t,i}^{(n)} \|^2 \to 0 \quad (n \to \infty) \text{ in probability.}
$$

Finally, we note that $X_{1,t+1}, \dots, X_{N,t+1}$ depend only on random variables independent from all random variables used up to time t provided we fix $X_{t,1}, \dots, X_{t,N}$. So the independence assumption in Corollary 5.4 is trivially fulfilled.

Corollary 5.4 implies the assertion. □

5.4. Proof of Theorem 3.2

Again we introduce the artificial sample $X_{t,i}$ of the real distribution, i.e., we set

$$X_{t,i} = (Z_{t,i}, \epsilon_{t,i}, \dots, \epsilon_{t-q+1,i}, h_{t,i}, \dots, h_{t-\max\{p,q\}+1,i})^T,$$

where

$$Z_{t,i} = \log(x_0) + rt - \frac{1}{2}\sum_{j=1}^{t} h_{j,i} + \sum_{j=1}^{t} \sqrt{h_{j,i}} \cdot \epsilon_{j,i},$$

$$h_{t,i} = a_0 + \sum_{j=1}^{q} a_j \cdot h_{t-j,i}(\epsilon_{t-j,i} - \lambda)^2 + \sum_{j=1}^{p} b_j \cdot h_{t-j,i},$$

for $t > 0$ and $h_{s,i} = 0$ for $s \leq 0$.

As in the proof of Theorem 3.1 it suffices to show

$$\mathbf{E}\{\gamma_n^2 \| X_{t,1} - \bar{X}_{t,1}^{(n)} \|^2\} \to 0 \quad (n \to \infty).$$

By the definition of the stochastic processes and the inequality of Jensen we get

$$\begin{aligned}
&\| X_{t,1} - \bar{X}_{t,1}^{(n)} \|^2 \\
&= |Z_{t,1} - \bar{Z}_{t,1}^{(n)}| + \sum_{j=1}^{\max\{p,q\}} |h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}|^2 \\
&= \left| \sum_{s=1}^{t} \frac{1}{2}(\bar{h}_{s,1}^{(n)} - h_{s,1}) + \sum_{s=1}^{t} \left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right) \cdot \epsilon_{s,1} \right|^2 + \sum_{j=1}^{\max\{p,q\}} |h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}|^2 \\
&\leq 2t \cdot \left(\sum_{s=1}^{t} \frac{1}{4}(\bar{h}_{s,1}^{(n)} - h_{s,1})^2 + \sum_{s=1}^{t} \left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2 \cdot \epsilon_{s,1}^2 \right) \\
&\quad + \sum_{j=1}^{\max\{p,q\}} |h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}|^2.
\end{aligned}$$

Because of $a_0 > 0$ and $a_j, \hat{a}_{j,n}, b_i, \hat{b}_{i,n}$ nonnegative, we know $\bar{h}_{s,1}^{(n)} \geq 0$ and $h_{s,1} \geq a_0$ for all $s \in \{0, \dots, t\}$. Therefore,

$$\left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2 = \left(\frac{h_{s,1} - \bar{h}_{s,1}^{(n)}}{\sqrt{h_{s,1}} + \sqrt{\bar{h}_{s,1}^{(n)}}} \right)^2 \leq \frac{1}{a_0}\left(h_{s,1} - \bar{h}_{s,1}^{(n)}\right)^2.$$

Using this and the independence of $\epsilon_{s,1}$ from $h_{s,1}$ and $\bar{h}_{s,1}^{(n)}$ we get

$$\begin{aligned}
&\mathbf{E}\{\gamma_{N(n)}^2 \cdot \| X_{t,1} - \bar{X}_{t,1}^{(n)} \|^2\} \\
&\leq 2t \cdot \left(\sum_{s=1}^{t} \frac{1}{4} \mathbf{E}\{\gamma_{N(n)}^2 \cdot (\bar{h}_{s,1}^{(n)} - h_{s,1})^2\} + \sum_{s=1}^{t} \mathbf{E}\left\{\gamma_{N(n)}^2 \cdot \left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2\right\} \cdot \mathbf{E}\epsilon_{s,1}^2 \right) \\
&\quad + \sum_{j=1}^{\max\{p,q\}} \mathbf{E}\{\gamma_{N(n)}^2 \cdot |h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}|^2\} \\
&\leq \left(\frac{t}{2} + \frac{2t}{a_0} + 1\right) \cdot \sum_{s=1}^{t} \mathbf{E}\{\gamma_{N(n)}^2 \cdot (\bar{h}_{s,1}^{(n)} - h_{s,1})^2\}.
\end{aligned}$$

So it remains to proof

$$\gamma_{N(n)}^2 \cdot \mathbf{E}\{(h_{s,1} - \bar{h}_{s,1}^{(n)})^2\} \to 0 \quad (n \to \infty) \tag{5.8}$$

for $s \in \{0, \ldots, t\}$. We will do this by induction over s. By definition we have for $s = 0$ that

$$\gamma_{N(n)}^2 \cdot \mathbf{E}\{(h_{s,1} - \bar{h}_{s,1}^{(n)})^2\} = 0.$$

Let $s \in \mathbb{N}_0$ and assume (5.8) holds for s (and all smaller indices). Then we have for $s + 1$, that

$$\begin{aligned}
&\gamma_{N(n)}^2 \cdot \mathbf{E}\left\{|h_{s+1,1} - \bar{h}_{s+1,1}^{(n)}|^2\right\} \\
&= \gamma_{N(n)}^2 \cdot \mathbf{E}\left\{\left|a_0 - \hat{a}_{0,n} + \sum_{j=1}^{q} \left(a_j h_{s+1-j,1}(\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n}\bar{h}_{s+1-j,1}^{(n)}(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2\right)\right.\right. \\
&\quad \left.\left. + \sum_{j=1}^{p} \left(b_j h_{s+1-j,1} - \hat{b}_{j,n}\bar{h}_{s+1-j,1}^{(n)}\right)\right|^2\right\} \leq (1 + p + q)\big(\gamma_{N(n)}^2 |a_0 - \hat{a}_{0,n}|^2 \\
&\quad + \sum_{j=1}^{q} \gamma_{N(n)}^2 \mathbf{E}\left\{|a_j h_{s+1-j,1}(\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n}\bar{h}_{s+1-j,1}^{(n)}(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2|^2\right\} \\
&\quad + \sum_{j=1}^{p} \gamma_{N(n)}^2 \mathbf{E}\left\{|b_j h_{s+1-j,1} - \hat{b}_{j,n}\bar{h}_{s+1-j,1}^{(n)}|^2\right\}\Big),
\end{aligned}$$

where the last inequality follows from the inequality of Jensen.

By the assumptions of the theorem we know

$$\gamma_{N(n)}^2 \left|a_0 - \hat{a}_{0,n}\right|^2 \to 0 \quad (n \to 0).$$

Using $(a+b+c)^2 \le 3a^2 + 3b^2 + 3c^2$ ($a, b, c \in \mathbb{R}$) and the independence of $\epsilon_{s+1-j,1}$ from $h_{s+1-j,1}$ and $\bar{h}^{(n)}_{s+1-j,1}$ we get

$$
\begin{aligned}
&\mathbf{E}\Big\{\big|a_j h_{s+1-j,1}(\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n}\bar{h}^{(n)}_{s+1-j,1}(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2\big|^2\Big\} \\
&\le 3a_j^2 \cdot \mathbf{E}h^2_{s+1-j,1} \cdot \mathbf{E}\Big\{\big((\epsilon_{s+1-j,1} - \lambda)^2 - (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2\big)^2\Big\} \\
&\quad + 3a_j^2 \cdot \mathbf{E}\Big\{\big(h_{s+1-j,1} - \bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} \cdot \mathbf{E}\big\{(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^4\big\} \\
&\quad + 3 \cdot (a_j - \hat{a}_{j,n})^2 \cdot \mathbf{E}\Big\{\big(\bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} \cdot \mathbf{E}\big\{(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^4\big\} \\
&= 3a_j^2 \cdot \mathbf{E}h^2_{s+1-j,1} \cdot (\lambda - \hat{\lambda}_n)^2 \cdot ((4 + (\lambda + \hat{\lambda}_n)^2) \\
&\quad + 3a_j^2 \cdot \mathbf{E}\Big\{\big(h_{s+1-j,1} - \bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} \cdot (6\hat{\lambda}_n^2 + \hat{\lambda}_n^4 + 3) \\
&\quad + 3 \cdot (a_j - \hat{a}_{j,n})^2 \cdot \mathbf{E}\Big\{\big(\bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} \cdot (6\hat{\lambda}_n^2 + \hat{\lambda}_n^4 + 3),
\end{aligned}
$$

where the last equality follows from $\mathbf{E}\epsilon_{s+1-j,1} = 0$, $\mathbf{E}\epsilon^2_{s+1-j,1} = 1$, $\mathbf{E}\epsilon^3_{s+1-j,1} = 0$, and $\mathbf{E}\epsilon^4_{s+1-j,1} = 3$.

From this, the induction hypothesis, $\mathbf{E}\{h^2_{s,1}\} < \infty$ (which follows by induction) and

$$\mathbf{E}\Big\{\big(\bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} \le 2 \cdot \mathbf{E}\Big\{\big(h_{s+1-j,1} - \bar{h}^{(n)}_{s+1-j,1}\big)^2\Big\} + 2 \cdot \mathbf{E}\big\{h^2_{s+1-j,1}\big\}$$

for all s, j, and the assumptions of the theorem we see

$$\gamma^2_{N(n)}\mathbf{E}\Big\{\big|a_j h_{s+1-j,1}(\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n}\bar{h}^{(n)}_{s+1-j,1}(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2\big|^2\Big\} \to 0 \quad (n \to \infty).$$

Similar arguments lead to

$$\gamma^2_{N(n)}\mathbf{E}\Big\{\big|b_j h_{s+1-j,1} - \hat{b}_{j,n}\bar{h}^{(n)}_{s+1-j,1}\big|^2\Big\} \to 0 \quad (n \to \infty),$$

from which we conclude the assertion. □

5.5. Proof of Theorem 3.3

The assertion depends only on the joint distribution of the random variables describing the discrete the price process used in the regression-based Monte Carlo method sampled at discrete points, so to prove the theorem we may assume without loss of generality that the random variables are generated in some special way. We do this in the same way for random variables describing the logarithms of the returns of the price process using the true parameter values. In both cases we use that values of a Poisson process sampled at discrete points can be generated as partial sums to a sequence of independent Poisson-distributed random variables.

Let

$$\varepsilon_{t,i}, \xi_{t,i}, \pi_{t,i}, \hat{\pi}^{(n)}_{t,i} \qquad (i, t \in \mathbb{N}),$$

be independent random variables, where $\epsilon_{t,i}$ and $\xi_{t,i}$ are standard normally distributed, $N_{t,i}$ is Poisson distributed with parameter λt, and $\hat{\pi}^{(n)}_{t,i}$ is Poisson distributed with parameter $|\lambda - \hat{\lambda}_n|t$.

At first we consider the case $\lambda < \lambda_n$. Because of the folding property of the Poisson distribution we can write $\bar{\pi}_{t,i}^{(n)}$ as

$$\bar{\pi}_{t,i}^{(n)} = N_{t,i} + \hat{\pi}_{t,i}^{(n)}.$$

For the logarithm of the price processes $X_{t,i}$ and $\bar{X}_{t,i}^{(n)}$ this means

$$X_{t,i} = \log(x_0) + \mu t + \sum_{s=1}^{t} \sigma \varepsilon_{s,i} + \sum_{j=1}^{\pi_{t,i}} (m + \delta \xi_{t,j})$$

and

$$\bar{X}_{t,i} = \log(x_0) + \hat{\mu}_n t + \sum_{s=1}^{t} \hat{\sigma}_n \varepsilon_{s,i} + \sum_{j=1}^{\pi_{t,i}+\hat{\pi}_{t,i}^{(n)}} (\hat{m}_n + \hat{\delta}_n \xi_{t,j}).$$

As in the proofs of Theorems 3.1 and 3.2, it is enough to show

$$\mathbf{E}\{\gamma_{N(n)}^2 \left| X_{t,1} - \bar{X}_{t,1}^{(n)} \right|^2\} \to 0 \qquad (n \to \infty).$$

Jensen's inequality implies

$$\begin{aligned} \left| X_{t,1} - \bar{X}_{t,1}^{(n)} \right|^2 = & \Bigg| (\mu - \hat{\mu}_n) t + (\sigma - \hat{\sigma}_n) \sum_{s=1}^{t} \varepsilon_{s,1} + \pi_{t,1}(m - \hat{m}_n) + (\delta - \hat{\delta}_n) \sum_{j=1}^{\pi_{t,1}} \xi_{t,j} \\ & - \sum_{j=\pi_{t,1}+1}^{\pi_{t,1}+\hat{\pi}_{t,1}^{(n)}} (\hat{m}_n + \hat{\delta}_n \xi_{t,j}) \Bigg|^2 \le 5 \cdot \left((\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 \left(\sum_{s=1}^{t} \varepsilon_{s,1} \right)^2 \right. \\ & \left. + (\pi_{t,1})^2 (m - \hat{m}_n)^2 + (\delta - \hat{\delta}_n)^2 \left(\sum_{j=1}^{\pi_{t,1}} \xi_{t,j} \right)^2 + \left(\sum_{j=\pi_{t,1}+1}^{\pi_{t,1}+\hat{\pi}_{t,1}^{(n)}} (\hat{m}_n + \hat{\delta}_n \xi_{t,j}) \right)^2 \right) \end{aligned}$$

and therefore,

$$\begin{aligned} & \mathbf{E}\left\{ \gamma_{N(n)}^2 \left| X_{t,1} - \bar{X}_{t,1}^{(n)} \right|^2 \right\} \\ & \le 5 \cdot \gamma_{N(n)}^2 \left((\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 \mathbf{E}\left\{ \left(\sum_{s=1}^{t} \varepsilon_{s,1} \right)^2 \right\} + \mathbf{E}\left\{ (\pi_{t,1})^2 \right\} (m - \hat{m}_n)^2 \right. \\ & \left. + (\delta - \hat{\delta}_n)^2 \mathbf{E}\left\{ \left(\sum_{j=1}^{\pi_{t,1}} \xi_{t,j} \right)^2 \right\} + \mathbf{E}\left\{ \left(\hat{\pi}_{t,1}^{(n)} \hat{m}_n + \hat{\delta}_n \sum_{j=\pi_{t,1}+1}^{\pi_{t,1}+\hat{\pi}_{t,1}^{(n)}} \xi_{t,j} \right)^2 \right\} \right). \end{aligned}$$

The random variables $\varepsilon_{1,1}, \ldots, \varepsilon_{t,1}$ are independent and standard normal distributed, so it holds

$$\mathbf{E}\left\{ \left(\sum_{s=1}^{t} \varepsilon_{s,1} \right)^2 \right\} = t,$$

and by definition of $\pi_{t,1}$ we have

$$\mathbf{E}\{(\pi_{t,1})^2\} = (\lambda t)^2 + \lambda t.$$

Because of the independence of $\pi_{t,1}, \xi_{t,1}, \xi_{t,2}, \dots$ and the identical distribution of $\xi_{t,1}, \xi_{t,2}, \dots$, this implies

$$\mathbf{E}\left\{\left(\sum_{j=1}^{\pi_{t,1}} \xi_{t,j}\right)^2\right\} = \mathbf{E}\{\pi_{t,1}\}\mathbf{Var}\{\xi_{t,1}\} + \mathbf{E}\{(\pi_{t,1})^2\}\,(\mathbf{E}\xi_{t,1})^2 = \lambda t$$

and so

$$(\delta - \hat{\delta}_n)^2 \mathbf{E}\left\{\left(\sum_{j=1}^{\pi_{t,1}} \xi_{t,j}\right)^2\right\} = (\delta - \hat{\delta}_n)^2 \lambda t.$$

From the independence of $\pi_{t,1}, \hat{\pi}^{(n)}_{t,1}, \xi_{t,1}, \xi_{t,2}, \dots$ and because of $\mathbf{E}\left(\xi_{t,j}\right) = 0$ for all i, j one gets

$$\begin{aligned}\mathbf{E}\left\{\left(\hat{\pi}^{(n)}_{t,1}\hat{m}_n + \hat{\delta}_n \sum_{j=\pi_{t,1}+1}^{\pi_{t,1}+\hat{\pi}^{(n)}_{t,1}} \xi_{t,j}\right)^2\right\} &= \mathbf{E}\left\{\left(\hat{\pi}^{(n)}_{t,1}\hat{m}_n\right)^2\right\} + \mathbf{E}\left\{\left(\hat{\delta}_n \sum_{j=\pi_{t,1}+1}^{\pi_{t,1}+\hat{\pi}^{(n)}_{t,1}} \xi_{t,j}\right)^2\right\} \\ &= \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot \hat{m}_n^2 + |\lambda - \hat{\lambda}_n| \cdot \hat{\delta}_n^2 \cdot t.\end{aligned}$$

To conclude this means in case of $\lambda < \lambda_n$, that

$$\begin{aligned}\mathbf{E}\left\{\gamma^2_{N(n)}\left|X_{t,1} - \bar{X}^{(n)}_{t,1}\right|^2\right\} \le 5 \cdot \gamma^2_{N(n)}\Big(&(\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 t + (\lambda^2 t^2 + \lambda t) \cdot (m - \hat{m}_n)^2 \\ &+ (\delta - \hat{\delta}_n)^2 \lambda t + (|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2) \cdot \hat{m}_n^2 + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2\Big).\end{aligned}$$

Similar argumentation implies for $\lambda \ge \lambda_n$, that

$$\pi_{t,i} = \bar{\pi}^{(n)}_{t,i} + \hat{\pi}^{(n)}_{t,i}.$$

So we get in this case

$$\begin{aligned}\mathbf{E}\left\{\gamma^2_{N(n)}\left|X_{t,1} - \bar{X}^{(n)}_{t,1}\right|^2\right\} \le 5 \cdot \gamma^2_{N(n)}\big(&(\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 t + \left(\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t\right) \cdot (m - \hat{m}_n)^2 \\ &+ (\delta - \hat{\delta}_n)^2 \hat{\lambda}_n t + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2 + \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot m^2\big).\end{aligned}$$

Altogether this means

$$
\begin{aligned}
&\mathbf{E}\Big\{\gamma_{N(n)}^2 \| X_{t,1} - \bar{X}_{t,1}^{(n)} \|^2\Big\} \\
&\leq 5\cdot\gamma_{N(n)}^2\bigg((\mu-\hat{\mu}_n)^2\, t^2 + (\sigma-\hat{\sigma}_n)^2\, t + \max\left\{\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t, \lambda^2 t^2 + \lambda t\right\}\cdot (m-\hat{m}_n)^2 \\
&\quad + (\delta-\hat{\delta}_n)^2 \max\{\lambda t, \hat{\lambda}_n t\} + |\lambda-\hat{\lambda}_n| t\cdot \hat{\delta}_n^2 + (|\lambda-\hat{\lambda}_n| t + |\lambda-\hat{\lambda}_n|^2 t^2)\cdot \max\left\{\hat{m}_n^2, m^2\right\}\bigg) \\
&\leq 5\cdot\gamma_{N(n)}^2\Big((\mu-\hat{\mu}_n)^2\, t^2 + (\sigma-\hat{\sigma}_n)^2\, t + \left(\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t + \lambda^2 t^2 + \lambda t\right)\cdot (m-\hat{m}_n)^2 \\
&\quad + (\delta-\hat{\delta}_n)^2 (\lambda t + \hat{\lambda}_n t) + |\lambda-\hat{\lambda}_n| t\cdot \hat{\delta}_n^2 + (|\lambda-\hat{\lambda}_n| t + |\lambda-\hat{\lambda}_n|^2 t^2)\cdot \left(\hat{m}_n^2 + m^2\right)\Big).
\end{aligned}
$$

By the assumptions of the theorem this expression converges to zero for $n \to \infty$. Here we have used that by the mean-value-theorem we have

$$
\left|\exp\left(\hat{m}_n + \frac{\hat{\delta}_n^2}{2}\right) - \exp\left(m + \frac{\delta^2}{2}\right)\right| \leq \left|\hat{m}_n + \frac{\hat{\delta}_n}{2} - m - \frac{\delta^2}{2}\right| \exp\left(|\hat{m}_n| + \frac{\hat{\delta}_n^2}{2} + m + \frac{\delta^2}{2}\right).
$$

The result follows as in Theorem 2.

REFERENCES

BELOMESTNY, D. (2011a): Pricing Bermudan Options Using Regression: Optimal Rates of Convergence for Lower Estimates, *Finance Stoch.* 15, 655–683.

BELOMESTNY, D. (2011b): On the Rates of Convergence of Simulation-Based Optimization Algorithms for Optimal Stopping Problems, *Ann. Appl. Probab.* 21, 215–239.

BELOMESTNY, D., and M. REIß (2010): Spectral Calibration of Exponential Levy, *Finance Stoch.* 10, 449–474.

BOLLERSLEV, T. (1986): Generalized Autoregressive Conditional Heteroscedasticity, *J Econ.* 31, 307–327.

CARRIÉR, J. (1996): Valuation of Early-Exercise Price of Options Using Simulations and Nonparametric Regression, *Insurance: Math. Econ.* 19 19–30.

CHOW, Y. S., H. ROBBINS, and D. SIEGMUND (1971): *Great Expectations: The Theory of Optimal Stopping*, Boston: Houghton Mifflin Company.

DE BOOR, C. (1987): *A Practical Guide to Splines*, New York: Springer.

DUAN, J.-C. (1995): The GARCH Option Pricing Model, *Math. Finance* 5, 13–32.

EGLOFF, D. (2005): Monte Carlo Algorithms for Optimal Stopping and Statistical Learning, *Ann. Appl. Probab.* 15, 1–37.

EGLOFF, D., M. KOHLER, and N. TODOROVIC (2007): A Dynamic Look-Ahead Monte Carlo Algorithm for Pricing American Options, *Ann. Appl. Probab.* 17(4), 1138–1171.

FROMKORTH, A., and M. KOHLER (2011): Analysis of Least Squares Regression Estimates in Case of Additional Errors in the Variables, *J. Stat. Plan. Infer.* 141, 172–188.

GILL, P., and M. MURRAY (1978): Mathematically Stable Methods for Quadratic Programming, *Math. Program.* 14, 349–372.

GLASSERMAN, P. (2004): *Monte Carlo Methods in Financial Engineering*, Vol. 53, Applications of Mathematics, New York: Springer.

GOLDFARB, D., and A. IDNANI (1983): A Numerically Stable Dual Method for Solving Strictly Convex Quadratic Programs, *Math. Program.* 27, 1–33.

GYÖRFI, L., M. KOHLER, A. KRZYŻAK, and H. WALK (2002): *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics, New York: Springer.

KARATZAS, I., and E.S. SHREVE (1998): *Methods of Mathematical Finance*, Vol. 21, Application of Mathematics–Stochastic Modelling and Applied Probability, New York: Springer.

KOHLER, M. (2008): A Regression-Based Smoothing Spline Monte Carlo Algorithm for Pricing American Options, *Adv. Statist. Anal.* 92, 153–178.

KOHLER, M. (2010): A Review on Regression-Based Monte Carlo Methods for Pricing American Options, in *Recent Developments in Applied Probability and Statistics. Dedicated to the Memory of Jürgen Lehn*, L. Devroye, B. Karasözen, M. L. Kohler, and R. Korn, eds. Heidelberg: Physica-Verlag, pp. 37–58.

KOHLER, M., and A. KRZYŻAK (2009): Pricing of American Options in Discrete Time Using Least Squares Estimates with Complexity Penalties. Preprint. To appear in *Journal of Statistical Planning and Inference*.

KOHLER, M., A. KRZYŻAK, and N. TODOROVICH (2010): Pricing of High-Dimensional American Options by Neural Networks, *Math. Finance* 20, 383–410.

LONGSTAFF, F. A., and E. S. SCHWARZ (2001): Valuing American Options by Simulation: A Simple Least-Squares Approach, *Rev. Financ. Stud.* 14, 113–147.

MERTON, R. (1976): Option Pricing When Underlying Stock Returns Are Discontinuous, *J. Financ. Econ.* 3, 125–144.

SHIRYAEV, A. N. (1978): *Optimal Stopping Rules*, *Applications of Mathematics*, Berlin: Springer.

TSITSIKLIS, J. N., and B. VAN ROY (1999): Optimal Stopping of Markov Processes: Hilbert Space Theory, Approximation Algorithms, and an Application to Pricing High-Dimensional Financial Derivatives, *IEEE Trans. Autom. Control* 44, 1840–1851.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 400–425

FROM SMILE ASYMPTOTICS TO MARKET RISK MEASURES

RONNIE SIRCAR

Princeton University

STEPHAN STURM

Worcester Polytechnic Institute

The left tail of the implied volatility skew, coming from quotes on out-of-the-money put options, can be thought to reflect the market's assessment of the risk of a huge drop in stock prices. We analyze how this market information can be integrated into the theoretical framework of convex monetary measures of risk. In particular, we make use of indifference pricing by dynamic convex risk measures, which are given as solutions of backward stochastic differential equations, to establish a link between these two approaches to risk measurement. We derive a characterization of the implied volatility in terms of the solution of a nonlinear partial differential equation and provide a small time-to-maturity expansion and numerical solutions. This procedure allows to choose convex risk measures in a conveniently parameterized class, distorted entropic dynamic risk measures, which we introduce here, such that the asymptotic volatility skew under indifference pricing can be matched with the market skew. We demonstrate this in a calibration exercise to market implied volatility data.

KEY WORDS: dynamic convex risk measures, volatility skew, stochastic volatility models, indifference pricing, backward stochastic differential equations.

1. INTRODUCTION

Risk measurement essentially conveys information about tails of distributions. However, that information is also contained in market prices of insurance securities that are contingent on a large (highly unlikely) downside, if we concede that those prices are mostly reflective of protection buyers' risk aversion. Examples are out-of-the-money put options that provide protection on large stock price drops, or senior tranches of collateralized debt obligations that protect against the default risk of say 15–30% of investment grade US companies over a 5-year period.

A central regulatory and internal requirement in recent years, in the wake of a number of financial disasters and corporate scandals, has been that firms report a measure of the risk of their financial positions. The industry-standard risk measure, value-at-risk,

R. Sircar's work partially supported by NSF grant DMS-0807440, and S. Sturm's work partially supported by NSF grant DMS-0739195.

S. Sturm completed this work while Postdoctoral Research Associate at Princeton University.

Manuscript received July 2011; final revision received July 2012.

Address correspondence to Ronnie Sircar, ORFE Department, Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA; e-mail: sircar@princeton.edu.

DOI: 10.1111/mafi.12015

is widely criticized for not being convex and thereby possibly penalizing diversification, and a number of natural problems arise:

1. How to construct risk measures with good properties.
2. Computation of these risk measures under typical financial models.
3. Choice: what is the "right" risk measure?

The first issue has been extensively studied in the static case (Artzner et al. 1999; Föllmer and Schied 2002) and recent developments in extending to dynamic risk measures with good time-consistency and/or recursive properties are discussed, e.g., in Barrieu and El Karoui (2009), Klöppel and Schweizer (2007), Musiela and Zariphopoulou (2009), and Föllmer and Schied (2011). However, concrete examples of dynamic, time-consistent convex risk measures are scarce, and they typically have to be defined abstractly, for example via the driver of a backward stochastic differential equation (BSDE) or as the limit of discrete time-consistent risk measures as in Stadje (2010). As a result, intuition is lost, and there is at present little understanding what the choice of driver says about the measure of risk. Or, to put it another way, how can the driver be constructed to be consistent with risk aversion reflected in the market?

Let ξ be a bounded random variable representing a financial payoff whose value is known at some future time $T < \infty$. A classical example of a convex risk measure, the entropic risk measure, is related to exponential utility:

$$\varrho(\xi) = \frac{1}{\gamma} \log(\mathbb{E}[e^{-\gamma\xi}]), \tag{1.1}$$

where $\gamma > 0$ is a risk-aversion coefficient. When extending to *dynamic* risk measures $\varrho_t(\cdot)$ adapted to some filtration $(\mathcal{F}_t)$, a desirable property is (strong) time-consistency

$$\varrho_s(-\varrho_t(\xi)) = \varrho_s(\xi), \quad 0 \le s \le t \le T.$$

This flow property is important if ϱ_t is used as a basis for a pricing system. The static entropic risk measure (1.1) generalizes to

$$\varrho_t(\xi) = \frac{1}{\gamma} \log(\mathbb{E}[e^{-\gamma\xi} \mid \mathcal{F}_t]). \tag{1.2}$$

The flow property follows simply from the tower property of conditional expectations. However, finding other directly defined examples is not easy, and to have a reasonable class of choices, we need to resort to more abstract constructions.

In a Brownian-based model, time-consistent dynamic risk measures can be built through BSDEs, as shown in Barrieu and El Karoui (2009), Klöppel and Schweizer (2007) (compare also Gianin 2006), extending the work of Peng (2004). That is, on a probability space with a d-dimensional Brownian motion W that generates a filtration $(\mathcal{F}_t)$, the risk measure of the $\mathcal{F}_T$-measurable random variable ξ (taking values in $\mathbb{R}$ for simplicity) is computed from the solution (R_t, Z_t), which takes values in $\mathbb{R} \times \mathbb{R}^d$, of the BSDE

$$\begin{aligned} -dR_t &= g(t, Z_t)\,dt - Z_t^*\,dW_t \\ R_T &= -\xi, \end{aligned}$$

where $*$ denotes transpose. Here the driver g, which defines the risk measure, is Lipschitz and convex in z and satisfies $g(t, 0) = 0$. The solution is a process R, taking values in $\mathbb{R}$ that matches the *terminal* condition $-\xi$ on date T (when ξ is revealed and the risk is known), and a process Z taking values in $\mathbb{R}^d$ that, roughly speaking, keeps the solution nonanticipating. Then $\varrho_t(\xi) := R_t$ defines a time-consistent dynamic convex risk measure. However, the possibility to offset risk by dynamically hedging in the market needs to be accounted for. Setting aside technicalities for the moment, this operation leads to a modification of the driver.

The left tail of the implied volatility skew observed in equity markets is a reflection of the premium charged for out-of-the-money put options. The bulk of the skew reveals the heavy left tail in the risk-neutral density of the stock price S_T at expiration, but the very far left tail, where investor sentiment and crash-o-phobia takes over, could be interpreted as revealing information about the representative market risk measure and its driver g, if we assume prices are consistent with this kind of pricing mechanism. The question then is to extract constraints on the driver from the observed tails of the skew, an inverse problem. In the application to equity options in Section 3, we assume the mid-market option prices reflect the premium a risk-averse buyer is willing to pay. We do not relate buyers' and sellers' prices to the bid-ask spread, since that is more likely related to the market maker's profit.

To put our analysis into a broader framework, we observe that the underlying structural question is the inference of preference structures from observable data. The idea of using (at least in theory) observable consumption and investment streams to reveal the preference structure of a rational utility maximizing investor dates back to Samuelson in the 1940s and Black in the 1960s—for a recent overview on this "backward approach" to utility theory we refer to Cox, Hobson, and Obłój (2011). The spirit of our presentation is a similar one, except we deal with dynamic risk measures rather than utility functions, and the observable data are not given as consumption and investment strategies but as readily available market implied volatilities.

The main goal of the current paper is to develop short-time asymptotics that can be used for the inverse problem of extracting information about the driver g from the observed skew. This could be used to construct an approximation to the driver and then to value more exotic derivatives in a way consistent with the risk measure. The information could also be used to quantify market perception of tail events, particularly when they depart from usual. Some studies have discussed the steepening of skew slopes in the run up to financial crises without a corresponding overall raise in volatility level. Inferring, fully or partially, a risk measure driver, could be used for detection of increased wariness of a crash.

Berestycki, Busca, and Florent (2004) presented short-time asymptotics for implied volatilities for *no arbitrage pricing* under a given risk-neutral measure in stochastic volatility models. Further work in this direction includes, among others, Feng, Forde, and Fouque (2010), Forde and Jacquier (2009), Forde, Jacquier, and Lee (2012), and references therein. In Section 2, we extend this analysis to the nonlinear partial differential equations (PDEs) characterizing indifference pricing under dynamic convex risk measures.

We find (Theorem 2.12) that the zero-order term in the short-time approximation is the same as found in no arbitrage pricing by Berestycki et al. (2004). The next order term is the solution of an inhomogeneous linear transport equation that sees only a particular slope of the partially Legendre-transformed driver, but is independent of the size of the options position (see equation (2.25)).

Section 3 illustrates the theoretical findings by focusing on a particular class of drivers, introducing distorted entropic convex dynamic risk measures. First we develop explicit calculations for the small time expansion in the Hull–White stochastic volatility model to illustrate the impact of the distortion parameter on the implied volatility skew. In Section 3.3, as a proof of concept, we perform a preliminary calibration exercise of the short-time asymptotics to S&P 500 implied volatilities close to maturity. This allows to estimate the stochastic volatility model parameters from the liquid central part of the skew, and to recover the distortion parameter of our family of dynamic convex risk measures from far out-of-the-money put options.

We illustrate the parameter impact for longer dated options in a numerical study (via the pricing PDE) of arctangent stochastic volatility driven by an Ornstein–Uhlenbeck process. Section 4 contains the conclusions and Section 5 gives the more technical proofs omitted in the exposition.

2. HEURISTICS AND STATEMENT OF RESULTS

We consider a model of a financial market consisting of a risk-free bond bearing no interest and some stock following the stochastic volatility model on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$

$$\begin{cases} dS_t = \mu(Y_t)S_t\,dt + \sigma(Y_t)S_t\,dW_t^1, & S_0 = S; \\ dY_t = m(Y_t)\,dt + a(Y_t)\big(\rho\,dW_t^1 + \rho'\,dW_t^2\big), & Y_0 = y, \end{cases} \tag{2.1}$$

where W^1, W^2 are two independent Brownian motions generating $(\mathcal{F}_t)$ and $\rho' = \sqrt{1-\rho^2}$.

ASSUMPTION 2.1. *We assume that:*

(i) $\sigma, a \in C_{loc}^{1+\beta}(\mathbb{R})$, *where* $C_{loc}^{1+\beta}(\mathbb{R})$ *is the space of differentiable functions with locally Hölder-continuous derivatives with Hölder-exponent* $\beta > 0$;

(ii) both σ *and* a *are bounded and bounded away from zero:*

$$0 < \underline{\sigma} < \sigma < \overline{\sigma} < \infty, \quad \text{and} \quad 0 < \underline{a} < a < \overline{a} < \infty;$$

(iii) $\mu, m \in C_{loc}^{0+\beta}(\mathbb{R})$, *and* $|\mu| < \overline{\mu} < \infty$.

The pricing will be done via the indifference pricing mechanism for dynamic convex risk measures, which are introduced in the next subsection.

2.1. Dynamic Convex Risk Measures, Indifference Pricing, and BSDEs

DEFINITION 2.2. We call the family $\varrho_t : L^\infty(\Omega, \mathcal{F}_T, P) \to L^\infty(\Omega, \mathcal{F}_t, P)$, $0 \le t \le T$, a convex dynamic risk measure, if it satisfies for all $t \in [0, T]$ and all $\xi, \xi^1, \xi^2 \in L^\infty(\Omega, \mathcal{F}_T, P)$ the following properties.

(i) Monotonicity: $\xi^1 \ge \xi^2$ P-a.s implies $\varrho_t(\xi^1) \le \varrho_t(\xi^2)$;
(ii) Cash invariance: $\varrho_t(\xi + m_t) = \varrho_t(\xi) - m_t$ for all $m_t \in L^\infty(\Omega, \mathcal{F}_t, P)$;
(iii) Convexity: $\varrho_t(\alpha\xi^1 + (1-\alpha)\xi^2) \le \alpha\varrho_t(\xi^1) + (1-\alpha)\varrho_t(\xi^2)$ for all $\alpha \in [0, 1]$;
(iv) Time-consistency $\varrho_t(\xi^1) = \varrho_t(\xi^2)$ implies $\varrho_s(\xi^1) = \varrho_s(\xi^2)$ for all $0 \le s \le t$.

We note that if the risk measure is additionally normalized, i.e., $\varrho_t(0) = 0$ for all $t \in [0, T]$, then (iv) is equivalent to the stronger property $\varrho_s(-\varrho_t(\xi)) = \varrho_s(\xi)$ for all $0 \le s \le t$ (Klöppel and Schweizer 2007, lemma 3.5). The risk measure $\varrho_t(\xi)$ should be understood as the risk associated with the position ξ at time t.

If ϱ_t is normalized, this is nothing else than the minimal capital requirement at time t to make the position riskless since $\varrho_t(\xi + \varrho_t(\xi)) = 0$. In this static setting, the certainty equivalent price of a buyer of a derivative $\xi \in L^\infty(\Omega, \mathcal{F}_T, P)$ at time t is just the cash amount for which buying the derivative has equal risk to not buying it.

In fact we are much more interested in the case where the buyer of the security is allowed to trade in the stock market to hedge her risk. In describing admissible strategies we follow the setting of continuous time arbitrage theory in the spirit of Delbaen–Schachermayer (for an overview, we refer to the monograph Delbaen and Schachermayer 2006). Denote therefore by Θ_t the set of all admissible hedging strategies from time t onwards, i.e., all progressive processes such that $\theta_t = 0$ and $\int_t^u \theta_s(\mu(Y_s)\,ds + \sigma(Y_s)\,dW_s^1)$ exists for all $u \in]t, T]$ and is uniformly bounded from below, and set

$$\mathcal{K}_t := \left\{ \int_t^T \theta_s\big(\mu(Y_s)\,ds + \sigma(Y_s)\,dW_s^1\big) \,:\, \theta \in \Theta_t \right\}.$$

The set of all superhedgeable payoffs is then given by $\mathcal{C}_t := (\mathcal{K}_t - L^0_+) \cap L^\infty$, where L^0_+ denotes the set of all almost surely nonnegative random variables.

The residual risk at time t of the derivative $\xi \in L^\infty(\Omega, \mathcal{F}_T, P)$ after hedging is given by

$$\hat{\varrho}_t(\xi) := \operatorname*{ess\,inf}_{h \in \mathcal{C}_t} \varrho_t(\xi + h). \tag{2.2}$$

Thus, assuming that the buyer's wealth at time t is x, her dynamic indifference price P_t, which can be viewed as the certainty equivalent after optimal hedging in the underlying market, is given via $\hat{\varrho}_t(x + \xi - P_t) = \hat{\varrho}_t(x)$, whence, using cash invariance,

$$P_t = \hat{\varrho}_t(0) - \hat{\varrho}_t(\xi). \tag{2.3}$$

We note, while restricting ourselves to the buyer's indifference price, all our considerations are easily adaptable to the seller's indifference price by a simple change of signs of ξ and P_t in (2.3).

A convenient class of dynamic convex risk measures to which we will stick throughout this paper is defined from solutions of BSDEs. Assume that $g : [0, T] \times \Omega \times \mathbb{R}^2 \to \mathbb{R}$ is a $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^2)$ predictable function which is continuous, convex, and quadratic (i.e., bounded in modulus by a quadratic function) in the $\mathbb{R}^2$-component. Next, let $\xi \in L^\infty(\Omega, \mathcal{F}_T, P)$ be a given bounded financial position. Then the BSDE

$$R_t = -\xi + \int_t^T g\big(t, \omega, Z_s^1, Z_s^2\big)\,ds - \int_t^T Z_s^1\,dW_s^1 - \int_t^T Z_s^2\,dW_s^2 \tag{2.4}$$

admits a unique $\mathcal{F}_t$-adapted solution (R_t, Z_t^1, Z_t^2), which defines a dynamic convex risk measure via $\varrho_t(X) := R_t$ (Barrieu and El Karoui 2009, theorem 3.21).

The existence of a solution of the BSDE (2.4) in this quadratic setting was first proved by Kobylanski (2000, theorem 2.3), with some corrections to their arguments given by Nutz (2007, theorem 3.6), whereas the uniqueness follows from the convexity of the driver as shown in Briand and Hu (2008, corollary 6). From a financial perspective, the components Z^1, Z^2 of the "auxiliary" process Z can be interpreted as risk sources, describing the risk stemming from the traded asset and the volatility process, respectively.

2.2. Transformed BSDE under Hedging

To assure the solvability of the BSDEs and PDEs that arise in our setting, we have to restrict the class of admissible drivers. Throughout, subscripts of functions indicate in the PDE context partial derivatives with the respect to the respective components.

DEFINITION 2.3. We call a function $g : \mathbb{R}^2 \to \mathbb{R}$ a strictly quadratic driver (normalized strictly quadratic driver) if it satisfies the following conditions (i)–(iii) (resp. (o)–(iii)):

(o) $g(0, 0) = 0$;
(i) $g \in C^{2,1}(\mathbb{R}^2)$;
(ii) $g_{z_1 z_1}(z_1, z_2) > 0$ for all $(z_1, z_2) \in \mathbb{R}^2$;
(iii) there exist constants $c_1, c_2 > 0$ such that

$$c_1\left(\frac{z_1^2}{4c_1^2} - (1 + z_2^2)\right) \le g(z_1, z_2) \le c_2(1 + z_1^2 + z_2^2) \text{ for all } (z_1, z_2) \in \mathbb{R}^2.$$

The normalization of the driver (condition (o)) corresponds to the normalization of the risk measure.

REMARK 2.4. To ease the presentation, we only work with drivers that do not depend explicitly on time or on ω. While any dependence on R_t would destroy the cash invariance, it is not difficult to add an additional dependence of g on time (requiring some local Hölder-continuity in the time component of g). The small time expansion Section 2.5 up to first order works in this case exactly as in the time-independent case (evaluating the driver at $t = T$), and the higher-order expansions can be adapted straightforwardly. For analysis of filtration-consistent, translation invariant nonlinear expectations, we refer to Coquet et al. (2002).

In passing from the principal risk measure defined by g to the residual risk measure after hedging, as in (2.2), we will need the Fenchel–Legendre transform of g in its first component, namely

$$\hat{g}(\zeta, z_2) := \sup_{z_1 \in \mathbb{R}} (\zeta z_1 - g(z_1, z_2)), \quad \zeta \in \mathbb{R}. \tag{2.5}$$

LEMMA 2.5. *Given that g is a (normalized) strictly quadratic driver, then the risk-adjusted driver $\hat{g}$ defined in (2.5) is also a (normalized) strictly quadratic driver.*

Proof. To show $\hat{g}$ satisfies condition (iii) of Definition 2.3, we fix z_2 and treat the function as classical Fenchel–Legendre transform in one variable. Therefore it holds for proper, continuous convex functions f, g, that $f \le g$ implies $\hat{f} \ge \hat{g}$ and $\hat{\hat{f}} = f$, (Hiriart-Urruty and Lemaréchal 2001b, proposition E.1.3.1 and corollary E.1.3.6). So the statement is proved by noting that

$$\sup_{z_1} \left(\xi z_1 - c\left(1 + z_1^2 + z_2^2\right)\right) = c\left(\frac{z_1^2}{4c^2} - \left(1 + z_2^2\right)\right)$$

for any positive constant c.

To show (*i*) and (*ii*) in Definition 2.3, we note that condition (*iii*) implies that g is 1-coercive in z_1, i.e., $g(z_1, z_2)/|z_1| \to \infty$ as $z_1 \to \pm\infty$ for fixed z_2. Now we can use the fact that the Fenchel–Legendre transform of any 1-coercive, twice differentiable function with positive second derivative is itself 1-coercive and twice differentiable with positive second derivative, cf. Hiriart-Urruty and Lemaréchal (2001a, corollary X.4.2.10). Thus it remains only to prove the differentiability of $\hat{g}$ with respect to z_2 which is a consequence of the differentiability properties of g: writing down the difference quotient and noting that the maximizer is differentiable, the positive second derivative with respect to the first component yields the existence of a finite limit. Finally $\hat{g}(0,0)=0$ follows from the definition if $g(0,0)=0$. □

In other words, the class of strictly quadratic drivers is invariant under the convex conjugation in the first component and the class of normalized strictly quadratic drivers is an invariant subclass thereof.

REMARK 2.6. We note that it is important in our setting to stick to the theory of quadratic drivers, since if g would be a Lipschitz driver, $\hat{g}$ would be no more a proper function. This fact is easily seen, since from the Lipschitz condition it follows that

$$g(z_1,z_2) \le L\left(1+\sqrt{z_1^2+z_2^2}\right) \le \sqrt{2}L(1+|z_1|+|z_2|)$$

for some constant L and hence

$$\begin{aligned}\hat{g}(\zeta,z_2) = \sup_{z_1}(\zeta z_1 - g(z_1,z_2)) \quad &\ge \sup_{z_1}(\zeta z_1 - \sqrt{2}L(1+|z_1|+|z_2|))\\ &= \begin{cases}\infty & \text{if } |\zeta| > \sqrt{2}L;\\ -\sqrt{2}L(1+|z_2|) & \text{if } |\zeta| \le \sqrt{2}L.\end{cases}\end{aligned}$$

From now on we will assume that g is convex as a function on $\mathbb{R}^2$ and an admissible driver. Our next step is to describe the dynamic hedging risk in terms of BSDEs. These results are in essence due to Toussaint (2007, section 4.4.1). Since his thesis is not easily available, we will nevertheless state the proofs here. It is convenient to introduce a notation for the Sharpe ratio:

$$\lambda(y) := \frac{\mu(y)}{\sigma(y)}. \tag{2.6}$$

PROPOSITION 2.7. *The risk of the financial position* $\xi \in L^\infty(\Omega,\mathcal{F}_T,P)$ *under hedging is* $\hat{\varrho}_t(\xi) = \hat{R}_t^{(\xi)}$ *where* $\hat{R}_t^{(\xi)}$ *is given via the unique solution of the BSDE*

$$\hat{R}_t^{(\xi)} = -\xi - \int_t^T \hat{Z}_s^1\lambda(Y_s) + \hat{g}\big(-\lambda(Y_s),\hat{Z}_s^2\big)\,ds - \int_t^T \hat{Z}_s^1\,dW_s^1 - \int_t^T \hat{Z}_s^2\,dW_s^2. \tag{2.7}$$

Moreover, $\hat{\varrho}_t$ is itself a dynamic convex risk measure.

Proof. It follows from the work of Klöppel and Schweizer (2007, theorem 7.17) that the risk is given via the BSDE

$$\hat{R}_t^{(\xi)} = -\xi + \int_t^T \tilde{g}\big(\hat{Z}_s^1,\hat{Z}_s^2\big)\,ds - \int_t^T \hat{Z}_s^1\,dW_s^1 - \int_t^T \hat{Z}_s^2\,dW_s^2,$$

where $\tilde{g}$ is given by the infimal convolution

$$\tilde{g}(z_1,z_2) := \inf_{v\in\mathbb{R}}(g(z_1+\sigma(Y_t)v,z_2)+\mu(Y_t)v). \tag{2.8}$$

To be precise, besides the differences in sign between our convex risk measures and their monetary concave utility functionals, our L^∞ framework is in line with the main part of their paper where they work in L^∞. However, the result (Klöppel and Schweizer 2007, theorem 7.17) is stated in the framework of L^2-BSDEs with Lipschitz drivers. Their detour to L^2 was due to their consideration that this is the natural framework for BSDEs. We have motivated that we have to work with quadratic drivers, for which there is yet no L^2-theory, but it is straightforward (though tedious) to check that their result (2.8) adapts to our setting due to the regularity enforced by the admissibility conditions in Definition 2.3.

Next, we rewrite the infimum in (2.8) to get

$$\tilde{g}(z_1, z_2) = \inf_{u\in\mathbb{R}} (g(u, z_2) - (z_1 - u)\lambda(Y_t)) = -z_1\lambda(Y_t) - \hat{g}(-\lambda(Y_t), z_2).$$

Finally, the uniqueness of the solution of the BSDE (2.7) follows again from Briand and Hu (2008, corollary 6) using the convexity of the driver, which is implied by the fact that $\hat{g}$ is concave in the second component. Moreover, this entails also that $\hat{\varrho}$ is a dynamic convex risk measure. □

REMARK 2.8.

1. We remark that in view of equation (2.7) of the above proposition, the notion of admissibility could be slightly extended: it is possible to replace the lower bound in Definition 2.3 (iii) by
$$c_1\big(f(z_1) - \big(1 + z_2^2\big)\big) \leq g(z_1, z_2)$$
for an arbitrary real-valued, convex, and 1-coercive function f. This is enough to get existence and uniqueness of equation (2.7), however it would clearly destroy the nice invariance property of Lemma 2.5 and we do not adopt it in the following.
2. As in (2.7) and the rest of the paper, the first argument of $\hat{g}$ is the negative of the Sharpe ratio. If we knew a priori that the Sharpe ratio was small relative to the minimal Lipschitz constant L in Remark 2.6, then we could allow BSDEs with Lipschitz drivers. However we choose to put restrictions on the driver, such as in Definition 2.3(iii), rather than on the range of the model parameters.

2.3. Indifference Valuation of European Claims

From the formula (2.3) for the indifference price P_t and Proposition 2.7, we have that

$$P_t = \hat{R}_t^{(0)} - \hat{R}_t^{(\xi)}. \tag{2.9}$$

From now on we will restrict ourselves to particular bounded payoffs, namely European put options with strike price K and maturity date T: $\xi = (K - S_T)^+$. Moreover, for the further treatment the substitutions

$$x := \log(S/K), \qquad \tau := T - t, \tag{2.10}$$

will be convenient and we introduce the following notation. Denote by L_T the layer $[0, T] \times \mathbb{R}^2$ and by $Q_{\tau_0,r}$ the open cylinder above the disk $B(m, r)$ with midpoint m, radius r, and height $0 < \tau_0 \leq T$: $Q_{\tau_0,r} :=]0, \tau_0[\times B(m, r)$. Since the location of the midpoint (once fixed) will play no further role, we skip it in the notation.

The following theorem characterizes the indifference price of a European put with respect to the dynamic convex risk measure with driver g under the stochastic volatility model (2.1).

THEOREM 2.9. *The buyer's indifference price of the European put option is given as*

$$P(\tau, x, y) = \tilde{u}(\tau, y) - u(\tau, x, y), \tag{2.11}$$

where $u \in C^{1+\beta/2,2+\beta}(Q_{T,r}) \cap C(L_T)$ for every cylinder $Q_{T,r}$ is the solution of the semi-linear parabolic PDE

$$\begin{cases} -u_\tau + Lu = \dfrac{1}{K}\hat{g}\big(-\lambda(y), \rho' K a(y) u_y\big); \\ u(0, x, y) = -(1 - e^x)^+, \end{cases} \tag{2.12}$$

with operator L given by

$$\begin{aligned} Lu &= \mathcal{M}u - \frac{1}{2}\sigma^2(y)u_x + (m(y) - \rho a(y)\lambda(y))u_y, \\ \mathcal{M}u &= \frac{1}{2}\sigma^2(y)u_{xx} + \rho\sigma(y)a(y)u_{xy} + \frac{1}{2}a^2(y)u_{yy}, \end{aligned} \tag{2.13}$$

and $\tilde{u}$ denotes the (x-independent) solution of (2.12), with altered initial condition $\tilde{u}(0, y) = 0$. Moreover there exists a solution of the Cauchy problem (2.12) (as well as to that with the altered initial condition) which is the unique classical solution that is bounded in L_T together with its derivatives.

Proof. By Ladyzhenskaya, Solonnikov, and Ural'tseva (1967, theorem V.8.1 and remark V.8.1), there exists a classical solution $v \in C^{1+\beta/2,2+\beta}(Q_{T,r}) \cap C(L_T)$ to the semilinear parabolic PDE

$$\begin{aligned} &v_t + \frac{1}{2}\sigma^2(y)S^2 v_{SS} + \frac{1}{2}a(y)^2 v_{yy} + \rho\sigma(y)a(y)S v_{Sy} + (m(y) - \rho a(y)\lambda(y))v_y \\ &\quad = \hat{g}(-\lambda(y), \rho' a(y) v_y), \end{aligned} \tag{2.14}$$

with terminal condition $v(T, S, y) = -(K - S)^+$ which is the unique classical solution to this PDE that is bounded in L_T together with its (first- and second-order) derivatives. Applying Itô's formula to $v(t, S_t, Y_t)$ and defining

$$\begin{aligned} \bar{Z}_t^1 &:= \sigma(Y_t)S_t v_s(t, S_t, Y_t) + \rho a(Y_t) v_y(t, S_t, Y_t), \\ \bar{Z}_t^2 &:= \rho' a(Y_t) v_y(t, S_t, Y_t), \\ \bar{R}_t &:= v(t, S_t, Y_t) \end{aligned}$$

shows that $(\bar{R}_t, \bar{Z}_t^1, \bar{Z}_t^2)$ solves the BSDE (2.7) for $(\hat{R}_t^{(\xi)}, \hat{Z}_t^1, \hat{Z}_t^2)$ with $\xi = (K - S_T)^+$, and therefore we identify $\hat{R}_t^{(\xi)} = v(t, S_t, Y_t)$. The transformation (2.10), together with $u(\tau, x, y) = v(t, s, y)/K$ leads to the Cauchy problem (2.12) for u. Finally, taking zero terminal condition for the PDE (2.14), and calling the solution $\tilde{v}(t, y)$ leads to $\hat{R}_t^{(0)} = \tilde{v}(t, Y_t)$. Therefore the indifference price in (2.9) is given by $P_t = \tilde{v}(t, Y_t) - v(t, S_t, Y_t)$, which, in transformed notation, leads to (2.11). □

Here u is the value function of the holder of the put option, and $\tilde{u}$ is related to the investment (or Merton) problem with trading only in the underlying stock and money market account. The nonlinearity in the PDE (2.12) is in its "Greek" u_y, that is, the Vega, and enters through the Legendre transform of the driver g in its first variable. For the familiar entropic risk measure, $g(z_1, z_2) = \gamma(z_1^2 + z_2^2)/2$, where $\gamma > 0$ is a risk-aversion parameter, we have $\hat{g}(\zeta, z_2) = (\zeta^2/\gamma - \gamma z_2^2)/2$. In this case, the nonlinearity is as u_y^2 (see, e.g., Benth and Karlsen 2005; Sircar and Zariphopoulou 2005).

2.4. Implied Volatility PDE

Our main goal is to establish an asymptotic expansion of the indifference price implied volatility in the limit of short time-to-maturity. To do so, we now adapt the approach of Berestycki et al. (2004) to establish a PDE satisfied by the Black–Scholes volatility $I(\tau, x, y)$ implied by the indifference pricing. Therefore, we note that in the Black–Scholes model with unit volatility, the no arbitrage pricing PDE is given by

$$\begin{cases} -U_\tau + \dfrac{1}{2}(U_{xx} - U_x) = 0; \\ \qquad\qquad U(0, x) = (1 - e^x)^+, \end{cases}$$

which can be represented explicitly as

$$U(\tau, x) = \Phi\left(-\frac{x}{\sqrt{\tau}} + \frac{\sqrt{\tau}}{2}\right) - e^x \Phi\left(-\frac{x}{\sqrt{\tau}} - \frac{\sqrt{\tau}}{2}\right),$$

where Φ is the cumulative density function of the standard normal distribution. Using the scaling properties of the Black–Scholes put price, the indifference pricing implied volatility $I(\tau, x, y)$ is hence given by the equation

$$P(\tau, x, y) = U(I^2(\tau, x, y)\tau, x). \tag{2.15}$$

Before we derive a PDE for the implied volatility, we give some a priori bounds on indifference prices and their associated implied volatilities.

PROPOSITION 2.10. *Denote by $P^{BS}(\tau, x; \sigma)$ the Black–Scholes price of the put calculated with constant volatility σ. Then*

$$P^{BS}(\tau, x; \underline{\sigma}) \leq P(\tau, x, y) \leq P^{BS}(\tau, x; \overline{\sigma}) \tag{2.16}$$

and

$$\underline{\sigma} \leq I(\tau, x, y) \leq \overline{\sigma}, \tag{2.17}$$

where $\underline{\sigma}$ and $\overline{\sigma}$ are the volatility bounds in Assumption 2.1.

The proof is given in Section 5.

To derive the PDE for the implied volatility, we plug (2.15) into the equation (2.11) and get after some calculations—detailed in Section 5—that the implied volatility I is subject to the nonlinear degenerate parabolic PDE

$$\begin{aligned} &-(\tau I^2)_\tau + \tau I \mathcal{M} I + I^2 \mathcal{G} \frac{x}{I} - \frac{1}{4}\tau^2 I^2 \mathcal{G} I + \tau q(y) I I_y \\ &\quad = 2\tau I I_y \frac{\hat{g}(-\lambda(y), \rho' K a(y) \tilde{u}_y) - \hat{g}(-\lambda(y), \rho' K a(y) u_y)}{K(\tilde{u}_y - u_y)}, \end{aligned} \tag{2.18}$$

with the square field type operator

$$\mathcal{G} I := \mathcal{M} I^2 - 2 I \mathcal{M} I \tag{2.19}$$

and

$$q(y) := \rho\sigma(y)a(y) + 2m(y) - 2\rho a(y)\lambda(y). \tag{2.20}$$

To motivate an initial condition for the Cauchy problem, we send formally τ to zero in (2.18). As the "Vega" $\nu = \tilde{u}_y - u_y$ tends also to zero as $\tau \downarrow 0$ (this is shown in Lemma 5.2), we observe that the quotient on the right side of (2.18)

$$\frac{\hat{g}(-\lambda(y), \rho' K a(y)\tilde{u}_y) - \hat{g}(-\lambda(y), \rho' K a(y) u_y)}{K(\tilde{u}_y - u_y)} \to \rho' a(y)\hat{g}_{z_2}(-\lambda(y), 0),$$

which is bounded by the definition of strictly quadratic drivers and Lemma 2.5. Dividing by I^2, this leads to the formal limit equation

$$\mathcal{G}\frac{x}{I(0, x, y)} = 1. \tag{2.21}$$

REMARK 2.11. Our Cauchy problem is similar to that derived for no arbitrage pricing implied volatilities in Berestycki et al. (2004), where they have the same equation (2.18), but (i) without the last term on the left side (which here is due to the change in measure from physical to a risk-neutral one); and (ii) without the right-side term (which here is due to the dynamic convex risk measure used for indifference pricing). However, our initial condition (2.21), which does not depend on $\hat{g}$ and the drift of the stochastic volatility model, is exactly the same as theirs.

Now we turn this heuristic argument into a precise statement.

THEOREM 2.12. *The implied volatility function $I(\tau, x, y)$ generated by the indifference pricing mechanism is the unique solution $I \in C^{1+\beta/2,2+\beta}(Q_{T,r}) \cap C(L_T)$ to the following nonlinear parabolic Cauchy problem*

$$\begin{aligned} -(\tau I^2)_\tau + \tau I \mathcal{M} I + I^2 \mathcal{G}\frac{x}{I} - \frac{1}{4}\tau^2 I^2 \mathcal{G} I + \tau q(y) I I_y \\ = 2\tau I I_y \frac{\hat{g}(-\lambda(y), \rho' K a(y)\tilde{u}_y) - \hat{g}(-\lambda(y), \rho' K a(y) u_y)}{K(\tilde{u}_y - u_y)}, \end{aligned} \tag{2.22}$$

where $\mathcal{M}$, $\mathcal{G}$, and q are given by (2.13), (2.19), and (2.20), respectively. The initial condition is given as $I(0, x, y) = x/\psi(x, y)$ where ψ is the unique viscosity solution of the eikonal equation

$$\begin{cases} \mathcal{G}\psi = 1; \\ \psi(0, y) = 0; \\ \psi(x, y) > 0 \quad \textit{for } x > 0. \end{cases} \tag{2.23}$$

The proof is given in Section 5.

It is worthwhile to note that indifference prices are not linear. Indeed, buying double the amount of securities will not lead to twice the price. In this way also the volatility implied by indifference prices is quantity-dependent as a consequence of the appearance of u_y and $\tilde{u}_y$ on the right side of equation (2.22). However, the nonlinearity in quantity is not observed in the zeroth- and first-order small time-to-maturity approximation as we show in the following subsection. Moreover, deriving the PDE for the indifference price implied volatility for buying n put options results in the same Cauchy problem as in Theorem 2.12, where one has only to replace K by nK in every appearance in equation (2.22).

2.5. Small-Time Expansion

In the short time limit the implied volatility under indifference pricing is equal to the usual one as calculated by Berestycki et al. (2004), as the initial conditions are the same (see Remark 2.11). The subtleties of the indifference pricing appear only away from maturity. This can be seen by methods of a formal small-time expansions. We make the *Ansatz* of an asymptotic expansion of the implied volatility:

$$I(\tau, x, y) = I^0(x, y)\big(1 + \tau I^1(x, y) + O(\tau^2)\big). \tag{2.24}$$

As seen above, the term $I^0(x, y) = I(0, x, y)$ is given via solution of the eikonal equation (2.23). To find the PDE for I^1, we plug the expansion (2.24) into the equation (2.22) and compare the first-order terms for $\tau \to 0$. This leads to the inhomogeneous linear transport equation

$$\begin{aligned} 2I^1(x, y) &+ \frac{\psi}{2}(\sigma^2(y)\psi_x + \rho\sigma(y)a(y)\psi_y)I^1_x(x, y) \\ &+ \frac{\psi}{2}(\rho\sigma(y)a(y)\psi_x + a^2(y)\psi_y)I^1_y(x, y) - \frac{\psi}{x}\mathcal{M}\frac{x}{\psi} \\ &= -(q(y) - \rho' a(y)\hat{g}_{z_2}(-\lambda(y), 0))\frac{\psi_y}{\psi}, \end{aligned} \tag{2.25}$$

where ψ is again the solution of the eikonal equation (2.23).

It is important to observe that the dependence of this first-order approximation on the risk measure (via its driver g) is given merely by the evaluation at $z_2 = 0$ of the derivative of its Fenchel–Legendre transform $\hat{g}$ with respect to the second component.

Comparing our PDE to the analogous equation in the arbitrage-pricing setting of Berestycki et al. (2004) (who, however, prescribe only the methodology in general and make explicit calculations just in one example), we note the presence of two additional terms in (2.25)—one due to the change to a risk-neutral probability measure and the second due to indifference pricing with a dynamic convex risk measure.

Furthermore in many cases we can obtain an interior boundary condition for the PDE at $x = 0$ by sending x formally to zero in (2.25), since by $\psi \to 0$ and ψ_x it follows that the first-order coefficients vanish. We will show this specific procedure on a concrete example in Section 3.2. Imposing higher regularity on the coefficients and on $\hat{g}$ one can obtain also higher-order terms in the expansion of the implied volatility. This is done by using Taylor expansions of $\hat{g}$ (in the second component) and $\tilde{u}_y$ and u_y (in τ).

3. EXAMPLES AND COMPUTATIONS

In this section, we introduce a family of dynamic risk measures within which to present the effect of risk aversion on implied volatilities, first using the asymptotic approximation in the Hull–White stochastic volatility model, and later using a numerical solution of the quasilinear option pricing PDE.

3.1. Distorted Entropic Dynamic Convex Risk Measures

To study the impact of the driver on the implied volatility, we will now turn to a nicely parameterized family of risk measures. Therefore we define the following class of drivers, generating distorted forms of the entropic risk measure:

$$g^{\eta,\gamma}(z_1, z_2) := \frac{\gamma}{2}\left(z_1^2 + z_2^2\right) + \eta\gamma z_1 z_2 + \frac{\eta^2\gamma}{2} z_2^2 = \frac{\gamma}{2}\left((z_1 + \eta z_2)^2 + z_2^2\right). \tag{3.1}$$

It is clear that in the case $\eta = 0$ this is the driver connected to the classical entropic risk measure, whereas η can be seen as a parameter which describes in which way volatility risk increases also the risk coming from the tradable asset. As we will see later in Section 3.4, η effectively plays the role of a *volatility risk premium*. In the case of the usual entropic risk measure every level set of the driver is a circle whose radius is governed by the parameter γ. In the distorted case it is now an ellipse where η determines additionally the eccentricity.

Turning to the Fenchel–Legendre transform, we have

$$\hat{g}^{\eta,\gamma}(\zeta, z_2) = \frac{1}{2\gamma}\zeta^2 - \frac{\gamma}{2} z_2^2 - \eta\zeta\, z_2. \tag{3.2}$$

Plugging this into (2.22), we see that the right-hand side now reads

$$\tau I I_y(2\eta\rho'\lambda(y)a(y) - \gamma K\rho'^2 a(y)^2(\tilde{u}_y + u_y))$$

and we remark in particular that γ scales with K and hence also with the number of securities bought (as mentioned at the end of Section 2.4). In particular, we see again that the term appearing in the first-order approximation of the implied volatility, $\hat{g}^{\eta,\gamma}_{z_2}(-\lambda(y), 0) = \eta\lambda(y)$, is independent of γ.

3.2. Short-Time Asymptotics for the Hull–White Model

In the following we look at an example which is an adaption of Example 6.1/6.3 of Berestycki et al. (2004). Let the stochastic volatility model be given as the Hull–White model

$$\begin{cases} dS_t = \mu(Y_t)S_t\, dt + Y_t S_t\, dW_t^1, & S_0 = s; \\ dY_t = \kappa\, Y_t\, dW_t^2, & Y_0 = y. \end{cases} \tag{3.3}$$

for two independent Brownian motions W^1, W^2. Obviously the model does not fall in the class considered above because the volatility $\sigma(Y) = Y$ is a geometric Brownian motion that is not bounded above or away from zero. Nevertheless, by a change of variables we will see that the results hold.

Writing down the pricing PDE in the case of the distorted entropic risk measure (3.1), we get

$$-u_\tau + \frac{1}{2}\left(y^2 u_{xx} + \kappa^2 y^2 u_{yy}\right) - \frac{1}{2} y^2 u_x = \frac{1}{2\gamma K}\frac{\mu(y)^2}{y^2} - \frac{\gamma K\kappa^2}{2} y^2 u_y^2 + \eta\kappa\mu(y) u_y,$$

and one sees that by the time change $\tau \mapsto \tau y^2$ (the boundary $y = 0$ is not hit when we start with $y_0 > 0$ given that Y_t is a geometric Brownian motion) and setting $\tilde{\mu}(y) := \mu(y)/y^2$

the equation becomes

$$-u_\tau + \frac{1}{2}(u_{xx} + \kappa^2 u_{yy}) - \frac{1}{2}u_x = \frac{1}{2\gamma K}\tilde{\mu}(y)^2 - \frac{\gamma K \kappa^2}{2}u_y^2 + \eta\kappa\tilde{\mu}(y)u_y.$$

This equation has a solution (again by Ladyzhenskaya et al. 1967, theorem V.8.1 and remark V.8.1), at least in the case that $\tilde{\mu}$ is locally β-Hölder continuous (which in turn implies that $\mu(y) = O(y^2)$ as $y \to 0$) .

In absence of global bounds on the volatility we are not able to derive results about existence and uniqueness of solutions for the PDE (2.22) for the implied volatility. Nevertheless we can postulate the small-time expansion to get

$$\text{(3.4)} \qquad \begin{cases} \psi_x^2 + \kappa^2\psi_y^2 = 1/y^2; \\ \quad \psi(0, y) = 0; \\ \quad \psi(x, y) > 0 \quad \text{for } x > 0 \end{cases}$$

as the PDE characterizing its zeroth-order term and

$$\text{(3.5)} \qquad \begin{cases} 0 = 2I^1 + y^2\psi\psi_x I_x^1 + \kappa^2 y^2 \psi\psi_y I_y^1 - \dfrac{x}{\psi}\mathcal{M}\dfrac{\psi}{x} - \eta\kappa\mu(y)\dfrac{\psi_y}{\psi}; \\ I^1(0, y) = \dfrac{\kappa^2}{12} + \eta\dfrac{\mu(y)}{2y} \end{cases}$$

for the first. The interior boundary condition for I^1 at $x = 0$ follows from the formal asymptotics as $x \to 0$. As derived in Berestycki et al. (2004), the zeroth-order term of the expansion is given via the solution of (3.4),

$$\text{(3.6)} \qquad \psi(x, y) = \frac{1}{\kappa}\ln\left(\frac{\kappa x}{y} + \sqrt{1 + \frac{\kappa^2 x^2}{y^2}}\right),$$

as $I^0(x, y) = x/\psi(x, y)$, whereas for (3.5) we can guarantee only a solution in the case where $\mu(y) = O(y^3)$ since $I_y^0/I^0 = -\psi_y/\psi = o(1/y)$ as $y \to 0$. Obviously this means practically that we need an extreme drift in the Hull–White model to compensate the very volatile volatility process. However, setting, e.g., $\mu(y) = \mu y^3$ for some constant μ, we are able to solve the PDE (3.5) explicitly by the method of characteristics to get

$$\text{(3.7)} \qquad I^1(x, y) = \frac{1}{\psi^2(x, y)}\left(\ln\left(\frac{y}{x}\psi(x, y)\Big(1 + \frac{\kappa^2 x^2}{y^2}\Big)^{\frac{1}{4}}\right) + \eta\frac{\mu x^2}{2}\right).$$

In Figure 3.1, we rely on the parameter set

$$\mu = 6; \qquad \kappa = 7; \qquad y = 0.3; \qquad \tau = 0.1.$$

Whereas the parameter γ does not appear in the first-order approximation, the distortion parameter η has a double effect. On the one hand, it shifts the smile at the money a small amount, on the other hand it changes more significantly the wing behavior of the smile, adding to the asymptotics the term $\eta\frac{\mu\kappa^2}{2}\frac{x^2}{(\ln|x|)^2}$ (since $x^2/\psi^2 \sim k^2x^2/(\ln|x|)^2$ as $x \to \pm\infty$). This changes the whole wing behavior, since $I^0 \sim \kappa\frac{|x|}{\ln|x|}$ and $I^1 \to 0$ for $\eta = 0$ as $x \to \pm\infty$. Of course $\eta = 0$ corresponds to the first-order term of martingale pricing as Berestycki et al. (2004). Positive η (hence a positive impact of the volatility risk on the risk of the traded asset) increases the implied volatility and steepens the wings.

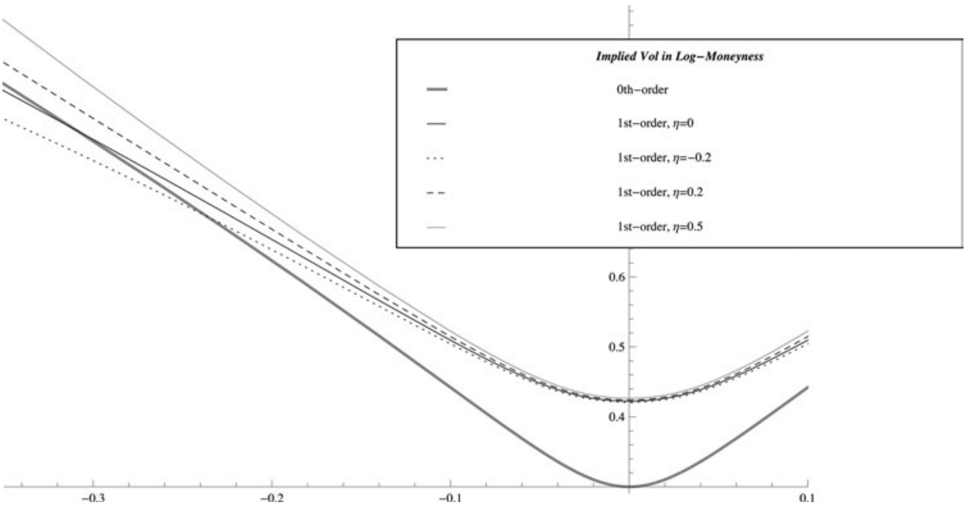


FIGURE 3.1. Implied volatility in terms of log-moneyness $\log(K/S_0)$ for the Hull–White model: zeroth- and first-order approximation in dependence of η.

3.3. Calibration Exercise

We perform a simple calibration exercise using S&P 500 implied volatilities to show how the approximation (2.24) could be used to recover some information about the market's risk measure. We work with the short-time approximation in the uncorrelated Hull–White model (3.3) of Section 3.2 and within the family of distorted dynamic convex risk measure (3.1).

We suppose that the near-the-money option prices are given by expectations under an equivalent martingale measure, specifically the minimal martingale measure. This corresponds to setting the distortion $\eta = 0$, and so the short-time asymptotic approximation for implied volatilities I is

$$I \approx \frac{x}{\psi(x, y)}(1 + \tau I^1(x, y)),$$

where $\psi(x, y)$ is given by (3.6) and $I^1(x, y)$ by (3.7) with $\eta = 0$. This gives

$$I \approx I_M := \frac{x}{\psi(x, y)}\left(1 + \frac{\tau}{\psi^2(x, y)} \ln\left(\frac{y}{x}\psi(x, y)\left(1 + \frac{\kappa^2 x^2}{y^2}\right)^{\frac{1}{4}}\right)\right),$$

where we see that the approximation I_M depends only on the parameter κ and the current volatility level y. Then having fit the liquid implied volatilities to recover estimates of κ and y, we can write the short-time approximation with distortion as

$$I \approx I_M + (\mu\eta)\frac{\tau x^3}{2\psi^3(x, y)}, \tag{3.8}$$

from which we can estimate the combination $\mu\eta$.

In Figure 3.2, we show the result of a fit to S&P 500 implied volatilities from options with nine days to maturity on October 7, 2010. We comment that the Hull–White model is not the best choice of stochastic volatility model, and our intention is primarily to

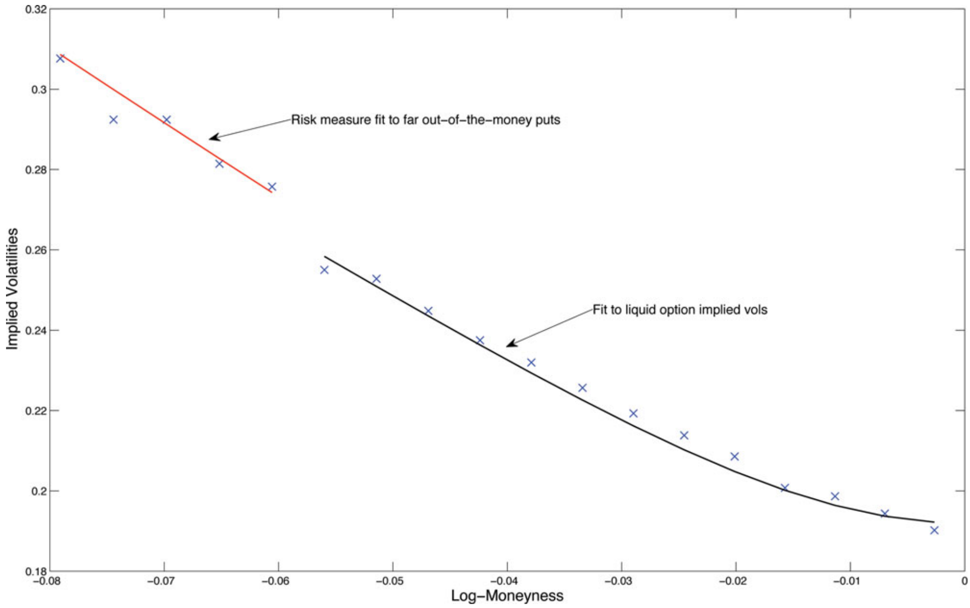


FIGURE 3.2. Fit of short-time asymptotics to S&P 500 implied volatility data nine days from maturity.

demonstrate the procedure. In particular, because our asymptotic formulas are for the uncorrelated model, which always has a smile with minimum implied volatility at the money, we restrict the fit to options with negative log-moneyness, which is mainly out-of-the-money put options. The first part of the least-squares fitting to options with log-moneyness between -0.06 and 0 gives $\kappa \approx 6.6$, and $y \approx 0.18$. Then fitting to (3.8) the options with log-moneyness less than -0.06, which we view as less liquid and more reflective of the market's risk measure, gives the estimate $\mu\eta \approx 35$. We note that the μ in this model is expected to be rather large because the composite drift of the stock is μy^3, and $y \ll 1$. So the estimate we get from the data of $\mu\eta$ is not unreasonable. In particular it reveals a *positive* distortion coefficient η.

3.4. Numerical Study

We consider the buyer's indifference price of one European put option with respect to the family of *distorted entropic risk measures* defined by (3.1). We work within the stochastic volatility model (2.1) and, for the numerical solution, we return to the primitive variables (t, S, y). Denote by $\mathcal{L}_y$ the generator of the Markov process Y:

$$\mathcal{L}_y = \frac{1}{2}a(y)^2\frac{\partial^2}{\partial y^2} + m(y)\frac{\partial}{\partial y},$$

and by $\mathcal{L}_{S,y}$ the generator of $(S,\ Y)$:

$$\mathcal{L}_{S,y} = \mathcal{L}_y + \frac{1}{2}\sigma(y)^2S^2\frac{\partial^2}{\partial S^2} + \rho a(y)\sigma(y)S\frac{\partial^2}{\partial S\partial y} + \mu(y)\frac{\partial}{\partial S}.$$

From (2.9), the buyer's indifference price of a put option with strike K and expiration date T at time $t < T$ when $S_t = S$ and $Y_t = y$, is given by

$$P(t, S, y) = \varphi(t, S, y) - \varphi_0(t, y),$$

where φ solves

$$\varphi_t + \left(\mathcal{L}_{S,y} - \rho a(y)\lambda(y)\frac{\partial}{\partial y}\right)\varphi = -\hat{g}(-\lambda(y), -\rho' a(y)\varphi_y), \tag{3.9}$$
$$\varphi(T, S, y) = (K - S)^+,$$

and $\varphi_0(t, y)$ solves the same PDE without the S-derivatives and with zero terminal condition. Note that $\varphi = -v$, where v was the solution to the PDE problem in (2.14), and $\varphi_0 = -\tilde{v}$ which was introduced in the proof of Theorem 2.9.

As $\hat{g}$ is given by (3.2), we can rewrite (3.9) as

$$\varphi_t + \left(\mathcal{L}_{S,y} - (\rho + \eta\rho')a(y)\lambda(y)\frac{\partial}{\partial y}\right)\varphi = -\frac{\lambda^2(y)}{2\gamma} + \frac{1}{2}(1 - \rho^2)\gamma a(y)^2\varphi_y^2. \tag{3.10}$$

This shows that η plays the role of a *volatility risk premium* in that it enters as a drift adjustment for the volatility-driving process Y. However the nonlinearity of the PDE is through a quadratic term in φ_y, as in the case of the entropic risk measure.

Moreover, introducing the transformation

$$\varphi_0(t, y) = -\frac{1}{\gamma(1 - \rho^2)} \log f(t, y),$$

leads to the *linear* PDE problem for f:

$$f_t + \left(\mathcal{L}_y - (\rho + \eta\rho')a(y)\lambda(y)\frac{\partial}{\partial y}\right) f - \frac{1}{2}\lambda^2(y)(1 - \rho^2) f = 0, \tag{3.11}$$

$$f(T, y) = 1. \tag{3.12}$$

Therefore the indifference price is given by

$$P(t, S, y) = \varphi(t, S, y) + \frac{1}{\gamma(1 - \rho^2)} \log f(t, y).$$

In the numerical solutions, we take the volatility-driving process (Y_t) to be an Ornstein–Uhlenbeck process with the dynamics:

$$dY_t = \alpha(m - Y_t)\,dt + \nu\sqrt{2\alpha}\left(\rho\, dW_t^1 + \rho'\, dW_t^2\right),$$

and we choose a function $\sigma(Y)$ that gives realistic volatility characteristics. For the OU process (Y_t), the rate of mean-reversion is α, the long-run mean-level is m, and the long-run variance is ν^2. For the computations, we will take $\alpha = 5$, $m = 0$, $\nu^2 = 1$, $\rho = -0.2$, and

$$\sigma(y) = \frac{0.7}{\pi}(\arctan(y - 1) + \pi/2) + 0.03,$$

so that $\sigma(m) = 0.2050$. The parameter ν measures approximately the standard deviation of volatility fluctuations. The values are chosen such that the interval of width one

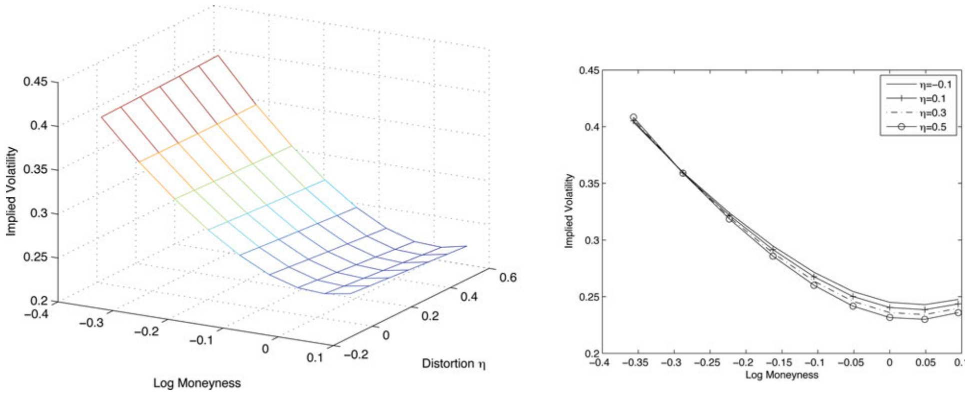


FIGURE 3.3. Implied volatility from the arctangent stochastic volatility model in terms of log-moneyness $\log(K/S_0)$ and η with fixed $\gamma = 0.5$.

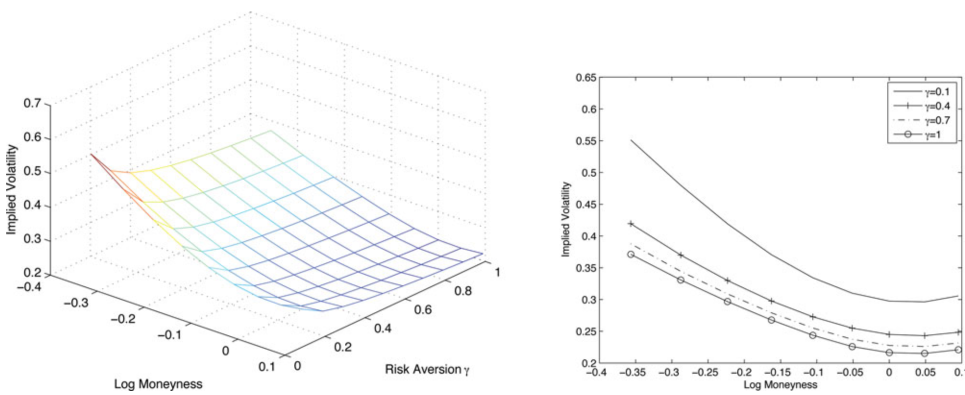


FIGURE 3.4. Implied volatility from the arctangent stochastic volatility model in terms of log-moneyness $\log(K/S_0)$ and γ with fixed $\eta = 0.2$.

standard deviation for Y under its stationary distribution is $(-1, 1)$, and this translates roughly to the interval (0.13, 0.38) for volatility σ. The two standard deviation interval for volatility is approximately (0.10, 0.56).

We first solve the quasilinear PDE (3.10) for φ using implicit finite-differences on the linear part, and explicit on the nonlinear part. Then we solve the linear PDE problem (3.11) for f. This procedure is done for a current stock price $S_0 = 100$ and $\sigma(Y_0) = 0.223$, calculating the solutions for European put options expiring in three months for strikes K in the range [70, 110] for different values of the distortion parameter η and the risk-aversion parameter $\gamma > 0$.

Figure 3.3 reveals a more complex picture regarding the effect of η away from the short maturity asymptotic approximation. We see, as in Figure 3.1 from the asymptotics, increasing η increases the skew slope; however it also shifts down the levels of implied volatility around the money (as opposed to the opposite effect we saw in Figure 3.1).

Figure 3.4 shows, as we would expect, that increasing risk aversion γ decreases the implied volatility skew which comes from the indifference price of the buyer who is willing to pay less for the risk of the option position. It also has a secondary effect of flattening the skew out of the money.

4. CONCLUSION

We have derived a nonlinear PDE for the implied volatility from indifference pricing with respect to dynamic convex risk measures defined by BSDEs under diffusion stochastic volatility models. Our asymptotic analysis has highlighted the principal effect of the risk measure on option implied volatility at short maturities, namely through the appearance of $\hat{g}_{z_2}$ in the first-order correction solving (2.25).

In the example of Section 3.2, this translates explicitly to a steepening effect on the implied volatility smile from the distortion parameter η. Numerical computations confirm this away from short maturity too, as well as quantifying the effect of risk aversion on the level of implied volatilities.

The analysis can be used to infer some information about the driver, for example η and γ in the family (3.1), from market implied volatilities. This is illustrated using the short-time asymptotics for the Hull–White model on S&P 500 implied volatilities in Section 3.3.

5. PROOFS

5.1. Price and Volatility Bounds

LEMMA 5.1 (First Price Comparison). *Suppose that* $\underline{P} = \underline{\tilde{u}} - \underline{u}$ *and* $\overline{P} = \tilde{\overline{u}} - \overline{u}$ *for some* $\underline{u}, \underline{\tilde{u}}, \overline{u}, \tilde{\overline{u}} \in C^{1,2,}(Q_{\tau_0,r}) \cap C(\overline{Q_{\tau_0,r}})$, $0 < \tau_0 \leq T$, *satisfy*

$$\overline{P}_\tau \geq L\overline{P} - \frac{1}{K}\big(\hat{g}\big(-\lambda(y), \rho' Ka(y)\tilde{\overline{u}}_y\big) - \hat{g}\big(-\lambda(y), \rho' Ka(y)\overline{u}_y\big)\big), \quad in\ Q_{\tau_0,r} \tag{5.1}$$

$$\underline{P}_\tau \leq L\underline{P} - \frac{1}{K}\big(\hat{g}\big(-\lambda(y), \rho' Ka(y)\underline{\tilde{u}}_y\big) - \hat{g}\big(-\lambda(y), \rho' Ka(y)\underline{u}_y\big)\big), \quad in\ Q_{\tau_0,r} \tag{5.2}$$

as well as

$$\overline{P} \geq \underline{P} \quad on \quad \{0\} \times \mathbb{R}^2 \cap \overline{Q_{\tau_0,r}} \quad and \quad]0, \tau_0] \times \mathbb{R}^2 \cap \partial Q_{\tau_0,r}. \tag{5.3}$$

Then $\overline{P} \geq \underline{P}$ *on* $Q_{\tau_0,r}$.

Proof. Even if the form of this comparison principle for sub- and superprices seems to be quite unusual, the proof follows directly along the lines of Friedman (1964, theorem 2.16) since the function $\hat{g}$ contains no second derivatives. To be precise: this argument leads a version where the inequalities in (5.1), (5.3) and the conclusion are strict. But setting $\overline{P}^{\varepsilon} = \overline{P} + \varepsilon(1 + \tau)$ one gets a strict superprice and sending ε to zero yields the stated version. □

5.2. Proof of Proposition 2.10

Proof. To prove (2.16), we intend to invoke the above comparison principle for the price process given in Theorem 2.9 since it is clear that the Black–Scholes pricing functions are sub- resp. supersolutions of the PDE. Unfortunately we have the indifference price only as solution of a Dirichlet problem which does not give rise to directly comparable lateral boundary conditions; thus we have to alter the argument a bit.

Denote for $N \in \mathbb{N}$ by $u^{N,\underline{\sigma}}$ the solution of the initial/boundary-value problems

$$\begin{cases} -u_\tau^{N,\underline{\sigma}} + Lu^{N,\underline{\sigma}} = \dfrac{1}{K}\hat{g}\big(-\lambda(y), \rho' Ka(y)u_y^{N,\underline{\sigma}}\big); \\ u^{N,\underline{\sigma}}(0,x,y) = -(1-e^x)^+; \\ u^{N,\underline{\sigma}}(\tau,x,y)|_{\partial B(0,N)} = -P^{BS}(\tau,x;\underline{\sigma}). \end{cases}$$

By a classical argument (Ladyzhenskaya et al. 1967, section V.§8), we can extract a subsequence $u^{N_k,\underline{\sigma}}$ of $u^{N,\underline{\sigma}}$ such that $u^{N_k,\underline{\sigma}}$ converges together with its derivatives to u and its derivatives pointwise in L_T. The same is true for $\tilde{u}^N$ given by

$$\begin{cases} -\tilde{u}_\tau^N + L\tilde{u}^N = \dfrac{1}{K}\hat{g}\big(-\lambda(y), \rho' Ka(y)\tilde{u}_y^N\big); \\ \tilde{u}^N(0,x,y) = 0; \\ \tilde{u}^N(\tau,x,y)|_{\partial B(0,N)} = 0. \end{cases}$$

Thus $P^{N,\underline{\sigma}}(\tau,x,y) = u^{N,\underline{\sigma}}(\tau,x,y) - \tilde{u}(\tau,x,y)$ satisfies

$$\begin{cases} -P_\tau^{N,\underline{\sigma}} + LP^{N,\underline{\sigma}} = \dfrac{1}{K}\hat{g}\big(-\lambda(y), \rho' Ka(y)\tilde{u}_y^N\big) - \dfrac{1}{K}\hat{g}\big(-\lambda(y), \rho' Ka(y)u_y^{N,\underline{\sigma}}\big); \\ P^{N,\underline{\sigma}}(0,x,y) = (1-e^x)^+; \\ P^{N,\underline{\sigma}}(\tau,x,y)\big|_{\partial B(0,N)} = P^{BS}(\tau,x;\underline{\sigma}), \end{cases} \tag{5.4}$$

and $P^{N,\underline{\sigma}} \to P$ along a subsequence.

Noting that $P^{BS}(\tau,x;\underline{\sigma})$ is a subprice on every cylinder $Q_{T,N}$ by writing it in the odd form $P^{BS}(\tau,x;\underline{\sigma}) = 0 - (-P^{BS}(\tau,x;\underline{\sigma}))$ to satisfy the comparison principle of Lemma 5.1, we have $P^{BS}(\tau,x;\underline{\sigma}) \le P^{N,\underline{\sigma}}$ on $Q_{T,N}$ and hence in the limit $P^{BS}(\tau,x;\underline{\sigma}) \le P$. The other direction of Theorem 2.9 is proved, of course, by the same argument using $P^{BS}(\tau,x;\overline{\sigma})$ as superprice.

Finally a reformulation of the achieved result reads

$$U(\underline{\sigma}^2\tau, x) \le U(I(\tau,x,y)^2\tau, x) \le U(\overline{\sigma}^2\tau, x)$$

and so the strict monotonicity of U with respect to the first variable yields (2.17). □

5.3. Deriving the PDE for the Implied Volatility

To derive the PDE for the implied volatility, we note that from (2.11) and (2.12), it follows that the indifference price P obeys the equation

$$-P_\tau + LP = \frac{1}{K}(\hat{g}(-\lambda(y), \rho' Ka(y)\tilde{u}_y) - \hat{g}(-\lambda(y), \rho' Ka(y)u_y)).$$

Substituting $P(\tau,x,y) = U(I^2(\tau,x,y)\tau, x)$ we note that the derivatives relate

$$\begin{aligned} P_\tau &= U_\tau \cdot (\tau I^2)_\tau; \qquad P_x = U_\tau \cdot 2II_x\tau + U_x; \qquad P_y = U_\tau \cdot 2II_y\tau; \\ P_{xx} &= U_{\tau\tau} \cdot 4I^2I_x^2\tau^2 + U_\tau \cdot 2I_x^2\tau + U_\tau \cdot 2II_{xx}\tau + U_{\tau x} \cdot 4II_x\tau + U_{xx}; \\ P_{xy} &= U_{\tau\tau} \cdot 4I^2I_xI_y\tau^2 + U_\tau \cdot 2I_xI_y\tau + U_\tau \cdot 2II_{xy}\tau + U_{\tau x} \cdot 2II_y\tau \\ P_{yy} &= U_{\tau\tau} \cdot 4I^2I_y^2\tau^2 + U_\tau \cdot 2I_y^2\tau + U_\tau \cdot 2II_{yy}\tau. \end{aligned} \tag{5.5}$$

Plugging this into (5.5) and dividing by U_τ, we derive the equation (2.22) by noting that

(5.6)
$$\frac{U_{xx}-U_x}{U_\tau}=2;\qquad \frac{U_{\tau\tau}}{U_\tau}=\frac{x^2}{2I^4\tau^2}-\frac{1}{2I\tau}-\frac{1}{8};\qquad \frac{U_{\tau x}}{U_\tau}=-\frac{x}{I\tau}+\frac{1}{2};\qquad \frac{1}{U_\tau}=\frac{2II_y\tau}{\tilde{u}_y-u_y}.$$

The initial condition follows by observing that Vega vanishes as τ tends to zero:

LEMMA 5.2. *It holds that $v=\tilde{u}_y-u_y\to 0$ uniformly on compacts as $\tau\to 0$.*

Proof. Choose the cylinder $Q_{T,r}$ in the layer L_T such that the compact set is contained. Thus $u,\tilde{u}\in C^{1+\beta/2,2+\beta}(Q_{T,r})\cap C(L_T)$ implies that u_y and $\tilde{u}_y$ are $\beta/2$-Hölder continuous with some Hölder constant c, whence

$$\begin{aligned}|v(\tau,x,y)|&=|\tilde{u}_y(\tau,y)-u_y(\tau,x,y)|\\&\le|\tilde{u}_y(\tau,y)-\tilde{u}_y(0,y)|+|u_y(\tau,x,y)-u_y(0,x,y)|\le 2c\tau^{\frac{\beta}{2}}\to 0\end{aligned}$$

as $\tau\to 0$ since $\tilde{u}_y(0,y)$ and $u_y(0,x,y)$ exist and are equal to zero by the definition of the initial conditions. □

5.4. Implied Volatility—Proof of the Main Theorem

LEMMA 5.3 (Second Price Comparison). *Recall that u is the solution of the Cauchy problem (2.12) and $\tilde{u}$ of the same problem with initial condition equal to zero. Suppose that $\overline{P},\underline{P}\in C^{1,2}(Q_{\tau_0,r})\cap C(\overline{Q_{\tau_0,r}})$, $0<\tau_0\le T$, satisfy*

$$\overline{P}_\tau\ge L\overline{P}-\frac{1}{K}(\hat{g}(-\lambda(y),\rho'Ka(y)\tilde{u_y})-\hat{g}(-\lambda(y),\rho'Ka(y)u_y)),\quad in\ Q_{\tau_0,r}\tag{5.7}$$

$$\underline{P}_\tau\le L\underline{P}-\frac{1}{K}(\hat{g}(-\lambda(y),\rho'Ka(y)\tilde{u_y})-\hat{g}(-\lambda(y),\rho'Ka(y)u_y)),\quad in\ Q_{\tau_0,r}\tag{5.8}$$

as well as

$$\overline{P}\ge P\ge\underline{P}\quad on\quad]0,\tau_0]\times\mathbb{R}^2\cap\partial Q_{\tau_0,r},\tag{5.9}$$

and

$$\overline{P}(0,x,y)=\underline{P}(0,x,y)=P(0,x,y)=(1-e^x)^+$$

then $\overline{P}\ge P\ge\underline{P}$ on $Q_{\tau_0,r}$.

Proof. We note that inequality (5.7) implies $(\overline{P}-P)_\tau\ge L(\overline{P}-P)$ which implies together with the lateral bound $(\overline{P}-P)\ge 0$ on $]0,\tau_0]\times\mathbb{R}^2\cap\partial Q_{\tau_0,r}$ and the initial condition $\overline{P}(0,x,y)-P(0,x,y)=0$ that $\overline{P}\ge P$ on $Q_{\tau_0,r}$ by the classical comparison principle. The second inequality is proved in the same way. □

LEMMA 5.4 (Volatility Comparison). *Suppose that $\overline{I},\underline{I}\in C^{1,2}(Q_{\tau_0,r})\cap C(\overline{Q_{\tau_0,r}})$, $0<\tau_0\le T$, satisfy*

$$\begin{aligned}(\tau\underline{I}^2)_\tau&\le\tau\underline{I}\mathcal{M}\underline{I}+\underline{I}^2\mathcal{G}\underline{I}+\tau q(y)\underline{II}_y\\&\quad-2\tau\underline{II}_y\left(\rho a(y)\lambda(y)+\frac{\hat{g}(-\lambda(y),\rho'Ka(y)\tilde{u}_y)-\hat{g}(-\lambda(y),\rho'Ka(y)u_y)}{K(\tilde{u}_y-u_y)}\right)\end{aligned}$$

resp.

$$(\tau\overline{I}^2)_\tau \ge \tau\overline{I}\mathcal{M}\overline{I} + \overline{I}^2\mathcal{G}\overline{I} + \tau q(y)\overline{I}\overline{I}_y$$
$$- 2\tau\overline{I}\overline{I}_y\left(\rho a(y)\lambda(y) + \frac{\hat{g}\big(-\lambda(y), \rho' Ka(y)\tilde{u}_y\big) - \hat{g}\big(-\lambda(y), \rho' Ka(y)u_y\big)}{K(\tilde{u}_y - u_y)}\right)$$

in $Q_{\tau_0,r}$ *together with the lateral comparison*

$$\underline{I}(\tau, x, y) \le I(\tau, x, y) \le \overline{I}(\tau, x, y) \quad on \quad]0, \tau_0] \times \mathbb{R}^2 \cap \partial Q_{\tau_0,r}$$

and the initial growth condition

$$\lim_{\tau\to 0} \tau\underline{I}^2(\tau, x, y) = \lim_{\tau\to 0} \tau\overline{I}^2(\tau, x, y) = 0 \quad on \quad \{0\} \times \mathbb{R}^2 \cap \overline{Q_{\tau_0,r}}. \tag{5.10}$$

Then it holds that

$$\underline{I}(\tau, x, y) \le I(\tau, x, y) \le \overline{I}(\tau, x, y) \quad in \quad Q_{\tau_0,r}.$$

Proof. Define first $\overline{P}(\tau, x, y) := U(\overline{I}^2(\tau, x, y)\tau, x)$ and $\underline{P}(\tau, x, y) := U(\underline{I}^2(\tau, x, y)\tau, x)$. Then by the same calculation as in the derivation of the PDE (2.22) of the implied volatility in the paragraph above we get

$$\overline{P}_\tau \ge L\overline{P} - \frac{1}{K}(\hat{g}(-\lambda(y), \rho' Ka(y)\tilde{u_y}) - \hat{g}(-\lambda(y), \rho' Ka(y)u_y)), \quad \text{in } Q_{\tau_0,r}$$

$$\underline{P}_\tau \le L\underline{P} - \frac{1}{K}(\hat{g}(-\lambda(y), \rho' Ka(y)\tilde{u_y}) - \hat{g}(-\lambda(y), \rho' Ka(y)u_y)), \quad \text{in } Q_{\tau_0,r}$$

as well as the lateral boundary condition

$$\underline{P}(\tau, x, y) \le P(\tau, x, y) \le \overline{P}(\tau, x, y) \quad \text{in} \quad Q_{\tau_0,r}.$$

Moreover, the growth condition (5.10) implies by the continuity of U that $\overline{P}(0, x, y) = \underline{P}(0, x, y) = P(0, x, y) = (1 - e^x)^+$. Thus we can use Lemma 5.3 to infer $\underline{P}(\tau, x, y) \le P(\tau, x, y) \le \overline{P}(\tau, x, y)$ in $Q_{\tau_0,r}$ and the strict monotonicity of the function U in the first component yields the result. □

5.5. Proof of Theorem 2.12

Proof. Remark first that if there exists a solution to the PDE with some fixed initial condition, it has to be unique by the smoothness and strict monotonicity of U (and the boundedness of I proven in Proposition 2.10) since otherwise the solution of the pricing PDE (Theorem 2.9) would not be unique. By the same reasoning we get also $I \in C^{1+\beta/2,2+\beta}(Q_{T,r}) \cap C(L_T)$. Moreover, the eikonal equation (2.23) has a unique viscosity solution as proved in Berestycki et al. (2004, section 3.2). To prove the theorem we will hence show that the solution of the eikonal equation is the only possible initial condition, i.e., that any solution of the PDE (2.22) has the eikonal equation as its small time limit. More precisely we will show that there exist parametrized families of (time-independent) local super- and subsolutions of (2.22) which converge locally uniformly to the eikonal equation. This is done in a similar way as in Berestycki et al. (2004, section 3.4), using an adapted vanishing viscosity method. However, in our setting the bounds on the volatility

σ enable us to simplify the proof and circumvent some obscurities in the local volatility argument in Berestycki et al. (2004).

Define $\overline{I}^{\varepsilon,\delta}(x, y)$ for $\varepsilon, \delta > 0$ as the solution of

$$\begin{cases} -\delta = -\left(\overline{I}^{\varepsilon,\delta}\right)^2 + \left(\overline{I}^{\varepsilon,\delta}\right)^2 \mathcal{G}\frac{x}{\overline{I}^{\varepsilon,\delta}} + \varepsilon\Delta\left(\ln\left(\overline{I}^{\varepsilon,\delta}\right)\right); \\ \overline{I}^{\varepsilon,\delta}|_{\partial B(m,r)} = \overline{\sigma}, \end{cases} \tag{5.11}$$

where $B(m, r)$ is an arbitrary disk. We will show that for $r, \delta, \varepsilon > 0$ there exists a solution to this equation and for fixed r and δ there exist $\varepsilon_0 > 0$, $\tau_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ this is a supersolution of (2.22) in $Q_{\tau_0,r}$. Moreover, we show that $\overline{I}^{\varepsilon,\delta} \to I^0$ locally uniformly as we send first ε and then δ to zero.

First Step: Making the change of variables $w := \ln(\overline{I}^{\varepsilon,\delta})$ one gets

$$\begin{aligned} \left(\overline{I}^{\varepsilon,\delta}\right)^2 \mathcal{G}\frac{x}{\overline{I}^{\varepsilon,\delta}} &= \left(\overline{I}^{\varepsilon,\delta}\right)^2 \left(\sigma^2(y)\left(\frac{x}{\overline{I}^{\varepsilon,\delta}}\right)_x^2 + 2\rho\sigma(y)a(y)\left(\frac{x}{\overline{I}^{\varepsilon,\delta}}\right)_x\left(\frac{x}{\overline{I}^{\varepsilon,\delta}}\right)_y + a^2(y)\left(\frac{x}{\overline{I}^{\varepsilon,\delta}}\right)_y^2\right) \\ &= \sigma^2(y)(1 - xw_x)^2 - 2\rho\sigma(y)a(y)x(1 - xw_x)w_y + a^2(y)x^2w_y^2 =: \tilde{\mathcal{G}}w \end{aligned}$$

and the equation (5.11) becomes

$$\begin{cases} -\delta = -e^{2w} + \tilde{\mathcal{G}}w + \varepsilon\Delta w; \\ w|_{\partial B(m,r)} = \ln\overline{\sigma}. \end{cases}$$

which admits a solution $w \in C^{2+\beta}(B(m, r))$ (Ladyzhenskaya and Ural'tseva 1968, Theorem 4.8.3) which is unique for sufficiently small r (Ladyzhenskaya and Ural'tseva 1968, theorem 4.2.1).

Second Step: By the Hölder property of the derivatives of w resp. $\overline{I}^{\varepsilon,\delta}$ implying the boundedness of the functions on the cylinder (as well as the boundedness of u_y and $\tilde{u}_y$ and the differentiability of $\hat{g}$) we can conclude that there exist constants c_1-c_4 solely depending on r such that

$$\begin{aligned} -\frac{1}{4}\tau^2\left(\overline{I}^{\varepsilon,\delta}\right)^2\mathcal{G}\overline{I}^{\varepsilon,\delta} &\le c_1(r)\tau^2 \\ \tau\overline{I}^{\varepsilon,\delta}\mathcal{M}\overline{I}^{\varepsilon,\delta} &\le c_2(r)\tau \\ \tau\left(q(y) - 2\frac{\hat{g}(-\lambda(y), \rho' Ka(y)\tilde{u}_y) - \hat{g}(-\lambda(y), \rho' Ka(y)u_y)}{K(\tilde{u}_y - u_y)}\right)\overline{I}^{\varepsilon,\delta}\overline{I}_y^{\varepsilon,\delta} &\le c_3(r)\tau \\ -\varepsilon\Delta\left(\ln\left(\overline{I}^{\varepsilon,\delta}\right)\right) &\le \varepsilon c_4(r) \end{aligned}$$

in $B(m, r)$. We can conclude that

$$\begin{aligned} \left((\tau\left(\overline{I}^{\varepsilon,\delta}\right)^2\right)_\tau &= \left(\overline{I}^{\varepsilon,\delta}\right)^2 = \delta + I\left(\overline{I}^{\varepsilon,\delta}\right)^2\mathcal{G}\frac{x}{\overline{I}^{\varepsilon,\delta}} + \varepsilon\Delta\left(\ln\left(\overline{I}^{\varepsilon,\delta}\right)\right) \\ &\ge \tau\overline{I}^{\varepsilon,\delta}\mathcal{M}\overline{I}^{\varepsilon,\delta} + \left(\overline{I}^{\varepsilon,\delta}\right)^2\mathcal{G}\frac{x}{\overline{I}^{\varepsilon,\delta}} - \frac{1}{4}\tau^2\left(\overline{I}^{\varepsilon,\delta}\right)^2\mathcal{G}\overline{I}^{\varepsilon,\delta} + \tau q(y)\overline{I}^{\varepsilon,\delta}\overline{I}_y^{\varepsilon,\delta} \\ &\quad - 2\tau I^{\varepsilon,\delta}I_y^{\varepsilon,\delta}\frac{\hat{g}(-\lambda(y), \rho' Ka(y)\tilde{u}_y) - \hat{g}(-\lambda(y), \rho' Ka(y)u_y)}{K(\tilde{u}_y - u_y)} \\ &\quad + \delta - (c_1(r)\tau^2 + c_2(r)\tau + c_3(r)\tau + \varepsilon c_4(r)), \end{aligned}$$

thus for given $\delta > 0$ and $r > 0$ we can find indeed positive bounds on ε_0, τ_0 such that $\overline{I}^{\varepsilon,\delta}$ is a supersolution of (2.22) for $0 < \varepsilon \le \varepsilon_0$ in $Q_{\tau_0,r}$. In the same way one proves that we get for $\underline{I}^{\varepsilon,\delta}$ of

$$\begin{cases} \delta = -(\underline{I}^{\varepsilon,\delta})^2 + \left(\underline{I}^{\varepsilon,\delta}\right)^2 \mathcal{G}\dfrac{x}{\underline{I}^{\varepsilon,\delta}} + \varepsilon\Delta\left(\ln\left(\underline{I}^{\varepsilon,\delta}\right)\right); \\ \underline{I}^{\varepsilon,\delta}|_{\partial B(m,r)} = \underline{\sigma} \end{cases}$$

subsolutions.

Third Step: Having now super- and subsolutions, we can (by sake of Proposition 2.10) invoke now the comparison principle Lemma 5.4 to conclude that

$$\underline{I}^{\varepsilon,\delta}(x, y) \leq I(\tau, x, y) \leq \overline{I}^{\varepsilon,\delta}(x, y) \quad \text{in} \quad Q_{\tau_0,r}$$

for all $0 < \varepsilon < \varepsilon_0$, ε_0 and τ_0 chosen as above. Thus we have

$$\underline{I}^{\varepsilon,\delta}(x, y) \leq \liminf_{\tau\to 0} I(\tau, x, y) \leq \limsup_{\tau\to 0} I(\tau, x, y) \leq \overline{I}^{\varepsilon,\delta}(x, y).$$

Next we want to send ε to zero. Therefore we note that the families of solutions $\overline{I}^{\varepsilon,\delta}$, $\underline{I}^{\varepsilon,\delta}$ are uniformly bounded (by the constants $\sqrt{\delta} \vee \overline{\sigma}$ and $\sqrt{\delta} \wedge \underline{\sigma}$ as a consequence of the comparison principle [Gilbarg and Trudinger 2001, theorem 10.7.(i)] applied to the equations in the log-variables) and equicontinuous in ε (since Hölder-continuous with the same Hölder constants). Thus by the Arzelà–Ascoli theorem $\overline{I}^{\varepsilon,\delta}$ converges along a subsequence uniformly on compacts to some limit function $\overline{I}^{\delta} \in C^{0+\beta}(Q_{\tau_0,r})$. This function is a viscosity solution of the PDE

$$\begin{cases} -\delta = -\left(\overline{I}^{\delta}\right)^2 + \left(\overline{I}^{\delta}\right)^2 \mathcal{G}\dfrac{x}{\overline{I}^{\delta}}; \\ \overline{I}^{\delta}|_{\partial B(m,r)} = \overline{\sigma}. \end{cases}$$

An analogous result holds true for the subsolutions. Now sending $\delta \to 0$, this gives by the same argument a solution of the PDE

$$\mathcal{G}\frac{x}{\overline{I}^{0}} = 1$$

which satisfies $I(0, y) = 0$. Thus for $\tau \to 0$, $I(\tau, x, y)$ converges locally uniformly to I^0 which is nothing else than the (by Berestycki et al. 2004, section 3.2 unique) viscosity solution of the eikonal equation (2.23) with $\psi = x/I^0$. □

REFERENCES

ARTZNER, P., F. DELBAEN, J. M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 9, 203–228.

BARRIEU, P., and N. EL KAROUI (2009): Pricing, Hedging, and Designing Derivatives with Risk Measures, in *Indifference Pricing. Theory and Applications*, R. Carmona, ed., Princeton, NJ: Princeton University Press, pp. 75–146.

BENTH, F., and K. KARLSEN (2005): A PDE Representation of the Density of the Minimal Entropy Martingale Measure in Stochastic Volatility Markets, *Stoch. Stoch. Rep.* 77(2), 109–137.

BERESTYCKI, H., J. BUSCA, and I. FLORENT (2004): Computing the Implied Volatility in Stochastic Volatility Models, *Comm. Pure Appl. Math.* 57(10), 1352–1373.

BRIAND, P., and Y. HU (2008): Quadratic BSDEs with Convex Generators and Unbounded Terminal Conditions, *Probab. Theory Related Fields* 141(3-4), 543–567.

COQUET, F., Y. HU, J. MÉMIN, and S. PENG (2002): Filtration-Consistent Nonlinear Expectations and Related g-Expectations, *Probab. Theory Relat. Fields* 123(1), 1–27.

COX, A. M. G., D. HOBSON, and J. OBŁÓJ (2011): Utility Theory Front to Back—Inferring Utility from Agents' Choices, Preprint.

DELBAEN, F., and W. SCHACHERMAYER (2006): *The Mathematics of Arbitrage*, Springer Finance. Berlin: Springer-Verlag.

FENG, J., M. FORDE, and J.-P. FOUQUE (2010): Short Maturity Asymptotics for a Fast Mean-Reverting Heston Stochastic Volatility Model, *SIAM J. Finan. Math.* 1, 126–141.

FORDE, M., and A. JACQUIER (2009): Small-Time Asymptotics for Implied Volatility under the Heston Model, *Int. J. Theor. Appl. Finance* 12(6), 861–876.

FORDE, M., A. JACQUIER, and R. LEE (2012): Small-Time Asymptotics for Implied Volatility under the Heston Model: Part 2, *SIAM J. Finan. Math.* 3(1), 690–708.

FÖLLMER, H., and A. SCHIED (2002): Convex Measures of Risk and Trading Constraints, *Finance Stoch.* 6(4), 429–447.

FÖLLMER, H., and A. SCHIED (2011): *Stochastic Finance: An Introduction in Discrete Time*, 3rd ed., Berlin: Walter de Gruyter.

FRIEDMAN, A. (1964): *Partial Differential Equations of Parabolic Type*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

GIANIN, E. R. (2006): Risk Measures via g-Expectations, *Insur. Math. Econ.* 39(1), 19–34.

GILBARG, D., and N. S. TRUDINGER (2001): *Elliptic Partial Differential Equations of Second Order*, Classics in Mathematics. Berlin: Springer-Verlag. Reprint of the 1998 edition.

HIRIART-URRUTY, J.-B., and C. LEMARÉCHAL (2001a): *Convex Analysis and Minimization Algorithms II*, Grundlehren der mathematischen Wissenschaften. Berlin: Springer-Verlag.

HIRIART-URRUTY, J.-B., and C. LEMARÉCHAL (2001b): *Fundamentals of Convex Analysis*, Grundlehren Text Editions. Berlin: Springer-Verlag. Abridged version of *Convex Analysis and Minimization Algorithms I and II*.

KLÖPPEL, S., and M. SCHWEIZER (2007): Dynamic Utility Indifference Valuation via Convex Risk Measures, NCRR FINRISK Working Paper No.209. ETH Zürich, 2005. Shorter version published in *Math. Finance* 17(4), 599–627.

KOBYLANSKI, M. (2000): Backward Stochastic Differential Equations and Partial Differential Equations with Quadratic Growth, *Ann. Probab.* 28(2), 558–602.

LADYZHENSKAYA, O. A., V. A. SOLONNIKOV, and N. N. URAL'TSEVA (1967): *Linear and Quasilinear Equations of Parabolic Type*, Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23. Providence, RI: American Mathematical Society.

LADYZHENSKAYA, O. A., and N. N. URAL'TSEVA (1968): *Linear and Quasilinear Elliptic Equations*, Translated from the Russian by Scripta Technica, Inc. Translation editor: L. Ehrenpreis. New York: Academic Press.

MUSIELA, M., and T. ZARIPHOPOULOU (2009): The Spot and Forward Dynamic Utilities and the Associated Pricing Systems: Case Study of the Binomial Model, in *Indifference Pricing. Theory and Applications*, R. Carmona, ed., Princeton, NJ: Princeton University Press, pp. 3–43.

NUTZ, M. (2007): *Quadratic PDE and Backward SDE*, Diploma thesis, ETH Zurich.

PENG, S. (2004): Nonlinear Expectations, Nonlinear Evaluations and Risk Measures, *Lect. Notes Math.* 1856, 165–253.

SIRCAR, R., and T. ZARIPHOPOULOU (2005): Bounds & Asymptotic Approximations for Utility Prices When Volatility Is Random, *SIAM J. Contl. Optim.* 43(4), 1328–1353.

STADJE, M. (2010): Extending Dynamic Convex Risk Measures from Discrete Time to Continuous Time: A Convergence Approach, *Insur. Math. Econ.* 47(3), 391–404.

TOUSSAINT, A. (2007): *Hedging and Pricing with L^2 Convex Risk Measures in Incomplete Markets*, PhD thesis, Princeton University.

MATHEMATICAL FINANCE
An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 25, No. 2 (April 2015), 426–456

CORRELATION UNDER STRESS IN NORMAL VARIANCE MIXTURE MODELS

MICHAEL KALKBRENER

Deutsche Bank AG

NATALIE PACKHAM

Frankfurt School of Finance & Management

We investigate correlations of asset returns in stress scenarios where a common risk factor is truncated. Our analysis is performed in the class of normal variance mixture (NVM) models, which encompasses many distributions commonly used in financial modeling. For the special cases of jointly normally and t-distributed asset returns we derive closed formulas for the correlation under stress. For the NVM distribution, we calculate the asymptotic limit of the correlation under stress, which depends on whether the variables are in the maximum domain of attraction of the Fréchet or Gumbel distribution. It turns out that correlations in heavy-tailed NVM models are less sensitive to stress than in medium- or light-tailed models. Our analysis sheds light on the suitability of this model class to serve as a quantitative framework for stress testing, and as such provides valuable information for risk and capital management in financial institutions, where NVM models are frequently used for assessing capital adequacy. We also demonstrate how our results can be applied for more prudent stress testing.

KEY WORDS: stress testing, risk management, correlation, normal variance mixture distribution, multivariate normal distribution, multivariate t-distribution.

1. INTRODUCTION

In risk management, stress testing encompasses various techniques used by financial firms to test bank capital adequacy in a strongly adverse market environment. Stress tests on bank portfolios are mandatory under the Basel II regulatory rules, and have gained importance as an integral part of risk management and banking supervision in the aftermath of the subprime crisis, see, e.g., Turner (2009), Larosière et al. (2009), Brunnermeier et al. (2009), BIS (2009). For example, bank supervisors conducted mandatory stress tests in the previous years to assess the adequacy of capital buffers of the largest US and European banks.

A standard technique of conducting stress tests is to calculate portfolio risk measures, such as expected loss, value-at-risk (VaR), economic capital, or regulatory capital under

The authors wish to thank Steffi Höse, Claudia Klüppelberg, Ludger Overbeck, Ursula Walther, and two anonymous referees for comments and discussions.

The views expressed in this paper are those of the authors and do not necessarily reflect the position of Deutsche Bank AG.

Manuscript received October 2010; final revision received October 2012.

Address correspondence to Natalie Packham, Department of Finance, Frankfurt School of Finance & Management, Sonnemannstr 9-11, 60314 Frankfurt am Main, Germany; e-mail: n.packham@fs.de.

DOI: 10.1111/mafi.12029

the assumption of adverse market conditions. Typically, stress scenarios are translated into constraints on risk factors, which are economic or market variables that affect portfolio losses. In our setup, the constraints are formalized by truncating the risk factor variables, that is, by conditioning on the range of values that a risk factor may attain. For example, a stress scenario for an equity portfolio may be defined as a decrease of an equity index (considered as a risk factor) by more than 10% within a time period of one year.

The truncation of risk factors is a commonly used stress-testing technique for the credit risk management and capital management of financial institutions, see, e.g., Bonti et al. (2006), Kalkbrener and Overbeck (2008), Duellmann and Erdelmeier (2009). In these papers, stress scenarios are implemented in structural credit portfolio models, which link the default of a firm to the relationship between its assets and liabilities. The dependence between assets is specified by systematic factors, which usually represent geographic regions or industries. Each stress scenario specifies a constraint on one or more systematic factors and thus quantifies the impact of a downturn in an economy or industry on the credit portfolio of a bank.[1]

The outcome of a stress test depends strongly on the correlations among portfolio components in the stress scenario. In general, the *conditional correlation* between asset returns—conditional on the truncated factor—is different from the unconditional correlation. However, the conditional correlations are the correlations that implicitly enter the calculation of risk measures under the stress scenario, because the risk measure is now calculated under a new probability measure obtained by conditioning. An analysis of conditional correlation may therefore contribute to understanding the behavior of portfolio models under stress and to refine the stress-testing methodology.

In this paper we analyze correlation under stress for a quite general class of distributions that encompasses many models commonly encountered in the financial industry. We provide closed formulas for the conditional correlation in the cases of jointly normally distributed asset returns and jointly t-distributed asset returns, which are standard distributions in portfolio modeling.[2] More generally, we consider normal variance mixture (NVM) distributions, a family of distributions with diverse tail behavior, ranging from light- to heavy-tailed. Essentially, a random vector follows an NVM distribution if it has a representation as a jointly normally distributed random vector multiplied with a strictly positive and independent "mixing" random variable.[3] NVM distributions belong to the class of elliptical distributions, whose key parameter for describing the dependence is the correlation matrix. These properties make NVM distributions attractive for high-dimensional applications, such as modeling the asset returns of a bank portfolio. For NVM models, we calculate the asymptotic limit of the conditional correlation. Here, we employ results and techniques from EVT.

It turns out that the limit of the conditional correlation depends on whether the variables in the NVM model are in the maximum domain of attraction (MDA) of a

[1]For further stress test methods, see also Kupiec (1998), Berkowitz (2000), and Breuer et al. (2009); for applications of extreme value theory (EVT) to stress testing, see, e.g., Longin (2000) and Schachter (2001).

[2]For example, in structural credit portfolio models, such as CreditMetrics^{TM}, Gupton, Finger, and Bhatia (1997), and Moody's KMV Portfolio Manager^{TM}, Crosbie and Bohn (2002), the multivariate normal distribution is still the de-facto standard for modeling log-returns of asset values.

[3]For example, the increments of certain time-changed Brownian motion models, such as the variance gamma or the inverse Gaussian processes and certain stochastic volatility models follow an NVM distribution, see, e.g., Madan and Seneta (1990), Barndorff-Nielsen (1998), Geman, Madan, and Yor (2001).

Fréchet distribution or the Gumbel distribution.[4] More precisely, consider asset returns A_i and A_j with (unconditional) correlation ρ_{ij} and a risk factor V, whose correlation with each of the asset returns is denoted by ρ_i and ρ_j, respectively. We show that the asymptotic limit of the conditional correlation of A_i and A_j equals

$$\frac{\rho_i\,\rho_j + (\rho_{ij} - \rho_i\,\rho_j)\,(\alpha - 1)}{\sqrt{\left(\rho_i^2 + \left(1 - \rho_i^2\right)(\alpha - 1)\right)\left(\rho_j^2 + \left(1 - \rho_j^2\right)(\alpha - 1)\right)}}, \quad \alpha > 2, \tag{1.1}$$

if the risk factor is stressed, i.e., if V is truncated at a threshold C, and C converges to $-\infty$. The parameter α specifies the tail index of the asset returns and the risk factor in the Fréchet case. The case when the variables are in the MDA of the Gumbel distribution corresponds to the limit as $\alpha \to \infty$, in which case the asymptotic limit of the conditional correlation between A_i and A_j is

$$\frac{\rho_{ij} - \rho_i\,\rho_j}{\sqrt{\left(1 - \rho_i^2\right)\left(1 - \rho_j^2\right)}}. \tag{1.2}$$

The limit formula (1.1) of the conditional asset correlation has a number of interesting consequences. First of all, it shows that for small α the correlations in the NVM model are insensitive to stress, i.e., the conditional correlations do not differ significantly from the unconditional correlations in heavy-tailed NVM models. In the Gumbel case, on the other hand, the impact of stress on the conditional correlation is typically much stronger than for heavy-tailed models, in particular when comparing to the heavy-tailed case with α sufficiently close to 2. The asymptotic conditional correlation may be either greater or smaller than the unconditional asset correlation depending largely on the correlations between the risk factor and the respective asset returns. In particular, when the assets in question are sufficiently correlated with the risk factor, the conditional correlation is typically smaller than the unconditional correlation. Loosely speaking, in many cases systematic risk is reduced by conditioning on the risk factor, whereas unsystematic risk remains (a detailed example is given in Section 5).

The implications of our results are threefold: first, they shed light on the impact of stress tests on the correlation of portfolio constituents and associated portfolio diversification within a model family that includes many commonly used light- and heavy-tailed distributions. Second, depending on how the conditional correlation obtained from the model reconciles with empirical observation, one may assess whether the model class is an appropriate model for the dependence structure. Furthermore, one may incorporate correlation as an additional risk factor and apply appropriate changes to the correlation itself when applying a stress scenario. We discuss these applications in our examples.

In a wider sense, our results may be of interest in the context of systemic risk measures, such as CoVaR and ΔCoVaR, suggested by Adrian and Brunnermeier (2011), and marginal expected shortfall, suggested by Acharya et al. (2010). Essentially, these measures gauge risk conditioned on certain stress events, such as a bank or the system being in distress. The (model) effect of such stress events on the correlation between institutions may give additional useful information when using these measures.

The model behavior of conditional correlation has also been studied for example by Boyer, Gibson, and Loretan (1999), where formulas for conditional asset correlations

[4]Essentially, the MDA of the Fréchet distribution consists of the heavy-tailed distributions, such as the t-distribution, while the MDA of the Gumbel distribution contains mostly light- to medium-tailed distributions, for example, the normal and the log-normal distributions.

are determined for bivariate normally distributed asset returns, when one of two asset returns is stressed. Malevergne and Sornette (2006) consider in addition bivariate t-distributed asset returns. These are special cases of our analysis of the stressed NVM model, in which the stressed risk factor coincides with one of the assets. In a stress test setup similar to ours, that is, by truncating a common risk factor, Bae and Iscoe (2010) consider a jump-diffusion model to achieve an increasing conditional correlation.

The remainder of the paper is structured as follows: In Section 2 we introduce the model setup involving asset returns and a single risk factor. We derive a general formula for conditional asset correlations in NVM models, which is based on some conditional expectations, in Section 3. In addition, we provide explicit formulas for the cases when the risk factor and the asset returns are normally distributed and t-distributed. In Section 4, we investigate the asymptotic limit of conditional correlation as the truncation of the risk factor tends to $-\infty$. More precisely, we derive formula (1.1) for the case when the NVM distribution is in the MDA of the Fréchet distribution (under some additional moderate assumption) and formula (1.2) when it is in the MDA of the Gumbel distribution. In Section 5, we analyze the relationship between correlations under stress and VaR in structural credit portfolios. We also provide examples where the correlation itself is considered to be a risk factor and subjected to change under stress. An empirical example involving DAX data is given. Finally, we conclude and discuss the results in Section 6. Appendix A provides a brief overview of the concepts from EVT that are applied in this paper.

2. DEFINITIONS AND PROBLEM

Let $V, A_1, \ldots, A_k$ be random variables on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. In our setting, V will be interpreted as a risk factor, and $A_1, \ldots, A_k$ may be interpreted as additional risk factors or as variables that specify asset returns or the credit quality of k assets.

We assume that $V, A_1, \ldots, A_k$ follow a multivariate normal variance mixture (NVM) distribution, i.e., there exist normally distributed variables $X, Y_1, \ldots, Y_k$ and a $\mathbf{P}$–a.s. strictly positive random variable W, independent of $X, Y_1, \ldots, Y_k$, such that

$$V := \sqrt{W}X, \quad A_i := \sqrt{W}Y_i, \quad i = 1, \ldots, k. \tag{2.1}$$

We assume in addition that W is integrable and that X and $Y_1, \ldots, Y_k$ are standardized.[5] For $i, j \in \{1, \ldots, k\}$ the correlations of X, Y_i, Y_j are denoted by

$$\rho_i := \mathrm{Corr}(X, Y_i), \quad \rho_{ij} := \mathrm{Corr}(Y_i, Y_j). \tag{2.2}$$

It is straightforward to show that $\mathrm{Corr}(V, A_i) = \rho_i$ and $\mathrm{Corr}(A_i, A_j) = \rho_{ij}$. We shall also sometimes use the following orthogonalization of Y_i on X:

$$Y_i = \rho_i X + \sqrt{1 - \rho_i^2} Z_i, \tag{2.3}$$

where Z_i is a standard normally distributed variable independent of X and W.

The NVM distribution encompasses a number of distributions that are commonly used to model asset returns. Some examples of NVM distributions are the multivariate normal distribution, in which case the mixing variable W is a constant, or the multivariate t-distributions, when W follows an inverse gamma distribution. When W follows a

[5] Working with standardized variables is not a restriction, as long as we are interested only in correlations.

so-called generalized inverse Gaussian distribution, then the NVM distribution is a symmetric generalized hyperbolic distribution. More generally, NVM distributions belong to the family of elliptic distributions, for which correlation appropriately describes the dependence among the variables. For a general review of NVM distributions we refer to McNeil, Frey, and Embrechts (2005, section 2.3) and Bingham and Kiesel (2002).

Let us consider the situation when V is truncated by $C \in \mathbb{R}$, that is, $V \leq C$ and write

$$\mathbf{P}^C(A) = \mathbf{P}(A \mid V \leq C), \quad A \in \mathcal{A}, \tag{2.4}$$

for the corresponding conditional distribution. In this setting, the random variable V is interpreted as a stressed risk factor with C the level of stress applied to V.

Let us rephrase the objectives of the paper: we wish to analyze the impact of the stress on the correlation of the A_i, $i = 1, \ldots, k$. In the special cases of a multivariate normal model ($W = 1$) and a jointly t-distributed model (W inverse gamma distributed), we derive formulas for conditional asset correlations $\mathrm{Corr}^C(A_i, A_j)$, where Corr^C is the correlation coefficient under the measure $\mathbf{P}^C$ (Section 3). For the general NVM model, we derive the limit of the conditional asset correlations as $C \to -\infty$ (Section 4).

3. CORRELATIONS UNDER STRESS

In this section we derive a general formula for the conditional correlation $\mathrm{Corr}^C(A_i, A_j)$. We further derive explicit formulas for the case when the variables are normally distributed (Section 3.1) and when the variables are t-distributed (Section 3.2).

PROPOSITION 3.1. *In the NVM model, and additionally writing* $\mathbb{E}^C$ *and* Var^C *for the expectation and variance under* $\mathbf{P}^C$, *respectively, the conditional correlation of* A_i, A_j *is given by*

$$\mathrm{Corr}^C(A_i, A_j) = \frac{\rho_i \, \rho_j \, \dfrac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} + \rho_{ij} - \rho_i \, \rho_j}{\sqrt{\left(\rho_i^2 \, \dfrac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} + \left(1 - \rho_i^2\right)\right)\left(\rho_j^2 \, \dfrac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} + \left(1 - \rho_j^2\right)\right)}}. \tag{3.1}$$

Observe that the conditional correlation depends only on the univariate quantity $\dfrac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)}$. In the remainder of the paper we shall thus concentrate on deriving explicit formulas for this quantity and for the limit of this quantity as $C \to -\infty$.

Proof. By the independence of Z_i and V, Z_i is a standard normal random variable under $\mathbf{P}^C$, so that, by equation (2.3),

$$\mathbb{E}^C(Z_i \, Z_j) = \mathbb{E}(Z_i \, Z_j) = \frac{\rho_{ij} - \rho_i \, \rho_j}{\sqrt{\left(1 - \rho_i^2\right)\left(1 - \rho_j^2\right)}}. \tag{3.2}$$

Also, by equation (2.3), we have $\mathbb{E}^C(A_i) = \rho_i \, \mathbb{E}^C(V)$, so that one obtains

$$\mathrm{Var}^C(A_i) = \rho_i^2 \, \mathrm{Var}^C(V) + \left(1 - \rho_i^2\right) \mathbb{E}^C(W),$$

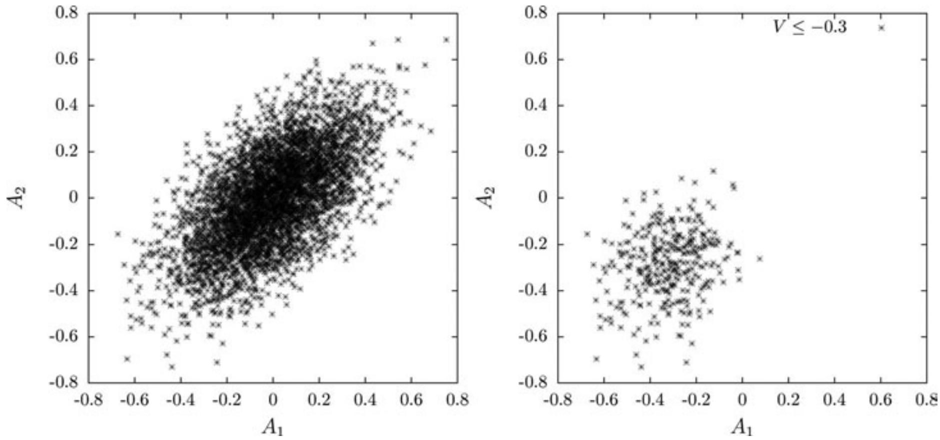


FIGURE 3.1. Left: Simulated asset returns of A_1 and A_2 that are jointly normally distributed with $\rho_{12} = 0.6$ and each with standard deviation 0.2. Right: Samples conditional on $V \leq -0.3$, where V is normally distributed with standard deviation 0.2, and $\rho_1 = 0.8$ and $\rho_2 = 0.7$.

$$\mathrm{Cov}^C(A_i, A_j) = \rho_i\,\rho_j\,\mathrm{Var}^C(V) + (\rho_{ij} - \rho_i\,\rho_j)\,\mathbb{E}^C(W).$$

The claim follows by dividing both the conditional covariance and the conditional variance by $\mathbb{E}^C(W)$. □

Let us illustrate the impact of truncating a risk factor on the conditional correlation: The left plot of Figure 3.1 shows a scatterplot of 5,000 simulated samples representing asset returns of two assets, A_1 and A_2. The asset returns are jointly normally distributed, each with a standard deviation of 20%, and with a correlation of $\rho_{12} = 0.6$. Let us assume that the assets are in addition correlated with a common risk factor V, which is also normally distributed with a standard deviation of 20%, and where $\rho_1 = 0.8$ and $\rho_2 = 0.7$. The right plot shows the samples conditional on $V \leq -0.3$. It appears that the conditional samples are less correlated than the unconditional samples; in fact, their correlation is approximately 0.25. We confirm this observation analytically below.

3.1. Special Case: Normal Distribution

Let us set $W = 1$, so that $V = X$ and $A_i = Y_i$ are normally distributed. Denote by ϕ the standard normal density function and by N the standard normal distribution function.

PROPOSITION 3.2. *Let $V, A_1, \ldots, A_k$ be standard normally distributed. Then,*

$$\mathrm{Corr}^C(A_i, A_j) = \frac{\rho_i\,\rho_j\,\mathrm{Var}^C(V) + \rho_{ij} - \rho_i\,\rho_j}{\sqrt{\left(\rho_i^2\,\mathrm{Var}^C(V) + 1 - \rho_i^2\right)\left(\rho_j^2\,\mathrm{Var}^C(V) + 1 - \rho_j^2\right)}}, \tag{3.3}$$

with

$$\mathrm{Var}^C(V) = 1 - \frac{C\,\phi(C)}{\mathrm{N}(C)} - \frac{(\phi(C))^2}{(\mathrm{N}(C))^2}. \tag{3.4}$$

TABLE 3.1
Example Parameters

Example	ρ_{12}	ρ_1	ρ_2	$\mathrm{Corr}(Z_1, Z_2)$
1	0.6	1	0.6	0
2	0.6	0.8	0.7	0.093
3	0.6	0.6	0.6	0.375
4	0.6	0.1	0.1	0.596
5	0.6	0.7	0.02	0.82

Proof. The first equation follows directly from Proposition 3.1, so it remains to prove equation (3.4). First, observe that $\mathbb{E}(V\mathbf{1}_{\{V\leq C\}}) = -\phi(C)$, which follows directly from $\phi(C) = \int_{-\infty}^{C} \phi'(x)\,\mathrm{d}x = \int_{-\infty}^{C} -x\,\phi(x)\,\mathrm{d}x$. Second, observe that $\mathbb{E}(V^2\,\mathbf{1}_{\{V\leq C\}}) = \mathrm{N}(C) - C\,\phi(C)$, which follows from $C\,\phi(C) = \int_{-\infty}^{C} (x\,\phi(x))'\,\mathrm{d}x = \mathrm{N}(C) - \int_{-\infty}^{C} x^2\,\phi(x)\,\mathrm{d}x$. The claim is now obtained via

$$\mathrm{Var}^C(V) = \mathbb{E}^C(V^2) - \left(\mathbb{E}^C(V)\right)^2 = \frac{\mathbb{E}(V^2\,\mathbf{1}_{\{V\leq C\}})}{\mathbf{P}(V\leq C)} - \frac{\mathbb{E}(V\mathbf{1}_{\{V\leq C\}})^2}{\mathbf{P}(V\leq C)^2}.$$ □

Let us illustrate conditional correlation for five examples. The data are given in Table 3.1. Here, ρ_{12} specifies the correlation between assets A_1 and A_2, and ρ_i is the correlation between the risk factor and asset A_i, $i = 1, 2$. Because all variables are standardized, the correlation $\mathrm{Corr}(Z_1, Z_2)$ of the orthogonalization of equation (2.3) is given by equation (3.2). In all examples, we have set $\rho_{12} = 0.6$. In example 1, the first asset coincides with the risk factor. In examples 2 and 3, both assets are quite strongly correlated with the risk factor. In example 4, both assets are only weakly correlated with the risk factor, and in example 5, A_1 is strongly correlated with the risk factor, while the correlation between A_2 and the risk factor is low. The conditional correlations as a function of the truncation level are shown in Figure 3.2. When the dependence of both assets on the risk factor is sufficiently high, the conditional correlation decreases as the truncation level decreases. For example, assume that V denotes the return on an equity index and let A_1 and A_2 be the returns on two index components. Clearly, a negative return smaller than C on the equity index return V translates into a strong negative return on two index constituents A_1 and A_2. However, conditional on $V \leq C$, the correlation between A_1 and A_2 is now driven mainly by the correlation between the specific components Z_1 and Z_2, as demonstrated in Figure 3.1. In fact, we show in Section 4.1 that the conditional correlation converges to $\mathrm{Corr}(Z_i, Z_j)$ as $C \to -\infty$. This also explains why in example 5 the conditional correlation increases with the stress level. For example, there could be a second risk factor driving the correlation of Z_1 and Z_2. This second risk factor becomes more dominant as the stress level increases.

3.2. Special Case: t-Distribution

The class of multivariate t-distributions is frequently used in the finance industry to model random variables with heavy tails. We derive an explicit formula for

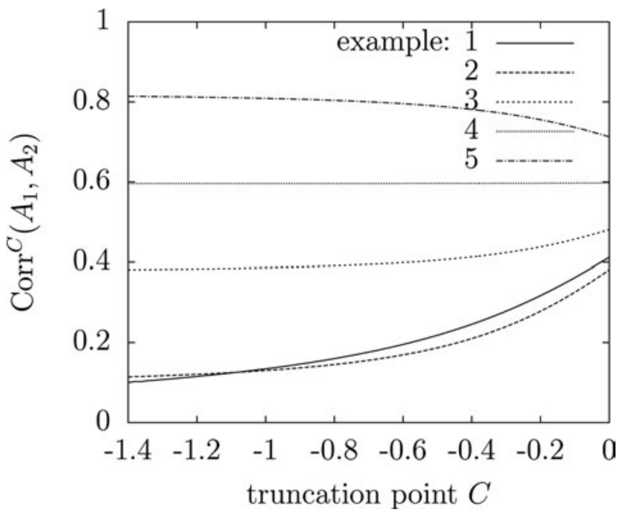


FIGURE 3.2. Conditional asset correlations when V, A_1, A_2 are normally distributed. The risk factor V has a standard deviation of 20%; for example, one may think of C as the truncation level of an equity index return with volatility 20%.

conditional correlations in t-distributed models. As outlined above, it is sufficient to calculate $\mathrm{Var}^C(V)/\mathbb{E}^C(W)$, which can be expressed in terms of beta functions.

PROPOSITION 3.3. *Let $V, A_1, \dots, A_k$ follow a multivariate t-distribution with parameter $\nu > 2$ denoting the degrees of freedom. Then, for $C < 0$,*

$$\frac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} = \frac{\mathbb{E}^C(V^2) - \mathbb{E}^C(V)^2}{\mathbb{E}^C(W)} = \frac{f(\nu, C)}{g(\nu, C)},$$

with

$$f(\nu, C) := B\left(\frac{\nu}{C^2+\nu}; \frac{\nu-2}{2}, \frac{3}{2}\right) - \frac{4\left(\frac{\nu}{C^2+\nu}\right)^{\nu-1}}{(\nu-1)^2 B\left(\frac{\nu}{C^2+\nu}; \frac{\nu}{2}, \frac{1}{2}\right)},$$

$$g(\nu, C) := \frac{B\left(\frac{1}{2}, \frac{\nu}{2}\right)}{\nu-2} - \frac{\left(B\left(\frac{\nu-2}{2}, \frac{1}{2}\right) - B\left(\frac{\nu}{\nu+C^2}; \frac{\nu-2}{2}, \frac{1}{2}\right)\right)}{\nu-1},$$

where $B(y; a, b) := \int_0^y t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$ is the incomplete beta function and $B(a, b) := B(1; a, b)$ is the beta function.

Proof.[6] For $n \in \mathbb{N}$, $n < \nu$, and for $C < 0$ we first show that the conditional moments of V are given by

[6]We are grateful to an anonymous referee whose suggestions significantly simplified the initial proof, which was based on hypergeometric functions.

$$\mathbb{E}^C(V^n) = \frac{(-1)^n\, \nu^{n/2}\, B\left(\dfrac{\nu}{C^2+\nu}; \dfrac{\nu-n}{2}, \dfrac{n+1}{2}\right)}{B\left(\dfrac{\nu}{C^2+\nu}; \dfrac{\nu}{2}, \dfrac{1}{2}\right)}. \tag{3.5}$$

Because V is t-distributed, we have

$$\mathbb{E}(V^n\, \mathbf{1}_{\{V\le C\}}) = \frac{\displaystyle\int_{-\infty}^{C} x^n \left(\frac{\nu}{x^2+\nu}\right)^{\frac{\nu+1}{2}} \mathrm{d}x}{\sqrt{\nu}\, B\left(\dfrac{\nu}{2}, \dfrac{1}{2}\right)}. \tag{3.6}$$

Integration by substitution with variable transformation $x = -\sqrt{(\frac{1}{y}-1)\nu}$ yields

$$\begin{aligned}\int_{-\infty}^{C} x^n \left(\frac{\nu}{x^2+\nu}\right)^{\frac{\nu+1}{2}} \mathrm{d}x &= \frac{1}{2}(-1)^n \nu^{\frac{n+1}{2}} \int_0^{\frac{\nu}{C^2+\nu}} (1-y)^{\frac{n-1}{2}}\, y^{\frac{\nu-n-2}{2}}\, \mathrm{d}y \\ &= \frac{1}{2}(-1)^n \nu^{\frac{n+1}{2}}\, B\left(\frac{\nu}{C^2+\nu}; \frac{\nu-n}{2}, \frac{n+1}{2}\right).\end{aligned} \tag{3.7}$$

By combining equations (3.6) and (3.7) we obtain

$$\mathbb{E}(V^n\, \mathbf{1}_{\{V\le C\}}) = \frac{(-1)^n \nu^{n/2}\, B\left(\dfrac{\nu}{C^2+\nu}; \dfrac{\nu-n}{2}, \dfrac{n+1}{2}\right)}{2B\left(\dfrac{\nu}{2}, \dfrac{1}{2}\right)}.$$

Dividing by $\mathbf{P}(V \le C) = B(\dfrac{\nu}{C^2+\nu}; \frac{\nu}{2}, \frac{1}{2})/(2B(\frac{\nu}{2}, \frac{1}{2}))$ yields equation (3.5).

A simple representation can be obtained for the first conditional moment $\mathbb{E}^C(V)$: it is an immediate consequence of the definition of incomplete beta functions that

$$B\left(\frac{\nu}{C^2+\nu}; \frac{\nu-1}{2}, 1\right) = \int_0^{\frac{\nu}{C^2+\nu}} t^{\frac{\nu-1}{2}-1}\, \mathrm{d}t = \frac{2\left(\dfrac{\nu}{C^2+\nu}\right)^{\frac{\nu-1}{2}}}{\nu-1},$$

and therefore

$$\mathbb{E}^C(V) = \frac{2\,(C^2+\nu)^{\frac{1-\nu}{2}}\, \nu^{\frac{\nu}{2}}}{(1-\nu)\, B\left(\dfrac{\nu}{C^2+\nu}; \dfrac{\nu}{2}, \dfrac{1}{2}\right)}. \tag{3.8}$$

Turning to the calculation of the conditional mean of W we now show that, for $C < 0$,

$$\begin{aligned}\mathbb{E}^C(W) = &\left(\frac{\nu B\left(\dfrac{\nu}{2}, \dfrac{1}{2}\right)}{\nu-2} - \frac{\nu\left(B\left(\dfrac{\nu-2}{2}, \dfrac{1}{2}\right) - B\left(\dfrac{\nu}{C^2+\nu}; \dfrac{\nu-2}{2}, \dfrac{1}{2}\right)\right)}{\nu-1}\right) \Bigg/ \\ &B\left(\frac{\nu}{C^2+\nu}; \frac{\nu}{2}, \frac{1}{2}\right).\end{aligned} \tag{3.9}$$

Because X and W are independent and ν/W follows a chi-square distribution with ν degrees of freedom, we have

$$
\begin{aligned}
\mathbb{E}(W\mathbf{1}_{\{V\le C\}}) &= \mathbb{E}\big(W\mathbf{1}_{\{X\le C/\sqrt{W}\}}\big)\\
&= \int_0^\infty \int_{-\infty}^{C\sqrt{\frac{y}{\nu}}} \left(\frac{\nu}{y}\right)\left(\frac{e^{-\dfrac{x^2}{2}}}{\sqrt{2\pi}}\right)\left(\frac{2^{-\nu/2}e^{-y/2}y^{\frac{\nu}{2}-1}}{\Gamma\left(\frac{\nu}{2}\right)}\right)\mathrm{d}x\,\mathrm{d}y\\
&= \frac{2^{(-\nu-1)/2}\nu}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \int_{-\infty}^{C\sqrt{\frac{y}{\nu}}} e^{\frac{-x^2-y}{2}}y^{\frac{\nu}{2}-2}\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}
$$

Integration by substitution with variable transformation $x=-\sqrt{z}$ yields

$$
\int_0^\infty \int_{-\infty}^{C\sqrt{\frac{y}{\nu}}} e^{\frac{-x^2-y}{2}}y^{\frac{\nu}{2}-2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^\infty \int_{\frac{C^2y}{\nu}}^\infty \frac{e^{\frac{-y-z}{2}}y^{\frac{\nu}{2}-2}}{2\sqrt{z}}\,\mathrm{d}z\,\mathrm{d}y.
$$

For the function

$$
F(t) := \int_0^\infty \int_{\frac{ty}{\nu}}^\infty \frac{e^{\frac{-y-z}{2}}y^{\frac{\nu}{2}-2}}{2\sqrt{z}}\,\mathrm{d}z\,\mathrm{d}y,
$$

we use the representation $F(t)=F(0)+\int_0^t F'(x)\,\mathrm{d}x$ and obtain

$$
\begin{aligned}
\int_0^\infty \int_{\frac{C^2y}{\nu}}^\infty \frac{e^{\frac{-y-z}{2}}y^{\frac{\nu}{2}-2}}{2\sqrt{z}}\,\mathrm{d}z\,\mathrm{d}y = F(C^2) &= 2^{\frac{\nu-3}{2}}\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}-1\right)\\
&\quad -\frac{2^{\frac{\nu-3}{2}}\Gamma\left(\dfrac{\nu-1}{2}\right)}{\nu}\int_0^{C^2}\frac{\left(\dfrac{x}{\nu}+1\right)^{\frac{1}{2}-\frac{\nu}{2}}}{\sqrt{\dfrac{x}{\nu}}}\,\mathrm{d}x.
\end{aligned}
$$

Integration by substitution with variable transformation $x=-(1-\frac{1}{z})\nu$ yields

$$
\int_0^{C^2}\frac{\left(\dfrac{x}{\nu}+1\right)^{\frac{1}{2}-\frac{\nu}{2}}}{\sqrt{\dfrac{x}{\nu}}}\,\mathrm{d}x = \int_{\frac{\nu}{C^2+\nu}}^1 \frac{z^{\frac{\nu-4}{2}}\nu}{\sqrt{1-z}}\,\mathrm{d}z = \nu\left(B\left(\frac{\nu-2}{2},\frac{1}{2}\right)-B\left(\frac{\nu}{C^2+\nu};\frac{\nu-2}{2},\frac{1}{2}\right)\right).
$$

By combining the equations above we obtain

$$
\begin{aligned}
E(W\mathbf{1}_{\{V\le C\}}) &= \frac{\nu}{4\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\\
&\cdot\left(\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}-1\right)-\Gamma\left(\frac{\nu-1}{2}\right)\left(B\left(\frac{\nu-2}{2},\frac{1}{2}\right)-B\left(\frac{\nu}{C^2+\nu};\frac{\nu-2}{2},\frac{1}{2}\right)\right)\right).
\end{aligned}
$$

The basic identities

$$
\Gamma(1/2)=\sqrt{\pi},\quad \Gamma(a+1)=\Gamma(a)a,\quad B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)
$$

can now be used to simplify the representation of $\mathbb{E}(W\mathbf{1}_{\{V\le C\}})$:

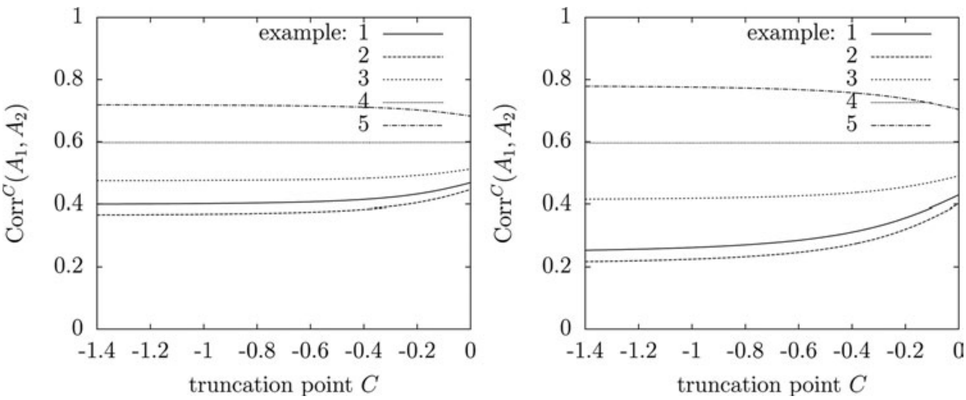


FIGURE 3.3. Conditional asset correlations when V, A_1, A_2 are t-distributed with parameter $\nu = 4$ (left) and $\nu = 10$ (right). The risk factor V has a standard deviation of 20%.

$$\mathbb{E}(W\mathbf{1}_{\{V\le C\}}) = \frac{\nu}{2\nu-4} - \frac{\nu\left(B\left(\frac{\nu-2}{2},\frac{1}{2}\right) - B\left(\frac{\nu}{\nu+C^2};\frac{\nu-2}{2},\frac{1}{2}\right)\right)}{2(\nu-1)B\left(\frac{\nu}{2},\frac{1}{2}\right)}.$$

Dividing by $\mathbf{P}(V \le C) = B(\frac{\nu}{C^2+\nu}; \frac{\nu}{2}, \frac{1}{2})/(2B(\frac{\nu}{2}, \frac{1}{2}))$ yields equation (3.9).

The representation $\mathrm{Var}^C(V)/\mathbb{E}^C(W) = f(\nu, C)/g(\nu, C)$ is now an immediate consequence of equations (3.5), (3.8), and (3.9). □

Examples of the conditional asset correlation for t-distributed models with parameter $\nu = 4$ and $\nu = 10$ are given in Figure 3.3. As before the examples refer to the data in Table 3.1. As in the normally distributed case, the conditional correlations increase or decrease depending on the correlations with the risk factor V. In the more heavy-tailed case ($\nu = 4$), the effect is much less pronounced than in the less heavy-tailed case ($\nu = 10$), which is in turn less pronounced than the normally distributed case. In other words, correlations in t-distributed models with rather heavy tails are quite insensitive to stress, whereas correlations in normally distributed models are much more affected by stress.

4. ASYMPTOTIC LIMIT OF CORRELATIONS UNDER STRESS

In this section we derive asymptotic limits for conditional correlations as the truncation level C tends to $-\infty$. Using Proposition 3.1, $\lim_{C\to-\infty}\mathrm{Corr}^C(A_i, A_j)$ is determined by

$$\lim_{C\to-\infty} \frac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)}, \tag{4.1}$$

if this limit exists. For normally distributed V the limit (4.1) can be computed using only properties of the normal distribution; see Section 4.1.

More generally, it turns out that essentially the limit exists whenever V is in the MDA of some EVT distribution.[7] A random variable X is in the MDA of the distribution H if H is nondegenerate and if a scaled and normalized distribution of the maximum M_n

[7]Basically all continuous distributions commonly used in statistics are in the MDA of some EVT distribution (see McNeil et al. 2005, section 7.1).

of n i.i.d. copies of X converges to H. The limit distribution H is one of the Fréchet, Gumbel, or Weibull EV distributions. We write $X \in \mathrm{MDA}(H)$ (or $F \in \mathrm{MDA}(H)$, where F is a distribution function) to mean that X (F) is in the MDA of H, cf. Appendix A.

The limit (4.1) depends on whether W is in the MDA of the Fréchet distribution, in which case V is in the Fréchet MDA, or whether V is in the MDA of the Gumbel distribution.[8] We treat the two cases separately in Sections 4.2 and 4.3, respectively.

The following result, required later, holds for all NVM-distributed variables.

LEMMA 4.1. *On $\mathbb{R} \setminus \{0\}$, V has a density, given by $f_V(x) = \mathbb{E}(\frac{1}{\sqrt{W}}\phi(\frac{x}{\sqrt{W}}))$.*

Proof. We have

$$F_V(x) = \mathbf{P}(\sqrt{W}X \le x) = \mathbb{E}\left(\mathbf{P}\left(X \le \frac{x}{\sqrt{W}}\,\middle|\, W\right)\right) = \mathbb{E}\left(\mathrm{N}\left(\frac{x}{\sqrt{W}}\right)\right).$$

For every $x \neq 0$, the family of derivatives of $\mathrm{N}(x/\sqrt{w})$, $w > 0$, is locally bounded (that is, there exist a neighborhood U of x and a constant M such that for all $u \in U$ and for all w the derivatives are bounded by M); this follows from the continuity of the derivatives in x and w and by observing that the limits tend to zero as $w \to 0$ and as $w \to \infty$. The claim now follows by the Fundamental Theorem of Calculus and by Dominated Convergence. □

4.1. Special Case: Normal Distribution

Let us first analyze the case of jointly normally distributed V and $A_1, \dots, A_k$ (i.e., $W = 1$). Although this is a special case of V in the MDA of the Gumbel distribution, we examine this special case separately as it requires only properties of the normal distribution, whereas the more general result relies on techniques from EVT.

PROPOSITION 4.2. *Let V be a standard normally distributed random variable. Then,*

$$\lim_{C \to -\infty} \mathrm{Var}^C(V) = 0.$$

By inserting this into the formula for asset correlations, equation (3.1), respectively, equation (3.3), one immediately obtains the limit of the conditional correlation:

COROLLARY 4.3. *Let $V, A_1, \dots, A_k$ be jointly normally distributed. Then, for $|\rho_i| < 1$ and $|\rho_j| < 1$,*

$$\lim_{C \to -\infty} \mathrm{Corr}^C(A_i, A_j) = \frac{\rho_{ij} - \rho_i \rho_j}{\sqrt{\left(1 - \rho_i^2\right)\left(1 - \rho_j^2\right)}} = \mathrm{Corr}(Z_i, Z_j). \tag{4.2}$$

If either $|\rho_i| = 1$ or $|\rho_j| = 1$ (but not both), then $\lim_{C \to -\infty} \mathrm{Corr}^C(A_i, A_j) = 0$. *If $\rho_i = \rho_j = \pm 1$, then* $\lim_{C \to -\infty} \mathrm{Corr}^C(A_i, A_j) = 1$. *If $\rho_i = -\rho_j = \pm 1$, then* $\lim_{C \to -\infty} \mathrm{Corr}^C(A_i, A_j) = -1$.

According to equation (4.2), the correlation under stress of A_i, A_j converges toward the correlation of the residuals Z_i and Z_j obtained by regressing Y_i and Y_j on X.

[8]The case that V is in the MDA of the Weibull distribution does not arise, as all distributions in this MDA have a finite right endpoint.

Proof of Proposition 4.2. The claim follows from equation (3.4), if we show that[9]

$$\lim_{C\to-\infty}\left[\frac{C\phi(C)}{\mathrm{N}(C)}+\frac{\phi(C)^2}{\mathrm{N}(C)^2}\right]=1. \tag{4.3}$$

Let $C < 0$. We shall use the identity $\mathrm{N}(C) = \phi(C)f(C)$, for $C < 0$, where f is the continued fraction

$$f(C)=\left[\cfrac{1}{-C+\cfrac{1}{-C+\cfrac{2}{-C+\cfrac{3}{-C+\cdots}}}}\right],$$

(see Abramowitz and Stegun 1972, equation 26.2.14). We obtain

$$\begin{aligned}
\frac{C\phi(C)}{\mathrm{N}(C)}+\frac{\phi(C)^2}{\mathrm{N}(C)^2} &= \frac{C}{f(C)}+\frac{1}{f(C)^2}\\
&= \frac{C}{\cfrac{1}{-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}}}+\left[\frac{1}{\cfrac{1}{-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}}}\right]^2\\
&= C\left[-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}\right]+\left[-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}\right]^2\\
&= \left[-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}\right]\cdot\left[C-C+\cfrac{1}{-C+\cfrac{2}{-C+\cdots}}\right]\\
&= \frac{-C}{-C+\cfrac{2}{-C+\cdots}}+\left[\frac{1}{-C+\cfrac{2}{-C+\cdots}}\right]^2\\
&= 1-\frac{\cfrac{2}{-C+\cdots}}{-C+\cfrac{2}{-C+\cdots}}+\left[\frac{1}{-C+\cfrac{2}{-C+\cdots}}\right]^2.
\end{aligned}$$

[9]It is known from Sampford (1953) that for $C < 0$, the expression in brackets is in (0, 1) and strictly decreasing.

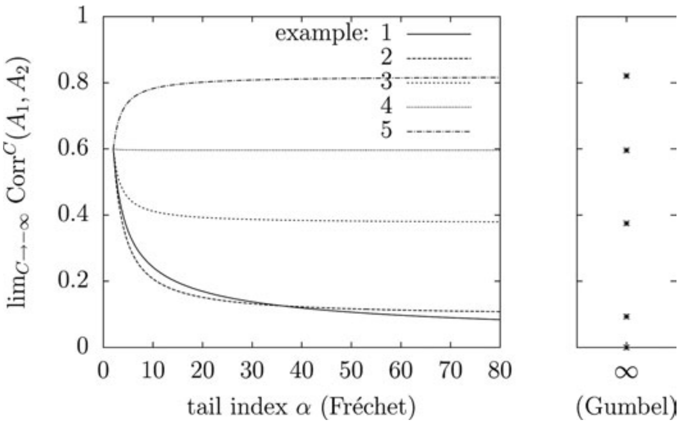


FIGURE 4.1. Asymptotic asset correlations as $C \to -\infty$.

Observing that $\frac{1}{-C+\frac{2}{-C+\cdots}} \le -\frac{1}{C}$, and $\frac{2}{-C+\cdots} \le -\frac{2}{C}$, for $C < 0$, equation (4.3) is obtained by taking the limit. □

Figure 4.1 shows the asymptotic asset correlation for the examples of Table 3.1, where the normally distributed case corresponds to the figure labeled "(Gumbel)."

4.2. W and V in the Fréchet MDA

Let us assume that $W \in \mathrm{MDA}(\Phi_{\alpha/2})$ (i.e., W is in the Fréchet MDA with tail index $\alpha/2$), with $\alpha > 2$. Lemma 4.4 below establishes that $V \in \mathrm{MDA}(\Phi_\alpha)$. Hence, the tail distribution function $\bar{F}_V(x) = 1 - F_V(x)$ is regularly varying (at ∞) with index $-\alpha$ and can be represented as $\bar{F}_V(x) = x^{-\alpha} L(x)$, for some slowly varying function L. Some members of the Fréchet MDA are the t-distribution, the Pareto distribution, and the inverse gamma distribution.

LEMMA 4.4. *If W is in* $\mathrm{MDA}(\Phi_{\alpha/2})$*, then $\sqrt{W}$ and V are in* $\mathrm{MDA}(\Phi_\alpha)$*. Furthermore,*

$$\lim_{x\to\infty} \frac{\bar{F}_V(x)}{\bar{F}_{\sqrt{W}}(x)} = \mathbb{E}\left(X^\alpha \mathbf{1}_{\{X\ge 0\}}\right).$$

Proof. By Theorem A.4 in the Appendix, $\bar{F}_W$ can be written as $\bar{F}_W(x) = x^{-\alpha/2} L(x)$, for $x > 0$ and some slowly varying function L. From $\bar{F}_{\sqrt{W}}(x) = \bar{F}_W(x^2) = x^{-\alpha} L(x^2)$, we obtain that $\sqrt{W} \in \mathrm{MDA}(\Phi_\alpha)$. The claim then follows directly from Breiman (1965, Proposition 3 and equation (3.1)) or McNeil et al. (2005, theorem 7.35). □

The main result of this section is the following proposition.

PROPOSITION 4.5. *Let* $W \in \mathrm{MDA}(\Phi_{\alpha/2})$, $\alpha > 2$. *Then,*

$$\lim_{C\to-\infty} \frac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} = \frac{1}{\alpha - 1}.$$

By inserting the previous result into equation (3.1) we obtain the asymptotic limit of $\mathrm{Corr}^C(A_i, A_j)$ for heavy-tailed factors:

COROLLARY 4.6. *Let* $W \in \mathrm{MDA}(\Phi_{\alpha/2})$, *with* $\alpha > 2$. *Then*

$$\lim_{C\to-\infty} \mathrm{Corr}^C(A_i, A_j) = \frac{\rho_i\,\rho_j + (\rho_{ij} - \rho_i\,\rho_j)(\alpha-1)}{\sqrt{\left(\rho_i^2 + \left(1-\rho_i^2\right)(\alpha-1)\right)\left(\rho_j^2 + \left(1-\rho_j^2\right)(\alpha-1)\right)}}.$$

If V is t-distributed with parameter ν (i.e., W follows an inverse gamma distribution with parameters $\nu/2$, $\nu/2$), then the formula above holds with $\alpha = \nu$. Observe that as α approaches the value 2, the limit correlation tends to ρ_{ij}, which is just the unconditional correlation. In other words, when the tail is very heavy, then correlations are unaffected by stress testing.

Figure 4.1 shows the asymptotic correlation (for the heavy-tailed case as derived in Corollary 4.6) as a function of the tail index α for the five examples of Table 3.1. This illustrates once again that the sensitivity of conditional correlation to stress increases with the tail index α. In other words, the greater the α, the greater the deviation from the unconditional correlation.

For the proof of Proposition 4.5 we need the following auxiliary result first.

LEMMA 4.7. *Let X be a nonnegative random variable with distribution function F and with* $\bar{F}(x) > 0$. *If* $\mathbb{E}(X^k) < \infty$, $k \geq 0$, *then*

$$\mathbb{E}(X^k \mid X > x) = x^k + k \int_x^\infty \frac{\bar{F}(y)\,y^{k-1}}{\bar{F}(x)}\,\mathrm{d}y, \qquad x \geq 0.$$

Moreover, if $X \in \mathrm{MDA}(\Phi_\alpha)$, *then, for* $k < \alpha$, $\mathbb{E}(X^k) < \infty$ *and*

$$\lim_{x\to\infty} \frac{\mathbb{E}(X^k \mid X > x)}{x^k} = \frac{\alpha}{\alpha - k}.$$

Proof. We have $\mathbb{E}(X^k \mid X > x) = \mathbb{E}(X^k\,\mathbf{1}_{\{X>x\}})/\bar{F}(x)$. For the numerator we obtain that

$$\mathbb{E}(X^k\,\mathbf{1}_{\{X>x\}}) = x^k\,\bar{F}(x) + \int_x^\infty k\,t^{k-1}\,\bar{F}(t)\,\mathrm{d}t,$$

e.g., by Kallenberg (2001, lemma 3.4), and the first claim follows.

For the second claim, observe first that $\mathbb{E}(X^k) < \infty$, if $k < \alpha$, see (Embrechts, Klüppelberg, and Mikosch 1997, proposition A 3.8). Let L be the slowly varying function defined by $\bar{F}(x) = x^{-\alpha} L(x)$. Applying Karamata's Theorem (theorem A.5), yields

$$\lim_{x\to\infty} \frac{\mathbb{E}(X^k \mid X > x)}{x^k} = \lim_{x\to\infty} \frac{k \displaystyle\int_x^\infty y^{k-1-\alpha} L(y)\,\mathrm{d}y}{x^{k-\alpha}\,L(x)} + 1 = \frac{\alpha}{\alpha - k}.$$

□

Proof of Proposition 4.5. First, let us show that

$$\lim_{C\to-\infty} \frac{\mathrm{Var}^C(V)}{C^2} = \frac{\alpha}{(\alpha-2)(\alpha-1)^2}. \tag{4.4}$$

Note that, $V = \sqrt{W}X$ is symmetric about 0, that is, $F_V(x) = \bar{F}_V(-x)$ for all $x \in \mathbb{R}$. Hence,

$$\begin{aligned}\lim_{C\to-\infty} \frac{\mathrm{Var}^C(V)}{C^2} &= \lim_{C\to-\infty} \frac{\mathbb{E}(V^2 \mid V \le C) - [\mathbb{E}(V \mid V \le C)]^2}{C^2} \\ &= \lim_{C\to\infty} \frac{\mathbb{E}(V^2 \mid V \ge C) - [\mathbb{E}(V \mid V \ge C)]^2}{C^2} \\ &= \frac{\alpha}{\alpha-2} - \left(\frac{\alpha}{\alpha-1}\right)^2,\end{aligned}$$

where the last step follows from Lemma 4.7, and equation (4.4) follows.

Second, let us show that

$$\lim_{C\to-\infty} \frac{\mathbb{E}^C(W)}{C^2} = \frac{\alpha}{(\alpha-2)(\alpha-1)}. \tag{4.5}$$

Observe that, for $C \ge 0$, $\mathbb{E}^{-C}(W) = \mathbb{E}(W \mid \sqrt{W}X \ge C)$. Now choose a nonnegative sequence $(c_n)_{n\ge1}$ such that $c_n \to \infty$ as $n \to \infty$, and write

$$\begin{aligned}\lim_{n\to\infty} \frac{\mathbb{E}^{-c_n}(W)}{c_n^2} &= \lim_{n\to\infty} \frac{\mathbb{E}(W \mid \sqrt{W}X \ge c_n)}{c_n^2} = \lim_{n\to\infty} \frac{\mathbb{E}(W\mathbf{1}_{\{\sqrt{W}X\ge c_n\}})}{c_n^2\,\bar{F}_V(c_n)} \\ &= \lim_{n\to\infty} \mathbb{E}\left[\frac{\mathbb{E}\left(W\mathbf{1}_{\{\sqrt{W}X\ge c_n\}} \mid X\right)}{c_n^2\,\bar{F}_V(c_n)}\right].\end{aligned} \tag{4.6}$$

To exchange the order of limit and expectation, we must show that the sequence of random variables $G_n(X)$, $n \ge 1$, defined **P**–a.s. by

$$G_n(X) := \frac{\mathbb{E}\left(W\mathbf{1}_{\{\sqrt{W}X\ge c_n\}} \mid X\right)}{c_n^2\,\bar{F}_V(c_n)}, \quad n \ge 1,$$

is uniformly integrable and that $G_n(X)$, $n \ge 1$, converges **P**–a.s. To establish uniform integrability we have to show that

$$\lim_{D\to\infty} \limsup_{n} \int_0^\infty G_n(x)\,\mathbf{1}_{\{G_n(x)>D\}}\,\mathrm{N}(\mathrm{d}x) = 0. \tag{4.7}$$

Let L be the slowly varying function that satisfies $\bar{F}_V(x) = x^{-\alpha}L(x)$, with α the tail index of V.

First, choose $\bar{C} > 0$ such that for all $c > \bar{C}$ the following four conditions are satisfied:

$$\frac{\mathbb{E}(W \mid \sqrt{W} > c)}{c^2} \le \frac{\alpha}{\alpha-2} + 1, \tag{4.8}$$

$$\bar{F}_{\sqrt{W}}(c)\,\mathbb{E}((X\mathbf{1}_{\{X>0\}})^\alpha) \le 2\bar{F}_V(c), \tag{4.9}$$

$$\frac{x^{\alpha-2}L(c/x)}{L(c)} < 2, \quad 0 < x \le 1, \tag{4.10}$$

$$\frac{x^{\alpha-2}L(c/x)}{L(c)} < 2\,x^{\alpha-1}, \quad 1 \le x \le c/\bar{C}. \tag{4.11}$$

The existence of $\bar{C}$ satisfying conditions (4.8) and (4.9) follows from Lemmas 4.4 and 4.7. Conditions (4.10) and (4.11) are a consequence of Theorem A.6 with parameters $\varepsilon = 1$ and $\delta = \alpha - 2$ for (4.10) and $\varepsilon = 1$ and $\delta = 1$ for (4.11).

Now, let $c_n \ge \bar{C}$ and $x \le c_n/\bar{C}$. Then, by equations (4.8) and (4.9),

$$\begin{aligned}
G_n(x) &= \frac{\mathbb{E}\big(W\mathbf{1}_{\{\sqrt{W}x \ge c_n\}}\big)}{c_n^2\,\bar{F}_V(c_n)} = \frac{\displaystyle\int_{c_n/x}^{\infty} w^2\,F_{\sqrt{W}}(\mathrm{d}w)}{c_n^2\,\bar{F}_V(c_n)} \\
&= \left(\frac{\displaystyle\int_{c_n/x}^{\infty} w^2\,F_{\sqrt{W}}(\mathrm{d}w)}{(c_n/x)^2\,\bar{F}_{\sqrt{W}}(c_n/x)}\right) \Bigg/ \left(\frac{c_n^2\,\bar{F}_V(c_n)}{(c_n/x)^2\,\bar{F}_{\sqrt{W}}(c_n/x)}\right) \\
&\le \left(\frac{\alpha}{\alpha-2} + 1\right)\frac{\bar{F}_{\sqrt{W}}(c_n/x)}{x^2\,\bar{F}_V(c_n)} \\
&\le H(\alpha)\frac{\bar{F}_V(c_n/x)}{x^2\,\bar{F}_V(c_n)} = H(\alpha)\frac{x^{\alpha-2}L(c_n/x)}{L(c_n)},
\end{aligned}$$

where $H(\alpha) := \dfrac{2}{\mathbb{E}\left((X\mathbf{1}_{\{X\ge 0\}})^{\alpha}\right)}\left(\dfrac{\alpha}{\alpha-2} + 1\right)$. Hence, by equations (4.10) and (4.11),

$$G_n(x) \le \begin{cases} 2H(\alpha) & 0 < x \le 1, \\ 2H(\alpha)\,x^{\alpha-1} & 1 < x \le c_n/\bar{C}. \\ \dfrac{\mathbb{E}(W)}{c_n^2\,\bar{F}_V(c_n)} & x > c_n/\bar{C}. \end{cases} \tag{4.12}$$

By equation (4.12) we obtain

$$\begin{aligned}
&\lim_{D\to\infty}\limsup_n \int_0^{c_n/\bar{C}} G_n(x)\,\mathbf{1}_{\{G_n(x)>D\}}\,\mathrm{N}(\mathrm{d}x) \\
&\le \lim_{D\to\infty}\int_0^1 2H(\alpha)\,\mathbf{1}_{\{2H(\alpha)>D\}}\,\mathrm{N}(\mathrm{d}x) \\
&\quad + \lim_{D\to\infty}\int_1^{\infty} 2H(\alpha)x^{\alpha-1}\,\mathbf{1}_{\{2H(\alpha)x^{\alpha-1}>D\}}\,\mathrm{N}(\mathrm{d}x) = 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\lim_{D\to\infty}\limsup_n \int_{c_n/\bar{C}}^{\infty} G_n(x)\,\mathbf{1}_{\{G_n(x)>D\}}\,\mathrm{N}(\mathrm{d}x) \\
&\le \lim_{D\to\infty}\limsup_n \mathbf{1}_{\{\frac{\mathbb{E}(W)}{c_n^2\,\bar{F}_V(c_n)}>D\}}\frac{\mathbb{E}(W)}{c_n^2\,\bar{F}_V(c_n)}\,[1-\mathrm{N}(c_n/\bar{C})] \\
&\le \limsup_n \frac{\mathbb{E}(W)}{c_n^2\,\bar{F}_V(c_n)}\big[1-\mathrm{N}(c_n/\bar{C})\big] \\
&= \limsup_n \frac{\mathbb{E}(W)\,\bar{\mathrm{N}}(c_n/\bar{C})\,e^{\lambda c_n}}{c_n^2\,\bar{F}_V(c_n)\,e^{\lambda c_n}} = 0,
\end{aligned}$$

where $\lambda > 0$, and where the last step follows, because the numerator vanishes in the limit, while the denominator tends to ∞, see Lemma A.7. This completes the proof of uniform integrability, that is, we have established equation (4.7).

It remains to evaluate $\lim_{n\to\infty} G_n(X)$. For any $x > 0$,

$$\lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \frac{\int_{c_n/x}^{\infty} w^2\, F_{\sqrt{W}}(\mathrm{d}w)}{c_n^2 \bar{F}_V(c_n)}$$

$$= \lim_{n\to\infty} \left[\left(\frac{\int_{c_n/x}^{\infty} w^2\, F_{\sqrt{W}}(\mathrm{d}w)}{(c_n/x)^2\, \bar{F}_{\sqrt{W}}(c_n/x)} \right) \Bigg/ \left(\frac{c_n^2\, \bar{F}_V(c_n)}{(c_n/x)^2\, \bar{F}_{\sqrt{W}}(c_n/x)} \right) \right].$$

By Lemma 4.7, we obtain for the limit of the numerator,

$$\lim_{n\to\infty} \left(\frac{\int_{c_n/x}^{\infty} w^2\, F_{\sqrt{W}}(\mathrm{d}w)}{(c_n/x)^2\, \bar{F}_{\sqrt{W}}(c_n/x)} \right) = \frac{\alpha}{\alpha - 2}.$$

For the denominator, application of Lemma 4.4 and the definition of regularly varying functions yield

$$\lim_{n\to\infty} \frac{c_n^2 \bar{F}_V(c_n)}{(c_n/x)^2\, \bar{F}_{\sqrt{W}}(c_n/x)} = \lim_{n\to\infty} \left[\frac{\bar{F}_V(c_n)}{\bar{F}_{\sqrt{W}}(c_n)} \cdot \frac{x^2\, \bar{F}_{\sqrt{W}}(c_n)}{\bar{F}_{\sqrt{W}}(c_n/x)} \right]$$
$$= \mathbb{E}((X\mathbf{1}_{\{X\geq 0\}})^{\alpha}) \cdot x^{-(\alpha-2)}.$$

Hence,

$$\lim_{n\to\infty} G_n(X) = \frac{\alpha\, \left(X\mathbf{1}_{\{X\geq 0\}}\right)^{\alpha-2}}{(\alpha-2)\, \mathbb{E}\left((X\mathbf{1}_{\{X\geq 0\}})^{\alpha}\right)} \quad \mathbf{P}\text{–a.s.}$$

Finally, taking expectation (continuing equation (4.6)), we obtain

$$\mathbb{E}\left[\lim_{n\to\infty} G_n(X)\right] = \frac{\alpha\, \mathbb{E}\left((X\mathbf{1}_{\{X\geq 0\}})^{\alpha-2}\right)}{(\alpha-2)\mathbb{E}\left((X\mathbf{1}_{\{X\geq 0\}})^{\alpha}\right)} = \frac{\alpha}{(\alpha-1)(\alpha-2)}, \tag{4.13}$$

where the last equality follows, because by partial integration

$$\int_0^{\infty} x^a\, \phi(x)\, \mathrm{d}x = \frac{1}{a+1} \int_0^{\infty} x^{a+2}\, \phi(x)\, \mathrm{d}x, \quad a > 0.$$

Because equation (4.13) holds for any sequence $(c_n)_{n\geq 1}$, we have established equation (4.5). Combining equations (4.4) and (4.5) completes the proof. □

4.3. *V* in the Gumbel MDA

Let us now consider the case when V is in the MDA of the Gumbel distribution. The MDA of the Gumbel distribution comprises a wide range of distribution functions, such as the log-normal, exponential, and the normal distributions.

The following proposition is the main result of this section.

PROPOSITION 4.8. *Let* $V \in \mathrm{MDA}(\Lambda)$. *Then,* $\lim_{C\to-\infty} \dfrac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} = 0$.

The corresponding formula for asset correlations is obtained by inserting this expression into equation (3.1).

COROLLARY 4.9. *Let* $V \in \mathrm{MDA}(\Lambda)$. *Then, for* $|\rho_i| < 1$ *and* $|\rho_j| < 1$,

$$\lim_{C\to-\infty} \mathrm{Corr}^C(A_i, A_j) = \frac{\rho_{ij} - \rho_i\,\rho_j}{\sqrt{\left(1-\rho_i^2\right)\left(1-\rho_j^2\right)}} = \mathrm{Corr}(Z_i, Z_j). \tag{4.14}$$

If either $|\rho_i| = 1$ *or* $|\rho_j| = 1$ *(but not both), then* $\lim_{C\to-\infty}\mathrm{Corr}^C(A_i, A_j) = 0$. *If* $\rho_i = \rho_j = \pm 1$, *then* $\lim_{C\to-\infty}\mathrm{Corr}^C(A_i, A_j) = 1$. *If* $\rho_i = -\rho_j = \pm 1$, *then* $\lim_{C\to-\infty}\mathrm{Corr}^C(A_i, A_j) = -1$.

The asymptotic correlations for the examples of Table 3.1 are given in Figure 4.1.

For the proof of Proposition 4.8 we need the following properties of the tail function $\bar{F}_V(x)$ of V. Observe that from the first part of Lemma 4.7 one obtains a "mean excess function for higher moments," in the sense that

$$\mathbb{E}(V^k - x^k \mid V > x) = k \int_x^\infty \frac{y^{k-1}\,\bar{F}_V(y)}{\bar{F}_V(x)}\,\mathrm{d}y, \quad x \ge 0. \tag{4.15}$$

Furthermore, it follows from Theorem A.9 that for $V \in \mathrm{MDA}(\Lambda)$

$$\lim_{C\to\infty} \frac{\displaystyle\int_C^\infty \bar{F}_V(x)\,\mathrm{d}x}{C\bar{F}_V(C)} = 0, \tag{4.16}$$

$$\lim_{C\to\infty} \int_C^\infty x\bar{F}_V(x)\,\mathrm{d}x = 0. \tag{4.17}$$

LEMMA 4.10. *Let* $V \in \mathrm{MDA}(\Lambda)$. *Then* $\lim_{C\to\infty} \frac{\int_C^\infty C\,\bar{F}_V(x)\,\mathrm{d}x}{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x} = 1$.

Proof. The denominator vanishes, see equation (4.17). Furthermore, the numerator is strictly smaller than the denominator, so it vanishes, too. By Lemma 4.1, $\bar{F}_V$ is continuous and therefore

$$\frac{\mathrm{d}}{\mathrm{d}C}\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x = -C\bar{F}_V(C) \neq 0, \quad C > 0,$$

so that the conditions for applying the rule of L'Hospital to $\frac{\int_C^\infty C\,\bar{F}_V(x)\,\mathrm{d}x}{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x}$ are satisfied:

$$\lim_{C\to\infty} \frac{\displaystyle\int_C^\infty C\,\bar{F}_V(x)\,\mathrm{d}x}{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x} = \lim_{C\to\infty} \frac{\displaystyle\int_C^\infty \bar{F}(x)\,\mathrm{d}x - C\bar{F}_V(C)}{-C\bar{F}_V(C)} = \lim_{C\to\infty} 1 - \frac{\displaystyle\int_C^\infty \bar{F}_V(x)\,\mathrm{d}x}{C\bar{F}_V(C)}.$$

But the last term vanishes, see equation (4.16). □

Proof of Proposition 4.8. By symmetry of V, we may equally show that

$$\lim_{C\to\infty} \frac{\mathrm{Var}(V \mid V \geq C)}{\mathbb{E}(W \mid V \geq C)} = 0. \tag{4.18}$$

Let $C > 0$ and write $\mathrm{Var}(V|V \geq C)$ in the form

$$\begin{aligned}\mathrm{Var}(V \mid V \geq C) &= C^2 + \mathbb{E}(V^2 - C^2 \mid V \geq C) - (C + \mathbb{E}(V - C \mid V \geq C))^2\\ &= \mathbb{E}(V^2 - C^2 \mid V \geq C) - 2C\,\mathbb{E}(V - C \mid V \geq C) - (\mathbb{E}(V - C \mid V \geq C))^2.\end{aligned}$$

We divide each component by $\mathbb{E}(V^2 - C^2 \mid V \geq C)$ and analyze the limit for $C \to \infty$: by equation (4.15) and Lemma 4.10,

$$\lim_{C\to\infty} \frac{2C\,\mathbb{E}(V - C \mid V \geq C)}{\mathbb{E}(V^2 - C^2 \mid V \geq C)} = \lim_{C\to\infty} \frac{\int_C^\infty C\,\bar{F}_V(x)\,\mathrm{d}x}{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x} = 1.$$

Furthermore, as $C \to \infty$

$$\frac{(\mathbb{E}(V - C \mid V \geq C))^2}{\mathbb{E}(V^2 - C^2 \mid V \geq C)} = \frac{\mathbb{E}(V - C \mid V \geq C)}{2C} \cdot \frac{2C\,\mathbb{E}(V - C \mid V \geq C)}{\mathbb{E}(V^2 - C^2 \mid V \geq C)} \to 0,$$

as $\dfrac{E(V - C \mid V \geq C)}{2C} = \dfrac{\int_C^\infty \bar{F}_V(x)\,\mathrm{d}x}{2C\bar{F}_V(C)}$ vanishes, cf. equation (4.16).

By combining the equations above we obtain

$$\lim_{C\to\infty} \frac{\mathrm{Var}(V \mid V \geq C)}{\mathbb{E}(V^2 - C^2 \mid V \geq C)} = 0.$$

For completing the proof of equation (4.18) we show that

$$\mathbb{E}(V^2 - C^2 \mid V \geq C) \sim 2\mathbb{E}(W \mid V \geq C). \tag{4.19}$$

From $\mathbb{E}(W\mathbf{1}_{\{V\geq C\}}) = \mathbb{E}(\mathbb{E}(W\mathbf{1}_{\{\sqrt{W}X\geq C\}} \mid W)) = \mathbb{E}(W\bar{\mathrm{N}}(\frac{C}{\sqrt{W}}))$ we obtain

$$\frac{\mathbb{E}(V^2 - C^2 \mid V \geq C)}{2\mathbb{E}(W \mid V \geq C)} = \frac{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x}{\mathbb{E}(W\mathbf{1}_{\{V\geq C\}})} = \frac{\int_C^\infty x\,\bar{F}_V(x)\,\mathrm{d}x}{\mathbb{E}\left(W\bar{\mathrm{N}}\left(\frac{C}{\sqrt{W}}\right)\right)}.$$

Both the numerator and the denominator vanish as $C \to \infty$; for the numerator this follows from equation (4.17), and for the denominator this follows by Dominated Convergence. We apply the rule of L'Hospital (where the derivative of the denominator is obtained by applying the Dominated Convergence Theorem) and further the identity (derived by partial integration)

$$\phi(x) = \int_x^\infty y\,\phi(y)\,\mathrm{d}y = x\,\bar{\mathrm{N}}(x) + \int_x^\infty \bar{\mathrm{N}}(y)\,\mathrm{d}y,$$

and obtain

$$\frac{C\,\bar{F}_V(C)}{\mathbb{E}\left(\sqrt{W}\phi\left(\frac{C}{\sqrt{W}}\right)\right)} = \frac{C\,\bar{F}_V(C)}{\mathbb{E}\left(C\,\bar{\mathrm{N}}\left(\frac{C}{\sqrt{W}}\right) + \sqrt{W}\int_{C/\sqrt{W}}^{\infty} \bar{\mathrm{N}}(y)\,\mathrm{d}y\right)}.$$

Finally, it follows from $\bar{F}_V(z) = \mathbb{E}(\bar{\mathrm{N}}(z/\sqrt{W}))$, $z \in \mathbb{R}$, that

$$\frac{C\,\bar{F}_V(C)}{\mathbb{E}\left(C\,\bar{\mathrm{N}}\left(\frac{C}{\sqrt{W}}\right) + \sqrt{W}\int_{C/\sqrt{W}}^{\infty} \bar{\mathrm{N}}(y)\,\mathrm{d}y\right)} = \frac{C\,\bar{F}_V(C)}{C\,\bar{F}_V(C) + \int_{C}^{\infty} \bar{F}_V(y)\,\mathrm{d}y} \to 1,$$

as $C \to \infty$, where the limit is a consequence of equation (4.16). This establishes equation (4.19) and completes the proof. □

REMARK 4.11. Because we have used only properties of the class of rapidly varying functions, to which the distribution functions with infinite right endpoint in the Gumbel MDA belong, Proposition 4.8 can be generalized to "If $\bar{F}_V$ is rapidly varying, then $\lim_{C\to-\infty} \frac{\mathrm{Var}^C(V)}{\mathbb{E}^C(W)} = 0$."

5. EXAMPLE: APPLICATION TO STRESS TESTING

Let us demonstrate how the results from the previous sections can be applied in practice. The initial motivation for analyzing correlations under stress is to better understand the results of stress tests performed in financial institutions. A typical example of a credit portfolio stress test is presented in Bonti et al. (2006), where a sample investment banking portfolio consisting of 25,000 loans is modeled in a structural credit portfolio model, see Merton (1974). In the Merton model, firm i defaults if its asset value A_i (at some fixed time horizon) falls below a threshold $D_i \in \mathbb{R}$, chosen such that $\mathbf{P}(A_i \le D_i)$ equals a given probability of default. Bonti et al. (2006) specify the dependence structure of the asset variables via 75 systematic factors that represent geographic regions and industries. The factors are assumed to be normally distributed, which is still industry standard in credit portfolio modeling. The stress scenario is implemented by truncating one of the systematic factors, the risk factor for the automotive industry.

The results of the stress test reported in Bonti et al. (2006) show a pattern that is often encountered: the relative impact of the stress on the expected loss of the portfolio, i.e., the mean of the portfolio loss distribution, is much stronger than on the VaR of the portfolio. In the specific stress test, the portfolio Expected Loss increased by 56% whereas the 99.98% VaR increased by 19% only.

The correlation formulas derived in this paper provide insight into the behavior of these risk measures under stress. First of all, note that the systematic country and industry factors typically show high positive (unstressed) correlations, which induce positive correlations of systematic factors and asset variables of the individual counterparties. In a normally distributed model, these correlations are significantly reduced under stress as illustrated in Figure 3.2. It is therefore not surprising that the increase of the stressed VaR, which strongly depends on the asset correlations in the stressed model, is less significant than the increase of the stressed EL, which only depends on exposure and stressed default probabilities.

In addition to the analysis of stress tests, the formulas for correlations under stress can be used to design new meaningful stress scenarios, where the conditional correlation is set to an appropriate value, such as an empirical conditional correlation. We provide two examples of credit portfolio stress tests for illustration.

Both stress tests are implemented in a structural credit portfolio model with NVM-distributed asset variables as specified in (2.1) and (2.2). We make two simplifying assumptions, which, however, do not affect the qualitative behavior of the results: First, we assume that the correlations of X and Y_i and the correlations of Y_i and Y_j are homogeneous, i.e., there exist ρ, $\bar{\rho}$ such that

$$\rho = \mathrm{Corr}(X, Y_i), \quad \bar{\rho}^2 = \mathrm{Corr}(Y_i, Y_j) \tag{5.1}$$

for all $i, j \in \{1, \ldots, k\}, i \neq j$. In this case, the asset variables $A_i := \sqrt{W} Y_i, i = 1, \ldots, k$, can be conveniently represented by the systematic factor $V = \sqrt{W}X$, an additional systematic factor $\sqrt{W}Y$ and k specific factors $\sqrt{W}\varepsilon_1, \ldots, \sqrt{W}\varepsilon_k$, where $X, Y, \varepsilon_1, \ldots, \varepsilon_k$ are standard normally distributed and independent:

$$A_i = \sqrt{W}\left(\bar{\rho}\left(\frac{\rho}{\bar{\rho}}X + \sqrt{1 - \rho^2/\bar{\rho}^2}\,Y\right) + \sqrt{1 - \bar{\rho}^2}\varepsilon_i\right). \tag{5.2}$$

It is easily seen that $\mathrm{Corr}(\sqrt{W}X, A_i) = \mathrm{Corr}(X, Y_i)$ and that $\mathrm{Corr}(A_i, A_j) = \mathrm{Corr}(Y_i, Y_j)$.

As a second simplification, we consider only credit portfolios consisting of loans with identical default probability p and we assume that the portfolios are infinitely granular with equally weighted loans with total value 1 (in some monetary unit). In the normally distributed case, this leads to the well-known Vasicek formula, Vasicek (1991), which is the basis for the Basel 2 formula for regulatory capital for credit risk. The Vasicek formula is easily generalized to NVMs, in which case the portfolio loss is given by

$$L := \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \mathbf{1}_{\{A_i \leq D\}} = \mathrm{N}\left(\frac{D/\sqrt{W} - \bar{\rho}(\rho/\bar{\rho}X + \sqrt{1 - \rho^2/\bar{\rho}^2}\ Y)}{\sqrt{1 - \bar{\rho}^2}}\right) \ \mathbf{P}\text{-a.s.}, \tag{5.3}$$

where N denotes the normal distribution function and D is the truncation threshold that yields a default probability p, i.e., $\mathbf{P}(A_i \leq D) = p$. Both the Vasicek formula and equation (5.3) follow from the Strong Law of Large Numbers by conditioning on W, X, and Y, and from Fubini's Theorem; for further details see, e.g., Bluhm, Overbeck, and Wagner (2003, proposition 2.5.4).

Note that equation (5.3) illustrates another interpretation of the general NVM model: instead of NVM-distributed systematic and specific components with a common mixing variable, which introduces dependence between the two components, one can think of normally distributed independent components and a stochastic default threshold $D/\sqrt{W}$.

5.1. Stressing the Correlation

In the first example, we choose a one-factor model with uncorrelated specific components, that is, the asset variables are given by equation (5.2) with $\bar{\rho}^2 = \rho^2 = \mathrm{Corr}(A_i, A_j)$ for any two assets i, j. We set the default probability to $p = 0.005$, and we examine VaR at the 99.9% level for different correlations and stress levels. The truncation threshold C

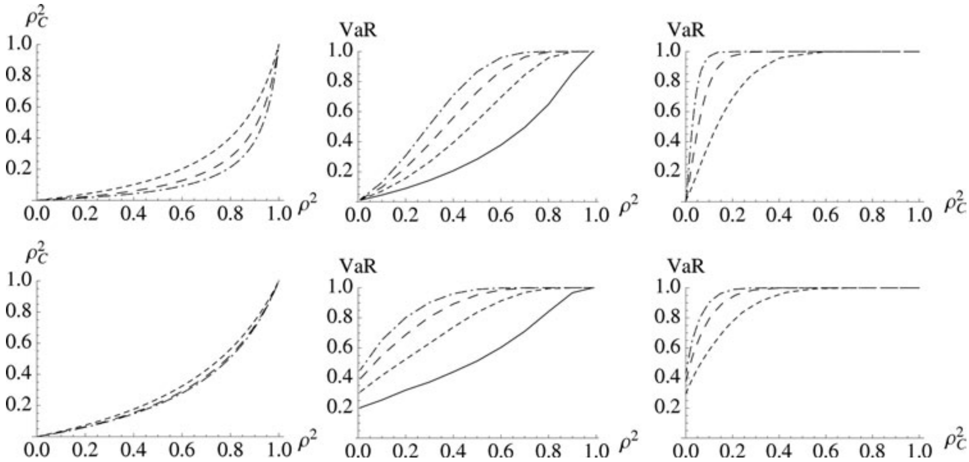


FIGURE 5.1. Top row: normally distributed asset returns; bottom row: t-distributed asset returns ($\nu = 5$). Left column: unconditional correlation ρ^2 versus conditional correlation ρ_C^2 for different stress probabilities; middle column: VaR for different stress probabilities; right column: VaR for different stress probabilities with correlation stressed. Stress probabilities: solid line: unstressed; small dashes: 10%; wide dashes: 1%; dash-dot-dash: 0.1%.

is chosen according to given stress probabilities of 10%, 1%, 0.1%, that is, $\mathbf{P}(V \leq C) \in \{10\%, 1\%, 0.1\%\}$. We consider both a normally distributed model and a t-distributed model with parameter $\nu = 5$.

In the middle column of Figure 5.1, the unstressed and stressed VaRs are shown as a function of the unstressed correlation ρ^2. As can be seen, VaR increases with the stress level and with ρ^2. Furthermore, VaR is greater in the t-distributed case than in the normally distributed case.

In these stress scenarios the dependence structure is specified via unstressed correlations. It is ignored that correlations under stress may behave differently than implied by the model. The formulas for conditional correlation derived in the previous sections can be applied to set the conditional correlation to a target value rather than relying on the values implied by the model. The left column of Figure 5.1 shows the mapping of asset correlation ρ^2 to the correlations under stress $\mathrm{Corr}^C(A_i, A_j) =: \rho_C^2$. This allows for choosing target correlations under stress by inserting appropriate unstressed correlations into the model. With the formulas for conditional correlation derived for normally and t-distributed random variables one can make the notion of a stressed correlation precise, whereas for the more general NVMs the asymptotic correlation may be used to derive the unstressed correlation.

Thus, in a second step, we consider the correlation itself as an additional risk factor subjected to stress. Here, we employ a "constant correlation" assumption in the sense that we require the target conditional correlation to equal the original unstressed correlation. The resulting VaRs under stress are shown in the right column of Figure 5.1. Because in our example conditional correlation decreases with increasing stress, these stressed VaRs are greater than the stressed VaRs in the first case. In particular, under high stress, VaR increases to 1, the notional amount of the portfolio, even for moderately low correlations. On the other hand, the differences in expected loss, which corresponds to the probability of default, between the two stress scenarios are relatively small compared to the differences in VaR.

The example gives an explanation for the observation of the relatively moderate increase of VaR under stress compared to EL mentioned at the beginning of the section. It confirms that decreasing conditional correlation in light-tailed models is a key driver of this behavior. Moreover, we demonstrate that choosing a heavy-tailed distribution or that stressing the correlation in an appropriate way mitigates the problem.

All calculations were done by simulating the risk factor. For each given stress probability, 100,000 simulations fulfilling the condition $\{V \leq C\}$ were drawn. To assess the quality of the estimates we calculate the sample standard deviation of the VaR estimator assuming a correlation of 0.5 by simulating the estimator 10 times. The resulting standard deviations are within 2% of the VaR figure in the normally distributed case and within 1.2% in the t-distributed case.

In the setup above, one may also consider one of the assets as a risk factor. Truncating the asset return at its default boundary and providing the correlation under stress may be useful for studying contagion effects.

5.2. Empirical Example

In contrast to the first example we now implement stress scenarios in the unrestricted two-factor model (5.3) and calibrate model parameters to empirical data, consisting of rating data and equity time series of the DAX and its 30 constituents. The default probability p that enters the Vasicek formula is set to 0.058% (i.e., 5.8 basis points). This value is derived from S&P credit ratings of the DAX companies.[10] Taking the DAX return to be the systematic factor, the correlation parameters are calibrated to historical equity data, i.e., to daily returns from the DAX index and its $n = 30$ constituents ranging from February 5, 2001 to December 20, 2011.[11] The average (unstressed) correlations are given by

$$\rho = \mathrm{Corr}(V, A_i) = \frac{1}{n}\sum_{i=1}^{n} \rho_i = 0.6161,$$

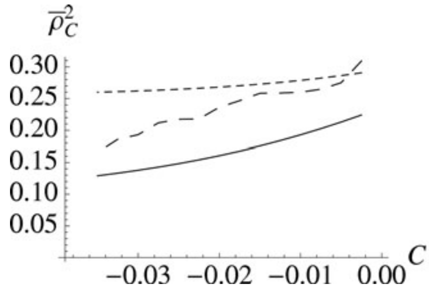
$$\bar{\rho}^2 = \mathrm{Corr}(A_i, A_j) = \frac{1}{n^2 - n(n-1)/2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} \rho_{ij} = 0.415.$$

Note that $\bar{\rho} = 0.6442$ and that $\rho^2 = 0.3796$.

We use the formulas for conditional correlation to calculate the correlations under stress for ρ^2 and $\bar{\rho}^2$ as estimated for both the normal distribution and the t-distribution,

[10] The credit ratings are mapped to one-year default probabilities, using cumulative default rates of European corporates from 1981 until 2010, S&P (2011). The default probabilities are in turn mapped to hazard rates, of which a weighted average is calculated. From this, an average default probability of 0.00058 is determined. Three names do not have a credit rating, making up a total of 9.62% of the DAX, which we simply ignore in our calculations.

[11] Because the time series data are typically not i.i.d., one can consider applying a GARCH filter to each time series. This entails fitting each time series to a GARCH model and dividing each return by its corresponding volatility estimate from the GARCH model, yielding the so-called innovations, which are assumed to be i.i.d. Each innovation is then multiplied with the volatility forecast for the next time period. Under the assumption of i.i.d. innovations, this yields a sample of i.i.d. returns. We refer to, e.g., Barone-Adesi, Bourgoin, and Giannopoulos (1998), McNeil and Frey (2000), Alexander and Sheedy (2008) for further details on GARCH filtering.



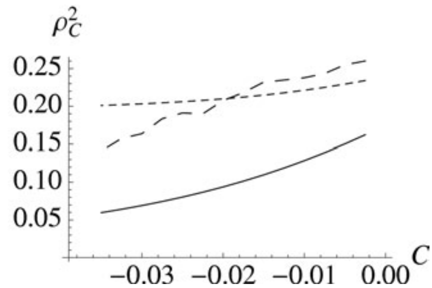


FIGURE 5.2. Average correlations under stress $\bar{\rho}_C^2 = \mathrm{Corr}^C(A_i, A_j)$ (left), respectively, $\rho_C^2 = (\mathrm{Corr}^C(V, A_i))^2$ (right); solid line (small dashes): assuming constant ρ^2 and $\bar{\rho}^2$ and normally distributed (t-distributed) asset returns; wide dashes: $\bar{\rho}_C^2$, respectively, ρ_C^2, implied by the data.

where in the latter case we set $\nu = 3.46$, which is the average tail index estimated from the DAX returns. In addition, we calculate average empirical correlations when the DAX is subjected to stress. As can be seen from Figure 5.2, the empirical conditional correlation behaves differently than the conditional correlations in both the normally distributed and the t-distributed case. For thresholds close to 0, the empirical conditional correlation is similar to the t-distributed case. For more severe stress the empirical conditional correlation is smaller than in the t-distributed case, but still greater than in the normally distributed case.[12] Of course, as the stress increases, the sample size decreases. The actual number of samples under conditioning ranges from 74 at $C = -0.035$ to 1069 at $C = -0.0025$ (the total number of returns is 2,769).

We calculate VaR under stress for ρ^2 and $\bar{\rho}^2$ as estimated for both the normal distribution and the t-distribution. We repeat the calculations, but now with the empirical conditional correlations. For this we back out unstressed correlations from the empirical stressed correlations ρ_C^2 and $\bar{\rho}_C^2$ that need to be fed into the model to produce the target correlations under stress.

The resulting ELs and VaRs are shown in Figure 5.3.[13] In the normally distributed case the lower conditional correlation (compared to the empirical conditional correlation) translates into lower stressed ELs and VaRs (see top graphs of Figure 5.3). Thus, imposing the empirical conditional correlations rather than using the unconditional estimates gives a more prudent estimate of risk under the stress scenario. For the t-distributed case, the impact of the correlation parameter turns out to be more moderate: in line with the correlation analysis in Figure 5.2, the stressed EL and VaR figures turn out to be similar for truncation levels greater than -2% (given the scarcity of data, a clear interpretation for

[12]Observe that one should not necessarily expect that conditional correlations increase with the level of stress applied. This so-called *correlation breakdown* refers to an increase of *dynamic* correlations in stressed markets and has been verified empirically in different markets, see for example, Longin and Solnik (2001), Loretan and English (2000), Ang and Chen (2002), Ang and Bekaert (2002), Kim and Finger (2000). It has gained wide interest as it annihilates the risk reducing effect of a well-diversified or a hedging portfolio when it is needed most, namely in a stressed market. The association of conditional correlation with correlation breakdown is not straightforward: one cannot necessarily deduce from the behavior of conditional correlations that correlations change with time or state. This point is discussed in detail in Boyer et al. (1999).

[13]All calculations are done by Monte Carlo simulation with 500,000 simulations fulfilling $\{\sqrt{W}X \leq C\}$ for each truncation level C.

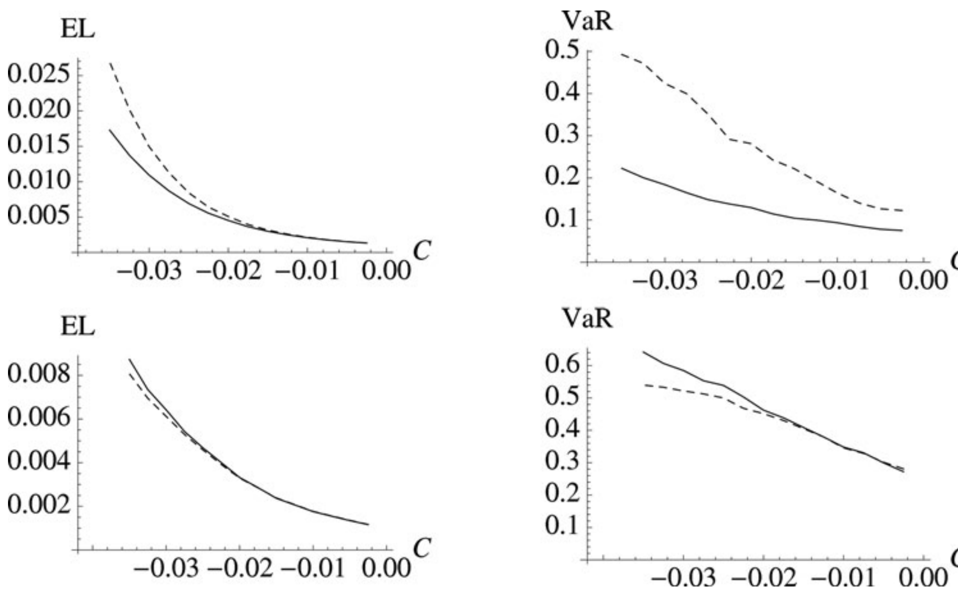


FIGURE 5.3. EL (left) and VaR (right) for the DAX portfolio under stress. Top: normal distribution case; bottom: t-distribution case. Solid line: assuming constant ρ^2 and $\bar{\rho}^2$; dashed line: setting ρ_C^2 and $\bar{\rho}_C^2$ to the values determined from the data.

lower truncation levels may be difficult). Hence, in this example the t-distributed model provides a consistent framework for modeling unstressed as well as realistic stressed correlations.

6. CONCLUSION

We study correlations in truncated NVM distributed models. The main motivation for our analysis comes from stress testing. Stress tests on bank portfolios have gained particular importance as a technique for assessing the adequacy of capital buffers in financial institutions, and as such form an integral part of risk management and banking supervision.

The class of NVM models is significant for practical applications as it encompasses many models commonly encountered in the financial industry. For example, our results can be applied to the most popular credit portfolio models such as CreditMetrics$^{\text{TM}}$ and Moody's KMV Portfolio Manager$^{\text{TM}}$, where log-returns of asset values follow a multivariate normal distribution. The appeal of the NVM distribution in the industry certainly comes from its simplicity combined with diverse tail behavior, ranging from light- to heavy-tailed.

From a technical point of view, our results are summarized as follows: for the special cases of jointly normally and t-distributed asset returns we have derived closed formulas for the correlation in the truncated model. More generally, for the NVM distribution, we have calculated the asymptotic limit of the correlation, which depends on whether the variables are in the MDA of the Fréchet or Gumbel distribution. It turns out that correlations in heavy-tailed NVM models are less sensitive to stress than in medium- or light-tailed models.

Our analysis allows for a critical review of the suitability of this model class to reproduce empirically observed conditional correlations, and as such provides valuable information for risk and capital management in financial institutions. Also, using our results, the conditional correlation can take the role of a further free parameter, which can be set to a target value, for example to design more prudent stress tests.

We illustrate the application of our correlation formulas by comparing VaR under stress for a stylized credit portfolio, both under different distribution and correlation assumptions, and using empirical conditional correlations derived from DAX data.

APPENDIX A: SOME RESULTS FROM EVT

We give a brief review of some results of EVT that are applied in this paper; standard references to EVT are Embrechts et al. (1997), Resnick (2007), and de Haan and Ferreira (2006).

The central result of classical EVT is the Fisher–Tippett Theorem, which specifies the form of the limit distribution for centered and normalized maxima.

THEOREM A.1 [Fisher–Tippett Theorem, limit laws for maxima]. *Let* (X_n) *be a sequence of i.i.d. random variables, and let* $M_n = \max(X_1, \ldots, X_n)$. *If there exist norming constants* $c_n > 0$, $d_n \in \mathbb{R}$ *and some nondegenerate distribution function* H *such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathcal{L}} H, \quad \text{as } n \to \infty, \tag{A.1}$$

then H *belongs to the type of one of the following three distribution functions:*

$$\text{Fréchet:} \quad \Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0,$$

$$\text{Weibull:} \quad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \le 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0,$$

$$\text{Gumbel:} \quad \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

If equation (A.1) holds, then the type of the limiting distribution is uniquely determined, and does not depend on the particular choice of norming constants c_n, d_n.

DEFINITION A.2. A random variable X with distribution function F belongs to the MDA of H if there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that equation (A.1) holds, written $X \in \text{MDA}(H)$ and $F \in \text{MDA}(H)$.

A different parameterization allows grouping the Fréchet, Weibull, and Gumbel distributions into one family of distribution functions, called the *generalized extreme value distributions*.

A.1. Maximum Domain of Attraction of the Fréchet Distribution

The distributions that are in MDA(Φ_α), $\alpha > 0$, have a particularly elegant and convenient representation via slowly varying functions or regularly varying functions.

DEFINITION A.3. A positive, Lebesgue-measurable function L on $(0, \infty)$ is slowly varying (at ∞) if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad t > 0.$$

A positive, Lebesgue-measurable function h on $(0, \infty)$ is regularly varying (at ∞) with index $\rho \in \mathbb{R}$ if

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\rho, \qquad t > 0.$$

We shall drop the specification "at ∞," because we only consider functions that are regularly or slowly varying at ∞. It is easily seen that h is regularly varying with index ρ if and only if there exists a slowly varying function L such that $h(x) = x^\rho L(x)$ (just choose $L(x) = h(x)/x^\rho$). For details on regularly varying functions, we refer to Bingham, Goldie, and Teugels (1987).

The following theorem completely characterizes those distribution functions that are in the MDA of the Fréchet distribution.

THEOREM A.4 [Gnedenko (1943)]. *The distribution function F belongs to* MDA(Φ_α), $\alpha > 0$, *if and only if* $\bar{F}(x) := 1 - F(x) = x^{-\alpha} L(x)$ *for some slowly varying function L.*

It follows directly that these distribution functions have infinite right endpoint, that is $\sup\{x : F(x) < 1\} = \infty$. Examples of distributions in MDA(Φ_α) are the t-distribution and the inverse gamma distribution. The parameter α is called the *tail index*. In the case of the t-distribution the degree of freedom is α.

An important result for regularly varying functions, which is used in Section 4.2, is Karamata's Theorem, see Bingham et al. (1987, theorem 1.5.11 (ii)) or McNeil et al. (2005, theorem A.5).

THEOREM A.5 [Karamata's Theorem]. *Let L be a slowly varying function that is locally bounded in* $[x_0, \infty)$, *for some* $x_0 \geq 0$. *Then,*

(i) *for* $\kappa > -1$,

$$\int_{x_0}^{x} t^\kappa L(t)\,\mathrm{d}t \sim \frac{1}{\kappa+1} x^{\kappa+1} L(x), \qquad \text{as} \quad x \to \infty,$$

(ii) *for* $\kappa < -1$,

$$\int_{x}^{\infty} t^\kappa L(t)\,\mathrm{d}t \sim -\frac{1}{\kappa+1} x^{\kappa+1} L(x), \qquad \text{as} \quad x \to \infty.$$

The following bound for regularly varying functions can be found, e.g., in Potter (1942) and de Haan and Ferreira (2006, proposition B.1.9):

THEOREM A.6 [Potter bound]. *Suppose h is regularly varying with index ρ. For arbitrary $\varepsilon > 0$ and $\delta > 0$ there exists $x_0 = x_0(\varepsilon, \delta)$ such that for $x \geq x_0$, $x t \geq x_0$,*

$$(1-\varepsilon)t^{\rho}\min(t^{\delta}, t^{-\delta}) < \frac{h(tx)}{h(x)} < (1+\varepsilon)t^{\rho}\max(t^{\delta}, t^{-\delta}).$$

Conversely, if h satisfies the property above then h is regularly varying with index ρ.

The following is a direct consequence of corollary 1.3.2 and lemma 1.3.5 of Embrechts et al. (1997):

LEMMA A.7. *If the tail distribution function $\bar{F}$ is regularly varying, then* $\lim_{x\to\infty} e^{\varepsilon x}\bar{F}(x) = \infty$, *for all $\varepsilon > 0$.*

A.2. Maximum Domain of Attraction of the Weibull Distribution

Distributions in MDA(Ψ_α), $\alpha > 0$, all have finite right endpoint. They include, for example, the uniform and the beta distributions.

A.3. Maximum Domain of Attraction of the Gumbel Distribution

The MDA of the Gumbel distribution covers a wide range of distribution functions, such as the log-normal, exponential, and the normal distributions. This class contains distribution functions with both finite and infinite right endpoint. If the distribution function $F \in$ MDA(Λ) has an infinite right endpoint, then $\bar{F}$ is rapidly varying.

DEFINITION A.8. A positive, Lebesgue-measurable function h on $(0, \infty)$ is rapidly varying with index $-\infty$ if

$$\lim_{x\to\infty}\frac{h(tx)}{h(x)} = \begin{cases} 0, & t > 1, \\ \infty, & 0 < t < 1. \end{cases}$$

For rapidly varying functions, there exists an analogous result to Karamata's Theorem, see e.g., Embrechts et al. (1997, theorem A3.12):

THEOREM A.9. *Let h be a nonincreasing, rapidly varying function. Then for some $z > 0$ and all $\kappa \in \mathbb{R}$*

$$\int_z^{\infty} t^{\kappa}\, h(t)\, \mathrm{d}t < \infty \quad \text{and} \quad \lim_{x\to\infty}\frac{x^{\kappa+1}\, h(x)}{\int_x^{\infty} t^{\kappa}\, h(t)\, \mathrm{d}t} = \infty.$$

REFERENCES

ABRAMOWITZ, M., and I. A. STEGUN (1972): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th Dover printing, 10th GPO printing ed., New York: Dover Publications.

ACHARYA, V., L. PEDERSEN, T. PHILIPPON, and M. RICHARDSON (2010): Measuring Systemic Risk. Working Paper, New York University, Stern School of Business.

ADRIAN, T., and M. BRUNNERMEIER (2011): CoVaR. Working Paper, Princeton University.

ALEXANDER, C., and E. SHEEDY (2008): Developing a Stress Testing Framework Based on Market Risk Models, *J. Banking Finance* 32(10), 2220–2236.

ANG, A., and G. BEKAERT (2002): International Asset Allocation with Regime Shifts, *Rev. Financ. Stud.* 15(4), 1137–1187.

ANG, A., and J. CHEN (2002): Asymmetric Correlations of Equity Portfolios, *J. Financ. Econ.* 63(3), 443–494.

BAE, T., and I. ISCOE (2010): Correlations under Stress, *Int. Rev. Appl. Financ. Issues Econ.* 2(2), 248–271.

BARNDORFF-NIELSEN, O. E. (1998): Processes of Normal Inverse Gaussian Type, *Finance Stoch.* 2, 41–68.

BARONE-ADESI, G., F. BOURGOIN, and K. GIANNOPOULOS (1998): Don't Look Back, *Risk* 11(8), 100–103.

BERKOWITZ, J. (2000): A Coherent Framework for Stress-Testing, *J. Risk* 2(2), 5–15.

BINGHAM, N. H., and R. KIESEL (2002): Semi-Parametric Modelling in Finance: Theoretical Foundations, *Quant. Finance* 2(4), 241–250.

BINGHAM, N. H., C. M. GOLDIE, and J. L. TEUGELS (1987): *Regular Variation*, Cambridge, MA: Cambridge University Press.

BIS (2009): *Principles for Sound Stress Testing Practices and Supervision*, Basel, Switzerland: Basel Committee on Banking Supervision, Consultative Document, Bank for International Settlements.

BLUHM, C., L. OVERBECK, and C. WAGNER (2003): *An Introduction to Credit Risk Modeling*, London: Chapman & Hall/CRC.

BONTI, G., M. KALKBRENER, C. LOTZ, and G. STAHL (2006): Credit Risk Concentrations under Stress, *J. Credit Risk* 2(3), 115–136.

BOYER, B. H., M. S. GIBSON, and M. LORETAN (1999): Pitfalls in Tests for Changes in Correlations, Board of Governors of the Federal Reserve System, International Finance Discussion Papers.

BREIMAN, L. (1965): On Some Limit Theorems Similar to the Arc-Sin Law, *Theory Probab. Appl.* 10, 323–331.

BREUER, T., M. JANDAČKA, K. RHEINBERGER, and M. SUMMER (2009): How to Find Plausible, Severe, and Useful Stress Scenarios, *Int. J. Central Banking* 5(3), 205–224.

BRUNNERMEIER, M., A. CROCKETT, C. GOODHART, A. PERSAUD, and H. SHIN (2009): The Fundamental Principles of Financial Regulation. Geneva Reports on the World Economy 11, Geneva, Switzerland: International Center for Monetary and Banking Studies.

CROSBIE, B., and J. BOHN (2002): Modelling Default Risk. Working Paper, KMV Corporation. www.kmv.com.

DE HAAN, L., and A. FERREIRA (2006): *Extreme Value Theory: An Introduction*, New York: Springer.

DUELLMANN, K., and M. ERDELMEIER (2009): Crash Testing German Banks, *Int. J. Central Banking* 5(3), 139–175.

EMBRECHTS, P., C. KLÜPPELBERG, and T. MIKOSCH (1997): *Modelling Extremal Events for Finance and Insurance*, Berlin: Springer.

GEMAN, H., D. B. MADAN, and M. YOR (2001): Time Changes for Lévy Processes, *Math. Finance* 11(1), 79–96.

GNEDENKO, B. (1943): Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.* 44(3), 423–453.

GUPTON, G. M., C. C. FINGER, and M. BHATIA (1997): CreditMetrics: Technical Document. Technical Report, JP Morgan & Co.

KALKBRENER, M., and L. OVERBECK (2008): Stress Testing in Credit Portfolio Models. Working Paper, University of Giessen, Department of Mathematics.

KALLENBERG, O. (2001): *Foundations of Modern Probability*, 2nd ed., New York: Springer.

KIM, J., and C. C. FINGER (2000): A Stress Test to Incorporate Correlation Breakdown, *J. Risk* 2, 5–20.

KUPIEC, P. H. (1998): Stress Testing in a Value at Risk Framework, *J. Derivatives* 6, 7–24.

LAROSIÈRE, J., L. BALCEROWICZ, O. ISSING, R. MASERA, C. MCCARTHY, L. NYBERG, J. PÉREZ, AND O. RUDING. (2009): Report. The High-Level Group on Financial Supervision in the EU, February 2009, Brussels: European Commission.

LONGIN, F. (2000): From Value at Risk to Stress Testing: The Extreme Value Approach, *J. Banking Finance* 24(7), 1097–1130.

LONGIN, F., and B. SOLNIK (2001): Extreme Correlation of International Equity Markets, *J. Finance* 56(2), 649–676.

LORETAN, M., and W. ENGLISH (2000): Evaluating Changes in Correlations during Periods of High Market Volatility, *BIS Quarterly Rev.* June 2000, 29–36.

MADAN, D. B., and E. SENETA (1990): The Variance Gamma (VG) Model for Share Market Returns, *J. Bus.* 63(4), 511–524. ISSN 0021-9398.

MALEVERGNE, Y., and D. SORNETTE (2006): Investigating Extreme Dependences, in *Extreme Financial Risks*. Heidelberg: Springer, pp. 227–270.

MCNEIL, A. J., and R. FREY (2000): Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach, *J. Empir. Finance* 7(3), 271–300.

MCNEIL, A. J., R. FREY, and P. EMBRECHTS (2005): *Quantitative Risk Management*, Princeton, NJ: Princeton University Press.

MERTON, R. C. (1974): On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *J. Finance* 29(2), 449–470.

POTTER, H. S. A. (1942): The Mean Values of Certain Dirichlet Series, II, *Proc. London Math. Soc.* 2(1), 1–19.

RESNICK, S. I. (2007): *Extreme Values, Regular Variation, and Point Processes*, New York: Springer Verlag.

SAMPFORD, M. R. (1953): Some Inequalities on Mill's Ratio and Related Functions, *Ann. Math. Statist.* 24(1), 130–132.

SCHACHTER, B. (2001): How Well Can Stress Tests Complement VaR, Working Paper, gloriamundi.org.

S&P (2011): Default, Transition, and Recovery: 2010 Annual European Corporate Default Study and Rating Transitions, Report, Standard & Poor's, March 2011.

TURNER, A. (2009): The Turner Review—A Regulatory Response to the Global Banking Crisis, London: FSA.

VASICEK, O. (1991): *Limiting Loan Loss Probability Distribution*, San Francisco, CA: KMV Corporation.