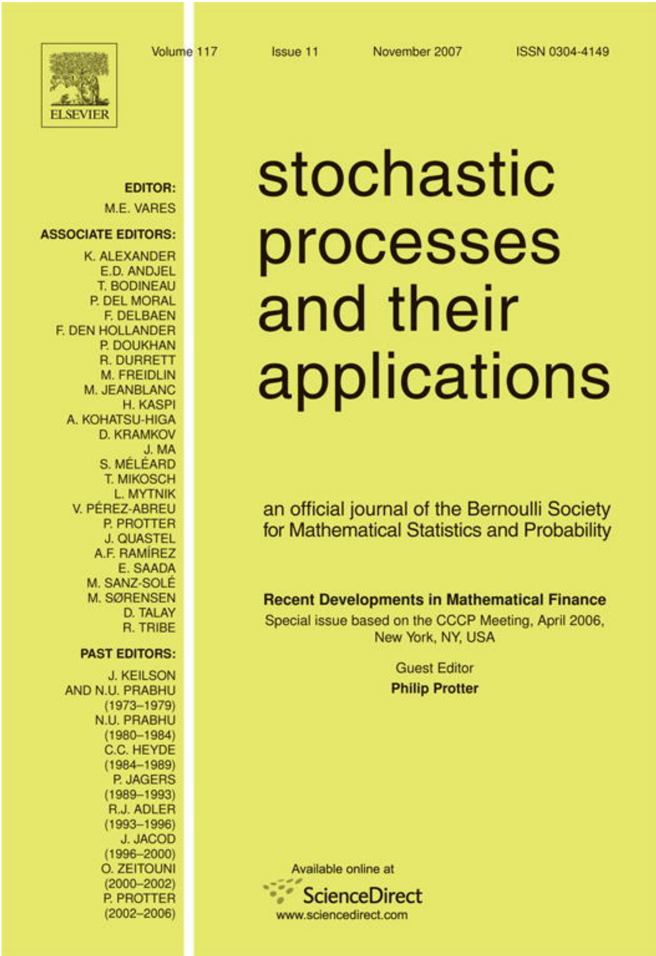
Volume 117
Issue 11
November 2007
ISSN 0304-4149
ELSEVIER
stochastic processes and their applications
EDITOR:
M.E. VARES
ASSOCIATE EDITORS:
K. ALEXANDER
E.D. ANDJEL
T. BODINEAU
P. DEL MORAL
F. DELBAEN
F. DEN HOLLANDER
P. DOUKHAN
R. DURRETT
M. FREIDLIN
M. JEANBLANC
H. KASPI
A. KOHATSU-HIGA
D. KRAMKOV
J. MA
S. MÉLÉARD
T. MIKOSCH
L. MYTNIK
V. PÉREZ-ABREU
P. PROTTER
J. QUASTEL
A.F. RAMÍREZ
E. SAADA
M. SANZ-SOLÉ
M. SØRENSEN
D. TALAY
R. TRIBE
PAST EDITORS:
J. KEILSON
AND N.U. PRABHU
(1973–1979)
N.U. PRABHU
(1980–1984)
C.C. HEYDE
(1984–1989)
P. JAGERS
(1989–1993)
R.J. ADLER
(1993–1996)
J. JACOD
(1996–2000)
O. ZEITOUNI
(2000–2002)
P. PROTTER
(2002–2006)
an official journal of the Bernoulli Society for Mathematical Statistics and Probability
Recent Developments in Mathematical Finance
Special issue based on the CCCP Meeting, April 2006, New York, NY, USA
Guest Editor
Philip Protter
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Stochastic Processes and their Applications 117 (2007) 1587–1605
stochastic processes and their applications
www.elsevier.com/locate/spa

Computing strategies for achieving acceptability: A Monte Carlo approach☆

Soumik Pal*

Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, United States

Received 14 July 2006; received in revised form 29 September 2006; accepted 29 September 2006
Available online 2 June 2007

Abstract

We consider a trader who wants to direct his or her portfolio towards a set of acceptable wealths given by a convex risk measure. We propose a Monte Carlo algorithm, whose inputs are the joint law of stock prices and the convex risk measure, and whose outputs are the numerical values of initial capital requirement and the functional form of a trading strategy for achieving acceptability. We also prove optimality of the capital obtained. Explicit theoretical evaluations of hedging strategies are extremely difficult, and we avoid the problem by resorting to such computational methods. The main idea is to utilize the finite Vapnik–Červonenkis dimension of a class of possible strategies.

Keywords: Measures of risk; VC dimension; Portfolio optimization; Neyman–Pearson lemma; Optimization algorithm

1. Introduction

1.1. Objective

In this paper we consider a T-period market model with a single stock and a money market. To model uncertainty in the stock price movements, we consider a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_T \subseteq \mathcal{F}$. At every time point $t = 0, 1, 2, \ldots, T$, the discounted price of the stock, S_t, is assumed to be an integrable random variable measurable with respect to $\mathcal{F}_t$.

☆ This work was partially supported by NSF grant 06-01774.

* Corresponding address: Departments of Mathematics and ORIE, 506 Malott Hall, Cornell University, Ithaca, NY 14853, United States. Tel.: +1 6463548342.

E-mail address: soumik@math.cornell.edu.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2006.09.012

Next, we consider a convex measure of risk. In the following subsection we briefly discuss the definition and significance of such a measure. Here it suffices to define it in the following way. Let $\{Q_i\}$, $i = 1, \ldots, m$, be a collection of probability measures on the sample space $(\Omega, \mathcal{F})$ which are absolutely continuous with respect to P, with Radon–Nikodým derivatives

$$\{f_i \overset{\triangle}{=} \mathrm{d}Q_i/\mathrm{d}P\}. \tag{1}$$

We are also given a collection $\{\alpha_i\}$ of real numbers. For every random variable $X \in \cap_i \mathbf{L}^1(Q_i)$, define

$$\rho(X) \overset{\triangle}{=} \sup_{1\le i\le m} [E^{Q_i}(-X) + \alpha_i] = \sup_{1\le i\le m} [-E(Xf_i) + \alpha_i]. \tag{2}$$

E here denotes taking expectation with respect to P. We call such a ρ a convex measure of risk.

Let us now introduce an agent who follows a self-financing portfolio by holding ξ_t number of shares in between time periods t and $(t+1)$. Due to the non-anticipative nature of trading, each ξ_t is an $\mathcal{F}_t$-measurable random variable. For any choice of initial capital w_0, and strategy $(\xi_0, \xi_1, \ldots, \xi_{T-1})$, let $V(w_0, \xi)$ denote the discounted terminal value of the portfolio, i.e.,

$$V(w_0, \xi) \overset{\triangle}{=} w_0 + W(\xi), \quad \text{where} \tag{3}$$

$$W(\xi) = \sum_{t=0}^{T-1} \xi_t (S_{t+1} - S_t). \tag{4}$$

In this paper we investigate an algorithm for computing a near-minimal w_0 and strategy ξ, such that $\rho(V(w_0, \xi)) \le 0$. We shall then say that $V(w_0, \xi)$ is *acceptable*.

Our objective is indeed numerical computation, and not just theoretical expressions. We do not impose any restrictions on the law of the price process S. However, we do assume the existence of $(\mathcal{F}_t, P)$-integrable random variables a_t and b_t such that the agent is forced to obey

$$a_t \le \xi_t \le b_t, \quad \forall t = 0, 1, \ldots, T-1. \tag{5}$$

This is often a natural assumption dictated by trading constraints. In any case, this is crucial for our analysis.

The literature on convex measures of risk is almost silent about computing strategies for achieving acceptability, the primary difficulty being that the terminal conditions on the portfolio are not given by almost-sure equalities/inequalities. This prevents the use of classical change-of-measure techniques. In this paper, we take a novel computational approach, combining the theory of the uniform law of large numbers with standard Monte Carlo simulations.

1.2. A brief history of the literature

In recent times, the theory of measures of risk has generated a lot of interest in the mathematical finance literature, partly because it makes a rigorous assessment of risks associated with random financial net worths, and partly because it generalizes no-arbitrage asset pricing and superhedging ideas in incomplete markets.

One of the first articles to define and study such measures is the seminal paper [1], which provides a definition and justifies a unified framework for analysis, construction and implementation of *measures of risk*. As the authors point out, these measures of risks, named *coherent* measures, can be used as extra capital requirements, to regulate the risk assumed by market participants, traders, insurance underwriters, as well as to allocate existing capital. The idea is twofold: first to stipulate axioms which define *acceptable* future random net worths, and secondly, to define the measure of risk of an unacceptable position as the minimum extra capital which, invested in a 'pre-specified reference investment instrument', makes the future discounted value of the position acceptable. The axioms defining acceptability do not specify a unique measure of risk; instead, they characterize a large class of risk measures. The choice of precisely which measure to use from this class has to be determined from additional economic considerations.

A significant extension was made by introducing convex measures of risk in [10]. A similar set-up, as in [1], is considered. However the authors argue that the positive homogeneity of the coherent risk measure is an undue requirement, because the risk of a position might increase in a non-linear way with the size of the position. They suggest relaxing the conditions of positive homogeneity and of subadditivity and to require the weaker property of convexity.

In both papers, the basic objects of study are random variables on the set of *states of nature* at a future date, interpreted as possible future (discounted) values of positions or portfolios currently held. A supervisor (e.g. regulator, exchange's clearing firm, or investment manager) decides on a subset of such future outcomes as *acceptable risks*. In other words, they choose a subset $\mathcal{A}$ of a suitable set of real functions, $\mathbf{L}^0$, on a set Ω, and call it the acceptance set. A measure of risk associated with $\mathcal{A}$ is a function $\rho_{\mathcal{A}} : \mathbf{L}^0 \to \mathbb{R}$, defined by

$$\rho_{\mathcal{A}}(X) \triangleq \inf\{m \mid m + X \in \mathcal{A}\}.$$

Conversely, for any function $\rho : \mathbf{L}^0 \to \mathbb{R}$, one can define a corresponding acceptance set by $\mathcal{A}_\rho \triangleq \left\{X \in \mathbf{L}^0 \mid \rho(X) \le 0\right\}$. Such a function, ρ, will be called a *convex measure of risk* if it satisfies the following axioms:

- *Translation invariance:* for all $X \in \mathbf{L}^0$, and $a \in \mathbb{R}$, we have $\rho(X + a) = \rho(X) - a$.
- *Monotonicity:* for all X and Y in $\mathbf{L}^0$ with $X \ge Y$, we have $\rho(X) \le \rho(Y)$.
- *Convexity:* for all X and Y in $\mathbf{L}^0$, and all $\lambda \in [0, 1]$, we have

$$\rho(\lambda X + (1 - \lambda)Y) \le \lambda\rho(X) + (1 - \lambda)\rho(Y). \tag{6}$$

Why these axioms are natural requirements for a measure of risk has been argued in [1, Section 2.2] and [10], and we skip such details.

The authors of [10] then prove a representation theorem, similar in spirit to one in [1], which shows that any convex measure of risk on a finite Ω is of the form

$$\rho(X) = \sup_{P \in \mathcal{P}} (E^P[-X] + \alpha(P)). \tag{7}$$

Here, the set $\mathcal{P}$ is the set of all probability measures on Ω. The function $\alpha(\cdot)$ is a certain *penalty function* on $\mathcal{P}$ which takes values in $\mathbb{R} \cup \{-\infty\}$. (Here we stray from the usual convention where α in (7) is replaced by $-\alpha$.) Representation (7) was independently proved by David Heath in [13]. As before, a convex measure of risk defines an associated acceptance set given by

$$\mathcal{A}_\rho = \{X \in \mathbf{L}^0 \mid \rho(X) \le 0\} = \{X \in \mathbf{L}^0 \mid E^P[X] \ge \alpha(P)\}. \tag{8}$$

Broad extensions of (7) can be found in [11], all of which exhibit the same structure.

Notions similar to the above started appearing simultaneously from very different contexts. In a now well-known paper, [4], the authors use the notion of acceptability to present a new approach for positioning, pricing, and hedging in incomplete markets that bridges standard arbitrage pricing and expected utility maximization. Also the theory of *no-good-deal pricing* (NGD), as a pricing technique based on the absence of attractive investment opportunities in equilibrium, was introduced in [5]. The description 'no-good-deal' is borrowed from an earlier paper with similar objectives, [6], where good deals were defined by a high sharp ratio of returns. The first paper which fully establishes the link between coherent risk measures and the NGD pricing theory is [14], which shows that convex risk measures are essentially equivalent to good-deal bounds. Relations between measures of risk and NGD are further extended by Staum in [17], where he proves fundamental theorem of asset pricing for good-deal bounds in incomplete markets.

All these diverse motivations can be assimilated by considering what the authors of [10] call *measure of risk in a financial market*. Several authors have recently contributed to the development of this theory, e.g., the authors of [2,3], who establish these risk measures as special cases of *inf-convolution* of risk measures. Consider the setting in the last subsection, in particular, the notation in (3) and (4). The minimum w_0 for which $\inf_\xi \rho(V(w_0, \xi))$ is non-positive can be thought of as a price one has to pay today for achieving acceptability in future. As is shown in [10], for any random variable Z, one can choose the penalty function suitably such that the minimum w_0 is the market measure of risk of Z. This duality between price and risk measures is also seen in NGD pricing. If a strategy ξ exists which achieves the infimum above, then, it can be thought of as a hedging strategy in the NGD setting. In any case, it can be thought as a strategy for achieving acceptability in the future, starting from a currently non-acceptable portfolio.

Our risk measure, ρ in (2), is clearly convex. We restrict ourselves to finite sets $\{Q_1, \ldots, Q_m\}$ to aid computation. This can be interesting either in its own right, or as an approximation to the general case. The assumption $Q_i \ll P$ is implied by the natural requirement: $\rho(X) = \rho(Y)$ if $P(X = Y) = 1$ (see, e.g., [11]).

1.3. Summary and organization

We propose our main result in the following section. First, we suppose that for a given w_0, the set of strategies ξ which satisfy (5), and for which $\rho(V(w_0, \xi)) \leq 0$, is non-empty. Then, Proposition 2.1 proves that the intersection of this set with a specific, much smaller family of strategies is also non-empty. This smaller set of strategies is indexed by a finite-dimensional space, and has nice combinatorial properties. This allows us to use the theory of the Uniform Law of Large Numbers (ULLN), and devise a Monte Carlo scheme for numerically computing a near-minimum w_0 and a corresponding strategy ξ for having $\rho(V(w_0, \xi))$ non-positive. In Section 3, we describe the method, and give precise error bounds on such approximations. In Section 4, we consider a natural example in which stock price follows discrete geometric Brownian motion, and show how our method leads to numerical values of both near-optimal capital and strategy for achieving acceptability.

2. Main results

Recall that m refers to the number of probability measures in the representation of ρ in (2). Let $\mathcal{L}$ be the collection of adapted processes $\xi = (\xi_0, \ldots, \xi_{T-1})$ which satisfy (5). Define the following set:

$$\mathcal{R} \stackrel{\triangle}{=} \{(E^{Q_1}(W(\xi)), \ldots, E^{Q_m}(W(\xi))), \xi \in \mathcal{L}\}. \tag{9}$$

The set $\mathcal{R}$ is hence the set of all possible expected discounted gains in wealth when a strategy is chosen within the restrictions (5).

For any $k \in \mathbb{N}$ and any $x \in \mathbb{R}^k$, define the *upper quantant* of x, denoted by Q_x, as the set

$$Q_x = \{y \in \mathbb{R}^k : y_j \ge x_j \text{ for } j = 1, 2, \ldots, k\}. \tag{10}$$

The dimension is suppressed in the notation for Q_x, since it is obvious from the dimension of x.

Proposition 2.1. *Fix a $w_0 \in \mathbb{R}$. Let $z_0 = (\alpha_1 - w_0, \ldots, \alpha_m - w_0)$. Assume that the convex set $Q_{z_0} \cap \mathcal{R}$ has a non-empty relative interior.*

For every $1 \le i \le m$, define the adapted sequence of random variables

$$v_t(f_i) \stackrel{\triangle}{=} (b_t - a_t)E\left[(S_{t+1} - S_t)f_i \mid \mathcal{F}_t\right], \quad t = 0, 1, \ldots, T - 1. \tag{11}$$

For every $\mathbf{r} \in \mathbb{R}^m$, consider the following weighted sum process:

$$\lambda_t(\mathbf{r}) \stackrel{\triangle}{=} \sum_{i=1}^{m} r_i v_t(f_i), \quad t = 0, \ldots, T - 1. \tag{12}$$

Now, let η be any continuous probability distribution function on the real line with finite first moment. Then, there exists a vector $\mathbf{r}^* \in \mathbb{R}^m$ such that the $\{\mathcal{F}_t\}$-adapted process

$$\xi_t^*(\omega) \stackrel{\triangle}{=} (b_t - a_t)\eta(\lambda_t(\mathbf{r}^*)) + a_t, \quad t = 0, \ldots, T - 1, \tag{13}$$

satisfies (5) and $\rho(W(\xi^*)) \le w_0$.

Remark 1. Note that assuming $Q_{z_0} \cap \mathcal{R}$ being non-empty is equivalent to assuming the existence of a strategy ξ such that $\rho(w_0 + W(\xi)) \le 0$. In the above proposition we assume a bit more than that.

Remark 2. The process $v_t(f_i)$ and the λ_t can be interpreted in the following way. Suppose m is one. Then our objective can be compared with the problem of maximizing $E^Q(W(\xi))$. The way to achieve this is economically natural. If $E^Q(S_{t+1} \mid \mathcal{F}_t) > S_t$, then we should buy shares, but if $E^Q(S_{t+1} \mid \mathcal{F}_t) < S_t$, then we should sell. When $m > 1$, it is intuitive that one should look at linear combinations of *expected increments*

$$E^{Q_i}(S_{t+1} - S_t \mid \mathcal{F}_t), \quad i = 1, 2, \ldots, m,$$

to make a decision at time t. Now, note that, by a change of measure

$$E\left[(S_{t+1} - S_t)f_i \mid \mathcal{F}_t\right] = E^{Q_i}\left[(S_{t+1} - S_t) \mid \mathcal{F}_t\right] \times E^{Q_i}(f_i \mid \mathcal{F}_t).$$

So, perhaps it is not so surprising that ξ_t^* turns out to be a function of λ_t which is a linear combination of $v_t(f_i)$.

The proof of Proposition 2.1 will follow after we have introduced some notation. Let $[T]$ denote the set $\{0, 1, \ldots, T - 1\}$. Enlarge the original sample space by considering

$$\Omega \times [T] = \Omega \times \{0, 1, \ldots, T - 1\}. \tag{14}$$

Let $\mathcal{P}^{[T]}$ be the power set of the finite collection $\{0, 1, \ldots, T - 1\}$ and let $\mathcal{F} \otimes \mathcal{P}^{[T]}$ denote the product σ-algebra of $\mathcal{F}$ and $\mathcal{P}^{[T]}$. Extract a sub-σ-algebra $\widehat{\mathcal{F}}$ by defining

$$\widehat{\mathcal{F}} \stackrel{\triangle}{=} \{A \in \mathcal{F} \otimes \mathcal{P}^{[T]} \mid \{\omega : (\omega, t) \in A\} \in \mathcal{F}_t, \forall t = 0, 1, \ldots, T - 1\}. \tag{15}$$

That $\widehat{\mathcal{F}}$ is a valid σ-algebra is straightforward to verify. Finally, let U_T denote the discrete uniform measure on $[T]$, and consider the product measure $P \otimes U_T$ on the σ-algebra $\widehat{\mathcal{F}}$. This gives us a probability space $\left(\Omega \times [T], \widehat{\mathcal{F}}, P \otimes U_T\right)$. The advantages of considering the above probability space is the following trivial lemma.

Lemma 2.1. *A process $(h_0, h_1, \ldots, h_{T-1})$ is adapted with respect to $(\Omega, \mathcal{F})$ if and only if the random variable $H(\omega, t) = h_t(\omega)$ is measurable with respect to the enlarged space $(\Omega \times [T], \widehat{\mathcal{F}})$.*

Proof. Follows from the definition of $\widehat{\mathcal{F}}$. □

For all sequences $\{\xi_t\}$ that satisfy (5) (i.e. $\xi \in \mathcal{L}$), let us make a change of variable $\phi = \pi(\xi)$, where

$$\phi_t = \pi(\xi)_t \overset{\triangle}{=} (\xi_t - a_t)/(b_t - a_t), \quad t = 0, \ldots, T-1, \tag{16}$$

then, each ϕ_t is $\mathcal{F}_t$-measurable and $P(0 \le \phi_t \le 1) = 1$.

Now, the discounted gained value of the portfolio in (4) can be expressed in terms of the $\phi = \pi(\xi)$ as

$$\begin{aligned} W(\xi) = W \circ \pi^{-1}(\phi) &= \sum_{t=0}^{T-1} (S_{t+1} - S_t)\,[(b_t - a_t)\phi_t + a_t] \\ &= \sum_{t=0}^{T-1} \left[(b_t - a_t)\,(S_{t+1} - S_t)\,\phi_t + a_t(S_{t+1} - S_t)\right]. \end{aligned} \tag{17}$$

Thus, for any suitably integrable f defined on $(\Omega, \mathcal{F}, P)$, one can write

$$\begin{aligned} \int W(\xi) f \mathrm{d}P &= E\,(W(\xi) f) = \sum_{t=0}^{T-1} E\left(\left[(b_t - a_t)\,(S_{t+1} - S_t)\,\phi_t + a_t(S_{t+1} - S_t)\right] f\right) \\ &= \sum_{t=0}^{T-1} E\left[(b_t - a_t)\,(S_{t+1} - S_t)\,\phi_t f\right] + \sum_{t=0}^{T-1} E\left[a_t(S_{t+1} - S_t) f\right] \\ &= \sum_{t=0}^{T-1} E\,[v_t(f)\phi_t] + c(f), \end{aligned} \tag{18}$$

where we have defined

$$v_t(f)(\omega) \overset{\triangle}{=} (b_t - a_t) E\left[(S_{t+1} - S_t)\, f | \mathcal{F}_t\right](\omega), \quad \text{and} \tag{19a}$$

$$c(f) = E\left[f \sum_{t=0}^{T-1} a_t(S_{t+1} - S_t)\right]. \tag{19b}$$

For $t \in [T]$, if we now look at ϕ and v as functions of two arguments (ω, t), i.e.,

$$\phi(\omega, t) \overset{\triangle}{=} \phi_t(\omega), \qquad v(f)(\omega, t) \overset{\triangle}{=} v_t(f)(\omega), \quad \omega \in \Omega, \tag{19c}$$

then, by Lemma 2.1, both ϕ and v are $\widehat{\mathcal{F}}$-measurable functions on $\Omega \times [T]$. Moreover, $v(f)$ is $P \otimes U_T$-integrable and $P \otimes U_T(\{0 \le \phi \le 1\}) = 1$. Thus, from (18), we can write

$$\int W \circ \pi^{-1}(\phi) f \mathrm{d}P - c(f) = \sum_{t=0}^{T-1} \int v_t(f)\phi_t \mathrm{d}P = T \int_{\Omega \times [T]} \phi v(f) \mathrm{d}(P \otimes U_T). \quad (20)$$

Proof of Proposition 2.1. Let η be any continuous probability distribution function on the real line with finite first moment. Consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \eta)$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. Consider the following product space:

$$\Omega \times [T] \times \mathbb{R}, \qquad \widehat{\mathcal{F}} \otimes \mathcal{B}(\mathbb{R}), \qquad P \otimes U_T \otimes \eta. \quad (21)$$

Let us recall here that $\Omega \times [T]$, and $\widehat{\mathcal{F}}$ are defined in (14) and (15), and U_T is the discrete uniform measure on the set $[T] = \{0, 1, 2, \ldots, T-1\}$. Let Z be a measurable map from this product space to $\mathbb{R}$, given by

$$Z(\omega, t, x) = x, \quad \omega \in \Omega, t \in [T], x \in \mathbb{R}.$$

Clearly, Z has distribution η, independent of the σ-algebra $\widehat{\mathcal{F}}$.

Consider the functions $f_i = \mathrm{d}Q_i/\mathrm{d}P$ appearing in (2), and define the following functions in $\mathbf{L}^1(P \otimes U_T \otimes \eta)$:

$$g_i(\omega, t, x) \stackrel{\triangle}{=} v(f_i)(\omega, t), \quad 1 \le i \le m, \qquad g_{m+1} \stackrel{\triangle}{=} -Z, \quad (22)$$

where the function v is defined in (19a) and (19c). Also define the constants

$$\gamma_i \stackrel{\triangle}{=} (\alpha_i - w_0 - c(f_i))/T, \quad i = 1, 2 \ldots, m.$$

The function c is defined above in (19b).

• Define Φ to be the convex collection of all $\widehat{\mathcal{F}} \otimes \mathcal{B}(\mathbb{R})$-measurable functions ϕ such that $P \otimes U_T \otimes \eta\ (0 \le \phi \le 1) = 1$. Let $\mathcal{M}$ denote the set of points

$$\mathcal{M} \stackrel{\triangle}{=} \left\{ \left(\int \phi g_1, \ldots, \int \phi g_{m+1} \right), \phi \in \Phi \right\},$$

where the integrations are with respect to $P \otimes U_T \otimes \eta$. Recall the assumption in the statement of the proposition that $Q_{z_0} \cap \mathcal{R}$ has a non-empty relative interior. Since every strategy $\xi \in \mathcal{L}$ corresponds to a ϕ by the linear mapping π defined in (16), it follows that there is a point $(q_1, \ldots, q_m)$ which is an interior point of $\mathcal{M} \cap Q_\gamma$.

We look at the following maximization problem: find the maximizer of

$$\int_{\Omega \times [T] \times \mathbb{R}} \phi g_{m+1} \mathrm{d}(P \otimes U_T \otimes \eta)$$

among all $\phi \in \mathcal{A} \subseteq \Phi$, where $\mathcal{A}$ is defined by

$$\mathcal{A} \stackrel{\triangle}{=} \left\{ \phi \in \Phi | \int_{\Omega \times [T] \times \mathbb{R}} \phi g_i \mathrm{d}(P \otimes U_T \otimes \eta) = q_i \right\}. \quad (23)$$

We use Theorem 5 on page 96 of [15]. Part (iv) of this theorem guarantees the existence of a solution ϕ^* of the above maximization problem which is of the form

$$\phi^* = \begin{cases} 1, & \text{if } \sum_{i=1}^m r_i v(f_i) + g_{m+1} > 0, \\ 0, & \text{if } \sum_{i=1}^m r_i v(f_i) + g_{m+1} < 0, \end{cases} \tag{24}$$

for some $(r_1, \ldots, r_m) \in \mathbb{R}^m$. Recall that $g_{m+1} = -Z$, and that from the definition of the function v in (19c), it is clear that each $v(f_i)$ is independent of Z. Thus

$$(P \otimes U_T \otimes \eta) \left(\sum_{i=1}^m r_i v(f_i) = Z \right) = \int \eta \left(Z = \sum_{i=1}^m r_i v(f_i) | \widehat{\mathcal{F}} \right) \mathrm{d}\,(P \otimes U_T) = 0,$$

the integrand being zero being the consequence of the continuity of η. Thus, the solution in (24) is actually

$$\phi^* = \begin{cases} 1 & \text{if } \sum_{i=1}^m r_i v(f_i) > Z \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Now from the constraint $\phi^* \in \mathcal{A}$, we also get that $\int \phi^* g_i \mathrm{d}\,(P \otimes U_T \otimes \eta) = q_i \geq \gamma_i, i = 1, 2, \ldots, m$. In other words, $\int \phi^* v(f_i) \mathrm{d}\,(P \otimes U_T \otimes \eta) \geq \gamma_i$, or, by expanding γ_i, equivalently

$$T \int_{\Omega \times [T]} E\left(\phi^* | \widehat{\mathcal{F}}\right) v(f_i) \mathrm{d}\,(P \otimes U_T) + c(f_i) \geq \alpha_i - w_0, \quad i = 1, \ldots, m. \tag{26}$$

We have used Fubini's theorem above, where

$$\begin{aligned} E\left(\phi^* | \widehat{\mathcal{F}}\right)(\omega, t) &= (P \otimes U_T \otimes \eta) \left(\sum_{i=1}^m r_i v(f_i) - Z > 0 | \widehat{\mathcal{F}} \right)(\omega, t) \\ &= \eta \left(\sum_{i=1}^m r_i v_t(f_i)(\omega) \right) = \eta\left(\lambda_t(r)\right)(\omega), \end{aligned}$$

where the $\{\mathcal{F}_t\}$-adapted process $\{\lambda_t\}$ is defined as in (12). Thus, if we let

$$\xi_t^* \triangleq (b_t - a_t)\eta(\lambda_t) + a_t, \quad t = 0, \ldots, T - 1,$$

then, $\xi^* = \pi^{-1}(E[\phi^* | \widehat{\mathcal{F}}])$ in the notation of (16). Thus, by (26) and (20), we conclude that

$$\int W(\xi^*) f_i \mathrm{d}P \geq \alpha_i - w_0, \quad i = 1, \ldots, m,$$

or in other words, $\rho(w_0 + W(\xi^*)) \leq 0$. This proves the proposition. □

3. Computations

For every $s = (s_1, \ldots, s_m) \in \mathbb{R}^m$, recall from Proposition 2.1, the $\mathcal{F}_t$-adapted process

$$\lambda_t(s) \triangleq \sum_{i=1}^m s_i v_t(f_i), \quad t = 0, 1, \ldots, T - 1, \tag{27}$$

and the derived process

$$\xi_t(s) \stackrel{\triangle}{=} (b_t - a_t)\eta(\lambda_t(s)) + a_t, \quad t = 0, \ldots, T-1. \tag{28}$$

For suitable w_0, Proposition 2.1 proves the existence of an $s \in \mathbb{R}^m$ via which the process $\xi(s)$ satisfies $\rho(w_0 + W(\xi(s))) \leq 0$, or equivalently, by translation invariance, $\rho(W(\xi(s))) \leq w_0$.

What we shall do now is like a partial converse. Suppose we can compute $\rho(W(\xi(s)))$ for every $s \in \mathbb{R}^m$. Then we can define w_0 by

$$w_0 := \inf_s \rho(W(\xi(s))).$$

If s^* achieves the above infimum, then, clearly $\rho(w_0 + W(\xi(s^*))) \leq 0$, and w_0 is near minimal by Proposition 2.1.

The above procedure would work if we could theoretically compute $\rho(W(\xi(s)))$ for every $s \in \mathbb{R}^m$. This is often impossible. However, for any fixed s, we can estimate $\rho(W(\xi(s)))$ via Monte Carlo simulations up to any desired level of accuracy. We show in this section that it is possible to carry out a Monte Carlo simulation to simultaneously approximate $\rho(W(\xi(s)))$ for *every* $s \in \mathbb{R}^m$ with a uniform error bound. The feasibility of our claim depends on the theory of the *uniform law of large numbers* and the related criteria of finite Vapnik–Červonenkis dimension which is a combinatorial property satisfied by the particular structure of $\{\xi_t\}$ in (28). This theory is well developed and we cherry-pick only the results necessary for our purpose. These have been stated in the appendix. Further references have also been provided for the interested reader.

Once we have our estimation procedure ready, we can construct a finite mesh $\mathbb{G}$ within $\mathbb{R}^m$ and approximate the value of $\rho(W(\xi(r)))$, by (say) $\hat{\rho}(r)$, for every $r \in \mathbb{G}$. Note that the error in approximation in our Monte Carlo procedure does not depend on the size of the grid, and we can make it as large and fine as we want. For that fine mesh $\mathbb{G}$, let r^* be a grid point which attains $\hat{\rho}(r^*) = \min_{r \in \mathbb{G}} \hat{\rho}(r)$.

Let $w_0^* = \hat{\rho}(r^*)$. Then, as we describe below, given any $\epsilon, \delta > 0$, with a very high probability greater than $(1 - \delta)$, the choice of $(w_0^*, \xi(r^*))$ satisfies

$$\rho(w_0^* + W(\xi(r^*))) \leq \epsilon.$$

This gives a near-minimal initial capital for the problem of finding (w_0, ξ) which satisfies (5) and $\rho(w_0 + W(\xi)) \leq \epsilon$.

Central to computing $\rho(W(\xi(s)))$, for any $s \in \mathcal{S}_{m+1}$, is computing $E(W(\xi(s)) \cdot f_i)$ for every f_i that defines ρ. Now, from Eq. (18), we can write

$$\begin{aligned} E(W(\xi(s))f_i) &= \sum_{t=0}^{T-1} \int_\Omega v_t(f_i)\eta(\lambda_t(s))\mathrm{d}P + c(f_i) \\ &= T \int_{\Omega \times [T]} \eta(\lambda(s))v(f_i)\mathrm{d}\,(P \otimes U_T) + c(f_i). \\ &= T \int_{\Omega \times [T] \times \mathbb{R}} \mathbb{I}\{\lambda(s) - Z > 0\}\, v(f_i)\mathrm{d}\,(P \otimes U_T \otimes \eta) + c(f_i). \end{aligned} \tag{29}$$

Here, as in the last section, Z is a random variable with law η independent of $\widehat{\mathcal{F}}$, and $\mathbb{I}\{\cdot\}$ denotes the indicator of an event.

We would now like to make a change of measure in (29) above with $v(f_i)$ as the 'Radon–Nikodým' derivative. This is not possibly directly, since $v(f_i)$ is not necessarily positive.

However, we can work separately with $v^+(f_i) = \max(v(f_i), 0)$ and $v^-(f_i) = \max(-v(f_i), 0)$, which denote the positive and the negative parts respectively. Hence, one obtains

$$\begin{aligned} E(W(\xi(s)) f_i) - c(f_i) &= T \int_{\Omega \times [T] \times \mathbb{R}} \mathbb{I}\{\lambda(s) - Z > 0\} v^+(f_i) \mathrm{d}(P \otimes U_T \otimes \eta) \\ &\quad - T \int_{\Omega \times [T] \times \mathbb{R}} \mathbb{I}\{\lambda(s) - Z > 0\} v^-(f_i) \mathrm{d}(P \otimes U_T \otimes \eta) \\ &= d_i^+ \cdot \left(\mu_i^+ \otimes \eta\right) \{\lambda(s) - Z > 0\} \\ &\quad - d_i^- \cdot \left(\mu_i^- \otimes \eta\right) \{\lambda(s) - Z > 0\}. \end{aligned} \tag{30}$$

Here we have introduced several probability measures on $(\Omega \times [T], \widehat{\mathcal{F}})$, defined by their corresponding *unnormalized* Radon–Nikodým derivatives:

$$\mathrm{d}\mu_i^+ / \mathrm{d}(P \otimes U_T) \propto v^+(f_i), \qquad \mathrm{d}\mu_i^- / \mathrm{d}(P \otimes U_T) \propto v^-(f_i), \tag{31a}$$

and the corresponding normalizing constants (multiplied by T):

$$d_i^+ \triangleq \sum_{t=0}^{T-1} E[v_t^+(f_i)], \qquad d_i^- \triangleq \sum_{t=0}^{T-1} E[v_t^-(f_i)], \quad i = 1, 2, \ldots, m. \tag{31b}$$

If any of the constants in (31b) is zero, the corresponding measure becomes the zero measure and can be dropped from our analysis. For efficiency in computation we would like to keep track of the number of non-zero measures above by defining

$$\aleph \triangleq \sum_{i=1}^{m} \left(1_{\{d_i^+ > 0\}} + 1_{\{d_i^- > 0\}} \right). \tag{32}$$

Assumption 3.1. Throughout the rest of this section, we shall assume that

(1) one can generate samples from the joint distribution of $(S_0, S_1, \ldots, S_T)$,
(2) the random variables $v_t(f_i)$ (and thus also λ_t) can be evaluated given the values of $(S_0, \ldots, S_T)$, and
(3) the constants $c(f_i)$, d_i^+ and d_i^- can be evaluated for every $1 \le i \le m$.

Now, by (30), evaluating $E(W(\xi(s)) f_i)$ boils down to evaluating the following two probabilities:

$$\left(\mu_i^+ \otimes \eta\right) \{\lambda(s) - Z > 0\}, \quad \text{and} \quad \left(\mu_i^- \otimes \eta\right) \{\lambda(s) - Z > 0\}, \quad s \in \mathbb{R}^m. \tag{33}$$

Instead, we use the Vapnik–Červonenkis theory, described in Appendix A.1, to set up a Monte Carlo scheme to estimate them for all $s \in \mathbb{R}^m$ with uniform precision. The key to this is observing the trivial equality

$$\{\lambda(s) - Z > 0\} = \left\{ \sum_{j=1}^{m} s_j v(f_j) - Z > 0 \right\} \tag{34}$$

and applying Dudley's theorem, Theorem A.1 in Appendix A, with $X = \Omega \times [T] \times \mathbb{R}$ and the vector space G to be linear space spanned by Z and $v(f_j)$, $j = 1, 2, \ldots, m$. Thus we infer that the collection of sets

$$\left\{\left\{\tilde{\omega} \in \Omega \times [T] \times \mathbb{R} : \sum_{j=1}^{m} r_j v(f_j)(\tilde{\omega}) + r_{m+1} Z(\tilde{\omega}) > 0\right\}, r \in \mathbb{R}^{m+1}\right\}, \tag{35}$$

has a VC dimension not more than $(m + 1)$. From (34), the collection of sets

$$\left\{\{\lambda(s) - Z > 0\}, s \in \mathbb{R}^m\right\}$$

is contained in (35), and hence also has a VC dimension not more than $(m + 1)$. It is hence possible to estimate the probabilities in (33), uniformly for all $s \in \mathbb{R}^m$, by drawing independent samples from distributions $\mu_i^+ \otimes \eta$ and $\mu_i^- \otimes \eta$.

Our aim now would be to apply Theorem A.3. We first have to choose two positive parameters, ϵ and δ, determining the precision of our estimates. Now, for every $i = 1, 2, \ldots, m$, choose κ_i^+ such that

$$4(\kappa_i^+)^{2(m+1)} \exp\left(-2\kappa_i^+ \left(\frac{\epsilon}{d_i^+}\right)^2 + 4\left(\frac{\epsilon}{d_i^+}\right) + 4\left(\frac{\epsilon}{d_i^+}\right)^2\right) \le \delta. \tag{36}$$

Generate κ_i^+ many iid samples $\{(\omega_j, t_j, z_j) \in \Omega \times [T] \times \mathbb{R},\ j = 1, 2, \ldots, \kappa_i^+\}$, from the joint distribution $\mu_i^+ \otimes \eta$.

Remark 3. It is fairly standard to generate samples from measures μ_i^+, defined through their unnormalized densities given in (31a). We can either directly identify the distribution, as we do in the next section. Or, under the assumption that one can generate perfect samples from the underlying distribution $(P \otimes U_T)$, one can use any of the standard Markov chain algorithms, from the simple rejection sampling, to the general Metropolis–Hastings algorithm to generate samples from μ_i^+. Several books, e.g. see [12, Chap. 11], describe the details of all these algorithms.

Let $\mathcal{E}_i^+(\cdot)$ denote the empirical estimates of probabilities from the sample frequency. For example, for any $s \in \mathbb{R}^m$, we have

$$\mathcal{E}_i^+\{\lambda_s - Z > 0\} = \frac{1}{\kappa_i^+} \sum_{j=1}^{\kappa_i^+} \mathbb{I}\{\lambda_{t_j}(s)(\omega_j) - z_j > 0\}. \tag{37}$$

We can now apply (A.4) from Theorem A.3 to claim that under the joint distribution of all the κ_i^+ many samples drawn,

$$\text{Prob}\left\{\sup_{s\in\mathbb{R}^m} d_i^+ |\mathcal{E}_i^+\{\lambda_s - Z > 0\} - (\mu_i^+ \otimes \eta)\{\lambda_s - Z > 0\}| > \epsilon\right\} \le \delta, \quad \forall i. \tag{38}$$

In exactly the same way, one can replace the μ_i^+ by μ_i^- above, compute κ_i^- by

$$4(\kappa_i^-)^{2(m+1)} \exp\left(-2\kappa_i^- \left(\frac{\epsilon}{d_i^-}\right)^2 + 4\left(\frac{\epsilon}{d_i^-}\right) + 4\left(\frac{\epsilon}{d_i^-}\right)^2\right) \le \delta, \tag{39}$$

and obtain estimates $\mathcal{E}_i^-$, analogous to (37), which satisfy

$$\text{Prob}\left\{\sup_{s\in\mathbb{R}^m} d_i^-|\mathcal{E}_i^-\{\lambda_s - Z > 0\} - (\mu_i^- \otimes \eta)\{\lambda_s - Z > 0\}| > \epsilon\right\} \le \delta, \quad \forall i. \tag{40}$$

From (38) and (40), it follows, by using (30), that one can estimate the quantity $E(-W(\xi(s))f_i) + \alpha_i$ using

$$\mathcal{D}_i(s) \stackrel{\triangle}{=} -d_i^+\mathcal{E}_i^+\{\lambda(s) - Z > 0\} + d_i^-\mathcal{E}_i^-\{\lambda(s) - Z > 0\} - c(f_i) + \alpha_i. \tag{41}$$

Since $\rho(W(\xi(s))) = \sup_{1\le i\le m}\{E(-W(\xi(s))f_i) + \alpha_i\}$, it follows that a good estimate of $\rho(W(\xi(s)))$ is

$$\hat{\rho}(s) \stackrel{\triangle}{=} \sup_i \mathcal{D}_i(s).$$

We can sum up this approximation with a simple union bound using (38) and (40) as follows.

Under the joint distribution of all the $\{\kappa_i^+, \kappa_i^-\}_{1\le i\le m}$ samples drawn from the distributions $\{\mu_i^+ \otimes \eta, \mu_i^- \otimes \eta\}_{1\le i\le m}$, one has

$$\text{Prob}\left\{\sup_{s\in\mathbb{R}^m} |\hat{\rho}(s) - \rho(W(\xi(s)))| \ge \epsilon\right\} \ge 1 - \aleph\delta.$$

Here, the number $\aleph$ ($\le 2m$) is described in (32). We use the number $\aleph$ and not the crude bound $2m$ to bring more efficiency into our estimate.

Now that we have estimated $\rho(W(\xi(s)))$ for every $s \in \mathbb{R}^m$ with uniform precision, we can carry out the grid searching procedure described at the beginning of this section. We minimize $\hat{\rho}(s)$ over the grid nodes (say $\mathbb{G}$) to obtain

$$w_0^* \stackrel{\triangle}{=} \inf_{s\in\mathbb{G}} \hat{\rho}(s) = \hat{\rho}(s^*).$$

Then, with a probability more than $(1 - \aleph\delta)$, we have

$$\rho(W(\xi(s^*))) \le \hat{\rho}(s^*) + \epsilon \le w_0^* + \epsilon.$$

In other words, with a high probability of being correct, we get

$$\rho(w_0^* + W(\xi(s^*))) \le \epsilon.$$

Thus one obtains a near-optimal pair (w_0, ξ) which satisfies (5) and $\rho(w_0 + W(\xi))$ is *almost* non-positive. The next section displays the entire method through an explicit example.

4. Examples

The previous theory is now applied to an explicit example where stock prices follow geometric Brownian motion, but observed only at finitely many time points.

We consider $T = 3$ and $\Omega = \mathbb{R}^T$, the σ-algebra $\mathcal{F}_t$ being generated by the first t coordinates of $\omega \in \Omega$. We take $\mathcal{F}_0$ to be the trivial σ-algebra $\{\emptyset, \Omega\}$. Take P to be the product probability measure of T many independent normal distributions with mean zero and variance one. In other words, we consider random variables $(Z_1, Z_2, \ldots, Z_T)$ such that each Z_i is independent and identically distributed as $N(0, 1)$. The discounted stock price movement, under P, is described by

$$S_0 = 4, \qquad S_{t+1} = S_t \exp\left[-\frac{1}{2} + Z_{t+1}\right], \quad t = 0, 1, \ldots, T-1. \tag{42}$$

In other words, we have

$$S_t = S_0 \exp\left[\sum_{i=1}^{t} Z_k - \frac{t}{2}\right], \quad t = 1, 2, \ldots, T. \tag{43}$$

However, the investor is not entirely certain of his or her modeling assumptions, and so considers other scenarios Q_1 and Q_2, where Q_1 and Q_2 are two probability measures defined on $(\Omega, \mathcal{F}_T)$ by

$$\text{under } Q_1, \quad Z_1, \ldots, Z_T \overset{iid}{\sim} N(1, 1),$$
$$\text{under } Q_2, \quad Z_1, \ldots, Z_T \overset{iid}{\sim} N(-1, 1).$$

For convenience we also introduce $Q_3 = P$.

Assume that various constraints dictate that his/her trading strategy is bounded between zero and one throughout, i.e., in the notation of (5), we have

$$a_t \equiv 0, \quad b_t \equiv 1, \quad \text{for all } 0 \le t \le T-1.$$

Now, the investor sets out to do the following: if the conditions are favorable, and the stock prices tend to go up under Q_1, he/she wants a large lower bound e^4 for his/her expected terminal wealth. On the other hand, if the stock prices tend to go down, under Q_2, he/she sets a lower bound for his/her expected losses, by setting that his/her final expected wealth should be more than e^{-1}. He/she has at least \$0.2 to invest, and would like to know an optimal initial capital, and a trading strategy for achieving his/her goals.

This requires us to define a measure of risk ρ: if X is measurable with respect to $\mathcal{F}_T$, then

$$\rho(X) \triangleq \max_{i=1,2,3} [E^{Q_i}(-X) + \alpha_i], \quad m = 3,$$

where

$$\alpha_1 = e^4, \qquad \alpha_2 = e^{-1}, \qquad \alpha_3 = 0.2.$$

Then, we would like to compute a near-optimal pair (w_0, ξ) of initial capital w_0 and $0 \le \xi_t \le 1$, for all $0 \le t \le T-1$, such that

$$\rho(w_0 + W(\xi)) \le 0 \Leftrightarrow w_0 + E^{Q_i}[W(\xi)] \ge \alpha_i, \quad i = 1, 2, 3.$$

Remark 4. Note, from (42), the effect of changing measure on the stock price movements. For Q_1, the geometric Brownian motion gets a positive drift, for Q_2 it gets a negative drift, while Q_3 is the same as P, where stock prices are a martingale. It is a known fact that computations regarding risk measures get more complicated if the convex hull of the scenario measures contain a martingale measure (see, e.g., [16]). This particular choice of the risk measure takes into account three widely different possible models, including one under which the price process is a martingale. Hence, we hope, it proves the point of being universally applicable.

Although the choice of this risk measure is somewhat arbitrary, this is probably almost always true for a choice of a specific risk measure. However, we would like to stress the fact that the choice of the models and the penalties do not affect anyway the computation scheme, although specific calculations sometimes get simplified.

The first step will be to compute the functions f_1, f_2, and f_3. They are straightforward since

$$\begin{aligned} f_1(z_1,\ldots,z_k) &= \mathrm{d}Q_1/\mathrm{d}P = \exp\left[\sum_{k=1}^{T} z_k - T/2\right] \\ f_2(z_1,\ldots,z_k) &= \mathrm{d}Q_2/\mathrm{d}P = \exp\left[-\sum_{k=1}^{T} z_k - T/2\right] \\ f_3(z_1,\ldots,z_k) &= \mathrm{d}Q_3/\mathrm{d}P \equiv 1. \end{aligned} \tag{44}$$

We can now compute the functions $v_t(f_i)$. These are given by

$$\begin{aligned} v_t(f_1) &= E\left[f_1(S_{t+1}-S_t)|\mathcal{F}_t\right] \\ &= S_t E\left[f_1\left(\exp(Z_{t+1}-1/2)-1\right)|\mathcal{F}_t\right], \quad \text{from (42)}, \\ &= S_t \exp\left(\sum_{k=1}^{t} Z_k\right) E\left(\exp\left\{\sum_{k=t+1}^{T} Z_k - T/2\right\}\left[\exp(Z_{t+1}-1/2)-1\right]\right), \end{aligned} \tag{45}$$

where the last equality is due to (44) and the independence of $\{Z_i\}$. Recall that if Z follows $N(0,1)$, then $E\left[\exp(\sigma Z)\right] = \exp(\sigma^2/2), \sigma \in \mathbb{R}$. Thus, for $\mathbf{z} = (z_1, z_2, \ldots, z_T) \in \Omega$, a straightforward computation leads to

$$\begin{aligned} v_t(f_1)(\mathbf{z}) &= S_t \exp\left[\sum_{k=1}^{t} z_k\right]\left\{\exp\left(1-\frac{t}{2}\right) - \exp\left(-\frac{t}{2}\right)\right\} \\ &= 4(\mathrm{e}-1)\exp\left\{2\sum_{1}^{t} z_k - t\right\}, \quad \text{by (43)}. \end{aligned} \tag{46}$$

In particular, we have $E(v_t(f_1)) = 4(\mathrm{e}-1)E\left[\exp\left(2\sum_{k=1}^{t} Z_k - t\right)\right] = 6.87\mathrm{e}^t$.

Similarly, we compute

$$\begin{aligned} v_t(f_2) &= E\left[f_2(S_{t+1}-S_t) \mid \mathcal{F}_t\right] \\ &= S_t E\left[f_2\left(\exp(Z_{t+1}-1/2)-1\right)|\mathcal{F}_t\right], \quad \text{from (42)}, \\ &= S_t \exp\left(-\sum_{k=1}^{t} Z_k\right) E\left(\exp\left\{-\sum_{k=t+1}^{T} Z_k - T/2\right\}\left[\exp(Z_{t+1}-1/2)-1\right]\right) \\ &= -S_0\exp(-t)\frac{\mathrm{e}-1}{\mathrm{e}} = -4(\mathrm{e}-1)\exp(-t-1). \end{aligned} \tag{47}$$

And obviously, since S_t is a martingale under Q_3, we have

$$v_t(f_3) = E\left[S_{t+1}-S_t \mid \mathcal{F}_t\right] = 0.$$

Hence, for $s = (s_1, s_2, s_3) \in \mathbb{R}^3$, the random variable $\lambda_t(s)$ is given by

$$\begin{aligned}\lambda_t(s) &= 4e^{-t}(e-1)\left[s_1 \exp\left\{2\sum_1^t z_k\right\} - s_2\exp(-1)\right]\\ &= 4e^{-t}(e-1)\left[s_1 e^t\left(\frac{S_t}{S_0}\right)^2 - s_2\exp(-1)\right]\\ &= 4(e-1)\left[s_1\left(\frac{S_t}{S_0}\right)^2 - s_2 e^{-t-1}\right].\end{aligned}$$

Thus, for $1 \le t \le 2$ and $\mathbf{z} = (z_1, z_2, z_3) \in \Omega$, we have the following table:

$$\begin{array}{llll} v^+(f_1)(t,\mathbf{z}) = v_t(f_1)(\mathbf{z}), & d_1^+ = 76.34, & v^-(f_1)(t,\mathbf{z}) = 0, & d_1^- = 0,\\ v^+(f_2)(t,\mathbf{z}) = 0, & d_2^+ = 0, & v^-(f_2)(t,\mathbf{z}) = 2.53e^{-t}, & d_2^- = 3.80,\\ v^+(f_3)(t,\mathbf{z}) = 0, & d_3^+ = 0, & v^+(f_3)(t,\mathbf{z}) = 0, & d_3^- = 0. \end{array}$$

From above and (32), we also have $\aleph = 2$. Clearly, we need to consider only two changes of measures, the one given by $v^+(f_1)$ and the other by $v^-(f_2)$. The rest are all zero measures. Finally, since $a_t \equiv 0$, from (19b), we get $c(f_i) = 0$, $i = 1, 2, 3$.

We take the precision parameters to be

$$\epsilon = 0.5, \qquad \delta = 0.05.$$

From (36) and (39), we determine that a sufficient number of samples for desired accuracy would be

$$\kappa_1^+ = 1{,}400{,}000, \qquad \kappa_2^- = 10{,}500.$$

Let us now analyze the probability measures μ_1^+ and μ_2^- on $\mathbb{R}^3 \times \{0, 1, 2\}$. If $\mathbf{z} \in \mathbb{R}^3$, and $0 \le t \le 2$, then from (31a) and (46) we get

$$\begin{aligned} d\mu_1^+(\mathbf{z}, t) &\propto v^+(f_1)(\mathbf{z}, t)\cdot d\,(P\otimes U_T)\,(\mathbf{z}, t)\\ &\propto \exp\left\{2\sum_1^t z_k - t\right\}\cdot\left(\frac{1}{\sqrt{2\pi}}\right)^3 \exp\left\{-\frac{1}{2}\sum_{k=1}^3 z_k^2\right\}\\ &\propto e^t\left(\frac{1}{\sqrt{2\pi}}\right)^3\exp\left\{-\frac{1}{2}\sum_{k=1}^t (z_k-2)^2 - \frac{1}{2}\sum_{k=t+1}^3 z_k^2\right\}. \end{aligned} \tag{48}$$

Thus generating a sample from μ_1^+ is the same as picking a $t \in (0, 1, 2)$ randomly with probability proportional to $\exp(t)$. Then, conditionally on t, we generate t independent samples $Z_1, \ldots, Z_t$ from $N(2, 1)$, and $3 - t$ samples from $N(0, 1)$.

Simulating from μ_2^- is even simpler, since, from (47), we get that

$$\begin{aligned} d\mu_2^-(\mathbf{z}, t) &\propto v^-(f_2)(\mathbf{z}, t)\cdot d\,(P\otimes U_T)\,(\mathbf{z}, t)\\ &\propto e^{-t}\cdot\left(\frac{1}{\sqrt{2\pi}}\right)^3\exp\left\{-\frac{1}{2}\sum_{k=1}^3 z_k^2\right\}. \end{aligned} \tag{49}$$

Here, we pick t from $\{0, 1, 2\}$ with probability proportional to $\exp(-t)$, and generate $(Z_1, \ldots, Z_T)$ as independent and identically distributed samples from $N(0, 1)$.

Finally, we take η to be $N(0, 1)$.

Result of simulations. We first generate the required number of samples from μ_1^+ and μ_2^- and set them aside. Now we choose a variety of grids, making them finer and more localized as we proceed, until $\hat{\rho}$ converges to a global minimum.

An estimate of the minimum capital is $w_0^* = 0.41$. The optimal grid point comes to $s_1 = 0.05, s_2 = 9.65$. Thus, an estimate of the trading strategy for this capital is $\xi_t^* = \Phi(\lambda_t)$, where Φ is the standard normal cumulative distribution function, and λ_t is the process given by

$$\lambda_t = 4(\mathrm{e} - 1)\left[0.05\left(\frac{S_t}{S_0}\right)^2 - 9.65\mathrm{e}^{-t-1}\right].$$

In other words, with a probability more than $(1 - \aleph\delta) = 0.9$, we will indeed have $\rho(w_0^* + W(\xi^*)) \leq \epsilon = 0.5$.

5. Conclusion

We devise a Monte Carlo algorithm for computing near-minimal initial capital requirement and a suitable trading strategy for achieving acceptability at a future date. The benefit of this approach is that it gives precise numerical values for portfolio optimization problems where purely theoretical methods (e.g. backward induction, linear programming) fail.

The primary shortcoming is that this approach requires intensive computing, mainly due to bound (A.4). However, the fault lies in the crudeness of the exact theoretical bound, and not in the method itself. In fact, there are better bounds (e.g. due to Talagrand [18]) which, unfortunately, lack exact constants.

A related problem (brought to the author's attention by Prof. Robert Jarrow at the CCCP conference, 2006) is the following. Suppose we have two risk measures ρ_1 and ρ_2 and an initial *constrained budget* w_0. Can we find a trading strategy ξ^* such that (w_0, ξ^*) minimizes ρ_1 among all strategies ξ for which $\rho_2(w_0, \xi)$ is non-positive? The author believes that the method in this paper can be suitably extended, and is currently involved in a project concerning this.

Acknowledgments

I thank Prof. Peter Bank for suggesting the particular example in Section 4. I am also grateful to Prof. Ioannis Karatzas for his several useful suggestions on the manuscript.

Appendix A

A.1. Uniform law of large numbers

We briefly mention here three basic theorems concerning the theory of the uniform law of large numbers and the related concept of Vapnik–Červonenkis dimensions. This is a subject in itself and we shall use very little of it for our purpose. Hence we shall skip all details and refer the reader to the excellent book [8, Chap. 12], from where our propositions in this section have been lifted.

Notation A.1. We consider a probability space $(\Theta, \Im, \varrho)$, where Θ is a complete, separable metric space. On Θ^n, let ϱ^n denote the product probability measure on the product σ-algebra. Similarly on $\Theta^\infty := \Theta^{\mathbb{N}}$, let ϱ^∞ denote the infinite product probability. For any $\theta \in \Theta^\infty$, and any $n \in \mathbb{N}$, define the random *empirical measure*: $\varrho_n(C) := 1/n \sum_{i=1}^n 1_{(\theta_i \in C)}$, $C \in \Im$, or, for any $\Im$-integrable function f, the corresponding random expectation $\varrho_n(f) := 1/n \sum_{i=1}^n f(\theta_i)$.

For any $C \in \Im$ and any $\epsilon > 0$, the law of large numbers dictates

$$\lim_{n\to\infty} \varrho^\infty(|\varrho_n(C) - \varrho(C)| > \epsilon) = 0. \tag{A.1a}$$

However, if we have a collection of $\{C_\alpha\}_{\alpha\in I}$ of sets in $\Im$, it is not always true that

$$\lim_{n\to\infty} \varrho^\infty(\sup_{\alpha\in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha)| > \epsilon) = 0. \tag{A.1b}$$

Equality above can be achieved under proper conditions on the collection $\{C_\alpha\}_{\alpha\in I}$, and then we say that the *Uniform Law of Large Numbers* (ULLN) holds. The Vapnik–Červonenkis theory provides one such condition. Its strength lies in the condition on $\{C_\alpha\}_{\alpha\in I}$ being combinatorial in nature, and hence independent from the choice of ϱ. (This sometimes can also be a weakness, since significant improvements can be made for specific choices of ϱ.) The theory begins with the concept of the *shatter coefficient*.

Definition A.1. Let $\{C_\alpha\}_{\alpha\in I}$ be a collection of $\Im$-measurable subsets of Θ. For $(\theta_1, \ldots, \theta_d) \in \Theta^d$, let $\mathcal{N}(\theta_1, \ldots, \theta_d)$ be the number of different sets in

$$\{\{\theta_1, \ldots, \theta_d\} \cap C_\alpha, \alpha \in I\}.$$

The dth shatter coefficient of the collection $\{C_\alpha\}_{\alpha\in I}$ is defined as

$$s_d \stackrel{\triangle}{=} \max_{(\theta_1,\ldots,\theta_d)\in\Theta^d} \mathcal{N}(\theta_1, \ldots, \theta_d).$$

In other words, the shatter coefficient is the maximal number of different subsets of d points that can be picked out by the class $\{C_\alpha\}_{\alpha\in I}$.

Remark 5. Note that we have deliberately suppressed mentioning the class $\{C_\alpha\}_{\alpha\in I}$ in the notation for the shatter coefficient. This is really for notational clarity. The shatter coefficient is clearly a property of the collection of sets that we consider.

The following theorem can be found in [8, Thm 12.5, p. 197].

Theorem A.1. *For any collection $\{C_\alpha\}_{\alpha\in I}$, and for any $n \in \mathbb{N}$, $\epsilon > 0$, we have*

$$\varrho^\infty\left\{\sup_{\alpha\in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha)| > \epsilon\right\} \le 8 s_n \exp(-n\epsilon^2/32), \tag{A.2}$$

where the constant s_n is the nth shatter coefficient of the collection $\{C_\alpha\}_{\alpha\in I}$ and is independent of the probability measure ϱ.

Hence (A.1b) will hold if the constant s_n grows at most polynomially. This is achieved for certain collections of sets which have a finite Vapnik–Červonenkis (VC) dimension. The following definition is from [8, p. 196].

Definition A.2. As before we consider the collection $\{C_\alpha\}_{\alpha\in I}$ of $\Im$-measurable subsets of Θ. The largest positive integer for which $s_d = 2^d$ is known as the VC dimension of the collection $\{C_\alpha\}_{\alpha\in I}$. If $s_d = 2^d$ for all integers $d \ge 1$, we then define the VC dimension to be ∞.

The next lemma [8, p. 218] describes a fundamental relationship between the VC dimension and the shatter coefficients.

Sauer's Lemma. *Let* $\{C_\alpha\}_{\alpha \in I}$ *be a subset of* $\Im$ *with finite VC dimension* $\mathcal{V} > 2$*. Then for all* $n > 2\mathcal{V}$*, we have* $s_n \le n^{\mathcal{V}}$*.*

Thus Theorem A.1 together with Sauer's lemma will yield the following.

Theorem A.2. *Let* $(\Theta, \Im)$ *be a measurable space. Let* $\{C_\alpha\}_{\alpha \in I}$ *be any collection of measurable subsets of* Θ *with a finite VC dimension* $\mathcal{V}$*. Then for any probability measure* ϱ *on* $(\Theta, \Im)$ *and any* $n \ge 2\mathcal{V}$*, we have*

$$\varrho^\infty \left\{ \sup_{\alpha \in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha)| > \epsilon \right\} \le 8n^{\mathcal{V}} \exp(-n\epsilon^2/32). \tag{A.3}$$

In particular, $\lim_{n\to\infty} \varrho^\infty \left\{ \sup_{\alpha \in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha) \mid > \epsilon \right\} = 0.$

The following better bound is from [7].

Theorem A.3. *In the setting of the previous Theorem A.2, we have*

$$\varrho^\infty \left\{ \sup_{\alpha \in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha)| > \epsilon \right\} \le 4 s_{n^2} \exp(-2n\epsilon^2 + 4\epsilon + 4\epsilon^2).$$

Hence, by Sauer's lemma,

$$\varrho^\infty \left\{ \sup_{\alpha \in I} |\varrho_n(C_\alpha) - \varrho(C_\alpha)| > \epsilon \right\} \le 4n^{2\mathcal{V}_C} \exp(-2n\epsilon^2 + 4\epsilon + 4\epsilon^2). \tag{A.4}$$

Finally, we shall need the following collection of sets with finite VC dimension.

Theorem (*[9, Thm 7.2]*). *Let* G *be a* d*-dimensional real vector space of real functions on an infinite set* X*. Define the class of sets*

$$\mathcal{C} = \{\{x \in X : g(x) > 0\} : g \in G\}.$$

Then the VC dimension of $\mathcal{C}$ *is not more than* d*.*

References

[1] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (1999) 203–228.

[2] P. Barrieu, N. El Karoui, Inf-convolution of risk measures and optimal risk transfer, Finance and Stochastics 9 (2005) 269–298.

[3] P. Barrieu, N. El Karoui, Pricing, hedging and optimally designing derivatives via minimization of risk measures, in: Volume on Indifference Pricing, Princeton University Press, 2006.

[4] P. Carr, H. Geman, D. Madan, Pricing and hedging in incomplete markets, Journal of Financial Economics 62 (2001) 131–167.

[5] A. Černy, S. Hodges, The theory of good-deal pricing in incomplete markets, in: Mathematical Finance — Bachelier Congress 2000, Springer-Verlag, Berlin, 2001, pp. 175–202.

[6] J.H. Cochrane, J. Saá-Requejo, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, Journal of Political Economy 108 (2000) 79–119.

[7] L. Devroye, Bounds for the uniform deviation of empirical measures, Journal of Multivariate Analysis 12 (1) (1982) 72–79.

[8] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, in: Applications of Mathematics, vol. 31, Springer-Verlag, New York, 1996.

[9] R.M. Dudley, Central limit theorems for empirical measures, The Annals of Probability 6 (1978) 899–929.
[10] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance and Stochastics 6 (2002) 429–447.
[11] H. Föllmer, A. Schied, Stochastic Finance: An Introduction in Discrete Time, second ed., in: Studies in Mathematics, vol. 27, de Gruyter, Berlin, 2004.
[12] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., Chapman & Hall/CRC, London, 2003.
[13] D. Heath, Back to the future, in: Plenary Lecture at the First World Congress of the Bachelier Society, Paris, 2000.
[14] S. Jaschke, U. Küchler, Coherent risk measures and good-deal bounds, Finance and Stochastics 5 (2001) 181–200.
[15] E.L. Lehmann, Testing Statistical Hypotheses, second ed., in: Wiley Series in Probability and Mathematical Statistics, 1986.
[16] K. Larsen, T. Pirvu, S. Shreve, R. Tütüncü, Satisfying convex risk limits by trading, Finance and Stochastics 9 (2004) 177–195.
[17] J. Staum, Fundamental theorems of asset pricing for good deal bounds, Mathematical Finance 14 (2004) 141–161.
[18] M. Talagrand, Sharper bounds for gaussian and empirical processes, The Annals of Probability 22 (1994) 28–76.