OPERATIONS RESEARCH
Vol. 58, No. 5, September–October 2010, pp. 1505–1521
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5805 | 1505

DOI 10.1287/opre.1090.0784

Stochastic Root Finding and Efficient Estimation of Convex Risk Measures

Jörn Dunkel
Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, United Kingdom,
jorn.dunkel@physics.ox.ac.uk

Stefan Weber
Institut für Mathematische Stochastik, Insurance and Financial Mathematics, Gottfried Wilhelm Leibniz Universität Hannover,
30167 Hannover, Germany, sweber@stochastik.uni-hannover.de

Reliable risk measurement is a key problem for financial institutions and regulatory authorities. The current industry standard Value-at-Risk has several deficiencies. Improved risk measures have been suggested and analyzed in the recent literature, but their computational implementation has largely been neglected so far. We propose and investigate stochastic approximation algorithms for the convex risk measure Utility-Based Shortfall Risk. Our approach combines stochastic root-finding schemes with importance sampling. We prove that the resulting Shortfall Risk estimators are consistent and asymptotically normal, and provide formulas for confidence intervals. The performance of the proposed algorithms is tested numerically. We finally apply our techniques to the Normal Copula Model, which is also known as the industry model CreditMetrics. This provides guidance for future implementations in practice.

Subject classifications: convex risk measures; shortfall risk; stochastic approximation; stochastic root finding; importance sampling; exponential twisting; portfolio credit risk management.
Area of review: Optimization.
History: Received November 2008; revisions received January 2009, June 2009; accepted July 2009. Published online in *Articles in Advance* May 11, 2010.

1. Introduction

The reliable measurement of financial risk is one of the key issues for financial institutions and regulatory authorities. Distributions of profits and losses (P&L) are very complex, and risk measures are needed to conveniently summarize the most important properties of the lower tail of a P&L distribution. Although very popular and widely used in practice, the industry standard Value-at-Risk (VaR) suffers from severe deficiencies if considered as a measure of the downside risk: First, VaR penalizes diversification in many situations; second, VaR does not take into account the structure of the P&L distribution beyond the VaR. These serious shortcomings stimulated intense research on alternative risk measures. The axiomatic theory of risk measures, initiated by the seminal paper of Artzner et al. (1999), has matured during the past 10 years. Various improved risk measures have been designed, including the risk measure Utility-Based Shortfall Risk (SR) considered in this paper. An excellent survey on the mathematical theory of risk measures is given by Föllmer and Schied (2004).

The axiomatic framework of risk measurement is already well developed. This paper considers computational aspects. The practical implementation of improved risk measures relies on the availability of efficient algorithms for their estimation. P&L distributions of realistic portfolio models are typically highly complex, and thus require Monte Carlo simulations when functionals of the P&L (like its mean, moments, or downside risk quantified by a risk measure) need to be computed. We introduce efficient algorithms for the direct estimation of SR. SR is a convex risk measure that does not share the deficiencies of VaR. Our method for estimating SR relies on a recursive stochastic approximation scheme.

The underlying ideas can be summarized as follows: The risk measure SR can be characterized as the unique root of a decreasing function $g\colon \mathbb{R} \to \mathbb{R}$ that typically needs to be computed by Monte Carlo simulation. A straightforward approach suggests splitting the problem into two parts; see Dunkel and Weber (2007):

(1) the design of a deterministic root-finding procedure that would converge to the sought-after root if g were known;

(2) the construction of efficient Monte Carlo algorithms that enable the estimation of $g(s)$ for each given argument $s \in \mathbb{R}$ with sufficient precision.

Variance reduction techniques can be used to accelerate the estimation of the function $g(s)$ at each given argument $s \in \mathbb{R}$. This approach is closely related to sample average methods in stochastic programming; see, for example, Kleywegt et al. (2001), Linderoth et al. (2006), Mak et al. (1999), Shapiro (2003), Shapiro and Nemirovski (2005), Verweij et al. (2003), and Shapiro et al. (2009).

An alternative to a combination of Monte Carlo estimators and deterministic root-finding schemes is provided

by integrated stochastic approximation algorithms that are presented in this paper. These exploit the whole structure of the root-finding problem by combining variance reduction techniques with *stochastic* root finding. Thereby, they take advantage of cross-averaging of errors for different arguments of the function g—an effect that is neglected whenever *deterministic* root-finding schemes are applied. Although we consider a one-dimensional stochastic root-finding problem, stochastic approximation schemes are also efficient tools for the solution of multidimensional optimization problems in stochastic programming whenever the objective function is given in terms of an expectation; cf. Nemirovski et al. (2009).

In this paper, we will concentrate on the estimation of the improved, convex risk measure SR. Our strategy, however, can easily be extended to the efficient estimation of the industry standard VaR. The recent literature on VaR focusses mostly on the efficient estimation of the probability of large losses; see, e.g., Glasserman (2004), Kang and Shahabuddin (2005), and Glasserman et al. (2008). VaR can indeed be determined by inverting the distribution function, i.e., by solving a deterministic root-finding problem. Clearly, this standard approach for the estimation of VaR does not exploit any cross-averaging effects. More efficient, direct VaR estimators can be constructed by combining the various variance reduction techniques for VaR with stochastic root-finding procedures analogous to those discussed below.[1]

This paper is organized as follows. Section 2 reviews the definitions and properties of VaR and SR. In §3, we present stochastic approximation algorithms for the estimation of SR. In §§3.1 and 3.2, we analyze these algorithms for SR. We show that the estimators are consistent and approximately normal, and we provide formulas for their rate of convergence and their asymptotic variance. We numerically investigate the performance of the proposed stochastic root-finding algorithms for various stylized P&L distributions in §4. In §5, we demonstrate that the algorithms can successfully be implemented for credit portfolio models that are used in practice. For simplicity, we focus on the Normal Copula Model and show how the algorithms can be combined with variance reduction techniques as discussed in Dunkel and Weber (2007). However, the proposed methods can easily be extended to more complex models like the t-copula model; see Kang and Shahabuddin (2005). Section 6 contains concluding remarks.

2. Value-at-Risk vs. Utility-Based Shortfall Risk

In §2.1, we recall the definition and basic properties of VaR. Subsequently, the convex risk measure SR, a useful alternative to VaR, will be defined and discussed in §2.2.

2.1. Value-at-Risk

We denote by X the overall P&L distribution of a financial position over a fixed time horizon T. Assuming that X is a random variable on some probability space $(\Omega, \mathcal{F}, P)$, the risk measure VaR at level $\lambda \in (0, 1)$ can be defined as

$$\mathrm{VaR}_\lambda(X) := \inf\{c \in \mathbb{R} \mid P[X + c < 0] \leqslant \lambda\}.$$

VaR corresponds to a quantile of the distribution of X. Equivalently, for any given level $\lambda \in (0, 1)$, the VaR of a position is the smallest monetary amount that needs to be added to the position such that the probability of a loss does not exceed λ. Typical values for λ that are used in practice are $\lambda = 0.05$ or $\lambda = 0.01$.

VaR has become very popular and is widely used in practice nowadays; cf. Jorion (2007). In particular, considerable effort has been devoted towards developing efficient Monte Carlo (MC) methods for estimating VaR in realistic credit risk and market models; see, e.g., Glasserman (2004), Glasserman et al. (2008), and the references therein. Unfortunately, VaR suffers from two drawbacks. First, it does not assess portfolio diversification as being beneficial. Mathematically, this is due to the fact that VaR is a nonconvex risk measure. Second, VaR does not take into account the size of very large losses that might occur in the case of a severe default event. The latter aspect can be illustrated by the following simple example. Consider two portfolios modeled by the random variables X_1 and X_2, respectively, where

$$X_1 = \begin{cases} +1\$, & \text{with probability } 99\%, \\ -1\$, & \text{with probability } 1\%, \end{cases}$$

and

$$X_2 = \begin{cases} +1\$, & \text{with probability } 99\%, \\ -10^{10}\$, & \text{with probability } 1\%. \end{cases}$$

A value $X_i \geqslant 0$ corresponds to the event "no loss," whereas $X_i < 0$ means "loss," $i = 1, 2$. Setting $\lambda = 0.01$, one finds

$$\mathrm{VaR}_\lambda(X_1) = \mathrm{VaR}_\lambda(X_2) = -1 \leqslant 0.$$

Hence, according to this VaR, both portfolios would be equally acceptable.[2] In this example, however, the first portfolio is clearly preferable. For more-complicated models, the amplitude of losses is usually much less obvious. Therefore, risk allocation based on VaR criteria may result in a concentration in the portfolio position with the smallest loss probability, even if the potential loss associated with this position is extremely large. The severe drawbacks of VaR have stimulated intense research on alternative risk measures, leading, among others, to the definition of convex SR as a useful alternative to VaR; see, e.g., Föllmer and Schied (2004), Weber (2006), Giesecke et al. (2008), and Föllmer et al. (2009).

2.2. Utility-Based Shortfall Risk

A useful alternative to VaR is provided by the convex risk measure SR. To define SR, we recall that an increasing, nonconstant function $l\colon \mathbb{R} \to \mathbb{R}$ is called a *loss function*.

DEFINITION 2.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, l a convex loss function, and λ a point in the interior of the range of l. We define an acceptance set by

$$\mathcal{A} := \{X \in L^\infty\colon E[l(-X)] \leqslant \lambda\}.$$

The corresponding risk measure

$$\mathrm{SR}_{l,\lambda}(X) := \inf\{m \in \mathbb{R}\colon m + X \in \mathcal{A}\}, \quad X \in L^\infty,$$

is called Utility-Based Shortfall Risk with loss function l and threshold level λ.

Let us recall some properties of SR. For a more detailed discussion, we refer the reader to Föllmer and Schied (2004), Weber (2006), Giesecke et al. (2008), and Föllmer et al. (2009).

REMARK 2.2. (1) Utility-Based Shortfall Risk is closely related to the expected utility theory of von Neumann and Morgenstern, and Savage. Setting $u(x) = -l(-x)$, we can define the von Neumann and Morgenstern expected utility of $X \in L^\infty$ by $U(X) := E[u(X)]$. Acceptability of a position X in terms of the risk measure $\mathrm{SR}_{l,\lambda}$ in Definition 2.1 amounts to requiring that the utility $U(X)$ is at least $-\lambda$.

(2) Utility-Based Shortfall Risk is a distribution-based convex risk measure that is continuous from above and below, see Föllmer and Schied (2004).

(3) If ϱ is Utility-Based Shortfall Risk associated with the convex loss function l and threshold level λ, i.e., $\varrho = \mathrm{SR}_{l,\lambda}$, then ϱ admits a robust representation in terms of the family $\mathcal{M}_1(P)$ of all probability measures that are absolutely continuous with respect to P,

$$\mathrm{SR}_{l,\lambda}(X) = \max_{Q \in \mathcal{M}_1(P)} (E_Q(-X) - \alpha_{\min}(Q)), \quad X \in L^\infty,$$

with minimal penalty function

$$\alpha_{\min}(Q) = \inf_{z>0} \frac{1}{z}\left(\lambda + E\left[l^*\left(z\frac{dQ}{dP}\right)\right]\right), \quad Q \in \mathcal{M}_1(P),$$

where $l^*(y) = \sup_{x \in \mathbb{R}}(yx - l(x))$ is the convex conjugate of l. Compared to the general case of convex risk measures, the calculation of the penalty function is substantially simplified.

(4) Utility-Based Shortfall Risk is *invariant under randomization*, which formalizes the following idea. Suppose that the financial positions $X_1, X_2 \in L^\infty$ are acceptable with respect to a given risk measure ϱ, i.e., $\varrho(X_1) \leqslant 0$ and $\varrho(X_2) \leqslant 0$. Let Y be an independent Bernoulli random variable that takes the value 0 with probability α and the value 1 with probability $1-\alpha$ for $\alpha \in (0,1)$. Consider the randomized position or compound lottery X given by

$$X(\omega) = \begin{cases} X_1(\omega) & \text{if } Y(\omega) = 0, \\ X_2(\omega) & \text{if } Y(\omega) = 1, \end{cases} \qquad (\omega \in \Omega).$$

If $\varrho = \mathrm{SR}_{l,\lambda}$ is Utility-Based Shortfall Risk, then $\varrho(X) \leqslant 0$, i.e., X is acceptable if X_1 and X_2 are acceptable. This implication might be economically plausible in many situations.

An analogous statement holds also if X_1 and X_2 are both not acceptable, i.e., $\varrho(X_1) > 0$ and $\varrho(X_2) > 0$. In this case, $\varrho(X) > 0$ and the compound lottery X is also not acceptable.

It turns out that *invariance under randomization* is closely related to the weak dynamic consistency of risk measurements; see Weber (2006).

(5) The domain of SR does not have to be restricted to L^∞, but can be extended to larger spaces, as long as suitable integrability conditions are satisfied.

The following implicit characterization of Utility-Based Shortfall Risk as the root of a function is particularly useful for the numerical estimation of the downside risk, cf. Föllmer and Schied (2004).

PROPOSITION 2.3. *Let* $\mathrm{SR}_{l,\lambda}$ *be Utility-Based Shortfall Risk associated with the convex loss function* l *and the threshold level* λ. *Suppose that* $X \in L^\infty$. *Then, the following statements are equivalent*:

(1) $\mathrm{SR}_{l,\lambda}(X) = s^*$,
(2) $E[l(-X - s^*)] = \lambda$.

Proposition 2.3 is the basis for the estimation algorithm that we propose in §3. $\mathrm{SR}_{l,\lambda}(X)$ is the unique zero s^* of the function

$$g(s) := E[l(-X - s)] - \lambda. \tag{1}$$

A stochastic approximation method for determining s^* can be subdivided into two partial tasks:

(1) Use a stochastic approximation algorithm that produces a random sequence $(s_k)_{k \in \mathcal{N}}$ which converges almost surely to the root s^*. The value s_{k+1} will be recursively chosen given s_k. This requires an estimator of $g(s_k)$. This estimator should have small variance.

(2) Given a model or certain statistics for X, find an estimator for $g(s)$ at any given point s that has small variance. In many cases, variance reduction techniques can be used to improve the estimator's efficiency. So far, almost all of the literature has focused on variance reduction techniques for VaR. We will, however, show in §5 how these methods can be adopted for the superior risk measure SR, see also Dunkel and Weber (2007).

3. Stochastic Approximation Algorithms

The aim of this section is to construct an algorithm that converges rapidly to the root s^* of the function g defined in (1). In contrast to the preceding section, we do not have to restrict attention to $X \in L^\infty$, but can allow for more general spaces of random variables.[3]

We analyze properties of two stochastic approximation algorithms: the standard Robbins-Monro algorithm and the Polyak-Ruppert algorithm, which averages the slowed iterates.

3.1. The Robbins-Monro Algorithm

We denote by $(\Omega, \mathcal{F}, P)$ some sufficiently rich probability space.[4] For $s \in [a, b]$, we will simulate random variables with expectation $g(s)$ to construct the sequence $(s_n)_{n\in\mathcal{N}}$ that converges to the unique zero s^* of the function g. A generic Robbins-Monro algorithm can be described as follows.

3.1.1. The Algorithm. For $s \in [a, b]$, we denote by $\hat{Y}_s\colon [0, 1] \to \mathbb{R}$ a measurable function such that

$$g(s) = E[\hat{Y}_s(U)] \tag{2}$$

for any on $[0, 1]$ uniformly distributed random variable U. The functions $\hat{Y}_s$ can be chosen appropriately to increase the efficiency of the stochastic root-finding scheme. A simple, but often not very efficient, choice is $\hat{Y}_s(\cdot) = l(-q(\cdot) - s) - \lambda$, where q denotes the quantile function of X. We will discuss in §3.1.2 how importance sampling techniques can be incorporated in the definition of $\hat{Y}_s$ and discuss consistency and asymptotic normality for this case.

For practical and mathematical reasons, we construct algorithms that are restricted to a bounded domain. Letting $-\infty < a < s^* < b < \infty$, we define for $x \in \mathbb{R}$ a projection $\Pi\colon \mathbb{R} \to [a, b]$ onto the interval $[a, b]$ by

$$\Pi(x) = \begin{cases} a, & x \leqslant a, \\ x, & a < x < b, \\ b, & b \leqslant x. \end{cases}$$

A Robbins-Monro algorithm for s^* can now be constructed as follows:

- Choose a constant $\gamma \in (\frac{1}{2}, 1]$, $c > 0$ and a starting value $s_1 \in [a, b]$.
- For $n \in \mathbb{N}$, we define recursively

$$s_{n+1} = \Pi\left[s_n + \frac{c}{n^\gamma} \cdot Y_n\right] \tag{3}$$

with

$$Y_n = \hat{Y}_{s_n}(U_n) \tag{4}$$

for a sequence (U_n) of independent, unif[0,1]-distributed random variables.

Remark 3.1. Defining the corresponding filtration

$$\mathcal{F}_n = \sigma(Y_i\colon i < n), \quad n \in \mathbb{N}, \tag{5}$$

we can rewrite the algorithm as

$$s_{n+1} = s_n + \frac{c}{n^\gamma} g(s_n) + \frac{c}{n^\gamma} \delta M_n + \frac{c}{n^\gamma} Z_n, \tag{6}$$

where $\delta M_n = Y_n - g(s_n)$, and Z_n is the correction term defined according to (3) and (6). Observe that $g(s_n) = E[Y_n \mid \mathcal{F}_n]$ implies that $(\delta M_n)_n$ is a martingale difference sequence.

The analysis of the consistency and asymptotic optimality of the Robbins-Monro algorithm above requires certain moment conditions. For this reason, we define the following quantities:

Definition 3.2. For $s \in [a, b]$, let $\hat{Y}_s\colon [0, 1] \to \mathbb{R}$ be defined as in (2), and let U be a random variable that is uniformly distributed on $[0, 1]$. We denote the variance of $\hat{Y}_s(U)$ by

$$\sigma^2(s) = \operatorname{var}(\hat{Y}_s(U)). \tag{7}$$

For $p > 0$, we also define the higher centered moments

$$m^{2+p}(s) = E(|\hat{Y}_s(U) - g(s)|^{2+p}). \tag{8}$$

3.1.2. Importance Sampling. To improve the speed of the stochastic approximation algorithm, the functions $\hat{Y}_s$, $s \in [a, b]$, in (2) are constructed from the distribution of $X + s$ by applying suitable variance reduction techniques. In this current paper, we focus on the case of *importance sampling*.

Denote by q the quantile function of X. Then,

$$g(s) = E[l(-q(U) - s)] - \lambda$$

for any on $[0, 1]$ uniformly distributed random variable. To define an equivalent importance-sampling measure change, we let $h_s\colon \mathbb{R} \to \mathbb{R}$ be a measurable function such that $h_s > 0$ P-almost surely and $E[h_s(q(U) + s)] = 1$. h_s defines an equivalent measure change by setting

$$\frac{d\hat{P}_s}{dP} = h_s(q(U) + s).$$

We denote expectation and variance under $\hat{P}_s$ by $\hat{E}_s$ and $\hat{\operatorname{var}}_s$, respectively.

Definition 3.3. Denote the distribution function of $q(U) + s$ under $\hat{P}_s$ by $\hat{F}_s$ and its right-continuous inverse by $\hat{q}_s$. We define

$$\hat{Y}_s(\cdot) := \frac{l(-\hat{q}_s(\cdot)) - \lambda}{h_s(\hat{q}_s(\cdot))}. \tag{9}$$

Notation 3.4. Observe that we defined $\hat{F}_s$ and $\hat{q}_s$ as the distribution and quantile function of

$$q(U) + s$$

under $\hat{P}_s$. These are not the distribution and quantile functions of $q(U)$ under $\hat{P}_s$, which are given by $\hat{F}_s(\cdot + s)$ and $\hat{q}_s - s$, respectively.

Remark 3.5. Defining

$$V_n = \hat{q}_{s_n}(U_n), \tag{10}$$

we can rewrite Equation (4) in the definition of the Robbins-Monro algorithm as $Y_n = l(-V_n) - \lambda/(h_s(V_n))$.

REMARK 3.6. Letting $p > 0$, it follows that for any random variable U that is uniformly distributed on $[0, 1]$ under the measure P,

$$g(s) = E[l(-q(U) - s) - \lambda] = E\big[\hat{Y}_s(U)\big] = \hat{E}_s\left[\frac{l(-q(U) - s) - \lambda}{h_s(q(U) + s)}\right],$$

$$\sigma^2(s) = \operatorname{var}\big(\hat{Y}_s(U)\big) = \hat{\operatorname{var}}_s\left(\frac{l(-q(U) - s) - \lambda}{h_s(q(U) + s)}\right),$$

$$m^{2+p}(s) = E(|\hat{Y}_s(U) - g(s)|^{2+p}) = \hat{E}_s\left(\left|\frac{l(-q(U) - s) - \lambda}{h_s(q(U) + s)} - g(s)\right|^{2+p}\right).$$

3.1.3. Consistency. In this section, we prove the consistency of the algorithm when the functions $\hat{Y}_s$ are defined according to (9). We make the following weak assumption.

ASSUMPTION 3.7. (1) *$g(s)$ is well-defined and finite for all $s \in [a, b]$;*
(2) $\sup_{s\in[a,b]} \sigma^2(s) < \infty$.

THEOREM 3.8. *Assume that $\hat{Y}_s$, $s \in [a, b]$, is given according to Definition* 3.3*, and suppose that Assumption* 3.7 *holds. Then, $s_n \to s^*$ P-almost surely.*

PROOF. g is decreasing, with s^* being its only zero. Thus, the only limiting point of the ordinary differential equation $\dot{s} = g(s)$ is s^*. Theorem 5.2.1 in Kushner and Yin (2003) implies that $s_n \to s^*$ as $n \to \infty$ P-almost surely if we can verify their conditions (A2.1)–(A2.5). (A2.1) is guaranteed by Assumption 3.7(2). (A2.2), (A2.4), and (A2.5) are obvious. Because l is convex, the function

$$[a, b] \to \mathbb{R}, \quad s \mapsto l(-q(U) - s) - \lambda$$

is continuous. Moreover, for $s \in [a, b]$, we have

$$l(-q(U) - b) - \lambda \leqslant l(-q(U) - s) - \lambda \leqslant l(-q(U) - a) - \lambda. \tag{11}$$

By Assumption 3.7(2), we have that $l(-q(U) - b) - \lambda$, $l(-q(U) - a) - \lambda \in L^1(P)$. Thus, $g(\cdot)$ is continuous on $[a, b]$ by Lebesgue's Dominated Convergence Theorem. □

REMARK 3.9. The consistency of the Robbins-Monro algorithm does not rely on the fact that we defined $\hat{Y}_s$, $s \in [a, b]$, by an importance-sampling procedure. Other variance reduction techniques can be used instead. We refer to Theorem 5.2.1. in Kushner and Yin (2003) for general sufficient conditions for consistency.

3.1.4. Asymptotic Normality. Under stronger conditions, it can also be shown that the sequence $(s_n)_{n\in\mathbb{N}}$ is asymptotically normal. If $X \in L^\infty$, then the following simple conditions suffice, which can easily be verified.

ASSUMPTION 3.10. (1) *The functions $(h_s)_{s\in[a,b]}$ are uniformly bounded away from zero on any compact set.*
(2) *l is continuously differentiable.*
(3) *h_s is continuous for any $s \in [a, b]$.*
(4) $c > (-2g'(s^*))^{-1}$.

In the general case where $X \in L^\infty$ is not necessarily satisfied, the following slightly more complicated condition replaces Assumption 3.10. Assumption 3.11 follows from Assumption 3.10 if $X \in L^\infty$.

ASSUMPTION 3.11. (1) *g is continuously differentiable.*
(2) *$\sigma^2(\cdot)$ is continuous at s^*.*
(3) $c > (-2g'(s^*))^{-1}$.
(4) *For some $p > 0$ and $\rho > 0$, we have*

$$\sup_{|s-s^*|<\rho} m^{2+p}(s) < \infty.$$

(5) *For some $\rho > 0$, the family $(Y_n \mathbf{1}_{\{|s_n - s^*| < \rho\}})_n$ is uniformly integrable.*

PROPOSITION 3.12. *If $X \in L^\infty$, then Assumption* 3.10 *implies Assumptions* 3.7 *and* 3.11.

PROOF. Assumption 3.7(1) is immediate. Also observe that g is continuous by Lebesgue's dominated convergence theorem because l is continuous. Because l is continuously differentiable and $|l'(-q(U) - s)| \leqslant \sup_{|x|\leqslant\|X\|_\infty,\, s\in[a,b]} |l'(-x - s)| < \infty$ for $s \in [a, b]$, Assumption 3.11(1) holds.

Observe further that for $p \geqslant 0$,

$$\hat{E}_s\left(\left|\frac{l(-q(U) - s) - \lambda}{h_s(q(U) + s)} - g(s)\right|^{2+p}\right) \leqslant \sup_{|x|\leqslant\|X\|_\infty,\, s\in[a,b]} \left(\left|\frac{l(-x - s) - \lambda}{h_s(x + s)}\right| + |g(s)|\right)^{2+p} < \infty,$$

because l and g are continuous and bounded on compacts, and the functions $(h_s)_{s\in[a,b]}$ are uniformly bounded away from zero on any compact set. This implies that Assumptions 3.11(2) and (4) hold. Analogously, the estimate

$$\|Y_n\| \leqslant \sup_{\|x\|\leqslant\|X\|_\infty,\, s\in[a,b]} \left|\frac{l(-x - s) - \lambda}{h_s(x + s)}\right| < \infty$$

implies Assumption 3.11(5). □

THEOREM 3.13. *Assume that $\hat{Y}_s$, $s \in [a, b]$, is given according to Definition* 3.3*, and suppose that Assumptions* 3.7 *and* 3.11 *hold.*

If $\gamma = 1$, then

$$\sqrt{n} \cdot (s_n - s^*) \to \mathcal{N}\left(0, \frac{-c^2\sigma^2(s^*)}{2cg'(s^*) + 1}\right).$$

If $\gamma \in (\frac{1}{2}, 1)$, then

$$\sqrt{n^\gamma} \cdot (s_n - s^*) \to \mathcal{N}\left(0, \frac{-c\sigma^2(s^*)}{2g'(s^*)}\right).$$

PROOF. We first verify that (A2.0)–(A2.7) on page 329 in Kushner and Yin (2003) hold. (A2.1) is satisfied by Assumption 3.11(5). (A2.2) is a consequence of Theorem 3.8. (A2.4) follows from Taylor's Theorem and Assumption 3.11(1). (A2.5) follows from the fact that $g(s^*) = 0$. For (A2.6), observe that $A = cg'(s^*)$. By Assumption 3.11(3), we have $A + 1/2 = (2cg'(s^*) + 1)/2 < 0$ because $g'(s^*) < 0$. Thus, A and $A + 1/2$ are negative. The first part of (A2.7) follows from Assumption 3.11(4), and the second part from Assumption 3.11(2). (A2.3) follows easily from Theorem 10.4.1 in Kushner and Yin (2003) because their Assumptions (A4.1)–(A4.5) are satisfied.

In summary, we have shown that Theorem 10.2.1 in Kushner and Yin (2003) applies, which implies the asymptotic normality. We finally need to verify the formula for the asymptotic variance. If $\gamma = 1$, the stationary variance of the process U in Theorem 10.2.1 in Kushner and Yin (2003) is

$$\int_0^\infty \exp((2cg'(s^*)+1)t) \cdot c^2\sigma^2(s^*)\,dt = \frac{-c^2\sigma^2(s^*)}{2cg'(s^*)+1}.$$

If $\gamma \in (\frac{1}{2}, 1)$, the stationary variance of the process U in Theorem 10.2.1 in Kushner and Yin (2003) is

$$\int_0^\infty \exp(2cg'(s^*)t) \cdot c^2\sigma^2(s^*)\,dt = \frac{-c\sigma^2(s^*)}{2g'(s^*)}. \quad \square$$

REMARK 3.14. (1) Note that the previous proof actually demonstrates that Theorem 10.2.1 in Kushner and Yin (2003) applies, if Assumptions 3.7 and 3.11 are satisfied. For practical applications we will not need this result in full generality. We will, however, need a similar result for the proof of Theorem 3.16 below.

(2) The asymptotic normality of the Robbins-Monro algorithm does not rely on the fact that we defined $\hat{Y}_s$, $s \in [a, b]$, by an importance-sampling procedure. Other variance reduction techniques can be used instead. We refer to Theorem 10.2.1. in Kushner and Yin (2003) for general sufficient conditions.

REMARK 3.15. (1) Asymptotically, the choice $\gamma = 1$ gives the best rate of convergence. However, if the step size decreases too quickly, convergence might be slow in practice for finite time horizons.

(2) The asymptotic variance in Theorem 3.13 involves an unknown constant c. For example, the optimal choice for the exponent $\gamma = 1$ is $c = -(2g'(s^*)+1)/g'(s^*)^2$, which depends on the unknown constant $g'(s^*)$. Adaptive procedures that choose the constant c dynamically by estimating $g'(s^*)$ adaptively have been suggested in the literature; see, e.g., Ruppert (1991), but are generally not as efficient as the Polyak-Ruppert averaging estimators discussed in §3.2.

3.2. Polyak-Ruppert Averaging

The optimal asymptotic variance of the Robbins-Monro algorithm relies on the choice of the constant c in (3), which is not known a priori. Ruppert (1988), Ruppert (1991), Polyak (1990), and Polyak and Juditsky (1992) suggest as an alternative estimation procedure the averaging of the results of the slowed Robbins-Monro algorithm. We next discuss a Polyak-Ruppert algorithm for Utility-Based Shortfall Risk and compare the Robbins-Monro algorithm and the Polyak-Ruppert algorithm in numerical case studies.

We will restrict our attention to the case where the value of the considered portfolio is a random variable in L^∞. This allows us to prove the asymptotic normality of a Polyak-Ruppert estimator that uses a maximum window of averaging; see Kushner and Yang (1995). However, the asymptotic properties of the estimator can also be studied if the boundedness assumption on the portfolio is dropped. For estimators that use a minimal window of averaging of fixed size, asymptotic normality holds under weaker conditions; see Theorem 11.1. in Kushner and Yin (2003). For estimators that use a maximal window of averaging, the asymptotic variance can still be identified; see, Theorem 11.3.1 in Kushner and Yin (2003).

The following theorem describes the Polyak-Ruppert algorithm for Utility-Based Shortfall Risk and states its consistency and asymptotic normality:

THEOREM 3.16. *Suppose that $X \in L^\infty$, $\gamma \in (\frac{1}{2}, 1)$, and that Assumption* 3.10 *holds. Assume that (s_n) is given according to* (3)*. For arbitrary $t > 0$, we define*

$$\bar{s}_n = \frac{1}{t \cdot n} \sum_{i=n}^{(t+1)n} s_i, \tag{12}$$

where any upper summation index $u \in \mathbb{R}_+$ is interpreted as the largest natural number that does not exceed u. For every $\epsilon > 0$, there exists another process $\breve{s}$ such that $P(\bar{s}_n = \breve{s}_n\ \forall n) \geqslant 1 - \epsilon$ and

$$\sqrt{tn} \cdot (\breve{s}_n - s^*) \to \mathcal{N}\left(0, \frac{\sigma^2(s^*)}{(g'(s^*))^2}\right). \tag{13}$$

PROOF. The convergence $\bar{s}_n \to s^*$ P-almost surely follows from the convergence of the sequence (s_n).

The process $\breve{s}$ is constructed as follows. Choose N large enough such that

$$P(|s_n - s^*| \leqslant \max(|s^* - a|, |s^* - b|)\ \forall n \geqslant N) \geqslant 1 - \epsilon.$$

Set

$$\tilde{s}_{n+1} = \begin{cases} s_{n+1}, & n < N, \\ \tilde{s}_n + \dfrac{c}{n^\gamma} \cdot \hat{Y}_{\tilde{s}_n}(U_n), & n \geqslant N. \end{cases}$$

We define $\breve{s}_n = (1/t \cdot n) \sum_{i=n}^{(t+1)\cdot n} \tilde{s}_i$. By construction, $P(\bar{s}_n = \tilde{s}_n\ \forall n) \geqslant 1 - \epsilon$. To prove the asymptotic normality (13), it suffices to verify (13) for $N = 1$ and $s_1 \in [a, b]$.

This is a consequence of Theorem 4.2 in Kushner and Yang (1995). We need to verify their Assumptions (A4.1)–(A4.5). (A2.1) follows from Theorem 4.5.13 in Benveniste et al. (1990). (A4.1) is satisfied with $q_n = n$, $a_n = n^{-\gamma}$, and $\gamma \in (1/2, 1)$. (A4.2) follows with $g(s,u) = Y_s(u)$ and $\bar{g}(s) := g(s)$ for $s \in [a,b]$, $g(s,u) := \bar{g}(s) := g'(a)(s-a) + g(a)$ for $s < a$, and $g(s,u) := \bar{g}(s) := g'(b)(s-b) + g(b)$ for $s > b$. This implies that $\psi_j(s) = 0$ for all $j \in \mathbb{N}$ and $s \in \mathbb{R}$. The first part of (A4.4) is, thus, also obvious. The second part follows from our definition of $g(\cdot,\cdot)$. □

Remark 3.17. The Polyak-Ruppert algorithm as described in Theorem 3.16 averages over slowed Robbins-Monro iterates with $\gamma < 1$. Apart from an additional factor of $1/t$, this Polyak-Ruppert algorithm achieves the asymptotically best convergence rate, corresponding to that of a Robbins-Monro algorithm with $\gamma = 1$. The asymptotic variance of the Polyak-Ruppert algorithm with $\gamma < 1$ equals the best-possible variance of the Robbins-Monro algorithm with $\gamma = 1$; see Theorem 3.13 and Remark 3.15. In particular, this optimal variance is independent of the parameter c. However, due to the convexity of the loss function l, the Polyak-Ruppert algorithm usually exhibits more bias towards values larger than s^* if c and t are chosen too large. The reason for this is as follows: The convexity of l implies that $|g(s-\epsilon)| > |g(s+\epsilon)|$ for $\epsilon > 0$. Hence, on average, the step sizes $|s_{n+1} - s_n|$ of the slowed Robbins-Monro algorithm is larger for values $s_n < s^*$ than for $s_n > s^*$. This effect induces a bias of the Robbins-Monro estimator toward values that are larger than s^*. However, this bias quickly vanishes in the case of the "pure" Robbins-Monro algorithm. By contrast, due to the additional averaging procedure, the nth iterate $\bar{s}_n$ of the Polyak-Ruppert algorithms will typically be more biased than the last iterate $s_{(t+1)n}$ of the Robbins-Monro algorithm that was used to compute $\bar{s}_n$. A possible remedy might be to employ regression-based root-finding schemes, as suggested in Ruppert (1991). In practice, numerical case studies for typical portfolio distributions can be used to choose reasonable parameters and to construct an efficient algorithm. The resulting numerical scheme can afterwards be applied to practically relevant portfolio distributions. The case studies in §4 consider, for example, a value of t on the order of 10%.

3.3. Estimator of the Asymptotic Variance

The asymptotic variance in (13) is given by the unknown quantity $\sigma^2(s^*)/(g'(s^*))^2$. In this section, we will discuss how $\sigma^2(s^*)$ and $g'(s^*)$ can be estimated based on the data that have already been generated in the Robbins-Monro algorithm (3). Except for the computation of the estimators, this procedure does not require any additional effort because the simulation results can be recycled for this purpose after s^* has been estimated. The proofs of the consistency of the estimators rely on a law of large numbers for martingale difference sequences or, more generally, mixingales; see McLeish (1975) and Chow (1967).

3.3.1. Confidence Intervals. Before we discuss these estimators, let us briefly recall how confidence intervals can be determined if $\sigma^2(s^*)$ and $g'(s^*)$ are known. In practice, these quantities are replaced by their estimators and $\check{s}_n$ is replaced by $\bar{s}_n$. For $\alpha \in (0,1)$, let $q := \inf\{x\colon \Phi(x) \geqslant 1-\alpha/2\}$ be the $(1-\alpha/2)$-quantile where Φ denotes the cumulative Gaussian distribution function. Then, $-q$ is the $\alpha/2$-quantile, and

$$\left[\check{s}_n - \left|\frac{\sigma(s^*)}{g'(s^*)\sqrt{tn}}\right| \cdot q,\ \check{s}_n + \left|\frac{\sigma(s^*)}{g'(s^*)\sqrt{tn}}\right| \cdot q\right]$$

is an approximate confidence interval for s^* with confidence coefficient $1-\alpha$.

3.3.2. Estimation.

Theorem 3.18. *Suppose that $X \in L^\infty$, $t > 0$, and that Assumption* 3.10 *holds. Then, we have*

$$\sigma^2(s^*) = \lim_{n\to\infty} \frac{1}{tn} \sum_{i=n}^{(t+1)n} Y_i^2, \tag{14}$$

$$g(s) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{l(-V_i + s_i - s) - \lambda}{h_{s_i}(V_i)}, \tag{15}$$

$$g'(s) = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{\varepsilon n} \cdot \sum_{i=1}^{n} \frac{l(-V_i + s_i - s - \varepsilon) - l(-V_i + s_i - s)}{h_{s_i}(V_i)}, \tag{16}$$

where $(Y_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ are defined according to (4) *and* (10)*, respectively, and convergence holds P-almost surely.*

Remark 3.19. For the definitions of $(Y_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ in (4) and (10), observe that $\hat{q}_s$ the quantile function of $q(U) + s$ under the measure $\hat{P}_s$; see Definition 3.3.

Proof. To prove (14), observe that for $i \in \mathbb{N}$ we have that

$$E[Y_i^2 \mid \mathcal{F}_i] - g(s_i)^2 = \sigma^2(s_i).$$

Thus, $M_i := Y_i^2 - g(s_i)^2 - \sigma^2(s_i)$, $i \in \mathbb{N}$, defines a martingale difference sequence. $X \in L^\infty$ together with Assumption 3.10 imply that $\sup_{i\in\mathbb{N}} \|M_i\|_\infty < \infty$. In particular, $(M_i)_{i\in\mathbb{N}}$ is a square-integrable martingale difference sequence and $\sum_i E[M_i^2]/i^2 < \infty$. By Corollary 1.9 in McLeish (1975),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} M_i^2 = 0.$$

Hence, $\lim_{n\to\infty} (1/tn) \sum_{i=n}^{(t+1)n} M_i^2 = 0$. Observe that

$$\lim_{n\to\infty} \frac{1}{tn} \sum_{i=n}^{(t+1)n} (g(s_i)^2 + \sigma^2(s_i)) = \sigma^2(s^*)$$

because g and σ are continuous, $g(s^*)=0$, and $\lim_n s_n = s^*$ P-almost surely. Thus,

$$\lim_{n\to\infty} \frac{1}{tn} \cdot \sum_{i=n}^{(t+1)n} Y_i^2 = \lim_{n\to\infty} \frac{1}{tn} \cdot \sum_{i=n}^{(t+1)n} (Y_i^2 - g(s_i)^2 - \sigma^2(s_i)) + \lim_{n\to\infty} \frac{1}{tn} \cdot \sum_{i=n}^{(t+1)n} (g(s_i)^2 + \sigma^2(s_i)) = \sigma^2(s^*).$$

For the proof of (15), observe that

$$E\left[\frac{l(-V_i+s_i-s)-\lambda}{h_{s_i}(V_i)} \,\middle|\, \mathcal{F}_i\right] = g(s).$$

Thus, $((l(-V_i+s_i-s)-\lambda)/(h_{s_i}(V_i))-g(s))_{i\in\mathcal{N}}$ is a martingale difference sequence that is uniformly bounded in L^∞ because of $X\in L^\infty$ and Assumption 3.10. Similar arguments to the above imply (15).

Finally, (16) is an immediate consequence of (15). □

4. Numerical Case Studies

In both this section and §5 below, we test the performance of the proposed simulation scheme for SR. First, we will analyze the stochastic approximation of SR for specific distributions without importance sampling. This will illustrate the key factors that determine the performance of the scheme. Later on, in §5 we will focus on the estimation of SR in a more realistic credit portfolio model. There, we will investigate the effects of importance sampling. As it turns out, variance reduction is often important to ensure a good performance of the stochastic approximation scheme.

4.1. Notation

The number of steps in one simulation run will be denoted by n'. Letting $\rho\in(0,1)$ be a constant, we will compare simulation results $s_{n'}$ for the Robbins-Monro algorithm (3), and

$$\bar{s}'_{n'} = \frac{1}{n'\rho} \sum_{i=n'(1-\rho)}^{n'} s_i \tag{17}$$

for the Polyak-Ruppert algorithm. The defining Equation (12) of the Polyak-Ruppert scheme is recovered from the latter Equation (17) by setting $n := n'(1-\rho)$ and $t=\rho/(1-\rho)$. Accordingly, we set

$$S_{n'} := \sqrt{n'^{\gamma}}(s_{n'} - s^*) \tag{18}$$

and

$$\bar{S}'_{n'} := \sqrt{n'\rho}(\bar{s}'_{n'} - s^*). \tag{19}$$

Under suitable conditions, $S_{n'}$ and $\bar{S}'_{n'}$ are asymptotically normal; see Theorems 3.13 and 3.16. The asymptotic variance of $S_{n'}$ is given by

$$\Sigma = \begin{cases} [-c\sigma^2(s^*)]/[2g'(s^*)], & \gamma\in(1/2,1), \\ [-c^2\sigma^2(s^*)]/[2cg'(s^*)+1], & \gamma=1. \end{cases} \tag{20}$$

We estimate this quantity according to Theorem 3.18, using one simulation run with n' steps.[5] The corresponding estimator is denoted by $\Sigma_{n'}$. The asymptotic variance of $\bar{S}'_{n'}$, given by

$$\bar{\Sigma} = \sigma^2(s^*)/[g'(s^*)]^2, \tag{21}$$

is estimated analogously. We denote the corresponding estimator by $\bar{\Sigma}_{n'}$.

4.2. Simulations

In applications, the loss amount L is often used as a starting point rather than the portfolio value X. L is simply the negative of X, i.e., $X=-L$. Positive (negative) values of L correspond to losses (gains).

As the first simple example, we consider an exponential loss function

$$l(-x) = \exp(-\beta x), \quad \beta>0, \tag{22}$$

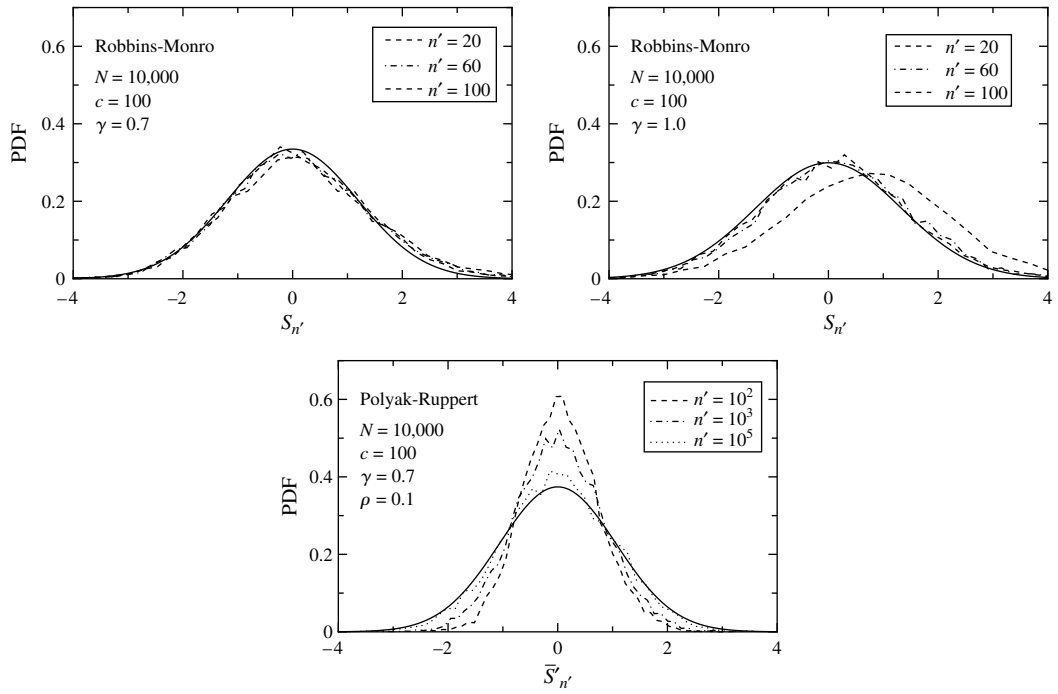
and Gaussian loss distribution, i.e., $X := -L$ where $L\sim\mathcal{N}(\mu_0,\sigma_0^2)$. In this case, $\mathrm{SR}_{l,\lambda}$ can be calculated in closed form:

$$\mathrm{SR}_{l,\lambda}(-L) = s^* = \mu_0 + \frac{\beta\sigma_0^2}{2} - \frac{\log\lambda}{\beta}, \tag{23}$$

which is useful for testing the algorithms. With $\beta=0.5$, $\lambda=0.05$, and $L\sim\mathcal{N}(\mu_0,\sigma_0^2)$ with $\mu_0=0$, $\sigma_0=1$, we obtain the exact value $\mathrm{SR}_{l,\lambda}=6.24146=:s^*$. For constant n', empirical means, variances, and histograms are based on $N=10{,}000$ simulation runs. At the beginning of each run, the initial value s_1 is uniformly sampled from the interval $[a,b]:=[s^*-10,s^*+10]$. We simulate and compare the Robbins-Monro (RM) algorithms with $\gamma=1.0$ and $\gamma=0.7$, respectively, and the Polyak-Ruppert (PR) algorithm with $\gamma=0.7$ and $\rho=0.1$. If ρ is too large, the performance of the PR algorithm decreases. We set $c=100$ which is close to the optimal choice for the RM algorithm. We increased the maximum number of steps $n'=20$ per run to $n'=10^5$. Figures 1 and 2 illustrate that $S_{n'}=\sqrt{n'^{\gamma}}(s_{n'}-s^*)$ and $\bar{S}'_{n'}=\sqrt{\rho n'}(\bar{s}'_{n'}-s^*)$ with $s^*=6.24146$ quickly converge to a Gaussian distribution, as predicted by Theorems 3.13 and 3.16.

The construction of confidence intervals requires estimates of the asymptotic variances Σ and $\bar{\Sigma}$. Figure 3 shows the corresponding numerically obtained histograms for the estimators $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$. This demonstrates that a reasonable confidence interval can be obtained for $n'=10^4$.

Figure 1. Gaussian loss distribution and exponential loss function.



Note. Asymptotically, the rescaled quantities $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ approach Gaussian distributions (cf. Theorems 3.13 and 3.16). The numerical results agree with the theoretical prediction (solid lines).

As a second example, we consider $L \sim \mathcal{N}(\mu_0, \sigma_0^2)$ with (piecewise) polynomial loss function

$$l_\eta(L-s) = \eta^{-1}(L-s)^\eta \mathbf{1}_{\{L>s\}} = \eta^{-1}\begin{cases} L-s)^\eta, & L \geqslant s, \\ 0, & L < s. \end{cases} \tag{24}$$

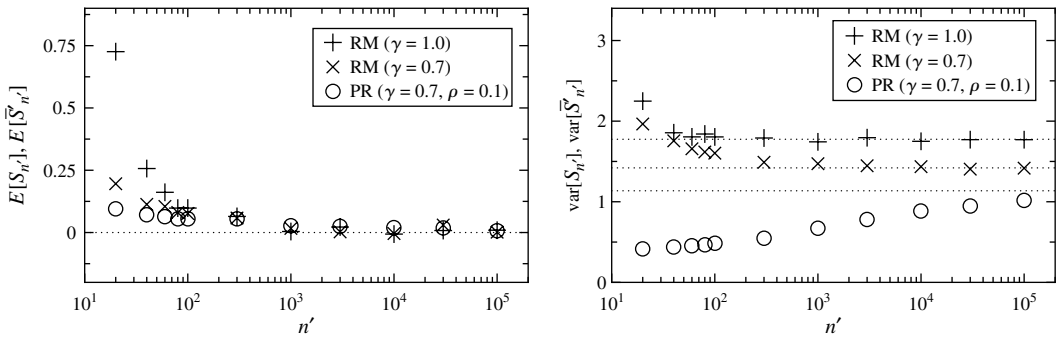
Setting $\eta = 2$, $\mu_0 = 0$, $\sigma_0 = 1$, and $\lambda = 0.05$, the numerically obtained SR value is given by $s^* = 0.86937$. We compare the same algorithms as before, but with $c > 15$ and $[a, b] = [s^* - 5, s^* + 5]$. The starting point s_1 of the simulation is uniformly sampled from this interval. The algorithms again show good convergence properties; cf. Figure 4. The PR algorithm shows better finite sample properties in this case than the RM algorithm with $\gamma = 1$.

For the case $\gamma = 0.7$, numerical simulations show that $S_{n'}$ and $\bar{S}'_{n'}$ are very close to normal for $n' = 10^5$, but also for smaller values of n', deviations from normality are not too large. Reasonable estimators $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$ of the asymptotic variances can be obtained for $n' = 10^5$.

As a third example, we consider a heavy-tailed loss distribution of the Frechet type, defined by the cumulative distribution function

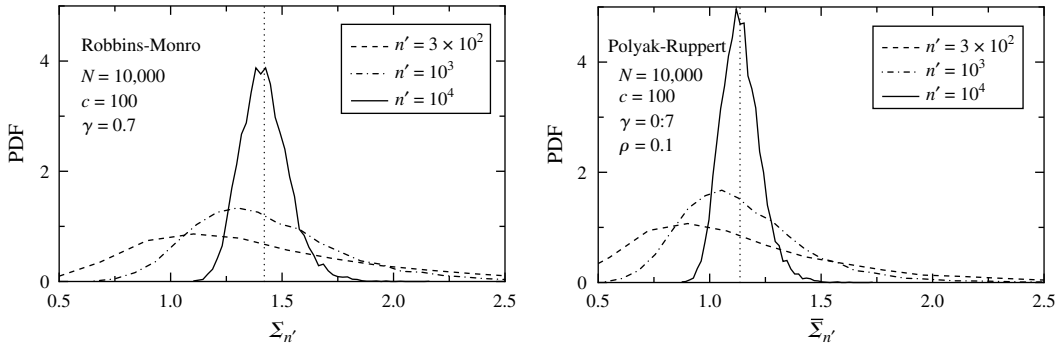
$$P[L < x] = \exp\left[-(1+\xi_0 x)^{-1/\xi_0}\right], \quad \xi_0 > 0, \tag{25}$$

Figure 2. Gaussian loss distribution and exponential loss function.



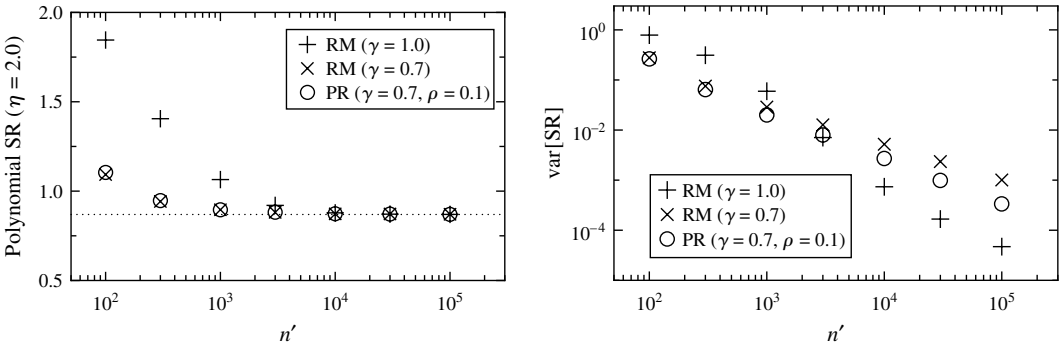
Note. Empirical mean values and variances of the rescaled quantities $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$, corresponding to the histograms from Figure 1. For large enough n', the variances converge to the values predicted by Theorems 3.13 and 3.16 (dotted lines).

Figure 3. Gaussian loss distribution and exponential loss function.



Note. Illustration of Theorem 3.18: Histogram of the variances $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$, using the estimators from Theorem 3.18, respectively. The vertical lines indicate the exact results.

Figure 4. Gaussian loss distribution and (piecewise) polynomial SR with $\eta = 2$.

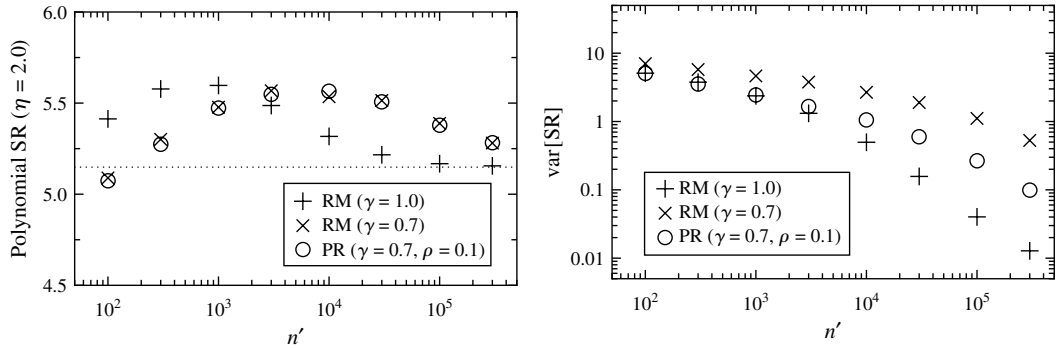


Note. Left: The different algorithms converge to the same SR value $s^* = 0.87$. Right: Sample variances of the SR estimates in the left diagram.

and, again, SR with polynomial loss function (24). For $\eta = 2$, $\lambda = 0.05$, and $\xi_0 = 0.1$, the SR value can be estimated numerically as $s^* = 5.1486$; cf. Figure 5. Choosing $c = 100 > (-2g'(s^*))^{-1}$ and $[a, b] = [s^* - 5, s^* + 5]$, the convergence of the algorithms is much slower than in the other two cases; cf. Figures 4 and 5. The distributions of $S_{n'}$ in the case of the RM algorithms and $\bar{S}'_{n'}$ become reasonably close to normal for $n' = 10^6$. However, convergence of the estimators $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$ of the asymptotic variances is extremely slow.

The examples illustrate that, without importance sampling, the algorithms work efficiently for P&L distributions that are close to a normal, whereas their performance decreases significantly for heavy-tailed distributions. In general, model-specific variance reduction techniques can be used to ensure a good

Figure 5. Heavy-tailed Frechet-type loss distribution (25) and (piecewise) polynomial SR with $\lambda = 0.05$.



Note. Left: The different algorithms converge to the same SR value $s^* = 5.1486$. Right: Sample variances of the SR estimates in the left diagram. The convergence is considerably slower than in the case of light-tailed Gaussian loss distributions; cf. Figure 4.

performance of the algorithms even for heavy-tailed distributions.

5. Application: Credit Portfolio Models

A standard example of a credit portfolio model is the Normal Copula Model (NCM), which was introduced in Gupton et al. (1997) and presents the foundation of *CreditMetrics*. Although frequently used in practice, the NCM should, of course, not be considered as a realistic model. The purpose of the this section is to illustrate how the proposed stochastic approximation techniques, when combined with importance sampling, can be used to compute the downside risk for models that are more complex than the simple examples in §4. Generally, our methods can be used in more general models like the t-copula model that exhibits tail dependence. The importance-sampling estimators for loss probabilities in Kang and Shahabuddin (2005) can be extended to the case of SR.

The basic equations and properties of the NCM are briefly reviewed in §5.1. Subsequently, we discuss the estimation of SR in the NCM and provide numerical examples.

5.1. Basic Equations

We consider a portfolio with m positions (or obligors) over a fixed time horizon T. Each position is subject to default risk. For each obligor $i = 1, 2, \ldots, m$, a random variable D_i with values in $\{0, 1\}$ indicates whether or not i has defaulted at horizon T. $D_i = 1$ corresponds to a default of position i. The partial net loss associated with a default of the obligor i is given by a positive constant $v_i > 0$. Assuming no recovery, the overall loss $L \geqslant 0$ of the portfolio over the horizon T can be written in the standard form

$$L = \sum_{i=1}^{m} v_i D_i. \tag{26}$$

Note that these models typically count losses as positive. If portfolio gains are not neglected, profits and losses are given by $X = -L$. If downside risk is measured by a risk measure ϱ, then the risk of L is given by $\varrho(-L)$.

The NCM is a threshold model describing a credit portfolio with m obligors, i.e., there exists an m-dimensional random vector $R = (R_1, R_2, \ldots, R_m)$ and threshold levels $r_1, r_2, \ldots, r_m \in \mathbb{R}$ such that

$$D_i = \mathbf{1}_{\{R_i > r_i\}}.$$

In the NCM it is specifically assumed that R is an m-dimensional normal random vector with standardized one-dimensional marginals. Denoting by $p_i = P\{D_i = 1\}$ the marginal default probability of the obligor i, we obtain that

$$r_i = \Phi^{-1}(1 - p_i), \tag{27}$$

where Φ is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. Thus, instead of directly choosing r_i, one could also specify the marginal default probabilities $p_1, \ldots, p_m$ and determine the threshold values $r_1, \ldots, r_m$ from Equation (27). In industry applications of the NCM, the covariance matrix of the Gaussian vector R is often specified through a factor model of the form

$$R_i = A_{i0}\varepsilon_i + \sum_{j=1}^{d} A_{ij} Z_j, \quad i = 1, \ldots, m,\ d < m; \tag{28}$$

$$1 = A_{i0}^2 + A_{i1}^2 + \cdots + A_{id}^2, \quad A_{i0} > 0,\ A_{ij} \geqslant 0. \tag{29}$$

The *systematic* risk variables $Z_1, \ldots, Z_d$ and the *idiosyncratic* risks variables $\varepsilon_1, \ldots, \varepsilon_m$ are chosen as independent standard normal random variables. The parameters (A_{ij}) determine the cross-coupling as well as the relative size (influence) of the different risk factors on the latent variables $R_1, \ldots, R_m$. The additional constraint (29) ensures that $R_i \sim \mathcal{N}(0, 1)$ holds.

If $R_1, \ldots, R_m$ are specified through the factor model above, the NCM obeys the following conditional independence structure. Conditionally on the common factors $Z = (Z_1, \ldots, Z_d)$, the default indicators D_i are independently distributed. Conditional on the vector of systematic factors Z, the default events $\{D_i = 1\}$ occur with probability

$$p_i(Z) := P[D_i = 1 \mid Z] = \Phi\left(\frac{\sum_{j=1}^{d} A_{ij} Z_j - r_i}{A_{i0}}\right). \tag{30}$$

In principle, it is straightforward to perform numerical MC studies on the basis of Equations (26)–(30). The NCM model is uniquely determined by the parameter vector

$$(m, d, p_1, \ldots, p_m, v_1, \ldots, v_m, A_{11}, \ldots, A_{md}).$$

In a naive MC simulation, one first draws the d independent random numbers $(Z_j)_{j=1,\ldots,d}$ from a standard normal distribution and then calculates L according to (26). Repeating this procedure several times, one can obtain estimators for functionals of L, e.g., the moments $E[L^n]$ of the loss distribution or the loss probabilities

$$P[L > c] = E[\mathbf{1}_{\{L > c\}}] \in [0, 1]. \tag{31}$$

In the NCM, the total portfolio loss is bounded from above, that is, $0 \leqslant L \leqslant L_+$, $L_+ := \sum_{i=1}^{m} v_i$; and it suffices therefore to consider $c \in [0, L_+]$. When measuring downside risk, one is typically interested in estimating rather rare events. In this case, the MC method outlined above becomes computationally expensive because the default events become very rare. Thus, naive MC estimators often do not provide good estimates unless very large sample sizes are considered. Accordingly, variance reduction techniques become very important in practical applications. For example, Glasserman (2004) and Glasserman et al. (2008) discuss efficient estimators for loss probabilities by applying the importance-sampling method *exponential twisting*. An analogous approach can be used to obtain numerically efficient MC estimators for SR if variance reduction techniques are combined with stochastic approximation techniques. We consider as an example the piecewise polynomial loss function

$$l_\eta(x) = \eta^{-1} x^\eta \mathbf{1}_{\{x \geqslant 0\}}(x), \quad \eta > 1.$$

5.1.1. Independent Default Events: Exponential Twisting. An intermediate step in the design of efficient stochastic approximation algorithms is the construction of importance-sampling estimators for $E_P[l(L-s)]$. If $\hat{P}_s$ is another probability measure that is equivalent to P with $\mathrm{d}\hat{P}_s/\mathrm{d}P = h_s(L-s)$, then $E_P[l(L-s)] = E_{\hat{P}_s}[l(L-s)/h_s(L-s)]$. It follows that $J_n^{h_s} = n^{-1}\sum_{k=1}^n (l(L_k-s)/h_s(L_k-s))$ is an unbiased, consistent estimator of $E_P[l(L-s)]$ if the random variables L_k are sampled independently from the distribution of L under $\hat{P}_s$. Because the estimator is unbiased, its mean square error can be expressed as the square root of its variance. Thus, the mean square error becomes small if and only if the variance $\mathrm{var}_{\hat{P}_s}[l(L-s)/h_s(L-s)]$ is small. In the present case, we are primarily interested in events that correspond to large L. To reduce the variance of the estimator, we need to transfer mass to these events. An *exponential twist* refers to a density g, which is exponential in L; i.e., we consider a class of measures $\hat{P}_s^\theta$, $\theta \geqslant 0$, with

$$\frac{\mathrm{d}\hat{P}_s^\theta}{\mathrm{d}P} = \frac{\exp(\theta L)}{\exp[\psi(\theta)]},$$

where

$$\psi(\theta) := \log E[\exp(\theta L)] = \sum_{i=1}^m \log\bigl[1 + p_i\bigl(e^{\theta v_i} - 1\bigr)\bigr] \tag{32}$$

is the cumulant generating function of the loss variable L, and $\exp[\psi(\theta)]$ is a normalizing constant. The twist parameter θ has to be determined such that a good variance reduction is achieved (see the discussion below).

For the NCM with independent default events, the discussed measure change is equivalent to a change of the individual default probabilities. The defaults are still independent under $\hat{P}_s^\theta$. For the individual default probabilities under $\hat{P}_s^\theta$, we obtain that

$$\hat{p}_i(\theta) := \hat{P}_s^\theta[D_i = 1] := \frac{p_i e^{\theta v_i}}{1 + p_i(e^{\theta v_i} - 1)}. \tag{33}$$

As evident from Equation (33), the new default probabilities $\hat{p}_i$ not only depend on the original default probabilities p_i, but also on the partial losses v_i. In general, for $\theta > 0$ the default probability of the ith portfolio position is increased (in particular, we have $\hat{p}_i(0) = p_i$). Hence, under the new measure $\hat{P}_s^\theta$ default events are more likely to occur. The inverse likelihood ratio for the change from P to $\hat{P}_s^\theta$ can be written as

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}\hat{P}_s^\theta} &= \prod_{i=1}^m \left(\frac{p_i}{\hat{p}_i(\theta)}\right)^{D_i} \left(\frac{1-p_i}{1-\hat{p}_i(\theta)}\right)^{1-D_i} \\ &= \exp[-\theta L + \psi(\theta)].\end{aligned} \tag{34}$$

Denoting by E and $\hat{E}s_s^\theta$ the expectations under P and $\hat{P}_s^\theta$, respectively, we can write

$$E[l(L-s)] = \hat{E}_s^\theta[l(L-s)\exp[-\theta L + \psi(\theta)]]. \tag{35}$$

Hence, in the case of the piecewise polynomial loss function, importance sampling for

$$E[l(L-s)] = E[\eta^{-1}(L-s)^\eta \mathbf{1}_{\{L \geqslant s\}}]$$

corresponds to generating samples of the quantity

$$\eta^{-1}(L-s)^\eta \mathbf{1}_{\{L \geqslant s\}} \exp[-\theta L + \psi(\theta)] \tag{36}$$

under the measure $\hat{P}_s^\theta$. The implementation of the sampling procedure is straightforward because of Equation (33). The probability distributions of the default indicators under $\hat{P}_s^\theta$ are known, which implies that L can easily be sampled.

It thus remains to discuss how the parameter θ can be determined such that the variance of the estimator based on Equation (36) is significantly smaller than the variance of the corresponding naive estimator for the left-hand side of (35). Because the estimator is unbiased, it is equivalent to consider the second moment,

$$\begin{aligned}M_2(s,\theta) &:= \frac{1}{\eta^2}\hat{E}_s^\theta\bigl[(L-s)^{2\eta}\mathbf{1}_{\{L\geqslant s\}}^2 \exp[-2\theta L + 2\psi(\theta)]\bigr] \\ &= \frac{1}{\eta^2}E\bigl[(L-s)^{2\eta}\mathbf{1}_{\{L\geqslant s\}}\exp[-\theta L + \psi(\theta)]\bigr] \\ &\leqslant M_2(s,0)\exp[-\theta s + \psi(\theta)].\end{aligned} \tag{37}$$

Here, $M_2(s,0) = E[(L-s)^{2\eta}\mathbf{1}_{\{L\geqslant s\}}]$ is the second moment "without" exponential twisting. Consequently, instead of directly minimizing $M_2(s,\theta)$, which is very difficult or even impossible in general, one can at least minimize the upper bound on the right-hand side of inequality (37). A promising choice for the twisting parameter is thus given by

$$\theta_s = \begin{cases} \text{u. s. of } \psi'(\theta) = s, & s > \psi'(0), \\ 0, & s \leqslant \psi'(0), \end{cases} \tag{38}$$

where the abbreviation "u. s." stands for "unique solution." As discussed in the next section, this approach is directly transferable to the case of nonindependent defaults.

5.1.2. Dependent Default Events: Conditional Exponential Twisting. Let us now return to the general case, where the default events of different portfolio positions may be coupled. In this case, exponential twisting can be applied to the conditional distribution $P[\cdot \mid Z]$ of the indicator variables D_i. Conditional on Z, we are in the situation of the preceding section because defaults are conditionally independent given Z.

The basic idea of *conditional exponential twisting* is thus to replace in the formulae of §5.1.1 the default probabilities p_i by the conditional default probabilities

$$p_i(Z) := P[D_i = 1 \mid Z] = \Phi\left(\frac{\sum_{j=1}^d A_{ij}Z_j - x_i}{A_{i0}}\right). \tag{39}$$

Analogous to Equation (32), we define the conditional cumulant generating function by

$$\psi(\theta, Z) := \log E[\exp(\theta L) \mid Z] = \sum_{i=1}^{m} \log\bigl[1 + p_i(Z)\bigl(e^{\theta v_i} - 1\bigr)\bigr]. \tag{40}$$

As in Equation (38), the parameter θ that governs the measure change can be determined. In the current case, θ depends on the factor Z, i.e.,

$$\theta_s(Z) = \begin{cases} \text{u. s. of } \psi'(\theta, Z) = s, & s > \psi'(0, Z), \\ 0, & s \leqslant \psi'(0, Z), \end{cases} \tag{41}$$

where

$$\psi'(\theta, z) := \frac{\partial}{\partial\theta}\psi(\theta, z), \qquad \psi'(0, Z) = E[L \mid Z] = \sum_{i=1}^{m} v_i p_i(Z). \tag{42}$$

With these definitions, the corresponding MC algorithm reads as follows:

(1) Generate a d-dimensional Gaussian random vector of factor variables, $Z \sim \mathcal{N}(0, 1_d)$, where 1_d denotes the $d \times d$-unity matrix.

(2) Calculate

$$\hat{p}_i(\theta_s(Z), Z) := \frac{p_i(Z)e^{v_i\theta_s(Z)}}{1 + p_i(Z)\bigl(e^{v_i\theta_s(Z)} - 1\bigr)} \tag{43}$$

with $\theta_s(Z)$ given by Equation (41) and $p_i(Z)$ given by Equation (39).

(3) Generate m Bernoulli-random numbers $D_i \in \{0, 1\}$, such that $D_i = 1$ with probability $\hat{p}_i(\theta_s(Z), Z)$.

(4) Calculate $\psi(\theta_s(Z), Z)$ from Equation (40) and $L = \sum_{i=1}^{m} v_i D_i$, and return the estimator

$$l(L - s)\exp[-L\theta_s(Z) + \psi(\theta_s(Z), Z)]. \tag{44}$$

Here, the exponential factor corresponds to the conditional likelihood ratio; cf. Equation (34).

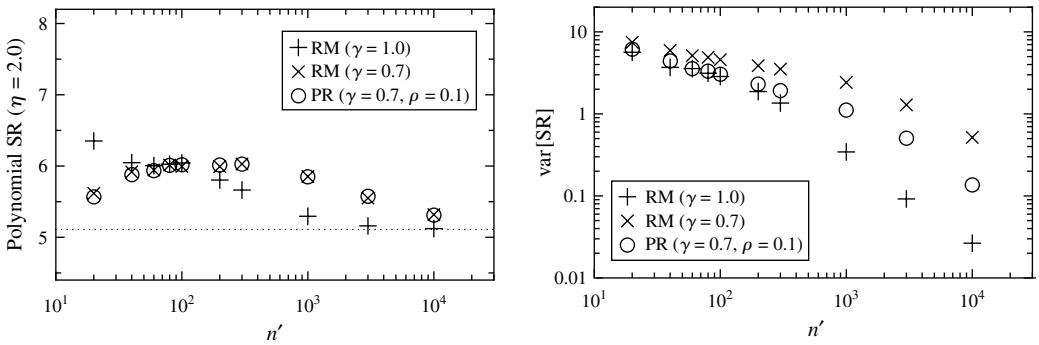
As in the case of VaR, this algorithm yields a significant variance reduction provided the default events are not too strongly correlated, i.e., if $A_{ij} \ll 1$ holds for $i \geqslant 1$. Otherwise, additional importance sampling of the factor variables Z may turn out to be helpful; cf. Glasserman et al. (2008).

5.2. Numerical Results

For a numerical case study, we consider a portfolio with $m = 25$ portfolio positions that are subdivided into 5 classes with partial losses $v_1 = \cdots = v_5 = 1.00$, $v_6 = \cdots = v_{10} = 1.25$, $v_{11} = \cdots = v_{15} = 1.50$, $v_{16} = \cdots = v_{20} = 1.75$, and $v_{21} = \cdots = v_{25} = 2.00$. The corresponding maximal loss is given by $L_+ = 37.5$. We assume that the marginal default probabilities are $p_i = 0.05$, corresponding to threshold values $x_i = 1.64488$. The number of common factors is given by $d = 6$. The coupling parameters $A_{ij} = A(i, j)$ are given by $A(1, 1) = A(2, 1) = A(3, 1) = A(4, 1) = A(5, 1) = 0.1$, $A(6, 2) = A(7, 2) = A(8, 2) = A(9, 2) = A(10, 2) = 0.1$, $A(11, 3) = A(12, 3) = A(13, 3) = A(14, 3) = A(15, 3) = 0.1$, $A(16, 4) = A(17, 4) = A(18, 4) = A(19, 4) = A(20, 4) = 0.1$, $A(21, 5) = A(22, 5) = A(23, 5) = A(24, 5) = A(25, 5) = 0.1$, $A(i, 6) = 0.1$, for $i = 1, \ldots, m$ and $A(i, j) = 0$ otherwise. That is, positions within the same class j are coupled with each other by means of the risk factor Z_j, where $j = 1, \ldots, 5$, and all positions are coupled by the common risk factor Z_6. SR will be defined in terms of a polynomial loss function with exponent $\eta = 2$ and threshold level $\lambda = 0.05$. The search interval is given by $[a, b] = [s^* - 5, s^* + 5]$, and we used $c = 100$. For constant n', empirical means, variances, and histograms are based on $N = 10{,}000$ simulation runs. The initial value s_1 is chosen according to a uniform distribution on $[a, b]$.

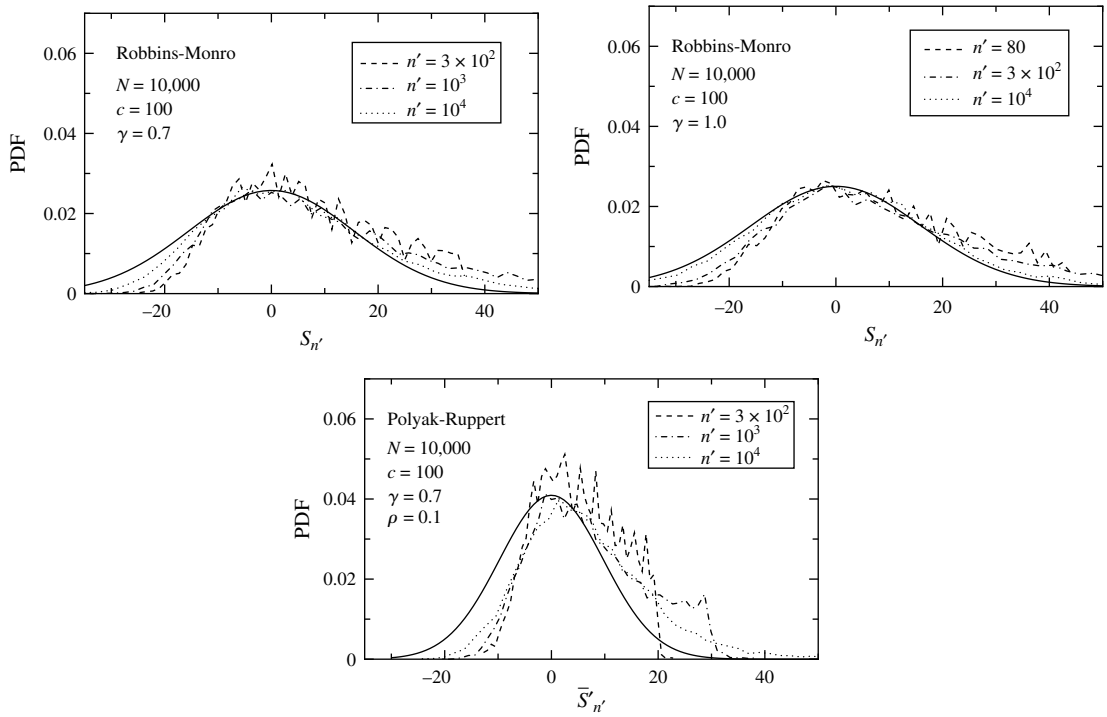
5.2.1. Case Studies Without Importance Sampling. Without importance sampling, the stochastic approximation algorithms are slow. Figure 6 shows the empirical mean and variance of the Robbins-Monro algorithms with $\gamma = 1$ and

Figure 6. NCM and polynomial SR with $\lambda = 0.05$ (without importance sampling).



Note. Left: The different algorithms converge slowly to the same SR value $s^* = 5.11$ (determined from IS). Right: Sample *variances* of the SR estimates in the left diagram. As is evident from the diagrams, without IS the convergence is very slow. The results are qualitatively similar to those obtained for the Frechet-type heavy-tail distribution; cf. Figure 5.

Figure 7. NCM and polynomial SR with $\eta = 2.0$ and $\lambda = 0.05$ (without importance sampling).

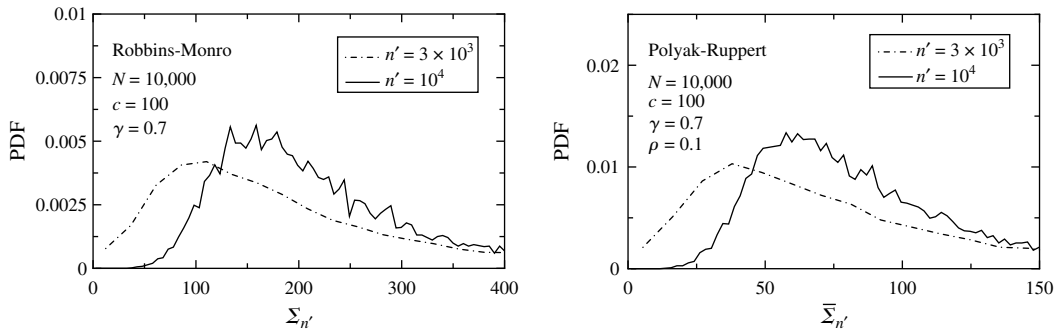


Note. Asymptotically, the rescaled quantities $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ approach Gaussian distributions (cf. Theorems 3.13 and 3.16). For large n', the numerical results approach the theoretical prediction (solid lines, obtained by using the estimators from Theorem 3.18), but again convergence is very slow.

$\gamma = 0.7$, and the Polyak-Ruppert algorithm with $\gamma = 0.7$ and $\rho = 0.1$. Figure 7 illustrates that $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ converge to a Gaussian distribution, as predicted by Theorems 3.13 and 3.16, but convergence is not particularly fast. The construction of confidence intervals relies on estimates of the asymptotic variances Σ and $\bar{\Sigma}$. Figure 8 shows the corresponding numerically obtained histograms for the estimators $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$. The convergence is poor, and the construction of reasonable confidence intervals requires more than $n' = 10^4$ steps.

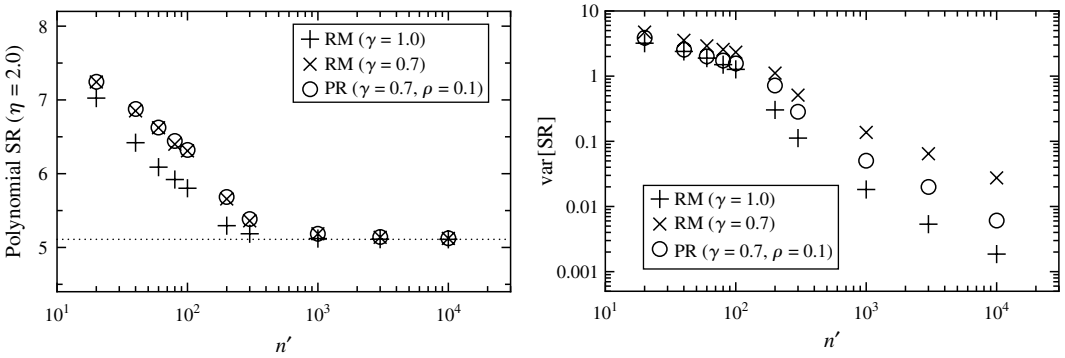
5.2.2. Case Studies with Importance Sampling. With importance sampling (IS), the stochastic approximation algorithms perform very well. We do not modify the distribution of the factor Z, but use importance sampling only conditional on Z. To reduce the numerical effort that is required to calculate the "optimal" value θ_s, we calculated θ_s when the IS criterion (38) was met for the first time; subsequently, the value of θ_s was repeatedly updated only after $0.1 \times n'_{\max}$ performed IS steps, where $n'_{\max}$ is the fixed maximum value n' per simulation run. Figure 9 shows the empirical mean and variance of the Robbins-Monro algorithms

Figure 8. NCM model and polynomial loss function (without importance sampling).



Note. Illustration of Theorem 3.18: Histogram of the variances $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$, using the estimators from Theorem 3.18, respectively. Importance sampling yields considerable variance reduction; cf. Figure 12.

Figure 9. NCM, polynomial SR ($\eta = 2.0$, $\lambda = 0.05$), and exponential twisting.



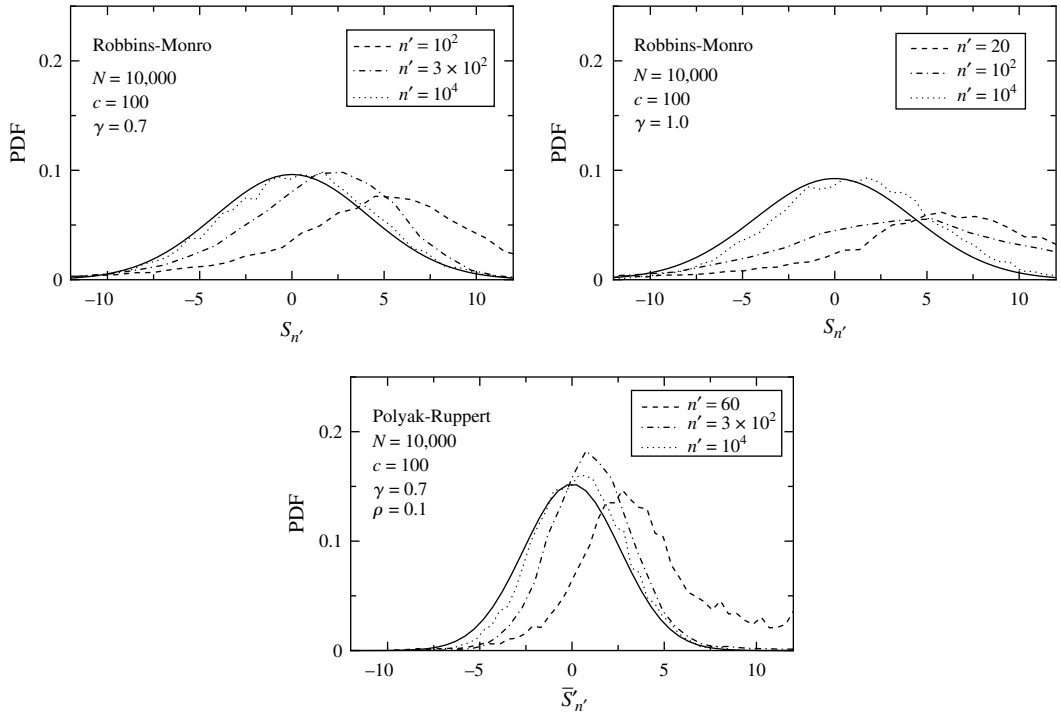
Note. Left: The different algorithms converge to the same SR value $s^* = 5.11$. Right: Sample variances of the SR estimates in the right diagram. Convergence has considerably improved; cf. Figure 6.

with $\gamma = 1$ and $\gamma = 0.7$, and the Polyak-Ruppert algorithm with $\gamma = 0.7$ and $\rho = 0.1$. Figure 10 illustrates that $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ converge to a Gaussian distribution, as predicted by Theorems 3.13 and 3.16. The approximation is excellent for $n' = 10^4$. The empirical means and variances of $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ are shown in Figure 11. The construction of confidence intervals relies on estimates of the asymptotic variances Σ and $\bar{\Sigma}$. Figure 12 shows that the construction of reasonable confidence intervals requires not more than $n' = 10^4$ steps.

6. Concluding Remarks

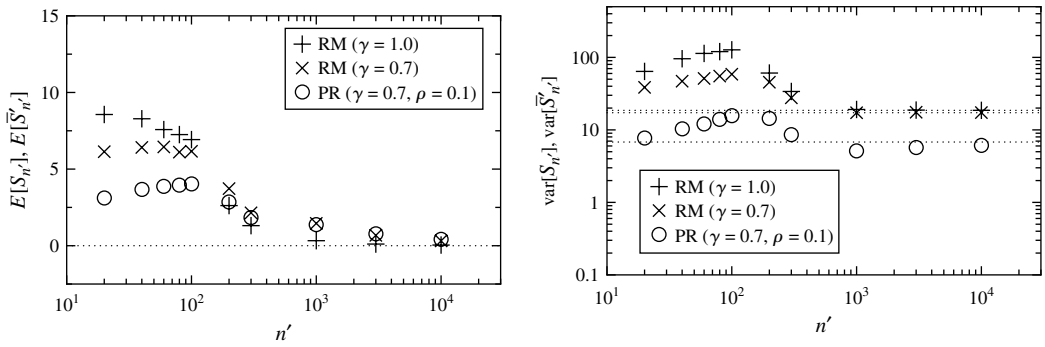
The proposed root-finding algorithms present an efficient tool for the numerical computation of the risk measure Utility-Based Shortfall Risk. We proved their consistency and asymptotic normality and provided estimators for confidence intervals. Our numerical case studies demonstrate that the stochastic approximation algorithms exhibit good finite sample properties if they are combined with importance sampling. The techniques were successfully implemented to estimate shortfall risk in the Normal Copula Model, which underlies the industry model CreditMetrics.

Figure 10. NCM, polynomial SR ($\eta = 2.0$, $\lambda = 0.05$), and exponential twisting.



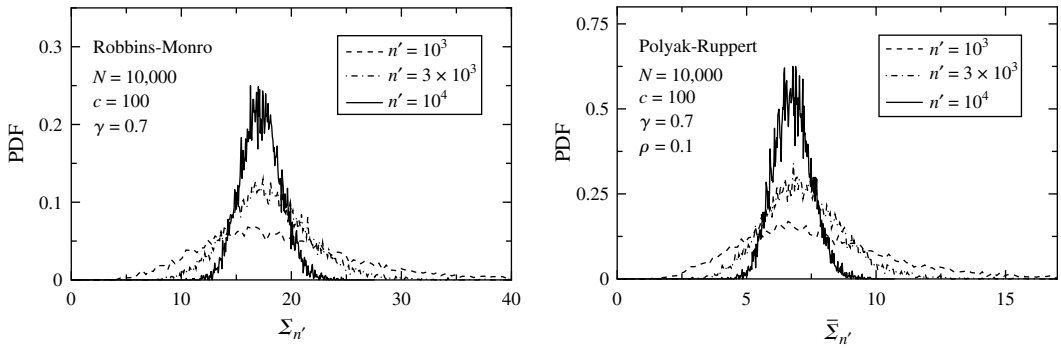
Note. Asymptotically, the rescaled quantities $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$ approach Gaussian distributions (cf. Theorems 3.13 and 3.16). For large n', the numerical results approach the theoretical prediction (solid lines, obtained by using the estimators from Theorem 3.18).

Figure 11. NCM, polynomial SR($\eta = 2.0$, $\lambda = 0.05$), and exponential twisting.



Note. Empirical mean values and variances of the rescaled quantities $S_{n'} = \sqrt{n'^{\gamma}}(s_{n'} - s^*)$ and $\bar{S}'_{n'} = \sqrt{\rho n'}(\bar{s}'_{n'} - s^*)$, corresponding to the histograms from Figure 11. For large enough n', the variances converge to the values predicted by Theorems 3.13 and 3.16 (dotted lines, obtained by using the estimators from Theorem 3.18).

Figure 12. NCM, polynomial SR ($\eta = 2.0$, $\lambda = 0.05$), and exponential twisting.



Note. Illustration of Theorem 3.18; histogram of the variances $\Sigma_{n'}$ and $\bar{\Sigma}_{n'}$, using the estimators from Theorem 3.18, respectively. Importance sampling has led to a considerable variance reduction; cf. Figure 8.

In numerical simulations, neither the Robbins-Monro algorithm nor the Polyak-Ruppert algorithm were clearly superior. Generally, our approach can easily be applied to more realistic models like the t-copula model. For future applications in practice, it is important to adapt and optimize the algorithms for typical portfolio distributions. The latter part is conceptually straightforward because the underlying root-finding schemes can be combined with various, model-specific variance reduction techniques. This enables the design of algorithms that provide an efficient tool for the fast measurement of the downside risk of real portfolios.

Endnotes

1. A similar approach has independently been taken in an interesting paper by Bardou et al. (2009). This working paper became available online at the end of December 2008 after the original submission of our manuscript.
2. Generally, a position is considered acceptable if its risk measure is negative.
3. However, if $X \in L^\infty$ is satisfied, the conditions for asymptotic normality can substantially be simplified.
4. The assumption that the true probability measure P is precisely known is, of course, an idealization. Real applications typically involve model uncertainty. The measure P should be interpreted as a probability measure that has been chosen for modeling purposes.
5. $\sigma^2(s^*)$ can be estimated according to Equation (14) with $n = n'(1-\rho)$ and $t = \rho/(1-\rho)$. $g'(s^*)$ can be estimated according to Equation (16) with $n = n'$.

Acknowledgments

The authors are grateful for helpful remarks of the editors and referees.

References

Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9**(3) 203–228.

Bardou, O. A., N. Frikha, G. Pages. 2009. Computation of VaR and CVaR using stochastic approximations and unconstrained importance sampling. Preprint, 4th version.

Benveniste, A., M. Métivier, P. Priouret. 1990. *Adaptive Algorithms and Stochastic Approximations*. Springer, Berlin.

Chow, Y. S. 1967. On a strong law of large numbers for martingales. *Ann. Math. Statist.* **38**(2) 610.

Dunkel, J., S. Weber. 2007. Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, R. R. Barton, eds. *Proc. 2007 Winter Simulation Conf.*, Washington, DC, 958–966.

Föllmer, H., A. Schied. 2004. *Stochastic Finance—An Introduction in Discrete Time*, 2nd ed. Walter de Gruyter, Berlin.

Föllmer, H., A. Schied, S. Weber. 2009. Robust preferences and robust portfolio choice. *Handbook of Numerical Analysis, XV, Mathematical Modeling and Numerical Methods in Finance*. North-Holland, Amsterdam, 29–89.

Giesecke, K., T. Schmidt, S. Weber. 2008. Measuring the risk of large losses. *J. Investment Management* **6**(4) 1–15.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*, Vol. 53. *Applications of Mathematics*. Springer, New York.

Glasserman, P., W. Kang, P. Shahabuddin. 2008. Fast simulation of multifactor portfolio credit risk. *Oper. Res.* **56**(5) 1200–1217.

Gupton, G., C. Finger, M. Bhatia. 1997. *CreditMetrics Technical Document*. J. P. Morgan & Co., New York.

Jorion, P. 2007. *Value at Risk*, 3rd ed. McGraw-Hill, New York.

Kang, W., P. Shahabuddin. 2005. Fast simulation for multifactor portfolio credit risk in the *t*-copula model. M. Kuhl, N. Steiger, F. Armstrong, J. Joines, eds. *Proc. 2005 Winter Simulation Conf., Orlando, FL,* 1859–1868.

Kleywegt, A. J., A. Shapiro, T. de Mello. 2001. The sample average method for stochastic discrete optimization. *SIAM J. Optim.* **12**(2) 479–502.

Kushner, H. J., J. Yang. 1995. Stochastic approximation with averaging and feedback: Faster convergence. K. Åström, G. Goodwin, P. Kumar, eds. *Adaptive Control, Filtering, and Signal Processing*. Springer, New York, 205–228.

Kushner, H. J., G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York.

Linderoth, J. T., A. Shapiro, S. Wright. 2006. The empirical behavior of sampling methods for stochastic programming. *Ann. Oper. Res.* **142**(1) 215–241.

Mak, W.-K., D. P. Morton, R. K. Wood. 1999. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Oper. Res. Lett.* **24**(1–2) 47–56.

McLeish, D. L. 1975. A maximal inequality and dependent strong laws. *Ann. Probab.* **3**(5) 829–839.

Nemirovski A., A. Juditsky, G. Lan, A. Shapiro. 2009. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19**(4) 1574–1609.

Polyak, B. T. 1990. New stochastic approximation type procedures. *Automat. Remote Control* **51** 937–946.

Polyak, B. T., A. B. Juditsky. 1992. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30**(4) 838–855.

Ruppert, D. 1988. Efficient estimators from a slowly convergent Robbins-Monro procedure. ORIE Technical Report 781, Cornell University, Ithaca, NY.

Ruppert, D. 1991. Stochastic approximation. B. Gosh, P. Sen, eds. *Handbook of Sequential Analysis*. Marcel Dekker, New York, 503–529.

Shapiro, A. 2003. Monte Carlo sampling methods. A. Ruszczynski, A. Shapiro, eds. *Stochastic Programming*. North-Holland, Amsterdam, 353–426.

Shapiro, A., A. Nemirovski. 2005. On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov, eds. *Continuous Optimization: Current Trends and Applications*. Springer, Berlin, 111–146.

Shapiro, A., D. Dentcheva, A. Ruszczynski. 2009. *Lectures on Stochastic Programming*. SIAM, Philadelphia.

Verweij, B., S. Ahmed, A. J. Kleywegt, G. Nemhauser, A. Shapiro. 2003. The sample average approximation method applied to stochastic routing problems: A computational study. *Comput. Optim. Appl.* **24**(2–3) 289–333.

Weber, S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance* **16**(2) 419–441.