EFFICIENT MONTE CARLO METHODS FOR CONVEX RISK MEASURES IN PORTFOLIO CREDIT RISK MODELS

Jörn Dunkel

Institut für Physik,
Universität Augsburg
Universitätsstraße 1, D-86135 Augsburg, Germany

Stefan Weber

School of Operations Research and Information Engineering
279 Rhodes Hall, Cornell University
Ithaca, N.Y. 14853, U.S.A.

ABSTRACT

We discuss efficient Monte Carlo (MC) methods for the estimation of convex risk measures within the portfolio credit risk model CreditMetrics. Our focus lies on the *Utility-based Shortfall Risk* (SR) measures, as these avoid several deficiencies of the current industry standard *Value-at-Risk* (VaR). It is demonstrated that the importance sampling method *exponential twisting* provides computationally efficient SR estimators. Numerical simulations of test portfolios illustrate the good performance of the proposed algorithms.

1 INTRODUCTION

The quantification of credit risks is a key issue for both financial institutions and regulating authorities. Modern risk analysis builds on portfolio models that capture the dependence among different obligors. Standard examples include *CreditMetrics* (Gupton, Finger, and Bhatia 1997) and *CreditRisk$^+$* (Credit Suisse Financial Products 1997). Due to the complexity of realistic models, risk estimation typically requires numerical simulations. These may become computationally expensive if events with low probability, like credit defaults, are investigated. In recent years considerable effort has been put into the development of efficient MC techniques for the popular risk measure VaR (Glasserman, Heidelberger, and Shahabuddin 2000a; Glasserman, Heidelberger, and Shahabuddin 2000b; Glasserman, Heidelberger, and Shahabuddin 2001). Although widely used in practice (Jorion 2000), VaR exhibits serious deficiencies: it penalizes diversification in many situations and is insensitive to the size of losses beyond a certain threshold level. These drawbacks stimulated intense research on alternative risk measures (Artzner et al. 1999; Acerbi and Tasche 2001; Tasche 2002). Accordingly, it is our main objective here to discuss efficient MC algorithms for risk measures that do not share the drawbacks of VaR. More precisely, we shall focus on the Utility-based SR which belongs to the class of *convex* risk measures. Even though SR does *not*

share the deficiencies of VaR, its construction parallels the definition of VaR. This similarity allows us to extend the numerical methods for VaR to the case of SR in a rather straightforward manner.

The paper is organized as follows. Section 2 summarizes the definitions and properties of VaR and Utility-Based SR. Section 3 is dedicated to the Normal Copula Model (NCM) underlying *CreditMetrics*. We will propose and test MC algorithms for estimating SR in this model, based on the importance sampling method *exponential twisting* (Glasserman and Li 2003; Glasserman and Li 2005). The paper concludes with a brief summary in Section 4.

2 RISK MEASURES

Section 2.1 recalls the basic properties of VaR. A general classification scheme of risk measures will be discussed in the Section 2.2. Utility-Based SR is defined in Section 2.3.

2.1 Value at Risk

By L we will always denote the overall loss of a credit portfolio over a fixed time horizon, say T. L is a random variable on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Given a random loss variable L, the risk measure VaR at level $\lambda \in (0,1)$ can be defined as

$$
\begin{aligned}
\mathrm{VaR}_\lambda(L) &:= \inf\{c \in \mathbb{R} \mid \mathbb{P}[L - c > 0] \leq \lambda\} \\
&= \inf\{c \in \mathbb{R} \mid \mathbb{P}[L > c] \leq \lambda\} \\
&= \inf\{c \in \mathbb{R} \mid \mathbb{E}[1_{\{L>c\}}] \leq \lambda\}. \quad (1)
\end{aligned}
$$

Here, $\mathbb{E}$ denotes the expected value with respect to the probability measure $\mathbb{P}$, and $1_{\{L>c\}}$ is the indicator function of the event $\{L > c\}$. Thus, VaR corresponds to the quantile of the losses at level λ. 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 choices are $\lambda = 0.05$ or $\lambda = 0.01$).

Due to its simple interpretation, VaR is very popular and widely used in practice nowadays. Nonetheless, VaR is suffering from two severe drawbacks: (i) VaR does not always assess portfolio diversification as beneficial (from the mathematical point of view, this is due to the fact that VaR is a non-convex risk measure, cf. Acerbi and Tasche 2001; Acerbi and Tasche 2002; Tasche 2002; Embrechts, McNeil, and Strautman 2002; Giesecke, Schmidt, and Weber 2005; and also the remarks below); (ii) VaR does not take into account the size of very large losses that might occur in the case of a default event. The latter fact is readily illustrated by the following simple example. Consider two portfolios with loss profiles L_1 and L_2, respectively, where

$$L_1 = \begin{cases} -1 \ \$, & \text{with probability } 99\,\%, \\ +1 \ \$, & \text{with probability } \ 1\,\%, \end{cases}$$

and

$$L_2 = \begin{cases} -1 \ \$, & \text{with probability } 99\,\%, \\ +10^{10} \ \$, & \text{with probability } \ 1\,\% \end{cases}$$

($L_{1,2} < 0$ corresponds to the event 'no loss', whereas $L_{1,2} > 0$ means 'loss'). Then, for $\lambda = 0.01$ one finds

$$\text{VaR}_\lambda(L_1) = \text{VaR}_\lambda(L_2) = -1 \leq 0.$$

Hence, according to this VaR, both portfolios would be equally acceptable. For this example, of course, it is easily seen that the first portfolio is preferable. For more realistic models, however, the amplitude of losses is usually much less obvious. Therefore, risk minimization based on VaR criteria may cause an investment concentration on the portfolio position with the smallest default probability, even if the potential absolute loss associated with this position is ruinously large. This severe shortcoming of VaR has stimulated an intense search for alternatives, leading to the introduction of convex risk measures (Föllmer and Schied 2002a; Föllmer and Schied 2002b; Fritelli and Gianin 2002; Heath and Ku 2004; Föllmer and Schied 2004).

2.2 General Classification of Risk Measures

Artzner et al. (1999) proposed an axiomatic approach, which allows a systematic classification of different risk measures according to rather general properties, as e.g. convexity, homogeneity, and coherence.

Let $\mathscr{D}$ denote the space of financial positions. A risk measure assigns to each financial position a number that quantifies its risk. We assume that $\mathscr{D}$ is some vector space of integrable random variables. For example, we could choose $\mathscr{D}$ as the space of bounded financial positions L^∞ or as the space of financial positions with finite variance L^2. A mapping $\rho : \mathscr{D} \to \mathbb{R}$ is called a distribution-invariant *risk measure* if it satisfies the following conditions for all $X_1, X_2 \in \mathscr{D}$: (i) *Inverse Monotonicity:* If $X_1 \leq X_2$, then $\rho(X_1) \geq \rho(X_2)$. (ii) *Cash Invariance:* If $m \in \mathbb{R}$, then $\rho(X_1 + m) = \rho(X_1) - m$. (iii) *Distribution-invariance:* If the distributions of X_1 and X_2 agree, then $\rho(X_1) = \rho(X_2)$.

Monotonicity refers to the property that risk decreases if the payoff profile is increased. Cash invariance formalizes that risk is measured on a monetary scale: if a monetary amount $m \in \mathbb{R}$ is added to a position X, then the risk of X is reduced by m. Distribution-invariance requires that two positions whose distributions agree have the same risk. VaR is a distribution-invariant risk measure, but it neither encourages diversification nor does it properly account for extreme loss events. Before considering alternative risk measures, we discuss some additional properties:

(a) Convexity formalizes the idea that diversification reduces risk. Let $X_1, X_2 \in \mathscr{D}$ and $\alpha \in [0, 1]$. The position $\alpha X_1 + (1 - \alpha)X_2$ is called diversified. The risk measure ρ is convex, if the risk of any diversified position does not exceed the weighted sum of the risk of the individual positions, i.e. for any $X_1, X_2 \in \mathscr{D}$ and any $\alpha \in [0, 1]$,

$$\rho(\alpha X_1 + (1 - \alpha)X_2) \leq \alpha\rho(X_1) + (1 - \alpha)\rho(X_2).$$

(b) Positive homogeneity is a property which is often required, but which is economically less meaningful. It states that if the size of a position is multiplied by a positive factor, then the associated risk is multiplied by the same factor, i.e. ρ is positively homogeneous if $\rho(\lambda X) = \lambda\rho(X)$ for any $X \in \mathscr{D}$ and any $\lambda \geq 0$. This property neglects the asymmetry between gains and losses. Increasing the size of a position by a factor λ may increase the associated risk by a factor larger than λ if the costs of bearing losses grow faster than their size. For example, the larger a position, the more costly it typically becomes to liquidate it.

(c) Coherence states that a risk measure is both convex and positively homogeneous.

(d) Invariance under randomization formalizes an idea also known as *invariance under compound lotteries* and *invariance under mixing*. Suppose $X_1, X_2 \in \mathscr{D}$ are acceptable with respect to a given risk measure ρ, i.e. $\rho(X_1) \leq 0$ and $\rho(X_2) \leq 0$. Consider the randomized position X given by

$$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1 - \alpha \end{cases} \qquad \alpha \in (0, 1)$$

where α is drawn independently of X_1 and X_2. Should this position be acceptable? After tossing a coin, a bank gets either the acceptable X_1 or the acceptable X_2. From the point of view of a financial institution, it seems reasonable that X should also be accepted. From a descriptive point of view,

for uncertainty averse individuals the answer might be no. From a normative perspective, the uncertainty associated with randomization should not matter. Similarly, if the individual positions are not acceptable with respect to risk measure ρ, i.e. $\rho(X_1) > 0$ and $\rho(X_2) > 0$, then X should also not be acceptable. A risk measure ρ is called invariant under randomization, if randomized positions of acceptable positions are again acceptable and if randomized positions of unacceptable positions are again unacceptable. This property is also closely related to the consistent dynamic measurement of risk (Weber 2006).

VaR is invariant under randomization and positively homogeneous. However, it is never convex, if $\mathscr{D}$ contains L^∞.

2.3 Utility-based Shortfall Risk

VaR does not appropriately measure the risk of the rarely occurring large losses. A useful alternative is provided by the Utility-based SR, defined as follows.

Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function (i.e. ℓ is increasing and not constant). Let λ be a point in the interior of the range of ℓ. To each financial position (payoff profile) X we assign an associated loss variable L by $L := -X$. The space of financial positions $\mathscr{D}$ is chosen such that for $X \in \mathscr{D}$ the expectation $\mathbb{E}[\ell(-X)]$ is well-defined and finite. Then SR with loss function ℓ at level λ is defined by

$$\mathrm{SR}_{\ell,\lambda}(L) := \inf\{c \in \mathbb{R} \mid \mathbb{E}[\ell(L-c)] \leq \lambda\}. \quad (2)$$

Typical examples of convex loss functions are exponential and piecewise polynomial functions, reading

$$\ell_\beta^{\mathrm{exp}}(x) = \exp(\beta x), \qquad \beta > 0; \quad (3a)$$

$$\ell_\gamma^{\mathrm{poly}}(x) = \gamma^{-1} x^\gamma 1_{\{x>0\}}, \qquad \gamma > 1, \quad (3b)$$

with levels $\lambda > 0$ in both cases. In the remainder, we primarily focus on these two examples.

We summarize the properties of SR. SR is convex and therefore encourages diversification. It is invariant under randomization. More generally, it can be shown that SR measures are essentially the only distribution-invariant convex risk measures that are invariant under randomization (Weber 2006). Thus, SR measures are the only distribution-invariant convex risk measures that should be used for the dynamic measurement of risk over time. This important advantage of SR over other risk measures provides a major motivation for developing MC methods for SR measures. At the same time, SR is well suited for measuring extreme events (Giesecke, Schmidt, and Weber 2005). SR measures are always convex, but in general *not* coherent. They are coherent, if and only if their loss functions are of the form $\ell(x) = \lambda - ax^- + bx^+$, $b \geq a \geq 0$ with x^- and x^+ denoting the negative and positive parts of x, respectively.

It is worthwhile to stress the close resemblance between the definitions of SR and VaR. Formally, the SR definition (2) is obtained by replacing the indicator function in Eq. (1) with the convex loss function ℓ. This has the effect that *SR is sensitive to the amplitude of losses, whereas VaR is merely indicating, whether or not the loss L exceeds a certain threshold c with a probability of at least λ.* Hence, risk evaluation based on SR may significantly reduce the risk of unexpected large losses, which might be overlooked by VaR (Giesecke, Schmidt, and Weber 2005).

It remains to discuss how SR can be calculated in practice. The definition (2) is unwieldy for direct numerical simulations, but there exists a comfortable way of determining $\mathrm{SR}_\lambda(L) = \mathrm{SR}_{\ell,\lambda}(L)$ for a given portfolio model. As shown in Prop. 4.104 of Föllmer and Schied (2004), the value $\mathrm{SR}_\lambda(L)$ is given by the unique root s_* of the function

$$f_\lambda(s) := \mathbb{E}[\ell(L-s)] - \lambda. \quad (4)$$

Thus, determining $s_* = \mathrm{SR}_\lambda(L)$ can be subdivided into two partial tasks: (i) Employ a recursive procedure in order to obtain a sequence $s_0, s_1, \ldots$, such that $s_k \to s_*$ as $k \to \infty$. Here, the choice of s_k will be based on the knowledge of the value of the function f_λ at some of the points $s_1, s_2, \ldots, s_{k-1}$. (ii) Given a model for L or for a certain statistics of L, calculate $f_\lambda(s_k)$ at a given point s_k. For this purpose, use MC methods to estimate the expected value $\mathbb{E}[\ell(L-s_k)]$.

The root-finding problem (i) can be treated by means of standard techniques as e.g. the bisection method or the secant method (Press, Vetterling, and Flannery 2002). In the case of the secant method, for example, one has to choose two initial values s_0 and s_1, such that $f_\lambda(s_0) \neq f_\lambda(s_1)$ holds, and the iterative sequence $\{s_k\}$ is obtained via the recursion rule

$$s_{k+1} = \frac{1}{2}\left[s_k + s_{k-1} - (s_k - s_{k-1})\frac{f_\lambda(s_k) + f_\lambda(s_{k-1})}{f_\lambda(s_k) - f_\lambda(s_{k-1})}\right], \quad (5)$$

where $k \geq 1$. Applying other iterative procedures, based on derivatives of $f_\lambda(s)$ with respect to s, as e.g. Newton's method, is less recommendable. The reason is that such derivatives usually involve additional expected values, and thus would require additional MC sampling.

In the remainder, we shall focus on the task (ii), i.e., on the numerically efficient estimation of the expectation value $\mathbb{E}[\ell(L-s_k)]$ for a given portfolio model.

3 NORMAL COPULA MODEL

The NCM (Gupton, Finger, and Bhatia 1997) provides the foundation of the standard credit risk model *CreditMetrics*. The basic equations of the NCM are summarized in Section 3.1. Importance sampling methods for SR will be

discussed in Sections 3.2 and 3.3. Numerical results are presented in Section 3.4.

3.1 Basic Equations

Consider a portfolio with m positions (obligors) over a fixed time horizon, say 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 $L \geq 0$ of the portfolio over the horizon T can be written in the standard form

$$L = \sum_{i=1}^{m} v_i D_i. \qquad (6)$$

The NCM is a particular threshold model, i.e. there exists an m-dimensional random vector $X = (X_1, X_2, \ldots, X_m)$ and threshold levels $x_1, x_2, \ldots, x_m \in \mathbb{R}$ such that

$$D_i = 1_{\{X_i > x_i\}}.$$

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

$$x_i = \Phi^{-1}(1 - p_i), \qquad (7)$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Therefore, instead of directly choosing x_i, one could also specify the marginal default probabilities $p_1, \ldots, p_m$ and determine the threshold values $x_1, \ldots, x_m$ according to Eq. (7). In industry applications, the covariance matrix of the Gaussian vector X is often specified through a factor model of the form

$$X_i = A_{i0} \varepsilon_i + \sum_{j=1}^{d} A_{ij} Z_j, \qquad 1 = A_{i0}^2 + A_{i1}^2 + \ldots + A_{id}^2, \quad (8)$$

where $i = 1, \ldots, m$, $d < m$ and $A_{i0} > 0$, $A_{ij} \geq 0$ are parameters. 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 $X_1, \ldots, X_m$ in the NCM. The second constraint in Eq. (8) ensures that $X_i \sim \mathcal{N}(0, 1)$ holds.

If $X_1, \ldots, X_m$ are specified through the above factor model, the NCM obeys the following conditional independence structure (Glasserman and Li 2005). 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

$$p_i(Z) := P[D_i = 1 | Z] = \Phi\left(\frac{\sum_{j=1}^{d} A_{ij} Z_j - x_i}{A_{i0}} \right). \qquad (9)$$

In principle, it is straightforward to perform numerical MC studies of the NCM on the basis of Eqs. (6)–(9). The NCM model is uniquely determined by the parameter vector $(m, d, p_1, \ldots, p_m, v_1, \ldots, v_m, A_{10}, \ldots, A_{md})$. In a naive MC simulation one first draws the $d + m$ independent random numbers $(\varepsilon_i)_{i=1,\ldots,m}$ and $(Z_j)_{j=1,\ldots,d}$ from a standard normal distribution and then calculates L according to (6). Repeating this procedure several times, one can obtain estimators for functionals of L, e.g., the moments $\mathbb{E}[L^n]$ of the loss distribution, or the loss probabilities $\lambda(c) := \mathbb{P}[L > c] = \mathbb{E}[1_{\{L>c\}}] \in [0, 1]$. Estimating the loss probabilities $\lambda(c)$ is closely related to determining the VaR. In the NCM the total portfolio loss is bounded from above, $0 \leq L \leq L_+$, $L_+ := \sum_{i=1}^{m} v_i$ and it suffices to consider $c \in [0, L_+]$. When measuring downside risk, one is typically interested in estimating $\lambda(c)$ for large values of c. In this case, the straightforward MC method outlined above becomes computationally demanding, since the default events become rare for $c \to L_+$. Glasserman and Li (2003) constructed more efficient estimators of $\lambda(c)$ by applying the importance sampling method exponential twisting. As we will show next, an analogous approach can be used to obtain numerically efficient MC estimators for SR.

3.2 Piecewise Polynomial Loss Function

According to Section 2.3, cf. Eq. (4), we have to construct an MC estimator for the expectation values $\mathbb{E}[\ell(L - s)]$, where $s \in (0, L_+)$ is a fixed loss value, ℓ a given convex loss function, and the distribution of L is determined by the rules of the NCM. Our variance reduction technique parallels the construction of Glasserman and Li (2003) and Glasserman and Li (2005) who considered the estimation of VaR. For practical applications this observation is highly significant: standard techniques for VaR can easily be extended to risk measures that do not share the deficiencies of VaR.

In this section we consider the piecewise polynomial loss function from Eq. (3b), $\ell_\gamma^{\text{poly}}(x) = \gamma^{-1} x^\gamma 1_{\{x \geq 0\}}$ with $\gamma > 1$. In this case, suitable initial values for the secant method are, e.g., given by $s_0 = 0$ and $s_1 = L_+$. An exponential loss function (3a) will be studied in Section 3.3.

3.2.1 Independent Defaults: Exponential Twisting

A particularly simple situation arises in the case of independent default events. In the NCM model, this corresponds to parameters $A_{i0} = 1$, $A_{ij} = 0$, $i = 1, \ldots, m$, $j = 1, \ldots, d$.

The total portfolio loss is $L = \sum_{i=1}^{m} v_i D_i$ with m independent Bernoulli-variables $D_i \in \{0,1\}$ with marginal probabilities $p_i = \mathbb{P}[D_i = 1]$. This case is useful for illustrating the basic idea of exponential twisting.

We aim at estimating $\mathbb{E}_{\mathbb{P}}[\ell(L-s)] = \mathbb{E}_{\mathbb{P}}[h(L)]$ with $h(L) = \ell(L-s)$. Here the subscript $\mathbb{P}$ was introduced to indicate that expectations are calculated with respect to the measure $\mathbb{P}$. If $\mathbb{Q}$ is another probability measure which is equivalent to $\mathbb{P}$ with $\frac{d\mathbb{Q}}{d\mathbb{P}} = g(L)$, then $\mathbb{E}_{\mathbb{P}}[h(L)] = \mathbb{E}_{\mathbb{Q}}\left[\frac{h(L)}{g(L)}\right]$. It follows that $J_n^g = n^{-1}\sum_{k=1}^{n}\frac{h(L_k)}{g(L_k)}$ is an unbiased, consistent estimator of $\mathbb{E}_{\mathbb{P}}[h(L)]$, if the random variables L_k are sampled independently from the distribution of L under $\mathbb{Q}$. Since 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 $\mathrm{var}_{\mathbb{Q}}\left[\frac{h(L)}{g(L)}\right]$ is small. In the present case, we are primarily interested in events which 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 $\mathbb{Q}_\theta$, $\theta \geq 0$, with

$$\frac{d\mathbb{Q}_\theta}{d\mathbb{P}} = \frac{\exp(\theta L)}{\exp[\psi(\theta)]},$$

where

$$\psi(\theta) := \log \mathbb{E}[\exp(\theta L)] = \sum_{i=1}^{m} \log\left[1 + p_i\left(e^{\theta v_i} - 1\right)\right] \quad (10)$$

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 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 $\mathbb{Q}_\theta$. For the individual default probabilities under $\mathbb{Q}_\theta$ we obtain that

$$q_i(\theta) := \mathbb{Q}_\theta[D_i = 1] := \frac{p_i e^{\theta v_i}}{1 + p_i(e^{\theta v_i} - 1)}. \quad (11)$$

As evident from Eq. (11), the new default probabilities q_i do 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 $q_i(0) = p_i$). Hence, under the new measure $\mathbb{Q}_\theta$ default events are more likely to occur. The inverse likelihood ratio for the change from $\mathbb{P}$ to $\mathbb{Q}_\theta$ can

be written as

$$\frac{d\mathbb{P}}{d\mathbb{Q}_\theta} = \prod_{i=1}^{m}\left(\frac{p_i}{q_i(\theta)}\right)^{D_i}\left(\frac{1-p_i}{1-q_i(\theta)}\right)^{1-D_i} \quad (12)$$
$$= \exp[-\theta L + \psi(\theta)].$$

Denoting by $\mathbb{E}$ and $\mathbb{E}_\theta$ the expectations under $\mathbb{P}$ and $\mathbb{Q}_\theta$, respectively, we can write

$$\mathbb{E}[\ell(L-s)] = \mathbb{E}_\theta\left[\ell(L-s)\exp[-\theta L + \psi(\theta)]\right]. \quad (13)$$

Hence, in the case of the piecewise polynomial loss function, importance sampling for $\mathbb{E}[\ell(L-s)] = \mathbb{E}[\gamma^{-1}(L-s)^\gamma 1_{\{L\geq s\}}]$ corresponds to generating samples of the quantity

$$\gamma^{-1}(L-s)^\gamma 1_{\{L\geq s\}} \exp[-\theta L + \psi(\theta)] \quad (14)$$

under the measure $\mathbb{Q}_\theta$. The implementation of the sampling procedure is straightforward because of Eq. (11). The probability distributions of the default indicators under $\mathbb{Q}_\theta$ 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 Eq. (14) is significantly smaller than the variance of the corresponding naive estimator for the lhs. of (13). Since the estimator is unbiased, it is equivalent to consider the second moment,

$$\begin{aligned} M_2(s,\theta) &:= \frac{1}{\gamma^2}\mathbb{E}_\theta\left[(L-s)^{2\gamma}1_{\{L\geq s\}}^2 \exp[-2\theta L + 2\psi(\theta)]\right] \\ &= \frac{1}{\gamma^2}\mathbb{E}\left[(L-s)^{2\gamma}1_{\{L\geq s\}}\exp[-\theta L + \psi(\theta)]\right] \\ &\leq M_2(s,0)\exp[-\theta s + \psi(\theta)]. \quad (15) \end{aligned}$$

Here $M_2(s,0) = \mathbb{E}[(L-s)^{2\gamma}1_{\{L\geq s\}}]$ 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 rhs. of inequality (15). A promising choice for the twisting parameter is thus given by

$$\theta_s = \begin{cases} \text{u. s. of } \psi'(\theta) = s, & s > \psi'(0); \\ 0, & s \leq \psi'(0), \end{cases} \quad (16)$$

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

3.2.2 Dependent Defaults: Conditional Twisting

We return to the general case, where the default events of different portfolio positions are coupled. On the one hand, in this case exponential twisting can be applied to

the conditional distribution $\mathbb{P}[\,\cdot\,|Z]$ of indicator variables D_i. Conditional on Z we are in the situation of the last section, since defaults are conditionally independent for a given Z. On the other hand, further variance reduction can be achieved by applying an additional importance sampling to the factor variables Z (two-step importance sampling).

One-step importance sampling. The basic idea of *conditional exponential twisting* is to replace in the formulae of Sec. 3.2.1 the default probabilities p_i by the conditional default probabilities

$$p_i(Z) := \mathbb{P}[D_i = 1|Z] = \Phi\left(\frac{\sum_{j=1}^d A_{ij}Z_j - x_i}{A_{i0}}\right). \qquad (17)$$

Analogous to Eq. (10), we define the conditional cumulant generating function by

$$\begin{aligned}\psi(\theta, Z) &:= \log \mathbb{E}[\exp(\theta L)|Z] \\ &= \sum_{i=1}^m \log\left[1 + p_i(Z)\left(e^{\theta v_i} - 1\right)\right].\end{aligned} \qquad (18)$$

Analogous to Eq. (16), a variance reducing parameter $\theta_s(Z)$ can be determined, yielding

$$\theta_s(Z) = \begin{cases} \text{u. s. of } \psi'(\theta, Z) = s, & s > \psi'(0, Z); \\ 0, & s \le \psi'(0, Z), \end{cases} \qquad (19)$$

where $\psi'(\theta, z) := \frac{\partial}{\partial \theta}\psi(\theta, z)$ and $\psi'(0, Z) = \mathbb{E}[L|Z] = \sum_{i=1}^m v_i p_i(Z)$. With these definitions, the one-step MC algorithm reads as follows: (i) 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. (ii) Calculate

$$q_i(\theta_s(Z), Z) := \frac{p_i(Z) e^{v_i \theta_s(Z)}}{1 + p_i(Z)\left(e^{v_i \theta_s(Z)} - 1\right)} \qquad (20)$$

with $\theta_s(Z)$ given by Eq. (19) and $p_i(Z)$ given by Eq. (17). (iii) Generate m Bernoulli-random numbers $D_i \in \{0, 1\}$, such that $D_i = 1$ with probability $q_i(\theta_s(Z), Z)$. (iv) Calculate $\psi(\theta_s(Z), Z)$ from Eq. (18) and $L = \sum_{i=1}^m v_i D_i$, and return the estimate

$$\ell(L - s)\exp\left[-L\theta_s(Z) + \psi\left(\theta_s(Z), Z\right)\right]. \qquad (21)$$

Here the exponential factor corresponds to the conditional likelihood ratio, cf. Eq. (12). Steps (i)–(iv) are repeated until the standard deviation of the sampling values (21) has become sufficiently small.

As in the VaR case (Glasserman and Li 2003; Glasserman and Li 2005), 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 $i \ge 1$). Otherwise,

additional importance sampling of the factor variables Z may turn out to be helpful (Glasserman, Heidelberger, and Shahabuddin 1999; Glasserman and Li 2005; Dunkel 2005; Glasserman 2004).

Two-step importance sampling. In addition to the above conditional exponential twisting, one can shift the mean value of the distribution of the factor vector Z from $0 \in \mathbb{R}^d$ to $\mu = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$, in order to achieve further variance reduction (Glasserman and Li 2005). Compared to the one-step algorithm this causes two slight modifications only: Generate in the first step a factor vector $Z \sim \mathcal{N}(\mu, 1_d)$ – instead of $Z \sim \mathcal{N}(0, 1_d)$; return – instead of (21) – the estimate

$$\ell(L - s)\exp\left[-L\theta_s(Z) + \psi\left(\theta_s(Z), Z\right) - \mu^\top Z + \frac{\mu^\top \mu}{2}\right],$$

where $z^\top z := \sum_{j=1}^d z_j^2$. The additional factor $\exp\left(-\mu^\top Z + \mu^\top\mu/2\right)$ is the likelihood ratio for the change from the d-dimensional standard normal distribution $\mathcal{N}(0, 1_d)$ to the d-dimensional normal distribution $\mathcal{N}(\mu, 1_d)$. A useful choice for the shift vector μ may be found by generalizing the arguments of Glasserman, Heidelberger, and Shahabuddin (1999) and Glasserman and Li (2005) to the case of convex loss functions, yielding (Dunkel 2005)

$$\mu \approx \underset{z \in \mathbb{R}^d}{\text{argmax}}\left(F_s(z) - \frac{z^\top z}{2}\right), \qquad (22)$$

where $F_s(z) := -\theta_s(z)\,s + \psi(\theta_s(z), z)$. In practical applications, the optimization problem (22) may be solved numerically.

3.3 Exponential Loss Function

As the second example, consider the exponential loss function $\ell_\beta^{\exp}(x) = \exp(\beta x)$ with $\beta > 0$. In this particular case, the corresponding SR can explicitly be calculated, and one finds

$$\text{SR}_\lambda(L) = \beta^{-1}\left\{\log \mathbb{E}[e^{\beta L}] - \log \lambda\right\}. \qquad (23)$$

It is thus not necessary to apply the iterative root-finding scheme when calculating this risk measure for the NCM.

In the case of independent defaults, Eq. (10) yields the following explicit representation

$$\text{SR}_\lambda(L) = \beta^{-1}\left\{\sum_{i=1}^m \log\left[1 + p_i\left(e^{\beta v_i} - 1\right)\right] - \log \lambda\right\}. \qquad (24)$$

In the case of dependent defaults, Eq. (23) can be rewritten as

$$\mathrm{SR}_\lambda(L) = \frac{1}{\beta}\left[\int_{\mathbb{R}^d} e^{\psi(\beta,z)}\mathrm{d}F(z) - \log\lambda\right], \qquad (25)$$

where $\psi(\beta,z) = \sum_{i=1}^m \log\left[1 + p_i(z)\left(e^{\beta v_i} - 1\right)\right]$ is the conditional cumulant generating function (18), and $\mathrm{d}F(z) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}\sum_{j=1}^d z_j^2\right)\mathrm{d}z_1\ldots\mathrm{d}z_d$. Accordingly, the estimator for the risk measure (25) reads

$$J_n = \frac{1}{n\beta}\sum_{k=1}^n\left\{\prod_{i=1}^m\left[1 + p_i(Z^{(k)})\left(e^{\beta v_i} - 1\right)\right]\right\} - \frac{\log\lambda}{\beta}, \quad (26)$$

where the random vectors $Z^{(k)}$ are sampled from a d-dimensional standard normal distribution, i.e. $Z^{(k)} \sim \mathcal{N}(0,1_d)$, and n is the sample size.

Further variance reduction can be achieved by importance sampling with respect to the factor vector Z. If we restrict attention to measure changes that shift only the mean of Z, a suitable choice of μ can be obtained as a solution of the maximization problem (Dunkel 2005)

$$\mu = \underset{z\in\mathbb{R}^d}{\mathrm{argmax}}\left(\psi(\beta,z) - \frac{z^\top z}{2}\right). \qquad (27)$$

The heuristics for this choice resembles the arguments leading to Eq. (22); cf. Glasserman, Heidelberger, and Shahabuddin (1999) and Glasserman and Li (2005). In practice, the shift-vector μ can be determined numerically from Eq. (27) for a given set of parameters (m,d,p_i,v_i,A_{ij}). Including the likelihood ratio of the measure change from $\mathcal{N}(0,1_d)$ to $\mathcal{N}(\mu,1_d)$, one finds the modified MC estimator

$$\begin{aligned}\hat{J}_n &= \frac{1}{n\beta}\sum_{k=1}^n\left\{\exp\left(-\mu^\top Z^{(k)} + \frac{\mu^\top\mu}{2}\right)\times\right.\\ &\left.\prod_{i=1}^m\left[1 + p_i(Z^{(k)})\left(e^{\beta v_i} - 1\right)\right]\right\} - \frac{\log\lambda}{\beta}\end{aligned} \qquad (28)$$

with $Z^{(k)}$ independently sampled from $\mathcal{N}(\mu,1_d)$.

3.4 Numerical Results

We shall now demonstrate the efficiency of the proposed importance sampling methods by means of numerical simulations of the NCM. More precisely, we will focus on the estimation of the expected values $\mathbb{E}[\gamma^{-1}(L-c)^\gamma 1_{\{L\geq c\}}]$ and $\mathbb{E}[e^{\beta L}]$, being relevant for the cases of SR with piecewise polynomial and exponential loss functions, respectively. In our simulations, we considered a simple test portfolio described by the following parameter set:

- Number of positions (obligors): $m = 10$.
- Size of partial net losses: $v_i = i$, where $i = 1,\ldots,m$ (i.e. financial losses are measured in units of v_1).
- Marginal default probabilities: $p_i = 0.05$, where $i = 1,\ldots,m$. This choice corresponds to threshold values $x_i = 1.645$.
- Number of common risk factors: $d = 3$.
- Coupling coefficients: $A_{ij} = 0.1$, $i = 1,\ldots,m$, $j = 1,\ldots,d$. This choice yields $A_{i0} = 0.985$ for the amplitude of the idiosyncratic risk factor.

For these parameters the maximum possible net loss is given by $L_+ = 55$. Although realistic credit portfolios may contain a larger number of obligors and risk factors, this simple benchmark portfolio suffices already in order to illustrate the efficiency of the importance sampling estimators constructed in the preceding sections In particular, it allows us to compare with results obtained by the naive MC approach (realistic parameter choices for the NCM are discussed by Haaf and Tasche 2002).

3.4.1 Polynomial Loss Function

Figure 1 shows estimates for the expected value $\mathbb{E}[\gamma^{-1}(L-c)^\gamma 1_{\{L\geq c\}}]$ for different sample sizes n and different threshold values c. In these simulations we have chosen the value $\gamma = 2$, and used the pseudo-random number generator of the computer algebra program Mathematica (Wolfram Research Inc. 2000). The results in Fig. 1 (a) were obtained *via* the naive MC method, where L is directly sampled according to the rules of the NCM. Figure 1 (b) shows the corresponding results for the one-step importance sampling method discussed in Sec. 3.2.2. The error bars give the sample standard deviation, which for an estimator of the form $I_n = n^{-1}\sum_{i=1}^n Y_i$ is defined by

$$\hat{s}(I_n) = \left[\frac{1}{n-1}\sum_{i=1}^n(Y_i - I_n)^2\right]^{1/2}.$$

By comparing the two diagrams, one readily observes the significantly improved convergence of the importance sampling estimator. This trend is amplified when increasing the loss threshold c. We mention that, in the case of the naive MC method, for $c \gtrsim 0.7L_+$ and sample sizes $n \leq 10^{4.75}$, as considered in our simulations, the rare event $\{L \geq c\}$ became practically unobservable. In contrast to this, the one-step importance sampling estimators showed a good convergence even in the range of large values $c \to L_+$.

3.4.2 Exponential Loss Function

In the case of SR with exponential loss function $\ell_\beta(x) = e^{\beta x}$ we are interested in determining the expected value $\mathbb{E}[e^{\beta L}]$,
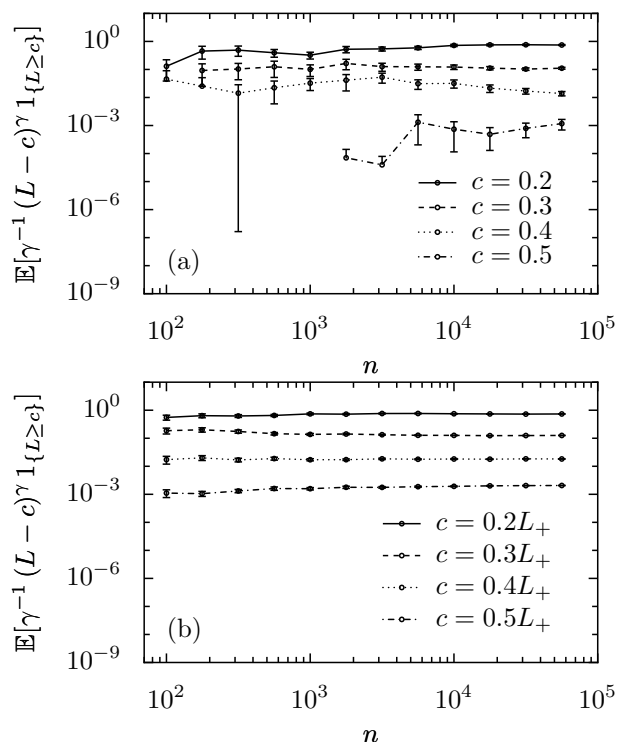
Figure 1: Numerical results for polynomial SR with $\gamma = 2$. The NCM parameters are given in Sec. 3.4. (a) Results based on direct sampling of $\mathbb{E}[\gamma^{-1}(L-c)^{\gamma}1_{\{L \geq c\}}]$ for different sample sizes n. We only included relevant data points corresponding to non-zero estimates. The error bars are determined by the sample standard deviation. (b) Results obtained via the one-step importance sampling method *exponential twisting*. One readily observes the superior convergence of the importance sampling estimator, in particular, for large loss threshold values c.

cf. Sec. 3.3. Figure 2 shows numerical estimates for this quantity, obtained for the same portfolio as before, using the parameter value $\beta = 1$ and different sample sizes n. The solid line corresponds to the naive MC estimator, i.e. direct sampling of $\exp(\beta L)$, while the dashed curve is based on the importance sampling estimator J_n from Eq. (26). As before, error bars indicate the sample standard deviation. This diagram illustrates that Eq. (26) can be used to efficiently estimate the exponential SR in the NCM.

4 CONCLUSION

Evaluating the downside risk of financial positions represents a major task for financial institutions and regulating authorities. In this context, Utility-based Shortfall Risk (SR) provides an excellent alternative to the industry standard Value-at-Risk (VaR). VaR is insensitive to the amplitude of losses beyond the threshold level and does not always value
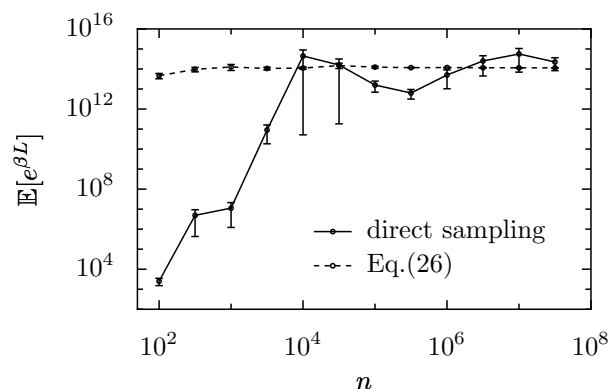

Figure 2: SR with exponential loss function for the NCM test portfolio described in the text. Numerical estimates for $\mathbb{E}[e^{\beta L}]$ with $\beta = 1$ for different sample sizes n. The solid line was obtained by direct sampling of $e^{\beta L}$, whereas the dashed curve is based on the estimator from Eq. (26).

the effect of diversification as beneficial. The convex risk measure SR overcomes these deficiencies. SR provides a sensitive and flexible tool for the measurement of extreme loss events. Moreover, SR allows for consistent dynamic risk evaluation if new information becomes available (Weber 2006).

In this paper, we have discussed how SR can be implemented into the standard credit portfolio model *CreditMetrics*. It has been demonstrated that importance sampling via *exponential twisting*, as proposed by Glasserman and Li (2005) for estimating VaR, can also be used to efficiently estimate SR in this model. Compared with naive Monte Carlo methods, the proposed importance sampling estimators exhibit a significantly improved convergence behavior. We also note that similar methods can be applied to efficiently estimate SR in other factor models as e.g. *CreditRisk*$^+$ (Credit Suisse Financial Products 1997, Dunkel 2005).

To summarize, standard techniques for VaR may easily be extended to convex SR measures that do not share the deficiencies of VaR.

REFERENCES

Acerbi, C., and D. Tasche. 2001. Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes* 31 (2): 379–388.

Acerbi, C., and D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26 (7): 1487–1503.

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

Credit Suisse Financial Products 1997. *CreditRisk$^+$: A Cred-itRisk Management Framework*. London: Credit Suisse Financial Products.

Dunkel, J. 2005, April. Effiziente Monte-Carlo-Methoden für konvexe Risikomaße. Diploma thesis, Institut für Mathematik, Humboldt-Universität zu Berlin.

Embrechts, P., A. McNeil, and D. Strautman. 2002. Correlation and dependency in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, ed. M. Dempster. Cambridge University Press.

Föllmer, H., and A. Schied. 2002a. Convex measures of risk and trading constraints. *Finance and Stochastics* 6 (4): 429–447.

Föllmer, H., and A. Schied. 2002b. Robust representation of convex measures of risk. In *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, 39–56. Springer-Verlag.

Föllmer, H., and A. Schied. 2004. *Stochastic finance - an introduction in discrete time*. 2. ed. Number 27 in de Gruyter Studies in Mathematics. Walter de Gruyter.

Fritelli, M., and E. R. Gianin. 2002. Putting order in risk measures. *Journal of Banking and Finance* 26:1473–1486.

Giesecke, K., T. Schmidt, and S. Weber. 2005. Measuring the risk of extreme events. In *Event Risk*, ed. M. Avellaneda. Risk Books.

Glasserman, P. 2004. *Monte carlo methods in financial engineering*. Number 53 in Applications of Mathematics. New York: Springer.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* 9:117–152.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000a. Importance sampling and stratification for value-at-risk. In *Computational Finance 1999*, ed. Y. S. Abu-Mostafa, B. L. Baron, A. W. Lo, and A. Weigend, 7–24. MIT Press.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000b. Variance reduction techniques for estimating value-at-risk. *Management Science* 46 (10): 1349–1364.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2001. Efficient Monte Carlo methods for value-at-risk. In *Mastering Risk*, Volume 2. Financial Times - Prentice Hall.

Glasserman, P., and J. Li. 2003. Importance sampling for a mixed Poisson model of portfolio credit risk. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice. Piscataway (NJ): IEEE Press.

Glasserman, P., and J. Li. 2005. Importance sampling for portfolio credit risk. *Management Science* 51:1643–1656.

Gupton, C., C. Finger, and M. Bhatia. 1997. *Creditmetrics technical document*. New York: J. P. Morgan & Co. www.riskmetrics.com.

Haaf, H., and D. Tasche. 2002, Juli/August. Calculating value-at-risk contributions in creditrisk$^+$. *GARP Risk Review* 7:43–47.

Heath, D., and H. Ku. 2004. Pareto equilibria with coherent measures of risk. *Mathematical Finance* 14:163–172.

Jorion, P. 2000. *Value at risk*. 2. ed. McGraw-Hill Companies.

Press, W. H., W. T. Vetterling, and B. P. Flannery. 2002. *Numerical recipes in c++: The art of scientific computing*. 2 ed. Cambridge: Cambridge University Press. 0-521-75033-4.

Tasche, D. 2002. Expected shortfall and beyond. *Journal of Banking and Finance* 26 (7): 1519–1533.

Weber, S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance* 16 (2): 419–442.

Wolfram Research Inc. 1988-2000. Mathematica 4.1.0.0. http://www.wolfram.com.

AUTHOR BIOGRAPHIES

JÖRN DUNKEL is a Doctoral Research Associate at the Institute of Theoretical Physics at Universität Augsburg. He previously was a member of the Max-Planck-Institute for Astrophyics. He has studied at Humboldt-Universität zu Berlin and University of Oxford. He holds diplomas in physics and mathematics from the Humboldt-Universität zu Berlin. His present research interests include Monte-Carlo methods in stochastic finance and relativistic stochastic processes. His web page can be found via <www.physik.uni-augsburg.de/~dunkeljo/>.

STEFAN WEBER is an Assistant Professor at the School of Operations Research and Industrial Engineering at Cornell University, Ithaca, New York. He has studied at the University of Hannover, Purdue University, New York University, the European University Institute, and Humboldt-Universität zu Berlin. He holds an M.Sc. in mathematics from the University of Hannover, an M.A. in Economics from the European University Institute in Florence, and a Ph.D. in mathematics from Humboldt-Universität zu Berlin. His research interests include Monte-Carlo methods in risk management, credit valuation, and optimal portfolio choice. His web page can be found via <people.orie.cornell.edu/~sweber/>.