

Importance Sampling Methods for Estimating Convex Risk Measures in Portfolio Credit Risk Models

Jörn Dunkel
Universität Augsburg

Stefan Weber
Cornell University

May 30, 2007*

Abstract

The importance sampling method *exponential twisting* is used to estimate *Utility-based Shortfall Risk* (SR) in two standard portfolio credit risk models. SR belongs to the class of convex risk measures and thus avoids the shortcomings of the industry standard *Value-at-Risk* (VaR). Our analysis demonstrates that standard Monte-Carlo (MC) techniques, originally developed for VaR, can be generalized to efficiently estimate SR in the framework of the portfolio models *CreditRisk⁺* and *CreditMetrics*. Numerical simulations of test portfolios illustrate the good performance of the proposed estimators.

Key words: Portfolio credit risk management, convex risk measures, shortfall risk, importance sampling, exponential twisting

1 Introduction

The reliable measurement of credit default risk represents one of the key issues for financial institutions and regulating authorities. Modern risk management employs portfolio models, as e.g. *CreditMetrics* (Gupton et al., 1997) and *CreditRisk⁺* (Cre, 1997), that take into account potential dependencies among the different obligors. Owing to the complexity of realistic models, quantitative risk analysis typically requires Monte-Carlo (MC) simulations (Glasserman, 2004). These numerical experiments may become computationally expensive, if one wishes to study rare events like credit defaults. Hence, aiming at reducing the simulation time, variance reduction techniques such as importance sampling play an increasingly important role when applying risk analysis to realistic credit portfolio models.

During the past decade an intense effort has been made to develop efficient MC techniques for the industry standard of risk measurement, *Value-at-Risk* (VaR), see Glasserman et al. (2000a,b, 2001); Glasserman and Li (2003, 2005). Although very popular and widely used in

*Jörn Dunkel is at the Institut für Physik, Universität Augsburg, Theoretische Physik I, Universitätsstraße 1, D-86135 Augsburg, Germany, email joern.dunkel@physik.uni-augsburg.de. Stefan Weber is at the School of Operations Research and Information Engineering, 279 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA, email sweber@orie.cornell.edu.

practice (Jorion, 2000), VaR suffers from two severe deficiencies if considered as a measure of downside risk:

- (i) VaR penalizes diversification in many situations;
- (ii) VaR is insensitive to the size of losses beyond the prespecified threshold level.

These serious shortcomings stimulated an intense search for alternative risk measures in recent years (Artzner et al., 1999; Acerbi and Tasche, 2001; Tasche, 2002). In this context, it is our main objective to discuss how the importance sampling methods recently developed for VaR can be generalized to an alternative class of risk measures that avoids the drawbacks of VaR. More precisely, we shall focus on *Utility-based Shortfall Risk* (SR), which belongs to the class of convex risk measures and is defined with help of a convex loss functions ℓ . Even though SR does not share the deficiencies of VaR, its construction parallels to some extent the definition of VaR. This formal analogy is the main reason why numerical methods for the estimation of VaR may be carried over to the case of SR.

The paper is organized as follows. In Section 2 the definitions and properties of VaR and SR are briefly summarized and compared. Section 3 is dedicated to the Normal Copula Model (NCM), representing the foundation of *CreditMetrics* (Gupton et al., 1997). Generalizing the approach of Glasserman and Li (2003, 2005) we will construct efficient MC estimators (which are based on importance sampling method *exponential twisting*) for measuring SR within the NCM of *CreditMetrics*. The good performance of the proposed estimators is illustrated by means of numerical simulations. In Section 4 a similar analysis will be carried out for the Mixed Poisson Model (MPM) which forms the basis of the portfolio model *CreditRisk*⁺ (Cre, 1997). The paper concludes with a brief summary in Section 5.

2 Value-at-Risk *vs.* Utility-based Shortfall Risk

In spite of some well-known deficiencies (Embrechts et al., 2002; Giesecke et al., 2005), VaR is among the most commonly used risk measures in the financial industry. In Section 2.1 we shall recall the definition and basic properties of VaR. Subsequently, the convex risk measure *Utility-based SR*, a useful alternative to VaR, will be defined and discussed in Section 2.2.

2.1 Value-at-Risk

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

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

Here, \mathbb{E} denotes the expected value with respect to the probability measure \mathbb{P} , and $\mathbf{1}_{\{L > c\}}$ 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 values for λ which are used in practice are $\lambda = 0.05$ or $\lambda = 0.01$.

Due its interpretation in terms of a loss probability threshold, VaR has become very popular and is widely used in practice nowadays, cf. Jorion (2000). In particular, considerable effort has been devoted towards developing efficient MC methods for estimating VaR in realistic credit risk and market models, see e.g. Glasserman et al. (2000a,b, 2001); Glasserman and Li (2003, 2005). Unfortunately, VaR suffers from two drawbacks. Firstly, it does not assess portfolio diversification as being beneficial. Mathematically, this is due to the fact that VaR is a non-convex risk measure, cf. discussion in Acerbi and Tasche (2001, 2002); Tasche (2002); Dunkel and Weber (2007). Secondly, 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 with loss profiles L_1 and L_2 , respectively, where

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

and

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

A negative loss value $L_{1,2} \leq 0$ corresponds to the event ‘no loss’, whereas $L_{1,2} > 0$ means ‘loss’. Setting $\lambda = 0.01$, 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.¹ 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 on the portfolio position with the smallest default probability, even if the potential loss associated with this position is extremely large. This severe drawback has stimulated an intense research in recent years, leading among others to the definition of convex SR measures as a useful alternative to VaR, see e.g. Föllmer and Schied (2002a,b, 2004); Frittelli and Gianin (2002); Heath and Ku (2004).

¹Generally, a position is considered acceptable if its risk measure is negative.

2.2 Utility-based Shortfall Risk

A useful alternative to VaR is provided by the convex risk measure *Utility-based Shortfall Risk* (SR), which is defined as follows: Let $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex loss function, i.e. a function that is increasing and not constant, and let λ be a point in the interior of the range of ℓ . Assuming that the expectation $\mathbb{E}[\ell(L)]$ is well-defined and finite, SR of L with loss function ℓ at level λ is defined by

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

Typical examples of convex loss functions are piecewise polynomial and exponential functions, i.e.

$$\ell_{\gamma}^{\text{poly}}(x) = \gamma^{-1}x^{\gamma}\mathbf{1}_{\{x>0\}}, \quad \gamma > 1; \quad (3a)$$

$$\ell_{\beta}^{\text{exp}}(x) = \exp(\beta x), \quad \beta > 0, \quad (3b)$$

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

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 et al., 2005).

We briefly summarize several properties of SR. In contrast to VaR, SR belongs to the class of convex risk measures and, therefore, always encourages diversification; however, even though SR measures are always convex, they are in general *not* coherent (for an axiomatic introduction to risk measures see Föllmer and Schied 2004). It can also be shown that SR measures are the only distribution-invariant, convex risk measures that are invariant under randomization (Weber, 2006).

SR is closely related to the von Neumann-Morgenstern theory of expected utility. Setting $u(x) := -\ell(-x)$, one obtains a concave Bernoulli utility function u , representing the central object in the von Neumann-Morgenstern theory, see e.g. Föllmer and Schied (2004). Defining the utility functional $\mathbb{U}(X) := \mathbb{E}[u(X)]$, where $X := -L$ is the *gain* of a financial position with loss L , Eq. (2) can be rewritten as

$$\text{SR}_{\ell,\lambda}(L) := \inf\{c \in \mathbb{R} \mid \mathbb{U}(-L + c) \geq -\lambda\}. \quad (4)$$

Thus, $\text{SR}_{\ell,\lambda}(L)$ is the smallest monetary amount that has to be added to the portfolio such that its utility is at least $-\lambda$.

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 $\text{SR}_{\lambda}(L) =$

$\text{SR}_{\ell,\lambda}(L)$ for a given portfolio model. As shown in Prop. 4.104 of Föllmer and Schied (2004), the value $\text{SR}_\lambda(L)$ is given by the unique root s_* of the function

$$f_\lambda(s) := \mathbb{E}[\ell(L - s)] - \lambda. \quad (5)$$

Thus, determining $s_* = \text{SR}_\lambda(L)$ can be subdivided into two partial tasks:

(i) Employ a recursive procedure in order to obtain a sequence s_0, s_1, \dots , such that $s_k \rightarrow s_*$ as $k \rightarrow \infty$. 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, \dots, s_{k-1} .

(ii) Given a model or 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 et al. (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

$$s_{k+1} = \frac{1}{2} \left[s_k + s_{k-1} - (s_k - s_{k-1}) \frac{f_\lambda(s_k) + f_\lambda(s_{k-1})}{f_\lambda(s_k) - f_\lambda(s_{k-1})} \right], \quad (6)$$

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

Risk measures like VaR or SR can be used to quantify the downside risk of profit and loss distributions of portfolios over a fixed time horizon. Realistic credit portfolio models capture the dependence among obligors. A standard example is the Normal Copula Model (NCM), which was introduced in Gupton et al. (1997) and provides the foundation of *CreditMetrics*. The basic equations and properties of the NCM are briefly reviewed in Section 3.1. Subsequently, we discuss importance sampling methods for the efficient estimation of SR in the NCM (Sections 3.2 and 3.3). Numerical simulations for simple benchmark portfolios will be presented in Section 3.4.

3.1 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 $i = 1, 2, \dots, m$, 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

$$L = \sum_{i=1}^m v_i D_i. \quad (7)$$

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

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

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

$$x_i = \Phi^{-1}(1 - p_i), \quad (8)$$

where Φ is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. Thus, instead of directly choosing x_i , one could also specify the marginal default probabilities p_1, \dots, p_m and determine the threshold values x_1, \dots, x_m from Eq. (8). In industry applications of the NCM the covariance matrix of the Gaussian vector X is often specified through a factor model of the form

$$X_i = A_{i0}\epsilon_i + \sum_{j=1}^d A_{ij}Z_j, \quad i = 1, \dots, m, \quad d < m; \quad (9a)$$

$$1 = A_{i0}^2 + A_{i1}^2 + \dots + A_{id}^2, \quad A_{i0} > 0, \quad A_{ij} \geq 0. \quad (9b)$$

The *systematic* risk variables Z_1, \dots, Z_d and the *idiosyncratic* risks variables $\epsilon_1, \dots, \epsilon_m$ 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, \dots, X_m . The additional constraint (9b) ensures that $X_i \sim \mathcal{N}(0, 1)$ holds.

If X_1, \dots, X_m are specified through the factor model above, the NCM obeys the following conditional independence structure. Conditionally on the common factors $Z = (Z_1, \dots, Z_d)$, 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

$$p_i(Z) := P[D_i = 1|Z] = \Phi\left(\frac{\sum_{j=1}^d A_{ij}Z_j - x_i}{A_{i0}}\right). \quad (10)$$

In principle, it is straightforward to perform numerical MC studies on the basis of Eqs. (7)–(10). The NCM model is uniquely determined by the parameter vector

$$(m, d, p_1, \dots, p_m, v_1, \dots, v_m, A_{10}, \dots, A_{md}).$$

In a naive MC simulation one first draws the $d + m$ independent random numbers $(\epsilon_i)_{i=1, \dots, m}$ and $(Z_j)_{j=1, \dots, d}$ from a standard normal distribution and then calculates L according to (7).

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

$$\lambda(c) := \mathbb{P}[L > c] = \mathbb{E}[\mathbf{1}_{\{L > c\}}] \in [0, 1]. \quad (11)$$

In the NCM the total portfolio loss is bounded from above, that is $0 \leq L \leq L_+$, $L_+ := \sum_{i=1}^m v_i$ and it suffices therefore to consider $c \in [0, L_+]$. Estimating the loss probabilities $\lambda(c)$ is closely related to determining the VaR. When measuring downside risk, one is typically interested in estimating $\lambda(c)$ for large values of $c \rightarrow L_+$. In this case, the MC method outlined above becomes computationally expensive, since the default events become very rare. Thus, naive MC estimators do not provide good estimates, unless very large sample sizes are considered. Accordingly, variance reduction techniques become very important in practical applications. Glasserman and Li (2003, 2005) constructed efficient estimators for $\lambda(c)$ 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

The SR of a loss variable L is given by the unique root of the function f_λ defined in Eq. (5). As described in Section 2.2, one can apply a recursive algorithm in order to obtain a sequence s_0, s_1, \dots , such that $s_k \rightarrow s_*$ as $k \rightarrow \infty$. The choice of s_k is based on the knowledge of the value of the function f_λ at some of the points s_1, s_2, \dots, s_{k-1} . Assuming that the sequence $s_k \rightarrow s_*$ is determined by the secant iteration rule (6), it remains to discuss how to estimate $\mathbb{E}[\ell(L - s_k)]$ for fixed values $s_k \in (0, L_+)$. Similar to the VaR case, naive MC estimators do not yield reliable results for $\mathbb{E}[\ell(L - s)]$, unless very large sample sizes are considered (to simplify the notation, we write s instead s_k from now on); thus, variance reduction techniques become important. Following Glasserman and Li (2003, 2005), who considered the estimation of VaR, we construct MC estimators for $\mathbb{E}[\ell(L - s)]$ in the NCM by employing the importance sampling method (*conditional exponential twisting*). For practical applications this analogy is highly significant: *Standard variance reduction techniques for VaR can be extended to SR measures that do not share the deficiencies of VaR.*

In this section we shall consider the piecewise polynomial loss (3a) function

$$\ell_\gamma^{\text{poly}}(x) = \gamma^{-1} x^\gamma \mathbf{1}_{[0, \infty)}(x), \quad \gamma > 1.$$

In this case, suitable initial values for the secant method are, e.g., given by $s_0 = 0$ and $s_1 = L_+$. The exponential loss function (3b) will be studied in Section 3.3.

3.2.1 Independent default events: exponential twisting

A particularly simple situation arises in the case of independent default events. In the NCM model, this corresponds to parameters $A_{i0} = 1$, $A_{ij} = 0$, $i = 1, \dots, m$, $j = 1, \dots, d$. 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 (Glasserman, 2004).

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 $\text{var}_{\mathbb{Q}}\left[\frac{h(L)}{g(L)}\right]$ 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}_{θ} , $\theta \geq 0$, with

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

where

$$\psi(\theta) := \log \mathbb{E}[\exp(\theta L)] = \sum_{i=1}^m \log[1 + p_i(e^{\theta v_i} - 1)] \quad (12)$$

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

$$q_i(\theta) := \mathbb{Q}_{\theta}[D_i = 1] := \frac{p_i e^{\theta v_i}}{1 + p_i(e^{\theta v_i} - 1)}. \quad (13)$$

As evident from Eq. (13), 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 i th portfolio position is increased (in particular, we have $q_i(0) = p_i$). Hence, under the new measure \mathbb{Q}_{θ} default events are more likely to occur. The inverse likelihood ratio for the change from \mathbb{P} to \mathbb{Q}_{θ} can be written as

$$\frac{d\mathbb{P}}{d\mathbb{Q}_{\theta}} = \prod_{i=1}^m \left(\frac{p_i}{q_i(\theta)}\right)^{D_i} \left(\frac{1-p_i}{1-q_i(\theta)}\right)^{1-D_i} = \exp[-\theta L + \psi(\theta)]. \quad (14)$$

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

$$\mathbb{E}[\ell(L - s)] = \mathbb{E}_{\theta}[\ell(L - s) \exp[-\theta L + \psi(\theta)]]. \quad (15)$$

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

$$\gamma^{-1}(L - s)^{\gamma} \mathbf{1}_{\{L \geq s\}} \exp[-\theta L + \psi(\theta)] \quad (16)$$

under the measure \mathbb{Q}_θ . The implementation of the sampling procedure is straightforward because of Eq. (13). The probability distributions of the default indicators under \mathbb{Q}_θ 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. (16) is significantly smaller than the variance of the corresponding naive estimator for the lhs. of (15). Since the estimator is unbiased, it is equivalent to consider the second moment,

$$\begin{aligned} M_2(s, \theta) &:= \frac{1}{\gamma^2} \mathbb{E}_\theta [(L - s)^{2\gamma} \mathbf{1}_{\{L \geq s\}}^2 \exp[-2\theta L + 2\psi(\theta)]] \\ &= \frac{1}{\gamma^2} \mathbb{E} [(L - s)^{2\gamma} \mathbf{1}_{\{L \geq s\}} \exp[-\theta L + \psi(\theta)]] \\ &\leq M_2(s, 0) \exp[-\theta s + \psi(\theta)]. \end{aligned} \quad (17)$$

Here $M_2(s, 0) = \mathbb{E}[(L - s)^{2\gamma} \mathbf{1}_{\{L \geq s\}}]$ 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 (17). A promising choice for the twisting parameter is thus given by

$$\theta_s = \begin{cases} \text{u. s. of } \psi'(\theta) = s, & s > \psi'(0); \\ 0, & s \leq \psi'(0), \end{cases} \quad (18)$$

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

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

$$p_i(Z) := \mathbb{P}[D_i = 1 | Z] = \Phi\left(\frac{\sum_{j=1}^d A_{ij} Z_j - x_i}{A_{i0}}\right). \quad (19)$$

Analogous to Eq. (12), we define the conditional cumulant generating function by

$$\psi(\theta, Z) := \log \mathbb{E}[\exp(\theta L) | Z] = \sum_{i=1}^m \log[1 + p_i(Z)(e^{\theta v_i} - 1)]. \quad (20)$$

As in Eq. (18), the parameter θ that governs the measure change can be determined. In the current case, θ depends on the factor Z , i.e.

$$\theta_s(Z) = \begin{cases} \text{u.s. of } \psi'(\theta, Z) = s, & s > \psi'(0, Z); \\ 0, & s \leq \psi'(0, Z), \end{cases} \quad (21a)$$

where

$$\psi'(\theta, z) := \frac{\partial}{\partial \theta} \psi(\theta, z), \quad \psi'(0, Z) = \mathbb{E}[L|Z] = \sum_{i=1}^m v_i p_i(Z). \quad (21b)$$

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

$$q_i(\theta_s(Z), Z) := \frac{p_i(Z) e^{v_i \theta_s(Z)}}{1 + p_i(Z) (e^{v_i \theta_s(Z)} - 1)} \quad (22)$$

with $\theta_s(Z)$ given by Eq. (21) and $p_i(Z)$ given by Eq. (19).

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

4. Calculate $\psi(\theta_s(Z), Z)$ from Eq. (20) and $L = \sum_{i=1}^m v_i D_i$, and return the estimator

$$\ell(L - s) \exp[-L\theta_s(Z) + \psi(\theta_s(Z), Z)]. \quad (23)$$

Here the exponential factor corresponds to the conditional likelihood ratio, compare Eq. (14).

As in the case of VaR (compare Glasserman and Li 2003, 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 $i \geq 1$. Otherwise, additional importance sampling of the factor variables Z may turn out to be helpful, cf. Glasserman et al. (1999); Glasserman and Li (2005); Glasserman (2004); Dunkel (2005).

Two-step importance sampling In addition to the conditional exponential twisting described above, one can shift the mean value of the distribution of the factor vector Z from $0 \in \mathbb{R}^d$ to $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$, 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 (23) – the estimator

$$\ell(L - s) \exp[-L\theta_s(Z) + \psi(\theta_s(Z), Z)] \exp\left(-\mu^\top Z + \frac{\mu^\top \mu}{2}\right), \quad (24)$$

where $z^\top z := \sum_{j=1}^d z_j^2$ denotes the Euclidean scalar product in \mathbb{R}^d .

Here, the additional factor $\exp(-\mu^\top Z + \mu^\top \mu/2)$ 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 $\mathcal{N}(\mu, 1_d)$.

It remains to be discussed how to determine the shift vector μ . For this purpose we may adapt the arguments of Glasserman et al. (1999) and Glasserman and Li (2005) to our setting. The key idea is that the variance of an estimator J_n for $\mathbb{E}[\ell(L - s)]$ can be separated into two summands,

$$\text{var}[J_n] = \mathbb{E}[\text{var}[J_n|Z]] + \text{var}[\mathbb{E}[J_n|Z]]. \quad (25)$$

Conditional exponential twisting reduces the first contribution on the rhs. of Eq. (25). Importance sampling with respect to Z should reduce the second contribution on the rhs. of Eq. (25). For the latter purpose, we consider changes of measure that shift the mean of Z to some $\mu \in \mathbb{R}^d$. By optimal choice of μ we wish to minimize the variance of the conditional mean of the corresponding importance sampling estimators. Applying the arguments of Glasserman et al. (1999) and Glasserman and Li (2005) to our case, we are led to the choice [compare Eq. (20) in Glasserman and Li (2005)]

$$\mu = \underset{z \in \mathbb{R}^d}{\text{argmax}} \mathbb{E}[\ell(L - s)|Z = z] \exp\left(-\frac{z^\top z}{2}\right). \quad (26)$$

However, exact solution of the optimization problem (26) is usually difficult, and one has to employ additional approximations. For our purpose the so-called *tail bound approximation* is particularly useful (for other types of approximations, see Section 5.1 in Glasserman and Li 2005). Defining

$$F_s(z) := -\theta_s(z) s + \psi(\theta_s(z), z) \quad (27)$$

and making use of the general inequality

$$\mathbf{1}_{\{y > x\}} \leq \exp[\theta(y - x)], \quad \theta \geq 0, \quad (28)$$

we find for the piecewise polynomial loss function from Eq. (3a)

$$\begin{aligned} \mathbb{E}[(L - s)^\gamma \mathbf{1}_{\{L > s\}}|Z = z] &\leq \mathbb{E}[|L - s|^\gamma \mathbf{1}_{\{L > s\}}|Z = z] \\ &\leq \mathbb{E}[|L - s|^\gamma e^{\theta_s(Z)(L-s)}|Z = z] \\ &\leq |L_+|^\gamma e^{F_s(z)}. \end{aligned} \quad (29)$$

Here we restricted ourselves to the relevant case $0 \leq s \leq L_+$. Inserting the upper boundary from the last inequality into the optimization problem (26), one obtains

$$\mu \approx \underset{z \in \mathbb{R}^d}{\text{argmax}} \left(F_s(z) - \frac{z^\top z}{2} \right). \quad (30)$$

In practical applications, the simplified optimization problem (30) must be solved numerically (which might still represent a formidable task for portfolios with $m > 2$).

3.3 Exponential loss function

As another example of SR, we consider the exponential loss function $\ell_\beta^{\text{exp}}(x) = \exp(\beta x)$ with $\beta > 0$. In this particular case, the corresponding SR measure can explicitly be calculated, i.e.

$$\text{SR}_\lambda(L) = \frac{1}{\beta} \log \left(\frac{\mathbb{E}[e^{\beta L}]}{\lambda} \right). \quad (31)$$

It is therefore not necessary to apply the iterative root finding scheme when calculating this particular risk measure.

Independent default events In the case of independent defaults we obtain the following explicit representation

$$\text{SR}_\lambda(L) = \frac{1}{\beta} [\psi(\beta) - \log \lambda] \quad (32)$$

with cumulant generating function

$$\psi(\beta) = \log \mathbb{E}[\exp(\beta L)] = \sum_{i=1}^m \log [1 + p_i (e^{\beta v_i} - 1)]. \quad (33)$$

Since $\psi(\beta)$ can be calculated explicitly in the NCM, numerical simulations are not necessary in this case.

Dependent default events In the case of dependent defaults, Eq. (31) can be rewritten as

$$\text{SR}_\lambda(L) = \frac{1}{\beta} \left[\int_{\mathbb{R}^d} e^{\psi(\beta, z)} dF(z) - \log \lambda \right], \quad (34a)$$

where

$$\psi(\beta, z) = \log \mathbb{E}[\exp(\beta L) | Z = z] = \sum_{i=1}^m \log [1 + p_i(z) (e^{\beta v_i} - 1)] \quad (34b)$$

is the conditional cumulant generating function, and the distribution F of the factor variables Z is given by the d -dimensional standard normal distribution

$$dF(z) = \left(\frac{1}{2\pi} \right)^{d/2} \exp \left(-\frac{1}{2} \sum_{j=1}^d z_j^2 \right) dz_1 \dots dz_d. \quad (34c)$$

An estimator for the risk measure (34a) can be obtained by sampling from a Gaussian random vector $Z = (Z_1, \dots, Z_d)$ and returning the value

$$\frac{1}{\beta} \left\{ \prod_{i=1}^m [1 + p_i(Z) (e^{\beta v_i} - 1)] - \log \lambda \right\}. \quad (35)$$

Accordingly, the estimator corresponding to n independent samples $(Z^{(k)})_{k=1, \dots, n}$ is given by

$$J_n = \frac{1}{n\beta} \sum_{k=1}^n \left\{ \prod_{i=1}^m [1 + p_i(Z^{(k)}) (e^{\beta v_i} - 1)] \right\} - \frac{1}{\beta} \log \lambda, \quad Z^{(k)} \sim \mathcal{N}(0, 1_d). \quad (36)$$

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

$$\mu = \operatorname{argmax}_{z \in \mathbb{R}^d} \left(\psi(\beta, z) - \frac{z^\top z}{2} \right). \quad (37)$$

The heuristics for this choice resembles the arguments described in Section 3.2.2. In practice, the optimal shift-vector μ can be determined numerically from Eq. (37) for a given set of parameters (m, d, p_i, v_i, A_{ij}) .

The likelihood ratio of the measure change from $\mathcal{N}(0, 1_d)$ to $\mathcal{N}(\mu, 1_d)$ modifies the MC estimator. The importance sampling estimator is thus given by

$$\tilde{J}_n = \frac{1}{n\beta} \sum_{k=1}^n \left\{ \exp \left(-\mu^\top Z^{(k)} + \frac{\mu^\top \mu}{2} \right) \prod_{i=1}^m [1 + p_i(Z^{(k)}) (e^{\beta v_i} - 1)] \right\} - \frac{1}{\beta} \log \lambda, \quad (38)$$

where now the $(Z^{(k)})$ are independently sampled from $\mathcal{N}(\mu, 1_d)$.

In principle, further improvements are possible. So far, we restricted the admissible measure change to shifts in the mean. By a saddle-point approximation in the vicinity of the maximum μ of (37), a measure change from $Z \sim \mathcal{N}(0, 1_d)$ to $Z \sim \mathcal{N}(\mu, B)$ with a modified covariance matrix B could be constructed, see Jensen (1995). In this case, of course, the likelihood ratio and the estimator had to be modified accordingly.

3.4 Numerical results

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

- Number of positions (obligors): $m = 10$.
- Size of partial net losses: $v_i = i$, where $i = 1, \dots, m$ (i.e. financial losses are measured in units of v_1).
- Marginal default probabilities: $p_i = 0.05$, where $i = 1, \dots, m$. This choice corresponds to threshold values $x_i = 1.645$.
- Number of common risk factors: $d = 3$.
- Coupling coefficients: $A_{ij} = 0.1$, $i = 1, \dots, m$, $j = 1, \dots, d$. 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, for the above parameter values it is still possible to compare with results obtained from naive MC simulations (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 $\mathbb{E}[\gamma^{-1}(L - c)^\gamma \mathbf{1}_{\{L \geq c\}}]$ for different sample sizes n and different threshold values c . In our simulations we have chosen the value $\gamma = 2$, and used the pseudo-random number generator of the computer algebra program Mathematica (Wolfram Research Inc., 1988-2000). The results in Figure 1 (a) were obtained *via* the naive MC method, where L is directly sampled according to rules of the NCM. For comparison, Figure 1 (b) shows the corresponding results for the one-step importance sampling method discussed in Section 3.2.2. The error bars are given by the sample standard deviation, which for an estimator

$$I_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

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 . Additionally, 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 $\{L \geq c\}$ could not be observed anymore. In contrast to this, the one-step importance sampling estimators showed a good convergence even in the range of large c .

In Table 1 we listed numerically obtained values of the sample variance ratio

$$R_n := \hat{s}(\hat{I}_n)^2 / \hat{s}(I_n)^2,$$

where \hat{I}_n refers to the importance sampling estimator and I_n to the naive estimator. The sample variance ratio R_n is a ‘good’ measure of the quality of different MC estimators, only if the sample size n is sufficiently large; e.g. the values for $c = 0.5L_+$ in Table 1 suggest that one requires $n \geq 10^4$ because of the slow convergence the naive estimator. Values $R_n < 1$ indicate a variance reduction due to importance sampling. Hence, analogous to Figure 1, the values in Table 1 illustrate the variance reduction gained by exponential twisting (note that values $R_n = \infty$ signal that the sample size n was not large enough for observing the rare event $\{L \geq c\}$ in the naive simulation).

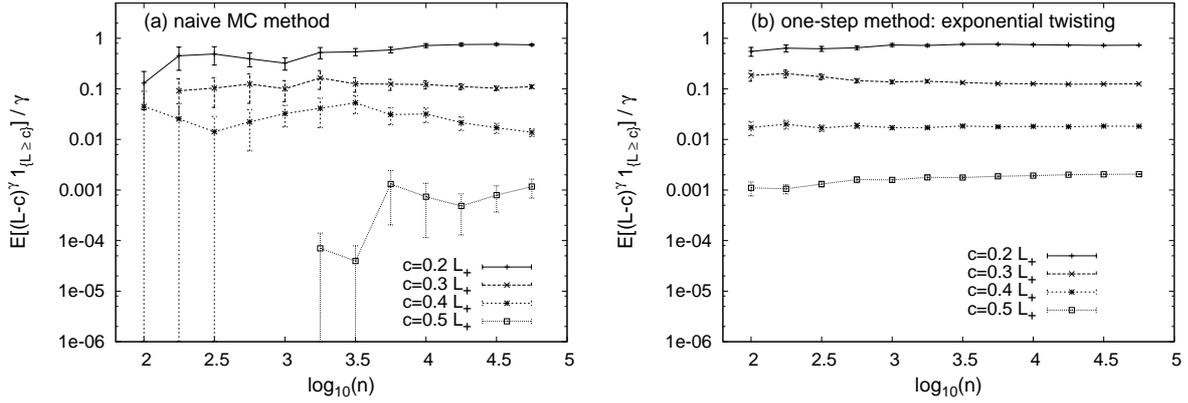


Figure 1: Numerical results for polynomial SR with $\gamma = 2$ in the NCM. The portfolio parameters are given in Section 3.4. (a) Results based on direct (naive) sampling of $\mathbb{E}[\gamma^{-1}(L - c)^\gamma \mathbf{1}_{\{L \geq c\}}]$ for different sample sizes n . We only included relevant data 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 .

3.4.2 Exponential loss function

In the case of SR with an exponential loss function $\ell_\beta(x) = e^{\beta x}$ we are interested in determining the expected value $\mathbb{E}[e^{\beta L}]$. Figure 2 depicts 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 estimate, whereas the dashed curve is based on the estimator (36). As before, error bars indicate the sample standard deviation.

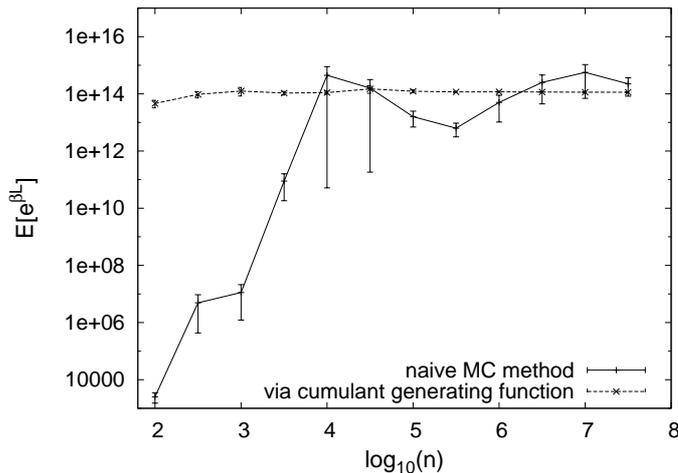


Figure 2: SR with exponential loss function for the NCM test portfolio described in Section 3.4: Numerical estimates for $\mathbb{E}[e^{\beta L}]$ with $\beta = 1$. The solid line was obtained by direct sampling of $e^{\beta L}$, whereas the dashed curve is based on the estimator from Eq. (36).

In view of these results we may briefly summarize: The MC techniques developed by Glasserman and Li (2003, 2005) for estimating VaR in the NCM may be extended to estimate the convex

risk measure SR in this model. As illustrated by our simulations, compared with naive MC methods, the proposed importance sampling algorithms exhibit a significantly improved convergence behavior.

In the next section, we are going to demonstrate that a similar approach can be applied to efficiently estimate SR in other credit risk models as well.

4 Mixed Poisson Model

The Mixed Poisson Model (MPM) forms the foundation of *CreditRisk+* (Cre, 1997). The basic equations of the MPM will be summarized in Section 4.1. Subsequently, we discuss important sampling methods for estimating polynomial SR in this model (Section 4.2), and demonstrate the efficiency of the proposed estimators by means of numerical simulations (Section 4.3). Last but not least, it will be shown in Section 4.4 that exponential SR in the MPM can be calculated analytically.

4.1 Basic equations

Similar to Eq. (7), the total portfolio loss L in the MPM is given by

$$L = \sum_{i=1}^m v_i \bar{D}_i, \quad v_i > 0. \quad (39)$$

However, in contrast to the NCM, the random variables \bar{D}_i are not interpreted as default indicators, but play the role of default counters, taking values in $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. This implies that the loss L will in general not be bounded anymore. Here, each index $i \in \{1, \dots, m\}$ represents a *class* of obligors, and all obligors from class i cause the same potential net losses v_i . That means, if an obligor of class i has defaulted at time horizon T , this leads to a partial loss v_i ; if two obligors of the same class default, then their contribution to the total loss is $2v_i$, etc..

The distributions of the counting variables \bar{D}_i are specified as follows. Given some random vector $X = (X_1, \dots, X_m)$, the variables $(D_i)_{i=1, \dots, m}$ are independent and conditionally Poisson distributed, i.e.

$$\mathbb{P}[\bar{D}_i = k | X] = \frac{X_i^k}{k!} e^{-X_i}, \quad k \in \mathbb{N}_0, \quad i = 1, \dots, m. \quad (40)$$

In industry applications, it is additionally assumed that the random vector X is specified through a factor model of the form

$$X_i = a_{i0} + \sum_{j=1}^d A_{ij} Z_j, \quad d < m. \quad (41)$$

Here, (a_{i0}) and (A_{ij}) are parameters that satisfy the restrictions

- $a_{i0} \geq 0$ for $i = 1, \dots, m$,
- $A_{ij} \geq 0$, $\sum_{j=1}^d A_{ij} > 0$, $\sum_{i=1}^m A_{ij} = 1$ for $i = 1, \dots, m$, $j = 1, \dots, d$.

The common risk factors $(Z_1, \dots, Z_d) = Z$ of the MPM are independent and obey Gamma-distributions with parameters $(\alpha_j, \beta_j) \in (0, \infty) \times (0, \infty)$; i.e., for each factor vector Z the probability density function is given by:

$$f(z) = \prod_{j=1}^d f_j(z_j), \quad f_j(z_j) = \frac{z_j^{\alpha_j-1}}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} \exp\left(-\frac{z_j}{\beta_j}\right), \quad z_j \geq 0. \quad (42)$$

We shall additionally impose the normalization

$$\alpha_j = \frac{1}{\sigma_j^2}, \quad \beta_j = \sigma_j^2, \quad \sigma_j > 0, \quad j = 1, \dots, d. \quad (43)$$

With this convention, we have

$$\mathbb{E}[Z_j] = \alpha_j \beta_j = 1, \quad \text{var}[Z_j] = \alpha_j \beta_j^2 = \sigma_j^2, \quad j = 1, \dots, d, \quad (44)$$

and

$$p_i := \mathbb{E}(X_i) = a_{i0} + \sum_{j=1}^d A_{ij}, \quad i = 1, \dots, m. \quad (45)$$

Thus, the MPM is characterized by the vector of parameters

$$(m, d, p_1, \dots, p_m, v_1, \dots, v_m, \sigma_1, \dots, \sigma_d, A_{11}, \dots, A_{md}).$$

4.2 Piecewise polynomial loss function

Glasserman and Li (2003) discuss how one can efficiently estimate VaR in the MPM. As in the case of the NCM, their methods can be extended to the case convex SR measures that avoid the deficiencies of VaR. In the present section we will outline the main aspects of the corresponding MC algorithm. For this purpose, we will focus on the piecewise polynomial loss function $\ell_\gamma(x) = \gamma^{-1} x^\gamma \mathbf{1}_{\{x \geq 0\}}$ from Eq. (3a). Conceptually, the approach is quite similar to the two-step method discussed in Section 3.2.2.

The first step of the algorithm is given by the conditional exponential twisting of L , using the likelihood ratio

$$h_1(\theta, X) = \exp[-\theta L + \psi(\theta, X)], \quad (46a)$$

where

$$\psi(\theta, X) := \log \mathbb{E}[e^{\theta L} | X] = \sum_{i=1}^m X_i (e^{\theta v_i} - 1) \quad (46b)$$

is the conditional cumulant generating function of L with respect to X . Note that for calculating $\psi(\theta, X)$ one has to make use of the fact that the default counters \bar{D}_i are independent conditional on $X = (X_1, \dots, X_m)$. Then under the changed measure the default counters are again independent Poisson random variables with

$$\mathbb{E}_\theta[\bar{D}_i | X] = X_i e^{\theta v_i}. \quad (47)$$

Evidently, choosing $\theta > 0$ increases the conditional mean of the default indicators \bar{D}_i . Thus, as desired, under the new measure the default events are more likely to occur. In order to achieve further variance reduction one can now additionally consider exponential twisting of the independent factor variables Z_j , by means of the likelihood ratio

$$h_2(\tau, Z) = \exp \left\{ - \sum_{j=1}^d [\tau_j Z_j + \alpha_j \log(1 - \beta_j \tau_j)] \right\}. \quad (48)$$

Here $\tau = (\tau_1, \dots, \tau_d)$ denotes the parameters of the second measure change, while

$$\psi_j(\tau_j) := \log \mathbb{E}[e^{\theta_j Z_j}] = -\alpha_j \log(1 - \beta_j \tau_j) \quad (49)$$

is the cumulant generating function of originally $\text{Gamma}(\alpha_j, \beta_j)$ -distributed variables Z_j . In order to guarantee that Eq. (48) defines a change of measure, it is additionally required that

$$\tau_j < 1/\beta_j, \quad j = 1, \dots, d. \quad (50)$$

Under this condition the following statement holds: With respect to the new measure, parameterized by $\tau = (\tau_1, \dots, \tau_d)$, each of the factor variables (Z_j) are again independent and Z_j obeys a Gamma-distribution with modified parameters

$$\left(\alpha_j, \frac{\beta_j}{1 - \beta_j \tau_j} \right);$$

i.e. exponential twisting maps a Gamma-distribution onto a Gamma-distribution with the same shape parameter but a modified scale parameter.

Combining the two individual changes of measure from Eqs. (46) and (48), the likelihood ratio of the resulting change of measure is given by the product

$$\begin{aligned} h_{12}(\theta, \tau, Z) &= h_1(\theta, X(Z)) h_2(\tau, Z) \\ &= \exp[-\theta L + \psi_{(1)}(\theta) + \psi_{(2)}(\tau) + \psi_{(3)}(\theta, \tau, Z)], \end{aligned} \quad (51)$$

where

$$\begin{aligned} \psi_{(1)}(\theta) &= \sum_{i=1}^m a_{i0} \left(e^{v_i \theta} - 1 \right), \\ \psi_{(2)}(\tau) &= - \sum_{j=1}^d \alpha_j \log(1 - \beta_j \tau_j), \\ \psi_{(3)}(\theta, \tau, Z) &= \sum_{j=1}^d Z_j \left[-\tau_j + \sum_{i=1}^m A_{ij} \left(e^{v_i \theta} - 1 \right) \right]. \end{aligned}$$

For simplicity, one would like to choose the parameters θ and $\tau = (\tau_1, \dots, \tau_d)$ such that the likelihood ratio (51) depends on the factors (Z_j) only through the loss variable L . This can be achieved by setting

$$\tau_j = \sum_{i=1}^m A_{ij} \left(e^{v_i \theta} - 1 \right). \quad (52)$$

Hence, inserting Eq. (52) into Eq. (51) yields the final form of the likelihood ratio for the combined change of measure

$$\frac{d\mathbb{P}}{d\mathbb{Q}_\theta} = \exp[-\theta L + \psi(\theta)], \quad (53a)$$

where

$$\psi(\theta) = \sum_{i=1}^m a_{i0} \left(e^{v_i \theta} - 1 \right) - \sum_{j=1}^d \alpha_j \log \left[1 - \beta_j \sum_{i=1}^m A_{ij} \left(e^{v_i \theta} - 1 \right) \right] \quad (53b)$$

is the cumulant generating function of L under the original measure \mathbb{P} . In contrast to the MPM, in the NCM such a closed representation for the two-step method does not exist (or, at least, has not yet been found).

In the case of the piecewise polynomial loss function $\ell_\gamma(x) = \gamma^{-1} x^\gamma \mathbf{1}_{\{x \geq 0\}}$ from Eq. (3a) we are interested in estimating expected values of the form $\mathbb{E}[\gamma^{-1} (L - s)^\gamma \mathbf{1}_{\{L \geq s\}}]$. By applying the same arguments as in Section 3.2, one finds that for a given value of s a sensible twisting parameter θ should be chosen according to [compare Eq. (18) above]

$$\theta_s = \begin{cases} \text{u. s. of } \psi'(\theta) = s, & s > \psi'(0); \\ 0, & s \leq \psi'(0). \end{cases} \quad (54)$$

Based on these considerations we are now in the position to summarize the main steps of the MC algorithm:

1. Determine θ_s from Eq. (54).
2. Calculate τ_j for $j = 1, \dots, d$ from Eq. (52), using θ_s .
3. Generate $Z_j \sim \text{Gamma} \left(\alpha_j, \frac{\beta_j}{1 - \beta_j \tau_j} \right)$ for $j = 1, \dots, d$.
4. Calculate for $i = 1, \dots, m$ the conditional mean values X_i from Eq. (41).
5. Generate $\bar{D}_i \sim \text{Poisson} (X_i e^{v_i \theta_s})$ for $i = 1, \dots, m$.
6. Calculate $L = v_1 \bar{D}_1 + \dots + v_m \bar{D}_m$ and return the estimator

$$\frac{1}{\gamma} (L - s)^\gamma \mathbf{1}_{\{L \geq s\}} \exp[-\theta_s L + \psi(\theta_s)]. \quad (55)$$

Reiterating the procedure several times and subsequently calculating the sample average gives the MC estimate for $\mathbb{E}[\gamma^{-1} (L - s)^\gamma \mathbf{1}_{\{L \geq s\}}]$. These MC calculations must, of course, again be combined with an iterative algorithm that solves for the root of the function f_λ defined in Eq. (5), cf. discussion at the end of Section 2.2.

4.3 Numerical results

To illustrate variance reduction by exponential twisting in the MPM we performed numerical simulations of a simple benchmark portfolio. The parameters of this portfolio, chosen in accordance with the rules of the MPM, read:

- Number of positions (obligors): $m = 10$.
- Size of partial net losses: $v_i = i$, where $i = 1, \dots, m$ (i.e., financial losses are measured in units of v_1).
- Expected value of the latent variables: $p_i = \mathbb{E}[X_i] = 0.1$ for $i = 1, \dots, m$.
- Number of (common) systematic risk factors: $d = 3$.
- Coupling coefficients: $A_{ij} = 0.01$, where $i = 1, \dots, m$ and $j = 1, \dots, d$. This yields $a_{i0} = 0.07$.
- Variance parameter for the distribution of the factor variables: $\sigma_j = 1$, where $j = 1, \dots, d$.

As in the case of the NCM, we used the pseudo-random number generator of Mathematica (Wolfram Research Inc., 1988-2000) in these simulations.

Figure 3 shows numerical results for the expected value $\mathbb{E}[\gamma^{-1}(L - c)^\gamma \mathbf{1}_{\{L \geq c\}}]$ for varying sample size $n \in [10^2; 10^{4.75}]$ and four different threshold values c . The length of the error bars corresponds to the sample standard deviation \hat{s} . As evident from the diagrams, compared to the naive estimator, the importance sampling algorithm based on exponential twisting is again characterized by significantly better convergence properties, in particular, when increasing the threshold value c . Thus, the algorithm proposed in Section 4.2 does indeed provide a suitable basis for the numerically efficient estimation of polynomial SR in the MPM.

4.4 Exponential loss function

As anticipated above, in the case of the MPM one can calculate analytically the SR associated with the exponential loss function $\ell_\beta(x) = \exp(\beta x)$ from Eq. (3b). Combining the explicit representation from Eq. (31) with the definition of the cumulant generating function we have

$$\text{SR}_\lambda(L) = \frac{1}{\beta} \left(\log \mathbb{E}[e^{\beta L}] - \log \lambda \right) = \frac{1}{\beta} [\psi(\beta) - \log \lambda], \quad (56a)$$

where, according to (53b), for the MPM

$$\psi(\beta) = \sum_{i=1}^m a_{i0} \left(e^{v_i \beta} - 1 \right) - \sum_{j=1}^d \alpha_j \log \left[1 - \beta_j \sum_{i=1}^m A_{ij} \left(e^{v_i \beta} - 1 \right) \right]. \quad (56b)$$

Hence, numerical simulation are not necessary for determining exponential SR in the MPM.

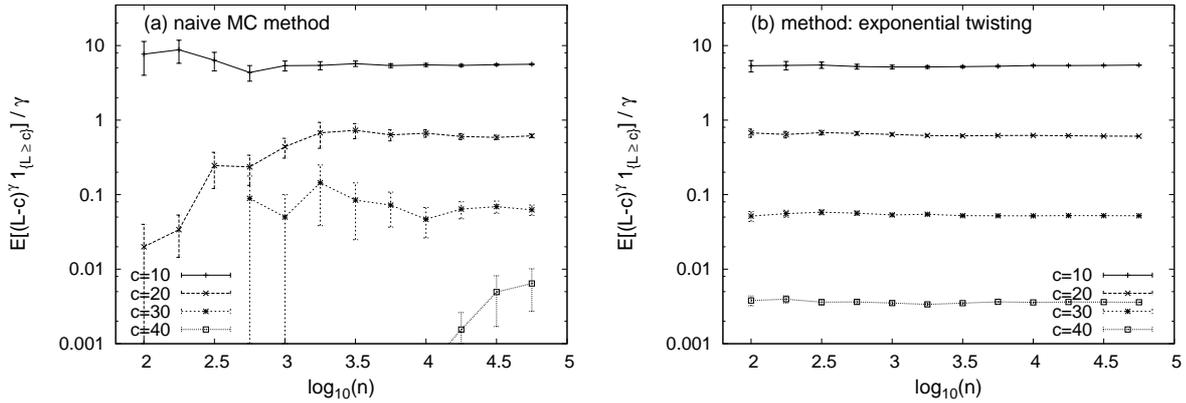


Figure 3: SR with piecewise polynomial loss function in the MPM. The simulation parameters are given in Section 4.3. (a) Numerical results based on the naive MC estimator for $\mathbb{E}[\gamma^{-1}(L - c)^\gamma \mathbf{1}_{\{L \geq c\}}]$ with $\gamma = 2$. (b) Corresponding results obtained by using the exponential twisting algorithm from Section 4.2. Evidently, over a wide range of threshold values c the importance sampling estimator works much more efficiently than the naive estimator.

5 Concluding remarks

Evaluating the downside risk of financial positions represents a major task for financial institutions and regulating authorities. The frequently used risk measure *Value-at-Risk* (VaR) is insensitive to the amplitude of losses beyond the threshold level and it often penalizes diversification. Convex risk measures like *Utility-based Shortfall Risk* (SR) do not share these drawbacks and thus present a useful alternative to VaR. SR provides a sensitive and flexible tool for the measurement of severe loss events, and it allows for consistent dynamic risk evaluation if new information becomes available (Weber, 2006).

In this paper, we have discussed important sampling techniques for determining polynomial and exponential SR in the standard credit portfolio models *CreditMetrics* (Gupton et al., 1997) and *CreditRisk+* (Cre, 1997). It was demonstrated that the variance reduction method *exponential twisting*, employed by Glasserman and Li (2003, 2005) for measuring VaR, can be extended to efficiently estimate SR in these models. Compared with naive Monte Carlo estimators, our importance sampling estimators exhibit a significantly improved convergence behavior. This was illustrated by means of numerical simulations.

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

References

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

- C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.
- P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- CreditRisk⁺: A CreditRisk Management Framework*. Credit Suisse Financial Products, London, 1997.
- J. Dunkel. Effiziente Monte-Carlo-Methoden für konvexe Risikomaße. Diploma thesis, Institut für Mathematik, Humboldt-Universität zu Berlin, April 2005.
- J. Dunkel and S. Weber. Efficient Monte Carlo Methods for Convex Risk Measures in Portfolio Credit Risk Models. In S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, editors, *Proceedings of the 2007 Winter Simulation Conference*, Washington, D.C., 2007. Accepted.
- P. Embrechts, A. McNeil, and D. Strautman. Correlation and dependency in risk management: Properties and pitfalls. In M. Dempster, editor, *Risk Management: Value at Risk and Beyond*. Cambridge University Press, 2002.
- H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002a.
- H. Föllmer and A. Schied. Robust representation of convex measures of risk. In *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, pages 39–56. Springer-Verlag, 2002b.
- H. Föllmer and A. Schied. *Stochastic Finance - An Introduction in Discrete Time*. Number 27 in de Gruyter Studies in Mathematics. Walter de Gruyter, 2. edition, 2004.
- M. Frittelli and E. Rosazza Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26:1473–1486, 2002.
- K. Giesecke, T. Schmidt, and S. Weber. Measuring the risk of extreme events. *Preprint*, 2005.
- P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Number 53 in Applications of Mathematics. Springer, New York, 2004.
- P. Glasserman and J. Li. Importance sampling for a mixed Poisson model of portfolio credit risk. In S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, editors, *Proceedings of the 2003 Winter Simulation Conference*, Piscataway (NJ), 2003. IEEE Press.
- P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Management Science*, 51:1643–1656, 2005.

- P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9:117–152, 1999.
- P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46(10):1349–1364, 2000a.
- P. Glasserman, P. Heidelberger, and P. Shahabuddin. Importance sampling and stratification for value-at-risk. In Y. S. Abu-Mostafa, B. Le Baron, A. W. Lo, and A.S. Weigend, editors, *Computational Finance 1999*, pages 7–24. MIT Press, 2000b.
- P. Glasserman, P. Heidelberger, and P. Shahabuddin. Efficient Monte Carlo methods for value-at-risk. In *Mastering Risk*, volume 2. Financial Times - Prentice Hall, 2001.
- C. Gupton, C. Finger, and M. Bhatia. *CreditMetrics Technical Document*. J. P. Morgan & Co., New York, 1997. www.riskmetrics.com.
- H. Haaf and D. Tasche. Calculating value-at-risk contributions in creditrisk⁺. *GARP Risk Review*, 7:43–47, Juli/August 2002.
- D. Heath and H. Ku. Pareto equilibria with coherent measures of risk. *Mathematical Finance*, 14:163–172, 2004.
- J. L. Jensen. *Saddlepoint approximations*. Oxford University Press, Oxford, 1995.
- P. Jorion. *Value at Risk*. McGraw-Hill Companies, 2. edition, 2000.
- W. H. Press, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C++: The art of scientific computing*. Cambridge University Press, Cambridge, 2 edition, 2002. 0-521-75033-4.
- D. Tasche. Expected shortfall and beyond. *Journal of Banking and Finance*, 26(7):1519–1533, 2002.
- S. Weber. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, 16(2):419–442, 2006.
- Wolfram Research Inc. Mathematica 4.1.0.0. <http://www.wolfram.com>, 1988-2000.

$\log_{10}(n)$	R_n for $c = 0.3L_+$	R_n for $c = 0.5L_+$
2.0	∞	∞
2.5	0.14	∞
3.0	0.06	∞
3.5	0.02	7.76
4.0	0.02	0.01
4.5	0.03	0.01

Table 1: Variance reduction for polynomial SR with $\gamma = 2$ in the NCM. The portfolio parameters were the same as in Figure 1. The variance ratio $R_n = \hat{s}(\hat{I}_n)^2 / \hat{s}(I_n)^2$ is given for different values of the sample size n and two different threshold values c , with I_n denoting the naive estimator and \hat{I}_n the one-step importance sampling estimator from Section 3.2.2. Values $R_n = \infty$ indicate that the rare event $\{L \geq c\}$ could not be observed in the naive simulation up to sample size n ; values $R_n < 1$ indicate variance reduction due to importance sampling.