Optimal Importance Sampling for Credit Portfolios
with Stochastic Approximation

Daniel Egloff    Markus Leippold
Stephan Jöhri

First version: November 2004
Current version: February 2005

This research has been carried out within the NCCR FINRISK project on
“Conceptual Issues in Financial Risk Management”.

National Centre of Competence in Research
Financial Valuation and Risk Management
Optimal Importance Sampling for Credit Portfolios with Stochastic Approximation*

Daniel Egloff
Zürcher Kantonalbank
Markus Leippold
Swiss Banking Institute, University of Zürich
Stephan Jöhri
Swiss Banking Institute, University of Zürich
Curdin Dalbert
Zürcher Kantonalbank

This Version: 12th April 2005. First Version: 19th September 2004

Abstract

We introduce an adaptive importance sampling method for the loss distribution of credit portfolios based on the Robbins-Monro stochastic approximation procedure. After presenting the subtle construction of the algorithm, we apply our adaptive scheme for calculating the risk figures of a typical medium-sized credit risk portfolio with 2000 obligors. Simulating the tail of the loss distribution, we can improve significantly the variance reduction and outperform other recently proposed importance sampling approaches that are based on deterministic methods providing asymptotically optimal importance sampling distributions. Furthermore, the simple structure of the algorithm not only allows a straightforward implementation, but also offers a lot of flexibility for extensions to more complex models. Therefore, our numerical results motivate interesting future research paths for the application of stochastic approximation methods in risk management.

JEL Classification Codes: C15, C61, G21.

Key Words: Credit risk, Monte-Carlo simulation, importance sampling, stochastic approximation, Robbins-Monro.

*Part of this research has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK). The NCCR FINRISK is a research program supported by the Swiss National Science Foundation.

1Correspondence Information: Zürich Kantonalbank, P.O. Box, CH-8010 Zürich, Switzerland. Phone: +41 1 292 45 33, mailto:daniel.egloff@zkb.ch
1 Introduction

In the finance and insurance industry, Monte Carlo simulation is a tool of paramount importance. For risk management, the focus lies on the tails of the distribution of the underlying processes and their calculation is the prerequisite for measuring risk, e.g., in terms of value-at-risk or expected shortfall. Unfortunately, for this task, standard Monte Carlo simulation is highly inefficient because it yields estimators of large variance. The resulting risk measures may be highly inaccurate or may even lead to wrong business decisions.

Therefore, there is a strong practical need for enhancing Monte Carlo simulation. Such enhancements are particularly challenging, when the underlying distribution exhibits rare events and complex dependence structures that influence the tails. These features are prominently present in credit risk management. Large losses from highly rated debtors and the need, not only for practitioners but also for regulators, to capture dependence within credit portfolios, increases significantly the complexity of the computational methods required to calculate sensible risk numbers in an efficient way.

While increasing the sample size is one technique for reducing the standard error of a crude Monte Carlo estimator, doing so can be computationally very expensive. A better solution is to employ some techniques of variance reduction. These techniques incorporate additional information about the analysis directly into the estimator to decrease the standard error of the Monte Carlo estimator.

In this paper, we focus on importance sampling which is one particular technique of variance reduction. In the credit risk literature, importance sampling has been suggested recently by Avranitis and Gregory (2001), Glasserman (2004), Glasserman and Li (2004), Kalkbrener, Lotter and Overbeck (2004), and Merino and Nyfeler (2004).
In contrast to crude Monte Carlo simulation, importance sampling places more weight on the sample range of interest, thereby making the simulation more efficient. However, it is generally far from obvious how such a change of measure should be obtained in a practical manner. When the distribution of interest is connected to a large deviation principle, a common procedure is to apply the change of measure suggested by the large deviation lower bound. Such an approach may yield an asymptotically optimal importance sampling distribution which, however, is not always guaranteed as shown in Glasserman and Wang (1997).

For Gaussian copula models of credit portfolios, a natural importance sampling measure is a shifted Gaussian, possibly combined with a Bernoulli twisting of the individual default probabilities as suggested by Glasserman and Li (2004), Kalkbrener et al. (2004), and Merino and Nyfeler (2004). To calculate a viable translation vector, Glasserman and Li (2004) derive an upper bound on the second moment of the importance sampling estimator of the tail probability. Then, they solve a deterministic optimization problem to find the translation vector that minimizes the upper bound. They show that the corresponding importance scheme is asymptotically optimal. However, to achieve asymptotic optimality, the number of obligors and the loss level for the tail probability needs to increase to infinity. For practical algorithms such assumptions might be unrealistic. Moreover, the importance sampling distribution is fixed, i.e., the variance reduction does not improve with an increasing sample size.

We suggest a Robbins-Monro stochastic approximation method to calculate the optimal translation vector by simulation. The method is adaptive. Every new sample is used to update the translation vector. Since the translation vector depends on the previous samples, we leave the i.i.d. setting. However, martingale limit theorems as presented in Hall and Heyde (1980) can be applied to derive empirical confidence intervals or to show asymptotic normality for the estimators. Using stochastic approximation for importance sampling, we can improve the variance reduction for a
sample credit portfolio by a multiple. Depending on the confidence level and the risk measure used, we outperform the deterministic methods applied in Glasserman and Li (2004) and Kalkbrener et al. (2004) by a factor that ranges between 2 and 40. Moreover, our adaptive scheme has the advantage that the variance reduction further improves with the sample size of the simulation. The above mentioned deterministic methods do not share this property.

Robbins-Monro methods and stochastic approximation date back to the historical work of Robbins and Monro (1951) and Kiefer and Wolfowitz (1952), see also Kushner and Clark (1978). More recent pointers to the literature are Benveniste, Metiver and Priouret (1990), Ljung, Pflug and Walk (1992), and Kushner and Yin (1997). Our adaptive algorithm is strongly guided by the work of Arouna (2003), who successfully applies the truncated Robbins-Monro algorithm of Chen, Guo and Gao (1988) for pricing options under different assumptions on the underlying process.

The paper is structured as follows. Section 2 introduces the credit risk model that underlies the subsequent Monte Carlo simulations. In Section 3, we present our stochastic approximation method for importance sampling. Section 4 applies our stochastic approximation method to a typical credit risk portfolio and compares our method with two other recent methods for importance sampling. Section 5 concludes.

2 Portfolio Credit Risk Model

To illustrate the effectiveness of our approach, we consider a simple credit risk model based on a multivariate Gaussian copula model. Such models are widely used in practice and can be found in many industry applications such as in J.P. Morgan’s CreditMetrics system (see, e.g., Gupton, Finger and Bhatia (1997)). The model assumes that an $N$-dimensional Gaussian latent variable, $A$, drives the creditworthiness of each single obligor $i$, $i = 1, \ldots, N$, within the credit portfolio. We
assume that we can represent $A$ as

$$
\mathbb{R}^N \ni A = \text{diag} \left( \sqrt{1 - v^2} \right) Y + \text{diag} \left( v \right) \epsilon. 
$$

(2.1)

The vector $Y \sim N(0, \Gamma) \in \mathbb{R}^N$, with $\Gamma \in \mathbb{R}^{N \times N}$ a positive definite correlation matrix, captures the dependence between the obligors. The obligor-specific risk is represented by the standard normal vector $\epsilon \sim N(0, I_N)$. We assume that $Y$ and $\epsilon$ are independent. Consequently, $Y$ serves as a conditional independence structure for the latent variable $A$. The vector $0 < v \leq 1$ assigns to each obligor the corresponding weight for the underlying factors $Y$ and the idiosyncratic risk, respectively. Note that every coordinate of $A$ is standard normal.

In the credit model (2.1), dependence between obligors may arise through industry or geographic dependence, but may also just arise from estimation without any direct economic interpretation. In general, the number of these influencing factors is much smaller than the number of debtors in the credit portfolio. Therefore, for our numerical implementation in Section 4, we simplify the structure of the model further and we group the debtors into a small number $K \ll N$ of different sectors according to some classification scheme. The resulting classification function

$$
s : \{1, \ldots, N\} \to \{1, \ldots, K\},
$$

(2.2)

maps every debtor to one of the $K$ sectors. To model the sector effects, we assume some systematic sector-specific risk factors that are given by

$$
Z = (Z_s)_{s=1,\ldots,K} \sim N(0, \Sigma), \quad \Sigma = (\sigma_{i,j})_{i,j=1,\ldots,K} \in \mathbb{R}^{K \times K}.
$$

(2.3)

Then, we can represent the vector of systematic factors on the debtor specific level and its covariance
matrix as
\[ Y = (Z_{s(i)})_{i=1,...,N} \sim N(0, \Gamma), \quad \Gamma = (\sigma_{s(i),s(j)})_{i,j=1,...,N}, \] (2.4)

and, for each debtor \( i \), we can rewrite the latent variable \( A_i \) as
\[ A_i = \sqrt{1 - \nu^2_{s(i)}} Z_{s(i)} + \nu_{s(i)} \epsilon_i, \quad A_i \sim N(0, 1). \] (2.5)

In the next step, we introduce default into our credit risk model. The default indicator for a debtor \( i \) is given by
\[ \chi_i = \mathbb{I}_{\{A_i < -\theta_i\}}. \] (2.6)

where \( \theta_i \) is the default threshold. The threshold values \( \theta = (\theta_i) \), the correlation matrix \( \Sigma \) and the specific risk weights \( \nu \) are free model parameters, which need to be calibrated to historical default data. We do not discuss calibration issues here. Instead, we refer to Gordy (2000) and Egloff, Leippold and Vanini (2003) for details. In our simple setting, the default indicator \( \chi_i \) is Bernoulli distributed with
\[ \mathbb{E}[\chi_i] = \Phi(\theta_i) = p_i, \] (2.7)

with \( p_i \) is the default probability of obligor \( i \). Moreover, conditional on the common factor \( Z \), the default indicators \( \chi_i \) are independent and
\[ p_i \mid Z = p_i(Z) = \Phi \left( \frac{\theta_i - \sqrt{1 - \nu^2_{s(i)}} Z_{s(i)}}{\nu_{s(i)}} \right). \] (2.8)

For notational convenience, we will drop the indices \( s(i) \) in the subsequent analysis. No confusion should occur. We note that we can extend the above credit risk model to a setting with multiple credit ratings. The adaptive importance sampling method presented in the next section can easily
be adopted to such a model setup. However, for expository purpose, we content ourselves with analyzing importance sampling for the model setup presented above.

### 3 Adaptive Importance Sampling for Credit Risk

Importance sampling rests on the simple fact that the expectation under one probability measure can be expressed as an expectation under a different measure, if the change of measure is compensated by the likelihood ratio given as the Radon-Nikodym derivative of the original measure with respect to the alternative measure. Often, an explicitly available likelihood ratio can be used to find a parametric family

$$
\mathbb{E}[G(X)] = \mathbb{E}[g(\mu, X)]
$$

for the expectation of interest. If the expectation is going to be calculated by a Monte Carlo estimator

$$
\frac{1}{N} \sum_{n=1}^{N} g(\mu, X_n),
$$

the most advantageous choice for $\mu$ is the variance minimizer

$$
\mu^* = \arg \min_{\mu} V(\mu),
$$

where

$$
V(\mu) = \mathbb{E} [g(\mu, X)^2] - \mathbb{E} [G(X)]^2.
$$

Then, $\mu^*$ minimizes the variance of the estimator (3.2) in the parametric family (3.1). However, in many cases and even if $X$ is a multivariate Gaussian, it is difficult or even impossible to derive a good approximation of the variance optimal solution $\mu^*$. Therefore, we suggest an adaptive Monte
Carlo algorithm which replaces the estimator in equation (3.2) by

\[ \frac{1}{N} \sum_{n=1}^{N} g(\mu_n, X_n), \]  

(3.5)

where \( \mu_n \) is a stochastic sequence constructed in such a way that

\[ \lim_{n \to \infty} \mu_n = \mu^* = \arg \min_{\mu} V(\mu), \]  

(3.6)

almost sure. Under suitable assumptions on \( g \) and the distribution of \( X \) we can construct such a sequence by solving the first order condition

\[ \nabla_{\mu} V(\mu) = 0 \]  

(3.7)

with a recursive Robbins-Monro type algorithm. If the gradient of \( V \) can be expressed as an expectation \( \nabla_{\mu} V(\mu) = \mathbb{E}[F(\mu, X)] \), the prototypical form of a Robbins-Monro algorithm is a recursive scheme

\[ \mu_{n+1} = \mu_n - \gamma_{n+1} F(\mu_n, X_{n+1}), \]  

(3.8)

where \( \gamma_n \) is a sequence of positive possibly random numbers satisfying

\[ \sum_{n \geq 0} \gamma_n = \infty, \quad \sum_{n \geq 0} \gamma_n^2 < \infty \]  

(3.9)

almost sure. For more details on Robbins-Monro and stochastic approximation algorithms we refer to Kushner and Clark (1978), Ljung et al. (1992), and Kushner and Yin (1997).

We now specialize the Robins-Monro algorithm to the estimation of tail measures of the credit portfolio loss distribution. Let \( L = L(Z, \epsilon) \) be the total loss from defaults over a given period as a
function of the common factors \( Z \sim N(0, \Sigma) \) and the obligor specific risk \( \epsilon \), i.e.,

\[
L(Z, \epsilon) = \sum_i c_i \mathbb{I}_{\{A_i \in (-\infty, \theta_i]\}},
\]

(3.10)

with \( c_i \) the credit exposure of obligor \( i \). For simplicity, we assume \( c_i \) to be constant, which gives rise to a discrete loss distribution. To estimate the tail probabilities, the function of interest is

\[
G(Z, \epsilon) = \mathbb{I}_{\{L(Z, \epsilon) > \lambda\}}.
\]

(3.11)

Other widely used measures are quantiles and conditional tail expectations, often labeled expected shortfall. For an important contribution to the theory of (coherent) risk measures, we refer to the seminal work of Artzner, Delbaen, Eber and Heath (1999). For precise definition of conditional tail expectations and expected shortfall we refer to, e.g., Acerbi and Tasche (2001).

Let \( q_\alpha(L) \) be the (lower) \( \alpha \)-quantile or inverse cumulative distribution function defined as

\[
q_\alpha(L) = \inf \{ l \mid P(L \leq l) \geq \alpha \}.
\]

(3.12)

A commonly used estimator is given by

\[
\hat{q}_{n, \alpha}(L) = L_{\lfloor \alpha n \rfloor: n},
\]

(3.13)

where \( L_{1:n} \geq \ldots \geq L_{n:n} \) is a decreasing orders statistics obtained from \( n \) i.i.d. samples of the loss distribution. Optimizing \( \mu \) based on the function (3.11) also leads to good results for the quantile estimator (3.13), although it reduces the variance of the estimator for the cumulative distribution function but not for its inverse. Note that, if \( X \) is a random variable with a density \( f \), the
asymptotic variance of the quantile estimator \( \hat{q}_{n,\alpha}(X) \) defined in (3.13) is given by \( \frac{a(1-\alpha)}{n f(q_{\alpha})^2} \), see for instance Schervish (1997, Theorem 7.25). However, the loss distribution \( L \) is discrete. Therefore, the above result on the asymptotic variance of the quantile estimator does not apply in our case.

The Monte Carlo estimator for the conditional tail expectation is

\[
\overline{TC}_{n,\alpha}(L) = \frac{1}{\alpha n} \sum_{i=1}^{\lfloor \alpha n \rfloor} L_{i:n}.
\] (3.14)

The function of interest to improve the estimation of \( \overline{TC}_{n,\alpha}(L) \) would be

\[
G(Z, \varepsilon) = L(Z, \varepsilon) \mathbb{1}_{\{L(Z, \varepsilon) > \lambda\}}.
\] (3.15)

for some \( \lambda \) in the range of \( q_{\alpha}(L) \). Note that using the function \( G \) as defined in (3.15) to find a suitable translation vector \( \mu \) just reduces the conditional variance of \( \overline{TC}_{n,\alpha}(L) \) given the sample quantile \( \hat{q}_{n,\alpha}(L) \). For additional details on the asymptotic variance of the estimator \( \overline{TC}_{n,\alpha}(L) \) we refer to Manistre and Hancock (2003).

Let \( \phi_{\Sigma}(z) \) be the density of \( Z \). Parameterizing by translates of \( Z \) leads to the representation

\[
\mathbb{E}[G(Z, \varepsilon)] = \mathbb{E}\left[ G(Z + \mu, \varepsilon) \frac{\phi_{\Sigma}(Z + \mu)}{\phi_{\Sigma}(Z)} \right],
\] (3.16)

which is of the form (3.1) with

\[
g(\mu, Z, \varepsilon) = G(Z + \mu, \varepsilon) \exp \left( -\mu \Sigma^{-1} Z - \frac{1}{2} \mu' \Sigma^{-1} \mu \right),
\] (3.17)
For the second moment of \(g(\mu, Z, \epsilon)\), we obtain

\[
V(\mu) \equiv \mathbb{E} \left[ g(\mu, Z, \epsilon)^2 \right] = \mathbb{E} \left[ G(Z + \mu, \epsilon)^2 \frac{\phi_\Sigma(Z + \mu)^2}{\phi_\Sigma(Z)^2} \right] = \mathbb{E} \left[ G(Z, \epsilon)^2 \frac{\phi_\Sigma(Z)}{\phi_\Sigma(Z - \mu)} \right]. \tag{3.18}
\]

A straightforward application of Hölder’s inequality shows that \(V\) is twice differentiable and strictly convex in \(\mu\). Furthermore \(V(\mu) \to \infty\) for \(\|\mu\| \to \infty\). Consequently, there exists a unique minimizer of

\[
\arg\min_{\mu} V(\mu), \tag{3.19}
\]

which yields a variance optimal translation \(\mu^*\). The first order condition for (3.19) becomes

\[
0 = \nabla_\mu \mathbb{E} \left[ G(Z, \epsilon)^2 \exp \left( -\mu \Sigma^{-1} Z + \frac{1}{2} \mu' \Sigma^{-1} \mu \right) \right] \tag{3.20}
\]
\[
= \mathbb{E} \left[ \sqrt{\Sigma^{-1}} (\mu - Z) G(Z, \epsilon)^2 \exp \left( -\mu' \Sigma^{-1} Z + \frac{1}{2} \mu' \Sigma^{-1} \mu \right) \right], \tag{3.21}
\]

where by \(\sqrt{\Sigma}\) we denote the Cholesky factorization of \(\Sigma\) such that \(\Sigma = \sqrt{\Sigma} \sqrt{\Sigma}\). Again, the exchange of integration and differentiation is justified by Hölder’s inequality. For a detailed proof, we refer to Arouna (2003, Proposition 2.1).

If we base our Robbins-Monro procedure on the first order condition (3.21), the algorithm may not converge due to numerical overflow problems caused by the exponential term \(\exp(1/2 \mu' \Sigma^{-1} \mu)\). Depending on the sign of the component of \(\mu\), this term may either become too large or too small. To resolve this numerical instability, we replace (3.21) by the equivalent first order condition

\[
0 = \mathbb{E} \left[ (\mu - Z) G(Z, \epsilon)^2 \exp \left( -\mu' \Sigma^{-1} Z \right) \right]. \tag{3.22}
\]
Given the expression (3.22) for the gradient, we obtain the following recursive procedure

\[ \mu_{n+1} = \mu_n - \gamma_{n+1} Y_{n+1}, \]  

(3.23)

with the stochastic gradient \( Y_{n+1} \) defined as

\[ Y_{n+1} = (\mu_n - Z_{n+1}) G(Z_{n+1}, \epsilon_{n+1})^2 \exp \left( -\mu_n^T \Sigma^{-1} Z_{n+1} \right). \]  

(3.24)

However, the assumptions generally imposed to prove convergence of Robbins-Monro algorithms do not cover the recursion (3.23). The first difficulty lies in the lack of the a-priori boundedness of the algorithm. This could be resolved by projecting to a fixed compact set. However, doing so is inherently difficult, because we have no prior information on the bounded region to which the optimal translation vector \( \mu^* \) belongs. The second and more severe problem is the growth of the stochastic gradient \( Y_{n+1} \).

To obtain a convergent algorithm, Chen et al. (1988) suitably modify the algorithm by truncating at randomly varying bounds. The resulting algorithm is given by

\[ \mu_{n+1} = (\mu_n - \gamma_{n+1} Y_{n+1}) 1_{J_n} + \bar{\mu}_n 1_{J_n^c}, \]  

(3.25)

with \( J_n^c \) the complement of the event \( J_n \). The event \( J_n \) is defined as

\[ J_n = \{ \|\mu_n - \gamma_{n+1} Y_{n+1}\| \leq R \epsilon_n \}. \]  

(3.26)
where $R_n$ is a sequence of positive numbers tending to $\infty$ and

$$\tau_0 = 0, \quad \tau_n = \sum_{i=0}^{n} \mathbb{1}_{I_i}$$

(3.27)

is the number of truncation up to iteration $n$. One possible truncation operation is a resetting procedure given by

$$\mu_n = \begin{cases} \tilde{\mu}_n & \text{if } \tau_n = 2n, \\ \tilde{\mu}_{n+1} & \text{if } \tau_n = 2n + 1, \end{cases}$$

(3.28)

where $\tilde{\mu}_1 \neq \tilde{\mu}_2$ are two arbitrary fixed points. The convergence proof for the truncated Robbins-Monro algorithm (3.25) relies on the fact that, if for some $\epsilon < 1$

$$\limsup_{n \to \infty} \gamma_n \sum_{i=1}^{n} Y_i - \mathbb{E}[Y_i | \mu_i] \leq \epsilon,$$

(3.29)

almost sure, then the truncation is only invoked a finite number of times. Furthermore, if (3.29) holds for $\epsilon = 0$ then the asymptotic properties of (3.25) are comparable to those of the Robbins-Monro algorithm under the classical assumptions and (3.25) converges to the unique optimizer $\mu^*$, see Chen et al. (1988), and Chen and White (2002). For the verification of (3.29) we refer to Arouna (2003, Theorem 3).

To implement the truncated Robbins-Monro algorithm (3.25), the crucial choice is the specification of the weighting sequence $\gamma_n$. Arouna (2003) proposes

$$\gamma_{n+1} = \frac{\alpha}{\beta + (n + 1)},$$

(3.30)

where $\alpha, \beta$ are two tunable parameters. Here, we suggest a random weighting sequence to improve the stability and convergence of the algorithm. In particular, we want to be sure that the denom-
inator in (3.30) increases by a small amount $\Delta$, only when the loss is larger than the threshold $\lambda$. Otherwise the denominator becomes too large at the beginning of the simulation and, as a consequence, the updated translation vector $\mu_n$ will not change fast enough. This consideration leads to a weighting sequence of the form

$$
\gamma_{n+1} = \frac{\alpha}{\beta + \Delta \cdot \sum_{i=1}^{n} 1\{L_i > \lambda\}}.
$$

(3.31)

We must assure that such a manipulation of $\gamma$ still fulfills conditions (3.9). This is straightforward to show. If $P(L > \lambda) > 0$, then it follows as a direct consequence of the strong law of large numbers that

$$
\sum_{i=1}^{n} 1\{L_i > \lambda\} \approx nP(L > \lambda).
$$

(3.32)

Therefore, equation (3.32) implies for the specification of $\gamma$ in (3.32) that the conditions (3.9) are valid almost sure.

We finally note that numerical overflows may also result from very large values coming from $G$. Therefore, instead of using (3.15), we do not consider absolute losses. Instead, we use percentage losses and normalize $L$ by $L/L_{\text{max}}$ to increase the numerical stability of the procedure.

4 Comparison with Previous Importance Sampling Schemes

We compare our adaptive method with two recently proposed importance sampling schemes. Glasserman and Li (2004) suggest a two step procedure changing the distribution of the conditional default probabilities of every obligor given the factors $Z$ by an exponential Bernoulli tilting, combined with a translation of the Gaussian factors $Z$. The translation is derived as the solution of an optimization problem relying on Glasserman, Heidelberger and Shahabuddin (1999), see also
Glasserman (2003) for further details and explanations.¹

The second method is due to Kalkbrener et al. (2004). They replace the original portfolio by a homogeneous portfolio of obligors with the identical exposure and default probability, and use this structurally simpler portfolio to determine an approximation of the optimal translation vector for the expected shortfall estimator stepwise by lifting the solution of the one-dimensional problem to the multi-dimensional case.

For the sake of comparability, we do not consider in our simulation study the full two-step importance sampling scheme of Glasserman and Li (2004) and omit the exponential Bernoulli tilting. At this point, we would like to stress that our adaptive scheme can also be combined with an exponential Bernoulli tilting at the individual obligor level as in Glasserman and Li (2004). This would lead to further variance reduction.

In contrast to our adaptive stochastic approximation, the deterministic methods of Glasserman and Li (2004), Kalkbrener et al. (2004), and Merino and Nyfeler (2004) justify their efficiency by asymptotic optimality. However, in practice credit portfolios are of finite size. Furthermore, the interest lies on a particular quantile, or expected shortfall above a specific level or quantile, and not on the asymptotic tail of the loss distribution. Hence, from a practitioner’s viewpoint, we are particularly interested in the performance of the above methods when applied to calculating the risk figures of a typical middle-sized credit portfolio, which shows rather skewed distribution of default probabilities and exposures.

The portfolio structure for our simulations has partially been motivated by the model portfolios presented in the work of Gordy (2000). There, the author makes reference to a study of the Federal Reserve Board analyzing the creditworthiness of large banks that presents three model portfolios

¹The Bernoulli tilting approach has been independently proposed in Merino and Nyfeler (2004).
consisting of mainly low, average and mainly highly rated companies. Gordy (2000) uses Standard & Poors rating from AAA to CCC. We choose the data from the low rated portfolio with a rating distribution as displayed in Table 1.

<table>
<thead>
<tr>
<th>Our Rating Classes</th>
<th>Standard &amp; Poors Rating Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion in Portfolio</td>
<td>AAA</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 1: Rating distribution for test portfolio.

Based on the above rating distribution we create a portfolio consisting of 2000 obligors. The total net-exposure of the portfolio is CHF 5.9 billion. For every obligor we draw at random, depending on his rating class, a net-exposures from a representable portfolio of commercial loans for firms based in the area of Zurich, Switzerland. The net-exposure distribution is given in Table 2 below.

<table>
<thead>
<tr>
<th>Min Net-Exposure</th>
<th>Max Net-Exposure</th>
<th>Proportion in Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 0$</td>
<td>$\leq 2 \times 10^5$</td>
<td>0.4108</td>
</tr>
<tr>
<td>$&gt; 2 \times 10^5$</td>
<td>$\leq 4 \times 10^5$</td>
<td>0.1140</td>
</tr>
<tr>
<td>$&gt; 4 \times 10^5$</td>
<td>$\leq 6 \times 10^5$</td>
<td>0.1174</td>
</tr>
<tr>
<td>$&gt; 6 \times 10^5$</td>
<td>$\leq 8 \times 10^5$</td>
<td>0.0780</td>
</tr>
<tr>
<td>$&gt; 8 \times 10^5$</td>
<td>$\leq 10 \times 10^5$</td>
<td>0.0606</td>
</tr>
<tr>
<td>$&gt; 10 \times 10^5$</td>
<td>$&lt; \infty$</td>
<td>0.2192</td>
</tr>
</tbody>
</table>

Table 2: Exposure distribution for test portfolio.

The construction of our credit portfolio reflects the characteristics of a typical credit risk portfolio for a middle-sized bank with local characteristics grouped into the BAK industry sectors from the Konjunkturforschung Basel AG. The BAK industry sectors are listed in Table 3.

Our simulations are based on a sample size of 50'000 runs. To emphasize the fact that deterministic methods based on asymptotic analysis may fail for finite samples, Figure 1 plots the result from applying the importance sampling method as proposed in Glasserman and Li (2004) (without Bernoulli tilting). The plot shows the tail of the loss distribution of the credit portfolio simulation.
<table>
<thead>
<tr>
<th>Sector</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Food, Agriculture</td>
</tr>
<tr>
<td>2</td>
<td>Textiles</td>
</tr>
<tr>
<td>3</td>
<td>Paper and Printing, Wood Manufacturing</td>
</tr>
<tr>
<td>4</td>
<td>Chemical</td>
</tr>
<tr>
<td>5</td>
<td>Plastics</td>
</tr>
<tr>
<td>6</td>
<td>Metal and Machine</td>
</tr>
<tr>
<td>7</td>
<td>Electronics</td>
</tr>
<tr>
<td>8</td>
<td>Other Manufacturing, Recycling</td>
</tr>
<tr>
<td>9</td>
<td>Construction</td>
</tr>
<tr>
<td>10</td>
<td>Retail, Commerce</td>
</tr>
<tr>
<td>11</td>
<td>Hotel and Restaurant</td>
</tr>
<tr>
<td>12</td>
<td>Traffic, Communication, Energy, Water</td>
</tr>
<tr>
<td>13</td>
<td>Finance and Insurance</td>
</tr>
<tr>
<td>14</td>
<td>Real Estate</td>
</tr>
</tbody>
</table>

Table 3: Sector Classification

The sector classification corresponding to BAK (Konjunkturforschung Basel AG) data.

The loss level is at $2.5 \times 10^8$. From the 95% confidence interval, we observe that the importance sampling method applied to a typical credit risk portfolio yields a higher variance than the crude Monte Carlo method. At first sight, this result may come at a surprise, but points at the severe problem of asymptotic methods. The failure of the method of Glasserman and Li (2004) for our test portfolio can be attributed to the fact that their method is optimal only in an asymptotic sense.

For our test portfolio, applying the method proposed in Kalkbrener et al. (2004) gives a similar picture. At the 95% confidence interval, Monte Carlo simulation with importance sampling gives a slightly higher variance than the crude Monte Carlo method. We illustrate our simulation results in Figure 2.

In Figure 3, we show the results using the adaptive method proposed in Section 3 to find the optimal shift for importance sampling. Again, the plot shows the tail of the loss distribution of the credit portfolio simulation. Clearly, at the 95% confidence level, the Robbins-Monro algorithm gives rise to a lower variance around the specified loss level than crude Monte Carlo simulation.
Figure 1: Importance sampling with the method of Glasserman and Li (2004), without Bernoulli tilting. The plot shows the tail of the loss distribution of the credit portfolio simulation and the corresponding 95% confidence intervals. The sample size of the simulation is 50'000 and the loss level is fixed at $2.5 \cdot 10^8$.

It is now interesting to see how the above results influence the calculation of risk calculations for the credit portfolio. Therefore, we consider next the impact of the effectiveness of the different importance sampling methods on the accuracy of risk figures such as value-at-risk and expected shortfall. Figure 4 plots the variance reduction for the three methods discussed above. As expected, it turns out that the Robins-Monro algorithm dominates by large the approaches of Glasserman and Li (2004) and Kalkbrener et al. (2004).

Compared to Glasserman and Li (2004), the Robbins-Monro algorithm improves the variance reduction for the value-at-risk figure by a factor of 4 to 14 depending on the confidence level.
Figure 2: Importance sampling with the method of Kalkbrener et. al (2004). The plot shows the tail of the loss distribution of the credit portfolio simulation and the corresponding 95% confidence intervals. The sample size of the simulation is 50,000 and the loss level is fixed at $2.5 \cdot 10^8$.

Similarly, in comparison with Kalkbrener et al. (2004), this factor is between 2 and 8. The domination of the Robbins-Monro algorithm becomes even more apparent for the (coherent) risk measure of expected shortfall. As an example, compared with Glasserman and Li (2004), the improvement is between a factor of 4 and 60. This factor may turn out to be even larger when we include an exponential Bernoulli tilting at the individual obligor level and compare it with the corresponding method of Glasserman and Li (2004). Finally, we recall that, since the Robbins-Monro algorithm is a stochastic approximation method, the variance reduction will improve even further when we increase the sample size for simulation.
Figure 3: Importance sampling with the Robbins-Monro algorithm developed in Section 3. The plot shows the tail of the loss distribution of the credit portfolio simulation and the corresponding 95% confidence intervals. The sample size of the simulation is 50’000 and the loss level is fixed at $2.5 \cdot 10^8$.

5 Conclusions and Outlook

There are many different ways to implement importance sampling for Monte Carlo simulations. Most methods apply an asymptotically optimal and deterministic change of drift. For practical purposes, since these methods are based on asymptotic arguments, they may sometimes fail. Indeed, we find that, for a typical middle-sized credit portfolio, some of these importance sampling techniques do not perform better than crude Monte Carlo simulation.

Motivated by this failure of asymptotic methods, we propose a method for importance sampling that is based on the adaptive stochastic approximation method of Robbins-Monro. To test
and compare the different importance sampling methods, we set up a typical medium-sized credit portfolio. We find that the stochastic approximation method gives rise to significant improvements for the tail approximation and, consequently, for the calculation of the relevant risk figures.

The simple structure of the algorithm allows not only a straightforward implementation, but also offers enough flexibility for extensions to more complex models. Therefore, our numerical results on the risk management of credit portfolios open interesting future research paths for the application of stochastic approximation methods in risk management. Finally, it would be interesting to consider extensions to computationally more involved models such as, e.g., the contagion model of Egloff et al. (2003), to asynchronous and distributed versions running on HPC clusters.
Figure 4: Variance reduction for value-at-risk and expected shortfall. The upper panel shows the variance reduction for value-at-risk calculations at different confidence levels using the importance sampling method in Glasserman and Li (2004) (dash-dotted line), the method used in Kalkbrener et al. (2004) (dashed line), and our method based on Robbins-Monro (solid line). The lower panel plots the corresponding variance reduction for the different methods when expected shortfall is calculated.
References


