Efficient Computation of Portfolio Credit Risk with Chain Default

Ikeda, Yuki

Individual

16 March 2021

Efficient Computation of Portfolio Credit Risk with Chain Default

Yuki Ikeda[*]

March 16, 2021

Abstract

Many banks consider the chain default or bankruptcy when they compute the credit loss distribution. One way to consider the chain default is the good-old Monte Carlo simulation, however, it is typically time-consuming. In this paper, we extend the efficient Monte Carlo simulation using the importance sampling introduced by Glasserman and Li (2005) to realize an efficient Monte Carlo simulation of the Value at Risk (VaR) that allows the chain defaults. In addition, we see that another method, the saddle point approximation, can also be modified for the case of the chain defaults. Moreover, we give a simple method of shifting the means of the multivariate factors using the well-known EM-algorithm to further reduce the variance of the simulated VaR. Simulation studies show that these proposed methods have superior numerical performance.

Keywords: Value-at-risk; Risk contributions; Importance sampling; Saddle point approximation; EM-algorithm.

1 Introduction

Most banks compute the credit loss distribution to conduct internal risk management and satisfy Basel regulations (Pillar two). When the chain reaction of defaults or bankruptcies within a corporate group happens, a bank that has credit exposures to the group may incur a large loss at a time. Therefore, many banks consider the possible chain reaction of defaults or bankruptcies when they compute the credit loss distribution.

Banks typically compute VaR or expected shortfall (ES) to capture the characteristics of their credit loss distributions and use them for their risk management. There are many ways to compute VaR and ES, and the most common one is to use Monte Carlo simulations. However, since it is typically time-consuming, there are many studies of efficient simulations of VaR, ES, and other risk indices. Among others, Glasserman and

---
[*]E-Mail: pt2y1003@gmail.com

Li (2005) devised an efficient Monte Carlo simulation using the importance sampling. Specifically, they proposed twisted default probability so that tail loss events would be no longer rare in the sense that the (conditional) expectation of the loss with respect to the twisted probabilities is identical to the pre-specified loss level. However, since they do not consider chain defaults within a corporate group, their method needs to be extended to allow chain defaults. In this paper, we modify the twisted default probability suggested by Glasserman and Li (2005) in order for the efficient Monte Carlo simulation to allow chain defaults. We show that our modification can be applied to compute the risk contributions (allocations) of individual credit exposures.

There are also some analytical methods that approximately derive VaR or ES to avoid time-consuming Monte Carlo simulation. One of the most common analytical methods is the saddle point approximation explored in Muromachi (2004), Martin and Ordovas (2006), Huang, *et al.* (2007), and others. We show that the method using the saddle point approximation can also be modified to allow chain default.

These modifications are to reduce the variance of the Monte Carlo simulation that comes from the exposure concentration. In addition, Glasserman and Li (2005) proposed to reduce the variance of the Monte Carlo simulation that comes from the high correlations among the obligers by shifting the mean of the univariate common factor in an additional importance sampling. When the common factors are multivariate, the situation becomes complicated, as shown by Glasserman, *et al.* (2007). They suggested a normal mixture distribution for the mean vector of the multivariate factors in the additional importance sampling. However, some complicated optimizations are required to derive the desired mixture normal distribution. Alternatively, we propose a simpler method to derive the desired mixture normal distribution using a well-known EM algorithm.

Lastly, for both of these proposed methods in this paper (twisting the default probability and shifting the factor mean), simulation studies show that these proposed methods have superior numerical performance.

The rest of this paper is organized as follows. Section 2 introduces the setting and briefly reviews the efficient Monte Carlo simulation using the importance sampling proposed by Glasserman and Li (2005). Section 3 proposes an efficient Monte Carlo simulation using the importance sampling that allows chain defaults. Section 4 proposes a method of saddle point approximations in the case of chain defaults. Section 5 proposes a simple method to shift multivariate factor means using (weighted) EM-algorithm to reduce the variance due to the correlations among the obligers. Section 6 conducts numerical experiments to verify the performance of the methods proposed in section 3 and 5. Section 7 concludes.

2   Tail probability of Credit Portfolio: Setting and Importance Sampling

Consider computing the tail probability and then VaR (Value at Risk) in the multivariate factor model founded by Merton (1974). In this model, the default (or bankruptcy)

indicators Y_k's are determined by latent variables X_k's that represent (the inverse of) the corporate values. Specifically, assume there are m corporations in our credit portfolio and

$$L = c_1 Y_1 + \cdots + c_m Y_m,$$

where for $1 \leq k \leq m$, c_k is a constant representing the credit exposure, or the loss resulting from the default of the corporation k, $Y_k = 1$ if $X_k > x_k$, and $Y_k = 0$ if $X_k \leq x_k$, $X_k \sim N(0,1)$, $x_k = \Phi^{-1}(1 - p_{0k})$, p_{0k} is the (unconditional) default probability of the corporation k, and Φ is the standard normal distribution function.

Our goal is to estimate the tail probability $P(L > L_0)$ for a given large constant $L_0 > 0$. X_k's are further specified to introduce correlations as

$$X_k = a_{k1} Z_1 + \cdots + a_{kd} Z_d, + \sqrt{1 - (a_{k1}^2 + \cdots a_{kd}^2)} \; \varepsilon_k,$$

where the systematic risk factors $Z_1, \ldots, Z_d$ and the idiosyncratic factors ε_k's are independent standard normal random variables, and $\boldsymbol{a}_k = (a_{k1}, \ldots, a_{kd})^T \in \boldsymbol{R}^d$ with $a_{k1}^2 + \cdots + a_{kd}^2 < 1$. In practice, $\boldsymbol{a}_k$ varies according to the industry or the nationality of the corporation k.

Upon fixed $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$, Y_k's are conditionally independent, and the conditional default probability is

$$p_k(\boldsymbol{Z}) = P(Y_k = 1|\boldsymbol{Z}) = \Phi \left(\frac{\boldsymbol{a}_k^T \boldsymbol{Z} + \Phi^{-1}(p_{0k})}{\sqrt{1 - (a_{k1}^2 + \cdots + a_{kd}^2)}} \right), \tag{2.1}$$

so one can estimate $P(L > L_0)$ by the Monte Carlo simulation; generate $\boldsymbol{Z}$ first, generate Y_k's next, and compute L. However, since L_0 is typically very large in real applications, it is rare that the computed L is larger than L_0. As a result, we need many number of iterations, and it is time-consuming. Glasserman and Li (2005) proposed to use the importance sampling to realize more efficient simulation. Namely, they introduced a parameter $\theta > 0$ and twisted the conditional probability $p_k(\boldsymbol{Z})$ to

$$q_k(\theta|\boldsymbol{Z}) := \frac{p_k(\boldsymbol{Z}) e^{\theta c_k}}{1 + p_k(\boldsymbol{Z})(e^\theta - 1)}.$$

Note that $q_k(\theta)$ is always larger than p_k. Then, the tail probability, which is our goal, is

$$P(L > L_0) = E_{\boldsymbol{Z}} \left[E_{\boldsymbol{q}(\theta|\boldsymbol{Z})} \left[1(L > L_0) LR(\{p_k(\boldsymbol{Z})\}, \{q_k(\theta|\boldsymbol{Z})\}) \right] \right], \tag{2.2}$$

where $E_{\boldsymbol{Z}}[\cdot]$ is the expectation with respect to $\boldsymbol{Z}$, $E_{\boldsymbol{q}(\theta|\boldsymbol{Z})}[\cdot]$ is the expectation with respect to the twisted probabilities $q_k(\theta|\boldsymbol{Z})$ conditioned on $\boldsymbol{Z}$, and $LR(\{p_k(\boldsymbol{Z})\}, \{q_k(\theta|\boldsymbol{Z})\})$ is the likelihood ratio relating the original distribution of Y_k's to the new one:

$$LR(\{p_k(\boldsymbol{Z})\}, \{q_k(\theta|\boldsymbol{Z})\}) = \Pi_{k=1}^m \left(\frac{p_k(\boldsymbol{Z})}{q_k(\theta|\boldsymbol{Z})} \right)^{Y_k} \left(\frac{1 - p_k(\boldsymbol{Z})}{1 - q_k(\theta|\boldsymbol{Z})} \right)^{1-Y_k}$$
$$= \exp(-\theta L + \psi(\theta|\boldsymbol{Z})).$$

Since the cumulant generating function of L is

$$\psi(\theta|\mathbf{Z}) := \log(E[e^{\theta L}|\mathbf{Z}]) = \Sigma_{k=1}^m \log(1 + p_k(\mathbf{Z})(e^{\theta c_k} - 1)),$$

it can be easily verified that the twisted conditional expectation of L is $E_{\mathbf{q}(\theta|\mathbf{Z})}[L] = (d/d\theta)\psi(\theta|\mathbf{Z})$. Noting that $(d/d\theta)\psi(\theta|\mathbf{Z})$ is monotonically increasing, by choosing $\theta^*(\mathbf{Z})$ satisfying $(d/d\theta)\psi(\theta^*|\mathbf{Z}) = L_0$, one can shift the distribution of L so that $E_{\mathbf{q}(\theta^*|\mathbf{Z})}[L] = (d/d\theta)\psi(\theta^*|\mathbf{Z}) = L_0$, i.e., L_0 is now the mean of L under $\mathbf{q}(\theta^*|\mathbf{Z})$. Then, it is no longer rare that $L > L_0$, and the resulting simulation that empirically calculates (2.2) becomes much efficient. Namely, one can compute the tail probability by first generating $\mathbf{Z}$, then generating each Y_k by the probability $q_k(\theta^*(\mathbf{Z})|\mathbf{Z})$, and taking the average of

$$1(L > L_0)LR(\{p_k(\mathbf{Z})\}, \{q_k(\theta^*(\mathbf{Z})|\mathbf{Z})\}) = 1(L > L_0)\exp(-\theta^*(\mathbf{Z})L + \psi(\theta^*(\mathbf{Z})|\mathbf{Z})).$$

From the next section, we abbreviate $\mathbf{Z}$ unless we emphasize the dependence on $\mathbf{Z}$ such as $\psi(\theta) := \psi(\theta|\mathbf{Z})$ and $q_k(\theta) := q_k(\theta|\mathbf{Z})$.

3 Extension to the Case of Chain Defaults: Importance Sampling

The derivation in the last section does not consider the chain default or bankruptcy. However, in practice, many banks consider the chain default or bankruptcy within the corporate group when they compute the credit loss distribution. This motivates us to modify the framework above to allow the chain defaults.

Specifically, suppose there are corporate groups in the credit portfolio. A corporate group is consisted of m^{par} parent companies indexed by $k = 1, \ldots, m^{par}$ with $m^{par} \leq m$ and their multiple subsidiary companies indexed by $j = 1, \ldots, J(k)$ with $\sum_{k=1}^{m^{par}} J(k) = m$.

Hereinafter, we denote the case of the chain defaults by attaching a **tilde** on the top of each notation in the last section, and by attaching **par** and **sub** on the right upper to refer to the **parent** and **subsidiary** companies respectively. Note that it is possible to represent an independent company k that has neither a parent company nor subsidiary companies by taking $c_{k,j}^{sub} = 0$ for all j. Also, in practice, the industries or the nationalities of the parent and of its subsidiary companies can be different, so we allow $\mathbf{a}_{k,j}^{sub} \neq \mathbf{a}_k^{par}$.

There are many ways to introduce the chain defaults, but we only consider the most conservative case here. In short, (i) the default of a parent company means the default of all of the subsidiary companies under the parent company. (ii) If the parent company does not default, then each subsidiary company defaults only due to its own inherent reason. Namely, the parent company k defaults by probability $\tilde{p}_k^{par}$ same with (2.1), and all of the subsidiary companies $j = 1, \ldots, J(k)$ default with probability $\tilde{p}_{k,j}^{sub} = 1$ if their parent company defaults. Otherwise, the subsidiary companies default with probability

$$\tilde{p}_{k,j}^{sub} := P(Y_{k,j}^{sub} = 1|\mathbf{Z}) = \Phi\left(\frac{(\mathbf{a}_{k,j}^{sub})^T\mathbf{Z} + \Phi^{-1}(p_{0k,j}^{sub})}{\sqrt{1 - ((a_{k,j1}^{sub})^2 + \cdots + (a_{k,jd}^{sub})^2)}}\right).$$
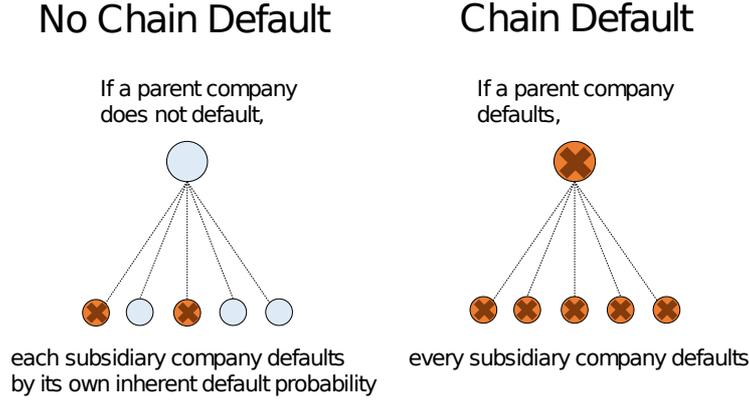
Figure 1: The notion of chain defaults

We also give Figure 1 to explain this. In real applications, this conservativeness is acceptable when banks compute their Value at Risk to verify the adequacy of their capital. Another popular setting is that the subsidiary companies share the idiosyncratic factor with their parent company, in which case, the subsidiary companies will not necessarily default when, for instance, their systematic factors are small even if their parent company defaults.

Our aim is to modify $\tilde{p}_k^{par}$ and $\tilde{p}_{k,j}^{sub}$ to conduct similar efficient Monte Carlo simulations in the last section. The direct calculation shows the following.

Lemma 3.1 *For $k = 1, \ldots, m^{par}$, and $j = 1, \ldots, J(k)$, let the modified twisted probabilities of the parent and its subsidiary companies be $\tilde{q}_k^{par}(\theta)$ and $\tilde{q}_{k,j}^{sub}(\theta)$. For $\theta > 0$, set*

$$\tilde{q}_k^{par}(\theta) = \frac{\tilde{p}_k^{par} \exp\left(\theta\left(c_k^{par} + \sum_{j=1}^{J(k)} c_{k,j}^{sub}\right)\right)}{\tilde{p}_k^{par} \exp\left(\theta\left(c_k^{par} + \sum_{j=1}^{J(k)} c_{k,j}^{sub}\right)\right) + (1 - \tilde{p}_k^{par})\Pi_{j=1}^{J(k)}\left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right) + 1 - \tilde{p}_{k,j}^{sub}\right)},$$

and

$$\tilde{q}_{k,j}^{sub}(\theta) = \begin{cases} 1 & (Y_k = 1), \\ \left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right)\right) / \left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right) + 1 - \tilde{p}_{k,j}^{sub}\right) & (Y_k = 0). \end{cases}$$

Then,

$$E_{\tilde{\boldsymbol{q}}(\theta)}[L] = \frac{d}{d\theta}\tilde{\psi}(\theta),$$

where $E_{\tilde{\boldsymbol{q}}(\theta)}[\cdot]$ is the expectation with respect to the modified twisted probabilities $\tilde{q}_k^{par}(\theta)$'s and $\tilde{q}_{k,j}^{sub}(\theta)$'s, and

$$\tilde{\psi}(\theta) = \log(E[e^{\theta L}])$$
$$= \sum_{k=1}^{m^{par}} \log\left(\tilde{p}_k^{par} \exp\left(\theta\left(c_k^{par} + \Sigma_{j=1}^{J(k)} c_{k,j}^{sub}\right)\right) + (1 - \tilde{p}_k^{par})\Pi_{j=1}^{J(k)}\left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right) + 1 - \tilde{p}_{k,j}^{sub}\right)\right).$$

Also, the likelihood ratio is

$$LR(\{\tilde{p}_k^{par}\}, \{\tilde{p}_{k,j}^{sub}\}, \{\tilde{q}_k^{par}(\theta)\}, \{\tilde{q}_{k,j}^{sub}(\theta)\}) = \exp(-\theta L + \tilde{\psi}(\theta)).$$

This lemma leads to the algorithm; (1) compute L by first generating $\boldsymbol{Z}$, (2) generating each Y_k^{par} and $Y_{k,j}^{sub}$ by the probabilities $\tilde{q}_k^{par}(\tilde{\theta}^*)$ and $\tilde{q}_{k,j}^{sub}(\tilde{\theta}^*)$ respectively, and (3) taking the average of

$$1(L > L_0)LR(\{\tilde{p}_k^{par}\}, \{\tilde{p}_{k,j}^{sub}\}, \{\tilde{q}_k^{par}(\tilde{\theta}^*)\}, \{\tilde{q}_{k,j}^{sub}(\tilde{\theta}^*)\}) = 1(L > L_0)\exp(-\tilde{\theta}^* L + \tilde{\psi}(\tilde{\theta}^*)),$$

with $\tilde{\theta}^*$ satisfying that

$$\frac{d}{d\theta}\tilde{\psi}(\tilde{\theta}^*) = L_0. \tag{3.1}$$

As in the case without the chain defaults, it is not difficult to see that $(d/d\theta)\tilde{\psi}(\theta)$ is monotonically increasing, so that the unique solution $\tilde{\theta}^*$ of (3.1) is assured. When $\theta = \tilde{\theta}^*$, as in the case without the chain defaults, the expected loss under the twisted probabilities is the same with the given loss level L_0.

The method above can also be applied to compute the marginal risk contributions to the VaR, to compute the expected shortfall (ES) and the marginal risk contributions to the ES, by using the equation (18) in Glasserman (2005). We illustrate this in the following numerical experiments. As in Glasserman and Li (2005), one can shift the means of $Z_1, \ldots, Z_d$ to further reduce the variance in the importance sampling. A simple alternative method is given in the section 5.

4   Case of Chain Defaults: Saddle Point Approximations

Muromachi (2004), Martin and Ordovas (2006), Huang, *et al.* (2007), and others explored to compute the Value at Risk and the marginal risk contributions to the Value at Risk by applying the saddle point approximations to the inversion formula of Laplace transform. The direct calculation shows that this method is also available in the case of chain defaults: Namely, for any $s > 0$,

$$P(L > L_0|\boldsymbol{Z}) = \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty}\frac{\exp\left(\tilde{\psi}(\theta) - \theta L_0\right)}{\theta}d\theta.$$

Therefore, one can carry out the saddle point approximations to the integration, and then take the expectations with respect to $\boldsymbol{Z}$ to get $P(L > L_0)$.

Similarly, one can compute the marginal VaR contribution of the parent CON_k^{par} and of its subsidiary company $CON_{k,j}^{sub}$ defined by $CON_k^{par} = c_k^{par}E[Y_k^{par}|L = L_0]$ and $CON_{k,j}^{sub} = c_{k,j}^{sub}E[Y_{k,j}^{sub}|L = L_0]$ by applying the saddle point approximation to the integration given in the following lemma.

6

Lemma 4.1 *For any $s > 0$,*

$$CON_k^{par} = c_k^{par} E\left[\int_{s-i\infty}^{s+i\infty} \frac{\partial \tilde{\psi}(\theta)}{\partial c_k^{par}} \frac{\exp\left(\tilde{\psi}(\theta) - \theta L_0\right)}{\theta} d\theta\right] E\left[\int_{s-i\infty}^{s+i\infty} e^{\tilde{\psi}(\theta) - \theta L_0} d\theta\right]^{-1}$$

$$= c_k^{par} E\left[\tilde{p}_k^{par} \int_{s-i\infty}^{s+i\infty} \exp\left[\tilde{\psi}_{-k}(\theta) - \theta\left(L_0 - (c_k^{par} + \Sigma_{j=1}^{J(k)} c_{k,j}^{sub})\right)\right] d\theta\right] E\left[\int_{s-i\infty}^{s+i\infty} e^{\tilde{\psi}(\theta) - \theta L_0} d\theta\right]^{-1},$$

and

$$CON_{k,j}^{sub} = c_{k,j}^{sub} E\left[\int_{s-i\infty}^{s+i\infty} \frac{\partial \tilde{\psi}(\theta)}{\partial c_{k,j}^{sub}} \frac{\exp\left(\tilde{\psi}(\theta) - \theta L_0\right)}{\theta} d\theta\right] E\left[\int_{s-i\infty}^{s+i\infty} e^{\tilde{\psi}(\theta) - \theta L_0} d\theta\right]^{-1}$$

$$= c_{k,j}^{sub} E\left[\tilde{p}_k^{par} \int_{s-i\infty}^{s+i\infty} \exp\left[\tilde{\psi}_{-k}(\theta) - \theta\left(L_0 - (c_k^{par} + \Sigma_{j=1}^{J(k)} c_{k,j}^{sub})\right)\right] d\theta\right] E\left[\int_{s-i\infty}^{s+i\infty} e^{\tilde{\psi}(\theta) - \theta L_0} d\theta\right]^{-1}$$

$$+ c_{k,j}^{sub} E\left[(1 - \tilde{p}_k^{par})\tilde{p}_{k,j}^{sub} \int_{s-i\infty}^{s+i\infty} \exp\left[\tilde{\psi}_{-(k\&j)}(\theta) - \theta\left(L_0 - c_{k,j}^{sub}\right)\right] d\theta\right] E\left[\int_{s-i\infty}^{s+i\infty} e^{\tilde{\psi}(\theta) - \theta L_0} d\theta\right]^{-1}$$

where

$$\tilde{\psi}_{-k_0}(\theta)$$
$$= \sum_{k \neq k_0} \log\left(\tilde{p}_k^{par} \exp\left(\theta\left(c_i^{par} + \Sigma_{j=1}^{J(k)} c_{k,j}^{sub}\right)\right) + (1 - \tilde{p}_k^{par})\Pi_{j=1}^{J(k)}\left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right) + 1 - \tilde{p}_{k,j}^{sub}\right)\right),$$

and

$$\tilde{\psi}_{-(k_0\&j_0)}(\theta)$$
$$= \sum_{k \neq k_0} \log\left(\tilde{p}_k^{par} \exp\left(\theta\left(c_i^{par} + \Sigma_{j=1}^{J(k)} c_{k,j}^{sub}\right)\right) + (1 - \tilde{p}_k^{par})\Pi_{j=1}^{J(k)}\left(\tilde{p}_{k,j}^{sub} \exp\left(\theta c_{k,j}^{sub}\right) + 1 - \tilde{p}_{k,j}^{sub}\right)\right)$$
$$+ \sum_{j \neq j_0, \ 1 \leq j \leq J(k_0)} \log\left(\tilde{p}_{k_0,j}^{sub} \exp(\theta c_{k_0,j}^{sub}) + 1 - \tilde{p}_{k_0,j}^{sub}\right).$$

These results, especially for $CON_{k,j}^{sub}$ are interesting. One can relate those two terms to the two cases in which the parent company defaults or not, respectively.

Also, when we use saddle point approximations to compute those integrations numerically, it is noteworthy that the required number of the saddle point approximations, namely the number of computation of the roots θ^*, is the same with the case without the chain defaults: $m + 1$ times total in both, leading to the rational expectation that the computation complexity is not so different. Note that the cumulant functions to which the approximation is applied are different: One cumulant function with all of the corporations in both cases (for the denominator) plus
(1) in the case without the chain defaults:
m cumulant functions leaving each one company out,

(2) in the case of the chain defaults:

  (i) m^{par} cumulant functions leaving each one corporate group out and

  (ii) $m - m^{par}$ cumulant functions leaving both each one parent company and each one of its subsidiary companies out.

5   Shifting the Factors Revisited: Simple Method Using EM-Algorithm

Both of the last two sections discuss how to reduce the variance of the Monte Carlo simulation that comes from the exposure concentration: the large loss resulting from some particular companies or corporate groups. Glasserman, *et al.*(2007) proposed a method to reduce the variance of the Monte Carlo simulation that comes from the high correlations among the obligers by shifting the mean vector of the multivariate common factors. They suggested asymptotically optimal shifts of the mean vector and used normal mixture distribution as the proposal distribution of the common factors in the importance sampling, but this method includes some complicated optimizations. Alternatively, we give a simple method to give the desirable normal mixture distribution. Although the asymptotic property of the proposed method is not investigated yet, the numerical performance shown in section 6 is favorable.

    Suppose we can easily sample from $f(\boldsymbol{Z})$. Let the probability density function of $\boldsymbol{Z}$ be $f(\boldsymbol{Z})$. In the discussions by the last section, $f(\boldsymbol{Z})$ is the multivariate standard normal distribution. Later, we consider the case in which it is difficult to sample from $f(\boldsymbol{Z})$.

    As Glasserman and Li (2005) pointed out, the optimal proposal distribution is $P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})$. Although it is unknown itself, we try to approximate it more directly by the normal mixture distribution using the EM-algorithm. For a given positive integer $R > 0$, we denote the resulting proposal normal mixture distribution by $g(\boldsymbol{Z}|\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R, w_1,\ldots,w_R)$. Namely, g is the mixture distribution of R multivariate normal distributions with mean vectors $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R$, identity covariance matrices, and positive weights $w_1,\ldots,w_R$ with $\sum_{r=1}^{R} w_r = 1$. Our goal is to decide $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R$ and $w_1,\ldots,w_R$ that approximate $P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})$ well.

    We search $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R$ and $w_1,\ldots,w_R$ that minimize the KL (kullback-Leibler) divergence from g to $P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})$:

$$\min_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R,w_1,\ldots,w_R} \int \log\left(\frac{P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})}{g(\boldsymbol{Z})}\right) P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})d\boldsymbol{Z}.$$

It is equivalent to the following maximization problem:

$$\max_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R,w_1,\ldots,w_R} \int \left(\log g(\boldsymbol{Z})\right) P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})d\boldsymbol{Z}.$$

Since $P(L > L_0|\boldsymbol{Z})$ is unknown, it is not feasible, so we employ normal approximation of

$P(L > L_0|\boldsymbol{Z})$ following the suggestion in Glasserman and Li (2005):

$$P(L > L_0|\boldsymbol{Z}) \approx \hat{P}(L > L_0|\boldsymbol{Z}) := 1 - \Phi\left(\frac{L_0 - E[L|\boldsymbol{Z}]}{\sqrt{\mathrm{Var}[L|\boldsymbol{Z}]}}\right). \tag{5.1}$$

Then, the maximization problem becomes

$$\max_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_R,w_1,\ldots,w_R} \int \left(\log g(\boldsymbol{Z})\right) \hat{P}(L > L_0|\boldsymbol{Z}) f(\boldsymbol{Z}) d\boldsymbol{Z}.$$

We estimate the expectation with respect to $f(\boldsymbol{Z})$ by empirical distribution:

$$\int \left(\log g(\boldsymbol{Z})\right) \hat{P}(L > L_0|\boldsymbol{Z}) f(\boldsymbol{Z}) d\boldsymbol{Z} \approx \sum_{n=1}^{N} \left(\log g(\boldsymbol{Z}_n)\right) \hat{P}(L > L_0|\boldsymbol{Z}_n) \tag{5.2}$$

for $\boldsymbol{Z}_n \sim$ i.i.d. $f(\boldsymbol{Z})$.

Observe that (5.2) is the typical log-likelihood maximization except that it is weighted by $\hat{P}(L > L_0|\boldsymbol{Z}_n)$. Then, one can use the well-known EM algorithm to solve it. Now write $g(\boldsymbol{Z}) = R^{-1} \sum_{r=1}^{R} w_r \phi(\boldsymbol{Z}|\boldsymbol{\mu}_r, \boldsymbol{I}_d)$, where $\phi(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the probability density function of the multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. By differentiating (5.2) with respect to $\boldsymbol{\mu}_r$ and w_r, one gets the optimal $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_R$ and $w_1, \ldots, w_R$ by applying the well-known EM algorithm:

New Mixed Importance Sampling Algorithm by EM algorithm
(1) Sample $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N$ from $f(\boldsymbol{Z})$.
(2) Randomly set the initial $\boldsymbol{\mu}_1^{(0)}, \ldots, \boldsymbol{\mu}_R^{(0)}$, and set $w_r^{(0)} \equiv 1/R$ for all r.
(3) For $t = 0, 1, \ldots$, update $\boldsymbol{\mu}_1^{(t)}, \ldots, \boldsymbol{\mu}_R^{(t)}$ and $w_1^{(t)}, \ldots, w_R^{(t)}$ by

$$\boldsymbol{\mu}_r^{(t+1)} = \frac{\sum_{n=1}^{N} \hat{P}(L > L_0|\boldsymbol{Z}_i) v_{nr}^{(t+1)} \boldsymbol{Z}_n}{\sum_{n=1}^{N} \hat{P}(L > L_0|\boldsymbol{Z}_n) v_{nr}^{(t+1)}},$$

and

$$w_r^{(t+1)} = \frac{\sum_{n=1}^{N} \hat{P}(L > L_0|\boldsymbol{Z}_n) v_{nr}^{(t+1)}}{\sum_{n=1}^{N} \hat{P}(L > L_0|\boldsymbol{Z}_n)},$$

where

$$v_{nr}^{(t+1)} = \frac{w_r^{(t)} \exp(-\parallel \boldsymbol{Z}_n - \boldsymbol{\mu}_r^{(t)} \parallel^2 /2)}{\sum_{r=1}^{R} w_r^{(t)} \exp(-\parallel \boldsymbol{Z}_n - \boldsymbol{\mu}_r^{(t)} \parallel^2 /2)}.$$

(4) Stop if $\boldsymbol{\mu}_1^{(t)}, \ldots, \boldsymbol{\mu}_R^{(t)}$ and $w_1^{(t)}, \ldots, w_R^{(t)}$ converge.

The required number of iteration is crucial for the computation time, but it should be noted that we have only to get approximate maximizers. As the number of factors

increases, we need more iterations. If the number of factors is small, we see that a small number of iterations, less than ten, is sufficient in the following numerical experiments.

The advantage of this method is that it is applicable to any $f(\boldsymbol{Z})$. Moreover, if it is difficult to sample from $f(\boldsymbol{Z})$ directly, then one can sample from proper proposal distribution $h(\boldsymbol{Z})$ instead, in which case (5.2) becomes

$$\sum_{n=1}^{N} \left(\log g(\boldsymbol{Z}_n)\right) \hat{P}(L > L_0|\boldsymbol{Z}_n) \frac{f(\boldsymbol{Z}_n)}{h(\boldsymbol{Z}_n)}$$

for $\boldsymbol{Z}_n \sim \text{i.i.d.} h(\boldsymbol{Z})$.

The EM algorithm above uses the samples from $f(\boldsymbol{Z})$ but by putting the weight $\hat{P}(L > L_0|\boldsymbol{Z}_n)$ on each sample $\boldsymbol{Z}_n$, that is, by placing more emphasis on $\boldsymbol{Z}_n$ with high $\hat{P}(L > L_0|\boldsymbol{Z}_n)$, the resulting fitted normal mixture distribution approximates $\hat{P}(L > L_0|\boldsymbol{Z}_n)f(\boldsymbol{Z})$. This idea is similar to SIR (Sampling Importance Re-sampling) where a sample is drawn from the empirical distribution of the proposal distribution with probabilities proportional to the importance ratio.

In the approximation given in (5.1), one can also employ the constant approximation that is used in Glasserman, *et al.*(2007). Under the constant approximation of $P(L > L_0|\boldsymbol{Z}_n)$, the weight of $\boldsymbol{Z}_n$ is 1 if $E[L|\boldsymbol{Z}_n] > L_0$ and 0 if $E[L|\boldsymbol{Z}_n] \leq L_0$ instead of $\hat{P}(L > L_0|\boldsymbol{Z}_n)$ under the normal approximation. Although the normal approximation is more precise in the case without the chain defaults, the precision becomes worse in the existence of chain defaults since the default of each company is no longer independent even if conditioned on $\boldsymbol{Z}$.

The rest of the algorithm is the same with the **MIS** (Mixed Importance Sampling method) introduced in Glasserman, *et al.*(2007):

New Mixed Importance Sampling Algorithm by EM algorithm (continued)
For $r = 1, \ldots, R$, and for $t = 1, \ldots, w_r \times$(number of replications),
(5) Sample $\boldsymbol{Z}$ from $\phi(\boldsymbol{Z}|\boldsymbol{\mu}_r, \boldsymbol{I})$.
(6) Find $\tilde{\theta}^*(\boldsymbol{Z})$ by solving (3.1), compute the twisted conditional default probabilities by the Lemma 2.1.
(7) Sample Y_k's (parent companies) and Y_{kj}'s (subsidiary companies), and compute L.
(8) Calculate

$$\frac{1(L > L_0) \exp(-\tilde{\theta}^* L + \tilde{\psi}(\tilde{\theta}^*))}{\sum_{r=1}^{R} w_r \exp(\boldsymbol{\mu}_r^T \boldsymbol{Z} - \boldsymbol{\mu}_r^T \boldsymbol{\mu}_r/2)}$$

(9) Take average of (8) over $r = 1, \ldots, R$ and $t = 1, \ldots, w_r \times$(number of replications).

We make some observations on this result.

First, we need to choose N, the sample size used for the weighted EM-algorithm. As N becomes larger, the empirical approximation in (5.2) becomes more precise. However, the computation time also increases. There is a trade-off, so we should decide the best N

that works well. Typically, as the confidence level increases, $P(L > L_0|\boldsymbol{Z}_n) \approx 0$ for most $\boldsymbol{Z}_n \sim f(\boldsymbol{Z})$, so we need more samples. It also relies on the dimension of $\boldsymbol{Z}$, namely on d. As d increases, the structure of $P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})$ may be more complicated. Moreover, if $a_{kl} < 0$ for some k and l, $P(L > L_0|\boldsymbol{Z})f(\boldsymbol{Z})$ may be multi-modal, and we need adequate samples around each mode to approximate it, which increases the required sample size N as well.

The choice of R, the number of normal mixture components, is also relevant. In the numerical experiments, we see that if we give sufficient but not too large R, some vectors $\boldsymbol{\mu}_r$'s converge to a smaller number of groups. In other words, virtually the weighted EM algorithm reduces the number of the components automatically. However, if we set too large R, the dimension of $\boldsymbol{\mu}_r$'s gets larger, and the estimation may become unstable. The computation cost will also increase. Further research is needed, such as using information criteria like the AIC.

Just as the orthodox EM-algorithm, one can also choose the desirable covariance matrix of the normally distributed components by setting $g(\boldsymbol{Z}) := R^{-1} \sum_{r=1}^{R} w_r \phi(\boldsymbol{Z}|\boldsymbol{\mu}_r, \sigma_k^2 \boldsymbol{I}_d)$ and maximizing (5.2) with respect to w_r, $\boldsymbol{\mu}_r$, and σ_k^2. However, as the number of parameters increases, the estimation may become more unstable, and the computation time will increase. Therefore, the numerical performance with covariance optimization needs to be confirmed in the future.

Lastly, the feasibility assumption imposed by Glasserman, *et al.*(2007), $G_{\mathscr{F}}^{(m)}$ is not empty, also applies to our method. If $a_{kl} < 0$ for some k and l, **MIS** may not work as they state on page 1205 (since $G_{\mathscr{F}}^{(m)}$ may be empty). A typical example is a case in which $c_i \equiv 1$ for $k = 1, \ldots, m$ (with no chain defaults), there are only two sectors with a strong negative correlation, half of the customers belong to each sector respectively, and L_0 is just above the half of the total exposure, $m/2$. Even in such a case, our method selects some mean shifts. However, it is no longer meaningful since the selected means concentrate on the origin: $\boldsymbol{\mu}_r \approx \boldsymbol{0}$. In fact, the multivariate common factors must be $\boldsymbol{Z} \approx \boldsymbol{0}$ whenever $L > L_0$. This is because if $\boldsymbol{Z}$ deviates from the origin, one sector has a large loss in very high probability, but at the same time, the other sector has a small loss in very high probability as well. Thus, roughly speaking, L always becomes smaller than L_0 (remember $L_0 > m/2$). On the other hand, if $\boldsymbol{Z} \approx \boldsymbol{0}$, all of the customers independently default, so $P(L > L_0|\boldsymbol{Z} \approx \boldsymbol{0})$ is small but not negligible. In short, in such a case, correlations do not dominate the tail probability. Therefore, shifting the means is meaningless.

6   Simulation studies

In this section, we confirm the performance of the proposed methods by numerical experiments.

Simulation 1 We confirm the efficiency of the importance sampling method given in the

section 3. The simulation setting is the following: The number of obligers is 100 indexed by from 1 to 100.

The credit exposures (c_k) and the default probabilities (p_{0k}) are:
(1) index 1-5: $(c_k, p_{0k}) = (1, 1\%)$,
(2) index 6-10: $(c_k, p_{0k}) = (1, 5\%)$,
(3) index 11-35: $(c_k, p_{0k}) = (1, 3\%)$,
(4) index 36-65: $(c_k, p_{0k}) = (5, 1\%)$,
(5) index 66-85: $(c_k, p_{0k}) = (10, 1\%)$,
(6) index 86-95: $(c_k, p_{0k}) = (20, 1\%)$,
(7) index 96-100: $(c_k, p_{0k}) = (50, 1\%)$.

The corporate groups are:
(1) corporate group 1: parent: index 1, subsidiary: index 86 and 96,
(2) corporate group 2: parent: index 2, subsidiary: index 87 and 97,
(3) corporate group 3: parent: index 3, subsidiary: index 88 and 98,
(4) corporate group 4: parent: index 4, subsidiary: index 89 and 99,
(5) corporate group 5: parent: index 5, subsidiary: index 90 and 100,
(6) corporate group 6: parent: index 6, subsidiary: index 91,
(7) corporate group 7: parent: index 7, subsidiary: index 92,
(8) corporate group 8: parent: index 8, subsidiary: index 93,
(9) corporate group 9: parent: index 9, subsidiary: index 94,
(10) corporate group 10: parent: index 10, subsidiary: index 95.

The correlation among the obligers is set to be zero to focus more on the effect of the chain defaults. The number of iteration is 10^5 for the plane Monte Carlo (plane MC) and 10^4 for the importance sampling method (**IS**). For all tail probabilities, the modified twisted probability given in Lemma 3.1 is computed by $L_0 = 144$ (which satisfies $P(L > L_0) = 10^{-3}$).

Figure 2 shows the pair of estimated tail probabilities (coupled with 95% confidence intervals) and their corresponding loss levels. The solid red line refers to **IS**, and the blue dotted one refers to the **plane MC**. The black dashed one is the case without the chain defaults (we omit the confidence interval). It is seen that the confidence intervals of **IS** are narrower than that of the **plane MC** even when the number of iteration is less than that of the **plane MC**. Figure 3 shows each obliger's marginal contribution to the expected shortfall (divided by its exposure). Each color refers to the same with figure 2. It is seen that the estimation by **IS** is more stable than that by the **plane MC**.

Simulation 2 Next, we confirm the efficiency of the mixed importance sampling method using EM algorithm given in the section 5 (**EM-MIS**). The simulation setting is exactly the same with the illustration given in Glasserman, *et al.*(2007): The number of obligers is 1000 indexed by from 1 to 1000.

The factor loadings of the obligers have two types:
(1) index 1-500 $(k = 1, \ldots, 500)$: $(a_{k1}, a_{k2}) = (0.7, 0)$,

(2) index 501-1000 ($k = 501, \ldots, 1000$): $(a_{k1}, a_{k2}) = (0, 0.65)$.

The credit exposures are all 1, the default probabilities are all 5%, and there are no corporate groups to focus more on the effect by the correlations. The loss level is 300 (to yield a tail probability around 1.1%). The number of iteration is 10^5 for the plane MC, **MIS** proposed by Glasserman, *et al.*(2007), and **EM-MIS**. In the simulations, we set $N = 10000$ and $R = 10$. The number of iteration of the EM-algorithm is 10.

Figure 4 shows the estimated cumulative tail probabilities of a single run. The solid red line refers to **EM-MIS**, the green dotted and dashed line refers to **MIS**, and the blue dotted line refers to the **plane MC**. It is seen that the estimation by **MIS** and **EM-MIS** is more stable than that by the **plane MC**. Figure 5 shows the sample variance of the single run. Each color refers to the same with figure 4. It is seen that the sample variance of **EM-MIS** is less than that of the **plane MC** and is comparable to that of **MIS**. Although we iterated this single run, we did not find any relevant change in this result.

7   Summary

In this paper, we have derived an efficient importance sampling method to compute the Value at Risk (VaR) and the expected shortfall with the chain default or bankruptcy. Similarly to Glasserman and Li (2005), by twisting the conditional default probability, we have shifted the loss distribution with its mean equal to the given large loss point, which makes the simulation much efficient. We have also derived a corresponding alternative method using the saddle point approximations. Moreover, a simple method to shift the multivariate factor means has been developed to further reduce the Monte Carlo variance due to the correlations generated by the factors. Simulation studies show that VaR or expected shortfall computed by these proposed methods has smaller variance.

References

[1] Glasserman, P., and Li. J. (2005). Importance sampling for portfolio credit risk. *Management Science*, **51**, 1643-1656.

[2] Glasserman, P. (2005). Measuring marginal risk contributions in credit portfolios. *Journal of Computational Finance*, **9**(2), 1-41.

[3] Glasserman, P., Kang W., and Shahabuddin P. (2007). Fast Simulation of Multifactor Portfolio Credit Risk. *Operations Research*, **56** (5), 1200-1217.

[4] Huang, X., Oosterlee, C., and Mesters, M. (2007) Computation of VaR and VaR Contribution in the Vasicek Portfolio Credit Loss Model: a Comparative Study. *Journal of Credit Risk*, **3** (3), 75-96.
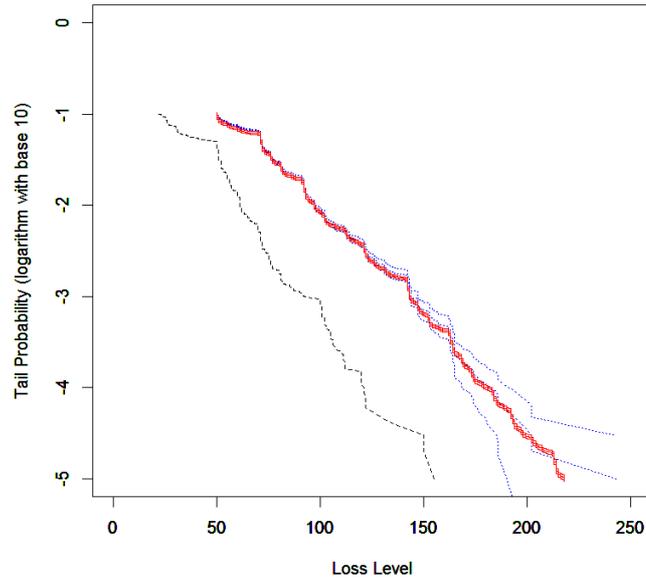
Figure 2: estimated tail probabilities (coupled with 95% confidence intervals) and their corresponding loss levels. The solid red line refers to **IS** (proposed method), and the blue dotted line refers to the **plane MC**. The black dashed one is the case without the chain defaults (we omit the confidence interval).
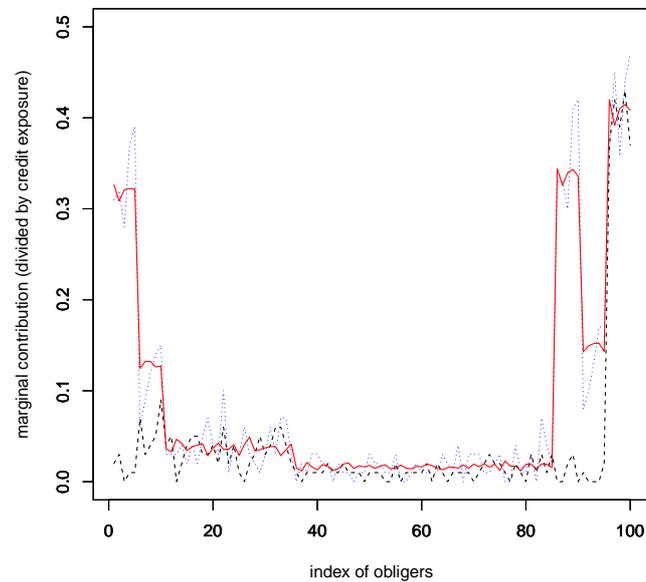

Figure 3: each obliger's marginal contribution to the expected shortfall (divided by its exposure). Each color refers to the same with figure 2
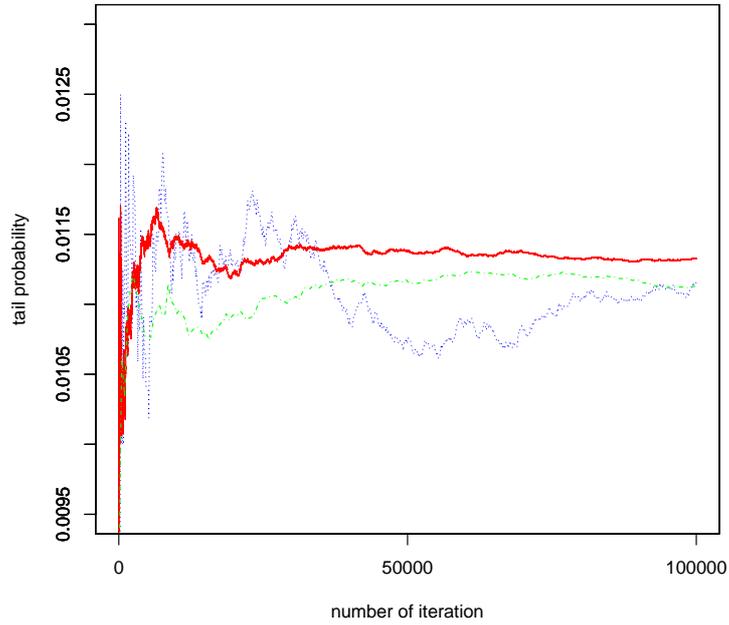
14

Figure 4: estimated cumulative tail probabilities of a single run. The red solid line refers to **EM-MIS** (proposed method), the green dotted and dashed line refers to **MIS** proposed by Glasserman, *et al.*(2007), and the blue dotted line refers to the **plane MC**.
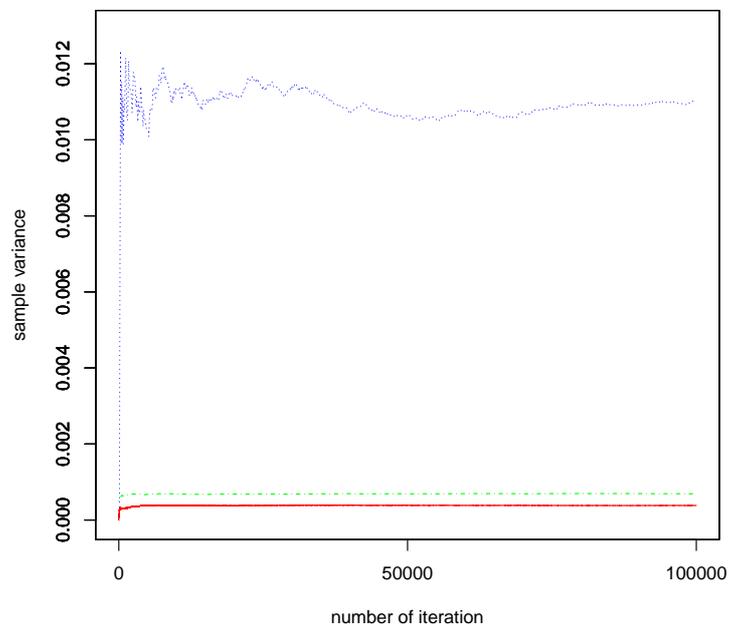

Figure 5: sample variance of the single run. Each color refers to the same with the figure 4.

[5] Martin, R., and Ordovas, R. (2006) An indirect view from the saddle. *Risk*, **19** (10), 94-99.

[6] Merton, R. (1974) On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance* **29** (2), 449-470.

[7] Muromachi, Y. (2004) A conditional independence approach for portfolio risk evaluation. *The Journal of Risk*, **7** (1), 27-53.