

Importance sampling with splitting for portfolio credit risk

Jinyoung Kim^a, Sunggon Kim^{1,b}

^aACUON Savings Bank Co., Ltd., Korea; ^bDepartment of Statistics, University of Seoul, Korea

Abstract

We consider a credit portfolio with highly skewed exposures. In the portfolio, small number of obligors have very high exposures compared to the others. For the Bernoulli mixture model with highly skewed exposures, we propose a new importance sampling scheme to estimate the tail loss probability over a threshold and the corresponding expected shortfall. We stratify the sample space of the default events into two subsets. One consists of the events that the obligors with heavy exposures default simultaneously. We expect that typical tail loss events belong to the set. In our proposed scheme, the tail loss probability and the expected shortfall corresponding to this type of events are estimated by a conditional Monte Carlo, which results in variance reduction. We analyze the properties of the proposed scheme mathematically. In numerical study, the performance of the proposed scheme is compared with an existing importance sampling method.

Keywords: importance sampling, credit portfolio, Bernoulli mixture model, skewed exposures

1. Introduction

Bernoulli mixture model for portfolio credit risk is a quite general framework for modeling the joint defaults of the obligors during a time period. In the model, some common factors or risk factors are introduced to explain the dependency of the joint defaults. The values of the factors determine the default probability of each obligor. By letting the default probabilities of some obligors have the same tendency for the change of the factors, one can model the clustered defaults of the obligors. However, an obligor defaults independently with the others. In other words, the conditional joint defaults of the obligors are independent Bernoulli events with probabilities specified by the factors. For the detailed discussion on the Bernoulli mixture model, we refer to McNeil *et al.* (2015).

For the Bernoulli mixture model of a credit portfolio, we propose a new importance sampling scheme to estimate efficiently the tail loss probability and the expected shortfall over a high threshold. The proposed scheme is a combination of the conditional Monte Carlo, the stratified sampling, and the importance sampling. In the proposed scheme, we split the sample space of the default events of the obligors into two subspaces so that the probability of the tail loss in a subspace is calculated exactly. The probability depends on the value of the factors, which are generated randomly. Thus, this part of the propose scheme is a conditional Monte Carlo (Rubinstein and Kroese, 2016). For the given value of factors, the conditional probability of the tail loss in the other subspace is estimated by an importance sampling of the default event. By combining these two conditional probabilities, we estimate the conditional probability of the tail loss when the factor variables are given. The tail loss probability is estimated efficiently by an importance sampling of the factor variables. In the same manner, we estimate the expected shortfall over a high threshold. We apply the two step importance

¹ Corresponding author: Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea. E-mail: sgkim@uos.ac.kr

sampling proposed by Glasserman and Li (2005) to our scheme. We also propose how to apply the cross entropy method (Chan and Kroses, 2010) for choosing the importance probability density function (pdf) of the factor variables, and how to apply the exponential twisting (Glasserman and Li, 2005) for choosing the importance default probability of each obligor optimally. We analyze the properties of the proposed scheme mathematically, and compare it with the two step importance sampling proposed by Glasserman and Li (2005).

The threshold model or the latent variable model, which is commonly used for modeling the portfolio credit risk, can be represented as a Bernoulli mixture model (Frey and McNeil, 2003). CreditRisk+ (Gundlach and Lehrbass, 2013) is an industry example of the Poisson mixture model, where an obligor might default more than once in a time period and the number of defaults follows the Poisson distribution. By considering the group of obligors with the same exposure and the default probability as an obligor, Bernoulli mixture model can be approximated by the Poisson mixture model.

In the Bernoulli mixture model, default events are easy to simulate by adopting a distribution from which the values of the factors are easy to generate. Examples of such distributions include the multi-variate normal distribution and the multi-variate student t . Despite the easy simulation of the joint defaults in the Bernoulli mixture model, it takes long time to obtain the reliable value of risk measure such as the value-at-risk or the expected shortfall through the crude Monte Carlo simulation. Importance sampling schemes, where related random variables are generated under a changed probability measure, are effective to obtain the risk measures (Brereton *et al.*, 2013; Glasserman, 2013; Merino and Nyfeler, 2004). The common approach to generate the joint default events with large losses is to shift the mean of the factors (Avrantis and Gregory, 2001; Egloff *et al.*, 2005; Kalkbrener *et al.*, 2004; Glasserman and Li, 2005). Exponential twisting has been studied for the estimation of the tail loss probability of the sum of independent or dependent random variables (Sadowsky, 1993; Bucklew *et al.*, 1990; Glasserman, 2013).

By integrating the exponential twisting of the default probabilities and the mean-shifting of the risk factors, Glasserman and Li (2005) proposed a two-step importance sampling for the threshold model, where the joint default probabilities are given by the normal copula, i.e. the risk factors are multivariate normal and a linear combination of the factors determines the default probabilities of obligors. They have shown that their algorithm is logarithmically efficient in the single factor case. The two-step importance sampling of Glasserman and Li (2005) can be easily implemented for the Bernoulli mixture model in the same manner, in which risk factors are not necessarily normally distributed, but generally distributed. An implementation is given in Başıoğlu *et al.* (2018).

Bassamboo *et al.* (2008) considered the case that the joint default probabilities are given by the t copula. They obtained the asymptotics of the tail loss probability and the expected shortfall in a large portfolio. They also proposed an importance sampling, and have shown that their method has bounded relative error in obtaining both the tail loss probability and the expected shortfall. For the same model as Bassamboo *et al.* (2008), Chan and Kroese (2010) proposed a fast simulation scheme to estimate the risk measures mentioned above. Their method is based on conditional Monte Carlo and has bounded relative error. Another fast Monte Carlo simulation scheme called the geometric shortcut is proposed by Sak and Hörmann (2012). The scheme can be used for threshold models with any copula, and can be integrated with the importance sampling of risk factors. Başıoğlu *et al.* (2013) developed a stratified importance sampling for the t -copula model with multiple loss thresholds, and they applied their method to the Bernoulli mixture model in Başıoğlu *et al.* (2018). They proposed to stratify the sample space of the factor variables into equi-probable subsets, and applied the sequential importance sampling scheme to reduce the variance of the estimation of the risk measure.

Differently from Başoğlu *et al.* (2018) in which the sample space of the factor variables is stratified, we stratify the sample space of the default events into two subsets. One of the subsets consists of the events that some of the obligors with very high exposures default simultaneously. The other subset is the collection of the remaining events. We expect that the typical occurrence of tail losses corresponds to the former set of events. In the Başoğlu *et al.* (2018)'s implementation of the two-step importance sampling, the default events are generated by the geometric shortcut proposed by Sak and Hörmann (2012). In this paper, the exponential twisting is adopted to generate the default events for ease of mathematical analysis. We have shown that the importance sampling with stratifying the sample space of the default events results in the variance reduction. This paper is organized as follows. In Section 2, we describe an implementation of the two-step importance sampling for the Bernoulli mixture model. In Section 3, we propose a two-step importance sampling for the model, and analyze the scheme mathematically. Numerical results are given in Section 4. Finally, we conclude the paper in Section 5.

2. Importance sampling for the Bernoulli mixture model

We consider a credit portfolio consisting of m obligors. Let Y_j , $j = 1, 2, \dots, m$, be the default indicator of obligor j in time interval $[0, T]$, i.e.

$$Y_j = \begin{cases} 1, & \text{obligor } j \text{ defaults in } [0, T], \\ 0, & \text{obligor } j \text{ does not default in } [0, T]. \end{cases}$$

We call $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ the default vector or the default event. The marginal default probability of obligor j , $j = 1, 2, \dots, m$, is $\Pr\{Y_j = 1\} = \bar{p}_j$. We denote by c_j , $j = 1, 2, \dots, m$, the deterministic exposure at the default of obligor j . Then, the loss of the portfolio during $[0, T]$ is given by

$$L(\mathbf{Y}) = \sum_{j=1}^m c_j Y_j. \quad (2.1)$$

In what follows, L is the abbreviation of $L(\mathbf{Y})$. We define $\theta = \Pr\{L > x\}$ for a threshold $x > 0$ and $\eta = E[I(L > x)L]$. We assume that $\theta > 0$. Then, the expected shortfall over the threshold $x > 0$ is computed as $E[L|L > x] = \eta/\theta$. Clearly, $\eta < \infty$ because we consider the case of finite number of obligors with deterministic exposures. The crude Monte Carlo (CMC) estimator of θ is

$$\hat{\theta}_n^C = \frac{1}{n} \sum_{i=1}^n I(L^{(i)} > x), \quad (2.2)$$

where $L^{(i)} = L(\mathbf{Y}^{(i)})$, $i = 1, \dots, n$, and $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(n)}$ are i.i.d. random samples of \mathbf{Y} . The CMC estimator of η is given by

$$\hat{\eta}_n^C = \frac{1}{n} \sum_{i=1}^n I(L^{(i)} > x) L^{(i)}. \quad (2.3)$$

Then, $E[L|L > x]$ is estimated as

$$\hat{\beta}_n^C = \frac{\hat{\eta}_n^C}{\hat{\theta}_n^C}. \quad (2.4)$$

In the following subsection, we introduce briefly the Bernoulli mixture model for modeling the dependency of Y . In the model, the default events of the obligors depend on some common factors. The performance of the CMC estimators in Equations (2.2) and (2.4) can be improved by employing an importance sampling of the factor variables and the default event. The two step importance sampling proposed by Glasserman and Li (2005) which was originally developed for the threshold model can be simply applied to the Bernoulli mixture model. We explain how to do the two step importance sampling in the Bernoulli mixture model in Sections 2.2 and 2.3. Chan and Kroese (2010) proposed the cross entropy method for the optimal selection of the sampling distribution of the factors in the t-copula model. The method can be applied to the Bernoulli mixture model. We explain the procedure in Section 2.4.

2.1. Bernoulli mixture model for dependent defaults of obligors

In the Bernoulli mixture model, the dependency of default events is modeled by introducing a number of d common factors $\Psi = (\Psi_1, \dots, \Psi_d) \in \mathbb{R}^d$. Let $p_j(\psi)$, $j = 1, 2, \dots, m$, be the conditional default probability of obligor j given $\Psi = \psi$. In the model, default indicators Y_1, \dots, Y_m are conditionally independent when the value of Ψ is given. Then, the conditional joint probability mass function (pmf) of the random vector Y given $\Psi = \psi$ is as follows: for $y = (y_1, \dots, y_m) \in \{0, 1\}^m$,

$$\begin{aligned} p(y|\psi) &= \prod_{j=1}^m \Pr\{Y_j = y_j | \Psi = \psi\} \\ &= \prod_{j=1}^m p_j(\psi)^{y_j} (1 - p_j(\psi))^{1-y_j}. \end{aligned} \quad (2.5)$$

The joint pmf of Y is given by

$$p(y) = \int \prod_{j=1}^m p_j(\psi)^{y_j} (1 - p_j(\psi))^{1-y_j} f(\psi) d\psi, \quad (2.6)$$

where $f(\psi)$ is the pdf of Ψ . We define \mathbb{P} as the probability measure on $\mathbb{R}^d \times \{0, 1\}^m$ induced from $f(\psi)p(y|\psi)$. Then, $\Pr\{Y_j = 1\}$ is equal to $E_{\mathbb{P}}[I(Y_j = 1)]$. Since $E_{\mathbb{P}}[I(Y_j = 1)] = E_f[p_j(\Psi)]$, we have that

$$E_f[p_j(\Psi)] = \bar{p}_j, \quad j = 1, 2, \dots, m, \quad (2.7)$$

where $E_f[\xi(\Psi)]$ is the expectation of $\xi(\Psi)$ for a real valued function ξ under the pdf $f(\psi)$ of Ψ . If Ψ follows a pdf $g(\psi)$, then the expectation of $\xi(\Psi)$ is denoted by $E_g[\xi(\Psi)]$.

2.2. Two step importance sampling

In the Bernoulli mixture model, we want to find an efficient importance sampling distribution for Ψ and a conditional importance sampling distribution Y given Ψ . Let $g(\psi)$ be an importance pdf of Ψ and $q_j(\psi)$, $j = 1, 2, \dots, m$, be an importance default probability for obligor j given $\Psi = \psi$. We denote by $q(y|\psi)$ the conditional joint pmf of Y given $\Psi = \psi$, i.e.

$$q(y|\psi) = \prod_{j=1}^m q_j(\psi)^{y_j} (1 - q_j(\psi))^{1-y_j}.$$

We assume that $g(\boldsymbol{\psi})q(\mathbf{y}|\boldsymbol{\psi}) > 0$ whenever $f(\boldsymbol{\psi})p(\mathbf{y}|\boldsymbol{\psi}) > 0$, and define \mathbb{Q} as the probability measure on $\mathbb{R}^d \times \{0, 1\}^m$ induced from $g(\boldsymbol{\psi})q(\mathbf{y}|\boldsymbol{\psi})$.

For a positive valued function h , we define $E_p[h(L)|\boldsymbol{\psi}](E_q[h(L)|\boldsymbol{\psi}])$ as the conditional expectation of $h(L)$ when the joint pmf of \mathbf{Y} is $p(\mathbf{y}|\boldsymbol{\psi})(q(\mathbf{y}|\boldsymbol{\psi}))$. We also define $E_{\mathbb{P}}[h(L)](E_{\mathbb{Q}}[h(L)])$ as the expectation of $h(L)$ when $(\boldsymbol{\Psi}, \mathbf{Y})$ follows the probability measure $\mathbb{P}(\mathbb{Q})$. Then, $E_{\mathbb{P}}[h(L)]$ has the following representation:

$$\begin{aligned} E_{\mathbb{P}}[h(L)] &= E_f \left[E_p[h(L)|\boldsymbol{\Psi}] \right] \\ &= E_g \left[E_q[w(\boldsymbol{\Psi}, \mathbf{Y})h(L)|\boldsymbol{\Psi}] \right], \end{aligned} \quad (2.8)$$

where

$$w(\boldsymbol{\Psi}, \mathbf{Y}) = \frac{f(\boldsymbol{\Psi})p(\mathbf{Y}|\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})q(\mathbf{Y}|\boldsymbol{\Psi})}.$$

Then, Equation (2.8) is rewritten as

$$E_{\mathbb{P}}[h(L)] = E_{\mathbb{Q}}[w(\boldsymbol{\Psi}, \mathbf{Y})h(L)]. \quad (2.9)$$

Suppose that we have random samples $(\boldsymbol{\Psi}^{(1)}, \mathbf{Y}^{(1)}), \dots, (\boldsymbol{\Psi}^{(n)}, \mathbf{Y}^{(n)})$ from \mathbb{Q} . Let $w^{(i)} = w(\boldsymbol{\Psi}^{(i)}, \mathbf{Y}^{(i)})$ and $L^{(i)} = L(\mathbf{Y}^{(i)})$, $i = 1, \dots, n$. Then, Equation (2.9) gives an estimator of $E_{\mathbb{P}}[h(L)]$ as

$$\hat{E}_{\mathbb{P}}[h(L)] = \frac{1}{n} \sum_{i=1}^n w^{(i)} h(L^{(i)}). \quad (2.10)$$

Since $\mathbb{P}(L > x) = E_{\mathbb{P}}[I(L > x)]$, Equation (2.10) says that importance sampling estimators of θ and η are given by

$$\hat{\theta}_n^T = \frac{1}{n} \sum_{i=1}^n w^{(i)} I(L^{(i)} > x), \quad (2.11)$$

and

$$\hat{\eta}_n^T = \frac{1}{n} \sum_{i=1}^n w^{(i)} I(L^{(i)} > x) L^{(i)}, \quad (2.12)$$

respectively. Note that $\hat{\theta}_n^T \rightarrow \theta$ almost surely, and $\hat{\eta}_n^T \rightarrow \eta$ almost surely as $n \rightarrow \infty$ due to the strong law of large numbers. Since $E_{\mathbb{P}}[L|L > x] = \eta/\theta$, $E_{\mathbb{P}}[L|L > x]$ is estimated by

$$\hat{E}_{\mathbb{P}}[L|L > x] = \frac{\hat{\eta}_n^T}{\hat{\theta}_n^T}. \quad (2.13)$$

Let $\hat{\beta}_n^T = \hat{\eta}_n^T / \hat{\theta}_n^T$. Then, we have

$$\hat{\beta}_n^T \rightarrow E_{\mathbb{P}}[L|L > x], \quad a.s.$$

2.3. Exponential twisting for the conditional sampling distributions of defaults

Glasserman and Li (2005) have shown that the exponential twisting is an efficient way to adopt the importance sampling pmf of Y . Let $q_j(t, \boldsymbol{\psi})$ be the exponential twisting of $p_j(\boldsymbol{\psi})$, i.e.

$$q_j(t, \boldsymbol{\psi}) = \frac{p_j(\boldsymbol{\psi}) \exp(c_j t)}{1 - p_j(\boldsymbol{\psi}) + p_j(\boldsymbol{\psi}) \exp(c_j t)}, \quad j = 1, \dots, m. \quad (2.14)$$

They consider $q_j(t, \boldsymbol{\psi})$ as an importance default probability of obligor j , $j = 1, \dots, m$, when $\boldsymbol{\Psi} = \boldsymbol{\psi}$ is given. Let $q_t(\mathbf{y}|\boldsymbol{\psi})$ be the corresponding joint pmf of Y , and $r(\mathbf{y}|\boldsymbol{\psi})$ the conditional likelihood $p(\mathbf{y}|\boldsymbol{\psi})/q_t(\mathbf{y}|\boldsymbol{\psi})$. Then, we have

$$E_p[h(L)|\boldsymbol{\psi}] = E_q[r(Y|\boldsymbol{\Psi})h(L)|\boldsymbol{\psi}], \quad (2.15)$$

where $E_q[r(Y|\boldsymbol{\Psi})h(L)|\boldsymbol{\psi}]$ is the conditional expectation of $r(Y|\boldsymbol{\Psi})h(L)$ with Y following $q_t(\mathbf{y}|\boldsymbol{\psi})$ and $\boldsymbol{\Psi} = \boldsymbol{\psi}$.

As mentioned in Glasserman and Li (2005), $r(\mathbf{y}|\boldsymbol{\psi})$ is represented in terms of $L(\mathbf{y})$ as follows:

$$r(\mathbf{y}|\boldsymbol{\psi}) = M_L(t|\boldsymbol{\psi}) \exp\{-L(\mathbf{y})t\}, \quad (2.16)$$

where $M_L(t|\boldsymbol{\psi})$ is the conditional moment generating function (m.g.f.) of L with Y following $p(\mathbf{y}|\boldsymbol{\psi})$. Since Y_1, \dots, Y_m are conditionally independent given $\boldsymbol{\Psi}$, we have that

$$M_L(t|\boldsymbol{\psi}) = \prod_{j=1}^m \left\{ (1 - p_j(\boldsymbol{\psi})) + p_j(\boldsymbol{\psi}) \exp(c_j t) \right\}.$$

Glasserman and Li (2005) obtained an upper bound of $V[\hat{\theta}_n^T|\boldsymbol{\psi}]$ in the case that $q_t(\mathbf{y}|\boldsymbol{\psi})$ is adopted as the importance sampling pmf of Y given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, and showed that the upper bound is minimized at $t = t(\boldsymbol{\psi})$ given by

$$t(\boldsymbol{\psi}) = \begin{cases} \text{the solution of } \sum_{j=1}^m c_j q_j(t, \boldsymbol{\psi}) = x, & x > E_p[L|\boldsymbol{\psi}], \\ 0, & x \leq E_p[L|\boldsymbol{\psi}]. \end{cases} \quad (2.17)$$

Then, the optimal default probability of obligor j for the importance sampling, is $q_j^*(\boldsymbol{\psi}) = q_j(t(\boldsymbol{\psi}), \boldsymbol{\psi})$, $j = 1, \dots, m$. The corresponding conditional joint pmf of Y is

$$q^*(\mathbf{y}|\boldsymbol{\psi}) = \prod_{j=1}^m q_j^*(\boldsymbol{\psi})^{y_j} (1 - q_j^*(\boldsymbol{\psi}))^{1-y_j}, \quad \mathbf{y} \in \{0, 1\}^m. \quad (2.18)$$

2.4. Cross entropy method for the sampling distribution of factors

Suppose that the pdf $f(\boldsymbol{\psi})$ has a parameter $\boldsymbol{\mu}$ which can be vector-valued. We denote the pdf by $f(\boldsymbol{\psi}; \boldsymbol{\mu})$ to specify the value of the parameter. The zero variance pdf for the importance sampling to estimate $E_{\mathbb{P}}[h(L)]$ is

$$\eta(\boldsymbol{\psi}; \boldsymbol{\mu}) = \frac{E_p[h(L)|\boldsymbol{\psi}] f(\boldsymbol{\psi}; \boldsymbol{\mu})}{E_{\mathbb{P}}[h(L)]}. \quad (2.19)$$

Chan and Kroese (2010) proposed to confine the parametric family of the importance sampling densities to

$$\mathcal{F} = \{f(\boldsymbol{\psi}; \boldsymbol{\nu})\}, \quad (2.20)$$

and to adopt $f(\boldsymbol{\psi}; \boldsymbol{\nu})$ as the one having the minimum Kullback-Leibler or cross-entropy distance to the zero variance density $\eta(\boldsymbol{\psi}; \boldsymbol{\mu})$ in Equation (2.19). Due to Rubinstein and Kroese (2016), the optimal value $\boldsymbol{\nu}^*$ is obtained by

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \int E_p[h(L)|\boldsymbol{\psi}] \log f(\boldsymbol{\psi}; \boldsymbol{\nu}) f(\boldsymbol{\psi}; \boldsymbol{\mu}) d\boldsymbol{\psi}. \quad (2.21)$$

Instead of normal approximation of $E_p[h(L)|\boldsymbol{\psi}]$ as done in Bařořlu *et al.* (2018), we consider the exponentially twisted joint default pmf $q^*(\mathbf{y}|\boldsymbol{\psi})$ given Equation (2.18). Then, the above equation is converted to

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \int E_{q^*} [r(\mathbf{Y}|\boldsymbol{\psi})h(L)|\boldsymbol{\psi}] \log f(\boldsymbol{\psi}; \boldsymbol{\nu}) f(\boldsymbol{\psi}; \boldsymbol{\mu}) d\boldsymbol{\psi}. \quad (2.22)$$

The maximization problem (2.22) can be solved by the simulation based optimization. We generate independently $\boldsymbol{\Psi}^{(1)}, \dots, \boldsymbol{\Psi}^{(M)}$ from $f(\boldsymbol{\psi}; \boldsymbol{\mu})$, and $\mathbf{Y}^{(i)}$ from $q^*(\mathbf{y}|\boldsymbol{\Psi}^{(i)})$, $i = 1, \dots, M$. Then, the generated i^{th} loss is $L^{(i)}$. It follows from Chan and Kroese (2010) that the problem (2.22) is converted to solve the following stochastic maximization problem:

$$\underset{\boldsymbol{\nu}}{\operatorname{maximize}} \frac{1}{M} \sum_{i=1}^M r(\mathbf{Y}^{(i)}|\boldsymbol{\Psi}^{(i)}) h(L^{(i)}) \log f(\boldsymbol{\Psi}^{(i)}; \boldsymbol{\nu}). \quad (2.23)$$

Let $\boldsymbol{\nu}^*$ be the solution of Equation (2.23). Then, the optimal importance pdf of $\boldsymbol{\Psi}$ is $f(\boldsymbol{\psi}; \boldsymbol{\nu}^*)$, and the likelihood ratio for an event $(\boldsymbol{\psi}, \mathbf{y})$ is

$$w(\boldsymbol{\psi}, \mathbf{y}) = \frac{f(\boldsymbol{\psi}; \boldsymbol{\mu})}{f(\boldsymbol{\psi}; \boldsymbol{\nu}^*)} M_L(t(\boldsymbol{\psi})|\boldsymbol{\psi}) \exp\{-L(\mathbf{y})t(\boldsymbol{\psi})\}.$$

3. A two step importance sampling with splitting

In this section, we propose a type of two step importance sampling to estimate efficiently the tail loss probability over a threshold x and the corresponding expected shortfall in the case that the distribution of $\{c_j\}_{j=1}^m$ is heavily skewed to the right.

3.1. The proposed scheme for the estimation of the tail loss probability

Without loss of generality, we assume that c_1, \dots, c_m are in their descending order, and let k be the integer satisfying

$$k = \min \left\{ i : \sum_{j=1}^i c_j > x \right\}. \quad (3.1)$$

We call the first k -obligors heavy obligors. Equation (3.1) implies that if all heavy obligors default simultaneously, then a tail loss over x occurs.

We define \mathcal{S} as the sample space of the default vector \mathbf{Y} , i.e. $\mathcal{S} = \{0, 1\}^m$. The sample space \mathcal{S} is split into two disjoint subspaces \mathcal{S}_0 and \mathcal{S}_1 . For a default vector $\mathbf{y} \in \mathcal{S}$, \mathbf{y} belongs to \mathcal{S}_0 if and only if the first k components of \mathbf{y} is equal to 1 (i.e., all heavy obligors default). If at least one of them does not default, then the corresponding default vector \mathbf{y} belongs to \mathcal{S}_1 . From the definition of the set \mathcal{S}_0 , we observe

$$L(\mathbf{y}) = \sum_{j=1}^k c_j + \sum_{j=k+1}^m c_j y_j, \quad \mathbf{y} \in \mathcal{S}_0. \quad (3.2)$$

Since $\sum_{j=1}^k c_j > x$, it follows from the above equation that $L(\mathbf{y}) > x$ for $\mathbf{y} \in \mathcal{S}_0$, which means $I(\mathbf{Y} \in \mathcal{S}_0, L > x) = I(\mathbf{Y} \in \mathcal{S}_0)$. Then,

$$I(L > x) = I(\mathbf{Y} \in \mathcal{S}_0) + I(L > x, \mathbf{Y} \in \mathcal{S}_1). \quad (3.3)$$

Suppose that Ψ is given to be ψ , and that \mathbf{Y} follows the pmf $p(\mathbf{y}|\psi)$. Then,

$$E_p[I(L > x)|\psi] = E_p[I(\mathbf{Y} \in \mathcal{S}_0)|\psi] + E_p[I(L > x, \mathbf{Y} \in \mathcal{S}_1)|\psi]. \quad (3.4)$$

In Equation (3.4), $E_p[I(L > x)|\psi]$ is the conditional tail loss probability over the threshold x given $\Psi = \psi$. In the equation, $E_p[I(\mathbf{Y} \in \mathcal{S}_0)|\psi]$ is the conditional loss probability due to the event $\mathbf{Y} \in \mathcal{S}_0$. We denote it by $p_a(\psi)$. Since Y_1, Y_2, \dots, Y_m are conditionally independent given $\Psi = \psi$, we have

$$p_a(\psi) = \prod_{j=1}^k p_j(\psi).$$

The last term of Equation (3.4) is the conditional tail loss probability due to the event $\mathbf{Y} \in \mathcal{S}_1$. By changing the conditional pmf of \mathbf{Y} from $p(\mathbf{y}|\psi)$ to $q(\mathbf{y}|\psi)$, we have the following form of it under the importance pmf $q(\mathbf{y}|\psi)$:

$$E_p[I(\mathbf{Y} \in \mathcal{S}_1, L > x)|\psi] = E_q[r(\mathbf{Y}|\psi)I(\mathbf{Y} \in \mathcal{S}_1, L > x)|\psi],$$

where $r(\mathbf{y}|\psi) = p(\mathbf{y}|\psi)/q(\mathbf{y}|\psi)$. For given $\Psi = \psi$, we let $q(\mathbf{Y} \in \mathcal{S}_1|\psi)$ be the conditional probability of the event $\{\mathbf{Y} \in \mathcal{S}_1\}$ with \mathbf{Y} following the pmf $q(\mathbf{y}|\psi)$. Under the same pmf of \mathbf{Y} , we let $E_q[r(\mathbf{Y}|\psi)I(L > x)|\mathbf{Y} \in \mathcal{S}_1, \psi]$ be the conditional expectation of $r(\mathbf{Y}|\psi)I(L > x)$ given $\mathbf{Y} \in \mathcal{S}_1$ and $\Psi = \psi$. By conditioning on the event $\{\mathbf{Y} \in \mathcal{S}_1\}$, the above equation is converted to

$$E_p[I(\mathbf{Y} \in \mathcal{S}_1, L > x)|\psi] = q(\mathbf{Y} \in \mathcal{S}_1|\psi)E_q[r(\mathbf{Y}|\psi)I(L > x)|\psi, \mathbf{Y} \in \mathcal{S}_1].$$

We denote $q(\mathbf{Y} \in \mathcal{S}_1|\psi)$ by $q_s(\psi)$ for simplicity of notation. Note that $q_s(\psi)$ is the probability that some of the heavy obligors do not default. Then, its values is computed as

$$q_s(\psi) = 1 - \prod_{j=1}^k q_j(\psi).$$

When it is given that $\Psi = \psi$ and $\mathbf{Y} \in \mathcal{S}_1$, the conditional joint pmf of \mathbf{Y} is given by

$$\tilde{q}(\mathbf{y}|\psi) = \frac{q(\mathbf{y}|\psi)}{q_s(\psi)}, \quad \mathbf{y} \in \mathcal{S}_1. \quad (3.5)$$

Then, $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$ is the conditional pmf of $q(\mathbf{y}|\boldsymbol{\psi})$ when the sample space is confined to \mathcal{S}_1 , which means

$$E_{\tilde{q}}[r(\mathbf{Y}|\boldsymbol{\psi})I(L > x)|\boldsymbol{\psi}, \mathbf{Y} \in \mathcal{S}_1] = E_{\tilde{q}}[r(\mathbf{Y}|\boldsymbol{\psi})I(L > x)|\boldsymbol{\psi}]. \quad (3.6)$$

From the above equations, it follows that Equation (3.4) is rewritten as

$$E_p[I(L > x)|\boldsymbol{\psi}] = p_a(\boldsymbol{\psi}) + q_s(\boldsymbol{\psi})E_{\tilde{q}}[r(\mathbf{Y}|\boldsymbol{\psi})I(L > x)|\boldsymbol{\psi}]. \quad (3.7)$$

In Equation (3.7), $E_p[I(L > x)|\boldsymbol{\psi}]$ is the conditional probability of the tail loss over x given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, and $p_a(\boldsymbol{\psi})$ is the tail loss probability due to the event $\{\mathbf{Y} \in \mathcal{S}_0\}$, which is calculated exactly. The tail loss probability due to the event $\{\mathbf{Y} \in \mathcal{S}_1\}$ is given by the last part of the equation, in which $q_s(\boldsymbol{\psi})$ is calculated exactly, and $E_{\tilde{q}}[r(\mathbf{Y}|\boldsymbol{\psi})I(L > x)|\mathbf{Y} \in \mathcal{S}_1, \boldsymbol{\psi}]$ is estimated by random sampling of the default vector $\mathbf{Y} \in \mathcal{S}_1$ from the importance pmf $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$. In this way, the conditional tail loss probability can be estimated by splitting the sample space \mathcal{S} into \mathcal{S}_0 and \mathcal{S}_1 . From the form of $E_p[I(L > x)|\boldsymbol{\psi}]$ in Equation (3.7), we obtain the following proposition.

Proposition 1. $\mathbb{P}(L > x)$ is represented in terms of $g(\boldsymbol{\psi})$ and $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$ as:

$$\mathbb{P}(L > x) = E_g \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} p_a(\boldsymbol{\Psi}) + q_s(\boldsymbol{\Psi}) E_{\tilde{q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}] \right]. \quad (3.8)$$

Proof: From the definition of the probability measure \mathbb{P} , it follows that

$$\mathbb{P}(L > x) = E_f[E_p[I(L > x)|\boldsymbol{\Psi}]].$$

Changing the pdf of $\boldsymbol{\Psi}$ from $f(\boldsymbol{\psi})$ to $g(\boldsymbol{\psi})$ gives

$$\mathbb{P}(L > x) = E_g \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} E_p[I(L > x)|\boldsymbol{\Psi}] \right].$$

By applying Equation (3.7) to above equation, we have

$$\mathbb{P}(L > x) = E_g \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} (p_a(\boldsymbol{\Psi}) + q_s(\boldsymbol{\Psi}) E_{\tilde{q}}[r(\mathbf{Y}|\boldsymbol{\Psi})I(L > x)|\boldsymbol{\Psi}]) \right].$$

If we note that $w(\boldsymbol{\Psi}, \mathbf{Y}) = (f(\boldsymbol{\Psi})/g(\boldsymbol{\Psi}))r(\mathbf{Y}|\boldsymbol{\Psi})$, then the proof completes. \square

We define $\tilde{\mathbb{Q}}$ as the probability measure on $\mathbb{R}^d \times \mathcal{S}_1$ induced from $g(\boldsymbol{\psi})\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$. It follows from the above equation that

$$\mathbb{P}(L > x) = E_{\tilde{\mathbb{Q}}} \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} p_a(\boldsymbol{\Psi}) + q_s(\boldsymbol{\Psi}) w(\boldsymbol{\Psi}, \mathbf{Y}) I(L > x) \right]. \quad (3.9)$$

Sampling from $\tilde{q}(\mathbf{y}; \boldsymbol{\psi})$ is done by the following procedure:

- i. sample Y_1, \dots, Y_k independently from the Bernoulli distribution with $\Pr(Y_j = 1) = q_j(\boldsymbol{\psi})$ for $1 \leq j \leq k$.
- ii. if $(Y_1, \dots, Y_k) = (1, \dots, 1)$, then resample Y_1, \dots, Y_k as described in (i) until $(Y_1, \dots, Y_k) \neq (1, \dots, 1)$.

- iii. sample Y_{k+1}, \dots, Y_m independently from the Bernoulli distribution with $\Pr(Y_j = 1) = q_j(\boldsymbol{\psi})$ for $k+1 \leq j \leq m$.

Then $\mathbf{Y} = (Y_1, \dots, Y_m)$ obtained by the above procedure is a sample from $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$. Suppose that we have samples $(\boldsymbol{\Psi}^{(1)}, \mathbf{Y}^{(1)}), \dots, (\boldsymbol{\Psi}^{(n)}, \mathbf{Y}^{(n)})$ from $\tilde{\mathbb{Q}}$. Then, from Equation (3.9), we can see that an importance sampling estimator for $\mathbb{P}(L > x)$ is given by

$$\hat{\theta}_n^S = \frac{1}{n} \sum_{i=1}^n \frac{f(\boldsymbol{\Psi}^{(i)})}{g(\boldsymbol{\Psi}^{(i)})} p_a(\boldsymbol{\Psi}^{(i)}) + q_s(\boldsymbol{\Psi}^{(i)}) w^{(i)} I(L^{(i)} > x), \quad (3.10)$$

where

$$w^{(i)} = \frac{f(\boldsymbol{\Psi}^{(i)})}{g(\boldsymbol{\Psi}^{(i)})} r(\mathbf{Y}^{(i)}|\boldsymbol{\Psi}^{(i)}).$$

In Equation (3.10), the sample average of $(f(\boldsymbol{\Psi}^{(i)})/g(\boldsymbol{\Psi}^{(i)}))p_a(\boldsymbol{\Psi}^{(i)})$, $i = 1, \dots, n$, is the conditional Monte Carlo estimator of $\mathbb{P}(L > x, \mathbf{Y} \in \mathcal{S}_0)$. The sample average of $q_s(\boldsymbol{\Psi}^{(i)})w^{(i)}I(L^{(i)} > x)$, $i = 1, \dots, n$, is a two-step importance sampling estimator of $\mathbb{P}(L > x, \mathbf{Y} \in \mathcal{S}_1)$.

Suppose that the typical occurrence of tail loss over the threshold x corresponds to the events that heavy obligors default simultaneously, i.e. the event $\{L > x\}$ is mostly due to the event $\{\mathbf{Y} \in \mathcal{S}_0\}$. Then, the event $\{\mathbf{Y} \in \mathcal{S}_1, L > x\}$ is rare compared to the event $\{\mathbf{Y} \in \mathcal{S}_0\}$. In this case, $q_s(\boldsymbol{\Psi}^{(i)})w^{(i)}I(L^{(i)} > x)$, $i = 1, \dots, n$, in Equation (3.10) has a very low value compared to $(f(\boldsymbol{\Psi}^{(i)})/g(\boldsymbol{\Psi}^{(i)}))p_a(\boldsymbol{\Psi}^{(i)})$ of the same equation in average. Thus, the estimator $\hat{\theta}_n^S$ is actually the conditional Monte Carlo estimator of $\Pr\{L > x\}$ in this case, and may have a lower variance than $\hat{\theta}_n^T$.

To compare the estimators $\hat{\theta}_n^T$ and $\hat{\theta}_n^S$ in terms of variance, we consider the case that the same importance pdf $g(\boldsymbol{\psi})$ for $\boldsymbol{\Psi}$ is adopted in both estimators, and assume that the same $q(\mathbf{y}|\boldsymbol{\psi})$ and the corresponding $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$ defined in Equation (3.5) are adopted as the importance pmf of \mathbf{Y} in $\hat{\theta}_n^T$ and in $\hat{\theta}_n^S$, respectively. It follows from Equation (2.11) that the variance of $\hat{\theta}_n^T$ is given by

$$V_{\mathbb{Q}}[\hat{\theta}_n^T] = \frac{1}{n} V_{\mathbb{Q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)].$$

By conditioning on $\boldsymbol{\Psi}$, we have

$$V_{\mathbb{Q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)] = V_g[E_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}]] + E_g[V_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}]].$$

Since $w(\boldsymbol{\Psi}, \mathbf{Y}) = (f(\boldsymbol{\Psi})/g(\boldsymbol{\Psi}))r(\mathbf{Y}|\boldsymbol{\Psi})$ and $E_q[r(\mathbf{Y}|\boldsymbol{\Psi})I(L > x)|\boldsymbol{\Psi}] = E_p[I(L > x)|\boldsymbol{\Psi}]$, the above equation is converted to

$$V_{\mathbb{Q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)] = V_g\left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})}p(L > x|\boldsymbol{\Psi})\right] + E_g[V_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}]], \quad (3.11)$$

where $p(L > x|\boldsymbol{\Psi}) = E_p[I(L > x)|\boldsymbol{\Psi}]$.

The following proposition gives a representation of $V[\hat{\theta}_n^S]$.

Proposition 2. *The variance of $\hat{\theta}_n^S$ is decomposed as:*

$$\begin{aligned} V[\hat{\theta}_n^S] = & \frac{1}{n} \left\{ V_g\left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})}p(L > x|\boldsymbol{\Psi})\right] + E_g[V_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(\mathbf{Y} \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]] \right\} \\ & - \frac{1}{n} E_{\mathbb{Q}}[q_a(\boldsymbol{\Psi})w(\boldsymbol{\Psi}, \mathbf{Y})^2 I(\mathbf{Y} \in \mathcal{S}_1, L > x)], \end{aligned}$$

where $q_a(\boldsymbol{\psi}) = \prod_{j=1}^m q_j(\boldsymbol{\psi})$.

Proof: For simplicity of presentation, we define $\theta(\boldsymbol{\psi}, \mathbf{y})$ as:

$$\theta(\boldsymbol{\psi}, \mathbf{y}) = \frac{f(\boldsymbol{\psi})}{g(\boldsymbol{\psi})} \{p_a(\boldsymbol{\psi}) + q_s(\boldsymbol{\psi})r(\mathbf{y}|\boldsymbol{\psi})I(L > x)\}, \quad \boldsymbol{\psi} \in \mathbb{R}^d, \mathbf{y} \in \{0, 1\}^d. \quad (3.12)$$

The number of n independent replications of $\theta(\boldsymbol{\psi}, \mathbf{y})$ constitute the summands of Equation (3.10). Then, it follows that

$$V_{\tilde{\mathbb{Q}}}[\hat{\theta}_n^S] = \frac{1}{n} V_{\tilde{\mathbb{Q}}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})].$$

By conditioning on the value of $\boldsymbol{\Psi}$, we have

$$V_{\tilde{\mathbb{Q}}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})] = V_g[E_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}]] + E_g[V_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}]]. \quad (3.13)$$

From Equations (3.7) and (3.12), we can see that

$$E_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}] = \frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} E_p[I(L > x)|\boldsymbol{\Psi}],$$

which gives a form of the first term in the right hand side of Equation (3.13) as:

$$V_g[E_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}]] = V_g\left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} E_p[I(L > x)|\boldsymbol{\Psi}]\right]. \quad (3.14)$$

From Equation (3.12), we also have that

$$V_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}] = q_s(\boldsymbol{\Psi})^2 V_{\tilde{q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}]. \quad (3.15)$$

Note that for $i = 1, 2, \dots$,

$$E_{\tilde{q}}[w(\boldsymbol{\Psi}, \mathbf{Y})^i I(L > x)|\boldsymbol{\Psi}] = \frac{E_q[w(\boldsymbol{\Psi}, \mathbf{Y})^i I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]}{q_s(\boldsymbol{\Psi})},$$

which gives

$$V_{\tilde{q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}] = \frac{E_q[w(\boldsymbol{\Psi}, \mathbf{Y})^2 I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]}{q_s(\boldsymbol{\Psi})} - \left(\frac{E_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]}{q_s(\boldsymbol{\Psi})} \right)^2.$$

Since $q_a(\boldsymbol{\Psi}) + q_s(\boldsymbol{\Psi}) = 1$, it follows from the above equation that

$$\begin{aligned} & q_s(\boldsymbol{\Psi})^2 V_{\tilde{q}}[w(\boldsymbol{\Psi}, \mathbf{Y})I(L > x)|\boldsymbol{\Psi}] \\ &= q_s(\boldsymbol{\Psi}) E_q[w(\boldsymbol{\Psi}, \mathbf{Y})^2 I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}] - E_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]^2 \\ &= V_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}] - q_a(\boldsymbol{\Psi}) E_q[w(\boldsymbol{\Psi}, \mathbf{Y})^2 I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}]. \end{aligned}$$

By applying the above equation to Equation (3.15), we have that the conditional variance of $\theta(\boldsymbol{\Psi}, \mathbf{Y})$ given $\boldsymbol{\Psi}$ is as follows:

$$V_{\tilde{q}}[\theta(\boldsymbol{\Psi}, \mathbf{Y})|\boldsymbol{\Psi}] = V_q[w(\boldsymbol{\Psi}, \mathbf{Y})I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}] - q_a(\boldsymbol{\Psi}) E_q[w(\boldsymbol{\Psi}, \mathbf{Y})^2 I(Y \in \mathcal{S}_1, L > x)|\boldsymbol{\Psi}].$$

Then, the last term of Equation (3.13) is represented as follows:

$$E_g \left[V_q [\theta(\Psi, Y) | \Psi] \right] = E_g \left[V_q [w(\Psi, Y) I(Y \in \mathcal{S}_1, L > x) | \Psi] \right] - E_{\mathbb{Q}} \left[q_a(\Psi) w(\Psi, Y)^2 I(Y \in \mathcal{S}_1, L > x) \right].$$

Combining the above equation with Equations (3.13) and (3.14) completes the proof. \square

From Proposition 2 and Equation (3.11), we derive an inequality on the difference between $V[\hat{\theta}_n^T]$ and $V[\hat{\theta}_n^S]$ as:

$$V[\hat{\theta}_n^T] - V[\hat{\theta}_n^S] \geq \frac{1}{n} \left\{ E_g \left[V_q [w(\Psi, Y) I(L > x) | \Psi] - V_q [w(\Psi, Y) I(Y \in \mathcal{S}_1, L > x) | \Psi] \right] \right\}. \quad (3.16)$$

In Equation (3.16), the term $V_q [w(\Psi, Y) I(L > x) | \Psi] - V_q [w(\Psi, Y) I(Y \in \mathcal{S}_1, L > x) | \Psi]$ is converted as follows:

$$\begin{aligned} & V_q [w(\Psi, Y) I(L > x) | \Psi] - V_q [w(\Psi, Y) I(Y \in \mathcal{S}_1, L > x) | \Psi] \\ &= \left(\frac{f(\Psi)}{g(\Psi)} \right)^2 \left(V_q [r(Y | \Psi) I(L > x) | \Psi] - V_q [r(Y | \Psi) I(Y \in \mathcal{S}_1, L > x) | \Psi] \right). \end{aligned}$$

Since $I(L > x) = I(Y \in \mathcal{S}_0) + I(Y \in \mathcal{S}_1, L > x)$, we have

$$\begin{aligned} & V_q [r(Y | \Psi) I(L > x) | \Psi] - V_q [r(Y | \Psi) I(Y \in \mathcal{S}_1, L > x) | \Psi] \\ &= V_q [r(Y | \Psi) I(Y \in \mathcal{S}_0) | \Psi] + 2\text{Cov}(r(Y | \Psi) I(Y \in \mathcal{S}_0), r(Y | \Psi) I(Y \in \mathcal{S}_1, L > x) | \Psi). \end{aligned}$$

The events $\{Y \in \mathcal{S}_0\}$ and $\{Y \in \mathcal{S}_1, L > x\}$ are mutually exclusive, which implies $I(Y \in \mathcal{S}_0) I(Y \in \mathcal{S}_1, L > x) = 0$. Then, the covariance term in the above equation is computed to be $-p_a(\Psi) E_p [I(Y \in \mathcal{S}_1, L > x) | \Psi]$. Then, it follows that

$$\begin{aligned} & V_q [r(Y | \Psi) I(L > x) | \Psi] - V_q [r(Y | \Psi) I(Y \in \mathcal{S}_1, L > x) | \Psi] \\ &= V_q [r(Y | \Psi) I(Y \in \mathcal{S}_0) | \Psi] - 2p_a(\Psi) E_p [I(Y \in \mathcal{S}_1, L > x) | \Psi]. \end{aligned}$$

We consider the case that the event $\{Y \in \mathcal{S}_1, L > x\}$ is rare compared to the event $\{Y \in \mathcal{S}_0\}$, i.e. $E_p [I(Y \in \mathcal{S}_1, L > x) | \Psi]$ is near 0, and $2p_a(\Psi) E_p [I(Y \in \mathcal{S}_1, L > x) | \Psi]$ is negligible compared to $V_q [r(Y | \Psi) I(Y \in \mathcal{S}_0) | \Psi]$. Then, the above equation says that, in this case, $V_q [r(Y | \Psi) I(L > x) | \Psi] \geq V_q [r(Y | \Psi) I(Y \in \mathcal{S}_1, L > x) | \Psi]$. Then, we can see from Equation (3.16) that $V[\hat{\theta}_n^T] \geq V[\hat{\theta}_n^S]$ (i.e., the proposed scheme works in this case).

3.2. The proposed scheme for the estimation of the expected shortfall

The following proposition gives a representation of $E_{\mathbb{P}} [I(L > x) L]$ in terms of a pdf $g(\psi)$ for Ψ and the conditional pmf $\tilde{q}(y | \psi)$ for Y given $\Psi = \psi$. The proposition tells us how to estimate the expected shortfall via an importance sampling of (Ψ, Y) from the probability measure $\tilde{\mathbb{Q}}$.

Proposition 3.

$$E_{\mathbb{P}} [I(L > x) L] = E_{\tilde{\mathbb{Q}}} \left[\frac{f(\Psi)}{g(\Psi)} p_a(\Psi) l(\Psi) + q_s(\Psi) w(\Psi, Y) I(L > x) L \right],$$

where $l(\Psi) = \sum_{j=1}^k c_j + \sum_{j=k+1}^m c_j p_j(\Psi)$.

Proof: Equation (3.3) says that the random variable $I(L > x)L$ is decomposed as:

$$I(L > x)L = I(Y \in \mathcal{S}_0)L + I(Y \in \mathcal{S}_1, L > x)L,$$

which gives

$$E_{\mathbb{P}}[I(L > x)L] = E_{\mathbb{P}}[I(Y \in \mathcal{S}_0)L] + E_{\mathbb{P}}[I(Y \in \mathcal{S}_1, L > x)L]. \quad (3.17)$$

It follows from Equation (3.2) that

$$E_p[I(Y \in \mathcal{S}_0)L|\boldsymbol{\psi}] = E_p \left[I(Y \in \mathcal{S}_0) \left(\sum_{i=1}^k c_j + \sum_{i=k+1}^m c_j Y_j \right) \middle| \boldsymbol{\psi} \right].$$

The event of $\{Y \in \mathcal{S}_0\}$ depends on only Y_1, Y_2, \dots, Y_k , and the r.v. $\sum_{j=k+1}^m c_j Y_j$ depends on only Y_{k+1}, \dots, Y_m . In the Bernoulli mixture model, Y_1, Y_2, \dots, Y_m are conditionally independent when $\boldsymbol{\Psi} = \boldsymbol{\psi}$ is given. Thus, for a given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $I(Y \in \mathcal{S}_0)$ and $\sum_{j=k+1}^m c_j Y_j$ are independent. Then, it follows that

$$\begin{aligned} E_p \left[I(Y \in \mathcal{S}_0) \left(\sum_{i=1}^k c_j + \sum_{i=k+1}^m c_j Y_j \right) \middle| \boldsymbol{\psi} \right] &= E_p[I(Y \in \mathcal{S}_0)|\boldsymbol{\psi}] E_p \left[\sum_{i=1}^k c_j + \sum_{i=k+1}^m c_j Y_j \middle| \boldsymbol{\psi} \right] \\ &= p_a(\boldsymbol{\psi})l(\boldsymbol{\psi}). \end{aligned}$$

Since $E_{\mathbb{P}}[I(Y \in \mathcal{S}_0)L] = E_f[E_p[I(Y \in \mathcal{S}_0)L|\boldsymbol{\Psi}]]$, the above equation says that

$$E_{\mathbb{P}}[I(Y \in \mathcal{S}_0)L] = E_f[p_a(\boldsymbol{\Psi})l(\boldsymbol{\Psi})].$$

By choosing $g(\boldsymbol{\psi})$ as the pdf of $\boldsymbol{\Psi}$ instead of $f(\boldsymbol{\psi})$, it follows from the above equation that

$$E_{\mathbb{P}}[I(Y \in \mathcal{S}_0)L] = E_g \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} p_a(\boldsymbol{\Psi})l(\boldsymbol{\Psi}) \right],$$

equivalently,

$$E_{\mathbb{P}}[I(Y \in \mathcal{S}_0)L] = E_{\tilde{Q}} \left[\frac{f(\boldsymbol{\Psi})}{g(\boldsymbol{\Psi})} p_a(\boldsymbol{\Psi})l(\boldsymbol{\Psi}) \right]. \quad (3.18)$$

It follows from Equation (2.9) that we have the following representation of the last term of Equation (3.17):

$$\begin{aligned} E_{\mathbb{P}}[I(Y \in \mathcal{S}_1, L > x)L] &= E_{\mathbb{Q}}[w(\boldsymbol{\Psi}, Y)I(Y \in \mathcal{S}_1, L > x)L] \\ &= E_g[E_q[w(\boldsymbol{\Psi}, Y)I(Y \in \mathcal{S}_1, L > x)L|\boldsymbol{\Psi}]]. \end{aligned}$$

By conditioning on the event $\{Y \in \mathcal{S}_1\}$, we have

$$E_q[w(\boldsymbol{\Psi}, Y)I(Y \in \mathcal{S}_1, L > x)L|\boldsymbol{\Psi}] = q_s(\boldsymbol{\Psi})E_q[w(\boldsymbol{\Psi}, Y)I(L > x)L|Y \in \mathcal{S}_1, \boldsymbol{\Psi}].$$

From the definition of $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$, we have

$$E_q[w(\boldsymbol{\Psi}, Y)I(Y \in \mathcal{S}_1, L > x)L|\boldsymbol{\Psi}] = q_s(\boldsymbol{\Psi})E_{\tilde{q}}[w(\boldsymbol{\Psi}, Y)I(L > x)L|\boldsymbol{\Psi}],$$

which gives

$$\begin{aligned} E_{\mathbb{Q}}[w(\Psi, Y)I(Y \in \mathcal{S}_1, L > x)L] &= E_g[q_s(\Psi)E_{\tilde{q}}[w(\Psi, Y)I(L > x)L|\Psi]] \\ &= E_{\tilde{q}}[q_s(\Psi)w(\Psi, Y)I(L > x)L]. \end{aligned}$$

Then, we have

$$E_{\mathbb{P}}[I(Y \in \mathcal{S}_1, L > x)L] = E_{\tilde{q}}[q_s(\Psi)w(\Psi, Y)I(L > x)L]. \quad (3.19)$$

Applying Equations (3.18) and (3.19) to Equation (3.17), we complete the proof. \square

Suppose that we have samples $(\Psi^{(1)}, Y^{(1)}), \dots, (\Psi^{(n)}, Y^{(n)})$ from $\tilde{\mathbb{Q}}$. Then, Proposition 3 says that an importance sampling estimator of $E_{\mathbb{P}}[I(L > x)L]$ is given by

$$\hat{\eta}_n^S = \frac{1}{n} \sum_{i=1}^n \left[\frac{f(\Psi^{(i)})}{g(\Psi^{(i)})} p_a(\Psi^{(i)}) l^{(i)} + q_s(Y^{(i)}) w^{(i)} I(L^{(i)} > x) L^{(i)} \right], \quad (3.20)$$

where

$$l^{(i)} = \sum_{j=1}^k c_j + \sum_{j=k+1}^m c_j p_j(\Psi^{(i)}).$$

Then, the expected shortfall $E[L|L > x]$ is estimated to be $\hat{\eta}_n^S / \hat{\theta}_n^S$.

3.3. The cross entropy method and exponential twisting for the proposed scheme

In Section 2.3, exponential twistings of $p_1(\psi), \dots, p_m(\psi)$ are considered as importance default probabilities of the obligors for given $\Psi = \psi$. To obtain the optimal value of the twisting parameter t , Glasserman and Li (2005) used the property that the second moment of $r(Y|\psi)I(L > x)$, which is an unbiased estimator of $p(L > x|\psi)$, is bounded above by $\exp\{-2tx + 2 \log M_L(t|\psi)\}$. When $E[L|\psi] > x$, minimizing the upper bound gives the pseudo-optimal value of t , which is given in Equation (2.17). In the same manner, we show that Equation (2.17) also gives a pseudo-optimal value of the twisting parameter $t(\psi)$ in our proposed scheme when we adopt an exponential twisting of $p_j(\psi)$, $j = 1, \dots, m$, as the importance default probability of obligor j .

Since $w(Y|\psi) = (f(\psi)/g(\psi))r(Y|\psi)$, it follows from Equation (3.10) that

$$\begin{aligned} nV_{\tilde{q}}[\hat{\theta}_n^S|\psi] &= \left(\frac{f(\psi)q_s(\psi)}{g(\psi)} \right)^2 V_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi] \\ &= \left(\frac{f(\psi)q_s(\psi)}{g(\psi)} \right)^2 \left(E_{\tilde{q}}[r(Y|\psi)^2 I(L > x)|\psi] - E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi]^2 \right). \end{aligned} \quad (3.21)$$

Note that $q_s(\psi)E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi] = E_p[I(L > x, Y \in \mathcal{S}_1)|\psi]$, which does not depend on the choice of $q(y|\psi)$. Since $r(Y|\psi) = \exp\{-tL + \log M_L(t|\psi)\}$, the value of t minimizing $V_{\tilde{q}}[\hat{\theta}_n^S|\psi]$ is the same as the t minimizing $E_{\tilde{q}}[r(Y|\psi)^2 I(L > x)|\psi]$. An upper bound of the latter term is given by

$$\begin{aligned} E_{\tilde{q}}[r(Y|\psi)^2 I(L > x)|\psi] &= E_{\tilde{q}}[\exp\{-2tL + 2 \log M_L(t|\psi)\} I(L > x)|\psi] \\ &\leq \exp\{-2xt + 2 \log M_L(t|\psi)\}. \end{aligned}$$

Algorithm 1 Importance sampling with splitting**Require:** $x, \bar{p}_1, \bar{p}_2, \dots, \bar{p}_m, c_1, c_2, \dots, c_m, p_1(\boldsymbol{\psi}), p_2(\boldsymbol{\psi}), \dots, p_m(\boldsymbol{\psi}), f(\boldsymbol{\psi}; \boldsymbol{\mu})$ **Ensure:** $\hat{\mathbb{P}}(L > x), \hat{E}_{\mathbb{P}}[L|L > x]$

- 1: Sample $\boldsymbol{\Psi}^{(1)}, \dots, \boldsymbol{\Psi}^{(M)}$ independently from $f(\boldsymbol{\psi}; \boldsymbol{\mu})$
- 2: **for** $i=1$ to M **do**
- 3: Find the value of $t(\boldsymbol{\Psi}^{(i)})$ defined in Equation (2.17)
- 4: Sample $\mathbf{Y}^{(i)}$ from $\tilde{q}(\mathbf{y}|\boldsymbol{\Psi}^{(i)})$
- 5: Compute $p_a(\boldsymbol{\Psi}^{(i)})$ and $q_s(\boldsymbol{\Psi}^{(i)})$
- 6: Compute $L^{(i)} = L(\mathbf{Y}^{(i)})$ and $r^{(i)} = r(\mathbf{Y}^{(i)}|\boldsymbol{\Psi}^{(i)})$
- 7: **end for**
- 8: Set $\boldsymbol{\nu}^*$ as the solution of the following stochastic maximization problem:

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \quad \frac{1}{M} \sum_{i=1}^M \left\{ p_a(\boldsymbol{\Psi}^{(i)}) + q_s(\boldsymbol{\Psi}^{(i)}) r^{(i)} I(L^{(i)} > x) \right\} \log f(\boldsymbol{\Psi}^{(i)}; \boldsymbol{\nu}) \quad (3.22)$$

- 9: Set $g(\boldsymbol{\psi}) = f(\boldsymbol{\psi}; \boldsymbol{\nu}^*)$
- 10: Sample $\boldsymbol{\Psi}^{(1)}, \dots, \boldsymbol{\Psi}^{(n)}$ independently from $g(\boldsymbol{\psi})$
- 11: **for** $i=1$ to n **do**
- 12: Find the value of $t(\boldsymbol{\Psi}^{(i)})$ defined in Eq. (2.17)
- 13: Sample $\mathbf{Y}^{(i)}$ from $\tilde{q}(\mathbf{y}|\boldsymbol{\Psi}^{(i)})$.
- 14: Compute $p_a(\boldsymbol{\Psi}^{(i)})$ and $q_s(\boldsymbol{\Psi}^{(i)})$
- 15: Compute $l^{(i)} = \sum_{j=1}^k c_j + \sum_{j=k+1}^m c_j p_j \boldsymbol{\Psi}^{(i)}$
- 16: Compute $L^{(i)} = L(\mathbf{Y}^{(i)})$ and $w^{(i)} = w(\mathbf{Y}^{(i)}|\boldsymbol{\Psi}^{(i)})$
- 17: **end for**
- 18: Compute

$$\hat{\theta}_n^S = \frac{1}{n} \sum_{i=1}^n \left[\frac{f(\boldsymbol{\Psi}^{(i)})}{g(\boldsymbol{\Psi}^{(i)})} p_a(\boldsymbol{\Psi}^{(i)}) + q_s(\boldsymbol{\Psi}^{(i)}) w^{(i)} I(L^{(i)} > x) \right]$$

- 19: Compute

$$\hat{\eta}_n^S = \frac{1}{n} \sum_{i=1}^n \left[\frac{f(\boldsymbol{\Psi}^{(i)})}{g(\boldsymbol{\Psi}^{(i)})} p_a(\boldsymbol{\Psi}^{(i)}) l^{(i)} + q_s(\boldsymbol{\Psi}^{(i)}) w^{(i)} I(L^{(i)} > x) L^{(i)} \right]$$

- 20: Return

$$\hat{\mathbb{P}}(L > x) = \hat{\eta}_n^S, \quad \hat{E}[L|L > x] = \frac{\hat{\eta}_n^S}{\hat{\theta}_n^S}.$$

The value of $t(\boldsymbol{\psi})$ given in Equation (2.17) minimizes the last term in the above equation, and it is a pseudo-optimal value of t . Then, for $j = 1, \dots, m$, the same $q_j^*(\boldsymbol{\psi}) = q_j(t(\boldsymbol{\psi}), \boldsymbol{\psi})$ as in Section 2.3 is the optimal importance default probability of obligor j , and the importance pmf $\tilde{q}(\mathbf{y}|\boldsymbol{\psi})$ of $\mathbf{Y} \in \mathcal{S}_1$ is given by

$$\tilde{q}(\mathbf{y}|\boldsymbol{\psi}) = \frac{\prod_{j=1}^m q_j^*(\boldsymbol{\psi})^{y_j} (1 - q_j^*(\boldsymbol{\psi}))^{1-y_j}}{1 - \prod_{j=1}^k q_j^*(\boldsymbol{\psi})}, \quad \mathbf{y} \in \mathcal{S}_1.$$

As we have done in Section 2.4, we assume that $\boldsymbol{\Psi}$ has the pdf $f(\boldsymbol{\psi}; \boldsymbol{\mu})$, and confine the parametric

family of the importance pdf of Ψ to $\mathcal{F} = \{f(\psi; \nu)\}$ in Equation (2.20). Using the cross entropy method described in Section 2.4, we can find the optimal value of ν . Note that $p(L > x|\psi) = p_a(\psi) + q_s(\psi)E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi]$, and that

$$\mathbb{P}(L > x) = \int \left(p_a(\psi) + q_s(\psi)E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi] \right) f(\psi; \mu) d\psi.$$

Then, the integrand $(p_a(\psi) + q_s(\psi)E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi])f(\psi; \mu)$ in the above integral is the zero-variance pdf for the estimation of $\mathbb{P}(L > x)$ up to a constant. As it is explained in Section 2.4, the optimal value of ν can be found by solving the following maximization problem:

$$\nu^* = \operatorname{argmax}_{\nu} \int \left\{ p_a(\psi) + q_s(\psi)E_{\tilde{q}}[r(Y|\psi)I(L > x)|\psi] \right\} \log f(\psi; \nu) f(\psi; \mu) d\psi. \quad (3.23)$$

Suppose that we have samples $(\Psi^{(i)}, Y^{(i)})$, $i = 1, 2, \dots, M$, from the distribution $f(\psi; \mu)\tilde{q}(y|\psi)$. Then, the stochastic optimization problem corresponding to Equation (3.23) is that

$$\operatorname{maximize}_{\nu} \frac{1}{M} \sum_{i=1}^M \left(p_a(\Psi^{(i)}) + q_s(\Psi^{(i)})r(Y^{(i)}|\Psi^{(i)})I(L^{(i)} > x) \right) \log f(\Psi^{(i)}; \nu). \quad (3.24)$$

Algorithm 1 shows how to obtain $\hat{\theta}_n^S$ and $\hat{\eta}_n^S$ by applying the cross entropy method and the exponential twisting described above.

Now, we consider the case that Ψ is a d -dimensional multivariate normal vector with mean μ and variance-covariance matrix Σ . We confine the family of the possible importance sampling distribution of Ψ to the family of the d -dimensional normal distributions with same variance-covariance matrix Σ , i.e.

$$\mathcal{F} = \{ \phi_d(\psi; \nu, \Sigma) : \nu \in \mathbb{R}^d \},$$

where $\phi_d(\psi; \nu, \Sigma)$ is the p.d.f of the d -dimensional normal distributions with mean ν and the variance-covariance matrix Σ . In this case, the solution of the optimization problem Equation (3.24) can be found analytically (Chan and Kroese, 2010) as follows:

$$\nu^* = \frac{\sum_{i=1}^M \left(p_a(\Psi^{(i)}) + q_s(\Psi^{(i)})r(Y^{(i)}|\Psi^{(i)})I(L^{(i)} > x) \right) \Psi^{(i)}}{\sum_{i=1}^M p_a(\Psi^{(i)}) + q_s(\Psi^{(i)})r(Y^{(i)}|\Psi^{(i)})I(L^{(i)} > x)}. \quad (3.25)$$

Now, Equation (3.22) in Algorithm 1 is replaced with Equation (3.25).

4. Numerical results

In this section, we give some numerical results on the performance of the proposed method. We estimate numerically tail loss probabilities over various thresholds and their corresponding estimated values of expected shortfalls. Crude Monte Carlo, two step importance sampling with cross entropy method and exponential twisting described in Section 2, and our proposed scheme are used for the estimation of each target value, respectively. We call the methods CMC, CE-ET, and IS-Sp in order. In obtaining of an estimate, the efficiency of a Monte Carlo simulation is taken as inversely proportional to the product of the sampling variance of the estimate and the simulation time to get the estimate (Glynn and Whitt, 1992). We compare the performance of the methods CMC, CE-ET, and IS-Sp in

terms of the product of the sampling variance and the simulation time. The smaller the value is, the better the performance is.

For the mixing distribution in the Bernoulli mixture model, we consider the probit-normal mixing distribution (Frey and McNeil, 2003, p.354). In the model, the factor vector Ψ follows the d -dimensional standard normal distribution $N_d(\mathbf{0}, \Sigma)$ with a correlation matrix Σ , and $p_j(\psi)$ has the following form:

$$p_j(\psi) = \Phi(\mathbf{a}'_j \psi + b_j), \quad j = 1, \dots, m,$$

where Φ is the c.d.f. of the standard normal, \mathbf{a}_j is a d -dimensional constant vector to adjust the effect of the factor vector on the conditional default probability of obligor i . By setting b_j as follows:

$$b_j = \sqrt{1 + \mathbf{a}'_j \Sigma \mathbf{a}_j} \Phi^{-1}(\bar{p}_j), \quad j = 1, \dots, m,$$

the marginal default probability condition of Equation (2.7) is satisfied.

In the numerical study, we set the correlation matrix Σ as

$$\sigma = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix},$$

and the constant vector as

$$\mathbf{a}_j = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad j = 1, \dots, m.$$

The number of obligors, m , are set to be 1000. The marginal default probabilities of obligors are randomly generated from the uniform distribution on $[0, 0.02]$. The once generated values are used for all simulations, i.e. the marginal default probabilities of obligors do not change from simulation to simulation.

Two sets of exposures of obligors are randomly generated. In order to get sets of heavily skewed exposures, we generate one set of exposures from Pareto distribution with shape parameter 0.8, and the other set from Pareto distribution with shape parameter 1.2. The former set of exposures contains larger extreme values and more skewed to the right. The generated exposures are also used for all simulations. For each set of exposures, we obtain the tail loss probabilities and the expected shortfalls over 4 different thresholds. The k^{th} threshold x_k , $k = 1, \dots, 4$, is set to be as:

$$x_k = \sum_{j=1}^k c_j - 0.1,$$

where c_1, c_2, \dots, c_m are assumed to be in their descending order. Then, we have $x_1 < \dots < x_4$. If the first k obligors default, then the loss is over x_k , $k = 1, \dots, 4$. It means that the threshold x_k corresponds to the case that the set of heavy obligors is $\{1, \dots, k\}$.

To obtain the tail loss probability $\hat{\theta}$ over the threshold x_k , we generate $n = 10^5$ of the default vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, and compute the corresponding losses $L^{(1)}, \dots, L^{(n)}$ with each of the methods, CMC, CE-ET, and IS-Sp. Table 1 shows the value of x_k , $k = 1, \dots, 4$, and the cumulative probability of x_k (i.e., the probability of which quantile is x_k). Exact values of the cumulative probabilities are hard to compute, and they are estimated to be $(1 - \hat{\theta}^T)$'s, in which the tail loss probability $\hat{\theta}^T$ over

Table 1: Thresholds, cumulative probabilities, and corresponding simulation times in the cases of $\alpha = 0.8$ and $\alpha = 1.2$

α	k	Threshold	Cumulative probability	Simulation time (sec.)		
				CMC	CE-ET	IS-Sp
(a) 0.8	1	11027.22	98.590%	14.53	70.12	82.58
	2	16525.18	99.784%	13.91	70.57	79.20
	3	20669.69	99.935%	13.95	72.28	80.75
	4	23529.80	99.973%	14.30	72.21	91.99
(b) 1.2	1	293.74	95.290%	13.63	43.86	46.10
	2	518.34	98.170%	13.55	42.71	43.60
	3	717.20	99.086%	13.60	41.26	43.82
	4	884.89	99.453%	13.49	44.05	46.64

Table 2: Estimation of the tail loss probabilities for various values of thresholds in the cases of $\alpha = 0.8$ and $\alpha = 1.2$

α	k	$\hat{\theta}^T$	S.E			Time * Variance		
			CMC	CE-ET	IS-Sp	CMC	CE-ET	IS-Sp
(a) 0.8	1	1.41×10^{-2}	3.67×10^{-4}	4.69×10^{-5}	1.65×10^{-5}	1.96×10^{-6}	1.54×10^{-7}	2.25×10^{-8}
	2	2.16×10^{-3}	1.48×10^{-4}	9.50×10^{-6}	7.40×10^{-6}	3.04×10^{-7}	6.36×10^{-9}	4.28×10^{-9}
	3	6.52×10^{-4}	7.55×10^{-5}	3.66×10^{-6}	3.05×10^{-6}	7.95×10^{-8}	9.70×10^{-10}	7.51×10^{-10}
	4	2.71×10^{-4}	4.58×10^{-5}	1.89×10^{-6}	1.23×10^{-6}	3.00×10^{-8}	2.57×10^{-10}	1.39×10^{-10}
(b) 1.2	1	4.71×10^{-2}	6.71×10^{-4}	1.86×10^{-4}	1.62×10^{-4}	6.13×10^{-6}	1.52×10^{-6}	1.22×10^{-6}
	2	1.83×10^{-2}	4.19×10^{-4}	8.12×10^{-5}	8.03×10^{-5}	2.38×10^{-6}	2.82×10^{-7}	2.81×10^{-7}
	3	9.14×10^{-3}	3.02×10^{-4}	4.40×10^{-5}	4.35×10^{-5}	1.24×10^{-6}	8.00×10^{-8}	8.29×10^{-8}
	4	5.47×10^{-3}	2.31×10^{-4}	2.77×10^{-5}	2.72×10^{-5}	7.19×10^{-7}	3.39×10^{-8}	3.46×10^{-8}

Table 3: Estimation of the expected shortfalls for various values of thresholds in the cases of $\alpha = 0.8$ and $\alpha = 1.2$

α	k	$E[\widehat{L} L > x]$	S.E			Time * Variance ratio		
			CMC	CE-ET	IS-Sp	CMC	CE-ET	IS-Sp
(a) 0.8	1	1.33×10^4	3.71×10^2	4.15×10^1	1.66×10^1	2.00×10^6	1.21×10^5	2.27×10^4
	2	1.98×10^4	1.36×10^3	8.06×10^2	6.32×10^2	2.57×10^7	4.59×10^5	3.16×10^5
	3	2.37×10^4	3.18×10^3	1.27×10^2	1.05×10^2	1.41×10^8	1.17×10^6	9.00×10^5
	4	2.63×10^4	5.68×10^3	1.75×10^2	1.14×10^2	4.61×10^8	2.21×10^6	1.21×10^6
(b) 1.2	1	5.56×10^2	9.19	1.82	1.54	1.15×10^3	1.45×10^2	1.09×10^2
	2	8.34×10^2	20.71	3.15	3.04	5.81×10^3	4.25×10^2	4.02×10^2
	3	1.06×10^3	37.06	4.37	4.33	1.87×10^4	7.89×10^2	8.22×10^2
	4	1.24×10^3	55.67	5.49	5.44	4.18×10^4	1.33×10^3	1.38×10^3

each threshold is given in Table 2. For each combination of the threshold, the sampling method, and the shape parameter, Table 1 also shows the simulation times to get the estimates of the tail loss probability and the expected shortfall which appear in Tables 2 and 3, respectively.

Table 1 shows that the simulation times of CE-ET and IS-Sp are much longer than those of CMC in each simulation with the same threshold and α . However, the simulation times of IS-Sp are only somewhat longer than those of CE-ET in the case of $\alpha = 0.8$, and the difference of two times are less than 3 seconds (or less than 10%) in the case of $\alpha = 1.2$. This means that finding the optimal twisting parameter $t(\psi)$ in Equation (2.17) is the bottleneck of CE-ET method and also IS-Sp method because the conditional default probabilities are not twisted in CMC method. This also explains the difference of simulation times of CE-ET (and also IS-Sp) between the cases of $\alpha = 0.8$ and $\alpha = 1.2$. Since the exposures in the case of $\alpha = 0.8$ are spread more widely than $\alpha = 1.2$, it needs more steps to find the

optimal twisting parameter numerically. We also can see that for a given $\Psi = \psi$, the generation of the default vector \mathbf{Y} from $\tilde{q}(\mathbf{y}|\psi)$ requires only a little additional time compared to the generation from $q^*(\mathbf{y}|\psi)$, and that the additional time is short in the case of $\alpha = 1.2$.

For each set of exposures, the estimated tail loss probabilities over the thresholds specified in Table 1 and the corresponding standard errors are given in Table 2. It also shows the time * variance for each estimation. The simulation time taken to get an estimate is inversely proportional to the variance of the estimate. Equations (2.2), (2.11), and (3.10) say that the variances of the estimators in the equations are proportional to $1/n$ since each estimator in the equations is an average of n i.i.d. random variables. The time taken to generate the random variables is clearly linear with n , while the time taken to compute the average of the generated random variables is negligible compared to the time of generation of the random variables. Since the total simulation time to get an estimate consists of the former and the latter times, the value of 'time*variance' of an estimate is nearly constant with respect to n . It implies that to reduce the variance of an estimator by r times, r times more simulation time is required. Suppose that the time * variance of an estimate is m times larger than that of the other. In order to make the variance of the two estimators the same, the required simulation time of the former is m times larger than the latter.

In Table 2(a), we can see that CE-ET and IS-Sp have about 100 times less value of time * variance compared to CMC in the case of $\alpha = 0.8$. CE-ET and IS-Sp are about 100 times faster than CMC in terms of simulation time to obtain the same estimation error (i.e., the variance of the estimate). We can see that the time * variances of CE-ET are from 1.3 to 6.8 times those of IS-Sp method, which means that IS-Sp is more efficient than CE-ET in the estimation of the tail loss probability.

Table 2(b) says that the performances of the methods have different behavior in the case of $\alpha = 1.2$. In this case, CE-ET and IS-Sp have about 10 times less value of time * variance compared to CMC. The time * variances of CE-ET and IS-Sp have the similar values to each other for $k > 1$. Therefore, there is no need to use IS-Sp instead of CE-ET for $k > 1$. When $k = 1$, IS-Sp method shows the better performance than CE-EC.

Tables 3(a) and (b) show the estimated expected shortfalls over the thresholds shown in Table 1 and the corresponding standard errors in the cases of $\alpha = 0.8$ and $\alpha = 1.2$, respectively. The values of time * variances for the estimated values are also shown. The same conclusion as the case of the tail loss probabilities in Table 2 are obtained. CE-ET and IS-Sp are faster than CMC in both cases of $\alpha = 0.8$ and $\alpha = 1.2$. In the former case, IS-Sp is more efficient than CE-ET in the estimation of the expected shortfall. In the latter case, there is no need to use IS-Sp instead of CE-ET for $k > 1$. When $k = 1$, IS-Sp method shows the better performance than CE-EC.

Tables 2 and 3 say that IS-Sp shows better performance than CE-EC as well as CMC for the case of $\alpha = 0.8$. However, the case of $\alpha = 1.2$, IS-Sp and CE-EC have the similar efficiency for $k > 1$. In the case of $\alpha = 0.8$, IS-Sp has better time efficiency than CE-EC with various values of thresholds; however this conclusion is only valid with low thresholds in the case of $\alpha = 1.2$. It implies that the proposed method may show better performance in the case of heavily skewed set of exposures, especially when the fitted value of α is less than 1.

5. Conclusion

We proposed an importance sampling scheme to estimate the tail loss probability over a threshold and the corresponding expected shortfall. The proposed scheme is for the Bernoulli mixture model with heavily skewed exposures. In the model, the typical occurrence of tail loss over a high threshold corresponds to the events that small number of the heavy obligors default simultaneously. In the

proposed scheme, the sample space of the default events is split into two subspaces, one of which consists of the described typical events. We developed a conditional Monte Carlo to get estimates of tail loss probability and the expected shortfall due to the events of this subspace, while an importance sampling was proposed for the estimation of the risk values due to the default events of the other subspace. The proposed scheme is a type of two-step importance sampling. For choosing optimally the importance distribution of the factor variables and the importance default probabilities of the obligors in the proposed scheme, we proposed how to apply the cross entropy method and the exponential twisting. We compared the proposed scheme and the two-step importance sampling without splitting of sample space in terms of variance, and discussed that the proposed scheme may work in the case that simultaneous defaults of small number of heavy obligors occur the tail loss over the threshold. In numerical study, we showed that the proposed scheme might be efficient compared to the two-step importance sampling without splitting in the described case.

The proposed scheme can be extended to the case of random exposures. In this case, the set of heavy obligors changes according to the set of exposures, and the typical default event is a random vector, while it is a fixed vector in the case of constant exposures. The approach to analyze the problem may be the same as that of constant exposures.

Acknowledgements

The author would like to thank the anonymous referees for their comments and suggestions on the first draft of this paper. Their valuable suggestions have greatly improve the representation of the paper.

References

- Arvanitis A and Gregory J (2001). *Credit: The Complete Guide to Pricing, Hedging and Risk Management*, Risk Books, London.
- Başoğlu İ, Hörmann W, and Sak H (2013). Optimally stratified importance sampling for portfolio risk with multiple loss thresholds, *Optimization: A Journal of Mathematical Programming and Operations Research*, **62**, 1451–1471.
- Başoğlu İ, Hörmann W, and Sak H (2018). Efficient simulations for a Bernoulli mixture model of portfolio credit risk, *Annals of Operations Research*, **260**, 113–128.
- Bassamboo A, Juneja S, and Zeevi A (2008). Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation, *Operations Research*, **56**, 593–606.
- Brereton TJ, Kroese DP, and Chan JC (2013). Monte Carlo methods for portfolio credit risk. *Credit Portfolio Securitization and Derivatives*, (H. Scheule and D. Rosch Eds), John Wiley & Sons, New York.
- Bucklew JA, Ney P, and Sadowsky JS (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains, *Journal of Applied Probability*, **27**, 44–59.
- Chan JCC and Kroese DP (2010). Efficient estimation of large portfolio loss probabilities in t -copula models, *European Journal of Operational Research*, **205**, 361–367.
- Egloff D, Leippold M, Jöhri S, and Dalbert C (2005). *Optimal importance sampling for credit portfolios with stochastic approximation*. Available from: <https://ssrn.com/abstract=693441>
- Frey R and McNeil AJ (2003). Dependent defaults in models of portfolio credit risk, *Journal of Risk*, **6**, 59–92.
- Glasserman P (2013). *Monte Carlo Methods in Financial Engineering* (Volume 53), Springer-Verlag, New York.
- Glasserman P and Li J (2005). Importance sampling for portfolio credit risk, *Management Science*,

- 51**, 1593–1732.
- Glynn PW and Whitt W (1992). The asymptotic efficiency of simulation estimators, *Operations Research*, **40**, 505–520.
- Gundlach M and Lehrbass F (2013). *CreditRisk+ in the Banking Industry*, Springer, Berlin.
- Kalkbrener M, Lotter H, and Overbeck L (2004). Sensible and efficient capital allocation for credit portfolios, *Risk*, **17**, S19–S24.
- McNeil AJ, Frey R, and Embrechts P (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton.
- Merino S and Nyfeler MA (2004). Applying importance sampling for estimating coherent credit risk contributions, *Quantitative Finance*, **4**, 199–207.
- Rubinstein RY and Kroese DP (2016). *Simulation and the Monte Carlo Method* (3rd ed), John Wiley & Sons, New York.
- Sadowsky JS (1993). On the optimality and stability of exponential twisting in Monte Carlo estimation, *IEEE Transactions on Information Theory*, **39**, 119–128.
- Sak H and Hörmann W (2012). Fast simulations in credit risk, *Quantitative Finance*, **12**, 1557–1569.

Received September 20, 2019; Revised February 19, 2020; Accepted March 18, 2020