Computation of Credit Portfolio Loss Distribution by a Cross Entropy Method

Xiaoying Han* Ruodu Wang†

July 23, 2015

Abstract

Quantification and management of credit risk is always crucial for the financial industry. Computing credit risk is generally a challenging task while correlated defaults exist. Traditional approaches such as exponential twisting are model specific and often involve difficult analysis, therefore computational methods are sought to estimate the credit risk when analysis is unavailable. The accurate measurement of credit risk is often a rare-event simulation problem, i.e., calculating probabilities (which are usually small) of extreme losses. It is well-known that the Monte Carlo (MC) method may become slow and expensive for such problems. Importance sampling (IS), a variance-reduction technique, can then be utilized for rare-event simulation for credit risk management. In this work, we propose the implementation of a special IS procedure, the cross-entropy (CE) method, to simulate credit risk models. More specifically, we obtain iteratively biasing probability density functions (PDF's) for credit portfolio losses by the CE method, and then combine the results from each stage by the technique of multiple importance sampling (MIS) to obtain a complete PDF. The main advantage of this method is that it can avoid the nontrivial analysis required by a general IS method, and therefore simplifies the estimation of loss distributions. Moreover, this approach is generic and can be applied to a wide variety of models with little modifications. In particular, we apply this approach to a normal copula model and a t-copula model to estimate the probabilities of extreme portfolio losses under the models. Numerical examples are provided to demonstrate the performance of our method.

Keywords: Credit risk; Monte Carlo method; Importance sampling; Cross-entropy method; Normal copula; Student t-copula.

1 Introduction

Credit risk, the risk of failure of an obligor to make contractual payments, is one of the major risks that financial institutions may encounter during business activities

*Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA, (xzh0003@auburn.edu).

†Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada, (wang@uwaterloo.ca).

[9]. The study of credit risk modeling has always been an active topic; a main focus of relevant studies is the computation of the loss distribution. In modern credit risk management where a portfolio view is taken, the models are expected to capture the effects of dependence across sources of credit risk to which a financial institution is exposed. Hence, the complexity of both the models used and the computational methods required to calculate outputs of a model is largely increased.

For the commonly applied models that take the dependence into account, such as the normal copula (NC) model [17, 22] or dependent risk factors for the CreditRisk+ model [33], it is difficult or impossible to calculate loss distributions analytically. Approximation via Monte Carlo (MC) simulation is widely used [13, 14] to obtain loss distributions numerically. In practice, since default probabilities are usually low for highly rated obligors and risk management is particularly concerned with rare but significant losses, the accurate measurement of the loss becomes a rare-event simulation where the standard MC method loses its efficiency. This issue potentially makes importance sampling (IS) attractive [1, 5, 15, 16, 29].

The key of IS is to design a good biasing PDF. We propose a new numerical method in this paper to calculate the portfolio loss distribution based on the cross-entropy (CE) [3, 18, 26, 27]. The CE method is a special IS procedure proposed by Rubinstein [26], and has been used in many applications in the field of rare-event simulation such as optical communications [23] and queuing networks [4]. Basically, the CE method finds a good biasing PDF by iteratively minimizing the Kullback-Leibler (K-L) distance [21] between the IS PDF and the zero-variance *optimal distribution*. The most interesting idea of this method is that the minimization procedure requires only little knowledge of the optimal distribution, apart from a crucial normalization constant.

While traditional approaches to compute credit risk under correlated defaults are generally model specific, our method is generic which can be applied to a wide variety of related models with little modifications. In this work we use a normal copula model and a t-copula model as test models to apply our method. The normal copula model, once popular, has been criticized for not being able to capture tail dependence, and it is known to be blamed by some people for the financial crisis in 2008 (e.g. [10]). However, it is still a good start point for illustrative and comparison purposes due to the viable factor decomposition of normal random variables. The t-copula model has also been increasingly popular to model vectors of risk factor log returns, due to its ability to model the extremal dependent of the risk factors and also the ease with which the parameters of the t-copula can be estimated from data. Both copulas belong to the family of elliptical copulas (see e.g. [11]) and their decomposition properties will be used in our method.

The rest of this paper is organized as follows. In section 2 we introduce the new technique of the IS-CE method. In section 3 we introduce the framework of portfolio credit risk followed by the normal and $t-$copula models. In section 4 we describe the procedure of applying the proposed IS-CE method to portfolio credit risk models. Numerical examples are provided in section 5 and closing remarks are given in section 6.

2 Importance sampling and the cross-entropy method

Before proceeding to discuss the importance sampling and cross-entropy method applied to calculating portfolio credit losses, we first introduce the general idea of combination of IS (see general references in [5, 13, 30]) and the CE method for rare probability estimation. Suppose that X is an n-dimensional random variable with probability density function (PDF) $p(x)$, and we are interested in the probability Q that a measurable function $f(X)$ falls in a specific region D. This probability can be expressed as

$$Q = \Pr(f(X) \in D) = \int_{f(x)\in D} p(x)\,dx = \mathbb{E}[\mathrm{I}_D(f(X))] = \int \mathrm{I}_D(f(x))p(x)dx, \tag{1}$$

where $\mathrm{I}_D(f)$ is the indicator function defined as $\mathrm{I}_D(f) = 1$ for $f \in D$ and $\mathrm{I}_D(f) = 0$ otherwise. Throughout this paper, we assume that all integrals are finite unless explicitly stated. Using the well-known MC method, one can estimate the probability by

$$\hat{Q} = \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_D(f(X^{(m)}))\,, \tag{2}$$

where $X^{(1)}, \cdots, X^{(M)}$ are i.i.d. copies of X and M is the total number of samples. However, if $Q \ll 1$, estimation by standard MC simulations becomes impractical due to the large number of samples needed. Importance sampling then can be used to resolve this problem. Note that Eq. (1) can be written as

$$Q = \int \mathrm{I}_D(f(x))p(x)\,dx = \int \mathrm{I}_D(f(x)) \cdot (p(x)/p^*(x)) \cdot p^*(x)\,dx\,, \tag{3}$$

where $p^*(x)$ is called a biasing pdf. We then estimate Q by

$$\hat{Q}^* = \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_D(f(X^{(m)}))R(X^{(m)})\,, \tag{4}$$

where samples are drawn according to $p^*(x)$ and $R(x) = p(x)/p^*(x)$ is the likelihood ratio.

The fundamental idea of IS is to design a good biasing strategy, i.e., to find a biasing PDF $p^*(x)$ that encourages the realizations to visit the *most likely* region of interest as frequently as possible. It is well known that an optimal biasing PDF

$$\tilde{p} = \mathrm{I}_D(f(x))p(x)/Q \tag{5}$$

exists in principle. Eq. (5) itself is not useful as it requires knowledge of the posterior probability Q. However, one can find a good biasing PDF by requiring it to be "close" to the optimal biasing PDF, in terms of some measure of distance. A particular convenient choice is the KL distance [21]:

$$\mathrm{dist}(g,h) = \mathbb{E}_g\left[\ln\left(g(X)/h(X)\right)\right] = \int g(x)\ln g(x)\,dx - \int g(x)\ln h(x)\,dx\,, \tag{6}$$

which is also known as the cross entropy between two probability distributions $g(x)$ and $h(x)$. Here and in the following, $\mathbb{E}_g$ means the random variable X is chosen from the density function g.

Minimizing the cross entropy between $\tilde{p}(x)$ and $p^*(x)$,

$$\mathrm{dist}(\tilde{p}, p^*) = \mathbb{E}_p \left[\ln\left(\tilde{p}(X)/p^*(X)\right)\right] = \int \tilde{p}(x)\ln\tilde{p}(x)\,dx - \int \tilde{p}(x)\ln p^*(x)\,dx\,, \tag{7}$$

is equivalent to maximizing $\int \tilde{p}(x)\ln p^*(x)\,dx$ since the first integral on the right-hand-side of Eq. (7), $\int \tilde{p}(x)\ln\tilde{p}(x)\,dx$, is fixed. According to (5),

$$\int \tilde{p}(x)\ln p^*(x)\,dx = \int \mathrm{I}_D(f(x))\cdot\ln\left(p^*(x)\right)\cdot p(x)/Q\;dx\,,$$

and therefore, minimizing the cross entropy between $\tilde{p}(x)$ and $p^*(x)$ is equivalent to maximizing the expectation $\mathbb{E}_p[\mathrm{I}_D(f(X))\,\ln p^*(X)]$.

Suppose that all the potential distributions of X are selected from a parametrized family of PDF $\{p(x;v)\}$, where v is the reference parameter, then maximizing $\int \tilde{p}(x)\ln p^*(x)\,dx$ becomes parametric with respect to v:

$$\max_v \mathcal{D}(v) := \max_v \int \tilde{p}(x)\ln p^*(x)\,dx\,. \tag{8}$$

Let μ be the parameter of unbiased distribution, $p(x;\mu)$, and denote by w_i the parameter of the empirical biasing distribution at step i, $\pi(x;w_i)$. Applying the importance-sampled Monte Carlo (ISMC) simulation, one obtains a stochastic maximization program

$$\max_v \hat{\mathcal{D}}(v) = \max_v \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_D(f(X_i^{(m)}))\,R_i(X^{(m)};\; u,\; w)\ln p(X^{(m)};\; v)\,, \tag{9}$$

where $R_i(X^{(m)};\; u,\; w) = p(X^{(m)};u)/\pi(X^{(m)};w_i)$ and the samples $\{X_i^{(m)}\}$ are generated according to $\pi(x;w_i)$. The optimal biasing PDF then can be adaptively traced by performing the following steps:

1. Set the initial distribution parameter of $\pi(x;w_i)$ to be $w_0 = \mu$.
2. At step i, generate samples $\{X_i^{(m)}\}$ according to $\pi(x;w_i)$.
3. Solve Eq. (9) for v.
4. If $\hat{\mathcal{D}}(v) \le \hat{\mathcal{D}}(w_i)$, proceed to step 5, otherwise set $w_{i+1} = v$, $i = i+1$, and reiterate from step 2.
5. Perform the ISMC simulation using the biasing PDF $\pi(x;w_i)$ to calculate the sought probability Q.

In summary, the above algorithm first searches a good biasing PDF $p^*(x) \approx \pi(x; w_i)$ by minimizing the CE distance between $p^*(x)$ and $\tilde{p}(x)$, and then implements IS using the determined distribution $p^*(x)$.

The major difficulty of this algorithm occurs at step 3, because the techniques for solving Eq. (9) depend on specific problems and is usually complicated. However, if function $\hat{\mathcal{D}}$ is convex and differentiable with respect to v, the solutions of (9) can be obtained by solving a system of equations [3]:

$$\frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_D(f(X^{(m)}))R_i(X^{(m)};\ u,\ w)\nabla \ln p(X^{(m)};\ v) = 0\,, \tag{10}$$

which can be solved *analytically* in many typical applications.

Another disadvantage of the above algorithm is that it fails when the probability of interest is very small, as most of the generated samples will fall out of the region of interest. To resolve this problem, we will introduce a multi-level algorithm. The idea is to construct a sequence of reference parameters corresponding to a sequence of sets on the space of f, and iterate both of them until the set coincides with D. For simplicity, we assume that $f(x)$ is a real-valued function and we seek the probability that $f(X)$ is larger than some fixed γ. In this problem-setting, the multi-level algorithm can be outlined as follows [3]:

1. Set the initial distribution parameter $w_0 = \mu$.

2. At step i, choose a threshold of tail probability $\rho \in (0, 1)$ and choose a number γ' as large as possible to satisfy $\Pr(f(X) > \gamma') \geq \rho$ under the density $p(x; w_i)$. If $\gamma' > \gamma$, let $\gamma' = \gamma$.

3. Find the optimal reference parameter v for estimating probability $\Pr(f(X) > \gamma')$.

4. If $\gamma' = \gamma$, return v as the optimal reference parameter; otherwise, let $w_{i+1} = v$, $i = i + 1$ and go to step 2.

In summary, each iteration of the algorithm consists of two phases: in the first phase γ is updated (step 2), and in the second phase v is updated (step 3). We will provide a more detailed algorithm for estimating specific loss distributions in the following sections.

3 Models of portfolio credit risk

In this section we first provide an introduction to portfolio credit risks and copulas, then we will specify the normal copula model and the t-copula model for portfolio credit risks.

Consider a portfolio with N obligors. The default indicator for n^{th} obligor is notated as

$$I_n = \begin{cases} 1, & \text{if the } n^{th} \text{ obligor defaults,} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

To describe dependence among obligors we model dependence among the default indicators $I_1, \cdots, I_N$, via a multivariate vector $(X_1, \ldots, X_N)$ of latent variables. Let X_n be a random variable with continuous distribution function $F_n(x) = \Pr(X_n \leq x)$ and let $\xi_n \in \mathbb{R}$ be such that $I_n = 1$ if and only if $I_n > \xi_n$. Then each default indicator can be represented as

$$I_n = \mathrm{I}\{X_n > \xi_n\}, \quad n = 1, \cdots, N,$$

with ξ_n chosen to match the marginal default probability, p_n (which is assumed to be known from credit ratings or other recourses) , i.e., $\xi_n = F_n^{-1}(1 - p_n)$. It is easy to see that

$$\Pr(I_n = 1) = \Pr(X_n > F_n^{-1}(1 - p_n)) = 1 - F_n(F_n^{-1}(1 - p_n)) = p_n, \quad n = 1, \cdots, N.$$

Denote by c_n the loss resulting from the default of the n^{th} obligor. The total loss of the N obligors from defaults is

$$L = \sum_{n=1}^{N} c_n I_n. \tag{12}$$

Here for simplicity, we assume that each c_n is non-random (the recovery rate is non-random) as in classic credit risk models. As was mentioned in [16], it would suffice to know the distribution of $c_n I_n$ instead of assuming c_n to be constants.

Our goal is to calculate the probability of a large portfolio loss, i.e., to calculate the probability $\Pr(L > \gamma)$ at large values of γ. Note that the correlations among X_n determine the dependence among I_n, and the underlying correlations can be specified through a particular model of copulas. Next we provide a brief introduction of the copulas.

Copulas, as one of the most powerful tools of modeling dependence, are now regarded as a common knowledge in mathematical finance and actuarial science. A copula, or the joint distribution function C of m uniform random variables $U_1, U_2, \cdots, U_m$ is defined as

$$C(u_1, u_2, \cdots, u_m) = \Pr(U_1 \leq u_1, U_2 \leq u_2, \cdots, U_m \leq u_m).$$

Copula functions can link uni-variate marginals to their full multivariate distribution via the relation

$$C(F_1(x_1), F_2(x_2), \cdots, F_n(x_m)) = F(x_1, x_2, \cdots, x_m)$$

for given uni-variate marginal distribution functions $F_1(x_1), F_2(x_2), \cdots, F_n(x_m)$. [28]'s theorem states that any multivariate distribution function F can be written in the form of a copula function, and thus copula functions provide a flexible way to study multivariate distributions. We refer to [19, 25] for more information on copulas in general, and [12] for the use of copulas in credit risk modeling.

3.1 The normal copula

An m-variate normal copula function is defined as

$$C(u_1, \cdots, u_m) = \Phi_\Sigma \left(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_m) \right)$$

where Φ^{-1} is the inverse function of the standard normal cumulative distribution function Φ, and Φ_Σ is the joint cumulative distribution function of a multivariate normal distribution with mean vector 0 and covariance matrix equal to the correlation matrix Σ (which is assumed to be known from credit ratings or other recourses).

The normal copula model was the first widely used copula model in the financial industry; it attempts to capture the dependence among obligors while maintaining mathematical tractability by assuming the vector of latent variables follows a multivariate normal distribution. Although it was criticized for not capturing the tail dependence (and thus it underestimates tail risks), we study the normal copula (NC) model as used in [16, 17, 22] for an illustrative purpose. It was mentioned in [22] that the CreditMetrics model [17] used by J.P. Morgan actually used a bi-variate normal copula function with the asset correlation as the correlation parameter in the copula function, although it does not use the concept of copula function explicitly. More specifically, it is shown that if the asset correlation is normally distributed with a binomial normal probability density function then the joint transformation probability is consistent with the result from using a normal copula function. Li then proposed to model the credit portfolio of any size by constructing high dimensional normal copula functions [22].

In the NC model for portfolio credit risk, the underlying correlations are specified through a factor model of the form

$$X_n = \boldsymbol{\alpha}_n^T \mathbf{Z} + \beta_n r_n \tag{13}$$

in which $\mathbf{Z} = (Z_1, \ldots, Z_K)$ is the column vector of *systematic risk* factor, each of $Z_k (k = 1, \ldots, K)$ having an standard normal distribution, $r_n \sim \mathcal{N}(0, 1)$ is the *idiosyncratic risk* associated with the nth obligor, $\boldsymbol{\alpha}_n^T = (\alpha_{n1}, \ldots, \alpha_{nK})$ is the row vector of *factor loadings* for the nth obligor with $\sum_{k=1}^K \alpha_{nk}^2 \leq 1$, and β_n is a constant determined by $\mathrm{Var}(X_n)$. The conditional default probability for the nth obligor given the factor loading $\mathbf{Z}$ can be written as

$$\begin{aligned} p_n(\mathbf{Z}) &= \Pr(I_n = 1 | \mathbf{Z}) = \Pr(X_n > \xi_n | \mathbf{Z}) \\ &= \Pr(\boldsymbol{\alpha}_n^T \mathbf{Z} + \beta_n r_n > \Phi^{-1}(1 - p_n) | \mathbf{Z}) = \Phi \left(\frac{\boldsymbol{\alpha}_n^T \mathbf{Z} + \Phi^{-1}(p_n)}{\beta_n} \right), \end{aligned} \tag{14}$$

where Φ denotes the standard normal distribution function. The factor loadings $\alpha_{n,k}$ $(k = 1, \cdots, K)$ are assumed to be nonnegative, to ensure all default indicators to be positively correlated so that larger values of the factors Θ_k lead to a larger number of defaults.

Note that for each X_n to have a standard normal distribution, it is required that $\sum_{k=1}^K \alpha_{n,k}^2 + \beta_n^2 = 1$ for all $n = 1, 2, \cdots, N$ to ensure the variance of X_n is equal to

1. From the model setup, we can see that $\mathbf{X}$ has a multivariate normal distribution, with covariance matrix Σ where the (i, j)-element of Σ is $\alpha_i\alpha_j^T$. Thus, since I_n is a non-decreasing function of X_n, the N-normal copula C with a correlation matrix Σ is a copula of $(I_1, \cdots, I_N)$ (although since $(I_1, \cdots, I_N)$ is discrete, its copula is not unique). That is why this model is called a normal copula model.

3.2 The *t*-copula

One of the potential problems of the normal copula model is that the event of many simultaneous defaults may be assigned to a probability which is too small. The t-copula has been increasingly popular in finance since it captures the tail-dependence, is not exchangeable (as opposed to Archimedean copulas), and in the meanwhile is still simple enough. In view of this, the t-copula model was introduced for credit risk models by assuming the underlying latent variables $\mathbf{X}$ follow a multivariate t distribution; see e.g. [2].

Let $\mathbf{Y} \sim \mathcal{N}_m(0, \Sigma)$, $S \sim \chi^2_\nu$ (a chi-square distribution with ν degrees of freedom), and $R = \sqrt{\nu/S}$. When $\mathbf{Y}$ and R are independent, the $\mathbb{R}^m$-valued random vector given by $\mathbf{X} = R\mathbf{Y}$ has a centered t-distribution with ν degrees of freedom. Note that for $\nu > 2$, $\mathrm{Cov}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$. By Sklar's theorem, the copula of $\mathbf{X}$ can be written as

$$C^t_{\nu,\boldsymbol{\rho}}(u_1, \ldots, u_m) = t^m_{\nu,\boldsymbol{\rho}}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_m))$$

where $t^m_{\nu,\boldsymbol{\rho}}$ denotes the multivariate t-distribution function with parameters $(\nu, \boldsymbol{\rho})$, $\rho_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$, and t_ν is the standard univariate t-distribution function with ν degrees of freedom.

As in the normal copula model, denote by $\alpha_{n1}, \ldots, \alpha_{nK}$ the factor loadings for the nth obligor and let each of $Z_k\,(k = 1, \ldots, K)$ and r_n have standard normal distribution. A general $K-$factor model was considered in [8] where the factors and idiosyncratic risks are modeled as independent t random variable. This is done by introducing independent shock variables $\Theta_k^2 \sim \mathrm{Gamma}(\nu/2, \nu/2)$, $(k = 1, \ldots, K)$ and $\Lambda_n^2 \sim \mathrm{Gamma}(\nu/2, \nu/2)$, $(n = 1, \ldots, N)$ for some $\nu > 0$. Define the model

$$X_n = \alpha_{n1}Z_1\Theta_1^{-1} + \cdots + \alpha_{nK}Z_K\Theta_K^{-1} + \beta_n r_n \Lambda_n^{-1}, \quad n = 1, \ldots, N, \tag{15}$$

then marginally $\mathbf{X} = (X_1, \ldots, X_N)$ follows a multivariate t distribution with degree of freedom ν (see e.g. [7]).

Similar to the NC model, it is required that $\sum_{k=1}^K \alpha_{n,k}^2 + \beta_n^2 = 1$ for all $n = 1, 2, \cdots, N$ to ensure the variance of X_n is equal to 1. The conditional default probability for the nth obligor given the factor loadings $\mathbf{Z} = (Z_1, \ldots, Z_K)$ can be calculated to be

$$\begin{aligned} p_n(\mathbf{Z}) &= \Pr\left(\alpha_{n1}Z_1\Theta_1^{-1} + \cdots + \alpha_{nK}Z_K\Theta_K^{-1} + \beta_n r_n\Lambda_n^{-1} > t_\nu^{-1}(1 - p_n)|\mathbf{Z}\right) \\ &= \Phi\left(\frac{\left(\sum_{k=1}^K \alpha_{nk}Z_k\Theta_k^{-1} + t_\nu^{-1}(p_n)\right)\Lambda_n}{\beta_n}\right). \end{aligned} \tag{16}$$

4 Cross-entropy method for the copula models

The main purpose of this study is to calculate the probability of a large portfolio loss, i.e., to calculate the probability $\Pr(L > \gamma)$, at large values of γ. The loss distribution of L under most credit risk models is often complicated and can not be solved explicitly. Numerical calculation such as Monte-Carlo simulation methods is needed in general. However, basic Monte-Carlo simulation can be very inefficient for such rare events and alternative algorithms are sought to estimate the small probability for large losses. For example, Glasserman *et. al.* provided a two-step IS method for the NC model of portfolio credit risk, where the biasing PDF is obtained via an analysis of the NC model involving a numerical optimization in each replication [16] and Chan and Kroese derived algorithms base on conditional Monte Carlo to estimate the probability that the portfolio incurs large losses under $t-$copula model [7].

The NC model has been studied in [20] by using MCMC (Markov chain Monte Carlo) methods, and related work was done by Chan in [7]. In this work we propose a general CE-MIS (cross-entropy-multiple importance sampling) method when the sought loss probability is relatively small. More specifically, we will draw samples according to multiple biasing PDF's resulted from each iteration of CE and combine their results by the technique of MIS [31, 32]. Note that MIS can be conveniently combined with the multi-stage CE algorithm. In fact, instead of using only the optimal distribution found with the final level, we can save distributions obtained at all levels, draw samples according to them and then combine the results by the end. Thus the samples are encouraged toward different levels of loss and one can calculate the complete loss distribution.

Let $\mathbf{x}$ be the $(K+N)$-dimensional i.i.d. random variable

$$\mathbf{x} = (z_1, z_2, ..., z_K, r_1, r_2, ..., r_N) = (x_1, x_2, \cdots, x_{K+N}),$$

each component following a standard normal distribution $N(0,1)$. The total loss L can then be regarded as a function of $\mathbf{x}$, and we are seeking the probability $Q = \Pr[L(\mathbf{x}) > \gamma]$, which can be calculated by a MC estimator:

$$\hat{Q} = \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))\,, \tag{17}$$

Applying the technique of IS to Eq. (17) one obtains

$$\hat{Q}^* = \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))R(\mathbf{x}^{(m)})\,, \tag{18}$$

where samples are drawn according to the biasing PDF $p^*(x)$.

Among various methods to construct the biasing PDF $p^*(x)$, the mean translation (MT) method [29] is one of the most direct ones for normal distributions. The basic idea of MT for this problem is to translate the mean of $\mathbf{x}$ from origin to a point determined *aprior* by the CE method. Specifically, the biasing PDF can be written as

$$p^*(x) = \frac{1}{(\sqrt{2\pi})^{N+K}} \prod_{i=1}^{N+K} \exp\left(-\frac{(x_i - x_i^*)^2}{2}\right), \tag{19}$$

where $\mathbf{x}^* = (x_1^*, ..., x_{N+K}^*) = (z_1^*, \cdots, z_K^*, r_1^*, \cdots, r_N^*)$ is the mean of $\mathbf{x}$ (and hence can be regarded as the reference parameter v in the CE framework). To apply the CE method to calculate the mean of $\mathbf{x}$, plug Eq (19) into Eq. (10) and solve Eq. (10) analytically for the optimal reference parameters:

$$x_i^* = \frac{\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))R(\mathbf{x}^{(m)})x_i^{(m)}}{\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))R(\mathbf{x}^{(m)})}, \tag{20}$$

for $i = 1, ..., N + K$.

However, the effectiveness of this strategy relies on the dimension of the state space, $N + K$. When $N + K$ is relatively large, the calculation may be beyond the capacity of the CE method. In practice, the number of common risk factors is usually not very large, thus the mean of $\mathbf{z} = (z_1, \cdots, z_K)$ can still be estimated by Eq. (20). For a large N, we apply an exponential twist [[16]] on $\mathbf{r} = (r_1, \cdots, r_N) = (x_{K+1}, \cdots, x_{K+N})$ by introducing a parameter θ and setting

$$p_{n,\theta} = \frac{p_n e^{\theta c_n}}{1 + p_n(e^{\theta c_n} - 1)}, \tag{21}$$

where p_n is the original default probability. The mean of each r_n can be calculated by

$$r_n^* = \xi_n - F_n^{-1}(1 - p_{n,\theta}), \tag{22}$$

and it follows immediately that

$$\sum_{n=1}^{N} r_n^* = \sum_{n=1}^{N} \left[\xi_n - F_n^{-1}(1 - p_{n,\theta})\right]. \tag{23}$$

Note that the left hand side of Eq. (23) can be simulated by

$$\sum_{n=1}^{N} r_n^* = \sum_{n=1}^{N} \frac{\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))R(\mathbf{x}^{(m)})r_n^{(m)}}{\sum_{m=1}^{M} \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m)}))R(\mathbf{x}^{(m)})}. \tag{24}$$

Substituting Eq. (24) into Eq. (23) then yields the optimal value of θ, and with θ known, the mean of $\mathbf{r}^* = (r_1^*, \cdots, r_N^*)$ can be obtained from Eq. (22).

In general, to construct loss distributions over a broad range, a single choice of biasing PDF is not enough to capture efficiently every region of sample space that give rise to the event of interest. With MIS we are able to combine multiple biasing PDF's resulted from each iteration of CE, by assigning each biasing PDF a weight $w_j(x)$ determined by balance heuristic [24, 31]:

$$w_j(x) = \frac{M_j p_j(x)}{\sum_{j'=1}^{J} M_{j'} p_{j'}(x)}. \tag{25}$$

The IS estimator becomes

$$\hat{Q} = \sum_{j=1}^{J} \frac{1}{M_j} \sum_{m=1}^{M_j} w_j(\mathbf{x}^{(m,j)}) \mathrm{I}_{(\gamma,\infty)}(L(\mathbf{x}^{(m,j)})) R(\mathbf{x}^{(m,j)}) \,, \tag{26}$$

where $p_1, \cdots, p_J$ are biasing PDF's, $\mathbf{x}^{(1,j)}, \cdots, \mathbf{x}^{(M_j,j)}$ is a sample drawn from $p_j(x)$, and the total number M_j is characterized by its mean x_j^*. In summary, the complete CE-IS algorithm proceeds as follows:

1. Choose a positive number $0 < \rho < 1$ (e.g., = 0.02) and a desired level of loss γ_*.
2. Set $\gamma_o = 0$, $x_1^* = 1$ (unbiased) and $t = 1$ (iteration level counter).
3. Generate M samples $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$, according to distribution $p(x; x_t^*)$. Let $t = t + 1$.
4. Compute the loss $L(\mathbf{x}^{(m)})$ for $m = 1 \dots M$ and sort the results ascendingly as $L_1, \dots, L_M$. Let $\gamma_t = L_{[(1-\rho)M]}$.
5. If $\gamma_t > \gamma_{t-1}$ and $\gamma_t < \gamma$, compute x_t^* from Eq. (20), compute r_t^* using Eqs. (22–24), and reiterate from step 3; otherwise proceed to step 6.
6. Generate samples according to distributions parameters $x_1^*, \dots, x_t^*$ respectively, and combine the results using Eq. (26).

In this section, we represented an CE-MIS method to calculate the probability of large portfolio losses. This method can avoid the non-trivial analysis required by a general IS method and outperforms traditional model-specific estimators when the parametric family in the CE method is chosen to be sufficiently large [6]. In addition, the method we propose is not model-specific, i.e., it can be applied to a large variety of models with little modification. Numerical examples will be provided to demonstrate our method in the next section.

5 Numerical examples

We provide some numerical examples for the CE-MIS method described in Section 4. In particular we will apply our method to the NC model and $t-$copula model described in Section 3. We adopt formulas used in [16] to generate the default probabilities and loss coefficients:

$$p_n = 0.01 \times \left(1 + \sin\left(\frac{16\pi n}{N}\right)\right) , \tag{27a}$$

$$c_n = \left(\frac{5n}{N}\right)^2 , \tag{27b}$$

for $n = 1, ...N$. The coefficients α_n, β_n are randomly generated and then normalized so that X_n follows standard normal distribution for the NC model and standard t distribution for the $t-$copula model.

5.1 A portfolio of 50 obligors with 10 common risk factors

We first consider a portfolio with 50 obligors and 10 common risk factors. In the numeric, it took 4 iterations with $1,000$ replications in each for the CE method to find 2 biasing PDF's. Fig. 1 shows the PDF constructed by using standard MC simulation with samples generated according to the biasing PDF's, and it illustrates the ability to put samples around different loss levels of the obtained PDF's. We can see that samples drawn according to the unbiased PDF are around small losses, while the biased pdfs can push the samples toward larger values of loss.

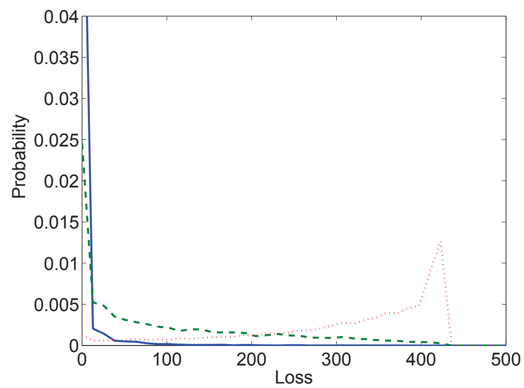

Figure 1: Loss distributions numerically reconstructed from standard MC simulations according to the PDF's of **x** obtained by using the CE method. The solid line represents the unbiased PDF and red and green dashed lines represent biased PDF's obtained from CE.

We then used the obtained PDF's to construct the tail probability $\Pr[L > \gamma]$ for loss level $\gamma \in [0, 450]$, where $2,000$ samples were generated according to each biasing PDF. In Figs. 2 and 3, we plot the tail probability $\Pr(L > \gamma)$ as a function of the loss level γ for both the NC model and $t-$copula model, respectively. We also put the results of standard MC simulations with the same number of samples in Figs. 2 and 3 for comparison purpose. We can see that the CE-MIS method allows us to estimate probabilities accurately to around 10^{-7} with several thousand samples.

5.2 A portfolio of 1000 obligors with 10 common risk factors

In our second example, we consider a situation of more obligors. Specifically, we choose $N = 1000$, again in a 10-factor model. The default probabilities and the loss coefficients are also determined by Eqs. (27), and the coefficients α_n, β_n are randomly generated and renormalized as before. In this example, the CE method finds two biasing PDF's using $3,000$ samples, and we draw 1,000 samples according to each distribution to construct the PDFs of loss in the range from 0 to 8,000 which are shown in Figs. 4 and 5.

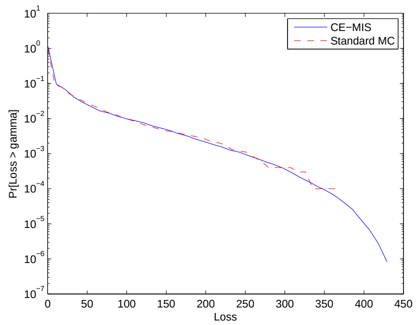

Figure 2: Tail probability plotted against the loss level for a 50-obligor-10-risk factor NC model

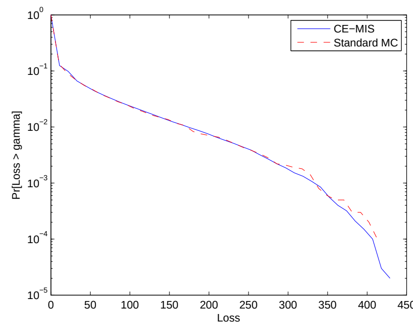

Figure 3: Tail probability plotted against the loss level for a 50-obligor-10-risk factor $t-$copula model

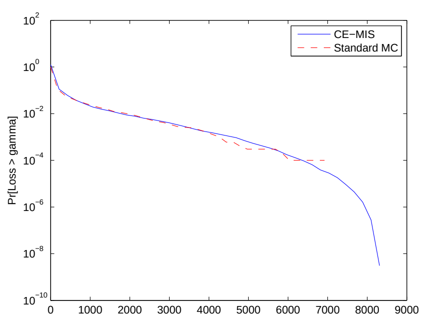

Figure 4: Tail probability plotted against the loss level for a 1000-obligor-10-risk factor NC model

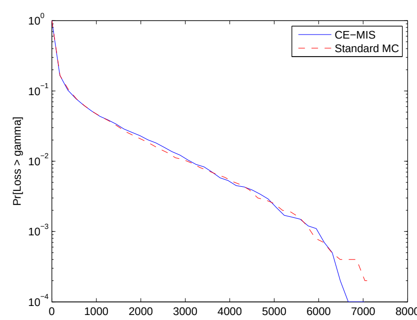

Figure 5: Tail probability plotted against the loss level for a 1000-obligor-10-risk factor $t-$copula model

Though the effectiveness of the CE method has been demonstrated by many applications, no theoretical justification of the optimality of this method is available up to date. Therefore, monitoring the estimator variance in numerical experiments may be the only way to verify the effectiveness of the CE method. A computation of the coefficient of variation (CV) of the MIS estimator for the loss distribution shows that the CE method does reduce the estimator variance significantly, which allows it to estimate very low probabilities with a relatively small number of samples. With this example, we show that the CE method remains effective for problems involving a large number of obligors.

5.3 A portfolio of 1000 obligors with 21 common risk factors

In the third example we consider a portfolio of 1000 obligors with 21 risk factors. The default probabilities and the loss coefficients are also determined by Eqs. (27), and the coefficients α_n, β_n are randomly generated and re-normalized as before. In this example, the CE method finds two biasing PDF's using $3,000$ samples, and we draw 1,000 samples according to each distribution to construct the PDF of loss in the range from 0 to 9,000 which are shown in Figs. 6 and 7.

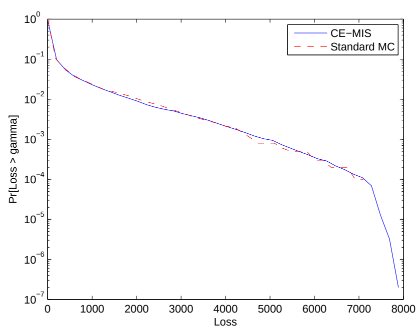

Figure 6: Tail probability plotted against the loss level for a 1000-obligor-21-risk factor NC model

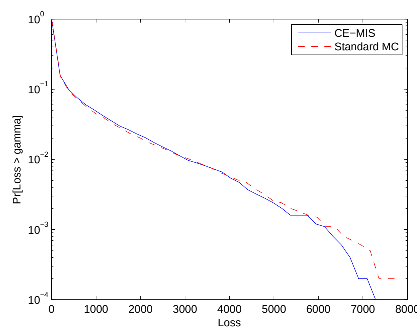

Figure 7: Tail probability plotted against the loss level for a 1000-obligor-21-risk factor $t-$copula model

Although there is no clear evidence that our CE-MIS method is superior than the numerical results in [16] by using IS, our proposed CE-MIS method can combine results from each iteration during multiple IS and obtain a complete and continuous PDF, than runs on a range larger than what standard MC can produce.

6 Conclusions

In this article we propose a new method to simulate rare extreme losses in credit risk management. This method can find the optimal biasing PDF's for both the obligors and the common risk factors simultaneously without requiring a numerical optimization step in every trial as in traditional IS method. Moreover, this method combines a multi-level CE algorithm with the MIS technique which makes it possible to construct complete loss distributions over a broad range. Specifically the CE algorithm is used to find biasing PDF's and MIS is used to estimate the probability of interest. The relative importance of the biasing of the common risk factors and that of the individual risk factors is automatically adjusted by the CE method.

In particular we first describe a general framework of the CE-MIS method, followed by the implementation of the proposed method to portfolio credit risk models. We then illustrate the effectiveness of the CE-MIS method by applying it to a normal copula model and a $t-$copula model.

Remarkably, a two-step IS approach has been proposed in [16], where the biasing PDF's are determined by analyzing the NC model and analytical approximations are required at the first stage. Our method provides an alternative without the requirement of initial analysis to the existing method and demonstrates the convenience of combining the CE-method with MIS to constructing a complete loss PDF. Moreover, our method is generic and can be applied to a large variety of different models with little modification.

References

[1] A. Arvanitis and J. Gregory. *Credit: the complete guide to pricing, hedging and risk management*. Risk Books, London, 2001.

[2] A. Bassamboo, S. Juneja, and A. Zeevi. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, 56:593–606, 2008.

[3] P.T. De Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein. A tutorial on the cross-entropy method. *Ann. Oper. Res.*, 134:19–67, 2005.

[4] P.T. De Boer, D.P. Kroese, and R.Y. Rubinstein. A fast cross-entropy method for estimating buffer overflows in queueing networks. *Manag. Sci.*, 50:883–895, 2004.

[5] J. A. Bucklew. *Introduction to rare event simulation*. Springer-Verlag, New York, 2004.

[6] J. Chan and D. Kroese. Improved cross-entropy method for estimation. *Statistics and Computing*, 22:1031–1040, 2012.

[7] J. C. C. Chan and D. P. Kroese. Efficient estimation of large portfolio loss probablities in *t*-copula models. *European Journal of Operational Research*, 205:361–367, 2010.

[8] J. C. C. Chan and D. P. Kroese. Rafe-event probability estimation with conditional monte carlo. *Annals of Operations Research*, 189:43–61, 2010.

[9] S. T. Charles. *Risk and financial management: mathematical and computational methods*. John Wiley & Son, UK, 2005.

[10] S. Felix. Recipe for disaster: the formula that killed wall street. *Wired Magazine*, 17(3), 2009.

[11] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas: applicability and limitations. *Statistics and Probability Letters*, 63(3):275–286, 2003.

[12] R. Frey, A. McNeil, and M. Nyfeler. Copulas and credit models. *Risk*, 10:111–114, 2001.

[13] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2006.

[14] P. Glasserman, W.M. Kang, and P. Shahabuddin. Fast simulation of multifactor portfolio credit risk. *Oper. Res.*, 56:1200–1217, 2008.

[15] P. Glasserman and J. Li. Importance sampling for a mixed poisson model of portfolio credit risk. *Proc. Win. Sim. Conf.*, pages 267–275, 2003.

[16] P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Manag. Sci.*, 51:1643–1656, 2005.

[17] G.M. Gupton, C.C. Finger, and M. Bhatia. CreditMetrics. Technical report, J.P. Morgan & Co., 1997.

[18] T. Homem-de-Mello. A study on the cross-entropy method for rare event probability estimation. *INFORMS J. Comp.*, 19:381–394, 2007.

[19] J. Joe. *Multivariate models and dependence concepts*. Chapman & Hall, London, 1997.

[20] D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2011.

[21] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[22] D. Li. On default correlation: A copula function approach. *J. Fixed Income*, 9:43–54, 2000.

[23] J. Li, G. Biondini, W.L. Kath, and H. Kogelnik. Anisotropic hinge model for polarization-mode dispersion in installed fibers. *Opt. Lett.*, 33:1924–1926, 2008.

[24] R.O. Moore, G.Biondini, and W.L. Kath. A method to compute statistics of large, noise-induced perturbations of nonlinear Schrödinger solitons. *SIAM J. Appl. Math.*, 67:1418–1439, 2007.

[25] R. Nelsen. *An introduction to copulas, 2nd ed.* Springer, New York, 2006.

[26] R.Y. Rubinstein. The simulated entropy method for combinatorial and continuous optimization. *Math. Comp. Appl. Prob.*, 1:127–190, 1999.

[27] R.Y. Rubinstein and D.P. Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Springer-Verlag, New York, 2004.

[28] A. Sklar. Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[29] R. Srinivasan. *Importance sampling: applications in communications and detection*. Springer, New York, 2002.

[30] S. T. Tokdar and R. E. Kass. Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2:54–60, 2010.

[31] E. Veach. *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, 1997.

[32] E. Veach and L. J. Guibas. Optimally combining sampling techniques for Monte Carlo rendering. *SIGGRAPH Proc.*, pages 419–428, 1995.

[33] T. Wilde. *CreditRisk+: a creditrisk management framework*. Credit Suisse First, London, UK, 1997.