Importance Sampling for Estimating Risk Measures in Portfolio Credit Risk Models

Zhao Li

October 2009

Abstract

This paper is the report of a Master's Degree project carried out at Royal Institute of Technology and in this paper we mainly apply the estimators and methods derived by P. Glasserman and J. Li (2003, 2005) of importance sampling methods in portfolio credit risk models. By using the exponential twisting method we will be able to compute the probability beyond one certain loss level (P(L>X)). We use the search method and a 'direct' method derived by Peter W. Glynn to estimate the Value-at-Risk (VaR) from the probability and Expected Shortfall (ES) in two portfolio credit risk models, and estimate a convex risk measure Shortfall Risk (SR) with the estimator given by J. Dunkel and S. Weber (2007) in the two models as well. We provide numerical simulation to show the good performance of importance sampling comparing with the plain Monte Carlo.

Table of Contents

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisor, Associate Professor Henrik Hult at Royal Institute of Technology, Department of Mathematics for his guidance and hospitality. His wide knowledge and abstract way of thinking together with his encouraging and constructive comments have been of greatest value for me in order to complete the present thesis.

I wish to express my warm and sincere thanks to graduate student Jens Svensson for his kindly help with the issues regarding the implementation of two-step importance sampling in Normal Copula Model.

I owe my loving thanks to my family, my girlfriend and my ace buddies, without their encouragement and understanding it would have been impossible for me to finish this thesis.

Zhao Li

October 2009, Stockholm

Chapter 1
Introduction

1.1 Background

Credit risk can be defined as the risk of loss due to a debtor's non-payment of a loan or other line of credit. The goal of credit risk management is to maximize an institution's risk-adjusted rate of return by maintaining credit risk exposure within acceptable limits. Financial institutions need to manage the credit risk inherent in the entire portfolio as well as the risk in individual credits or transactions. The effective management of credit risk is a critical component of a comprehensive approach to risk management and essential to the long-term success of any financial organization. A firm measurement of credit default risk is one of the key issues for financial institutions, and it's the driving force leading the financial industry to develop new models to measure and manage this risk. An important feature of modern credit risk management models is that it tries to capture the effect of dependence across sources of credit risk to which a bank or financial institution is exposed. These models are now in widespread use, e.g. CreditMetrics (Gupton et al.,1997), originally developed by JP Morgan, and CreditRisk+ (Cre,1997), Credit Suisse Financial Products, which take into account dependence structure between obligors. Capturing dependence adds complexity both to the models used and to the computational methods required to calculate outputs of a model.

Monte Carlo simulation is widely used in financial institutions and is easy to implement on a computer. The Monte Carlo method relies on repeated replication scenarios that determine which obligors default and the losses given default. For trusty obligors, default is a rare event associated. The computation cost may become large for the rare-event simulation with complex

dependence between obligors. Hence, importance sampling, a variance reduction technique could improve the simulation process efficiency.

During the past decade, P. Glasserman and J. Li had made intense efforts to develop a general approach of importance sampling (IS) based on applying a change of the distribution to the factors and a change of distribution to the default indicators conditional on the factors, see P. Glasserman and J. Li (2003,2005). We are going to combine this approach with numerical method to estimate the industry standard of risk measurement, Value-at-Risk (VaR). Due to the inherent deficiencies of VaR, we are going to estimate two additional risk measures, Expected Shortfall (ES) and utility-based Shortfall Risk (SR).

1.2 Outline

The paper is organized as follows. In the Chapter 2 we give definition and properties of risk measures and credit risk models. Chapter 3 is dedicated to show how risk measures are estimated by plain MC and importance sampling method in Normal Copulas Model. Chapter 4 introduced the estimation of risk measures in Mixed Poisson Model. We give the numerical simulations in Chapter 5. The conclusion is given in Chapter 6.

Chapter 2
Risk Measures and Portfolio Credit Risk Models

2.1 Risk Measures

Normal distributions are widely used in the traditional tools for assessing and optimizing portfolio risk. In that case two statistical quantities, the mean and the standard deviation, could be used to estimate the return and risk. However, often, the distributions of losses are far from normal; for instance, heavily tailed, skewed, high kurtosis. A measure of risk is needed to compare the riskiness of different portfolios. A scalar value is important for the sake of risk comparison.

Here we give definitions of three scalar risk measures, Value-at-risk, Expected-shortfall and Shortfall-Risk.

2.1.1 Value-at-Risk

Value-at-Risk is by far the most popular and most accepted risk measure among financial institution. But it suffers from two severe deficiencies if considered as a measure of downside risk:

(i) VaR penalizes diversification which means it is not subadditive

(ii) VaR is insensitive to the size of loss beyond the pre-specified threshold level

First we recall the definition of VaR. We denote L as the potential loss of a credit portfolio over a fixed time horizon T. Assuming that L is random variable on some probability space (Ω, F, P). The $\mathrm{VaR}_\alpha(\mathrm{L})$ is defined by the smallest number l such that the probability that the loss L exceeds l is no larger than 1-α, i.e.

$$\mathrm{VaR}_\alpha(\mathrm{L}) \coloneqq \inf(l \in \mathrm{R}: \mathrm{P}(\mathrm{L} > l) \leq 1 - \alpha) = \inf(l \in \mathrm{R}: \mathrm{E}[\mathrm{I}_{\mathrm{L}>l}] \leq 1 - \alpha) \qquad (2.1)$$

Here, E denote the expected value with respect to the probability measure P, and $I_{L>l}$ is the indicator function of the event $\{L>l\}$. Thus, VaR corresponds to the quantile of the losses at level α. Typical value for α which are used in practice are α = 0.95 or α = 0.99, but higher value may also be of interest.

As we told VaR suffers from two drawbacks. Firstly, VaR is not a subadditive measures, it means VaR does not assess portfolio diversification as being beneficial which violates common sense. Secondly, VaR does not take into account the size of very large losses that might occur over the certain threshold. This defect can be illustrated by a simple example. Consider two portfolios with loss L_1 and L_2 respectively, where L_1 equals to -1 with probability 99% and +1 with probability 1%, which means portfolio 1 has 99% probability to earn 1 unit of money and 1% to loss 1 unit of money, and L_2 equals to -1 with 99% probability and +1000 with probability 1%, which means portfolio 2 has 99% probability to earn 1 unit of money and 1% to loss 1000 unit of money. It is easy to find out that $\mathrm{VaR}_{0.99}(L_1) = \mathrm{VaR}_{0.99}(L_2) = -1$. Hence, according to typical 95% or 99% VaR, both portfolios are equal risk. However, we could easily see the portfolio 1 is preferable.

Although VaR has these severe drawbacks, it is still the most widely used risk measures, and efficient method to compute VaR is meaningful in the industry.

2.1.2 Expected-Shortfall (ES)

As VaR has two serious limitations, we also consider an alternative risk measure called Expected-Shortfall (ES). This measure is also known as Mean Excess Loss, Conditional Value-at-Risk (CVaR) or Tail VaR. However, for discrete distributions, Expected-Shortfall may differ from CVaR. By definition, the ES is the expected loss exceeding VaR at α level, i.e.

$$\mathrm{ES}_\alpha(L) := \mathrm{E}[L|L > \mathrm{VaR}_\alpha(L)]$$

$$= \frac{1}{1-\alpha}\int_p^1 \mathrm{VaR}_p\,(L)\mathrm{d}p \qquad (2.2)$$

We could easily show several reasons why ES is preferred risk measure to VaR. Firstly, ES is a sub-additive measure which VaR is not. Secondly, ES provides information about the amount of loss exceed VaR which acts as an upper bound of VaR. Therefore, portfolio with a low ES

should also have a low VaR. Thirdly, under general conditions, ES is a convex function and it is a coherent measure of risk as well (Föllmer and Schied, 2008).

2.1.3 Shortfall-Risk (SR)

Another alternative to VaR is provided by the convex risk measure called Utility-based Shortfall Risk (SR). The definition is as follow: take a convex loss function f: **R**→**R**, and let λ be a point in the interior of the range of f. Assuming the expectation of f(L) is well defined and finite, we define SR with function f at level λ as

$$\mathrm{SR}_{f,\lambda}(\mathrm{L}) \coloneqq \inf\{\mathrm{s} \in \mathrm{R} | \mathrm{E}[f(\mathrm{L}-\mathrm{s})] \leq \lambda\}, \quad \lambda > 0; \tag{2.3}$$

We will use typical convex loss functions piecewise polynomial and exponential functions, i.e.

$$f_{\gamma}^{\mathrm{poly}}(\mathrm{x}) = \gamma^{-1}\mathrm{x}^{\gamma}\mathbf{1}_{\{\mathrm{x}>0\}}, \quad \gamma > 1; \tag{2.4a}$$

$$f_{\beta}^{\exp}(\mathrm{x}) = \exp(\beta \mathrm{x}), \quad \beta > 0; \tag{2.4b}$$

We see that the SR definition (2.3) is obtained by replacing the indicator function in VaR definition (2.1) with the convex loss function f, which makes SR be affected by large losses, as the loss L exceeds a certain threshold s with a probability of at least λ. Hence SR might take into account the risk of unexpected large losses, which may ignored by VaR.

Why we say SR is utility-based? That's because SR is related to the von Neumann-Morgenstern theory of expected utility. If we set u(x) = -f(-x), we can get a concave Bernoulli utility function u, representing the central object in the von Neumann-Morgenstern theory (Föllmer and Schied, 2004). Defining the utility function U(X) := E[u(X)] where X = -L, we can rewrite the (2.3) into

$$\mathrm{SR}_{f,\lambda}(\mathrm{L}) \coloneqq \inf\{\mathrm{s} \in \mathrm{R} | \mathrm{U}(-\mathrm{L}+\mathrm{s}) \geq -\lambda\}, \quad \lambda > 0;$$

We could calculate the SR within the two steps (Dunkel and Weber, 2007) as follows:

(i) SR equals to the solution s* of $\mathrm{E}[f(\mathrm{L}-\mathrm{s})] = \lambda$. Employ a recursive procedure in order to obtain a sequence of $\mathrm{s_k}$ that convergence to s*. To generate this sequence we need to have the knowledge of $\mathrm{E}[f(\mathrm{L}-\mathrm{s})]$

(ii) Given a model or a certain statistics of Ls, have a initial guess for s*, calculate $\mathrm{E}[f(\mathrm{L}-\mathrm{s})]$. For this purpose we need to use the MC method to estimate the expectation.

We could use this recursive method to generate the sequence as

$$s_{k+1} = \frac{1}{2}\left[s_k + s_{k-1} - (s_k - s_{k-1})\frac{E[f(L-s_k)]+E[f(L-s_{k-1})]-2\lambda}{E[f(L-s_k)]-E[f(L-s_{k-1})]}\right]; \quad (2.5)$$

where $s_k \to s^*, \text{as } k \to \infty$.

2.2 Credit Risk Models

We consider a portfolio with m obligors over a fixed time horizon T (e.g. one year); Let Y_i denote the default indicator (or counter) of the ith obligor. When $Y_i = 1$, it means this obligor defaults within the horizon and $Y_i = 0$ otherwise. The net loss associated with the default of the ith obligor is given by c_i which is positive constant, $c_i > 0$. In some models c_is are modeled as random variables, but here we take the simple approach. The portfolio loss over the horizon T is

$$L = \sum_{i=1}^{m} c_i * Y_i \quad (2.6)$$

The marginal default probabilities $p_i = P(Y_i = 1)$ may be obtained from published credit ratings (e.g. S&P). Different models have different mechanisms in capturing dependence among c_is. In the following sections we give a brief description of two models.

2.2.1 Normal Copula Model

In the normal copula model (NCM), the dependence is modeled through a multivariate normal vector $(X_1, \dots, X_m)$ of latent variables. The threshold x_i is determined such that each latent variable is chosen to match the marginal default probability p_i. Each default indicator is represented as follow:

$$Y_i = I\{X_i > x_i\}, \ i = 1, \dots, m \quad (2.7)$$

and $x_i = \Phi^{-1}(1 - p_i)$. Thus,

$$P(Y_i = 1) = P(X_i > x_i) = 1 - \Phi\left(\Phi^{-1}(1 - p_i)\right) = p_i, \quad i = 1, \dots, m$$

Through the construction, the dependence among Y_is is determined by the correlation among the X_is. The underlying correlations of X_is are often specified through a factor model of the form:

$$X_i = \sum_{k=1}^{d} A_{ik}Z_k + A_{i0}\varepsilon_i, \quad i = 1, \dots, m \quad (2.8)$$

$$\sum_{k=0}^{d} A_{ik}^2 = 1, \quad A_{i0} > 0, A_{ik} \geq 0 \tag{2.9}$$

$Z_1, \dots, Z_d$ are the systematic risk variables with independent standard normal distribution. And the idiosyncratic risks variables $\varepsilon_1, \dots, \varepsilon_d$ are chosen as standard normal as well and independent from the systematic risk variables. As we assume the factors A_{ik} are nonnegative. This condition simplifies our discussion by ensuring that larger values of the factors Z_ks lead to a larger number of defaults. Nonnegative of the A_{ik} is often imposed in practice as a conservative assumption ensuring that all the default indicators are positively correlated. The constraint (2.9) ensures that X_is are standard normal distribution.

Conditionally on the common factors $Z = (Z_1, \dots, Z_d)$, the default indicators Y_is are independent. It is easy to prove that conditioned on the systematic vector Z, the default event $\{Y_i = 1\}$ occurs with the probability:

$$\begin{aligned} p_i(Z) &:= P(Y_i = 1|Z) = P(X_i > x_i|Z) \\ &= P\left(\sum_{k=1}^{d} A_{ik}Z_k + A_{i0}\varepsilon_i > \Phi^{-1}(1-p_i)\middle| Z\right) = \Phi\left(\frac{\sum_{k=1}^{d} A_{ik}Z_k + \Phi^{-1}(p_i)}{A_{i0}}\right) \end{aligned} \tag{2.10}$$

2.2.2 Mixed Poisson Model

In CSFP's (1997) CreditRisk+ model, an alternative way to introduce the dependence is provided by using a mixed Poisson model. This implies that each default counter Y_i is conditionally Poisson distributed. This may be viewed as a Poisson approximation to a Bernoulli random variable (based on the fact that a Poisson random variable with a very small mean that has a very small probability of taking value larger than 1); or it can be viewed as the loss L will in general not be bounded anymore, each i represents a class of obligors, and all obligors from class i cause the same potential net losses c_i. In this interpretation, values of Y_i greater than 1 are meaningful.

The distributions of counter variables Y_is are specified as follows. Given some random vectors $Y = (Y_1, \dots, Y_m)$, the variables Y_is are independent and conditionally Poisson distributed:

$$P(Y_i = k|Z) = \frac{X_i^k}{k!} e^{-X_i}, \qquad k \in N_0, i = 1, \dots, m \tag{2.11}$$

We choose independent Gamma-distributed random variables $Z = (Z_1, \dots, Z_d)$ as the common risk factors in this model. Additionally it is assumed that the random vector X is specified through a factor model as follows:

$$X_i = A_{i0} + \sum_{k=1}^{d} A_{ik} Z_k, \qquad i = 1, \dots, m \tag{2.12}$$

with the constraints:

$$A_{i0} \geq 0, A_{ik} \geq 0, \sum_{i=1}^{m} A_{ik} = 1, \sum_{k=1}^{d} A_{ik} > 0 \quad i = 1, \dots, m \; k = 1, \dots, d$$

For the risk factor vector Z, the probability density function is as below:

$$f(z) = \prod_{k=1}^{d} f_k(z_k), \quad f_k(z_k) = \frac{z_k^{\alpha_k - 1}}{\beta_k^{\alpha_k} \Gamma(\alpha_k)} \exp\left(-\frac{z_k}{\beta_k}\right), \quad z_k > 0$$

We impose the normalization:

$$\alpha_k = \frac{1}{\sigma_k^2}, \qquad \beta_k = \sigma_k^2, \qquad k = 1, \dots, d$$

Then the risk factors $Z_1, \dots, Z_d$ have the mean 1 and variance $\sigma_1^2, \dots, \sigma_d^2$. With these constraints, we have:

$$p_i := E(X_i) = A_{i0} + \sum_{k=1}^{d} A_{ik}$$

Then we can evaluate the portfolio loss by generating Y_i from Poisson(X_i).

2.3 Monte Carlo Method

The Monte Carlo method has been widely used in several industries and academic research. It's the method which solves a problem by generating suitable random numbers and observing the fraction of the numbers obeying some property or properties. The method is useful for obtaining

numerical solutions to problems which are too complicated to solve analytically. It was named by S. Ulam, who in 1946 became the first mathematician to dignify this approach with a name, in honor of a relative having a propensity to gamble (Hoffman 1998, p. 239). Nicolas Metropolis also made important contributions to the development of such methods.

The Monte Carlo Method encompasses any technique of statistical sampling employed to approximate solutions to quantitative problems. Essentially, the Monte Carlo method solves a problem by directly simulating the underlying physical process and then calculating the (average) result of the process. This very general approach is valid in areas such as physics, chemistry, computer science etc. Monte Carlo methods were first introduced to finance in 1964 by David B. Hertz in "Risk Analysis in Capital Investment" (Harvard Business Review), discussing their application in Corporate Finance. In 1977, Phelim Boyle pioneered the use of simulation in derivative valuation in his seminal paper "Options: A Monte Carlo Approach". In finance and mathematical finance, Monte Carlo methods are used to value and analyze complex instruments, portfolios and investments by simulating the various sources of uncertainty affecting their value, and then determining their average value over the range of resultant outcomes. The advantage of Monte Carlo methods over other techniques increases as the dimensions (sources of uncertainty) of the problem increase.

Although Monte Carlo methods provide flexibility, and can handle multiple sources of uncertainty, the use of these techniques is nevertheless not always appropriate. In general, simulation methods are preferred to other valuation techniques only when there are several state variables (i.e. several sources of uncertainty).

We give detailed algorithm of Monte Carlo method in solving the risk measures under assumption of the two credit risk models which we mentioned.

Chapter 3
Estimate risk measures in the Normal Copula Model

As we mentioned in Chapter 2, plain Monte Carlo method is widely used in risk management. Due to high dimensional uncertainty, it would take too much computation time in industry if we want to get a result with high accuracy. We first show the algorithm of calculate the risk measures by using the MC method under the assumption of the Normal Copula Model. Later, we introduce two efficient Monte Carlo method based on importance sampling. Then we would compare the efficiency between importance sampling and the MC method.

3. 1 Plain Monte Carlo method in NCM

In the Normal Copula Model, the indicator Y_i is related to a multivariate normal vector X, according to (2.8) and (2.9). We can implement the MC method as follows:

1. Generate Z~N(**0**,**I**) which is a d-vector of independent normal random variables.
2. Generate ε~N(**0,I**) which is m-vector of independent normal random variable.
3. Calculate $X = (X_1, \dots, X_m)$ via (2.8) and calculate threshold x_i of each latent variable by the function $x_i = \Phi^{-1}(1 - p_i)$.
4. Generate indicator Y_i by $Y_i = 1$ if $X_i > x_i$ and $Y_i = 0$ otherwise
5. Compute the loss by using the (2.6) and repeat the procedure N times. N should be relative large to protect the accuracy.
6. Sort the N losses by descending order, and output as vector L. L_i is the ith largest loss value in the repeated MC simulations.

We could get the α-VaR and α-ES estimator according to empirical function:

$$\widehat{\mathrm{VaR}}_{\alpha}(\mathrm{L}) = \mathrm{L}_{[\mathrm{N}(1-\alpha)]+1}$$

$$\widehat{\mathrm{ES}}_{\alpha}(\mathrm{L}) = \frac{1}{[\mathrm{N}(1-\alpha)]+1} \sum_{\mathrm{k}=1}^{[\mathrm{N}(1-\alpha)]+1} \mathrm{L}_{\mathrm{k}}$$

From these two empirical functions, we can tell that an efficient Monte Carlo method should also get the VaR and ES estimators by using same simulation output N-vector L, otherwise the method would not be considered as efficient. So we simulate VaR and ES together in the future.

We could get SR by the recursive procedure given by (2.5), it's easy to tell that the efficiency and accuracy of estimating $\mathrm{E}[f(\mathrm{L}-\mathrm{c})]$ would have great effect on the result of SR. By inserting the (2.4a) and (2.4b), we are interested in estimation of

$$p(c) = \mathrm{E}\left[\gamma^{-1}(\mathrm{L}-\mathrm{c})^{\gamma}\mathbf{1}_{\{(\mathrm{L}-\mathrm{c})>0\}}\right]$$

and

$$e(c) = \mathrm{E}[\exp(\beta(\mathrm{L}-\mathrm{c}))]$$

Here is the MC algorithm for these two functions:

1. Generate N losses by descending order in one vector L (same method as for VaR and ES)
2. For p(c) we have the estimator:

$$\hat{p}(c) = \frac{1}{\mathrm{n}}\sum_{\mathrm{k}=1}^{\mathrm{n}} \gamma^{-1}(\mathrm{L}_{\mathrm{k}}-\mathrm{c})^{\gamma}$$

Where L_{n} is the minimal loss in vector L which is larger than c.

And for e(c) we have the estimator:

$$\hat{e}(c) = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}} \exp(\beta(\mathrm{L}_{\mathrm{k}}-\mathrm{c}))$$

3.2 Two-step importance sampling for estimate risk measures in NCM

3.2.1 Search Method based on two-step IS for VaR and ES

Importance sampling is a variance reduction technique that can be used in the Monte Carlo method. The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these "important" values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in importance sampling is to choose a distribution

which "encourages" the important values. This use of "biased" distributions will result in a biased estimator if it is applied directly in the simulation. However, the simulation outputs are weighted to correct for the use of the biased distribution, and this ensures that the new importance sampling estimator is unbiased. The weight is given by the likelihood ratio, that is, the Radon-Nikodym derivative of the true underlying distribution with respect to the biased simulation distribution. We give an example to illustrate the method.

Consider a random variable X on some probability space (Ω, F, P), and *f* is the probability density function of X. Take *h*(•) as some function of random variable X. Then the expectation of *h*(X) can be written as:

$$\theta = \mathrm{E}[h(\mathrm{X})] = \int_{-\infty}^{+\infty} h(\mathrm{x}) f(\mathrm{x}) \mathrm{dx} \tag{3.1}$$

Then we can easily get the MC estimator $\hat{\theta}_{\mathrm{n}}^{\mathrm{MC}}$ as follow:

$$\hat{\theta}_{\mathrm{n}}^{\mathrm{MC}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} h(\mathrm{X_i})$$

where $\mathrm{X_i}$s are generated from density *f* independently. Now we consider a second probability density *g* and define the likelihood ratio *r*(x) by *r*(x):= *f*(x)/*g*(x) whenever *g*(x)>0, and *r*(x)=0 otherwise. The integral (3.1) can be written as:

$$\theta = \int_{-\infty}^{+\infty} h(\mathrm{x}) g(\mathrm{x}) r(\mathrm{x}) \mathrm{dx} = \mathrm{E_g}[h(\mathrm{X}) r(\mathrm{X})]$$

where $\mathrm{E_g}$ denote the expectation with respect to density *g*. Then we can have the IS estimator:

$$\hat{\theta}_{\mathrm{n}}^{\mathrm{IS}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} h(\mathrm{X_i})\, r(\mathrm{X_i})$$

Here $\mathrm{X_i}$s are generated from density *g* independently. The density *g* is the biased distribution.

The fundamental issue in implementing importance sampling simulation is the choice of the biased distribution which encourages the important regions of the input variables. Choosing or designing a good biased distribution is the "art" of importance sampling. The rewards for a good distribution can be huge run-time savings; the penalty for a bad distribution can be longer run times than for a plain Monte Carlo simulation without importance sampling.

To estimate Value-at-Risk in NCM, we based on the algorithm of P. Glasserman and J. Li (2005) which used IS method exponential twisting to get the estimator of P(L>x). As the VaR can be taken as a quantile, we use the search method to get the quantile from the probability.

Under a simplified setting of NCM where the obligors are independent, we improve our estimate efficiency of a tail probability P(L>x) with a well established approach. We try to replace each default probability p_i by some other default probability q_i. Here is the unbiased estimate of P(L>x) if the default indicators are sampled using the new default probabilities.

$$P(L > x) = E_{q_i}\left[I\{L > x\}\prod_{i=1}^{m}\left(\frac{p_i}{q_i}\right)^{Y_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-Y_i}\right]$$

where the product inside the expectation taken new probabilities is the likelihood ratio relating the original distribution to new one.

3.2.1.1 Exponential twisting to estimate probability

We choose a parameter θ and set:

$$q_i = p_{i,\theta} = \frac{p_i(Z)e^{\theta c_i}}{1 + p_i(Z)(e^{\theta c_i} - 1)}$$

If $\theta > 0$, this does increase the default probabilities; a larger exposure c_i result in a greater increase in the default probability. If $\theta = 0$, then it comes back to the original probabilities.

With the choice of probabilities, straightforward calculation of likelihood ratio simplifies to:

$$\prod_{i=1}^{m}\left(\frac{p_i}{q_i}\right)^{Y_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-Y_i} = \exp\left(-\theta L + \psi(\theta)\right) \tag{3.2}$$

where

$$\psi(\theta) = \log E\left[e^{\theta L}\right] = \sum_{i=1}^{m}\log\left(1 + p_i(Z)\left(e^{\theta c_i} - 1\right)\right) \tag{3.3}$$

is the cumulant generating function (CGF) of L. For any θ, the unbiased estimator of P(L>x) is:

$$\widehat{P}(L > x) = \frac{1}{n}\sum_{j=1}^{n} I\{L_j > x\}\, e^{-\theta L_j + \psi(\theta)}$$

where n is the number of simulation, L_j stands for the total loss at jth simulation.

It remains to discuss how to determine the parameter θ. A good importance sampling density should be, for fixed n, the variance of the IS estimator is considerably smaller than that of the standard Monte Carlo estimator. We can find the upper bound of second moment:

$$M_2(x, \theta) = E_{p_{i,\theta}}\left[I\{L > x\}e^{-2\theta L + 2\psi(\theta)}\right] \le e^{-2\theta x + 2\psi(\theta)} \tag{3.4}$$

where the upper bound holds for all $\theta > 0$. Minimizing the second moment is difficult, but minimizing the upper bound is same as maximize $\theta x - \psi(\theta)$ over $\theta > 0$. As the function ψ is strictly convex and passes through the origin point, the maximum is attained at

$$\theta_x = \begin{cases} \text{unique solution of } \psi'(\theta) = x, & x > \psi'(0) \\ 0, & x \le \psi'(0) \end{cases} \tag{3.5}$$

So we twist by θ_x to estimate P(L>x).

Now we come for applying IS for a more complicated setting, where the obligors are dependent. Due to the obligors are dependent, we need to shift the mean value of distribution of the factor vector Z from $0 \in R^d$ to $\mu = (\mu_1, \dots, \mu_d) \in R^d$. It means we need to do an importance sampling step with respect to Z. We use the arguments from P. Glasserman and J. Li (2005) to our setting. For any estimator $\hat{p}_x$ of P(L>x), there is a decomposition:

$$\mathrm{Var}[\hat{p}_x] = \mathrm{E}\big[\mathrm{Var}[\hat{p}_x|Z]\big] + \mathrm{Var}\big[\mathrm{E}[\hat{p}_x|Z]\big] \tag{3.6}$$

The exponential twisting of the Bernoulli random variables reduces the first contribution of (3.6), and exponential twisting with respect to $\boldsymbol{Z}$ minimize the second contribution of (3.6). Through the tail bound approximation method (P. Glasserman and J. Li, 2005), we can get the mean value μ:

$$\mu = \arg\max_z \left\{ F_x(z) - \frac{1}{2} z^T z \right\} \tag{3.7}$$

where

$$F_x(z) = -\theta_x(z)x + \psi(\theta_x(z), z)$$

This is the logarithm of the likelihood ratio in (3.2) evaluated at L=x.

We can conclude that the importance sampling algorithm procedure for estimating the loss probability in a NCM with dependent obligors is as follow:

1. Take x
2. Generate $Z \sim N(\mu, I)$, a d-vector of independent normal random variables, where μ is the solution of (3.7).
3. Calculate the new conditional default probabilities

$$q_i(\theta_x(Z), Z) = \frac{p_i(Z) e^{\theta_x(Z) c_i}}{1 + p_i(Z)(e^{\theta_x(Z) c_i} - 1)}$$

with $\theta_x(Z)$ given by (3.5) and $p_i(Z)$ given by (2.10).

4. Generate the default indicator $Y_1, \dots, Y_m$ from Bernoulli random numbers Bin(1,p) with the probability $q_i(\theta_x(Z), Z)$.
5. Calculate the loss L from the (2.6) and return the estimator of P(L>x)

$$\frac{1}{n}\sum_{k=1}^{n} I\{L_k > x\}\, e^{-\theta_x(Z)L_k+\psi(\theta_x(Z),Z)} e^{-\mu^T Z+\frac{\mu^T\mu}{2}} \tag{3.8}$$

The factor $e^{-\mu^T Z+\frac{\mu^T\mu}{2}}$ is the likelihood ratio for the change from density of the N(0,I) distribution to that of N(μ,I) distribution.

3.2.1.2 Estimator for VaR

α-VaR is equivalent to α-quantile, which is clear from the definition. From the two-step IS procedure, we can obtain an estimator of P(L>x), as follows, let:

$$f(x) = P(L > x) - (1 - \alpha) \tag{3.9}$$

The α-VaR (α-quantile) is the unique solution x^* of $f(x) = 0$. We could have $x_{-1} = 0$ and $x_0 = \sum c_i$ as the initial guess. We can easily prove that $f(x_{-1}) > 0$ and $f(x_0) > 0$, we can employ a recursive procedure in order to obtain a sequence $x_1, x_2, \dots, x_k$ such that $x_k \to x^*$ when $k \to \infty$. Here $x_1 = 0.5 * (x_{-1} + x_0)$, if $f(x_1) > 0$, then $x_{-1} = x_1$, else if $f(x_1) < 0$, then $x_0 = x_1$and we can have the recursion rule as:

$$x_{k+1} = \frac{1}{2}(x_{-1} + x_0), \quad where\ x_{-1} = \max(x_k | f(x_k) > 0), x_0 = \min(x_k | f(x_k) < 0) \tag{3.10}$$

This is the basic idea of search method. It's an easy iterative method, and it could be efficient if we can set the iteration terminate condition properly. The two-step IS estimator of the $P(L > x)$ is asymptotically optimal (P. Glasserman and J. Li (2005). From the numerical simulations showing later, we can see search method convergent quickly if the terminate condition has been properly set (less than 10 times).

In principle, further improvements are possible, such as we could add stratified sampling during the process (P. Glasserman, P. Heidelberger and P. Shahabuddin (2000a), and other iterative procedure may be more efficient in some occasions.

3.2.1.3 Estimator for ES

We would like to estimate the ES by using the information of the sequence $x_1, x_2, \dots, x_K$ where x_K is a good estimator of VaR and the relative probability $P(L > x_i)$ for each element in

sequence. From figure 1, ES is the graphic area from α to 1 under the curve. It can be easily proved by the definition of ES.

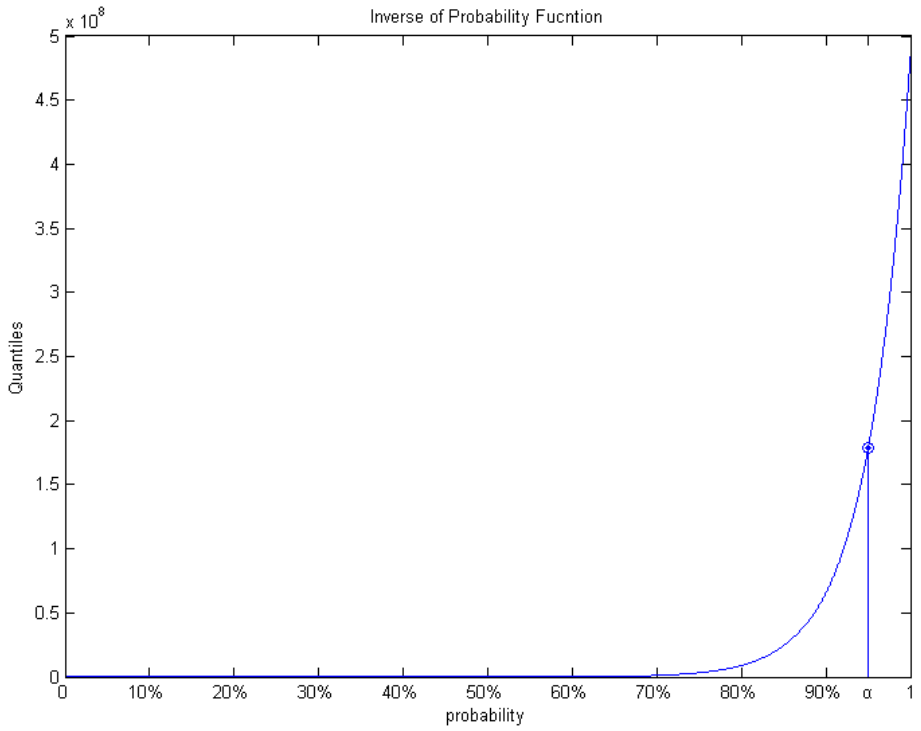

Figure 1 Inverse of Probability function

The point on the graph is the α-quantile or VaR.

We would like to employ the rectangle element approach to estimate the area under the curve from α to 1. So we sort the vector x according to descend order. Refer to the figure 2, it enlarges the graph from α to 1 and shows one way of estimation by rectangle element. Suppose the x_t is the smallest value larger than the VaR. $1 - \mathrm{F}(x_t)$ equals to $P(L > x)$.

We can have the estimator as:

$$\widehat{\mathrm{ES}}_1 = \frac{1}{\alpha}(\mathrm{VaR} \times \Delta \mathrm{p_t} + x_t \times \Delta \mathrm{p_{t-1}} + \cdots + x_1 \times \Delta \mathrm{p_0}) \tag{3.11}$$

where $\Delta \mathrm{p_0} = 1 - P(L > x_1)$, $\Delta \mathrm{p_1} = P(L > x_1) - P(L > x_2), \ldots,$ $\Delta \mathrm{p_t} = P(L > x_t) - \alpha$, and we can have another way of estimation as follow:

$$\widehat{\mathrm{ES}}_2 = \frac{1}{\alpha}\left(x_t \times \Delta \mathrm{p_t} + \cdots + x_1 \times \Delta \mathrm{p_1} + \sum_{i=1}^{m} c_i \times \Delta \mathrm{p_0}\right)$$

where $\Delta \mathrm{p_k}$ as same as in (3.11).

It's easy to see that $\widehat{\mathrm{ES}}_2 > ES > \widehat{\mathrm{ES}}_1$, $\widehat{\mathrm{ES}}_1$ and $\widehat{\mathrm{ES}}_2$ are the lower and upper bound of the estimation of ES by the rectangle element approach. We could take

$$\widehat{\mathrm{ES}} = \frac{1}{2}\left(\widehat{\mathrm{ES}}_1 + \widehat{\mathrm{ES}}_2\right)$$

as a better estimation or choose $x_{[t/2]+1}$ as a demarcation point, and use the function below:

$$\widehat{\mathrm{ES}}_3 = \frac{1}{\alpha}\left(\mathrm{VaR} \times \Delta \mathrm{p}_{\mathrm{t}} + \cdots + x_{[t/2]-1} \times \Delta \mathrm{p}_{[t/2]} + x_{[t/2]+1} \times \Delta \mathrm{p}_{[t/2]+1} \ldots + \sum_{i=1}^{m} c_i \times \Delta \mathrm{p}_0\right)$$

We are going to use $\widehat{\mathrm{ES}}_3$ as ES estimator in our paper, it's more suitable in the situation which the inverse probability function grows faster in the tail part.

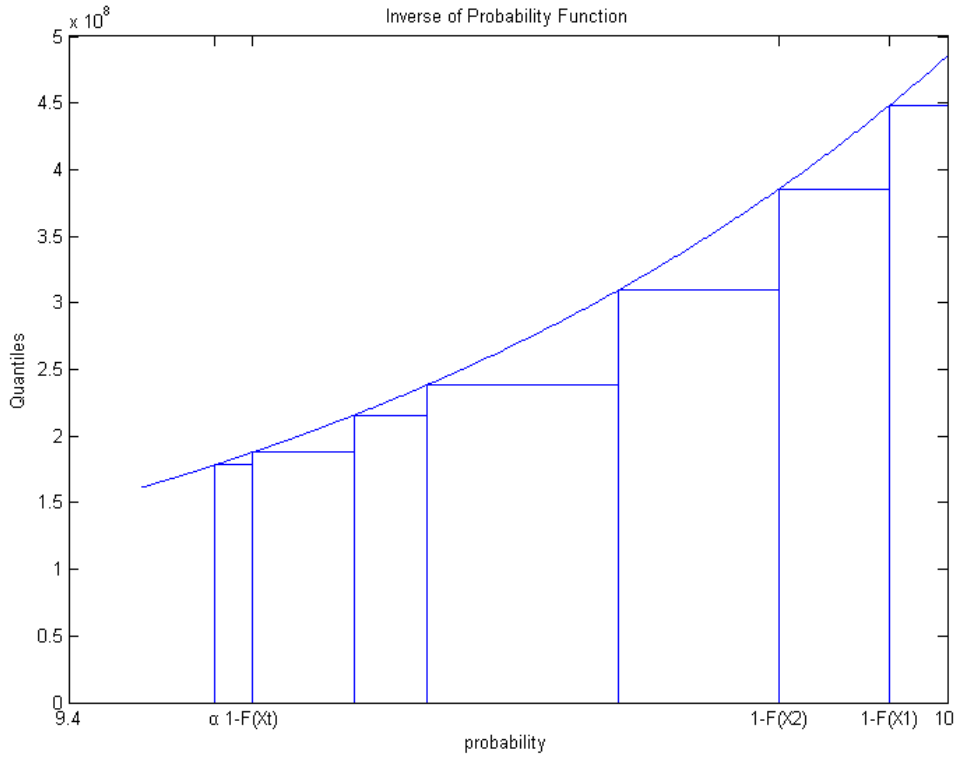

Figure 2 one approach for estimate ES

3.2.2 Search Method for SR

3.2.2.1 Piecewise polynomial loss function

As described in Section 2.1.3, one can apply the recursive algorithm based on (2.5) to estimate s. It remains to discuss how to estimate $\mathrm{E}[f(\mathrm{L}-\mathrm{s})]$ for fixed value $\mathrm{s} \in (0, \sum_{i=1}^{m} c_i)$. Plain MC method does not yield reliable estimator for $\mathrm{E}[f(\mathrm{L}-\mathrm{s})]$, unless very large sample size are

considered. Follow J.Dunkel and S.Weber (2007), we employ the exponential twisting importance sampling method to construct estimators for $\mathrm{E}[f(\mathrm{L}-\mathrm{s})]$.

As the same exponential twisting procedure described in Section 3.2 and the (2.4a), we can have the estimator of $\mathrm{E}[\gamma^{-1}(\mathrm{L}-\mathrm{s})^{\gamma}\mathbf{1}_{\{L-s>0\}}]$ as:

$$\widehat{\mathrm{E}}[\gamma^{-1}(\mathrm{L}-\mathrm{s})^{\gamma}\mathbf{1}_{\{L-s>0\}}] = \frac{1}{N}\sum_{i=1}^{N}\gamma^{-1}(\mathrm{L_i}-\mathrm{s})^{\gamma}\mathbf{1}_{\{L_i-s>0\}}\exp[-\theta\mathrm{L_i}+\psi(\theta)] \qquad (3.12)$$

where N is the number of simulations and $\psi(\theta)$ is the cumulant generating function (3.3).

Similar to (3.4), we should have that the variance of the estimators based on (3.12) is significantly smaller than the variance of the corresponding plain MC estimator. Since the estimator is unbiased, it's equivalent to consider the second moment,

$$\mathrm{M_2(s,\theta)} = \frac{1}{\gamma^2}\mathrm{E_{p_{i,\theta}}}\left[(\mathrm{L}-\mathrm{s})^{2\gamma}\mathrm{I}\{\mathrm{L}>s\}\mathrm{e}^{-2\theta\mathrm{L}+2\psi(\theta)}\right] \le \mathrm{M_2(s,0)e}^{-2\theta\mathrm{x}+2\psi(\theta)} \qquad (3.13)$$

Here $\mathrm{M_2(s,0)} = \mathrm{E}[(\mathrm{L}-\mathrm{s})^{2\gamma}\mathrm{I}\{\mathrm{L}>s\}]$ is the second moment without exponential twisting. Consequently, instead of directly minimize $\mathrm{M_2(s,\theta)}$, we can generally minimize the upper bound on the rhs. of inequality (3.13). The choice for the twisting parameter is given by:

$$\theta_{\mathrm{s}} = \begin{cases} \text{unique solution of } \psi'(\theta) = \mathrm{s}, & \mathrm{s} > \psi'(0) \\ 0, & \mathrm{s} \le \psi'(0) \end{cases} \qquad (3.14)$$

Similar to the Search Method for VaR, we can have the Search Method procedure based on two-step importance sampling as follow:

1. Take s
2. Generate $\mathrm{Z}\sim\mathrm{N}(\mu,\mathrm{I})$, a d-vector of independent normal random variables shift from N(0,I)
3. Calculate the new conditional default probabilities

$$\mathrm{q_i}(\theta_{\mathrm{s}}(\mathrm{Z}),\mathrm{Z}) = \frac{\mathrm{p_i(Z)e}^{\theta_{\mathrm{s}}(\mathrm{Z})\mathrm{c_i}}}{1+\mathrm{p_i(Z)}(\mathrm{e}^{\theta_{\mathrm{s}}(\mathrm{Z})\mathrm{c_i}}-1)}$$

   with $\theta_{\mathrm{s}}(\mathrm{Z})$ given by (3.14) and $\mathrm{p_i(Z)}$ given by (2.10).
4. Generate the default indicator $\mathrm{Y_1},\ldots,\mathrm{Y_m}$ from Bernoulli random numbers Bin(1,p) with the probability $\mathrm{q_i}(\theta_{\mathrm{s}}(\mathrm{Z}),\mathrm{Z})$.
5. Calculate the loss L from the (2.6) and return the estimator of $\mathrm{E}[\gamma^{-1}(\mathrm{L}-\mathrm{s})^{\gamma}\mathbf{1}_{\{L-s>0\}}]$

$$\frac{1}{N}\sum_{i=1}^{N}\gamma^{-1}(L_i-s)^{\gamma}\mathbf{1}_{\{L_i-s>0\}}\,exp[-\theta L_i+\psi(\theta)]\,exp\,(-\mu^T Z+\frac{\mu^T\mu}{2})$$

where the factor $e^{-\mu^T Z+\frac{\mu^T \mu}{2}}$ is the likelihood ratio for the change from density of the N(0,I) distribution to that of N(μ,I) distribution.

6. Employ the recursive algorithm (2.5) get the estimator of SR_p

Now we need to discuss how to determine the shift vector μ. Similar as the Glasserman and Li (2005) did, we have

$$\mu \approx \arg\max_z \left\{ F_s(z) - \frac{1}{2} z^T z \right\}$$

In simulations given in Chapter 5, μ is obtained by modified Newton's Method.

3.2.2.2 Exponential loss function

As another example of SR, we use the (2.4b). Luckily, the corresponding SR measure can be explicitly be calculated, i.e.

$$SR_\lambda(L) = \frac{1}{\beta} \log\left(\frac{E\left[e^{\beta L}\right]}{\lambda} \right), \tag{3.15}$$

It is therefore not necessary to apply (2.5) when calculating this particular risk measure.

In the case of dependent defaults, (3.15) can be rewritten as

$$SR_\lambda(L) = \frac{1}{\beta} \left[\int e^{\psi(\beta,z)} dF(z) - log\lambda \right], \tag{3.16}$$

where

$$\psi(\beta, z) = \log(E[\exp(\beta L) \,|Z = z]) = \sum_{i=1}^{m} \log\left[1 + p_i(z)\left(e^{\beta c_i} - 1\right)\right],$$

$\psi(\beta, z)$ is the conditional cumulant generating function, and the distribution F of the factor variables Z is given by the d-dimensional standard normal distribution.

The estimator for the risk measure (3.16) can be obtained by sampling from N Gaussian random vector $Z = (Z_1, \ldots, Z_d)$ and returning the value

$$\widehat{SR}_\lambda(L) = \frac{1}{\beta} \log\left(\frac{1}{N} \sum_{k=1}^{N} \left\{ \prod_{i=1}^{m} \left[1 + p_i(z^k)\left(e^{\beta c_i} - 1\right)\right] \right\} \right) - \frac{1}{\beta} log\lambda, \qquad z^k \sim N(0, I)$$

Variance reduction can be achieved by importance sampling with respect to the factor vector Z. If we restrict attention to measure changes that shift only the mean of Z, a suitable choice of μ can be obtained as a solution of the maximization problem

$$\mu \approx \arg\max_z \left\{ \psi(\beta, z) - \frac{1}{2} z^T z \right\}$$

The likelihood ratio of the measure change from N(0, I) to N(μ, I) modifies the MC estimator. The importance sampling estimator is thus given by

$$\widehat{SR}_{\lambda}(L) = \frac{1}{\beta}\log\left(\frac{1}{N}\sum_{k=1}^{N}\left\{\exp\left(-\mu^{\mathrm{T}}z^{k} + \frac{\mu^{\mathrm{T}}\mu}{2}\right)\prod_{i=1}^{m}\left[1 + p_i(z^{k})\left(e^{\beta c_i} - 1\right)\right]\right\}\right)$$
$$-\frac{1}{\beta}log\lambda, \qquad z^{k}\sim N(\mu, I)$$

3.3 Direct method for estimate risk measures in NCM

3.3.1 Direct Approximation Method based on IS for VaR and ES

As we showed, the key for applying IS method to calculate VaR is how to get quantile from probability, as α-VaR equals the α-quantile. We develop this direct approximation method from the work of Peter W. Glynn in 1996.

Our goal is to compute the quantile $F^{-1}(\alpha) = \inf\{x: F(x) \geq \alpha\}$. We would like to use the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}$ represent the F(x). And estimator of quantile $F_n^{-1}(\alpha) = \inf\{x: F_n(x) \geq \alpha\}$.

As the definition of the importance sampling, we insert a known function g, and get the likelihood ratio. We can rewrite the density as follow:

$$F(dx) = g(x)\tilde{F}(dx),$$
$$P(X_i > x) = \tilde{E}\left[r_i 1_{\{X_i > x\}}\right],$$

where r_i is the likelihood ratio associated with X_i. The two above equalities motivate the following approximation of F:

$$\tilde{F}_n = 1 - \frac{1}{n}\sum_{i=1}^{n} r_i 1_{\{X_i > x\}}$$

valid in tail where X_i is generated from $\tilde{F}$ rather than F. The corresponding quantile estimator is then defined by:

$$\tilde{F}_n^{-1}(\alpha) = inf\left\{x: \tilde{F}_n(x) \geq \alpha\right\}$$

According to the large deviations theory, the tail approximation

$$P(X > x) \approx \exp\left(-x\theta_x + \psi(\theta_x)\right) \tag{3.17}$$

is valid for x ≫ E(x), where the θ_x is same as (3.5) in Section 3.2.1, and ψ is the cumulant generating function of X. Not surprisingly, the tail approximation (3.17) suggests a quantile approximation.

Let $\tilde{\theta}_p$ be the root of the equation

$$-\psi'(\tilde{\theta}_p)\tilde{\theta}_p + \psi(\tilde{\theta}_p) = \log(1-p) \tag{3.18}$$

for p close to 1, it implies x ≫ E(x),

$$P\left(X > \psi'(\tilde{\theta}_p)\right) \approx 1 - p \tag{3.19}$$

(3.19) suggests that $\psi'(\tilde{\theta}_p)$ can be used as an approximation to the quantile $\tilde{F}_n^{-1}(p)$. Of course, this approximation is crude.

If we set $g(x) = \exp\left(-x\tilde{\theta}_p + \psi(\tilde{\theta}_p)\right)$, the relation (3.19) indicates that, under $\tilde{F}$, sampling from the appropriate tail event associated with the quantile $F^{-1}(p)$ is no longer a rare event, suggesting the possibility of a variance reduction.

Now we could apply the tail approximation importance sampling method into the NCM. Here we only employ the one-step importance procedure as follows:

1. Set vector p = α,
2. Generate $Z \sim N(0, I)$, a d-vector of independent normal random variables,
3. Generate $\tilde{\theta}_p$ from (3.18), where ψ is refer to (3.3),
4. Calculate the new conditional default probabilities with $\tilde{\theta}_p$

$$q_i(\tilde{\theta}_p, Z) = \frac{p_i(Z)e^{\tilde{\theta}_p c_i}}{1 + p_i(Z)\left(e^{\tilde{\theta}_p c_i} - 1\right)}$$

   $p_i(Z)$ given by (2.10),
5. Generate the default indicator $Y_1, \ldots, Y_m$ from Bernoulli random numbers Bin(1,p) with the probability $q_i(\tilde{\theta}_p, Z)$,
6. Calculate the loss vector L by (2.6) and likelihood ratio vector r by $\exp\left(-x\tilde{\theta}_p + \psi(\tilde{\theta}_p)\right)$.

Replicate the procedure for n times, we get a set of simulated values $(L_1, r_1), \ldots, (L_n, r_n)$. We sort the L_is in descending order, which thereby forming the ordered sample $(L_{(1)}, \ldots, L_{(n)})$. We could get the α quantile by the value $L_{(t)}$ associated with the first integer t for which

$$\sum_{j=1}^{t} r_{(j)} \geq (1-p)n \tag{3.20}$$

where $r_{(j)} = \exp\left(-L_{(j)}\tilde{\theta}_{p(j)} + \psi(\tilde{\theta}_{p(j)})\right)$.

Similar to the empirical function, we could have the α-VaR and α-ES estimator as

$$\widehat{\mathrm{VaR}}_{\alpha}(\mathrm{L}) = \mathrm{L}_{\mathrm{t}}$$

$$\widehat{\mathrm{ES}}_{\alpha}(\mathrm{L}) = \frac{1}{\mathrm{N}(1-\mathrm{p})}\sum_{\mathrm{k}=1}^{\mathrm{t}} \mathrm{L}_{\mathrm{k}}\mathrm{r}_{\mathrm{k}}$$

where t is from the VaR estimator.

We should mention that we only employ one step importance sampling estimator for the Direct Method. For the model with high underlying correlation, the two-step importance estimator (3.8) is more reasonable. But the tail bound approximation affects the efficiency very much, and we demonstrate this in the numerical part. It's obviously that this direct method would be faster to get the result of the VaR and ES. In Section 3.2, the Search Method based on importance sampling is a two-stage method. In the first stage, it gets the estimator of probability $P(L > x)$, then the VaR could be calculated by iteration procedure by (3.10), and ES is calculated by the information of the VaR. Due to the high computation cost of the MC method, so the direct method is more efficient than the two-stage procedure.

3.3.2 Direct Approximation Method based on IS for Shortfall Risk

3.3.2.1 Piecewise polynomial loss function

Similar to plain Monte Carlo method, a sequence of loss value L could be get by Direct Method without setting the threshold value *x* or *s*. We prefer this one stage algorithm rather than the two-stage method, Search Method. Combine the Direct Method procedure with the recursive procedure (2.5) to calculate s_{k}, the algorithm procedure is as follow:

1. Set vector p from $p_1 \sim p_2$ with n elements, where p_1 is relative small, and p_2 is close to 1,
2. Generate $\mathrm{Z} \sim \mathrm{N}(0, \mathrm{I})$, a d-vector of independent normal random variables,
3. Generate $\tilde{\theta}_p$ from (3.18), where ψ is refer to (3.3),
4. Calculate the new conditional default probabilities with $\tilde{\theta}_p$

$$\mathrm{q}_{\mathrm{i}}\left(\tilde{\theta}_p, \mathrm{Z}\right) = \frac{\mathrm{p}_{\mathrm{i}}(\mathrm{Z})\mathrm{e}^{\tilde{\theta}_p \mathrm{c}_{\mathrm{i}}}}{1 + \mathrm{p}_{\mathrm{i}}(\mathrm{Z})\left(\mathrm{e}^{\tilde{\theta}_p \mathrm{c}_{\mathrm{i}}} - 1\right)}$$

   $\mathrm{p}_{\mathrm{i}}(\mathrm{Z})$ given by (2.10),
5. Generate the default indicator $\mathrm{Y}_1, \ldots, \mathrm{Y}_{\mathrm{m}}$ from Bernoulli random numbers Bin(1,p) with the probability $\mathrm{q}_{\mathrm{i}}\left(\tilde{\theta}_p, \mathrm{Z}\right)$,
6. Calculate the loss vector L by (2.6) and likelihood ratio vector r by $\exp\left(-x\tilde{\theta}_p + \psi\left(\tilde{\theta}_p\right)\right)$

7. Set initial guess s_0 and s_1, calculate the expected value $E[\gamma^{-1}(L-s)^{\gamma}\mathbf{1}_{\{L-s>0\}}]$ by

$$\frac{1}{N}\sum_{i=1}^{N}\gamma^{-1}(L_i-s_k)^{\gamma}\mathbf{1}_{\{L_i-s_k>0\}}\,exp\left[-\tilde{\theta}_p L_i+\psi\left(\tilde{\theta}_p\right)\right] \tag{3.21}$$

8. Insert (3.21) into the recursive procedure (2.5), stop the recursive when error between $E[\gamma^{-1}(L-s)^{\gamma}\mathbf{1}_{\{L-s>0\}}]$ and parameter λ is smaller enough, the value s^* is take as the estimator of utility based Shortfall Risk with piecewise polynomial loss function

Only one step importance sampling method is employed, and the defects have been discussed in Section 3.3.1.

3.3.2.2 Exponential loss function

As we mentioned in 3.2.2.2, we can get SR with exponential loss function explicitly. No need Direct Method estimator.

Chapter 4
Estimate risk measures in Mixed Poisson Model

4. 1 Plain Monte Carlo method in MPM

In the Mixed Poisson Model, the counter Y_i is conditionally Poisson distributed; this could be seen as a Poisson approximation to a Bernoulli random variable. The approximation is based on the fact that a Poisson random variable with a very small mean has a very small probability of taking a value other than 0 or 1. According to (2.11) and (2.12), we have the MC procedure as follows:

1. Generate $Z_k \sim \text{Gamma}(\alpha_k, \beta_k)$ which $\alpha_k = 1/\sigma_k^2, \beta_k = \sigma_k^2, k = 1,2, \dots, d$;
2. Calculate the $X = (X_1, \dots, X_m)$ via (2.12) ;
3. Generate indicator Y_i from Poisson(X_i) with X_i calculated in step 2;
4. Compute the loss by using the (2.6) and repeat the procedure N times. N should be relative large to protect the accuracy.
5. Sort the N losses by descending order, and output as vector L. L_i is the ith largest loss value in the repeated MC simulations.

We could get the α-VaR and α-ES estimator according to empirical function:

$$\widehat{\text{VaR}}_\alpha(L) = L_{[N(1-\alpha)]+1}$$

$$\widehat{\text{ES}}_\alpha(L) = \frac{1}{[N(1-\alpha)]+1} \sum_{k=1}^{[N(1-\alpha)]+1} L_k$$

Same as in NCM, we simulate VaR and ES together in this chapter.

Also same as in NCM, we focus on estimating $\mathrm{E}[f(\mathrm{L}-\mathrm{c})]$ for piecewise polynomial loss function. And we could get explicit solution for exponential loss function which would be introduced later.

Insert (2.4a), we have

$$p(c) = \mathrm{E}\left[\gamma^{-1}(\mathrm{L}-\mathrm{c})^{\gamma}\mathbf{1}_{\{(\mathrm{L}-\mathrm{c})>0\}}\right]$$

Here is the MC algorithm for piecewise polynomial loss functions:

1. Generate N losses by descending order in one vector L (same as for VaR and ES)
2. For p(c) we have the estimator:

$$\hat{p}(c) = \frac{1}{\mathrm{n}}\sum_{\mathrm{k}=1}^{\mathrm{n}}\gamma^{-1}(\mathrm{L}_{\mathrm{k}}-\mathrm{c})^{\gamma}$$

Where L_{n} is the minimal loss in vector L which is larger than c.

4.2 Two-step importance sampling for estimate risk measures in MPM

We consider the estimation of risk measures for the Mixed Poisson Model. As in the Normal Copulas Model, we firstly use two-step IS method to get the probability (P. Glasserman and J. Li (2003), then we use the search method to get VaR and ES based on the information of VaR. Then we would show the two-step importance sampling estimator for SR. Due to the special structure of the mixed Poisson model, we can easily combine two-step IS method conveniently and efficiently. Last, we would employ the direct method again to estimate VaR and ES, as this two risk measures are most popular in industry.

4.2.1 Search Method based on two-step IS for VaR and ES

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

$$\mathrm{r}_1(\theta,\mathrm{X}) = \exp\left(-\theta\mathrm{L}+\psi(\theta,\mathrm{X})\right) \tag{4.1}$$

where

$$\psi(\theta,\mathrm{X}) = \log\left(\mathrm{E}\left[\mathrm{e}^{\theta\mathrm{L}}\middle|\mathrm{X}\right]\right) = \sum_{\mathrm{i}=1}^{\mathrm{m}}\mathrm{X}_{\mathrm{i}}\left(\mathrm{e}^{\theta\mathrm{c}_{\mathrm{i}}}-1\right)$$

is the conditional cumulant generating function of L with respect to X. Y_is are independent conditional on $X = (X_1, \dots, X_m)$. Under the changed measure Y_is are independent Poisson random variables with

$$E_\theta[Y_i|X] = X_i e^{\theta c_i}$$

Obviously, choosing $\theta > 0$ increase the conditional mean of the default indicator Y_i. Thus, the default events are more likely to occur under the new measure as desired. Now we employ exponential twisting of the independent factor variables Z_k in order to achieve further variance reduction. We have the likelihood ratio

$$r_2(z, Z) = \exp\left(-\sum_{k=1}^{d} \{z_k Z_k + \alpha_k \log(1 - \beta_k z_k)\} \right) \tag{4.2}$$

Here z_k denotes the parameter of the second measure change, where

$$\psi_k(z_k) = \log\left(E\left[e^{\theta_k z_k}\right]\right) = -\alpha_k \log(1 - \beta_k z_k)$$

is the cumulant generating function of original $\text{Gamma}(\alpha_k, \beta_k)$ distributed variable Z_k. With respect to new measure, each of the factor variables Z_k are again independent and Z_k obeys $\text{Gamma}(\alpha_k, \beta_k/(1 - \beta_k z_k))$.

Combine the two measure change from (4.1) and (4.2). The likelihood ratio is given by the product of the two equations

$$r_3(\theta, z, Z) = r_1(\theta, X) r_2(z, Z) = \exp\left(-\theta L + \psi^{(1)}(\theta) + \psi^{(2)}(z) + \psi^{(3)}(\theta, z, Z)\right) \tag{4.3}$$

where

$$\psi^{(1)}(\theta) = \sum_{i=1}^{m} A_{i0}\left(e^{\theta c_i} - 1\right) \tag{4.3a}$$

$$\psi^{(2)}(z) = -\sum_{k=1}^{d} \alpha_k \log(1 - \beta_k z_k) \tag{4.3b}$$

$$\psi^{(3)}(z) = \sum_{k=1}^{d} Z_k[-z_k + \sum_{i=1}^{m} A_{ik}\left(e^{\theta c_i} - 1\right)] \tag{4.3c}$$

For simplicity, we could choose the z_k and θ such that (4.3c) equals to 0, which means the likelihood ratio (4.3) depends on factors θ.

$$z_k = \sum_{i=1}^{m} A_{ik}\left(e^{\theta c_i} - 1\right) \tag{4.4}$$

Hence the final form of the likelihood ratio is

$$\frac{dP}{dQ_\theta} = \exp\left(-\theta L + \psi(\theta)\right) \tag{4.5}$$

where

$$\psi(\theta) = \sum_{i=1}^{m} A_{i0}\left(e^{\theta c_i} - 1\right) - \sum_{k=1}^{d} \alpha_k \log\left(1 - \beta_k \sum_{i=1}^{m} A_{ik}\left(e^{\theta c_i} - 1\right)\right) \tag{4.5a}$$

is the cumulant generating function of L under the original measure P.

For the choice of θ, we use the same idea as in the NCM case. Choose

$$\theta_x = \begin{cases} \text{unique solution of } \psi'_{L,m}(\theta_x) = x, & x > \psi'(0) \\ 0, & x \le \psi'(0) \end{cases} \tag{4.6}$$

Thus the two-step IS estimator of P(L>x) for MPM is:

$$\widehat{P}(L > x) = \frac{1}{N}\sum_{k=1}^{N} I\{L_k > x\} \exp\left(-\theta_x L_k + \psi_{L,m}(\theta_x)\right) \tag{4.7}$$

Here we summarize the IS algorithm as follow:

1. Set $\psi_{L,m}(\theta)$ and solve for θ_x as in (4.6)
2. Generate $Z_k \sim \Gamma\left(\alpha_k, {\beta_k}/{1 - \beta_k z_k}\right), k = 1, \dots, d$, where z_k from (4.4)
3. Compute the conditional mean $X_i, i = 1, \dots, m$, as in (2.12)
4. Generate $Y_i \sim \text{Poisson}\left(X_i e^{\theta_x c_i}\right), i = 1, \dots, m$
5. Calculate loss L according to (2.6)
6. Repeat step 1~6 for N times and return the two-step IS estimator as (4.7)

4.2.1.1 Estimator for VaR

α-VaR is the unique solution of the function as same as (3.9) in Section 3.2.1.2.

$$f(x) = P(L > x) - (1 - \alpha)$$

We could take $x_{-1} = 0$ and $x_0 = \sum c_i$ as the initial guess and we use the same recursive procedure:

$$x_{k+1} = \frac{1}{2}(x_{-1} + x_0), \quad where\ x_{-1} = \max(x_k | f(x_k) > 0), x_0 = \min(x_k | f(x_k) < 0)$$

The IS two-step IS estimator of the $P(L > x)$ is asymptotic optimal (P. Glasserman and J. Li (2003).

4.2.1.2 Estimator for ES

As we mentioned, we use the square element approximation to estimate the ES in Section 3.2.1.3. Here we would like to employ the same method to estimate ES by using the same simulation result as the VaR. We would have the same estimator as in Section 3.2.1.3 (3.11).
We have the estimator as:

$$\widehat{\mathrm{ES}}_1 = \frac{1}{\alpha}(\mathrm{VaR} \times \Delta \mathrm{p_t} + x_t \times \Delta \mathrm{p_{t-1}} + \cdots + x_1 \times \Delta \mathrm{p_0})$$

where $\Delta \mathrm{p_0} = 1 - P(L > x_1),\ \Delta \mathrm{p_1} = P(L > x_1) - P(L > x_2), \ldots,\ \Delta \mathrm{p_t} = P(L > x_t) - \alpha$, and we can have another way of estimation as follow:

$$\widehat{\mathrm{ES}}_2 = \frac{1}{\alpha}(x_t \times \Delta \mathrm{p_t} + \cdots + x_1 \times \Delta \mathrm{p_1} + \sum_{i=1}^{m} c_i \times \Delta \mathrm{p_0})$$

where $\Delta \mathrm{p_k}$ as same as in (3.20).

It's easy to see that $\widehat{\mathrm{ES}}_2 > ES > \widehat{\mathrm{ES}}_1$, $\widehat{\mathrm{ES}}_1$ and $\widehat{\mathrm{ES}}_2$ are the lower and upper bound of the estimation of ES by the rectangle element approach. We could take

$$\widehat{\mathrm{ES}} = \frac{1}{2}\left(\widehat{\mathrm{ES}}_1 + \widehat{\mathrm{ES}}_2\right)$$

as a better estimation or choose $x_{[t/2]+1}$ as a demarcation point, and use the function below:

$$\widehat{\mathrm{ES}}_3 = \frac{1}{\alpha}(\mathrm{VaR} \times \Delta \mathrm{p_t} + \cdots + x_{[t/2]-1} \times \Delta \mathrm{p_{[t/2]}} + x_{[t/2]+1} \times \Delta \mathrm{p_{[t/2]+1}} \ldots + \sum_{i=1}^{m} c_i \times \Delta \mathrm{p_0})$$

We are going to use $\widehat{\mathrm{ES}}_3$ to estimate the ES in the numerical simulation part; it's more suitable in the situation which the inverse probability function grows faster in the tail part.

4.2.2 Search Method based on IS for Shortfall Risk

4.2.2.1 Piecewise polynomial loss function

Here we would outline the main aspects of the importance sampling algorithm for estimating $\mathrm{E}\left[\gamma^{-1}(\mathrm{L} - \mathrm{c})^{\gamma}\mathbf{1}_{\{(\mathrm{L}-\mathrm{c})>0\}}\right]$. Conceptually, the approach is quite similar to the two-step method discussed in Section 3.2.2.1.

Firstly, we assume the values of the common risk factors $Z_1, \dots, Z_d$ are given, so that Y_is are independent Poisson random variables with parameters X_i condition on factors Z. In analogy with the procedure for case in NCM, we can easily get likelihood ratio of loss L as:

$$\exp\left(-\theta L + \sum_{i=1}^{m} X_i(e^{\theta c_i} - 1)\right) \tag{4.8}$$

And $\sum_{i=1}^{m} X_i(e^{\theta c_i} - 1)$ is the conditional cumulant generating function of L given the risk factors $Z_1, \dots, Z_d$.

Secondly, we apply the importance sampling to the common risk factors. We consider exponentially twisting each Z_k by some z_k. We get a change of distribution through likelihood ratio:

$$\exp\left(-\sum_{k=1}^{d}\{z_k Z_k + \alpha_k \log(1 - \beta_k z_k)\}\right) \tag{4.9}$$

where $-\alpha_k \log(1 - \beta_k z_k)$ is the cumulant generating function of Z_k, which has a $\text{Gamma}(\alpha_k, \beta_k)$ distribution. It's obviously that $z_k < 1/\beta_k$. Under the distribution defined by z_k, we can get $Z_k \sim \Gamma(\alpha_k, {\beta_k}/{1 - \beta_k z_k})$. It shows exponential twisting to a gamma distribution produces another gamma distribution with same shape parameter and a different scale parameter. From the product of (4.8) and (4.9), we obtain the likelihood ratio for the two-step change of distribution. X_is are determined via (2.12). The likelihood ratio for two-step IS method can be written as:

$$\exp\left(-\theta L + \psi^{(1)}(\theta) + \psi^{(2)}(z)\right)$$

where

$$\psi^{(1)}(\theta) = \sum_{i=1}^{m} A_{i0}\left(e^{\theta c_i} - 1\right)$$

$$\psi^{(2)}(z) = -\sum_{k=1}^{d} \alpha_k \log(1 - \beta_k z_k)$$

when we choose

$$z_k = \sum_{i=1}^{m} A_{ik}\left(e^{\theta c_i} - 1\right) \quad k = 1, \dots, d \tag{4.10}$$

Choosing z_k in this way, we can eliminate the Z_k from the likelihood ratio, leaving only the dependence of L.

The rest is how to choose θ. It can be proved the $\psi_{L,m}(\theta) = \psi^{(1)}(\theta) + \psi^{(2)}(z)$ is the cumulant generating function of L. Similar as in the NCM case, we choose

$$\theta_s = \begin{cases} \text{unique solution of } \psi'_{L,m}(\theta_s) = s, & s > \psi'(0) \\ 0, & s \le \psi'(0) \end{cases} \tag{4.11}$$

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

1. Set s and solve for θ_s as in (4.11)
2. Generate $Z_k \sim \Gamma\left(\alpha_k, {\beta_k}/{1 - \beta_k z_k}\right), k = 1, \dots, d$, where z_k from (4.10)
3. Compute the conditional mean $X_i, i = 1, \dots, m$, as in (2.12)
4. Generate $Y_i \sim \text{Poisson}\left(X_i e^{\theta_s c_i}\right), i = 1, \dots, m$
5. Calculate loss L according to (2.6)
6. Repeat step 1~6 for N times and return the two-step IS estimator for $E\left[\gamma^{-1}(L - c)^{\gamma}\mathbf{1}_{\{(L-c)>0\}}\right]$ as

$$\frac{1}{N}\sum_{k=1}^{N}\frac{1}{\gamma}(L_k - s)^{\gamma}1_{\{L_k > s\}}\exp\left(-\theta_s L_k + \psi_{L,m}(\theta_s)\right)$$

7. Compare the estimator value with λ

4.2.2.2 Exponential loss function

In the case of MPM, one can calculate analytically the SR associated with the exponential loss function from (2.4b). Combining the explicit representation from (3.16) and the definition of the cumulant generating function we showed

$$SR_{\lambda}(L) = \frac{1}{\beta}\log\left(\frac{E\left[e^{\beta L}\right]}{\lambda}\right) = \frac{1}{\beta}[\psi(\beta) - \log(\lambda)]$$

where according to (4.5a), in MPM

$$\psi(\beta) = \sum_{i=1}^{m} A_{i0}\left(e^{\beta c_i} - 1\right) - \sum_{k=1}^{d}\alpha_k \log\left(1 - \beta_k \sum_{i=1}^{m} A_{ik}\left(e^{\beta c_i} - 1\right)\right)$$

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

4.3 Direct method for estimate risk measures in MPM

4.3.1 Direct Approximation Method based on IS for VaR and ES

As we showed in Section 3.3.1, by taking the following approximation of F

$$\tilde{F}_n = 1 - \frac{1}{n}\sum_{i=1}^{n} r_i 1_{\{X_i > x\}}$$

where X_i is generated from $\tilde{F}$ rather than F. And employ the large deviation theory, we a quantile approximation

$$P(X > x) \approx \exp\left(-x\theta_x + \psi(\theta_x)\right)$$

Let $\tilde{\theta}_p$ be the root of the equation

$$-\psi'\left(\tilde{\theta}_p\right)\tilde{\theta}_p + \psi\left(\tilde{\theta}_p\right) = \log(1-p) \tag{4.12}$$

for p close to 1, it implies x ≫ E(x),

$$P\left(X > \psi'\left(\tilde{\theta}_p\right)\right) \approx 1 - p$$

suggesting that $\psi'\left(\tilde{\theta}_p\right)$ can be used as an approximation to the quantile $\tilde{F}_n^{-1}(p)$. Of course, this approximation is crude.

Now we could apply the tail approximation importance sampling method into the MPM. We can have the procedure as follow:

1. Set vector p = α,
2. Generate $\tilde{\theta}_p$ from (4.12), where ψ is refer to (4.5a),
3. Generate $Z_k \sim \Gamma\left(\alpha_k, {\beta_k}/{1-\beta_k z_k}\right), k = 1, \ldots, d$, where z_k from (4.10)
4. Compute the conditional mean $X_i, i = 1, \ldots, m$, as in (2.12)
5. Generate $Y_i \sim \text{Poisson}\left(X_i e^{\tilde{\theta}_p c_i}\right), i = 1, \ldots, m$
6. Calculate loss L according to (2.6)
7. Calculate the loss vector L by (2.6) and likelihood ratio vector r by $\exp\left(-x\tilde{\theta}_p + \psi\left(\tilde{\theta}_p\right)\right)$.

We get a set of simulated values $(L_1, r_1), \ldots, (L_n, r_n)$. We sort the L_is in descending order, which thereby forming the ordered sample $(L_{(1)}, \ldots, L_{(n)})$. We could get the α quantile by the value $L_{(t)}$ associated with the first integer t for which

$$\sum_{j=1}^{t} r_{(j)} \geq (1-p)n \tag{4.13}$$

where $r_{(j)} = \exp\left(-L_{(j)}\tilde{\theta}_{p(j)} + \psi\left(\tilde{\theta}_{p(j)}\right)\right)$.

Similar to the empirical function, we could have the α-VaR and α-ES estimator as

$$\widehat{\mathrm{VaR}}_\alpha(L) = L_t$$

$$\widehat{\mathrm{ES}}_\alpha(L) = \frac{1}{N(1-p)} \sum_{k=1}^{t} L_k r_k$$

where we get t by (4.13).

It's obviously that this direct method would be faster to get the result of the VaR and ES, cf. the discussion at the end of Section 3.3.

4.3.2 Direct Approximation Method based on IS for Shortfall Risk

4.3.2.1 Piecewise polynomial loss function

Similar to plain Monte Carlo method, a sequence of loss value L could be get by Direct Method without setting the threshold value *x* or *s*. We prefer this one stage algorithm rather than the two-stage method, Search Method. Combine the Direct Method procedure with the recursive procedure (2.5) to calculate s_k, the algorithm procedure is as follow:

1. Set vector p from $p_1 \sim p_2$ with n elements, where p_1 is relative small, and p_2 is close to 1,
2. Generate $\tilde{\theta}_p$ from (4.12), where ψ is refer to (4.5a),
3. Generate $Z_k \sim \Gamma\left(\alpha_k, {\beta_k}/{1-\beta_k z_k}\right), k = 1, \ldots, d$, where z_k from (4.10)
4. Compute the conditional mean $X_i, i = 1, \ldots, m$, as in (2.12)
5. Generate $Y_i \sim \mathrm{Poisson}\left(X_i e^{\tilde{\theta}_p c_i}\right), i = 1, \ldots, m$
6. Calculate loss L according to (2.6) and likelihood ratio vector r by $\exp\left(-x\tilde{\theta}_p + \psi\left(\tilde{\theta}_p\right)\right)$.
7. Set initial guess s_0 and s_1, calculate the expected value $E[\gamma^{-1}(L-s)^\gamma \mathbf{1}_{\{L-s>0\}}]$ by

$$\frac{1}{N}\sum_{i=1}^{N} \gamma^{-1}(L_i - s_k)^\gamma \mathbf{1}_{\{L_i - s_k > 0\}} exp\left[-\tilde{\theta}_p L_i + \psi\left(\tilde{\theta}_p\right)\right] \tag{4.14}$$

8. Insert (4.14) into the recursive procedure (2.5), stop the recursive when error between $E[\gamma^{-1}(L-s)^{\gamma}\mathbf{1}_{\{L-s>0\}}]$ and parameter λ is smaller enough, the value s^* is take as the estimator of utility based Shortfall Risk with piecewise polynomial loss function

Only one step importance sampling method is employed, and the defects has been discussed in Section 3.3.1.

4.3.2.2 Exponential loss function

As we mentioned in Section 4.2.2.2, we can analytically solve SR with exponential loss function. No need Direct Method estimator.

Chapter 5
Numerical Simulations

We are going to show the numerical simulation result of Monte Carlo method and importance sampling here. And compare the efficiency of methods which have been mentioned in the first four chapters. It's obviously that the importance sampling could be more efficient in Mixed Poisson Model as importance sampling step respect to risk variables Z doesn't require additional approximation procedure. So we do the simulation relate to MPM first.

5. 1 Estimate risk measures in MPM

We now show the performance of our procedure to estimate risk measures in a multi-factor model through numerical experiments. We firstly show the result of search method and direct method for VaR and ES respectively, and then result for SR would be introduced as well. To illustrate variance reduction performance of importance sampling algorithms in MPM, we demonstrate a numerical simulation of a simple portfolio with the following parameters:

1. Number of obligors m=10
2. Number of common risk factors d=3
3. Size of exposures $c_i = i, \ \ i = 1, \dots, m$
4. Expected value of latent variables: $p_i := E(X_i) = 0.1$ for i=1,…,m
5. Coupling coefficient: $A_{ik} = 0.01, i = 1, \dots, m, k = 1, \dots, d$ and this yields $A_{i0} = 0.07$
6. Variance parameter for the distribution of the common risk factor variables: $\sigma_k = 1$

From these parameters, we could take the initial guess at $x_{-1} = 0$ and $x_0 = \sum c_i = 55$ for search method. Although the realistic credit portfolios may contain much larger numbers of obligors and risk factors, even the exposure may be random variables, this simple portfolio is sufficient

for illustrating the efficiency of the importance sampling procedure. We will compare with results obtained from plain MC simulations under the above parameters. We replicate each simulation 100 times.

Table 1 α=95%

Number of Simulation N	One Simulation (Time in seconds)	Mean Value of VaR	Std. of VaR	Value of ES	Std. of ES
	Plain Monte Carlo				
100	0.1696	17.1500	1.9142	21.2833	2.5418
200	0.3485	17.8200	1.5595	21.8709	1.8942
500	1.0354	17.7700	0.8391	21.9819	1.0839
1000	2.3156	17.7800	0.7860	22.1933	0.8531
2000	3.5058	17.7900	0.4984	22.2648	0.5843
	Search Method				
400(8*50)	0.9190	17.5400	0.9211	22.0694	0.6642
900(9*100)	1.8247	17.7800	0.6305	22.3034	0.4704
	Direct Method				
100	0.2761	17.9000	0.7587	23.0118	0.6537
500	1.1691	17.9600	0.3897	22.4553	0.3290
1000	2.6534	17.9400	0.2429	22.4524	0.1646

Source: simulation data, by Matlab R2007a

Table 1 is the result of estimating VaR and ES at 95% in Mixed Poisson Model. For the search method, the number in the bracket is the number of x times the replicated value n for each x. We run a large number $\mathrm{N} = 1 \times 10^6$ by plain Monte Carlo method. We would like to use the result of this large MC simulation as the 'true' value of VaR and ES. It takes 3 hours and 8 minutes to finish the simulation (11317.9126 sec). The 'true' value of VaR is 18 and ES is 22.3810. The first column shows the number of generated losses, second column is the computation time to estimate the risk measures. The third and fifth column shows the mean value of estimated VaR and ES of 100 replications, and the standard deviation of the 100 estimated VaR and ES are given in column five and seven.

As showed in table 1, Search Method and Direct Method convergence to the 'true' value much faster than that of plain Monte Carlo Method. It's clear that for roughly same computation time among the three methods, the Direct Method gets the smallest standard deviation for the two risk measures, and has the best estimated result as well. The Search Method could have rather good accuracy when the replicated value n for each x is large enough. Table 2 shows the simulation result when α=99%. From the large deviation theory, we could figure out that the Direct Method would work better when $x \gg E[x]$, and the estimated result of plain Monte Carlo

Method varies a lot from another replication. Because the value 99% represents quite rare events, the plain MC method requires large simulation numbers.

Table 2 α=99%

Number of Simulation N	One Simulation (Time in seconds)	Mean Value of VaR	Std. of VaR	Value of ES	Std. of ES
	Plain Monte Carlo				
500	0.9354	25.0300	1.8719	28.3719	2.3260
1000	1.7315	25.2500	1.3661	28.8309	1.7254
2000	3.2747	25.2500	1.0860	29.2792	1.2537
	Direct Method				
100	0.2340	25.2300	1.4809	31.8301	2.3719
500	1.1712	25.2300	0.5006	29.7665	0.7077
1000	2.2134	25.2800	0.4120	29.6090	0.4067

Source: simulation data, by Matlab R2007a

The 'true' value of VaR and ES at 99% is 25 and 29.6542 respectively. As evident from table 2, compared to plain MC estimators, the estimators based on importance sampling is characterized by significantly better convergence properties, in particular, when increasing the value α. Figure 3 and Figure 4 show the numerical result for VaR and ES at level 95%.

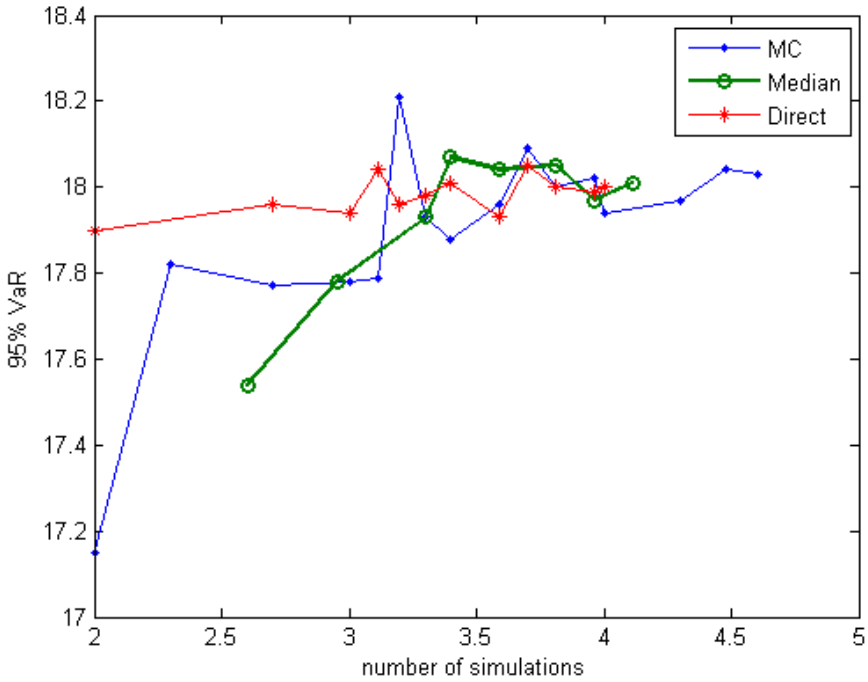

Figure 3 95% VaR

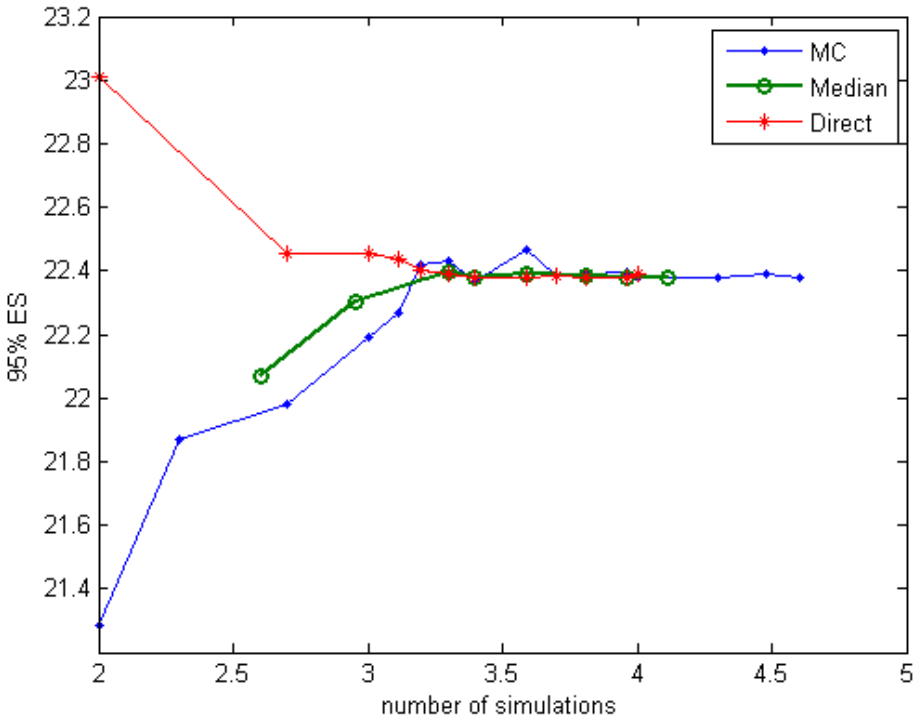

Figure 4 95% ES

The typical value for α is 95% and 99%, but sometimes higher value may also be of interest. We compare the performance of estimating VaR and ES based on $N = 10^4$ samples for α equals to 0.999, 0.9999 and 0.99999. The simulation is repeated 100 times and the mean and standard deviation are shown in Table 3.

Table 3 Number of simulations N = 1e+04

1-α	Time of Simulation (in seconds)	Value of VaR 100 replications	Std. of VaR	Value of ES 100 replications	Std. of ES
	Plain Monte Carlo				
1e-3	16.5612	34.7200	1.2640	38.3800	1.5779
1e-4	17.6072	45.0200	4.4585	45.0200	4.4585
1e-5	17.3187	44.9500	4.2696	44.9500	4.2696
	Direct Method				
1e-3	26.0263	35.0000	0.0000	39.0390	0.0558
1e-4	24.5167	44.0000	0.0000	47.7125	0.0582
1e-5	25.2904	52.3600	0.4824	56.0324	0.0733

Source: simulation data, by Matlab R2007a

It clearly shows in Table 3, plain MC method doesn't work for very rare event, meanwhile the Direct Method have quite good estimation for both VaR and ES.

Due to the highly computation cost when calculate SR with Search method, Table 4 doesn't show the simulation results for utility based Shortfall Risk with piecewise polynomial loss function by search method.

Table 4 $\gamma = 2, \lambda = 1$

Number of Simulation N	One Simulation (Time in seconds)	Mean Value of SR_λ^p (piecewise polynomial)	Std. of SR_λ^p
	Plain Monte Carlo		
500	0.8395	18.0938	2.3935
1000	1.8782	17.8354	1.8062
2000	3.4920	17.7118	0.8884
	Direct Method		
100	0.5529	17.6619	1.1021
500	2.7771	17.8123	0.4890
1000	5.4834	17.7823	0.3028

Source: simulation data, by Matlab R2007a

Evidently, the importance sampling estimator works much more efficiently than the plain Monte Carlo estimator. The standard techniques for VaR could be extended to convex SR measures that do not share the deficiencies of VaR. Larger number of simulation results could be found in the following figure 5.

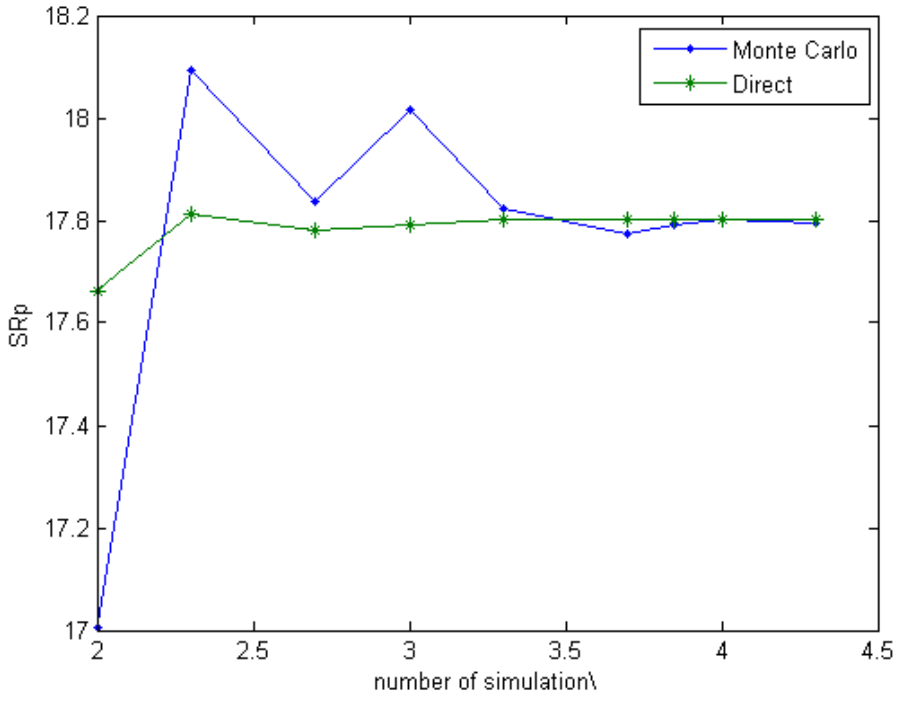

Figure 5 SR_λ^p with piecewise polynomial loss function with parameter $\gamma = 2, \lambda = 1$

5. 2 Estimate risk measures in NCM

Refer to the Glasserman and Li (2003), when underling correlations are not too large, one step importance sampling is also effective. We first give a low correlation example for NCM. We only employ the one step importance sampling for Search Method in this model, which doesn't require the tail bound approximation. The parameters are as follow:

1. Number of obligors m=10
2. Number of common risk factors d=3
3. The marginal default probabilities $p_i = 0.05, \text{where } i = 1, \dots, m$
4. Size of exposures $c_i = i, \text{ where } i = 1, \dots, m$
5. The coupling coefficient $A_{ik} = 0.1, i = 1, \dots, m, k = 1, \dots, d$ and $A_{i0} = 0.985$ for the amplitude of the idiosyncratic risk factor

From these parameters, we could take the initial guess at $x_{-1} = 0$ and $x_0 = \sum c_i = 55$ for search method. This simple portfolio is sufficient for illustrating the efficiency of the importance sampling procedure. Results obtained from plain MC simulations and importance sampling methods would be listed under the above parameters. Simulations would be replicated for 100 times.

Table 5 α=95%

Number of Simulation N	Time of Simulation (in seconds)	Value of VaR	Std. of VaR	Value of ES	Std. of ES
	Plain Monte Carlo				
1000	0.9846	11.1400	0.8530	15.1890	0.7099
2000	1.8701	11.1200	0.6077	15.2224	0.5318
4000	3.9999	11.1000	0.5025	15.2828	0.4268
	Search Method				
1000(10*100)	4.9904	10.95	0.7043.	15.1750	0.5375
1800(9*200)	8.5152	10.93	0.5675	15.2757	0.3600
	Direct Method				
500	1.6042	11.1000	0.6577	15.3722	0.3907
1000	2.0589	11.0700	0.3968	15.3065	0.2402
2000	3.5426	11.1100	0.2532	15.3283	0.1871

Source: simulation data, by Matlab R2007a

Table 5 is the result of estimating VaR and ES at 95% in Normal Copulas Model. We run a large number $N = 1 \times 10^6$ by plain Monte Carlo method. We would like to use the result of this large

MC simulation as the 'true' value of VaR and ES. It takes 2 hours and 48 minutes to finish the simulation (10086.5320 sec). The 'true' value of VaR is 11 and ES is 15.2627.

From the table 5, it is clearly that Direct Method is a rather good method for estimating in NCM. For each simulation, it takes almost double time for each simulation in plain MC, but VaR estimator convergences almost 4 times and ES estimator convergences almost 8 times faster than those of the plain MC. From the table 6 below, which estimate at 99%, shows the direct method convergence even faster, as the large deviation theory would work better when the $x \gg E[x]$. We don't simulate for search method for 99%, as it's not that efficient in NCM.

Table 6 α=99%

Simulations N	Time of Simulation (in seconds)	Value of VaR 100 replications	Std. of VaR	Value of ES 100 replications	Std. of ES
	Plain Monte Carlo				
1000	1.0349	18.5600	1.2053	20.9236	1.1436
2000	1.9372	18.4326	0.8234	20.4683	0.7891
4000	4.1077	18.0321	0.6724	20.5643	0.5632
	Direct Method				
500	1.6086	18.1200	0.8808	20.6115	0.4750
1000	2.1375	18.1000	0.4201	20.6107	0.3512

Source: simulation data, by Matlab R2007a

Similar as in MPM, we compare the performance of estimating VaR and ES for α equals to 0.999, 0.9999 and 0.99999 in NCM. The simulation is repeated 100 times and the mean and standard deviation are shown in Table 7.

Table 7 Number of simulations N = 1e+4

1-α	Time of Simulation (in seconds)	Value of VaR 100 replications	Std. of VaR	Value of ES 100 replications	Std. of ES
	Plain Monte Carlo				
1e-3	9.4104	24.5100	0.9045	26.8282	1.0569
1e-4	9.4366	31.6200	2.5615	31.6200	2.5615
1e-5	9.6320	32.1500	3.1346	32.1500	3.1346
	Direct Method				
1e-3	29.0778	24.9400	0.2387	27.2729	0.1015
1e-4	28.8738	30.8600	0.3766	33.1625	0.3146
1e-5	29.0275	35.9300	0.3828	38.6157	1.3372

Source: simulation data, by Matlab R2007a

Figure 6 and Figure 7 show estimates of the 95% VaR and ES for different sample size N (in log). Y label shows the mean value of VaR and ES with the preceding parameters and X label illustrate the number of replications $\log_{10}(N)$.

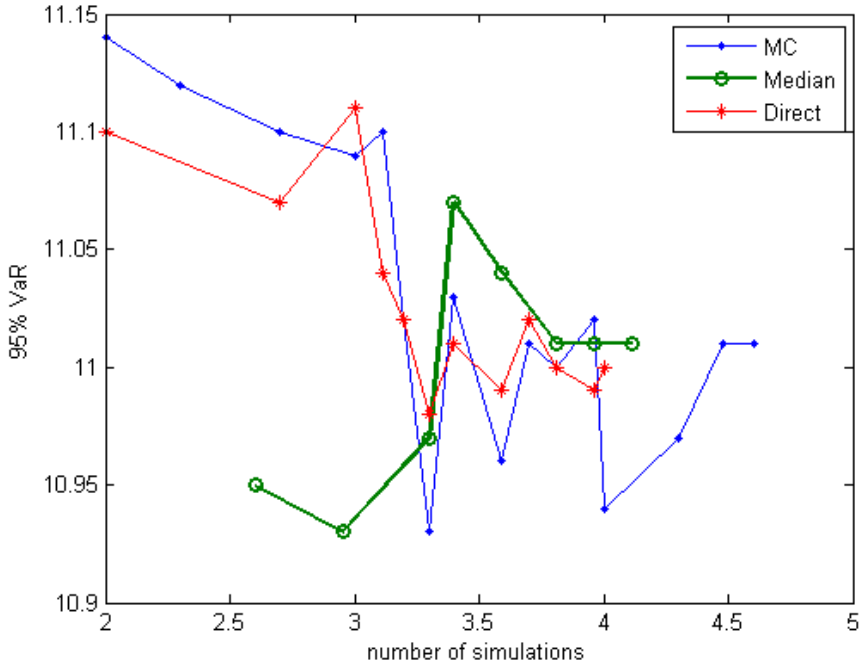

Figure 6 95% VaR

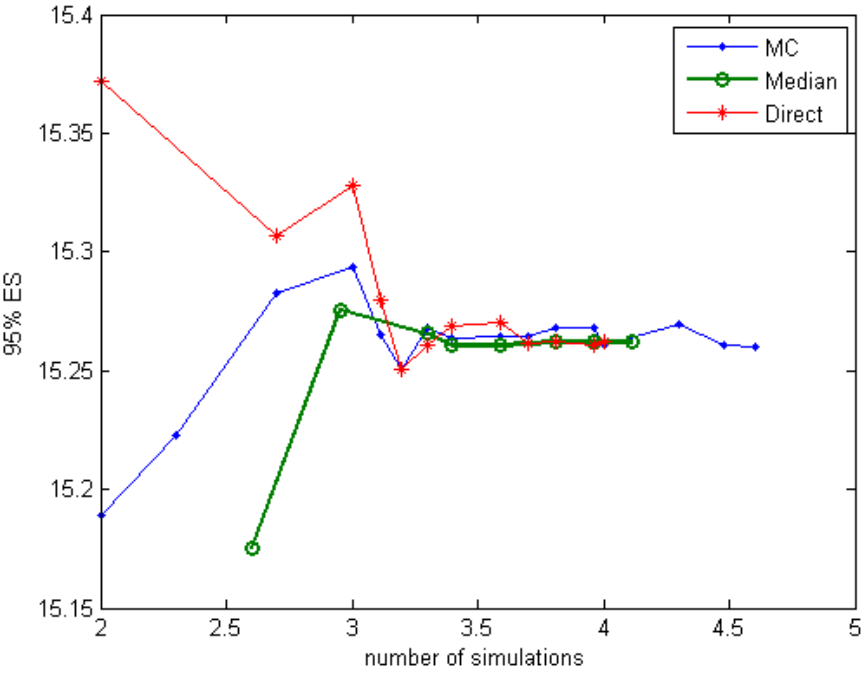

Figure 7 95% ES

The Search Method is a two-stage algorithm when estimate utility based SR with piecewise polynomial loss function, and compared with direct method, it is obviously not that efficient to do the simulation. Table 8 shows the simulation results by plain Monte Carlo estimator and

Direct Method estimator. We take λ=1, β=1 and γ=2 as the parameter to estimate utility based SR.

Table 8 $\gamma = 2, \lambda = 1$

Number of Simulation N	One Simulation (Time in seconds)	Mean Value of SR_{λ}^{p} (piecewise polynomial)	Std. of SR_{λ}^{p} (piecewise polynomial)
	Plain Monte Carlo		
100	0.1030	9.4278	2.6224
500	0.5495	9.8918	1.2811
1000	1.0978	9.8268	0.8885
5000	6.0292	10.0321	0.4481
	Direct Method		
100	0.2329	9.7598	1.0721
500	1.2171	9.8943	0.6321
1000	2.3875	9.9432	0.4323

Source: simulation data, by Matlab R2007a

Table 9 $\beta=1, \lambda = 1$

Number of Simulation N	One Simulation (Time in seconds)	Mean Value of SR_{λ}^{e} (exponential)	Std. of SR_{λ}^{e} (exponential)
	Plain Monte Carlo		
1000	0.9505	19.3210	3.5924
5000	5.4957	23.0918	2.2811
10000	10.0978	34.2268	1.8885
	One-step IS		
100	0.2775	32.1757	0.5790
500	1.4045	32.1933	0.3691
1000	2.9649	32.2378	0.2836

Source: simulation data, by Matlab R2007a

We use algorithm based on one-step importance sampling for estimating SR_{λ}^{p} and SR_{λ}^{exp}. From table 8, we can see that the Direct Method takes approximately twice as long per replication as plain Monte Carlo simulation, but Direct Method offers dramatic increase in precision. As evident from table9, plain Monte Carlo convergence very slow due to the actual value of SR_{λ}^{e} under the inputs β=1, $\lambda = 1$ is quite big. This indicates that Monte Carlo method needs large sample size to estimate the rare event, meanwhile importance sampling algorithms work much better at the tail. We could have more visualized result as figure 8 and 9 illustrated.

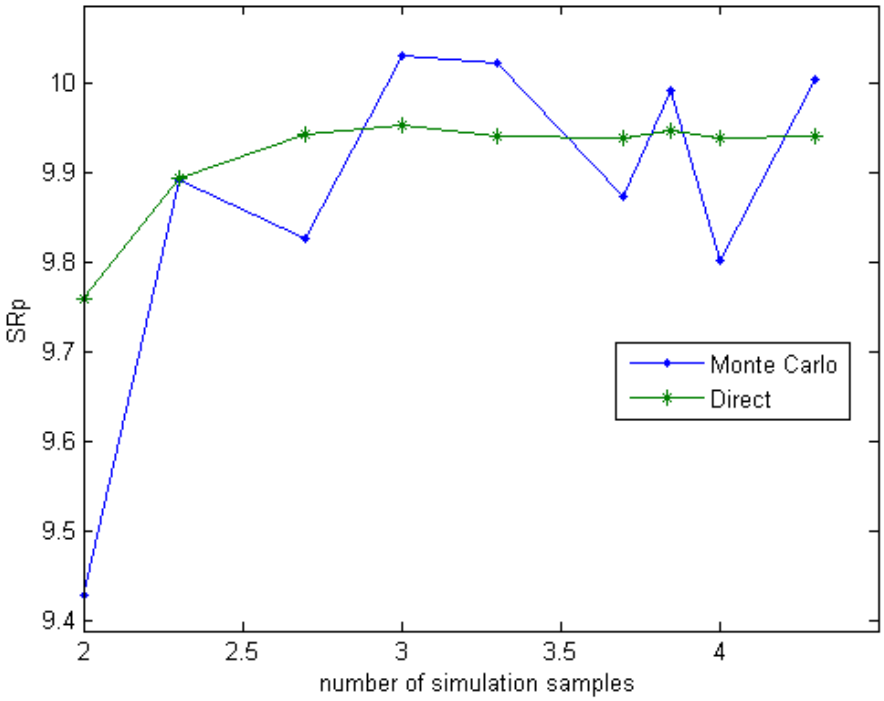

Figure 8 $\mathrm{SR}_{\lambda}^{\mathrm{p}}$, $\gamma = 2, \lambda = 1$

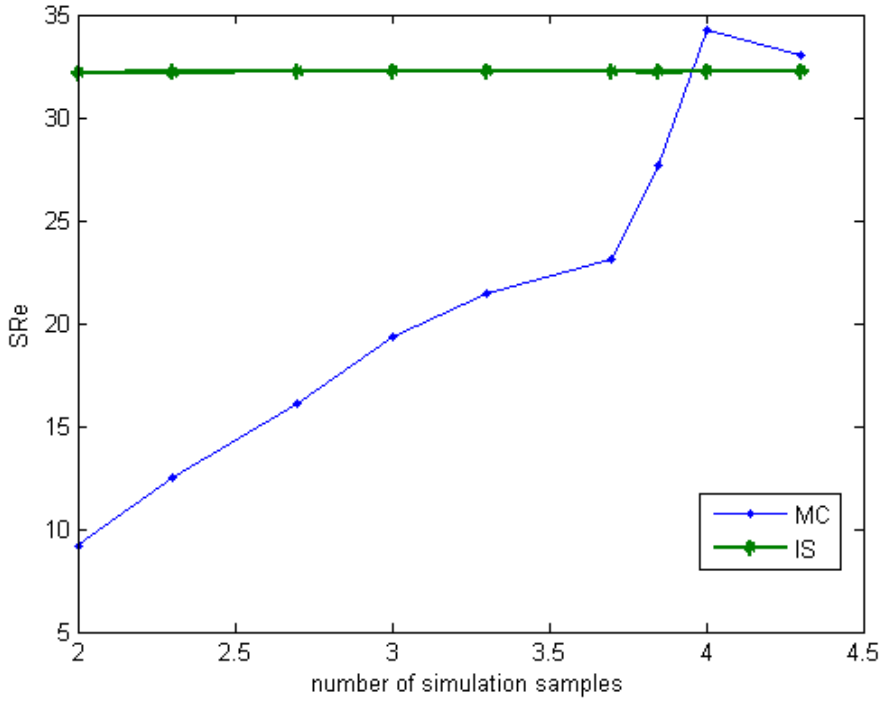

Figure 9 $\mathrm{SR}_{\lambda}^{\mathrm{e}}$, β=1, $\lambda = 1$

Here we give the second example for NCM, which have high underlying correlations. We need to employ the two-step importance sampling, and we would like to figure out that tail bound approximation would have significant influence the efficiency for the estimation. The factors are as follows:

1. Number of obligors m=100
2. Number of common risk factors d=10
3. The marginal default probabilities $p_i = 0.01 * (1 + \sin(8\pi i/m)),\ \ i = 1, \dots, m$
4. Size of exposures $c_i = ([5i/m])^2,\ \ i = 1, \dots, m$
5. The coupling coefficient A_{ik} is chosen independently and uniformly from the interval $(0, 1/\sqrt{d})$, where d=10; it ensure the sum of square A_{ik} for each i does not exceed 1, and A_{i0} for each i gets from (2.9).

The marginal default probabilities vary between 0 and 2% with mean 1%, and the possible exposures are 1, 4, 9, 16 and 25, with 20 obligors at each level.

From the parameters, we could take the initial guess at $x_{-1} = 0$ and $x_0 = \sum c_i = 1100$ for search method. The two-step importance sampling estimator (3.8) is given for Search Method; meanwhile in the Direct Method, we still keep the one step importance sampling estimator. We could see the simulation result as follows:

Table 10 α=95%

Number of Simulation N	Time of Simulation (in seconds)	Mean Value of VaR	Std. of VaR	Mean Value of ES	Std. of ES
	Plain Monte Carlo				
1000	9.5558	51.5892	8.2053	84.7453	15.0436
2000	18.6596	50.9873	6.1234	83.2881	12.7889
9100	74.8451	50.5432	3.2124	81.6184	9.6531
	Search Method				
2000(10*200)	378.0990	50.2235	1.4808	81.8296	1.9328
9100(13*700)	649.7052	50.0124	0.8634	81.8731	1.0308
	Direct Method				
1000	21.8399	51.3426	10.6483	82.8761	15.3150
2000	49.3452	50.8721	7.4392	82.2109	13.0013
9100	201.7592	50.4632	4.8243	81.9880	9.7453

Source: simulation data, by Matlab R2007a

We run a large number $N = 5 \times 10^6$ by plain Monte Carlo method. We use this large simulation result the 'true' value of VaR and ES. It takes 3 days and 21 hours to finish the simulation (337027.1620 sec). The 'true' value of VaR is 50 and ES is 81.9407. Evidently, the Direct

Method estimator doesn't show the good convergence property as in last model, because the one step IS method is employed here. The Search Method which apply the two-step IS estimator (3.8) cost too much computation time, but with quite good accuracy. It shows the theoretical validity of (3.8) for the model with high underlying correlation.

We could have a brief summery here. Compare with plain Monte Carlo method, importance sampling method exhibits a significantly improved convergence behavior, and the computation cost could be reduced by employ better algorithm for solving the tail bound approximation function (3.7) in two-step importance sampling for NCM.

Chapter 6
Concluding Remarks

6. 1 Conclusion

Evaluating the risk of financial positions is an important task for financial institutions and regulating authorities. The frequently used approach Monte Carlo method would cost too much time and computation resources when simulations focus in the rare events or the tail. We tailored two variance reduction methods to address the applicable way for improving the efficiency of Monte Carlo method.

We have developed, analyzed and tested these algorithms based on importance sampling procedure for estimating risk measures in two standard portfolio credit risk models CreditMetrics (Gupton etal., 1997) and CreditRisk plus (Cre, 1997). It was demonstrated that the variance reduction method exponential twisting, employed by Glasserman and Li (2003, 2005) for measuring probability over threshold x, could be extend to efficiently estimate risk measures (VaR, ES and utility based Shortfall Risk) in these models. Compare with plain Monte Carlo estimators, the importance sampling estimators, especially Direct Method estimators exhibit a significantly improved convergence behavior. This was illustrated by means of numerical examples.

Conclude with, variance reduction method exponential twisting may be extended to estimate risk measures that do not require so large sample size as plain Monte Carlo method.

6. 2 Discussion and Further Development

In the year 2008, the subprime crisis attacked the whole world's banks, funds, investment banks and etc. Three of the top five investment banks had been broken or bought by other commercial banks in US. The governments in Europe offered hundreds of thousands salvage money to those banks with a long standing reputation. The great depression also appeared in other industries. It is the biggest crisis after the 'Great Depression' which happened in 1930s. Everything began when the equities prices touched the roof, and the defaults of bonds, CDOs, CDSs and other equities credit derivatives. By put too many eggs in one bracket almost destroyed the world financial system. Many people criticized that it's the normal copulas which blew up the Wall Street. Felix Salmon published one essay on Wired Magazine in February 2009. He stated the normal copulas formula derived by David Li is the recipe of disaster.

Normal Copulas Model has its natural defect. It is a light tail model, and low dependence property between obligors has long academic distance from reality. But the attitude when using the model is much more important. What we have done with the Normal Copulas Model here is trying to give out an idea of accelerating the velocity of convergence. We may extend the variance reduction idea to the heavy tail models, such as t copulas model and a better algorithm to solve the tail bound approximation would be great promotion for the efficiency.

References

1. C. Gupton, C. Finger, and M. Bhatia *CreditMetrics Technical Document.* J. P. Morgan & Co., New York, 1997. www.riskmetrics.com.

2. *CreditRisk+: A CreditRisk Management Framework.* Credit Suisse Financial Products, London, 1997.

3. H. Föllmer and A. Schied *Convex and coherent risk measures.* October 2008

4. H. Föllmer and A. Schied *Stochastic Finance - An Introduction in Discrete Time.* Number 27 in de Gruyter Studies in Mathematics. Walter de Gruyter, 2 edition, 2004.

5. J. Dunkel, and S. Weber *Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models.* Winter Simulation Conference ISBN:1-4244-1306-0 Pages 958-966 Washington D.C. 2007

6. John C. Hull *Options, futures and other derivatives (4th ed.)*. Prentice Hall. ISBN 0-13-015822-4.

7. Panos M. Pardalos, and Vassilis K. Tsitsiringos *Financial engineering, E-commerce, and supply chain.* Kluwer Academic Publishers, pp.21-22,2002

8. P. Glasserman *Monte Carlo methods in financial engineering*. Springer-Verlag. ISBN 0-387-00451-3. 2003

9. P. Glasserman and J. Li *Importance sampling for a mixed Poisson model of portfolio credit risk.* In S. Chick, P. J. S´anchez, D. Ferrin, and D. J. Morrice, editors, Proceedings of the *2003 Winter Simulation Conference*, Piscataway (NJ), 2003. IEEE Press.

10. P. Glasserman and J. Li *Importance sampling for portfolio credit risk.* Management Science, 51:1643–1656, 2005.

11. P. Glasserman, P. Heidelberger, and P. Shahabuddin *Variance reduction techniques for estimating value-at-risk.* Management Science, 46(10):1349–1364, 2000a.

12. P. Glasserman, P. Heidelberger, and P. Shahabuddin *Importance sampling and stratification for value-at-risk.* In Y. S. Abu-Mostafa, B. Le Baron, A. W. Lo, and A.S. Weigend, editors, Computational Finance 1999, pages 7–24. MIT Press, 2000b.

13. P. Glasserman, P. Heidelberger, and P. Shahabuddin *Efficient Monte Carlo methods for valueat-risk. In Mastering Risk,* volume 2. Financial Times - Prentice Hall, 2001.

14. P. Glynn *Importance Sampling for Monte Carlo Estimation of Quantiles.* Publishing House of Saint Petersburg University,1996

15. P. Hoffman *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth.* New York: Hyperion, pp. 238-239, 1998.