University of London
Imperial College of Science, Technology and Medicine
Department of Mathematics

An application of Importance Sampling to the evaluation of the Economic capital

Inass El Harrak
CID number: 01287672

Thesis submitted as part of the requirements for the award of the MSc in Mathematics and Finance, Imperial College London, 2016-2017

Acknowledgements

My dissertation was made possible thanks to the opportunity to work in EBRD (European Bank for Reconstruction and Development).

I would like to express my special admiration and thanks to my advisor and mentor Piotr Karasinski; thank you for your priceless advice and guidance.

I would like to express next my gratitude to my supervisor, Dr. Antoine Jacquier; I thank you for your encouragements, patience and support throughout my year at Imperial College.

My time in EBRD wouldn't have been as pleasant as it was if not for Guglielmo De Maria and Kaitai Chan; many thanks for your company which was full of brilliant suggestions and advice that were of tremendous help to me.

I also want to thank the Quantitative Risk Analytics team who've made me feel welcome and who've answered my questions whenever needed.

Finally, no words can describe the gratitude I have for my parents and my sister, your endless sacrifices and support have always made me overcome obstacles.

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Signature and date:

Inass El Harrak

September 12^{th}, 2017

Abstract

My dissertation is on a new approach to evaluate the Economic Capital. A brief introduction of the latter is thus essential.

When financial institutions trade with counterparties, there is always a chance (a positive probability, however small) that these can default, failing to pay back the principal or the interests of the trade. Being regulated, banks are obliged to hold an amount of risk capital so that their balance sheets remain solvent over the measurement period. The Economic Capital is basically the amount of money needed to survive in the case of a worst scenario. This amount is assessed internally in Risk departments and covers all sorts of risks the bank may encounter. Amongst these risks: Credit Risk -which is the uncertainty that debtors will not honor their obligations. Market Risk -which is associated to fluctuations of the market prices of exchange rates, financial assets, commodities, etc. One can use quantitative skills to simulate scenarios which output different possible values of losses to have an idea of the risk one is facing. There are other risks, harder to simulate, a financial firm could encounter: Operational Risk which includes IT failures, human errors, fraud, wars... and Liquidity Risk, which is the risk an asset cannot be bought or sold quickly in the market, or the risk of not being able to acquire sufficient funding.

For simplicity, and to introduce the motivation behind the approach we're taking, we define the Economic Capital to be the Value at Risk or the Expected Shortfall at high confidence levels of a given portfolio. The natural way to estimate both risk measures is to do a simple Monte Carlo simulation to generate scenarios of losses and have an idea on their frequencies. The default probabilities of many banks' counterparties are very low (can reach the order of 0.01%), and most of the banks use a confidence measurement level of the order of 99.96% to 99.99%, which is the insolvency rate for an institution with an AA or AAA credit rating.

The motivation behind avoiding to use the classic Monte Carlo method stems from the fact that a confidence level of 99.99% -which is equivalent to a quantile of 1e-04, requires a big number

of simulations. If we were to use 1 million scenarios, only 100 scenarios will exceed the 99.99% level. However, we are mostly interested in what is happening beyond this threshold. Also, if the probability of default is 0.01%, scenarios that give the biggest values of the loss will be rare if not non-existent, whereas in real life, they could be reachable with very low probability. This will make our estimation of risk inaccurate.

The Importance Sampling, as we will detail in the first chapter, is a method used to simulate rare events. It will prove to be a very efficient method to reduce the variance of our estimator. Chapter 2 will detail the Normal Copula model (introduced by J.P. Morgans CreditMetrics system), which will be used to model the dependence between the obligors of a certain portfolio. It is based on modeling the dependence by stating they all depend on the same common factors. The first Importance Sampling procedure consists on changing the distribution of these factors in order to have a smaller variance. Since the dependence of our counterparties is expressed through these systematic factors, once we condition on them, the oligors become independent. The second Importance Sampling procedure will consist in changing the conditional (on the common factors) default probabilities. The next chapters show the gain of applying Importance Sampling on the accuracy of our estimator, and the last chapter shows that our method can be extended to more general cases.
My paper is mainly an implementation on the paper by Glasserman on Importance Sampling for Portfolio Credit Risk.

Notations

The notations below will be used throughout the paper.

$\Phi(x)$	The cumulative distribution function of a standard normal random variable
$\phi(x)$	The density of a standard normal distribution
$\mathbb{1}_{x \in A}$	Indicator function: equal to 1 if $x \in A$ and 0 otherwise
$\mathbf{E}_g[X]$	The expected value of the random variable X generated with the density function $g(x)$
$var(X)$	The variance of the random variable X
VaR	The Value at Risk of a given portfolio

Contents

Chapter 1

Importance Sampling

1.1 Motivation and Objectives

When we are estimating a variable, we are constantly looking for a way to reduce the variance of our estimator. That is in the aim of increasing the accuracy and the rate of convergence. Variance reduction methods are effective and easy to implement. Amongst the techniques used, we have: control variates, partial integration, systematic sampling, re-weighting and Importance Sampling.

Importance Sampling is the core of this paper. If the problem is to estimate:

$$\theta = \mathbf{E}_f[h(X)]$$

The idea is to change the estimator of θ, requiring a change of the density function of the random variable we're simulating. This change is efficient only if the new estimator has a lower second moment, in more details:

Let $P_X(.)$ and $Q_X(.)$ be two probability densities such that P is absolutely continuous with respect to Q (i.e. $P_X(x) = 0$ whenever $Q_X(x) = 0$) and such that $\mathbf{E}_{Q_X}\left[\frac{P_X}{Q_X}\right] = 1$, let X be a random variable under the density function P_X. The likelihood ratio $R = \frac{P_X}{Q_X}$ is thus well

defined. The idea behind IS is based on the following identity:

$$\mathbf{E}_{P_X}\left[h(X)\right] = \int h(x)P_X(dx) = \int h(x)\frac{dP_X}{dQ_X}(x)dP_X(dx) = \mathbf{E}_{Q_X}\left[h(X)R(X)\right].$$

This means that instead of generating X under P_X, we generate it under Q_X, and the change only requires a multiplication by a likelihood ratio.
The estimator will thus change from:

$$\widehat{\theta_n^P} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

where $(X_i)_{1\leq i\leq n}$ are generated using P_X, to:

$$\widehat{\theta_n^Q} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}$$

where $(X_i)_{1\leq i\leq n}$ are generated using Q_X.
This change is effective only if the variance is reduced, i.e.

$$var_{Q_X}\left[h(X)R(X)\right] < var_{P_X}\left[h(X)\right]$$

which is equivalent to:

$$\mathbf{E}_{Q_X}\left[h(X)^2R(X)^2\right] < \mathbf{E}_{P_X}\left[h(X)^2\right]$$

Importance Sampling becomes efficient in the cases where the events we are interested in are rare using the P_X density. One example of a rare event , as we stated in the introduction, is simulating a default indicator with very low default probability (the default indicator gives 1 on average every 10000 simulations for a default probability of 0.01%).

1.2 Some examples of likelihood ratios:

For future use, we provide two examples of likelihood ratios.

-First example:

The likelihood ratio for changing the density from a multivariate standard normal $\mathcal{N}(0, I)$ to a multivariate normal variable with mean μ and covariance matrix I, $\mathcal{N}(\mu, I)$:

$$R(x) = \frac{f(x)}{g(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^T x}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^T (x-\mu)}{2}}} = e^{\frac{\mu^T \mu}{2} - \mu^T x} \tag{1.1}$$

- Second example:

The likelihood ratio for changing a random variable Y with a Bernoulli distribution with probability p to a Bernoulli with probability :

$$p_\theta = \frac{p e^{c\theta}}{1 + p(e^{c\theta} - 1)} \tag{1.2}$$

The parameters θ and c are to be taken as deterministic constants for now, we will see later what they correspond to. Unlike the first example, this is a discrete distribution, the likelihood ratio will have two different outputs depending on the value of Y:

- $R(1) = \dfrac{p}{p_\theta}$
- $R(0) = \dfrac{1-p}{1-p_\theta}$

This can be put in one expression:

$$R(Y) = \left(\frac{p}{p_\theta}\right)^Y \left(\frac{1-p}{1-p_\theta}\right)^{1-Y} \tag{1.3}$$

In the case of N independent Bernoulli variables, $Y_1, ..., Y_N$, with probabilities $p_1, ..., p_N$, the likelihood ratio would be the product of the individual likelihood ratios, i.e.

$$\begin{aligned} R(Y_1, ..., Y_N) &= \prod_{i=1}^{N} \left(\frac{p_k}{p_{k,\theta}} \right)^{Y_k} \left(\frac{1-p_k}{1-p_{k,\theta}} \right)^{1-Y_k} \\ &= \frac{1 + p_k(e^{\theta c_k} - 1)}{e^{\theta c_k Y_k}} \\ &= exp(-\theta L + \Psi(\theta)) \end{aligned} \tag{1.4}$$

where :

$$\Psi(\theta) = \sum_{n=1}^{N} log\left(1 + p_k(e^{\theta c_k} - 1)\right) \tag{1.5}$$

Chapter 2

Normal Copula Default Model

2.1 Introduction to the model

Over a fixed horizon of time $[0, T]$, we would like to estimate the Value at Risk and the Expected Shortfall of a credit risk portfolio. To do so, one needs to simulate the loss of the portfolio.
Let N be the number of counterparties to which the portfolio is exposed. And let p_k denote the unconditional probability of default of the k^{th} obligor. These unconditional probabilities are inputs of the model and are linked to Credit Ratings (depending on the obligor's past history of borrowing and paying off debts and to its future economic potential), or to the market prices of corporate bonds or Credit Default Swaps.
The counterparties of a portfolio are usually dependent, which makes the problem more complex. A common way to express this dependence is to use a Normal Copula model, which we will introduce shortly.
The loss of a portfolio is the sum of losses over all its obligors. We simulate the default of counterparty k by a default indicator, which is a random variable defined as follows:

$$D_k = \begin{cases} 1 \text{ with probability } p_k \\ 0 \text{ with probability } 1 - p_k. \end{cases}$$

As a first approach, the exposures of the obligors will be supposed deterministic. We will discuss the case of the stochastic exposures later in paper. We denote c_k the exposure of the k^{th} obligor. Thus $c_k D_k$ is the loss due to the k^{th} obligor, and

$$L = \sum_{n=1}^{N} c_k D_k$$

is the loss of the portfolio at time T.

For the k^{th} counterparty, we introduce the Credit Worthiness index X_k defined as follows:

- X_k is a standard normal variable
- $D_k = \mathbb{1}_{\{X_k < t_k\}}$

t_k is the threshold that determines whether there's a default or not, it is linked to the unconditional default probability by:

$$P(Y_k = 1) = p_k = P(X_k < t_k) = \Phi(t_k) \tag{2.1}$$

Thus, at time T, the counterparty defaults when the value of that index falls below t_k:

$$X_k < \Phi^{-1}(p_k) = t_k$$

For example, if the probability of default is 1%, the credit index must fall below -2.32 for default.

The idea behind the Normal Copula model is to express the dependence between the default indicators by passing it to the correlation between the X_ks. We simplify the model by stating that there are d common factors that explain this dependence. Mathematically, this can be modeled by writing that the credit index for obligor k is the sum of contributions from d

systemic factors $F_1, ..., F_d$ and an idiosyncratic factor I_k:

$$X_k = \rho_{k,1}F_1 + ... + \rho_{k,d}F_d + a_k I_k \tag{2.2}$$

The idiosyncratic factor I_k, which is a standard normal variable is independent of $\{I_1, I_2, ..., I_{k-1}, I_{k+1}, ..., I_N\}$. This is typically an internal factor of default. We also know that the D_ks are correlated, they are all linked to economic variables such as stock market level and volatility, Non-Performing Loans (NPL) or other credit cycle indicators.

Our main assumption here is that the systematic and idiosyncratic factors, alongside the Credit Worthiness index are distributed according to the standard normal distribution and are uncorrelated as a first approach. Thus, in order to insure that the variance of the Credit Worthiness index is 1, we must have, for k=1,...,N:

$$\rho_{k,1}^2 + \rho_{k,2}^2 + ... + \rho_{k,N}^2 + a_k^2 = 1 \tag{2.3}$$

Since the common factors are supposed mutually independent, and are all independent from the idiosyncratic factor.

Also, the factors $\{\rho_1, ..., \rho_N\}$ will be supposed non-negative, which is a consequence of the positive correlation between the defaults of the obligors (in a financial crisis, a financial institution's default can cause other institutions' default).

We will see later that the common factors can be correlated, and in the real world, in fact, are.

The systematic factor has a correlation ρ_k with the credit index X_k.

Remark 1.

In the case of a one-factor model, simple calculation gives that the correlation between X_k and X_l is equal to $\rho_k \rho_l$.

2.2 Conditional default probabilities

From the expression (2.2), conditioning on the common factors, i.e. for a given value of $F_1, ..., F_d$, (we will denote $F = (F_1, ..., F_d)^T$ and $\rho_k = (\rho_{k,1}, ..., \rho_{k,d})$) the counterparty defaults if the idiosyncratic factor falls below a threshold:

$$I_k < \frac{N^{-1}(p_k) - \rho_k F}{a_k}. \tag{2.4}$$

Meaning that the conditional default probability of the obligor k is given by:

$$p_k(F) = p_k(Y_k = 1|F) = p_k(X_k < t_k|F) = N\left(\frac{N^{-1}(p_k) - \rho_k F}{a_k}\right) \tag{2.5}$$

Note that once we condition on the common factors, the default indicators become independent Bernoulli variables with probability given by (2.5).

Stopping to analyze this function is essential here. We want to see how the common factors change the conditional default probability. For this, we will plot the value of the systematic factor against the conditional probability.

In the case of a one-factor model, i.e. all the counterparties depend one only one factor, the conditional probability becomes:

$$p_k(F) = N\left(\frac{N^{-1}(p_k) - \rho_k F}{\sqrt{1 - \rho_k^2}}\right)$$

We plot the values of the common factor in the X-axis, and the values of the conditional default probability in the Y-axis.

We see from Figure 2.1 that when the normal variable takes negative values, the conditional default probabilities go up. This will motivate the Importance Sampling on the common factors, which we will detail in the next chapter.

In the case of two systematic factors, the conditional default probability is:

$$p_k(F_1, F_2) = N\left(\frac{N^{-1}(p_k) - \rho_{k,1}F_1 - \rho_{k,2}F_2}{\sqrt{1 - (\rho_{k,1}^2 + \rho_{k,2}^2)}}\right)$$

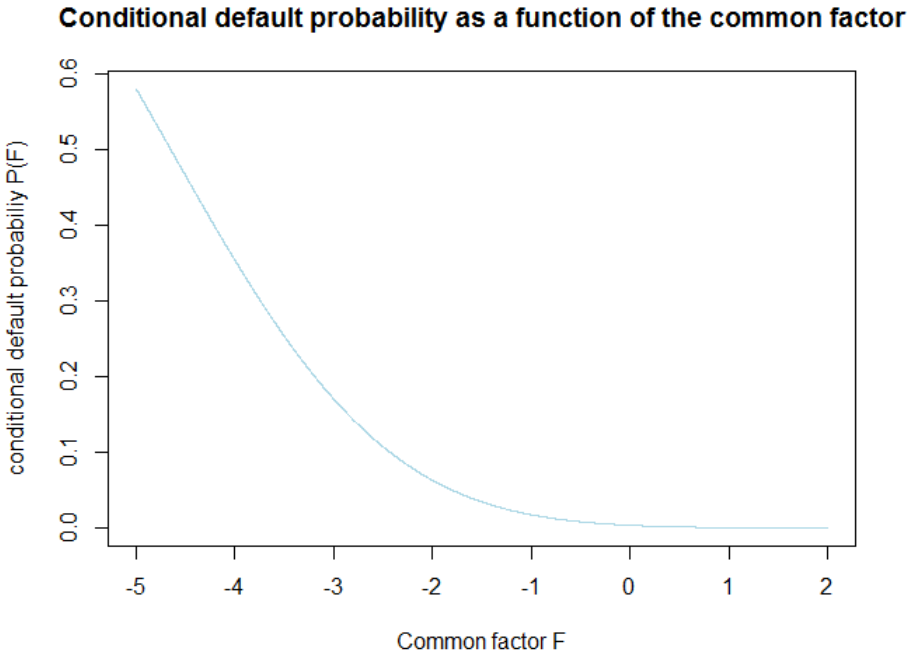

Figure 2.1: **The conditional default probability as a function of one systematic factor, unconditional default probability $p_k = 1\%$ and $\rho_k = 50\%$**

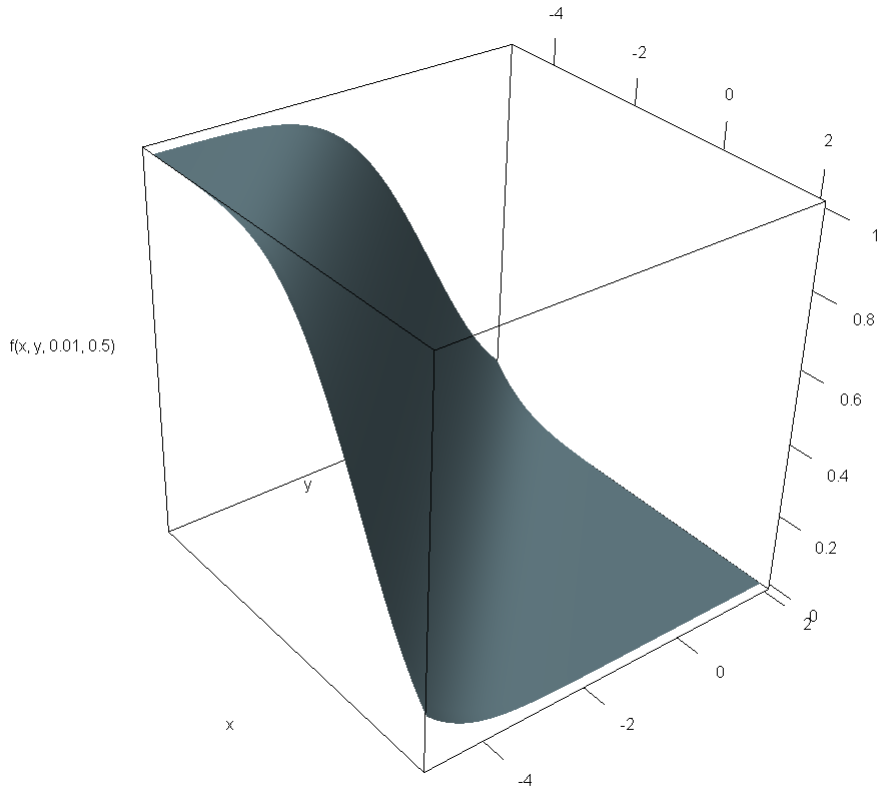

Figure 2.2: **The conditional default probability as a function of two independent systematic factors, unconditional default probability $p_k = 1\%$ and $\rho_k = 50\%$**

We plot the values of the first common factor in the X-axis, the second common factor in the Y-axis and the values of the conditional default probability in the Z-axis.

We see from figure 2.2 that since the figure the factor loadings corresponding to the common factors are similar, the figure is symmetric. And as before, the more negative the values of the common factors, the bigger the conditional probability gets.

Chapter 3

Importance Sampling on the Conditional Default Probabilities

3.1 The problematic

We want to draw scenarios of the loss in order to evaluate the two risk measures: the Value-at-Risk and the Expected Shortfall. To do so, we will need to estimate:

$$\theta_x = P(L > x) = \mathbf{E}[\mathbb{1}(L > x)].$$

We want to use an unbiased estimator $\widehat{\theta_x}$ that will have the optimal variance. The loss in our model is conditional on the common factors, thus the variance of our estimator is:

$$var(\widehat{\theta_x}) = var(\mathbf{E}[\widehat{\theta_x}|F]) + \mathbf{E}[var(\widehat{\theta_x}|F)].$$

Since the variance of any estimator is positive. We look to minimize both $var(\mathbf{E}[\widehat{\theta_x}|F])$ and $\mathbf{E}[var(\widehat{\theta_x}|F)]$.

3.2 Changing the Conditional Default Probabilities

We are now interested in minimizing $\mathbf{E}[var(\widehat{\theta_x}|F)]$. Now that the model is settled, we will move on to applying the Importance Sampling on the conditional default probabilities. Since these are small, making the event of the default very rare, we will be looking for a probability density that makes them bigger. Glasserman suggested a method called Exponential Twisting in his Importance Sampling paper, and the method is based on twisting the conditional default probabilities in the following way, given a common factor F:

$$p_{k,\theta}(F) = \frac{p_k(F)e^{\theta c_k}}{1 + p_k(F)(e^{\theta c_k} - 1)} \tag{3.1}$$

This twisting method was introduced in the first chapter, we have calculated the likelihood ratio (see equation (1.4)).
Equation (3.1) is dependent on the parameter θ, we will be choosing this parameter in a way to optimize the variance of our estimator, this will be detailed later on.

$$\frac{p_{k,\theta}(F)}{p_k(F)} = \frac{e^{\theta c_k}}{1 + p_k(F)(e^{\theta c_k} - 1)}$$

and

$$e^{\theta c_k} - \big(1 + p_k(F)(e^{\theta c_k} - 1)\big) = (1 - p_k(F))(e^{\theta c_k} - 1)$$

We see from these last two equations that the new twisted probability is bigger than the real one if $\theta > 0$, smaller if $\theta < 0$, and equal if $\theta = 0$

Another important remark is that when the default probabilities are equal to 1, they remain at 1 no matter the θ we choose. This remark is important because the banks usually model a defaulted counterparty by giving it a default probability equal to 1.

Remark. For real portfolios, the exposures are usually very high (in millions or billions). If we compute naively the twisted probabilities, we will see that the term $e^{\theta c}$ will be infinite in

all programming languages. A good method to tackle this issue is to scale everything by 10^{-9} and scale everything back at the end of the computations.

3.3 The optimal θ

In order to estimate the VaR and the Expected Shortfall of the estimator, we will be looking to estimate the value of $P(L > x)$. We will be concentrating our study on large values of thresholds x. Consequently, our main objective would be to choose a θ that minimizes the variance of this estimator, or simply that minimizes its second moment:

$$\theta_m = \arg\min_{\theta} M_2(\theta) = \arg\min_{\theta} \mathrm{E}\left[\mathbb{1}\{L > x\} e^{-2\theta L + 2\Psi(\theta)}\right] \tag{3.2}$$

We see from (3.2) that the optimal θ does not have a closed formula. We will, as done in [ ] optimize the upper bound of $M_2(\theta)$:

$$M_2(\theta) \leq e^{-2\theta x + 2\Psi(\theta)}$$

which is equivalent to maximizing the function :

$$f(\theta, x) = \theta x - \Psi(\theta)$$

We remind the expression of Ψ:

$$\Psi(\theta, F) = \sum_{n=1}^{N} log\left(1 + p_k(F)(e^{\theta c_k} - 1)\right)$$

The function Ψ is strictly convex as a function of θ. Its first two derivatives w.r.t. θ are:

$$\Psi^{'}(\theta, F) = \sum_{n=1}^{N} \frac{p_k(F) c_k e^{\theta c_k}}{1 + p_k(F)(e^{\theta c_k} - 1)} = \sum_{n=1}^{N} c_k p_{k,\theta}(F) \qquad \Psi^{''}(\theta, F) = \sum_{n=1}^{N} \frac{c_k e^{\theta c_k}(1 - p_k(F))}{(1 + p_k(F)(e^{\theta c_k} - 1))^2} > 0 \tag{3.3}$$

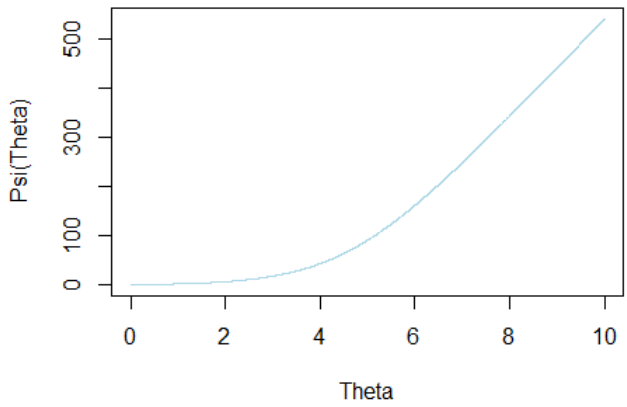

Figure 3.1: $\Psi(\theta)$ **for** $p_k = 1\%$ **and** $c_k = 1$ for all obligors

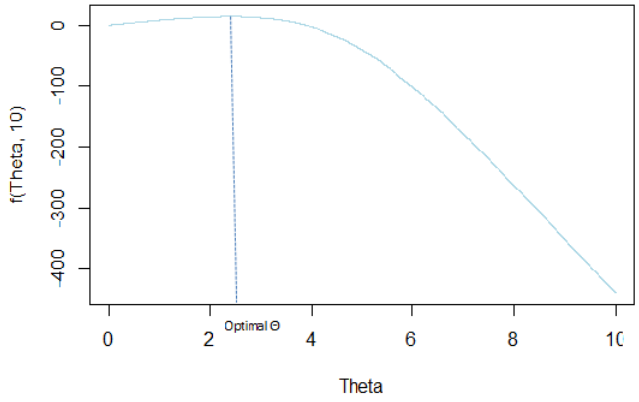

Figure 3.2: $f(\theta, 10)$ **for** $p_k = 1\%$ **and** $c_k = 1$ for all obligors

The optimal θ for this new approach has a closed form:

$$\theta_m = max\{0, \text{ solution to } \Psi^{'}(\theta) = x\}$$

We make sure that θ is positive, since a negative one would make the probabilities smaller. It is easy to see from the optimal value of θ and the expression giving the derivative of the Ψ function that:

$$\mathbf{E}_\theta[L] = \sum_{n=1}^{N} c_k p_{k,\theta} = \Psi^{'}(\theta_m) = x$$

This means that the Exponential Twisting shifts the distribution of the loss, so that its new mean becomes a high value of a threshold of our choice.

We see from the histogram that the effect of the exponentially twisting is indeed shifting the

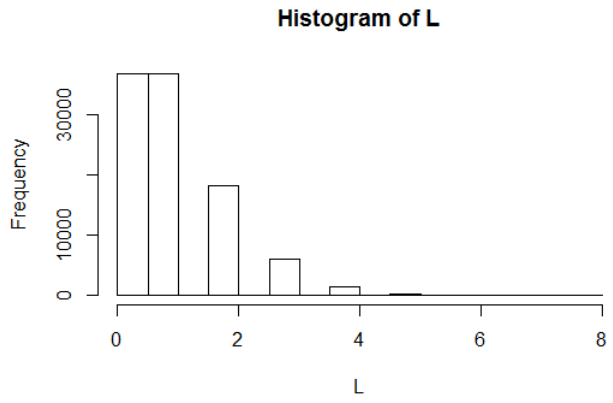

Figure 3.3: **Histogram of the loss simulated with real default probabilities** $p_k = 1\%$ **and** $c_k = 1$ **for all** 100 **obligors and for** 10^5 **simulations**

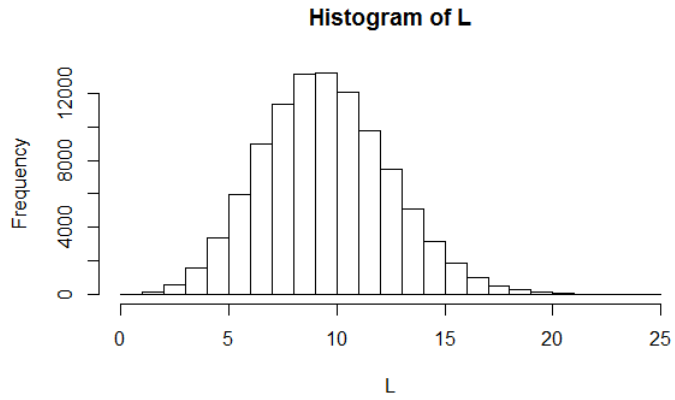

Figure 3.4: **Histogram of the loss simulated with exponentially twisted default probabilities by the optimal** θ **for** $p_k = 1\%$**,** $c_k = 1$**, threshold** $= 10$ **for all** 100 **obligors and for** 10^5 **simulations**

loss to a threshold of our choice. In Figure 5, we have chosen a threshold = 10, and we see that the mean of the loss under the new distribution is indeed around 10. This means that the threshold will be an input of our implementation, thus has to be chosen in an optimal way.

The motivation behind using the exponential twisting method stems from the fact that it gives the asymptotic optimal estimator. We will not detail this here, more details can be found in section 3.2 of [1].

3.4 The optimal threshold

Since we will be interested in calculating the VaR and the Expected Shortfall of a portfolio at high levels, we would want to shift the loss distribution to where the VaR is. Indeed, if the loss

is now distributed around the VaR, calculating the Expected Shortfall becomes more accurate. However, this means that we have a prior idea of what the approximate value of the VaR is. In simple portfolios, one can have a vague idea, but in diversified ones, one has to calculate.. One way to do this is to do a Brute Monte Carlo method with a relatively small number of simulations. This is will give an order of the threshold one has to choose.

Chapter 4

Shifting the common factors

4.1 Motivation

If we look now at the random variable $\mathbf{E}[\widehat{\theta}_x|F]$, we want to apply Importance Sampling to change the distribution the common factors are drawn from in the aim of reducing the variance of our estimator. If we remember the figures 2.1 and 2.2, an intuitive way would be to shift the mean of the common factors to negative values if the factor loadings are positive. In our model, the common factors are standard normal, which means that there is approximately a 70% chance that the common factors lie in $[-1, 1]$. This makes the conditional default probabilities low, making the event of default rare, especially if correlations between default indicators are high.

In order to tackle this issue, we will shift the mean of the common factors towards negative values. Basically change the distribution of F from $\mathcal{N}(0, I)$ to $\mathcal{N}(\mu, I)$.

Glasserman in [1] chapter 5.1 gives several methods to optimize the μ, these are only approximations to the optimal value and require to look for the optimal vector that maximizes or minimizes a certain function.

As an example, *the Tail bound approximation* defined in [1] page 12 is the commonly used

method, because it is easily applicable to our IS. The method states that:

$$\mu = \max_f\{F(f,x)\} = \max_f\{-\theta_x(f)x + \Psi(\theta_x(f),f) - \frac{1}{2}z^Tz\}$$

However, since the optimization of μ is only approximate, and in some cases, as we will see in the case of a real portfolio, the dimension of the common factors vector can be high (more than 20 for some real cases). The optimization problem becomes tricky and time consuming. This makes the objective of this paper, which is to have the optimal accuracy for the run time aimed less attractive.

4.2 Simulation steps

Once the value of μ is set up, the steps to follow to estimate the estimator are the following, for each simulation:

- Generate d independent $\mathcal{N}(\mu,1)$ for the common factors $F = (F_1,..,F_d)$
- Calculate the conditional default probabilities $p_k(Z)$ for all $k \in \{1,..,N\}$
- Calculate the optimal θ and deduce $p_{k,\theta}$ for all $k \in \{1,..,N\}$
- Calculate $\Psi(\theta,F)$
- Generate N independent Bernoulli random variables with probability $p_{k,\theta}$: Y_k for all $k \in \{1,..,N\}$
- Calculate the loss $L^{(i)} = \sum_{i=1}^{N} c_kY_k$
- Calculate the likelihood ratio for each i^{th} simulation :

$$R^{(i)} = e^{(\frac{1}{2}\mu^T\mu - \mu^T*F)} - \theta * L^{(i)} + \Psi(\theta,F))$$

- Sort the $L^{(i)}$ in a decreasing order
- For each $L^{(i)}$, the estimator for $P(L > L^{(i)})$ is:

$$\hat{P}(L > L^{(i)}) = \frac{1}{NS} \sum_{i=1}^{NS} \mathbb{1}\{L^{(i)} > x\} R^{(i)} \ ^{(*)}$$

$^{(*)}$ NS stands for the number of simulations

Since the $L^{(i)}$ are sorted in a decreasing order, a quick way to estimate $\hat{P}(L > L^{(i)})$ is to do a cumulative sum of the likelihood ratios over the indexes smaller than i.

We introduce next an example of a portfolio we will be working on. For simplicity, we will be using a one-factor homogeneous portfolio with 1000 obligors. The portfolio is defined such that for all $k \in \{1, .., 1000\}$:

- $p_k = 0.01$
- $c_k = 1$
- threshold $= 80$
- Number of simulations $= 10^4$
- $\rho_k = 30\%$

We use the *Tail bound approximation* to get the optimal μ, we obtain $\mu = -0.98$.
In order to see the importance of shifting the common factor, we estimate the variance of the $99.99\% VaR$ for $\mu \in \{-2.5, -0.98, 0\}$, we use the unbiased estimator of the variance with 100 different seeds:

$$\widehat{\sigma^2_{100}} = \frac{1}{99} \sum_{i=1}^{100} \left(VaR_{99.99\%,i} - \frac{1}{100} \sum_{j=1}^{100} VaR_{99.99\%,j} \right)^2$$

We have the following results:

Value of μ	-2.5	-0.98	0
Variance of the estimator	0.6138963	2.69635	20.66016

The table above confirms the gain we have from shifting the common factor. When $\mu = 0$ (no shifting), the variance is about 20, we can reduce it by a factor of 8 using the optimal $\mu = -0.98$. However, we notice that when $\mu = -2.5$, the variance is further reduced, this stems from the fact that the optimization we did is only an approximation, and that the range from -2.5 to -1 give approximately the same outputs for the function $F(f, x)$(see figure 4.1). That is why we have decided to choose the common factors mean using a heuristic method based on a simulation approach.

When we are looking at the Expected Shortfall, we want to estimate the tail of the loss distri-

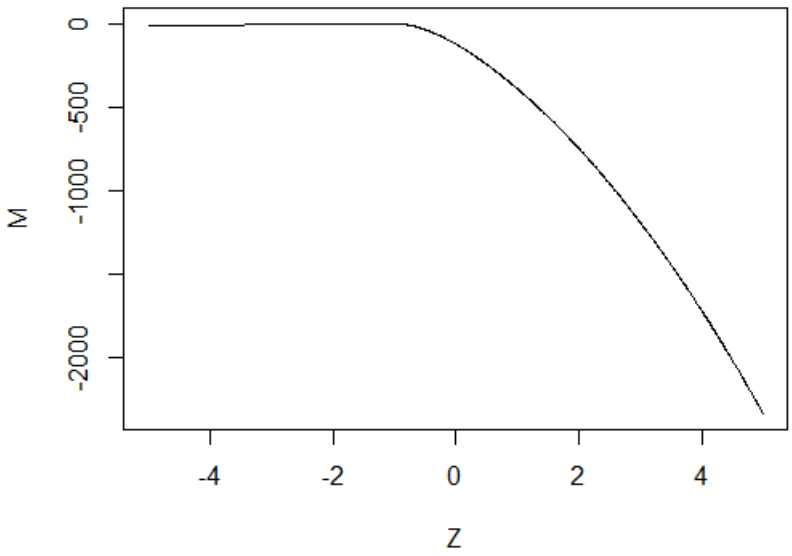

Figure 4.1: **Function F(f, 80) against f**

bution accurately. We look at the convergence of the probability of default plotting $P(L > x)$ against x for the portfolio above.

Remark 2. The 99.99% VaR corresponds to $log(10^{-4}) \simeq -9.2$ in figure 4.2

In an equivalent way, if we were to shift the common factors to positive values, the variance of the estimator would be bigger than the one with $\mu = 0$ (see equation (2.5)).

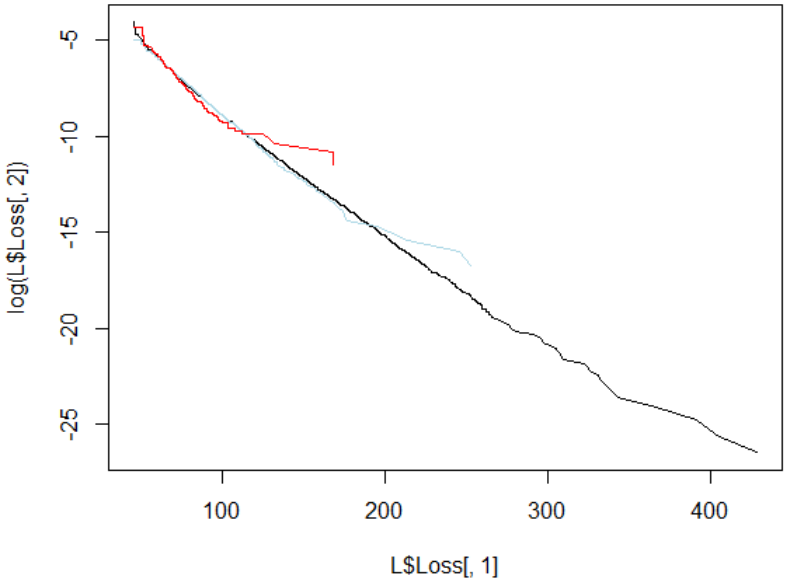

Figure 4.2: **The convergence of the tail probability, black curve is for $\mu = -2.5$, blue for $\mu = -1$ and red for $\mu = 0$**

Chapter 5

The marginals

In order to evaluate the economic capital, we will mainly be working on the two famous risk measures; the Value-at-Risk and the Expected Shortfall.
The overall loss of a portfolio is the sum of the losses over all its counterparties. Thus, we will be interested in estimating the contribution of each counterparty in the overall risk amount, and use this to estimate the error of our estimator.
This type of decomposition is used for capital allocation and for measuring risk-adjusted performance.

5.0.1 Reminder on the risk measures

- The Value-at-Risk

  The most prominent loss-distribution-based risk measure in Risk Management is the value-at- risk. Formally, it is defined as follows:

  Let $\alpha \in (0, 1)$, The VaR of loss L at a confidence level α is given by:

$$VaR_\alpha(L) := inf\{x \in \mathbb{R} : P(L > x) \leq (1 - \alpha)\}$$

  Intuitively, for a large probability α, it is the level such that the probability that the loss exceeds this level is less or equal to $1 - \alpha$, which is a small probability.

One of the limitations of the VaR is that it does not see the severity of losses and tail risk", and that it isn't subadditive, meaning that the sum of the VARs of two portfolios may be less than the VAR for the combined portfolio.

In order to address the shortcomings of the VaR, we introduce a new risk measure:

- The expected shortfall:

  Let $\alpha \in (0,1)$, the expected shortfall of loss L at confidence level α is defined by:

$$ES_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 q_u(L)du$$

  The ES is related to the VaR by :

$$ES_\alpha(L) =:= \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du$$

  Remark

  If L has continuous cdf F_L and $\mathbb{E}[max\{L,0\}] < \infty$, then, for any $\alpha \in (0,1)$:

$$ES_\alpha(L) = \frac{\mathbb{E}[L\mathbb{1}_{L\geq VaR_\alpha(L)}]}{1-\alpha} = \mathbb{E}[L|L \geq VaR_\alpha(L)]$$

  The expected shortfall is a coherent measure, it is subadditive and sensitive to the severity of losses.

5.1 Marginal expected shortfalls

For the k^{th} obligor, we define the marginal expected shortfall to be:

$$ES_{k,\alpha}(L) = \mathbb{E}[c_k Y_k | L \geq VaR_\alpha(L)]$$

So that the overall expected shortfall is:

$$\sum_{k=1}^N ES_{k,\alpha}(L) = \sum_{k=1}^N \mathbb{E}[c_k Y_k | L \geq VaR_\alpha(L)] = \mathbb{E}[\sum_{k=1}^N c_k Y_k | L \geq VaR_\alpha(L)] = ES_\alpha(L)$$

Another way to estimate the error of our estimation would be to calculate the error of the marginals. Meaning that we will be looking at the outputs of a homogeneous portfolio, which are supposed to be the same, and estimate the error from the difference we obtain.
Using the IS, the estimator for the k^{th} marginal expected shortfall is:

$$\widehat{ES_{k,\alpha}}(L) = \frac{\sum_{i=1}^{NS} c_k Y_k^{(i)} \mathbb{1}_{L^{(i)} \geq VaR_\alpha(L^{(i)})}}{\sum_{k=1}^{NS} \mathbb{1}_{L^{(i)} \geq VaR_\alpha(L^{(i)})}}$$

Consider the following portfolio of 100 obligors (same as the one in [3]):

- $p_k = 0.01$ for all $k \in \{1, ..., 100\}$
- $\rho_k = 30\%$
- $$c_k = \begin{cases} 1 \text{ for } k = 1..20 \\ 4 \text{ for } k = 21..40 \\ 9 \text{ for } k = 41..60 \\ 16 \text{ for } k = 61..80 \\ 25 \text{ for } k = 81..100 \end{cases}$$

Each 20 obligors are supposed to have the same marginal contribution to the overall Expected Shortfall since they have the same inputs. Yet, we see from Figure 4.2 that it is not the case, especially for the Monte Carlo case. Set the error to be:

$$err(\hat{X}) = \frac{\max\limits_{1,...N} X_i - \min\limits_{1,...N} X_i}{\frac{1}{N}\sum_{i=1}^{N} X_i} * 100$$

For 10^5 simulations, we calculate the error for the last 20 identical obligors:

Brute Monte Carlo	58.27506%
Importance Sampling	4.168058%

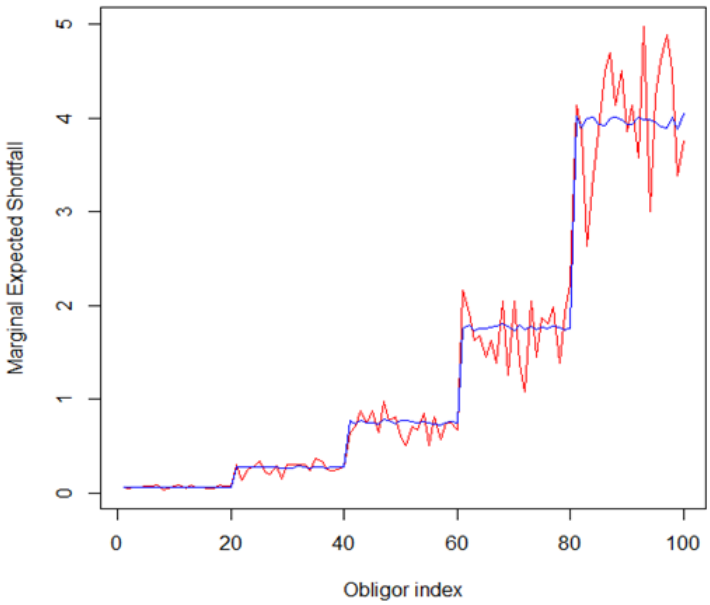

Figure 5.1: **Marginal Expected Shortfall, red curve is for Brute Monte Carlo, blue for IS**

We see that the gain is very important for a relatively small number of simulations (approximately 12 times better!).

Chapter 6

Some generalizations

6.1 The case of guarantors

This section is to show that IS can be applied for a more general case, where there is a presence of guarantors.

For example, when we are looking at a portfolio of loans, these can be guaranteed. In this case, default happens when both the counterparty and the guarantor default.

We have introduced the normal copula model in chapter 2 where we have defined the credit worthiness index X_k. In the case of a guarantor, the loss on the loan will be dependent on more than one credit worthiness index. For the counterparty k, we will set X_{k_c} to be the credit worthiness index for the default of the counterparty, and we will set X_{k_g} to be credit worthiness index for the default of the guarantor. In this setting, the probability of losing the exposure of the k^{th} obligor with probability:

$$P(D_{k_g} = 1) = P(\{X_{k_c} < x_{k_c}\} \cap \{X_{k_g} < x_{k_g}\})$$

The loss now becomes:

$$L = \sum_{i=1}^{N} c_{k_g} D_{k_g}$$

Note that the exposure c_k changes to c_{k_g} because the LGD (Loss Given Default) changes in the case of a guarantor. Set LGD_{k_c} and LGD_{k_g} to be the LGDs corresponding to the k^{th} counterparty and guarantor respectively. The overall LGD becomes:

$$LGD_k = LGD_{k_c} LGD_{k_g}$$

If we want to apply IS on the case of guarantors, we need to express the correlation between the k^{th} obligor and its counterparty (these are not independent, because the default of the counterparty can cause a systematic risk, causing the insurer to default too, creating a domino effect).

For simplicity, we will be expressing this dependence by the same common factors of the different counterparties, i.e.

$$X_{k_c} = \rho_{k,1} F_1 + ... + \rho_{k,d} F_d + a_k I_{k_c}$$

$$X_{k_g} = \beta_{k,1} F_1 + ... + \beta_{k,d} F_d + b_k I_{k_c}$$

Denote $\rho := (\rho_1, ..., \rho_d)$ and $\beta := (\beta_1, ..., \beta_d)$ Thus, the conditional default probability for the k^{th} obligor given the factor loadings $F = (F1, ..., Fd)^T$ is:

$$\begin{aligned} P(D_{k_g} = 1|F) &= P(\{X_{k_c} < x_{k_c}\} \cap \{X_{k_g} < x_{k_g}\}|F) \\ &= P(\{\rho_{k,1} F_1 + ... + \rho_{k,d} F_d + a_k I_{k_c} < x_{k_c}\} \cap \{\beta_{k,1} F_1 + ... + \beta_{k,d} F_d + b_k I_{k_c} < x_{k_g}\}|F) \\ &= N\left(\frac{x_{k_c} - \rho_k F}{a_k}\right) N\left(\frac{x_{k_g} - \beta_k F}{b_k}\right) \end{aligned}$$

The last equality stems from the fact that once we condition on the common factors, the events $\{X_{k_c} < x_{k_c}\}$ and $\{X_{k_g} < x_{k_g}\}$ become independent.

Consequently, the only difference here is that the conditional default probabilities change, the IS method described before follows easily.

The figure below shows the difference between Brute Monte Carlo and applying Importance Sampling in the case of presence of obligors. As before, we see that the Monte Carlo doesn't

capture big values of losses, and it starts diverging at low probabilities (here $e^{-9} \approx 10^{-4}$)

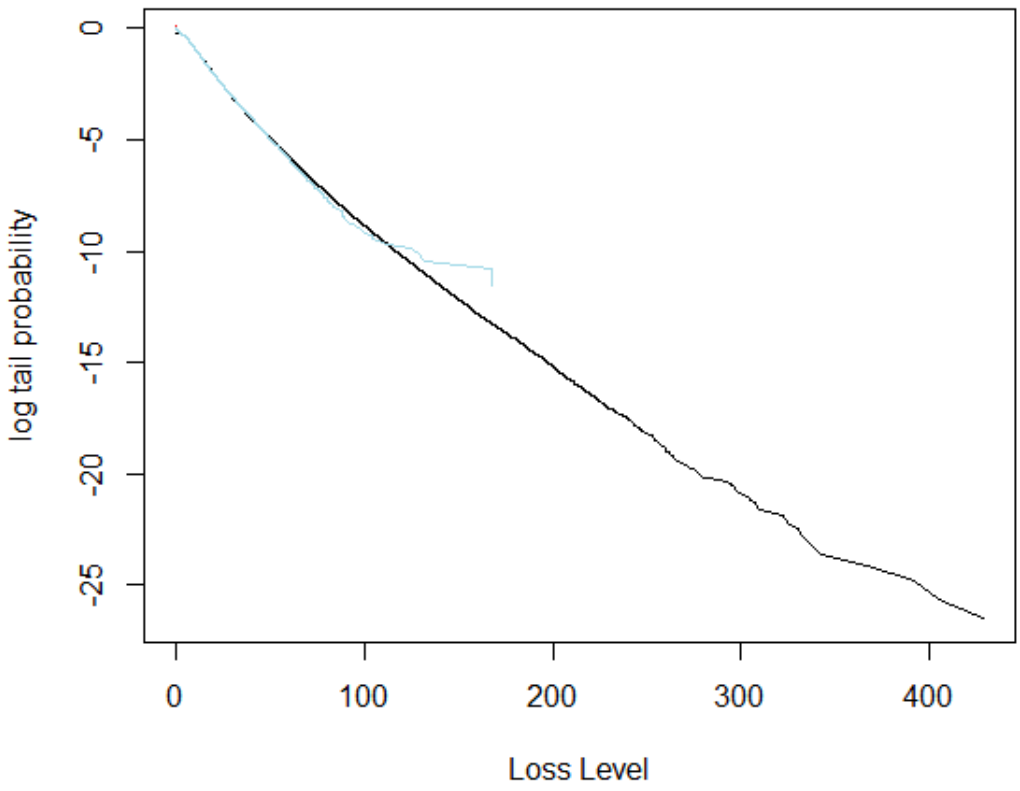

Figure 6.1: **Comparison between Monte Carlo(blue curve) and Importance Sampling (black curve) in the case of 1000 obligors and 1000 identical guarantors**

6.2 The case where the common factors are dependent

As a numerical example, we introduce the case of a real Banking portfolio. The inputs are the default probabilities ranging from 0.01% to 1 (the obligor has defaulted), and the exposures taking into account the LGDs.

For simplicity and without loss of generality, we suppose that there are no guarantors.

There are two main common factors, one common factor of the region of the counterparty, and the second one is the sector of the counterparty. The model we apply is the following:

$$X_k = \rho(a_{k,R}F_R + a_{k,S}F_S) + \sqrt{1-\rho^2}I_k \tag{6.1}$$

where :

- $F_R = (F_{R,1}, ..., F_{R,NR})$ and $F_S = (F_{S,1}, ..., F_{S,NS})$ (NR and NS are the number od regions and sectors respectively that our portfolio handles)

- $a_{k,R}$ is a row vector of dimension NR which has only one non-null component, the non-null element will be equal to $\frac{2}{\sqrt{5}}$

- $a_{k,S}$ is a row vector of dimension NS which has only one non-null component, the non-null element will be equal to $\frac{1}{\sqrt{5}}$

- ρ expresses the dependence of the counterparties on the common factors, in our model, $\rho = 0.68$ for all the obligors.

- F_R is a multi-normal vector with correlation matrix Σ_R (The different regions are correlated between them)

- F_S is a multi-normal vector with correlation matrix Σ_S (The different sectors are correlated between them)

In our example, there are 10 regions and 13 sectors. This makes the dimension of the common factors equal to 23.

As we have mentioned before, the optimization problem for shifting the common factors becomes heavy, since it is a maximization over a high-dimension vector. The idea now is to minimize the dimension of the optimization.

We will be using the Principal Components Analysis to reduce the dimension from 23 to, for example, 2 dimensions. In oder to generate the standard multi-normal vectors for the region and sector common factors, we will avoid using the Cholesky decomposition, because the latter method doesn't give information on the principal directions of the common factors.

Let:

$$\Sigma_R = U_R D_R U_R^T$$

and:

$$\Sigma_S = U_S D_S U_S^T$$

The common factors are thus generated as follows:

$$F_R = \left(U_R^T D_R^{\frac{1}{2}}\right) Z_R$$

$$F_S = \left(U_S^T D_S^{\frac{1}{2}}\right) Z_S$$

Z_R and Z_S are two independent standard multi-normal variables (with dimension NR and NS respectively).

Once we have done this, the PCA approach tells us that the principal directions are the eigen vectors matching the biggest eigen values. So we will be shifting only the components of the eigen vectors with the biggest eigen vectors. If we want to reduce our dimension to 2, then we will be shifting the principal component of the region factor and that of the sector factor.

For our code to be clear and easy to read, we will adjust the inputs so that the common factors are always generated independently. Equation (6.1) becomes:

$$X_k = \rho(a_{k,R}\left(U_R^T D_R^{\frac{1}{2}}\right) Z_R + a_{k,S}\left(U_S^T D_S^{\frac{1}{2}}\right) Z_S) + \sqrt{1-\rho^2} I_k$$

The difference here is that the components of Z_R and Z_S are mutually independent, which simplifies the code.

The IS follows easily once the above is settled.

For a real case numerical example, we apply a simple Brute Monte Carlo using low number of simulations. We have an approximate value of the Value-at-Risk, here, around 4 millions. We ask our code to shift the loss distribution so that most of the simulations are centered around 4 millions:

If the Monte Carlo value of the Value-at-Risk isn't very accurate, for example, 3 millions, we would have asked our program to shift the loss so that more outcomes are around 3 millions. This indeed is not optimal, yet, it has still a big gain compared to Monte Carlo:

If we calculate the variance σ^2 of our estimator (evaluated for 100 values of the Value-at-Risk

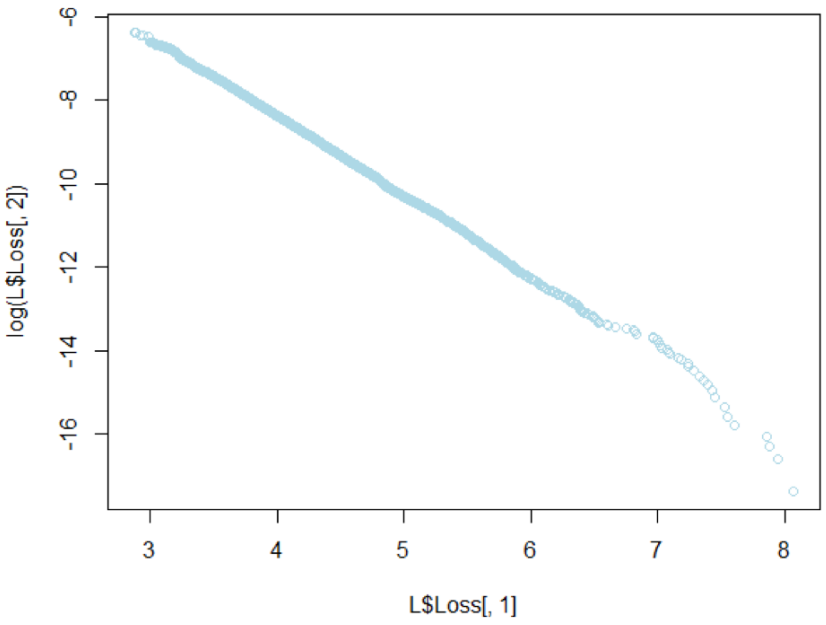

Figure 6.2: **Tail of Probability of the Loss centered around 4 for 1 million simulations**

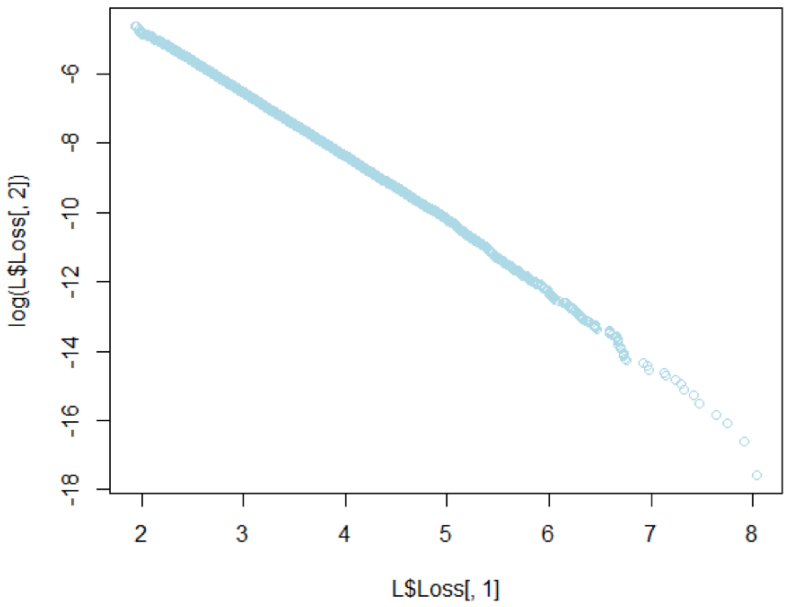

Figure 6.3: **Tail of Probability of the Loss centered around 3 for 1 million simulations**

using 100 different seeds) and set the error to be:

$$err(\widehat{\theta}) = \frac{3\sigma}{VaR_1}$$

We get as numerically for the threshold 4 a value around 1%, which is relatively a good error.

Bibliography

[1] Paul Glasserman and Jingyi Li, *Importance Sampling for Portfolio Credit Risk*, Columbia Business School December 2003; revised July 2004, February 2005

[2] Paul Glasserman, Wanmo Kang and Perwez Shahabuddin, *Fast Simulation of Multifactor Portfolio Credit Risk*, July 2005, revised January 2007

[3] Paul Glasserman *Fast Simulation of Multifactor Portfolio Credit Risk*, 403 Uris Hall, Columbia Business School, New York, NY 10027, USA

[4] Gosia Wasiewicz *Internal presentation, Economic Model for Banking Book*, Quantitative Risk Analytics, EBRD

[5] Guglielmo De Maria *Monte Carlo Simulation for the evaluation of the Economic Capital*, MSc thesis Imperial College

[6] Jonathan Goodman - *Lecture Notes on Monte Carlo Methods, Chapter 3: Variance Reduction*, Courant Institute of Mathematical Sciences, NYU

[7] Glasserman, Heidelberger, and Shahabuddin *Asymptotically optimal importance sampling and stratification for pricing path-dependent options, Mathematical Finance 9:117152*