Importance Sampling for Portfolio Credit Risk in Factor Copula Models

Parrini, Alessandro

Vrije Universiteit Amsterdam

24 August 2013

Master's Thesis in
Econometrics and Mathematical Economics

Importance Sampling for Portfolio Credit Risk in Factor Copula Models

August 24, 2013

Author:
Alessandro Parrini[1]

Supervisors:
Dr. L. Hoogerheide
Dr. A.A.N. Ridder

[1]Student number: 2517549, e-mail: a.parrini@student.vu.nl

Abstract

This work considers the problem of the estimation of Value at Risk contributions in a portfolio of credits. Each risk contribution is the conditional expected loss of an obligor, given a large loss of the full portfolio. This rare-event framework makes it difficult to obtain accurate and stable estimations via standard Monte Carlo methods. Moreover, the factor copula models employed to capture the dependence among obligors, poses an additional challenge to this problem. Such a challenge is tackled by applying the importance sampling algorithms developed in Glasserman and Li (2005). The aim of this study is to improve the importance sampling scheme currently in use at the Royal Bank of Scotland, where this project has been carried out[2]. By conveniently modifying Glasserman and Li's methods, we develop importance sampling schemes which lead to significant variance reduction, both in single and multi-factor models.

Keywords: Monte Carlo Methods, Importance Sampling, Value-at-Risk, Portfolio Credit Risk, Gaussian Copula Models.

[2]I am greatful to all the members of the Economic Capital Modeling team, for their help and precious advises. Special thanks go to Denis for always pointing me in the right direction and for providing an important support to this work.
I also acknowledge my accademic supervisors for their valuable comments and suggestions.
Finally, a warm thanks and hug to all the friends I had the pleasure to meet this year at the "Fu".

Contents

Preface

In financial industry firms, *Economic Capital* (EC) acts as a buffer to absorb large unexpected losses, thereby protecting depositors and claim-holders, as well as providing confidence to external investor and rating agencies on the financial health of the firm. In contrast with *Regulatory Capital*, which refers to the minimum capital requirements which institutions are required to hold[3], economic capital is so called because it measures risk in terms of *economic* realities. In fact, it is designed to cover all the risks (market, credit, operational etc.) that may force a financial firm into insolvency. Once these risks have been properly modeled and aggregated, a firm determines the probability of default that is acceptable, in line with its target financial strength (e.g. credit rating)[4]. While most of the concepts and methodologies developed have broader applicability, we focus on *economic credit capital*, the buffer against those risk associated with obligor credit events such as default, credit migrations (downgrades or upgrades) and credit spread changes.

Economic capital is quantified to absorb *unexpected losses* at a specified confidence level, while credit reserves are set aside to cover *expected losses*. Therefore, EC is typically defined as the difference between the portfolio's Value at risk (VaR) and the expected loss of the portfolio. The VaR level (i.e., quantile) is chosen in a way that trades off providing a high return on capital for shareholders with protecting debt holders and maintaining a desired credit rating.

The measurement of economic credit capital is often just a first step in a more extensive process of allocating capital among the various components of the credit portfolio. This process requires decomposing the total credit risk in a portfolio into individual risk contributions, called *Economic Capital contributions* (ECC). This is vital for several management decisions such as risk based compensations, pricing, profitability assessment, as well as building optimal risk-return portfolios and strategies. Given the common definition of EC, it is natural to allocate capital based on the portfolio components' contributions to VaR.

The ECC associated with VaR can be represented as conditional expectations of

[3]For example, regulations for internationally active banks are given in the *Basel Accord*, framework created by the Basel Committee on Banking Supervisions ([3]).

[4]For instance, a firm that capitalizes to Moody's Aa standard over a one-year horizon determines its economic capital as the "cushion" required to keep the firm solvent over a one-year period with 99.97% probability. In fact, firms rated Aa by Moody's have historically defaulted with a 0.03% frequency over a one-year horizon.

individual losses, given that the full portfolio loss equals the quantile at level α, typically for high values of α. The rarity of these events presents an obstacle to the practical calculation of these conditional expectations. In fact, each contribution depends on the probability of a rare event (an obligor's default) conditional on an even rarer event (an extreme loss of the whole portfolio).
These features pose a particular challenge to the computation of ECC via standard Monte Carlo simulations, and motivate research on methods to accelerate simulations through variance reduction.
Various numerical tools have been proposed to improve the quality of simulation-based risk contributions. The most appealing one, in a rare-event framework, is importance sampling (IS), which can generate a greater number of observations from the tail of the loss distribution. However, capturing dependence among obligors adds complexity to the econometric models employed, and does not easily lend itself to rare-event simulations techniques used in other settings. Throughout this work, we capture the dependence across sources of credit risk by employing the widely used copula factor models (Gupton et al. 1997[14]; Li 2000[24]).
Importance sampling procedures for rare-event simulations for credit risk measurement have been proposed by a number of authors. In the following, we review the most relevant literature.
Kalkbrener et al. (2004)[20] introduce an IS algorithm to calculate expected shortfall (ES) of credit portfolios modeled by means of Merton-type threshold models (Merton 1974[29]). Since defaults occur when the firms' returns fall below a certain threshold, they propose to shift the factors mean by adding a negative number in order to generate more scenarios with a sufficiently high number of defaults.
Morokoff (2004)[30] proposes to increase the variance of the systemic factors instead of shifting their mean. The framework is again the one introduced by Merton, but the quantity of interest is here the quantile of the portfolio loss distribution. Yet, Morokoff's method implicitly assumes that the credit portfolio is roughly *homogeneous*, i.e. the portfolio loss is the sum of a large number of similar random losses. This assumption is far from being realistic, and makes this approach less applicable in practice.
Other applications of importance sampling in this context include Merino and Nyfeler (2004)[28] and, more recently, Huang et al. (2010)[18]. All these methods are however largely heuristic, in the sense that the authors lack in providing a theoretical support and a rigorous analysis of the effectiveness of the proposed algorithms. Conversely, this aspect is extensively covered in most of the works by Glasserman et al. (2005[9], 2007[11] and 2008[12]).
Glasserman and Li (2005)[9] develop and analyze a *two-step* IS estimator in the context of factor copula models. In order to generate scenarios with a higher number of losses, first they shift the mean of the systemic factors, and then they increase the default probabilities conditional on these factors by means of an exponential change of measure. This procedure is shown to be asymptotically optimal for the estimation of the tail of the portfolio's loss distribution. Numerical results show a variance reduction (compared to standard MC) which remarkably increases with the

loss level.
The asymptotical optimality of the two-step estimator in Glasserman and Li (2005) is provided only for the single-factor homogeneous model. Glasserman et al. (2008)[12] develop an alternative approach which lends itself more readily to the selection of multiple mean shifts for multi-factor heterogeneous models[5]. However, the numerical results show that for large quantiles, a single shift of the common factor mean outperforms the multiple mean shifts. In the light of these considerations, we will base our research study mainly on the work of Glasserman and Li (2005).

This project has been carried out during a graduation internship at the Royal Bank of Scotland (Economic Capital Modeling Team, Amsterdam, The Netherlands). At present, the bank uses an importance sampling scheme for the ECC estimation which relies only on a shift in the systemic factor's mean. The goal of this study is to identify and to overcome the weaknesses of the current scheme. In fact, while it is not difficult to obtain reliable ECC for frequently defaulting portfolios, this task is much harder for high quality credit portfolios. We thus aim at improving the accuracy of the estimator with a particular focus on obligors with low probability of default. To address this challenge, we will attempt to apply the approach proposed in Glasserman and Li (2005) to the problem at hand. Throughout this work, however, the non-optimality of this approach in our context will become apparent. Nevertheless, we will employ tailor-made modifications which will lead to a notable improvement of the current IS procedure.

The paper is structured as follows. Chapter 1 introduces the theoretical basis of the work, namely general importance sampling and Merton-type factor models. Chapter 2 describes the importance sampling scheme for portfolio credit risk developed by Glasserman and Li. Finally, Chapter 3 presents our simulation study. Here, the proposed IS algorithms are discussed and their performance are tested with several numerical examples.

[5]This paper is built on the insights gleaned from the analysis of Glasserman et al. (2007)[11] who investigate rare event asymptotics for the loss distribution in the Gaussian copula model. Their analysis reveals a qualitative distinction between two cases: in the rare-default regime, the tail of the loss distribution decreases exponentially, but with increasingly high loss thresholds the decay follows a power law. This suggests that the dependence between defaults imposed by the Gaussian copula is qualitatively different for portfolios of high-quality and lower-quality credits.

Chapter 1

Theoretical Background

In this chapter we present the theoretical foundations which have been necessary milestones for the development and the understanding of this work. First, a general introduction to importance sampling is provided. Next, an introduction to copulas and a detailed description of factor models is presented.

1.1 Importance Sampling

One of the fundamental methods for increasing the efficiency of Monte Carlo simulations is *importance sampling*. Importance sampling attempts to reduce estimators' variance by changing the sampling distribution to give more weight to "important" samples. This method often leads to a dramatic variance reduction, in particular when estimating *rare event* probabilities.

To fix ideas, let

$$\ell = \mathbb{E}_f\left[h(X)\right] = \int h(x)f(x)dx, \tag{1.1}$$

where x is a random variable in $\mathbb{R}^d$ with probability density function (pdf) f, $h : \mathbb{R}^d \to \mathbb{R}$, and $\mathbb{E}_f[\cdot]$ indicates that the expectation is taken with respect to f (called *nominal pdf*).
The standard (or *crude*) Monte Carlo estimator (CMC) is

$$\hat{\ell}_{CMC} = \frac{1}{n}\sum_{i=1}^{n} h(X_i). \tag{1.2}$$

with $X_1, \ldots, X_n$ independent draws from f.
Let g be another density function on $\mathbb{R}^d$ which is not dominated by f, i.e.

$$f(x) > 0 \Rightarrow g(x) > 0, \quad \forall\, x \in \mathbb{R}^d$$

Then, if $X_1, \ldots, X_n \overset{iid}{\sim} g$, we can write (1.1) as

$$\ell = \mathbb{E}_g\left[h(X)\frac{f(X)}{g(X)}\right], \tag{1.3}$$

where g is referred to as *importance density*. The importance sampling (IS) estimator is thus,

$$\hat{\ell}_{IS} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)\gamma(X_i), \tag{1.4}$$

where the ratio of densities $\gamma(X_i) := f(X_i)/g(X_i)$ is called *likelihood ratio* (LR) or *Radon-Nikodym derivative*[1].
Clearly, in the case where there is no change of measure, i.e. $g \equiv f$, we have $w \equiv 1$, and the importance sampling estimator reduces to the standard Monte Carlo estimator.

From (1.3) and (1.4) it follows that $\mathbb{E}_g[\hat{\ell}_{IS}] = \ell$ and thus that $\hat{\ell}_{IS}$ is unbiased. Hence, to compare the variances of the estimator with and without importance sampling, we only need to compare the second moments. With importance sampling we have

$$\mathbb{E}_g\left[\left(h(X)\frac{f(X)}{g(X)}\right)^2\right] = \mathbb{E}_f\left[h(X)^2\frac{f(X)}{g(X)}\right]. \tag{1.5}$$

Depending on the choice of g this could result smaller (or even infinitely larger!) than $\mathbb{E}_f\left[h(X)^2\right]$. Successful importance sampling schemes are crucially linked to the selection of an effective importance density.

Minimum variance density

From (1.5) it is clear that in order to obtain variance reduction, we should look for an appropriate pdf g for which $\mathbb{E}_f\left[h(X)^2\,\gamma(X)\right]$ is minimal. If we consider the problem of minimizing the variance of $\hat{\ell}_{IS}$ with respect to g, that is

$$\min_g \mathbb{V}ar_g\left[h(X)\frac{f(X)}{g(X)}\right], \tag{1.6}$$

it is not difficult to show (for instance, Rubinstein and Melamed[33]) that the solution of this problem is

$$g^*(x) = \frac{|h(x)|f(x)}{\int |h(x)|f(x)dx}. \tag{1.7}$$

In particular, consider the case in which h is nonnegative, then the optimal importance sampling density is $g^*(x) = h(x)f(x)/\ell$, and

$$\mathbb{V}ar_{g^*}[\hat{\ell}_{IS}] = \mathbb{V}ar_{g^*}\left[h(x)\frac{f(x)}{g^*(x)}\right] = \mathbb{V}ar_{g^*}[\ell] = 0.$$

[1] This name is due to the fact that the *change of measure* in (1.3) is justified by the well-known result in measure theory introduced by Radon (1913) and then proved in its general case by Nikodym (1930).
The name likelihood ratio is a slight abuse of nomenclature, since the likelihood is usually seen in statistics as a function of the parameters. In spite of that, we will adopt this term since it is most commonly used.

Thence, in this case the IS estimator provides a *zero-variance* estimator. An obvious practical difficulty for the evaluation of g^* is that it depends on the target ℓ. Nevertheless, the optimal IS density provide some useful guidance on the choice of g: a good importance sampling density g should be proportional to the product of h and f.

In the case that h is an indicator function of a set, then the optimal importance density is the nominal density f conditioned on the set. Suppose $h(x) = 1\{X \in B\}$, for some $B \subset \mathbb{R}^d$. Then $\ell = \mathbb{P}(X \in B)$, and the optimal IS density is the conditional density of X, given $X \in B$ (assuming $\ell > 0$). It follows that, to estimate a probability, one should look for an importance sampling pdf that approximates the conditional pdf. In other words, one should choose g to make the event $\{X \in B\}$ more likely, especially if B is a rare set under f.

Importance Sampling pitfalls

One of the main considerations for choosing a good importance sampling density is that the IS estimator (1.4) should have finite variance. This is equivalent to require that

$$\mathbb{E}_g \left[\left(h(X)^2 \frac{f(X)^2}{g(X)^2} \right) \right] = \mathbb{E}_f \left[h(X)^2 \frac{f(X)}{g(X)} \right] < \infty. \tag{1.8}$$

This indicates that g should not be lighter-tailed than f and that the likelihood ratio should be bounded. Equation (1.8) also suggests that wherever $h(x)^2$ is large, the likelihood ratio should be small. That is, when $h(X)^2$ is large we should sample more often X under g than under f. However, such shifts of probability density (also called *biasing*) may lead to bad estimates. For simple problems (where an analytic solution is available) it is easily seen that shifting too much density towards the importance set causes many observations to have too small original probabilities: this in turn results in too small likelihood ratios which lead to underestimate ℓ. This is sometimes called *over-biasing*; conversely, if the shift does not give enough weight to the set of interest, the sampled observations are too few (*under-biasing*).
A somewhat more serious issue may also arise if the proposed density g concentrates too much density in a small part of the support of X. In such cases, the original density may be much larger in a big part of the support and this may blow up the variance because of the squaring of the likelihood ratio.
In the light of these considerations, it should be clear that the choice of the importance density is a very delicate issue.

Another importance sampling weakness is represented by a somewhat pathological behavior of the LR as the dimensionality d of X increases. In fact, the distribution of $\gamma(X)$ under g may become increasingly skewed, taking values close to 0 with high probability, but also taking very large values with a small but significant probability. As a consequence, the variance of $\gamma(X)$ under g may become very large for large d.
Suppose the components X_i of $X = (X_1, \ldots, X_d)$ are iid under both f and g and suppose the marginal pdfs of each X_i are f_1 and g_1 respectively. The likelihood

ratio, can thus be written as

$$\gamma(X) = \prod_{i=1}^{d} \frac{f_1(X_i)}{g_1(X_i)} = \exp\left\{\sum_{i=1}^{d} \log \frac{f_1(X_i)}{g_1(X_i)}\right\} \approx \exp(d\mu), \tag{1.9}$$

where $\frac{1}{d}\sum_{i=1}^{d} \log \frac{f_1(X_i)}{g_1(X_i)} \approx \mathbb{E}_g\left[\log \frac{f_1(X)}{g_1(X)}\right] \equiv \mu$ by applying the (strong) law of large numbers (SLLN). By Jensen's inequality we have

$$\mu \leq \mathbb{E}_g\left[\log \frac{f_1(X)}{g_1(X)}\right] = \log \int \frac{f(x)}{g(x)} g(x)dx = 0,$$

with strict inequality (if $f_1(X_i)/g_1(X_i)$ is not a fixed constant) being the logarithm strictly concave. But if $\mu < 0$, this implies

$$\sum_{i=1}^{d} \log \frac{f_1(X_i)}{g_1(X_i)} \longrightarrow -\infty,$$

and thus, $\gamma(x) \approx \mathrm{e}^{d\mu} \xrightarrow{d\to\infty} 0$, under g. Thus, in high-dimensional problems the likelihood ratio converges to zero albeit its expectation is 1 for all d, by definition. This again confirm that the variance of $\hat{\ell}_{IS}$ can be very large if the change of measure is not chosen carefully[2].

Remark. Since likelihood ratios are often skewed, the sample standard deviation of $\hat{\ell}_{IS}$ typically underestimates the true standard deviation. Therefore, a very large sample size may be needed for confidence intervals based on central limit theorem to provide a reliable coverage.

1.1.1 Exponential Change of Measure

Let $X_1, \ldots, X_n \overset{iid}{\sim} f$, with f being a univariate standard normal pdf. If we apply a change of measure that preserves the independence of the X_i but changes their common density to $g = N(\mu, 1)$, then the corresponding likelihood ratio is

$$\prod_{i=1}^{n} \frac{f(X_i)}{g(X_i)} = \exp\left(-\mu \sum_{i=1}^{n} X_i + \frac{n}{2}\mu^2\right). \tag{1.10}$$

And, if we let g_i have mean μ_i,

$$\prod_{i=1}^{n} \frac{f(X_i)}{g_i(X_i)} = \exp\left(-\sum_{i=1}^{n} (\mu_i X_i + \frac{1}{2}\mu_i^2)\right). \tag{1.11}$$

This simple example is a special case of a more general class of convenient measure transformation, called *exponential change of measure* (ECM).

[2] In order to prevent this degeneracy of the likelihood ratio, a solution could be to reduce the dimension of X by one of the numerous dimension-reduction techniques.

Let X be a random variable (or vector) with pdf $f(x;\theta)$, where θ is an m-dimensional parameter vector. X is said to belong to an *exponential family* if there exists real-valued functions $t_i(x)$, $h(x) > 0$ and a normalizing function $c(\theta)$, such that

$$f(x;\theta) = c(\theta)\mathrm{e}^{\theta \cdot t(x)} h(x),$$

where $\theta \cdot t(x)$ is the inner product $\sum_{i=1}^{m} t_i(x)\theta_i$. Note that the representation of an exponential family is generally not unique.
Consider the univariate, single-parameter case with $t(x) = x$. The resulting family of densities $\{f(x;\theta), \theta \in \Theta \subset \mathbb{R}\}$ is given by

$$f(x;\theta) = c(\theta)\mathrm{e}^{\theta x} h(x). \tag{1.12}$$

If $h(x)$ is a pdf, then $c^{-1}(\theta)$ is the corresponding *moment generating function*:

$$c^{-1}(\theta) = \int \mathrm{e}^{\theta x} h(x) dx. \tag{1.13}$$

It is convenient to introduce the logarithm of the moment generating function (*cumulant generating function*):

$$\psi(\theta) := \log \int_{-\infty}^{+\infty} \mathrm{e}^{\theta x} h(x) dx. \tag{1.14}$$

Thus (1.12) can be written in the following convenient form:

$$f(x;\theta) = \mathrm{e}^{\theta x - \psi(\theta)} h(x).$$

Starting from any pdf f_0 we can easily generate an exponential family of the form (1.12) in the following way. Let $\Theta = \{\theta : \psi(\theta) < \infty\}$ and suppose that Θ is nonempty. Now define

$$f(x;\theta) = \mathrm{e}^{\theta x - \psi(\theta)} f_0(x). \tag{1.15}$$

Then $\{f(\cdot;\theta), \theta \in \Theta\}$ is an exponential family. The transformation from f_0 to $f(x;\theta) \equiv f_\theta$ is referred to as exponential twist or exponential change of measure with a twisting or tilting parameter θ. The likelihood ratio for such a transformation is

$$\prod_{i=1}^{n} \frac{f_0(X_i)}{f_\theta(X_i)} = \exp\left(-\theta \sum_{i=1}^{n} X_i + n\psi(\theta)\right). \tag{1.16}$$

Clearly, when $\theta = 0$, we recover the original pdf. In the standard normal case, $\psi(\theta) = \theta^2/2$, which confirms that this generalizes (1.10).
An important feature of ECM is that the likelihood ratio (which is in principle a function of all $X_1, \ldots, X_n$) reduces to a function of the sum of the X_i, i.e. to a *sufficient statistic* for θ.

The cumulative generating function ψ has some noteworthy properties. Let us denote by $\mathbb{E}_\theta[\cdot]$ and $\mathbb{E}_0[\cdot]$ the expectation under f_θ and f_0 respectively. Then,

$\psi(\theta)' = \mathbb{E}_\theta[X]$. To see this, note first that $\psi(\theta) = \log \mathbb{E}_0[\exp(\theta X)]$. Differentiating, we get

$$\psi(\theta)' = \frac{\mathbb{E}_0\left[X\mathrm{e}^{\theta X}\right]}{\mathbb{E}_0\left[\mathrm{e}^{\theta X}\right]} = \mathbb{E}_0\left[X\exp\left(\theta X - \psi(\theta)\right)\right] = \mathbb{E}_\theta[X].$$

A similar calculation shows that $\psi(\theta)'' = \mathbb{V}ar_\theta[X]$ which also implies that $\psi(\theta)'$ is increasing in θ. Finally, the function ψ passes through the origin, and Hölder inequality shows that it is strictly convex.

1.2 Copula Factor Models

1.2.1 An introduction to Copulas

If we assumed that the risk factors underlying our credit portfolio were mutually independent, we could easily study any problem associated with the portfolio. However, this assumption is far from being realistic. Clearly, the default rate of a group of obligors tends to be higher during a recession and lower when the economy is in a good state. In other words, different credit risks are subject to the same set of macroeconomic factors, and some positive dependence among these risks is typically observed.

In order to introduce a correlation structure into the portfolio, we need to specify the joint distribution of the random vector describing our (credit) risk factors. Generally speaking, knowing the joint distribution of random variables, allows us to derive the marginal distributions and the correlation structure, but the vice versa is not true. Among the many different techniques which permit to specify a joint distribution given marginals and dependence structure, copula function is a simple and convenient approach. First, copulas help in the understanding of dependence at a deeper level. Indeed, they express dependence on a quantile scale, which is natural in a risk-management context and helps to describe the dependence of extreme outcomes. Furthermore, copulas facilitate a bottom-up approach to multivariate model building, and this is particularly useful when we have a better idea about marginal behavior of individual risk factors than we do about their dependence structure[3].

The use of the Gaussian copulas to model defaults' correlation was first proposed by Li (2000)[24]. In what follows, a brief introduction to the concept of copulas is provided.

[3]This is typically the case in a credit risk context, where the individual default risk of an obligor is available (although not easy to estimate), while the dependence among defaults is hard to handle.

Basic properties. The word "copula" is a Latin noun that means "link"; in fact, a copula is a function that links univariate marginals to their multivariate distribution. A formal definition is provided below.

Definition 1 (Copula)**.** *A d-dimensional copula is a cumulative distribution function (cdf) on* $[0,1]^d$ *such that all the marginal distributions are standard uniform.*

We use the notation $C(u) = C(u_1, \ldots, u_d)$ for multivariate cdfs that are copulas. Thence C is a mapping of the form $C : [0,1]^d \to [0,1]$ with the following three properties.

1. $C(u_1, \ldots, u_d)$ is increasing in each component u_i.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i \ \forall i \in \{1, \ldots, d\}, u_i \in [0,1]$.
3. For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leq b_i$,

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j \ \forall j \in \{1, \ldots, d\}$.

The first property is clearly required for any multivariate cdf, and the second for uniform marginal distributions. The third property is the so called *rectangle inequality* and ensures that, if the random vector U has cdf C, $\mathbb{P}(a_1 \leq U_1 \leq b_1, \ldots, a_d \leq U_1 \leq b_d) \geq 0$.
Notice also that, for $2 \leq k < d$, the k-dimensional marginals of a d-dimensional copula are themselves copulas.

Let us introduce the following two lemmas, which constitute the basis for the construction of a copula.
Let X be a random variable with cdf F . Since F is a nondecreasing function, the inverse function F^{-1} may be defined as $F^{-1}(y) = \inf\{x : F(x) \geq y\}$.

Lemma 1. *Let U be an uniform random variable in (0,1) and let F be a distribution function. Then the random variable* $X = F^{-1}(U)$ *has distribution F.*

Lemma 2. *Let X be a random variable with continuous distribution F. Then* $F(X) \sim U(0,1)$.

Note that Lemma 1 is the key to stochastic simulations: if we can generate an uniform variate U and compute the inverse of a cdf F^{-1}, then we can sample from that cdf[4].

[4]This method for generating one-dimensional random variables is called *Inverse-Transform Method*.

Sklar's Theorem. The importance of copulas in the study of multivariate distributions is summarized by the following well-known theorem.

Theorem 1 (Sklar 1959)**.** *Let F be a joint cdf with marginals $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $X_1, \ldots, X_d$ in $\bar{\mathbb{R}} = [-\infty, +\infty]$,*

$$F(X_1, \ldots, X_d) = C(F_1(X_1), \ldots, F_d(X_d)). \tag{1.17}$$

If the marginals are all continuous, then the copula is unique. Otherwise, C is uniquely determined on Ran $F_1 \times$ Ran $F_2 \times \cdots \times$ Ran F_d, where Ran $F_i = F_i(\bar{\mathbb{R}})$ denotes the range of F_i.
Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distributions, then the function F in (1.17) is a joint distribution function with marginals $F_1, ..., F_d$.

The existence of a copula can be easily shown by using Lemma 1 and the definition of copula. The uniqueness was established by Sklar (1959)[37]. He proved that evaluating (1.17) at the arguments $x_i = F^{-1}(u_i)$, $0 \le u_i \le 1$, $i = 1, \ldots, d$, we obtain

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)), \tag{1.18}$$

which gives an explicit representation of C in terms of F and its marginals.
Notice that formula (1.17) shows how joint distributions are formed by *coupling* marginal distributions with copulas C, and formula (1.18) how copulas are extracted from multivariate cdfs. Moreover, (1.18) highlights how copulas express dependence on a quantile scale.

The concept of copula is slightly less natural than the concept of multivariate distribution because there is more than one copula which can be used to join marginal distributions. However, copulas have some very useful properties that makes them very convenient to represent the dependence structure of continuous marginals.

Proposition 1 (Invariance)**.** *Let $(X_1, \ldots, X_d)$ be a random vector with continuous marginals and copula C and let $T_1, ..., T_d$ be increasing functions. Then $(T_1(X_1), ..., T_d(X_d))$ also has copula C.*

Theorem 2 (Fréchet–Hoeffding Bounds)**.** *For every copula $C(u_1, \ldots, u_d)$ we have the bounds*

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \le C(u) \le \min\{u_1, \ldots, u_d\}. \tag{1.19}$$

Notice that Fréchet–Hoeffding bounds hold for any multivariate distribution function.

Fundamental Copulas. In this paragraph we provide some examples of copulas which represent special dependence structure. Next, the widely used Gaussian copula is presented. This is the copula function used in CreditMetrics (although not explicitly), and it will also be adopted throughout most of this thesis work.

The *independence copula* is given by

$$\Pi(u_1, \dots, u_d) := \prod_{i=1}^{d} u_i. \tag{1.20}$$

It is clear from (1.17) that random variables (with continuous distribution) are independent if and only if their dependence structure is given by (1.20).

The *comonotonicity copula* is the Fréchet upper bound in (1.19), i.e.

$$M(u_1, \dots, u_d) := \min\{u_1, \dots, u_d\}. \tag{1.21}$$

Notice that this copula is the joint distribution of the uniform random vector $(U, \dots, U)$ where $U \sim (0, 1)$.

Finally, the *countermonotonicity copula* is the two-dimension Fréchet lower bound from (1.19):

$$W(u_1, u_2) := \max\{u_1 + u_2 - 1, 0\}. \tag{1.22}$$

This copula is the joint cdf of the uniform random vector $(U, 1 - U)$.

Gaussian copula. If $Y \sim N_d(\mu, \Sigma)$, then its copula is the so called *Gaussian copula*. By the invariance property it follows that this copula is the same as the copula of $X \sim N_d(0, P)$, with P correlation matrix of Y. We thus have

$$C_P^{Ga}(u) := \mathbb{P}(\Phi(X_1) \leq u_1, \dots, \Phi(X_d) \leq u_d) = \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)), \tag{1.23}$$

where $\Phi(\cdot)$ denotes the standard normal cdf and $\Phi_P(\cdot)$ denotes the joint cdf of X. The notation C_P^{Ga} emphasizes that the copula is parametrized by the $\frac{1}{2}d(d-1)$ parameters of the correlation matrix.
The Gaussian copula does not have a closed form, but it is possible to express it as an integral over the density of X which, in the two-dimensional case, is given by

$$C_\rho^{Ga} = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{\frac{-(s_1^2 - 2\rho s_1 s_2 + s_2^2)}{2(1-\rho^2)}\right\} ds_1 ds_2.$$

If $P = I_d$ we obtain the independence copula as a special case; if $P = J_d$, i.e. the $d \times d$ matrix consisting entirely of ones, then we obtain the comonotonicity copula. The countermonotonicity copula is obtained setting the two off-diagonal elements of P to $\rho = -1$ (and $d = 2$). Thus the bivariate Gaussian copula can be thought of as a dependence structure that interpolates between perfect positive and negative dependence, with ρ representing the strength of the dependence.

Simulation of Gaussian copula. In general, it is fairly easy to simulate from any copula, provided that we can sample from the distribution from which the copula is extracted. In fact, if we can generate a vector X with distribution function F, we can transform each component with its own marginal cdf to obtain a vector $U = (U_1, \dots, U_d)' = (F_1(X_1), \dots, F_d(X_d))'$ whose cdf is C, the copula of F.

Algorithm 1. (Simulation of a Gaussian copula).

1. *Generate* $Z \sim N_d(0, P)$*;*
2. *Return* $U = (\Phi(Z_1), \dots, \Phi(Z_d))'$*.*

The random vector $U = (\Phi(Z_1), \dots, \Phi(Z_d))'$ has cdf C_P^{Ga}.
To generate normal variates in step 1, modern scientific software libraries use the *Ziggurat method*, a special case of the acceptance-rejection method which is extremely fast in comparison to the classical methods (see, for instance, Marsaglia and Tsang, 2000[26]).

Similarly to the way we extract a copula from the multivariate normal distribution, we can extract a copula from any other distribution with continuous marginals. As an example, the d-dimensional t copula takes the form

$$C_{\nu,P}^t(u) = t_{\nu,P}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)), \tag{1.24}$$

where t_ν is the cdf of a standard univariate Student-t distribution and $t_{\nu,P}$ is the joint cdf of the vector $X \sim t_d(0, P, \nu)$ where P is the correlation matrix of X[5].
Finally, note that the converse statement of Sklar's Theorem provides a very useful technique to construct multivariate cdfs with arbitrary marginals and copulas. This in fact permits to separate the description of the marginal behavior of individual risk factors and that of their dependence structure. Starting with a copula C and marginal distributions $F_1, \dots, F_d$, then $F(x) := C(F_1(x_1), \dots, F_d(x_d))$ defines a multivariate cdf with marginals $F_1, \dots, F_d$. We can thus build a distribution with say a Gaussian copula C_P^{Ga} but with arbitrary margins[6]; such a model is known as *meta-Gaussian* distribution.
Further examinations of copulas can be found in Nelsen (1999)[32] and McNeil et al. (2005)[27].

[5] In contrast with the Gaussian case, if $P = I_d$ we do not obtain the independence copula. This is due to the fact that uncorrelated multivariate t random variables are not independent. However, if $P = J_d$ we do obtain the comonotonicity copula.

[6] In credit risk modeling, for instance, the Gaussian copula is used to join together exponential marginals to model the dependence of obligors' times of default. See, for instance, Li (2000)[24].

1.2.2 Merton-type Factor Models

Over the last fifteen years, many credit portfolio models have been proposed for measuring economic credit capital. In fact, the development of credit risk models is still an active area of research[7]. Here we give an introduction to factor copula models, also known in this context as *conditional independent credit risk models*. These models are widely used in practice due to their ability to replicate realistic correlated defaults behavior while retaining a certain degree of tractability.
Dependence between the defaults of different obligors can be caused by several different factors. There may be direct links between the obligors (e.g. one obligor is creditor of the other), or indirect (but still potentially strong) links. For instance, industrial firms may use the same inputs (and hence be exposed to the same price shocks) or sell to the same market sectors. The general state of an industry and the economic cycle in a certain region also affect the obligor credit quality[8].
The first reason why a structural theoretical model is fundamental for a successful assessment of the risk of default correlation[9] is the availability of data. In fact, since joint defaults are rare events, historical data are typically sparse. Furthermore, the relation between equity prices (which are often used as proxy for correlation between firms) and credit risk is not obvious, and this link can only be established by using a theoretical model[10].
Another reason for the need of default models is the fact that the specification of a full joint default probability model is too complex: with n obligors, there are 2^n joint defaults events, and a realistic number of obligors makes it too costly to enumerate all these probabilities.
A framework which translates information in equity prices to information about credit risk was first developed by Merton (1974)[29]. According to Merton's structural model, a firm finances itself by equities (i.e. by issuing shares) and by debt (taken as zero coupon bonds with face value B and maturity T). Hence, the value of a firm's assets at time t is simply given by the sum of the value of equity S_t and debt B_t: $V_t = S_t + B_t$. A firm defaults if its assets value is lower than the value of its debt payable at time T: $V_T < B$. Therefore, the probability of default is given by $\mathbb{P}(V_T < B)$. The firm's assets value V_t is modeled as a geometric Brownian motion of the form

$$dV_t = \mu V_t dt + \sigma V_t dW_t \tag{1.25}$$

for constant drift $\mu \in \mathbb{R}$, volatility $\sigma \in \mathbb{R}^+$ and a standard Wiener process W_t. Equation (1.25) implies that $V_T = V_0 \exp\{(\mu - \frac{1}{2}\sigma^2)T + \sigma W_T\}$, and in particular that

[7] There is a growing econometrics literature on models for credit risk: some noteworthy works include Gagliardini and Gourieroux (2005)[7], Koopman et al. (2008)[22].

[8] The relation between default rates and economic growth has been addressed in several studies (e.g. Koopman and Lucas 2005[21]).

[9] Although the term *default correlation* has become a standard, it might be misleading. In fact, the classical linear correlation is a very inadequate measure of dependence between defaults and rank correlation measures (such as Spearman's rho or Kendal's tau) should be used.

[10] Furthermore, while fluctuation of market prices provide a nearly continuous flow of information, firms' credit worthiness is typically measured at an annual frequency.

$\ln V_T \sim N(\ln V_0 + (\mu - \frac{1}{2}\sigma^2)T, \sigma^2 T)$. Thus, under dynamics (1.25), the probability of default of a firm is given by

$$\mathbb{P}(V_T < B) = \mathbb{P}(\ln V_T < \ln B) = \Phi\left(\frac{\ln(B/V_0) - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right). \tag{1.26}$$

Notice how the probability of default in (1.26) is increasing in B and σ and decreasing in V_0 (for $V_0 > B$) and μ, in agreement with the economic intuition.
We are interested in one-year default probabilities, i.e. we set the maturity to $T = 1$. We can express the one-year probability of default in terms of the *standardized log-returns* $X := \left(\ln(V_1/V_0) - (\mu - \frac{1}{2}\sigma^2)\right)/\sigma$ as

$$p := \mathbb{P}(X < x),$$

where $x = -DD := \left(\ln(V_0/B) + (\mu - \frac{1}{2}\sigma^2)\right)/\sigma$ is the so called *distance to default*, as it indicates the minimal assets return to avoid default.
Similar to the (one-year) default probabilities one could define probabilities that obligor's year-end rating is category k, $k = 1, \ldots, K$[11]. Then one calibrates critical values x_k accordingly. These represent obligor's minimal one-year assets return to obtain a year-end credit rating equal to or better than rating k. This so called *migration model*, is a generalization of the *default-only* model, in so far as a default threshold is set to $x_{K+1} = -DD$. In our study we will use the default-only model, where we observe a loss only if during the one-year horizon, an obligor has defaulted.

Our interest is thus focused on the distribution of losses from default over one year. The choice of this horizon is justified by the availability of model parameters which are typically estimated on annual-basis. In order to specify the loss distribution, let us introduce the following notation.

- n = number of obligors to which the credit portfolio is exposed;
- p_i = probability of default (PD) of obligor i;
- c_i = loss at default (LAD) of obligor i[12].

Notice that both LGD and EAD, and so the loss at default c_i, are assumed exogenously given and hence uncorrelated with the default event[13].
Let now Y_i be the default indicator of ith obligor, i.e. $Y_i = 1$ if obligor i defaults and 0 otherwise. The total portfolio loss is then given by

$$L = c_1 Y_1 + \cdots c_n Y_n.$$

[11] Usually $K = 26$ credit rating categories are used.
[12] This is computed as the product of two parameters: the exposure at default (EAD), i.e. the amount to which the bank was exposed to the borrower at the time to default, and the loss given default (LGD), that is the percentage of loss over the total EAD.
[13] This assumption could be relaxed by assuming, for instance, LGD distributed as a Beta (e.g. Morokoff (2004)[30]).

Let us begin from the simple case where the credit portfolio is *homogeneous*, i.e. all the obligors have identical LAD $c_i \equiv 1$ and PD $p_i \equiv p$, and where the defaults are independent. In this situation the portfolio loss distribution is binomial, $L \sim \text{Bin}(n, p)$. It is a well-known fact that the binomial density has an extremely thin tail and the quantiles do not substantially change when different individual default probabilities are used.
The other extreme case is observed when defaults are perfectly dependent. This means that either all obligor defaults (with probability p) or no obligor defaults (with probability $1 - p$).
It is clear that, in order to obtain a realistic portfolio credit risk model, modeling the dependences among the obligors is crucial. In the following we present the framework with the so called *normal copula model*, associated with CreditMetrics (Gupton et al. 1997[14]) and related settings (e.g. Li 2000[24]).

One-factor dependence. In Merton's model, the default is triggered by the change of the value of its firm's assets V_t. The simplest approach to model the stochastic interdependence among the firms is to decompose the stochastic component dW_t in (1.25) in two terms: a systematic part Z that influences all obligors within the portfolio, and an obligor-specific (idiosyncratic) part ϵ. Thus, the one-factor model specifies the standardized assets log-return X of each obligor as

$$X_i = \rho_i Z + \sqrt{1 - \rho_i^2}\epsilon_i, \quad i = 1, \dots, n \tag{1.27}$$

where both Z and ϵ_i are assumed, without loss of generality, iid standard normal for all $i = 1, \dots, n$.
The default thresholds x_i are calibrated to reflect the individual default probabilities, i.e. $x_i = \Phi^{-1}(p_i)$. It is easy to see that each obligor indeed defaults with probability p_i:

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i < x_i) = \mathbb{P}(X_i < \Phi^{-1}(p_i)) = \Phi(\Phi^{-1}(p_i)) = p_i$$

being X_i standard normal. The correlation between obligors arises from their factor loadings. In fact, it is immediately seen that the correlation between obligor i and j is given by $cov(X_i, X_j) = \rho_i \rho_j$.
The key feature of this model is that, conditional on the realization of the systemic factor Z, the firm's return, and hence the defaults, are independent. In fact, given a realization of the common systemic factor, the individual conditional probability of default is given by

$$\begin{aligned} p_i(Z) := \mathbb{P}(X_i < x_i | Z = z) &= \mathbb{P}(\rho_i Z + \sqrt{1 - \rho_i^2}\epsilon_i < \Phi^{-1}(p_i) | Z = z) \\ \mathbb{P}\left(\epsilon_i < \frac{\Phi^{-1}(p_i) - \rho_i Z}{\sqrt{1 - \rho_i^2}} \middle| Z = z \right) &= \Phi\left(\frac{\Phi^{-1}(p_i) - \rho_i z}{\sqrt{1 - \rho_i^2}} \right). \end{aligned} \tag{1.28}$$

Thus, given Z, the obligors default independently with probability $p_i(Z)$.
Assume now that all the obligors have same PD p, unitary LAD $c_i \equiv 1$ and common

factor loading ρ. It follows, by the law of iterated expectations, that the probability of m defaults is given by

$$\mathbb{P}(L=m)=\int_{\mathbb{R}}\binom{n}{m}\left(\Phi\left(\frac{\Phi^{-1}(p)-\rho Z}{\sqrt{1-\rho^2}}\right)\right)^m\left(1-\Phi\left(\frac{\Phi^{-1}(p)-\rho Z}{\sqrt{1-\rho^2}}\right)\right)^{n-m}\phi(z)dz.$$

Figure 1.1 shows the histogram[14] of the default losses for a portfolio of 1000 obligors with individual default probability $p=0.05$ and unitary LAD, under different asset correlations. It is clear that, increasing the correlation parameter ρ, the density function is shifted to the left (towards the 'good' events) and its tail to the right. In fact, good events become equally more likely than bad events.
The increased mass in the tail of the loss distribution is the most significant effect from the risk management perspective. A higher assets correlation (and hence, a higher default correlation) leads indeed to a dramatic increase in the quantiles of the distribution (Value at Risk).

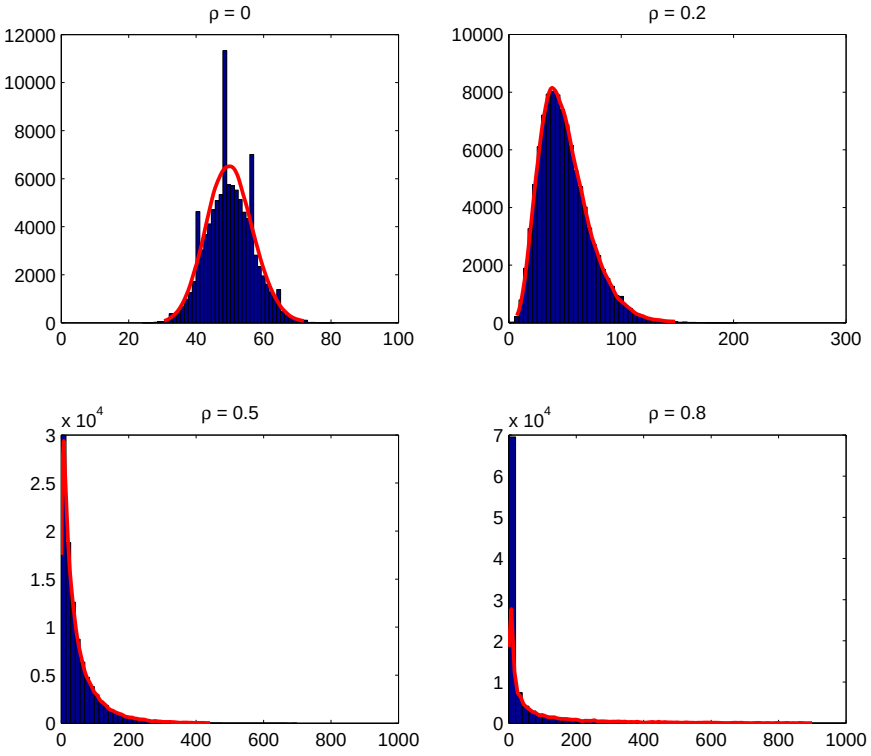

Figure 1.1: Histogram of the default losses under one-factor model for a portfolio with $n=1000, p=0.05$ and $c=1$, for different asset correlations.

[14]Obtained over 10^5 simulations, by means of the matrix language Matlab.

Large portfolio approximation. Even more tractability can be achieved when the number of obligors n tends to infinity. Let us stick to the homogeneous portfolio assumption with unitary LAD, and denote by Q the fraction of defaulted obligors, i.e. $Q = \frac{1}{n}\sum_{i=1}^{n} L_i$. If the defaults were independent, the portfolio loss distribution would converge by the central limit theorem (CLT) to a normal distribution. However, since defaults are not independent, the conditions of the CLT are not satisfied, and the portfolio loss is not asymptotically normal. Nevertheless, Vasicek (1987)[40] derives a useful limiting form for the portfolio loss distribution. In fact, given a value of Z, the fraction of portfolio losses conditional on Z converges, by the law of large numbers to $p(Z)$ as $n \to \infty$. Vasicek uses this fact to show that

$$\mathbb{P}(Q \leq x) \approx N\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)^{15}.$$

Obviously, infinitely large (and homogeneous) portfolios do not occur in practice. However, if the portfolio is sufficiently large and it is not dominated by a few exposures, the limiting distribution is shown to provide a very good approximation for the portfolio loss[16].

Multi-factor dependence. The results of the previous sections can be extended to include more than one systemic factor describing the firm's assets' value. Through this construction, the correlations among the X_i, which drive the correlations among defaults, have the form

$$X_i = \rho_{i1}Z_1 + \cdots + \rho_{id}Z_d + a_i\epsilon_i \quad i = 1, \ldots, n \tag{1.29}$$

where

- $Z_1, \ldots, Z_d$ are the systematic factors, each having standard normal distribution, and joint distribution $N(0, \Sigma)$;
- ϵ_i is the idiosyncratic risk associated with obligor i, iid-$N(0,1)$ distributed;
- $\rho_{i1}, \ldots, \rho_{id}$ are the factor loadings for obligor i, with $\rho_{i1}^2 + \cdots + \rho_{id}^2 \leq 1$, and $a_i = \sqrt{1 - (\rho_{i1}^2 + \cdots + \rho_{id}^2)}$ so that X_i is also $N(0,1)$.

The latent factors Z_i are sometimes derived from (macro-)economic variables, such as industry or regional risk factors[17]. We assume that the factor loadings are non-negative, thus ensuring that smaller values of the factors lead to smaller assets return

[15] Notice that the one-factor model used in Vasicek (1987) is $X_i = \sqrt{\rho}Z + \sqrt{1-\rho}\epsilon_i$, which is slightly different from (1.28) but also ensures the standard normality of the returns X_i.

[16] The approximation is not reliable for very low correlation between the assets and for very small number of defaults. In these cases, in fact, the discreteness of the real-world is too far from the continuous approximation of the infinitely large portfolio distribution.

[17] An alternative to this multi-factor model specification is proposed in Lucas et al. (1999)[25] who use a linear regression model to describe the assets return, namely $X_i = \sum_{j=1}^{d} \beta_{ij} Z_j + \epsilon_i$.

(and thence, to a higher number of defaults)[18].
Denoting by ρ the row vector $(\rho_{i1}, \ldots, \rho_{id})$ of factor loadings of obligor i, the correlation between two assets return X_i and X_j is $\rho_i \rho_j'$. Given an outcome of the risk factors $Z = (Z_1, \ldots, Z_d)'$, the conditional probability of default is now

$$p_i(Z) = \Phi\left(\frac{\Phi^{-1}(p_i) - \rho_i Z}{a_i}\right). \tag{1.30}$$

Similar formulas can naturally be obtained for the corresponding migration model (see, for instance Gupton et al. [14]).

A straightforward generalization of factor model (1.29) consists in changing the distribution of the firm assets value X_i from normal to, for instance, Student's t. This can be done by introducing the following dynamics

$$X_i = \frac{1}{W}(\rho_{i1} Z_1 + \cdots + \rho_{id} Z_d + a_i \epsilon_i) \quad i = 1, \ldots, n$$

in which W is χ^2 with ν degrees of freedom, independent of Z and ϵ.
More generally, changing the distribution function of systematic factors and noise does not affect the linear correlation between assets returns. However, it may have a large impact on the default risk of the portfolio. For a generic cdf G, the conditional probability of default becomes

$$p_i(Z) = G\left(\frac{G^{-1}(p_i) - \rho_i Z}{a_i}\right),$$

provided that its inverse G^{-1} exists.

[18] Non-negativity of the factor loadings is often imposed in practice as a conservative assumption ensuring that all defaults are positively correlated.

Chapter 2

Importance Sampling for Portfolio Credit Risk

Monte Carlo simulation is one of the most widely used computational tools in risk management. It has the advantage of being very general and the disadvantage of being rather slow, especially if a high accuracy (i.e. many reliable digits) is required. In the credit risk setting, there are two features that pose a particular challenge and motivate research on methods to accelerate simulations. First, accurate estimation of low-probability events of large losses are required. Importance sampling (Section 1.1) then stands out as a natural method to face this problem. Second, capturing dependence across sources of credit risk adds complexity both to the models employed and to the computational method used to calculate outputs of that model. In our study, the dependence between obligors in the portfolio are described by the normal copula model, as presented in Section 1.2. This chapter presents the importance sampling methods developed in this context by Glasserman and Li (2005)[9], hereafter GL.

2.1 Importance Sampling for Independent Defaults

Before discussing importance sampling in the normal copula model, it is instructive to consider the simpler problem of estimating the distribution of the portfolio loss L when its obligors are independent[1]. We thus take the default indicators $Y_1, \dots, Y_n$ to be independent, or equivalently, we take all the factor loadings $\rho_{ij} = 0$.
In this context, the problem of efficiently estimating rare event events such as $\mathbb{P}(L > x)$ for large x reduces the application of importance sampling to a sum of independent (but not identically distributed) random variables.
In order to estimate a portfolio loss statistic $h(L)$ more effectively[2], we want to

[1] Notice that with these simplifications, simulations are not necessary.

[2] GL consider the estimation of $\mathbb{P}(L > x)$; in our context, the estimators of interest are several (see Chapter 3), therefore we consider here a general (rare event) statistic of the portfolio loss $h(L)$. Particular attention should be paid to the likelihood ratio, as it will be addressed in Chapter 3.

increase the individual default probabilities. If we replace each p_i by some other probability of default q_i, the importance sampling method consists of estimating

$$\mathbb{E}_q\left[h(L)\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{Y_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-Y_i}\right] \tag{2.1}$$

where $\mathbb{E}_q[\cdot]$ denotes the expectation taken using the default probabilities q_i, $i = 1,\dots,n$, and the product inside the expectation is the likelihood ratio relating the original distribution of $(Y_1,\dots,Y_n)$ to the new one. It follows that, as long as the default indicators Y_i are sampled under the new probability of default q_i, expression (2.1) is an unbiased estimator of the portfolio loss statistic $h(L)$.
Instead of increasing the probability of default arbitrarily, we now apply an exponential change of measure introduced in subsection 1.1.1. In this simplified setting, L is a sum of independent random variables and its moment generating function, defined in (1.13), is given by

$$\mathbb{E}\left[\mathrm{e}^{\theta L}\right] = \prod_{i=1}^{n}\mathbb{E}\left[\mathrm{e}^{\theta Y_i c_i}\right] = \prod_{i=1}^{n}\left(p_i\mathrm{e}^{c_i\theta} + (1-p_i)\right) \quad \forall\,\theta\in\mathbb{R},$$

with $Y_i \sim \mathrm{Be}(p_i)$. Thus, the cumulant generating function in (1.14) is

$$\psi(\theta) = \log\mathbb{E}\left[\mathrm{e}^{\theta L}\right] = \sum_{i=1}^{n}\log\left(p_i\mathrm{e}^{c_i\theta} + (1-p_i)\right) \equiv \sum_{i=1}^{n}\psi_i(\theta)$$

and the likelihood ratio resulting from this change of measure is

$$\exp\left(\theta L - \psi(\theta)\right).$$

Exponentially twisting the portfolio loss L is equivalent to applying an ECM to every single default indicator Y_i. In fact, for each obligor the probability of default is the probability that $Y_i = 1$, which is also the mean of Y_i. We know from the property of the cumulative generating function (subsection 1.1.1) that $\psi_i'(\theta) = \mathbb{E}_\theta[Y_i c_i]$. Since under the new distribution each obligor defaults with probability $\mathbb{E}_\theta[Y_i = 1]$, the new individual default probabilities are given by

$$p_{i,\theta} := \mathbb{E}_\theta[Y_i = 1] = \psi_i'(\theta)/c_i = \frac{p_i\mathrm{e}^{\theta c_i}}{1 + p_i(\mathrm{e}^{\theta c_i} - 1)} \quad i = 1,\dots,n \tag{2.2}$$

for a chosen $\theta\in\mathbb{R}$. And since the default indicators are independent, if we apply ECM to all the Y_i the resulting likelihood ratio is the product of the individual likelihood ratios, namely

$$\prod_{i=1}^{n}\exp\left(\theta Y_i c_i - \psi_i(\theta)\right) = \exp\left(\theta L - \psi(\theta)\right).$$

Notice that once a twisting parameter is chosen, the same θ is used to determine the default probabilities of each obligor.

Although for any θ, IS yields an unbiased estimator[3], we shall investigate what choice of θ is the best one. From (2.2) we immediately see what is the most convenient direction. Since we want to increase the number of defaults, we need to choose $\theta > 0$. In fact, (2.2) can be written equivalently (Glasserman 2003[8]) as

$$\frac{p_{i,\theta}}{1-p_{i,\theta}} = \left(\frac{p_i}{1-p_i}\right) \mathrm{e}^{\theta c_i} \tag{2.3}$$

which confirms that a $\theta > 0$ increases the odds ratio for every obligor. It is worth noticing that a larger loss at default c_i leads to a larger increase in the odds ratio. Now, we would like to choose θ so as to minimize the variance (and hence, the second moment) of the estimator. Let us conform to GL, whose goal is the estimation of the probability of the tail, so that $h(L) = \mathbb{P}(L > x)$, for large x. This will substantially simplify the asymptotic analysis. The second moment of the estimator is given by

$$M_2(x;\theta) = \mathbb{E}_\theta \left[\mathbb{1}\{L > x\}\mathrm{e}^{-2\theta L + 2\psi(\theta)}\right] \leq \mathrm{e}^{-2\theta x + 2\psi(\theta)} \tag{2.4}$$

where it is understood that $\mathbb{E}_\theta[\cdot]$ indicates expectation taken using the θ-twisted probabilities and the upper bound holds for all $\theta \geq 0$. The minimization of $M_2(x;\theta)$ is difficult, therefore GL propose to minimize the upper bound. This is equivalent to maximize $\theta x - \psi(\theta)$ over $\mathbb{R}^+$. Since the function $\psi(\cdot)$ is strictly convex and passes through the origin, the maximum is attained at

$$\theta_x = \begin{cases} \text{unique solution to } \psi'(\theta) = x, & x > \psi'(0) \\ 0, & x \leq \psi'(0) \end{cases} \tag{2.5}$$

As for $\theta = 0$, $\mathbb{E}[L] = \psi'(0)$, the two cases in (2.5) correspond to $x > \mathbb{E}[L]$ and $x \leq \mathbb{E}[L]$. Therefore in the first case the choice of the twisting parameter implies

$$\mathbb{E}_{\theta_x}[L] = \psi'(\theta_x) = x,$$

i.e., we shift the distribution of L so that x is now its mean. An interpretation to the second case is that, when $x \leq \psi'(\theta)$, the event is not rare, so no change of measure is needed.

Asymptotic Optimality

In what follows, we briefly introduce the asymptotic framework and the first optimality results. For the sake of brevity, no proofs will be reported. Interested readers may refer to Glasserman and Li (2005)[9].
In order to establish the effectiveness of a simulation estimator in rare event settings, a standard way is to show that, as the event becomes increasingly rare, the estimator's second moment decreases at the fastest possible rate (among all the unbiased

[3] This is true only theoretically: in many practical examples, under/over-biasing can be observed (see Section 1.1).

estimators).
Glasserman and Li develop optimality results for the estimator $\mathbb{P}(L > x)$, based on the approximations

$$\mathbb{P}(L > x) \approx \mathrm{e}^{-\gamma x}, \quad M_2(x) \approx \mathrm{e}^{-2\gamma x} \tag{2.6}$$

for some $\gamma > 0$. In fact, since by Jensen's inequality $M_2(x) \geq (\mathbb{P}(L > x))^2$, if the second moment decays at twice the rate of the estimator itself, it decays at the fastest possible rate.
Let the number of obligors n in our portfolio increase along with the threshold x[4]. Let then $(p_i, c_i), i = 1, 2, \ldots$ be an infinite sequence of obligor parameters, and let us use the subscript n to indicate quantities corresponding to a portfolio with n obligors.

Theorem 3. *Assume that, for all $\theta \in \mathbb{R}$,*

$$\psi_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + p_i(\mathrm{e}^{\theta} c_i - 1)) \longrightarrow \bar{\psi}(\theta)$$

for some strictly convex $\bar{\psi}(\theta)$, and that there exists a $\bar{\theta}_x > 0$ at which $\bar{\psi}'(\bar{\theta}_x) = x$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n > xn) = -\gamma$$
$$\lim_{n\to\infty} \frac{1}{n} \log M_2(xn, \theta_n) = -2\gamma$$

where $\gamma = \sup_\theta = \{\theta x - \psi(\bar{\theta})\}$.

Therefore, the IS estimator with ECM parameter chosen to satisfy (2.5) is asymptotically optimal. Moreover, the limits in Theorem 3 make precise approximations (2.6). Also, these results, as well as its extensions in the next sections, continue to hold if we replace $Y_i c_i$ with a sequence of independent thin-tailed random variables.

2.2 Dependent Defaults: Conditional Importance Sampling

We now want to study the more interesting (and realistic) case in which the default indicators Y_i are dependent. As a first step in extending the method to the normal copula model presented in Section 1.2, we apply importance sampling conditional on the systematic factors Z. In fact, given $Z = z$, the default indicators are independent, with probability of default given by (1.28). Thus, we can proceed similarly to what we did in Section 2.1. Define

$$\psi(\theta, Z) := \log \mathbb{E}\left[\mathrm{e}^{\theta L} \middle| Z\right] = \sum_{i=1}^{n} \log\left(p_i(Z)\mathrm{e}^{c_i\theta} + (1 - p_i)\right) \tag{2.7}$$

[4]Notice that this is practically meaningful, being bank portfolios possibly exposed to many thousands of obligors.

Now, let $\theta_{x(Z)}$ solve

$$\psi'(\theta_{x(Z)}, Z) = x.$$

Let the new conditional probabilities of default

$$p_{i,\theta_{x(Z)}}(Z) = \frac{p_i(Z)\mathrm{e}^{\theta_{x(Z)}c_i}}{1 + p_i(Z)(\mathrm{e}^{\theta_{x(Z)}c_i} - 1)} \tag{2.8}$$

Conditional on Z, the default indicators Y_i are independent and, under the $\theta_{x(Z)}$-twisted distribution, Y_i takes value 1 with probability $p_{i,\theta_x(Z)}(Z)$. The default indicators are therefore easy to generate. Setting L equal to the sum of the $Y_i c_i$, the estimator of the tail of the distribution is given by

$$\mathbb{1}\{L > x\}\mathrm{e}^{-\theta_x(Z)L + \psi'(\theta_{x(Z)}, Z)} \tag{2.9}$$

The steps presented so far are summarized in the following

Algorithm 2 (Conditional Importance Sampling)**.** *For each replication:*

1. *Generate* $Z \sim N_d(0, 1)$.
2. *Compute the conditional default probabilities* $p_i(Z)$, $i = 1, \ldots, n$ *as in (2.8). If*

$$\mathbb{E}[L|Z] = \sum_{i=1}^{n} p_i(Z)c_i \geq x,$$

   set $\theta_x(Z) = 0$*; else, set* $\theta_x(Z)$ *equal to the unique solution of* $\psi'(\theta_{x(Z)}, Z) = x$*, with* $\psi'(\theta_{x(Z)}, Z)$ *defined in (2.7)*[5].
3. *Generate indicators* $Y_1, \ldots, Y_n$ *from the twisted conditional default probabilities in (2.8) and calculate the loss* $L = c_1 Y_1 + \ldots, c_n Y_n$.
4. *Return the estimator (2.9).*

The efficiency of Algorithm 2 depends on the degree of dependence among obligors. If the dependence is weak, this IS scheme is effective, that is, increasing the conditional probabilities of defaults is sufficient to achieve substantial variance reduction. However this method is much less effective when the defaults correlation is high. In fact, in that case, large losses occur primarily because of small outcomes of the systemic factor(s), and so far we have not applied IS to its distribution.
To illustrate these points more precisely, GL consider the special case of a single-factor, homogeneous portfolio. In such a simple situation Z is scalar, $p_i \equiv p$ and

[5] In fact, for small values of the factor returns Z, the conditional default probabilities may be sufficiently large that the expected loss given Z exceeds x. In such a case θ_x would be negative, and therefore, we set $\theta_x = 0$ to avoid to decrease the default probabilities. There is indeed no advantage to reducing the conditional default probabilities if the aim is to generate large values of L. Negative values of θ_x could however be useful, for instance, in estimating conditional expectations given $L = x$.

$c_i \equiv c$, and all obligors have common loading ρ. Thus, all the X_i have the form (1.27), with conditional probabilities of default $p(Z) = \Phi\left(\frac{\Phi^{-1}(p)-\rho Z}{\sqrt{1-\rho^2}}\right)$.
In parallel with Theorem 3, we analyze the behavior of the estimator when both the number of obligors n and the loss level x increase. Now, let ρ depend on n by writing $\rho = a/n^\alpha$, for $a \in \mathbb{R}$ and $\alpha \in \mathbb{R}^+$. Hence, the strength of the correlation is smaller for higher α. First, notice that keeping ρ fixed, the (S)LLN implies that the normalized loss L_n/n converges to $p(Z)$, and for $0 < q < 1$,

$$\mathbb{P}(L_n > qn) \longrightarrow \mathbb{P}(p(Z) > q) = 1 - \Phi\left(\frac{\Phi^{-1}(p) + \Phi^{-1}(q)\sqrt{1-\rho^2}}{\rho}\right) > 0.$$

Thus, with ρ fixed, the event $\{L_n > qn\}$ does not become vanishingly rare. Conversely, by letting ρ decrease with n, we get a probability that decays exponentially fast. GL then distinguish three cases, corresponding to different strength of asymptotic correlation: $\alpha < 1/2, \alpha = 1/2$, and $\alpha > 1/2$. It is shown that, asymptotic optimality is achieved only when $\alpha > 1/2$ (i.e. when the correlation vanishes relatively fast). Indeed, only in this case the second moment decreases at twice the rate of the first moment. At $\alpha = 1/2$, the second moment decreases faster than the first moment, but not twice as fast. And with $\alpha < 1/2$ the two decrease at the same rate, meaning that the importance sampling is (asymptotically) no more effective than standard Monte Carlo.
The failure of asymptotic optimality for estimator (2.9) is justified by the impact of the common risk factor Z on the occurrence of a large number of defaults. This suggests that, if the defaults are not weakly correlated, an application of IS to Z is necessary to achieve asymptotic optimality.

2.3 Dependent Defaults: Two-Step Importance Sampling

We want to apply, now, further importance sampling to the distribution of the factors Z, as well as to the conditional probabilities of default. To motivate this approach, observe that the variance of any estimator of $h(L)$ can be decomposed as

$$\mathbb{V}ar[\hat{h}(L)] = \mathbb{E}[\mathbb{V}ar[\hat{h}(L)|Z]] + \mathbb{V}ar[\mathbb{E}[\hat{h}(L)|Z]]$$

Twisting the default probabilities conditional on Z as in Section 2.2, reduces the first term of the variance decomposition, but does not affect the second term. This suggests that, in order to improve the IS scheme further, we need to focus on the variance of the conditional expectation. Being $\mathbb{E}[\hat{h}(L)|Z]$ a function of Z, we should apply importance sampling to Z to reduce the contribution of the second term to the total variance.
As in Glasserman and Li (2005), we restrict ourselves to changes of measure that shift the mean of Z from 0 to some other vector μ[6].

[6] A different approach which increases the variance of Z can be found in Morokoff (2004)[30].

Assume, for now, that the mean of the approximating normal distribution has been selected (and that factor loadings and probabilities of default have been estimated). The two-step IS algorithm can be summarized as follows.

Algorithm 3 (Two-Step Importance Sampling)**.** *For each replication:*

1. *Draw the systematic factors Z from $N_d(\mu, 1)$.*
2. *Apply step 2-3 of Algorithm 2.*
3. *Return the estimator*

$$\mathbb{1}\{L > x\}\mathrm{e}^{-\theta_{x(Z)}L+\psi'(\theta_{x(Z)},Z)}\mathrm{e}^{-\mu' Z+\frac{1}{2}\mu'\mu} \tag{2.10}$$

The only new factor in the estimator (2.10) is the likelihood ratio's term relating the density of the $N(0,1)$ to that of the $N_d(\mu,1)$ [7]. Thus, once an appropriate mean shift has been determined, this two-step IS scheme is no more difficult than the one-step procedure of Section 2.2.
To gain insights about the magnitude of the mean shift[8], let us consider again the case of a homogeneous portfolio with default probabilities $p_i \equiv p$ and loss at default $c_i \equiv 1$ for all $i = 1, \ldots, n$. The one-factor model then takes the form (1.27). If $\rho = 0$, the default indicators Y_i become independent, and $\mathbb{P}(L > x|Z) = \mathbb{P}(L > x)$. As Z drops out of the problem, the optimal choice of μ in this case is simply $\mu = 0$. At the other extreme, when $\rho = 1$, either all obligors default ($L = n$) or no defaults occur ($L = 0$). Let us further assume that $0 < x < n$, it follows that

$$\mathbb{P}(L > x|Z) = \mathbb{1}\{L > x\} = \mathbb{1}\{Y_i = 1\} = \mathbb{1}\{Z \leq x_i\},$$

for all $i = 1, \ldots, n$. Glasserman (2003) shows that, following the same steps that led to (2.5), we arrive at the parameter value satisfying

$$\mathbb{E}_\mu[Z] = x_i$$

that is, $\mu = x_i$. And since the default threshold x_i equals $\Phi^{-1}(p)$, this argument suggests that μ should increase from 0 to $\Phi^{-1}(p)$ as ρ increases from 0 to 1.
Recall from (1.7) that the minimum variance density is the normalized product of the integrand and the nominal density. Hence, for the problem at hand, the optimal IS density is proportional to

$$z \mapsto \mathbb{P}(L > x|Z)\mathrm{e}^{-\frac{1}{2}z'z} \tag{2.11}$$

Sampling from this density is generally infeasible, since the normalization constant is the value that we seek. However, we may try to select μ so that $N_d(\mu, I)$ approximates the optimal distribution. For instance, we could choose μ to be the mode

[7] Exponential change of drift for normal distributions has been already discussed in subsection 1.1.1.

[8] The direction is in fact straightforward: if Z is interpreted as liability, as in GL, μ has to be positive; in our case, we take $\mu < 0$.

of the optimal density (2.11). If our target is the tail of the distribution of L, the mode occurs at the solution to the optimization problem

$$\max_z \left\{ \mathbb{P}(L > x | Z = z)\, \mathrm{e}^{-\frac{1}{2}z'z} \right\} \tag{2.12}$$

Exact solution of (2.12) is usually difficult, thus Glasserman and Li propose several methods to approximate this problem. The one adopted for deriving the optimality results, as well as for the numerical examples, is the so called *tail bound approximation*[9]. Define

$$F_x(z) := \max_{\theta \geq 0} \{\psi(\theta, z) - \theta x\} = \psi(\theta_{x(z)}, z) - \theta_{x(z)} x \tag{2.13}$$

This provides an upper bound on the conditional tail of L, in the sense that

$$\mathbb{P}(L > x | Z = z) \leq \mathrm{e}^{F_x(z)}$$

This upper bound is then used as an approximation in (2.12), which leads to choose μ as the solution[10] to

$$\max_z \{F_x(z) - \frac{1}{2} z'z\} \tag{2.14}$$

In order to assess the efficiency of this combined estimator, GL consider again a single-factor homogeneous portfolio where the loss level x increases with the size of the portfolio n. In contrast with conditional IS (Section 2.2), here we do not need to assume that the correlation parameter ρ decreases asymptotically. It is shown that the two-step IS estimator is asymptotically optimal, thus indicating that it is effective in estimating probabilities of large losses in large portfolios[11]. It is important to highlight the difference with the optimality results of the conditional IS estimator. In fact, to establish its efficiency we were forced to let ρ decrease with the size of the portfolio in order to get a meaningful limit for the second moment. This indicates that twisting the probabilities of default without shifting the systematic factor mean is effective only when the correlation parameter is small.
Furthermore, GL show that the combined IS scheme is effective also when large losses are rare because individual probabilities of default are small[12]. By considering the usual single-factor homogeneous portfolio in which all obligors have default probability $p = \Phi(-c\sqrt{n})$, $c > 0$, i.e. p is $\mathcal{O}(\exp(-nc^2/2))$, the two-step IS estimator proves to be again asymptotically optimal.

Glasserman, Kang, Shahabuddin (2008)[12] develop an alternative approach for selecting shifts in factor means which lends itself more readily to the selection of shifts

[9] This, of course, does not preclude that other methods are asymptotically optimal as well.

[10] Solutions to this problem, as well as to the optimal θ search in (2.5), will be handled via Matlab (numerical) optimization routines. See Chapter 3 for details.

[11] However, since the limits are normalized by $\log n$, the actual decay rates are polynomial rather than exponential.

[12] Notice that this setting is typically relevant for banks whose portfolios have many highly rated obligors. It is also relevant to measure risk over short time horizons.

for multi-factor heterogeneous models[13]. Nevertheless, the numerical results provided suggest that for large loss thresholds x, the two step IS in GL outperforms the method based on multiple mean shifts. This is corroborated by the numerical examples in Glasserman and Li (2005), which show that the two-step importance sampling procedure is effective even in more general cases than the single-factor homogeneous portfolio, for which optimality is analytically proved.

[13]This method divides the space of factor outcomes into sets that lead to the default of different combinations of obligors; the possible mean shifts are the minimal points in these sets which corresponds to the "most likely" factor outcome leading to the default of a particular set of obligors.

Chapter 3

Simulation study

In financial institutions, *Economic Capital* (EC) acts as a buffer to absorb large unexpected losses, thereby protecting depositors and claim-holders, as well as providing confidence to external investor and rating agencies on the financial health of the firm.
Traditionally, economic capital is designed to cover *unexpected* losses up to a specified confidence level α, while credit reserves are set aside to absorb *expected* losses. Therefore, EC is typically estimated as the difference between the α-quantile of the portfolio loss distribution[1] (VaR_α) and the expected loss of the portfolio L, over a specified time horizon:

$$\text{EC}_\alpha = \text{VaR}_\alpha - \mathbb{E}[L]. \tag{3.1}$$

Once the amount of capital has been determined, it must be allocated among the various components of the credit portfolio. This process requires decomposing the total credit risk in the portfolio into individual risk contributions, referred to as *Economic Capital contributions* (ECC).
Given the definition of EC, it is natural to allocate capital based on contributions to VaR, defined as

$$\mathbb{E}[L_i | L = \text{VaR}_\alpha]. \tag{3.2}$$

The capital charged to an obligor i is then $\mathbb{E}[L_i | L = \text{VaR}_\alpha] - \mathbb{E}[L_i]$. Since an obligor's contribution to EL is simply its expected loss, which is easily computed analytically, the capital allocation essentially reduces to measuring VaR contributions.
Value at Risk has several shortcomings since it is not a coherent risk measure in the sense of Artzner et al. (1999)[1]. Specifically, it is sub-additive only for normal distributions, which means that in general, the sum of VaRs for two portfolios may be less than VaR for the combined portfolio. Expected Shortfall (ES) is a coherent measure of risk and represents an alternative to VaR both for measuring and allocating capital. ES contributions are defined as

$$\mathbb{E}[L_i | L \geq \text{VaR}_\alpha].$$

[1] The regulatory proposal in the Basel II agreement is based on the 99.9%-quantile (Basel Committee on Bank Supervision 2004[3]).

However, VaR (and VaR contributions) are more widely used. The rationale under this choice is that VaR is more representative of the risk of a financial institution. Furthermore, VaR refers to one particular point in the loss distribution, which makes it more challenging to obtain accurate and stable risk contributions via Monte Carlo simulations.
The estimation of ECC by Monte Carlo can be thought of as a two-phase procedure. During "Phase I", the VaR (and possibly also the ES) is estimated; then, the EC contributions are computed using the estimated VaR in place of the true VaR in the conditional expectation ("Phase II").
EC contributions are conditioned on events in the tail of the portfolio loss distribution. The rarity of these tail events presents a big obstacle to the calculation of these conditional expectations: each contribution depends on the probability of a rare event (a default) conditional on an even rarer event (an extreme loss for the whole portfolio). To address the difficulties associated with simulating rare events we will employ importance sampling algorithms, introduced in Section 1.1. Dependence among defaults are captured using the widespread Gaussian copula model of credit risk, presented in Section 1.2.

This work has been carried out during a thesis-internship at the Royal Bank of Scotland (RBS)[2]. The bank has already at its disposal an importance sampling scheme for EC contribution estimation. This scheme, which will be referred to as *current* IS, will constitute the starting point for our analysis. It consists of perturbing the mean of the factor returns Z by means of a Householder transformation lifting a univariate mean shift to the multivariate Z-space. From extensive simulation studies it turned out that, for α between 0.999 and 0.9997, the best results in terms of stability are obtained by choosing a shift of $\mu = -2$. Therefore, instead of having as benchmark a standard Monte Carlo estimator, we aim at developing importance sampling schemes which may further reduce the variability of the current IS estimator.
In order to face this challenge, we will first analyze the stability of the estimator to investigate which class of obligors leads to the highest instability. Then, we will develop efficient IS schemes starting from the approach of Glasserman and Li (2005)[9], discussed in detail in Chapter 2.

The chapter is organized as follows. Section 3.1 presents some measurements used to compare the efficiency of estimators. In Section 3.2, the test portfolio used for the simulation study is described. Estimator and results of Phase I and Phase II are illustrated in Section 3.3 and 3.4 respectively. Section 3.5 concludes.

[2]Economic Capital Modeling Team, Amsterdam, The Netherlands.

3.1 Efficiency measurements of estimators

The aim of this study is to improve the current importance sampling scheme. In order to discuss improvements, we first need to explain the criteria that will be used to compare alternative estimators[3]. Three aspects have to be considered: bias, variance, and computing time.
Let us begin by assuming that we are comparing unbiased estimators, i.e. that we can apply the (strong) law of large numbers to ensure that

$$\hat{\ell}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

converges to the true value ℓ as the number of iid draws of the r.v. X_i increases. Assume further that $\mathbb{V}ar[\hat{\ell}_n] = \sigma^2/n < \infty$. The central limit theorem provides information about the likely magnitude of the error in the estimate. In fact, the standardized estimator $(\hat{\ell}_n - \ell)/(\sigma/\sqrt{n})$ converges in distribution to the standard normal, i.e.

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\hat{\ell}_n - \ell}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall\, x \in \mathbb{R}.$$

It is a well-known fact that the same limit holds when σ is replaced with the sample standard deviation $\hat{\sigma}$, as $\hat{\sigma}$ is a consistent estimator of σ.
The quantity $\sigma/\sqrt{n}$ is called *standard error* (SE) of the estimator. Although the standard error is an important quality criteria, it is heavily affected by the scale of the property of interest (and its estimator). Therefore, in order to compare the accuracy of different estimators we use the *coefficient of variation* (CV) (sometimes referred to as *relative error*), defined as

$$\mathrm{CV}[\hat{\ell}_n] = \frac{SE[\hat{\ell}_n]}{\mathbb{E}[\hat{\ell}_n]}.$$

The coefficient of variation can be estimated by

$$\widehat{\mathrm{CV}}[\hat{\ell}_n] = \frac{\widehat{\mathrm{SE}}[\hat{\ell}_n]}{\hat{\ell}_n} = \frac{\hat{\sigma}/\sqrt{n}}{\hat{\ell}_n}. \tag{3.3}$$

The CLT justifies the confidence intervals

$$\hat{\ell}_n \pm z_{\alpha/2}\widehat{\mathrm{SE}}[\hat{\ell}_n],$$

where $\Phi(z_\alpha) = 1 - \alpha$. As $n \to \infty$, the probability that this interval contains the true value ℓ is $1 - \alpha$. Another way to see this is that our simulation error $\hat{\ell} - \ell$ is approximately distributed as $N(0, \sigma^2/n)$. This makes clear the intuition that, other things being equal, when we compare two estimators of the same quantity we should

[3] This section was mainly inspired by Glasserman (2003)[8].

prefer the one with the smaller variance.
But what if other things were not equal? In particular, if we are comparing two estimators and the one with smaller variance takes longer to compute, how shall we balance the variance reduction and the computational effort? Suppose that to generate a replication ℓ_i takes a (fixed) amount of computational time T. Let B be our computational budget, measured in the same units as T. Then, the number of replications we can complete given the available budget is $\lfloor B/T \rfloor$, assuming that the computing time of the initialization (i.e., the preparations before the simulations are performed) is negligible compared with the computing time of the simulations themselves. Let us denote the resulting estimator as $\ell_{\lfloor B/T \rfloor}$. It follows that

$$\sqrt{\lfloor B/T \rfloor}(\ell_{\lfloor B/T \rfloor} - \ell) \xrightarrow{d} N(0, \sigma^2)$$

as the computational budget B increases to infinity. Since $\lfloor B/T \rfloor / B \to 1/T$, we have that $\sqrt{B}(\ell_{\lfloor B/T \rfloor} - \ell) \xrightarrow{d} N(0, \sigma^2 T)$, as $B \to \infty$.
This formulation suggests that, as the computational budget increases, we should prefer the estimator with the smaller value of $\sigma^2 T$. This, in fact, will produce the more precise estimate (i.e. narrower confidence interval) for a given budget B.
A nice feature of the product $\sigma^2 T$ (variance per replication times computer time per replication) as a measure of efficiency is that it is insensitive to bundling replications into a single replication. In fact, suppose we redefined a replication as the average of two independent replications. This would halve the variance per replication and double the computing time, thus leaving the product unchanged. Clearly, this measure requires that the computing time per replication is constant, as it is the case in our study. For generalizations to cases where time can vary across simulations[4], the method derived in Glasserman (2003) leads to replace T with its expected value.
In the light of this, we will consider an estimator $\hat{\ell}_1$ more efficient than an alternative estimator $\hat{\ell}_2$ if the *relative time variance ratio*

$$\frac{\hat{\sigma}_2^2 T_2}{\hat{\sigma}_1^2 T_1} > 1.$$

In order to compare estimators of potentially different quantities, we replace the variance by the corresponding (squared) coefficient of variation, i.e.

$$\frac{\widehat{\mathrm{CV}}_2^2 T_2}{\widehat{\mathrm{CV}}_1^2 T_1}. \tag{3.4}$$

Notice that this generalizes (3.3) to situations where computing time are different. Therefore, unless otherwise stated, (3.4) will be used to compare the efficiency of two estimators.

[4] As an example, consider the pricing of a *barrier option*, whose simulation path ends the first time a certain barrier is crossed.

For unbiased estimators, estimator variability and computational effort are the most important considerations. However, reducing the variability or the computing time would be pointless if it accelerated the convergence to an incorrect value.
While bias in small samples is sometimes inevitable, some simulation estimators are *consistent* (i.e. asymptotically unbiased, provided that the expectation exists), in the sense that the estimator converges to the true value as the number of simulations increases.
Consider for example the problem of estimating a ratio of expectations $\mathbb{E}[X]/\mathbb{E}[Y]$ from iid replications $(X_i, Y_i), i = 1, \ldots, n$ of the pair (X, Y). The ratio of sample means $\bar{X}/\bar{Y}$ is biased for all n, as $\mathbb{E}[X/Y] \neq \mathbb{E}[X]/\mathbb{E}[Y]$. Nevertheless, $\bar{X}/\bar{Y}$ converges to $\mathbb{E}[X]/\mathbb{E}[Y]$ with probability 1 as $n \to \infty$.
But not all types of bias vanish in large samples. If the bias persists as the number of simulations increases, to measure an estimator performance we need to balance bias and variance. A standard measure is the mean square error, given by

$$\mathrm{MSE}[\hat{\ell}] = \mathbb{E}[(\hat{\ell} - \ell)^2] = (\mathbb{E}[\hat{\ell}] - \ell)^2 + \mathbb{E}[(\hat{\ell} - \mathbb{E}[\hat{\ell}])^2] = \mathrm{Bias}^2[\hat{\ell}] + \mathbb{V}ar[\hat{\ell}].$$

As it will be discussed later, our estimator of interest (which has the form of a ratio of sample means) is asymptotically unbiased.

3.2 Test portfolio

For the simulation study, a test portfolio representative of the true bank portfolio has been constructed. This permits to prevent obvious confidential data issues, as well as to work on a smaller portfolio, more suitable to carry out statistical analysis via matrix languages such as Matlab. The correctness of the developed Matlab codes (a sample of which is shown in Appendix A) has been verified by comparing the results with the original bank's model coded in C++.
The test portfolio consists of $n = 1300$ loans with a highly inhomogeneous exposure and default probability distribution. The obligors are equally spread in 26 buckets according to their credit quality. In each bucket we have 50 obligors, all with unit LGD and with EAD (and hence LAD) of five different sizes[5], namely

$$c_i = \begin{cases} 1, & i = 1, \ldots, 10 \\ 10, & i = 11, \ldots, 20 \\ 100, & i = 21, \ldots, 30 \\ 1000, & i = 31, \ldots, 40 \\ 10000, & i = 41, \ldots, 50 \end{cases} \tag{3.5}$$

The probability of default p_i varies between 0.00005 for the obligors in the best rating category and 0.4096 for the obligors with the lowest credit worthiness. In each category the PD is the same among the obligors. This deliberately heterogeneous lumpy

[5]The loans with smallest exposures can be thought of representing retail clients, the ones with bigger exposure are typically financial institutions clients.

profile (Figure 3.1) will lead to significant variation in marginal risk contributions. The resulting expected loss is 119.51 with standard deviation 496.61. Notice that these can be easily determined without simulations: the expected loss is in fact given by $\mathbb{E}[L] = \frac{1}{n}\sum_{i=1}^{n} p_i c_i$, and the standard deviation by $\sigma(L) = \sqrt{\mathbb{E}(L^2) - \mathbb{E}(L)^2}$.
Notice that the portfolio size does not appear to have a crucial effect on the quality

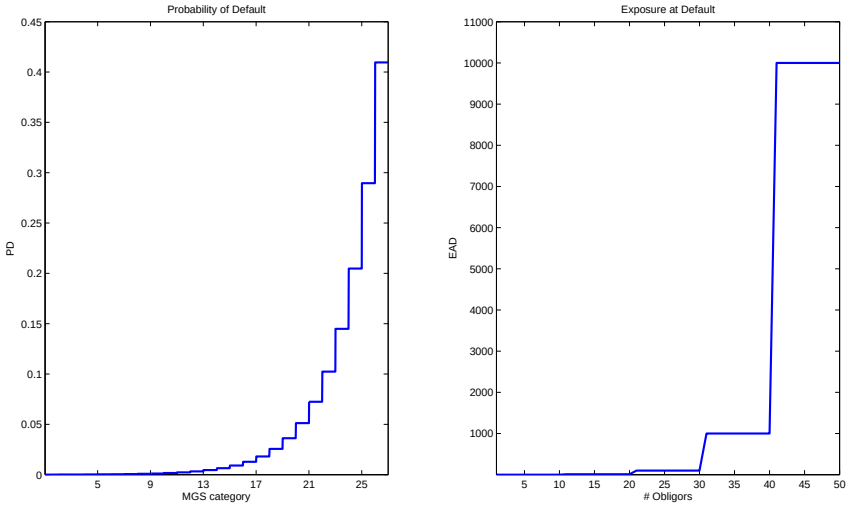

Figure 3.1: Test portfolio: 26 probability of default groups (left panel) and five exposure sizes (right panel). In each PD category the portfolio has 10 obligors for each EAD size.

of the estimators. Indeed, the asymptotic optimality results in Glasserman and Li (2005) assume $n \to \infty$, suggesting that the method is even more effective in larger portfolios.
Finally, we first assume that firms' assets returns are described by a single-factor model with constant factor loading $\rho = 0.6$ for each firm.

3.3 Phase I: Value at Risk

3.3.1 Estimator

Consider a portfolio of risky assets and a fixed time horizon, and denote by $f_L(l)$ and $F_L(l)$ respectively pdf and cdf of the corresponding loss distribution. The Value at Risk at level $\alpha \in (0,1)$ is the lowest amount l not exceeded by the portfolio loss L with probability $(1-\alpha)$, i.e.

$$\mathrm{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : \mathbb{P}(L > l) \leq 1-\alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.$$

In other words, $\mathrm{VaR}_\alpha(L)$ is simply the α-quantile of the portfolio loss distribution. Phase I is concerned with the estimation of such quantiles, first via standard Monte

Carlo, then by applying importance sampling methods. In fact, given the success observed in applying IS to rare-event simulations, the marriage between importance sampling and quantile estimation is natural, especially for α close to 1 (as required by the Basel II capital-adequacy framework).
Suppose we want to estimate the Value at Risk at level $100\alpha\%$ (e.g. $\alpha = 0.999$). We are seeking q such that $\int_q^\infty f_L(l)dl = 1-\alpha$, i.e.

$$\mathbb{E}[\mathbb{1}\{L > q\}] = 1 - \alpha, \tag{3.6}$$

where $\mathbb{E}[\cdot]$ denotes the expectation taken under the f_L density function.
Standard Monte Carlo estimation of (3.6) involves the simulations of N iid draws $L^j \sim f_L$, $j = 1, \ldots, N$. For N large enough we have

$$\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{L^j > q\} \approx 1 - \alpha. \tag{3.7}$$

In fact, the law of large numbers implies that $\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{L^j > q\}$ converges to $1-\alpha$ as $N \to \infty$. Sort now the losses L^j in ascending order and denote by $L^{(m)}$ the m-th loss in the ordering. Then, under (weak) regularity conditions,

$$L^{(\lceil \alpha N \rceil)} \longrightarrow \mathrm{VaR}_\alpha(L),$$

since $\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{L^j > L^{(\lceil \alpha N \rceil)}\} \approx 1 - \alpha$.

Notice that, once VaR_α has been estimated, to estimate Expected Shortfall it suffices to take only those L^j that are greater than $\widehat{\mathrm{VaR}}_\alpha$ and average across them.

To apply importance sampling to Phase I problem, we choose a distribution function $\tilde{F}_L$ from which variate generation is possible and for which the probability measure associated with F_L is absolutely continuous[6] with respect to the measure associated with $\tilde{F}_L$. This guarantees the existence of a density $\tilde{f}_L$, under which we simulate N iid draws $\tilde{L}^j \sim \tilde{f}_L$, $j = 1, \ldots, N$. Then, for N large enough we have

$$\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{\tilde{L}^j > q\}\frac{f_L(\tilde{L}^j)}{\tilde{f}_L(\tilde{L}^j)} \approx 1 - \alpha, \tag{3.8}$$

since

$$\tilde{\mathbb{E}}\left[\mathbb{1}\{\tilde{L}^j > q\}\frac{f_L(\tilde{L}^j)}{\tilde{f}_L(\tilde{L}^j)}\right] = \int_q^\infty \tilde{f}_L(l)\frac{f_L(l)}{\tilde{f}_L(l)}dl = \int_q^\infty f_L(l)dl = 1 - \alpha,$$

where it is understood that $\tilde{\mathbb{E}}[\cdot]$ is the expectation under the IS pdf $\tilde{f}_L$. To simplify the notation, let us denote the likelihood ratio as $\gamma_j = \frac{f_L(\tilde{L}^j)}{\tilde{f}_L(\tilde{L}^j)}$. The (S)LLN implies

[6] I.e., F_L has a derivative f_L almost everywhere which is Lebesgue integrable, and $F_L(x) = F_L(a) + \int_a^x f_L(l)dl$, for all x in a compact interval $[a, b]$.

that $\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{\tilde{L}^{(j)}\gamma_j > q\}$ converges to $1-\alpha$ as $N \to \infty$.
The equalities above motivate the following estimator for the quantile q:

$$\hat{q} = \tilde{L}^{(k)}, \quad k = \sup\left\{ k : \sum_{j=k}^{N} \gamma_j > N(1-\alpha) \right\}. \tag{3.9}$$

Notice that if $f_L = \tilde{f}_L$, then $\gamma_j = 1$, and the standard MC estimator is recovered.

An importance sampling estimate of ES can be naturally be obtained as

$$\sum_{j=k}^{N} \gamma_j \tilde{L}^{(j)} / \sum_{j=k}^{N} \gamma_j.$$

Glynn (1996)[13] derives a central limit theorem for four different IS estimators of extreme quantiles. It is shown that, under regularity conditions[7] estimator (3.9) is the preferred estimator (in terms of convergence rate) when α is close to 1. Obviously, when α is close to 0, it is more convenient to start summing from the beginning of the ordered series. In fact, the best estimator turns out to be

$$\hat{q} = \tilde{L}^{(k)}, \quad k = \inf\left\{ k : \sum_{j=1}^{k} \gamma_j > N\alpha \right\}.$$

The CLT derived in Glynn can be used as the basis for large-sample confidence intervals for $\hat{q}$. An alternative way of computing numerical standard errors for $\hat{q}$ can be found in Hoogerheide and Van Dijk (2010)[17][8].

3.3.2 Numerical results

Here, we present results from Phase I obtained with crude Monte Carlo (CMC) and with the current IS scheme (IS), which perturbates the systemic factor's mean by a shift of $\mu = -2$.
Table 3.1 presents quantile estimates and coefficient of variation obtained with these two methods, for various choices of α. All results in Phase I are based on 50 runs (used to compute Monte Carlo standard errors) of $N_1 = 2 \times 10^5$ simulations. Results have been obtained using a computer with processor Intel Xeon, Core Duo CPU @ 3.00GHz; one run with Matlab takes around 2 minutes and 50 seconds. Notice that, since differences in computational times between the two methods are negligible, in calculating the reduction factor (3.4) computer time has not been taken into account.

[7] These amount to the differentiability of F_L and the requirement that $\tilde{\mathbb{E}}[\gamma_j^3] < \infty$.

[8] Note that, in our framework, the separation between "high loss draws" and "regular draws" is not required. This is in fact used to recover the exact posterior density needed for the Bayesian estimation of a model. In our application, it suffices to focus on the high loss draws solely.

While the estimates are roughly the same, the variability of CMC estimator increases substantially as α approaches 1. Conversely, IS yields a relative error which does not to depend on the chosen confidence level (differences are just due to statistical noise). As a result, IS proves to be more effective as the event becomes rarer, leading to a variance reduction higher than 160% for $\alpha = 0.9997$.
The four different confidence levels are motivated by Royal Bank of Scotland's business choices. The level $\alpha = 0.9997$ is required to be rated as Aa by rating agencies which can be regarded as RBS's target; $\alpha = 0.999$ is the level dictated by Basel Committee on Bank Supervision (2004)[3]; $\alpha = 0.98$ is used for stress testing, and $\alpha = 0.96$ is used to set up the bank's risk appetite.
However, even for smaller α, because under importance sampling L is centered at the quantile VaR_α, IS performs better than CMC, as long as $\text{VaR}_\alpha > \mathbb{E}_\mu[L]$.
Figure 3.2 plot the sampling distributions of standard Monte Carlo and current IS before correcting by the likelihood ratio. As desired, importance sampling draws more scenarios in the tail of the portfolio loss distribution.

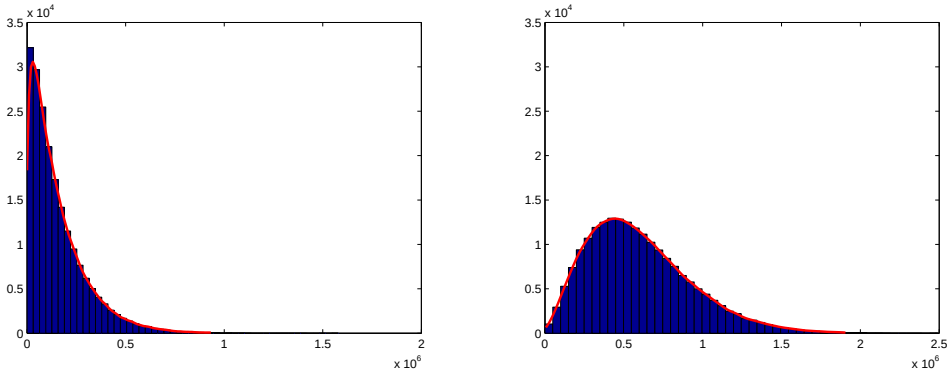

Figure 3.2: Sampling distributions of standard Monte Carlo (left panel) and current IS before correcting by the likelihood ratio (right panel).

α	0.96	0.98	0.999	0.9997
CMC	479.50	575.36	972.75	1128.28
	(3.39e-3)	(3.91e-3)	(9.30e-3)	(14.20e-3)
IS	479.33	575.06	972.20	1112.49
	(0.84e-3)	(0.87e-3)	(0.89e-3)	(0.84e-3)
Red. factor	4.0357	4.4943	10.9412	16.1364

Table 3.1: Quantile estimates (€×1000) and CV (within parentheses) based on 50 runs of 2×10^5 simulations. Crude MC and current Importance Sampling for various α.

As long as VaR estimation is concerned, current IS performs very well and further improvements are not needed for practical purposes. However, it is worth trying to implement Glasserman and Li approach to compare its performance with the one

obtained with the current IS scheme.
The optimization programs have been solved by means of Matlab routines. The conditional importance sampling (CIS), described in Section 2.2, requires the search of an optimal twisting parameter θ^* which has been found using the function "fzero". The algorithm upon which this function is based, uses a combination of bisection, secant, and inverse quadratic interpolation methods. The optimal mean shift μ^* for the two-step algorithm (2IS) has been found by means of the routine "fminbnd". Its algorithm is based on golden section search and parabolic interpolation. More details about these algorithms can be found, for instance, in Brent (1973)[4].
Table 3.2 shows the VaR estimates using the optimal mean shift μ^* in place of $\mu = -2$, to which the CV is then compared. The method yields a further variance reduction as long as $\mu^* < -2$. In fact, this shift is optimal provided that also an optimal twisting parameter θ^* is obtained. Yet, while μ^* has to be found only once (and the computer time to obtain it is negligible), the search of an optimal θ requires that a non-linear equation is solved in each simulation. This increases significantly the computing time and, as a result, GL's 2IS method proves to be less effective than current IS, as long as (3.4) is used to compare the efficiency[9].

α	0.96	0.98	0.999	0.9997
μ^*	-1.7708	-2.075	-3.1138	-3.4645
IS (μ^*)	4.7935e+5 (0.93e-3)	5.7511e+5 (0.72e-3)	9.7220e+5 (0.62e-3)	11.1249e+5 (0.60e-3)
Red. factor	0.9032	1.2083	1.4355	1.4000

Table 3.2: Quantile estimates (€×1000) and CV (within parentheses) using the optimal factor's mean shift. The reduction factor uses the current IS as benchmark.

We noticed, however, that the optimal twisting parameter θ^* closely depends on the realization of the common systemic factor Z. It is natural indeed that, if a small systemic return is drawn, a small θ suffices to observe a total loss at quantile level. One could thus think to infer θ^* from the realization of Z.
A *naive* way, which proves to be quite effective, is to estimate θ^* as $\hat{\theta}^* = d + sZ$, where d is the difference in the mean between θ^* and Z and s the ratio of their standard deviation ratio. Means and standard deviations can be easily obtained from a pilot run at a very low computing cost. The right panel of Figure 3.3 shows 200 draws of the systemic factor and the corresponding optimal θ; the left panel plots the optimal twisting parameter θ^* and its $\hat{\theta}^*$, based on the draws of Z.
Avoiding the search of θ^* at each iteration, makes the computational time of the IS scheme comparable with that of CMC. However, since the focus of this study is the ECC estimator (Phase II), this will be further analyzed in future studies.

[9] A similar result can be found in Heitfield et al. (2006)[16]. In their application it turns out that greater accuracy can be achieved (for a fixed computation time) by CMC than GL's conditional IS method.

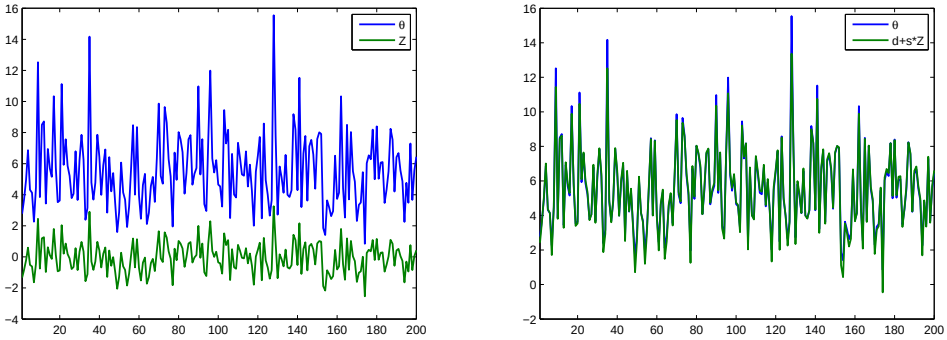

Figure 3.3: 200 draws of the systemic factor Z and corresponding θ^* (left panel); in the right panel, 200 draws of θ^* and of its naive estimate $\hat{\theta}^*$.

3.4 Phase II: VaR Contributions

3.4.1 Estimator

Once we have estimated the quantile q at the desired confidence level α (VaR_α), we focus on the second-phase problem of estimating the marginal contribution of each obligor i introduced in (3.1). Because an obligor's contribution to $\mathbb{E}[L]$ is simply its expected loss, which is straightforward to calculate, capital allocation reduces to the more challenging task of measuring the VaR contributions (3.2). That is, we aim at estimating the expected value of each loss L_i when the portfolio loss L equals the α-quantile.

To determine the VaR contribution of each obligor, we would need to enumerate all combinations of obligors' default that lead to the same quantile portfolio loss. For large portfolio, this is clearly an infeasible task. Therefore, we replace the condition $L = VaR_\alpha$ with $|L - VaR_\alpha| < \epsilon$, for some $\epsilon > 0$, which also ensures that $\mathbb{P}(|L - VaR_\alpha| < \epsilon) > 0$. Let us denote the interval around the quantile by $Q = [VaR_\alpha - \epsilon, VaR_\alpha + \epsilon]$, and let K be the number of portfolio losses in Q. The quantity of interest is thus

$$\ell_i = \mathbb{E}\left[L_i | L \in Q\right]. \tag{3.10}$$

Assuming that the ith obligor's loss is represented by a discrete random variable L_i, (3.10) can be written as

$$\ell_i = \sum_{j=1}^{\infty} L_i^j \mathbb{P}(L_i = L_i^j | L^j \in Q) = \sum_{j=1}^{\infty} L_i^j \frac{\mathbb{P}(L_i = L_i^j \cap L^j \in Q)}{\mathbb{P}(L^j \in Q)}, \tag{3.11}$$

where L_i^j denotes a realization of L_i in scenario j, for $j = 1, \ldots, N$ iid replications. We first consider the estimation of $\ell_i, i = 1, \ldots, n$, by ordinary Monte Carlo. Next,

we present the importance sampling estimator under the "current" scheme, consisting in a shift of the systemic factor's mean. Finally we derive a new estimator where both the systemic and in the idiosyncratic factors' mean are shifted. This extension is needed to apply GL's approach, which is derived for the estimation of whole portfolio loss tail, to the ECC estimator.

- **Crude Monte Carlo.** Based on a set of N iid replications, we can estimate (3.11) with standard MC as

$$\hat{\ell}_i = \sum_{j=1}^{N} L_i^j \frac{\mathbb{1}\{L^j \in Q\}/N}{K/N} = \frac{L_i^1 + \cdots + L_i^K}{K},$$

  where L_i^j indicates a realization of L_i in scenario j (for $j = 1, \ldots, N$) and where $L_i^1, \ldots, L_i^K$ indicate ith obligor's K losses that fall in Q.
  Applying the SLLN to both numerator and denominator, it is easily seen that

$$\hat{\ell}_i \longrightarrow \ell_i, \quad i = 1, \ldots, n$$

  with probability 1 (assuming that the denominator is positive).

- **IS schemes shifting the mean of the systemic factor Z.**
  Let us denote by $\tilde{L}_i^j$, a draw of ith obligor's loss under the new probability measure $\tilde{\mathbb{P}}(\cdot)$ in scenario j, and let γ_{Z^j} be the likelihood ratio's term corresponding to the mean shift of Z. We have:

$$\ell_i = \sum_{j=1}^{\infty} \tilde{L}_i^j \gamma_{Z^j} \frac{\tilde{\mathbb{P}}(\tilde{L}_i = \tilde{L}_i^j \cap \tilde{L}^j \in Q)}{\tilde{\mathbb{P}}(\tilde{L}^j \in Q)}.$$

  This can be estimated using a set of N iid replications as

$$\hat{\ell}_i = \frac{\sum_{j=1}^{K} \tilde{L}_i^j \gamma_{Z^j}}{\sum_{j=1}^{K} \gamma_{Z^j}}. \tag{3.12}$$

  since the numerator $\tilde{\mathbb{P}}(\tilde{L}_i = \tilde{L}_i^j \cap \tilde{L}^j \in Q) \approx \sum_{j=1}^{N} \mathbb{1}\{\tilde{L}^j \in Q\}/N$ and the denominator is estimated by $\sum_{j=1}^{N} \gamma_{Z^j} \mathbb{1}\{\tilde{L}^j \in Q\}/N = \sum_{j=1}^{K} \gamma_{Z^j}/N$.
  Thus, by shifting only the factor mean, the conditional expectation is a weighted average with weights equal to the likelihood ratios of the losses around the quantile.

- **IS schemes shifting both systemic and idiosyncratic factor means.**
  Because each individual loss has its idiosyncratic term shifted, it is now convenient to indicate by $\tilde{L}_i^*$ a realization of the loss of obligor i corrected by the idiosyncratic part of the likelihood ratio $\gamma_{\epsilon_i^j}$, i.e., for each replication $j = 1, \ldots, N$, $\tilde{L}_i^{*j} = \tilde{L}_i^j \gamma_{\epsilon_i^j}$. We stress that writing the individual losses in

this fashion, is not just meant to simplify the notation, but it has proven to be necessary in practice. In fact, it permits to avoid numerical problems due to a too small likelihood ratio term at the denominator. Now, we can write the conditional expectation (3.10) as

$$\ell_i = \sum_{j=1}^{\infty} \tilde{L}_i^{*j} \gamma_{Z^j} \frac{\tilde{\mathbb{P}}(\tilde{L}_i^* = \tilde{L}_i^{*j} \cap \tilde{L}^j \in Q)}{\tilde{\mathbb{P}}(\tilde{L}^j \in Q)},$$

and since the ratio of probabilities is again estimated as $\frac{\mathbb{1}\{\tilde{L}^j \in Q\}}{\gamma_{Z^j} \mathbb{1}\{\tilde{L}^j \in Q\}}$, we have

$$\hat{\ell}_i = \frac{\sum_{j=1}^{K} \tilde{L}_i^{*j} \gamma_{Z^j}}{\sum_{j=1}^{K} \gamma_{Z^j}}. \tag{3.13}$$

Remark. A Phase II procedure for the estimation of ES contributions is much simpler than the estimator derived in (3.13). In fact, it suffices to store all the losses $\tilde{L}_i^{*j}$ for which the total loss $\tilde{L}^{*j} \geq \text{VaR}_\alpha$, and then averaging across them using γ_{Z^j} as weights.

Glasserman (2005)[10], develops importance sampling schemes for the estimation of marginal risk contributions (such as VaR contributions). It is claimed that estimators of type (3.12)-(3.13) converge to their true values. Formally, denoting the quantity of interest as

$$\ell_i = \frac{\tilde{\mathbb{E}}\left[L_i \gamma \mathbb{1}\{\tilde{L} \in Q\}\right]}{\tilde{\mathbb{E}}\left[\gamma \mathbb{1}\{\tilde{L} \in Q\}\right]},$$

it can be shown that, with probability 1,

$$\hat{\ell}_i \longrightarrow \ell_i, \quad i = 1, \ldots, n.$$

Furthermore, it can be proved that confidence intervals can be calculated using the following standard errors[10]:

$$\hat{\sigma}_i = \left(\frac{N \sum_{j=1}^{N} (\tilde{L}_i^{j*} \gamma_Z^j - \hat{\ell}_i \gamma_Z^j)^2 \mathbb{1}\{\tilde{L}^j \in Q\}}{(\sum_{j=1}^{N} \gamma_Z^j \mathbb{1}\{\tilde{L}^j \in Q\})^2} \right)^{1/2} \tag{3.14}$$

Then, for both estimators (3.12)-(3.13),

$$\frac{\hat{\ell}_i - \ell_i}{\hat{\sigma}_i / \sqrt{N}} \xrightarrow{d} N(0,1)$$

[10] In our application, we have used Monte Carlo standard errors, which require the run of many replications, but ensure a higher accuracy.

and

$$\hat{\ell}_i \pm z_{\alpha/2}\frac{\hat{\sigma}_i}{\sqrt{N}}$$

is a valid $1-\alpha$ confidence interval for $\hat{\ell}_i$[11].

The algorithm. In order to program an importance sampling scheme for the estimation of the Value at Risk (Phase I), it suffices to follow the steps that lead to estimator (3.9). On the other hand, the implementation of Phase II deserves some more attention[12].

Phase I returns an estimate of the quantile at a chosen confidence level α. It is worth noticing that, once the quantile of interest has been stored, for a simulation study of the ECC estimator, there is no need to run Phase I again. The separation between the two phases also permits to avoid "out of memory" errors[13]: in fact, the only quantity we need to save from Phase I is the estimated quantile $\hat{q}$.

First of all, the number of observations around the quantile is set: we typically use $K = 0.01 \times N_2$, with N_2 being the number of simulations for the second phase (we set $N_2 = 2 \times 10^5$ for each run). The larger K, the smaller the variability of the ECC estimator. However, as K increases, the estimator may be less precise as the losses in the interval may be not representative of the loss at quantile level (that is, the sum of the individual losses may be very far from $\hat{q}$). The trade-off between large number of scenarios and narrow interval will be addressed in more detail in the next section.

Since we take the interval around the quantile to be symmetric (i.e., we store $K/2$ vector of losses smaller than $\hat{q}$, and $K/2$ greater or equal $\hat{q}$), with crude Monte Carlo it is practically useless to set $K > 2(1-\alpha) \times N_2$. In fact, the law of large numbers implies that, for a high number of simulations, roughly $(1-\alpha) \times N_2$ observations will fall to the right of the quantile. Hence, importance sampling is crucial to cluster a high number of scenarios around $\hat{q}$, while retaining a narrow interval Q.

Our Phase II algorithm proceeds as follows. Once the obligors' losses are generated, we first store $K/2$ vectors of losses for which $L = \sum_{i=1}^n L_i < \hat{q}$, and $K/2$ for which $L = \sum_{i=1}^n L_i \geq \hat{q}$; when both the left and the right semi-interval are filled, we efficiently sort them, according to the corresponding portfolio loss L. This permits to compare every new generated portfolio loss only with the furthest loss from the quantile. Suppose a portfolio loss $L^j < \hat{q}$ is drawn; if $L^j > L^1$, where L^1 indicates the smallest loss stored in Q, then we replace L^1 with L^j and the semi-interval is sorted again. Similarly for scenarios where $L^j \geq \hat{q}$ (see the code for further details). Once all vectors of losses in Q (and the corresponding likelihood ratios) have been stored, the estimator is computed as described in subsection 3.4.1.

[11] This follows from a general result on the asymptotic normality of nonlinear functions of sample means (e.g., Serfling 1980[36]).

[12] A Matlab code for both Phase I and II can be found in Appendix A.

[13] Matlab generates an out of memory message whenever it requests a segment of memory from the operating system that is larger than what is currently available. Therefore, an efficient use of memory is required to help optimize the available memory.

Finally, because the total risk (VaR) decomposes in VaR contributions, i.e.

$$\sum_{i=1}^{n} \mathbb{E}[L_i | L = \text{VaR}_\alpha] = \mathbb{E}\left[\sum_{i=1}^{n} L_i | L = \text{VaR}_\alpha\right] = \mathbb{E}[L | L = \text{VaR}_\alpha] = \text{VaR}_\alpha, \quad (3.15)$$

we multiply each obligor's loss L_i^j, $i = 1, \ldots, n$, to a scalar a_j in order to ensure that (3.15) holds for each scenario j. Namely,

$$a_j = L^j / \hat{q}, \quad j = 1, \ldots, K.$$

Interestingly, as soon as an adequate importance sampling scheme is implemented, a_j is typically in the range [0.999-1.001].

3.4.2 Sensitivity results

In this section, we aim at studying the variability of the ECC estimator under the current IS scheme. In fact, before developing a new scheme, it is important to investigate how the parameters of the model may affect the variability of the estimator. This will then guide us to prototype new and more efficient IS algorithms.
For Phase II we focus on the 99.9%-quantile of the portfolio loss distribution. Results are based on 50 runs, each with 200000 simulations. The Matlab codes have been verified by comparing the results with the C++ program available at RBS. Matlab and C++ outputs perfectly match each other (albeit one run in Matlab, with $K = 2000$ scenarios in Q, takes around 7 minutes and 50 seconds, while C++ takes only 10 seconds). For this reason, and because the new IS schemes are implemented in Matlab, only results obtained with Matlab will be reported.
Table 3.3 shows the average coefficient of variation of the ECC estimator obtained by using both CMC and current IS, for different number of observations around the quantile estimated in Phase I. The table also reports the quantile intervals. In fact, we are not only interested in reducing the variability of the estimator, but also in getting a high precision in terms of capturing the exact quantile contribution. As discussed before, stability and precision compete: averaging across a large number of tail scenarios reduces the variability of the VaR contributions and makes the estimator more stable; however, since these scenarios span a large domain, may not be all representative of the α-quantile. Vice versa, using few scenarios clustered around the quantile, leads to a heavily unstable estimate, as it crucially depends on whether an obligor has defaulted in that particular scenario.
We see that, the higher K (computed as percentage of the total number of simulations), the more the variability is reduced. On the other hand, the estimator becomes less specific to the α-quantile scenarios (that is, the quantile interval is less narrow[14]). This holds especially with ordinary Monte Carlo: from the table, the difficulty faced by CMC in filling the semi-interval on the right of the quantile is apparent.

[14]The estimated 99.9%-quantile in Phase I is $\hat{q} = 971724$ €.

For a given number of observations in Q, importance sampling concentrates the scenarios around the quantile thus boosting the precision. For a given range around the quantile, the variability is reduced as more scenarios are drawn within the range. Moreover, since by increasing K crude Monte Carlo does not manage to fill the whole interval[15], the higher the number of scenarios, the more IS outperforms CMC, as can be glanced also from Figure 3.4.

K	Method	mean(CV)	Q (€×1000)
$0.005N_2$	CMC	0.1360	[813.838, 1560.003]
	IS	0.1104	[966.472, 976.833]
$0.01N_2$	CMC	0.1092	[738.145, 1575.232]
	IS	0.0777	[961.332, 981.678]

Table 3.3: Avarage ECC coefficient of variation and quantile interval Q under CMC and IS methods, for different number of observations in Q.

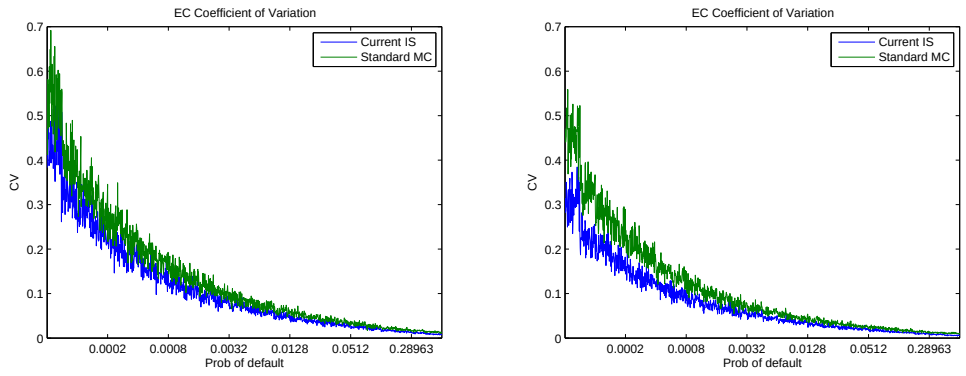

Figure 3.4: ECC's coefficient of variation estimated with CMC (green) and current IS (blu) plotted against the obligors' default probability.

What drives the ECC estimator variability? The charts in Figure 3.4 plot the coefficient of variation (3.3) of the ECC estimated with both CMC and IS. The estimated CV's are ordered accordingly to the obligors' probability of default, which is reported on the x-axis.
Few considerations, although obvious, are noteworthy. The probability of default crucially affects the variability of the estimator. It is natural indeed that the higher the credit quality of an obligor, and so the lower its PD, the harder it is to estimate

[15]This holds because the program stops when the total number of simulations N_2 is reached (see Appendix A). If instead the program stopped when the number of observations in Q was reached, we would obtain comparable results in terms of stability, but obviously CMC would take an enormous time to fill the interval. Moreover, the precision would still be very far from the one obtained with IS.

its EC contribution, since defaults are rare events. It follows that, for low PD, current IS is more effective than CMC, while for high PD ($p > 0.05$), CMC and IS yield roughly the same performances. Notice also that there is a jump in the CV from the first rating category ($p = 0.00005$) and the second rating category ($p = 0.0001$). The graphs in Figures 3.5-3.6 show that the only driver of ECC's instability is the probability of default. In particular, the loss at default does not seem to play any role in the variability of EC contributions, just determining its scale. Thence, as soon as we scale by the estimator mean, there is no relation between the magnitude of obligors' exposure and their ECC variability. The y-axis reports the LAD's scale (on the right) and the PD scale (on the left), while the CV is now shown on the x-axis in descending order. Figure 3.5 depicts the 130 obligors with the highest CV, while in Figure 3.6 the complete portfolio is plotted. It can be seen that, while the PD shows a clearly increasing pattern, the LAD does not have any relation with the ECC's coefficient of variation.
In light of these considerations, to further reduce the variability of the ECC estimator, focusing on the obligors with high credit quality is crucial. This finding will constitute the guideline to attempt to improve the IS scheme for Phase II.

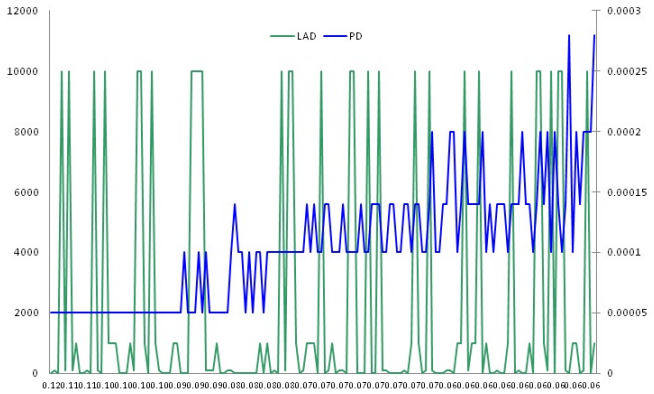

Figure 3.5: Loss at default (scale reported on the right vertical axis) and probability of default (scale on the left vertical axis) of the 10% most unstable obligors, whose corresponding ECC's CV is shown on the x-asis in descending order.

3.4.3 GL's approach and our modifications

In this section, we present the results obtained by applying the approach developed by Glasserman and Li to the estimator introduced in (3.13). We start from the "conditional IS" (CIS), which shift only the idiosyncratic term by selecting a twisting parameter to increase the individual conditional default probabilities. Next, we combine this approach to a shift in the systemic factor mean (as in the current IS scheme), referred to as "two-step IS" (2IS). It turns out that GL's approach is not effective for ECC estimation. However, inspired by their work, we will apply various modification which will prove to remarkably reduce the estimator's variability.

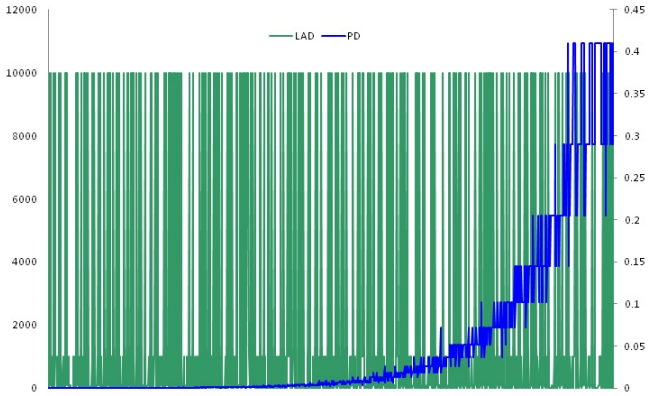

Figure 3.6: Loss at default (scale reported on the right vertical axis) and probability of default (scale on the left vertical axis) of the all portfolio plotted against the corresponding ECC's CV in descending order.

Due to the many simulation experiments carried out in a short time span, the results in this section are based on 50 runs with 20000 simulations. Since all results confirm that estimators of type (3.13) are consistent (see subsection 3.4.4), we will base the comparison on the coefficient of variation (3.3).

Conditional IS. By shifting only the conditional probabilities of default, as presented in Section 2.2, results in terms of variance reduction are better when compared to CMC, but not as good as current IS (see Figure 3.7, which plots the CV in descending order). In fact GL show that, for high correlation, CIS is not more effective than standard MC. However, this seems to hold even for low correlation ($\rho = 0.2$). Actually, in our framework, when the correlation is low the variability is even higher. In fact, despite the quantile is smaller[16], conditional PD become smaller as well. While this may facilitate the quantile estimation, it makes it harder to estimate ECC, especially for the firms with very low PD.
In addition, our portfolio is very different from the simple portfolio used in GL. In their numerical examples (which are claimed to show that their approach works also for more general cases) the PD and LAD are much more homogeneous than the one in our portfolio[17]. This could also explain why the "optimal" θ does not yield better results than an arbitrary theta, as can be seen from Figure 3.7. But the main reason for this is that GL's estimator is different from the ECC estimator in (3.13). Indeed, GL aim at estimating the tail's probability of the whole portfolio loss. Conversely, our focus is on the individual loss of each obligor, conditional on the portfolio loss. This makes the search for θ^* a useless effort (also) for the sake of Phase II estimator's stability.

[16]When the losses are normally distributed, in single factor models, quantiles are constant multiples of the correlation parameter ρ. This holds more generally for elliptic distributions.

[17]The probabilities of default in GL vary between 0 and 2% with a mean of 1 %, and the exposures are 1, 4, 9, 16, 25 with 200 obligors at each level.

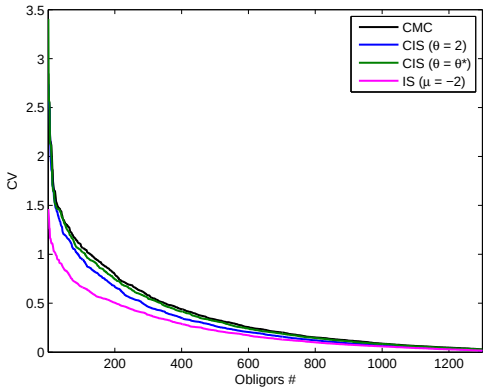

Figure 3.7: Descending ordered ECC's CV obtained using different methods: CMC (black), current IS (magenta), Conditional IS with θ^* (green) and Conditional IS with $\theta = 2$ (blue).

Two-step IS. Because conditional IS does not lead to any variability reduction (compared to current IS), we now combine CIS with both the current factor's mean shift ($\mu = -2$) and the optimal shift (μ^*) presented in Section 2.3.
When we add a shift of -2 in the factor mean to a twist in the conditional default probabilities with $\theta = 2$ (chosen arbitrarily), we notice a tremendous reduction in the CV for the firms with highest LAD size (Figure 3.8). In fact, $p_i(\theta, Z)$ depends exponentially on the exposure. Recall the discussion surrounding (2.3). Since GL's estimator aims at the whole portfolio loss distribution, it is indeed natural to give more emphasis to the instances with higher LAD, which, in case of default, may quickly bring the total loss to the desired quantile level. As we wish to reduce the variability of each contribution (in particular for low PD obligors), we keep the LAD uniform in the computation of the twisted conditional PD, which now become:

$$p_i^*(\theta, Z) = \frac{p_i(Z)\mathrm{e}^{\theta}}{1 + p_i(Z)(\mathrm{e}^{\theta} - 1)} \quad i = 1, \ldots, n. \tag{3.16}$$

By applying this simple modification, an arbitrary shift of $\theta = 2$ and $\theta = 4$ leads to a significant improvement for all the class of obligors, as illustrated in Figure 3.9.

On the other hand, the "optimal" shift of the common factor mean, as expectable from our previous discussions, does not improve the stability of the EC contribution estimator. This is in fact optimal only for Phase I. However, it helps to improve the precision of the estimator by reducing the quantile interval, as it will be illustrated below (Table 3.4).

Minimum variance estimator

Figure 3.9 suggests that we can further decrease the variability by arbitrarily increasing $p_i^*(\theta, Z)$. In fact, the higher the twisting parameter θ, the lower the CV

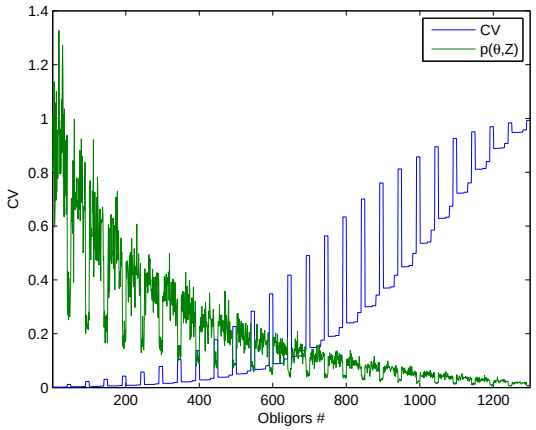

Figure 3.8: Conditional IS with $\theta = 2$: coefficient of variation (green) and corresponding conditional twisted PD (blue).

of the resulting ECC estimator. Since the ECC estimator is most unstable for the obligors with the lowest default probability ($p = 0.00005$ in our portfolio), we could choose θ such that their twisted conditional PD is 1, that is, we set all $p_i^*(\theta, Z) \equiv 1$. In other words, we let each obligor default in every scenario. It follows that the corresponding likelihood ratio term γ_{ϵ_i} amounts to the conditional default probability $p_i(Z)$. In fact, for any i,

$$\gamma_{\epsilon_i} = \left(\frac{p_i(Z)}{p_i^*(\theta, Z)} \right)^{Y_i} \left(\frac{1 - p_i(Z)}{1 - p_i^*(\theta, Z)} \right)^{1 - Y_i},$$

and since the default indicator $Y_i = 1$ for all $i = 1, \ldots, n$, then $\gamma_{\epsilon_i} = \frac{p_i(Z)}{p_i^*(\theta, Z)} = p_i(Z)$, by setting $p_i^*(\theta, Z) = 1$.
The variance reduction we obtain is striking, as it is evident from Figure 3.10. Figure 3.11 zooms the estimator's CV, which shows 26 decreasing steps. These correspond to the 26 PD categories: obligors in the same category turn out to have the same CV, in agreement with our finding that the probability of default is the only driver of the ECC's variability. Notice that a jump in the CV can still be seen between the first and the second category. This is however extremely small compared to the one we aimed to reduce (Figure 3.4).
It is interestingly to investigate how the estimator can retain consistency (see subsection 3.4.4). In fact, a too large shift towards the "importance area", may lead to a biased estimator in finite samples. By making all obligors default, the total loss would indeed be beyond the quantile level. However, weighting each individual loss with its obligor-specific likelihood ratio γ_{ϵ_i}, as shown in the derivation of estimator (3.13), brings the portfolio loss around the quantile level. Differently from the other methods, we collect a full vector of losses at each replication, thus boasting the stability of the estimator.

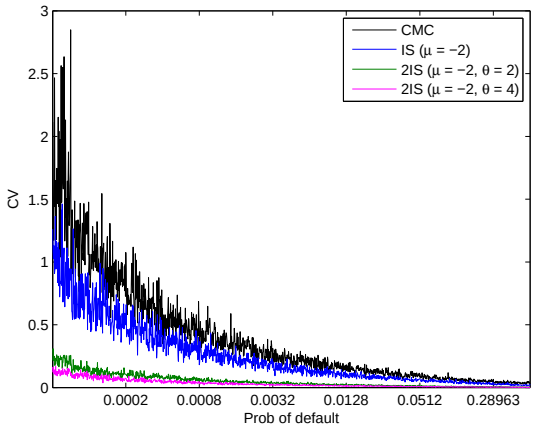

Figure 3.9: ECC's coefficient of variation obtained using the two-step IS with $\theta = 2$ (green) and $\theta = 4$ (magenta) compared with CMC (black) and current IS (blue). The x-axis shows the corresponding probability of default.

Finally, it is worth noting that the shift in the systemic factor's mean has a crucial role in this framework. A (negative) shift in μ makes the conditional default probabilities $p(Z)$ bigger. Hence, the required shift to lead all the obligors to default is relatively smaller. It follows that the resulting likelihood ratio terms γ_{ϵ_i} are less small. This helps to reduce finite-sample bias as well as to decrease the variability and to increase the precision of the estimator.

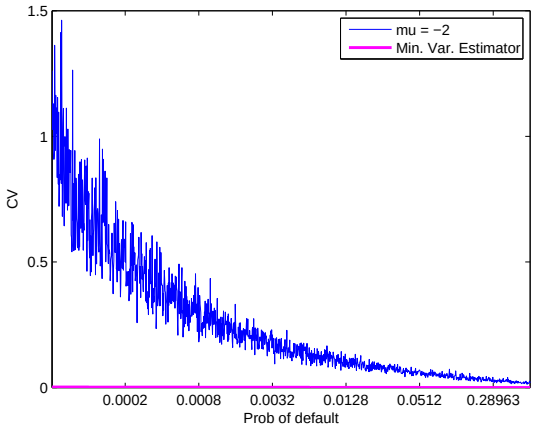

Figure 3.10: ECC's coefficient of variation obtained using the current IS (blue) and the "Min. variance" estimator (magenta). The x-axis shows the corresponding probability of default.

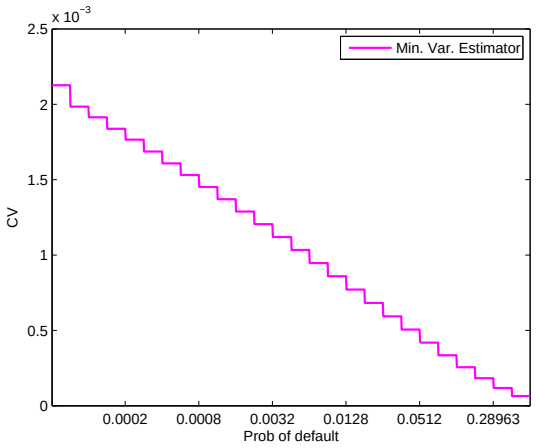

Figure 3.11: ECC's coefficient of variation obtained with the "Min. variance" estimator. The x-axis shows the corresponding probability of default.

3.4.4 Normal idiosyncratic factor

Let us now diverge from GL's approach and attempt to retain the normality framework of the Gaussian copula factor model. Namely, instead of using a twisted Bernoulli to shift the conditional default probabilities, we aim at shifting the standard normal idiosyncratic factor ϵ_i by choosing an adequate normal distribution $N(\theta_i, 1)$, for all $i = 1, \dots, n$.
If we shift the idiosyncratic factor's mean by an arbitrary shift $\theta = 2$ for all the obligors, we do not give enough stress to the firms with the highest credit quality, whose EC contribution remain very unstable. Furthermore, by choosing an arbitrary shift, it is easy to obtain a biased estimator. The mean is in fact strongly over-biased for low PD, and under-biased for high PD, and so the opposite holds for the CV. This causes the unusual pattern in the estimator's coefficient of variation illustrated in Figure 3.12.

Since the lower the default probability p_i, the more we need to shift θ_i to let obligor i default, one could proceed as follows. Recall that the default indicator in one-factor models has the form

$$Y_i = \mathbb{1}\{\rho Z + \sqrt{1-\rho^2}\epsilon_i \leq \Phi^{-1}(p_i)\}.$$

Once a systemic factor's realization $Z = z$ has been drawn, we can take the shift in the mean of the idiosyncratic term ϵ_i as

$$\theta_i^* = \min\left\{\frac{\Phi^{-1}(p_i) - \rho Z}{\sqrt{1-\rho^2}}, 0\right\}. \tag{3.17}$$

It is easy to see that in this way we obtain $\mathbb{E}_{\theta_i^*}[Y_i|Z] = 1$. We truncate θ_i^* in 0 to avoid obligors with high PD to have a large positive shift, which would negatively

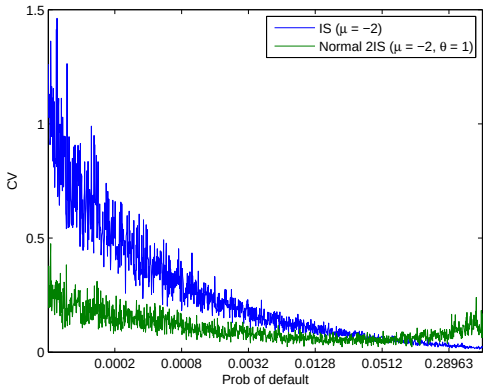

Figure 3.12: ECC's coefficient of variation when the estimator is biased. The increasing CV for higher probabilities of default shows the under-bias in the corresponding estimates.

affect their ECC stability.
The resulting estimator's CV is plotted in Figure 3.13. It can be seen that the variability is not as low as in the minimum variance estimator case. In fact, the default indicator Y_i is one only on average, and this does not imply that every obligor defaults in each scenario. Nevertheless, this method leads to a notable variance reduction; in addition, the normality framework may be more convenient for many practical applications.
Remark. Similarly to the reasoning that led to the minimum variance estimator, we could directly set ϵ_i equal to (3.17) and compute the corresponding likelihood ratio. This would make all obligors default at each iteration and would substantially reduce the estimator's variability. However, in this framework, the resulting estimator is heavily biased, therefore results are not reported.

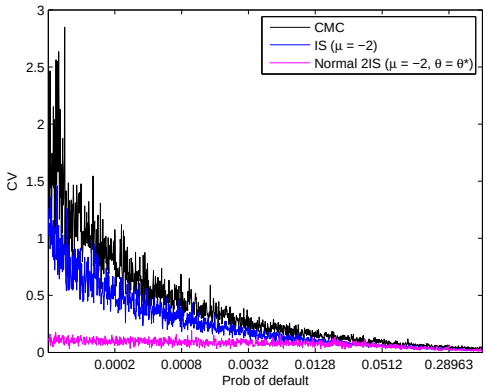

Figure 3.13: ECC's coefficient of variation obtained with $\epsilon \sim N(\theta^*, 1)$ (magenta), compared with CMC (black) and current IS (blue). The x-axis shows the corresponding probability of default.

Shift in the variance. It is worth considering whether changing the variance of the factors may improve the IS schemes. To experiment this, we go back to considering the change in the systemic factor's pdf. In such a framework, a shrink in the factor's variance could be appealing to have more samples clustered around $L = \text{VaR}_\alpha$.
In order to select an appropriate first and second moment for the conditional importance density of Z, we can employ a pilot run[18] and store the systemic factor's draws that yield to portfolio loss around the interval Q. In the second step, we can use the sample mean and variance of the stored factors to update the normal distribution, which becomes a normal $N(\hat{\mu}, \hat{\sigma}^2)$, where $\hat{\mu}$ and $\hat{\sigma}^2$ indicates the estimates of $\mathbb{E}[Z|L \in Q]$ and $\sigma^2[Z|L \in Q]$, respectively.
When we shift both mean and variance, simple algebra shows that the likelihood ratio introduced in (1.10) becomes[19]

$$\gamma_Z = \sigma \exp\left\{ \frac{\mu^2/2 - Z\mu + \frac{1}{2}(1-\sigma^2)Z^2}{\sigma^2} \right\}.$$

A pilot run with with 10000 simulations takes 4.9 seconds with Matlab. This additional computing time becomes negligible as the number of simulations in the second step increases. The estimated conditional moments are $\hat{\mu} = -3.0865$, $\hat{\sigma}^2 = 0.1166$. Interestingly, $\hat{\mu}$ is remarkably closed to shift obtained by solving GL's optimization program $\mu^* = -3.1138$. In fact, it leads to a very precise estimator (i.e. narrow interval Q). On the other hand, the estimator's variance is not significantly reduced. This was however expectable for our previous discussions, as this method does not focus on obligors with low PD. Nonetheless, Figure 3.14, which compares this method with current IS, shows a less volatile CV. This suggests that a shift in the variance of Z may help to control the estimator's fourth moment.

Using linear regression models. Along with the systemic factors, during the pilot run we could store the idiosyncratic terms of those losses whose sum falls in Q. Then, we can perform an OLS regression of each ϵ_i on the stored draws of Z. That is, we estimate the linear regression model

$$\epsilon_i = \beta_{1i} + \beta_{2i} Z + \eta_i, \quad i = 1, \dots, n$$

where the two parameters can be estimated as

$$\hat{\beta}_i = (X'X)^{-1} X' \epsilon_i,$$

where X is a matrix with K rows (the losses in Q) and 2 columns (a column of ones and a column with the Z values), and ϵ_i is a $n \times K$ matrix[20]. We could then specify

[18] The starting distribution could be, for instance, the current IS distribution, $N(-2, 1)$.

[19] Note that when we cope with one single-factor models, because $X = \rho Z + \sqrt{1-\rho^2}\epsilon$, in order to compute the likelihood ratio we need to take into account the factor loading ρ. In fact, if we take $Z \sim N(\mu, \sigma^2)$, the importance density is actually $N(\rho\mu, (\rho^2\sigma^2 + \sqrt{1-\rho^2}))$.

[20] In this example, ϵ_i is (1300×2000).

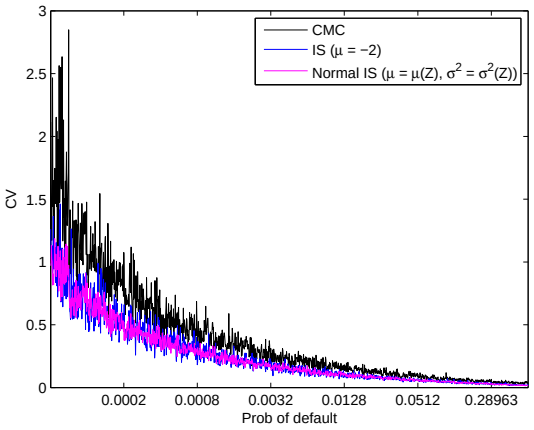

Figure 3.14: ECC's coefficient of variation obtained using $Z \sim N(\hat{\mu}, \hat{\sigma}^2)$ (magenta), compared with CMC (black) and current IS (blue). The x-axis shows the corresponding probability of default.

the conditional importance density for each ϵ_i given Z, as a normal distribution with first moment

$$\mathbb{E}[\epsilon_i|Z] = \hat{\beta}_{1i} + \hat{\beta}_{2i}Z.$$

Next, one could compute the residuals $\hat{\eta}_i = \epsilon_i - X\hat{\beta}_i$ and use $s_i^2 = (\hat{\eta}_i'\hat{\eta}_i)/(n-2)$ as conditional variance. Therefore, for every ϵ_i, the conditional importance density given Z is computed as $N(\hat{\beta}_{1i} + \hat{\beta}_{2i}Z, s_i^2)$. Alternatively, one could include also the squared Z in the regression, and take the conditional IS density to be $N(\hat{\beta}_{1i} + \hat{\beta}_{2i}Z + \hat{\beta}_{3i}Z^2, s_i^2)$, with s_i^2 updated accordingly.
Since we want to increase the number of defaults, there is no advantage in specifying a negative conditional mean. Thus, we take $\mathbb{E}[\epsilon_i|Z]$ as the minimum between the fitted values $\hat{\beta}_i X$ and 0. After this modification, this method leads to a reduction in the CV (using the current IS as benchmark) of 9%, and 11% when Z^2 is added in the model.
More accuracy could be achieved by replacing the normal with a Student-t distribution with low degrees of freedom. This allows to sample more idiosyncratic draws in the tail of the distributions, and thus to observe more defaults. In fact, specifying the idiosyncratic factor's distribution as t_8 and t_4, with conditional mean and variance as described above (with squared Z in the regression), permits to reduce the CV by a factor of 38% and 58% respectively (see Figure 3.15).

Quantile intervals. Table 3.4 compares some of the developed methods not only in terms of stability (here measured as the average of ECC estimator's CV) but also in terms of precision, measured as width of the quantile interval.
These results are based on 50 $\times 10^4$ simulations with $K = 2000$ losses around the

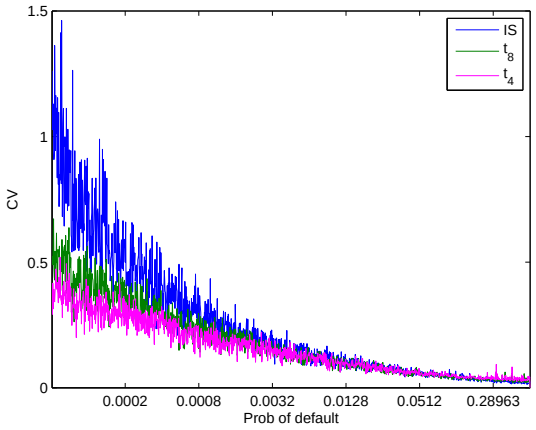

Figure 3.15: ECC's coefficient of variation obtained combining $Z \sim N(\hat{\mu}, \hat{\sigma}^2)$ with $\epsilon \sim t_8$ (green) and t_4 (magenta) with moments estimated by means of an OLS regression with squared terms Z^2.

quantile. For a fair comparison, the 99.9%-quantile estimated in Phase I is set $\hat{q} = 971724$ for all the methods.

Method	mean(CV)	Q (€×1000)
CMC	0.3885	[761.840, 1334.862]
IS ($\mu = -2$)	0.2466	[960.895, 983.107]
2IS ($\mu = -2, \theta = 2$)	0.0861	[961.834, 980.308]
2IS ($\mu^*, \theta = 2$)	0.0881	[966.031, 977.420]
2IS (μ^*, θ^*)	0.0003	[966.803, 976.823]
IS norm ($\mu = -2, \theta^*$)	0.0749	[961.077, 981.194]
IS norm (μ^*, θ^*)	0.0766	[967.354, 976.144]
IS $N(\hat{\mu}, \hat{\sigma}^2)$	0.2406	[970.761, 972.609]
IS $N(\hat{\mu}, \hat{\sigma}^2)$, OLS	0.2220	[969.854, 973.641]
IS $N(\hat{\mu}, \hat{\sigma}^2)$, OLS ($t_4$)	0.1560	[970.609, 972.661]

Table 3.4: ECC estimator's stability (measured as average CV) and precision (width of the quantile interval) for different methods.

Table 3.4 is a nutshell of the most relevant IS schemes implemented in this work. It can be seen that current IS ($\mu = -2$) manages to substantially reduce the width of CMC interval, as well as to increase the stability. The average CV can be made three times smaller by combining current IS with a twist of $\theta = 2$ in the modified conditional PD given in (3.16) which however have no significant impact on the precision. Conversely, the optimal shift in the systemic factor mean derived in GL (μ^*) permits to straddle the quantile in a narrower interval.
The Table confirms that our minimum variance estimator is by far the most accurate

one. Combining it with the optimal GL's shift (2IS (μ^*, θ^*)) improves the performances also in terms of precision. A further improvement in the precision can be achieved with a normal distribution for Z, with the conditional mean and variance estimated in a pilot run. In fact, this method proves to yield the narrowest interval around the quantile.

On the (asymptotic) unbiasedness

To provide evidence on the consistency of the developed estimators, we show a graph of the ECC estimates averaged ten by ten. In fact, for each bucket of PD in the test portfolio, there are 10 obligors with the same EAD and LGD. These clearly constitute replications of the same obligor profile. The graphs in Figure 3.16 and 3.17 depict the ECC estimated with the "Min. Var. IS" estimator, the most critical one in terms of possible sample bias. In order to be able to compare the means in the pictures, we have split the obligor with lower exposure (Figure 3.16) and the one with higher exposure (Figure 3.17). One run with 200000 simulations already suggests the asymptotic unbiasedness. By increasing the number of simulations the sample bias tends to vanish, thus corroborating the consistency of the estimator.

To further enhance the asymptotic unbiasedness of our estimators, one could esti-

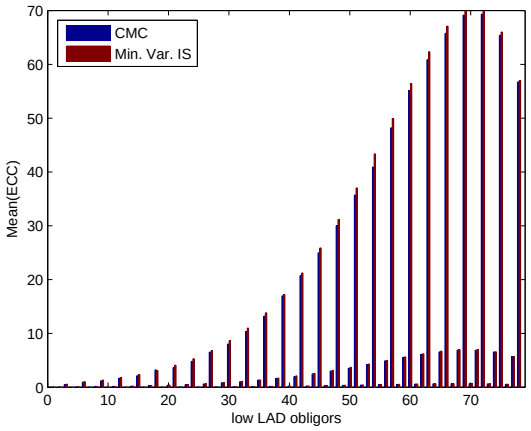

Figure 3.16: Economic Capital Contributions (averaged ten by ten) for "low LAD" obligors estimated with Standard MC (blue) and Min. Var. estimator (red).

mate the mean square error (MSE) presented in Section 3.1. The estimated MSE should approach the variance of the ECC estimator. The classic way of estimating MSE is to generate N replications of the K vectors of individual losses $(L^1, \ldots, L^K)$ in Q. This, however, would require an enormous effort. An alternative and more efficient way is the bootstrap method, which consists on re-sampling with replacement from the vectors of losses in Q already stored in the memory until we get another set of K vectors around the quantile. This is repeated B times and the outputs, acting as replications of the original losses, can be used to estimate the MSE. The

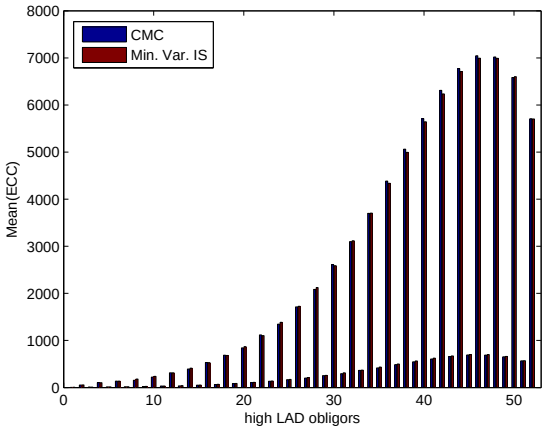

Figure 3.17: Economic Capital Contributions (averaged ten by ten) for "high LAD" obligors estimated with Standard MC (blue) and Min. Var. estimator (red).

underlying distribution function is however the empirical distribution rather than distribution of the real data. However, bootstrap method in importance sampling framework is not straightforward and will be left for future research.

3.4.5 Multi-factor Models

Finally, we test the applicability of our minimum variance method to multi-factor models, whose dynamics are given in (1.29).
Following the setting in use at RBS, we assume that each obligor's return is captured only by one factor; the sensitivity of each return to a certain factor is given by the factor loading ρ. We split each group of obligors with the same profile (10 obligors) into two groups of 5 obligors, whose return is then captured by a different factor. The difference with the single-factor case is that now different factors may be correlated through their covariance (matrix). Notice that this, first of all, affects the shift in the vector of means. In fact, if we change the distribution of Z from $N(0,\Sigma)$ to $N(\mu,\Sigma)$, the corresponding likelihood ratio is given by

$$\gamma_Z = \frac{\exp\left\{\frac{1}{2}z'\Sigma^{-1}z\right\}}{\exp\left\{\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)\right\}} = \exp\left\{\frac{1}{2}\mu'\Sigma^{-1}\mu - z'\Sigma^{-1}\mu\right\}.$$

In order to study how Σ influences the VaR and the VaR contributions, we first consider a two-factor model for four different covariances: $\sigma_{12} = \{0, 0.25, 0.75, 0.999\}$. Note that we cannot take the covariance to be exactly one, otherwise Σ could not be inverted, being the factors' variance also unitary. We then investigate the role of the factor loading ρ, carrying out all the experiments in a situation of low asset correlation ($\rho = 0.2$) and high asset correlation ($\rho = 0.6$). Results are illustrated in Table 3.5.

As discussed before, the Value at Risk is very sensitive to ρ. Similarly to one-factor models, where (for elliptical distributions) the quantile is a multiple of ρ, also in the two-factor case a lower asset correlation corresponds to a thinner tail of the portfolio loss distribution, and hence to a smaller VaR[21]. Interestingly, the (average) CV of Phase II, proves to increase with ρ. This may be due to the higher sensitivity of returns (and hence of defaults) to movements in the assets return, which is reflected in a higher variability of the likelihood ratios. See the wider bands (corresponding to obligors modeled with different systemic factors) in the right panel of Figure 3.18 compared with those in the left panel.
For a given ρ, the lower the correlation between the two factors the smaller the quantile, meaning that some diversification effects is observed. On the other hand, the average CV increases, because of the more variability in the systemic factors. When σ_{12} approaches 1, the two draws of Z are almost the same, therefore the model reduces to a single-factor model and results are similar to the one previously obtained. Finally, notice that the increased volatility in the estimator's CV for high PD obligors might be caused by a small under-bias in the ECC estimates which vanishes by increasing the number of simulations. However, we already stressed that, for such high PD instances, IS methods would not be needed.

	σ_{12}	$\hat{q}$ (€×1000)	mean(CV_{ECC})	Q (€×1000)
$\rho = 0.2$	0	317.217	0.0569	[312.068, 321.746]
	0.25	331.616	0.0302	[326.875, 336.259]
	0.75	354.562	0.0135	[349.951, 358.865]
	0.999	367.653	0.0010	[364.252, 371.992]
$\rho = 0.6$	0	665.905	0.1505	[662.587, 669.212]
	0.25	738.644	0.0726	[734.199, 742.262]
	0.75	895.488	0.0291	[887.652, 902.029]
	0.999	971.563	0.0015	[961.730, 981.349]

Table 3.5: Quantile estimates (Phase I), average ECC's CV and quantile interval (Phase II) under two-factor models with $\rho = 0.2$ (left) and $\rho = 0.6$ (right), for various factors' covariances σ_{12}.

Next, we employ a five-factor model, with four different Σ: the 5×5 identity matrix, a matrix with uniform correlations in $(0.3, 0.5)$, referred to as "low" in Table 3.6, a matrix with uniform correlations in $(0.6, 0.8)$ (which we call "high"), and a matrix with 1 in the diagonal and all other elements equal to 0.999 ("0.999"). Here the asset correlations are set to $\rho = 0.4$, for the obligor with lower exposures (EAD=1,10,100), and to $\rho = 0.6$ for higher exposure loans (EAD=1000, 10000). It is in fact plausible that clients such as banks or other financial institutions have a

[21] Due to the short time available and the fact that our focus is on Phase II, we run only one replication for Phase I. Thus, the VaR's CV could not be calculated.

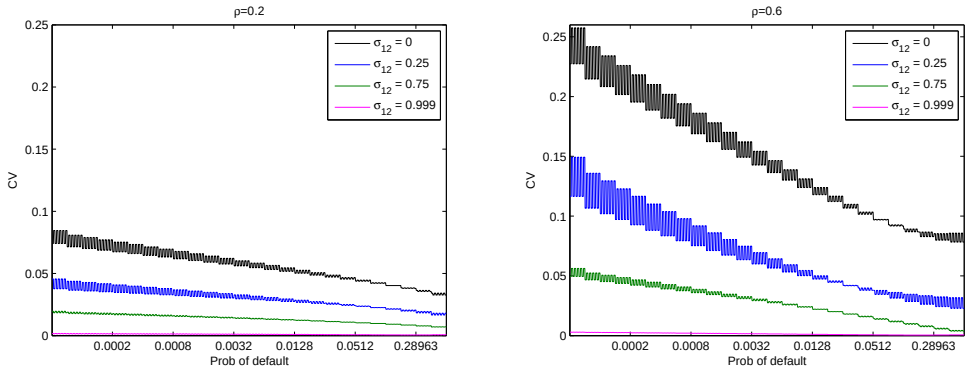

Figure 3.18: ECC's coefficient of variation obtained with the minimum variance estimator under two-factor models for various asset correlation (ρ) and factors' covariance σ_{12}.

higher correlation with the systemic factor.
Notice that since we have a larger number of factors, we have a higher diversification effect and hence a bigger difference in the quantiles between low and high correlation. Note that when the factors are (almost) perfectly correlated, the estimated quantile is smaller than the one in the single-factor case because we have taken 3/5 of the returns with loading $\rho = 0.4$. Obviously, setting $\rho = 0.6$, we obtain a quantile around 970000 €. For what concerns Phase II, the model parameters prove to affect the CV in the same way as in the two-factor case, but with a more pronounced effect.
Finally, we compare Phase II's coefficient of variation between current IS and our minimum variance estimator. Firstly, in the two-factor model with $\rho = 0.6$ and $\sigma_{12} = 0.75$ (left panel of Figure 3.20), and secondly, in the five-factor case with "high" correlation (right panel of Figure 3.20). It is seen that, even in the multi-factor case, the new IS scheme largely outperforms the current one.

Σ	$\hat{q}$ (€×1000)	mean(CV_{ECC})	Q (€×1000)
I_5	413.371	0.4123	[411.710, 414.851]
low	562.614	0.0630	[560.220, 565.399]
high	664.904	0.0360	[660.132, 669.174]
0.999	773.782	0.0017	[766.458, 779.651]

Table 3.6: Quantile estimates (Phase I), average ECC's CV and quantile interval (Phase II) under five-factor models with various covariance matrices Σ.

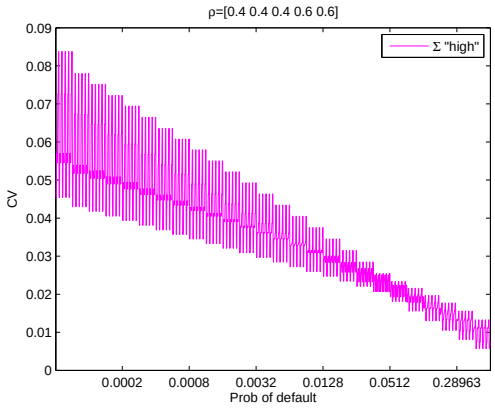

Figure 3.19: ECC's coefficient of variation obtained with the minimum variance estimator under a five-factor model with "high" factors' covariance matrix.

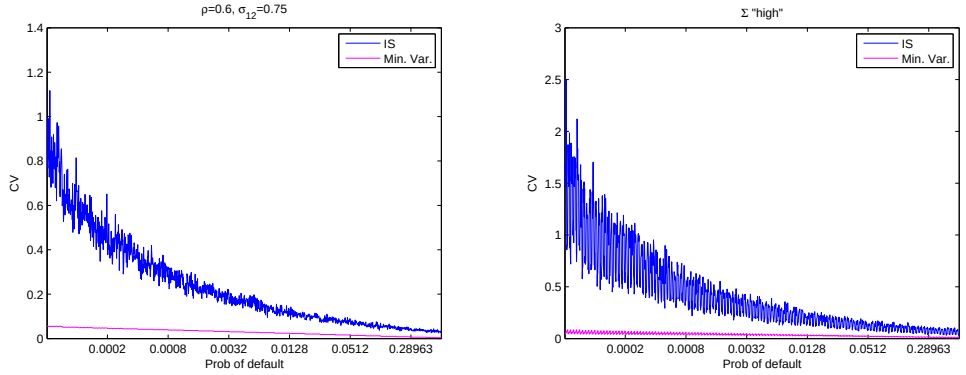

Figure 3.20: ECC's coefficient of variation obtained with the "Min. variance" estimator (magenta) and current IS (blue) under a two-factor model with $\rho = 0.6, \sigma_{12} = 0.75$ (left), and a five-factor model with "high" Σ (right).

3.5 Conclusions

In this work, we extensively reviewed the importance sampling procedures developed in Glasserman and Li (2005) and applied them to the estimation of economic capital contributions based on VaR. The objective was to analyze the current RBS importance sampling scheme in order to identify its weaknesses, and propose a method to improve it. Although GL's approach proved not to immediately lend itself to the estimation of VaR contributions, we used it a basement for our study. In fact, their procedure is optimal only for the estimation of the tail of the whole portfolio's loss. Nevertheless, we proposed some tailor-made modifications which turned out to yield significant variance reduction.

Among the methods we proposed, the so called "minimum variance estimator" proved to largely outperform the others. This method, which lets all the oblig-

ors default at each iteration and conveniently corrects with their obligor-specific likelihood ratio, constitutes a notable improvement to the current scheme. In fact, with the test portfolio utilized in the simulation experiments, it led to an average variance reduction higher than 800% compared to current IS, and about 1300% compared to crude Monte Carlo.
In addition, this method proved to yield promising results even in the more complicated case of multi-factor models.
As a final remark, we notice that all the experiments have been performed in a default-only framework. When migrations models are concerned, shifting all the obligors to default might not be convenient. The applications of these methods to migrations models will be left for future research.

Appendix A

MATLAB Code

A.1 Phase I

```
R = 10;          % # of replications
N1 = 2e+4;       % # of simulations Phase 1
alpha = 0.999;   % confidence level
rho = 0.6;       % factor loading
muz = -2;        % current IS shift
QL = zeros(R,1);
k = N1*(1-alpha);

for j=1:R;

Z = muz + randn(1, K1);          % (common) factor return
LR = exp(muz^2/2 - muz*Z);       % likelihood-ratio
L = [];
i = 1;

while i <= K1;
    % Obligor return
    X = rho.*Z(i) + sqrt(1-rho.^2).*randn(N, 1);    % N=1300

    % Default Indicator
    Y = X < norminv(PD);

    % Portfolio losses
    Li = EAD.*LGD.*Y;
    L = [L sum(Li, 1)];
    i = i + 1;
end
```

```
% Quantile estimation
A = [L' LR'];
B = sortrows(A, [-1]);
C = [B cumsum(B(:,2))];
q = min(find(C(:,3) >= k));
QL(j) = C(q, 1)
end

QL = mean(QL)
```

A.2 Phase II

```
R = 50;        % # replications
N2 = 2e+5;   % # simulations Phase 2
K = round(0.01*N2);  % # losses around QL
muz = -2;    % mean shift Z
EC = zeros(N, R);

for j = 1:R;

Ll = []; Lr = [];
Lls = []; Lrs = [];
LRl = []; LRr = [];
LRls = []; LRrs = [];
i = 1;

while i <= N2;

    % Draw factor returns and likelihood ratio
    Z = muz + randn;
    LR = exp(muz^2/2 - Z*muz);

    % Compute pd conditional on Z and twisted cond. pd
    pZ = normcdf((-Z*rho + norminv(PD))/sqrt(1-rho^2));
    p_theta = 1;  % Min. Var. IS

    % Generate default indicators, LR, Li and L
    Y = rand(N,1) < p_theta;
    g = ((pZ./p_theta).^Y).*(((1-pZ)./(1-p_theta)).^(1-Y));
    Li = EAD.*LGD.*Y.*g;
    L = sum(Li);
```

```
    if L <= QL;
        if size(Ll, 2) < B/2
            Ll = [Ll [Li; LR]]; % Store loss vector and LR
            % Sort according to portfolio loss
            Ll = [Ll; sum(Ll(1:end-1,:))];
            [y1, I1] = sort(Ll(end,:));   % or (sort(Ll', [end])'
            Lls = Ll(1:end-1,I1);
            Ll = Ll(1:end-1,:);
        end

        if L > sum(Lls(1:end-1,1));
            Lls(:,1) = [Li; LR];
            % Sort again Lls
            Lls = [Lls; sum(Lls(1:end-1,:))];
            [y2, I2] = sort(Lls(end,:));
            Lls = Lls(1:end-1,I2);
        end

    else
        if size(Lr, 2) < B/2
            Lr = [Lr [Li; LR]];
            Lr = [Lr; sum(Lr(1:end-1,:))];
            [y3, I3] = sort(Lr(end,:));
            Lrs = Lr(1:end-1,I3);
            Lr = Lr(1:end-1,:);
        end

        if L < sum(Lrs(:,end));
            Lrs(:,end) = [Li; LR];
            Lrs = [Lrs; sum(Lrs(1:end-1,:))];
            [y4, I4] = sort(Lrs(end,:));
            Lrs = Lrs(1:end-1,I4);
        end
    end

    i = i + 1;

    if mod(i, N2/10) == 0
       display(i)
    end

end

Lp = [Lls(1:end-1,:) Lrs(1:end-1,:)];    % Losses in Q
```

```
[sum(Lp(:,1)) sum(Lp(:,end))];              % Quantile interval
gg = [Lls(end,:)  Lrs(end ,:)];             % Lik-ratios
%
display([sum(Lp(:,1)) sum(Lp(:,end))])
display(j)
%
a = repmat(QL./sum(Lp,1), N, 1);    % Ensure VaR decomposition
Lp = Lp.*a;
M = sum(Lp.*repmat(gg, N, 1), 2)./sum(gg);
EC(:,j) = M - EAD.*LGD.*PD;         % ECC
end
```

Bibliography

[1] Artzner P., Delbaen F., Eber J.M., Heath D. (1999), "Coherent measures of risk", *Mathematical Finance*, **9**, 203-228.

[2] Avranitis, A., Gregory, J. (2001), *Credit: The Complete Guide to Pricing, Hedging and Risk Management*, Risk Books, London.

[3] Basel Committee on Bank Supervision (2004), "International Convergence of Capital Measurement and Capital Standards: A Revised Framework". Available at www.bis.org.

[4] Brent, R.P. (1973), *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey.

[5] Durbin, J., Koopman S.J. (2012), *Time Series Analysis by State Space Methods: Second Edition*, Oxford University Press, Oxford.

[6] Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, Third Edition, Wiley, New York.

[7] Gagliardini P., Gourieroux, C. (2005), "Stochastic Migration Models with Application to Corporate Risk", *Journal of Financial Econometrics*, **3**, 188-226.

[8] Glasserman, P. (2003), *Monte Carlo Methods in Financial Engineering*, Springer, New York.

[9] Glasserman, P., Li, J. (2005), "Importance sampling for portfolio credit risk", *Management Science*, **51**(11), 1643-1656.

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

[11] Glasserman, P., Kang, W., Shahabuddin, W. (2007), "Large deviations in multifactor portfolio credit risk", *Mathematical Finance*, **17**(3), 345-379.

[12] Glasserman, P., Kang, W., Perwez, S. (2008), "Fast simulation of multifactor Portfolio Credit Risk", *Operation Research*, **56**(5), 1200-1217.

[13] Glynn, P.W. (1996), "Importance sampling for Monte Carlo estimation of quantiles", In Mathematical Methods in Stochastic Simulation and Experimental Design: Proc. 2nd St. Petersburg Workshop on Simulation, 180-185.

[14] Gupton, G., Finger, C., Bhatia, M. (1997), "CreditMetrics Technical Document", *J.P. Morgan & Co.*, New York.

[15] Hamilton, J. (1994) *Time Series Analysis*, Princeton University Press, New York.

[16] Heitfield, E., Burton, S., Chomsisengphet, S. (2006), "Systematic and idiosyncratic risk in syndicated loan portfolios", *Journal of Credit Risk*, **2**(3).

[17] Hoogerheide, L.F., Van Dijk, H.K. (2010), "Bayesian forecasting of Value at Risk and Expected Shortfall using adaptive importance sampling", *International Journal of Forecasting*, **26**, 231-247.

[18] Huang, P., Subramanian, D., Xu, J. (2010), "An importance sampling method for portfolio CVaR estimation with Gaussian copula models", 2010 Winter Simulation Conference, 2790-2800.

[19] Hull, J.C. (2011), *Options, Futures, and Other Derivatives*, Prentice Hall.

[20] Kalkbrener, M., Lotter, H., Overbeck, L. (2004), "Sensible and efficient capital allocation for credit portfolios", *RISK*, **17**(1), S19-S24.

[21] Koopman, S.J., Lucas, A. (2005), "Business and Default Cycles for Credit Risk", *Journal of Applied Econometrics*, **20**, 311-323.

[22] Koopman, S.J., Lucas A., Monteiro, A. (2008), "The Multi-State Latent Factor Intensity Model for Credit Rating Transitions", *Journal of Econometrics*, **142**(1), 399-424.

[23] Kroese, D.P., Taimre, T. , Botev Z.I. (2011), *Handbook of Monte Carlo Methods*, John Wiley & Sons, New York.

[24] Li, D.X. (2000), "On Default Correlation: A Copula Function Approach", *Journal of Fixed Income*, **9**(4), 43-54.

[25] Lucas, A., Klaassen, P., Spreij, P., Staetmans, S. (1999), "An analytic approach to credit risk of large corporate bond and loan portfolios", *Research Memorandum*, Vrije Universiteit Amsterdam.

[26] Marsaglia, G., Tsang, W.W. (2000), "The Ziggurat Method for Generating Random Variables", *Journal of Statistical Software*, **5**(8).

[27] McNeil, A.J., Frey, R., Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, New York.

[28] Merino, S., Nyfeler, M.A. (2004), "Applying importance sampling for estimating coherent credit risk contributions", *Quantitative Finance*, **4**, 199-207.

[29] Merton, R.C. (1974), "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, **29**, 449-470.

[30] Morokoff, W.J. (2004), "An importance sampling method for portfolios of credit risky assets". Winter Simulation Conference, 596-604.

[31] Mausser, H., Rasen, D. (2007), "Economic Credit Capital Allocation and Risk Contributions", *Handbooks in Operations Research and Management Science* **15**, 681-726.

[32] Nelsen, R.B. (1999), *An Introduction to Copulas*, Springer-Verlag, New York.

[33] Rubinstein, R.Y.,Melamed, B. (1998), *Modern Simulation and Modeling*, John Wiley & Sons, New York.

[34] Rubinstein, R.Y., Kroese, D.P. (2008), *Simulation and the Monte Carlo Method*, 2nd Edition, John Wiley & Sons, New York.

[35] Schönbucher, P.J. (2000), "Factor Models for Portfolio Credit Risk", *Bonn Economic Discussion Papers*, University of Bonn, Bonn.

[36] Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*. Wiley, New York

[37] Sklar, A, (1959), "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris, **8**, 229-231.

[38] Taleb, N. (2007), "Black Swans and the Domains of Statistics", *The American Statistician*, **61**(3), 198-200.

[39] Tsay, R.S. (2010), *Analysis of Financial Time Series*, Third Edition, John Wiley & Sons, New Jersey.

[40] Vasicek, O. (1897), "Probability of Loss on Loan Portfolio", KMV Corporation, San Francisco.