Importance Sampling for Credit Risk Monte Carlo Simulations using the Cross Entropy method

Master Thesis Computer Science

"Prediction is very difficult, especially about the future"
Niels Bohr (1885-1962)

Student Name/Author:	Floris de Vooys
Studentnr:	839055253
Graduation Committee (afstudeercommissie):	dr. Pieter Spronck, prof. dr. Marko van Eekelen, ir. Evert van de Vrie
Date	17 April 2012
Version	6

www.ou.nl

Acknowledgments

I knew from the moment I started the studies that I wanted to graduate in Artificial Intelligence, as I found – and still find – this a fascinating research area. The two AI courses were easily the most fun to do of all courses in the master.

At the Open University, you are encouraged to think early about what you want as a topic to graduate on. It is also encouraged to select a topic that is related to your job, as this helps to set the relevancy and get support in terms of time and resources. However, I work at a bank, generally considered to be among the most boring institutions on the planet, and the mathematics that are involved are mostly statistical, in my limited mind-set among the least interesting of the mathematical disciplines. The work-related subject on credit risk that I found had everything: relevancy, manageable scope, on the edge of the field. The only problem is that it appeared quite dull. But then, as the physicist Richard Feynman wrote: “everything is interesting if you go into it deeply enough”, and this is absolutely true. There is a great satisfaction in getting to really understand a subject, the underlying mechanisms and interrelationships, and then build something that actually works.

Many people helped me to make the right decision and keep on the straight path. My coach at the master programme Harry Passier urged me to investigate the different options early and consider all the advantages of a work-related assignment despite my reluctance. At ING, Menno van Tongeren and Danny Dieleman provided valuable background on the feasibility of the thesis subject and the possible first steps. Credit Risk management, David Day and Joe Katz, gracefully allowed the use of the ING portfolio data in Vortex which is usually guarded so vigilantly.

I would of course like to thank my graduation committee and especially Pieter Spronck for all the detailed readings of my sometimes lengthy work. In addition to the valuable feedback on the content, I really enjoyed working together. I always felt we were on the same page and agreed on where the project should be going.

But most of all I need to thank my wife Sandra, who never failed to support me during five years of study, both morally and in the review of my work, and my children Linde and Olivia whose joy and laughter gave me the energy to go on.

Abstract

One of the most important risks that a bank runs is credit risk: the risk of default of its customers. Credit risk can threaten the existence of the bank if there are concentrations in the bank portfolio and the different loans are closely correlated. An important question for a bank is how much buffer is needed to absorb credit related losses, typically with a very high level of confidence (e.g. 99.95%).

Since the correlations in the portfolio cannot be assessed analytically without making strong assumptions, one often resorts to stochastic simulation or “Monte Carlo”. In this type of simulation, many different possible future economic scenarios are generated and the losses per scenario are determined. The problem with this approach is that it is very slow – a typical run can require hundreds of thousands of scenarios and can take hours or even days.

A potential method to increase the speed of the calculation is to apply Importance Sampling (IS). With IS, the future economic scenarios are not generated randomly, but the “bad” scenarios have a higher chance of being selected than the “good” scenarios and the bias that is thus introduced is corrected later. The underlying rationale is that the bad scenarios that lead to high losses define the area of interest. A method to select the appropriate “bad” scenarios is the Cross Entropy method.

For this thesis, we applied the Cross Entropy method on a credit risk model for the ING wholesale lending portfolio and some synthetically created realistic portfolios. The Cross Entropy method is found to be able to find appropriate Importance Sampling parameters within a relative modest resource budget. With the new parameters, the standard deviation of the estimate that the losses will exceed the available buffer can be decreased with more than 95%. A similar reduction with regular Monte Carlo would require the number of scenarios to increase four hundred times. Alternative methods provide similar reductions, but these use numerical methods that are more complex to implement and require more resources to calculate.

Further tests show that the method is robust to the parameters used in the Cross Entropy method (within reasonable limits), it is not influenced significantly by the constitution of the portfolio and that none of the problems occur that the scientific literature warns about (in particular the “degeneracy of the likelihood ratio”).

Samenvatting

Eén van de belangrijkste risico's die een bank loopt is kredietrisico: het risico dat een klant zijn lening niet terugbetaalt. Kredietrisico kan het voortbestaan van de bank in gevaar brengen als er concentraties zijn in de kredietportfolio en de leningen onderling zijn gecorreleerd. Een belangrijke vraag die een bank zich hoort te stellen is hoeveel buffer nodig is om onverwachte verliezen op te vangen die gerelateerd zijn aan slechte leningen. Deze buffer moet typisch voldoende zijn om in verreweg de meeste gevallen (99,95%) de verliezen te kunnen absorberen.

De correlaties in een portfolio zijn in de regel niet analytisch te bepalen zonder beperkende aannames te maken. Daarom valt men vaak terug op stochastische simulatie, of "Monte Carlo". Bij dit type simulatie worden diverse toekomstige "toestanden van de economie" gegenereerd en bij elke toestand berekend wat de kredietgerelateerde verliezen zijn. Het probleem van deze benadering is dat deze erg traag is – voor een typische run zijn honderdduizenden scenario's nodig en deze kan uren of zelfs dagen duren.

Een mogelijke methode om de snelheid van de berekening te verbeteren is "Importance Sampling" (IS). Bij IS worden de toekomstige toestanden van de economie niet willekeurig gegenereerd, maar "slechte" economische condities zullen vaker voorkomen dan "goede", waarbij de voorselectie die zo wordt geïntroduceerd wordt achteraf gecompenseerd. Men is immers vooral geïnteresseerd in de scenario's die leiden tot hoge verliezen. Een methode om de "slechte" scenario's te selecteren is de Cross Entropy methode.

Voor dit onderzoek hebben we de Cross Entropy methode toegepast op een kredietrisicomodel voor de ING portfolio van zakelijke kredieten. Het bleek dat de Cross Entropy methode in staat was om goede Importance Sampling parameters kon vinden met relatief beperkte middelen. Met deze parameters kon de standaarddeviatie van de schatting van de kans op een verlies dat groter is dan de buffer worden verkleind met ongeveer 95%. Voor een vergelijkbare reductie met reguliere Monte Carlo zou het aantal scenario's moeten worden vervierhonderdvoudigd. Alternatieve methoden bereiken een vergelijkbare verbetering, maar via numerieke methoden die lastiger zijn te implementeren en meer rekentijd kosten om uit te voeren.

Verdere testen tonen aan dat de methode robuust is onder verschillende keuzes voor parameters van de Cross Entropy methode (binnen redelijke grenzen) en dat geen van de problemen optreden waarvoor in de wetenschappelijke literatuur wordt gewaarschuwd, in het bijzonder het probleem met de "degeneratie van de Likelihood Ratio".

Table of Contents

Revision History

Version	Date	Description
1	12 October 2011	First draft
2	30 November 2011	Included comments Pieter Spronck on sections 1-4
3	3 January 2012	Included approach and results sections. Included more information about related research.
4	20 January 2012	included first draft discussion and conclusion
5	7 February 2012	First complete version
6	6 March 2012	Change reference style and lay-out. Added clarification regarding the different factors in the multi-factor model.

1. Introduction

1.1. *Portfolio Credit Risk*

The years 2007 to 2009 saw the worst financial crisis since the 1930s, the apex of which was the collapse of the Lehman's Brothers investment bank. Called among other things the subprime-crisis, liquidity crisis, credit crunch or banking crisis, the integrity of several banks and other financial institutions was severely compromised in this period, and several of these either defaulted or required external/government support. The crisis demonstrates the relative fragility of the worldwide banking system, as well as the importance of proper credit risk management[1].

Institutions that have exposure to credit risk[2] need a comprehensive set of tools to measure and mitigate credit risk. The measurement of credit risk should take place on two distinct levels:

- On an **individual** level, the characteristics of the transaction and the company taking the credit is scrutinised for risks. These risks are for instance related to the profitability of a company, the strength of the competitors in the industry or the competency of the management of the company. On individual transactions, the nature of risk mitigants like guarantees and mortgages also play an important role.

- On a **portfolio** level, the risk analysis focuses on how risks to individual companies are related. For instance, even if all the individual creditors are of good credit quality, many of them could potentially be affected by the same event, like the explosion of a volcano or a decline in American house prices.

This project will focus on the second type of analysis, looking at a complete portfolio of credit exposures and analysing the interactions. More specifically, the project will consider the buffer a financial institution wants/needs to keep in order to absorb unexpected[3] credit losses and prevent insolvency. This measure is usually called *Economic Capital*. A related measure is the *Expected Shortfall*, which measures the losses in excess of the Economic Capital, given that the losses exceed the Economic Capital level. The level of Economic Capital depends on the desired level of confidence that the company will remain solvent, in other words the desired credit rating of the institution. This confidence level is typically in the range of 99.9% to 99.99%, corresponding with a credit rating of BBB to AA on the Standard & Poor scale. Figure 1 illustrates these different concepts. It shows the probability density function (pdf) of the Loss Distribution of a portfolio. The area under the pdf indicates

[1] Credit risk is defined in this context as the risk that a customer or counterparty will default on its financial obligations, loosely speaking that a creditor will not repay its loan. It is a subcategory of default risk, because defaults can also occur because of non-financial obligations like delivery of raw materials or a service. Conversely, bankruptcy is a specific credit event, but others are also possible, like an extension of payment.
[2] These institutions include – but not limited to – banks. For instance, any fund or investor that owns bonds is exposed to credit risk.
[3] Not all credit losses are unexpected. Given a portfolio of sufficient size (like a typical credit card portfolio) there will be a relatively constant level of defaults and corresponding losses. These losses will also occur in "usual" circumstances and are cumulatively called the *Expected Loss*.

the likelihood that losses between two values occur. The total area under the graph is 1 (100%).

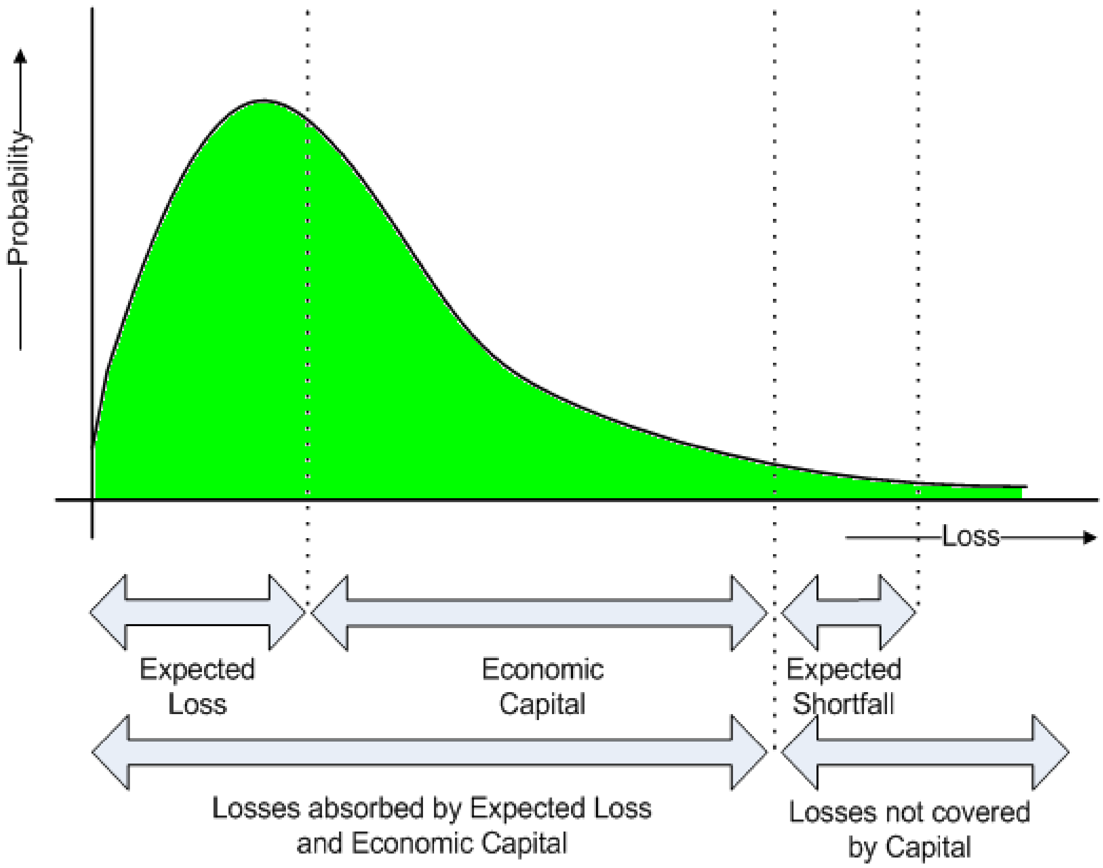

Figure 1: Probability Density Function showing the Expected Loss, Economic Capital and Expected Shortfall.

What is "capital" and what is the difference between the different types?

The term "capital" can mean different things in different contexts. This section is aimed to explain the differences between these different types.

In the context of this project, the term "capital" refers to Economic Capital. This is the buffer that a financial institution *wants* to keep as a reserve against unexpected losses. How high this buffer is depends on how risky and concentrated the loans and investments are, but also on the desired risk rating of the company. A company that wants to be rated AAA by external agencies (corresponding with a default probability of 0.01%) will require a higher capital than a company that desires a rating of A- (with a default probability of 0.1%). Thus, aiming for a lower rating may be a good thing for a bank, because it allows more money to be used for making loans (and thus generating revenue) than when keeping it as a buffer. However, a lower rating also means that you have to pay more on your liabilities (bonds and deposits), as you will be perceived as more risky by investors and peers.

A bank regulator like DNB is responsible for the stability of the financial system, and also wishes to prevent insolvency of banks. It therefore poses a Regulatory Capital requirement to banks. For a bank regulator, it is important to keep a "level playing field" between banks, so it is important that all banks follow the same rules and formulas. In practice this means that the formulas must be relatively simple to implement (because all banks must implement them). To prevent that the lack of sophistication of the simplified models leads to capital requirements that are too low, the regulator usually applies some arbitrary multipliers to the capital number.

The actual available capital of a company is in principle equal to the amount that the shareholders provided to the company, plus any retained earnings and minus any losses. The available capital should normally be greater than both the Economic and Regulatory Capital. If the available capital is insufficient, a bank must reduce the risk of the loans it provides, or issue new shares.

Note that the capital buffers mentioned above relate to the risk of insolvency, i.e. the chance that a bank is not able to pay off its debts. Another major risk for banks is the risk of illiquidity, which means that the bank is not able to meet immediate demands for cash. The typical scenario where a bank becomes illiquid is a "bank run", where for instance depositors start claiming their savings, which the bank cannot pay because the money is invested in long-term loans like mortgages. Managing liquidity is not covered by the models used in this project.

Because there is typically a complex interrelation between the creditworthiness of companies to which loans are extended, Economic Capital and Expected Shortfall are hard to calculate analytically. Therefore traditionally the calculation of Economic Capital and Expected Shortfall is often performed using stochastic simulation, also called Monte Carlo analysis. In this context, Monte Carlo analysis consists of generating scenarios that reflect a "state of the economy" and determining how many defaults[4] occur (and the corresponding losses) given this scenario. The chance that defaults occur that are greater than a certain amount can then be simply determined by determining the fraction of all scenarios where the losses exceed this level. A more detailed explanation is provided in section 2.1. Monte Carlo analysis is a robust and versatile method, but a major disadvantage of it is that it is typically very slow.

The essential ingredient that is necessary to perform the Monte Carlo analysis accurately is the estimate of the correlation of defaults between different parties. These correlations are hard to observe directly, so they are typically calculated indirectly, for instance from equity (option) returns. More details are given in sections 2.3 and 2.4 .

If Monte Carlo analysis is used to find "rare events", a commonly used technique to improve performance is *Importance Sampling*. With this technique, a bias is introduced in the original distribution used to generate the scenarios in the Monte Carlo analysis, so that the rare events that are of interest occur more frequently. This bias is then corrected later in the process.

[4] One could also include rating migrations in this model as an extension

There are several ways to introduce the bias. The traditional one that is used for this particular Monte Carlo analysis is variance minimisation (see also section 3.1). This is a theoretically optimal model. However, the bias that must be introduced can often not be calculated analytically so one must resort to numerical methods, which introduce considerable complexity and loss of performance.

An alternative method to perform Importance Sampling is the Cross Entropy method. While not theoretically optimal, the bias can often be calculated analytically with this method, making the method much faster and more transparent. The Cross-Entropy method is successfully applied in many different contexts, including buffer allocation; scheduling; routing; max-cut and bipartition problems (de Boer, Kroese et al. 2005). The application to the field of finance is so-far relatively limited, a notable exception being the work of Chan and Kroese (2010a). However, the latter is restricted to synthetic and homogeneous credit portfolios, not realistic ones.

A specific point of attention is that the Cross Entropy method can suffer from what is called "degeneracy of the likelihood ratio" when the number of dimensions is high. This degeneracy can lead to a high error in the final estimate. More background on the Cross Entropy method and degeneracy will be given in Section 3.

1.2. *Problem Statement and Research Questions*

The principal problem statement that this assignment intends to address is the following:

> ***PS:*** *To what extent can the Cross-Entropy method be applied to reduce the error-margin in Monte Carlo-based portfolio analysis of Credit Risk?*

This main problem can be split into several research questions, as indicated below:

> ***RQ1:*** *How can the Cross-Entropy method be applied to Monte Carlo analysis of credit portfolios, and ING's credit portfolio in particular?*
>
> ***RQ2:*** *How big is the problem of degeneracy and to what extent is it influenced by the number of levels in the multi-level approach and the number of factors in the multi-factor model?*
>
> ***RQ3:*** *What performance improvements can be achieved as a result of applying the Cross-Entropy method to this Monte Carlo Credit-Risk analysis?*
>
> ***RQ4:*** *How do the results on the Cross-Entropy method compare to other results found in academic publications, including results on other acceleration techniques like exponential twisting, particle filters or conditional Monte Carlo?*

1.3. *Benefits and relevance*

The benefits of increased performance are several:

- It saves computation time, thus saving hardware costs.
- It allows the credit portfolio analysis to be performed on a more frequent basis, and the results to be available sooner. Especially in times of severe market volatility, like the recent banking crisis, this can be important.
- It allows refinements to the model that could otherwise be not feasible because of performance constraints. Examples include
  - calculation of correlated LGD values
  - "what if" scenarios, which can be used to determine the optimal mitigants[5] to reduce portfolio concentrations.

The research topic is one of computational efficiency, and belongs to the discipline of Computational Science. Within the Open University this is an area of the *Mathematics and Artificial Intelligence* domain of the Computer Science faculty. This project will therefore be supported from that domain.

Other researchers on this topic like as for instance done by Chan and Kroese (2010a) and Bassamboo et al. (2005) simulate a credit portfolio by using synthetic portfolios to measure the efficiency of the Importance Sampling algorithm, based on assumptions regarding the distribution of individual exposure sizes and default probabilities. For this research, a reflection of the banking portfolio of ING will be used, making it possible to measure the effect of applying the Cross-Entropy method in a manner that reflects realistic use in the industry instead of based on purely academic research. Realistic synthetic portfolios are also included to verify the support of the method across different portfolios.

1.4. *Document layout*

The rest of this document is laid out as follows:

- Section 2 describes the general background and methods used to estimate the risk in credit risk portfolios.
- Section 3 discusses acceleration techniques to Monte Carlo calculations in general and the Cross Entropy method in particular.
- Section 4 describes the experimental set up used in the thesis.
- Section 5 describes the results that are found. These are compared against findings from other researchers
- Section 6 discusses the consequences and limitations of the results and provides suggestions for further results.
- Section 7 contains the conclusions of this thesis.

[5] A mitigant is a measure that reduces the risk in the portfolio. An example is buying a Credit Default Swap with acts as an insurance policy in case of a default.

Throughout the document there will be grey sections that provide more background information on certain topics. These sections are relatively loosely written and intended to give a better "feel" of the overall subject matter, without attempting a rigorous discussion. They are not part of the main text, and a reader who is already informed on the subject may skip these as desired.

The limitation of models and throwing away the child with the bathwater

The Monte Carlo calculation for portfolio Credit Risk has been under a lot of criticism lately, especially in the light of the banking crisis. After all, using these models banks claim that insolvency could only occur once every, say, two thousand years. How is it then possible that they need to be saved less than 80 years after the previous crisis? Obviously the model is flawed! A particularly strong-worded argument is made by Nassim Nicholas Taleb in his popular book "The Black Swan" (2007), where he argues among other things that the use of normal distributions is inappropriate. This point is elaborated more academically in (Taleb 2010) and (Taleb 2009).

So if this kind of model is so limited, why perform more research on optimising it? The answer in the context of this project is twofold:

First, every model is an approximation of reality to some extent, so any model is "flawed" in this respect. In this case for instance, the calibration of the model is based on historical factors, so any predictive value is based on events that occurred in the past. The proliferation of securitisations[6], the lack of understanding of these products and the market practices that it buoyed were unprecedented. Therefore, these models were not able to predict the associated risks accurately.

This does not mean that the model is useless. The model is very useful to find concentrations and risks in portfolios with products and companies that have long, relatively stable histories. This is true for traditional banking and investment products. But it does mean that one should be very careful when interpreting the results of such a model, and always be on the lookout for new developments and the related risks. A model is no substitute for common sense and keeping an open mind.

Second, the subject of the project is to increase the performance of the model and relates to the computational efficiency. It does not address the accuracy or appropriateness of the model itself. For instance, if the use of the normal distribution is found to be inappropriate by economists, the model should be amended (as for instance proposed by Chan and Kroese (2010a) and Bassamboo et al. (2005)) but the methods to increase computational efficiency can still remain valid.

[6] Securitization is the repackaging of loans or other assets in the form of bonds, so they can be sold to investors.

2. The Credit Risk Monte Carlo framework

2.1. *Credit Portfolios*

Default risk is the risk that a party does not honour its obligations. Such an obligation may be a payment obligation, but it is also a default if a supplier does not deliver the parts he promised to deliver, or if a contractor does not render the services that he promised. In this wide context, default risk is everywhere, and the only transactions that do not involve default risk deliver are the ones that are performed on the spot.

For a financial institution the largest and most important component of default risk refers to payment obligations such as loans, bonds, and payments arising from over-the-counter derivatives transactions. This risk of a payment default, in particular when it refers to loans and bonds, is called credit risk.

Credit risk is typically divided in three separate components:

1) The Probability of Default (PD). This indicates the chance that an individual customer will default in a certain time frame (typically one year).
2) The Exposure at Default (EAD). This indicates the amount that will be outstanding with the customer at the moment that a default occurs. For loans this is equal to the principal amount plus any accrued interest.
3) The Loss Given Default (LGD). This indicates the percentage of the EAD that could not be recovered. The complement of the LGD, the Recovery Rate, indicates the percentage of the EAD that could be recovered, after taking into account recovery costs and the time value of money.

A common approximation is to assume that the three components are independent, and that the last two are fixed (i.e. that there is no uncertainty in the EAD and LGD). This way, the analysis of the risk in the portfolio is limited to defaults occurring simultaneously, not including for instance effects of correlated recoveries[7].

For a bank, individual credit defaults of customers are not unusual events and – although painful and inconvenient – these events are part of the normal course of business. If the exposure is not too large it can be buffered using normal operating cash flows. But when multiple defaults occur simultaneously (or within a short time span) this can threaten the existence of a financial institution. Thus, a major task of the credit risk manager is to measure and control the risk of losses from a whole portfolio of credits. For instance, suppose the credit losses are correlated, so that a default from one obligor makes default of

[7] That is not to say that recoveries are not correlated to each other or to the probability of default. In practice, both probabilities of default are higher and recoveries are lower in times of economic stagnation. However, data on recoveries is notoriously scarce and unreliable, so it is hard to build statistical models for this.

other customers more likely[8], then there is a *concentration risk* in the portfolio that needs to be managed.

In practice, companies do not default independently, but defaults come in “waves”, as illustrated by the following number of defaults per sector in the United States (taken from Schönbucher (2000)):

- Oil industry 22 companies defaulted between 1982 and 1986
- Railroad Conglomerates: One default each year between 1970 and 1977
- Airlines: 3 defaults in 1970–1971, 5 defaults in 1989–1990
- Savings and Loans: 19 defaults in 1989-1990
- Casinos / Hotel Chains: 10 defaults in 1990
- Retailers: more than 20 defaults between 1990 and 1992
- Construction / Real Estate: 4 defaults in 1992

The effect of concentration risk depends highly on the correlation between defaults in the portfolio. Figure 2 shows the probability density functions of the loss for different asset correlations[9] (see also sections 2.3 and 2.4). The pdfs in this diagram are based on an analytical approximation designed by Vasicek (1987) that is also used for regulatory purposes (BIS 2006). As can be seen from the picture, higher correlations lead to a higher chance of exceptionally high or exceptionally low losses, i.e. a “heavier tail” or higher skewness of the loss distribution.

[8] This correlation can be the result of a direct relationship, for instance because the defaulting company is an important customer of the second company, or the result of the fact that both companies are vulnerable to the same factors, for instance if they are in the same country or industry.
[9] Default correlations and asset correlations are different but related concepts. Asset correlations are usually easier to observe and can be used to deduce default correlations. More details are provided in Section 2.3.

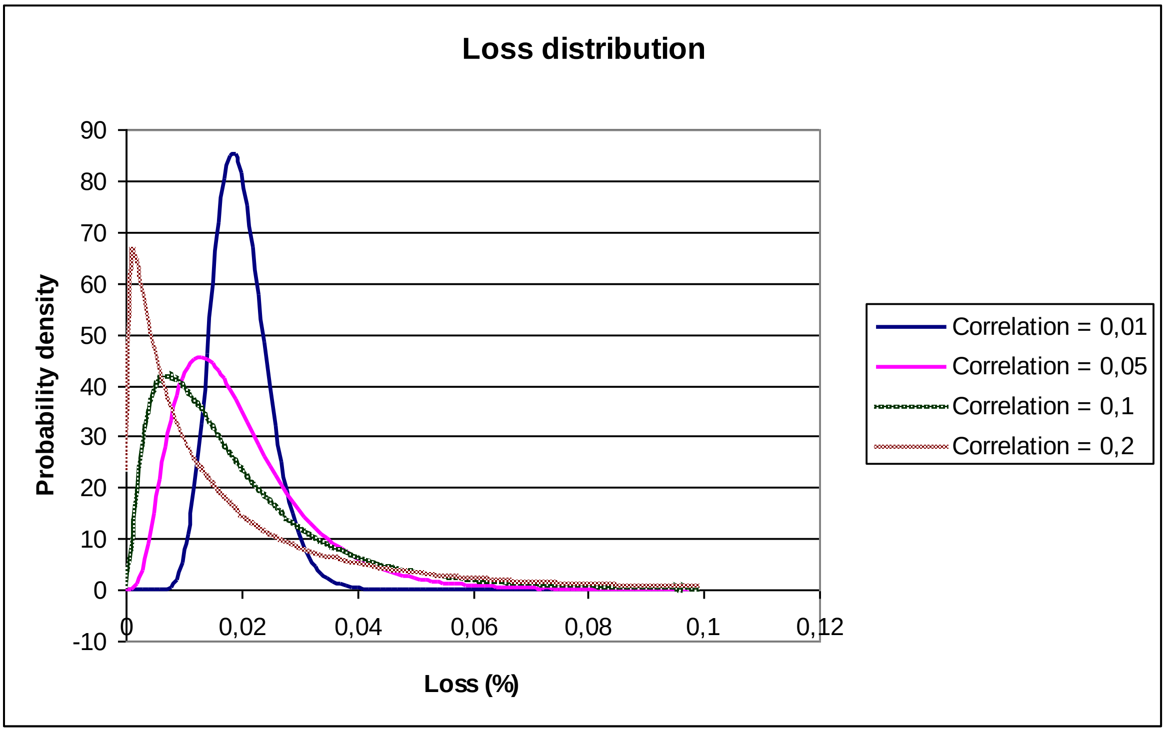

Figure 2: Probability density function of losses, dependent on the asset correlation (rho).

The main interest of managers of credit risky portfolios is the behaviour of the "tail" of this distribution, as these indicate events of a severe nature for the institution.

2.2. *Credit Portfolio Approaches*

Several models exist to determine the credit risk in portfolios. The bullets below provide a high-level description of the components of the most common models. More (technical) details and comparisons between the models can be found in the Professional Risk Managers' handbook (Alexander and Sheedy 2004) and are also described by Gordy (2000) and Crouhy et al. (2000). These models include:

- The Credit Migration Framework, implemented in the Credit Metrics model (Gupton 1997) The model is based on historical risk rating migrations[10] to determine the Probability of Default[11] of a company. It uses concepts from the Merton model (described in more detail in section 2.3) to determine the default correlations, i.e. the chance that two companies default simultaneously. The Credit Metrics technical document describes an effort by JP Morgan in cooperation with other financial institutions to develop an industry standard way to calculate portfolio credit risks.

[10] The rating migration information is typically kept in the form of historical migration matrices. Each cell in such a matrix represents the number of companies that migrated from one rating (e.g. AAA) to another (e.g. AA-) in a specific period. Usually, the numbers on the diagonal are the highest, as ratings tend to be relatively stable.

[11] A default in this context occurs when a customer does not pay its financial obligations, e.g. does not repay its loan or interest. The probability of default reflects the chance that this occurs for a specific company in the next year. A typical probability of default for an average company or private individual is 2-3%.

Although the document is relatively dated (1997), it has become a seminal document in this field on which many other models are based.

- The Conditional Transition Probabilities Approach, implemented in CreditPortfolioView by McKinsey and Company. Like Credit Metrics, this model uses historical rating migrations but it links these migration matrices to macro-economic factors like the unemployment rate, rate of growth in GDP, etc. While this is a potentially useful addition, this model introduces additional complexity which is not required for the research project. Also, the required dependent rating migrations are not available at ING, where the research is performed.

- The Contingent Claims Approach, which is implemented in Portfolio Manager by Moody's KMV. The Contingent Claims Approach is also based on the Merton model, and it bases the Probability of Default on direct analysis of a company's balance sheet. This is contrary to the two previous approaches that are based on historical rating migrations. The Contingent Claims Approach requires an individual analysis of the balance sheets of all involved companies. Such an analysis is performed regularly by Moody's KMV and is kept in a proprietary database. This data is not available for this project and the individual analysis is hard to perform for most companies (it involves the analysis of balance sheets of individual companies) and impossible for others (if no balance statements are available). Therefore this approach will not be taken in this research project.

- The Actuarial Approach, which is implemented in CreditRisk+ by Credit Suisse (Products 1997). The Actuarial approach is inspired by mortality models from the insurance industry. It treats the firm's default as purely exogenous, i.e. it does not try to explain the reason for or the mechanism behind the default, contrary to the structural models discussed before. The timing of the default is assumed to take the creditors "by surprise" and the probability of such a surprise is assumed to be known and to follow a Poisson distribution. The model has an analytical solution and is not based on a Monte Carlo approach. As such, it is not possible to increase performance by applying the Cross Entropy method. Also, the model does not account for default correlations, which is an important limitation. For these reasons, this approach will not be used in this research project.

As already implied in the above summary descriptions, the Rating Migration Approach/Credit Metrics is the favoured approach for this thesis. This model is sophisticated enough to capture the dynamics involved in portfolio concentrations, it is based on a Monte Carlo approach, the calculations are tractable, and it is simple enough to develop within the context of this thesis.

2.3. *Credit Metrics*

One of the classical models used to calculate the risk in credit portfolios is Credit Metrics (Gupton 1997). As already indicated in section 2.1, a key element to predict large losses in a credit portfolio lies in the understanding and prediction of default correlations. There are three methods to model the default correlation:

- The default correlations can be observed directly, as done by Zhou (Lucas 1995). The problem with direct observation of default is that defaults or rating changes are often coarse-grained and relatively rare. As a result, it is hard to find statistical evidence for default correlations.

- The default correlation can be modelled based on the price of Credit Default Swaps (CDSs). Through a CDS, a party can insure himself against a default of a third party. The level of the "insurance premium" that he has to pay indicates the credit-riskiness of the party for whom the insurance is bought. By measuring the covariance of the premium changes of different organisations, the default correlation can be observed indirectly.

- The credit-riskiness can also be modelled through the correlation of the assets of the companies, which can be indirectly observed through the equity price of Corporate Debt (Zhou 2001). This model is a dominant market standard and will be the basis of this project.

The last model is inspired by an article of Robert C. Merton (1974) and is often referred to as the "Merton model". This model is also used by Credit Metrics. In addition, the Credit Metrics model makes the following assumptions:

- The yearly returns (relative changes) on the assets of a company follow a normal distribution;
- The value of the liabilities is constant.

A key element of the Merton model is that a company defaults if the value of the assets is lower than the value of the liabilities. Taking this into account and the above assumptions, it is possible to define a "default threshold" for each company. If the value of the assets drops below this threshold, the company is in default. This principle is illustrated in Figure 3. The default threshold for a company can be determined by simply taking the inverse cumulative normal distribution corresponding with the Probability of Default. Thus, there is a chance equal to the Probability of Default that the value of the assets will be lower than the default threshold, which indicates that the value of the assets is lower than the liabilities, which in turn constitutes a default of the company.

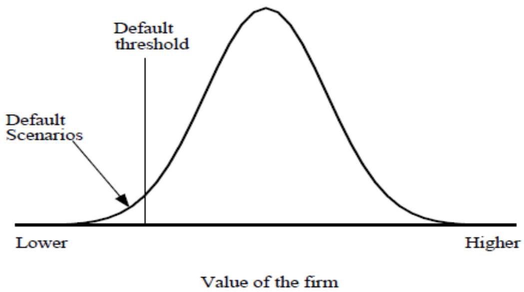

Figure 3: Model of firm value and default threshold (from: Credit Metrics Technical Document)

What is there to the Merton model?

In the Merton model, the economic value of a default is represented as a put option on the firm's assets. This put option can consequently be priced using an option pricing model, which he proposes in the same article.

Put options are financial instruments that give the buyer the right to sell an asset at a pre-agreed price to the seller of the option. Conversely, the seller has the obligation to buy for the agreed price. In order to compensate the seller, he usually receives a premium. The transaction can be seen as insurance, where the buyer acquires an insurance policy against low prices. For instance, a farmer that grows oranges could buy a put option on the price of oranges to make sure that a price drop would not cause large revenue losses when he has to sell his crop. The buyer of the option retains the upside potential: there is no obligation to sell if prices rise.

How does this relate to companies and defaults? A company can be financed by equity or by debt (loans and bonds). The debt holders receive interest on their investment, and have first claim over the assets in case of a company default. The equity holders benefit from the profits of the company directly, either through dividends or share price increase. In case of a default they are subordinated to the debt holders (they leave with nothing).

The equity holders own the company, but in case of a default they can walk away and the debt holders will be left with any additional losses caused by a shortfall of the value of the assets over the value of the liabilities. In other words, the debt holders lose the value of their investment and acquire the assets, which have now a lower value than their investment. This reflects exactly the situation of a put option, where the bond/loan holders have sold a put option to the owners of the company: the equity holders. The premium of this option is reflected in the fact that corporate bonds pay a higher interest rate than "risk free" bonds like certain government treasury paper.

The benefit of this reasoning is that it allows relationships to be made between for instance a company's creditworthiness, the "leverage" (ratio between debt and equity), the interest rate, the asset volatility, the remaining term of the debt and the "price" of the credit risk. In other words, it allows something that is relatively poorly understood econometrically (corporate defaults) to be explained in terms that are well understood (market factors and balance sheet). The details are not required for this project, but are provided in the background documentation (Alexander and Sheedy 2004) for the interested reader.

In its full form, the Merton Model is quite complex to understand, but in the context of the Credit Metrics model only the relationship between asset value and default needs to be understood.

The Credit Metrics approach shifts the problem of determining correlated defaults to correlated asset values. The Credit Metrics model proposes to use equity returns for this purpose. In the present research, an internal model is available to estimate asset correlations which I will use without further direct verification. The structure of this model is explained in section 2.4.

Other key inputs in the model are the Probability of Default, Exposure at Default and Loss Given Default. These values are all determined based on statistical or expert models of the credit institution (in this case ING) that provides the portfolio data.

The portfolio model that is applied in this project consists of two states: default and non-default. This means that the model does not take the losses in value that occur because of rating degradation into account. The reasons for this choice are:

- Banks do not take the loss in value on loans because of rating migrations into account in their Profit and Loss statement[12]. The new IFRS accounting standards do not allow banks to account for losses until they are actually occurred[13].
- No properly calibrated migration data are available.
- The losses incurred because of default are much greater than the losses caused by rating migration.
- The model is easier to calibrate and run with two states than with for instance twenty.

While Credit Metrics proposes to use a stochastic Loss Given Default (LGD) based on a beta-distribution, it is a common approximation to use a fixed LGD instead. In order to determine the effect of this additional uncertainty in the model, gauging the effect of this choice on the performance of the Cross Entropy method will be part of this thesis.

2.4. *Internal Model for Asset Correlations*

ING has an internal model to model Asset Correlations. The following section will describe the underlying principles behind this model. The model will be considered a “given”: it will not be separately validated or compared to other models.

What is an asset correlation, and how would I model it?

As explained in Section 2.3, the Merton model is based on estimating the (change in) value of the assets of a company. The assets of a company are everything the company uses to generate revenue - like machines, inventory and real estate – and financial assets – like

[12] It does have an effect on the P&L through specific “Incurred But Not Recognized” (IBNR) loan loss provisions, but this reflects a different concept than the actual losses in value. The background behind these provisions and simulating the level of these is beyond the scope of this project.

[13] This is true for the assets that the bank intends to hold until maturity, like most loans. It is not true for assets that the bank holds for trading purposes. However, the latter are usually treated in a market risk framework, not a credit risk model. The assets held for trading purposes are thus also excluded from the portfolio used in this project.

receivables, invested securities and positive cash balances on current accounts. For instance, for an airline company, the value of an airplane for the company is related to the number of passengers it can transport, for which price and at which cost.

If there is stress in the market environment, this typically impacts the value – or revenue-generating capacity – of the asset. For instance, a period of economic downturn can lead to a lower demand of long-distance flights and a lower value of airplanes. For an inventory of luxury goods this relationship is even more direct, as the prices of the product being sold can be under pressure. Whether the company is able to withstand the deteriorating conditions depends on the available buffer, which is the reflected by the equity of the company.

From the above it is also intuitive that assets from different companies do not move independently. If the revenue-generating capacity of airplanes decreases because of a decreased demand, this will impact all airline companies. For financial assets like receivables, the correlation is less obvious as the relevant underlying parties will be different. However, in stressed economic conditions the number of defaults on the receivables will be higher in both cases, so the changes in the value of the assets will be correlated: they are more likely to move in the same direction than in opposite directions.

In practice, the correlation between assets is not modelled directly from the type of asset. Instead, the value of the equity of the company is observed in the market and the total change in the asset value is derived from that. Based on this, there are methods to extract “explaining variables” from the information, for instance using Principal Component Analysis. The explaining variables can be for instance the state of the global economy, the state of the economy of a specific country or region, or the state of a specific sector.

The internal model used within ING to model asset correlations is based on three levels of explaining factors:

1) A part of the asset return is explained by global factors. These can usually be related to specific large regions (e.g. Western Europe) or industry sectors (e.g. high-tech).
2) A remaining part of the return is explained by the specific country or industry.
3) The remainder of the asset return is considered idiosyncratic, i.e. specific for that company.

Each customer has a different sensitivity to each of these factors. This depends to a large extent on the country, the industry and size (i.e. turnover or asset size). For instance, industries like the financial industry can be negatively affected by a decline in the state of the global economy, while the food packaging industry would barely be affected, and the tobacco industry could actually be positively affected. The individual sensitivity to each factor is driven by:

- The country/region of the customer;
- The (main) industry/sector of the customer;
- The dependency of the individual company on the world economy, also called the R-squared. The R-squared is different for each company. A major driver for the R-squared is the size of the company: larger companies tend to be more affected by the general state of the economy, while the smaller companies are more dependent on individual circumstances.

The model is calibrated in such a manner that all factors are independent. Because of this there is no need to generate correlated random samples through e.g. Cholesky decomposition. The actual method of calibration and the underlying data used to perform the calibration are not further investigated as part of this master thesis.

Given the factors and the individual sensitivities of the customer above, the total Asset Return of a customer can be defined as follows:

$$AssetReturn = \alpha_1 GFR_1 + \alpha_2 GFR_2 + ... + \alpha_{Country} FR_{Country} + \alpha_{Industry} FR_{Industry} + \alpha_{Idiosyncratic} FR_{Idiosyncratic}$$

Where:

Asset Return	= the relative return on the assets of the customer
α_x	= the sensitivity of the customer on global factor x.
$\alpha_{Country}$	= the sensitivity of the customer on the Country Factor.
$\alpha_{Industry}$	= the sensitivity of the customer on the Industry Factor.
$\alpha_{Idiosyncratic}$	= the sensitivity of the customer on the Idiosyncratic Factor.
GFR_x	= the Global Factor Returns, or the relative change in the global factor, for the global factor x.
$FR_{Country}$	= the Country Factor Return for the country of the customer.
$FR_{Industry}$	= the Industry Factor Return for the country of the customer.
$FR_{Idiosyncratic}$	= the Idiosyncratic Factor Return for the customer.

And subject to the restrictions that:

1) $\sum \alpha = 1$
2) $\alpha_{Idiosyncratic} = 1 - (R\text{-}squared)$

The first restriction ensures that the total asset volatility is a correctly weighted average of the volatility of the factors. The second restriction ensures that the asset returns are appropriately correlated with “common” (global, country and industry) factors and the company-specific “idiosyncratic” factor. The common factors are scaled linearly to comply with both restrictions.

Example

The multi-factor model provides the following factors for a company:

$\alpha_{1,\text{ Original}} = 0.25$
$\alpha_{2,\text{ Original}} = -0.10$
$\alpha_{\text{Country, Original}} = 0.85$
$\alpha_{\text{Industry, Original}} = 0.05$
R-squared = 0.45

$\alpha_{\text{Idiosyncratic}}$ can be directly derived from the R-squared and is equal to 0.55. The sum of the other factors must be equal to 1 – 0.55 = 0.45. The original sum of the factors is equal to 0.25 – 0.10 + 0.85 + 0.05 = 1.05, so all factors must be scaled down and multiplied with 0.45/1.05. The resulting factors for this company are:

$\alpha_{1,} = 0.11$
$\alpha_{2} = -0.04$
$\alpha_{\text{Country}} = 0.36$
$\alpha_{\text{Industry}} = 0.02$
$\alpha_{\text{Idiosyncratic}} = 0.55$

Using this model, the correlation between different customers can be explained by the factors that they have in common. Typically, the more the customers are alike in terms of country and industry, the more they will be sensitive to the same factors. This is illustrated in Figure 4.

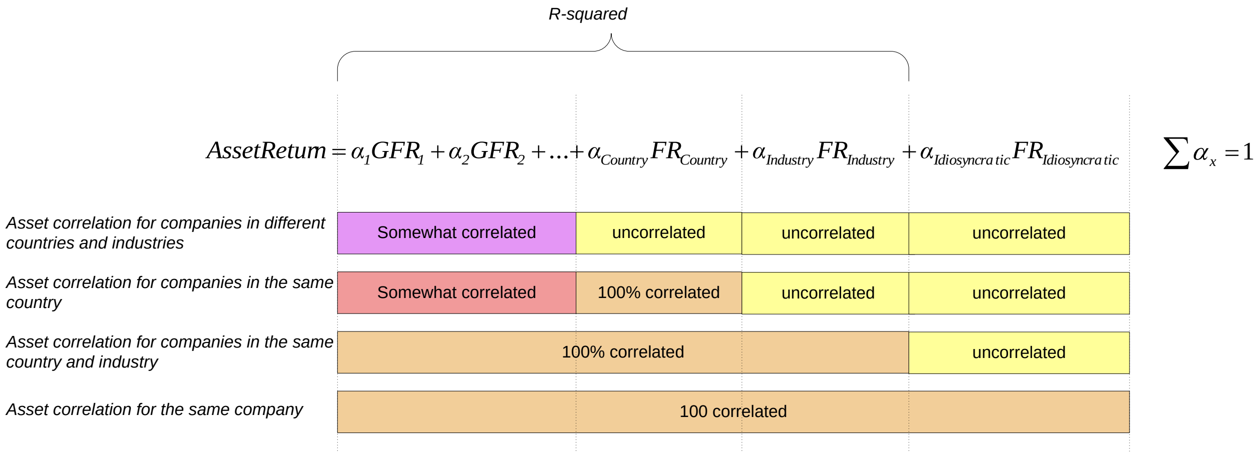

Figure 4: asset correlation as a result of dependency on common factors

Since a linear combination of normally distributed variables is itself normally distributed, it is possible to generate appropriately correlated asset returns with a normal distribution. This can be done by generating the common factors using a normal distribution and applying the appropriate sensitivities of the asset value of the company to each factor.

2.5. *The Monte Carlo method*

The relationship between the factors in the asset correlation model and the loss in the portfolio is such that it does not allow an analytical solution. The typical resolution is to resort to stochastic simulation, or Monte Carlo analysis. The Monte Carlo analysis for credit portfolios consists of three steps:

1) scenario generation
2) valuation of the portfolio/determination of the losses
3) aggregation

In the first step, different "states of the economy" are generated. In our case, this is equal to a value of each of the factors in the asset correlation model, which we assumed are normally distributed. These factors affect the value of the assets of each company in the portfolio and determine whether a company defaults. The losses are then determined based on the Exposure at Default and Loss Given Default. In the aggregation step the losses are summed.

Economic Capital can now be simply determined by sorting the scenarios in increasing amount of total loss and taking the desired percentile. Similarly, the Expected Shortfall can be determined by selecting the average loss in excess of the Economic Capital for those scenarios where the loss exceeds the Economic Capital.

3. Acceleration techniques, Importance Sampling and the Cross Entropy method

3.1. *Acceleration Techniques*

A “naïve” or “crude” Monte Carlo analysis that is used to estimate a rare event typically takes very long to run, because a sufficient number of scenarios need to be generated that lead to the rare event to make a statistically relevant estimate. For instance, if the chance needs to be estimated of an event that occurs once every 10,000 tries, you would need to generate on average 100,000 scenarios to get only 10 “interesting” results. Fortunately, different methods exist to improve the performance of naïve Monte Carlo calculations[14] (Atzberger ; Rubinstein and Kroese. 2007):

- In the *antithetic variates* method, for every scenario that is generated, the antithetic scenario is also selected. In the context of the problem under observation, this can be thought of that for every improvement in the economy an opposite deterioration with equal likelihood is considered and vice versa. This method is easy to implement and has few material drawbacks.

- Another method is the use of *control variates*. The method uses the errors in the estimates of known quantities to reduce the errors in the estimate of a similar unknown quantity. An application in the context of the research problem would be to make a simplified model of the credit portfolio under consideration for which the probability of insolvency can be analytically determined (in the context of this subject for instance CreditRisk+ (Products 1997)) and use this distribution as a control variate. The underlying idea is that the variance in the estimate that is caused by behaviour that is analytically understood can be “backed out” of the estimator, leaving a more accurate result. This approach is taken by for instance Tchistiakov et al. (2004).

- A commonly used method to improve performance of Monte Carlo calculations when trying to determine the chance of rare events is *Importance Sampling*. In general, Importance Sampling consists of the following steps:

    1. First, the rare event is made less rare, so it will occur more frequently. This step is optional, but is common if the event is sufficiently rare (e.g. less than one in 100,000 chance of occurring).
    2. Several scenarios are generated.
    3. The original distribution is shifted so that the “rare” events will occur more frequently.
    4. If the event was made less rare in step 1, the event is made rarer.
    5. The process is repeated with the shifted distribution until the event has reached the desired “rarity”.

[14] Aside from “technical” solutions like parallel processing.

There are different ways to shift the original distribution in step 4, as will be outlined later in this section.

- *Conditional Monte Carlo* (Chan and Kroese 2010c) uses estimators where the distribution is conditional on a certain situation. As with importance sampling, the chance of this conditional situation occurring must be corrected in the estimator. Conditional Monte Carlo is for instance useful in the following circumstances:

  - If the expected value and variance of the conditional distribution can be calculated analytically, simulation is only required to determine the conditioning factors (Rubinstein and Kroese. 2007). This can lead to significant performance improvements.
  - If one or some of the input factors have a considerably larger impact on the result than the others, this factor can be "stressed" and the simulation can take place on the remaining factors. The estimator has to correct for the probability of the stressed condition occurring, but this can often be achieved analytically (Asmussen 2004).

  In the context of credit portfolios, the second approach is for instance used by Chan and Kroese (2010a). The "stressed" factor then represents the general state of the economy.

- *Stratified Sampling* (described for instance by Glasserman et al. (2000)) is similar to Conditional Monte Carlo simulation. One of the input factors is divided in sub regions (*strata*, single *stratum*) and for each stratum a conditional Monte Carlo is performed. The number of samples in each stratum is not constant, but depends on the variance in each stratum. Using a dynamic sampling algorithm, the samples can thus be guided to the most "interesting" stratum. The complexity in this method lies with the correct choice of strata.

- *Particle Filters* or *Sequential Monte Carlo Methods*[15] (Moral and Garnier 2005; Carmona, Fouque et al. 2009; Creal 2009) simulate the (time-dependent) path to the horizon at which the event occurs. It starts with a number of possible paths ("particles") and these particles can be periodically *re-sampled* based on the distance to the rare event. Particles that are closer to the rare event (in the context of credit risk: that lead to a higher loss), are more likely to be re-sampled than particles that are farther away. In this sense the method is similar to a genetic algorithm. Particle filters are successfully applied to the problem of Credit Portfolio Monte Carlo (Deng, Giesecke et al. 2011). A drawback of the method is that the loss of the portfolio must be calculated much more frequently than in the traditional Monte Carlo approach. Also, the method suffers from the same degeneracy problem in high-dimensional cases as the multi-level Importance Sampling approach described by Rubinstein and Glynn (2009) (see also section 3.5: "Degeneracy").

[15] Sequential Monte Carlo methods are not strictly an acceleration technique, but a different simulation approach., but since t allows for different acceleration techniques it is included in this section.

This project will focus on the application of Importance Sampling, for the following reasons:

- The use of antithetic variables or stratified sampling is useful and simple to implement, but the effect of the method is usually more limited than importance sampling.
- In order to apply control variates, a second simulation model needs to be developed with an analytical solution to the problem. This makes it more an econometrical solution than a computational one.
- The use of particle filters requires different background knowledge than is currently available at the institution where the project is performed.
- Conditional Monte Carlo is of similar interest and complexity as Importance Sampling. Both methods can actually be combined, as for instance done by Chan and Kroese (2010a). The choice between both methods is arguably arbitrary.

As mentioned, the shifting of the sampling distribution in the Importance Sampling method can be done in various ways. The "traditional" way is to try to find the shifted distribution that will lead to the lowest variance in the estimator. This is referred to as variance minimisation or exponential twisting (Glasserman and Li 2005; Glasserman, Kang et al. 2008). While theoretically optimal, the distribution with minimal variance can often not be determined analytically and requires numerical methods (Egloff, Leippold et al. 2005). An alternative approach is the Cross Entropy approach, which often results in similar improvements (Asmussen, Kroese et al. 2005), but in many situations has an analytical solution. The Cross Entropy approach is explained in more detail in the next section.

3.2. *The Cross-Entropy method for Rare Event estimation*

A formal explanation of the Cross-Entropy method is provided by Rubinstein and Kroese (2007) and the Cross Entropy Tutorial (de Boer, Kroese et al. 2005). In order to provide some insight in the practical mechanics of the method, the principles of Importance Sampling and the Cross-Entropy method will be illustrated by a simple example in the following sections.

<u>Example 1: sum of standard normal variables</u>

Consider a simple situation, where we would like to estimate the chance that the sum of five factors that follow a standard normal distribution exceeds a certain threshold, say 10. This example is illustrated in Figure 5. The figure shows the probability density functions of five standard normal distributions and five selected values. The values selected are shown on the x-axis. The y-axis shows the relative likelihood of that sample value occurring. In this case, the threshold will most likely be exceeded if values are sampled on the right-hand side of the distribution.

The solution can easily be found analytically, because the sum of normally distributed factors is itself normally distributed with a mean equal to the sum of the individual means and a variance equal to the sum of the individual variances. Since standard normal distributions have a variance of one, the sum of five standard normal distributions has a mean of zero and a variance of 5. Selecting the value of the normal cumulative distribution with variance five at the threshold value of 10[16] provides the answer of $3.872*10^{-6}$. However, for the sake of this illustration we want to estimate this through Monte Carlo estimation. Since this chance is relatively low, we will need to generate millions of scenarios to find sufficient samples where the threshold is exceeded when using naïve Monte Carlo.

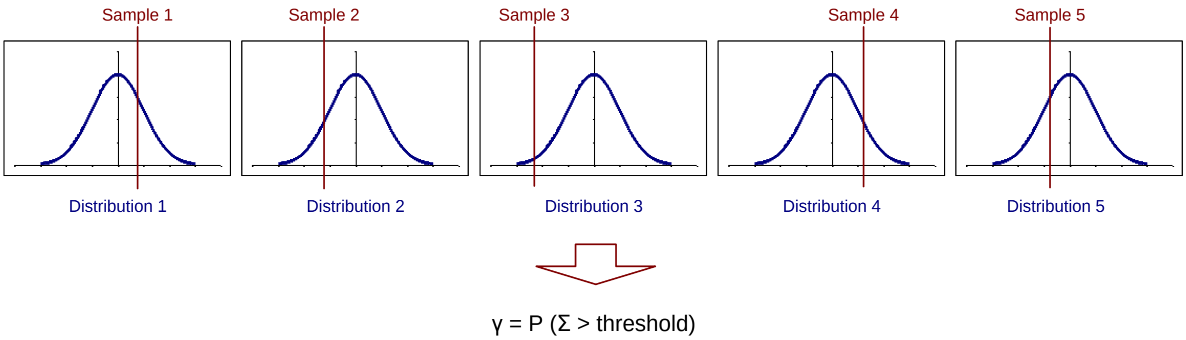

Figure 5: illustrative example – estimating the chance that a sum of normally distributed values exceeds a threshold

3.3. *Importance Sampling and the Likelihood Ratio*

The purpose of Importance Sampling is to draw the samples from a new distribution that will make the rare event much more common, and then correct for the introduced bias afterwards. It is often convenient to use the same type of distribution for the shifted sample as the original distribution. For our example, we will use a normal distribution with a shifted mean and a fixed standard deviation of 1.

In the example of the sum of the normally distributed variables, an intuitively appealing new mean to use for the different factors would be 2. With this value, the new mean would be 10, so the samples would be distributed around the sought threshold value. The probability density functions (pdf) of the original and shifted distributions are shown in Figure 6.

[16] For instance, in MS Excel this can be achieved by the formula =1-NORMDIST(10;0;SQRT(5);1)

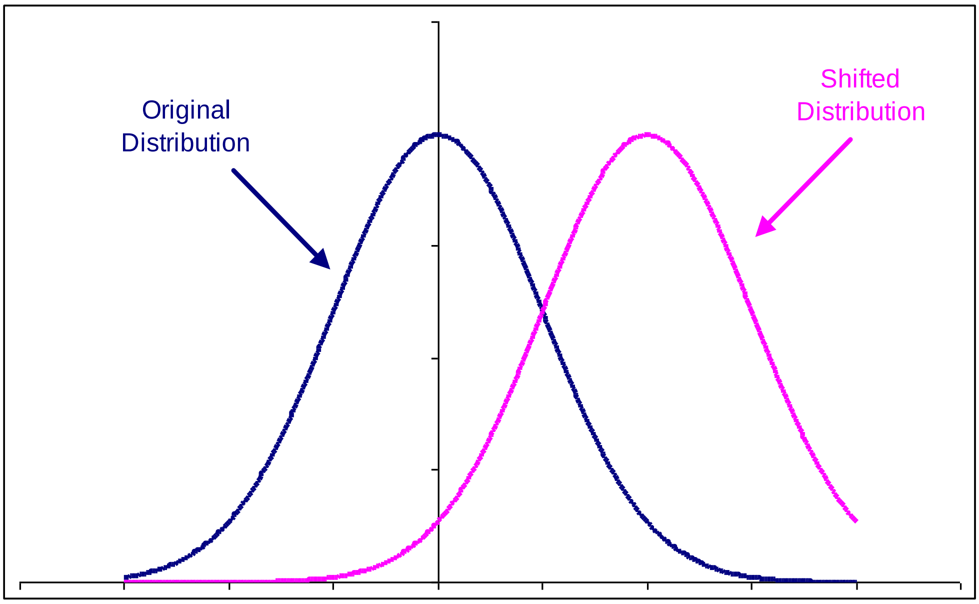

Figure 6: shifted standard normal distribution with mean = 2.

The found samples must be corrected for the introduced bias, because otherwise the chance that the rare event occurs would be overstated. This correction is equal to the Likelihood Ratio, which is the ratio between the value of the original pdf at the chosen sample and the value of the shifted pdf. This is illustrated in Figure 7 for a sample value of 2.1, which is relatively likely value for a normal distribution with mean 2, but relatively rare for a distribution with mean 0.

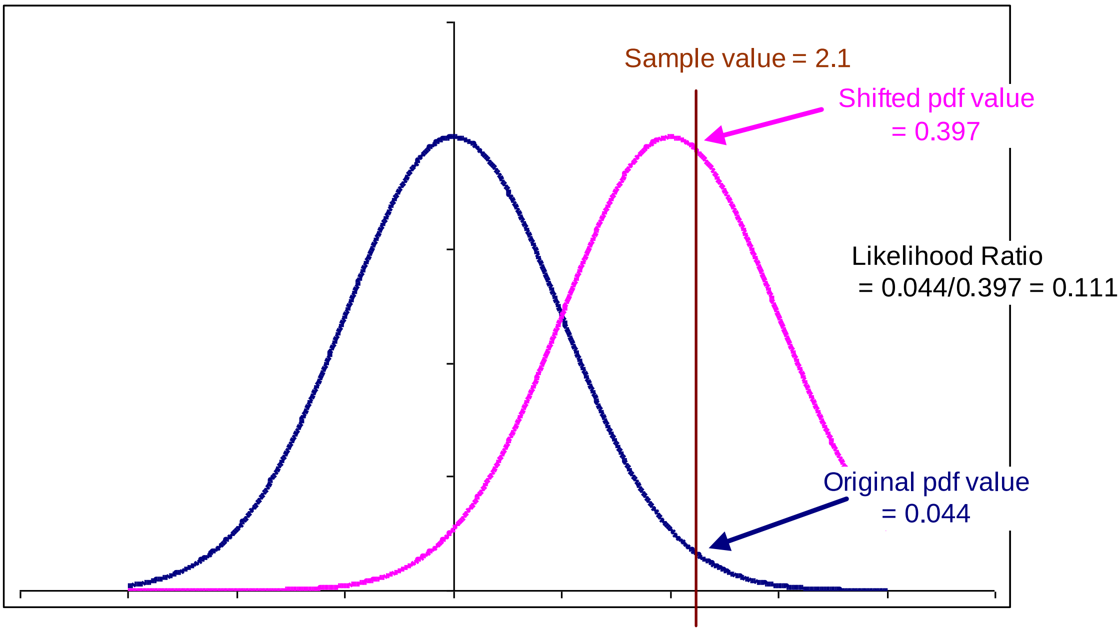

Figure 7: determining the likelihood ratio

Some things can be observed about the likelihood ratio:

- By definition, the expected value of the likelihood ratio is 1.
- While the likelihood ratio will typically have a value lower than 1, it can potentially have very large values as well. This divergence of very high and very low values is higher when the shift of the means increases. This is illustrated in Figure 8 to Figure 10, which shows the log (base 10) of the likelihood ratio for different sample values and different shifts. For instance, a sample value of -2.5 with a mean shift of +2 would lead to a likelihood ratio of 1097. Since the likelihood ratio acts as a “weight” for the found samples that exceed the threshold, samples with high likelihood ratios can potentially affect the ultimate estimator significantly. Because of their high impact, these samples could lead the estimator away from the optimal solution.

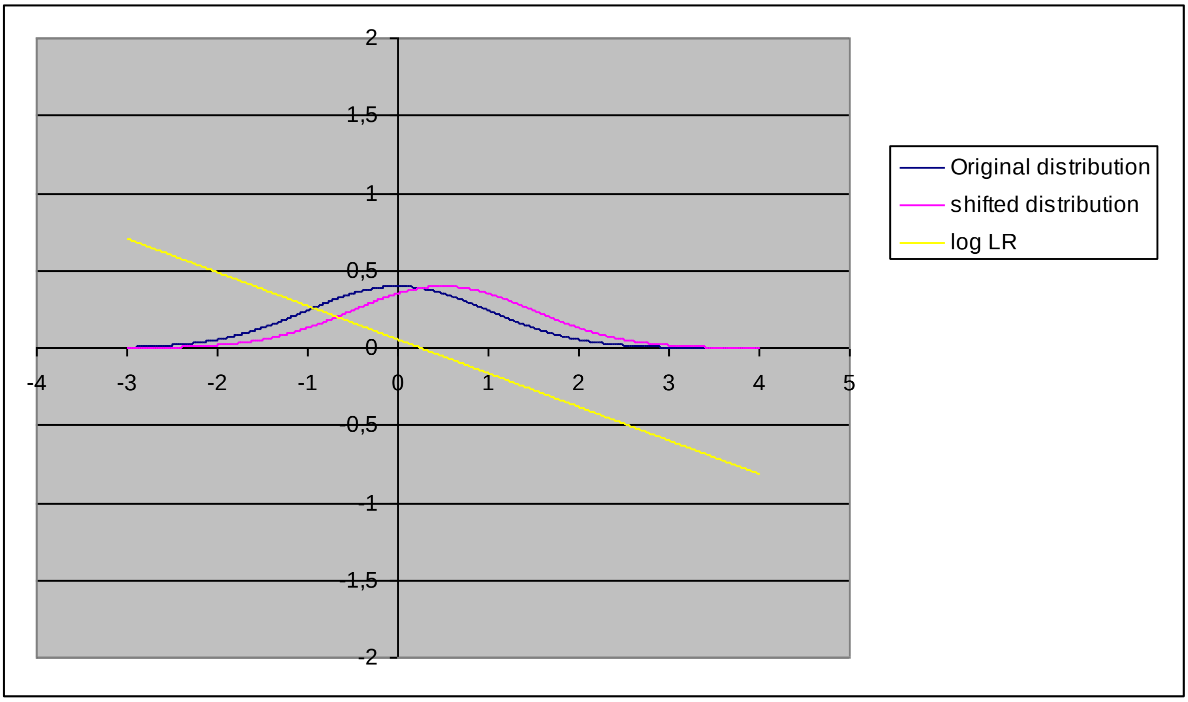

Figure 8: log of the Likelihood Ratio with mean shift 0.5

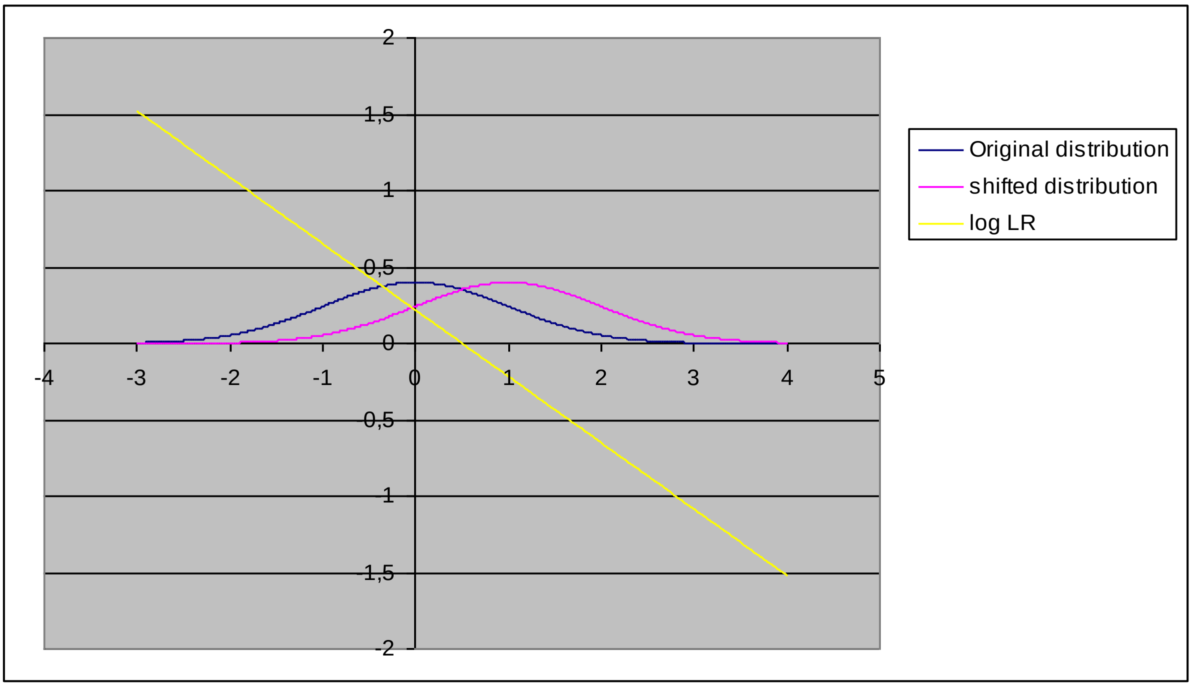

Figure 9: log of the Likelihood Ratio with mean shift 1

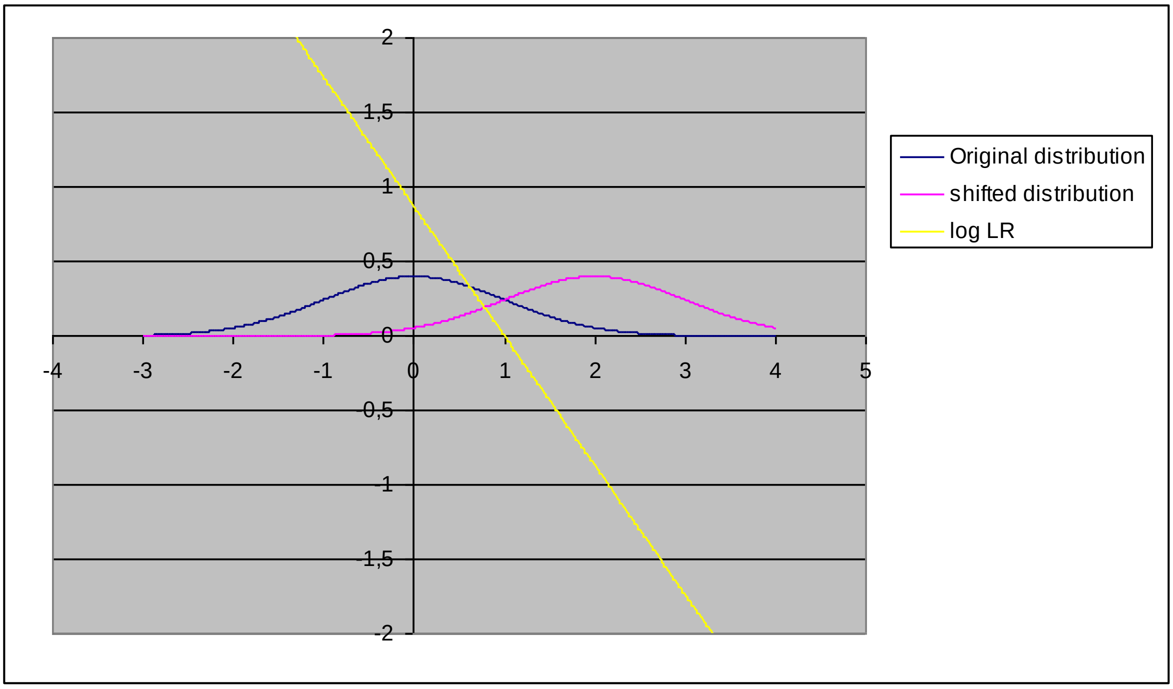

Figure 10: log of the Likelihood Ratio with mean shift 2

3.4. *The Cross-Entropy method*

For the case of example 1 we can make an educated guess regarding a decent shifted distribution, i.e. a new mean of 2. However, in general such a convenient measure will not be available, and an optimal distribution must be determined by other means. One method to determine such an optimal distribution is the Cross-Entropy method. The steps that are taken in the general Cross-Entropy method are the following:

1) Define a set of starting parameters for the distribution.

2) Generate N1 samples from the probability density and determine the top ρ-percentile of the found values, the “elites”. ρ is typically a value between 0.1 and 0.01.

3) Use the same sample found in step 2 to solve the basic Cross-Entropy program as described in the tutorial (de Boer, Kroese et al. 2005). Informally, this means finding the shifted distribution that most closely matches the distribution of the samples that led to the elites, where “most closely” is defined by the Kullback-Leibler distance:

$$D(g,h) = \int g(x) \ln g(x) dx - \int g(x) \ln h(x) dx$$

   For the normal distribution parameterised by the mean, the solution to the Cross-Entropy program is to take the mean of the samples that led to the elites.

4) Repeat steps 2 and 3 with the shifted distribution, until the top ρ-percentile exceeds the desired threshold.

5) Generate N2 samples with the shifted distribution to perform a final run.

As can be seen from the description above, the important parameters of the Cross-Entropy method are:

- N1: The number of scenarios used to estimate the optimal shifted distribution.
- ρ: The percentage of “elites” used to estimate the optimal shifted distribution.
- N2: The number of scenarios used to estimate the value.

The process above is illustrated for a single factor of Example 1 below:

Step 1: define a set of starting parameters.

In the example, all distributions are standard normal distributions, so the starting mean is zero.

Step 2: Generate samples and select the top ρ-percentile.

Assume that we generate 6,000 samples and take a ρ of 0.1, so 600 elites. Figure 11 shows the found distributions of the samples and the distribution of the factors that led to elite values.

Step 3: solve the Cross-Entropy program

For the normal distribution parameterised by the mean, solving the Cross-Entropy program is equivalent to taking the average of the samples that led to the elites. This value is 0.82 in this example, so this will be taken as the mean of the shifted distribution. The fitted distribution is shown in Figure 12.

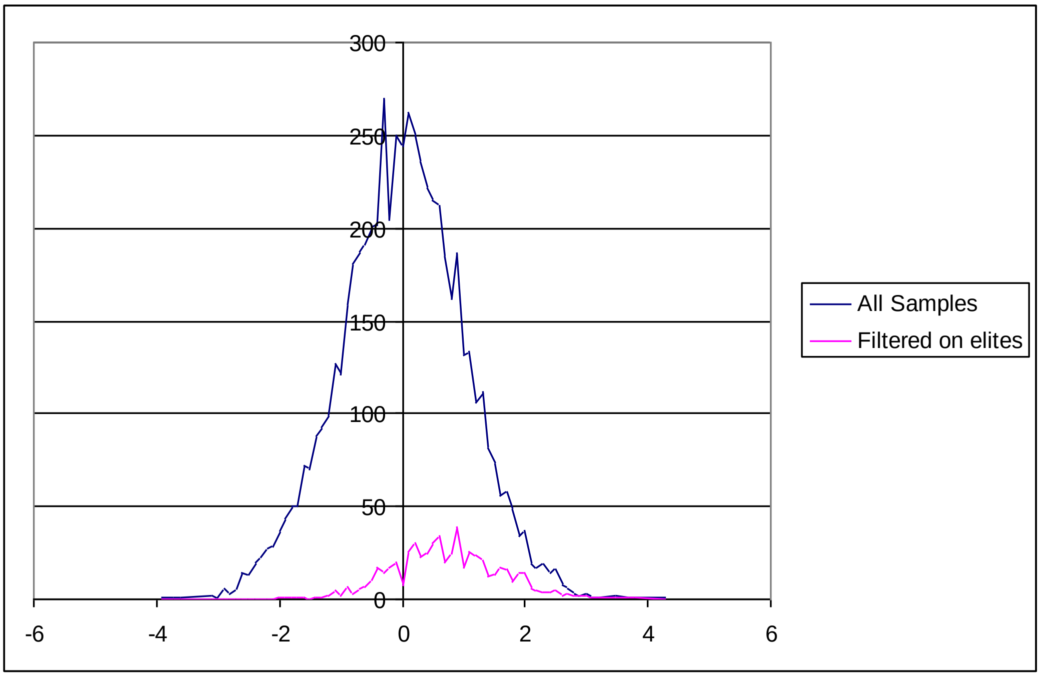

Figure 11: histogram (breadth 0.1) of the sampled factors showing all sampled factors and the samples leading to elites

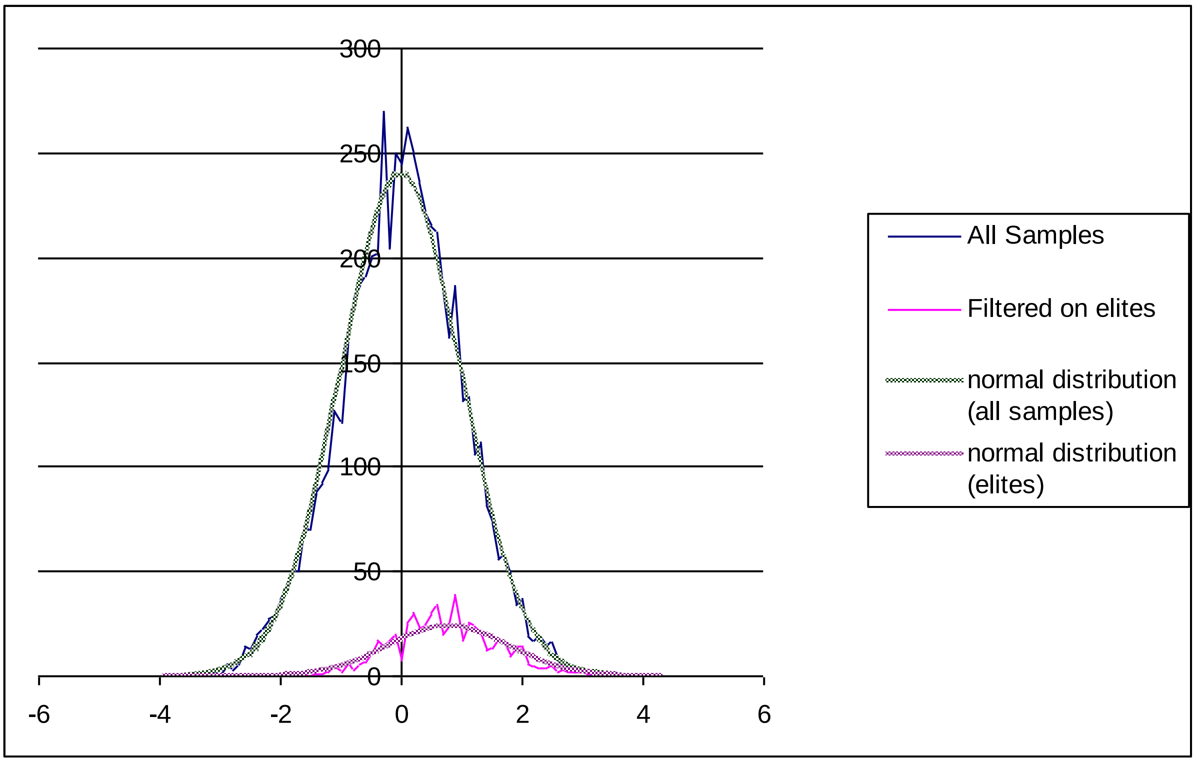

Figure 12: histogram from Figure 11 with fitted normal distributions

<u>Step 4: repeat steps 2-3 until the desired threshold is reached</u>

Each iteration will increase the mean with a certain amount. Reaching the desired threshold of 10 took 5 iterations in this example. The evolution of the mean is shown in Figure 13.

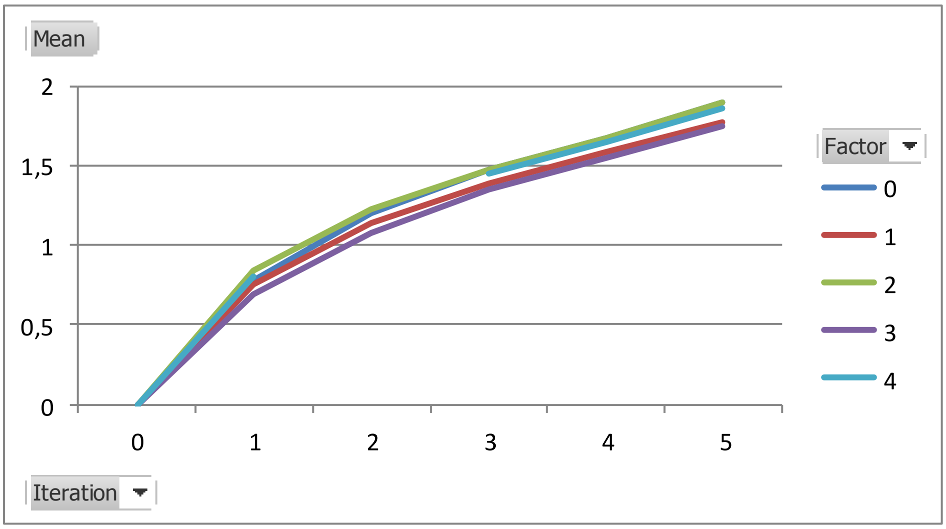

Figure 13: evolution of the mean of the five different factors

<u>Step 5: perform the final run with the shifted distributions</u>

A run of 60,000 samples gave a final estimate of $3.893*10^{-6}$, with a relatively error of 1%. This is approximately 0.5% higher than the actual value of $3.872*10^{-6}$. Note that the estimate took in total 90.000 samples, which for naïve Monte Carlo would normally not even be enough to generate a single rare event.

3.5. *Degeneracy*

The problem of degeneracy of the likelihood ratio is explained as follows by Rubinstein and Kroese (Rubinstein and Kroese. 2007), p. 133:

> The likelihood ratio estimator suffers from a form of degeneracy in the sense that the distribution of the Likelihood Ratio *W*(X) under the importance sampling density *g* may become increasingly skewed as the dimensionality *n* of X increases. That is, *W*(X) may take values close to 0 with high probability, but may also take very large values with a small but significant probability. As a consequence, the variance of *W*(X) under *g* may become very large for large *n*. As an example of this degeneracy, assume for simplicity that the components in X are identically and independently distributed (iid) under both the original distribution *f* and *g*. Hence, both *f*(x) and *g*(x) are the products of their marginal pdfs. Suppose the marginal pdfs of each component X_i are f_1 and g_1 respectively. We can then write *W*(X) as
>
> $$W(X) = \exp \sum_{i=1}^{n} \ln \frac{f_1(X_i)}{g_1(X_i)}$$
>
> Using the law of large numbers, the random variable $\sum_{i=1}^{n} \ln \frac{f_1(X_i)}{g_1(X_i)}$ is approximately equal to $nE_{g_1}\left[\ln \frac{f_1(X_i)}{g_1(X_i)}\right]$ for large n. Hence,
>
> $$W(X) \approx \exp\left\{-nE_{g_1}\left[\ln \frac{f_1(X_i)}{g_1(X_i)}\right]\right\}$$
>
> Since $E_{g_1}\left[\ln \frac{f_1(X_i)}{g_1(X_i)}\right]$ is nonnegative, the likelihood ratio *W*(X) tends to 0 as n tends to infinity. However, by definition the expectation of *W*(X) under *g* is always 1. This indicates that the distribution of *W*(X) becomes increasingly skewed when n gets large.

The implied assumption is that a higher skewness also leads to a higher variance, and thus a large relative error in the estimator.

The screening method (Rubinstein and Glynn 2009) is a relatively simple solution that can be used to prevent degeneracy. The principle behind the screening method is that only certain "bottleneck elements" are shifted. For instance, to estimate the reliability of a network you would stress the most vulnerable network elements. The method is described in more detail in Appendix A.

3.6. *Other improvements to the Cross-Entropy method*

Other published improvements to the Cross-Entropy method include the following:

- Combine the Cross-Entropy method with Conditional Monte Carlo (Chan and Kroese 2010a).
- Introduce concepts from the field of Markov Chain Monte Carlo to prevent the necessity to apply likelihood ratios, and thus prevents degeneracy of the likelihood ratio. This approach is taken by Chan and Kroese (2010b).

Both Conditional Monte Carlo and Markov Chain Monte Carlo are large research areas in their own respects. Since each of these improvements extends the research beyond the scope of the master thesis project, these options will not be pursued further as part of this project.

3.7. *Summary*

Figure 14 shows an overview of the relative contexts of the method used in this project.

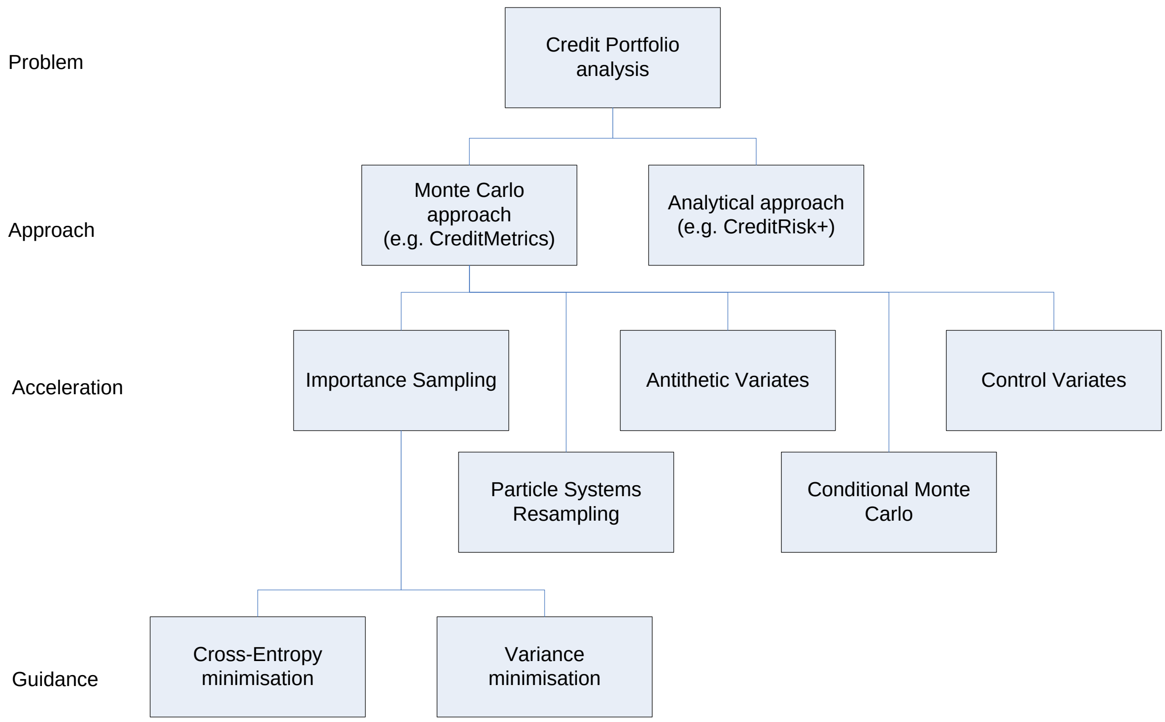

Figure 14: overview of the problem domain

4. Experimental setup

4.1. *Software framework*

The software that is developed as part of this project consists of the following main classes:

- An abstract Cross Entropy class, which performs the generic steps of the Cross Entropy defined in section 3.4.

- An abstract "factor distribution" class, which creates drawings from a distribution and determines the likelihood ratio for each (shifted) factor.

- Two concrete implementations of the factor distribution that are developed in this project for the normal distribution and the exponential distribution.

- A Credit Portfolio class, which calculates the losses in a given portfolio. It will perform the import of the portfolio file and load the different factors. It supports either stochastic LGD or fixed LGD calculations.

- A concrete Cross Entropy Credit Portfolio class, which implements the generic Cross Entropy procedure and uses the Credit Portfolio class when required. It creates and initialises the global, country and industry factors.

- A user interface to start a Cross Entropy Credit Portfolio run and display the results.

The dependencies between the different classes and their main functions are provided in Figure 15.

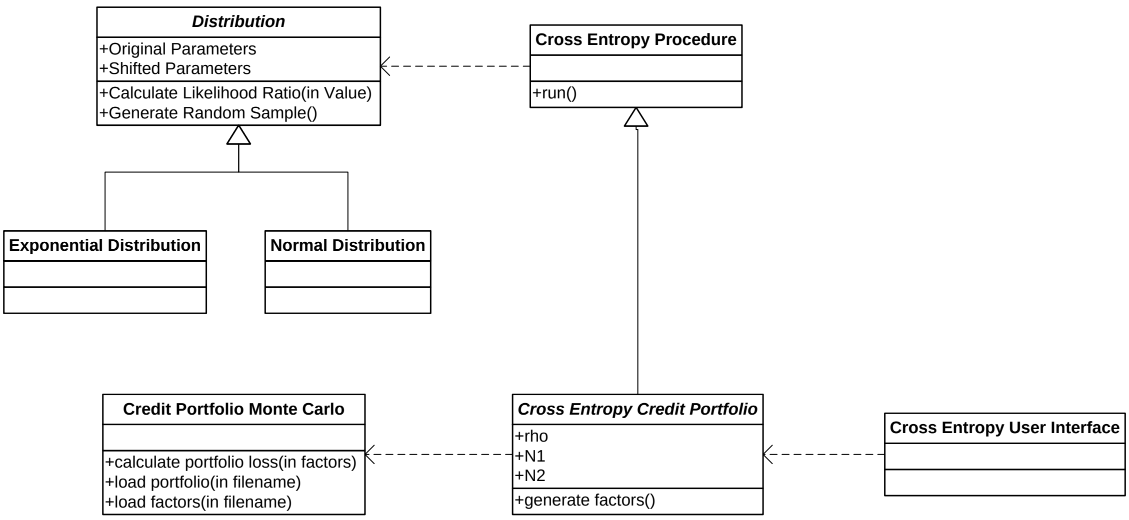

Figure 15: Class diagram of the software framework

The generic Cross Entropy component not only supports the calculation of the chance that the losses exceed a given capital threshold, but can also calculate the required capital threshold given a desired chance. The latter is a more typical use of the Credit Portfolio model, although most research on the Cross Entropy method is based on the former estimate.

In order to determine the threshold, it is required to select the appropriate percentile from the loss distribution. A regular percentile is selected by sorting the values from high to low, and selecting the n^{th} value that corresponds with the desired percentage (n = number of samples * [1- percentile]). This is illustrated in Figure 16[17]. When Importance Sampling is applied, the desired threshold is the value where the cumulative Likelihood Ratio corresponds with the desired percentile. This is illustrated in Figure 17.

Loss amount	sequence	
158	1	
128	2	
117	3	
108	4	
98	5	← 95% percentile
89	6	
87	7	
…		
5	100	

Figure 16: regular percentile

[17] The examples shown ignore the problem of required interpolation to determine the percentile for the sake of simplicity. Instead the value closest to the desired percentile is selected.

Loss amount	Likelihood Ratio	Cumulative Likelihood Ratio	sequence
158	0.1	0.1	1
128	0.7	0.8	2
127	0.5	1.3	3
121	1.2	2.5	4
117	0.1	2.6	5
115	0.4	3.0	6
111	1.2	4.2	7
110	0.2	4.4	8
108	0.4	4.8	9 ← 95% percentile
98	0.8	5.6	10
…			
5			100

Figure 17: percentile under Importance Sampling

The software and the underlying source code are available for download at http://www.de-vooys.nl/Monte%20Carlo.zip .

4.2. *Credit portfolios*

The main credit portfolio that will be used as part of this project is the ING portfolio as of the end of 2008. Some descriptive measures of this portfolio are provided in Appendix C.

In order to test the robustness of the measure relative to the underlying portfolio, three synthetic portfolios will also be considered. These portfolios are based on the Markit iTraxx Europe Series 16 index (Markit 2011). These indices are used for financial instruments that allow investors to transfer the credit risk of a synthetic credit portfolio defined by 125 corporations. This index is used in this project because the 125 companies are major institutions with publicly available rating information, and because the purpose of the index (mitigating risk or investing in credit portfolios) matches the context of this research.

Different indices exist for different regions. The one selected is specific for Western Europe. This choice is made because the European index provides a good diversification across industries and countries, while maintaining concentrations that are characteristic of many real credit portfolios.

The Probability of Default is implied from the external rating from the three main rating agencies - Standard&Poor's, Fitch and Moody's – linked to the published default data by Standard&Poor's (Rafat Khan 2011). No recovery data was available, so this is assumed to be

40% across all transactions. For the R-squared, the ING estimates are used. These parameters are provided in full per customer in Appendix D.

The iTraxx index has an equal contribution of the 125 corporations in the index. In practice, the contribution of each corporation to the total portfolio will not be equal. Therefore, additional synthetic portfolios are generated with different contributions of each corporation, based on a power-law distribution with exponential cut-off. The distribution has been calibrated manually to fit the distribution found in the ING portfolio. The cumulative density function of the distribution thus found and used is:

$$c(x) = 1 - \frac{e^{-\lambda x} * x^{\alpha}}{\beta}$$

Where:

$\lambda = 2*10^{-8}$
$\alpha = 2.27$
$\beta = 8*10^{-4}$

Actual credit portfolios consisting of only 125 corporations are rare. Therefore additional synthetic portfolios are created where each corporation is cloned 20 times to create a portfolio of realistic size. The full portfolios are included in the software download referred in Section 4.2.

4.3. *Performance measure*

The number of scenarios generated relative to the relative error is chosen as a performance measure rather than timing or hardware measure for the following reasons:

- The number of scenarios and relative error are not sensitive to the environmental factors like the programming language or underlying hardware.

- Other researchers (Heidelberger 1995; Asmussen 2004; Morokoff 2004; Asmussen, Kroese et al. 2005; Juneja 2007; Rubinstein and Glynn 2009; Chan and Kroese 2010a; Chan and Kroese 2010b; Chan and Kroese 2010c) in the field also use these measures, so this allows a comparison between the findings of this project and those in other articles.

The error in the estimate is formally based on the difference between the values found in the simulation and the actual value. However, since the real Economic Capital is not known (this is the value that is to be estimated), the sample variance is used instead. The advantage of using the relative error instead of a performance measure like variance is that it does not depend on total volume of the portfolio and thus allows better comparison between results by different researchers. The relative error is defined as:

$$RE = \frac{S}{l\sqrt{N}}$$

Where S is the sample standard deviation (S^2 is the sample variance), l is the current estimate and N is the number of scenarios.

5. Results

5.1. *Comparison of naïve Monte Carlo and Cross Entropy Monte Carlo*

Figure 18 shows the evolution of the estimate and relative error as a function of the number of scenarios for a naïve Monte Carlo run and a Cross Entropy Monte Carlo run. The run parameters for the Cross Entropy run are ρ=0.1, N1=10,000 and N2=30,000. Note that the Cross-Entropy method only provides results from the 20,000th scenario onwards, because the first 20,000 runs are required to determine the shifted probability density functions, i.e. for the "training"-part of the Cross Entropy method.

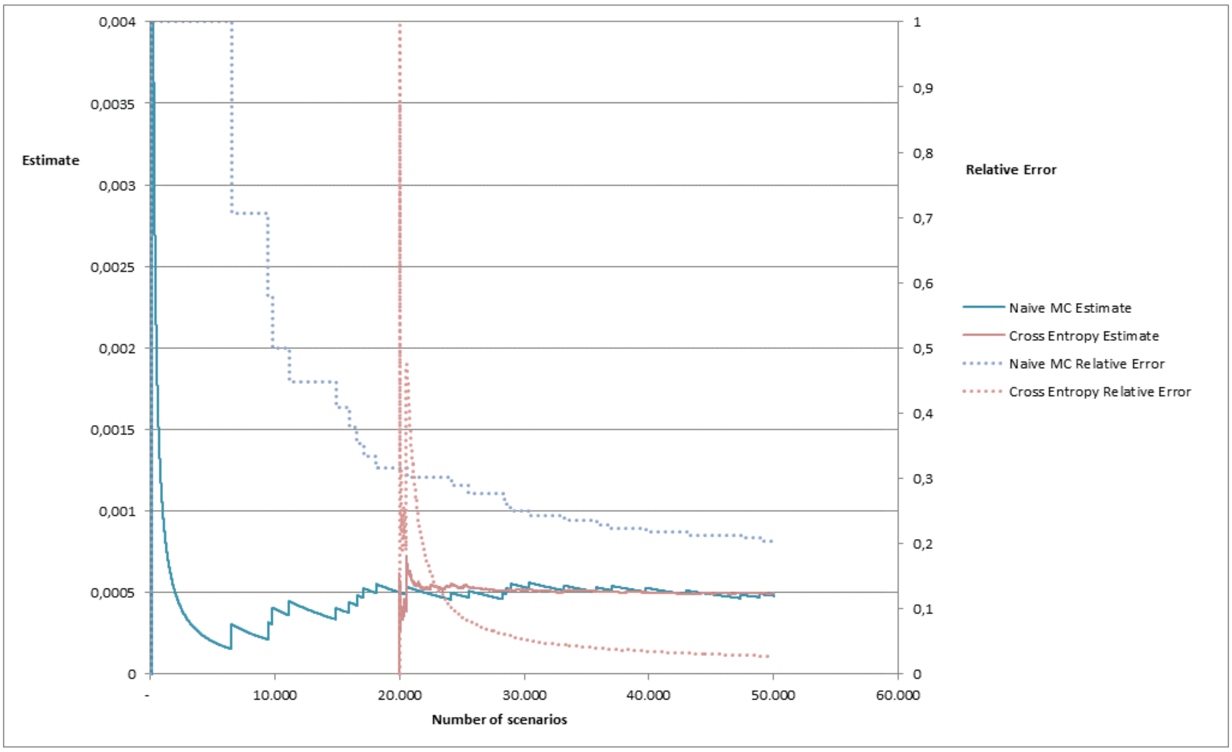

Figure 18: Estimate and Relative Error for naïve Monte Carlo and the Cross Entropy method

As can be seen from Figure 18, the relative error after 50,000 scenarios is reduced by approximately 90%, with a break-even after 21,000 scenarios. To reach the same level of accuracy with a naïve Monte Carlo approach would require more than 5 million scenarios to be evaluated.

The effect of the Cross-Entropy based importance sampling is also clear from the loss distributions. Figure 19 and Figure 20 show the parts of the loss distribution for low and high losses respectively. As can be seen from these pictures, the naïve Monte Carlo distribution is more stable (fewer spikes) for the low losses, but the Cross-Entropy distribution is smoother for the high losses, which is the area of interest for the business problem underlying this research.

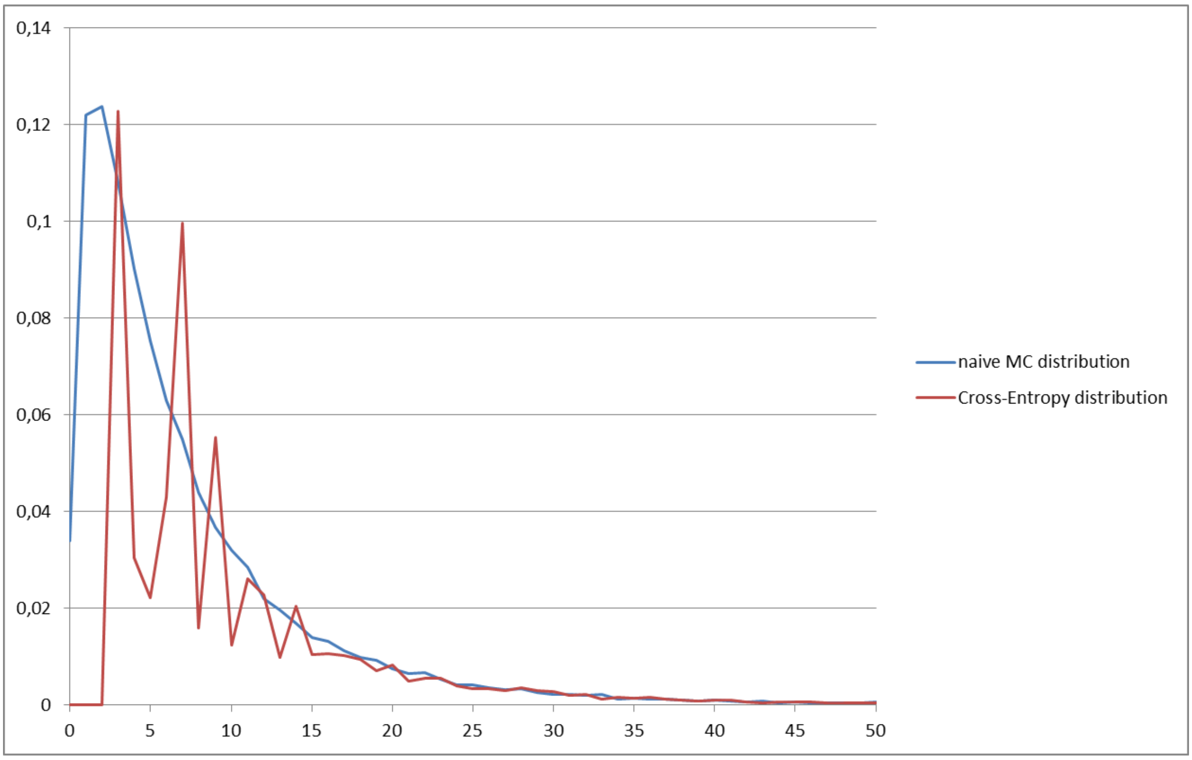

Figure 19: loss distribution for naïve MC and Cross Entropy for low losses

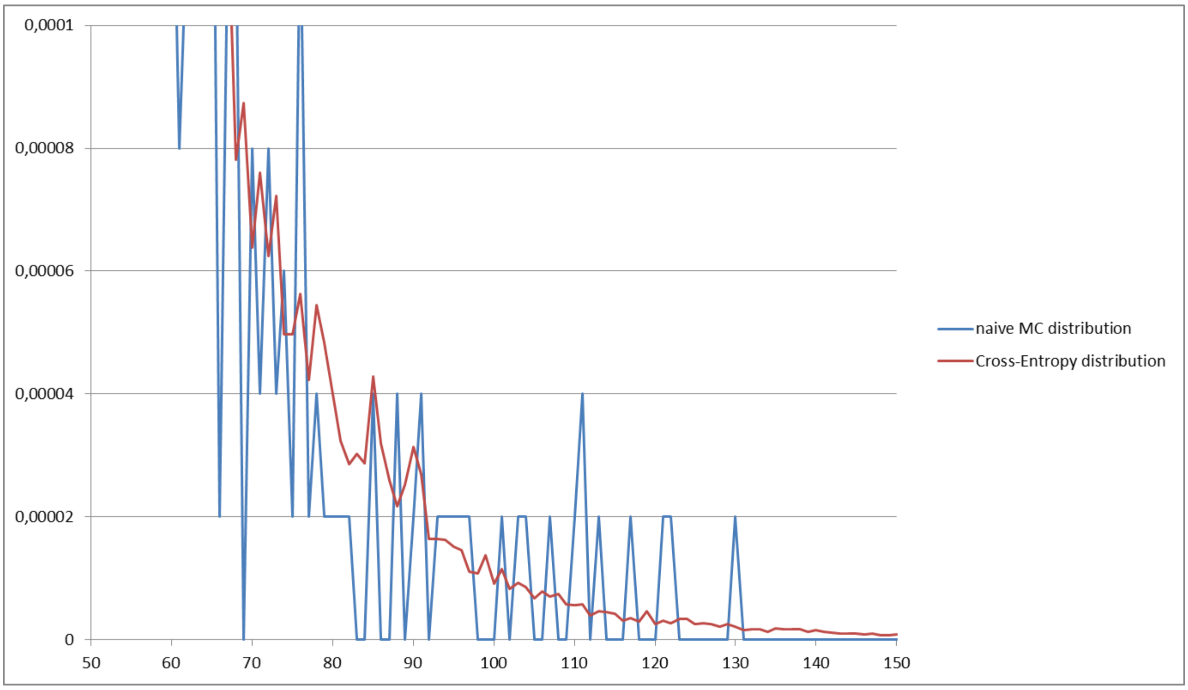

Figure 20: loss distribution for naïve MC and Cross Entropy for high losses

5.2. *Sensitivity of the Cross-Entropy calculation for choices of ρ, N1 and N2*

The Cross Entropy method has three input parameters: ρ, N1 and N2. It is relevant to understand the sensitivity of the method to the choice of these parameters for following reasons:

- It provides a starting point for future developments and implementations in this area as to what reasonable parameters settings are;
- It determines how much effort should be spent in finding the optimal set;
- It determines the “downside potential” of the method if the method performs badly with certain settings.

Of the three parameters, N2 has a predictable effect on the relative error, as the relative error will decrease with the square root of N2. Hence, doubling N2 will decrease the relative error by a factor of 1.4 (square root of 2). The parameters rho and N1 will impact the standard deviation in a much less predictable way. Figure 21 shows the relative reduction of the standard deviation of the estimate based on the Cross Entropy method relative to the naïve Monte Carlo, based on a desired confidence level of 0.05%. The reduction is an average of five runs. As the diagram shows, the reduction is more than 90% as long as the total number of “elites” per step is greater than approximately 300. Increasing N1 above this has an increasingly smaller effect. If the number of “elites” per step drops below 10, the effect is detrimental and the error in the estimate will even increase.

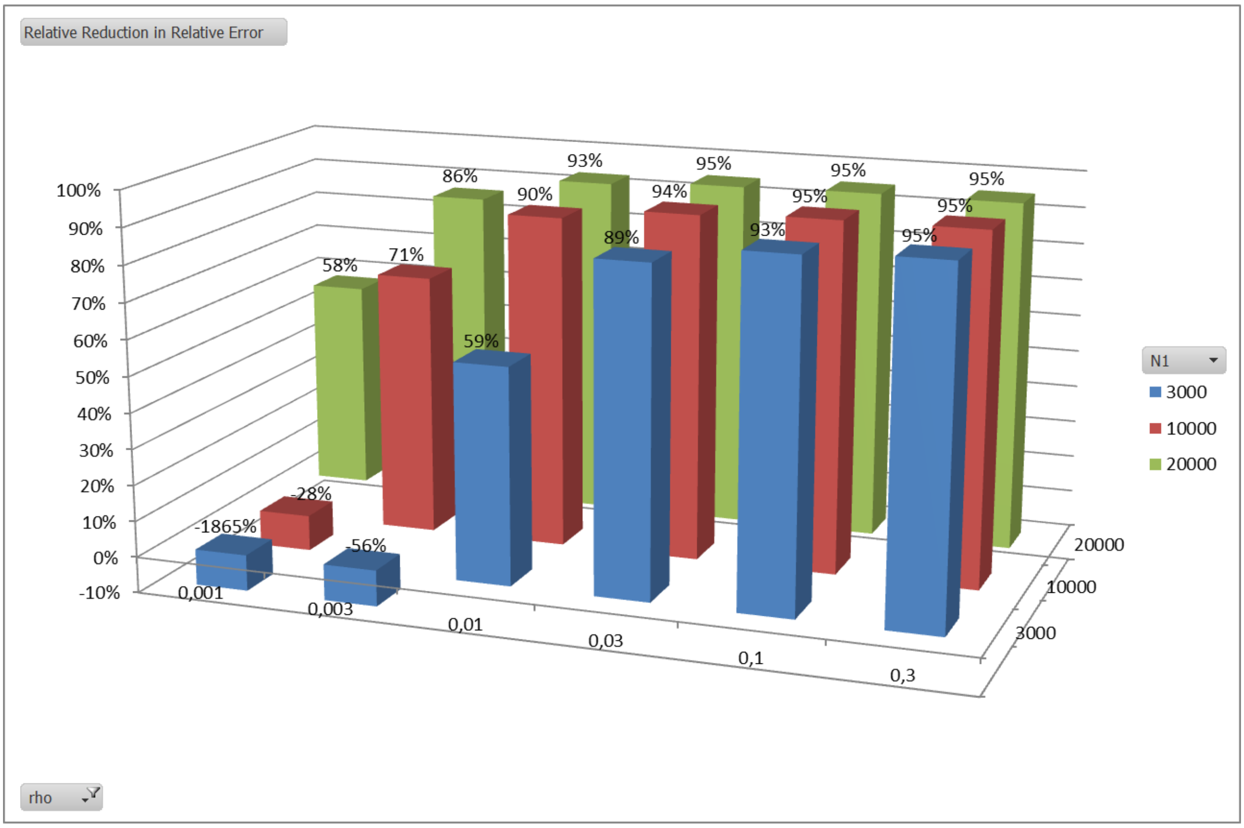

Figure 21: effect of rho and N1 on the standard deviation of the estimate

All else being the same, a higher N1 or a lower ρ will lead to a higher number of scenarios. These additional calculation costs should also be considered in the comparison. Figure 22 shows the required number of scenarios required for training used for the estimates in Figure 21.

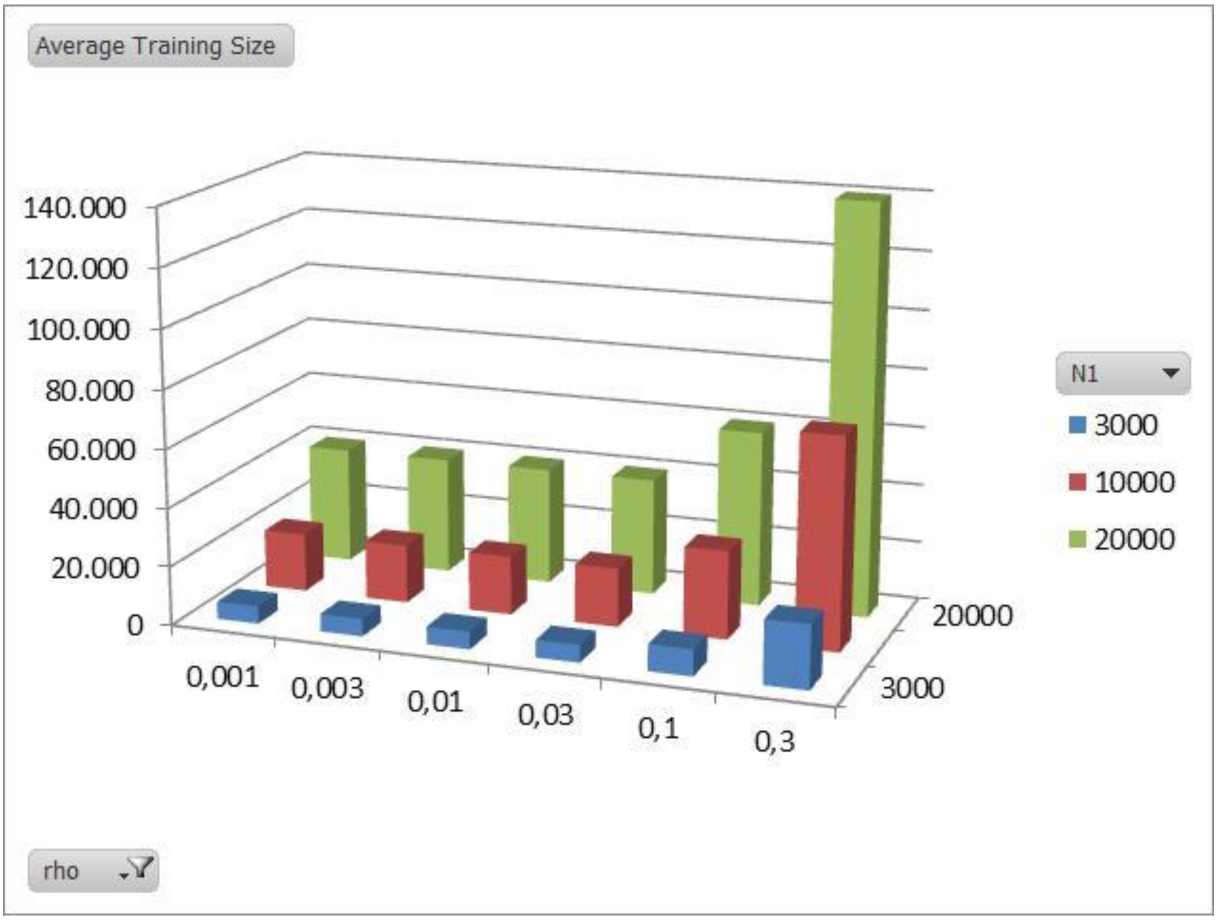

Figure 22: required number of scenarios for training for different rho and N1

Whether the increased cost in terms of training outbalances the reduced standard deviation is a consideration that must be made on a case by case basis, as this depends on the desired accuracy and available calculation budget. In general, however, Figure 21 and Figure 22 suggest choosing a relatively high ρ and low N1, i.e. taking small steps. Note that these findings are consistent with the recommendations in the literature, which recommend a ρ between 0.1 and 0.01.

5.3. *Comparison to results found by other researchers*

5.3.1. The Cross-Entropy method in other contexts

Table 1 shows a comparison of the results from Figure 21 with results from other researchers using the Cross-Entropy method. From this table it can be seen that the improvement in the relative error that we found is similar or better than improvements found in other contexts. This is despite the fact that our model contains 120 dimensions, so would a priori be expected to suffer from degeneracy of the likelihood ratio, as models with more than 50 dimensions are expected to (see Section 3.5). In this case, however, there appears to be no such problem.

Table 1: comparison between naïve Monte Carlo and Cross Entropy found in the academic literature

Case	Value sought /Confidence level	Reduction in Relative Error	Reference
Bridge network	$7.0787*10^{-5}$	87%	(Hui, Bean et al. 2005)
3x3 unreliable grid network	$4.0120*10^{-6}$	70%	(Hui, Bean et al. 2005)
6x6 unreliable grid network	$4.008*10^{-6}$	-20% (deterioration)	(Hui, Bean et al. 2005)
2x2 bridge model (Bernoulli)	$3.2 *10^{-6}$ (approx.)	91%	(Rubinstein and Glynn 2009)
Credit Portfolio	*$5*10^{-4}$*	*95%*	*Figure 21*

5.3.2. Alternatives to the Cross-Entropy method for credit portfolio analysis

Table 2 shows a comparison of the Cross-Entropy method against other results that are used to optimise credit portfolio analyses/importance sampling distributions. As can be seen from the table, the improvements of the Cross-Entropy method are comparable to those of other methods. The main advantage of the Cross-Entropy method however is that it is easy to implement and faster than numerical methods.

Table 2: comparison between naïve Monte Carlo and credit portfolio optimisations found in the academic literature

Method	Value sought / Confidence level	Reduction in Relative Error	Notes	Reference
Variance minimisation (*)	99.98%	95%	Focuses on Expected Shortfall, but mentions similar results for Economic Capital	(Kalkbrener, Lotter et al. 2003)
Variance minimisation (*)	99.95%	94%		(Glasserman and Li 2005)
Variance minimisation (*)	99.95%	93%	This research is based on earlier findings by the same researcher (Glasserman and Li 2005). There are different levels of improvement reported, depending on the model used. The value taken is the median of these improvements across the different models.	(Glasserman, Kang et al. 2008)

Method	Value sought / Confidence level	Reduction in Relative Error	Notes	Reference
Conditional Monte Carlo	99.96%	94%		(Chan and Kroese 2010c)
Cross Entropy	*99.95%*	*95%*		*Figure 21*

(*) Note that the different "variance minimisation" methods are not identical: since finding the real minimum variance is analytically intractable numerical methods must be used; the different articles use different techniques to approximate the minimum variance importance distribution.

5.4. *Sensitivity to inclusion of stochastic LGD*

As mentioned in section 2.4, the Loss Given Default can be used as a fixed variable, or can be an independent stochastic factor. Since the latter would include additional noise in the calculation, it is relevant to ensure that the Cross Entropy method is not adversely impacted by this. The settings used in both estimations are the ones that provided the best results within a reasonable calculation budget in section 0: ρ=0.1, N1=10,000, N2=30,000 and a confidence level of 99.95%. The results are shown in Table 3.

Table 3: difference reduction in relative error between choosing a fixed or stochastic LGD

LGD approach	Reduction in Relative Error
Fixed LGD	95%
Stochastic LGD	88%

As can be seen from this table, the improvement is less when a stochastic LGD is used, which is expected because an additional random element is included that is not affected by the Cross-Entropy method. The reduction is still material, however.

5.5. *Sensitivity to shifting the mean vs. shifting both mean and standard deviation*

If the normal distribution is used for the underlying parameters, there are three choices for parameterisation:

- Normal distribution parameterised by the mean (standard deviation is fixed at 1).
- Normal distribution parameterised by the standard deviation (mean is fixed at 0).
- Normal distribution parameterised by both mean and standard deviation.

Of these, parameterisation of the mean makes more sense than parameterisation of the standard deviation, as the former will introduce a bias towards "bad" economic conditions. Parameterisation of the standard deviation would only increase the implied volatility of future economic conditions. Theoretically the best result should come from a combined parameterisation, as this distribution will have the lowest Cross-Entropy distance from the optimal distribution.

Table 4 shows the Relative Error for two sets of runs: one where only the mean is shifted and one where both the mean and the standard deviation are. The used parameter settings are ρ=0.1, N1=10,000, N2=30,000 and a confidence level of 99.95%. As the table shows, only shifting the mean provides slightly better results than shifting both mean and standard deviation. While unexpected, these results suggest that the optimal distribution has a standard deviation that is very close to the original value (i.e. optimal standard deviation is close to 1). In that case, any perturbation of the standard deviation is likely to bring it further away from the optimal one that it was already very close to.

Table 4: difference reduction in relative error if only the mean is shifted vs. shifting both mean and standard deviation

Shifted Parameters	Reduction in Relative Error
Mean	95%
Mean and Standard Deviation	94%

5.6. *Sensitivity to the portfolio composition*

In order to determine whether the methods works on portfolios of different compositions, the method is also tested on synthetic portfolios as described in Section 4.2. The reductions in relative error are shown Figure 23 for the homogeneous portfolio (with equal contributions of each position) and in Figure 24 for the heterogeneous portfolio (with unequal contributions of each position). While the reductions are slightly lower than for the ING portfolio, they are still significant.

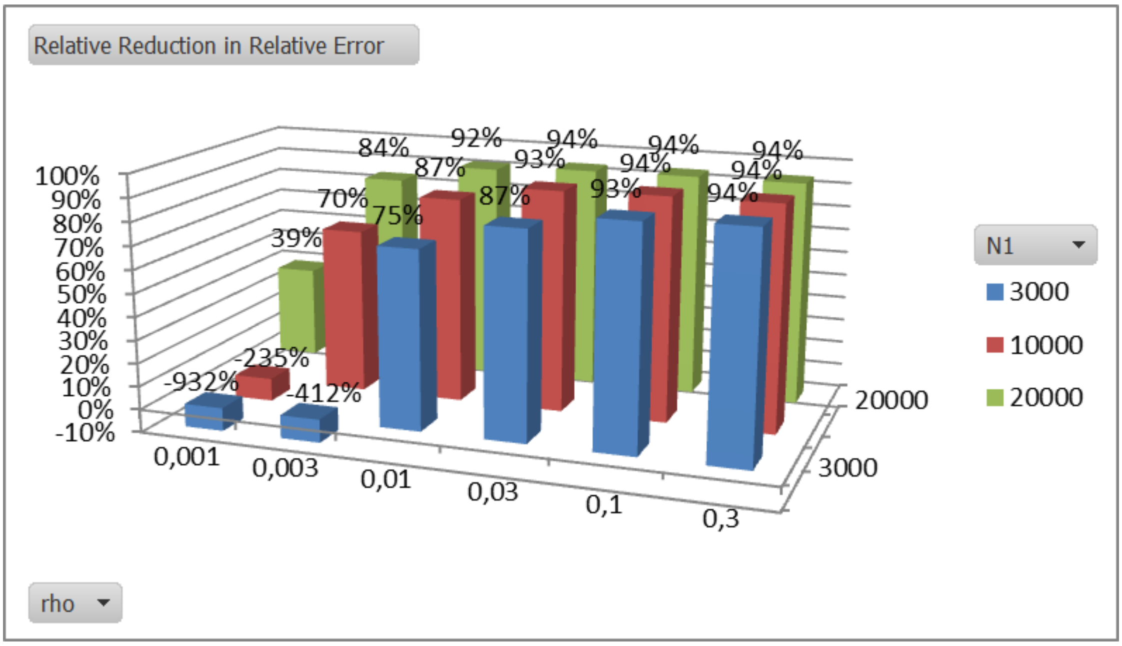

Figure 23: Relative reduction in Relative Error for the homogeneous synthetic portfolio with different Cross Entropy parameter settings

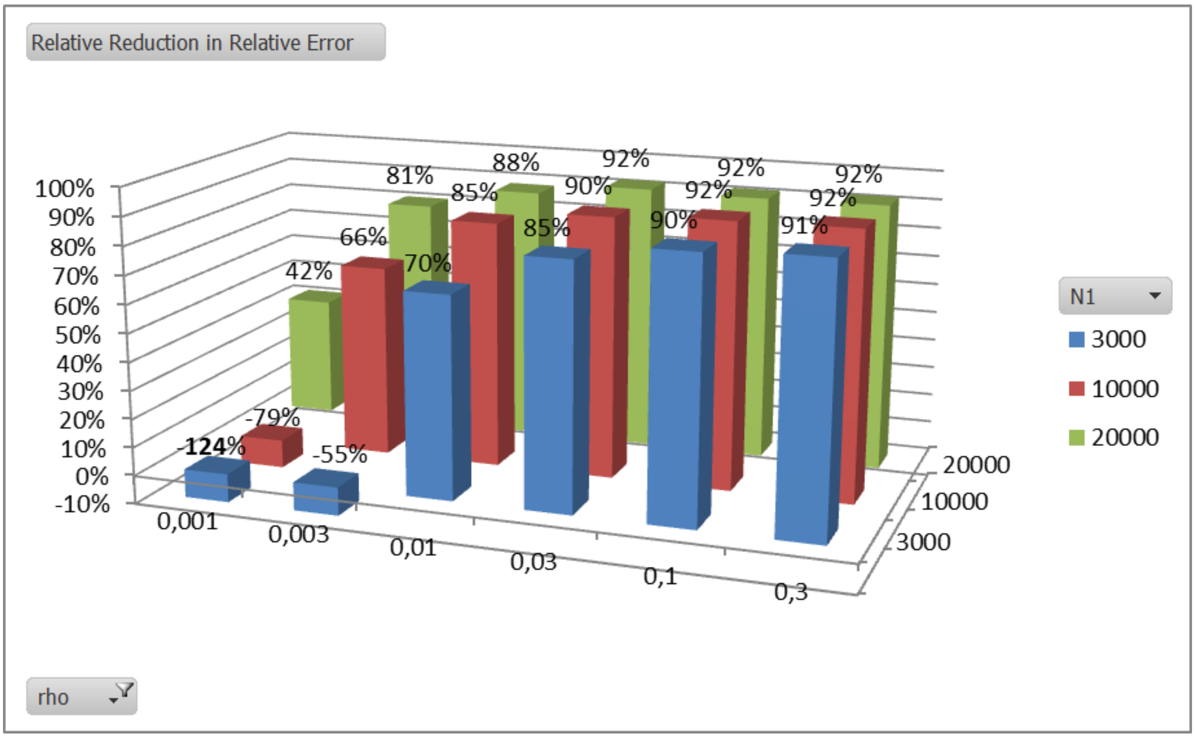

Figure 24: Relative reduction in Relative Error for the heterogeneous synthetic portfolio with different Cross Entropy parameter settings

5.7. *Variance in the Economic Capital number*

The previous sections focused on estimating the Relative Error in the estimation of the chance that the rare event occurs. In practice, the estimate is often reversed: how much capital would be required to ensure that losses do not exceed this amount more than a given percentage of the time. The reduction in the standard deviation of this measure when compared to naïve Monte Carlo is shown in Figure 25. As can be seen, similar reductions are seen in this measure as in the relative error of the chance that the rare event occurs.

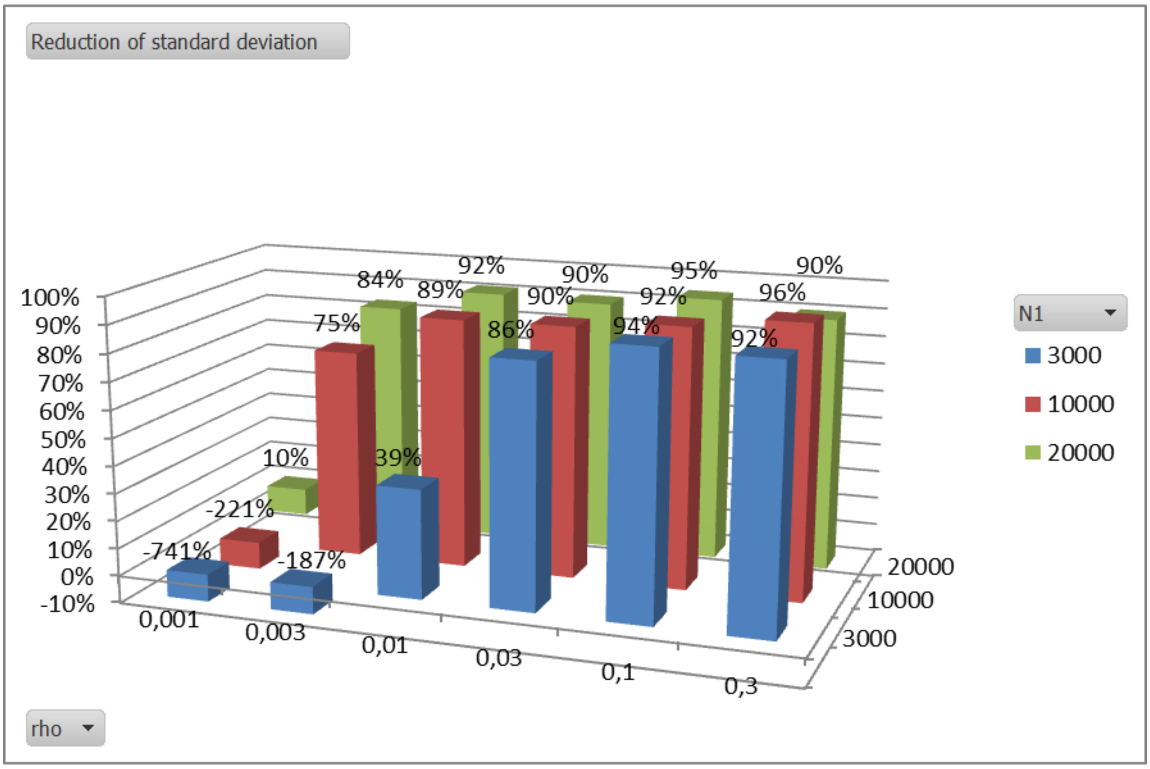

Figure 25: reduction in the standard deviation of the Economic Capital when using the Cross Entropy method relative to naïve Monte Carlo

"Real" validation: comparing the model predictions to observations

While comparing the results of the improved model with naïve Monte Carlo or other researchers gives some indication of the model performance, the real proof of the pudding is in the eating: how do the results of the model compare to observations in the real world? Does the improved model predict the future more accurately than the original model? Did it predict the credit crisis?

In the context of the research performed in this project, the answer to the last two questions will unfortunately be "no". The model will not predict the future more accurately, as it has the same fundamental flaws as the model it is based on. It will reach the same – incorrect – conclusions, but will reach these much more quickly.

The validation of rare event models against real observations is problematic in most cases, and in particular in this instance where there is only one observable system available (the world economy) and the time horizon is large (typically one year). This project will not try to address the modelling problem directly, but will hopefully provide tools that increase the efficiency of the calculation and that allow other researcher to refine their models.

6. Discussion and suggestions for further research

The results provided in Section 5 indicate that the Cross Entropy method is very suitable to determine Importance Sampling distributions for Credit Portfolio Monte Carlo calculations. The results are similar to those found by the more commonly used variance minimisation techniques, but the analytical solution can be implemented and calculated with significantly less effort and statistical background knowledge.

While the method is only applied to three underlying portfolios, the consistency of the results suggests that the portfolio composition is not a major influence on the applicability of the method. As such, Credit Portfolio Monte Carlo calculations can be added to the already extensive list of domains where the Cross Entropy method is applied.

The method is only applied to one multi-factor model, while different models or different calibrations are possible. However, it can be noted that one of the most important factors is the sensitivity of the company to the world economy, which is reflected in the R-squared. The R-squared does not depend on the factor model as such, but is a characteristic of the borrower. The rest of the model is essentially only taking linear combinations of normally distributed values and it is unlikely that other variations of such models will render the Cross Entropy method useless. However, a more thorough test of this assumption could be part of a further study.

The canonical Cross Entropy method only supports distributions from the exponential family of distributions. While this is an extensive list, it is not complete and misses relevant “heavy tailed” distributions like the Cauchy and Student-t distributions. For this, the Cross Entropy method must be extended, for instance by using the Generalised Cross Entropy method (Botev, Kroese et al. 2007) that suggests using difference distance measures than the Kullback-Leibler divergence and the variation mentioned by Chan and Kroese (2010b) that does not use Importance Sampling, but samples directly from the distribution with the lowest divergence from the optimal distribution. Both variations are beyond the scope of this project, but validation and application to the field of credit portfolios could be part of further research.

The ease of implementation of the Cross Entropy method makes it available for smaller projects where there are no resources available to develop a numerical solution. These implementations are also more likely to work with the normal distribution for asset returns instead of more exotic ones. Also the calibration of the parameters of the method is manageable without extensive statistical knowledge, because most settings will provide adequate improvements provided that they are within reasonable limits. Also larger projects will benefit from the performance improvement, but these may have to invest in the more advanced and experimental techniques of the Cross Entropy method if certain “heavy-tailed” distributions are to be used.

A particular aspect that this research shows is that the alleged “degeneracy of the likelihood ratio” did not occur for the problem at hand, even though the number of dimensions exceeds the threshold that is normally used for applicability of the method. In order to further advance the usability of the Cross Entropy method, it could be investigated under which conditions this degeneracy actually occurs and under which conditions it does not.

7. Conclusion

7.1. *Research Questions*

Given the results found, let us return to the research questions at the beginning of the document:

RQ1: *How can the Cross-Entropy method be applied to Monte Carlo analysis of credit portfolios, and ING's credit portfolio in particular?*

The ING portfolio can be modelled with an industry standard asset correlation model. The factors in this model are distributed according to a distribution that is supported in the Cross Entropy method (the normal distribution) and allows a straightforward implementation.

RQ2: *How big is the problem of degeneracy and to what extent is it influenced by the number of levels in the multi-level approach and the number of factors in the multi-factor model?*

Degeneracy did not seem to occur at any parameter setting, or at least did not give rise to substantial errors in the final estimate.

RQ3: *What performance improvements can be achieved as a result of applying the Cross-Entropy method to this Monte Carlo Credit-Risk analysis?*

Depending on the calculation budget for training, the relative error in the estimate that the losses will exceed the Economic Capital can be reduced by 90%-95%. Similar reductions in the relative error would require increasing the number of scenarios by more than a hundredfold.

RQ4: *How do the results on the Cross-Entropy method compare to other results found in academic publications, including results on other acceleration techniques like exponential twisting, particle filters or conditional Monte Carlo?*

The results of the Cross-Entropy method found compare very favourably with those found in the academic literature. While the total reduction in the error is similar, it is reached with significantly less effort in terms of implementation and calculation.

7.2. *Problem Statement*

We can now also look to which extent we addressed the problem statement:

PS: *To what extent can the Cross-Entropy method be applied to reduce the error-margin in Monte Carlo-based portfolio analysis of Credit Risk?*

The Cross Entropy method has proven to be an excellent candidate to find an Importance Sampling distribution that reduces the error margin. The results are significant – reductions are found up to 95% in relative error – and comparable to variance minimisation techniques that are more commonly used in the industry today.

While the canonical Cross Entropy method that is used in this project is limited in the sense that it only supports (thin-tailed) distributions from an exponential family, generalisations of the method exist that expand the applicability of the method further.

8. References

8.1. *Books*

Alexander, C. and E. Sheedy (2004). The professional risk managers' handbook: A comprehensive guide to current theory and best practices. Wilmington, DE, Prmia Publications.

Taleb, M. (2007). The Black Swan, New York: Random House.

Rubinstein, R. Y. and D. P. Kroese. (2007). Simulation and the Monte Carlo Method Second Edition. New York, John Wiley & Sons.

8.2. *Scientific publications*

Asmussen, S., D. P. Kroese, et al. (2005). "Heavy Tails, Importance Sampling and Cross Entropy." Stochastic Models **21**(1): 57-76.

Botev, Z. I., D. P. Kroese, et al. (2007). "Generalized Cross-entropy Methods with Applications to Rare-event Simulation and Optimization." Simulation **83**(11): 785-806.

Carmona, R., J.-P. Fouque, et al. (2009). "Interacting particle systems for the computation of rare credit portfolio losses." Finance and Stochastics **13**(4): 613-633.

Chan, J. C. C. and D. P. Kroese (2010a). "Efficient estimation of large portfolio loss probabilities in t-copula models." European Journal of Operational Research **205**(2): 361-367.

Chan, J. C. C. and D. P. Kroese (2010b). "Improved Cross-Entropy Method for Estimation." Statistics and Computing.

Chan, J. C. C. and D. P. Kroese (2010c). "Rare-event probability estimation with conditional Monte Carlo." Annals of Operations Research **To appear**.

Crouhy, M., D. Galai, et al. (2000). "A comparative analysis of current credit risk models." Journal of Banking & Finance **24**(1-2): 59-117.

de Boer, P.-T., D. Kroese, et al. (2005). "A Tutorial on the Cross-Entropy Method." Annals of Operations Research **134**(1): 19-67.

Deng, S., K. Giesecke, et al. (2011). Sequential Importance Sampling And Resampling for Dynamic Portfolio Credit Risk Operations Research **To appear**.

Glasserman, P., P. Heidelberger, et al. (2000). "Variance Reduction Techniques for Estimating Value-at-Risk." Management Science **46**(10): 1349-1364.

Glasserman, P., W. Kang, et al. (2008). "Fast Simulation of Multifactor Portfolio Credit Risk." Operations Research **56**(5): 1200-1217.

Glasserman, P. and J. Li (2005). "Importance Sampling for Portfolio Credit Risk." Management Science **51**(11): 1643-1656.

Gordy, M. B. (2000). "A comparative anatomy of credit risk models." Journal of Banking & Finance **24**(1-2): 119-149.

Heidelberger, P. (1995). "Fast simulation of rare events in queueing and reliability models." ACM Trans. Model. Comput. Simul. **5**(1): 43-85.

Hui, K. P., N. Bean, et al. (2005). "The Cross-Entropy Method for Network Reliability Estimation." Annals of Operations Research **134**(1): 101-118.

Juneja, S. (2007). "Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error." Queueing Systems **57**(2): 115-127.

Lucas, D. J. (1995). "Default correlation and credit analysis." Journal of Fixed Income **4**(4): 76-87.

Merton, R. C. (1974). "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." The Journal of Finance **29**: 449-470.

Moral, P. D. and J. Garnier (2005). "Genealogical Particle Analysis of Rare Events." The Annals of Applied Probability **15**(4): 2496-2534.

Rubinstein, R. Y. and P. W. Glynn (2009). "How to deal with the curse of dimensionality of likelihood ratios in Monte Carlo simulation." Stochastic Models **25**: 547 – 568.

Taleb, M. (2007). The Black Swan, New York: Random House.

Taleb, N. N. (2009). "Finiteness of variance is irrelevant in the practice of quantitative finance." Complexity **14**(3): 66-76.

Zhou, C. (2001). "An Analysis of Default Correlations and Multiple Defaults." The Review of Financial Studies **14**(2): 555-567.

8.3. *Industry papers, conference proceedings, reports and regulation*

Asmussen, S., and D.P. Kroese (2004). Improved algorithms for rare event simulation with heavy tails, Research Report. Aarhus University, Denmark, , Department of Mathematical Sciences

Atzberger, P. J. "Strategies for Improving the Efficiency of Monte-Carlo Methods." from http://www.math.ucsb.edu/~atzberg/finance/improvingMonteCarloMethods.pdf.

Bassamboo, A., S. Juneja, et al. (2005). "Portfolio Credit Risk with Extremal Dependence." SSRN eLibrary.

BIS (2006). International Convergence of Capital Measurement and Capital Standards, Bank for International Settlements.

Creal, D. (2009). A survey of sequential Monte Carlo methods for economics and finance, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Egloff, D., M. Leippold, et al. (2005). Optimal Importance Sampling for Credit Portfolios with Stochastic Approximation Available at SSRN: http://ssrn.com/abstract=693441.

Gupton, G. M., Finger, C.C., Bhatia, M. (1997). CreditMetrics-Technical Document. New York, J.P. Morgan & Co. Incorporated.

Kalkbrener, M., H. Lotter, et al. (2003). Sensible and efficient capital allocation for credit portfolios, Deutsche Bank AG, .

Markit (2011). "Markit iTraxx Europe Series 16 Final Membership List." Retrieved 10 January 2012, 2012, from http://www.markit.com/assets/en/docs/products/data/indices/credit-and-loan-indices/iTraxx/Markit%20iTraxx%20Europe%20Final%20List%20Series%2016.pdf.

Morokoff, W. J. (2004). An importance sampling method for portfolios of credit risky assets. Proceedings of the 36th conference on Winter simulation. Washington, D.C., Winter Simulation Conference: 1668-1676.

Products, C. S. F. (1997). CreditRisk+: A CreditRisk Management Framework. London.

Rafat Khan, N. M. R. (2011). "2010 Annual Global Corporate Default Study And Rating Transitions." Retrieved 10 January 2012, 2012, from http://www.standardandpoors.com/ratings/articles/en/us/?assetID=1245302234237.

Schönbucher, P. J. (2000). Factor Models for Portfolio Credit Risk. Bonn University, Department of Statistics.

Taleb, N. N. (2010). "Why Did the Crisis of 2008 Happen?" SSRN eLibrary.

Tchistiakov, V., J. de Smet, et al. (2004). "A credit loss control variable." Risk (July): 81-85.

Vasicek, O. (1987). Probability of loss on loan portfolio, KMV Corporation, .

Appendix A: Acronyms

cdf	cumulative density function
CE	Cross Entropy
EAD	Exposure at Default
iid	independent and identically distributed
IS	Importance Sampling
LGD	Loss Given Default
LR	Likelihood Ratio
MC	Monte Carlo
PD	Probability of Default
pdf	probability density function
RE	Relative Error

Appendix B: Formal application of the Cross Entropy method to credit portfolios

The following sections are largely taken from the Cross Entropy Tutorial (de Boer, Kroese et al. 2005), but with the addition of the interpretation of the parameters in the context of the specific problem domain of Credit Portfolios.

Let $X = (X_1,\ldots,X_n)$ be a random vector taking values in some space ***X***. In the context of Credit Portfolio measurement, *X* represents the state of the economy represented by the relative change in asset values in particular country and industry segments. Let $\{f(\cdot;v)\}$ be a family of probability density functions (pdfs) on ***X***. For instance, following the approach of Credit Metrics (Gupton 1997), f could be the family of normal distributions, since Credit Metrics assumes that asset returns are normally distributed.

Let *S* be some real-valued function on ***X***, in this case the total credit loss in the portfolio given the state of the economy. Suppose we are interested in the probability that *S(x)* is greater than or equal to some real number γ, under $\{f(\cdot;u)\}$. This probability can be expressed as

$$l = P(S(X) \geq \gamma) = E_u I_{\{S(X)\geq\gamma\}}$$

If this probability is very small, say smaller than 10^{-5}, we call $\{S(X) \geq \gamma\}$ a rare event. A straightforward way to estimate is to use naïve Monte-Carlo simulation: Draw a random sample $(X_1,\ldots,X_n)$ from $f(\cdot;u)$; then

$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(X)\geq\gamma\}}$$

is an unbiased estimator for *l*. However, this poses serious problems when $\{S(X) \geq \gamma\}$ is a rare event. In that case a large simulation effort is required in order to estimate *l* accurately, i.e., with small relative error or a narrow confidence interval. An alternative is based on importance sampling: take a random sample $(X_1,\ldots,X_n)$ from an importance sampling (different) density *g* on ***X***, and evaluate using the Likelihood Ratio estimator:

$$l' = \frac{1}{N}\sum_{i=1}^{N} I_{\{S(X)\geq\gamma\}} \frac{f(X_i;u)}{g(X_i)} \qquad (9)$$

The best way to estimate *l* is to use the change of measure with density

$$g^* := \frac{I_{\{S(X)\geq\gamma\}} f(x;u)}{l} \qquad (10)$$

Namely, by using this change of measure we have in (9):

$$I_{\{S(X)\geq\gamma\}} \frac{f(X_i;u)}{g^*(X_i)} = l \tag{11}$$

for all i. In other words, the estimator has zero variance and we only need N=1 sample.

The difficulty is of course that this *g** depends on the unknown parameter *l*. Also, it is often convenient to choose g from the family of densities $\{f(\cdot;v)\}$. The idea now is to choose the parameter vector, called the reference parameter (sometimes called tilting parameter) *v* such that the distance between the densities *g* and $\{f(\cdot;v)\}$ is minimal. A particular convenient measure of distance between two densities g and h is the Kullback-Leibler distance, which is also termed the cross-entropy between g and h. The Kullback-Leibler distance is defined as:

$$D(g,h) = E_g \ln\frac{g(X)}{h(X)} = \int g(x)\ln g(x)dx - \int g(x)\ln h(x)dx$$

Minimizing the Kullback-Leibler distance between g* in (10) and $\{f(\cdot;v)\}$ is equivalent to choosing *v* such that $-\int g^*(x)\ln(f(\cdot;v))dx$ is minimised, because $\int g^*(x)\ln(g^*(x))dx$ is constant relative to *v*. As explained in the Cross Entropy Tutorial (de Boer, Kroese et al. 2005), the optimal *v* can be obtained by solving the following system of equations:

$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(X_i\geq\gamma)\}} W(X_i;u;w)\nabla \ln f(X_i;v) = 0 \tag{12}$$

where $W(X_i;u;w)$ is the likelihood ratio

$$W(x;u;w) = \frac{f(x;u)}{f(x;w)}$$

and the gradient is relative to *v*. The parameter *u* represents the original pdfs, and the parameter *w* any shifted distribution that is applied to generate *X*. Note that if this algorithm is applied to the original distribution, *u* and *w* are the same and *W* is always equal to 1. Note also that the $I_{\{S(X_i\geq\gamma)\}}$ factor implies that only observations that cross the "rare event" threshold need to be considered when solving the system.

If *f* is a normal distribution parameterised by the mean, then $\nabla \ln f(X_i;v)$ can be rewritten as:

$$\frac{d}{dv}\ln\left[\frac{1}{2\pi\sigma^2}e^{-\frac{(X_i-v)^2}{2\sigma^2}}\right] = \frac{d}{dv}\left[\ln\frac{1}{2\pi\sigma^2} - \frac{(X_i-v)^2}{2\sigma^2}\right] = \frac{X_i - v}{\sigma^2}$$

Now the system of equations in (12) can be solved as follows:

$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(X_i \geq \gamma)\}} W(X_i;u;w) \frac{X_i - v}{\sigma^2} = 0$$
$$\Leftrightarrow \sum_{i=1}^{N} I_{\{S(X_i \geq \gamma)\}} W(X_i;u;w) X_i - I_{\{S(X_i \geq \gamma)\}} W(X_i;u;w) v = 0 \; (if \;\; N > 0 \wedge \sigma > 0)$$
$$\Leftrightarrow \frac{\sum_{i=1}^{N} I_{\{S(X_i \geq \gamma)\}} W(X_i;u;w) X_i}{\sum_{i=1}^{N} I_{\{S(X_i \geq \gamma)\}} W(X_i;u;w)} \qquad (13)$$

Note that this formula as an intuitive appeal, as the new mean of the pdf is set to the "weighted average" of the draws that led to the rare event.

Appendix C: description of the ING portfolio

In this research project, a representation of a commercial banking portfolio is used. This portfolio represents the Commercial Banking positions (including small companies, but excluding private retail clients, intercompany positions and defaulted customers) of ING. Certain positions with known data quality issues have been removed from these data, so the total position and results from the calculation will not reconcile to publicly published numbers. The data is selected as of the end of 2008. This date is selected because:

- End-of-year data is usually the most reliable;
- The date is relatively recent, but not recent enough to be commercially sensitive.

Figure 26 to Figure 29 provide overviews of the distribution of the exposure of this portfolio across Countries, Industries, Ratings and Exposure Class[18]. Figure 30 and Figure 31 show how the R-squared (which is comparable to the correlation in the Regulatory Capital formula) and Loss Given Default are distributed. These diagrams are histograms, but for representation purposes they are shown as line graphs. The selected interval is 1%. The peaks in the distribution are caused by the underlying parameterisation that is used to derive these values. Finally, Table 5 shows the relative contribution of the top 20 highest exposures. As can be seen from these diagrams and tables, ING has material concentrations, for instance in the Benelux and on the financial industry. Also name concentration is relevant, with the top 20 customers accounting for more than 16% of the total portfolio exposure.

[18] Exposure Class is a classification that is used by the regulator and also frequently used in official reporting. It is a combination of product and customer attributes. For details, see BIS (2006). International Convergence of Capital Measurement and Capital Standards, Bank for International Settlements.

.

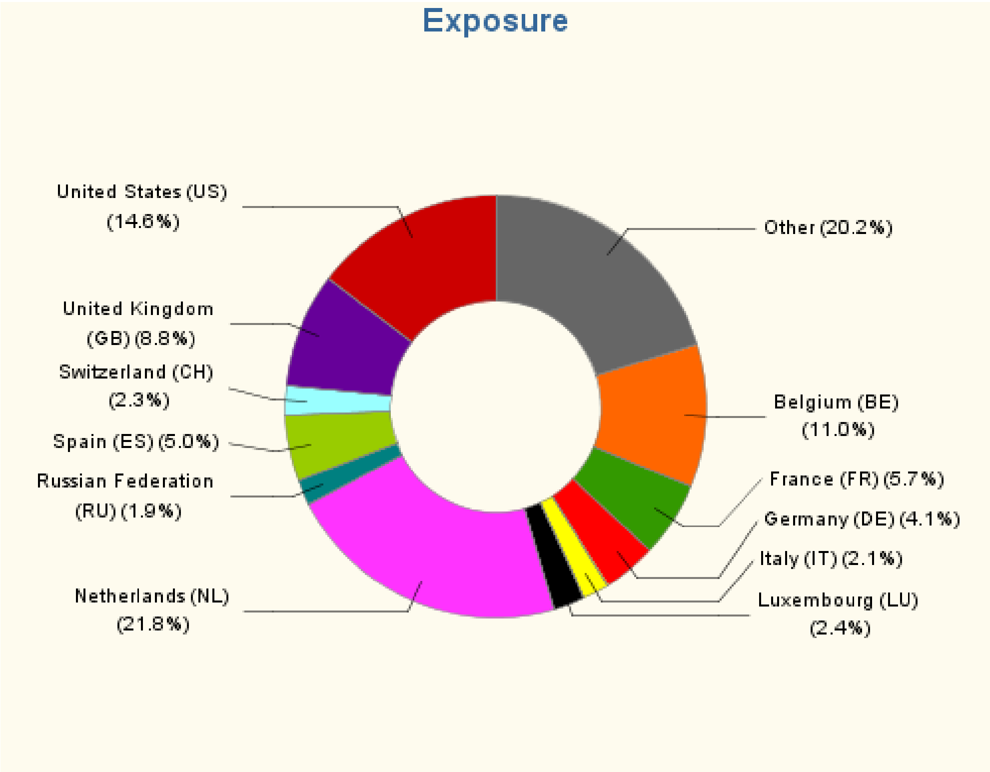

Figure 26: Exposure per Country

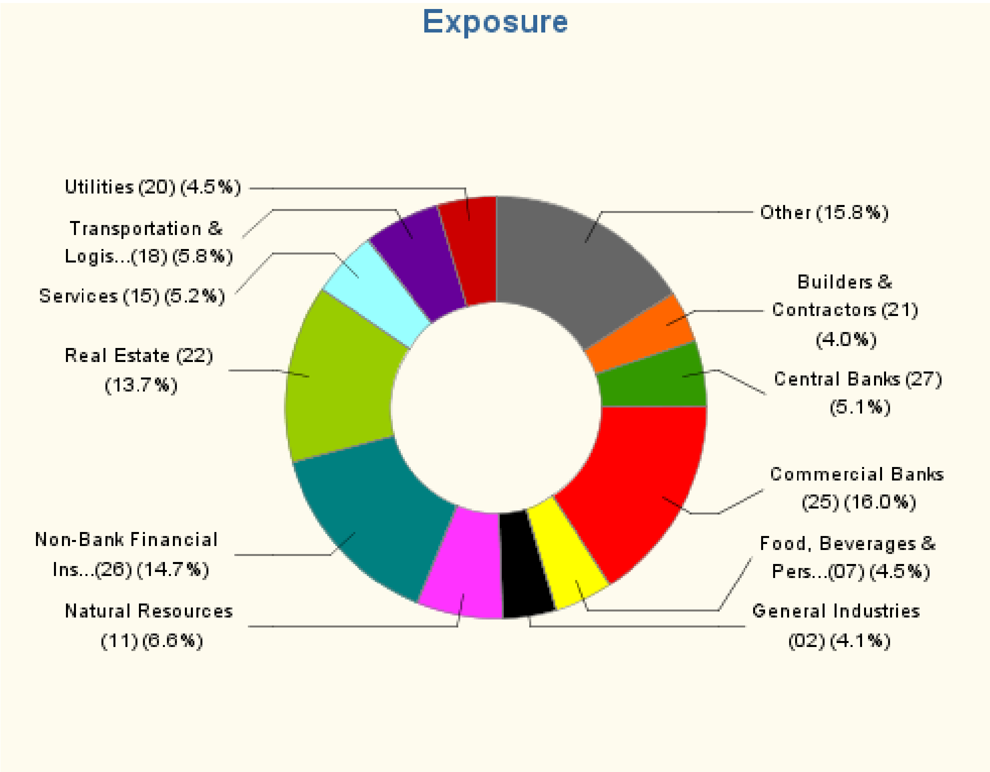

Figure 27: Exposure per Industry

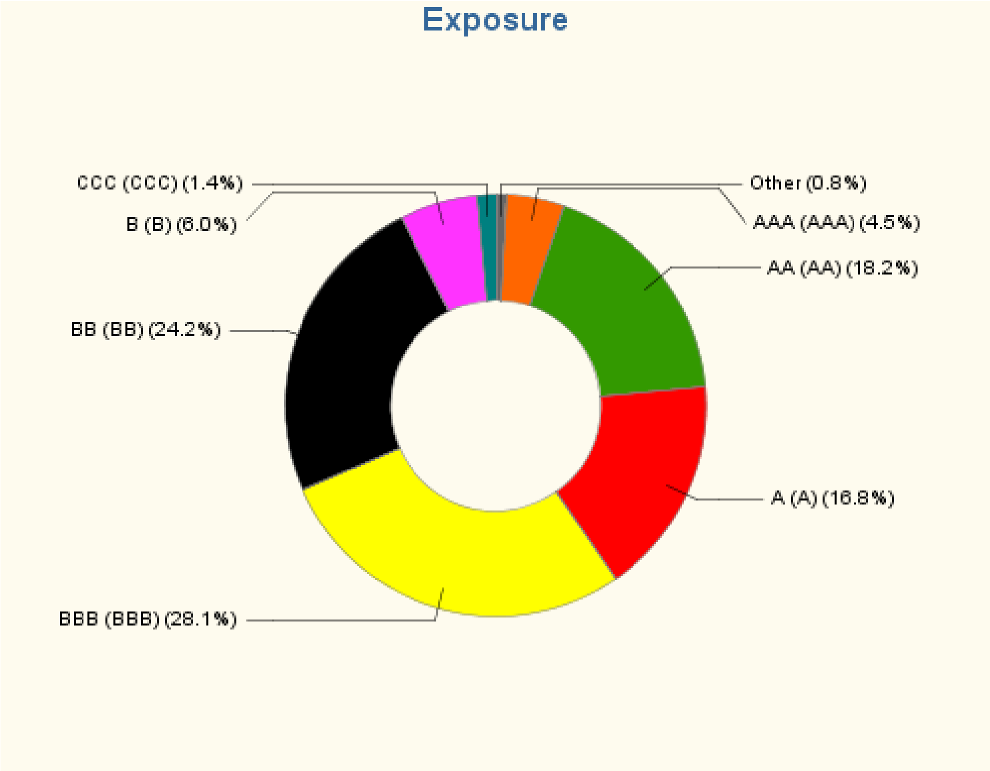

Figure 28: Exposure per Risk Rating

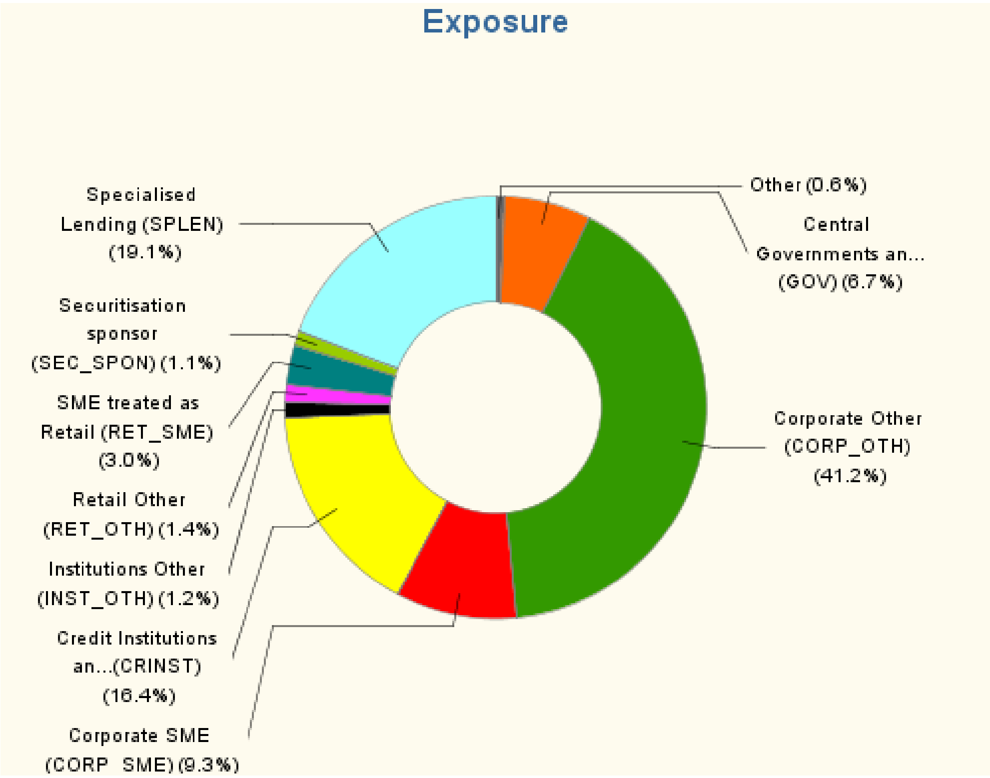

Figure 29: Exposure per Exposure Class

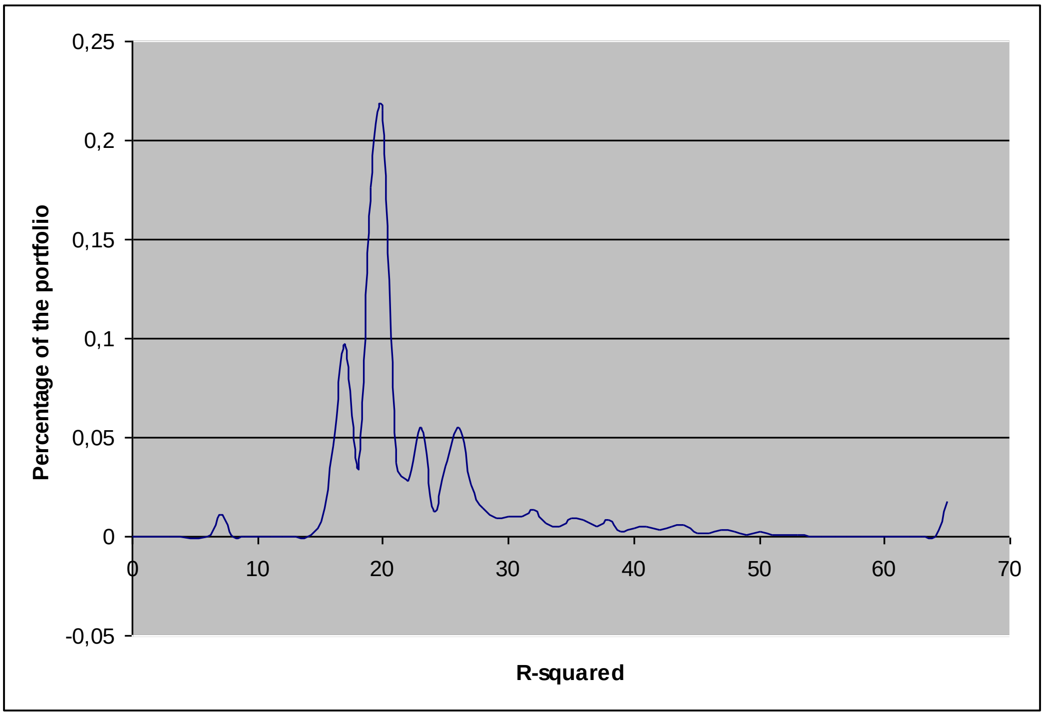

Figure 30: Distribution of the R-squared (histogram with interval size of 1%)

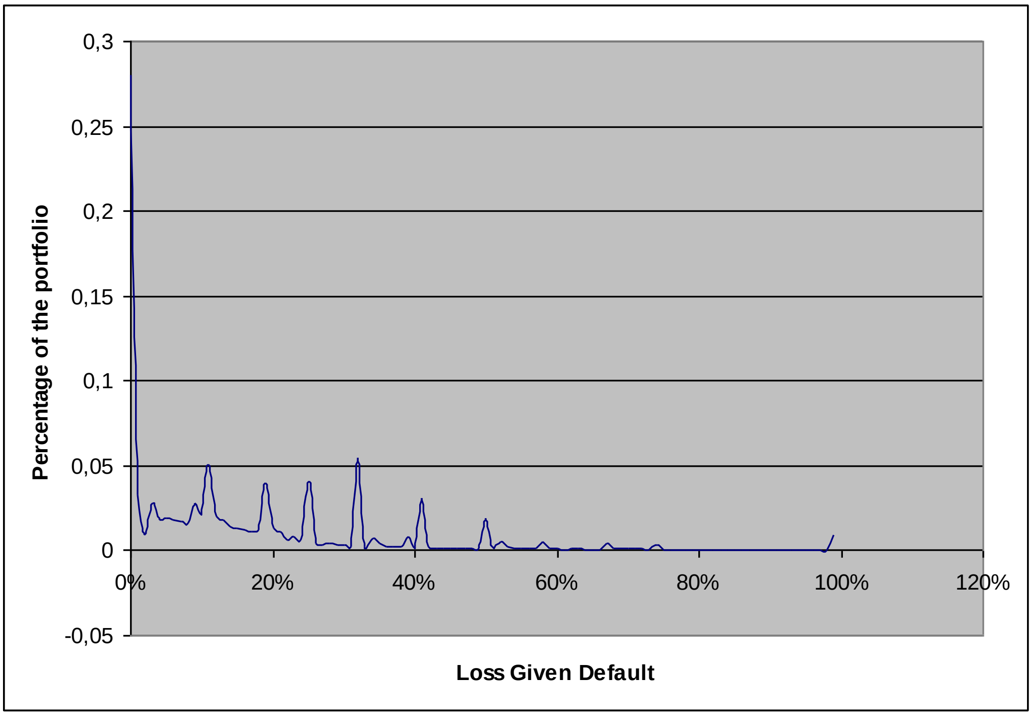

Figure 31: Distribution of the Loss Given Default (histogram with interval size of 1%)

Table 5: Top 20 highest exposures expressed as percentage of the total exposure

Rank	Percentage of Total exposure
1	2,42%
2	2,40%
3	2,15%
4	1,27%
5	0,95%
6	0,84%
7	0,63%
8	0,60%
9	0,57%
10	0,56%
11	0,55%
12	0,52%
13	0,48%
14	0,46%
15	0,41%
16	0,39%
17	0,36%
18	0,34%
19	0,32%
20	0,32%
Total	**16,53%**

Appendix D: description of the parameters of the synthetic portfolios

Table 6 shows the synthetic portfolios that are used as part of this research next to the real portfolio. The first three columns are taken directly from the Markit iTraxx index. The next descriptive columns ("R-squared" to "Country") are enrichments based on public data or assumptions. The final three columns indicate the contribution of each obligor to the total portfolio. Portfolio 1 is a homogeneous portfolio, while Portfolio 2 follows a power law distribution. All data used is as of end 2008.

Table 6: description of the synthetic portfolios

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
Autos & Industrials	ADO	Adecco S.A.	44,69	BBB	BAA3		BBB	0,26%	40,00%	Temporary Help Services (561320)	561320	CH
Autos & Industrials	VLVY	Aktiebolaget Volvo	36,33	BBB	BAA2	BBB	BBB	0,26%	40,00%	Heavy Duty Truck Manufacturing (336120)	336120	CE
Autos & Industrials	AKZO	AKZO Nobel N.V.	31,22	BBB+	BAA1		BBB+	0,16%	40,00%	Petrochemical Manufacturing (325110)	325110	NL
Autos & Industrials	ALSTOM	ALSTOM	30,75	BBB	BAA1		BBB	0,26%	40,00%	Mechanical Power Transmission Equipment Manufacturing (333613)	333613	FR
Autos & Industrials	AAUK	Anglo American plc	45,72	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Metal Service Centers and Other Metal Merchant Wholesalers (423510)	423510	GB
Autos & Industrials	ARMLL	ArcelorMittal	38,05		BAA3		BBB	0,26%	40,00%	Iron and Steel Mills (331111)	331111	LU
Autos & Industrials	ATSPA	ATLANTIA S.P.A.	42,89	A-	A3		A-	0,07%	40,00%	Highway, Street, and Bridge Construction (237310)	237310	IT
Autos & Industrials	BAPLC	BAE SYSTEMS PLC	35,88	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Aircraft Manufacturing (336411)	336411	GB
Autos & Industrials	BASFSE	BASF SE	42,11	A+	A1	A+	A+	0,05%	40,00%	All Other Plastics Product Manufacturing (326199)	326199	DE
Autos & Industrials	BYIF	Bayer Aktiengesellschaft	35,01	A-	A3	A-	A-	0,07%	40,00%	Petrochemical Manufacturing (325110)	325110	DE
Autos & Industrials	BMW	Bayerische Motoren Werke Aktiengesellschaft	41,35	A-	A2		A-	0,07%	40,00%	Automobile Manufacturing (336111)	336111	DE
Autos & Industrials	STGOBN	COMPAGNIE DE SAINT-GOBAIN	35,01	BBB	BAA2	BBB+	BBB	0,26%	40,00%	Other Construction Material Merchant Wholesalers (423390)	423390	FR
Autos & Industrials	MICH-CoFinMich	Compagnie Financiere Michelin	31,27	BBB	BAA2	BBB	BBB	0,26%	40,00%	All Other Rubber Product Manufacturing (326299)	326299	CH

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
Autos & Industrials	DAMLR	Daimler AG	47,10	BBB+	A3	A-	BBB+	0,16%	40,00%	Automobile Manufacturing (336111)	336111	DE
Autos & Industrials	EAD	European Aeronautic Defence and Space Company EADS N.V.	36,91	A-	A1		A-	0,07%	40,00%	Aircraft Manufacturing (336411)	336411	NL
Autos & Industrials	FINMEC	FINMECCANICA S.P.A.	38,28	BBB-	BAA2	BBB	BBB-	0,31%	40,00%	Other Aircraft Parts and Auxiliary Equipment Manufacturing (336413)	336413	IT
Autos & Industrials	GLCORE	Glencore International AG	48,21	BBB	BAA2		BBB	0,26%	40,00%	Coal and Other Mineral and Ore Merchant Wholesalers (423520)	423520	CH
Autos & Industrials	HOLZSW	Holcim Ltd	43,84	BBB	BAA2	BBB	BBB	0,26%	40,00%	Cement Manufacturing (327310)	327310	CH
Autos & Industrials	KDSM	Koninklijke DSM N.V.	17,29	A	A3	A-	A	0,07%	40,00%	Paint and Coating Manufacturing (325510)	325510	NL
Autos & Industrials	LNX	LANXESS Aktiengesellschaft	29,48	BBB	BAA2	BBB	BBB	0,26%	40,00%	Synthetic Rubber Manufacturing (325212)	325212	DE
Autos & Industrials	LINDE	Linde Aktiengesellschaft	16,98	A-	A3		A-	0,07%	40,00%	Plumbing, Heating, and Air-Conditioning Contractors (238220)	238220	DE
Autos & Industrials	PNL	PostNL N.V.	31,31	BBB	BAA-		BBB	0,26%	40,00%	Postal Service (491110)	491110	NL
Autos & Industrials	RNTKIL	RENTOKIL INITIAL PLC	30,77	BBB-			BBB-	0,31%	40,00%	Exterminating and Pest Control Services (561710)	561710	GB
Autos & Industrials	SANFI	SANOFI	32,99	AA-	A2	AA-	AA-	0,03%	40,00%	Pharmaceutical Preparation Manufacturing (325412)	325412	FR
Autos & Industrials	SIEM	Siemens Aktiengesellschaft	47,37	A+	A1	A+	A+	0,05%	40,00%	Electric Housewares and Household Fan Manufacturing (335211)	335211	DE
Autos & Industrials	SOLVAY	Solvay	29,71	BBB+	BAA1	A-	BBB+	0,16%	40,00%	Drugs and Druggists' Sundries Merchant Wholesalers (424210)	424210	BE
Autos & Industrials	VLOF	VALEO	39,54		BAA3		BBB	0,26%	40,00%	All Other Motor Vehicle Parts Manufacturing (336399)	336399	FR
Autos & Industrials	VINCI	VINCI	30,10	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Power and Communication Line and Related Structures Construction (237130)	237130	FR
Autos & Industrials	VW	VOLKSWAGEN AKTIENGESELLSCHAFT	46,86	A-	A3	A-	A-	0,07%	40,00%	Automobile Manufacturing (336111)	336111	DE
Autos & Industrials	XSTR	XSTRATA PLC	41,87	BBB+	BAA2		BBB+	0,16%	40,00%	Lead Ore and Zinc Ore Mining (212231)	212231	GB
Consumers	ELTLX	Aktiebolaget Electrolux	42,21	BBB+		BBB	BBB+	0,16%	40,00%	Electric Housewares and Household Fan	335211	SE

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
										Manufacturing (335211)		
Consumers	BATSLN	BRITISH AMERICAN TOBACCO p.l.c.	29,97	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Tobacco and Tobacco Product Merchant Wholesalers (424940)	424940	GB
Consumers	CDBRYH	CADBURY HOLDINGS LIMITED	40,00	BBB	BAA2	BBB-	BBB	0,26%	40,00%	Chocolate and Confectionery Manufacturing from Cacao Beans (311320)	311320	GB
Consumers	CARR	CARREFOUR	35,57	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Department Stores (except Discount Department Stores) (452111)	452111	FR
Consumers	GROUPE	CASINO GUICHARD-PERRACHON	32,26	BBB-		BBB-	BBB-	0,31%	40,00%	All Other Miscellaneous Food Manufacturing (311999)	311999	FR
Consumers	CPGLN	COMPASS GROUP PLC	30,03	A-	BAA1	BBB+	A-	0,07%	40,00%	Food Service Contractors (722310)	722310	GB
Consumers	DANONE	DANONE	29,94	A-	A3		A-	0,07%	40,00%	All Other Miscellaneous Food Manufacturing (311999)	311999	FR
Consumers	DIAG	DIAGEO PLC	29,56	A-	A3	A-	A-	0,07%	40,00%	Distilleries (312140)	312140	GB
Consumers	EXPGRL-EXPFIN	EXPERIAN FINANCE PLC	26,17	A-	BAA1		A-	0,07%	40,00%	All Other Information Services (519190)	519190	GB
Consumers	AUCHAN	GROUPE AUCHAN	33,92	A			A	0,07%	40,00%	Warehouse Clubs and Supercenters (452910)	452910	FR
Consumers	HENAGK	Henkel AG & Co. KGaA	34,60	A	A2	A	A	0,07%	40,00%	Adhesive Manufacturing (325520)	325520	DE
Consumers	IMPTOB	IMPERIAL TOBACCO GROUP PLC	43,33	BBB	BAA3	BBB-	BBB	0,26%	40,00%	Cigarette Manufacturing (312221)	312221	GB
Consumers	JTI	JTI (UK) FINANCE PLC	48,88	A+	AA3	A+	A+	0,05%	40,00%	Commodity Contracts Brokerage (523140)	523140	GB
Consumers	KINGFI	KINGFISHER PLC	33,92	BBB-	BAA3	BBB-	BBB-	0,31%	40,00%	Warehouse Clubs and Supercenters (452910)	452910	GB
Consumers	AHOLD	Koninklijke Ahold N.V.	31,20	BBB	BAA3	BBB	BBB	0,26%	40,00%	Supermarkets and Other Grocery (except Convenience) Stores (445110)	445110	NL
Consumers	PHG	Koninklijke Philips Electronics N.V.	31,49	A-	A3	A	A-	0,07%	40,00%	Electric Lamp Bulb and Part Manufacturing (335110)	335110	NL
Consumers	MOET	LVMH MOET HENNESSY LOUIS VUITTON	33,79	A			A	0,07%	40,00%	Wineries (312130)	312130	FR
Consumers	MKS-M+SPlc	MARKS AND SPENCER p.l.c.	43,55	BBB-	BAA3		BBB-	0,31%	40,00%	Department Stores (except Discount Department Stores)	452111	GB

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
										(452111)		
Consumers	METFNL	METRO AG	43,68	BBB	BAA2	BBB	BBB	0,26%	40,00%	Offices of Other Holding Companies (551112)	551112	DE
Consumers	NESTLE	Nestle S.A.	47,06	AA	AA1	AA+	AA	0,01%	40,00%	All Other Miscellaneous Food Manufacturing (311999)	311999	CH
Consumers	NXT	NEXT PLC	26,58	BBB	BAA2	BBB	BBB	0,26%	40,00%	Other Clothing Stores (448190)	448190	GB
Consumers	PPR	PPR	31,13	BBB-			BBB-	0,31%	40,00%	Department Stores (except Discount Department Stores) (452111)	452111	FR
Consumers	SABLN	SABMILLER PLC	29,66	BBB+	BAA1		BBB+	0,16%	40,00%	Breweries (312120)	312120	GB
Consumers	AYLL	SAFEWAY LIMITED	45,46		A3		A-	0,07%	40,00%	Supermarkets and Other Grocery (except Convenience) Stores (445110)	445110	GB
Consumers	DEXO	SODEXO	29,51	BBB+		BBB+	BBB+	0,16%	40,00%	Full-Service Restaurants (722110)	722110	FR
Consumers	SUEDZU	Suedzucker Aktiengesellschaft Mannheim/Ochsenfurt	26,40	BBB	BAA2		BBB	0,26%	40,00%	Beet Sugar Manufacturing (311313)	311313	DE
Consumers	SCACAP	Svenska Cellulosa Aktiebolaget SCA	34,30	BBB+	BAA1		BBB+	0,16%	40,00%	Pulp Mills (322110)	322110	SE
Consumers	TATELN	TATE & LYLE PUBLIC LIMITED COMPANY	27,31	BBB	BAA3		BBB	0,26%	40,00%	All Other Miscellaneous Food Manufacturing (311999)	311999	GB
Consumers	TSCO	TESCO PLC	40,48	A-	A3	A-	A-	0,07%	40,00%	Department Stores (except Discount Department Stores) (452111)	452111	GB
Consumers	ULVR	Unilever N.V.	42,59	A+	A1	A+	A+	0,05%	40,00%	All Other Miscellaneous Food Manufacturing (311999)	311999	NL
Energy	BPLN	BP P.L.C.	41,77	A	A2		A	0,07%	40,00%	Petroleum Refineries (324110)	324110	GB
Energy	CENTRI	Centrica Plc	24,87	A-	A3	A	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	GB
Energy	EON	E.ON AG	35,69	A	A3	A	A	0,07%	40,00%	Other Electric Power Generation (221119)	221119	DE
Energy	EDF	ELECTRICITE DE FRANCE	33,82	AA-	AA3	A+	AA-	0,03%	40,00%	Electric Power Distribution (221122)	221122	FR
Energy	BAD	EnBW Energie Baden-Wuerttemberg AG	42,58	A-	A2	A-	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	DE
Energy	ENEL	ENEL S.P.A.	42,85	A-	A3	A-	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	IT

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
Energy	ENI	ENI S.P.A.	24,30	A+	A1	A+	A+	0,05%	40,00%	Petroleum Refineries (324110)	324110	IT
Energy	FORTUM	Fortum Oyj	29,55	A	A2	A-	A	0,07%	40,00%	Other Electric Power Generation (221119)	221119	FI
Energy	GASSM	GAS NATURAL SDG, S.A.	38,37	BBB	BAA2	A-	BBB	0,26%	40,00%	Natural Gas Distribution (221210)	221210	ES
Energy	GDFS	GDF SUEZ	24,49	A	A1		A	0,07%	40,00%	Natural Gas Distribution (221210)	221210	FR
Energy	IBERDU	IBERDROLA, S.A.	43,92	A-	A3	A-	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	ES
Energy	NGP	NATIONAL GRID PLC	29,83	A-	BAA1	BBB	A-	0,07%	40,00%	Electric Power Distribution (221122)	221122	GB
Energy	REP	REPSOL YPF, S.A.	47,16	BBB	BAA1	BBB+	BBB	0,26%	40,00%	Petroleum Refineries (324110)	324110	ES
Energy	RDSPLC	ROYAL DUTCH SHELL PLC	42,57		AA1	AA	AA	0,01%	40,00%	All Other Metal Ore Mining (212299)	212299	NL
Energy	RWE	RWE Aktiengesellschaft	40,34	A-	A3	A+	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	DE
Energy	STOL	Statoil ASA	46,39	AA-	AA2		AA-	0,03%	40,00%	Crude Petroleum and Natural Gas Extraction (211111)	211111	NO
Energy	TOTALN	TOTAL SA	36,87	AA-	AA1	AA	AA-	0,03%	40,00%	Petroleum Refineries (324110)	324110	FR
Energy	UU	UNITED UTILITIES PLC	26,60	BBB-	BAA1	BBB	BBB-	0,31%	40,00%	Water Supply and Irrigation Systems (221310)	221310	GB
Energy	VATFAL	Vattenfall Aktiebolag	33,68	A-	A2	A-	A-	0,07%	40,00%	Other Electric Power Generation (221119)	221119	SE
Energy	VEOLIA	VEOLIA ENVIRONNEMENT	34,45	BBB+	A3	A-	BBB+	0,16%	40,00%	Water Supply and Irrigation Systems (221310)	221310	FR
Financials	AEGON	Aegon N.V.	22,28	A-	A3	A	A-	0,07%	40,00%	Direct Life Insurance Carriers (524113)	524113	NL
Financials	ALZSE	Allianz SE	43,20	AA	AA3	AA-	AA	0,01%	40,00%	Miscellaneous Financial Investment Activities (523999)	523999	DE
Financials	ASSGEN	ASSICURAZIONI GENERALI - SOCIETA PER AZIONI	31,47	AA-	A1	A+	AA-	0,03%	40,00%	Direct Life Insurance Carriers (524113)	524113	IT
Financials	AVLN	AVIVA PLC	43,13	A	A2		A	0,07%	40,00%	Direct Health and Medical Insurance Carriers (524114)	524114	GB
Financials	AXAF	AXA	41,65	A	A2	A	A	0,07%	40,00%	All Other Insurance Related Activities (524298)	524298	FR
Financials	MONTE	BANCA MONTE DEI PASCHI DI SIENA S.P.A.	46,92	BBB+	BAA-		BBB+	0,16%	40,00%	Commercial Banking (522110)	522110	IT

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
Financials	BBVSM	BANCO BILBAO VIZCAYA ARGENTARIA, SOCIEDAD ANONIMA	50,24	A+		A+	A+	0,05%	40,00%	Commercial Banking (522110)	522110	ES
Financials	BPSC	BANCO POPOLARE SOCIETA COOPERATIVA	46,76	BBB	A2		BBB	0,26%	40,00%	Commercial Banking (522110)	522110	IT
Financials	SANTNDR	BANCO SANTANDER, S.A.	47,22	AA-	AA3	AA-	AA-	0,03%	40,00%	Commercial Banking (522110)	522110	ES
Financials	BACR-Bank	BARCLAYS BANK PLC	44,37	A+	AA3	A-	A+	0,05%	40,00%	Commercial Banking (522110)	522110	GB
Financials	BNP	BNP PARIBAS	42,81	AA-	AA2	AA-	AA-	0,03%	40,00%	Commercial Banking (522110)	522110	FR
Financials	CMZB	COMMERZBANK Aktiengesellschaft	47,25	A		A+	A	0,07%	40,00%	Commercial Banking (522110)	522110	DE
Financials	ACAFP	CREDIT AGRICOLE SA	42,81	A+	AA2	AA-	A+	0,05%	40,00%	Commercial Banking (522110)	522110	FR
Financials	CSGAG	Credit Suisse Group Ltd	50,69	A	AA2	AA-	A	0,07%	40,00%	Commercial Banking (522110)	522110	CH
Financials	DB	DEUTSCHE BANK AKTIENGESELLSCHAFT	47,25	A+	AA3	AA-	A+	0,05%	40,00%	Commercial Banking (522110)	522110	DE
Financials	HANRUE	Hannover Rueckversicherung AG	45,60	AA-		A+	AA-	0,03%	40,00%	All Other Insurance Related Activities (524298)	524298	DE
Financials	SANPAO	INTESA SANPAOLO SPA	50,28	A	A2		A	0,07%	40,00%	Commercial Banking (522110)	522110	IT
Financials	LLOYDS-Bank	LLOYDS TSB BANK plc	44,36	A	A1	A	A	0,07%	40,00%	Commercial Banking (522110)	522110	GB
Financials	MUNRE	Muenchener Rueckversicherungs-Gesellschaft Aktiengesellschaft in Muenchen	26,58	AA-	AA3	AA-	AA-	0,03%	40,00%	Reinsurance Carriers (Regulated) (524130)	524130	DE
Financials	SOCGEN	SOCIETE GENERALE	42,81	A+	AA3	A+	A+	0,05%	40,00%	Commercial Banking (522110)	522110	FR
Financials	SWREL	Swiss Reinsurance Company Ltd	48,09	AA-	A1		AA-	0,03%	40,00%	All Other Insurance Related Activities (524298)	524298	CH
Financials	RBOS-RBOSplc	THE ROYAL BANK OF SCOTLAND PUBLIC LIMITED COMPANY	44,37	A	A2		A	0,07%	40,00%	Commercial Banking (522110)	522110	GB
Financials	UBS	UBS AG	50,53	A	AA3	A	A	0,07%	40,00%	Commercial Banking (522110)	522110	CH
Financials	USPA	UNICREDIT, SOCIETA PER AZIONI	47,28	A	A2	A	A	0,07%	40,00%	Commercial Banking (522110)	522110	IT
Financials	ZINCO	Zurich Insurance Company Ltd	42,07	AA-		A	AA-	0,03%	40,00%	Miscellaneous Intermediation (523910)	523910	CH
TMT	BERTEL	Bertelsmann AG	42,03	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Periodical Publishers (511120)	511120	DE
TMT	BSY	BRITISH SKY BROADCASTING GROUP PLC	29,45	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Television Broadcasters (515120)	515120	GB

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
TMT	BRITEL-BritTel	BRITISH TELECOMMUNICATIONS public limited company	33,54	BBB	BAA2	BBB	BBB	0,26%	40,00%	Wired Telecommunications Carriers (517110)	517110	GB
TMT	DT	Deutsche Telekom AG	38,39	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Wired Telecommunications Carriers (517110)	517110	DE
TMT	FRTEL	FRANCE TELECOM	35,59	A-	A3	A-	A-	0,07%	40,00%	Wired Telecommunications Carriers (517110)	517110	FR
TMT	KPN	Koninklijke KPN N.V.	31,57	BBB+	BAA2	BBB+	BBB+	0,16%	40,00%	Telecommunications Resellers (517911)	517911	NL
TMT	PSON	PEARSON plc	43,51	BBB+	BAA1		BBB+	0,16%	40,00%	Newspaper Publishers (511110)	511110	GB
TMT	PUBFP	PUBLICIS GROUPE SA	30,43	BBB+	BAA2		BBB+	0,16%	40,00%	Advertising Agencies (541810)	541810	FR
TMT	REEDLN	REED ELSEVIER PLC	43,59			A-	A-	0,07%	40,00%	Periodical Publishers (511120)	511120	GB
TMT	STM	STMicroelectronics N.V.	22,64	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Other Electronic Component Manufacturing (334419)	334419	NL
TMT	TIIMN	TELECOM ITALIA SPA	41,31	BBB	BAA2	BBB	BBB	0,26%	40,00%	Wired Telecommunications Carriers (517110)	517110	IT
TMT	LMETEL	Telefonaktiebolaget L M Ericsson	45,86	BBB+	A3	BBB+	BBB+	0,16%	40,00%	Radio and Television Broadcasting and Wireless Communications Equipment Manufacturing (334220)	334220	SE
TMT	TELEFO	TELEFONICA, S.A.	46,98	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Wired Telecommunications Carriers (517110)	517110	ES
TMT	TKA	Telekom Austria Aktiengesellschaft	38,46	BBB	A3		BBB	0,26%	40,00%	Wired Telecommunications Carriers (517110)	517110	AT
TMT	TELNOR	TELENOR ASA	33,18	A-	A3		A-	0,07%	40,00%	Wired Telecommunications Carriers (517110)	517110	NO
TMT	TLIASS	TeliaSonera Aktiebolag	34,56	A-	A3	A-	A-	0,07%	40,00%	Wired Telecommunications Carriers (517110)	517110	SE
TMT	VIVNDI	VIVENDI	32,40	BBB	BAA2	BBB	BBB	0,26%	40,00%	Motion Picture and Video Production (512110)	512110	FR
TMT	VOD	VODAFONE GROUP PUBLIC LIMITED COMPANY	41,13	A-	A3	A-	A-	0,07%	40,00%	Wireless Telecommunications	517210	GB

Sector	Markit Ticker	Markit Long Name	R-squared	S&P Rating	Moody's Rating	Fitch Rating	Implied S&P Rating	Implied PD	Assumed LGD	Industry	Industry Code	Country
										Carriers (except Satellite) (517210)		
TMT	WOLKLU	Wolters Kluwer N.V.	27,66	BBB+	BAA1	BBB+	BBB+	0,16%	40,00%	Book Publishers (511130)	511130	NL
TMT	WPPGRP-2005	WPP 2005 LIMITED	26,01	BBB	BAA3	BBB+	BBB	0,26%	40,00%	Advertising Agencies (541810)	541810	GB