Royal Institute of Technology

Master Thesis In Financial Mathematics

Credit Risk Management in Absence of Financial and Market Data

Author:
Sepehr Yousefi
sepehry@kth.se

Supervisor:
Jimmy Olsson

May 10, 2016

Abstract

Credit risk management is a significant fragment in financial institutions' security precautions against the downside of their investments. A major quandary within the subject of credit risk is the modeling of simultaneous defaults. Globalization causes economises to be affected by innumerous external factors and companies to become interdependent, which in turn enlarges the complexity of establishing reliable mathematical models. The precarious situation is exacerbated by the fact that managers often suffer from the lack of data. The default correlations are most often calibrated by either using financial and/or market information. However, there exists circumstances where these types of data are inaccessible or unreliable. The problem of scarce data also induces difficulties in the estimation of default probabilities. The frequency of insolvencies and changes in credit ratings are usually updated on an annual basis and historical information covers 20-25 years at best. From a mathematical perspective, this is considered as a small sample and standard statistical models are inferior in such situations.

The first part of this thesis specifies the so-called entropy model which estimates the impact of macroeconomic fluctuations on the probability of defaults, and aims to outperform standard statistical models for small samples. The second part specifies the CIMDO, a framework for modeling correlated defaults without financial and market data. The last part submits a risk analysis framework for calculating the uncertainty in the simulated losses.

It is shown that the entropy model will reduce the variance of the regression coefficients but increase its bias compared to the OLS and Maximum Likelihood. Furthermore there is a significant difference between the student's t CIMDO and the t-Copula. The former appear to reduce the model uncertainty, however not to such extent that evident conclusions were carried out.

Sammanfattning

Kreditriskhantering är den enskilt viktigaste delen i bankers och finansiella instituts säkerhetsåtgärder mot nedsidor i deras investeringar. En påtaglig svårighet inom ämnet är modelleringen av simultana konkurser. Globalisering ökar antalet parametrar som påverkar samhällsekonomin, vilket i sin tur försvårar etablering av tillförlitliga matematiska modeller. Den prekära situationen förvärras av det faktum att analytiker genomgående saknar tillräcklig data. Konkurskorrelation är allt som oftast kalibrerad med hjälp av information från årsrapporter eller marknaden. Dessvärre existerar det omständigheter där sådana typer av data är otillgängliga eller otillförlitliga. Samma problematik skapar även svårigheter i skattningen av sannolikheten till konkurs. Uppgifter såsom frekvensen av insolventa företag eller förändringar i kreditbetyg uppdateras i regel årligen, och historisk data täcker i bästa fall 20-25 år. Syftet med detta examensarbete är att ge ett övergripande ramverk för kreditriskhantering i avsaknad av finansiell information och marknadsdata. Detta innefattar att estimera vilken påverkan fluktueringar i makroekonomin har på sannolikheten för konkurs, modellera korrelerade konkurser samt sammanfatta ett ramverk för beräkning av osäkerheten i den estimerade förlustdistributionen.

Den första delen av examensarbetet specificerar den såkallade entropy modellen. Denna skattar påverkan av makroekonomin på sannolikheterna för konkurs och ämnar att överträffa statistiska standardmodeller vid små datamängder. Den andra delen specificerar CIMDO, ett ramverk för beräkning av konkurskorrelation när marknads- och företagsdata saknas. Den sista delen framlägger ett ramverk för riskanalys av förlustdistributionen.

Det visas att entropy modellen reducerar variansen i regressionskoefficienter men till kostnad av att försämra dess bias. Vidare är det en signifikant skillnad mellan student's t CIMDO och t-Copula. Det förefaller som om den förstnämnda reducerar osäkerheten i beräkningarna, men inte till den grad att uppenbara slutsatser kan dras.

Acknowledgements

I would like to take this opportunity to express gratitude to my supervisor at the Royal Institute of Technology, Jimmy Olsson, for his excellent guidance and feedback. I also want to thank my family and friends for their support throughout my studies.

Contents

1 Introduction **1**
1.1 Background 1
1.2 Literature Review 2
1.2.1 Probability of Default Modeling 2
1.2.2 Modeling Correlated Defaults 3
1.3 Purpose 5

2 Preliminaries **7**
2.1 Basel II and its Risk Parameters 7
2.2 Credit Portfolio Losses 8
2.3 Categorization of Borrowers 9
2.4 Set-Up for PD modeling 10
2.5 Set-Up for Modeling Correlated Defaults 10
2.5.1 Economic Interpretation 11
2.5.2 The Vasicek Model 11
2.5.3 Additional Comments 14

3 Mathematical Background **15**
3.1 Calculus of Variations 15
3.2 Entropy Measure 16
3.3 Kullback-Liebler Divergence 16
3.4 Mahalanobis Distance 17
3.5 The Rearrangement Algorithm 18
3.6 Risk Measures 19
3.6.1 Value-at-Risk 19
3.6.2 Expected Shortfall 19

4 Econometric Modeling **20**
4.1 Ordinary Least Squares 20
4.2 Maximum Likelihood 21
4.3 The Entropy Model 21
4.3.1 Principle of Maximum Entropy 23
4.3.2 Additional Comments on Econometric Modeling 25

5 Modeling Correlated Defaults **26**
5.1 Copula Approach 26
5.1.1 Additional Comments 27
5.2 Consistent Information Multivariate Density Optimization . . 28
5.2.1 Additional Comments 30

6 Risk Analysis of the Loss Distribution **31**
6.0.1 Distribution Uncertainty 32
6.0.2 Defining Upper and Lower Bounds 33
6.0.3 Preparation for Approximation of Bounds 34
6.0.4 Bounds on Variance 35
6.0.5 Bounds on Expected Shortfall 36
6.0.6 Bounds on Value-at-Risk 36
6.0.7 Additional Comments 37

7 Results **38**
7.1 Data 38
7.2 Econometric Models 39
7.3 Comparison of Prior and Posterior in Two Dimensional Case 42
7.4 Analysis of Variance, ES and VaR Bounds 44

8 Conclusions **51**

Appendices **55**

A Nontechnical Comments **56**
A.1 Pros and Cons of Econometric Models 56
A.2 Stress-Testing 57
A.3 Extensions of the Vasicek Model 58
A.4 Short Comment on LGD and EAD Estimation 58

B Mathematical Details **59**
B.1 Maximum Likelihood for Asset Correlation 59
B.1.1 Proof of Maximum Likelihood 60
B.2 General Solution of Maximum Entropy 60
B.3 Obtaining the Posterior by Calculus of Variations 61
B.4 Application to Credit Portfolios 63

C Figures and Tables **65**
C.1 Entropy Model vs Maximum Likelihood 65
C.2 Additional Plots for Comparing Prior and Posterior Distributions 66
C.3 QQ-plots of Prior And Posterior for Different Scenarios . . . 69
C.4 VaR and ES bounds for normal CIMDO 72

Chapter 1

Introduction

Credit risk is defined as the risk of a lender incurring losses due to a credit downgrade or default of a counterparty. It is of paramount importance that these losses are calculated correctly so that banks and financial institutions can protect themselves from potential downsides in investments, hence contributing to the economic stability.

Over the past two decades, the subject of credit risk has developed rapidly from being interdisciplinary to become purely quantitative. The greatest advance occurred in 2004, when the Basel Committee on Banking Supervision (BCBS) released Basel II. Among others, the accords contain the Internal Rating Based approach, allowing banks to autonomously calculate the regulatory capital to provide buffer against the risks emerging from credit activities. For this reason the main target for any investor is to calculate the potential loss distribution of their loan portfolio, from where this buffer capital is obtained.

The execution is usually done in two separate steps, where it starts by estimating the *Probability of default* (PD), *Loss given default* (LGD) and *Exposure at default* (EAD). The subsequent step is developing a model which properly captures the *default correlation* among the risk components making up the portfolio. This is particularly important in order to correctly simulate simultaneous defaults among the counterparties, hence accurately estimate the credit losses.

1.1 Background

Despite the progress from the conventions by BCBS, the US subprime mortgages crisis was triggered in 2007, causing innumerous bankruptcies and worldwide financial paralysis whose adverse effects are still present today. In retrospect, experts have identified incorrect model assumptions regarding the dependence structure, rather than erroneous calculations as the main reason for the crisis.

The literature is far from consentaneous about what should be included in the concept of default modeling and how it should be accomplished. From a broad perspective it is obvious that common factors such as recessions or changes in government and monetary policies are affecting the likelihood of default. Furthermore, today's globalized world expands the amount of parameters impacting the default correlation of counterparties, thereby driving the complexity further and making it even more difficult to reliably calibrate mathematical models aimed for the purpose.

The precarious situation is exacerbated by the fact that risk managers often suffer from lack of data. In most published frameworks the default correlation is calibrated by either using information in financial reports or from the secondary market. However, there exists circumstances where these types of data are inaccessible or unreliable. Such scenarios could arise when publicly unlisted companies, SME's and obligors in emerging markets or developing countries are making up the portfolio. Another example is when the lender and borrower sign the loan deal through a financial intermediary.

The trivial solution in these scenarios is to make assumptions on the interaction of the counterparties. But this was precisely such unsubstantial claims that led to the escalation of the crisis. Consequently, there is a need to accomplish adequate default modeling based on the data that is actually available.

The problem of scarce data also induces difficulties in the estimation of the PD's, however in a different way. The frequency of insolvencies and changes in credit ratings are usually updated on an annual basis and historical information covers 20-25 years at best. From a mathematical perspective, this is considered as a small sample and standard statistical models are inferior in such situations.

The information considered available in this thesis will only be the credit rating of each company, the number of defaults per year in each rating class, the number of companies per year in each rating class and macroeconomic variables. Hence, the information on accounting and market data is considered inaccessible. The following section submits a literature review on PD estimation and modeling correlated defaults, partly from a broad perspective but also in the light of what has been presented in this section.

1.2 Literature Review

1.2.1 Probability of Default Modeling

Several methods for estimating the PD of an obligor have been developed over the years. These are mainly divided into two broad groups; *market-based models* and *fundamental-based models*.

Market-based models are applicable whenever there exists a secondary market of publicly traded securities connected to the obligors. One technique is the utilization of the high correlation between the CDS spread and PD's. The PD is derived from the insurance premium and the expected recovery rate.

The fundamental-based models are useful when market information is unavailable. This group is in turn divided into three categories; *econometric*, *credit-scoring* and *hybrid models*. The first category attempts to calibrate the macro economy with the default rate movements. Usually the obligors are clustered into sectoral or rating groups. Credit-scoring models use accounting data such as sales growth and liquidity ratios to calculate a score which subsequently is transformed into a PD. A famous example is EDF^{TM} developed by Moody's KMV [12]. Finally, hybrid models are as the name suggests, hybrids between the econometric and scoring models. Details on all these methods are found in the two surveys by Chan-Lau (see [10] and [11]).

In Section 1.1, the foundation of this thesis was outlined by assuming that information on accounting and public market data are inaccessible. It is therefore inevitable that almost all of the models above are omitted, leaving the econometric models as the only alternative to predict the PD's. Focusing on this particular category, the articles by Wilson ([37], [38]) present one of the first attempts of linking the PD's to the business cycle. Wilson's model is based on a distributed lag regression model where optional macroeconomic variables are inputs and an index is the response variable. The index in turn is obtained by a logistic transformation of the historical default rates of companies clustered by their sectorial belonging.

Several extensions has been made after Wilson's model. Virolainen [36] introduces univariate time-series modelling on each of the exogenous variables and connects these by correlating their error terms. Thereafter the parameters are estimated by seemingly unrelated regression. Breuer et. al. [7] go further by an ARMAX set-up, i.e. modelling the PD's with lagged time dependence along with exogenous macro variables and additional disturbance. Wong et. al. [39] develop a dynamic framework where the default rates also affect the macro variables. Nevertheless, none of these articles address the problems and consequences of small samples.

1.2.2 Modeling Correlated Defaults

Recall from Section 1.1 that default modeling is the last step in the procedure of obtaining the credit losses. This part is particularly important since simultaneous defaults must be correctly simulated, so that the losses could be accurately calculated. Opinions differ as to how modeling correlated defaults should be accomplished, most methods utilize in some way market data for this purpose.

However, the *mixture models*[1] are close at hand since frameworks within this category typically requires minimal input data. What generally applies for all mixture models is the PD of a borrower is linked through a function to a set of common macroeconomic and/or market variables. Given a realization of these factors, defaults of the counterparties are assumed to be independent [32]. Thus, if $\mathbf{\Theta} = [\theta_1, \ldots, \theta_M]$ represents the set of common factors and $F : \mathbb{R}^M \to [0,1]$ is the link function, then the PD of borrower i is

$$\Pr\Big[\text{Borr. } i \text{ defaults } |\mathbf{\Theta}\Big] = F(\mathbf{\Theta}). \tag{1.1}$$

The various frameworks based on this representation differ in the choice of the link function and common factors. The most famous model in the financial industry, CreditRisk+ by Credit Suisse, uses an exponential link function along with market sector weights [19]. Tiwari [35] develops the original CreditRisk+ version further by introducing dependence among the considered markets. The CreditRisk+ framework assumes low PD's for large portfolios, which naturally is a major drawback [13] in cases of high risk investments or if a small portfolios is considered.

Denuit et. al. [15] insert a max function in the argument of the link function to capture whether class specific or global factors mostly affect the PD. The authors claim that their method will minimize the risk of underestimating the extreme losses.

Bae and Iscoe [1] utilize the class of double mixtures where they fit a joint distribution describing the likelihood of simultaneous defaults. The correlation between any two borrowers is dynamic and dependent on the common factors. However, the framework presumes homogeneous credit portfolios and the correlation structure is calibrated with market data.

Another class of default modeling framework tries to correlate default events through the *asset processes* of the counterparties. This methodology stems from Merton's paper issued in 1974 [33]. Several modifications of the original model has been made afterwards, see for instance Jakubik [27] or Hashimoto [24]. The idea is that any borrower is unable to meet its obligations whenever its liabilities exceed the assets. The asset process methodology simplifies the correlation structure and is closely related to the mixture models, which will be shown in Section 2.5.2.

Frey and McNeil [20] go further by fitting *copulas* to the univariate assets in order to simultaneously generate outcomes. However, the copula approach does not account for the data that is actually available and is more or less an assumption on the multivariate distribution of the assets. Nevertheless, the copula approach will serve as a benchmark model for this thesis.

[1] The notation refers to mixture of normals. Also known as static reduced-form models.

1.3 Purpose

The aim of this thesis is to bypass the obstacles presented in Section 1.1 for loan portfolios exclusively. In other words, the work will focus on how to overcome the absence of financial and market data in the context of credit risk modeling. Estimation of the LGD and EAD are excluded, hence PD estimation and modeling correlated defaults are the two main topics, see Figure 1.1 below.

It is immediately announced that no real data was available during the work and therefore the overall report should be perceived as informative rather than evidential. What is considered to be known is stated in the last paragraph in Section 1.1. The focal points are

- to specify and analyze an econometric model for estimating the PD's which outperforms standard statistical techniques for small samples,
- to specify and analyze a default modeling framework which is not only based on assumptions (such as the copulas), but also takes into account the data that is actually available.

Furthermore, all models have shortcomings which bring insecurity in the estimation of the loss distribution. If this uncertainty is quantified, there will be opportunities to make judgments on which default modeling framework that performs best. Hence, the final goal of the thesis is

- to specify a risk analysis method for quantifying the uncertainty in the loss distribution. This method will be used to compare the specified default modeling framework with the copula approach.

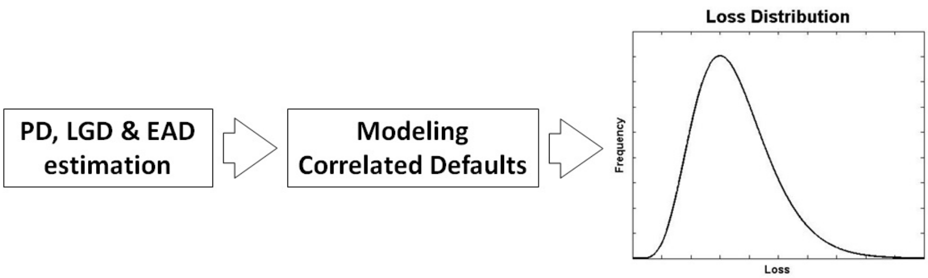


Figure 1.1: A general scheme showing the separate steps of credit risk management. The goal is to obtain the loss distribution. The PD, LGD and EAD is estimated first. The next step is to model the default dependency among the counterparties. LGD and EAD estimation is excluded in this thesis.

The rest of this paper is organized as follows. In the subsequent chapter assumptions, delimitations and fundamental parts of credit risk are stated, set-ups for PD estimation and modeling correlated defaults are presented as well. Chapter 3 highlights the mathematical details used within the thesis. Chapter 4 describes the econometric models used the for estimating the impact of macroeconomic changes on the PD's. Chapter 5 describes the selected frameworks for modeling correlated defaults. Chapter 6 provides a risk analysis method for quantifying the risks in the loss distribution. Chapter 7 reports results, discussions and analyzes. Chapter 8 concludes.

Chapter 2

Preliminaries

This chapter describes the fundamental parts of credit risk more closely and presents predetermined assumptions and delimitations. Credit losses are defined, and the set-up for PD modeling as well as modeling correlated defaults are presented.

2.1 Basel II and its Risk Parameters

As aforementioned, BCBS released Basel II in 2004, which consists of recommendations for banking supervision and risk management. The most essential part is the minimum capital requirement a financial institution must hold to protect against the risks due to business activities. Within the accords there are three risk parameters explicitly mentioned to be estimated, namely the PD, EAD and LGD. The definitions are listed below.

- As the name suggests, PD is the likelihood of a borrower being unable to meet its financial obligations over a predetermined time period (usually set to one year).
- The EAD is the gross exposure a creditor faces in case of default. EAD is divided into two parts, outstanding and commitments. The first is often treated as deterministic while the latter is calibrated, usually by the creditworthiness of the borrower. EAD is measured in currency.
- LGD is the actual loss incurred by the creditor. It is determined by the recovery rate, which in turn is affected by the type and quality of the collateral, potential insurances, additional costs due to repurchase etc. LGD is measured in percent of the EAD.

For more details on the risk parameters see Bluhm et. al. [5]. BCBS also released Basel III in 2010. However, the updated version focuses on capital allocation and is not concerning this thesis.

2.2 Credit Portfolio Losses

Financial institutions are keen to estimating the loss distribution in order to calculate capital requirements and supervise the overall risk within the business. The loss distribution describes the potential losses a lender may incur over a fixed time period due to simultaneous defaults of the counterparties.

Consider L as the random variable representing the losses. Furthermore, let i denote the i:th borrower in the portfolio and let t be a predetermined future time point. According to Huang and Oosterlee [26], the total credit loss is defined as

$$L(t) = \sum_{i=1}^{N} EAD_{it} \cdot LGD_{it} \cdot \mathbb{1}_{it}, \tag{2.1}$$

where N is the total amount of borrowers in the portfolio. $\mathbb{1}_{it}$ is the default indicator taking value 1 if the borrower i defaults up to time t, and 0 otherwise. The *Expected loss* (EL) refers to the expectation of the total losses. From business point of view, EL is interpret as the normal cost of doing credit business and is calculated as

$$\mathrm{EL}(t) = \sum_{i=1}^{N} EAD_{it} \cdot LGD_{it} \cdot p_{it}, \tag{2.2}$$

where p_{it} is the PD up to time t of borrower i. The *Unexpected losses* (UL) are larger losses occurring more occasionally. UL is defined by the BCBS as the Value-at-Risk (VaR) at level 99.9% of the loss distribution, see Section 3.6.1 for the definition of VaR. The difference between UL and EL is equal to the *Economic capital* (EC), i.e. the amount a creditor must put aside to manage financial distressed periods. Figure 2.1 gives an illustration of EL, UL and EC as fractions of the loss distribution. The shaded area reflects the losses exceeding the UL. These are losses which are excluded from the EC since they are considered too expensive to hold.

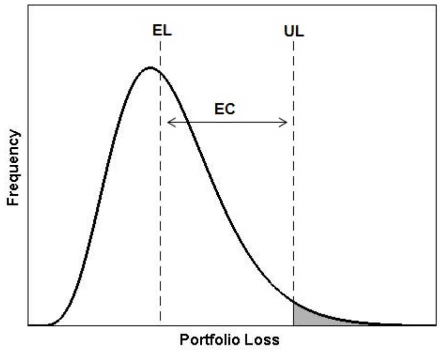


Figure 2.1: Illustration of the loss distribution along with the EL, UL and EC.

Thus, proper estimation of LGD, EAD, PD as well as default correlation is a necessity to accurately calculate the loss distribution.

The remainder of the chapter presents, for this thesis specifically, the set-up for PD modeling as well as modeling correlated defaults.[2] However, first the borrowers must be categorized.

2.3 Categorization of Borrowers

In most credit risk frameworks some sort of classification of the borrowers making up the portfolio is implemented. This varies across the literature, in most applications the borrowers are divided into groups by either their market sector, geographical location and/or rating class. Subsequently, assumptions concerning certain properties shared among all borrowers within the same group are assigned. From a mathematical point of view, the classification is crucial since it significantly reduces the model parameters and thereby making the calculations feasible.

In this thesis the borrowers will be divided into 6 rating classes. All borrowers within the same rating will have equal PD. Hence, instead of calculating the PD for each individual counterparty, the number of parameters to be estimated is reduced to six. Rating 1 contains borrowers less likely to default (lowest PD), whereas rating 6 contains the borrowers associated with the greatest risk (highest PD). Furthermore, the ratings are regarded as fixed through time, i.e. a borrower either jumps directly into default or remains in the same rating.[3] Therefore, all methods based on rating migrations are excluded.

[2] LGD and EAD estimation are excluded in this thesis. These will be set as constants. See Appendix A.4 for a short comment on these parameters.

[3] In some credit rating frameworks the companies could move from one rating to another, i.e. downgrading/upgrading in creditworthiness.

2.4 Set-Up for PD modeling

Recall from Section 1.1 that it was assumed that accounting and public market data are inaccessible, leaving the econometric models as the only alternative to predict the PD's (see the Literature review 1.2 for details). Moreover, recall the assumption that each borrower i has the same PD as all other counterparties within the same rating class. Hence, the endogenous variables in the model, i.e. the PD's corresponding to rating r at time t, are given by

$$p_{it} = p_t^r = \frac{d_t^r}{n_t^r}, \quad \text{if } i \in r, \;\; r = 1, \ldots, 6. \tag{2.3}$$

Here d_t^r is the number of defaults in rating r at time t, and n_t^r is the total number of borrowers in rating r at time t. d_t^r and n_t^r are assumed to be known, see the last paragraph in Section 1.1.

Furthermore, let $\mathbf{X}$ be a vector containing optional macroeconomic variables and consider the set of functions $g : \mathbb{R} \to [0, 1]$.[4] Then one may link the PD's and macroeconomic variables by

$$p_t^r = g(\beta^r \cdot \mathbf{X}), \qquad r = 1, \ldots, 6 \tag{2.4}$$

where β^r is the vector of regression coefficients.

The choice of econometric models are by no means uncontroversial. These models have some appealing properties but also some serious drawbacks. For detailed explanation on this issue, the interested reader is referred to Appendix A.1. Section 1.3 outlined the goals of this thesis, in which one is to specify an econometric model outperforming standard statistical methods for small samples. This model is presented in Chapter 4 and further on compared to the standard methods in Chapter 7.

2.5 Set-Up for Modeling Correlated Defaults

The idea of the default modeling is to estimate the number of counterparties that are simultaneously unable to meet their obligations. This is the final step and enables to forecast credit losses, see Figure 1.1 and Equation 2.1. The PD's have been calculated at an earlier stage and therefore the potential defaults could theoretically be determined by using Bernoulli random variables, or some sort of corresponding multivariate representation. However, such approaches are rudimentary and perhaps impossible in some aspects. For instance, if marginal Bernoulli's are selected then no account is taken on the correlation between the obligors in the portfolio. Furthermore, on a multivariate level the joint PD of two or more companies are required for

[4]The set of functions must map its arguments to the unit interval due to the dependent variable in Equation (2.4) is indeed a probability.

calibration. Loan portfolios vary heavily in size and content as loans are continuously refunded and new contracts are signed, which is why the joint PD's are almost impossible to estimate from empirical data.

Thus, there is a demand to enforce a model facilitating the correlation among the borrowers. Only then opportunities arise to develop sophisticated methods for modeling correlated defaults. To completely grasp the selected framework, the so-called *Vasicek model*, the underlying economic interpretation must first be declared.

2.5.1 Economic Interpretation

Intuitively, all companies have assets emerging through various business activities. Likewise, just as indisputable is that companies have liabilities such as provisions and loans. If the liabilities exceed the assets, the company becomes insolvent and hence incapable to meet its obligations. This applies irrespective of whether the information regarding assets and liabilities are available. Therefore, this conception will be accepted in the further work.

It is of great importance to emphasize that this does not, by any means, implies the actual asset process is estimated. Instead the asset return should be viewed as a latent variable, or a parametric assumption describing an unknown but real event.

2.5.2 The Vasicek Model

To visualize the Vasicek model, contemplate Figure 2.2 below. The black curve represents the asset process between two time points of borrower i belonging to a rating class with PD=3%. One could simulate random numbers from a normal distribution[5] to represent the asset returns at the end of the time period and compare these to the threshold value $\Phi^{-1}(0.03) \approx -1.88$, pictured as the blue line in Figure 2.2. If a specific sample is below the quantile, then borrower i is considered defaulted, whereas if the sample is above then the same borrower is considered solvent. Certainly 3% of all the outcomes will fall into the default zone, pictured as the blue shaded area, if sufficient amount of samples are generated.

Now consider PD=10% with the corresponding threshold value -1.28. The increase of simulated defaults are visualised by an expanding of the default zone where the red shaded area now is included.

[5] There is no explicit reason for normal distribution other than its simplicity and manageability. Student's t distributions are also used, see Appendix A.3 for details.

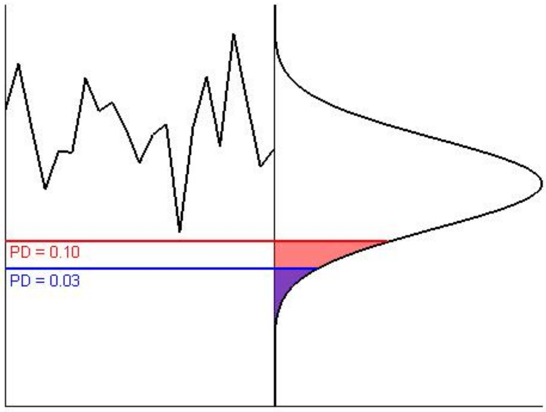


Figure 2.2: The black curve shows the asset process between two time points. If the PD=3% then the default zone constitutes of the blue shaded area. If the PD increases to 10%, the default zone is expanded to also include the red shaded area as well.

The actual values from the normal distribution representing the asset returns are not of interest, but rather whether these falls above or below the quantile. Thus, the entire concept is in fact a two state model such as Bernoulli random variables.

Until now no account has been taken on the correlation between borrowers. This is where the Vasicek model becomes useful. According to Hashimoto [24], the asset return of obligor i at time t is

$$A_{it} = \sqrt{\rho^r} S_t + \sqrt{1 - \rho^r} U_{it} \quad i \in r, \tag{2.5}$$

where S_t represents the common *systematic risk*[6] while U_{it} is the *idiosyncratic risk*[7], both assumed to be standard normally distributed and independent of each other. Here ρ^r denotes the asset correlation of rating r and explains to what extent the asset return is affected by the risk factors respectively. A_{it} will also be standard normal because the asset correlation is defined on the unit interval.

The asset correlation could be estimated by using for instance Maximum Likelihood, see Appendix B.1 for details. However, the estimation of the asset correlation is not the main focus in this thesis and the absence of real data makes it irrelevant to calculate them. Instead, the asset correlations

[6] *Systematic risk* is also known as market risk, affects all borrowers and cannot be prevented by diversification of the held portfolio. Recessions and changes in monetary policies are examples of systematic risk.

[7] *Idiosyncratic risk* is commonly known as the risk connected to a single borrower and it is prevented by diversification.

will further on be varied to examine different scenarios. The logic behind estimating the asset correlation by rating class, and not for instance by market sector, is also discussed in Appendix B.1.

The covariance between two asset returns of the borrowers (i, j) belonging to the rating classes (r, r') is given by

$$\begin{aligned}\sigma_{i,j} &= \mathrm{Cov}[A_i, A_j] = \mathrm{E}[A_i \cdot A_j] - \mathrm{E}[A_i] \cdot \mathrm{E}[A_j] \\ &= \mathrm{E}\Big[\big(\sqrt{\rho^r} S_t + \sqrt{1-\rho^r} U_{it}\big)\big(\sqrt{\rho^{r'}} S_t + \sqrt{1-\rho^{r'}} U_{jt}\big)\Big] = \sqrt{\rho^r \rho^{r'}}.\end{aligned} \tag{2.6}$$

Since all asset returns have variance equal to 1 the correlation between two asset returns is equal to the covariance. Equation (2.6) shows how the Vasicek model simplifies the establishment of correlation structure. Instead of finding the correlation of each individual pair of counterparties, there is a reduction in parameters which is significantly lower.

Moreover, recall that all borrowers within same rating have the same PD. Let γ_{it} denote the liabilities at time t for borrower i belonging to rating r, then

$$\begin{aligned}&p_{it} = \{\text{Eq. (2.3)}\} = p_t^r = \Pr(A_{it} < \gamma_{it}) = \Phi(\gamma_{it}) \Rightarrow \\ &p_t^r = \Phi(\gamma_t^r),\end{aligned} \tag{2.7}$$

where $\Phi(\cdot)$ is the *cumulative distribution function* (CDF) of a standard normal variable. Equation (2.7) shows that the liabilities will be equal for all borrowers belonging to the same rating class[8] and is in fact equal to the quantile pictured in Figure 2.2. For this reason, henceforth the liabilities will be referred to as the threshold value, and denoted γ_t^r instead. Comparing Equations (2.7) and (2.4) reveals that the threshold value is through the PD indirectly affected by the macroeconomic state. Thus, if an adverse macroeconomic shock causes the PD of any rating class to rise, it will in the Vasicek model be equivalent to an increase of the threshold value and an expansion of the default zone, see Figure 2.2.

Lastly, the *conditional* probability of default is defined. Conditioned on the realization $S_t = s$, the conditional PD for obligor i in rating r is

$$\begin{aligned}&p_{it}(s) = p_t^r(s) = \Pr[A_{it} < \gamma_t^r \mid S = s] = \Pr[s\sqrt{\rho^r} + U_{it}\sqrt{1-\rho^r} < \gamma_t^r] \\ &p_t^r(s) = \Phi\left(\frac{\Phi^{-1}(p_t^r) - s\sqrt{\rho^r}}{\sqrt{1-\rho^r}}\right).\end{aligned} \tag{2.8}$$

If the systematic risk is realized, then the only remaining risk is U_{it} and therefore all obligors will be independent of each other. By comparing Equations (2.8) and (1.1) it becomes apparent that the Vasicek model has a mixture representation. Here, the gaussian CDF is the link function and

[8] Obviously is this not true in reality. However, since the parameters in the Vasicek model are latent factors it is regarded as accepted.

the common factors are S_t and also the macroeconomic variables through $\Phi^{-1}(p_t^r)$ (see Equation (2.3)). Hence, the Vasicek model and other mixture models are closely related, the difference is roughly the distributional assumption.

2.5.3 Additional Comments

For various reasons the literature concerning credit risk has modified the Vasicek model. For more details on this particular subject, the interested reader is referred to Appendix A.3.

Given the information presented until now, it is feasible to simulate the losses by generating joint asset movements using Equation (2.5), or by calculating the conditional PD from the mixture representation. However these are poor approaches in practice. Long before the aftermath of the crisis Frey and McNeil [20] emphasized that asset correlation certainly affects the default correlation, nevertheless, these terms should not be equated. Furthermore, mixture models have been heavily criticized for their tenuous assumption regarding the default dependency merely stems from the dependence of individual PD's on the set of common variables.

Therefore, to simulate the losses, the idea is to calibrate a multivariate distribution for the univariate Vasicek asset returns described in previous section. The Vasicek model provides a manageable foundation to estimate a correlation matrix for this multivariate distribution. From the distribution, the assets of the borrowers could be *simultaneously* generated and thence compared to their corresponding threshold value. These threshold values are in turn obtained by first estimating the relationship between the PD's and the macro economy by econometric modeling. Thereafter, new PD's could be predicted and inserted into Equation (2.7) to obtain the threshold values for each rating class respectively.

Chapter 5 presents the methods for fitting a multivariate distribution to the Vasicek asset returns, and Chapter 4 presents the methods for PD estimation. Chapter 6 provides a method for quantifying the uncertainty in the generated loss distribution. Therefore, the next chapter is dedicated for explaining the mathematical details behind these methods.

Chapter 3

Mathematical Background

This chapter outlines the fundamental mathematical parts used in this thesis. The objective is to provide the reader with a deeper understanding of the formulas and derivations within the models in the subsequent chapters. Readers who are familiar with these concepts could skip to the next Chapter.

3.1 Calculus of Variations

The field of *calculus of variations* is aimed to find a minimum or maximum of any functional, i.e. integrals containing functions with corresponding derivatives and arguments. The optimization is performed by applying the *Euler-Lagrange equations*, where the solution is expressed as an extremal function. The theorem of Euler-Lagrange is given below.

Theorem 3.1. *Let* $\mathbf{x} = [x_1, x_2, \ldots, x_m]$ *be a vector of variables. Consider the set of functions* $f_1, \ldots, f_n$ *with corresponding derivatives* $f'_{j,i} = \partial f_j / \partial x_i$*. Furthermore, let* $H(\cdot)$ *and* $G(\cdot)$ *be any functional on some sample space* Ω*. Then the integral*

$$\int_\Omega H\big(f_1, .., f_n, f'_{1,1}, .., f'_{n,m}, \mathbf{x}\big) d^m \mathbf{x}$$

subject to the constraints

$$\lambda_k \int_\Omega G_k(f_1, .., f_n, f'_{1,1}, .., f'_{n,m}, \mathbf{x}) d^m \mathbf{x} \quad k = 0, \ldots, K < \infty$$

attains a minimum if and only if the following condition holds

$$\frac{\partial H}{\partial f_j} - \sum_{i=1}^{m} \frac{\partial}{\partial x_i} \frac{\partial H}{\partial f'_{j,i}} + \sum_{k=0}^{K} \left[\lambda_k \left(\frac{\partial G_k}{\partial f_j} - \frac{\partial}{\partial x_i} \frac{\partial G_k}{\partial f'_{j,i}} \right) \right] = 0,$$

for $j = 1, \ldots, n$*. The system of equations is commonly known as the Euler-Lagrange equations.*

The Euler-Lagrange equations stated in Theorem 3.1 is an extended version of the original definition. First, the theorem allows several functions to be included in the functional. Second, there exists a finite set of constraints which is not the case in the initial formulation. For more information, the interested reader is referred to [8].

3.2 Entropy Measure

Entropy is a measure of the unpredictability in a random variable or model. Higher randomness is equivalent to higher entropy. A simple example is a coin toss. If the coin has two heads, the randomness is zero and consequently the entropy is at its minimum. Whereas if the coin has both head and tail with the equal probability, it is impossible to predict the next toss and hence the entropy is maximized. Depending on usage, the definition of entropy is different. However this thesis will work with the following notation

$$H(X) = -\sum_i \Pr(x_i) \ln \big[\Pr(x_i) \big], \tag{3.1}$$

where X is the random variable, the x_i's are the possible outcomes and $\Pr(x_i)$ is the probability of being in state x_i.

3.3 Kullback-Liebler Divergence

Suppose the distribution Q is used for approximating another distribution P. The *Kullback-Leibler* (KL) divergence[9] measures the information lost due to the approximation. Another interpretation of KL-divergence is that it measures the distance between Q and P. By definition [6] the KL-divergence for one-dimensional continuous distributions is formulated as

$$\mathcal{D}(p \mid q) = \int_{-\infty}^{\infty} p(x) \ln \left[\frac{p(x)}{q(x)} \right] dx, \tag{3.2}$$

where p and q are the density functions of P and Q respectively. The KL-divergence is demonstrated by the following example. Consider a standard normal distribution and a logistic distribution with location parameter 2 and scale parameter 0.7. The left plot in Figure 3.1 shows the density functions. In the right plot the integrand function in Equation (3.2) is displayed for two cases. The solid red function is the situation where the standard normal is used for approximating the logistic distribution. The dashed blue function is the opposite case, the logistic is used for approximating the standard normal. The red and blue areas under the curves are equal to the KL-divergence, for each case respectively.

[9] Also known as the *cross entropy distance* or *relative entropy*.

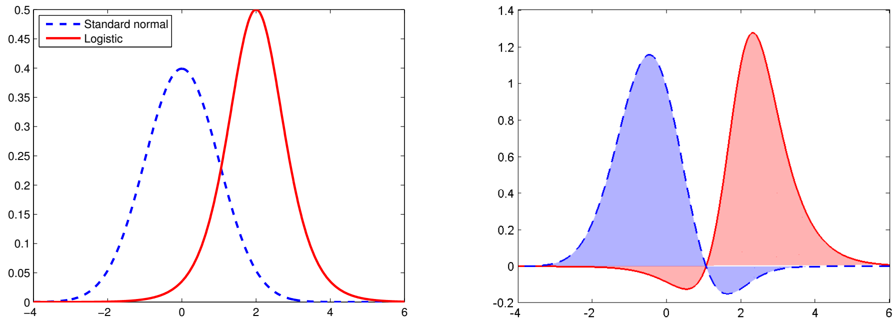


Figure 3.1: *Left*: The density functions of the standard normal and the logistic distributions. *Right*: Demonstration of the Kullback-Leibler divergence. The red solid curve corresponds to the case where the logistic distribution is approximated by the standard normal. The dashed blue curve is the opposite scenario, the standard normal is approximated by the logistic.

From the two curves in the right plot in Figure 3.1 it is obvious that the KL-divergence is non-symmetric. More interestingly, the areas under these curves are not equal. When the logistic is approximated, the area (i.e. the KL-divergence) is 2.02, whereas in the reversed case the area is 2.06. In fact, generally $\mathcal{D}(p \mid q) \neq \mathcal{D}(q \mid p)$. Consequently the KL-divergence does not fulfill the criterion of being a distance in the formal sense.

3.4 Mahalanobis Distance

Suppose $\mathbf{x} = (x_1, \ldots, x_N)$ is an arbitrary point on the real coordinate space $\mathbb{R}^N$ generated by the multivariate distribution $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, with covariance matrix $\mathbf{\Sigma}$ and possible location parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)$. The *Mahalanobis Distance* (MD) is the length, measured in standard deviations, between the point $\mathbf{x}$ and the average of $\mathbf{X}$. It is defined as

$$\mathcal{M}(\mathbf{x}) = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}. \tag{3.3}$$

If $\mathbf{X}$ is multivariate normal, then $\mathcal{M} \in \chi^2(N)$. Similarly, if $\mathbf{X}$ is student's t-distributed with ν degrees of freedom, then $\frac{\nu}{N(\nu-2)}\mathcal{M} \in F(N, \nu)$. MD also works ex post, where in such case $\mathbf{x}$ is an outcome and $\mathbf{X}$ is a fitted or predetermined multivariate distribution.

For illustration, let (X, Y) have a bivariate normal distribution with covariance equal to 0.5 and mean $\boldsymbol{\mu} = [2, 2]^T$. The MD for the points $(x_1, y_1) = (7, 5)$, $(x_2, y_2) = (-1, 3)$, $(x_3, y_3) = (-1, 1)$ and $(x_4, y_4) = (2, 2)$ are shown in Figure 3.2 together with 10'000 other samples from the distribution.

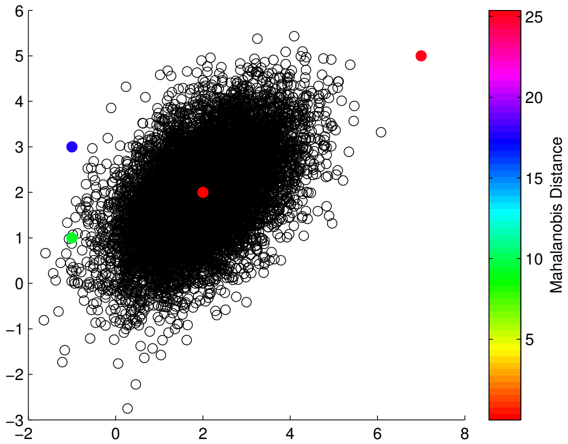


Figure 3.2: Samples from a bivariate normal distribution with mean value 2 and covariance 0.5. The plot also shows the Mahalanobis distance for the points $(7, 5), (-1, 3), (-1, 1), (2, 2)$.

3.5 The Rearrangement Algorithm

The Rearrangement Algorithm (RA) made by Embrechts et al. [17] is purely a computational tool for minimizing the variance of the row sums in a matrix. The algorithm is given below.

Algorithm 3: The Rearrangement Algorithm	
1	Given a $R \times K$ dimensional matrix M, exclude column k.
2	Calculate the row sums of the remaining columns.
3	Sort each element in column k in inversely descending order from the row sums.
4	Repeat (1)-(3) for all columns $1 \leqslant k \leqslant K$.

To get an intuition of how the RA works, consider the 3×3 matrix given below. From the beginning the row sums are 16, 10 and 7. After applying the RA the row sums become 12, 11 and 10. Hence the variance has decreased.

$$
\begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{pmatrix} 1 & 7 & 8 \\ 0 & 2 & 8 \\ 2 & 4 & 1 \end{pmatrix} \end{array}
\begin{array}{c} c_2 + c_3 \\ \begin{pmatrix} 15 \\ 10 \\ 5 \end{pmatrix} \end{array}
\longrightarrow
\begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{pmatrix} 0 & 7 & 8 \\ 1 & 2 & 8 \\ 2 & 4 & 1 \end{pmatrix} \end{array}
\begin{array}{c} c_1 + c_3 \\ \begin{pmatrix} 8 \\ 9 \\ 3 \end{pmatrix} \end{array}
$$

$$
\longrightarrow
\begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{pmatrix} 0 & 4 & 8 \\ 1 & 2 & 8 \\ 2 & 7 & 1 \end{pmatrix} \end{array}
\begin{array}{c} c_1 + c_2 \\ \begin{pmatrix} 4 \\ 3 \\ 9 \end{pmatrix} \end{array}
\longrightarrow
\begin{array}{c} \begin{array}{ccc} c_1 & c_2 & c_3 \end{array} \\ \begin{pmatrix} 0 & 4 & 8 \\ 1 & 2 & 8 \\ 2 & 7 & 1 \end{pmatrix} \end{array}
$$

3.6 Risk Measures

According to Lindskog et al. [31], by using different risk measures a manager is capable of comparing different portfolios, make decisions regarding position changes, and determining the amount of buffer capital (EC, see Figure 2.1) that should be hold to prevent insolvency in case of financial distress.

3.6.1 Value-at-Risk

Value-at-Risk (VaR) is the most used risk measure in the financial industry. If L represents the random loss variable and $F_L(\ell)$ is the corresponding CDF, then the VaR at confidence level $p \in (0, 1)$ is defined as

$$\mathrm{VaR}_p(L) = F_L^{-1}(p) = \inf\{\ell \in \mathbb{R} \mid F_L(\ell) \geqslant p\}. \tag{3.4}$$

The interpretation is as follows. Suppose $p = 0.99$, then VaR describes the value where greater losses will occur only 1% of the time.

VaR has endured a lot of criticism because of its drawbacks. For instance it is not sub-additive, meaning that the merging of two portfolios might have larger VaR than the sum of the VaR of the same portfolios separately. Consequently VaR ignores the effect of diversification. Furthermore, VaR neglects the losses exceeding above p, which could be catastrophic if these losses are extreme. To compensate for these disadvantages, the Expected Shortfall is introduced.

3.6.2 Expected Shortfall

The Expected Shortfall (ES) is defined as

$$ES_p(L) = \mathrm{E}[L \mid L \geqslant VaR_p(L)] = \frac{1}{1-p}\int_p^1 VaR_u(L)du. \tag{3.5}$$

ES is interpreted as the average of the losses greater than the loss obtained by VaR at level p. ES is sub-additive and consequently it satisfactorily compensates VaR as a risk measure. Similarly, the Left Tail Value-at-Risk (LTVaR) is the average of the losses lower than the loss obtained by VaR at level p, i.e.

$$LTVaR_p(L) = \mathrm{E}[L \mid L < VaR_p(L)] = \frac{1}{p}\int_0^p VaR_u(L)du. \tag{3.6}$$

In this thesis, LTVaR will solely be used as a tool for computing bounds of VaR, see Section 6.0.6.

Chapter 4

Econometric Modeling

This chapter outlines the econometric model used to capture the impact of macroeconomic fluctuations on the PD's. The first two sections describe standard techniques for estimating regression coefficients. These will later on be compared to the *entropy model*.

To contemplate how the endogeneous variables are obtained, see Section 2.4. Also recall from Section 2.3 that all counterparties within the same rating have equal PD. Therefore, all the econometric models are performed on each rating class separately.

4.1 Ordinary Least Squares

Let $t = (0, \ldots, T)$ be the time vector and let $\boldsymbol{\gamma}^r = [\gamma_T^r, \ldots, \gamma_0^r]$ denote the vector of dependent variables. Here $\gamma_t^r = \Phi^{-1}(p_t^r)$, where γ_t^r is the threshold value, $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal[10] and p_t^r is the PD of rating class r at time t, see Equations (2.3) and (2.7). Furthermore let $\mathbf{X}$ be a $T \times K$ matrix containing K optional explanatory macroeconomic variables, $\boldsymbol{\beta}_{\text{OLS}}^r = [\beta_1^r, \ldots, \beta_K^r]$ denotes the vector of coefficients to be determined and $\mathbf{e}^r = [e_T^r, \ldots, e_0^r]$ is the vector of error terms. The linear function

$$\boldsymbol{\gamma}^r = \mathbf{X}\boldsymbol{\beta}_{\text{OLS}}^r + \mathbf{e}^r, \qquad r = 1, \ldots, 6 \tag{4.1}$$

has the solution

$$\widehat{\boldsymbol{\beta}}_{\text{OLS}}^r = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\gamma}^r. \tag{4.2}$$

OLS has the advantages of being comprehensible and easily manageable. Unfortunately the method has several disadvantages. In situations where the sample size is small the regression coefficients will be sensitive to small

[10] This is true under the assumptions of assets being standard normal. If another distribution is selected, then $\Phi(\cdot)$ is substituted to the corresponding CDF.

changes in the data set and both have large standard errors as well as mean square errors (MSE).

4.2 Maximum Likelihood

If the link function in Equation (2.3) is the CDF of a standard normal variable, then the equation is commonly known as the *probit model*. Typically the maximum likelihood technique is the used to obtain the coefficients in such situations. Let $\boldsymbol{\gamma}^r$ and $\mathbf{e}^r$ be defined as in the previous section. Furthermore, define $\mathbf{x}_t$ as the vector containing the outcomes of the selected macroeconomic variables at time t. If the elements in $\mathbf{e}^r$ are assumed be independent and follow a standard normal distribution, then the Maximum Likelihood (ML) function is formulated as

$$\begin{aligned} \mathcal{K}(\mathbf{x}_t, \gamma_t^r) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(\gamma_t^r - \mathbf{x}_t \boldsymbol{\beta}_{\mathrm{ML}}^r)^2}{2} \right] \\ \mathcal{ML} &= \prod_{t=0}^{T} \mathcal{K}(\mathbf{x}_t, \gamma_t^r), \quad r = 1, \dots, 6 \end{aligned} \tag{4.3}$$

where the coefficients $\boldsymbol{\beta}_{\mathrm{ML}}^r$ are obtained by maximizing $\mathcal{ML}$ for each rating class r respectively. The ML procedure is easily implemented in statistical programmes and has similar advantages and drawbacks as the OLS.

4.3 The Entropy Model

A more sophisticated model originally developed by Golan et. al. [22] aims to minimize the shortcomings of regular regression models in situations where the sample size is small. This section explains the method and the theory behind it.

The essential equation of the entropy model is still the linear function in Equation (4.1). The main difference is that each coefficient $\beta_1^r, \dots, \beta_K^r$ are treated as random variables instead of constants. To obtain the distribution of each regression coefficient it is suitable to perform bootstrap. Fox [18] describes the procedure as Algorithm 4 below.

Algorithm 4: The Bootstraping Procedure	
1	Calculate the error $e_i = Y_i - \widehat{Y}_i$, where $i = 1, \ldots, n$. $\widehat{Y}_i$ are estimates from OLS while Y_i are the original samples.
2	Produce a new vector $\mathbf{e}^* = [e_1^*, \ldots, e_n^*]$ by drawing with replacement from $\mathbf{e} = [e_1, \ldots, e_n]$.
3	Compute a new sample vector, $Y_i^* = \widehat{Y}_i + e_i^*$.
4	A new coefficient vector $\boldsymbol{\beta}^*$ is estimated by the OLS, $\boldsymbol{\beta}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^*$.
5	Store $\boldsymbol{\beta}^*$ and repeat (2)-(4) N times. $\mathbf{X}$ is fixed through all iterations.

The question now is how to find the optimal β_k^r's from the bootstrapped distributions which will result in the smallest MSE. The most trivial way is to set β_k^r to be equal to the mean of its corresponding distribution. However, since the bootstrap procedure is based on the OLS it will not be expected that such simple solution would make any improvement over the OLS.

Bootstrap on smaller sample size is likely to cause larger variance of the generated distribution and/or making it skewed. To capture this diffusion, an alternative way is to select $U > 2$ outcomes from the bootstrapped distribution, and find the optimal coefficient by weighting these outcomes. This is the approach of the entropy model.

Bootstrap is not performed for the error terms. One can instead use the standard deviation of the dependent variable, $\sigma_{\boldsymbol{\gamma}^r}$, and use it as outcomes of the errors.[11] For computational purposes let $\mathbf{Z}^r$ and $\mathbf{V}^r$ be matrices containing the selected outcomes, namely

$$\mathbf{Z}^r = \begin{bmatrix} z_{11}^r & z_{12}^r & \cdots & z_{1U}^r \\ \vdots & \vdots & \ddots & \vdots \\ z_{k1}^r & z_{k2}^r & \cdots & z_{kU}^r \\ \vdots & \vdots & \ddots & \vdots \\ z_{K1}^r & z_{K2}^r & \cdots & z_{KU}^r \end{bmatrix}, \quad \mathbf{V}^r = \begin{bmatrix} v_{11}^r & v_{12}^r & \cdots & v_{1J}^r \\ \vdots & \vdots & \ddots & \vdots \\ v_{t1}^r & v_{t2}^r & \cdots & v_{tJ}^r \\ \vdots & \vdots & \ddots & \vdots \\ v_{T1}^r & v_{T2}^r & \cdots & v_{TU}^r \end{bmatrix} \tag{4.4}$$

For the regression coefficients, if U is set to 3, then the outcomes could simply be chosen to be the mean and one standard deviation from the mean, i.e. $z_{k1}^r = -\sigma$, $z_{k2}^r = \mu$ and $z_{k3}^r = \sigma$. See Figure 4.1 for illustration. For the error terms, if J is set to 2 then the outcome could for instance be $v_{t1}^r = -2\sigma_{\boldsymbol{\gamma}^r}$ and $v_{t2}^r = 2\sigma_{\boldsymbol{\gamma}^r}$.

[11] The statement is founded on the following basic theory. If the explanatory variables and the error terms are independent then $\mathrm{Var}[Y] = \mathrm{Var}[X\beta + e] = \mathrm{Var}[X\beta] + \mathrm{Var}[e]$. In addition, if X is fixed then $\mathrm{Var}[X\beta] = 0$ is indeed true.

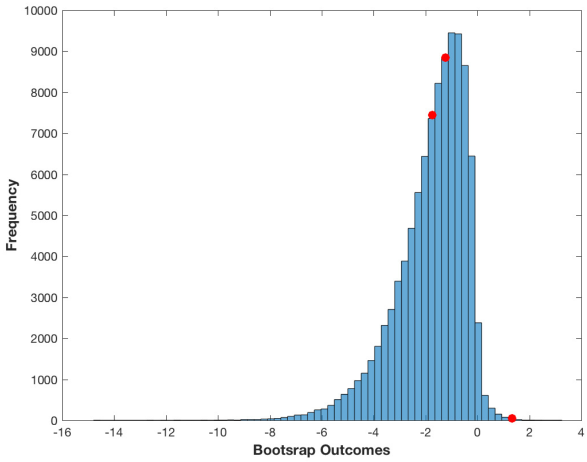


Figure 4.1: Example of a coefficient distribution generated by the bootstrap algorithm. The mean value and one standard deviations are also plotted. These outcomes is obtained from the distribution could be used in the entropy model.

Moreover, Golan et. al. [22] defines weights $q_{ku}^r \in [0,1]$ and $w_{tj}^r \in [0,1]$ such that each β_k^r and e_t^r are expressed as a linear combination like

$$\beta_k^r = \sum_{u=1}^{U} z_{ku}^r q_{ku}^r \quad e_t^r = \sum_{j=1}^{J} v_{tj}^r w_{tj}^r. \tag{4.5}$$

q_{ku}^r and w_{tj}^r could be viewed as the probabilities of being in state z_{ku}^r and v_{tj}^r respectively, although it is formally incorrect. In any case, the reformulation in Equation (4.5) of the model parameters causes the linear regression function in Equation (4.1) to be rewritten as

$$\boldsymbol{\gamma}^r = \mathbf{X}\mathbf{Z}^r\mathbf{q}^r + \mathbf{V}^r\mathbf{w}^r, \quad r = 1, \ldots, 6. \tag{4.6}$$

The only unknown parameters remaining are the weights. The principle of maximum entropy will be used to obtain them.

4.3.1 Principle of Maximum Entropy

To obtain $\mathbf{q}^r$ and $\mathbf{w}^r$ the idea is to maximize their entropy. At first sight this may seem confusing since this infer to maximize the uncertainty of the weights. However, as Penfield [28] points out, the principle of maximum entropy is to avoid redundant assumptions and restrictions on the distribution of the considered parameters, consequently removing additional bias. Thus, in resemblance to Section 3.2 the weights are calculated maximizing the following equation

$$F[\mathbf{q}^r, \mathbf{w}^r] = -\Big[\sum_{k=1}^{K}\sum_{u=1}^{U} q_{ku}^r \ln[q_{ku}^r]\Big] - \Big[\sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj}^r \ln[w_{tj}^r]\Big]. \tag{4.7}$$

Some limitations must be taken into consideration. The most obvious is that the linear function in Equation (4.6) must be satisfied. Moreover, $\mathbf{q}^r$ and $\mathbf{w}^r$ are viewed as probabilities in the context of entropy measure and therefore they must sum to 1. To summarize, the constraints of F are

$$\begin{aligned} &\gamma_t^r = \sum_{k=1}^{K}\sum_{u=1}^{U} x_{tk} z_{ku}^r q_{ku}^r + \sum_{j=1}^{J} v_{tj}^r w_{tj}^r, \quad t = 0, \ldots, T \\ &\sum_{u=1}^{U} q_{ku}^r = 1, \quad \sum_{j=1}^{J} w_{tj}^r = 1. \end{aligned} \tag{4.8}$$

With everything stated, the Lagrangian $\mathcal{L}$ to be maximized is

$$\begin{aligned} \mathcal{L}(\mathbf{q}^r, \mathbf{w}^r; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\psi}) = &- \left[\sum_{k=1}^{K}\sum_{u=1}^{U} q_{ku}^r \ln[q_{ku}^r]\right] - \left[\sum_{t=0}^{T}\sum_{j=1}^{J} w_{tj}^r \ln[w_{tj}^r]\right] \\ &+ \sum_{t=0}^{T} \lambda_t \left[\gamma_t^r - \sum_{k=1}^{K}\sum_{u=1}^{U} x_{tk} z_{ku}^r q_{ku}^r + \sum_{j=1}^{J} v_{tj}^r w_{tj}^r\right] \\ &+ \sum_{k=1}^{K} \eta_k \left[1 - \sum_{u=1}^{U} q_{ku}^r\right] + \sum_{t=0}^{T} \psi_t \left[1 - \sum_{j=1}^{J} w_{tj}^r\right], \end{aligned} \tag{4.9}$$

where $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_T)$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)$ and $\boldsymbol{\psi} = (\psi_0, \ldots, \psi_T)$ are vectors of Lagrangian multipliers. The solution of the weights are (see proof in Appendix B.2)

$$\widehat{q}_{ku}^r(\widehat{\boldsymbol{\lambda}}) = \frac{\exp\big[-\sum_{t=0}^{T} \widehat{\lambda}_t x_{tk} z_{ku}^r\big]}{\Xi(\widehat{\boldsymbol{\lambda}})} \qquad \widehat{w}_{tj}^r(\widehat{\boldsymbol{\lambda}}) = \frac{\exp\big[-\sum_{t=0}^{T} \widehat{\lambda}_t v_{tj}^r\big]}{\Psi(\widehat{\boldsymbol{\lambda}})}, \tag{4.10}$$

where

$$\begin{aligned} \Xi(\widehat{\boldsymbol{\lambda}}) &= \sum_{u=1}^{U} \left[\exp\big[-\sum_{t=0}^{T} \widehat{\lambda}_t x_{tk} z_{ku}^r\big]\right] \\ \Psi(\widehat{\boldsymbol{\lambda}}) &= \sum_{j=1}^{J} \left[\exp\big[-\sum_{t=0}^{T} \widehat{\lambda}_t v_{tj}^r\big]\right]. \end{aligned} \tag{4.11}$$

The weights are obtained by first finding the vector $\boldsymbol{\lambda}$ which maximizes $\mathcal{L}$, and subsequently insert these multipliers into the solution stated in Equation (4.10). However, Equation (4.9) is somewhat complicated to compute numerically. For this reason, Golan [30] rewrites $\mathcal{L}$ into a *dual concentrated formulation*. The "new" function, denoted as $C(\boldsymbol{\lambda})$, is

$$C(\boldsymbol{\lambda}) = \sum_{t=0}^{T} \Big[\lambda_t \gamma_t^r + \ln\big[\Psi(\boldsymbol{\lambda})\big]\Big] + \sum_{k=1}^{K} \ln\big[\Xi(\boldsymbol{\lambda})\big]. \tag{4.12}$$

Minimizing Equation (4.12) is equivalent to maximizing Equation (4.7). The concentrated version has two main advantages. First, $C(\boldsymbol{\lambda})$ has closed-form expressions of its first and second derivatives. Secondly, the function is strictly convex, i.e. C is increasing in $\boldsymbol{\lambda}$ which results in an unique global solution. A minimum of Equation (4.12) is obtained by numerical methods, e.g. Newton-Raphson.[12]

4.3.2 Additional Comments on Econometric Modeling

The econometric models are flexible in the sense that they all permit lag on any explanatory variable. Furthermore, it is feasible to extend the entropy model and the OLS to also have time dependence on the dependent variable. The latter is commonly known as the Autoregressive Distributed Lag (ADL) model.

The OLS, ML and the entropy model are compared during a simulation exercise to study which of them performs best (giving the smallest MSE) for small samples. The results are presented and discussed in Section 7.2.

[12]Although a function has a unique solution, Newton's method does not guarantee convergence to this point. This happens when the initial value is poor. However, to choose a good starting guess could be impossible, specially in cases where the function depends on several variables. Instead, one could iterate a starting guess using a different numerical method with guaranteed convergence, for instance the Gradient descent method.

Chapter 5

Modeling Correlated Defaults

This chapter presents two techniques for fitting multivariate distributions to the univariate Vasicek asset returns described in Section 2.5.2. The copula approach is regarded as a benchmark model while the main model is the CIMDO.

5.1 Copula Approach

According to Lindskog et al. [31], a common problem for risk managers is when a random vector $\mathbf{X} = (X_1, \dots, X_N)$ has well known marginal distributions but whose multivariate representation is merely partially or not understood. A useful solution in such situations is the construction of copulas. Copulas are multivariate probability distributions substantiated on two properties, namely the probability integral transform and the quantile transform. The probability integral transform says that if X is a random variable with continuous distribution function $F_X(x)$, then $F_X(X)$ is standard uniformly distributed. The quantile transform says that if U is standard uniformly distributed and if G is any distribution function, then $G^{-1}(U)$ has distribution function G. In other words, if the marginals of $\mathbf{X}$ have continuous distribution functions $F_{X_1}, \dots, F_{X_N}$, then the random vector

$$\mathbf{Y} = \Big(G_1^{-1}\big(F_{X_1}(X_1)\big), \dots, G_N^{-1}\big(F_{X_N}(X_N)\big)\Big) \tag{5.1}$$

is indeed corresponding to a multivariate model with predetermined marginals. The distribution function C whose components are standard uniform is called copula, i.e.

$$C(u_1, \dots, u_N) = P(U_1 \leqslant u_1, \dots, U_N \leqslant u_N). \tag{5.2}$$

The Gaussian and student's t copula are defined as follows

$$\begin{aligned} C^{\mathrm{G}}_{\mathbf{\Sigma}}(\mathbf{u}) &= \Phi_{\mathbf{\Sigma}}\Big(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\Big) \\ C^{\mathrm{t}}_{\mathbf{R},\nu}(\mathbf{u}) &= t_{\mathbf{R},\nu}\Big(t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d)\Big), \end{aligned} \tag{5.3}$$

where $\mathbf{\Sigma}$ and $\mathbf{R}$ is the covariance and correlation matrix respectively, ν is the degrees of freedom.

The whole concept of copulas is best understood by example. Using same notations as above, consider $\mathbf{X}$ being two dimensional with student's t marginals with 3 degrees of freedom and correlation equal to 0.2. Now a Gaussian copula and t_3-Copula are applied on $\mathbf{X}$. The scatter plots below show the outcomes of the copulas.

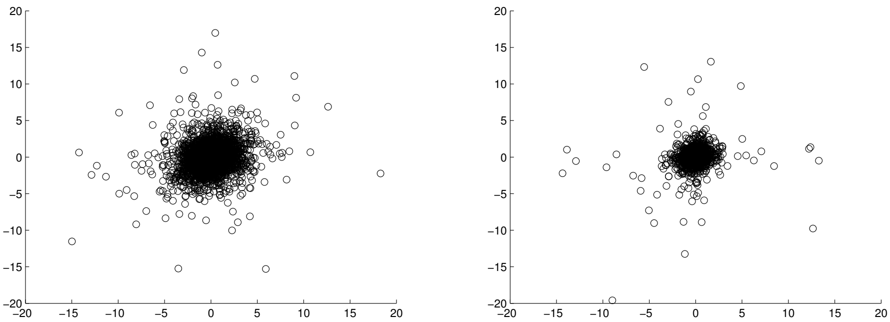


Figure 5.1: *Left plot*: t_3-Copula with student's marginals. *Right plot*: Gaussian copula with student's marginals.

Figure 5.1 clearly reveals that the t_3-Copula exhibits both heavier left and right tails than the Gaussian copula. This is regardless of the marginal being student's t-distributed.

5.1.1 Additional Comments

Copulas are a simple and comprehensible method for calibrating a multivariate distribution to asset returns. With Vasicek model as the basis, the covariance and correlation matrix become easy to compute. Having said that, copulas are still only an eloquent guess on the multivariate distribution of the asset returns.

Note that a Gaussian copula with normal marginals is equal to a multivariate normal distribution. Same applies for the t-Copula with student's t marginals, if the these have the same degree of freedom.

5.2 Consistent Information Multivariate Density Optimization

The Consistent Information Multivariate Density Optimizing (CIMDO) methodology by Segoviano [3] proceed from the premises of the Vasicek model. In contrast to the copula approach, the CIMDO methodology endeavour to minimize the assumptions concerning the multivariate distribution of the asset returns. To show this, let $\mathbf{A} = [A_1, \ldots, A_r]$ represent the random vector of asset returns in each rating class. Denote the unknown but true multivariate density function as $p(\mathbf{a})$, representing the likelihood of joint movement of $\mathbf{A}$. Now capitalizing the fact that the following system of equations $(\mathcal{S})$ must hold[13]

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\mathbf{a})d^r\mathbf{a} &= 1 \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\gamma_t^1} p(\mathbf{a})d^r\mathbf{a} &= p_t^1 \\
&\vdots \\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\gamma_t^r} p(\mathbf{a})d^r\mathbf{a} &= p_t^r
\end{aligned} \tag{5.4}$$

or alternatively,

$$\begin{aligned}
\int_{\Omega} p(\mathbf{a})d^r\mathbf{a} &= 1 \\
\int_{\Omega} p(\mathbf{a})\mathbb{1}\{a_1 < \gamma_t^1\}d^r\mathbf{a} &= p_t^1 \\
&\vdots \\
\int_{\Omega} p(\mathbf{a})\mathbb{1}\{a_r < \gamma_t^r\}d^r\mathbf{a} &= p_t^r.
\end{aligned} \tag{5.5}$$

Here γ_t^r is the threshold value of rating r defined in Section 2.5.2, $\mathbb{1}\{a_r < \gamma_t^r\}$ is the default indicator function taking the value 1 if $a_r < \gamma_t^r$ (i.e. default) and zero otherwise, and p_t^r is the PD of borrowers belonging to rating r at time t.

Although the system of equations $(\mathcal{S})$ is true, it is in solitude insufficient to find or compute the true multivariate density $p(\mathbf{a})$. The reason is as follows. The only certain information available is the PD's of each rating class, which are either known beforehand or forecasted by using some econometric

[13] The integral $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\gamma_t^y}\int_{-\infty}^{\gamma_t^x} p(\mathbf{a})d^r\mathbf{a} = p_t^{x,y}$, where $p_t^{x,y}$ is the joint probability of default of ratings x and y at time t, must also hold. However in this context it is impractical to include this condition since it will lead to an overdetermined system of equations.

model (see Section 4). If the Vasicek model is presumed to hold, then the PD's reveal the frequency of the realized asset returns passing above/below the threshold value for each rating class respectively, see Figure 2.2 and Equation (2.7). What is of interest now is the asset returns *jointly* passing above or below the threshold value, which is embedded in $p(\mathbf{a})$. However, the probabilities of the asset returns jointly taking particular values and the particular values themselves are unknown. From the supposed available data described in Section 1.3, there are no possibilities of estimating these outcomes nor probabilities. Therefore, the amount of potential densities $p(\mathbf{a})$ satisfying $(\mathcal{S})$ is indeed innumerous.

Instead of solving $(\mathcal{S})$ directly, one could approximate $p(\mathbf{a})$, henceforth referred to as the *posterior*, by using a known density. Let $q(\mathbf{a})$ denote this arbitrary chosen density, henceforth referred to as the *prior*. Through any approximation it will naturally emerge information losses. Thus, it appears logical to minimize the KL-divergence (described in Section 3.3) between the prior and posterior.

However, $(\mathcal{S})$ must yet be satisfied. If these equations are viewed as constrains of KL-divergence it is possible to formulate the Lagrangian function $\mathcal{L}$ as

$$\begin{aligned}
\mathcal{L}(p \mid q) = \int_{\Omega} p(\mathbf{a}) \ln \left[\frac{p(\mathbf{a})}{q(\mathbf{a})}\right] d^r\mathbf{a} + \lambda_0 \left[\int_{\Omega} p(\mathbf{a}) d^r\mathbf{a} - 1 \right] \\
+ \lambda_1 \left[\int_{\Omega} p(\mathbf{a}) \mathbb{1}\{a_1 < \gamma_t^1\} d^r\mathbf{a} - p_t^1 \right] \\
\vdots \\
+ \lambda_r \left[\int_{\Omega} p(\mathbf{a}) \mathbb{1}\{a_r < \gamma_t^r\} d^r\mathbf{a} - p_t^r \right],
\end{aligned} \tag{5.6}$$

where $\lambda_0, \ldots, \lambda_r$ are the Lagrangian multipliers. The calculus of variations (described in Section 3.1) is utilized in order to find an optimal solution for the Lagrangian. In Equation (5.6) there are no derivatives on the functions and therefore the Euler-Lagrange equations are automatically reduced. Moreover, since the prior is arbitrarily chosen and consequently known, it will in the context of Theorem 3.1 be treated as a constant. Thus, the Euler-Lagrange equations are actually a single equation depending on one function. After applying the calculus of variations the posterior is easily obtained by

$$\begin{aligned}
&\ln \left[\frac{p(\mathbf{a})}{q(\mathbf{a})}\right] + 1 + \lambda_0 + \lambda_1 \mathbb{1}\{a_1 < \gamma_t^1\} + \cdots + \lambda_r \mathbb{1}\{a_r < \gamma_t^r\} = 0 \Rightarrow \\
&p(\mathbf{a}) = q(\mathbf{a}) \exp \Big[-1 - \lambda_0 - \lambda_1 \mathbb{1}\{a_1 < \gamma_t^r\} - \cdots - \lambda_r \mathbb{1}\{a_r < \gamma_t^r\} \Big].
\end{aligned} \tag{5.7}$$

Equation (5.7) is the optimal solution which minimizes the KL-divergence between the prior and posterior, and is at the same time consistent with the restrictions composed in ($\mathcal{S}$). The Lagrangian multipliers are obtained by inserting the solution into ($\mathcal{S}$) and thereby using optional numerical approximation technique on the integrals. In Appendix B.3 the solution is derived by using the very definition of calculus of variations.

5.2.1 Additional Comments

To summarize, the CIMDO methodology starts from an assumption on the multivariate distribution of the asset returns. However, in contrast to the copula approach which does not make any further interventions, the CIMDO proceeds by incorporating the available data to influence the conjecture. This is accomplished by minimizing the KL-divergence between the starting guess (prior) and the true but unknown density (posterior).

Thus, from this thesis starting point one could for instance estimate the correlation matrix of the asset returns by using the Vasicek model (specially Equations (2.6) and (B.1)), thence selecting a Gaussian or student's t multivariate distribution as the prior.

Nevertheless, this does not imply that the selection of prior will not affect the final result. By a quick look at the optimal solution in Equation (5.7), it is inevitable that properties of the prior will be inherited by the posterior. For instance, a choice of multivariate Gaussian as prior is likely to cause the posterior to presumably underestimate the extreme losses. A comparison between the prior and posterior in two dimensional case is presented in the Result Section 7.3.

The CIMDO in its original formulation is affirmative to any prior as long as the premises of the Vasicek model is maintained. Thus, if additional data is available this could be included indirectly, either by extending the Vasicek or changing the prior. Furthermore, it may be plausible to adapting the CIMDO to be acceptable for mixture models such as the CreditRisk+, however this will not be investigated in this thesis.

Chapter 6

Risk Analysis of the Loss Distribution

As mentioned earlier, the losses can be forecasted after PD modeling is completed and a default modeling framework is in place[14], see Figure 1.1. Recall Section 1.3, the last goal in the thesis is to define a risk analysis method for quantifying the uncertainty in the generated loss distribution. The motivation is to be capable of form perceptions and make judgments on which default modeling framework that performs best, for instance comparing the CIMDO with the Copula approach.

This chapter presents the method by Bernard and Vanduffel [4]. They claim that the major model risk emerges due to the complexity of fitting a multivariate distribution to asset returns, which also was pointed out in Section 2.5.2. There it was mentioned that the latent factor representation, i.e. the simulation from a normal distribution to symbolize the outcome of an asset return, works well for each obligor separately. The difficulty is to *simultaneously* generate asset returns to properly reflect multiple defaults. Misspecifications in the multivariate distribution will lead to inaccurate calculations of the losses.

Thus, it is desirable to quantify the inaccuracy emerging due to multivariate distribution fitting. The method by Bernard and Vanduffel estimates bounds on the variance, the VaR and the ES of the loss distribution.[15] The wider these bounds are, the more uncertainty there is on the multivariate distribution. The remainder of this chapter explains the method by Bernard and Vanduffel.

[14] See Equation 2.1 to see how losses are calculated. Modeling of LGD and EAD are excluded. However, these parameters will be set to constants.

[15] See Section 3.6 for the definition of VaR and ES.

6.0.1 Distribution Uncertainty

Let $\mathbf{a}^m = (a_1^m, \ldots, a_N^m)$ represent the m:th sample from an N-dimensional multivariate distribution. In context of this thesis, suppose that this multivariate distribution has been fitted by the CIMDO approach or some Copula, and that the vector $\mathbf{a}^m$ represents a simultaneous simulation of the asset returns of N obligors.

With regard to what has been aforementioned, there is an awareness of misspecification. For this reason, the space $\mathbb{R}^N$ is divided into two parts; a trusted area $\mathcal{T}$ and an untrusted area $\mathcal{U}$. The samples considered sufficiently creditable to have been generated from the fitted distribution are included in $\mathcal{T}$, whereas the rest belongs to $\mathcal{U}$. Although the samples comprised in $\mathcal{U}$ are generated from the fitted distribution, these will be treated as if they are from an *unknown* distribution. By definition

$$\mathbb{R}^N = \mathcal{T} \bigcup \mathcal{U}, \qquad \emptyset = \mathcal{T} \bigcap \mathcal{U}.$$

Now the question is how to determine whether a specific sample belongs to the subset $\mathcal{T}$ or $\mathcal{U}$. A realized vector from any arbitrary distribution is intuitively most likely to be located nearby the expected value. Thus, it is suitable to utilize the Mahalanobis Distance (MD), described in Section 3.4. Therefore, the trusted area will be defined as

$$\mathcal{T} \equiv \left\{ \mathbf{a}^m \in \mathbb{R}^N,\ \ m = 1, \ldots, M \mid \mathcal{M}(\mathbf{a}^m) \leqslant c(p_{\mathcal{T}}) \right\}, \tag{6.1}$$

where M is the total number of samples and $\mathcal{M}(\mathbf{a}^m)$ is the MD of $\mathbf{a}^m$. $p_{\mathcal{T}} = \Pr[\mathbf{a}^m \in \mathcal{T}]$ is the level of trustworthiness one has on the distribution and is arbitrarily selected. The closer $p_{\mathcal{T}}$ is to 1, the more confidence one has on the distribution. The reverse is true when $p_{\mathcal{T}}$ is close to zero. $c(p_{\mathcal{T}})$ is the threshold value and is equal to the quantile at level $p_{\mathcal{T}}$ of the distribution of $\mathcal{M}$.

To visualize all this, Figure 6.1 below shows the qq-plot of samples presumed to come from a two-dimensional normal distribution against the corresponding MD-distribution.

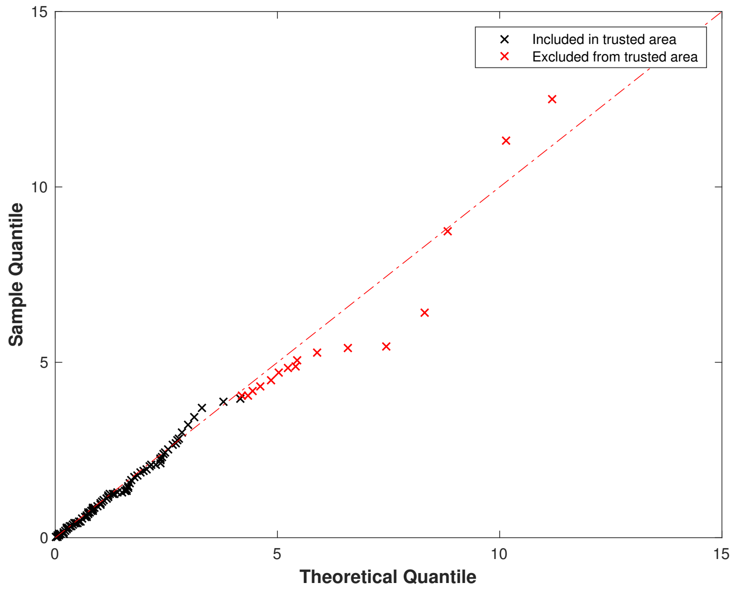


Figure 6.1: Samples presumed to come from a two-dimensional normal distribution, against the theoretical MD-distribution. $p_{\mathcal{T}}$ is set to 0.85 giving $c(p_{\mathcal{T}})$ around 4. Black samples are included in $\mathcal{T}$ while reds are included in $\mathcal{U}$.

6.0.2 Defining Upper and Lower Bounds

Before defining the upper and lower bounds on the variance and risk measures, see Equation 2.1 to recall how the losses are calculated. Recall also that obligor i is considered defaulted when the asset return falls below the corresponding threshold value, see Section 2.5.2 and Figure 2.2. Consider a vector $\mathbf{R}$ representing *one* random simulation of the asset returns. For clearness, if this particular sample belongs to $\mathcal{T}$ then $\mathbf{R}$ is renamed $\mathbf{A}$, while if it belongs to denote $\mathcal{U}$ then $\mathbf{R}$ is renamed $\mathbf{Z}$. For simplicity, the LGD and EAD are set to 1 for all obligors. The loss could now be expressed as

$$L \overset{d}{=} \mathbb{1}_{[\mathbf{R}\in\mathcal{T}]} \sum_i \mathbb{1}\{A_i < \gamma_i\} + \mathbb{1}_{[\mathbf{R}\in\mathcal{U}]} \sum_i \mathbb{1}\{Z_i < \gamma_i\}, \tag{6.2}$$

where the indicator function $\mathbb{1}_{[\mathbf{R}\in\mathcal{T}]}$ is equal to 1 if $\mathbf{R}$ belongs to $\mathcal{T}$. Here $\mathbb{1}\{A_i < \gamma_i\}$ is equal to 1 if the asset return falls below the threshold.

Once again it is important to emphasize, any vector $\mathbf{Z}$ in $\mathcal{U}$ is treated as having an unknown multivariate distribution. Hence it is inevitable that certain characteristics regarding the loss distribution is inherited from $\mathbf{Z}$. In financial mathematics, the term *comonotonicity* is a central concept describing the existence of maximum correlation between stochastic variables.

In general, a portfolio is most risky if a comonotonic dependence structure is prevailing among its components. Let $\mathbf{Z}^{\text{com}}$ denote the comonotonic representation of $\mathbf{Z}$. According to Lindskog et. al. [31] for any convex function $f(x)$ the following statement holds

$$\mathrm{E}\Big[f\Big(\sum_i Z_i\Big)\Big] \leqslant \mathrm{E}\Big[f\Big(\sum_i Z_i^{\text{com}}\Big)\Big]. \tag{6.3}$$

Equation (6.3) is intuitively reasonable. In general, stronger correlation will yield outcomes jointly having larger magnitude, hence the sum of realized random variables inserted into a convex function will naturally be greater. Thus, implementing a comonotonic dependence among the risk components in the untrusted area will yield an upper bound for any convex function. Moreover, Dhaene et. al. [16] prove the following statement

$$\mathrm{E}\Big[f\Big(\sum_i Z_i\Big)\Big] \geqslant \mathrm{E}\Big[f\Big(\sum_i E[Z_i]\Big)\Big] = \mathrm{E}\Big[f\Big(\sum_i E[Z_i^{\text{com}}]\Big)\Big]. \tag{6.4}$$

The inequality sign on the left in Equation (6.4) comes directly from Jensen's inequality. The equality sign is intuitive. Commonotonicity depicts maximum correlation among random variables but does not change the corresponding marginal distributions. Therefore the sum of expected values of any random variables is in fact equal to the sum of expected values of their comonotonic representation.

Hence the Equations (6.3) and (6.4) yield the upper and lower bounds of any convex risk measure ρ applied on credit losses, i.e.

$$\begin{aligned}
\rho^- &= \rho\Big[\sum_i \Big(\mathbb{1}_{[\mathbf{R}\in\mathcal{T}]}\mathbb{1}\{A_i < \gamma_i\} + \mathbb{1}_{[\mathbf{R}\in\mathcal{U}]}\mathbb{1}\{\mathrm{E}[Z_i^{\text{com}}] < \gamma_i\}\Big)\Big] \\
\rho^+ &= \rho\Big[\sum_i \Big(\mathbb{1}_{[\mathbf{R}\in\mathcal{T}]}\mathbb{1}\{A_i < \gamma_i\} + \mathbb{1}_{[\mathbf{R}\in\mathcal{U}]}\mathbb{1}\{Z_i^{\text{com}} < \gamma_i\}\Big)\Big].
\end{aligned} \tag{6.5}$$

Next section outlines how the upper and lower bound is obtained in practice.

6.0.3 Preparation for Approximation of Bounds

Let $m_{\mathcal{T}}$ be the total number of samples allocated to $\mathcal{T}$ while $m_{\mathcal{U}}$ is the remaining samples allocated to $\mathcal{U}$, thus $m_{\mathcal{T}}+m_{\mathcal{U}} = M$. Furthermore, denote $\mathbf{a}_j = (a_{j,1}, \ldots, a_{j,N})$ as the j:th sample in $\mathcal{T}$, where $j = 1, \ldots, m_{\mathcal{T}}$. Similarly, $\mathbf{z}_v = (z_{v,1}, \ldots, z_{v,N})$ is the v:th sample in $\mathcal{U}$, where $v = 1, \ldots, m_{\mathcal{U}}$.

All samples will be inserted into an $M \times N$ matrix $\mathbf{V}$. The first $m_{\mathcal{T}}$ rows of $\mathbf{V}$ are the samples from $\mathcal{T}$. For simplicity, let the LGD and EAD be equal to 1 for all obligors. Let $s_j = \sum_{i=1}^{N} \mathbb{1}\{a_{j,i} < \gamma_i\}$ be the loss generated from sample j. All the generated losses will be sorted in a column vector $\mathbf{S}_{\mathcal{T}}$.

For future computational purposes, the losses will be sorted in descending order, i.e. $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_{m_\mathcal{T}}$. The sample with the highest loss will be inserted into the first row in $\mathbf{V}$, the sample with the second highest loss will be inserted into the second row in $\mathbf{V}$ etc. The sorting will neither violate the dependence structure nor the distribution.

The sorting procedure of the last $m_\mathcal{U}$ rows in $\mathbf{V}$ is different. Recall that samples in the untrusted area are treated as if they where coming from some unknown multivariate distribution, or in other words, there are no assumptions regarding the dependence structure between the random variables in $\mathbf{z}_v$. However, from Equation (6.5) it is obvious that comonotonic dependence between the random variables in $\mathbf{z}_v$ is desirable. To obtain this, all samples from $\mathcal{U}$ is first arbitrarily inserted into $\mathbf{V}$, thereafter the elements in each column is sorted in descending order, i.e. $z_{1,n} \geqslant z_{2,n} \geqslant \cdots \geqslant z_{m_\mathcal{U},n}$ for $n = 1, \ldots, N$. After the sorting is complete, the loss $\widetilde{s}_v = \sum_{i=1}^{N} \mathbb{1}\{z_{v,i} < \gamma_i\}$ is calculated for $v = 1, \ldots, m_\mathcal{U}$ and stored in the column vector $\mathbf{S}_\mathcal{U}$. Hence, matrix $\mathbf{V}$ with corresponding matrices of the credit losses are structured as

$$\mathbf{V} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_\mathcal{T},1} & a_{m_\mathcal{T},2} & \cdots & a_{m_\mathcal{T},N} \\ z_{1,1} & z_{1,2} & \cdots & z_{1,N} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m_\mathcal{U},1} & z_{m_\mathcal{U},2} & \cdots & z_{m_\mathcal{U},N} \end{bmatrix} \quad \begin{matrix} \mathbf{S}_\mathcal{T} = \begin{bmatrix} s_1 = \sum_{i=1}^{N} \mathbb{1}\{a_{1,i} < \gamma_i\} \\ s_2 \\ \vdots \\ s_{m_\mathcal{T}} \end{bmatrix} \\ \mathbf{S}_\mathcal{U} = \begin{bmatrix} \widetilde{s}_1 = \sum_{i=1}^{N} \mathbb{1}\{z_{1,i} < \gamma_i\} \\ \widetilde{s}_2 \\ \vdots \\ \widetilde{s}_{m_\mathcal{U}} \end{bmatrix} \end{matrix} \tag{6.6}$$

To fully grasp this section, the reader is encouraged to study the example provided in Appendix B.4. The rest of this chapter presents how the bounds on the variance and risk measures are obtained from the matrices in (6.6).

6.0.4 Bounds on Variance

The implementation of comonotonic dependence between the marginals in the untrusted area makes the calculation of the upper bound of the portfolio variance straightforward. The upper bound is given by

$$\rho^+_{\text{variance}} = \frac{1}{M}\left(\sum_{i=1}^{m_\mathcal{T}} \left(s_i - \bar{s}\right)^2 + \sum_{i=1}^{m_\mathcal{U}} \left(\widetilde{s}_i - \bar{s}\right)^2 \right), \tag{6.7}$$

where $\bar{s}$ is the total average of $\mathbf{S}_\mathcal{T}$ and $\mathbf{S}_\mathcal{U}$. Obtaining the lower bound is trickier, Equation (6.5) shows that the expected value of each comonotonic random variable must be calculated. The information available is the sums in $\mathbf{S}_{\mathcal{U}}$. However, for any random vector $\mathbf{X} = (X_1, \ldots, X_n)$, $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathrm{E}[X_i]$ is true if the sum of the marginals have constant quantile on $(0, 1)$. This is commonly known as *joint mixability* and is asymptotically obtained as the elements in $\mathbf{S}_{\mathcal{U}}$ gets closer to each other. Recall the RA described in Section 3.5, the tool used for decreasing the variance of the row sums in any matrix. In general, minimization of variance is equivalent to minimizing the spread of the sample. Thus, to obtain joint mixability in practice, the RA is applied on the untrusted area of $\mathbf{V}$. The lower bound is subsequently calculated by

$$\rho^{-}_{\text{variance}} = \frac{1}{M}\left(\sum_{i=1}^{m_{\mathcal{T}}}\left(s_i - \bar{s}\right)^2 + \sum_{i=1}^{m_{\mathcal{U}}}\left(\widetilde{s}_i^{\,\mathrm{RA}} - \bar{s}\right)^2\right), \tag{6.8}$$

where $\widetilde{s}_i^{\,\mathrm{RA}}$ is the elements of $\mathbf{S}_{\mathcal{U}}$ after the RA is implemented.

6.0.5 Bounds on Expected Shortfall

To calculate the bounds on ES at level p, all the values in $\mathbf{S}_{\mathcal{T}}$ and $\mathbf{S}_{\mathcal{U}}$ are first sorted in descending order. To obtain the upper bound, pick out the k highest values, where $k = M(1-p)$, and calculate the ES from these values.

For the lower bound, apply the RA on $\mathbf{S}_{\mathcal{U}}$ and proceed just like the upper bound. The argument for applying the RA is the same as for the variance.

6.0.6 Bounds on Value-at-Risk

As mentioned in Section 3.6.1, VaR is not sub-additive and for this reason not convex. Therefore Equation (6.5) can not be applied directly. Fortunately, obtaining the upper and lower bound are similar to the variance and ES. For the upper bound, the sum of the untrusted comonotonic marginals are replaced with the sum of the ES of the untrusted comonotonic marginals. Similarly, for the lower bound, the sum of the expected value of untrusted comonotonic marginals are replaced with the sum of the LTVaR of the untrusted comonotonic marginals. In Algorithm 5 below yields the approximation of the bounds on VaR. For details and derivations see Bernard and Vanduffel [4].

Algorithm 5: Computing bounds on Value-at-Risk at level p

1 Compute $\alpha_1 = \max\left\{0, \frac{p+p_{\mathcal{T}}-1}{p_{\mathcal{T}}}\right\}$, $\alpha_2 = \min\left\{1, \frac{p}{p_{\mathcal{T}}}\right\}$ and $k = M(1-p)$.

2 For the upper bound, calculate
$\alpha_* = \inf\Big\{\alpha \in (\alpha_1, \alpha_2) \mid \mathrm{VaR}_\alpha(\mathbf{S}_{\mathcal{T}}) \geqslant \mathrm{ES}_{\frac{p-p_{\mathcal{T}}\alpha}{1-p_{\mathcal{T}}}}(\mathbf{S}_{\mathcal{U}})\Big\}$,
and for the lower bound
$\alpha_{**} = \inf\Big\{\alpha \in (\alpha_1, \alpha_2) \mid \mathrm{VaR}_\alpha(\mathbf{S}_{\mathcal{T}}) \geqslant \mathrm{LTVaR}_{\frac{p-p_{\mathcal{T}}\alpha}{1-p_{\mathcal{T}}}}(\mathbf{S}_{\mathcal{U}})\Big\}$.

3 Calculate $\beta_* = \frac{p-p_{\mathcal{T}}\alpha_*}{1-p_{\mathcal{T}}}$ and $\beta_{**} = \frac{p-p_{\mathcal{T}}\alpha_{**}}{1-p_{\mathcal{T}}}$. For the upper bound, perform the RA on the first $\lfloor(1-\beta_*)m_{\mathcal{T}}\rfloor$ rows of the untrusted part of $\mathbf{V}$. For the lower bound, perform the RA on the last $\lceil\beta_{**}m_{\mathcal{U}}\rceil$ rows of $\mathbf{V}$.

4 Calculate the losses and sort them in descending order $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_M$. The maximum/minimum VaR is s_k.

6.0.7 Additional Comments

The method by Bernard and Vanduffel [4] provides a satisfactorily platform to measure the uncertainty in the losses generated by the CIMDO and Copula approach. The idea is to compare the variance, ES and VaR bounds of the loss distribution generated by the copula approach and the CIMDO at different level of trustworthiness. The smaller the bounds are the more certainty there is on the modeling framework.

Since $p_{\mathcal{T}}$ is arbitrarily selected, the modeller is able to shift the level of trustworthiness for investigation purposes. The bounds will certainly get wider as $p_{\mathcal{T}}$ get closer to 0.

The method could also be applied under various scenarios to study how the uncertainty in the loss distribution changes. For instance, one could test the CIMDO and Copula approach during adverse macroeconomic conditions, see Appendix A.2 for detailed explanation of stress-testing.

Chapter 7

Results

In this chapter results, analyses and discussions of the selected models are presented.

7.1 Data

Unfortunately, no real data was available for this thesis. Instead, the process of generating data was as follows. First Japan's GDP annual growth rate and Nikkei 225[16] annual growth rate were selected as the exogenous variables. The reason for this particular choice was that the correlation between these macro indicators is relatively weak (about 0.2), i.e. the motive was to minimize the effects of multicollinearity as much as possible.

Six rating classes were considered, where rating 1 had the lowest expected default rate whereas rating 6 had the highest. The regression coefficients of the macro indicators plus an intercept were fixed, in such way that if a normal economic scenario corresponding to a GDP growth of +2% and an increase of Nikkei 225 by +15% were inserted into the OLS solution in Equation (4.1), the default probabilities became 1%, 3%, 5%, 8%, 14% and 30% for each rating class respectively. The fixed coefficients are presented in Table 7.1 in the subsequent section.

To obtain the yearly default rates for each rating class, historical outcomes over the past 15 years of the chosen macro variables (see Figure 7.1) was inserted into Equation (4.1). Note that in statistical terms 15 data points is considered as a small sized sample. Finally, a normally distributed disturbance where added with standard deviation equal to the standard deviation of the generated default rates.

[16]The Nikkei 225 Stock Average is a price-weighted index of the 225 best blue chip companies on the Tokyo Stock Exchange.

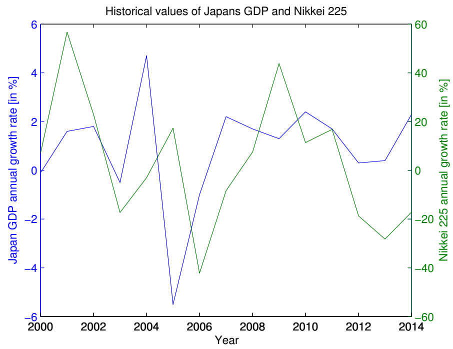
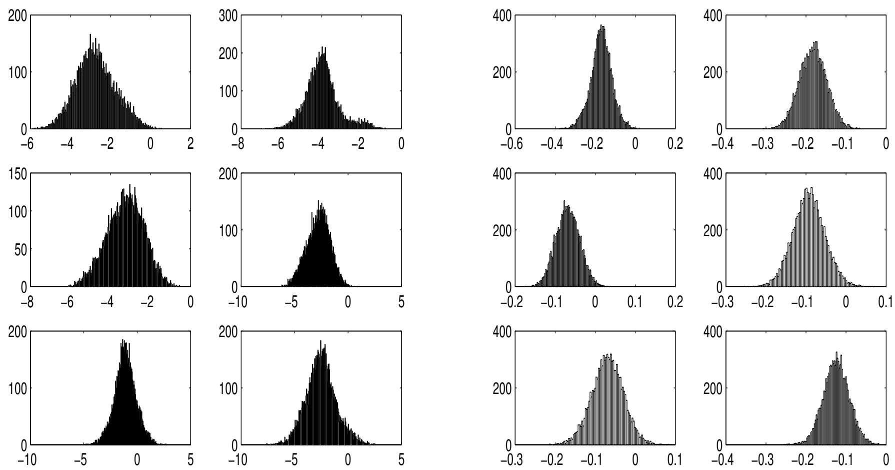


Figure 7.1: Historical values from year 2000 to 2014 of Japan's GDP annual growth rate and Nikkei 225 annual growth rate.

7.2 Econometric Models

Recall Chapter 4, where the econometric models were explained. Bootstrap was performed 10'000 times in order to obtain the distribution of each coefficient. Seven points were selected for the entropy model, namely the mean and $\pm 3, \pm 2, \pm 1$ standard deviations from the mean. The resulting coefficients along with the OLS and ML estimates are given in Table 7.1 below. The bootstrapped distributions are shown in Figure 7.2.



(a) GDP coefficients (b) Nikkei coefficients

Figure 7.2: The distributions of all coefficients generated by bootstrapping 10'000 times. *Upper left plots* corresponds to rating 1, *upper right plots* corresponds to rating 2, *middle left plots* for rating 3 etc.

Rating		1	2	3	4	5	6
Fixed β	Intercept	-2.30	-1.83	-1.60	-1.35	-1.01	-0.45
	GDP	-1.7	-2.0	-2.30	-2.80	-3.0	-3.2
	Nikkei	-0.2	-0.15	-0.12	-0.10	-0.08	-0.07
$\widehat{\beta}_{OLS}$	Intercept	-2.3130	-1.8257	-1.6053	-1.3507	-1.0141	-0.4466
	GDP	-1.5994	-1.9495	-2.9921	-2.7066	-3.2886	-3.5998
	Nikkei	-0.2600	-0.1458	-0.1375	-0.1240	-0.1002	-0.010
$\widehat{\beta}_{ML}$	Intercept	-2.3123	-1.8251	-1.6049	-1.3502	-1.0140	-0.4465
	GDP	-1.5710	-1.9400	-2.9243	-2.6969	-3.2380	-3.5729
	Nikkei	-0.2576	-0.1448	-0.1375	-0.1269	-0.1017	-0.019
$\widehat{\beta}_{ent.}$	Intercept	-2.3129	-1.8257	-1.6054	-1.3506	-1.0140	-0.4466
	GDP	-1.6087	-1.9494	-2.9893	-2.7124	-3.2897	-3.6003
	Nikkei	-0.2599	-0.1459	-0.1375	-0.1238	-0.1003	-0.011

Table 7.1: Table showing the fixed coefficients of all exogenous variables and rating classes along with the estimates from OLS, ML and the entropy model.

Table 7.1 reveals several interesting things. The OLS, ML and the entropy model have almost the same size on all coefficients. Most importantly, all models correctly estimate the signs. The estimates of the intercepts are sufficiently accurate in all models, however, some coefficients of the macro variables are not. For instance the Nikkei coeff. in rating class 6 is about 85% lower than correct value, and the GDP coeff. in rating 3 is 30% greater than the fixed coefficient. Note that due to the added disturbance the outcomes will be moderately different each time one estimates the coefficients. Therefore, the results from the historical values should be viewed as representatively illustrating. Same applies for the bootstrapped distributions in Figure 7.2.

The result from the historical values does not establish whether the entropy model actually reduces the MSE compared to the OLS and ML. For this reason a simulation exercise was carried out. First, the GDP and the Nikkei 225 were fitted to a multivariate AR(2)-process by utilizing historical values from the past 40 years. 15 outcomes were simulated from this time-series model and subsequently inserted to the regression model (fixed coefficients with additional disturbance) to generate new default rates. Thereafter new coefficients were estimated by OLS, ML and the entropy model. The points on the bootstrapped distributions were selected exactly as before. The procedure was repeated 100'000 times in order to calculate the variance, bias and MSE of each coefficient. Recall that MSE is defined as the sum of variance and squared bias. The result for the OLS and entropy model is presented in Table 7.2 below.

Rating		1	2	3	4	5	6
Var.	$\beta_{OLS}^{Int.}$	$1.9 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$
	β_{OLS}^{GDP}	2.3058	1.7456	1.7321	2.1556	2.1330	2.1824
	$\beta_{OLS}^{Nikk.}$	$2.15 \cdot 10^{-2}$	$1.68 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$
	$\beta_{ent.}^{Int.}$	$1.6 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$
	$\beta_{ent.}^{GDP}$	1.0395	0.7231	0.7143	0.9123	0.8838	0.9425
	$\beta_{ent.}^{Nikk.}$	$7.6 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$7.0 \cdot 10^{-3}$	$7.1 \cdot 10^{-3}$	$7.5 \cdot 10^{-3}$
	$\Delta_{Int.}$	20.77%	16.29%	13.38%	11.39%	6.07%	-11.74%
	Δ_{GDP}	121.81%	141.41%	142.50%	136.27%	141.35%	131.56%
	$\Delta_{Nikk.}$	183.81%	211.63%	214.19%	206.10%	204.08%	198.55%
Bias sqr.	$\beta_{OLS}^{Int.}$	$1.16 \cdot 10^{-4}$	$6.06 \cdot 10^{-5}$	$5.24 \cdot 10^{-5}$	$5.49 \cdot 10^{-5}$	$3.42 \cdot 10^{-5}$	$7.88 \cdot 10^{-6}$
	β_{OLS}^{GDP}	$1.52 \cdot 10^{-2}$	$8.27 \cdot 10^{-3}$	$7.03 \cdot 10^{-3}$	$1.48 \cdot 10^{-2}$	$6.88 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$
	$\beta_{OLS}^{Nikk.}$	$1.78 \cdot 10^{-4}$	$5.18 \cdot 10^{-5}$	$2.40 \cdot 10^{-5}$	$1.20 \cdot 10^{-5}$	$3.46 \cdot 10^{-6}$	$1.38 \cdot 10^{-6}$
	$\beta_{ent.}^{Int.}$	$8.1 \cdot 10^{-6}$	$3.24 \cdot 10^{-7}$	$4.55 \cdot 10^{-7}$	$1.83 \cdot 10^{-5}$	$1.07 \cdot 10^{-4}$	$1.01 \cdot 10^{-3}$
	$\beta_{ent.}^{GDP}$	$6.68 \cdot 10^{-2}$	0.213	0.344	0.504	0.618	0.541
	$\beta_{ent.}^{Nikk.}$	$4.52 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$	$1.28 \cdot 10^{-3}$	$5.43 \cdot 10^{-4}$	$1.77 \cdot 10^{-4}$	$8.43 \cdot 10^{-6}$
MSE	$\beta_{OLS}^{Int.}$	$2.03 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$2.05 \cdot 10^{-3}$	$2.22 \cdot 10^{-3}$
	β_{OLS}^{GDP}	2.321	1.754	1.739	2.170	2.140	2.183
	$\beta_{OLS}^{Nikk.}$	$2.16 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$
	$\beta_{ent.}^{Int.}$	$1.59 \cdot 10^{-3}$	$1.31 \cdot 10^{-3}$	$1.34 \cdot 10^{-3}$	$1.76 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$3.52 \cdot 10^{-3}$
	$\beta_{ent.}^{GDP}$	1.106	0.936	1.058	1.416	1.502	1.484
	$\beta_{ent.}^{Nikk.}$	$1.2 \cdot 10^{-2}$	$7.84 \cdot 10^{-3}$	$6.65 \cdot 10^{-3}$	$7.55 \cdot 10^{-3}$	$7.28 \cdot 10^{-3}$	$7.55 \cdot 10^{-3}$
	$\Delta_{Int.}$	27.46%	20.90%	17.25%	13.35%	2.11%	-36.84%
	Δ_{GDP}	109.80%	87.40%	64.41%	53.27%	42.29%	47.19%
	$\Delta_{Nikk.}$	79.11%	115.12%	154.07%	184.27%	196.72%	198.24%

Table 7.2: The variance, bias squared and the MSE of each coefficient obtained from the simulation.

Looking at the MSE difference in Table 7.2, one can conclude that the entropy model is most efficient for all coefficients except one. For many coefficients the improvement is above 100%.

The reason lies in the reduction of the variances. By using the OLS, the variances become extremely large, specially for the GDP coefficients. The presumed explanation is that the small sample size leads the OLS to cause *overfitting*, a term outlining when a statistical model describes the error in the data set rather than the underlying relationship. Some textbooks mention an informal rule to at least have 15-20 data points per independent variable in order to prevent against overfitting. Thus, if for example 3 macroeconomic indicators with yearly update are chosen, one has to collect data up to 60 years. This is obviously unrealistic and shows the huge drawback of the OLS.

The squared biases, however, are worse in most cases for the entropy model. This has probably to do with the so-called *bias-variance tradeoff*. In general, if one of these variables decrease, the other will increase. This phenomenon becomes visually apparent by looking at the definitions. If θ is the parameter to be estimated and $\widehat{\theta}$ is the actual estimation, then the bias is defined as $\mathrm{E}[\widehat{\theta}] - \theta$ and the variance is $E[(\widehat{\theta} - E[\widehat{\theta}])^2]$. From the result one can conclude that the entropy model reduces the variance, but at the cost of increasing the bias. The worst case is the GDP coefficient in rating 5, yielding approximately 25% wrong estimation from the true value. For the intercepts and the Nikkei coefficients, the biases are sufficiently small to be acceptable. Nevertheless, the overall picture is that the entropy significantly improves the estimation.

Furthermore, the number and allocation of the selected points on the bootstrap distribution for the entropy model are strongly affecting its performance. Segoviano [2] claims that the selection of five points is optimal, but from an examination it turned out not be true in this case. Nevertheless, no attempts were made to find the best solution and thus it is possible to decrease the variance and/or bias further and thereby making the entropy model even more efficient.

Due to absence of real data, no attempts were made on testing an ADL set-up nor time dependency lag on the explanatory variables. However, the result is expected to be similar as above. Finally, the entropy model also outperforms the ML. The results are similar and presented in Appendix C.

7.3 Comparison of Prior and Posterior in Two Dimensional Case

The actual difference between the prior and posterior distribution will now be analyzed. Two rating classes (X, Y) is considered, and the bivariate standard normal and student's t-distribution with 3 degrees of freedom will be selected as prior. The PD's are set to 5% and 10% respectively, and the (asset) correlation is 0.15. The Lagrangian multipliers are given in Table 7.3 below.

LM	Prior Normal	Prior Student
$\widehat{\lambda}_0$	-1.000	-0.910
$\widehat{\lambda}_1$	-0.001	-1.017
$\widehat{\lambda}_2$	-0.003	-1.024

Table 7.3: The Lagrangian multipliers.

In order to visualize the difference, the qq-plots for each rating (X, Y) and each scenario were plotted, see Figure 7.3 below.

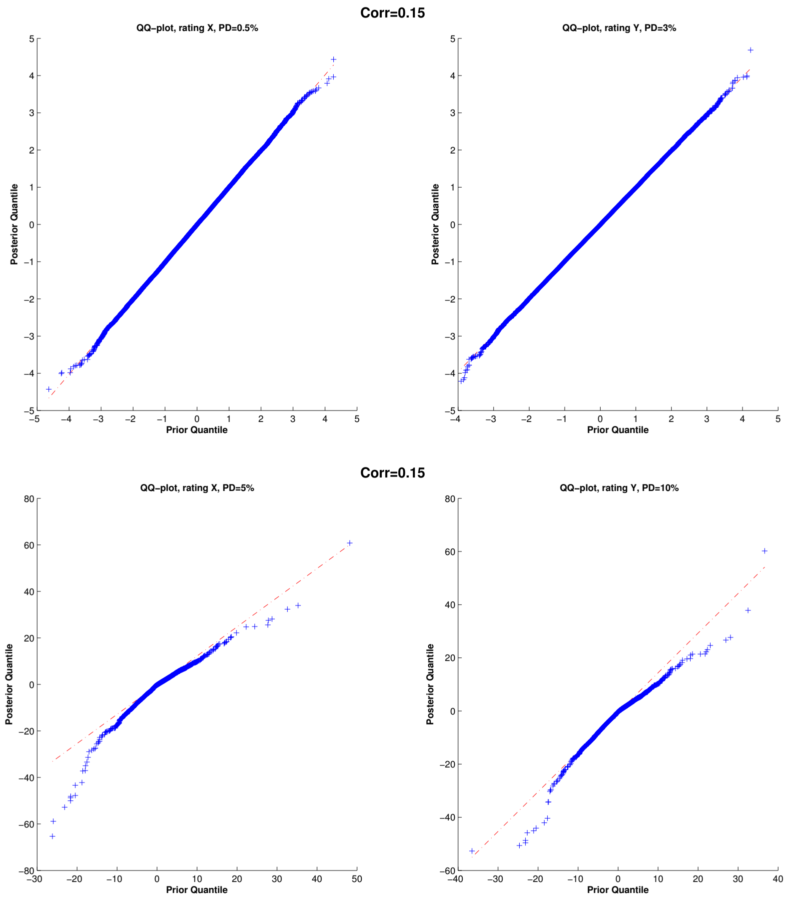


Figure 7.3: QQ-plots of each asset return (x, y) generated by the posterior and prior distribution. *Upper left*: Asset return x when prior is normally distributed. *Upper right*: Asset return y when prior is normally distributed. *Lower left*: Asset return x when prior is student's t. *Lower right*: Asset return y when prior is student's t.

From the upper two plots in Figure 7.3 one can see that the qq-plots are linear, which is a indication of no difference between the prior and posterior.

In the lower two plots, when the prior is student's t, the qq-plots are curving downwards for both low and high quantiles. This demonstrates that the posterior has heavier left tail but lighter right tail compared to the prior. This is further visualized by scatter plots and figures of the densities in Appendix C.2. Also, similar results are obtained when the PD's and correlation are varied, see Appendix C.3.

The conclusion is that, depending on the prior, there could be a significant difference between the posterior and prior distributions. Hence, the choice of prior is directly affecting the size of disparity between the two distributions. Although it may be apparent from the optimal solution in Equation (5.7), these figures establish that the posterior inherits the characteristics of the prior. Thus the risk manager must carefully select the prior, for instance in the light of economic theory or preferably empirical data.

7.4 Analysis of Variance, ES and VaR Bounds

To perform a risk analysis on the loss distribution, a portfolio of 55 obligors was considered; 5 from rating 1, 20 from rating 2, 15 from rating 3, 10 from rating 4, 4 from rating 5 and 1 from rating 6. The portfolio was simulated under the following four scenarios

- **Scenario I**: The asset correlations for the ratings are set to [0.02 0.018 0.015 0.01 0.008 0.001] and the macroeconomic condition is regarded as normal, i.e. the PD's are equal to those stated in Section 7.1.
- **Scenario II**: Same asset correlations as in Scenario I, however the macroeconomic condition is adverse. The PD's are 1.5%, 4%, 7%, 11%, 19% and 38% for each rating class respectively.
- **Scenario III**: The asset correlations for all ratings are set to [0.06 0.054 0.045 0.03 0.024 0.015] and the PD's are same as in Scenario I.
- **Scenario IV**: Same asset correlations as in Scenario III and the PD's are the same as in Scenario II.

Four distributions were considered, namely standard normal marginals with Gaussian copula, standard normal marginals with normal CIMDO, student's t marginals with 3 degrees of freedom with student's t copula, and student's t marginals with 3 degrees of freedom with student's t CIMDO.

The procedure was as follows. From the multivariate distributions, the joint movement of the borrowers' asset returns were simulated 1'000'000 times. By utilizing the MD the outcomes were inserted into $\mathcal{T}$ and $\mathcal{U}$, see Equation (6.1). The borrowers were checked whether they defaulted or not by comparing each simulated asset return to their threshold value[17]. Finally,

[17] The threshold value is obtained by taking the inverse CDF of the PD's, see Equation (2.7).

for obtaining the risk measure bounds, the losses were inserted into matrix $\mathbf{V}$ exactly as in the example described in Appendix B.4. Losses from each counterparty were randomly predefined between 0.15 and 0.75.

The standard deviation of the Gaussian copula and the CIMDO with normal prior under the four scenarios are shown in Figure 7.4. The level of trustworthiness, $p_{\mathcal{T}}$, is varied from 1 (no uncertainty) to 0.5.

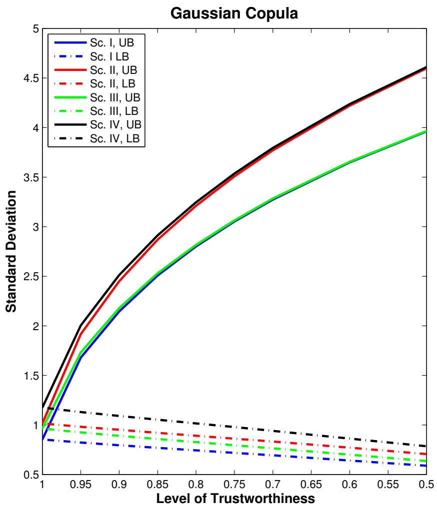


Figure 7.4: Standard deviation bounds of the loss distribution at different levels of trustworthiness. *Left plot*: Gaussian copula. *Right plot*: CIMDO with normal prior.

Not surprisingly, the left and right graphs are almost identical and thus they confirm former statements regarding no practical difference between the Gaussian copula and the CIMDO when multivariate normal is selected as prior.

For all four scenarios the upper bounds are increasing rapidly while the lower bounds declines much slower. For example, already at $p_{\mathcal{T}} = 95\%$ the upper bound in Scenario II is approximately twice as large as when no model uncertainty is concerned. As $p_{\mathcal{T}}$ approaches 50%, the same upper bound increases to be almost 4.6 times larger than the benchmark value. This strongly indicates that the risk of underestimation is overwhelming.

Furthermore, estimating the variance is more doubtful during adverse macroeconomic set-ups than normal conditions. The conclusion predicates from the end points of the curves in Figure 7.4. The bounds in Scenario II and IV compared to Scenario I and III are disproportionate over the levels of trustworthiness. More precisely, the former two scenarios have faster increase of the upper bounds.

The standard deviation of the t-Copula and the CIMDO with student's

t prior are shown in Figure 7.5.

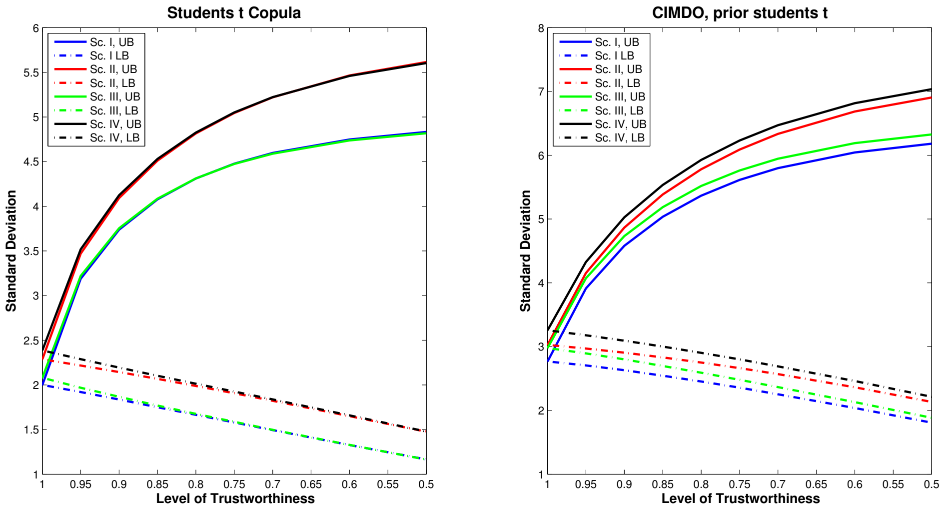


Figure 7.5: Standard deviation bounds of the loss distribution at different levels of trustworthiness. *Left plot*: Student's t copula with 3 degrees of freedom. *Right plot*: CIMDO with student's t prior.

The shape of the curves in Figure 7.5 are similar to those in Figure 7.4. However, the upper bounds for the t-Copula have shrank. For Scenario II, the upper bound at $p_{\mathcal{T}} = 0.5$ is 2.3 times larger than when no model uncertainty is concerned, which is a decrease by half compared to the corresponding case when using the Gaussian copula.

The student's t CIMDO also appear to decrease the model risk further. The ratio of the upper bound when $p_{\mathcal{T}} = 0.95$ and no model uncertainty is 8% smaller for the CIMDO in comparison to the t-Copula. Moreover, the curves seem to be more proportionate when CIMDO is used, suggesting that the stressed scenarios are not riskier in the CIMDO framework.

We now turn to the bounds on VaR and ES. The result of the Gaussian copula is presented in Table 7.4 below.

Gaussian Copula							
$p_{\mathcal{T}}$		1	0.99	0.98	0.97	0.96	0.95
Sc. I	$\text{ES}_{0.995}$	4.60	(4.48, 8.15)	(4.41, 11.50)	(4.35, 14.69)	(4.29, 17.72)	(4.24, 20.39)
	$\text{ES}_{0.99}$	4.28	(4.18, 6.07)	(4.12, 7.77)	(4.07, 9.41)	(4.02, 10.99)	(3.98, 12.50)
	$\text{ES}_{0.98}$	3.94	(3.86, 4.83)	(3.81, 5.68)	(3.77, 6.51)	(3.72, 7.31)	(3.69, 8.08)
	$\text{VaR}_{0.995}$	4.15	(4.06, 5.06)	(4.00, 9.14)	(3.96, 13.11)	(3.90, 16.38)	(3.86, 19.31)
	$\text{VaR}_{0.99}$	3.80	(3.73, 4.13)	(3.68, 4.88)	(3.65, 6.66)	(3.61, 8.72)	(3.59, 10.67)
	$\text{VaR}_{0.98}$	3.45	(3.40, 3.60)	(3.36, 3.78)	(3.33, 4.09)	(3.29, 4.56)	(3.26, 5.32)
Sc. II	$\text{ES}_{0.995}$	5.62	(5.52, 9.89)	(5.44, 13.88)	(5.39, 17.71)	(5.32, 21.32)	(5.27, 23.73)
	$\text{ES}_{0.99}$	5.27	(5.18, 7.43)	(5.12, 9.48)	(5.06, 11.48)	(5.00, 13.40)	(4.96, 15.23)
	$\text{ES}_{0.98}$	4.90	(4.82, 5.98)	(4.76, 7.02)	(4.72, 8.03)	(4.67, 9.01)	(4.63, 9.96)
	$\text{VaR}_{0.995}$	5.12	(5.04, 6.35)	(4.99, 11.64)	(4.93, 16.13)	(4.86, 20.36)	(4.71, 22.73)
	$\text{VaR}_{0.99}$	4.74	(4.67, 5.14)	(4.63, 6.11)	(4.58, 8.46)	(4.54, 11.24)	(4.49, 13.58)
	$\text{VaR}_{0.98}$	4.34	(4.29, 4.51)	(4.24, 4.77)	(4.21, 5.14)	(4.17, 5.74)	(4.14, 6.92)
Sc. III	$\text{ES}_{0.995}$	5.29	(5.18, 8.77)	(5.11, 11.76)	(5.03, 14.82)	(4.98, 17.61)	(4.93, 20.22)
	$\text{ES}_{0.99}$	4.86	(4.78, 6.63)	(4.72, 8.15)	(4.65, 9.74)	(4.60, 11.25)	(4.56, 12.69)
	$\text{ES}_{0.98}$	4.45	(4.37, 5.33)	(4.32, 6.10)	(4.27, 6.90)	(4.23, 7.66)	(4.19, 8.40)
	$\text{VaR}_{0.995}$	4.66	(4.59, 5.72)	(4.52, 8.86)	(4.49, 13.05)	(4.37, 16.12)	(4.51, 19.09)
	$\text{VaR}_{0.99}$	4.26	(4.20, 4.61)	(4.16, 5.28)	(4.10, 6.70)	(4.07, 8.60)	(4.05, 10.53)
	$\text{VaR}_{0.98}$	3.83	(3.79, 4.01)	(3.74, 4.20)	(3.69, 4.47)	(3.65, 4.89)	(3.62, 5.48)
Sc. IV	$\text{ES}_{0.995}$	6.60	(6.51, 10.73)	(6.44, 14.31)	(6.37, 17.98)	(6.30, 21.34)	(6.24, 23.97)
	$\text{ES}_{0.99}$	6.11	(6.04, 8.23)	(5.98, 10.08)	(5.91, 11.99)	(5.86, 13.80)	(5.80, 15.51)
	$\text{ES}_{0.98}$	5.61	(5.54, 6.68)	(5.49, 7.63)	(5.43, 8.61)	(5.38, 9.55)	(5.34, 10.45)
	$\text{VaR}_{0.995}$	5.92	(5.83, 7.25)	(5.78, 11.58)	(5.74, 16.22)	(5.72, 20.13)	(5.68, 23.02)
	$\text{VaR}_{0.99}$	5.40	(5.33, 5.86)	(5.28, 6.70)	(5.24, 8.61)	(5.19, 11.14)	(5.17, 13.51)
	$\text{VaR}_{0.98}$	4.86	(4.81, 5.07)	(4.78, 5.34)	(4.72, 5.71)	(4.69, 6.24)	(4.65, 7.12)

Table 7.4: The bounds on VaR and ES for various scenarios and different levels of trustworthiness when using the Gaussian copula.

Regardless of risk measure, confidence level and scenario, the lower bounds are slowly declining as $p_{\mathcal{T}}$ decreases. On the other hand, the upper bounds grows rapidly as the model uncertainty increases, specially for confidence level 99% and 99.5%. VaR and ES at 98% appear to have somewhat tight bounds for all four scenarios. Thus, a pattern are distinguished from the results in Table 7.4 and Figure 7.4, namely the likelihood of inaccurate forecasts of the losses is increasing as one moves closer to the extreme values.

Adverse macroeconomic condition seems to increase the upper bounds but stronger correlation appears to not impacting the uncertainty.

The normal CIMDO yields similar bounds as the Gaussian Copula and is therefore presented in Appendix C.4. The VaR and ES bounds for the student's t copula is presented in Table 7.5 below.

Student's t Copula							
$p_{\mathcal{T}}$		1	0.99	0.98	0.97	0.96	0.95
Sc. I	$ES_{0.995}$	12.35	(10.69, 22.60)	(9.67, 28.08)	(8.92, 28.63)	(8.33, 28.71)	(7.82, 28.74)
	$ES_{0.99}$	11.20	(10.37, 16.13)	(9.41, 20.66)	(8.72, 24.67)	(8.16, 27.24)	(7.69, 27.57)
	$ES_{0.98}$	9.88	(9.75, 12.07)	(9.19, 14.19)	(8.56, 16.16)	(8.04, 18.01)	(7.60, 19.73)
	$VaR_{0.995}$	10.77	(9.19, 19.36)	(8.78, 27.16)	(8.16, 27.94)	(7.58, 27.98)	(7.13, 28.07)
	$VaR_{0.99}$	9.46	(8.14, 10.88)	(7.36, 17.09)	(6.79, 23.58)	(6.60, 26.17)	(6.55, 26.50)
	$VaR_{0.98}$	7.85	(7.00, 8.14)	(6.38, 9.40)	(5.91, 11.68)	(5.52, 14.45)	(5.18, 17.41)
Sc. II	$ES_{0.995}$	13.12	(11.61, 23.72)	(10.71, 28.54)	(9.99, 28.72)	(9.47, 28.80)	(9.02, 28.84)
	$ES_{0.99}$	12.35	(10.69, 22.60)	(9.67, 28.08)	(8.92, 28.63)	(8.33, 28.71)	(7.82, 28.74)
	$ES_{0.98}$	11.20	(10.37, 16.13)	(9.41, 20.66)	(8.72, 24.67)	(8.16, 27.24)	(7.69, 27.57)
	$VaR_{0.995}$	11.67	(10.34, 20.71)	(10.24, 27.74)	(9.63, 27.96)	(9.34, 28.11)	(8.49, 28.11)
	$VaR_{0.99}$	10.77	(9.19, 19.36)	(8.78, 27.16)	(8.08, 27.94)	(7.84, 27.98)	(7.68, 28.08)
	$VaR_{0.98}$	9.46	(8.14, 10.88)	(7.36, 17.09)	(6.79, 23.58)	(6.60, 26.17)	(6.55, 26.50)
Sc. III	$ES_{0.995}$	13.47	(11.57, 23.17)	(10.42, 28.09)	(9.72, 28.64)	(9.04, 28.72)	(8.47, 28.75)
	$ES_{0.99}$	12.05	(10.88, 16.84)	(9.85, 21.15)	(9.16, 24.82)	(8.56, 27.21)	(8.05, 27.57)
	$ES_{0.98}$	10.51	(10.23, 12.68)	(9.49, 14.74)	(8.82, 16.64)	(8.27, 18.40)	(7.80, 20.03)
	$VaR_{0.995}$	11.47	(9.96, 19.41)	(9.65, 27.08)	(9.50, 27.97)	(8.78, 27.98)	(8.21, 28.02)
	$VaR_{0.99}$	9.89	(8.70, 11.64)	(7.88, 17.23)	(7.25, 23.77)	(7.12, 25.89)	(7.04, 26.49)
	$VaR_{0.98}$	8.19	(7.28, 8.70)	(6.63, 9.89)	(6.18, 12.01)	(5.78, 14.43)	(5.44, 17.54)
Sc. IV	$ES_{0.995}$	14.31	(12.79, 24.20)	(11.75, 28.47)	(11.19, 28.72)	(10.57, 28.80)	(10.04, 28.84)
	$ES_{0.99}$	12.99	(11.85, 18.04)	(10.94, 22.57)	(10.35, 26.36)	(9.80, 27.80)	(9.33, 28.16)
	$ES_{0.98}$	11.54	(11.16, 13.92)	(10.42, 16.18)	(9.83, 18.31)	(9.33, 20.27)	(8.92, 22.15)
	$VaR_{0.995}$	12.46	(11.25, 20.64)	(10.35, 27.55)	(9.83, 28.07)	(9.24, 27.97)	(8.77, 28.07)
	$VaR_{0.99}$	11.00	(9.98, 12.71)	(9.26, 18.90)	(8.76, 25.44)	(8.23, 26.70)	(8.04, 27.09)
	$VaR_{0.98}$	9.38	(8.52, 9.98)	(7.97, 11.02)	(7.54, 13.28)	(7.15, 16.32)	(6.86, 19.87)

Table 7.5: The bounds on VaR and ES for various scenarios and different levels of trustworthiness when using the student's t copula.

As expected, the extreme losses at $p_{\mathcal{T}} = 1$ (no model risk) are much higher for the student's t copula than in the Gaussian Copula. However, what is somewhat unexpected is the fact that the bounds are extremely wider for the t-Copula in compassion to the Gaussian Copula and normal CIMDO (for the latter see Appendix C.4). The upper bounds rapidly converge to 30 as the level of trustworthiness decreases.

Finally, the results for the student's t CIMDO is presented in Table C.3 below. The Metropolis-Hastings algorithm[18] was used for generating samples.

[18] This is a Markov chain Monte Carlo method for generating random samples. For definition and details, see Gentle et. al. [21].

CIMDO, prior student's t							
$p_{\mathcal{T}}$		1	0.99	0.98	0.97	0.96	0.95
Sc. I	$ES_{0.995}$	14.84	(13.44, 25.90)	(12.55, 28.71)	(11.83, 28.85)	(11.22, 28.92)	(10.71, 28.96)
	$ES_{0.99}$	13.81	(12.90, 19.41)	(12.05, 24.54)	(11.40, 28.13)	(10.86, 28.64)	(10.41, 28.71)
	$ES_{0.98}$	12.58	(12.35, 15.21)	(11.71, 17.78)	(11.12, 20.22)	(10.63, 22.58)	(10.22, 24.78)
	$VaR_{0.995}$	13.47	(12.14, 23.71)	(12.00, 27.91)	(11.23, 28.14)	(10.90, 28.18)	(10.38, 27.97)
	$VaR_{0.99}$	12.21	(11.08, 13.72)	(10.33, 21.77)	(10.04, 27.15)	(9.96, 27.75)	(9.88, 27.91)
	$VaR_{0.98}$	10.63	(9.80, 11.08)	(9.17, 12.26)	(8.68, 15.09)	(8.23, 19.64)	(7.82, 23.46)
Sc. II	$ES_{0.995}$	15.38	(14.22, 26.41)	(13.58, 28.76)	(12.88, 28.90)	(12.37, 28.97)	(11.94, 29.00)
	$ES_{0.99}$	14.39	(13.53, 20.11)	(12.86, 25.40)	(12.25, 28.52)	(11.80, 28.72)	(11.40, 28.78)
	$ES_{0.98}$	13.27	(12.97, 16.04)	(12.39, 18.76)	(11.87, 21.34)	(11.44, 23.82)	(11.08, 26.09)
	$VaR_{0.995}$	13.99	(13.05, 24.39)	(12.14, 28.01)	(11.39, 28.19)	(11.20, 28.20)	(11.02, 28.33)
	$VaR_{0.99}$	12.89	(12.02, 14.33)	(11.39, 23.03)	(11.19, 27.60)	(11.05, 27.92)	(10.85, 28.01)
	$VaR_{0.98}$	11.53	(10.84, 11.99)	(10.30, 13.14)	(9.85, 16.15)	(9.47, 20.93)	(9.12, 24.99)
Sc. III	$ES_{0.995}$	16.24	(15.02, 26.40)	(14.24, 28.73)	(13.56, 28.85)	(12.93, 28.93)	(12.32, 28.98)
	$ES_{0.99}$	15.05	(13.98, 20.44)	(13.20, 25.29)	(12.56, 28.25)	(11.99, 28.66)	(11.47, 28.74)
	$ES_{0.98}$	13.65	(13.19, 16.30)	(12.51, 18.83)	(11.92, 21.25)	(11.41, 23.51)	(10.96, 25.78)
	$VaR_{0.995}$	14.62	(13.54, 24.13)	(12.54, 28.04)	(11.93, 28.12)	(11.33, 28.24)	(10.96, 28.22)
	$VaR_{0.99}$	13.19	(12.18, 14.87)	(11.45, 22.65)	(11.29, 26.92)	(11.16, 27.92)	(10.67, 27.97)
	$VaR_{0.98}$	11.47	(10.72, 12.13)	(10.08, 13.39)	(9.56, 16.03)	(9.11, 20.44)	(8.67, 24.29)
Sc. IV	$ES_{0.995}$	16.95	(15.95, 27.01)	(15.29, 28.77)	(14.67, 28.90)	(14.15, 28.98)	(13.71, 29.01)
	$ES_{0.99}$	15.84	(14.89, 21.26)	(14.25, 26.10)	(13.64, 28.59)	(13.12, 28.74)	(12.71, 28.79)
	$ES_{0.98}$	14.54	(13.95, 17.25)	(13.36, 19.87)	(12.82, 22.39)	(12.35, 24.73)	(11.98, 26.84)
	$VaR_{0.995}$	15.47	(14.53, 25.13)	(13.93, 28.13)	(13.37, 28.23)	(12.89, 28.23)	(12.49, 28.36)
	$VaR_{0.99}$	14.15	(13.23, 15.83)	(12.62, 23.84)	(12.61, 27.84)	(12.16, 27.97)	(11.67, 27.98)
	$VaR_{0.98}$	12.49	(11.76, 13.14)	(11.24, 14.33)	(10.78, 17.07)	(10.37, 21.92)	(9.98, 25.76)

Table 7.6: The bounds on VaR and ES for various scenarios and different levels of trustworthiness when using CIMDO with student's t prior.

The loss outcomes from the student's t CIMDO, when there is no uncertainty, are greater in the tails than the t-Copula. Unfortunately, the fast convergence of the upper bounds remain unchanged. On the other hand, the bounds have been reduced. Take for instance $VaR_{0.98}$ in Scenario IV. The ratio of the upper bound when $p_{\mathcal{T}} = 0.98$ and no model uncertainty is 2.3% smaller for the CIMDO in comparison to the t-Copula. Another examples are $ES_{0.98}$ in Scenario II and $VaR_{0.98}$ in Scenario I, the ratios of the upper bound when $p_{\mathcal{T}} = 0.99$ and no model uncertainty are 19% and 4.8% respectively. This ratio will be greater as the confidence level increases and/or when $p_{\mathcal{T}}$ decreases. However, this has presumably its explanation in the fact that the student's t CIMDO generates larger losses during no model uncertainty, and not that it is performing remarkably better. The extreme convergence as $p_{\mathcal{T}}$ decreases is still present, which is an indication of poor performance. Hence, no distinct conclusions can be established.

The overall results from this section should be treated cautiously. It should rather be viewed as an intimation that both the copula approach and CIMDO are in solitude presumably insufficient to reliably forecast future losses.

Moreover, the results are demonstrating that the Gaussian copula is more accurate in the tails than the t-Copula. However, this is not indicating that the Gaussian copula is making *correct* predictions as such statement is in direct contradiction to the well-known fact that the model underestimates the extreme losses. Thus, the whole model risk framework should not be utilized for comparing distributions with different characteristics, but rather within the same distribution where additional information has been added or further interventions has been carried out.

Chapter 8

Conclusions

By simulating artificial samples it is shown that the entropy model reduces the variances of the regression coefficients when compared to traditional statistical methods such as the OLS and Maximum Likelihood. In other words, in environment of small samples the risk of overfitting will be remarkably lower when using the entropy model. On the other hand, due to the bias-variance trade-off the biases become worst in the entropy model. However, no attempts were made in this thesis to optimize the outcomes of the entropy model and hence the performance could plausibly be improved further.

Depending on the initial distributional guess (prior) there could be a significant difference between the prior and posterior in the CIMDO methodology. For multivariate normal priors the differences are negligible, whereas using a student's t causes the posterior to have fatter right tail and thinner left tail than the prior.

The overall conclusion from the risk analysis framework is that the uncertainty of the simulated losses become more extensive as one moves closer to the extreme values of the distribution. The student's t CIMDO appears to reduce the uncertainty in comparison to the t-Copula, however the results are by no means articulate. Instead, the conclusion is that the CIMDO is rather a commencement towards modeling credit risk without unsubstantial assumptions, and not in solitude sufficient for estimating the potential credit losses. No model risk reduction were found when the normal CIMDO was compared to the Gaussian copula.

Bibliography

[1] Taehan Bae and Ian Iscoe. "Sum of Bernoulli Mixtures: Beyond Conditional Independence". In: *Journal of Probability and Statistics* (2014).

[2] Miguel Angel Segoviano Basurto. "Conditional Probability of Default Methodology". In: *Financial Markets Group, London School of Economics and Political Science* Discussion paper, 558 (2006).

[3] Miguel Angel Segoviano Basurto. "Consistent Information Multivariate Density Optimizing Methodology". In: *Financial Markets Group, London School of Economics and Political Science* 322.Discussion paper, 557 (2006).

[4] Carole Bernard and Steven Vanduffel. "A new approach to assessing model risk in high dimensions". In: *Journal of Banking* & *Finance* 58 (2015), pp. 166–178.

[5] Christian Bluhm, Ludger Overbeck, and Christoph Wagner. *An Introduction to Credit Risk Modeling*. Chapman & Hall/CRC Financial Mathematics Series. CRC Press, 2002. ISBN: 9781420057362.

[6] Pieter-Tjerk De Boer et al. "A Tutorial on the Cross-Entropy Method". In: *Springer Science + Business Media, Inc.* 134 (2005), pp. 19–67.

[7] Thomas Breuer et al. "A systematic approach to multi-period stress testing of portfolio credit risk". In: *Journal of Banking & Finance* 36 (2012), pp. 332–340.

[8] John A. Burns. *Introduction to the Calculus of Variations and Control with Modern Applications*. Chapman & Hall/CRC Applied Mathematics & Nonlinear Science Series. CRC press, 2013. ISBN: 978-1466571396.

[9] Carlos Castro. "Asset correlation for portfolio credit risk: Sector, region and structured product estimates". In: *ECARES, Universite Libre de Bruxelles* (2008).

[10] Jorge A. Chan-Lau. "Fundamentals-Based Estimation of Default Probabilities: A Survey". In: *IMF working papers* 149 (2006).

[11] Jorge A. Chan-Lau. "Market-Based Estimation of Default Probabilities and Its Application to Financial Market Surveillance". In: *IMF Working Papers* 104 (2006).

[12] Peter Crosbie. "Modeling Default Risk". In: *Moody's KMV Company* (2003).

[13] Michel Crouhy, Dan Galai, and Robert Mark. "A Comparative Analysis of Current Credit Risk Models". In: *Journal of Banking & Finance* 24.10 (2000), pp. 59–117.

[14] Paul Demey et al. "Maximum Likelihood Estimate of Default Correlations". In: *Risk Magazine* (2004), pp. 104–108.

[15] Michel Denuit, Anna Kiriliouk, and Johan Segers. "Max-factor individual risk models with application to credit portfolios". In: *Insurance: Mathematics and Economics* 62 (2015), pp. 162–172.

[16] Jan Dhaene et al. "The Concept of Comonotonicity in Actuarial Science and Finance: Theory". In: *Insurance : Mathematics & Economics* 31.1 (2002), pp. 3–33.

[17] Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf. "Model uncertainty and VaR aggregation". In: *Journal of Banking and* & *Finance* 37 (8 2013), pp. 2750–2764.

[18] John Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, Inc., 2008. ISBN: 978-0761930426.

[19] Rüdiger Frey and Alexander J. McNeil. "Dependent Defaults in Models of Portfolio Credit Risk". In: *Journal of Risk* (2003).

[20] Rüdiger Frey and Alexander J. McNeil. "Modelling Dependent Defaults: Asset Correlations Are Not Enough!" In: *ETH E-Collection, Department of Mathematics, ETH Zurich* (2001).

[21] James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori. *Handbook of Computational Statistics*. Springer Handbooks of Computational Statistics. Springer, 2012. ISBN: 978-3-642-21550-6.

[22] Amos Golan, George G. Judge, and Douglas Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Series In Financial Economics and Quantitative Analysis. Wiley, 1996. ISBN: 978-0-471-95311-1.

[23] Michael B. Gordy and Eric Heitfield. "Small-Sample Estimation of Models of Credit Risk". In: *Recent Advances in Financial Engineering: Proceedings of the Kier-Tmu International Workshop on Financial Engineering* (2009), pp. 44–63.

[24] Takashi Hashimoto. "Asset correlation for credit risk analysis - Empirical study of default data for Japanese companies". In: *Bank of Japan Working Paper Series* (2011).

[25] Markus Höchstötter and Abdolreza Nazemi. "Analysis of loss given default". In: *Investment Management and Financial Innovations* 10 (4 2013).

[26] Xinzheng Huang and Cornelis W. Oosterlee. “Generalized beta regression models for random loss given default”. In: *The Journal of Credit Risk* 7.4 (2011), pp. 1–26.

[27] Petr Jakubík. “Stress Testing Credit Risk: Comparison of the Czech Republic and Germany”. In: *Bank for International Settlements* (2008).

[28] Paul Penfield Jr. “Information and Entropy, lecture notes”. In: *Massachusetts Institute of Technology* (2008).

[29] Robert Lamb and William Perraudin. “Dynamic Default Rates”. In: *Imperial College Working Papers* (2008).

[30] Wolfgang von der Linden et al., eds. *Maximum Entropy and Bayesian Methods Garching, Germany 1998*. Fundamental Theories of Physics. Springer Science + Business Media, B.V., 1998. ISBN: 978-94-010-5982-4.

[31] Filip Lindskog et al. *Risk and Portfolio Analysis- Principles and Methods*. Springer Series in Operations Research and Financial Engineering. Springer, 2012. ISBN: 978-1-4614-4102-1.

[32] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005. ISBN: 0-691-12255-5.

[33] Robert C. Merton. “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”. In: *The Journal of Finance* 29.2 (1974), pp. 449–470.

[34] Roger M. Stein. “The role of stress testing in credit risk management”. In: *Journal Of Investment Management* 10.4 (2012), pp. 64–90.

[35] Sentendra Tiwari. “Enhancements in CreditRisk+ Model”. In: *Journal of Financial Risk Management* 7.3 (2010), pp. 25–46.

[36] Kimmo Virolainen. “Macro stress testing with a macroeconomic credit risk model for Finland”. In: *Bank of Finland: Discussion Papers* (2004).

[37] Thomas C. Wilson. “Portfolio Credit Risk I”. In: *Risk* 10.9 (1997), pp. 111–117.

[38] Thomas C. Wilson. “Portfolio Credit Risk II”. In: *Risk* 10.10 (1997), pp. 56–61.

[39] Jim Hock-Yuen Wong, Ka-Fai Choi, and Pak-Wing Fong. “A framework for stress-testing banks’ credit risk”. In: *The Journal of Risk Model Validation* 2.1 (2008), pp. 3–23.

Appendices

Appendix A

Nontechnical Comments

A.1 Pros and Cons of Econometric Models

The advantages of the econometric models are that macroeconomic variables are numerous, easily accessible, well understood and recorded over a long period of time. Due to their structure, the econometric models are well suited for stress-testing[19] on regional, national and global level. Moreover, these models do usually not require complex mathematical tools for calibration.

On the other hand, to capture the whole picture data must be accumulated during at least one business cycle, which is a problem since it is not unusual that the length of one cycle could exceed ten years.

Furthermore, most macro variables are highly correlated, which in the context of regression analysis are referred to as multicollinearity. The phenomenon causes the standard errors of the coefficients to become larger and the impact of a single exogenous variable on the response variable, ceteris paribus, to be inaccurate.

Many macro indicators are updated each quarter at best. Consequently, the PD's could only be updated with the same time period if one solely chooses these kind of variables. This is a disadvantage if the defaults in a certain sector or geographical area fluctuates heavily for a relatively short period of time. In such situations the econometric models are unlikely to predict these deviations. In this regard the marked-based models are having an advantage since the market is more likely to instantaneously respond to the fluctuations, or perhaps even react ahead of them.

According to Chan-Lau [10], another major drawback is the so-called Lucas critique. Changes in government policy rules and the fact that every business cycle is unlike the other, are making econometric models improbable of remaining trustworthy over time. In other words, although one estimates the regression coefficients with data collected over the past ten years, it is not a guarantee that the calibrated model will show correct predictions over

[19] See Appendix A.2 for definition and details.

the next ten years.

A.2 Stress-Testing

There is no formal definition of what stress-testing implies. Nonetheless, the consensus is that it is the process of purposely inserting extreme values into a system with the intention to examine the response. Stress-testing is used in all the financial risk categorizations. For example in market risk, factors such as yield curves, currencies and volatility are tested since these parameters are instantaneously affecting the price process of the underlying.

Within credit risk applications, the concept is better known as *macro* stress-testing. The added term refers to exposing the portfolio to abnormal but plausible macroeconomic scenarios. The objective is to study how the loss distribution reshapes by various levels of negative shocks. In contrast to market stress-testing, the procedure is indirect in the sense that the probability of default is first affected, thereby the losses. Macro stress-testing should be viewed as a complimentary tool to the risk measures such as VaR. For instance the modeller could examine how the EC increases at different levels of financial distress. Thus, the combination of both yields a reliable quantitative instrument for risk management.

The difficulty lies in the very design of the adverse macroeconomic scenarios. The reasons are many. For example, is the invented scenario adverse enough and at the same time plausible to happen? And, what macro variables should be stressed?

Depending on situation, the second question could have a straightforward answer. For example, the choice would naturally fall on oil price if the portfolio exclusively consisted of companies heavily exposed to the transport industry. However, scenario designing becomes more complicated as the portfolio gets more diversified. Even if some candidates are selected for a specific portfolio it is difficult to sort them from worst scenario to best ex ante. Furthermore, financial distress affects different portfolios to different degrees, making the designing of universal stress tests troublesome ex ante, according to Stein [34].

Multicolinearity, the phenomenon of macro variables being highly correlated, also causes uncertainty. This is due to the difficulty of predicting how selected macro variables will interact in reality during abnormal conditions. Misspecified scenarios could lead to incorrect forecasts.

In addition, managers and authorities usually want to formalize such adverse scenarios where it is unlikely that the available data contains similar situations. Since the calibration is on less painful situations there will be uncertainties on whether the model will respond correctly, and hence the risk of results being inaccurate is huge.

A.3 Extensions of the Vasicek Model

The literature concerning credit risk has modified the Vasicek model without changing its premise. Among others, Castro [9] extends the one-factor model into a K-factor model where a global risk component affecting all borrowers has been introduced along with multi-index risk factors, allowing the modeller to incorporate several systematic risk components. In the dynamic factor model by Lamb and Perraudin [29] the initial assumption of standard normal risk factors is substituted into a model where the systematic factor is an AR(1)-process.

The unanimous argument for all the extended versions is that the original Vasicek model leads to erroneous perception of true correlation and does not correctly imitate the actual sequence of asset movements. Nevertheless, these models induce calibration difficulties and sometimes also cause inaccurate results. The problems occur since as the models become more advanced, the number of parameters to be estimated grows rapidly, but the amount of data remains unchanged [14]. Furthermore, due to absence of real data it is impossible to test these different versions, hence this thesis will maintain to the formulations above.

Empirical evidence are showing asset processes are closer to being student's t-distributed than normal. There is a simple way to compensate for this without significantly changing the equations in the previous section. Let W be an independent Wald-distributed random variable with location and scale parameter equal to $\nu/2$ and let A_{it} be defined as in the previous section. Then $Z_{it} = \sqrt{W}A_{it}$ is in fact a student's t-distributed asset return with ν degrees of freedom. Using this notation, the correlation between borrowers remains the same as when the asset returns are normally distributed.

A.4 Short Comment on LGD and EAD Estimation

There exists substantial evidence on the connection between the LGD and the economic environment, see Höchstötter and Nazemi [25]. However, this is not necessarily concordant with the PD's correlation to buisness cycle. For example, if the collateral of the loan is a commodity with a volatile secondary market, it is doubtful whether adverse periods coincide with financial distress. However, the losses will be enormous if the LGD is high during a period of multiple defaults, which is why proper LGD estimation is vital.

Quantitative research on the EAD is less extensive. Some researchers claim that the commitment part of EAD is linked to external factors such as balance sheets and market indicators. Nevertheless, modeling of LGD and EAD was outside the scope of the thesis.

Appendix B

Mathematical Details

B.1 Maximum Likelihood for Asset Correlation

There exist two alternatives for estimating the asset correlation in the Vasicek factor model, namely maximum likelihood (ML) and method of moments (MM). Gordy and Heitfield [23] conclude in their article that in environment of small samples, the ML is more accurate than the MM as the latter lacks of downward bias. If the asset returns are assumed to be normally distributed, then according to Hashimoto [24] the general ML equation for obtaining the asset correlations is

$$\mathcal{ML} = \prod_{t=0}^{T} \int_{-\infty}^{\infty} \prod_{r=1}^{R} \binom{n_t^r}{d_t^r} \Big(p_t^r(s)\Big)^{d_t^r} \Big(1 - p_t^r(s)\Big)^{n_t^r - d_t^r} \phi(s) ds, \tag{B.1}$$

where t is each time point in which data has been collected, d_t^r is the number of defaults in rating r at time t and n_t^r is the total number of borrowers in rating r at time t. $p_t^r(s)$ is the conditional default probability defined in Equation (2.8) and R is the total number of rating classes. The proof of Equation (B.1) is given in the subsequent section.

It may appear most logical to estimate the asset correlation for each market sector considered. This is justified by the argument that companies belonging to the same sector should be affected equally by the systematic risk. On the other hand, the financial performance of larger companies are stronger correlated to the business cycle than smaller companies, which in turn are more vulnerable to individual risk factors, according to Hashimoto [24]. Since larger companies have in general higher credit rating it is commonly accepted to estimate the asset correlation for each rating class separately. Furthermore, the borrowers have already been categorized into homogeneous groups by their the rating, causing the estimation of asset correlation for each market sector to be complicated.

B.1.1 Proof of Maximum Likelihood

The conditional PD of rating r at time t is given by

$$p_t^r(s) = \Phi\left(\frac{\Phi^{-1}(p_t^r) - s\sqrt{\rho^r}}{\sqrt{1-\rho^r}}\right) \tag{B.2}$$

Obligors are independent if mixture representation is considered and the common factors are realized. Hence, the probability of d_t^r defaults at rating r and time t is given by a binomial distribution

$$\Pr\left[d_t^r \mid \text{com. fact.}\right] = \binom{n_t^r}{d_t^r}\Big(p_t^r(s)\Big)^{d_t^r}\Big(1 - p_t^r(s)\Big)^{n_t^r - d_t^r}. \tag{B.3}$$

Obligors are not only independent within same rating, but also across all the R classes. Hence the probability of total amount of defaults d_t at time t is

$$\Pr\left[d_t \mid \text{com. fact.}\right] = \prod_{i=1}^{R}\binom{n_t^r}{d_t^r}\Big(p_t^r(s)\Big)^{d_t^r}\Big(1 - p_t^r(s)\Big)^{n_t^r - d_t^r}. \tag{B.4}$$

Integrating out the systematic risk and multiplying over all time points yields the unconditional likelihood function

$$\mathcal{ML} = \prod_{t=0}^{T}\int_{-\infty}^{\infty}\prod_{r=1}^{R}\binom{n_t^r}{d_t^r}\Big(p_t^r(s)\Big)^{d_t^r}\Big(1 - p_t^r(s)\Big)^{n_t^r - d_t^r}\phi(s)ds. \tag{B.5}$$

B.2 General Solution of Maximum Entropy

Suppose there exists a random variable X with $x_1, \ldots, x_n$ outcomes. Let $\mathbf{p} = (p_1, \ldots, p_n)$ contain the probabilities of each outcome. Define H as the entropy of X, i.e. $H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \ln p_i$. Furthermore, consider a finite set of functions $f_k(X), 1 \le k \le K < n$ such that

$$f_k(x_i) = C_k, \quad 1 \le k \le K < \infty \tag{B.6}$$

where the C_k's are fixed constants. The entropy measure subject to the constrains yields the Lagrangian

$$\mathcal{L} = H(\mathbf{p}) - \sum_{k=1}^{K}\lambda_k\Big(\sum_{i=1}^{n} f_k(x_i)p_i - C_k\Big) - \underbrace{\mu}_{\lambda_0\text{-}1}\sum_{i=0}^{n} p_i. \tag{B.7}$$

Differentiating the Lagrangian with respect to p_i and set equal zero to find the maximum

$$\frac{\partial\mathcal{L}}{\partial p_i} = 0. \quad \text{for } i = 1, \ldots, n \tag{B.8}$$

Thus we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p_i} &= -\ln[p_i] - 1 - \sum_{k=1}^{K} \lambda_k f_k(x_i) - (\lambda_0 - 1) = 0 \Rightarrow \\ p_i &= \exp\Big(- \lambda_0 - \sum_{k=1}^{K} \lambda_k f_k(x_i)\Big). \quad \text{for } i = 1, \dots, n \end{aligned} \tag{B.9}$$

By definition $\sum_i^n p_i = 1$. Inserting Equation (B.9) into this statement yields

$$\begin{aligned} 1 = \sum_{i=1}^{n} p_i &= \sum_{i=1}^{n} \exp\Big(- \lambda_0 - \sum_{k=1}^{K} \lambda_k f_k(x_i)\Big) \Rightarrow \\ \exp(-\lambda_0) &= \frac{1}{\sum_{i=1}^{n} \exp\Big(- \sum_{k=1}^{K} \lambda_k f_k(x_i)\Big)}. \end{aligned} \tag{B.10}$$

Thus, the solution is given by

$$p_i = \frac{\exp\Big(- \sum_{k=1}^{K} \lambda_k f_k(x_i)\Big)}{\sum_{i=1}^{n} \exp\Big(- \sum_{k=1}^{K} \lambda_k f_k(x_i)\Big)}. \tag{B.11}$$

B.3 Obtaining the Posterior by Calculus of Variations

The Lagrangian $\mathcal{L}$ was stated as

$$\begin{aligned} \mathcal{L}(p, q) = \int_{\Omega} p(\mathbf{a}) \ln \Big[\frac{p(\mathbf{a})}{q(\mathbf{a})}\Big] d^r\mathbf{a} &+ \lambda_0 \Bigg[\int_{\Omega} p(\mathbf{a}) d^r\mathbf{a} - 1 \Bigg] \\ &+ \lambda_1 \Bigg[\int_{\Omega} p(\mathbf{a}) \mathbb{1}\{a_1 < \gamma_t^1\} d^r\mathbf{a} - p_t^1 \Bigg] \\ &\vdots \\ &+ \lambda_r \Bigg[\int_{\Omega} p(\mathbf{a}) \mathbb{1}\{a_r < \gamma_t^r\} d^r\mathbf{a} - p_t^r \Bigg], \end{aligned} \tag{B.12}$$

where $\lambda_0, \dots, \lambda_r$ are the Lagrangian multipliers. We will now use the definition of calculus of variations to obtain the optimal solution of the posterior distribution. In the context of calculus of variations, p and q are viewed as variables rather than functions. Furthermore, since the prior is known and fixed the variation will only be implemented on p. Formally we have

$$\mathcal{L}(p + \epsilon f, q), \tag{B.13}$$

where ϵ is a small quantity (approximately zero) and $f(\mathbf{a})$ is an arbitrary continuous function which fulfills the properties of being at least once differentiable and bounded with value zero at the endpoints of the integral in question. Applying the variations in Equation (B.12) yields

$$\begin{aligned}\mathcal{L}(p+\epsilon f, q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big(p(\mathbf{a}) + \epsilon f(\mathbf{a})\Big) \ln\left[\frac{p(\mathbf{a}) + \epsilon f(\mathbf{a})}{q(\mathbf{a})}\right] d^r\mathbf{a} \\ + \lambda_0 \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big(p(\mathbf{a}) + \epsilon f(\mathbf{a})\Big) d^r\mathbf{a} - 1\right] \\ + \lambda_1 \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big(p(\mathbf{a}) + \epsilon f(\mathbf{a})\Big) \mathbb{1}\{a_1 < \gamma_t^1\} d^r\mathbf{a} - p_t^1\right] \\ \vdots \\ + \lambda_r \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big(p(\mathbf{a}) + \epsilon f(\mathbf{a})\Big) \mathbb{1}\{a_r < \gamma_t^r\} d^r\mathbf{a} - p_t^r\right].\end{aligned} \tag{B.14}$$

Taking the derivative of $\mathcal{L}$ with respect to ϵ yields

$$\begin{aligned}&\frac{d\mathcal{L}(p+\epsilon f, q)}{d\epsilon} = \\ &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[f(\mathbf{a}) \ln\left[\frac{p(\mathbf{a}) + \epsilon f(\mathbf{a})}{q(\mathbf{a})}\right] + \Big(p(\mathbf{a}) + \epsilon f(\mathbf{a})\Big)\left(\frac{q(\mathbf{a}) f(\mathbf{a})/q(\mathbf{a})}{p(\mathbf{a}) + \epsilon f(\mathbf{a})}\right)\right] d^r\mathbf{a} \\ &+ \lambda_0 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{a}) d^r\mathbf{a} \\ &+ \lambda_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{a}) \mathbb{1}\{a_1 < \gamma_t^1\} d^r\mathbf{a} \\ &\vdots \\ &+ \lambda_r \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{a}) \mathbb{1}\{a_r < \gamma_t^r\} d^r\mathbf{a} \\ &\iff \\ &\frac{d\mathcal{L}(p+\epsilon f, q)}{d\epsilon} = \\ &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\ln\left[\frac{p(\mathbf{a} + \epsilon f(\mathbf{a})}{q(\mathbf{a})}\right] + 1 + \lambda_0 + \lambda_1 \mathbb{1}\{a_1 < \gamma_t^1\} + \dots \right. \\ &\left. + \lambda_r \mathbb{1}\{a_r < \gamma_t^r\}\right] f(\mathbf{a}) d^r\mathbf{a}.\end{aligned} \tag{B.15}$$

Since the interest is in minimizing $\mathcal{L}$, the derivative in Equation (B.15) is set to zero. Using the fundamental lemma of calculus of variations[20] and evaluating for $\epsilon \to 0$ yields

$$
\begin{aligned}
&\ln\left[\frac{p(\mathbf{a})}{q(\mathbf{a})}\right] + 1 + \lambda_0 + \lambda_1 \mathbb{1}\{a_1 < \gamma_t^1\} + \cdots + \lambda_r \mathbb{1}\{a_r < \gamma_t^r\} = 0 \Rightarrow \\
&p(\mathbf{a}) = q(\mathbf{a}) \exp\Big[-1 - \lambda_0 - \lambda_1 \mathbb{1}\{a_1 < \gamma_t^1\} - \cdots - \lambda_r \mathbb{1}\{a_r < \gamma_t^r\}\Big].
\end{aligned}
\tag{B.16}
$$

B.4 Application to Credit Portfolios

Suppose a six dimensional distribution has been fitted from which eight samples have been generated. Furthermore, assume that it has already been determined whether the outcomes belong to $\mathcal{T}$ or $\mathcal{U}$. In Matrix (B.17) below the samples are shown. They are not sorted by any particular rule, however outcomes belonging to $\mathcal{U}$ are highlighted in red.

$$
\begin{pmatrix}
-4.77 & -1.54 & -9.33 & -9.18 & -11.30 & 1.12 \\
-6.54 & -1.47 & -6.68 & 7.22 & 11.11 & -3.18 \\
-0.70 & -4.53 & -5.81 & -4.88 & 10.29 & 0.51 \\
-6.58 & 9.72 & 2.47 & -9.95 & -6.31 & 4.30 \\
-7.90 & 11.51 & 5.07 & -5.70 & -0.99 & -2.51 \\
-6.47 & 10.16 & -2.19 & -4.35 & 5.53 & -6.44 \\
8.26 & -1.68 & 2.28 & -1.82 & -0.27 & -0.27 \\
-7.33 & -7.56 & -5.71 & 0.19 & 1.88 & 2.98
\end{pmatrix}
\tag{B.17}
$$

For credit portfolios, the manager is interested in the losses due to defaults of the conterparties. If the numbers in Matrix (B.17) are viewed as asset outcomes, then the first step is determining which borrowers have defaulted. Each separate outcome is compared to the threshold value. If the borrower defaulted (asset falls below the threshold) the asset outcome is replaced by the product of the exposure and loss given default. If the borrower did not default, the outcome is replaced by 0.

For simplicity, the threshold value is set to -3 for all six borrowers. The Matrix (B.18) below shows how Matrix (B.17) is reformulated after determining whether a borrower defaulted or not. The losses are randomly assigned.

[20] In general , if g is the set of continuous functions fulfilling $\{g : g(x) \in C^k[a,b],\ g(a) = g(b) = 0\}$, and $\int_a^b f(x)g(x) = 0$, then $f(x) = 0$ on $[a,b]$.

$$\begin{pmatrix} 0.16 & 0 & 0.21 & 0.44 & 0.15 & 0 \\ 0.13 & 0 & 0.27 & 0 & 0 & 0.72 \\ 0 & 0.22 & 0.26 & 0.34 & 0 & 0 \\ 0.08 & 0 & 0 & 0.31 & 0.19 & 0 \\ 0.14 & 0 & 0 & 0.45 & 0 & 0 \\ 0.15 & 0 & 0 & 0.4 & 0 & 0.77 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.12 & 0.3 & 0.1 & 0 & 0 & 0 \end{pmatrix} \tag{B.18}$$

$\mathbf{V}$, $\mathbf{S}_{\mathcal{T}}$ and $\mathbf{S}_{\mathcal{U}}$ are now obtained as

$$\mathbf{V} = \begin{pmatrix} 0.15 & 0 & 0 & 0.4 & 0 & 0.77 \\ 0.16 & 0 & 0.21 & 0.44 & 0.15 & 0 \\ 0 & 0.22 & 0.26 & 0.34 & 0 & 0 \\ 0.12 & 0.3 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.14 & 0 & 0.27 & 0.45 & 0.19 & 0.72 \\ 0.13 & 0 & 0 & 0.31 & 0 & 0 \\ 0.08 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{matrix} \mathbf{S}_{\mathcal{T}} = \begin{pmatrix} 1.32 \\ 0.96 \\ 0.82 \\ 0.52 \\ 0 \end{pmatrix} \\ \mathbf{S}_{\mathcal{U}} = \begin{pmatrix} 1.77 \\ 0.44 \\ 0.08 \end{pmatrix} \end{matrix} \tag{B.19}$$

The rows in the trusted area has been sorted by their total row sums. For the highlighted part of $\mathbf{V}$, each column has been sorted in descending order. Thereafter the row sums have been calculated and inserted into $\mathbf{S}_{\mathcal{U}}$.

Appendix C

Figures and Tables

C.1 Entropy Model vs Maximum Likelihood

The variance, bias and mean square error for the entropy model and ML are presented in Table C.1 below. Note that the values for the entropy model are somewhat different from those in Table 7.2. The reason is that two separate simulations were performed.

The results are similar to those in Table 7.2, however it is inevitable that the ML seems to perform better than the OLS. Nevertheless, the entropy model reduces the MSE in all cases except for three coefficients.

Rating		1	2	3	4	5	6
Var.	$\beta^{\text{Int.}}_{\text{ML}}$	$1.5\cdot10^{-3}$	$1.3\cdot10^{-3}$	$1.4\cdot10^{-3}$	$1.7\cdot10^{-3}$	$1.9\cdot10^{-3}$	$2.2\cdot10^{-3}$
	$\beta^{\text{GDP}}_{\text{ML}}$	1.6589	1.4550	1.4895	1.8719	1.9601	2.1085
	$\beta^{\text{Nikk.}}_{\text{ML}}$	$1.53\cdot10^{-2}$	$1.39\cdot10^{-2}$	$1.46\cdot10^{-2}$	$1.19\cdot10^{-2}$	$2.01\cdot10^{-2}$	$2.15\cdot10^{-2}$
	$\beta^{\text{Int.}}_{\text{ent.}}$	$1.6\cdot10^{-3}$	$1.3\cdot10^{-3}$	$1.3\cdot10^{-3}$	$1.7\cdot10^{-3}$	$1.9\cdot10^{-3}$	$2.5\cdot10^{-3}$
	$\beta^{\text{GDP}}_{\text{ent.}}$	1.0521	0.7387	0.7195	0.9337	0.8965	0.9475
	$\beta^{\text{Nikk.}}_{\text{ent.}}$	$7.6\cdot10^{-3}$	$5.4\cdot10^{-3}$	$5.4\cdot10^{-3}$	$7.2\cdot10^{-3}$	$7.1\cdot10^{-3}$	$7.4\cdot10^{-3}$
	$\Delta_{\text{Int.}}$	-7.29%	1.05%	1.21%	-0.54%	-0.9%	-14.35%
	Δ_{GDP}	57.68%	96.96%	107.02%	100.48%	118.63%	122.54%
	$\Delta_{\text{Nikk.}}$	101.31%	158.84%	169.32%	164.12%	183.75%	189.26%
Bias sqr.	$\beta^{\text{Int.}}_{\text{ML}}$	$3.47\cdot10^{-6}$	$2.26\cdot10^{-6}$	$1.63\cdot10^{-6}$	$2.39\cdot10^{-6}$	$1.85\cdot10^{-6}$	$1.42\cdot10^{-7}$
	$\beta^{\text{GDP}}_{\text{ML}}$	$1.39\cdot10^{-4}$	$3.18\cdot10^{-4}$	$9.63\cdot10^{-5}$	$1.46\cdot10^{-4}$	$2.29\cdot10^{-4}$	$3.09\cdot10^{-4}$
	$\beta^{\text{Nikk.}}_{\text{ML}}$	$1.50\cdot10^{-6}$	$7.83\cdot10^{-7}$	$2.47\cdot10^{-6}$	$2.62\cdot10^{-7}$	$6.46\cdot10^{-9}$	$2.05\cdot10^{-8}$
	$\beta^{\text{Int.}}_{\text{ent.}}$	$9.17\cdot10^{-6}$	$2.62\cdot10^{-7}$	$9.78\cdot10^{-7}$	$1.77\cdot10^{-5}$	$1.02\cdot10^{-4}$	$1.00\cdot10^{-3}$
	$\beta^{\text{GDP}}_{\text{ent.}}$	$6.92\cdot10^{-2}$	0.210	0.339	0.510	0.617	0.534
	$\beta^{\text{Nikk.}}_{\text{ent.}}$	$4.50\cdot10^{-3}$	$2.42\cdot10^{-3}$	$1.24\cdot10^{-3}$	$5.28\cdot10^{-4}$	$1.75\cdot10^{-4}$	$7.69\cdot10^{-6}$
MSE	$\beta^{\text{Int.}}_{\text{ML}}$	$1.48\cdot10^{-3}$	$1.31\cdot10^{-3}$	$1.37\cdot10^{-3}$	$1.73\cdot10^{-3}$	$1.89\cdot10^{-3}$	$2.16\cdot10^{-3}$
	$\beta^{\text{GDP}}_{\text{ML}}$	1.659	1.455	1.490	1.872	1.960	2.109
	$\beta^{\text{Nikk.}}_{\text{ML}}$	$1.53\cdot10^{-2}$	$1.39\cdot10^{-2}$	$1.46\cdot10^{-2}$	$1.90\cdot10^{-3}$	$2.01\cdot10^{-2}$	$2.15\cdot10^{-2}$
	$\beta^{\text{Int.}}_{\text{ML}}$	$1.60\cdot10^{-3}$	$1.30\cdot10^{-3}$	$1.35\cdot10^{-3}$	$1.75\cdot10^{-3}$	$2.01\cdot10^{-3}$	$3.52\cdot10^{-3}$
	$\beta^{\text{GDP}}_{\text{ML}}$	1.121	0.945	1.058	1.444	1.513	1.482
	$\beta^{\text{Nikk.}}_{\text{ML}}$	$1.21\cdot10^{-2}$	$7.80\cdot10^{-3}$	$6.65\cdot10^{-3}$	$7.72\cdot10^{-3}$	$7.24\cdot10^{-3}$	$7.44\cdot10^{-3}$
	$\Delta_{\text{Int.}}$	-7.60%	1.21%	1.25%	-1.4%	-5.81%	-38.76%
	Δ_{GDP}	47.97%	54.07%	40.76%	29.63%	29.55%	42.32%
	$\Delta_{\text{Nikk.}}$	26.54%	78.46%	119.15%	146.06%	176.89%	188.96%

Table C.1: The variance, bias squared and the MSE of each coefficient obtained from the simulation.

C.2 Additional Plots for Comparing Prior and Posterior Distributions

The upper two graphs in Figure C.1 show the posterior density and the subtraction between posterior and prior, when the prior was normally distributed. Similarly, the lower two plots show the posterior and the subtraction between posterior and prior, in the case when the prior is Student's t distributed.

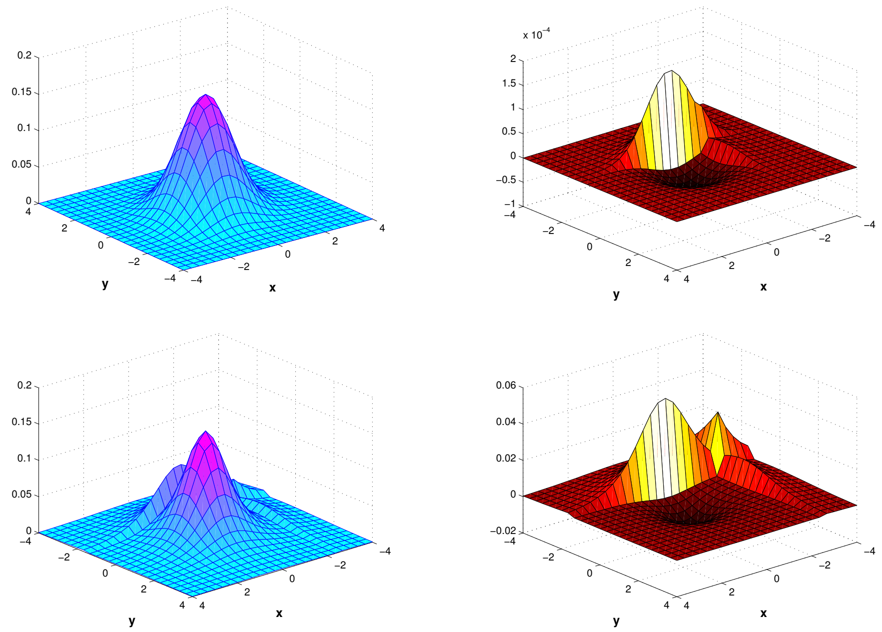


Figure C.1: All graphs correspond to the case when PD=(5%, 10%) and the correlation is 0.15. *Upper left*: The posterior density when prior is normally distributed. *Upper right*: Difference between posterior and prior, when prior is normally distributed. *Lower left*: The posterior density when prior is student's t distributed with $\nu = 3$. *Lower right*: Difference between posterior and prior, when prior is student's t distributed.

By looking at the upper right plot in Figure C.1 the difference between the posterior and prior seems small. In the scenario where the prior is student's t distributed however, it appears to be a significant difference. To visualize this further, approximately 10'000 samples were generated from both the posterior and prior distributions from the same scenario. Because of the complex structure of the posterior, rejection sampling was used to generate samples.

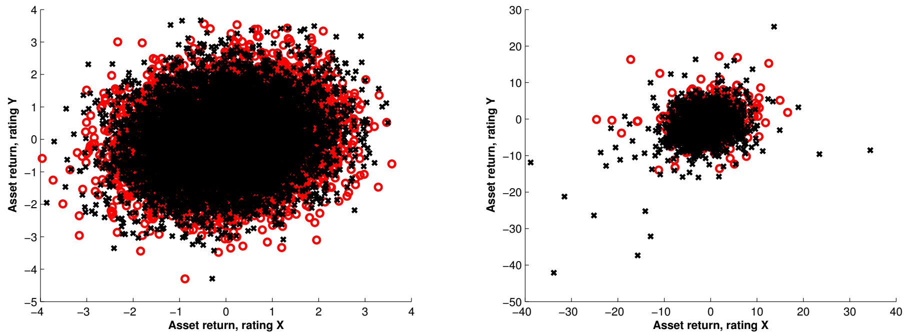


Figure C.2: All graphs correspond to the case when PD=(5%, 10%) and correlation is 0.15. Samples generated from the posterior is marked with × and samples from the prior is marked with o. *Left plot*: Posterior and prior correspond to the case when prior is normally distributed. *Right plot*: Posterior and prior corresponding to the case when prior is student's t distributed with $\nu = 3$.

Focusing on the right scatter plot in Figure C.2 one can perceive that the posterior distribution exhibits heavier left tail than the prior. In the left plot, where the prior is assumed to have normal distribution, the differences are not as obvious. Furthermore, the samples in the right plot are taking larger values than the outcomes in the right plot.

C.3 QQ-plots of Prior And Posterior for Different Scenarios

This section presents the difference between the prior and posterior in various scenarios. The qq-plots in Figure C.3 shows that there is no difference between the prior and posterior when the prior is selected to be multivariate normal. A prior as multivariate student's t will cause the posterior to have fatter left tail but thinner right tail in comparison to the prior. The Lagrangian multipliers are presented in Table C.2.

Correlation		0		0.15		0.5	
PD (X, Y)	LM	Normal	Student	Normal	Student	Normal	Student
(0.5%, 3%)	$\hat{\lambda}_0$	-1.00	-0.975	-1.00	-0.976	-1.00	-0.976
	$\hat{\lambda}_1$	0.022	-1.046	0.022	-0.911	0.022	-0.581
	$\hat{\lambda}_2$	0.002	-1.381	0.002	-1.376	0.001	-1.379
(5%, 10%)	$\hat{\lambda}_0$	-1.00	-0.906	-1.00	-0.910	-1.00	-0,919
	$\hat{\lambda}_1$	-0.001	-1.097	-0.001	-1.017	0.00	-0.812
	$\hat{\lambda}_2$	-0.003	-1.055	-0.003	-1.024	-0.003	-0.969

Table C.2: The Lagrangian multipliers for the different scenarios.

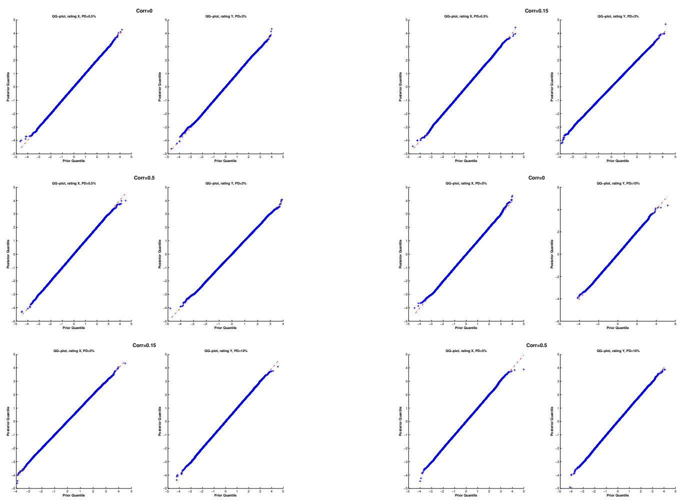


Figure C.3: QQ-plots of asset returns (X, Y) corresponding to the case when prior is standard normal. Left column is when $PD = (0.5\%, 3\%)$ and right column is when $PD = (5\%, 10\%)$. *Upper plots* is the case when the correlation is 0, *middle plots* for correlation=0.15 and *lower plots* for correlation=0.5.

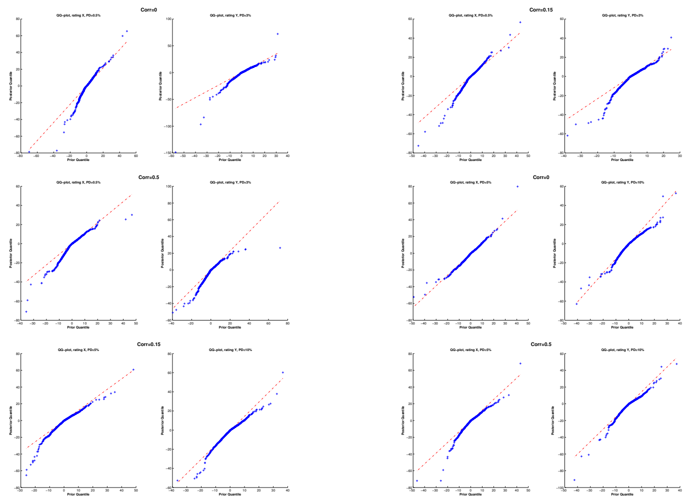


Figure C.4: QQ-plots of asset returns (X, Y) corresponding to the case when prior is student's t distributed with $\nu = 3$ degrees of freedom. Left column is when $PD = (0.5\%, 3\%)$ and right column is when $PD = (5\%, 10\%)$. *Upper plots* is the case when the correlation is 0, *middle plots* for correlation=0.15 and *lower plots* for correlation=0.5.

C.4 VaR and ES bounds for normal CIMDO

CIMDO, prior normal							
$p_{\mathcal{T}}$		1	0.99	0.98	0.97	0.96	0.95
Sc. I	$\mathrm{ES}_{0.995}$	4.57	(4.47, 8.10)	(4.41, 11.40)	(4.35, 14.45)	(4.31, 17.39)	(4.28, 20.00)
	$\mathrm{ES}_{0.99}$	4.27	(4.18, 6.03)	(4.12, 7.71)	(4.07, 9.28)	(4.03, 10.82)	(4.00, 12.30)
	$\mathrm{ES}_{0.98}$	3.93	(3.85, 4.81)	(3.81, 5.56)	(3.77, 6.44)	(3.73, 7.22)	(3.70, 7.80)
	$\mathrm{VaR}_{0.995}$	4.14	(4.06, 5.07)	(4.00, 8.83)	(3.96, 12.81)	(3.92, 16.17)	(3.86, 18.88)
	$\mathrm{VaR}_{0.99}$	3.80	(3.72, 4.11)	(3.68, 4.85)	(3.65, 6.52)	(3.61, 8.46)	(3.58, 10.38)
	$\mathrm{VaR}_{0.98}$	3.43	(3.38, 3.58)	(3.35, 3.77)	(3.31, 4.06)	(3.28, 4.50)	(3.25, 5.23)
Sc. II	$\mathrm{ES}_{0.995}$	5.63	(5.52, 9.90)	(5.44, 13.89)	(5.39, 17.54)	(5.34, 21.07)	(5.32, 23.75)
	$\mathrm{ES}_{0.99}$	5.27	(5.17, 7.43)	(5.11, 9.48)	(5.07, 11.39)	(5.02, 13.28)	(4.99, 15.11)
	$\mathrm{ES}_{0.98}$	4.89	(4.82, 5.97)	(4.76, 7.01)	(4.72, 7.99)	(4.69, 8.95)	(4.66, 9.90)
	$\mathrm{VaR}_{0.995}$	5.12	(5.05, 6.34)	(4.99, 11.60)	(4.93, 16.00)	(4.86, 19.95)	(4.79, 22.84)
	$\mathrm{VaR}_{0.99}$	4.74	(4.66, 5.12)	(4.62, 6.13)	(4.58, 8.30)	(4.55, 11.03)	(4.51, 13.37)
	$\mathrm{VaR}_{0.98}$	4.35	(4.29, 4.52)	(4.25, 4.76)	(4.22, 5.10)	(4.19, 5.72)	(4.16, 6.78)
Sc. III	$\mathrm{ES}_{0.995}$	5.38	(5.25, 8.93)	(5.17, 12.03)	(5.12, 15.02)	(5.08, 17.68)	(5.03, 20.28)
	$\mathrm{ES}_{0.99}$	4.95	(4.85, 6.76)	(4.78, 8.34)	(4.73, 9.89)	(4.69, 11.33)	(4.65, 12.75)
	$\mathrm{ES}_{0.98}$	4.51	(4.43, 5.41)	(4.37, 6.21)	(4.33, 7.00)	(4.29, 7.74)	(4.25, 8.47)
	$\mathrm{VaR}_{0.995}$	4.78	(4.71, 5.72)	(4.62, 9.20)	(4.58, 13.08)	(4.54, 16.21)	(4.50, 19.02)
	$\mathrm{VaR}_{0.99}$	4.32	(4.25, 4.73)	(4.21, 5.31)	(4.16, 6.75)	(4.13, 8.63)	(4.10, 10.58)
	$\mathrm{VaR}_{0.98}$	3.87	(3.80, 4.04)	(3.77, 4.26)	(3.72, 4.53)	(3.69, 4.93)	(3.66, 5.52)
Sc. IV	$\mathrm{ES}_{0.995}$	6.66	(6.56, 10.86)	(6.49, 14.54)	(6.42, 18.10)	(6.36, 21.41)	(6.32, 23.87)
	$\mathrm{ES}_{0.99}$	6.16	(6.07, 8.31)	(6.01, 10.20)	(5.94, 12.07)	(5.90, 13.84)	(5.85, 15.55)
	$\mathrm{ES}_{0.98}$	5.65	(5.57, 6.73)	(5.51, 7.70)	(5.46, 8.66)	(5.41, 9.58)	(5.37, 10.47)
	$\mathrm{VaR}_{0.995}$	5.96	(5.88, 7.24)	(5.80, 11.71)	(5.73, 16.25)	(5.70, 20.42)	(5.66, 22.94)
	$\mathrm{VaR}_{0.99}$	5.44	(5.35, 5.90)	(5.32, 6.77)	(5.25, 8.63)	(5.22, 11.17)	(5.17, 13.53)
	$\mathrm{VaR}_{0.98}$	4.90	(4.83, 5.10)	(4.78, 5.37)	(4.75, 5.71)	(4.71, 6.26)	(4.67, 7.14)

Table C.3: The bounds of VaR and ES for various scenarios and different levels of trustworthiness when using CIMDO with normal prior.