COMPUTATION OF PORTFOLIO CREDIT RISK MEASURES USING COPULA-BERNOULLI MIXTURE MODELS

MISILE KUNENE

MASTER OF SCIENCE IN MATHEMATICS

(Financial Option)

PAN AFRICAN UNIVERSITY INSTITUTE FOR BASIC SCIENCES, TECHNOLOGY AND INNOVATION

2021

Computation of Portfolio Credit Risk Measures using Copula-Bernoulli Mixture Models

Misile Kunene
Registration No: MF300-0005/2019

A research work submitted to the Pan African University Institute for Basic Sciences, Technology and Innovation in partial fulfilment of the requirement for the award of the degree of Master of Science in Mathematics (Financial Option)

2021

Declaration

This research is my original work and has not in part or in whole been presented for a degree award in any other university.

Signature.......................................Date............../............../...............

Misile Kunene

This research work has been submitted with our approval as university supervisors.

Signature.......................................Date............../............../...............

Dr. Joseph Mung'atu

Jomo Kenyatta University of Agriculture and Technology, Kenya

Signature.......................................Date............../............../...............

Dr. Euna Nyarige

Machakos University, Kenya

Dedication

To my family and friends.

Acknowledgement

I am very delighted to express my sincere appreciation to the African Union for sponsoring my degree at this level. I extend my appreciation and thanks to Pan African University and Jomo Kenyatta University of Agriculture and Technology for offering a conducive environment for learning. I would like to express my sincere gratitude to my supervisors, Dr Joseph Mung'atu and Dr Euna Nyarige.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract

The new Basel II capital-adequacy framework for market risk allows banks to use internal models to assess regulatory capital to both general market risk and credit risk for their trading book. In the first part of the work, we made a review of Generalized Linear Mixed Models (GLMMs) and then highlighted the usefulness of modelling portfolio credit default risk using a Bernoulli mixture model, a class of GLMMs. The study provided a general overview of a credit portfolio and portfolio notations. Thereafter, we made a discussion of some important concepts on copulas with focus on their implicit applications in GLMMs. Lastly, a general mathematical framework of GLMMs and their relation to Bernoulli mixture models was presented. An empirical study of the model was conducted on S & P 500 data as a way of investigating the applicability of the model. The second part of the discussion considered the problem of estimating Value-at-Risk (VaR) and Expected Shortfall (ES) of a credit portfolio from an estimated loss distribution. Given default of obligors in a credit portfolio, estimation of VaR and ES using standard Monte Carlo can result in high computational cost. A review of importance sampling concepts has been provided, a common method to make estimations more efficient, and then investigated its application assuming Bernoulli mixture models. Finally, the research study demonstrated the usefulness of using Bernoulli mixture models together with Importance Sampling algorithms in computing the VaR and ES for a financial institution.

CHAPTER 1

Introduction

1.1 Background Information

Quantitative risk management has become a relevant topic in corporate finance theory and in managerial practice. The modern society at large relies on the functioning of banking and insurance systems and has a collective interest in the stability of such systems. According to McNeil et al. (2015a), the regulatory process culminating in Basel II framework has been strongly motivated by the fear of systemic risk, i.e. the danger that problems in a single financial institution may spill over and, in extreme situations, disrupt the normal functioning of the entire financial system. The 2008 financial crisis provided us with an example of the devastating financial and economic consequences of overly naive and simplistic assumptions about default contagion. Default contagion refers to the propagation of economic distress from one firm to another which may cause default of other firms. It is very important for quantitative analysts in financial institutions and banks to properly understand the modelling of the dependences between financial assets. Practitioners, regulators and academics are all in the quest of developing and analysing quantitative models for credit losses in large lending portfolios.

From the 21^{st} century, there has been a rapid increase in financial derivatives that are used in the financial market. According to Frey and McNeil (2001), the conceptual advances in pricing options and other complex financial products, along with improvements in computer and telecommunications technologies has posed a number of risks in the financial industry that need to be hedged. Moreover, the counterparty credit risk which is associated with the use of derivative instruments has been mitigated by legally enforceable netting and through the growing use of collateral agreements. This is a major threat to financial credit providers since without an efficient quantitative model of credit risk, financial credit providers may go bankrupt due to high default instances from obligors. A company typically has strong incentives to strictly limit the probability of default in order to avoid the associated bankruptcy costs. This is directly linked to the notion of economic capital as observed by McNeil et al. (2015a). In a narrow sense, economic capital is the capital that shareholders should invest in the company in order to limit the probability of default to a given confidence level over a given time horizon. Economic capital offers a firm-wide language for discussing and pricing risk that is related directly to the principal concerns of management and other key stakeholders, namely institutional solvency and profitability.

The calculation of economic capital is a process that begins with the quantification of the risks that any given company faces over a given time period (see McNeil et al., 2015a). Efficient quantitative models can help a great deal in capturing the loss potential due to defaulting counterparties on the portfolio level; they are intended to be used for the measurement of the overall risk in a large loan portfolio, the active management of credit portfolios under risk-return considerations, or the pricing of credit insurance. As such, industry models for credit risk management need to be revisited in order to assess their relevancy and extended, if need be, for effective risk mitigation purposes.

1.2 Problem Statement

Currently, many popular credit risk models for calculating risk measures used for capital allocation purposes in a financial institution, such as the CreditMetrics model of J.P Morgan, Moody's KMV model, CreditRisk+ of Credit Suisse Financial Product and the CreditPortfolioView of McKinsely, are based on the Gaussian copula. The facts that catastrophic or extreme events do occur leaves us to casts some doubts whether the aforementioned industry credit risk models which are based on Gaussian copula are necessarily the best choice for modelling credit risk during catastrophic or extreme events since these models give very small or negligible probability of the occurrence of these extreme losses. Financial risk managers are in fear that problems in financial system may disrupt the whole economic system and hence they are more interested in modelling extreme losses of their portfolios, with value at risk and expected shortfall as their extremal risk measures. In this study, instead of using a Gaussian copula model commonly used in industry, we used a $t-$copula model to calculate portfolio risk measures since the $t-$copula model allows the occurrence of catastrophic or extreme events.

1.3 Objectives

1.3.1 General Objective

The main objective of this research work is to develop a credit risk model that can be used to calculate the value-at-risk and the expected shortfall for a given credit portfolio of a financial institution.

1.3.2 Specific Objective

(i) To develop a Bernoulli mixture model that can be used to model the occurrences of default events in a credit portfolio

(ii) To find the default distribution of a credit portfolio and

(iii) To find the loss distribution given default of a credit portfolio.

1.4 Significance of the Study

A number of empirical studies have shown that the normal distribution - univariate or multivariate - is a poor model for financial data. A further observation is that extreme, synchronized rises and falls in financial markets occur infrequently and traditional models which assumes normality for credit returns do not assign a high enough chance of occurrence of the scenario in which many things go wrong at the same time, that is, extreme events. As such, there is a need to develop credit risk models that help cater for the nature of credit returns, and assigns high enough chance of the occurrence of extreme events or extreme losses in the financial industry.

CHAPTER 2

Literature Review

Losses resulting from the failure of an obligor to make a contractual payment (default), generally associated to credit risk, are one of the major concerns for financial institutions. Default correlation and default dependency are a topic of high interest in the banking and investment community.

There are a number of credit risk models that have been proposed in literature that aim to capture the dependence of defaults (or credit migration) and these are categorised into structural and dynamic models. Credit risk models that aim to explain the mechanism by which default takes place are called structural models. A detailed description of structural models that have been proposed in literature can be found in Crouhy et al. (2000). Morgan (1997) from JP Morgan developed the credit migration approach (CreditMetrics) to model credit risk. The CreditMetrics model estimates the forward distribution of the change in the value of the portfolio of loan and bond type of product. These changes in the value are assumed to be related to the eventual migrations in credit quality of the obligor (both up and downgrades) as well as default events. The model also assumes that credit returns are normally distributed. However, McNeil et al. (2015a) argues that, while it is legitimate to assume normality with market returns, it is no longer the case with credit returns which are by nature highly-skewed and fat-tailed.

According to Mönkkönen (1998), the weakness of the CreditMetrics is its reliance on transition probabilities based on average historic frequencies of defaults and credit migration and this view has been challenged by KMV. KMV argue that in the sense of the CreditMetrics framework, the assumptions that all firms within the same rating class have the same default rate and that actual default rate is equal to the historical average default rate, cannot be true since default rates are continuous while the ratings are adjusted in a discrete fashion. KMV does not use Moody's or Standards and Poor's statistical data to assign a probability of default, as in the CreditMetrics model, which only depends on the rating of the obligor. Instead, KMV derives the actual probabilities of default, the Expected Default Frequency (EDF), for each obligor based on Merton (1974). The probability of default is thus a function of the firm's capital structure, the volatility of the asset returns and the current asset value of the firm. The EDF is firm-specific, and can be mapped into any rating system to derive the equivalent rating of the obligor.

Contrary to the CreditMetrics and the KMV models, the CreditRisk+ as proposed by Suisse (1997) from Credit Suisse Financial Product only focusses on default. The default probablity for individual bonds or loans is assumed to follow an exogenous Poisson process. In this model, the assumption is that the default risk is not related to the capital structure of the firm. Unlike in the KMV model, in the CreditRisk+ framework there is no assumption made about the causes of default: an obligor A is either in default with probability $\mathbb{P}_A$ or it is not in default with probability $1-\mathbb{P}_A$. It is assumed that for a loan, the probability of default in a given period, say a month, is the same for any other month, and for large number of obligors, the probability of default by any particular obligor is small, and the number of defaults that occur in any given period is independent of the number of defaults that occur in any other period.

Dependence of different obligors is very important in order to understand and capture default probabilities. As mentioned earlier, the problem we are facing is that we may know everything about the marginal distribution of each obligor but we have very limited information about the joint distribution of the obligors. Embrechts et al. (2002) emphasizes that the concept of copulas and notion of extreme dependence is very important in the field of quantitative risk management for capital allocation purposes and hence avoiding bankruptcy. In general, a copula is a function that links n-dimensional distribution functions to its one dimensional margins and is itself a continuous distribution function which characterises the model's dependence structure. This means that copula methods for modelling dependency provide a way to isolate the dependence structure of a portfolio from the individual margins of the assets (McNeil et al., 2015a).

Copula based methods have been found to be more efficient and reliable in modelling quantitative risk over traditional models that are based on linear correlation (McNeil et al., 2015a). Groundbreaking work on the application of copula functions has been done by a number of researchers some of whom include Embrechts et al. (1999, 2002); Straumann (2001); Bouyé et al. (2001). These authors clarify the important concepts of dependence and correlation and certainly this will have a great influence in the risk management industry. The copula theory tells us that every joint distribution function for a random vector of risk factors implicitly contains a description of the marginal behaviour of individual risk factors and a description of their dependence structure. It is of course only one way of treating dependence in multivariate risk models and is perhaps most natural in a static distributional context rather than a dynamic time series one.

According to Trivedi et al. (2007), copulas express dependence on a quantile scale, which is useful for describing the dependence of extreme outcomes and is natural in a risk-management context, where Value-at-Risk (VaR) and Expected Shortfall (ES) have led us to think of risk in terms of quantiles of the loss distribution. Moreover, copula facilitates a bottom-up approach to multivariate model building. This is particularly useful in quantitative risk management, where most of the time we have a much better understanding about the marginal behaviour of individual risk factors than we do about their dependence structure. For example, in credit risk where the individual default risk of an obligor is difficult to estimate, is at least something we can get a better handle on than the dependence among default risks for several obligors (Mai and Scherer, 2014). The copula methods allow us to combine our developed marginal models with a variety of possible specification.

In order to model the dependence of simultaneous default events observed empirically, a dependence structure is imposed on the multivariate default distribution. The most popular choice of such a structure is the multivariate normal distribution. This gives rise to the celebrated normal copula model, which is widely used in the financial industry and forms the basis of the CreditMetrics and other management systems (Gupton et al., 1997). These models are an extension of the Merton (1974) kind of model. In these models, which are often called latent variable models or threshold models, default occurs if a latent variable, often interpreted as the value of the obligors' assets, falls below some threshold, often interpreted as the value of the obligors' liability. Dependence between defaults is caused by dependence between the latent variables.

For multivariate-normally distributed risk factors such as asset values for different firms, the occurrence of many joint large movements of risk factors is a rare event. In particular, one of the most prominent features of financial variables is that they exhibit extreme dependence, i.e, they are asymptotically dependent. Financial variables take up large values (in absolute terms) simultaneously with a non-negligible probability. This in turn casts some doubts in whether the latent variable models based on the normal copula are the best choice for modelling dependent defaults. While most of the models that are used in industry are based explicitly or implicitly on the normal copula, there is no reason why we have to assume a normal copula.

Even though there are other copula functions that can be used, in this work, we chose to use the t copula due to the fact that the t distribution or t copula has fat tails than the normal copula, hence it can accommodate the modelling of rare events (by giving a higher chance of the occurrence of extreme events). Also, the symmetry properties of the t distribution and its few parameters makes it easier to handle during calibration.

Moreover, many popular credit portfolio risk measure models, such as Credit-Metrics of J. P. Morgan, KMV Portfolio Manager, CreditRisk of Credit Suisse First Boston, and McKinsey's Credit Portfolio View, widely rely on Monte Carlo simulation techniques for calculating the tail probability of credit portfolio loss distribution, or its VaR for a given confidence level over a fixed time horizon. However, credit default events of companies are rare, the threshold value of default is large and thus the tail probability of credit portfolio loss distribution is small. Therefore, standard Monte Carlo simulation is a quite time-consuming computation process for portfolios with many counter parties, and it is inefficient to calculate the rare events with small probabilities. To correct for this, a number of models that aim to improve the Monte Carlo simulations have been proposed in literature. The well known CreditMetrics portfolio model was extended by Grundke (2007) and implored a Fourier-based approach to calculate risk measures for a credit portfolio. Clustering algorithms have also been used to evaluate financial risks in Kou et al. (2014). In a paper by Merino and Nyfeler (2004), it has clearly been demonstrated that Monte Carlo simulations can be combined with Fast Fourier Transform (FFT) in order to reduce the computational time when estimating the tail probabilities of loss distribution of a credit portfolio within the conditional independence framework.

In this study, we provide a credit risk model that can be used to model the loss distribution of default credit risk in credit portfolios using Bernoulli mixture model, incorporating $t-$copulas implicitly, and importance sampling algorithms.

CHAPTER 3

Methods

In this chapter, we provide brief discussions of the methods that have been used to accomplish this work. We provide some definitions, propositions and theorems necessary for the smooth understanding of the methods used. We begin by giving a brief discussion on credit portfolio notations.

3.1 Portfolio Notations

We consider a portfolio of d obligors for $d > 0$ and fix the time horizon T. Let $\boldsymbol{X} = (X_1, \ldots, X_d)'$ be an d-dimensional vector with continuous distribution functions $F_i(x_i) = \mathbb{P}(X_i \leq x_i)$. Let the random variable X_i for $1 \leq i \leq d$ to be the asset value of the i^{th} obligor and we assume that default occurs when the asset X_i is less than the total liabilities D_i of the i^{th} obligor. We assume that the default dependence among obligors stems from the dependence among the components of the vector $\boldsymbol{X}$. We introduce another random variable S_i which is a state indicator for the i^{th} obligor. Assume that S_i take integer values in the set $\{0, 1, \ldots, n\}$ representing credit rating classes; we interpret the value 0 as default and non-zero representing states of increasing credit worthiness. We assume that at time $t = 0$ all obligors are at non-default state. We concentrate on the binary outcomes of default and non-default and ignore the fine categories of the non-defaulted obligors. We write Y_i for the default indicator variables; $Y_i = 1 \iff S_i = 0$ and $Y_i = 0 \iff S_i > 0$. Then the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ is a vector of default indicators for the portfolio and our interest is to devise a mathematical model for the joint probability function

$$p(\boldsymbol{y}) = \mathbb{P}(Y_1 = y_1, \ldots, Y_d = y_d), \quad \boldsymbol{y} \in \{0,1\}^d, \tag{3.1}$$

and the marginal default probability $\bar{p}_i = \mathbb{P}(Y_i = 1)$. In particular, we want to model the default correlation $\rho(Y_i, Y_j)$ for $i \neq j$. Since

$$\mathbb{V}(Y_i) = \mathbb{E}(Y_i^2) - \bar{p}_i^2 = \mathbb{E}(Y_i) - \bar{p}_i^2 = \bar{p}_i - \bar{p}_i^2,$$

where $\mathbb{V}(Y_i)$ is the variance of Y_i and

$$Cov(Y_i, Y_j) = \mathbb{E}[(Y_i - \mathbb{E}(Y_i))(Y_j - \mathbb{E}(Y_j))] = \mathbb{E}(Y_i Y_j) - \bar{p}_i \bar{p}_j$$

for $i \neq j$, we have that

$$\rho(Y_i, Y_j) = \frac{\mathbb{E}(Y_i Y_j) - \bar{p}_i \bar{p}_j}{\sqrt{(\bar{p}_i - \bar{p}_i^2)(\bar{p}_j - \bar{p}_j^2)}}. \tag{3.2}$$

At the portfolio level, default risk is largely driven by correlated defaults. The various factors that are pertinent to default risk contribute a lot to default correlation. In a paper by Molins and Vives (2016), it is well demonstrated that at a certain threshold, a small shift in default correlation can trigger credit portfolios or the whole market system to suffer a phase transition, such as the collapse of credit portfolios or the whole market system. We illustrate this by considering arbitrary homogenous portfolios (portfolio A and portfolio B) both consisting of 1000 obligors. We assume that in portfolio A, default events occur independently while in portfolio B, we introduce some negligible dependence between the default events, ie, we assume that the default correlation between the default events is 0.05 %. We also assume that the default probability in both the portfolios for each obligor is 0.5 %, so that on average, we expect 5 defaults. Figure 3.1 shows that the loss distribution of portfolio B is more skewed than that of A and that its right tail is significantly

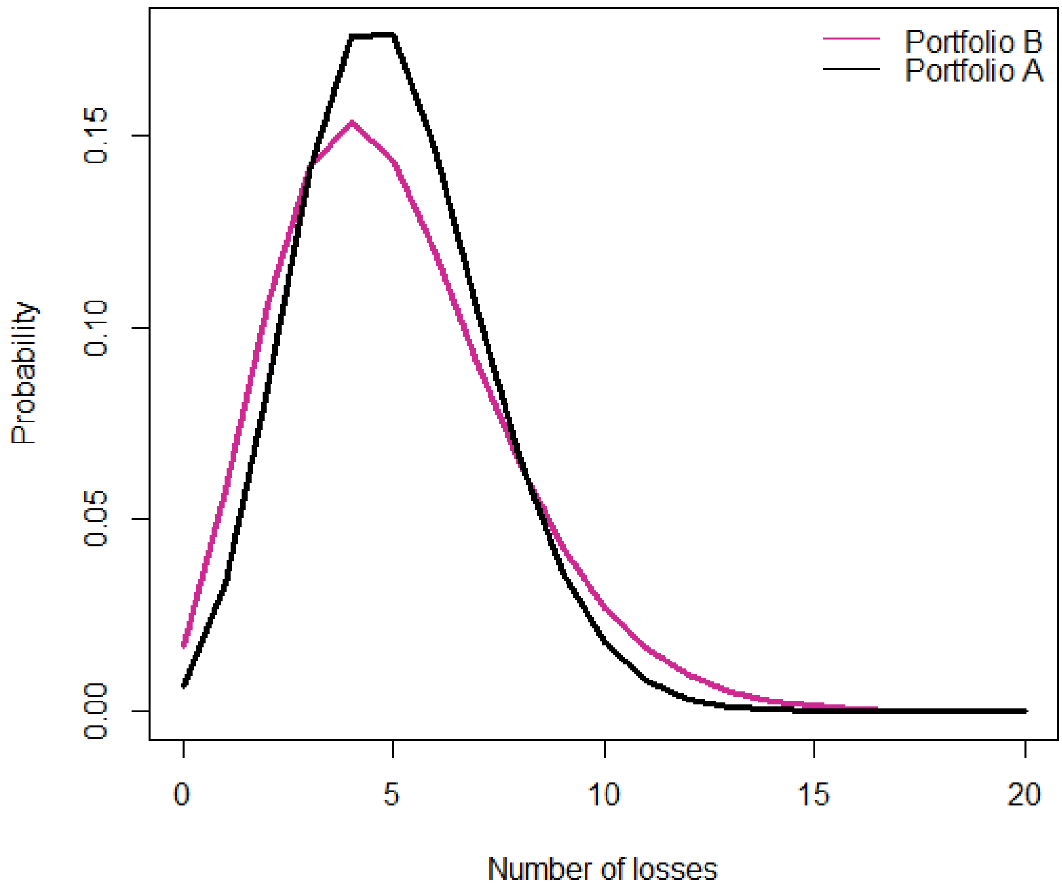

Figure 3.1: Comparison of the loss distribution of two homogeneous portfolios of 1000 loans with a default probability of 0.5 % and different dependence structures. In portfolio A, there is no default correlation while portfolio B has 0.05 % default correlation.

heavier than the right tail of portfolio A's loss distribution. This shows that default dependency has a significant influence on credit loss distributions. Typically for this reason, we devote a large part of our exposition on the analysis of credit portfolios and dependent defaults.

While numerous theoretical models have been developed to estimate default correlation (dependency measure), in the next section, we continue with our discussion by giving some concepts on copulas.

3.2 Copulas

In this section, we look closely at the issue of modelling dependence among components of a random vector of financial risk factors using the concept of copula. In a sense, every joint distribution function for a random vector of risk factors implicitly contains both a description of the marginal behaviour of individual risk factors and description of their dependence structure; the copula approach provides a way of isolating the description of the dependence structure. Copulas facilitate a botton-up approach to multivariate building of models. This is particularly very important in quantitative risk management, where we very often have a much better understanding of the marginal behaviour of individual risk factors than we do about their dependence structure. Readers interested on more concepts on copulas can refer to McNeil et al. (2015a); Nelsen (2007); Mai and Scherer (2014),

Definition 3.2.1 (Copula)**.** *A d-dimensional copula is a distribution function on $[0,1]^d$ with standard uniform marginal distribution functions (Nelsen, 2007).*

We use the notation $C(\boldsymbol{u}) = C(u_1, \cdots, u_d)$ for the multivariate distribution functions that are copulas. We see a copula C as a mapping of the form $C : [0,1]^d \to [0,1]$ satisfying the following properties:

(1) $C(u_1, \cdots, u_d)$ is increasing in each component u_i.

(2) $C(1, \cdots, 1, u_i, 1, \cdots, 1) = u_i$ for all $i \in \{1, \cdots, d\}, u_i \in [0,1]$.

(3) For all $(a_1, \cdots, a_d), (b_1, \cdots, b_d) \in [0,1]^d$ with $a_1 \leq b_i$ we have

$$\sum_{i_1}^{2} \cdots \sum_{i_d}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \cdots, u_{di_d}) \geq 0, \tag{3.3}$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \cdots, d\}$.

In working with copulas, it is very important to know the operations of probability and quantile transformations, which are concepts widely used in statistics, and these are summarized in the proposition below.

Proposition 3.2.1. *Let G be a distribution function and let $G^{\leftarrow}$ denote its generalized inverse ($G^{\leftarrow}(y) = \inf\{x : G(x) \geq y\}$).*

(1) Quantile transformation: if $U \sim U(0,1)$ has standard uniform distribution, then $\mathbb{P}(G^{\leftarrow}(U) \leq x) = G(x)$.

(2) Probability transformation: If Y has a distribution function G, where G is a continuous univariate distribution function, then $G(Y) \sim U(0,1)$ (McNeil et al., 2015a).

Proposition (3.2.1) allows us to transform and study risks with a particular continuous distribution function to have any other continuous distribution function. The Sklar's theorem (Sklar, 1959), also known as the fundamental theorem of copula, is one of the most important results in the study of multivariate distributions. Firstly, it shows that all multivariate distribution functions contain copulas and secondly, that all copulas may be used in conjunction with univariate distribution functions to construct multivariate distribution functions.

Theorem 3.2.1 (Sklar theorem.)**.** *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exist a Copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d \in \bar{\mathbb{R}}$,*

$$F(x_1, \ldots, x_d) = C(F(x_1), \ldots, F(x_d)). \tag{3.4}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times \cdots \times RanF_d$, where $RanF_i = F_i(\bar{\mathbb{R}})$ denotes the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the func-

tion F defined in (3.4) is a joint distribution function with marginals $F_1, \ldots, F_d$ (Sklar, 1959).

A full proof of the Sklar's theorem can be found in Nelsen (2007); Schweizer and Sklar (2011).

Definition 3.2.2 (Copula of F.). *If the random vector $\boldsymbol{X}$ has joint distribution function F with continuous marginal distribution functions $F_1, \ldots, F_d$, then the copula of F (or $\boldsymbol{X}$) is the distribution function C of $F_1(X_1), \ldots, F_d(X_d)$ (McNeil et al., 2015a).*

One of the most important results in copula theory is the invariance principle under strictly increasing transformations of the marginals. This result is presented below and a sketch of its proof can be found in McNeil et al. (2015a).

Proposition 3.2.2 (Invariance principle.). *Let $X_1, \ldots, X_d$ be a random vector with continuous margins and copula C and let $T_1, \ldots, T_d$ be strictly increasing functions. Then $T_1(X_1), \ldots, T_d(X_d)$ has copula C (McNeil et al., 2015a).*

A formal proof of the invariance principle can be found in Mai and Scherer (2014); McNeil et al. (2015a). The first part of Sklar's theorem (together with the invariance principle) allows us to decompose any distribution function F into its margins and copula. This in-turn allows us to study dependence independently of the margins with the margin-free $\boldsymbol{U} = (F_1(X_1), \ldots, F_d(X_d))$ instead of using $\boldsymbol{X} = (X_1, \ldots, X_d)$ since they both have the same copula by the invariance principle.

Theorem 3.2.2 (Fréchet bound.). *For every copula $C(u_1, \ldots, u_d)$ we have*

$$\max\left(\sum_{i=1}^{d} u_i + 1 - d, 0\right) \leq C(\boldsymbol{u}) \leq \min\{u_1, \ldots, u_d\} \tag{3.5}$$

(Mai and Scherer, 2014).

Fréchet bounds simply say that every copula lies between a minimum and a maximum value (see McNeil et al. (2015b) for a more detailed discussion, especially their application in dependence modelling).

Example 3.2.1 (Examples of copulas). *If $Y \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian random vector, then its copula is the Gauss copula. By Proposition (3.2.1), the copula of $\boldsymbol{Y}$ is the same as the copula of $X \sim N_d(\mathbf{0}, P)$, where P is the correlation matrix of $\boldsymbol{Y}$. Then by definition, this copula is given by*

$$\begin{aligned} C_P^{Ga}(\boldsymbol{u}) &= \mathbb{P}(\Phi(X_1) \leq u_1, \ldots, \Phi(X_d) \leq u_d) \\ &= \boldsymbol{\Phi}_P(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \end{aligned}$$

where Φ denotes the standard univariate normal distribution function amd $\boldsymbol{\Phi}_P$ is the joint distribution function of $\boldsymbol{X}$. For $d = 2$, we have

$$C_\rho^{Ga}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[\frac{-(s_1^2 - 2\rho s_1 s_2 + s_2^2)}{2(1-\rho^2)}\right] ds_1 ds_2, \tag{3.6}$$

where $\rho = \rho(X_1, X_2)$. The t copula takes the form

$$C_{\nu,P}^t(\boldsymbol{u}) = \boldsymbol{t}_{\nu,P}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)), \tag{3.7}$$

where t_ν is the standard univariate t distribution, $\boldsymbol{t}_{\nu,P}$ is the joint distribution function of the vector $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, P)$ and P is the correlation matrix.

Using Sklar's theorem, we obtain that the density of $C(\boldsymbol{u})$ is given by

$$\begin{aligned} c(u_1, \cdots, u_d) &= \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d} \\ &= \frac{f(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))}{\prod_{i=1}^d f_i(F_i^{\leftarrow}(u_i))}, \end{aligned}$$

where f is the joint density of F, f_i for $1 \leq i \leq d$ are the marginal densities and F_i

for $1 \leq i \leq d$ are the generalized inverses of the marginal distributions. Interested readers can find more details on the t copula and related copulas in Demarta and McNeil (2005).

Definition 3.2.3 (Exchangeability.)**.** *We say that a random vector $\boldsymbol{X}$ is exchangeable if $(X_1, \ldots, X_d) \stackrel{d}{=} (X_{\Pi(1)}, \cdots, X_{\Pi(d)})$ for any permutation $(\Pi(1), \ldots, \Pi(d))$ (Mai and Scherer, 2014).*

A copula is exchangeable if it is a distribution function of an exchangeable random vector and this property is very useful in modelling default dependence for homogeneous groups of companies in the context of credit risk management.

Other than the Pearson linear correlation, for $d = 2$, there is another type of correlation known as the rank correlation. Rank correlations are simple scalar measures of dependence that depend only on the copula of a bivariate distribution and not on the marginal distributions. There are a number of rank correlation measures, but in our context, we define two of the widely used rank correlations.

Definition 3.2.4 (Kendall's tau.)**.** *For random variables X_1 and X_2, Kendall's tau is given by*

$$\rho_\tau(X_1, X_2) = \mathbb{E}(sign((X_1 - \tilde{X_1})(X_2 - \tilde{X_2})))$$

where $(\tilde{X_1}, \tilde{X_2})$ is an independent copy of (X_1, X_2) (Mai and Scherer, 2014).

Definition 3.2.5 (Spearman's rho.)**.** *For random variables $\boldsymbol{X} = (X_1, \ldots, X_d)$ with X_i for $1 \leq i \leq d$ with marginal distribution functions F_i, Spearman's rho is given by*

$$\rho_S(\boldsymbol{X}) = \rho(F_1(X_1), \ldots, F_d(X_d))$$

(Mai and Scherer, 2014).

It is easy to see that Spearman's rho is just the linear correlation of the probability transformed random variables, which for continuous random variables is the linear correlation of their unique copula. For extreme financial risks, our interest is to measure extremal dependences, that is, to measure the strength of dependence in the tails of bivariate distributions (in the case where $d = 2$). This is very important in the risk management context because risk managers would want to know the joint movement of obligors. We give the following definitions.

Definition 3.2.6 (Coefficients of tail dependence.). *Let X_1 and X_2 be random variables with distribution functions F_1 and F_2 respectively. The coefficient of upper tail dependence of X_1 and X_2 is given by*

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \to 1^-} \mathbb{P}(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)),$$

provided a limit $\lambda_u \in [0, 1]$ exists. If $\lambda_u \in (0, 1]$, then X_1 and X_2 are said to show upper tail dependence or extremal dependence in the upper tail; if $\lambda_u = 0$, then they are asymptotically independent in the upper tail. Analogously, the coefficient of lower tail dependence is

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \to 0^+} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q)),$$

provided a limit $\lambda_l \in [0, 1]$ exists (McNeil et al., 2015a).

Equiped with the notion of copulas, we are now in a good position to fully describe or specify a mathematical model for default correlation.

3.3 Generalized Linear Mixed Models

Bernoulli mixture models are a specific class of generalized linear mixed models and are well known models in statistics. This family of models have the potential to deal with data in continuous, discrete, or binary format involving multiple sources of random error. The CreditRisk+ Boston (1997) industry model fits in this general framework where the number of defaults, conditionally on gamma-distributed latent factors, is Poisson distributed. GLMM are characterized by (i) random effects $\boldsymbol{\varrho}_t$ with distribution H_ϱ and hyperparameters θ, (ii) a distribution from the exponential family for the conditional response variable $Y_{i,t}$ given $\boldsymbol{\varrho}_t$, and (iii) a response function h (its inverse is known as link function) relating $\boldsymbol{X}'_{i,t}\boldsymbol{\Theta} + \boldsymbol{W}'_{i,t}\boldsymbol{\varrho}_t$ to the responses. In the absence of $\boldsymbol{\varrho}_t$, the model is simply a Generalized Linear Model (GLM), see Lindsey (2000) for concepts.

We study the responses $y_{i,t}$ and the covariates variables $x_{i,t}$ for obligor $i = 1, \dots, m_t$ and year $t = 1, \dots, T$. We let $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{m_t,t})$. Given a random effect $\boldsymbol{\varrho}_t$ of an arbitrary dimension p and covariates $\boldsymbol{x}_t$, we assume that the conditional density of the responses $\boldsymbol{y}_{i,t}$ belong to the exponential family, such as Bernoulli or Poisson, with conditional mean

$$\mathbb{E}(Y_{i,t}|\boldsymbol{\varrho}_t) = h(\omega_{i,t}), \quad \omega_{i,t} = \boldsymbol{X}'_{i,t}\boldsymbol{\Theta} + \boldsymbol{W}'_{i,t}\boldsymbol{\varrho}_t \tag{3.8}$$

for $i = 1, \dots, m_t$ and $t = 1, \dots, T$. In a GLMM model, the vector $\boldsymbol{X}_{i,t}$ is designed to specify fixed effects through the fixed parameter vector $\boldsymbol{\Theta}$ and $\boldsymbol{W}_{i,t}$ is designed to specify random effects through the vector $\boldsymbol{\varrho}_t$, hence the quantity $\boldsymbol{X}'_{i,t}\boldsymbol{\Theta} + \boldsymbol{W}'_{i,t}\boldsymbol{\varrho}_t$ determines the systemic risk of the model.

Fixed effects may be fully obligor-specific or shared for (parts of) the portfolio. Shared covariates that vary with time, such as macroeconomic variables or other observed risk factors,induce time-inhomogeneity in default rates. Obligor-specific covariates, such as balance sheet data, create heterogeneity among obligors. The design vectors $\boldsymbol{X}_{i,t}$ and $\boldsymbol{W}_{i,t}$ may contain time-dependent covariates that are known at the start of time period t or covariates that are realized during time period t, contemporaneously with the default indicators. In this setting, the vector-valued random effect $\boldsymbol{\varrho}_t$ could have each component interpreted as the general state of the economy according to industry sector and/or geographical location hence its components would then typically be strongly correlated and the (observable) design vector $\boldsymbol{W}_{i,t}$ holds the corresponding (possibly weighted) exposures of obligor i.

For the purpose of this paper, it is enough to consider the case where $W_{i,t} = 1$. Conditional on the random effect $\boldsymbol{\varrho}_t$ the responses $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{m_t,t})$ on unit t are treated as independent. By applying the De Finetti theorem (see Frey and McNeil, 2001), the unconditional joint distribution of $\boldsymbol{y}_t$ is obtained by integrating out the effect of $\boldsymbol{\varrho}_t$ and thus creates dependence among the responses $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{m_t,t})$ on unit t.

In the following section, we write the Bernoulli mixture model as a GLMM and then apply the ML estimation to obtain the model parameters.

3.3.1 Maximum Likelihood Estimation of GLMMs

Models that are used in the industry (such as the KMV, CreditMetrics and CreditRisk+) are less formally statistical in the estimation of the model parameters McNeil et al. (2015a). The reason for this is that there is not enough default data for higher rated firms to give reliable approximations of the model parameters. In this section, we discuss about the general framework of the maximum likelihood (ML) method for fitting GLMMs. There are other methods that have been proposed in literature (such as Bayesian estimation methods) but in this work, we focus on the ML method.

We recall the notations used in Section (3.3). The unconditional density or mass function f of response vector $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{m_t,t})$ is given by

$$f(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\Theta}, \theta) = \int_{\mathbb{R}^p} \mathbb{P}(Y_{i,t} = y_{i,t}|\boldsymbol{x}_t, \boldsymbol{\Theta}, \theta) h_{\varrho_t}(\boldsymbol{\varrho}_t) d\boldsymbol{\varrho}_t \tag{3.9}$$

where $p = \dim(\boldsymbol{\varrho}_t)$ and f_{ϱ_t} is the density of $\boldsymbol{\varrho}_t$. In order to catch the between year dependence of default, we assume that the random effects $\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_T$ are dependent. Since the $Y_{i,t}$ are conditionally independent (knowing $\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_T$), the the likelihood can be written in the form

$$L(\boldsymbol{\Theta}, \theta|D) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{t=1}^{T} \prod_{i=1}^{m_t} \mathbb{P}(Y_{i,t} = y_{i,t}|\boldsymbol{x}_t, \boldsymbol{\Theta}, \theta) f_{\varrho_t}(\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_T) d\boldsymbol{\varrho}_1 \cdots d\boldsymbol{\varrho}_T \tag{3.10}$$

which is an $\mathbb{R}^{p\times T}$ dimensional integral. This is a high-dimensional integration and it is difficult to master numerically and results are often inaccurate. In that context, numerical approximation methods such as the penalized quasi-likelihood (PQL) and the marginal quasi-likelihood (MQL) usually come in handy for solving this problem (Breslow and Clayton, 1993). However, models with correlated random effects are generally much too cumbersome to fit by numerical maximization of the likelihood. For correlated random effects, alternatives such as the expectation maximization (EM) algorithm or simulation of the full likelihood function using the importance sampling technique might be used (Gourieroux et al., 1996).

3.4 Threshold Models

These kinds of models are widely used in industry, some of which include the famous CreditMetrics and the KMV models. It has been proven in Frey and McNeil (2001) that these models use a similar mechanism to model the joint distribution of default; they only differ from the approach they use for the determination of individual default probabilities of obligors. To begin with, we give a general definition of a threshold model.

Definition 3.4.1 (Threshold Model)**.** *Let* $\boldsymbol{X} = (X_1, \dots, X_d)'$ *be an* $d-$*dimensional random vector and* $D^{d\times n}$ *be a deterministic matrix with elements* D_{ij} *for* $1 \le i \le d$ *and* $0 \le j \le n$ *such that, for every* i*, the elements of the* i^{th} *row form a set of increasing thresholds satisfying* $D_{i1} < \cdots < D_{in}$*. Set* $D_{i0} = -\infty$ *and* $D_{i(n+1)} = \infty$ *and set*

$$S_i = j \iff D_{ij} < X_i \le D_{i(n+1)}.$$

The pair $(\boldsymbol{X}, D)$ *is said to be a threshold model for the state vector* $\boldsymbol{S} = (S_1, \dots, S_d)'$ *(Frey et al., 2001).*

We call the vector $\boldsymbol{X}$ as the vector of critical variables and its marginal distributions are given by $F_i(x) = \mathbb{P}(X_i \le x)$. We set the i^{th} row to be the critical thresholds for the i^{th} obligor. Following Merton (1974), we assume that the i^{th} obligor default when $S_i = 0 \iff X_i \le D_{i1}$, hence the default probability for the i^{th} obligor is given by $\bar{p}_i = F_i(D_{i1})$. Assuming that we know the default distribution, then the default correlation $\rho(Y_i, Y_j)$ depends on $\mathbb{E}(Y_iY_j) = \mathbb{P}(X_i \le D_{i1}, X_j \le D_{j1})$ as presented in (3.2). So in general, the default correlation depends on the joint distribution of X_i and X_j for $i \ne j$. To establish a relationship between threshold models and copulas, we provide the following result (with proof).

Lemma 3.4.1. *Let* $(\boldsymbol{X}, D)$ *and* $(\tilde{\boldsymbol{X}}, \tilde{D})$ *be a pair of threshold models with state vectors* $\boldsymbol{S} = (S_1, \dots, S_d)'$ *and* $\tilde{\boldsymbol{S}} = (\tilde{S}_1, \dots, \tilde{S}_d)'$ *respectively. Then the models are equivalent if the following conditions hold:*

(i) The marginal distributions of the random vectors $\boldsymbol{S}$ *and* $\tilde{\boldsymbol{S}}$ *coincide, that is, if we have*

$$\mathbb{P}(S_i = j) = \mathbb{P}(\tilde{S}_i = j), j \in \{1, \dots, n\}, i \in \{1, \dots, d\}.$$

(ii) $\boldsymbol{X}$ *and* $\tilde{\boldsymbol{X}}$ *have the same copula* C*.*

Proof. We recall from Definition (3.2.3) that $\boldsymbol{S} \overset{d}{=} \tilde{\boldsymbol{S}}$ if and only if, for all $(j_1, \ldots, j_d) \in (0, \ldots, n)$, hence we can write

$$\begin{aligned} \mathbb{P}(D_{ij_1} < X_1 &\leq D_{1(j_1+1)}, \ldots, D_{dj_d} < X_d \leq D_{m(j_m+1)}) \\ &= \mathbb{P}(\tilde{D}_{ij_1} < \tilde{X}_1 \leq \tilde{D}_{1(j_1+1)}, \ldots, \tilde{D}_{dj_d} < \tilde{X}_d \leq \tilde{D}_{m(j_m+1)}) \end{aligned}$$

. By standard measure-theoretical arguments this holds if, for all $(j_1, \ldots, j_d) \in (1, \ldots, n)$,

$$\mathbb{P}(X_1 \leq D_{1(j_1+1)}, \ldots, X_d \leq D_{m(j_m+1)}) = \mathbb{P}(\tilde{X}_1 \leq \tilde{D}_{1(j_1+1)}, \ldots, \tilde{X}_d \leq \tilde{D}_{m(j_m+1)}).$$

By Sklar's theorem, this is equivalent to say that

$$C(F_1(D_{1j_1}), \ldots, F_d(D_d(j_d))) = C(\tilde{F}_1(\tilde{D}_{1j_1}), \ldots, \tilde{F}_d(\tilde{D}_d(j_d))),$$

where C is the copula of $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ using condition (ii). Condition (i) implies that $F_i(D_{ij}) = \tilde{F}_i(\tilde{D}_{ij})$ for all $i \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, n\}$ and the claim follows. □

Now, using the above lemma for a model of default and non-default of a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, d\}$ with individual default probabilities $\bar{p}_{i_1}, \ldots, \bar{p}_{i_k}$, we can write

$$\begin{aligned} \mathbb{P}(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) &= \mathbb{P}(X_{i_1} \leq D_{i_1 1}, \ldots, X_{1_k} \leq D_{i_k 1}) \\ &= C_{1\cdots k}(\bar{p}_{i_1}, \ldots, \bar{p}_{i_k}), \end{aligned}$$

where $C_{1\cdots k}$ denotes the corresponding k dimensional copula.

An example of a threshold model with the same framework used in industry by practitioners to price basket credit derivatives is the one proposed by Li (2000).

In this model, the critical variable X_i is interpreted as the default time of the i^{th} company and it is assumed that X_i is exponentially distributed. So obligor i default by time T if and only if $X_i \leq T$, hence $\bar{p}_i = F_i(T)$. For the multivariate distribution of $\boldsymbol{X}$, Li (2000) assumes that $\boldsymbol{X}$ has a Gauss copula C_P^G for some correlation matrix P, so that

$$\mathbb{P}(X_1 \leq t_1, \ldots X_d \leq t_d) = C_P^G(F_1(t_1), \ldots, F_d(t_d)).$$

While most threshold models used in industry are based on the Gaussian copula, there is no reason for such a choice of copula. In fact, it has been shown in Frey et al. (2001) that the choice of the copula is very critical to the tail of the distribution of the number of defaults. In this study, we are mainly concerned with the sensitivity of the distribution of the number of defaults with respect to the assumption of the Gaussian dependence structure. Our interest is driven by the fact that the Gaussian dependence structure may underestimate the probability of default due to large movement of risk factors, with potentially drastic implications for the performance of risk management models.

3.5 Bernoulli Mixture Models

In these types of models, we assume that the default risk of an obligor depend on a set of risk factors (eg. climatic, geographic, economic, etc) and these factors can also be modelled stochastically. We also assume that if we know the realization of the risk factors, then the default of the individual obligors are independent. The dependence between defaults is introduced by the dependence of the individual default probabilities on the set of risk factors. Before we continue with our discussion, we first give the following definition.

Definition 3.5.1 (Bernoulli mixture model.)**.** *Given some $p < d$ and a p- dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ follows* *a Bernoulli mixture model with factor vector $\boldsymbol{\Psi}$ if there are functions $h_i : \mathbb{R}^p \to [0,1]$ for $1 \le i \le d$, such that conditional on $\boldsymbol{\Psi}$ the components of $\boldsymbol{Y}$ are independent Bernoulli random variables satisfying $h_i(\boldsymbol{\psi}) = \mathbb{P}(Y_i = 1|\boldsymbol{\Psi} = \boldsymbol{\psi})$ (Frey and McNeil, 2003).*

Now, let $A = \{1, \ldots, \kappa\}$ be a set of credit rating classes where higher values indicate high creditworthiness. We suppose that we can collect historical default data over T time periods (yearly data) for κ different credit rating classes indexed by $r = 1, ..., \kappa$. We denote by $m_{r,t}$ the number of obligors for the t^{th} year cohort in the credit rating class r and $M_{r,t}$ to be the number of defaulted obligors in the credit rating class r in the t^{th} year, so that $m_t = \sum_{r=1}^{\kappa} m_{r,t}$ and $M_t = \sum_{r=1}^{\kappa} M_{r,t}$ denotes the number of obligors and number of defaulted obligors in the period t respectively. We write $Y_{i,t}$ for the default indicator variables; $Y_{i,t} = 1$ if and only if the i^{th} obligor defaults on the t^{th} year and $Y_{i,t} = 0$ otherwise, where $1 \le i \le m_t$. With this setting, for $s \neq t$, the indicator variables $Y_{i,t}$ and $Y_{i,s}$ do not refer to the same obligor and hence the vectors $\boldsymbol{Y}_t = (Y_{1,t}, \ldots, Y_{m_t,t})'$ and $\boldsymbol{Y}_s = (Y_{1,s}, \ldots, Y_{m_s,s})'$ may not be of the same length.

Now, given the random vector $\boldsymbol{\Psi}_t = (\Psi_{1,t}, \ldots, \Psi_{m_t,t})'$ with distribution G_{Ψ}, we assume that the default indicators $(Y_{1,t}, \ldots, Y_{m_t,t})$ are conditionally independent Bernoulli random variables with success probability

$$\mathbb{P}(Y_{i,t} = 1|\boldsymbol{\Psi}_t = \boldsymbol{\psi}_t) = g(\mu_r + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t). \tag{3.11}$$

The vectors $\boldsymbol{w}_{i,t}$ and $\boldsymbol{x}_{i,t}$ are known vectors corresponding to the covariates of the i^{th} obligor in period t, $\boldsymbol{\beta}$ and $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_\kappa)$ are vectors of unknown regression parameters and g is a smooth, strictly increasing function taking values from $\mathbb{R}$ to the unit interval, called the response function. We can easily show that the probit response function is a natural choice for the response function, and for other choices, interested readers can see Joe (1997).

One of the stylized facts about default events is that they are dependent. However, when we assume that the defaults events are dependent it is difficult to come up with a tractable model (it is not easy to compute the joint probability of default under dependence). A solution to this or the simplest kind of dependence (tractability) is through the concept of conditional independence, that is, we assume that conditional on a set of risk factors, the default events are independent. To clarify this, we give the following definition.

Definition 3.5.2 (Conditional independence structure)**.** *Let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ be a random vector. We say that the random vector $\boldsymbol{Y}$ has a $d-$dimensional conditional independence structure with conditioning vector $\boldsymbol{\Psi}$ if there is some $p < d$ and a $p-$dimensional random vector $\boldsymbol{\Psi} = (\Psi_i, \ldots, \Psi_p)'$ such that, conditional on $\boldsymbol{\Psi}$, the random variables $Y_1, \ldots, Y_d$ are independent (Frey and McNeil, 2003).*

Using Definition 3.5.2, for $\boldsymbol{y} \in \{0,1\}^{m_t}$, the conditional joint default probability for the random variables $Y_{1,t}, \ldots, Y_{m_t,t}$ is given by

$$\mathbb{P}(\boldsymbol{Y}_t = \boldsymbol{y} | \boldsymbol{\Psi}_t = \boldsymbol{\psi}) = \prod_{i=1}^{m_t} \mathbb{P}(Y_{i,t} = 1 | \boldsymbol{\Psi}_t = \boldsymbol{\psi})^{y_i} (1 - \mathbb{P}(Y_{i,t} = 1 | \boldsymbol{\Psi}_t = \boldsymbol{\psi}))^{1-y_i}. \quad (3.12)$$

To continue with our discussion, we give the following theorem.

Theorem 3.5.1 (De Finetti)**.** *For any infinite sequence $Y_1, Y_2, \ldots$ of exchangeable Bernoulli random variables there is a probability distribution G on $[0,1]$ such that for all $k \leq m \in \mathbb{N}$, we have*

$$\mathbb{P}(Y_1 = 1, \ldots, Y_k = 1, Y_{k+1} = 0, \ldots, Y_m = 0) = \int_0^1 q^k (1-q)^{m-k} dG(q)$$

(Lindsey, 2000).

Using the De Finetti theorem, then the unconditional joint default probability is found by integrating equation (3.12) over the distribution of $\boldsymbol{\Psi}$, that is,

$$\mathbb{P}(\boldsymbol{Y}_t=\boldsymbol{y})=\int_0^1\prod_{i=1}^{m_t}\mathbb{P}(Y_{i,t}=1|\boldsymbol{\Psi}_t=\boldsymbol{\psi})^{y_i}(1-\mathbb{P}(Y_{i,t}=1|\boldsymbol{\Psi}_t=\boldsymbol{\psi}))^{1-y_i}dG_{\Psi}(\boldsymbol{\Psi}_t). \tag{3.13}$$

By definition of Bernoulli mixture models, we can show that threshold models inspired by the Merton (1974) type model can be interpreted as Bernoulli mixture models. We show this in the following discussion.

For simplicity, we recall the notations given in Section 3.1 and Section 3.4, we will not consider the time period and the credit rating in which obligors belong since this will not change the results. We still assume that the i^{th} obligor defaults when $S_i=0 \iff X_i\leq D_{i1}$. Under certain conditions, we want to show that Bernoulli mixture models can still be written as threshold models and this will help us to introduce the copula concepts in our model (implicitly).

Now, the following condition ensures that a Bernoulli mixture model can be written as a threshold model, and this is very important for the development of our model.

Lemma 3.5.1. *Let $(\boldsymbol{X},D)$ be a threshold model for an $d-$dimensional random vector $\boldsymbol{X}$. If $\boldsymbol{X}$ has a $d-$dimensional conditional independence structure with conditioning variable $\boldsymbol{\Psi}$, then the default indicators $Y_i=\boldsymbol{I}_{\{X_i\leq D_{i1}\}}$ follow a Bernoulli mixture model with factor vector $\boldsymbol{\Psi}$, where the conditional default probabilities are given by $h_i(\boldsymbol{\psi})=\mathbb{P}(X_i\leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi})$.*

Proof. Let $\boldsymbol{y}\in\{0,1\}^d$ and define the set $A:=\{1\leq i\leq d: y_i=1\}$ and $A^c:=$

$\{1,\ldots,d\} - A$. Now consider

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}|\boldsymbol{\Psi}=\boldsymbol{\psi}) &= \mathbb{P}\left(\bigcap_{i\in A}\{X_i \leq D_i 1\}\bigcap_{i\in A^c}\{X_i > D_{i1}\}\middle|\boldsymbol{\Psi}=\boldsymbol{\psi}\right) \\
&= \prod_{i\in A}\mathbb{P}(X_i \leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi})\prod_{i\in A^c}(1-\mathbb{P}(X_i \leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi})) \\
&= \prod_{i=1}^{d}\mathbb{P}(X_i \leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi})^{y_i}(1-\mathbb{P}(X_i \leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi}))^{1-y_i} \\
&= \prod_{i=1}^{d} h_i(\boldsymbol{\psi})^{y_i}(1-h_i(\boldsymbol{\psi}))^{1-y_i}
\end{aligned}$$

showing that conditional on $\boldsymbol{\Psi}=\boldsymbol{\psi}$, the Y_i for $1 \leq i \leq d$ are independent Bernoulli random variables with success probability $h_i(\boldsymbol{\psi}) = \mathbb{P}(X_i \leq D_{i1}|\boldsymbol{\Psi}=\boldsymbol{\psi})$. □

So far we have not answered the question: How do we describe dependency coming from the factors? To provide a solution for this question, we require the following definition from basic statistics.

Definition 3.5.3 (Normal variance mixture distribution)**.** *Let $\boldsymbol{X} = (X_1,\ldots,X_d)'$ be a random vector. We say that the random vector $\boldsymbol{X}$ has a normal variance mixture distribution if*

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{v} + \sqrt{W}B\boldsymbol{Z}, \tag{3.14}$$

where

1. *$Z \sim N_p(\mathbf{0}, I_p)$*
2. *$W \geq 0$ is a non-negative, scalar-valued random variable which is independent of Z and*
3. *$B \in \mathbb{R}^{d\times p}$ and $\boldsymbol{v} \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively.*

(McNeil et al., 2015a)

This definition will help us to specify the dependency structure of the model, and at the same time introduce (implicitly) the t copula.

Now, we let the critical variables $X_1, \ldots, X_d$ to have a normal mean-variance mixture distribution so that we can write $\boldsymbol{X} = \boldsymbol{l}(W) + \sqrt{W}\boldsymbol{Z}$ where W is independent of $\boldsymbol{Z}$. We write $\boldsymbol{Z} = B\boldsymbol{F} + \boldsymbol{\epsilon}$ where the random vector $\boldsymbol{F} \sim N_p(\boldsymbol{0}, \Omega)$, $B \in \mathbb{R}^{d \times p}$ is the factor loading matrix, and the random variables ϵ_i are independent, normally distributed random variables which are also independent of $\boldsymbol{F}$. We define $\boldsymbol{\Psi} = (F_1, \cdots, F_p, W)$, then conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the random vector $\boldsymbol{X} \sim N_d(\boldsymbol{l}(w) + \sqrt{w}B\boldsymbol{f}, w\Upsilon)$ (where Υ is the diagonal covariance matrix of $\boldsymbol{\epsilon}$) has a $(p+1)-$dimensional conditional independence structure. Note that

$$\frac{X_i - l_i(w) - \sqrt{w}\boldsymbol{b}'\boldsymbol{f}}{\sqrt{w\eta_i}} \sim N(0,1) \tag{3.15}$$

where $l_i(w)$ is the i^{th} component of $\boldsymbol{l}(w)$, $\boldsymbol{b}_i$ is the i^{th} row of B, and η_i is the i^{th} diagonal element of Υ. We deduce that for the threshold model $(\boldsymbol{X}, D)$ has an equivalent Bernoulli mixture model with conditional default probabilities

$$h_i(\boldsymbol{\psi}) = \mathbb{P}(X_i \leq D_{i1} | \boldsymbol{\Psi} = \boldsymbol{\psi}) = \Phi\left(\frac{D_{i1} - l_i(w) - \sqrt{w}\boldsymbol{b}_i'\boldsymbol{f}}{\sqrt{w\eta_i}}\right). \tag{3.16}$$

Industry credit risk models such as the KMV and CreditMetrics use this type of model. To see this, if we assume that the critical variables are Gaussian so that $\boldsymbol{X} = \boldsymbol{Z}$ and $\boldsymbol{\Psi} = \boldsymbol{F}$. Recall that the individual default probabilities are given by $\mathbb{P}(X_i \leq D_{i,1}) = F_i(D_{i1}) = \bar{p}_i$, hence we have that $D_{i1} = F_i^{-1}(\bar{p}_i)$. We standardize the critical variables $X_1, \ldots, X_d$ in a way that they have a variance of one and then reparametrize the formula in (3.16) in terms of the individual default probabilities and the systemic variance component $\beta_i = \boldsymbol{b}_i'\Omega\boldsymbol{b}_i = 1 - \eta_i$, then we have

$$h_i(\boldsymbol{\psi}) = \Phi\left(\frac{\Phi^{-1}(\bar{p}_i) - \boldsymbol{b}_i{}'\boldsymbol{\Psi}}{\sqrt{(1-\beta_i)}}\right) \tag{3.17}$$

where $D_{i1} = \Phi^{-1}(\bar{p}_i)$. Given that a large number of empirical studies have found that financial variables often show high kurtosis and heavy tails, there is no reason to assume that the critical variables are Gaussian. Studies such as those from Ankudinov et al. (2017); Behr and Pötter (2009); D'Amico and Petroni (2018) show that the dependence between financial variables is generally extremal. In view of this, using the model in (3.17) as a stepping stone, we use the t copula (instead of the Gaussian copula) to model credit defaults of obligors. To achieve the t copula model, we assume that the critical variables $X_1, \ldots, X_d$ have a multivariate t distribution so that we can write $\boldsymbol{X} = \sqrt{W}\boldsymbol{Z}$ and $W \sim Ig(\frac{\nu}{2}, \frac{\nu}{2})$. In this setting, $\boldsymbol{X}$ has a multivariate t distribution by definition. We then standardize the margins to be univariate t-distributed with ν degrees of freedom. We let the proportion of the critical variables X_i explained by the factors $\boldsymbol{\Psi} = \boldsymbol{F}$ to be $\beta_i = \boldsymbol{b}_i'\Omega\boldsymbol{b}_i = 1 - \eta_i$. It is easy to see from (3.16) that the t copula (by Sklar's Theorem) model is of the form

$$h_i(\boldsymbol{\psi}) = \Phi\left(\frac{t_\nu^{-1}(\bar{p}_i)W^{-1/2} - \boldsymbol{b}_i{}'\boldsymbol{\Psi}}{\sqrt{(1-\beta_i)}}\right). \tag{3.18}$$

Remark 3.5.1. *If we recall the definition of a copula function, it is easy to see that the quantity $t_\nu^{-1}(\bar{p}_i)$ is a component of a copula, hence we say that the the model in equation (3.18) is an implicit copula model.*

Equation (3.18) can be viewed as the probit-normal distribution model.

3.6 Bernoulli Mixture Model as GLMM

We continue with the notations provided in Section 3.5 (we also include the time periods). We choose the response function g such that

$$\mathbb{E}(Y_{i,t}|\boldsymbol{\Psi}_t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\zeta_{i,t}} \exp\left(-\frac{u^2}{2}\right)du =: g(\zeta_{i,t}), \quad \zeta_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\Pi} + \boldsymbol{w}_{i,t}'\boldsymbol{\Psi}_t$$

and it is easy to see that this is the probit model where $g(\zeta_{i,t}) = \Phi(\boldsymbol{x}_{i,t}'\boldsymbol{\Pi} + \boldsymbol{\Psi}_t)$, the cumulative standard normal distribution function. In this setting, the fixed and random effects are represented by $\boldsymbol{x}'_{i,t}\boldsymbol{\Pi}$ and $\boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t$ respectively. Note that this is the same representation as in equation (3.11) where we rewrite $g(\zeta_{i,t}) = g(\mu_r + \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t) = g(\boldsymbol{x}'_{i,t}\boldsymbol{\Pi} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t)$ and for $i = 1, \ldots, m_t$ and $t = 1, \ldots, T$. Using equation (3.11), we have that conditional on $\boldsymbol{\Psi}_t$, the conditional default probability for the i^{th} obligor in the t^{th} year is given by

$$\mathbb{P}(Y_{i,t} = 1|\boldsymbol{\Psi}_t = \boldsymbol{\psi}_t) = \Phi(\boldsymbol{x}'_{i,t}\boldsymbol{\Pi} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t), \tag{3.19}$$

and from equation (3.12) the conditional joint distribution is given by

$$\mathbb{P}(\boldsymbol{Y}_t = \boldsymbol{y}_t|\boldsymbol{\Psi}_t = \boldsymbol{\psi}_t) = \prod_{i=1}^{m_t} \Phi(\boldsymbol{x}'_{i,t}\boldsymbol{\Pi} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t)^{y_{i,t}}(1 - \Phi(\boldsymbol{x}'_{i,t}\boldsymbol{\Pi} + \boldsymbol{w}'_{i,t}\boldsymbol{\Psi}_t)^{1-y_{i,t}}. \tag{3.20}$$

In GLMMs, the unconditional distribution of the responses is obtained by integrating out the effect of the random effects (by the De Finetti theorem), and this greatly complicates the use of standard maximum likelihood (ML) estimation as mentioned in Section 3.3. The general likelihood function for the responses on different time periods $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ follows from the conditional independence property, and is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\mu}, \alpha|D) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \left(\prod_{t=1}^{T} \prod_{i=1}^{m_t} \mathbb{P}(Y_{i,t} = y_{i,t}|\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{x}_{i,t}, \boldsymbol{\Psi}_t, \boldsymbol{w}_{i,t}) \right) f_{\Psi}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_T|\alpha) d\boldsymbol{\Psi}_1 \cdots d\boldsymbol{\Psi}_T \tag{3.21}$$

where f_{Ψ} is the joint density of the random effect, $D = \{\boldsymbol{y}_t, \boldsymbol{\Psi}_t, \boldsymbol{x}_t\}_{t=1}^{T}$ denotes the observed data. Full ML for GLMM is only a viable option for the simplest models. In that regard, we follow the work of Vasicek (1997) and consider a one factor Bernoulli mixture model of the form

$$\mathbb{P}(Y_{i,t} = 1|\Psi_t) = \Phi(\tau_{i,t} + w_{i,t}\Psi_t) \tag{3.22}$$

where $\Psi_t \sim N(0,\sigma^2)$ and $\Psi_1,\ldots,\Psi_T$ are identically independently distributed and we see that equation (3.22) is as a GLMM form.

Remark 3.6.1. *Note that equation (3.22) is the same as equation (3.18) where* $\tau_{i,t} = \boldsymbol{x}_{i,t}\boldsymbol{\Pi} = \dfrac{t_\nu^{-1}(\bar{p}_{i,t})W^{-1/2}}{\sqrt{1-\beta_{i,t}}}$ *and* $w_{i,t} = \dfrac{\beta_{i,t}}{\sqrt{1-\beta_{i,t}}}$.

Under these assumptions, the likelihood is easy to compute and equation (3.21) is reduced to

$$L(\boldsymbol{\beta},\boldsymbol{\mu},\alpha|D) = \prod_{t=1}^{T}\left(\int_{\mathbb{R}}\prod_{i=1}^{m_t}\mathbb{P}(Y_{i,t}=y_{i,t}|\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{x}_{i,t},\boldsymbol{\Psi}_t)f_{\Psi}(\psi_t)d\psi_t\right) \tag{3.23}$$

Now, recall that a bond or loan portfolio consist of firms with different credit ratings and suppose that we have historical default data over specific time periods (yearly data) for κ different classes of credit rating classes $1,\ldots,\kappa$ and let $r = 1,...,\kappa$ be the credit rating classes. We denote by $m_{r,t}$ the number of obligors for the t^{th} year cohort in the credit rating class r and $M_{r,t}$ to be the number of defaulted obligors in the t^{th} year.

We assume that in the t^{th} year, the conditional distribution of the vector $\boldsymbol{M}_t = (M_{1,t},\cdots,M_{\kappa,t})'$ is of the form

$$\mathbb{P}(\boldsymbol{M}_t = \boldsymbol{l}_t|\Psi_t) = \prod_{t=1}^{T}\prod_{i=1}^{\kappa}\begin{pmatrix} m_{i,t} \\ l_{i,t}\end{pmatrix}\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t)^{l_{i,t}}(1-\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t))^{m_{i,t}-l_{i,t}} \tag{3.24}$$

and the unconditional distribution is obtained by integrating over the factor variable. Also, assume that the vectors $\boldsymbol{M}_j$ and $\boldsymbol{M}_{j'}$ are conditionally independent given Ψ_t and $\Psi_{t'}$. With these assumptions, then the joint distribution of $\boldsymbol{M}_1,\ldots,\boldsymbol{M}_T$ is the product of the marginal distributions of the vector $\boldsymbol{M}_t$ so that we have

$$L(\boldsymbol{\mu}, \boldsymbol{\beta}; \boldsymbol{M}_1, \ldots, \boldsymbol{M}_T) = \sum_{t=1}^{T} \sum_{i=1}^{\kappa} \log \begin{pmatrix} m_{i,t} \\ M_{i,t} \end{pmatrix} + \sum_{t=1}^{T} \log S_t \tag{3.25}$$

where

$$S_t = \int_{-\infty}^{\infty} \exp \left[\sum_{r=1}^{\kappa} M_{r,t} \log(\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t))+(m_{r,t}-M_{r,t}) \log(1-\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t)) \right] f_{\Psi} d\psi_t, \tag{3.26}$$

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\kappa)'$ and $\boldsymbol{\beta}$ are the unknown model parameters. The estimates of the unconditional default probabilities $\hat{\pi}_1^{(r)}$ for each obligor in the credit rating r are obtained by taking the expectation of $\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi} + w_{i,t}\Psi_t)$, hence given by

$$\hat{\pi}_1^{(r)} := \mathbb{E}(\Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t)) = \int_{-\infty}^{\infty} \Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi}+w_{i,t}\Psi_t) f_{\Psi}(\psi_t) d\psi_t, \quad 1 \le r \le \kappa, \tag{3.27}$$

and the within-credit rating and between-credit rating default correlations for each of the i^{th} and j^{th} obligors are given by

$$\begin{aligned} \hat{\pi}_2^{(r,s)} : &= \mathbb{E}(\Phi(\boldsymbol{x}_{i,t}\mathbf{\Pi} + w_{i,t}\Psi_t)\Phi(\boldsymbol{x}_{j,t}\boldsymbol{\Pi} + w_{j,t}\Psi_t)) \\ &= \int_{-\infty}^{\infty} \Phi(\boldsymbol{x}_{i,t}\boldsymbol{\Pi} + w_{i,t}\Psi_t)\Phi(\boldsymbol{x}_{j,t}\boldsymbol{\Pi} + w_{i,t}\Psi_t) f_{\Psi}(\psi_t) d\psi_t, \quad 1 \le r, s \le \kappa \end{aligned}$$

So the matrix of estimated within and between groups default correlations is given by

$$\hat{\rho}_Y^{(r,s)} = \frac{\hat{\pi}_2^{(r,s)} - \hat{\pi}_1^{(r)} \hat{\pi}_1^{(s)}}{\sqrt{(\hat{\pi}_1^{(r)} - \hat{\pi}_1^{(r)2})(\hat{\pi}_1^{(s)} - \hat{\pi}_1^{(s)2})}} \tag{3.28}$$

CHAPTER 4

Loss Distribution Estimation

We begin by giving a discussion about the loss distribution and some risk measures associated with it. We consider a Bernoulli mixture model for a loan or bond portfolio with d obligors and we follow the ideas of Artzner et al. (1999) but with a different notation tailored for an application in the field of credit risk management. We assume that the overall loss of the portfolio is given by $L = \sum_{i=1}^{d} L_i = \sum_{i=1}^{d} e_i Y_i$ where L_i is the loss incurred by the i^{th} obligor for $1 \leq i \leq d$, and e_i and Y_i are the exposure (loss given default) and indicator variables respectively. We also assume that the L_i are conditionally independent given the risk factors $\boldsymbol{\Psi}$. Our main focus is the distribution function $F_L = \mathbb{P}(L > c), \quad c \in \mathbb{R}$ of the loss L. In particular, we are interested in two risk measures which are based on the loss distribution F_L, namely Value-at-Risk (VaR) and expected shortfall (ES).

Definition 4.0.1 (Value-at-Risk)**.** *Given some confidence level $\alpha \in (0,1)$, the VaR of a portfolio at the confidence level α is the smallest number $l \in \mathbb{R}$ such that the probability that the loss L exceeds l is no greater than $1-\alpha$. Formall*

$$VaR_\alpha = \inf\{l \in \mathbb{R} : \mathbb{P}(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : 1 - F_l(l) \leq 1 - \alpha\}. \qquad (4.1)$$

(Artzner et al., 1999)

It is easy to see that VaR is the α-quantile of the distribution of L in terms of a generalised inverse of the distribution function F_L since

$$\inf\{l \in \mathbb{R} : 1 - F_l(l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_l(l) \geq \alpha\}$$

by the definition of the generalized inverse.

Definition 4.0.2 (Expected shortfall). *Consider the loss L with a continuous distribution function F_L satisfying $\int_{\mathbb{R}} |l| dF_L(l) < \infty$. Then the expected shortfall at the confidence level $\alpha \in (0,1)$ is defined to be*

$$ES_\alpha = \mathbb{E}(L|L \geq VaR_\alpha) = \frac{\mathbb{E}(L; L \geq VaR_\alpha)}{\mathbb{P}(L \geq VaR_\alpha(L))}. \tag{4.2}$$

(Artzner et al., 1999)

A general definition of expected shortfall has been presented in Acerbi and Tasche (2002).

Definition 4.0.3 (Generalized expected shortfall). *Given an integrable random variable L and $\alpha \in (0,1)$, the generalized expected shortfall at confidence level α is given by*

$$GES_\alpha = \frac{1}{1-\alpha}\Big[\mathbb{E}(L; L \geq VaR_\alpha(L)) + q_\alpha(1 - \alpha - \mathbb{P}(L \geq VaR_\alpha(L)))\Big] \tag{4.3}$$

A possible method for calculating risk measures and other related quantities such as capital allocations is to use Monte Carlo (MC) simulations. However, for large values of c, very few samples actually have $L > c$, leading to the outcome that most samples are wasted because $\boldsymbol{I}_{\{L>c\}} = 0$, where $\boldsymbol{I}$ is the indicator function. Fortunately, for this type of problem there is a fairly well-established approach, which we now describe. It matches the VaR problem with a small tail probability (see Fernández et al., 2012; Glasserman et al., 2000, 2002; Glasserman and Li, 2005).

4.1 Importance Sampling

For importance sampling and related subsections, our discussions will follow the ideas of Glasserman and Li (2005). Let Y be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assume that it has an absolutely continuous distribution function

$f_Y^{\mathbb{P}}$. We consider the problem of computing the expected value

$$I(g(Y)) = \mathbb{E}^{\mathbb{P}}[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y^{\mathbb{P}}(y) dy \tag{4.4}$$

for some known function g. The notation $\mathbb{E}^{\mathbb{P}}$ emphasizes that the expectation is with respect to the measure $\mathbb{P}$. For independent random variables $Y_1, \ldots, Y_d$ sampled from the distribution function $f_Y^{\mathbb{P}}$, a typical Monte Carlo estimator of $I(g(Y))$ is given by

$$\hat{I}_{\mathbb{P}}(g) = \frac{1}{d} \sum_{i=1}^{d} g(Y_i) \tag{4.5}$$

and it converges to $I(g(Y))$ by the strong law of large numbers. However, because we are dealing with rare-event simulation, the speed of convergence may not be very fast. The importance sampling technique is based on the alternative representation of equation (4.4). Now, let $\mathbb{Q}$ be another probability measure equivalent to $\mathbb{P}$ and let $f_Y^{\mathbb{Q}}$ be the probability density function of Y (whose support should contain that of $f_Y^{\mathbb{P}}$, i.e. on the set $\{y \in \mathbb{R}^d : f_Y^{\mathbb{P}}(y) \neq 0\}$) with respect to $\mathbb{Q}$. It is easy to show that $I(g(Y)) = \mathbb{E}^{\mathbb{Q}}[g(Y)l(Y)]$ where $l(Y) = f_Y^{\mathbb{P}}(Y)/f_Y^{\mathbb{Q}}(Y)$ is called the likelihood ratio, hence we can define another unbiased estimator of $I(g)$ by

$$\hat{I}_{\mathbb{Q}}(g) = \frac{I}{d} \sum_{i=1}^{d} l(Y_i) g(Y_i). \tag{4.6}$$

Now, we consider the case of efficiently simulating a rare event probability $\mathbb{P}(L > c)$ corresponding to $g(y) = \boldsymbol{I}_{\{y>c\}}$ for c significantly larger than the mean of Y, where $\boldsymbol{I}_{(.)}$ is the indicator function. In particular, if $f_Y^{\mathbb{Q}}$ is a new probability density function of Y_i then we can write

$$\mathbb{P}(L > c) = \mathbb{E}^{\mathbb{P}}[\boldsymbol{I}_{\{L>c\}}] = \mathbb{E}^{\mathbb{Q}}[\boldsymbol{I}_{\{L>c\}} l(Y)]. \tag{4.7}$$

The quantity $\boldsymbol{I}_{\{L>c\}} l(Y)$ is an unbiased importance sampling estimator for $\mathbb{P}(L > c)$. We encounter the problem of how to choose $f_Y^{\mathbb{Q}}$ such that $l(Y)$ is small for $L > c$. In other words, we want the variance of $\boldsymbol{I}_{\{L>c\}}l(Y)$ to be the least possible. Fortunately, for this type of problem there is a fairly well-established approach, which we now describe.

4.2 Exponential Twisting

Definition 4.2.1 (Asymptotically efficient change of measure)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \mathbb{Q})$ be two equivalent probability spaces and Y be a random variable on them. Also, let the probability density function of Y be $f_Y^{\mathbb{P}}$ and $f_Y^{\mathbb{Q}}$ with respect to $\mathbb{P}$ and $\mathbb{Q}$ respectively. Then we say that the change of measure from $f_Y^{\mathbb{P}}$ to $f_Y^{\mathbb{Q}}$ is asymptotically efficient if*

$$\limsup_{c\to\infty} \left(\frac{\ln \mathbb{E}^{\mathbb{Q}}[\boldsymbol{I}_{\{L>c\}} l^2(L)]}{2 \ln \mathbb{P}(L > c)} \right) \geq 1. \tag{4.8}$$

This means that the exponential rate of decrease of the second moment is twice the exponential rate of decrease of the probability one is trying to estimate. Non-negativity of the variance implies that this is the fastest possible rate for any unbiased estimator.

For light-tailed random variables one may use a special change of measure that is obtained by exponential twisting. The exponential twisting technique is one of the widely used methods to find the IS probability density function $f_Y^{\mathbb{Q}}$ when Y has a light tailed density (see Huang et al., 2003). For the case of $f_Y^{\mathbb{Q}}$, the exponentially twisted density function by the amount z, $z > 0$, is

$$f_{z,Y}^{\mathbb{Q}}(y) := \frac{e^{zy} f_Y^{\mathbb{P}}(y)}{M_Y(z)}, \tag{4.9}$$

where $M_Y(z)$ is the moment generating function for Y and is given by

$$M_Y(z) = \mathbb{E}^{\mathbb{P}}[e^{zY}] = \int_{-\infty}^{\infty} e^{zY} f_Y^{\mathbb{P}}(y)dy \tag{4.10}$$

and we assume that it is finite for $z \in \mathbb{R}$. We recall that our goal is to find an IS density function so that one "achieves" exponential twisting on L by an amount z. If we are able to do that, then from our definition of exponential twisting (see equation (4.9)), then $l(y) = f_Y^{\mathbb{P}}(y)/f_{z,Y}^{\mathbb{Q}}(y) = e^{-zy}M_Y(z)$. Note that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[\boldsymbol{I}_{\{L>c\}}l^2(y)] &= \mathbb{E}^{\mathbb{Q}}[\boldsymbol{I}_{\{L\geq c\}}e^{-2zy}M_Y^2(z)] &(4.11)\\
&\leq e^{-2zy}M_Y^2(z) &(4.12)
\end{aligned}$$

Instead of solving the difficult problem of minimizing equation (4.11) over z, we choose z so that this bound becomes small (see the inequality in (4.12)) so that we can choose z that minimizes $e^{-2zy}M_Y^2(z)$ or equivalently $\ln M_Y(z) - zc$. Appropriate convexity properties hold (see Bucklew, 1990) so that the optimal solution, z^* is a solution of the equation

$$\frac{M_Y'(z)}{M_Y(z)} = c. \tag{4.13}$$

Define μ_z to be the mean of Y with respect to the density $f_{z,Y}^{\mathbb{Q}}$, i.e.

$$\mu_z := \mathbb{E}^{\mathbb{Q}}[Y] = \int_{-\infty}^{\infty} \frac{ye^{zy}f_Y^{\mathbb{P}}(y)}{M_Y(z)}dy = \frac{\mathbb{E}^{\mathbb{P}}[Ye^{zY}]}{M_Y(z)}. \tag{4.14}$$

We have

$$\frac{d}{dz}(\ln M_Y(z) - zc) = \frac{\mathbb{E}^{\mathbb{P}}[Ye^{zY}]}{M_Y(z)} - c = \mu_z - c \tag{4.15}$$

suggesting that $z^* = z(c)$ as a solution of the equation $\dfrac{M_Y'(z)}{M_Y(z)} = c$. Although c is in the tail of the original probability measure, it is near the centre of the new distribution when the IS technique is applied. In other words, the rare events $\{L > c\}$ are more probable to occur under the new probability measure $\mathbb{Q}$ and are no longer rare events.

4.3 Bernoulli Mixture Models and Exponential Twisting

We assume that the random vector $\boldsymbol{Y}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ follows a Bernoulli mixture model. Denote the individual default probabilities by $\bar{p}_i$, take $\Omega = \{0,1\}^d$ and define the probability measure $\mathbb{P}$ to be

$$\mathbb{P}(\{\boldsymbol{y}\}) := \prod_{i=1}^{d} \bar{p}_i^{y_i}(1-\bar{p}_i)^{1-y_i}, \quad \boldsymbol{y} \in \Omega = \{0,1\}^d. \tag{4.16}$$

In applying importance sampling, our main goal is to change the conditional marginal default probabilities from $\bar{p}_i$ to exponentially twisted one $p_{i,z}$, where z is the exponential twisting parameter. We focus on the measure changes that result from the exponential twisting of the loss L. The moment generating function of L is given by

$$M_L(z) = \mathbb{E}^{\mathbb{P}}\left[\exp\left(z\sum_{i=1}^{d} e_i Y_i\right)\right] = \prod_{i=1}^{d} \mathbb{E}^{\mathbb{P}}[e^{ze_iY_i}] = \prod_{i=1}^{d}(e^{ze_i}\bar{p}_i + 1 - \bar{p}_i).$$

Define the measure $\mathbb{Q}$ to be

$$\mathbb{Q}(\{\boldsymbol{y}\}) = \mathbb{E}^{\mathbb{P}}\left[\frac{e^{zL}}{M_L(z)}; \boldsymbol{Y} = \boldsymbol{y}\right], \tag{4.17}$$

so that we have

$$
\begin{aligned}
\mathbb{Q}(\{\boldsymbol{y}\}) &= \frac{\exp(z\sum_{i=1}^{d} e_i y_i)}{M_L(z)} \mathbb{P}(\{\boldsymbol{y}\}) \\
&= \prod_{i=1}^{d} \frac{\exp(ze_i y_i)}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i} \bar{p}_i^{y_i}(1-\bar{p}_i)^{1-y_i} \\
&= \prod_{i=1}^{d} \left(\frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{y_i} \left(\frac{1-\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{1-y_i} \\
&= \prod_{i=1}^{d} \left(\frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{y_i} \left(\frac{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i - \exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{1-y_i} \\
&= \prod_{i=1}^{d} \left(\frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{y_i} \left(\frac{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i} - \frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{1-y_i} \\
&= \prod_{i=1}^{d} \left(\frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{y_i} \left(1 - \frac{\exp(ze_i)\bar{p}_i}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\right)^{1-y_i} \\
&= \prod_{i=1}^{d} p_{i,z}^{y_i}(1-p_{i,z})^{1-y_i}
\end{aligned}
$$

where

$$
p_{i,z} := \frac{\exp(ze_i)}{\exp(ze_i)\bar{p}_i + 1 - \bar{p}_i}\bar{p}_i \tag{4.18}
$$

are the new default probabilities. If $z > 0$, then this does indeed increase the default probabilities; a larger exposure e_i results in a greater increase in the default probability. We also note that as $z \to -\infty$, $p_{i,z} \to 0$ and as $z \to \infty$, $p_{i,z} \to 1$, the original probabilities correspond to the value $z = 0$.

Conditionally independent default indicators. If we are given realizations of the economic risk factors $\boldsymbol{\psi}$ then we can estimate the conditional exceedance probability $q(\boldsymbol{\psi}) := \mathbb{P}(L > c)$ using the approach for independent default indicators. The algorithm is presented below.

Algorithm 4.3.1 (IS for conditional loss distribution)**.** *(1) Given $\boldsymbol{\psi}$ calculate the* *conditional default probabilities $h_i(\boldsymbol{\psi})$, and solve the equation*

$$\sum_{i=1}^{d} e_i \frac{\exp(ze_i)h_i(\boldsymbol{\psi})}{\exp(ze_i)h_i(\boldsymbol{\psi}) + 1 - h_i(\boldsymbol{\psi})} = c; \tag{4.19}$$

the solution $z = z(c, \boldsymbol{\psi})$ gives the optimal degree of tilting.

(2) Generate N conditional realization of the default indicator vector $(Y_1, \dots, Y_d)$. The defaults of the companies are simulated independently.

(3) Denote by $M_L(z, \boldsymbol{\psi}) := \prod_{i=1}^{d}\{\exp(ze_i)h_i(\boldsymbol{\psi}) + 1 - h_i(\boldsymbol{\psi})\}$ the conditional moment generating function of L. From the simulated default data construct N conditional realizations of $L = \sum_{i=1}^{d} e_i Y_i$ and label these $L^{(1)}, \dots, L^{(N)}$. Determine the IS estimator for the conditional loss distribution:

$$\hat{I}_N(\boldsymbol{\psi}) = M_L(z(c, \boldsymbol{\psi}), \boldsymbol{\psi}) \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{I}_{\{L \geq c\}} \exp(-z(c, \boldsymbol{\psi}) L_j). \tag{4.20}$$

IS for the risk factors: Variations in the loss distribution are caused by the economic risk factors, $\boldsymbol{\Psi}$. Since our main focus is on estimating the tail probabilities of the loss distribution, it is also reasonable to apply the importance sampling algorithm on the economic risk factors. Note that for the estimator $\hat{I}_N$ of $\mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi})$, we have the variance decomposition

$$\mathbb{V}(\hat{I}_N) = \mathbb{E}[\mathbb{V}(\hat{I}_N|\boldsymbol{\Psi})] + \mathbb{V}(\mathbb{E}[\hat{I}_N|\boldsymbol{\Psi}]). \tag{4.21}$$

Conditional on $\boldsymbol{\Psi}$, the obligors are independent and now we know how to apply an asymptotically optimal importance sampling. This makes $\mathbb{V}(\hat{I}_N|\boldsymbol{\Psi})$ small and suggests that in applying importance sampling to $\boldsymbol{\Psi}$, we should focus on the second term of the decomposition above. Now, since $\hat{I}_N$ is an estimator for $\mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi})$, then we have that $\mathbb{E}[\hat{I}_N] = \mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi})$, hence we must choose an importance sampling distribution for $\boldsymbol{\Psi}$ that will reduce the variance in estimating the integral of $\mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi})$ against the density of $\boldsymbol{\Psi}$. Also, applying importance sampling to the factors requires finding a good importance sampling density for the function $\boldsymbol{\psi} \to \mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi})$. From (4.9), the optimal importance sampling density $f^{\mathbb{Q}}_{z,\psi}$ must satisfy

$$f^{\mathbb{Q}}_{z,\psi} \propto \mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\boldsymbol{\psi}'\Omega^{-1}\boldsymbol{\psi}/2) \tag{4.22}$$

where $\propto$ means proportional to. But sampling from this density is generally infeasible- the normalization constant required to make it a density is the value $\mathbb{P}(L > c)$ we are looking for. This is the same problem that was faced by Glasserman et al. (1999) in the context of option pricing, hence we will not deviate away from their method. We also use the multivariate normal density $f^{\mathbb{P}}_{\psi}$ with the same mode as the optimal density $f^{\mathbb{Q}}_{z,\psi}$ as an approximation of the optimal importance sampling density. This mode occurs at the solution to the optimization problem

$$\max_{\psi} \mathbb{P}(L > c|\boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\boldsymbol{\psi}'\Omega^{-1}\boldsymbol{\psi}/2), \tag{4.23}$$

which is then also the mean vector $\boldsymbol{\mu}$ of the approximating multivariate normal distribution. This is to say that $f^{\mathbb{Q}}_{z,\psi}$ should be a density of $N_p(\boldsymbol{\mu}, \Omega)$ so that the likelihood ratio is given by

$$l_{\boldsymbol{\mu}}(\boldsymbol{\psi}) = \frac{\exp(-\boldsymbol{\psi}'\Omega^{-1}\boldsymbol{\psi}/2)}{\exp(-(\boldsymbol{\psi} - \boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\psi} - \boldsymbol{\mu})/2)} = \exp(-\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\psi} + \boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\psi}\boldsymbol{\mu}). \tag{4.24}$$

Now we are in a position to give the overall importance sampling algorithm.

Algorithm 4.3.2 (Full importance sampling). *(1) Generate* $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n \sim N(\boldsymbol{\mu}, I_p)$.

(2) For each $\boldsymbol{\psi}_i$ *calculate* $\hat{I}_N(\boldsymbol{\psi})$

(3) determine the overall importance sampling estimator:

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} l_{\boldsymbol{\mu}}(\boldsymbol{\psi})\hat{I}_N(\boldsymbol{\psi})$$

.

Our main interest is to estimate VaR and ES, hence the estimation of loss probabilities is not enough. We need to estimate quantiles of the loss distribution. The empirical quantile function is the quantile function of F_L and thus given by

$$\hat{F}_L^{-1}(\alpha) = \inf\{l \in \mathbb{R} : \hat{F}_L(l) \geq \alpha\}, \tag{4.25}$$

for $\alpha \in (0,1)$. The main challenge we are facing here is how to calculate $\hat{F}_L^{-1}(\alpha)$. There are a number of methods already set in place that can be used to solve this problem. The method proposed by Glasserman et al. (2002) can be used, but in this work, we choose to use a more direct approach proposed by Glynn (1996). We continue with the notations given in Algorithm 4.3.1. We order the losses $L^{(1)}, \ldots, L^{(N)}$ in an increasing form such that we have $L_N^{(1)} \geq L_N^{(2)} \geq \cdots \geq L_N^{(N)}$. Then we can write

$$\begin{aligned}
\hat{F}_L^{-1}(\alpha) &= \inf\left\{ l \in \mathbb{R} : \sum_{i=1}^{N} \boldsymbol{I}_{\{L_N^{(i)} \leq l\}} \geq N\alpha \right\} \\
&= L_N^{(j)} \quad \text{for some } j \in \{1, \ldots, N\}
\end{aligned}$$

Note that in the above equation, we used the fact that the estimator of F_L is $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}_{\{L_N^{(i)} \leq l\}}$. Since for $j \in \{1, \ldots, N\}$

$$\sum_{i=1}^{N} \boldsymbol{I}_{\{L_N^{(i)} \leq L_N^{(j)}\}} = \sum_{j=1}^{N} \boldsymbol{I}_{\{L_N^{(i)} \leq L_N^{(j)}\}} = N - j + 1 \tag{4.26}$$

we need to find j such that $N - j + 1 \geq N\alpha$, hence $j = [N(1-\alpha)+1]$. Therefore

$$\hat{F}_L^{-1}(\alpha) = L_N^{(N(1-\alpha)+1)}. \tag{4.27}$$

We recall that $\hat{F}_L^{-1}$ is a piecewise constant function on $(0,1)$, so we have that $\hat{F}_L^{-1}(\alpha) = L_N^{(1)}$ if $\alpha \in (1-1/N, 1)$ or $\hat{F}_L^{-1}(\alpha) = L_N^{(i)}$ if $\alpha \in (1-i/N, 1-(i-1)/N]$ for some $i \in \{2, \ldots, N\}$. Using (4.27), we estimate VaR and ES of our portfolio by

$$V\hat{a}R_\alpha = \hat{F}_L^{-1}(1-\alpha) = L_N^{(N\alpha+1)} \tag{4.28}$$

and

$$\hat{ES}_\alpha = \frac{1}{\alpha}\int_{1-\alpha}^{\alpha} \hat{F}_L^{-1}(p)dp \tag{4.29}$$

respectively, for $\alpha \in (0,1)$.

CHAPTER 5

Empirical Studies

In this chapter, we present numerical examples and show how the theoretical ideas of GLMM can be implemented in practice. We also give some results on the estimation of VaR and Expected Shortfall of a specific portfolio. We stick to the assumptions and notations presented in Chapter 3 and Chapter 4.

Numerical Example 1. We give one example of a five-different credit rating model with a binomial default counts using R. The credit default data that has been used in this paper was retrieved from the Standard and Poor's database. The default counts have been collected for one year periods, ranging from January 1981 to December 2000 ($T = 20$). The credit rating classes that were considered include A, B, BB, BBB and CCC hence $\kappa = 5$. Obligors in high rating classes hardly default, as such, including them in this paper would not serve the purpose of the study. Table 5.1 shows the total number of obligors for each credit rating class together with the number of defaulted obligors in the 20-year period.

Table 5.1: Parameter estimates for a one-factor Bernoulli mixture model fitted to historical S & P 500 one-year default data

Rating class	Number of obligors	Number of defaulted obligors
A	14 857	6
BBB	10 258	23
BB	7 231	91
B	7 606	403
CCC	784	172

We recall the Bernoulli mixture model discussed in Section 3.5. We consider a simple Bernoulli mixture model and we choose the response function g to be the probit function so that

$$\mathbb{E}(Y_{i,t}|\Psi_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta_{i,t}} \exp\left(-\frac{u^2}{2}\right) du, \quad \zeta_{i,t} = \mu_r + \Psi_t,$$

and it is easy to see that this in line with the one factor model of Vasicek (1997). We begin by plotting the default rates of credit rating over the time period (20 years) and Figure 5.1 shows plots of time periods against the default rates of each obligor. We calculated the default rate (DefRate) of each obligor for all the periods using

$$DefRate = \frac{M_{r,t}}{m_{r,t}}.$$

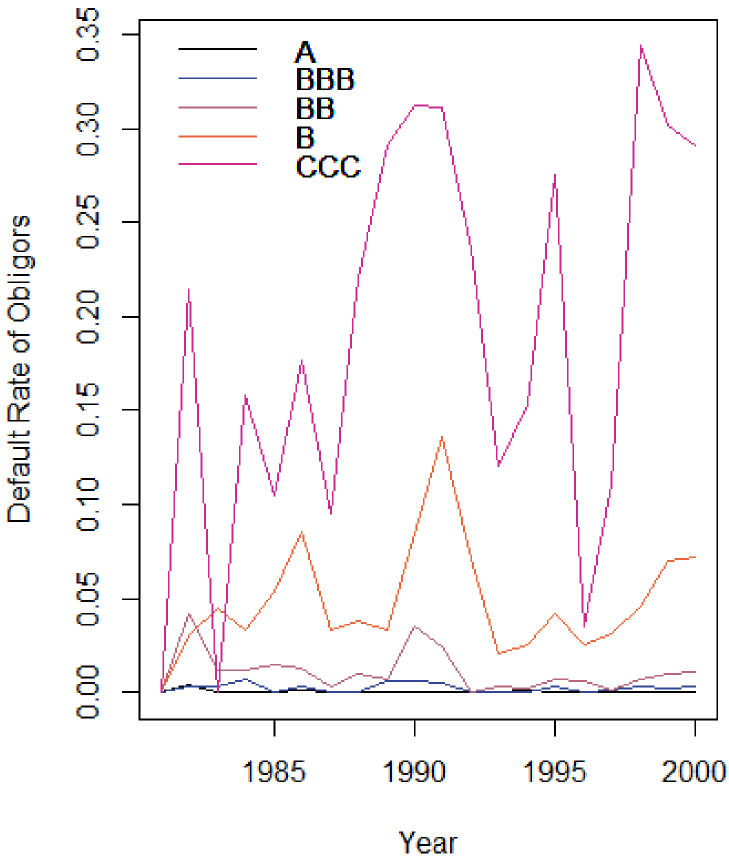

Figure 5.1: Empirical default probabilities according to S & P 500 data for 20 years (1981 - 2000)

We observe that there are cyclic variations in the default occurrences in each of the credit rating classes. These variations are as a result of non-obligor specific risk factors. To put it in another way, the increased number of defaults between the recessions of 1990 and 1993 can be attributed to global risk factors. As expected, the lower rated classes such as CCC possess high default rates compared to high rated credit rating classes (such as A). Over the years, the default rate of the A credit rating class is almost zero and this is the reason why we considered only 5 credit rating classes since obligors with higher credit ratings (than the credit rating A) rarely default, so we were not going to observe anything for their plots.

The model parameter estimate $\hat{\boldsymbol{\mu}}_r = (\hat{\mu}_{CCC}, \hat{\mu}_B, \hat{\mu}_{BB}, \hat{\mu}_{BBB}, \hat{\mu}_A)$ are presented in Table 5.2, with p- values less than 2×10^{-16}. We have $\boldsymbol{\mu}_r = (\hat{\mu}_{CCC}, \hat{\mu}_B, \hat{\mu}_{BB}, \hat{\mu}_{BBB}, \hat{\mu}_A) = (-0.84, -1.69, -2.40, -2.92, -3.43)$. We see that the estimates decrease with an increase in creditworthiness. For the credit rating class A, the $\hat{\mu}_A = -3.43$ and this means that the credit rating class A is associated with 3.43 lower log-odds than the other credit rating categories for default, compared to non-default.

Table 5.2: Bernoulli mixture model parameter estimates

Parameters	A	BBB	BB	B	CCC
$\hat{\mu}_r$	-3.43	-2.92	-2.40	-1.69	-0.84
S.E	0.13	0.09	0.07	0.06	0.08
Odd-ratio	3.2 %	5.4 %	9.0 %	18.5 %	43.3 %
p-values	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$

To find the odd ratios, we exponentiate estimated value as presented in Table 5.2. The odds ratio for the A credit rating class was found to be 3.2 %, which means that for "1 unit increase" of the credit rating class A to a higher rating class, we expect to see (approximately) 96.8 % decrease in the odds of the total number of defaulting obligors. The same analysis applies to the other credit rating classes. As expected, low rated credit rating classes have very high odd ratios.

The default probability estimates were calculated and are presented in Table 5.3. We can see that the default probabilities decrease with increasing creditworthiness.

Table 5.3: Estimated default probabilities

Parameter	A	BBB	BB	B	CCC
$\hat{\pi}_1^{(r)}$	0.0042	0.0230	0.0965	0.0492	0.2078

Default probabilities for low rated credit classes are higher than those that are highly rated. Obligors with a credit rating C are more likely to default compared to those with a credit rating on A. Table 5.4 gives the calculated estimates for the within credit rating and between credit rating default correlations.

Table 5.4: Estimates of within and between credit rating default correlations

$\hat{\rho}_Y^{(r,s)}$	A	BBB	BB	B	CCC
A	0.0004				
BBB	0.0021	0.0015			
BB	0.0013	0.0037	0.0043		
B	0.0021	0.0043	0.0057	0.0143	
CCC	0.0029	0.0157	0.0108	0.0061	0.0284

From the table, we see that the within credit rating default correlations increase with decrease in credit rating class and visa versa. For the between credit rating default correlations, the estimates vary depending on the pair of credit classes considered. Note that the default correlations are correlations between event indicators for very low probability events and are necessarily very small.

The hyperparameter estimate $\alpha = \hat{\sigma} = 0.24$ suggests that there is some significant variation within the random effects Ψ_t. In our model, we have assumed that the variance of the random effects is the same for all firms in all years and as such, we might be concerned that the model does not allow for enough heterogeneity in the variance of the systemic risk of the portfolio. Moreover, while controlling credit rating classes and the repeated measures within the time periods, we have found evidence that there is an association of credit rating with credit default of obligors. The p values presented in Tale 5.2 are statistically significant, so we can say that if there was no association of credit default and credit rating, then the probability of observing the S & P 500 data we have used in this work is less than 2×10^{-16} (almost zero).

In McNeil and Wendin (2007) the same data was fitted using the Gibbs sampler and virtually yielding similar results.

Numerical Example 2. In this example, we show how the algorithms from Section 4 can be applied to a specific loan or bond portfolio. We construct a portfolio that can be used to illustrate importance sampling and these constructions are frequently used in credit risk management studies where historical data is scarce to find. We follow the construction given in Glasserman et al. (1999). We consider a homogenous portfolio with 100 obligors, with each obligor having an exposure of 1 (this can be in any currency, for example, Ksh 1 million). We also assume that each obligor has a default probability of 0.05. We set the rare event threshold c to be 20. The aim is to calculate the tail probability $\mathbb{P}(L \geq 20)$, and in such a simple model, the tail probability can be calculated analytically and is 0.00112. We compare the Monte Carlo/Importance Sampling methods; (i) naive Monte Carlo (no importance sampling) (ii) importance sampling on the risk factor distribution (outer importance sampling) (iii) importance sampling on the conditional default distribution (inner importance sampling) and (iv) importance sampling on both risk factor distribution and conditional default distribution (full importance sampling).

To start with, we began by finding the optimal mean (solution to the optimization problem in equation 4.23). Figure 5.2 shows the effects of the importance sampling algorithm in estimating the rare event probability. Figure 5.2 (a) shows us the convergence of the Monte Carlo method without applying the importance sampling technique. In Figures 5.2 (b) and 5.2 (c), we show the convergence of the Monte Carlo method when the importance sampling algorithm is applied to the losses and the risk factors respectively. In Figure 5.2 (d), we show the convergence of the Monte Carlo method when the importance sampling technique is applied to both the losses and the risk factors. We observe that fast convergence is attained in the case where we have applied importance something in both the losses and the risk factors.

In the case where we did not apply importance sampling, there was no convergence in 5000 simulations. We also observe that convergence is much faster in the case when we apply the importance sampling on the losses compared to the case of risk factors.

In Figure 5.3 , we show the comparison on the convergence of the different cases we have considered. It is clear that in the case where we have applied the importance sampling on the losses and risk factors, the rate of convergence is faster than the other cases.

Table 5.5: Standard errors for the estimates of the tail probability

Method	S.E.
Naive MC	1.7×10^{-4}
Outer IS	1.3×10^{-4}
Inner IS	1.2×10^{-4}
Full IS	8.8×10^{-6}

This provides us with clear evidence that the importance sampling methods can indeed help speed up the Monte Carlo method during the estimation of a rare event probability. In Table 5.5 we provide the standard errors of the different methods used for calculating the tail probability $\mathbb{P}(L \geq 20)$ and we observe that the full importance sampling method has the least standard error.

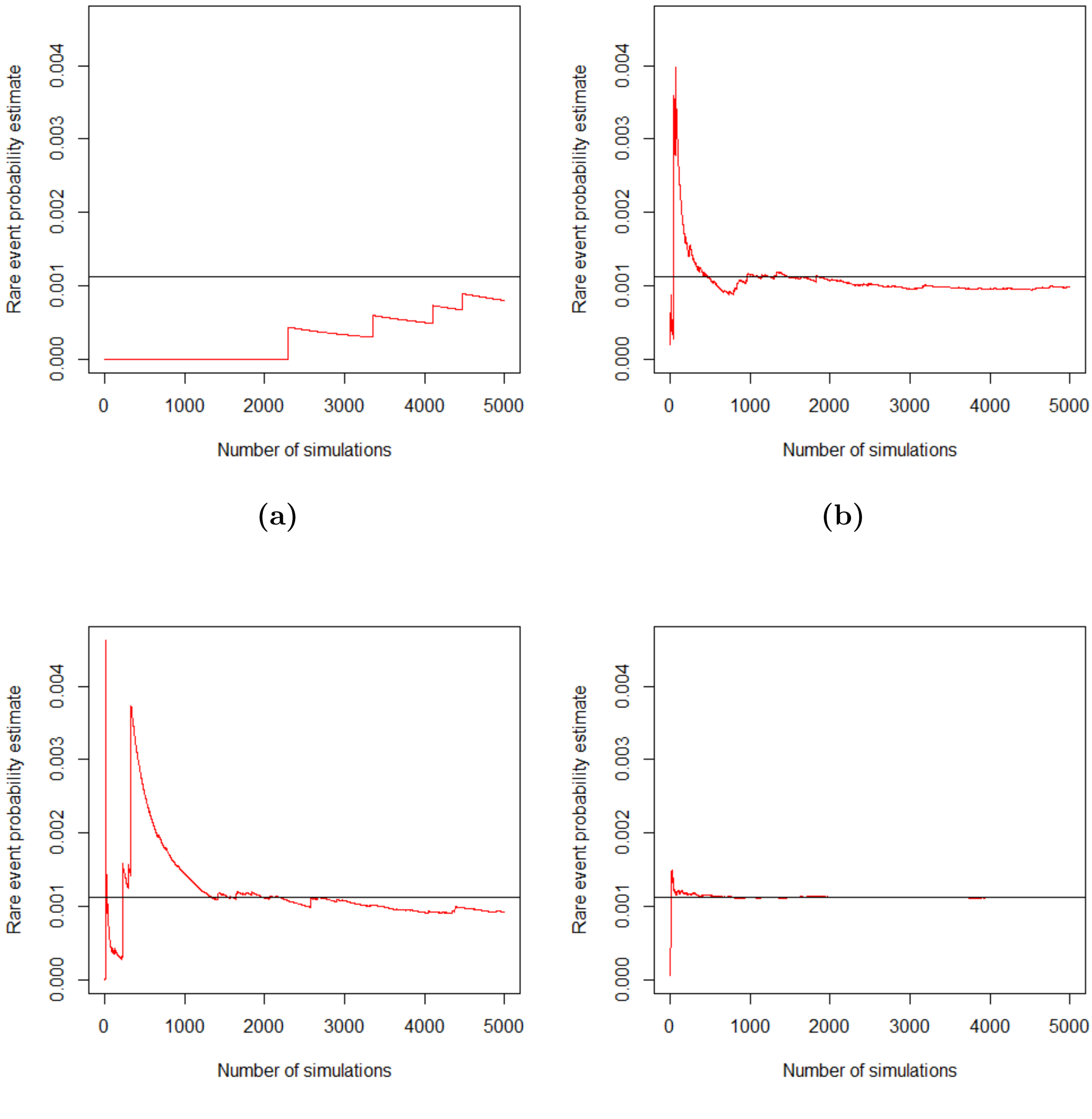

Figure 5.2: Plots for the Monte Carlo/Importance Sampling methods

In Figure 5.4 and Figure 5.5, we compare the loss distributions under the Gaussian and the t copula models. We observe that under the t copula model (with 5 degrees of freedom), the probability of getting a loss $L = 5, 10, \ldots, 100$ is much higher compared to the Gaussian copula model. These results tell us that the t copula model is indeed a suitable to model rare event probabilities compared to the Gaussian copula

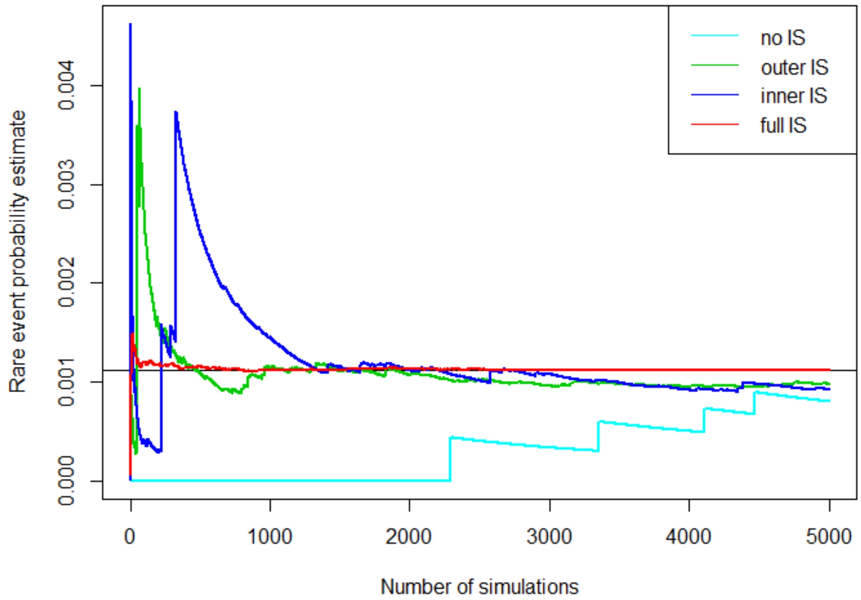

Figure 5.3: Comparison of the convergence of the Monte Carlo/Importance Sampling methods in calculating rare events probability

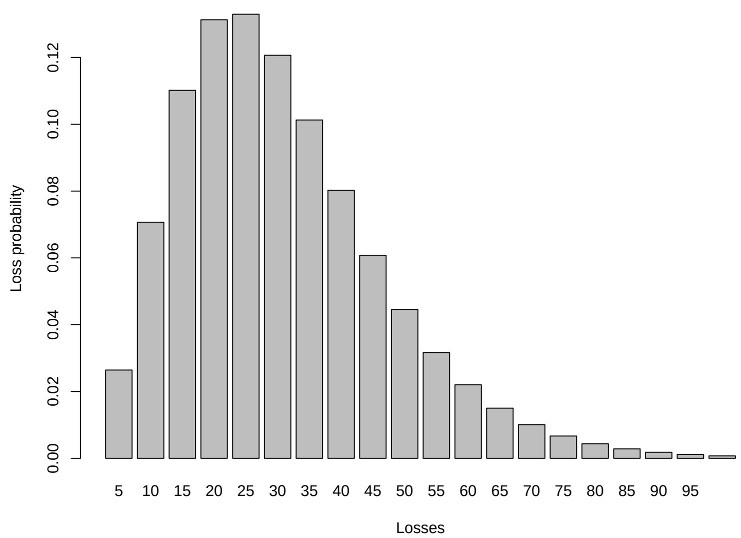

Figure 5.4: Loss distribution for the Gaussian copula model

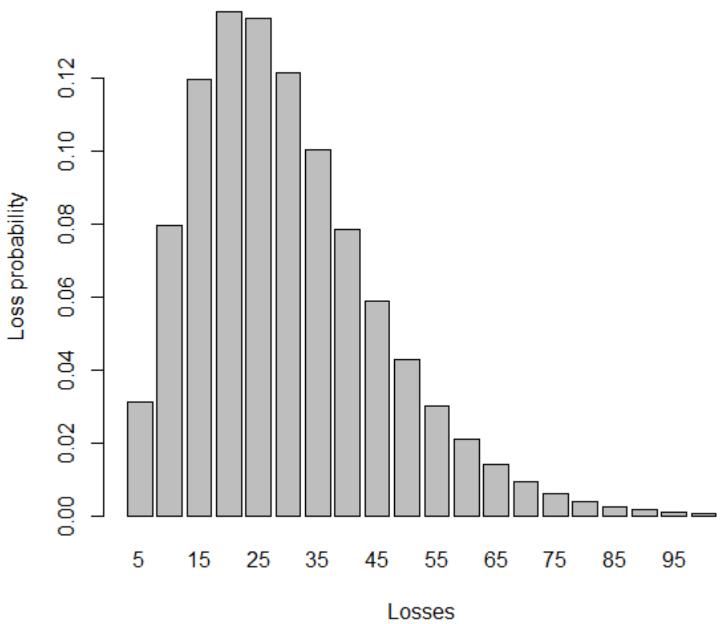

Figure 5.5: Loss distribution for the t copula model

as it can give high enough chances for the occurrence of extreme events. Note that the maximum loss we can have is 100.

Using equation (4.28), we obtained the value at risk of our portfolio at 95% confidence and we used equation (4.29) to find the expected shortfall of our portfolio under the Gaussian and also under the $t-$copula models.

Table 5.6: Portfolio risk measures for Gaussian and $t-$copula models

Model	$\widehat{VaR}$	$\widehat{ES}$
Gaussian copula	20	25
$t-$copula	25	30

We see that both the VaR and ES estimates are higher in the $t-$copula model compared to the Gaussian copula model. Under the Basel II capital-adequacy framework, a bank is required to hold 8% of the so-called risk-weighted assets of its credit portfolio as risk capital. The risk weighted assets are considered to be the asymptotic contribution of risk of 95% VaR of the overall portfolio. As such, we see that under the $t-$copula model, a bank will have to allocate more money that will act as risk capital during extreme events, compared to the Gaussian copula model.

CHAPTER 6

Conclusion and Recommendations

6.1 Conclusion

From the results we presented in Section 5 (numerical example 1), we can say that Bernoulli mixture models can be used to model dependent credit defaults events. The default probabilities and the default correlations are very important in credit risk management for the calculation of the capital allocations for firms. We observe that default probabilities increase with a decrease in the credit rating of a particular firm, that is, higher rated firms are unlikely to default while firms with a low credit rating are more prone to default. We also observe that if a firm with a low credit rating defaults, other firms with a low credit rating are more likely to default, and this is explained the values of the estimates of within and between credit rating default correlations presented in Table 5.4.

In the numerical example 2, our main focus was to estimate the loss distribution so that we can be able to find estimates of the portfolio risk measures. We see that under the t copula model, the loss distribution has fat tails compared to the Gaussian copula model. This means that under the t copula model, risk managers can be able to model extreme events since the model can give high enough chances in which everything goes wrong. This translates directly to cases of catastrophic events such as, for example, the COVID-19 pandemic. The estimated VaR of our portfolio was 30 and the estimated ES of our portfolio was 45.

6.2 Recommendations

The Bernoulli mixture model we looked at is relatively simple and used the probit model as the response function, but it allows for a handy formulation of systematic portfolio risk in terms of observed fixed effects and unobserved random effects in order to capture inhomogeneities in default rates throughout the portfolio and across time. Instead of the probit model, it is still possible to use a logit response function. We solely took into account default and non-default results in this study, however, it is still possible to extend the model by incorporating cases of credit migration of obligors from one credit rating to another (upgrades and downgrades). We have also mentioned that GLMMs apply to a number of well-known industry models used to model credit defaults. It would be reasonable to incorporate additional random effects in the model and allow more heterogeneity provided we had more information on the industrial and geographical sectors to which the obligors belonged.

6.3 Recommendations for Further Research

In reality, obligors can default more than once during a specific time period, as a result, further studies may focus on approximating the binomial distribution by the Poisson distribution thereby allowing obligors to default more than once within a specific time period. Moreover, in this study we only considered the case where the IS density is the same at each iteration. However, it is still possible to extend this method by considering the case where at each iteration, we update the IS density, ie, instead of using the same importance sampling density for all the simulations, it may be reasonable to use $f^{\mathbb{Q}}_{X_i}(X_i)$ to draw X_i for $i = 1, \ldots, d$. In this way it is our intention to make this literature more accessible to researchers in the field of quantitative risk management.

References

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

Ankudinov, A., Ibragimov, R., and Lebedev, O. (2017). Heavy tails and asymmetry of returns in the Russian stock market. *Emerging Markets Review*, 32:200–219.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228.

Behr, A. and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models. *Annals of Finance*, 5(1):49–68.

Boston, C. S. F. (1997). Creditrisk+: A credit risk management framework. Technical report, Technical report, Credit Suisse First Boston.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2001). Copulas: an open field for risk management. *Available at SSRN 1032557*.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Bucklew, J. A. (1990). Large deviation techniques in decision, simulation, and estimation. *Wiley Series in Probability and Mathematical Statistics*.

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

D'Amico, G. and Petroni, F. (2018). Copula based multivariate semi-markov models with applications in high-frequency finance. *European Journal of Operational Research*, 267(2):765–777.

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, 73(1):111–129.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

Embrechts, P., Mcneil, E., and Straumann, D. (1999). Correlation: pitfalls and alternatives. In *Risk Magazine*. Citeseer.

Fernández, A., Rumí, R., and Salmerón, A. (2012). Answering Queries In Hybrid Bayesian Networks Using Importance Sampling. *Decision Support Systems*, 53(3):580–590.

Frey, R. and McNeil, A. J. (2001). Modelling dependent defaults. Technical report, ETH Zurich.

Frey, R. and McNeil, A. J. (2003). Dependent defaults in models of portfolio credit risk. *Journal of Risk*, 6:59–92.

Frey, R., McNeil, A. J., and Nyfeler, M. (2001). Copulas and credit models. *Risk*, 10(111114.10).

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (1999). Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical finance*, 9(2):117–152.

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

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656.

Glynn, P. W. (1996). Importance sampling for Monte Carlo estimation of quantiles. In *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, pages 180–185. Citeseer.

Gourieroux, M., Gourieroux, C., Monfort, A., Monfort, D. A., et al. (1996). *Simulation-based econometric methods*. Oxford university press.

Grundke, P. (2007). Computational aspects of integrated market and credit portfolio models. *OR Spectrum*, 29(2):259–294.

Gupton, G. M., Finger, C. C., and Bhatia, M. (1997). *Creditmetrics: technical document*. JP Morgan & Co.

Huang, Z. et al. (2003). Rare-event, heavy-tailed simulations using hazard function transformations, with applications to value-at-risk. In *Proceedings of the 2003 Winter Simulation Conference, 2003.*, volume 1, pages 276–284. IEEE.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Kou, G., Peng, Y., and Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences*, 275:1–12.

Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income*, 9(4):43–54.

Lindsey, J. K. (2000). *Applying generalized linear models*. Springer Science & Business Media.

Mai, J. and Scherer, M. (2014). *Financial engineering with copulas explained*. Springer.

McNeil, A. J., Frey, R., and Embrechts, P. (2015a). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

McNeil, A. J., Frey, R., and Embrechts, P. (2015b). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

McNeil, A. J. and Wendin, J. P. (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance*, 14(2):131–149.

Merino, S. and Nyfeler, M. A. (2004). Applying importance sampling for estimating coherent credit risk contributions. *Quantitative Finance*, 4(2):199–207.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2):449–470.

Molins, J. and Vives, E. (2016). Model risk on credit risk. *Risk and Decision Analysis*, 6(1):65–78.

Mönkkönen, H. M. (1998). *Modeling default risk: Theory and empirical evidence*.

Morgan, J. P. (1997). Creditmetrics-technical document. *JP Morgan, New York*.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Schweizer, B. and Sklar, A. (2011). *Probabilistic metric spaces*. Courier Corporation.

Sklar, M. (1959). Distribution functions in dimensions and their margins. *Publ. inst. statist. univ. Paris*, 8:229–231.

Straumann, D. Y. (2001). Correlation and dependency in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*. Cambridge University Press.

Suisse, C. (1997). Creditrisk+, a credit risk management framework. credit suisse financial products, london.

Trivedi, P. K., Zimmer, D. M., et al. (2007). Copula modeling: an introduction for practitioners. *Foundations and Trends® in Econometrics*, 1(1):1–111.

Vasicek, O. (1997). The loan loss distribution. *Preprint, KMV Corporation.*