The University of New South Wales
School of Mathematics and Statistics

Honours Thesis

On Exploring Dependence Structures for Operational Risk Models

Author
Chen Yang

Supervised by
Gareth W. Peters

22nd October 2010

Abstract

In this thesis we explore various facets of modelling dependencies between operational risk factors through the use of copula models, which are multivariate distributions with uniform marginals. We employ the loss distribution approach to capture the characteristics of risk factors, in which losses are described by a compound process. The implications of imposing different dependence structures in this setting are explored with respect to induced correlations between annual losses and their joint value-at-risk. In particular, we observe under what circumstances capital reduction can be achieved and when negative diversification benefits are realized. Investigations into methods of model selection and parameter estimation for dependence models in this context are also carried out using maximum likelihood estimation, Bayesian inference and selected goodness-of-fit tests.

Acknowledgments

I am grateful for my supervisor Gareth's help throughout the year, without which I would have not been able to complete this thesis. I know he is a busy man so I really appreciate the time he has taken out for our weekly meetings as well as the time he has spent proof reading my work.

I would also like to thank my work colleagues who introduced me to the concept of copulas before I took on this project. I guess it was part of the reason why I wanted to research this area in the first place.

A special thanks goes out to my lecturers this year, especially to Donna who has kept coursework manageable so I never felt overburdened.

Finally my family deserves my greatest gratitude for supporting me, not just through this year, but all throughout my life.

Introduction

Motivation and Aims

The regulatory framework Basel II dictates the modelling of operational risk for all banks. Dependence modelling is allowed in this context to obtain a capital reduction which has therefore incentivised financial institutions to incorporate dependence modelling into their business. In the area of operational risk, data is typically rare and so dependence modelling is not an subject that has been extensively explored.

The main aims of this thesis are:

- To investigate the effects of multi-risk dependence modelling on the induced correlation between the underlying operational risk factors and the consequent implications on the capital requirements and diversification benefits of the bank.
- To analyze statistical dependence models for multi-risk processes in the context of OpRisk.
- To explore effective methods of parameter estimation and model selection for different marginal and dependence models.

Chapter Overview

Chapter 1

We introduce the concept of operational risk as defined by the regulatory framework Basel II and outline its main properties. We also look at areas of Basel II which are relevant to the modelling of operational risks and describe the quantitative risk measure that is prescribed in this setting.

Chapter 2

We set up the basic model for the annual loss process associated with operational risk and present some relevant statistical distributions that will be used throughout the thesis. We then turn our attention to dependence models and, in particular, focus on the concept of copula modelling. We look at the historical development of copula theory as as well as its advantages and shortcomings. We present the properties of several basic and advanced parametric copula models and discuss their relevance for operational risk.

Chapter 3

We first explore the characteristics of the copula models presented in Chapter 2 through a visual approach. Scatterplots of bivariate and trivariate copulas are generated to give intuition for the types of dependence shapes that are modelled by each copula. In the second part of the chapter, we introduce dependence into the annual loss process using a variety of copula models and study the implications that dependence modelling has on the capital requirements of banks through Monte Carlo simulations. Matlab 2009b was used as the software package for all programming done in this thesis.

Chapter 4

We describe two approaches to model fitting that can be carried out in operational risk settings: maximum likelihood estimation and Bayesian inference. For each of these methods, we outline algorithms that can be used for parameter estimation. We will focus on the Monte Carlo Markov Chain approach for Bayesian estimation as models for operational risk typically have complex forms which makes simulation using simpler techniques such as inversion and acceptance-rejection impossible.

Chapter 5

We study the accuracy of estimators corresponding to the methods outlined in Chapter 4 by fitting models to data that have been generated from known distributions. For maximum likelihood estimation, we perform a 2-step procedure outlined by Harry Joe in (**?**). In the Bayesian approach, we develop and utilize models based on a slice sampling simulation technique proposed by Radford Neal in (**?**). The results from the two techniques are then compared and assessed from an operational risk modelling perspective.

Chapter 6

We first describe currently available model selection techniques in the form of goodness-of-fit tests as seen in (**?**). We then perform a case study for copula model selection by fitting known data from copula models to misspecified copulas using maximum likelihood estimation. We apply one of the goodness-of-fit tests outlined in this chapter to examine both the flexibility of the copula model and the power of the test.

Terminology and Notation

The table below lists most of the common notation and abbreviations that will appear throughout this thesis.

Term	Definition
OpRisk	Operational Risk
VaR	Value at risk
LDA	Loss Distribution Approach
$E[X]$	Expectation of the random variable X
$Var[X]$	Variance of the random variable X
$Cov[X, Y]$	Covariance between the random variables X and Y
τ	Kendall's Tau rank correlation
$F_X(x)$	The cumulative distribution function of the random variable X
$F_i(x)$	The ith marginal distribution function of the random variable X
$f_X(x)$	The probability density function of the random variable X
$f_i(x)$	The ith marginal density function of the random variable X
$C(u)$	The distribution function of a copula model
$c(u)$	The density function of a copula model
$X^{(j)}$	A quantity X that corresponds to risk cell j
$L(Y\|\gamma)$	The likelihood of data Y parameterized by the vector $\boldsymbol{\gamma}$
$\pi(\theta)$	The prior density of the random variable θ
$\pi(\theta\|Y)$	The posterior density of the random variable θ given data Y

Contents

Chapter 1

Operational Risk and Basel II

1.1 Operational Risk

Financial regulators require banks to hold a minimum amount of capital to guard against losses associated with particular risks. The main risks faced by banks in descending order of exposure are credit risk, operational risk and market risk.

Figure 1.1: Risk exposure of typical banks

The international regulatory requirements for the banking industry Basel II defines operational risk (OpRisk) in its Section V-A as:

The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk but excludes strategic and reputation risk.

It has only been in the last decade or so that operational risk has been identified as an area that should be quantified and managed actively. Basel II recognizes 7 distinct event types within 8 business lines for a total of 56 OpRisk factors. These are listed below:

Event Types

- Internal Fraud
- External Fraud

- Employment Practices and Workplace Safety
- Clients, Products, & Business Practice
- Damage to Physical Assets
- Business Disruption & Systems Failures
- Execution, Delivery, & Process Management

Business Lines

- Corporate finance
- Trading & Sales
- Retail banking
- Commercial banking
- Payment & Settlement
- Agency Services
- Asset management
- Retail brokerage

Some operational risks, such as employment practices and internal fraud, can be managed through methods like improved internal control systems and staff training whilst others like infrastructure failure and natural disasters cannot be managed at all. However, while risk management procedures exist, they cannot prevent all losses and companies will need to hold sufficient capital to endure any damages. The amount of capital that is required to be held needs to be quantitatively modelled and must satisfy the guidelines set out by Basel II, which is administered by APRA in Australia.

Losses associated with OpRisks possess several common characteristics that need to be considered when they are modelled. These include:

- Positivity - Losses are positive by definition. However losses can also be truncated at amounts above zero due to reporting standards, for instance losses of small magnitude may not always be recorded.
- Heavy tailedness - OpRisks losses can have significant magnitudes so models used must have the capability of accounting for extreme events, for example natural disasters can cause millions of dollars worth of damage to physical assets.
- Low frequency - Due to the infrequent nature of OpRisks, data is limited and this presents a potential problem in model fitting.

1.2 Basel II

Basel II (**?**) is a regulatory framework which banks must follow in order to manage the losses they potentially face through exposure to various risks. Although Basel II covers almost all facets of a bank's operations, only the section relevant to dealing with operational risk will be described, and only the quantitative aspects will be focused upon. The Basel II framework consists of three pillars:

1. **Pillar I: Minimum Capital Requirements.** This pillar requires banks to allocate a minimum amount of capital at all times for operational risk. This amount can be calculated using various approaches which will be detailed below.

2. **Pillar II: Supervisory Review Process.** This pillar focuses on the regulator's supervision of banks systems and capital adequacy.

3. **Pillar III: Market Discipline.** The pillar establishes market discipline by requiring that all relevant risk measures and methods of risk management be publicly disclosed.

The main focus of this thesis is the statistical modelling required under Pillar I. Under this Pillar, three approaches can be adopted to model the annual charge C for operational risk:

1. **The Basic Indicator Approach:**

$$C = \alpha \frac{1}{n} \sum_{j=1}^{n} \max\big(GI(j), 0\big) \tag{1.1}$$
$$n = \sum_{j=1}^{n} \mathbf{1}_{GI(j)>0}$$

where $GI(j)$ is the gross annual income over the previous years, n is the number of those previous years with positive gross income, $\mathbf{1}$ is the indicator function and $\alpha = 0.15$. $n = 3$ is specified by Basel II.

2. **The Standardized Approach:**

$$C = \frac{1}{n} \sum_{j=1}^{n} \max\left(\sum_{i=1}^{8} \beta_i GI_i(j), 0 \right) \tag{1.2}$$

where β_i are factors for eight business lines (BL) and $GI_i(j)$ are the annual gross incomes of the i^{th} BL in the previous years.

3. **The Advanced Measurement Approaches (AMA):**
The bank can use their own internally developed model to calculate the capital charge provided it is approved by the regulator.

The internal model developed by banks must satisfy the following criteria:

- The model must use internal and external data as well as scenario analysis and factors reflecting the internal and external business environment.
- The capital charge must be within the 99.9% confidence level of the risk measure used for a one year holding period.
- Diversification benefits can only be incorporated if an adequate level of dependence modelling is demonstrated.
- Insurance related capital reduction is limited to a maximum of 20%.

Activities which have exposure to operational risk can be mapped to the Basel II Matrix. The matrix is divided into seven risk types and eight business lines, making up a 56 cell matrix. The BLs in the matrix are known as Level 1 BLs and can be broken down into Level 2 and Level 3 BLs. The most popular approach for AMA is the Loss Distribution Approach (LDA), which models the losses associated with each risk factor cell as a compound process. The total loss of a bank due to OpRisk is the aggregate of losses across all of the 56 risk factor cells.

This thesis will only consider the LDA framework. Much scope for statistical model development is afforded to the quantitative analyst in this area of risk modelling by regulation. One of the challenges addressed by this thesis involves a detailed study of basic and advanced statistical dependence models in this domain.

1.3 Risk Measures

A risk measure is used to determine the amount of capital that needs to be kept in reserve in order to make the risks taken by financial institutions, such as banks and insurance companies, acceptable to the regulator. Ideally, a risk measure, denoted by the function M, should be coherent, meaning it possesses the following properties (**?**):

- Subadditive: $M[X+Y] \leq M[X] + M[Y]$,
- Monotone: if $X \leq Y$ then $M[X] \leq M[Y]$,
- Positive Homogeneous: if $c > 0$ then $M[cX] = cM[X]$,
- Translation Invariant: if $c > 0$ then $M[X+c] = M[X] + c$.

The two most commonly used risk measures in risk modelling are the Value-at-Risk (VaR) and Expected Shortfall (ES).

The α% VaR is calculated as the αth quantile of the loss distribution.

$$VaR_\alpha(X) = F_X^{-1}(\alpha) \tag{1.3}$$

The 99.9% VaR is the prescribed risk measure according to Basel II however it is not a coherent risk measure, in particular the sum of individual VaRs can exceed the VaR of the sum of individuals.

The α% ES is the mean of the loss distribution above the αth quantile. For continuous distributions, this is given as:

$$ES_\alpha[X] = E[X|X \geq VaR_\alpha(X)] \tag{1.4}$$

The ES is a popular alternative and is a coherent risk measure. However, calculating the ES requires the tail information of the loss distribution which are often inaccurate and this thesis will only deal with capital estimation in terms of VaR.

In general, the distribution of total losses cannot be determined so a common approach to evaluate the VaR and ES is through the use of Monte Carlo simulation. This approach is employed in this thesis and some of the advantages and disadvantages are outlined below:

Advantages of Monte Carlo Estimation

- Easy to implement as long as the underlying random variables can be sampled from without difficulty.
- Converges towards the true results as the number of simulations increase.

Disadvantages of Monte Carlo Estimation

- Potentially time consuming when generating a large number of simulations to obtain an accurate result.
- Requires the underlying random variables to be appropriately specified, which is not trivial due the general lack of OpRisk data.
- Does not approximate the tails of distributions well, especially for heavy-tailed distributions. This is especially relevant in a risk management context because the VaR and ES required in are all in the extreme upper tail.

These challenges can be used to motivate the need for a strong understanding of the effects of different dependence structures on the capital reserve.

Chapter 2

Loss Distribution Approach

2.1 Basic Model

The most commonly used Loss Distribution Approach (LDA) model for a bank's annual losses, and the one which will be adopted throughout this thesis, is formulated as follows:

$$Z_t = \sum_{j=1}^{J} Z_t^{(j)} \tag{2.1}$$

$$Z_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} X_i^{(j)}(t), \tag{2.2}$$

where

- J is the number of risk cells,
- t is the discrete time in annual units,
- Z_t is the total annual loss in year t across all risk cells,
- $Z_t^{(j)}$ is the annual loss for risk cell j in year t,
- $N_t^{(j)}$ is the frequency of events for risk cell j in year t,
- $X_i^{(j)}$ is the severity of events for risk cell j in year t.

The frequency, $N_t^{(j)}$, is a random variable distributed according to a counting process, typically Poisson, which can either be modelled as dependent on constant parameters λ or time dependent parameters λ_t. Similarly, the severity, $X_t^{(j)}$ is also a random variable and it is distributed according to a severity distribution, typically Lognormal, Weibull or Generalized Pareto, characterized either by constant parameters ψ or time dependent parameters ψ_t. We will call parameters of the frequency and severity random variables their risk profile. Basic models in literature treat frequency and severity as independent with potentially different distributions for each risk cell j.

The total annual loss across all risk cells is just the sum of all the annual losses of each risk cell. This then does not represent a comprehensive description of the total losses as reductions (or increases) can be realized through diversification benefits (either positive or negative). This then motivates the thesis as a more robust model would require the dependence across risk cells to be quantified.

2.2 Frequency and Severity Distributions

Although there is a wide range of distributions available for modelling the frequencies and severities of losses, we will confine our studies to the ones described in this section.

Frequency - Poisson Distribution
The Poisson distribution is a discrete distribution that is commonly used to model the number of particular events that will occur. Its probability mass function is given by:

$$Pr(X = k) = \frac{\lambda^k \exp^{-\lambda}}{k!},\ \lambda > 0 \tag{2.3}$$

It is parameterized by the intensity parameter λ which, in our context, can be interpreted as the mean number of losses occurring within a certain time frame, usually a year.

Severity - Lognormal Distribution
A popular distribution for the magnitude of losses is the Lognormal distribution. It is obtained via an exponential transformation of the Normal distribution and consequently retains the simplicity of the Normal distribution whilst having desirable properties for modelling OpRisk losses such as possessing a heavy tail and strictly positive support. The density function is:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\, e^{-\dfrac{(\ln x - \mu)^2}{2\sigma^2}},\ x \in (0, \infty),\ \sigma > 0 \tag{2.4}$$

where μ and σ are its parameters which corresponding to the mean and standard deviation of the underlying Normal distribution respectively.

Severity - Normal Inverse Gaussian Distribution
A relatively unexplored distribution, especially in the context of OpRisk and risk modelling, which has recently gained attention in the area of risk modelling is the Normal Inverse Gaussian (NIG) distribution. It is found to fit well to many cases of loss data after truncating the negative portion of the distribution. The NIG distribution is obtained by mixing a Normal and Inverse Gaussian distribution. If

$$Y \sim IG(\delta, \sqrt{\alpha^2 - \beta^2})$$

and

$$X|Y \sim N(\mu + \beta Y, Y)$$

then the marginal obtained by integrating Y from $\int f(x, y)dy$ has a closed form density function given by:

$$f(x) = \frac{\alpha\delta}{\pi} \frac{\exp\big(p(x)\big)}{q(x)} K_1[\alpha q(x)],\ 0 < |\beta| \leq \alpha,\ \delta > 0 \tag{2.5}$$

where $K_1[x]$ is a modified Bessel function of the second kind with index 1 and

$$p(x) = \delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)$$
$$q(x) = \sqrt{(x - \mu)^2 + \delta^2}$$

and we say the severity X has a Normal Inverse Gaussian distribution.

$$X \sim NIG(\alpha, \beta, \mu, \delta)$$

The heaviness of the tail is controlled by the α parameter, where smaller values of α lead to heavier tails. β is the skewness parameter, μ is the location or shift parameter and δ is the scale parameter. The following plots show how the density of the NIG distribution behaves as its tail index and skewness parameters vary. In each case $\delta = 3$ and $\mu = 4$. In particular it can be seen that reducing α or increasing

β increases the heaviness of the upper tail.

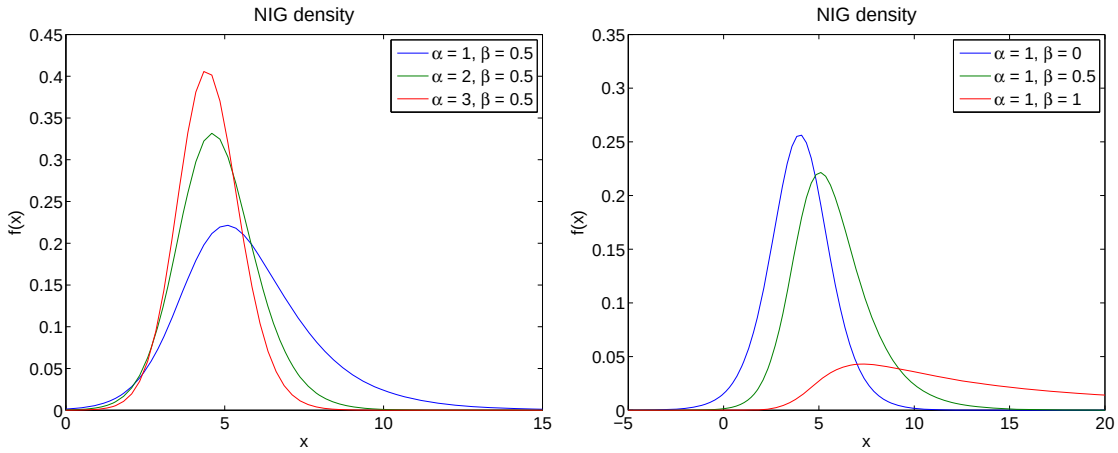


Figure 2.1: Density functions of the NIG distribution

2.3 Dependence Models

In the LDA compound distributional models detailed above there is no specification of the loss process or its associated variables. It is unrealistic to assume independence between frequencies and severities and doing so can have significant impacts on the reporting of the adequate capital required by banks, which will be demonstrated later sections. In order to address this issue, a wide scale of dependence models have been invented for more advanced modelling methodology. Dependence models are particularly important in an OpRisk setting as potential losses are of significant magnitude and are often correlated, for instance, IT failures will inevitably lead to transaction errors. Examples of OpRisk losses include Merrill Lynch's $250 million legal settlement for gender discrimination in 1999, which falls under Employment Practices and Workplace Safety, and Bank of New York's $140 million loss related to damage of facilities due to the September 11 attack, classed under Damage to Physical Assets. However, due to the scarcity in OpRisk data, dependence in this area is relatively unexplored and this is the core motivation behind this thesis.

In this section we will look at some of the dependence models for OpRisk presented in (**?**) and Chapter 7 of (**?**). These models are typically used in a finance setting and we will explore their properties, advantages and shortcomings. In particular, we will focus on copula models because of their flexibility and mathematical convenience.

Linear Correlation

Linear correlation measures the linear dependence between two random variables. It is given by:

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} \tag{2.6}$$

One of its key properties is:

$$\rho[\alpha X + \beta, \gamma Y + \delta] = \text{sign}(\alpha\gamma)\rho(X,Y)$$

for $\alpha, \beta \in R$. As a result, it is invariant under strictly increasing linear transformations.

Linear correlation is popular in many modelling settings because it is easy to calculate and is a natural scalar measure of dependence for elliptical distributions such as the Normal distribution. However, there are many problems with using linear correlation as an universal measure of dependence. Some problems include, but are not restricted to:

- Not invariant under non-linear transformations.

- The majority of distributions are not elliptical and so results and inference may prove to be unreliable as linear relationships do not describe their dependence structures appropriately.
- Zero linear correlation does not imply zero dependence except under the Normal distribution setting.
- Not all distributions have finite moments, meaning that linear correlation is not defined in all settings.

Rank Correlations

Rank-based correlations are a popular measure of dependence for general distributions as they are not a measure of linear relationships. The two most commonly used ones are described below.

Spearman's Rank Correlation, also known as **Spearman's rho**, is a rank based dependence measure. If X and Y are random variables with marginal distributions $F_X(x)$ and $F_Y(y)$ respectively, then:

$$\rho_S[X,Y] = \rho[F_X(X), F_Y(Y)] \tag{2.7}$$

where ρ is the linear correlation coefficient. Spearmans ρ can be seen as the linear correlation of the CDFs of random variables.

Kendall's tau rank correlation is another popular rank-based dependence measure.

$$\begin{aligned}\rho_\tau[X,Y] &= P\Big(\big((X-\tilde{X})(Y-\tilde{Y})>0\big) - P\big((X-\tilde{X})(Y-\tilde{Y})<0\big)\Big) \\ &= \text{Cov}[\text{sign}(X-\tilde{X})\text{sign}(Y-\tilde{Y})]\end{aligned} \tag{2.8}$$

where $(X,Y)^T$ and $(\tilde{X},\tilde{Y})^T$ are independent random vectors from the same distribution. Kendall's tau can also be approximated from linear correlation by:

$$\rho[X,Y] \approx \sin\big(\frac{1}{2}\pi\tau(X,Y)\big) \tag{2.9}$$

Common Shock

Common shock processes can be used to model events affecting multiple risk cells. Consider J risks with frequencies:

$$N^{(j)}(t) = N_C(t) + \tilde{N}^{(j)}(t), \qquad j = 1,...J \tag{2.10}$$

where $\tilde{N}^{(j)}(t)$ and $N_C(t)$ come from independent Poisson processes with intensity parameters $\tilde{\lambda}^{(j)}$ and λ_C respectively. It is clear that $N^{(j)}(t)$ has a Poisson marginal distribution with parameters $\lambda^{(j)} = \tilde{\lambda}^{(j)} + \lambda_C$ and are dependent on the common events $N_C(t)$. The linear correlation and covariances between the frequencies of risks are respectively given as:

$$\begin{aligned}\rho[N^{(i)}(t), N^{(j)}(t)] &= \frac{\lambda_C}{\sqrt{\lambda^{(i)}\lambda^{(j)}}} \\ Cov[N^{(i)}(t), N^{(j)}(t)] &= \lambda_C\end{aligned}$$

The covariance between any two pairs of risk is always the same, however their linear correlations will be different. Under this model, only positive dependence on the counts can be modelled. Furthermore, counts with negative dependence will be treated as having positive dependence as the correlation is only dependent on the intensity parameters. In this case, unreliable results will be produced. Consequently, the common shock model should only be used when the counts between the losses are definitively positively correlated.

Tail Dependence

Tail dependence is associated with the amount of dependence in the upper right or the lower left quadrant of a bivariate distribution. It is useful for studying the relationships between extreme values, for instance, the relationship between severe losses associated with two risk factors. This is important in a risk modelling setting as regulatory capital is almost always a risk measure of a loss distribution in its upper right tail, for example 99.9% VaR.

If the random variables X and Y have distribution functions F and G respectively, the upper tail dependence coefficient between them is:

$$\lambda_U = \lim_{u \to 1} P\big(Y > G^{-1}(u) | X > F^{-1}(u)\big) \tag{2.11}$$

provided the limit $\lambda_u \in [0, 1]$ exists. Similarly, their lower tail dependence coefficient is:

$$\lambda_L = \lim_{u \to 0} P\big(Y < G^{-1}(u) | X < F^{-1}(u)\big) \tag{2.12}$$

For copulas models discussed below, the tail dependence coefficients can be expressed as follows:

$$\lambda_U = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u} \tag{2.13}$$

$$\lambda_L = \lim_{u \to 0} \frac{C(u, u)}{u} \tag{2.14}$$

2.4 Copula Models

The concept of a copula model is a method of dependence modelling by the construction of multivariate distributions. This approach has become popular in recent years and will be the main focus of this thesis in the context of OpRisk.

In this section we will first look at the historical development of copulas, as outlined in (**?**). Then the advantages and disadvantages of utilizing copula models are discussed, as seen in (**?**), (**?**) and (**?**). Finally the properties of selected copula families will be described.

Definition

A copula, C, is a d-dimensional multivariate distribution on $[0, 1]^d$ with uniform marginal distributions. It can be used to uniquely represent a joint distribution $F_{\boldsymbol{X}}(\boldsymbol{x})$ of a continuous random variable $X \in R^d$ in terms of its marginals.

Historical Context

The theory of copulas can first be attributed to the work of Frecht *"Sue les tableaux de correlation don't les marges son donnees"* in 1951, who studied the relationship between the two random variables and their joint distribution. However it was not until 1959 that Sklar, in (**?**), introduced the concept and the name of a *copula* and proposed the most defining theorem in copula history, later named after him. For many years all work on copulas was done in a probabilistic metric space context. The first presentation of the concept to the statistical community was in 1976 by Scheiwer and Wolff in their paper *"Sur une mesure de dependance pour les variables aleatoires"*.

In 1990, the first conference for copulas, titled *"Advances in probability distributions with given marginals"*, was held by Dall'Aglio. This later led to a series of conferences which helped the development of copula theory significantly. At the end of the 1990's, the notion of copulas had grown quite popular and two books addressing copulas were published and have since become the standard reference texts even until today. In 1997, Joe released his book *"Multivariate models and dependence concepts"*, in which a large section was dedicated towards copulas and different families of copulas. In 1999, Nelsen published his book *"An introduction to copulas"* which was referenced heavily in industry groundbreaking papers such as (**?**).

Currently, copulas are used in many fields and interest continues to grow rapidly. The reason for the widespread development of modelling involving copulas is due to its introduction into the applied fields, such as hydrology and particularly finance. Perhaps the most defining paper in risk modelling is David Li's *"On Default Correlation: A Copula Function Approach"* (**?**), where he first applies the copula concept to model the association between the default times of financial instruments. As a result of this publication, copula modelling has quickly spread through the financial world where it was widely adopted. However, some have tried to argue that it was the extensive use of copulas that was responsible for the Global Financial Crisis in 2008, such as in (**?**). This demonstrates the importance of understanding underlying assumptions of copula modelling before it is applied in practice. For instance, Gaussian copulas used in the modelling of credit swap defaults assumes the dependence structure of the modelled variables are elliptical in shape and have no relationships in their extreme values. This assumption is blatantly false when modelling securities in the financial market because if the market collapses, defaults will likely occur simultaneously. As a result adopting this particular copula model was inappropriate for its intended purpose and ultimately resulted in world recession. However this does not mean that the methodology was at fault, merely that it was applied inappropriately. It is also important to note that the theory of copulas is not a complete solution to the dependence problem and the area is still in its infancy.

Copula Models and Applications

Copula models have gained much popularity in recent years, especially in the area of finance. The deficiencies in classic linear correlation, outlined in the previous section, motivated the need for a more robust modelling technique for dependence structures which were non-linear in nature. In general, multivariate distributions are a much more powerful tool for dependence analysis than just a single correlation coefficient as they represent the entire structure of the dependence, not just the overall strength of association. Consequently, copulas largely became the methodology of choice because of the relative mathematical simplicity and convenient properties. However, some might argue that copulas do not provide a complete and comprehensive solution to how all association should be modelled, for example in (**?**). There are many disadvantages to this approach and like every other modelling technique, caution should be exercised in its application. Some advantages and criticisms of the copula approach will be discussed in this section.

One of the key defining features and advantages of copula modelling is the ability to separate the marginal distributions of underlying variables from their dependence structures. This is achieved by transforming the marginal random variables into (typically) Uniform $[0, 1]$ variables through their marginal distribution functions. These transformed $[0, 1]$ random variables then serve as inputs into an appropriate copula function. In this way, the distribution of each variable can be modelled independently of each other first and a copula function, acting as the dependence relationship, can then be applied to all of the marginal variables to obtain their full multivariate distribution. It should be noted that although this transformation is convenient and easy to understand, we do not know whether it is optimal.

This approach gives simplicity and flexibility to the modelling process as changes to the marginal distributions will not necessarily affect the choice of our copula model, only its parameters, and vice versa. This property is relevant in OpRisk modelling because we usually have comprehensive knowledge of individual risk factors, but struggle to understand the interdependence relationships between multiple risks. Furthermore, unlike much of the multivariate theory existing today, the marginal distributions do not have to be identically distributed. Efforts of modellers can also be divided when constructing the model as the marginal and copula distributions can be fit separately. Simulation from these models is also mostly trivial if the marginal distributions can easily be simulated from. However, a common criticism this approach has received is its tendency to increase model risk as estimation of each of the components can contribute bias to the overall model. One way to overcome this inaccuracy is to estimate the marginal and copula distributions jointly, but this approach neglects the separation benefit of copulas. In addition, the dimensionality of the problem will increase severely.

Another major advantage of copula modelling is that it is a distribution based model of dependence

and not just a scalar quantifier. This method constructs a full multivariate distribution with respect to the variables of interest to represent the dependence structure between these underlying variables and hence brings many benefits from an analytical point of view. In particular, the multivariate distribution allows us to calculate risk measures such as VaR and ES for groups of risks easily as they are both quantile based measures. Furthermore, both linear and non-linear correlations are able to be modelled. Tail dependence between pairs of variables can also be established, which is particularly useful from a risk modelling perspective where we usually want to model losses with heavy tails.

Flexibility of copula modelling is another desirable property. A broad range of dependence structures can be formed through the many types of known copula families. These include both parametric copulas, like the (in)famous Gaussian copula that was wildly popular in modelling credit swap defaults, and non-parametric copula, such as the empirical copula. Combined with the myriad selection of univariate distributions, again parametric and non-parametric, available for the marginals, it is easy to see the flexibility copula modelling brings and why its applications have developed. However, a disadvantage to flexibility is the increased risk of model misspecification. A notable shortcoming in the copula area is the lack of procedures when it comes to model selection. Although standard estimation techniques, such as maximum likelihood or Bayesian inference, can be employed to determine the relevant parameters of a copula model after it has been chosen, there is no standard criteria for choosing which copula model should be used in the first place. Instead many such criteria are currently being developed. Furthermore, there is a lack of literature on goodness-of-fit tests, with some notable exceptions outlined later in this paper, so there is substantial difficulty in judging how good the fitted model is.

Traditionally, copula literature has not focused much on the context of dynamic models, i.e. models where the underlying dependence structure changes over time. The parameters for usual copula models are assumed to be static and so the implied associate is only valid for the particular instance in time the copula model was derived from. As such, common frameworks such as stochastic processes and time series were avoided. Moreover, results derived from standard copula approaches should not be treated as an indication of the future for the reasons given above. For example, the VaR value computed from a static copula model is an instantaneous measure. Some may argue that this is a weakness in the copula methodology, such as in (**?**), however, recent developments to copula modelling have been made that explores the concept of dynamic copulas. In papers such as (**?**), (**?**) and (**?**), underlying parameters of marginal distributions and/or dependence parameters are modelled under a stochastic setting. This sets up a framework where the dependence relationship is allowed to change over time. These works have contributed to the extension of copula models beyond the usual static interpretations and applications.

Despite the flaws in the copula concept, the flexibility and convenience it offers has elevated it to being one of the most applied dependence modelling techniques today, especially in the area of finance. However modellers must be wary of its shortcomings or risk facing drastic consequences.

Sklar's Theorem

The following theorem proposed by Sklar in (**?**), is the underlying principle beneath the theory and application of copula models.

Let $F : R^n \mapsto (0,1)$ *be a joint distribution with margins* $F_1, ..., F_d$. *Then there exists a copula* $C : (0,1)^n \mapsto (0,1)$ *such that for all* $\boldsymbol{x} \in R^n, \boldsymbol{u} \in (0,1)^n$

$$F(\boldsymbol{x}) = C\big(F_1(x_1), ..., F_d(x_d)\big) = C(\boldsymbol{u}) \tag{2.15}$$

Conversely, if $C : (0,1)^n \mapsto (0,1)$ *is a copula and* $F_1, ..., F_d$ *are distribution functions, then there exists a joint distribution function* F *with margins* $F_1, ..., F_d$ *such that for all* $\boldsymbol{x} \in R^n, \boldsymbol{u} \in (0,1)^n$

$$F(\boldsymbol{x}) = F\big(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)\big) = C(\boldsymbol{u})$$

Furthermore, if $F_1, ..., F_d$ *are all continuous, then the copula* C *is unique.*

Proof

Let $\boldsymbol{X} = (X_1, ..., X_d)$ be a random vector with distribution function F and let $F_1, ...F_d$ be the marginals of $X_1, ..., X_d$ respectively. Now we have $U_i = F_i(x_i)$, where U_i is Uniform(0,1), and $X_i = F_i^{-1}(U_i)$ for $i = 1, ..., d$. By defining C to be the distribution function of $(U_1, ..., U_d)$ we get

$$\begin{aligned} F(x) &= P\big(\boldsymbol{X} \leq (x_1, ..., x_d)\big) \\ &= P\big(F_1^{-1}(U_1) \leq x_1, ..., F_d^{-1}(U_d) \leq x_d\big) \\ &= P\big(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)\big) \\ &= C\big(F_1(x_1), ..., F_d(x_d)\big) \end{aligned}$$

Hence C is a copula of F. The joint density can then be written as:

$$f(x_1, ...x_d) = c\big(F_1(x_1), ..., F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i) \tag{2.16}$$

where c is the copula density corresponding to distribution C when it exists.

2.4.1 Standard Copula Models

In this section we will examine the properties of several parametric copula models that have started to find widespread use in OpRisk as discussed in (**?**). In particular, their distribution and/or density functions will be given along with associated dependence measures. Their pros and cons will also be discussed in an OpRisk context.

Elliptical Copulas

A popular family of copulas used in finance is the Elliptical copulas. They are characterized by their symmetric distributions and tails. The most commonly used members of this family are the Gaussian and t copula described below.

Gaussian Copula

The Gaussian copula is a distribution derived from the multivariate Normal. Its relative simplicity in estimation and simulation has made it the most used copula model in finance. Its density function is given by:

$$c(u_1, ...u_d) = \frac{f_N^{\Sigma}\big(F_N^{-1}(u_1), ..., F_N^{-1}(u_d)\big)}{\prod_{i=1}^{d} f_N\big(F_N^{-1}(u_i)\big)} \tag{2.17}$$

where

- f_N and F_N are the standard univariate Normal density and distribution respectively.
- f_N^{Σ} is a multivariate Normal density with zero means, unit variances and a linear correlation matrix Σ.

The Gaussian copula is characterized by its correlation matrix $\boldsymbol{\Sigma}$, which determines the linear correlations between the variates. It shares many similarities with a multivariate Normal distribution. It has a symmetric distribution with light concentrations of the density in both of its tails. The Gaussian copula does not allow tail dependences unless there is perfect correlation between variates. This is quite a serious deficiency in the context of OpRisk modelling as losses are typically heavy tailed and associations between them do exist in extreme events.

t Copula

The t copula is a generalization of the Gaussian copula based on the multivariate Student's t distribution. Although mathematically more complex than the Gaussian copula, its ability to model tail dependences makes it more attractive for practical use. To study its properties, we define the following notation:

- $\boldsymbol{Z} \sim \text{Normal}(0, \boldsymbol{I})$,
- $S \sim \chi^2_\nu$,
- $\sqrt{\nu/S} \sim G_\nu$,
- $V \sim$ Uniform(0,1) independent of $\boldsymbol{Z}$,
- $W = G_\nu^{-1}(V)$.

The random vector $\boldsymbol{X} = W\boldsymbol{Z}$ comes from the multivariate Student's t distribution and the random vector $\boldsymbol{U} = \big(t_\nu(X_1), ..., t_\nu(X_d)\big)'$ is distributed according to the standard t-copula.
The upper and lower tail dependences of any pair of variates in the t-copula can be expressed as:

$$\lambda_U = \lambda_L = 2t_{\nu+1}\left(\sqrt{\nu+1}\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\right) \tag{2.18}$$

where ρ is the linear correlation between the two variates and ν is the degree-of-freedom parameter.

We can generalize the t-copula to use multiple degrees of freedom parameters, like in (**?**), which can give more flexibility when fitting a dependence structure. The copula of the random vector $\boldsymbol{X} = (W_1 Z_1, ..., W_d Z_d)$, where $W_k = G^{-1}_{\nu_k}(V)$, is a t-copula with multiple degrees of freedom, $\boldsymbol{U} = \big(t_{\nu_1}(X_1), ..., t_{\nu_d}(X_d)\big)'$

The distribution and density functions of the generalized t-copula are given below:

$$C^\Sigma_\nu(\boldsymbol{u}) = \int_0^1 \phi_\Sigma\big(z_1, (u_1, s), ..., z_d(u_d, s)\big)ds \tag{2.19}$$

$$\begin{aligned} c^\Sigma_\nu(\boldsymbol{u}) &= \frac{\partial^d C^\Sigma_\nu(\boldsymbol{u})}{\partial u_1, ..., \partial u_d} \\ &= \frac{1}{\prod_{k=1}^d f_{\nu_k}(u_k)} \int_0^1 \varphi_\Sigma\big(z_1(u_1, s), ...z_d(u_d, s)\big) \prod_{k=1}^d (w_k(s))^{-1} ds \end{aligned} \tag{2.20}$$

where

- $z_k(u_k, s) = t^{-1}_{\nu_k}(u_k)/w_k(s),\ k = 1, ...d$,
- $w_k(s) = G^{-1}_{\nu_k}(s)$,
- φ_Σ is the multivariate normal density,
- $x_k = t^{-1}_{\nu_k}(u_k),\ k = 1, ...d$.

The t copula is characterized by a degree of freedom parameter ν, or a vector of ν in the generalized case, which determine tail dependence between variates, and its linear correlation matrix $\boldsymbol{\Sigma}$, which defines the linear dependence between variates. It shares common features with a multivariate t distribution such as symmetric shape and heavy tails. The Gaussian copula is a special case of the t copula when $\nu = \infty$. As such, the t copula is a more robust model to use in practice despite extra estimation and simulation efforts required.

Archimedean Copulas

Another popular class of copulas is the Archimedean family of copulas. Several members of this family

have strong tail dependence properties which makes their use preferable in certain contexts such as survival analysis (lower tail) and risk management (upper tail). Archimedean copulas have a distribution function of the form:

$$C(u_1, ..., u_d) = \phi^{-1}\big(\phi(u_1) + ... + \phi(u_d)\big) \tag{2.21}$$

where ϕ is a decreasing function called the generator for the copula.

The tail dependence coefficients of Archimedean copulas can be expressed in terms of their generator functions. If $\phi^{-1}(0)$ is finite, then the corresponding copula has no tail dependence. A copula has upper tail dependence if $\phi^{-1}(0) = -\infty$ and the coefficient of the upper tail dependence is:

$$\lambda_U = 2 - 2 \lim_{s \to 0} \frac{\phi^{-1\prime}(2s)}{\phi^{-1\prime}(s)} \tag{2.22}$$

Its lower tail dependence coefficient is:

$$\lambda_L = 2 \lim_{s \to \infty} \frac{\phi^{-1\prime}(2s)}{\phi^{-1\prime}(s)} \tag{2.23}$$

There are three prominent members of the Archimedean family used in practice; Clayton, Gumbel and Frank. Their properties are discussed below as they will be utilized in this thesis as building blocks for mixture copulas.

Clayton copula

The Clayton copula is popular in survival studies due to its lower tail dependence properties. However in an OpRisk context, it is generally unsuitable as we are interested in distributions with heavy upper tails. The distribution and density functions of the Clayton copula are given respectively as:

$$C(u_1, ..., u_d) = \left(1 - d + \sum_{i=1}^{n} u_i^{-\theta}\right)^{-1/\theta} \tag{2.24}$$

$$c(u_1, ..., u_d) = \left(1 - d + \sum_{i=1}^{d} (u_i)^{-\rho}\right)^{-d-\frac{1}{\rho}} \prod_{i=1}^{d} \Big((u_i)^{-\rho-1}\big((i-1)\rho + 1\big)\Big) \tag{2.25}$$

where $\rho > 0$ is the dependence parameter. The Kendall's Tau rank correlation can be expressed in terms of the dependence parameter:

$$\tau = \frac{\rho}{\rho + 2} \tag{2.26}$$

The Clayton copula does not have upper tail dependence. Its lower tail dependence can be expressed as:

$$\lambda_L = 2^{-1/\rho} \tag{2.27}$$

The generator for the Clayton copula is:

$$\phi_C(t) = (t^{-\rho} - 1) \tag{2.28}$$

Gumbel copula

The Gumbel copula is appropriate for use in OpRisk modelling due to its upper tail dependence. Since OpRisk losses are typically heavy tailed, the Gumbel copula acts as a suitable dependence structure for modelling their interactions under extreme events. However one of the limitations of this model is the

inability to account for negative dependence. Its distribution and density functions respectively are:

$$C(u_1, ..., u_d) = \exp\left(-\left[\sum_{i=1}^{d}\big(-\log(u_i)\big)^{\rho}\right]^{\frac{1}{\rho}}\right) \tag{2.29}$$

$$c(u_1, ..., u_d) = \frac{\partial^d}{\partial u_1, ...\partial u_d} C(u_1, ..., u_d) \tag{2.30}$$

where $\rho \geq 1$ is the dependence parameter. The corresponding Kendall's Tau measure is:

$$\tau = 1 - \frac{1}{\rho} \tag{2.31}$$

In the bivariate case, the closed form expression for the density of the Gumbel copula is

$$\begin{aligned} c(u_1, u_2) &= \frac{\partial^2}{\partial u_1, \partial u_2} C(u_1, u_2) \\ &= C(u_1, u_2) u_1^{-1} u_2^{-1} \left[\sum_{i=1}^{2}\big(-\ln(u_i)\big)^{\rho}\right]^{2(\frac{1}{\rho}-1)} [\ln(u_1)\ln(u_2)]^{\rho-1} \\ &\times \left[1 + (\rho - 1)\left[\sum_{i=1}^{2}\big(-\ln(u_i)\big)^{\rho}\right]^{-\frac{1}{\rho}}\right]. \end{aligned} \tag{2.32}$$

The generator for the Gumbel copula is:

$$\phi_G(t) = (-\ln t)^{\rho} \tag{2.33}$$

The upper tail dependence of the Gumbel copula can be expressed as:

$$\lambda_U = 2 - 2^{1/\rho} \tag{2.34}$$

Hence for $\rho > 1$, the Gumbel copula has upper tail dependence. The Gumbel copula does not have lower tail dependence.

Frank copula

The Frank copula possesses similar properties to the Gaussian copula including symmetry and the lack of tail dependence. It may be an convenient alternative to the Gaussian copula in higher dimensions as only 1 parameter is required to be estimated. The distribution and density of the Frank copula is:

$$C(u_1, ..., u_d) = \frac{1}{\rho}\ln\left(1 + \frac{\prod_{i=1}^{d}(e^{\rho u_i} - 1)}{(e^{\rho} - 1)^{d-1}}\right) \tag{2.35}$$

$$c(u_1, ..., u_d) = \frac{\partial^d}{\partial u_1, ...\partial u_d} C(u_1, ..., u_d) \tag{2.36}$$

where $\rho \in R/\{0\}$ is the dependence parameter. The Kendall's Tau measure can be written as:

$$\tau = 1 - \frac{4}{\rho}\big(D_1(-\rho) - 1\big) \tag{2.37}$$

where D is the Debye function defined as:

$$D_n(x) = \frac{n}{x^n}\int_0^x \frac{t^n}{e^t - 1}\, dt. \tag{2.38}$$

The generator for the Frank copula is:

$$\phi_F(t) = -\ln t \frac{e^{-\rho t} - 1}{e^{-\rho} - 1} \tag{2.39}$$

The Frank copula does not have upper or lower tail dependence.

2.4.2 Advanced Copula Models

Standard parametric copulas can be restrictive because they are constructed using a pre-determined set of distribution functions and parameters. This limits the shapes of the dependence structures that can be modelled as parametric copulas do not cover the full spectrum of dependence relationships. For instance, if the observed dependence structure contains heavy upper tail dependence, no lower tail dependence and an elliptical body, then using either a pure elliptical or Archimedean copula does not capture all of the main features. This means that the estimated parameters of the chosen parametric model is sub-optimal and can lead to misleading risk measure estimates and consequently incorrect diversification benefits for a bank when modelling OpRisk. This is a concern because OpRisk involves risk factors with a wide variety of characteristics, for instance fraud and IT failure are vastly different, which can lead to a diverse range of dependence structures. Furthermore, risks can potentially incur very heavy losses and any erroneous quantifications means either the bank has to hold unnecessary excess capital or the bank holds inadequate capital. These factors motivate the need for more robust copula models to be developed in order to capture and quantify a larger array of dependence models more accurately. However, adopting complex models over basic ones usually comes at a cost, such as heightened difficulties in estimation and simulation, and these should be weighed with the benefits to determined whether the model should be employed under a particular framework.

Mixture Copulas

To address the issue of flexibility of dependence structures, one may combine different copula models. A mixture of copulas can be used to allow more flexibility when modelling dependence compared to using single copulas. By using a weighted mixture of copula densities, a more diverse set of tail dependence and correlation structures can be modelled. A linear combination of copulas result in the same linear combination of their tail dependence and correlation coefficients.

One trade-off for the additional flexibility is the increased complications in estimation, simulation and inference. By adding more copulas into the model, the required number of parameters to be estimated increases. For example, models using a Gumbel copula only requires one parameter to be obtained. However if we add a bivariate Gaussian copula, then we also need to estimate the correlation parameter of the Gaussian as well as the weights of the copula mixture. Simulations will also be more complex as more copulas will need to be simulated from. However, the time of the projections will not necessarily differ if the same number of samples is required.

Let $C_{mix}(\boldsymbol{u}) = \sum_{i=1}^{n} w_i C_i(\boldsymbol{u})$ be a mixture copula formed from the linear combination of n other copulas $C_i(u)$, where $\sum_{i=1}^{n} w_i = 1$. Then the associated Kendall's tau correlation coefficient is just the same linear combination of the Kendall's taus of the component copulas (**?**) and it can be shown that the tail dependence also follow this behaviour (see Appendix).

- $\tau_{mix} = \sum_{i=1}^{n} w_i \tau_i$, where τ is the Kendall's Tau rank correlation coefficient,
- $\lambda_{L_{mix}} = \sum_{i=1}^{n} w_i \lambda_{L_i}$, where λ_L is the lower tail dependence coefficient,
- $\lambda_{U_{mix}} = \sum_{i=1}^{n} w_i \lambda_{U_i}$, where λ_U is the upper tail dependence coefficient.

As an example, consider a mixture copula that consists of three bivariate components, Gaussian, Gumbel and Clayton. This particular mixture exhibits the upper tail dependence of the Gumbel, the lower tail dependence of the Clayton and the linear correlation of the Gaussian copula. The degree of these properties depend on the weights used. Hence this new copula covers a much broader range of dependence scenarios than using any one of the components individually. It will therefore be interesting in this thesis

to study the implications on risk measures and estimation of OpRisk LDA models in this context.

OpRisk Model Construction Example
Consider a setting where we are only interested in two risks; IT failure and transaction errors. Suppose we know that these two risk factors roughly have a linear dependence relationship. However, we also know that in a case where there is an extreme number of IT failures, the number of transaction errors will also likely be large in magnitude as obtained from expert opinions or experience. This implies that there is reason to consider upper tail dependence between the frequency of the two risks processes. In this scenario, we may wish to select a mixture of a Gaussian and Gumbel copula to model the dependence structure of the counts of the events. The linear correlation for the body of the distribution is modelled by the Gaussian copula while the upper tail dependence is accounted for by the Gumbel. We avoid using the t copula in this case as we suspect there is no lower tail dependence. The following LDA model can then be formulated for this simplified example:

$$Z_t = Z_t^{(I)} + Z_t^{(T)}$$
$$Z_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} X_i^{(j)}(t)$$

where

- t is the discrete time in annual units,
- Z_t is the total annual loss in year t across both risk cells,
- $Z_t^{(j)}$ is the annual loss for risk cell j in year t,
- $N_t^{(j)}$ is the frequency of events for risk cell j in year t,
- $X_i^{(j)}$ is the severity of events for risk cell j in year t,
- $j \in \{I, T\}$ where I and T correspond to IT failure and transaction error respectively.

The bivariate distribution between the frequencies $\boldsymbol{N} = (N_I, N_T)$ can then be expressed as:

$$\begin{aligned} F_{\boldsymbol{N}}(n_I, n_T) &= C_{mix}\big(F_{N_I}(n_I), F_{N_T}(n_T)\big) \\ &= W_{Ga} \cdot C_{Ga}\big(F_{N_I}(n_I), F_{N_T}(n_T)\big) + (1 - W_{Ga}) \cdot C_{Gu}\big(F_{N_I}(n_I), F_{N_T}(n_T)\big) \end{aligned} \tag{2.40}$$

where

- C_{mix} is the distribution function of the mixed copula,
- C_{Ga} is the distribution function of the Gaussian copula component,
- C_{Gu} is the distribution function of the Gumbel copula component,
- W_{Ga} is the weight of the Gaussian copula in the mixture.

Adopting such a model allows a greater range of dependence structures to be modelled as it can account for all the shapes between the Gaussian and Gumbel copula. Furthermore, if one of the copulas in the mixture was not appropriate for the data, then it will be reflected with a corresponding weight close to zero. However fitting this model requires estimation of 3 parameters whilst if a Gaussian or Gumbel was used individually, only 1 parameter needs to be estimated. Due to the lack of data in OpRisk, the inaccuracy of parameter estimates may lead to an higher error in capital estimation than choosing a less flexible dependence structure. We will study an example of this numerically in Chapter 3.

Vine Copulas

Other structures that capture more intricate dependence relationships between random variates include vine copulas. Multivariate distributions are in general extremely difficult to specify and this has motivated the need for simpler approaches to model high dimensional systems. The notion of vine copulas was introduced to ease the specification of multivariate distributions by decomposing them into the product of a series of bivariate distributions. First we introduce the different vine based dependence structures based on a D vine and a canonical vine decomposition, as seen in (**?**), then we discuss their properties and pros and cons followed by an illustration of this concept in a novel application in OpRisk modelling.

We first show how all multivariate distributions can be decomposed into a product of bivariate distributions, as per (**?**). Note that:

$$f(x_1|x_2) = c_{12}\big(F_1(x_1), F_2(x_2)\big) \cdot f_1(x_1)$$
$$f(x_1|x_2, x_3) = c_{13|2}\big(F(x_1|x_2), F(x_3|x_2)\big) \cdot f(x_1|x_2)$$

It follows that for every j:

$$f(x|v_1, v_2, ..., v_d) = c_{xv_j|v_{-j}}\big(F(x|v_{-j}), F(v_j, |v_{-j})\big) \cdot f(x|v_{-j}) \tag{2.41}$$

Combining this with:

$$f(x_1, ..., x_n) = f(x_n) \cdot f(x_{n-1}|x_n)... \cdot f(x_2|x_3, ..., x_n) \cdot f(x_1|x_2, ..., x_n) \tag{2.42}$$

we can write any joint density in terms of marginals and bivariate copulas. For high dimensional distributions, the concept of vines has been introduced to order the pair-copula decompositions.

A vine copula denoted V can be thought of as a collection of trees that provide a graphical representation of the relationship between marginal, conditional and joint distributions. Formally, it can be represented according to (**?**) as follows:

- $V = (T_1, ..., T_m)$, where T_i is a tree,
- T_1 is a tree with nodes $N_1 = 1, ..., n$ and a set of edges E_1,
- $T_i, i = 2, ..., m$ are trees with nodes N_i which forms a subset of $N_1 \cup E_1 \cup E_2 \cup ... \cup E_i$ and edge set E_i.

Each node in the vine represents a random variable and each edge represents an assumed dependence relationship between the nodes it is linking. A vine is a regular vine on n elements if:

- $m = n$,
- T_i is a connected tree with edge set E_i and node set $N_i = E_{i-1}$, with $|N_i| = n-(i-1)$ for $i = 1, ..., n$, where $|N_i|$ is the cardinality of the set N_i,
- The proximity condition holds, i.e. for $i = 2, ..., n-1$, if $a = (a_1, a_2)$ and $b = (b_1, b_2)$ are two nodes in the set N_i connected by an edge $(a_1, a_2, b_1, b_2 \in N_{i-1})$ then $|a \cap b| = 1$.

Intuitively, every node in a tree is an edge in the previous tree. The nodes in the first tree represent the marginal distributions of the variates while the edges represent the conditional joint distribution between them. The distribution of the whole vine copula is the product of all of the marginal and conditional joint distributions across all trees.

The degree of a node is the number of edges it possesses. If each node in the base tree of the vine has a degree of at most two, then the vine is a d-vine. If each tree has a unique node of degree $n - i$, then the vine is a canonical vine.

Consider the joint distribution of a quad-variate distribution, $f_{1234}(x_1, x_2, x_3, x_4)$. A d-vine representation of the joint distribution is:

$$\begin{aligned} f_{1234}(x_1, x_2, x_3, x_4) &= f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \\ &\times C_{12}\big(F_1(x_1), F_2(x_2)\big) C_{23}\big(F_2(x_2), F_3(x_3)\big) \\ &\times C_{34}\big(F_3(x_3), F_4(x_4)\big) C_{13|2}\big(F_{1|2}(x_{1|2}), F_{3|2}(x_{3|2})\big) \\ &\times C_{24|3}\big(F_{2|3}(x_{2|3}), F_{4|3}(x_{4|3})\big) C_{14|23}\big(F_{1|23}(x_{1|23}), F_{4|23}(x_{4|23})\big) \end{aligned} \tag{2.43}$$

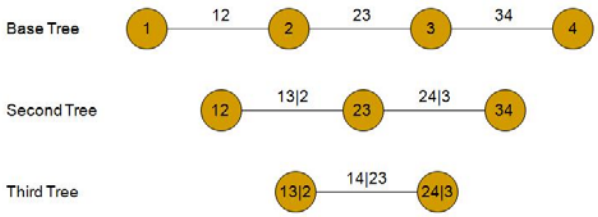


Figure 2.2: D-vine decomposition

A canonical vine representation of the same joint distribution is:

$$\begin{aligned} f_{1234}(x_1, x_2, x_3, x_4) &= f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \\ &\times C_{12}\big(F_1(x_1), F_2(x_2)\big) C_{13}\big(F_1(x_1), F_3(x_3)\big) \\ &\times C_{14}\big(F_1(x_1), F_4(x_4)\big) C_{23|1}\big(F_{2|1}(x_{2|1}), F_{3|1}(x_{3|1})\big) \\ &\times C_{24|1}\big(F_{2|1}(x_{2|1}), F_{4|1}(x_{4|1})\big) C_{34|12}\big(F_{3|12}(x_{3|12}), F_{4|12}(x_{4|12})\big) \end{aligned} \tag{2.44}$$

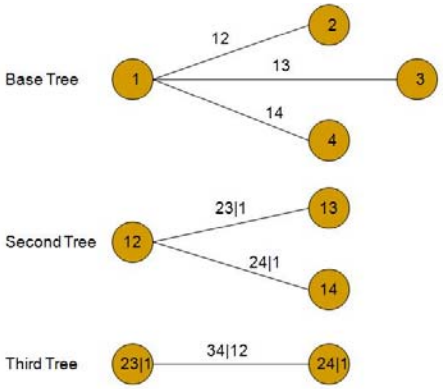


Figure 2.3: Canonical vine decomposition

In instances where one variable is significant and we wish to specify its relationship with others, the canonical vine is more desirable. However in general, the d-vine is more flexible as we have more freedom to select the pairs we wish to model.

In order to construct a vine copula, one must determine the selection of a specific factorization or arrangement of the trees, the choice of the pair-copula models and the parameters that correspond to these models. Due to the large number of possible combinations of decompositions and copula types, the most feasible way to establish the trees is to consider which bivariate relationships are important to model correctly and set up a decomposition accordingly. Then given the data and the decomposition, each pair-copula needs to be assigned a parametric shape. The copula families can be chosen by looking at the pairwise plots of the corresponding variables for each tree. Common estimation techniques such as maximum likelihood and Bayesian inference can be carried out to estimate the parameters of all the

pair-copulas given their types. goodness-of-fit tests, such as the ones outlined in Chapter 6, can then be used to determine the appropriateness of the chosen model.

The main advantage of using vine copula is that it allows complex multivariate data to be modelled using pair-copulas as simple building blocks without assuming unnecessary conditional independence. High dimensional copulas are generally difficult to build, however bivariate copulas have been studied extensively which makes vine copula an attractive approach. From an OpRisk perspective, it allows us to simplify the specification of dependences between up to 56 risk factors into pair-wise models. *The lack of data makes modelling high dimensional associations difficult and this method alleviates this issue. The distribution of the entire vine can then be interpreted as the full joint distribution of all risk factors.*

Despite the advantages, there are some issues associated with building vine copulas. Choosing which copula family to use for each pair-copulas is not a trivial task. Furthermore as the dimensions grow, the number of pairs increase rapidly and there are a huge number of potential decompositions to choose from. However, it is worth noting that not all of the conditional joint distribution have to be specified if certain risk factors are believed to be independent.

OpRisk Model Construction Example

Let us now consider an example of a vine copula decomposition applied to OpRisk. Suppose, under the basic LDA framework, we have three risk factors; IT failure, transaction errors and fraud. So our loss distribution is given as:

$$Z_t = Z_t^{(I)} + Z_t^{(T)} + Z_t^{(F)}$$
$$Z_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} X_i^{(j)}(t)$$

where

- t is the discrete time in annual units,
- Z_t is the total annual loss in year t across all three risk cells,
- $Z_t^{(j)}$ is the annual loss for risk cell j in year t,
- $N_t^{(j)}$ is the frequency of events for risk cell j in year t,
- $X_i^{(j)}$ is the severity of events for risk cell j in year t,
- $j \in \{I, T, F\}$ where I, T and F correspond to IT failure, transaction error and fraud respectively.

We wish to model dependence between these factors by imposing a dependence structure between their frequencies $N^{(j)}$. Under a trivariate setting, the d-vine and the canonical vine are equivalent but we will arrange our decompositions in the shape of a d-vine. One possible pairwise decomposition is illustrated below:

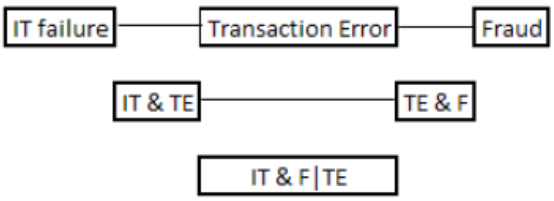


Figure 2.4: D-Vine arrangement 1

In addition to the marginal distributions of the three risk factors, to fully determine the joint distribution according to the vine, we also need to specify the bivariate copulas between IT failure and transaction errors, transaction errors and fraud, and IT failure and fraud given transaction errors. The multivariate density of $\boldsymbol{N} = (N_I, N_T, N_F)$ according to the vine above is:

$$\begin{aligned} f_{\boldsymbol{N}}(n_I, n_T, n_F) = & f_I(n_I) f_T(n_T) f_F(n_F) \\ & \times C_{IT}\big(F_I(x_I), F_T(x_T)\big) C_{TF}\big(F_T(x_T), F_F(x_F)\big) \\ & \times C_{IF|T}\big(F_{I|T}(x_{I|T}), F_{F|T}(x_{F|T})\big) \end{aligned} \tag{2.45}$$

Suppose now we are more confident with our copula model between fraud and IT failure than between IT failure and transaction errors. We could use a different decomposition of the vine as shown below:

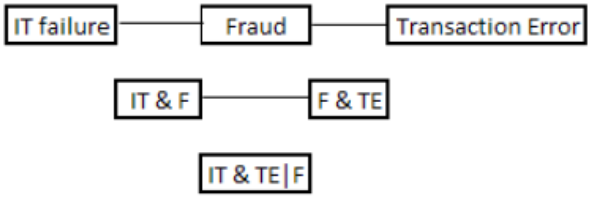


Figure 2.5: D-vine arrangement 2

Here the multivariate distribution of $\boldsymbol{N}$ is given by:

$$\begin{aligned} f_{\boldsymbol{N}}(n_I, n_F, n_T) = & f_I(n_I) f_F(n_F) f_T(n_T) \\ & \times C_{IF}\big(F_I(x_I), F_F(x_F)\big) C_{FT}\big(F_F(x_F), F_T(x_T)\big) \\ & \times C_{IT|F}\big(F_{I|F}(x_{I|F}), F_{T|F}(x_{T|F})\big) \end{aligned} \tag{2.46}$$

In this case the pairwise dependencies we need to specify are different from the first vine due the different ordering of risk factors on the base tree. Hence we can see that vine copula decomposition is very flexible due to the wide spectrum of dependence structures it can model. Furthermore, it achieves this by using simple bivariate copulas. This is useful in an OpRisk setting as we usually lack the data to calibrate dependence structures between multiple risk factors accurately, but can relatively easily determine the dependence structure between particular pairs of risks. However the number of pairwise copulas required increases drastically as the number of risks increase and in practice, such as in Basel II where there are 56 risk factors, it is not always feasible to specify the entire vine.

Chapter 3

Copula Model Simulation Studies

In this chapter we will explore the properties of copula models and their effects on capital estimation through Monte Carlo simulations.

3.1 Properties of Copula Models

We will study the behaviour of selected copula models under a bivariate setting (trivariate for vine copulas) by looking at the pairwise scatterplots of the uniform samples generated from the copula. We will also look at the corresponding joint multivariate distribution assuming the marginals are both Normal(0,1) random variables. In the graphs below, U will denote the copula samples and X will denote Normal samples.

3.1.1 Standard Copulas

Independent

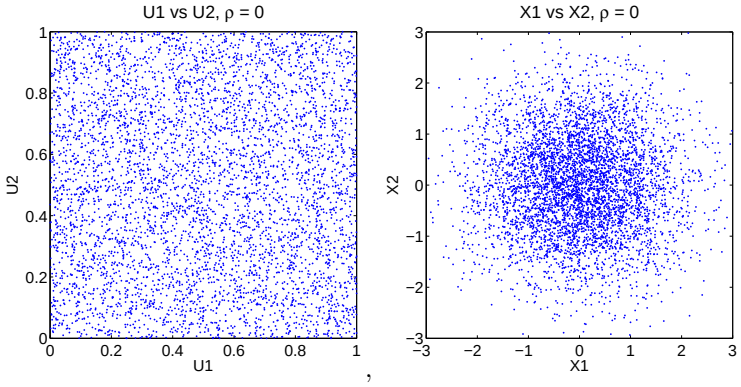


Figure 3.1: Independent samples

The independence case will be a benchmark of our analysis. The uniform samples are spread out evenly on $[0,1]^2$ while the corresponding Normal samples form a circular shape centred at $(0,0)$.

Gaussian Copula

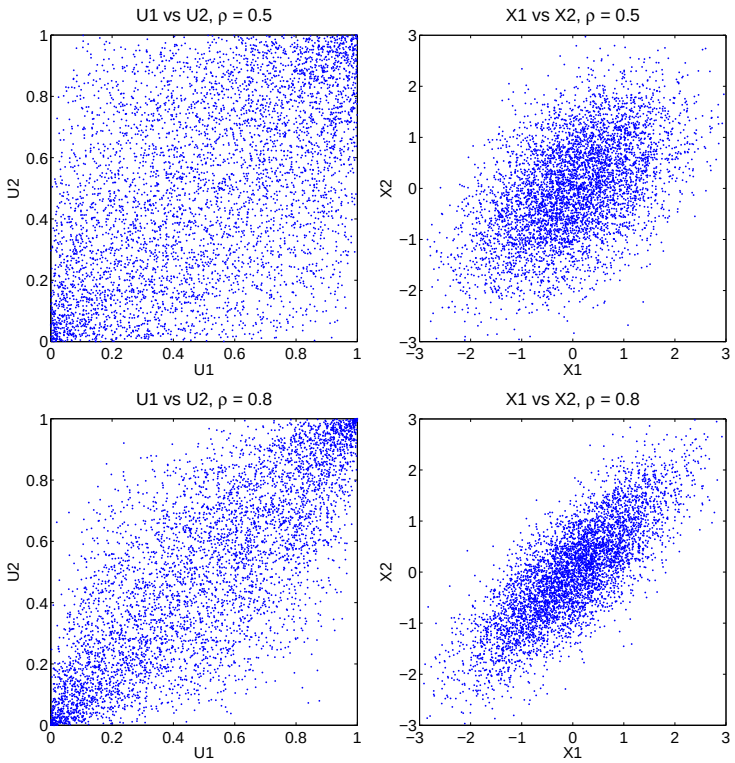


Figure 3.2: Gaussian copula samples

As the dependence parameter increases, a clear increase in linear dependence between the uniform samples can be seen. The circular shape of the Normal sample stretches to become an ellipse as the correlation increases. However, we observe almost no Normal samples in the upper right and lower left tails which generally makes modelling OpRisk inappropriate since we want to also account for dependence between extreme events. However, the Gaussian copula is mathematically simple to understand and apply so it is commonly used in practice.

Standard t Copula

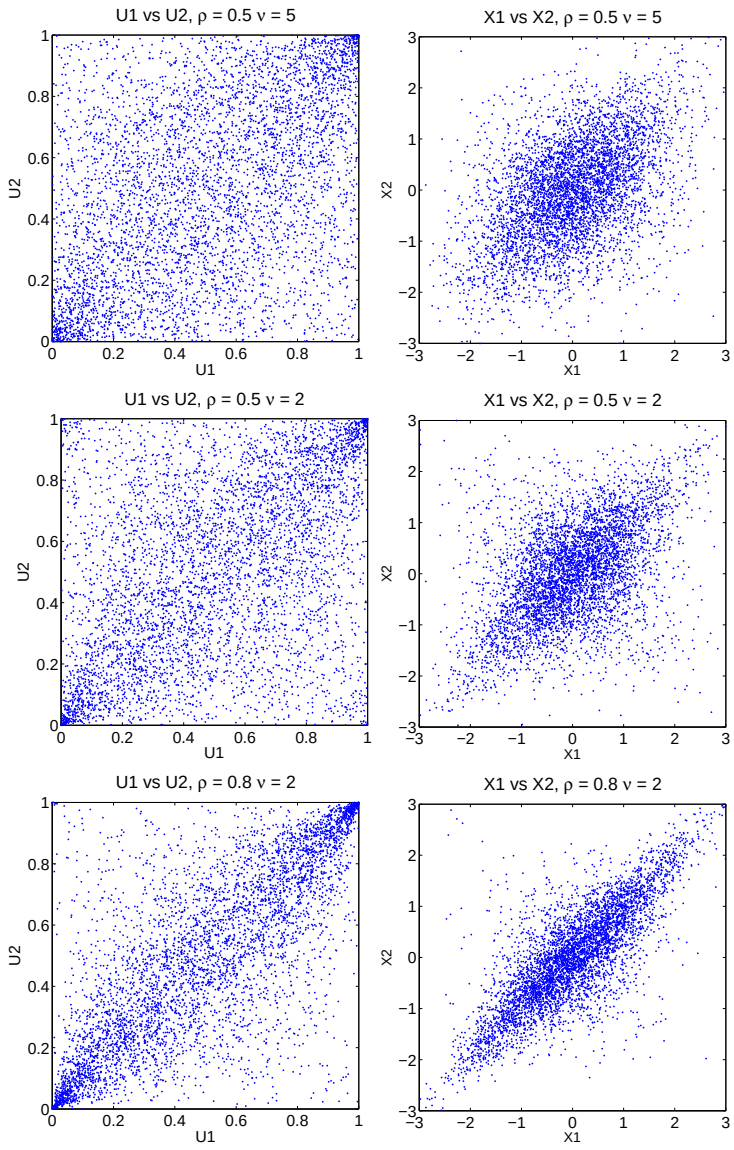


Figure 3.3: Standard t copula samples

The t copula is quite similar to the Gaussian copula in terms of its elliptical shape, however the t copula has distinct upper and lower tails. This makes it suitable to use under a risk modelling framework as it can capture tail dependences. The linear correlation parameter ρ affects the main body of the samples. We can see that by increasing ρ from 0.5 to 0.8, the shape of the sample becomes more stretched like in the Gaussian case. The ν parameter governs the tail behaviour of the samples. A ν of 2 exhibits stronger presence in the tails than a ν of 5.

Gumbel Copula

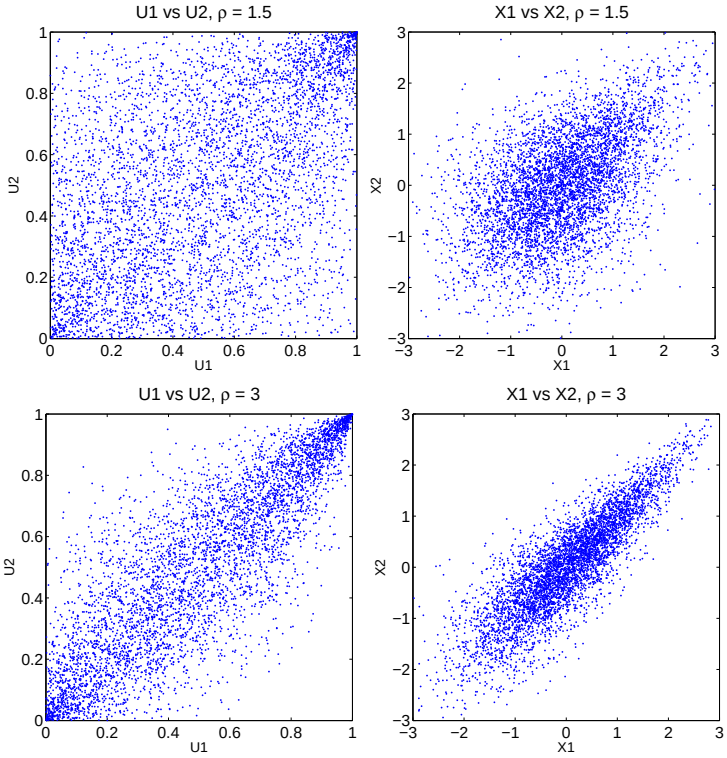


Figure 3.4: Gumbel copula samples

The Gumbel copula is not symmetric like the Gaussian and t copulas. A strong upper tail dependence is observable and the samples are skewed to the upper right tail. This too makes it a suitable copula to model OpRisk under the Basel II framework as we are interested in the 99.9% VaR, which is a measure in the upper tail.

3.1.2 Advanced Copulas

Mixture Copulas

Bivariate scatterplots for several mixture copula models are presented below. We will analyze how changing the weights and parameters of the component copulas affect the shape of the sampled distributions.

Gaussian & Gumbel

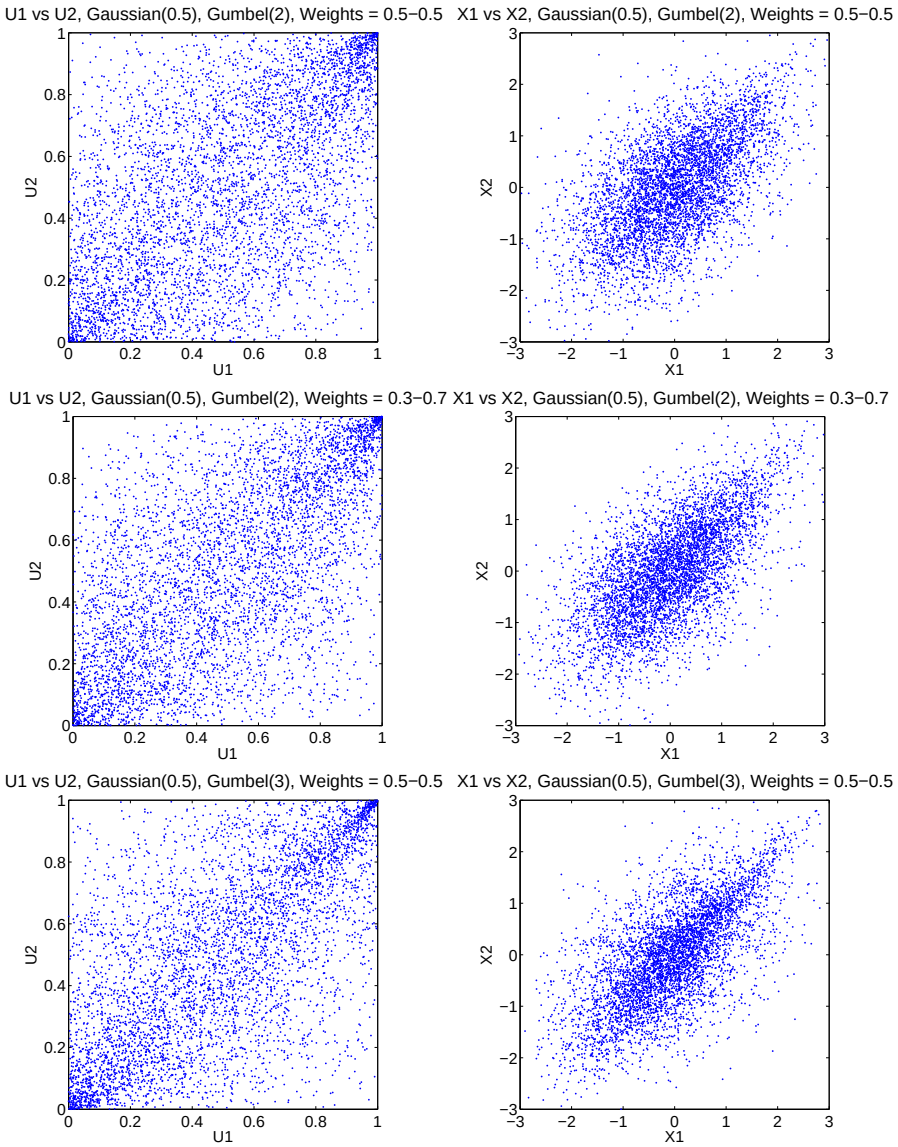


Figure 3.5: Gaussian & Gumbel mixture copula samples

The first row of plots show the Uniform copula samples and their respective Normal samples of a generated using a mixture of a Gaussian copula with $\rho_{Gaussian} = 0.5$ and a Gumbel copula with $\rho_{Gumbel} = 2$ in equal proportions. We notice a bit of tail dependence in the Uniform plot and a skew to the upper tail in the Normal one. However, the upper tail dependence is much less observable than in a pure Gumbel copula with the same parameter. If we increase the weight of the Gumbel copula from 0.5 to 0.7, as shown in the second row of graphs, we notice that the tail dependence increases slightly, however it is not a significant change. In contrast, if we increase the parameter of the Gumbel copula from 2 to 3 while keeping the weights at 50-50, there is a much stronger increase in both the tail dependence and the

overall dependence. In all of Normal plots, the body of the distributions are all elliptical in shape, while still exhibiting upper tail dependence. This mixed copula is useful if we wish to model losses that have upper tail dependence, but still want to keep the body of the distribution Normal.

Gumbel & Clayton

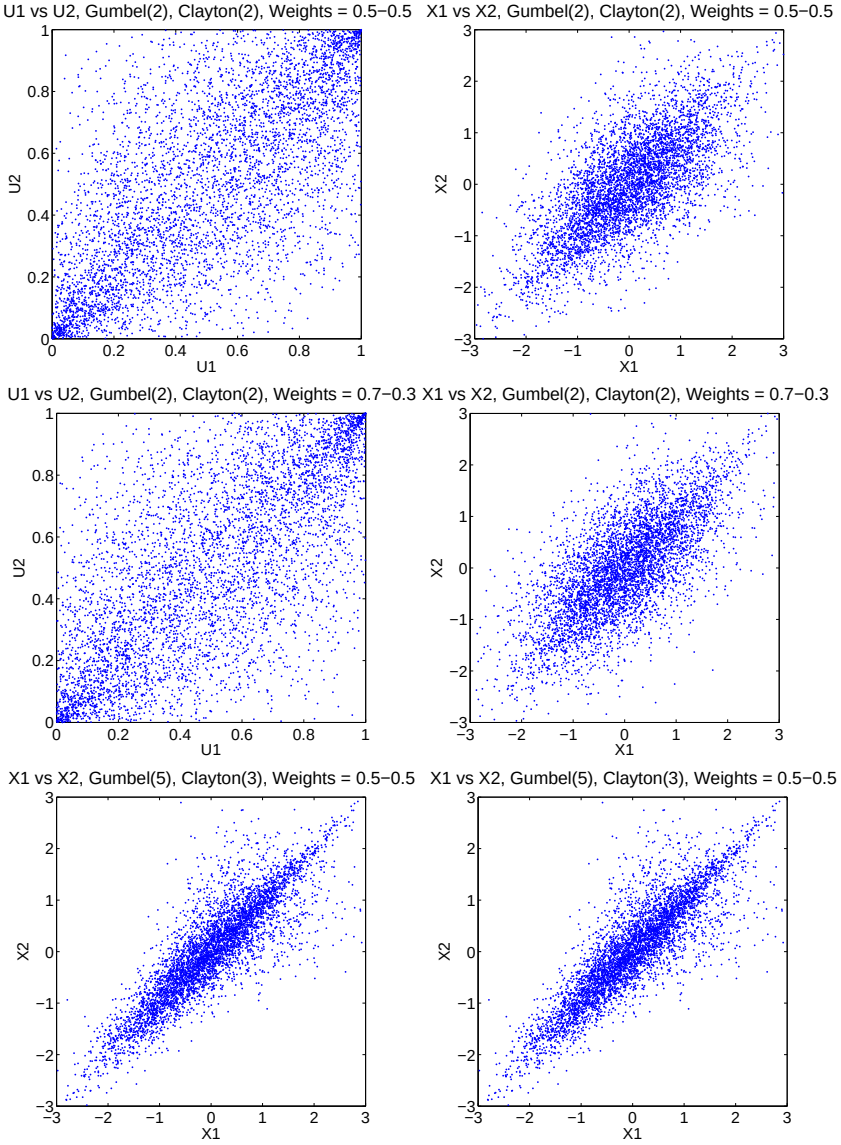


Figure 3.6: Gumbel & Clayton mixture copula samples

When a Gumbel copula with $\rho_{Gumbel} = 2$ is mixed with a Clayton copula with $\rho_{Clayton} = 2$ in equal proportions, the resulting uniform samples have an both an upper and lower tail, a combination of the characteristic features of the component copulas. Increasing the weight of the Gumbel copula heightens the upper tail dependence, but decreases the lower tail dependence of the sample. Changing the Gumbel parameter from 2 to 3 and the Clayton parameter from 2 to 5 while keeping the weights at 50-50 greatly increases both the overall dependence as well as the upper and lower tail dependence.

The behaviours of mixing a Gumbel copula with both Gaussian and Clayton models are consistent. The prominent property of the Gumbel copula, upper tail dependence, is reflected in the sampled distribution

after mixing. From a OpRisk modelling perspective, this addresses the main deficiency of the Gaussian copula and makes a mixture of Gaussian and Gumbel desirable to use. Since we are not so interested in lower tail dependence of losses, the mixture of Clayton and Gumbel does not really add any further advantages to a pure Gumbel copula and is less preferable due to the additional effort required in estimation.

Vine Copulas

Uniform random numbers sampled from selected trivariate vine copulas are studied below. In each case, the following D vine decomposition is used:

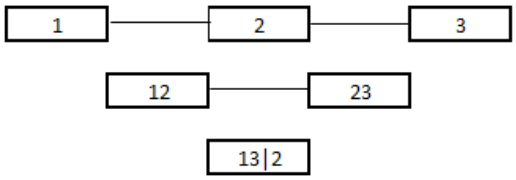


Figure 3.7: D-vine structure

The pairwise dependence structures that need to be specified according to the vine above are:

- C_{12}, the copula between X1 and X2,
- C_{23}, the copula between X2 and X3,
- $C_{13|2}$, the conditional copula between X1 and X3 given X2.

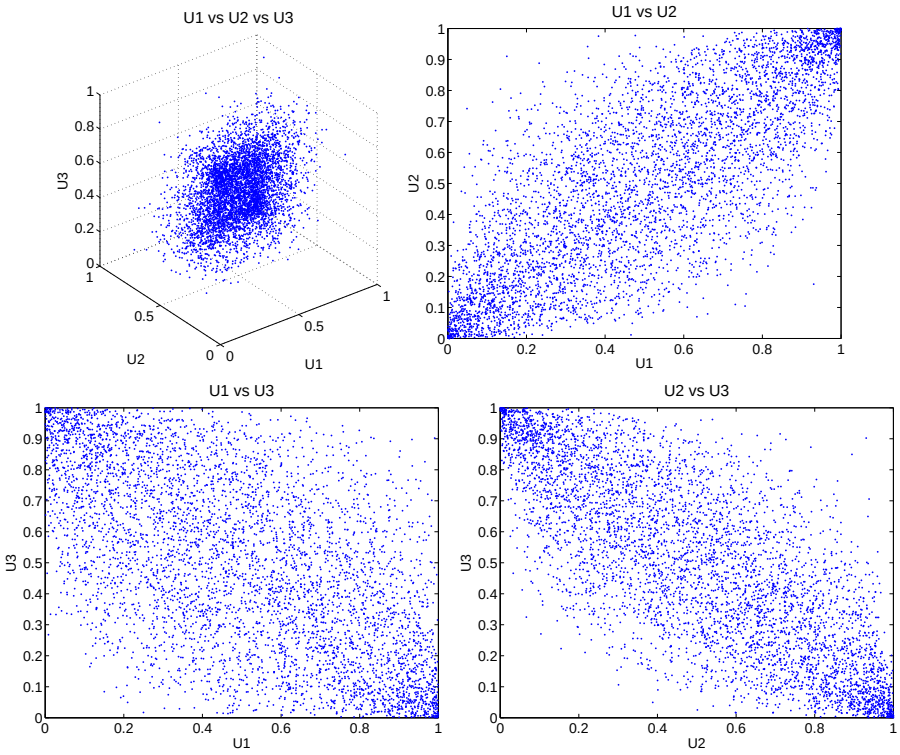


Figure 3.8: D-vine copula samples

The sample above is generated with the following dependence structure:

- C_{12} is a Gaussian copula with $\rho = 0.8$,

- C_{23} is a Gaussian copula with $\rho = -0.8$,
- $C_{13|2}$ is a Gaussian copula with $\rho = 0$.

A comparison of the pairwise scatterplots with the one presented in section 2.1, we can see that the dependence structure between X1 and X2 and between X2 and X3 are consistent with what is specified. However, despite X1 and X3 being conditionally independent given X2, we can see that there is still significant negative correlation between them. This is effect is due to induced correlation as X1 and X3 are both correlated with X2.

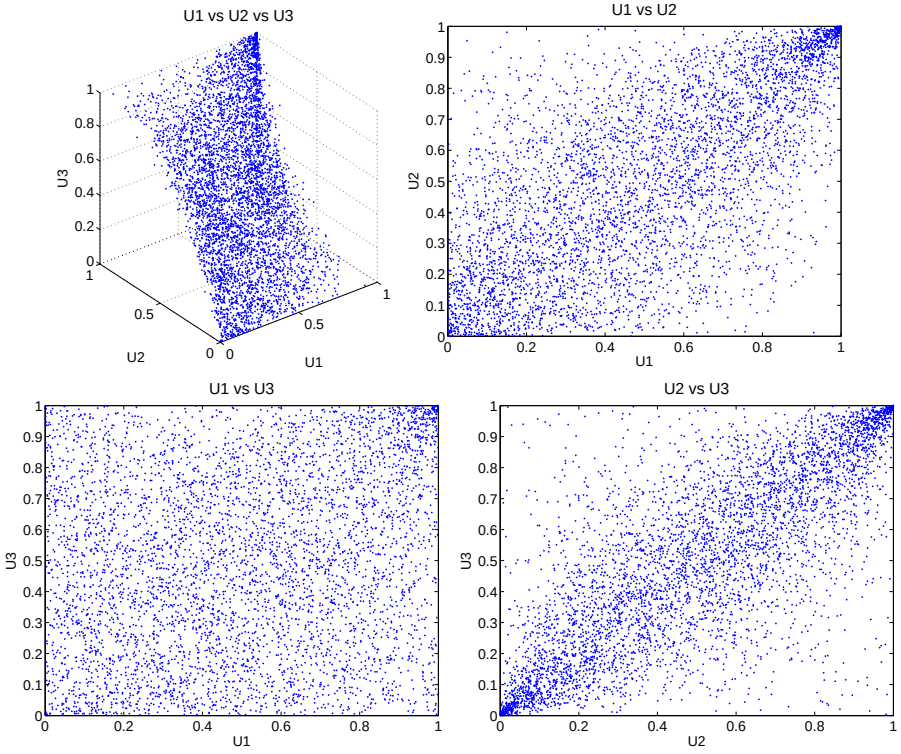


Figure 3.9: D-vine samples

In this second sample presented above, the following dependence structure was used:

- C_{12} is a Gumbel copula with $\rho = 2$,
- C_{23} is a t copula with $\rho = 0.8, \nu = 3$,
- $C_{13|2}$ is a Gaussian copula with $\rho = -0.8$.

The declared dependence relationship between the marginals X2-X3 and X1-X2 can clearly be seen in the pairwise scatterplots. However, Although a Gaussian copula with negative correlation is declared for the conditional copula between X1 and X3, our sampled bivariate distribution appears to show independence between X1 and X3, with perhaps a slight tail dependence in the upper left and right tails. Again this is shows the effect of induced correlation between X1 ans X3 from X2 and highlights the difficulty of properly specifying the desired relationship between variables via conditional copulas.

In OpRisk, vine copulas allow us to model pairwise dependences between risk cells individually. For instance, in the first example above, if we know that the dependence between X1 and X2 and between X2 and X3 are both Gaussian but have different correlation parameters, then a vine copula approach is more desirable as it allows us to directly capture this instead of calibrating a trivariate model. This convenience is amplified as the number of dimensions increase as multivariate models are in general difficult to fit

especially in rare event settings. The trade-off is the potentially number of pairwise dependences we need to specify. Providing the distribution of conditional copulas may also pose a challenge as we need to account for induced correlations as shown in the example above.

3.2 Copula Model Application in the Loss Distribution Approach

Basic Loss Model

We will now look at the effects of using various copula models in the context of a Loss Distribution Approach. In each of the studies below, the losses will be modelled using the following compound distribution framework:

$$Z_t = \sum_{j=1}^{J} Z_t^{(j)}$$
$$Z_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} X_i^{(j)}(t)$$

where

- J is the number of risk cells,
- t is the discrete time in annual units,
- Z_t is the total annual loss in year t across all risk cells,
- $Z_t^{(j)}$ is the annual loss for risk cell j in year t,
- $N_t^{(j)}$ is the frequency of events for risk cell j in year t,
- $X_i^{(j)}$ is the severity of events for risk cell j in year t.

Capital Estimation

The VaR of the compound distribution above with dependence between the frequencies and/or severities can be determined using the Monte Carlo simulation procedure outlined below. Let:

$$N_t^{(j)} \sim F(\lambda_t^{(j)})$$
$$X_t^{(j)} \sim G(\varphi_t^{(j)})$$

For each year/iteration t:

1. Simulate K sets of T x J dependent Uniform(0,1) random numbers from selected copula models, where K is the number correlated items and T is the total number of years simulated. These random numbers are used to generate samples from the dependent frequencies and/or severities (profiles).
2. Generate (dependent) samples for frequency parameter(s) $\lambda_t^{(j)}$ and severity parameters $\varphi_t^{(j)}$ if needed. If we assume the risk parameters are constant then this step is not necessary.
3. Generate (dependent) samples for frequencies $N_t^{(j)}|\lambda_t^{(j)}$ and severities $X^{(j)}(t)|\varphi_t^{(j)}$
4. Aggregate all of the losses at each t for each risk cell to get $Z_t^{(j)}$
5. Aggregate the losses across all risk cells to get Z_t
6. Take the n^{th} percentile value of Z_t across all t to obtain the $n\%$ VaR.

3.2.1 Standard Copulas

The effects of imposing selected dependence structures on both the frequency/severity level and the frequency/severity profile level of various loss distributions have been investigated. In particular the induced Kendall's Tau rank correlation between the risk cells, the 95% and 99.9% Value-at-Risk are calculated. The test cases below are all carried out under a bivariate setting where the distributions of both risk cells are identical. Furthermore, where applicable, the same copula model is applied to frequencies and severities (profiles).

When dependence is imposed on the frequency and severities levels, loss distribution has the following properties:

- $J = 2$,
- $N_t \sim$Poisson(10),
- $X_t \sim$Lognormal$(5, 1)$.

When dependence is imposed on the frequency and severities profiles, loss distribution has the following properties:

- J = 2,
- $N_t \sim$ Poisson(λ_t), $\lambda_t \sim$ Gamma(5,2),
- $X_t \sim$ Lognormal$(\mu_t, 1)$, $\mu_t \sim$ Normal(5,1).

The following copula models were considered:

- Gaussian copula,
- Standard t copula with 3 degrees of freedom,
- Gumbel copula.

The Gumbel and t copula both have upper tail dependence so they are suitable for modelling dependences in an OpRisk setting. The Gaussian copula is a popular model in financial applications and is used for comparatively purposes.

Application of dependence to various levels of the loss distribution

In this study, the Gaussian copula is applied to the frequency, severity and both the frequency and severity levels as well as their corresponding profiles. The graphs below illustrate the induced Kendall's tau between the risk factors when the Gaussian copula is applied to different levels and profiles. The 95% and 99.9% VaRs are also shown. The error bars shown on the graphs are calculated by sub-sampling (see Appendix). The results are based on 100,000 years of simulation. We will discuss the appropriateness of applying dependence at different levels of the loss distribution.

Frequency and Severity Levels

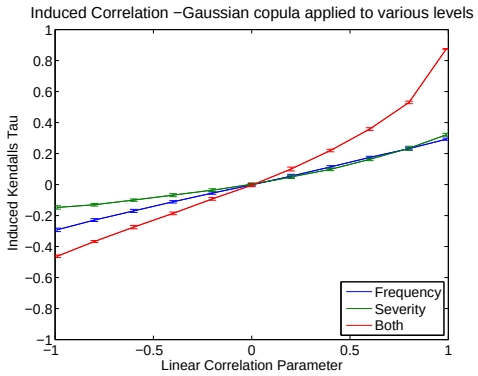


Figure 3.10: Induced Kendall's τ between losses when Gaussian copula is applied at various levels.

In scenarios of positive correlation between risk factors, there is no distinguishable difference between applying dependence on the frequency level or the severity level in terms of the induced Kendall's tau on the losses. Imposing dependence on both levels reinforce the induced correlation. We can see that modelling dependence using frequency or severity only restricts the range of the induced correlations between losses and so it is theoretically more flexible to account for associations in both the frequencies and the severities of losses. However imposing dependence in two areas means increased chance of model misspecification which can lead to compounded bias in estimation. This is heightened by the fact that the limited availability of OpRisk data means this is likely to occur.

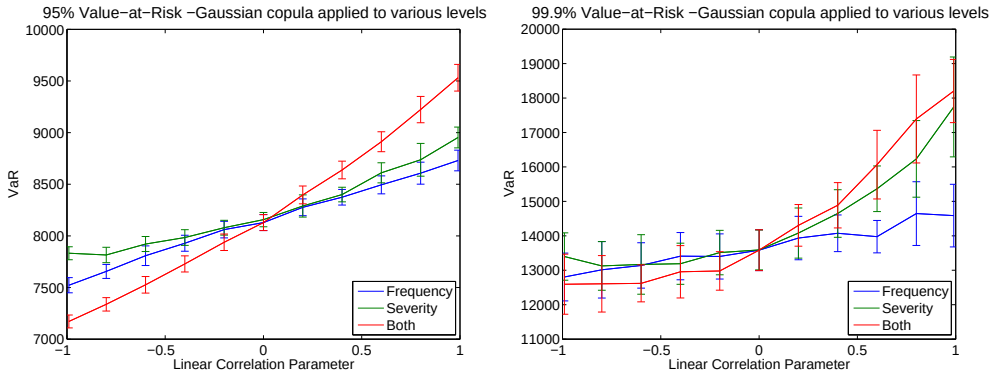


Figure 3.11: VaR when Gaussian copula is applied at various levels.

The trend of the VaRs are similar to those of the induced correlations. However for the 99.9% VaRs, correlated severities have a significant effect. It is important to note that the 99.9% VaR is in the extreme tail of the distribution and hence capital estimation will be relatively less accurate than a lower quantile. This is evident in the larger error bars in the 99.9% results than the 95% results. For the methodology to on how the error bars were calculated, see the Appendix. This means that more data as well as more simulations are required to approximate the 99.9% VaR well which is a problem in OpRisk with the scarcity of data.

Frequency and Severity Profiles

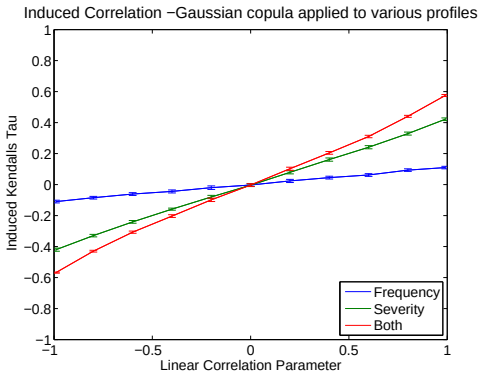


Figure 3.12: Induced Kendall's τ when Gaussian copula is applied at various risk profiles.

The dependent severity profiles induce higher correlation between losses than frequency profiles, which barely induces any correlation between losses at all.

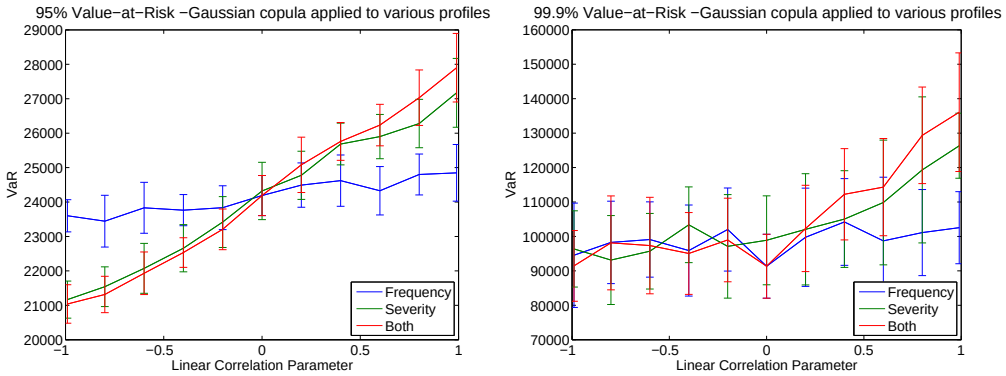


Figure 3.13: VaR when Gaussian copula is applied at various risk profiles.

The VaR graphs confirm that frequency profile dependence only results in minuscule changes in the VaR values compared to the independent, or zero correlation, cases so it may be advisable to exclude frequency dependences in a model where parameters of risk factors are modelled as random variables.

In general, the magnitude of the induced correlation when dependences are imposed on the risk profiles are larger for cases of positive dependence and smaller when the correlation is negative. Furthermore, the error bars of VaR in risk profile dependence modelling is noticeably wider than when dependence is applied at the frequency and severity levels. This reflects the extra layer of simulation required when frequency and severity parameters are modelled as random processes and has negative implications on OpRisk modelling because the extra uncertainty translates to the need for more data. Furthermore, to achieve a similar level of error, more iterations must be run in the simulation and hence more time is required for capital estimations of a satisfactory quality.

Application of various dependence structures

The Gaussian, t and Gumbel copulas were applied to both the frequency and severity levels as well as both corresponding profiles. The graphs below illustrate the induced Kendall's tau between the risk factors as well as the 95% and 99.9% VaRs. The results are based on 100,000 years of simulation. We will look at the effects of different dependence structures on the induced correlation and the VaR of the joint losses and observe when diversification benefits are realized (if at all).

Note that for comparative purposes, we derive a approximate linear correlation parameter of the Gumbel copula by first finding its Kendall's tau approximation:

$$\tau = 1 - \frac{1}{\rho_{Gumbel}} \tag{3.1}$$

and then transforming it into a linear correlation coefficient using (15).

Frequency and Severity Levels

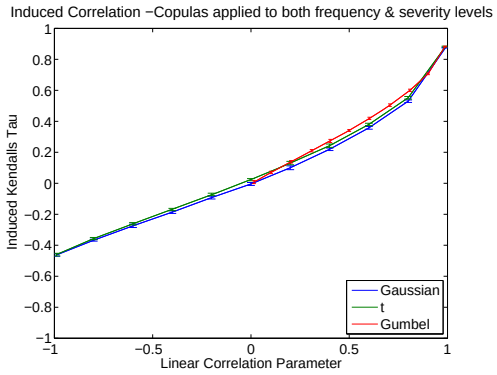


Figure 3.14: Induced Kendall's τ when various copulas are applied to both frequencies and severities.

Although it is clear that the Gumbel copula induces the highest correlation between losses, no real significant differences between the copula models are observed. However, we cannot naively take this to mean that the joint loss behaviour is the same across dependence copulas as the Kendall's tau rank correlation does not take the magnitudes of the losses into account and for risk processes with heavy tails, significant differences can occur. We must turn to the VaR results to see the real implications of the different copula models.

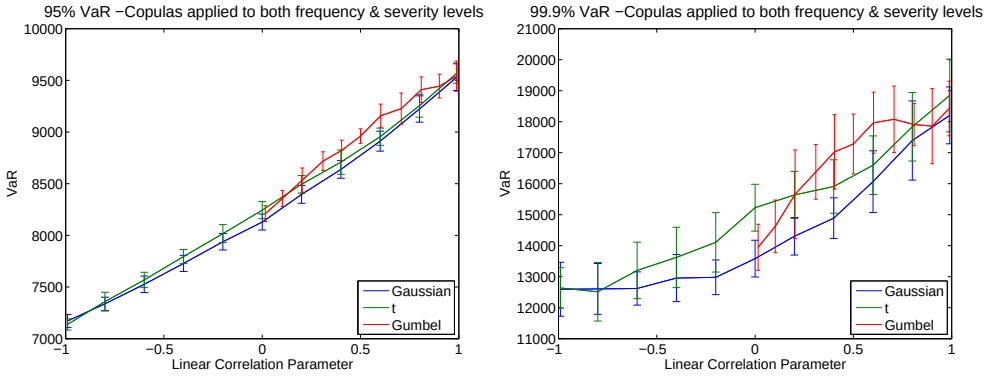


Figure 3.15: VaR when various copulas are applied to both frequencies and severities.

The VaR graphs show significant differences between the VaR results of different copulas. For all three copulas, we see an increasing trend with respect to the linear correlation parameter. This means that negative diversification benefits are realized when losses are positive correlated, in other words the VaR is higher when losses are positively correlated than when they are independent. This can be attributed to the fact that the Lognormal severity distributions are heavy tailed so higher positive dependence leads to a higher number of large scale losses. On the other hand, capital reduction can be achieved when the correlation between losses are negative, with high magnitude losses offset by less severe ones.

The Gaussian copula consistently produces lower VaR results than the t and Gumbel copula, due to having zero tail dependence, while the Gumbel copula produces the highest VaR results except in the most extreme case of positive correlation. While modelling using a Gaussian copula can reduce the reported capital requirements, the true VaR can actually be substantially more, leading to potential insolvency. Modelling via Gumbel copula on the other hand, will never lead to capital reduction as it is restricted to positive dependences.

Frequency and Severity Profiles

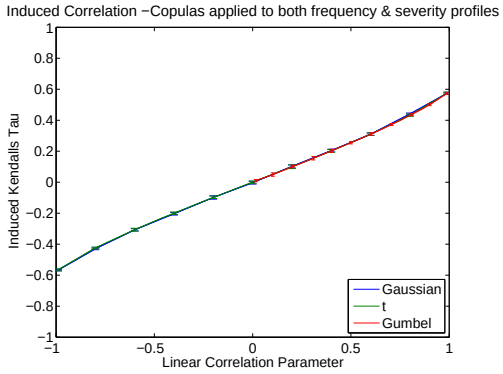


Figure 3.16: Induced Kendall's τ when various copulas are applied to both frequency and severity profiles.

The difference between copula models on the induced correlation is non-existent when dependence is applied on the risk profiles. This is also reflected somewhat in the VaRs.

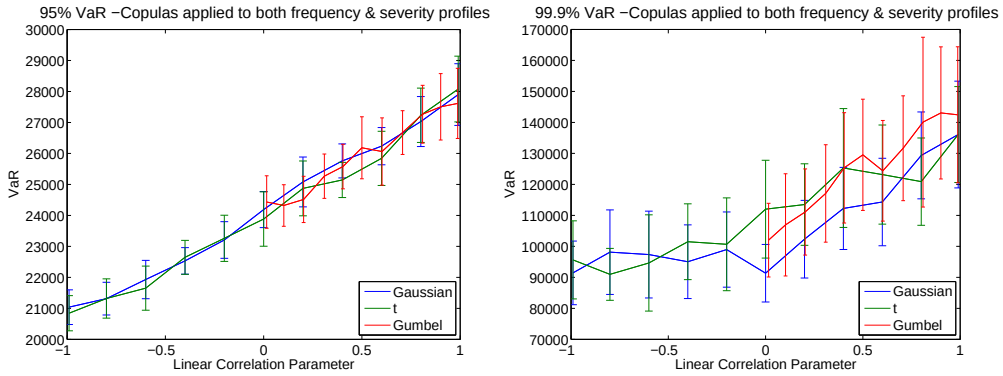


Figure 3.17: VaR when various copulas are applied to both frequency and severity profiles.

Unlike the previous model, significant differences between copula models cannot be observed when dependence is applied to the risk profiles in both the induced correlations at the 95% VaR level. Although the 99.9% VaR shows a pattern similar to the previous scenario, where using a Gumbel copula yields the highest results and the Gaussian the lowest except at extreme correlations, the large errors make this claim inconclusive. This can be due to the extra layer of randomness eliminating any distinct effects of the different dependence structures. From an OpRisk modelling point of view, it means we can choose the simplest dependence model to fit as there are no real differences between the impacts of the models. However, we should be wary of the large uncertainty in the extreme tail estimates. Nevertheless, the increasing trend of VaR with respect to correlation still holds. Risk factors that are negatively correlated will result in capital reductions for the bank while positively correlated risk factors will require the bank to hold more capital when dependence is accounted for in our model.

3.2.2 Advanced Copulas

Mixture copulas

Mixed copulas were applied to both the frequency and severity levels. The resulting correlation and VaR are compared to cases where only a pure copula was used. The investigated loss distributions have the following properties:

- $J = 2$,
- $N \sim$Poisson(10),
- $X \sim$Lognormal(5,1).

The following mixed copulas were applied:

- Gaussian copula with $\rho_{Gaussian} = 0.5$ and Gumbel copula with varying parameters, 50-50 weights,
- Gaussian copula with varying parameters and Gumbel copula with $\rho_{Gumbel} = 3$, 50-50 weights,
- Gaussian copula with $\rho_{Gaussian} = 0.5$ and Gumbel copula with $\rho_{Gumbel} = 3$, varying weights.

The graphs below illustrate the induced Kendall's Tau between the risk cells and the 99.9% VaR. The results are based on 100,000 years of simulation.

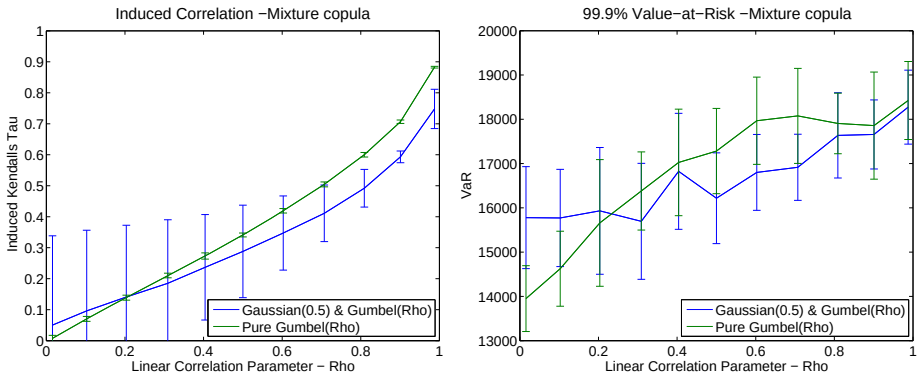


Figure 3.18: Mixture copula compared to Gumbel copula

Gumbel copulas of varying dependence were mixed with a Gaussian copula with linear correlation 0.5. Comparison of the induced correlation between losses and joint VaR are made with a pure Gumbel copula case.

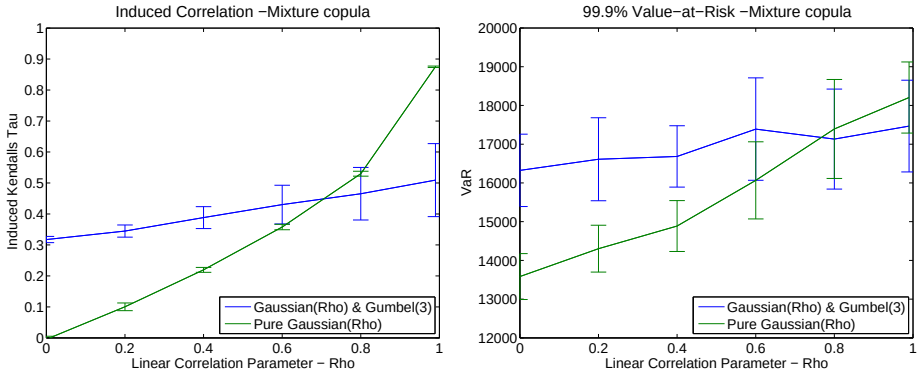


Figure 3.19: Mixture copula compared to Gaussian copula

Gaussian copulas of varying dependence were mixed with a Gumbel copula with dependence parameter 3. This is compared to a pure Gaussian copula case.

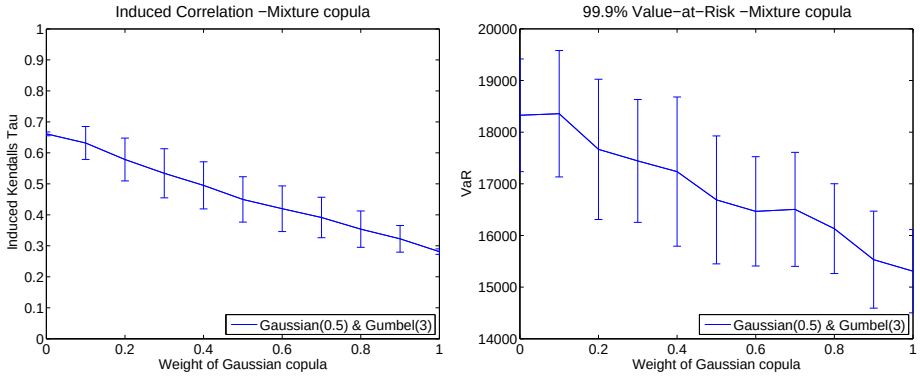


Figure 3.20: Mixture copula with varying weights

Behaviour of the induced correlation and VaR are shown for varying weights of the component copulas of fixed parameters.

We observe from the first graph that adding a Gaussian(0.5) copula to a Gumbel copula will increase the induced correlation in low dependence cases but will decrease the induced correlation as the dependence parameter of the Gumbel copula increases. Similarly, in the third graph, the Gumbel copula added to the Gaussian copula will bring up the induced Kendall's tau when the linear correlation parameter of the Gaussian copula is low but will decrease it in for high correlation parameters, although the crossing point is at a much higher correlation level than the first case. This implies that the overall induced correlation of mixed copulas are a weighted average of the component correlations. Consequently, we can see that combining a Gaussian copula with a Gumbel copula can add flexibility or control to the induced correlations. For instance, in the first graph, by mixing a Gaussian(0.5) copula with a weight of 0.5 to Gumbel copulas, we can keep the induced correlation within a consistent range. In the second figure, mixing a Gumbel(3) copula with Gaussian copulas will consistently increase the induced correlation compared to a pure Gaussian except at the highest levels of dependence.

The relationship between induced correlations of the mixture copula and pure copula are a good reflection of the behaviour of VaR of the mixture model compared to the pure model. As the second and fourth graphs demonstrate, the VaR lines between the pure and mixed copula intersect roughly where the corresponding induced correlation lines do and the general trend and relationship in the VaRs between using a mixture and a pure copula is similar to the induced correlations. Furthermore, the induced correlation and VaRs also vary linearly with the weights of the component copulas as illustrated in the fifth and sixth figures. The limits of mixture copula is its component copulas as the weights approach the boundary of [0,1]. We can see from these results that if we have prior knowledge or assumptions regarding either the correlation or VaR range of losses, then combining appropriate copula distributions can help achieve that target.

From a OpRisk modelling point of view, mixture copulas allows us to model a more diverse range of dependence structures by using only the basic copula models. However, scarcity of OpRisk data may have negative implications on model fitting due to the increased number of parameters.

Vine copulas

In this section we study the effects of vine copulas when applied to both the frequency and the frequency profile levels. Resulting VaR values were computed and analyzed for varying dependence parameters. The pair copula decomposition employed was in the shape of a d-vine structure as follows:

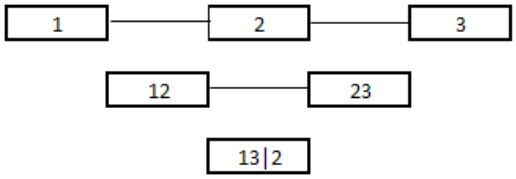


Figure 3.21: D-vine structure

We will study the following four different vine decompositions:

- All pair copulas are t with $\nu = 3$ and the same correlation parameter.
- All pair copulas are Gumbel with the same dependence parameter.
- C_{12}, C_{23} are t copula with a fixed parameters $\rho = 0.5, \nu = 3$ and $C_{13|2}$ is Gumbel. This case will be referred to as "t-t-Gumbel".
- C_{12}, $C_{13|2}$ are t copula with a fixed parameters $\rho = 0.5, \nu = 3$ and C_{23} is Gumbel. This case will be referred to as "t-Gumbel-t".

We want to examine the effects of choosing different pair copulas in the vine. First we will look at a comparison between choosing a vine that is entirely made up of identically distributed t copula and one that is entirely Gumbel, corresponding to decompositions 1 and 2 above. Next we will analyze the sensitivities of the VaR given a change in one particular pair copula within the structure, reflected in structures 3 and 4. The 95% and 99.9% VaR plotted as a function of the linear copula parameter are shown below.

Frequency Dependence
The investigated loss distributions have the following properties:

- $J = 3$,
- $N \sim$ Poisson(10),
- $X \sim$ Lognormal(5,1).

The selected graphs below show the 95% and 99.9% VaRs for each of the cases described above when the vine copula is applied to the frequencies. The results are based on 250,000 years of simulation. More simulations are carried out here than for the basic copula models as there is more uncertainty in the capital estimation due to the increased dimensionality and amount of copula models.

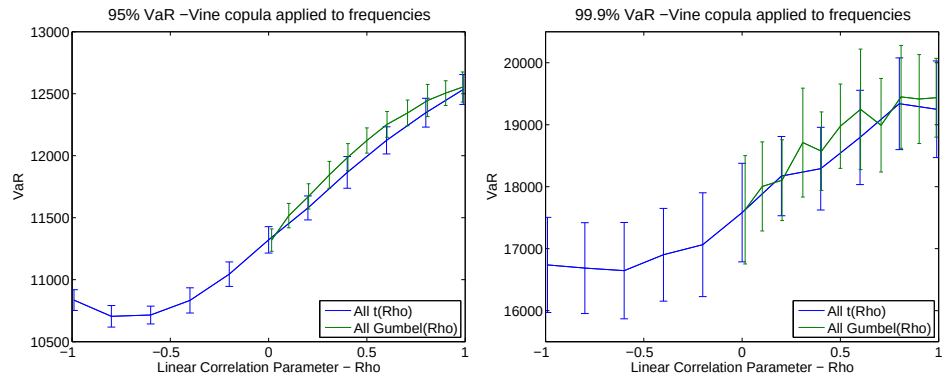


Figure 3.22: VaR when vine copulas are applied to frequencies

Vine structures with only t and Gumbel copulas are plotted against their linear correlation parameter. The behaviour of both the all-t and all-Gumbel VaRs are similar to their corresponding basic copulas.

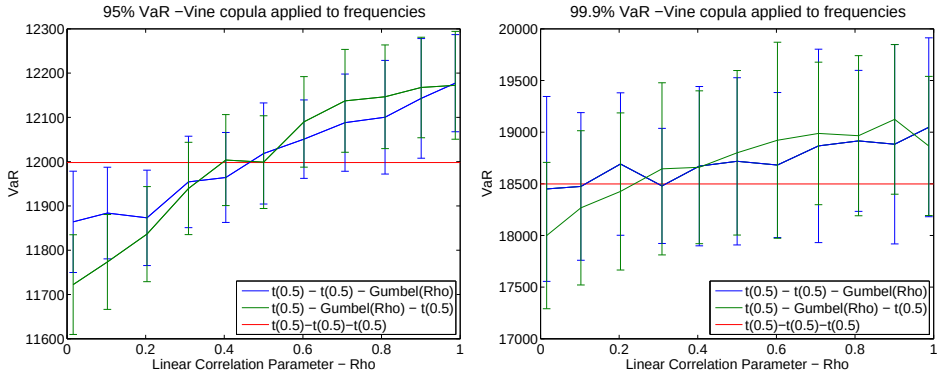


Figure 3.23: VaR when vine copulas are applied to frequencies

The effects of changing one pair-copula in the vine are shown in these graphs. While a difference in VaR can be witnessed, it is not substantial.

Frequency Profile Dependence

The investigated loss distributions have the following properties:

- $J = 3$,
- $N \sim \text{Poisson}(\lambda),\ \lambda \sim \text{Gamma}(5,2)$,
- $X \sim \text{Lognormal}(\mu, 1),\ \mu \sim \text{Normal}(5,1)$.

The selected graphs below show the 95% and 99.9% VaRs for each of the cases described above when the vine copula is applied to the frequency profiles. The results are based on 500,000 years of simulation, twice the amount used when dependence is applied on frequencies as the extra layer of simulation will cause more uncertainty in estimates.

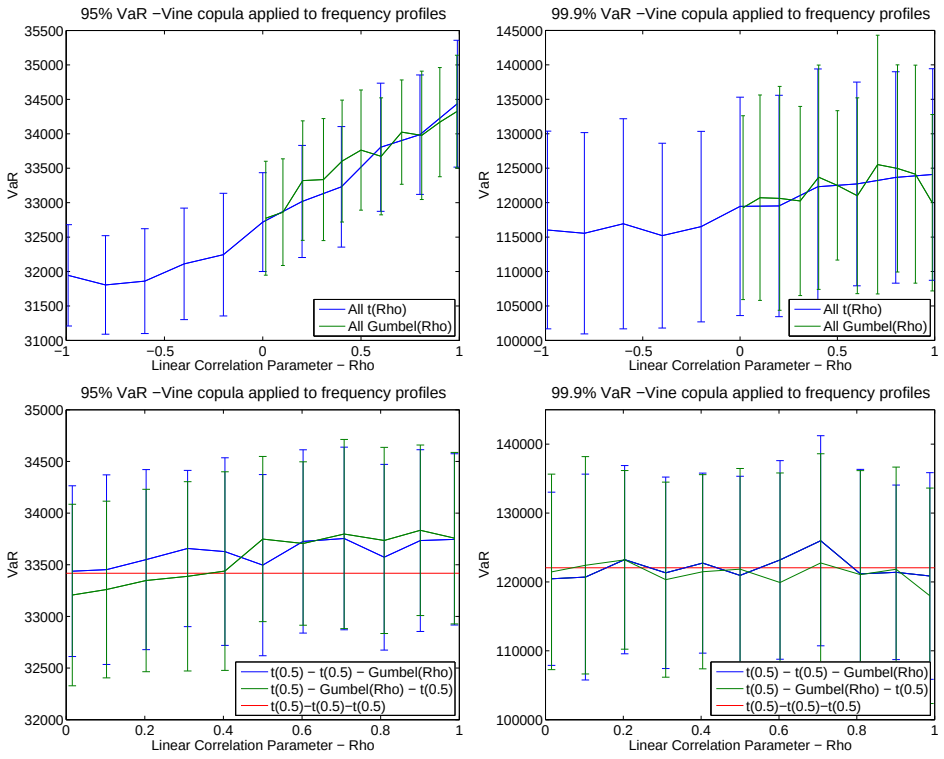


Figure 3.24: VaR when vine copulas are applied to frequency profiles

The results of applying the vine copula to frequency profiles yielded VaR curves with extremely high error bars. This is consistent with the results for the standard copulas due to increased layers of simulation. In the first row of graphs of both the correlated frequency and frequency profile cases, we observe that the behaviour of the VaR in the all-t and all-Gumbel cases match the corresponding basic copulas. The VaR exhibits an steadily increasing trend as the correlation parameter increases from -1 to 1 and the Gumbel copula causes higher VaR for low to medium correlation than the t copula but converge towards it for extreme positive correlations. This result reassures risk practitioners that consistency of results can be expected from vine copulas when all pair-copulas have the same structure when compared to using the corresponding basic copula. This allows large amounts of risks to be modelled without performing high dimensional estimation since bivariate copulas are used to construct the vine.

From the second row of graphs in each set, we speculate that while changing one pair-copula in a vine can have noticeable efforts on the VaR, the impact is not significant, especially when dependence is applied to the frequency profile. The horizontal line in each graph represents a benchmark VaR corresponding to a model where all pair-copulas are t with $\rho_{Linear} = 0.5$, $\nu = 3$. In all instances, changes to the VaR do not exceed 5% when a pair-copula was changed. The simulations also suggest that changes in the conditional pair copula do not induce as much of a change in VaR as changes made to the direct copula between two risk factor. This is apparent as the curve for the t-t-Gumbel case is usually further away from the benchmark than the curve for the t-Gumbel-t model.

Although individual pair-copula misspecifications may be insignificant to capital estimation, there is a good chance of multiple misspecifications in higher dimensional models as the number of pair-copulas increase rapidly. This is especially likely in an OpRisk context due to the lack of data. Furthermore, conditional copula distributions are generally hard to determine accurately and even though the effect is not as great as misspecifying a direct copula, differences in accumulated inaccuracies can result in significant VaR errors. Nevertheless, vine copulas are very flexible and can capture the wide range of dependence structures encountered in OpRisk if estimation and application are implemented correctly. Selecting and fitting models for a series of bivariate distributions is also generally more accurate than for multivariate distributions with scarce data sets.

Chapter 4

Parameter Estimation Procedures

Estimation of capital cannot be done until risks have been fit to appropriate models. In the parametric case, relevant model parameters need to be estimated. This section will discuss some popular techniques used to fit copula based parametric models.

4.1 Maximum Likelihood Estimation

A popular frequentist approach to parameter estimation is the maximum likelihood method, where the estimated parameters are chosen to maximize the likelihood function (how likely a set of data will be observed assuming those parameters). If samples of the dataset $\boldsymbol{Y}$ are independently identically distributed, then their likelihood function is just the product of their densities, given as:

$$L(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{d} f(\boldsymbol{Y}|\boldsymbol{\theta}) \tag{4.1}$$

where f is the density of the samples of $\boldsymbol{Y}$ with the set of underlying parameters $\boldsymbol{\theta}$. The Maximum Likelihood Estimator (MLE) in such a case is:

$$\tilde{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta}}[L(\boldsymbol{Y}|\boldsymbol{\theta})] \tag{4.2}$$

Asymptotically, $\boldsymbol{\theta} \xrightarrow{d} \text{Normal}(\boldsymbol{\theta}, I^{-1})$ as sample size grows, meaning that the MLE estimator is consistent. However, small sample sizes are commonly a problem, especially in operational risk, as they lead to biased estimates of the true model parameters.

Inference Function for Margins

The Inference Function for Margins (IFM) method is a two-step maximum likelihood procedure used in copula parameter estimation introduced in (**?**). Its main advantage is that it reduces high dimensional multivariate problems to univariate or bivariate problems where their solutions have been extensively researched and are computationally feasible.

Consider a multivariate distribution decomposed into a copula model parameterized by $\boldsymbol{\theta}$ and d marginals parameterized by $\boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}$ respectively. The distribution and density functions are given as:

$$F(\boldsymbol{y}; \boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}, \boldsymbol{\theta}) = C\big(F_1(y_1; \boldsymbol{\alpha_1}), ..., F_d(y_d; \boldsymbol{\alpha_d}); \boldsymbol{\theta}\big)$$

$$f(\boldsymbol{y}; \boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}, \boldsymbol{\theta}) = c\big(F_1(y_1; \boldsymbol{\alpha_1}), ..., F_d(y_d; \boldsymbol{\alpha_d}); \boldsymbol{\theta}\big) \prod_{j=1}^{d} f_j(y_j; \boldsymbol{\alpha_j})$$

The log-likelihood functions of the margins are:

$$L_j(\boldsymbol{\alpha_j}) = \sum_{i=1}^{n} \log f_j(y_{ij}; \boldsymbol{\alpha_j}),\ j = 1, ...d \tag{4.3}$$

and the log-likelihood for the joint distribution is:

$$L(\boldsymbol{\theta}, \boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}, \boldsymbol{\theta}) \tag{4.4}$$

An IFM algorithm for estimating all of the parameters in the above model is as follows:

1. The log-likelihood functions of the margins $L_j(\boldsymbol{\alpha_j})$ are maximized separately to obtain estimates for $\boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d}$
2. The log-likelihood function for the joint distribution given the marginals $L(\boldsymbol{\theta}|\boldsymbol{\alpha_1}, ..., \boldsymbol{\alpha_d})$ is maximized to obtain an estimate for the dependence parameter θ

Despite its advantage of simplicity, the IFM method tends to produce biased estimates of the true model parameters. There is increased model risk when using this approach as there is a higher chance of model misspecifications, since more components are modelled and biases can come from each individual estimation. This problem is amplified in OpRisk there is usually only limited amounts of data available for modelling so the convergence property of the MLE estimator does not usually apply for more sophisticated distributions. Consequently, other estimation techniques should be explored when trying to fit a copula model.

4.2 Bayesian Inference

A significantly different approach to the estimation problem is Bayesian inference. The main concept behind this method is the use of both data and prior beliefs regarding specific random process, in our case risk factors, to formulate a posterior distribution for the parameters of interest. The basic formula underpinning this framework is based on Baye's Theorem:

$$f(\boldsymbol{Y}, \boldsymbol{\theta}) = L(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\boldsymbol{Y})f(\boldsymbol{Y}) \tag{4.5}$$

where

- $f(\boldsymbol{Y}, \boldsymbol{\theta})$ is the density of the observed data $\boldsymbol{Y}$ and parameters$\boldsymbol{\theta}$,
- $L(\boldsymbol{Y}|\boldsymbol{\theta})$ is the likelihood of the data given the parameters,
- $\pi(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$,
- $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ is the posterior density of $\boldsymbol{\theta}$ given the data,
- $f(\boldsymbol{Y}) = \int L(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal density of Y

The posterior density is then given by:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) \propto L(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{4.6}$$

Unlike MLE, Bayesian inference determines a complete distribution for the parameter of interest rather than just a point estimate. For example, in an OpRisk context, if the frequencies of frauds $N \sim \text{Poisson}(\lambda)$, the MLE will find a point estimate for λ whilst the Bayesian approach will find the distribution of λ. One of the main advantages of being able to determine a distribution for the parameter is the ability to form a full predictive distribution of the annual loss process. The distribution of λ in this case is called

the posterior distribution. It does not typically have a closed form equation and must be approximated, usually via simulation methods. We will look at several ways to compute the posterior distribution in the next few sections.

The prior density in this setting typically comes from some initial knowledge of the interested variables, such as from external data and expert opinions. MLE can also be used to determine the parameters of the prior distribution. In this way, the Bayesian framework allows for the combination of different sources of data, as demanded by Basel II, and is an effective tool for model fitting.

Estimation of the Prior Distribution

There are certain cases where the choice of a certain family of priors will lead to closed form solutions for the corresponding posterior distributions. These are called conjugate pairs but due to their limited amount of coverage, will not be considered in this thesis. Two general approaches to formulating the prior distributions of risk parameters, which are described in (**?**), are the subjective approach (pure Bayesian) and the empirical approach. In an OpRisk context, a pure Bayesian approach corresponds to using expert opinions while the empirical approach can be seen as using external data. These are natural ways to satisfy the criteria for OpRisk modelling set out by Basel II which requires model be constructed through a combination of expert opinions, external data and internal data.

Pure Bayesian Approach

Under this method the prior distribution is subjectively specified, usually by industry experts. Several ways this can done are outlined below according to (**?**):

- Form a density using a smoothed histogram.
- Compare likelihoods given certain selections of parameters and distributions and find the best choice.
- Subjectively construct the CDF through curving sketching.
- Find the distribution assuming a specific functional form by choosing parameters that match prior beliefs.

The advantage of this approach lies in the fact that the use of empirical data alone does not allow as much intuition and common sense to be used.

Empirical Bayesian Approach

The determination of the prior distribution in this context is typically done through a frequentist approach. The parameter of interest θ is treated as a random variable from the prior distribution. Observations of losses, which may not be from the risk corresponding to θ, are used to estimate the parameters of the prior. Popular procedures include:

- Maximum Likelihood Estimation - where the (log)likelihood of the data given a choice of marginal distributions and parameters is maximized.
- Method of Moments - where the parameters of the chosen distribution are determined by equating the theoretical moments with the sample moments.

In an OpRisk context, due to the limited data available for individual banks, industry data and expert opinions are typically used to formulate the prior distribution. The bank's internal data can then be incorporated to find the posterior distribution.

Estimation of the Posterior Distribution

The posterior distribution of parameter $\boldsymbol{\theta}$ can be seen as its resultant estimate after both our prior believes and data have been taken into account. The posterior density given data $\boldsymbol{Y}$ is given as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) \propto L(\boldsymbol{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

Generally the posterior distribution is unknown so numerical estimation through simulation is required to determine its properties. Even if the posterior density has a closed form, it may be difficult to analytically derive the desired statistical properties such its mean and mode. In the latter case, simple methods such as the Inverse Transformation method or the Acceptance-Rejection method can be used to simulate samples for analysis.

Inverse Transformation Method
If $U \sim$ Uniform(0,1), then $X = F^{-1}(U)$ has distribution F. As a result, simulating from a distribution F can be done by simulating a Uniform(0,1) random variable and then taking $F^{-1}(U)$ as a sample from F.

For estimating the posterior density, the Inverse method cannot be used if the normalization constant is not known and so can only be applied if the full posterior distribution is known.

Acceptance-Rejection Method
If, for a density $g(x)$, there exists a constant M such that $Mg(x) \geq f(x)$, then to simulate X from $f(x)$:

1. Draw X from $g(x)$ and U from Uniform$\big(0, Mg(x)\big)$
2. Accept the sample X if $U \leq f(X)$, else repeat from step 1

The Acceptance-Rejection method cannot be used if the bounds for the desired density cannot be easily obtained.

Monte Carlo Markov Chain (MCMC) Methods

In general, when the density is unknown, MCMC methods can be used to simulate from the posterior distribution. The idea behind MCMC methods is that to obtain an approximation for an integral based on a distribution $\pi(\theta|x)$, we can sample from a Markov chain that has a limiting distribution $\pi(\theta|x)$.

Markov Chains
A Markov chain is a sequence of variables or vectors $\Theta^{(0)}, \Theta^{(1)}, ..., \Theta^{(l)}, ...$ where the distribution of $\Theta^{(l+1)}$ only depends on $\Theta^{(l)}$, that is:

$$P(\Theta^{(l+1)} = \theta^{(l+1)}|\Theta^{(0)} = \theta^{(0)}, ..., \Theta^{(l)} = \theta^{(l)}) = P(\Theta^{(l+1)} = \theta^{(l+1)}|\Theta^{(l)} = \theta^{(l)}).$$

The possible values of $\theta^{(l)}$ form a set called the state space of the Markov chain. The stationary or limiting distribution of a Markov chain is a distribution $\pi(\theta|x)$ such that if $\Theta^{(l)}$ is distributed from it, then so is $\Theta^{(l+1)}$. MCMC methods iteratively sample a Markov chain until it converges to have a stationary distribution, which in this case is designed to be the target posterior of model parameters conditional on observed OpRisk losses.

Metropolis-Hastings Algorithm
The Metropolis-Hasting algorithm is a generic algorithm to generate Markov chains with a stationary distribution. Given a target density $\pi(\theta|x)$ known up to a normalization constant and a proposal density $q(\theta^*|\theta)$, the algorithm generates the Markov chain $\theta^{(l)}$ via the following steps:

1. Initialize $\theta^{(0)}$ with any value and set $l = 1$
2. Set $\theta^{(l)} = \theta^{(l-1)}$
3. Generate the proposal θ^* from $q(\theta^*|\theta^{(l)})$

4. Accept the proposal with the acceptance probability

$$p(\theta^{(l)}, \theta^*) = \min\left(1, \frac{\pi(\theta^*|x)q(\theta^{(l)}|\theta^*)}{\pi(\theta^{(l)}|x)q(\theta^*|\theta(l))}\right)$$

5. If $l = L$ end, else set $l = l + 1$ and repeat from step 2

Gibbs Sampler

The Gibbs sampler is an algorithm for generating random variables from a multivariate distribution without having to calculate its joint density. Consider a random vector $\boldsymbol{X}$ with a joint density $f(\boldsymbol{x})$ and denote the full conditionals $f_x(x_i|x_{-i}) = f(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_N)$. Assume sampling X from $f(x)$ cannot be directly done but sampling X from the conditionals is easy, then the Gibbs sampler generates random samples of X via the following steps:

1. Initialize $x_2^{(0)}, ..., x_N^{(0)}$ with any value and set $l = 1$
2. Simulate $x_1^{(l)}$ from $f_1\left(x_1|x_2^{(l-1)}, ..., x_N^{(l-1)}\right)$
   Simulate $x_2^{(l)}$ from $f_2\left(x_2|x_1^{(l)}, x_3^{(l-1)}..., x_N^{(l-1)}\right)$
   $\vdots$
   Simulate $x_N^{(l)}$ from $f_N\left(x_N|x_2^{(l-1)}, ..., x_{N-1}^{(l-1)}\right)$
3. If $l = L$ end, else set $l = l + 1$ and repeat from step 2

Under general conditions, the Markov chain that is generated, $x^{(l)}$, has a stationary distribution that corresponds to the joint density $f(\boldsymbol{x})$.

Slice Sampler

In order to utilize the approaches outlined above, a suitable proposal distribution is needed. However, this requires tuning of the target distribution which is quite computationally intensive especially in high dimensional problems. If the full conditional distribution of the posteriors are known up to the normalizing constant, then an alternative which does not involve tuning is the Slice sampler, outlined in (**?**). This method is attractive because in most cases it is more efficient than MH and does not need a closed form solution to the normalization constant, which in OpRisk models are generally impossible to obtain.

The concept behind univariate Slice sampling is that sampling uniformly from the region under a univariate distribution $f(\theta)$ can always yield samples from $f(\theta)$. The algorithm for univariate Slice sampling is given as follows:

1. Initialize $\theta^{(0)}$ with value within the support of the target density $f(\theta)$ and set $l = 1$
2. Sample $u^{(l)}$ from Uniform$\big(0, p(\theta^{(l-1)})\big)$
3. Sample $\theta^{(l)}$ uniformly from the level set $A_l = \big(\theta : p(\theta) > u^{(l)}\big)$, i.e $\theta^{(l)} = U(A_l)$
4. Set $l = l + 1$ and repeat from 2

This concept can be extended into a multivariate context by using a Gibbs sampling framework. Each of the conditional marginal distributions can be sampled using the univariate Slice sampler. Consequently, sampling from the multivariate distribution can be achieved in a Bayesian framework if we know the form of the conditional posterior density up to the normalization constant. The general multivariate algorithm for generating from $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_d)$ is outline below:

1. Initialize $l = 0$ and the parameters to be estimated $\theta_{1,0}, \theta_{2,0}, ...\theta_{d,0}$.
2. For $l = 1, ..., L$

- For $i = 1, ..., d$ sample $\theta_{i,l}$ from the full conditional distribution $\pi(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_d)$ using the univariate Slice sampler.

3. Set $l = l + 1$ and return to Step 2

Chapter 5

Parameter Estimation Case Studies

In this chapter we will study the performance of the MLE and Bayesian parameter estimation techniques outlined above by looking at a series of test cases.

5.1 Maximum Likelihood Estimation

The precision of MLE estimation procedures when applied to selected copula models will be analyzed in this section. Copula structures will be applied to severities between risk cells under a bivariate setting. The severities distributions that are considered are:

- Lognormal,
- Normal Inverse Gaussian.

The properties of these distribution are outlined in Chapter 2. The copula models that are imposed on the severities are:

- Gaussian,
- Standard t,
- Gumbel,
- Mixture of Standard t and Gumbel.

Estimation of the parameters will be carried out using IFM, where the severity marginals are fitted first and the copula parameters are determined afterwards.

5.1.1 Likelihood Functions

Suppose we have data points $\boldsymbol{X} = (\boldsymbol{X_1}, ..., \boldsymbol{X_n})$, where $\boldsymbol{X_i} = (x_{i1}, ..., x_{id})$. The likelihood functions of the above marginal severity and copula distributions are listed below. These functions are maximized to obtained a point estimate for the underlying parameters of the respective distributions. Where explicit forms cannot be derived for MLE estimators, numerical optimization is employed to determine them.

Severity Distributions

Lognormal

$$L(\boldsymbol{X}|\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{x_i\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x_i-\mu)^2}{2\sigma^2}} \tag{5.1}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n} \tag{5.2}$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}{n} \tag{5.3}$$

Closed form solution to the MLEs of the Lognormal solution are available making estimation simple.

Normal Inverse Gaussian

$$L(\boldsymbol{X}|\alpha,\beta,\mu,\delta) = \prod_{i=1}^{n} \frac{\delta\alpha \exp\big(\delta\sqrt{\alpha^2-\beta^2} + \beta(x_i-\mu)\big)}{\pi\sqrt{\delta^2+(x_i-\mu)^2}} K_1\big(\alpha\sqrt{\delta^2+(x_i-\mu)^2}\big) \tag{5.4}$$

where $K_1(x)$ is a modified Bessel function of the second kind with index 1. No closed form solution exists for the MLEs so numerical methods must be employed to approximate them.

Copula Models

Closed form solutions to the MLEs of copula parameters are generally not available so numerical methods are used to maximize the likelihood functions.

Gaussian

$$L(\boldsymbol{X}|\boldsymbol{\Sigma}) = \prod_{i=1}^{n} \frac{f_N^{\Sigma}\big(F_N^{-1}(x_{i1}),...,F_N^{-1}(x_{id})\big)}{\prod_{j=1}^{d} f_N\big(F_N^{-1}(x_{ij})\big)} \tag{5.5}$$

Generalized t

$$L(\boldsymbol{X}|\boldsymbol{\Sigma},\boldsymbol{\nu}) = \prod_{i=1}^{n} \frac{1}{\prod_{k=1}^{d} f_{\nu_k}(x_{ik})} \int_0^1 \varphi_{\Sigma}\big(z_1(x_{i1},s),...z_n(x_{id},s)\big) \prod_{k=1}^{d} \big(w_k(s)\big)^{-1} ds \tag{5.6}$$

The standard t copula is a special case of the generalized t copula where $\nu_k = \nu$ for all $k = 1,...,d$.

Gumbel - Bivariate Case

$$L(\boldsymbol{X}|\rho) = \prod_{i=1}^{n} C(x_{i1},x_{i2}) x_{i1}^{-1} x_{i2}^{-1} \left[\sum_{j=1}^{2} (-\ln(x_{ij}))^{\rho}\right]^{2(\frac{1}{\rho}-1)} [\ln(x_{i1})\ln(x_{i2})]^{\rho-1}$$
$$\times \left[1 + (\rho-1)\left[\sum_{j=1}^{2} \big(-\ln(x_{ij})\big)^{\rho}\right]^{-\frac{1}{\rho}}\right] \tag{5.7}$$

where

$$C(x_{i1}, x_{i2}) = \exp\left(- \left[\sum_{j=1}^{2} \big(-\log(x_{ij}) \big)^{\rho} \right]^{\frac{1}{\rho}} \right)$$

There is no general closed form expression for the Gumbel copula density in an arbitrary number of dimensions so derivation using partial differentiation must be used to find an analytical form for the likelihood function in each dimension. Alternatively, numerical differentiation can also be employed.

Generalized t & Gumbel Mixture - Bivariate Case

$$\begin{aligned} L(\boldsymbol{X}|\rho_t, \boldsymbol{\nu}, \rho_G, \boldsymbol{W}) = \prod_{i=1}^{n} \Bigg(& W_1 \frac{1}{\prod_{k=1}^{d} f_{\nu_k}(x_{ik})} \int_0^1 \varphi_{\Sigma}\big(z_1(x_{i1}, s), \ldots z_n(x_{id}, s)\big) \prod_{k=1}^{d} \big(w_k(s)\big)^{-1} ds \\ & + W_2 C_G(x_{i1}, x_{i2}) x_{i1}^{-1} x_{i2}^{-1} \left[\sum_{j=1}^{2} \big(-\ln(x_{ij}) \big)^{\rho} \right]^{2(\frac{1}{\rho} - 1)} [\ln(x_{i1}) \ln(x_{i2})]^{\rho - 1} \\ & \times \left[1 + (\rho - 1) \left[\sum_{j=1}^{2} \big(-\ln(x_{ij}) \big)^{\rho} \right]^{-\frac{1}{\rho}} \right] \Bigg) \end{aligned} \tag{5.8}$$

where

$$C_G(x_{i1}, x_{i2}) = \exp\left(- \left[\sum_{j=1}^{2} (-\log(x_{ij}))^{\rho} \right]^{\frac{1}{\rho}} \right)$$

The density of the mixture copula is a linear combination of the component copulas and the corresponding likelihood function is derived in the usual way, as the product of densities.

5.1.2 Testing Procedure

Data is generated from known distributions so we can examine the accuracy of estimating various severity and copula distributions. The following procedure was carried out:

1. Generate bivariate uniform data from the copula models below. Each variable represents the severity of losses for a risk cell.
   - Gaussian copula with $\rho = 0.8$,
   - Standard t copula with $\rho = 0.8,\ \nu = 3$,
   - Gumbel copula with $\rho = 2$,
   - Mixture copula with $\rho_t = 0.8,\ \nu = 3,\ \rho_G = 2,\ W_1 = 0.5,\ W_2 = 0.5$.
2. Using each set of uniform data, generate the following dependent severity loss data assuming each risk cell has identical marginal distributions.
   - Lognormal(5,1),
   - Normal Inverse Gaussian(2,1,3,4).
3. For each dependent severity loss dataset generated, estimate the parameters of the marginal distributions using MLE.
4. Based on the fitted marginals, estimate the copula parameter using MLE.

The above procedure was carried out 20 times and the MLE reported is the mean of the MLEs across all 20 iterations. The error bars displayed correspond to the standard deviations of the MLEs across these 20 iterations.

5.1.3 Results

The Lognormal, Gaussian copula, t copula and Gumbel copula parameters were fitted using the *lognfit* and *copulafit* functions in Matlab, which are MLE procedures. The NIG and mixed copula parameters were determined by minimizing their respective negative-log-likelihood functions using Matlab's *ga* function (Genetic optimization algorithm). The weight parameter in the mixture model shown below corresponds to the weight of the t copula in the mixture. The estimated parameter values were plotted as a function of the number of data points used for estimation. Since the marginal distributions were identical, only the results for the first risk factor were presented. A table of approximated convergence is also provided where we consider parameter estimates to have converged when their error bars are within roughly 5% of their true values.

Lognormal Severities

The results of the Lognormal marginal severity distributions are shown below.

Estimation of Marginal Distribution Parameters

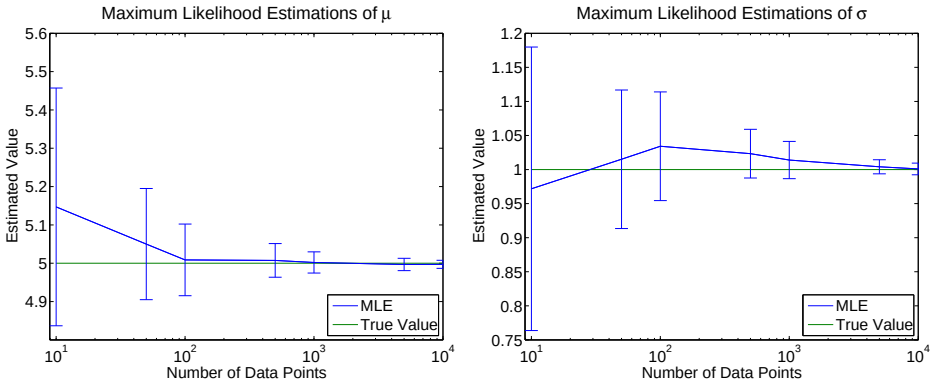


Figure 5.1: Lognormal parameter MLEs

Estimation of Copula Parameters

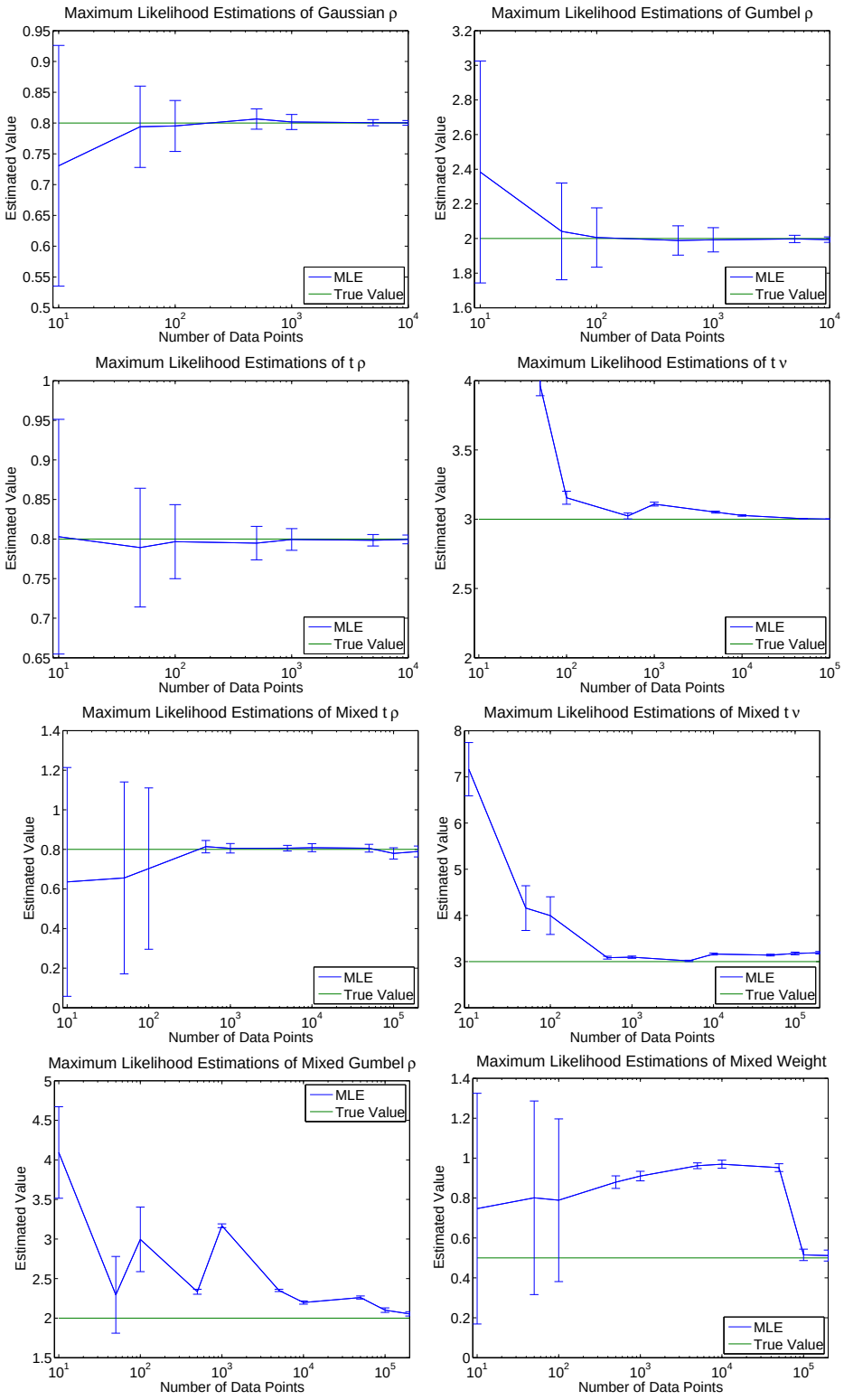


Figure 5.2: Copula parameter MLEs for Lognormal marginals

Parameter	Data Points Required for Convergence
Lognormal μ	50
Lognormal σ	1000
Gaussian copula ρ	100
t copula ρ	100
t copula ν	50,000
Gumbel copula ρ	500
Mixture copula t ρ	500
Mixture copula t ν	50,000
Mixture copula Gumbel ρ	$> 200,000$
Mixture copula t weights	$100,000$

Table 5.1: Approximate convergence table for MLEs for Lognormal severities

Convergence of the parameter estimates to their true values are seen in all cases except the mixed copula. However, by OpRisk standards, most of the parameters require a substantial amount of data points for estimation accuracy to be within 5% of their true values. Many operational risk factors have low frequencies and high severities, such as system failures and natural disasters, so data requirements of 1000 or over is unrealistic. In these cases caution should be exercised if MLE is used for model fitting. Nonetheless, high amounts of data are available for some risks, such as data entry error, so the MLE approach can still be applied well in certain situations.

The location parameter μ of the Lognormal severity distribution converged rapidly after only 50 data points, however the scale parameter σ converged after 1000 data points were used. This presents a problem in capital estimation as it is the scale parameter that drives the tail behaviour in the Lognormal distribution.

In terms of the copula model fitting, the parameters driving linear correlation Gaussian ρ, t ρ and Mixture t ρ all required significantly less data for accurate estimation than other parameters, with none of them exceeding 500 data points for convergence. The Gumbel ρ also drives linear correlation in the non-mixture model so it too took 500 data points to converge. The tail dependence related parameters t ν and mixture Gumbel ρ converged slowly, if at all, due to the limited information on tail data and the slow convergence of the marginal scale parameter σ.

Only 2 parameters converged in the mixture copula model, the linear correlation parameter corresponding to the t copula component and the weight parameter, despite point estimates of all parameters coming close to the true value for more than 50,000 data points. We can observe that the t copula influences the mixture model more than the Gumbel copula as high weights for the t copula were estimated for low amounts of data. Despite the added flexibility of the mixture model, the extra data requirements for estimation means it is infeasible to fit under the MLE approach.

Both the marginal and copula parameter estimates reinforce the idea that tail behaviour is hard to fit to without an abundance of data when using MLE. From an OpRisk point of view, this means another approach should be sought or there will be negative implications on the accuracy of capital estimation.

Normal Inverse Gaussian Severities

The results of the NIG marginal severity distribution are shown below. For our estimation, we fixed the location and scale parameters δ and μ and estimation was only performed on the tail and skewness parameters α and β. This is not unrealistic in practice as scenario analysis and expert opinions can provide reasonable estimates for the location and scale of the distribution. Since we are mainly interested in VaR, which is a tail measure, only the parameters affecting the tail of the distribution, α and β, were chosen for estimation.

Estimation of Marginal Distribution Parameters

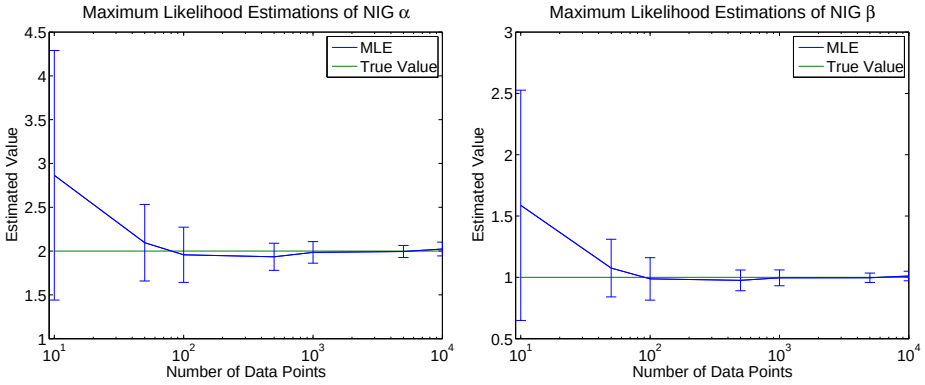


Figure 5.3: Normal Inverse Gaussian parameter MLEs

Estimation of Copula Parameters

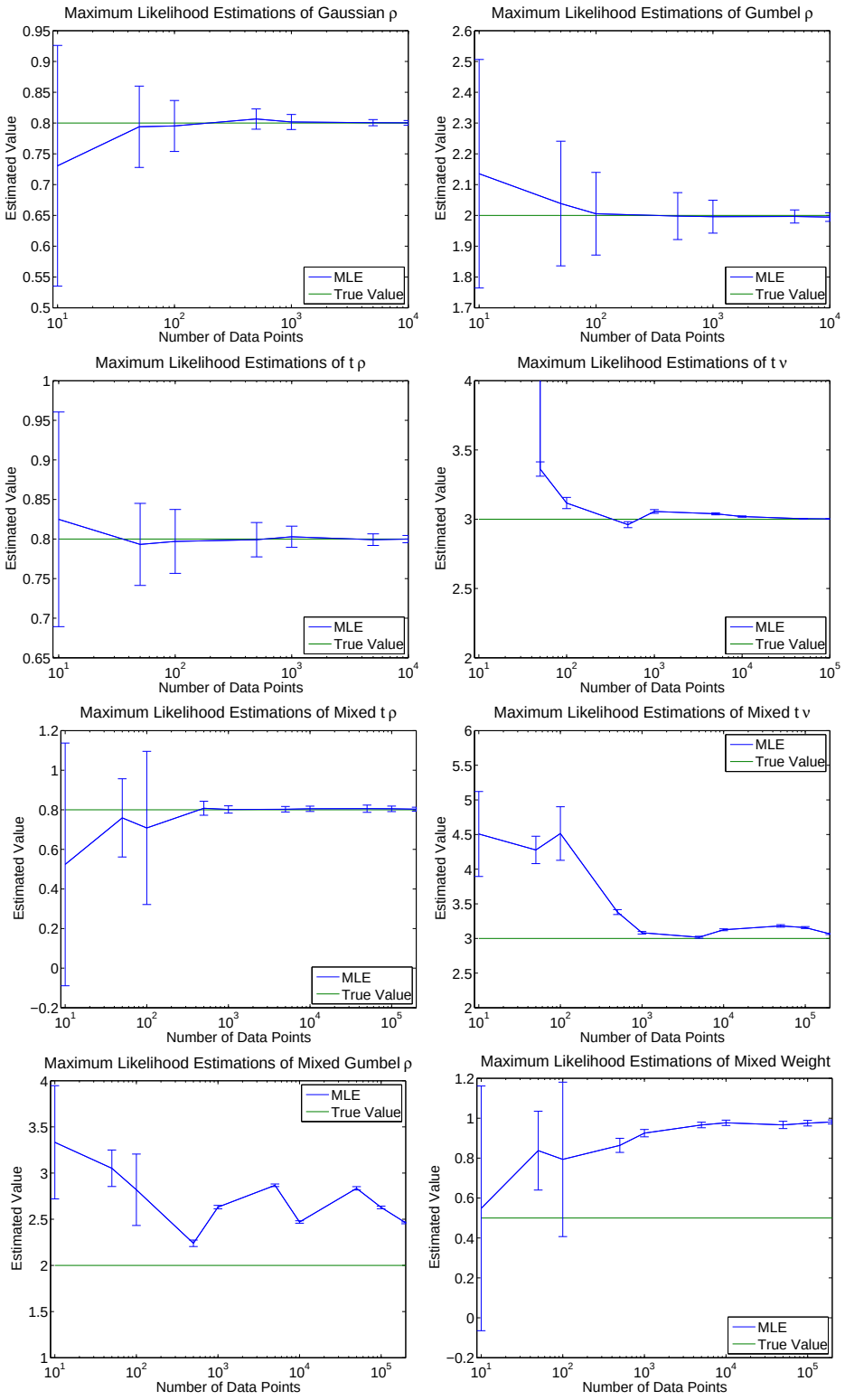


Figure 5.4: Copula parameter MLEs for Normal Inverse Gaussian marginals

Parameter	Data Points Required for Convergence
NIG α	5000
NIG β	5000
Gaussian copula ρ	100
t copula ρ	100
t copula ν	50,000
Gumbel copula ρ	500
Mixture copula t ρ	500
Mixture copula t ν	50,000
Mixture copula Gumbel ρ	$> 200,000$
Mixture copula weights	$> 200,000$

Table 5.2: Approximate convergence table for MLEs for NIG severities

The results for the NIG severity distribution are consistent with the Lognormal case. The tail and skewness parameters of the NIG distribution α and β require a large of amount of data, converging after 5000 points are used. These are tail related parameters and due to limited tail information, yields poor estimates when data is lacking. The copula parameter estimates exhibit almost the same behaviour as in the previous case, requiring the same amount of data points to be used to converge within 5% of their true values.

The Gumbel and weight parameters were again poorly estimated in the mixture copula. High estimated weights for the t copula indicate its dominance in the combination, most likely because the t copula used has high linear and tail dependence compared to the Gumbel component. This could suggest that the characteristics of the Gumbel copula are being made redundant by the t copula in this particular model.

Since we are working in an OpRisk framework, data availability is generally rare and requirements such as 1000 data points, which may seem low in other contexts such as market risk, is considered significant and usually infeasible. We can conclude from these two studies the characteristics of MLEs with respect to the number of data needed for accurate estimation. Location and linear correlation parameters are typically well estimated without plenty of data as they define the body of their corresponding distributions where most of the information will lie. However parameters such as tail, scale and degree-of-freedom which are directly related to tail heaviness of their respective distributions are poorly estimated without an abundant source of information. This arises from the fact that tail regions are not densely populated even for heavy tailed distributions. Hence using MLE parameters to estimate VaR may lead to large errors and conservative results should be reported to avoid potential insolvency.

5.2 Bayesian Estimation

Estimation of the model parameters for the Lognormal and Normal Inverse Gaussian marginal distributions were carried out using a Bayesian approach. An MCMC approach is adopted for simulating from the posterior distributions of the required parameters. The Minimum Mean Square Error (MMSE) estimators are taken as point estimates for the desired parameters. For a parameter θ, the MMSE estimator is the mean of its posterior density, given as:

$$MMSE(\theta) = E[\theta|\boldsymbol{X}] \tag{5.9}$$

where $\boldsymbol{X}$ is the observed data. The formulation of the Lognormal and NIG severity marginal models are given below.

5.2.1 Lognormal Severity Model

- $\boldsymbol{X}_t^{(j)} = \left(X_{1,t}^{(j)}, ..., X_{N_t^{(j)},t}^{(j)}\right)$ are i.i.d loss severities for the jth risk cell in year t and $\boldsymbol{X}_t = \left(\boldsymbol{X}_t^{(1)}, ..., \boldsymbol{X}_t^{(J)}\right)$.
- $(X_{i,t}^{(j)}|\mu^{(j)}, \sigma^{(j)}) \sim$ Lognormal$(\mu^{(j)}, \sigma^{(j)})$.
- $\boldsymbol{X}$ are correlated across the J risk cells with copula distribution and density $C(\cdot|\boldsymbol{\theta})$ and $c(\cdot|\boldsymbol{\theta})$ respectively.
- The parameters $\boldsymbol{\mu} = (\mu^{(1)}, ..., \mu^{(J)})$ and $\boldsymbol{\sigma} = (\sigma^{(1)}, ..., \sigma^{(J)})$ are constant over time, have independent components and are independent of each other. The prior distributions used for their estimation are:
  - $\pi(\mu^{(j)}) \sim$ Normal$(a_\mu^{(j)}, b_\mu^{(j)})$,
  - $\pi(\sigma^{(j)}) \sim$ Inverse Chi Square$(a_\sigma^{(j)}, b_\sigma^{(j)})$.
- $\boldsymbol{\theta} = (\theta_1, ..., \theta_l)$ where θ_i are copula parameters. They are independent of each other and are independent of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. Each copula model will have a different prior specification depending on its parameterization. The ones used for our case study will be defined later.

For convenience, we will define the subscript $1:T$ to mean the set of all values from year 1 to T. For example:

- $\boldsymbol{X}_{1:T} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_T)$ is the set of all severity observations from years 1 to T.

We will also define the superscript $(-j)$ to be exclusion parameters. For example:

- $\boldsymbol{\mu}^{(-j)} = (\mu^{(1)}, ..., \mu^{(j-1)}, \mu^{(j+1)}, ..., \mu^{(J)})$ is the set of μ parameters for all risk cells except the jth.

Sampling from the posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\theta}|\boldsymbol{X}_{1:T})$ is needed to determine the required MMSE estimators. However, we do not know the normalization constant so we utilize an univariate Slice sampler within a Gibbs sampling framework as the sampling mechanism. This requires knowledge of the full conditional posteriors up to the normalization constant. We perform estimation using a 2-step approach akin to the IFM approach, as well as a full joint estimation.

2-step Estimation
This approach can be seen as a parallel of the IFM method in a Bayesian context. First we sample from $\pi(\mu^{(j)}, \sigma^{(j)}|\boldsymbol{X}_{1:T}^{(j)})$, for all j, without considering dependence to obtain estimates of our marginal parameters, then we sample from $\pi(\boldsymbol{\theta}|\boldsymbol{X}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ to obtain the copula parameter. The required conditional posterior densities are given below:

$$\begin{aligned}
\pi(\mu^{(j)}|\boldsymbol{\mu}^{(-j)}, \boldsymbol{\sigma}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\mu}^{(-j)}, \mu^{(j)}, \boldsymbol{\sigma})\ \pi(\mu^{(j)}) \\
&\propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\mu^{(j)}, \sigma^{(j)})\ \pi(\mu^{(j)}) \qquad (5.10) \\
\pi(\sigma^{(j)}|\boldsymbol{\sigma}^{(-j)}, \boldsymbol{\mu}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\sigma}^{(-j)}, \sigma^{(j)}, \boldsymbol{\mu})\ \pi(\sigma^{(j)}) \\
&\propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\mu^{(j)}, \sigma^{(j)})\ \pi(\sigma^{(j)}) \qquad (5.11) \\
\pi(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\theta})\ \pi(\boldsymbol{\theta}) \\
&\propto \prod_{t=1}^{T} \left(c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right) \prod_{i=1}^{J} \pi(X_t^{(i)}|\mu^{(i)}, \sigma^{(i)}) \right) \\
&\quad \times \pi(\boldsymbol{\theta}) \\
&\propto \prod_{t=1}^{T} c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right) \pi(\boldsymbol{\theta}) \qquad (5.12)
\end{aligned}$$

Joint Estimation

We sample from $\pi(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\theta}|\boldsymbol{X}_{1:T})$ to obtain estimates for our marginal and copula parameters jointly. Although this approach seems preferable due to the decreased risk of bias due to the potential misspecification of the marginals in the first model, more computational effort is required to evaluate the joint likelihood of the data. The required conditional posteriors are given below:

$$\begin{aligned}\pi(\mu^{(j)}|\boldsymbol{\mu}^{(-j)}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\mu}^{(-j)}, \mu^{(j)}, \boldsymbol{\sigma}, \boldsymbol{\theta})\ \pi(\mu^{(j)}) \\ &\propto \prod_{t=1}^{T}\left(c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right)\prod_{i=1}^{J}\pi(X_t^{(i)}|\mu^{(i)}, \sigma^{(i)})\right) \\ &\quad \times \pi(\mu^{(j)}) \\ &\propto \prod_{t=1}^{T} c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right)\ \pi(X_t^{(j)}|\mu^{(j)}, \sigma^{(j)})\ \pi(\mu^{(j)})\end{aligned} \tag{5.13}$$

$$\begin{aligned}\pi(\sigma^{(j)}|\boldsymbol{\sigma}^{(-j)}, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\sigma}^{(-j)}, \sigma^{(j)}, \boldsymbol{\mu}, \boldsymbol{\theta})\ \pi(\sigma^{(j)}) \\ &\propto \prod_{t=1}^{T} c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right)\pi(X_t^{(j)}|\mu^{(j)}, \sigma^{(j)})\ \pi(\sigma^{(j)})\end{aligned} \tag{5.14}$$

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\theta})\ \pi(\boldsymbol{\theta}) \\ &\propto \prod_{t=1}^{T} c\left(F_1(X_t^{(1)}|\mu^{(1)}, \sigma^{(1)}), ..., F_J(X_t^{(J)}|\mu^{(J)}, \sigma^{(J)})|\boldsymbol{\theta}\right)\pi(\boldsymbol{\theta})\end{aligned} \tag{5.15}$$

5.2.2 Normal Inverse Gaussian Severity Model

- $\boldsymbol{X}_t^{(j)} = \left(X_{1,t}^{(j)}, ..., X_{N_t^{(j)},t}^{(j)}\right)$ are i.i.d loss severities for the jth risk cell in year t and $\boldsymbol{X}_t = \left(\boldsymbol{X}_t^{(1)}, ..., \boldsymbol{X}_t^{(J)}\right)$.
- $(X_{i,t}^{(j)}|\alpha^{(j)}, \beta^{(j)}, \delta^{(j)}, \mu^{(j)})$ ~Normal Inverse Gaussian$(\alpha^{(j)}, \beta^{(j)}, \delta^{(j)}, \mu^{(j)})$.
- $\boldsymbol{X}$ are correlated across the J risk cells with copula distribution and density $C(\cdot|\boldsymbol{\theta})$ and $c(\cdot|\boldsymbol{\theta})$ respectively.
- The parameters $\boldsymbol{\alpha} = (\alpha^{(1)}, ..., \alpha^{(J)})$, $\boldsymbol{\beta} = (\beta^{(1)}, ..., \beta^{(J)})$, $\boldsymbol{\delta} = (\delta^{(1)}, ..., \delta^{(J)})$ and $\boldsymbol{\mu} = (\mu^{(1)}, ..., \mu^{(J)})$ are constant over time and have independent components. $\boldsymbol{\beta}$ is dependent on $\boldsymbol{\alpha}$ via the relationship $0 < |\beta^{(j)}| \leq \alpha^{(j)}$ but all the other parameters are independent of each other. The prior distributions used for estimation are:
  - $\pi(\alpha^{(j)})$ ~Inverse-Chi-Square$(a_\alpha^{(j)}, b_\alpha^{(j)})$,
  - $\pi(\beta^{(j)}|\alpha_t^{(j)})$ ~Uniform$(a_\beta^{(j)}, b_\beta^{(j)})$,
  - $\pi(\delta^{(j)})$ ~Inverse-Chi-Square$(a_\delta^{(j)}, b_\delta^{(j)})$.
  - $\pi(\mu^{(j)})$ ~Normal$(a_\mu^{(j)}, b_\mu^{(j)})$
- $\boldsymbol{\theta} = (\theta_1, ..., \theta_l)$ where θ_i are independent copula parameters whose priors will differ according to the copula model which will be defined later.

Similar to the Lognormal model, we want to sample from the posterior $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\theta}|\boldsymbol{X}_{1:T})$ using the univariate Slice sampler within a Gibbs framework. Only the 2-step estimation approach was considered

for this model. The conditional posteriors required are given as:

$$\pi(\alpha^{(j)}|\boldsymbol{\alpha}^{(-j)},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\alpha^{(j)},\beta^{(j)},\delta^{(j)},\mu^{(j)})\ \pi(\alpha^{(j)}) \tag{5.16}$$

$$\pi(\beta^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta}^{(-j)},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\alpha^{(j)},\beta^{(j)},\delta^{(j)},\mu^{(j)})\pi(\beta^{(j)}|\alpha^{(j)}) \tag{5.17}$$

$$\pi(\delta^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta}^{(-j)},\boldsymbol{\mu},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\alpha^{(j)},\beta^{(j)},\delta^{(j)},\mu^{(j)})\ \pi(\delta^{(j)}) \tag{5.18}$$

$$\pi(\mu^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu}^{(-j)},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}^{(j)}|\alpha^{(j)},\beta^{(j)},\delta^{(j)},\mu^{(j)})\ \pi(\mu^{(j)}) \tag{5.19}$$

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{X}_{1:T}) &\propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta})\ \pi(\boldsymbol{\theta}) \\ &\propto \prod_{t=1}^{T} c\left(F_1(X_t^{(1)}|\alpha^{(1)},\beta^{(1)},\delta^{(1)},\mu^{(1)}),...,F_J(X_t^{(J)}|\alpha^{(J)},\beta^{(J)},\delta^{(J)},\mu^{(J)})|\boldsymbol{\theta}\right) \\ &\quad \times \pi(\boldsymbol{\theta})\end{aligned} \tag{5.20}$$

We also present the conditional posteriors required for a joint estimation. Simplifications can be made using the same approach as the Lognormal case. This model is not explored in this thesis and is subject to future work.

$$\pi(\alpha^{(j)}|\boldsymbol{\alpha}^{(-j)},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha}^{(-j)},\alpha^{(j)},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta})\ \pi(\alpha^{(j)}) \tag{5.21}$$

$$\pi(\beta^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta}^{(-j)},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha},\boldsymbol{\beta}^{(-j)},\beta^{(j)},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta})\pi(\beta^{(j)}|\alpha^{(j)}) \tag{5.22}$$

$$\pi(\delta^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta}^{(-j)},\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta}^{(-j)},\delta^{(j)},\boldsymbol{\mu},\boldsymbol{\theta})\ \pi(\delta^{(j)}) \tag{5.23}$$

$$\pi(\mu^{(j)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu}^{(-j)},\boldsymbol{\theta},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu}^{(-j)},\mu^{(j)},\boldsymbol{\theta})\ \pi(\mu^{(j)}) \tag{5.24}$$

$$\pi(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{X}_{1:T}) \propto \pi(\boldsymbol{X}_{1:T}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\theta})\ \pi(\boldsymbol{\theta}) \tag{5.25}$$

5.2.3 Testing Procedure

The following procedure was carried to evaluate the performance of estimation:

1. Generate bivariate uniform data from the univariate copula models below. Each variable represents the severity of losses for a risk factor.
   - Gaussian copula with $\rho = 0.8$,
   - Gumbel copula with $\rho = 1.5$.
2. Using each set of uniform data, generate dependent severity loss data assuming each risk factor has identical marginal distributions.
   - Lognormal(5,1),
   - Normal Inverse Gaussian(2,1,3,4).
3. For each dependent severity loss dataset generated, estimate the parameters of the marginal distributions and copulas using the Bayesian models set up above with a Slice Sampler.

The above procedure was carried out 20 times and the estimator reported is the average of the MMSE estimators across all 20 iterations. The error bars calculated correspond to the standard deviation of the MMSEs across the 20 iterations. Here we only explore estimation with 1 parameter copula models, however extensions to multi-parameter copula models are straightforward if the parameters are independent, for example in the case of a bivariate t copula where ρ and ν are independent the conditional posterior can easily be derived. However estimation involving copulas with dependent parameters are subject to further work, such as the multivariate Gaussian copula where the correlation parameters must form a positive definite matrix.

We do not explore prior elicitation in this project and instead assume informative priors are given. The prior distributions used for the Lognormal severity model were:

- $\mu^{(1)} \sim$ Normal(3,1),
- $\mu^{(2)} \sim$ Normal(7,1),
- $\sigma^{(1)} \sim$ Inverse-Chi-Square$(8, \sqrt{4/5})$,
- $\sigma^{(2)} \sim$ Inverse-Chi-Square$(12, \sqrt{4/5})$.

The priors for the Normal Inverse Gaussian model were:

- $\alpha^{(1)} \sim$ Inverse-Chi-Square$(4, \sqrt{4/3})$,
- $\alpha^{(2)} \sim$ Inverse-Chi-Square$(8, \sqrt{4/3})$,
- $\beta^{(1)} \sim$ Uniform$(-\alpha, \alpha)$,
- $\beta^{(2)} \sim$ Uniform$(-\alpha, \alpha)$,
- $\delta^{(1)} \sim$ Inverse-Chi-Square$(4, \sqrt{2})$,
- $\delta^{(2)} \sim$ Inverse-Chi-Square$(8, \sqrt{2})$,
- $\mu^{(1)} \sim$ Normal(3,1), $\mu^{(2)} \sim$ Normal(7,1).

Priors specified for the copula models were:

- $\rho_{Gaussian} \sim$ Uniform(0,1),
- $\rho_{Gumbel} \sim 1 + \gamma$ where $\gamma \sim$ Gamma(2,1).

5.2.4 Results

The univariate slice sampler used is a modified version of Matlab's *slicesample* function. The Bayesian estimates obtained are the Minimum Mean Squared Errors (MMSE) calculated by averaging across the 20 datasets. These are plotted as a function of the number of data points used for estimation alongside the corresponding MLEs to compare their performance. Diagnostic plots for the convergence of the Markov chain for one case of estimation using 500 data points are also included.

Lognormal Severities

10,000 samples were generated from the Markov chain for the 2-step approach and MMSEs of the posterior distributions were taken as the mean of the last 20% of samples. Due to the quick convergence of the chain observed for the 2-step approach, only 5,000 iterations were run for the joint estimation and the last 20% of samples were used for calculating MMSEs.

2-Step Estimation of Marginal Distribution Parameters

Estimates of the Lognormal severity parameters are shown here. Only the results for the first risk factor are shown since identical behaviour is observed for the second risk (included in the Appendix).

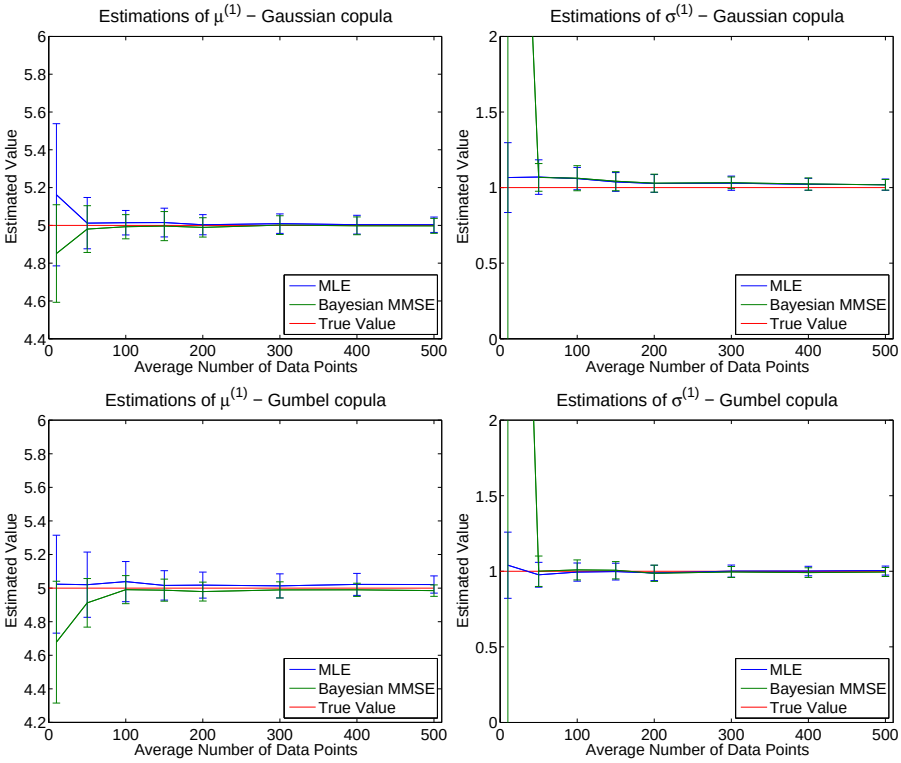


Figure 5.5: Lognormal parameter estimates using the 2-step Bayesian approach

2-Step Estimation of Copula Parameters

Copula parameter estimates for both the Gumbel and Gaussian copulas are shown below.

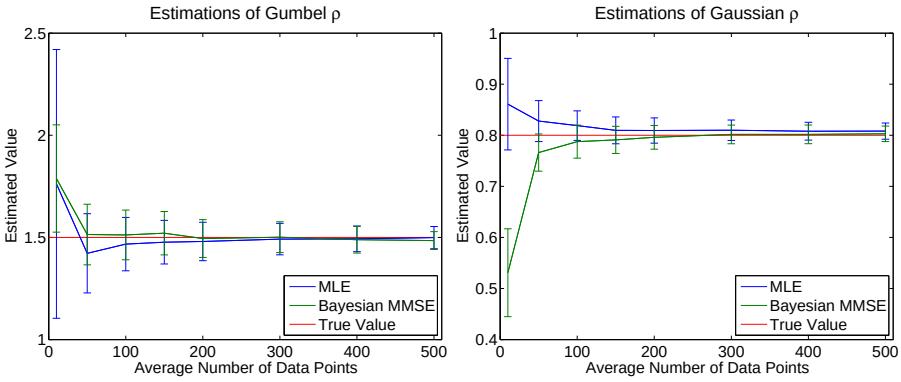
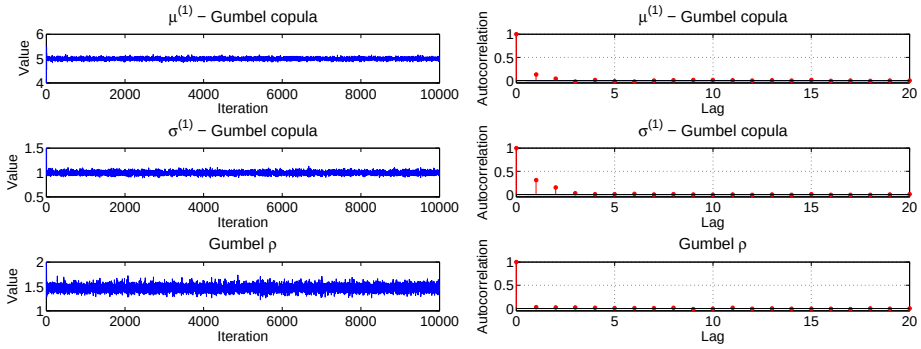


Figure 5.6: Copula parameter estimates using the 2-step Bayesian approach for Lognormal marginals

Trace and Autocorrelation Plots

MCMC diagnostic plots are shown for the first risk factor in the Gumbel copula case. The graph on the left is the trace plot of the sampled marginal and copula parameters over 10,000 iterations of the Markov chain and the graph on the right the is autocorrelation function for the same parameters. The results for

the second risk factor and the Gaussian copula case are similar and are included in the Appendix. The trace plots and autocorrelations shown are for the estimation of one dataset with 500 points.



MCMC diagnostic plots The Bayesian estimates exhibit similar performance to the MLEs but often with smaller errors. This means that the MLE approach is more preferable for a simple model like Lognormal without many parameters, as it requires less computational time but achieves the accuracy of the Bayesian approach with the same amount of data. Furthermore, it does not require the specification of a prior structure which is usually challenging in OpRisk due to the lack of data.

We observe that the Lognormal μ converged after 50 data points in for both types of dependence structures, while the Lognormal σ converged after 300 data points for Gumbel and did not converge after 500 points for Gaussian. This again demonstrates the difficulties of determining tail related parameters as with the MLE approach. The dependence parameters for Gaussian and Gumbel copulas converged after 100 and 150 points respectively which is within feasible data constraints for some OpRisks. However, it should be noted that different dependence models require different amounts of data to estimate correctly even if they have the same amount of parameters as shown here. In this case, since the Gumbel copula accounts for tail dependence, more data is needed for an accurate estimate than Gaussian.

Although the data amount for this model seems reasonable for some OpRisks, we note that in higher dimensional models, accuracy is expected to deteriorate for the same amount of information provided. An example of this can be seen in the next section where the estimates for the NIG severity model, which is a 4 parameter marginal distribution, are presented. Increasing the number of risk factors of the model is also expected to adversely affect its performance.

The autocorrelation plots show the Markov chains are rapidly exploring their corresponding posteriors and the chain is mixing well. The trace plots show that the Markov chains for all parameters have converged to their stationary distributions as the samples are all oscillating around a common region. The convergence in this case was also very rapid as the oscillation behaviour commenced after only a small number of iterations. Though this means we could have used substantially less time for the estimation, a priori we do not the behaviour of the Markov chain so a large amount of iterations is recommended. However subsequent estimations can be run with less iterations when we re-estimate the parameters of risk factors given new data using this model.

Joint Estimation of Marginal Distribution Parameters

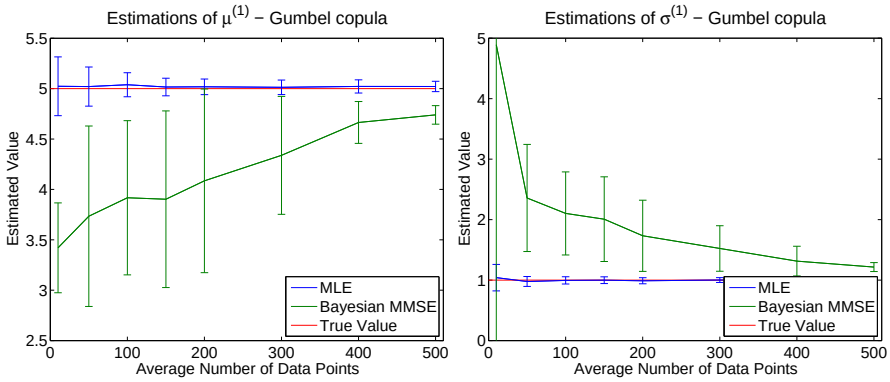


Figure 5.7: Lognormal parameter estimates using the joint Bayesian approach

Joint Estimation of Copula Parameters

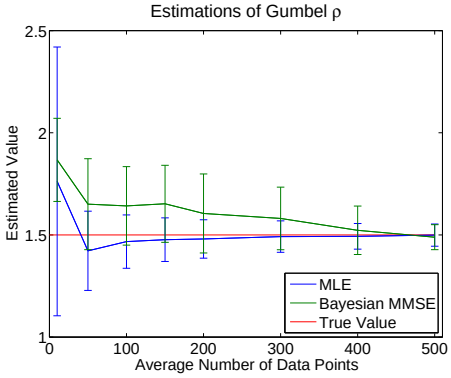


Figure 5.8: Copula parameter estimates using the joint Bayesian approach for Lognormal marginals

The joint estimation produced results that are less consistent than the 2-step procedure as seen from the larger error bars. Furthermore, although convergence to the true value is clearly seen with respect to increasing data amounts, the accuracy of the results are worse for the same amount of data points. No parameters apart from the Gumbel ρ reached within 5% of their true values for 500 data points or less. We can conclude from these results that the 2 step approach is more preferable in this context as it is more accurate and less computationally intensive. However, theoretically a joint estimation should be more accurate as it takes all of the information into account at once and ways to improve the model is subject to future work.

Normal Inverse Gaussian Severities

10,000 iterations were run for the Markov chain and the last 20% of samples were used to determine MMSEs.

Estimation of Marginal Distribution Parameters

The marginal parameter estimates for both risk factors in the Gumbel copula case are displayed below. Results for the Gaussian copula case are similar and can be found in the Appendix).

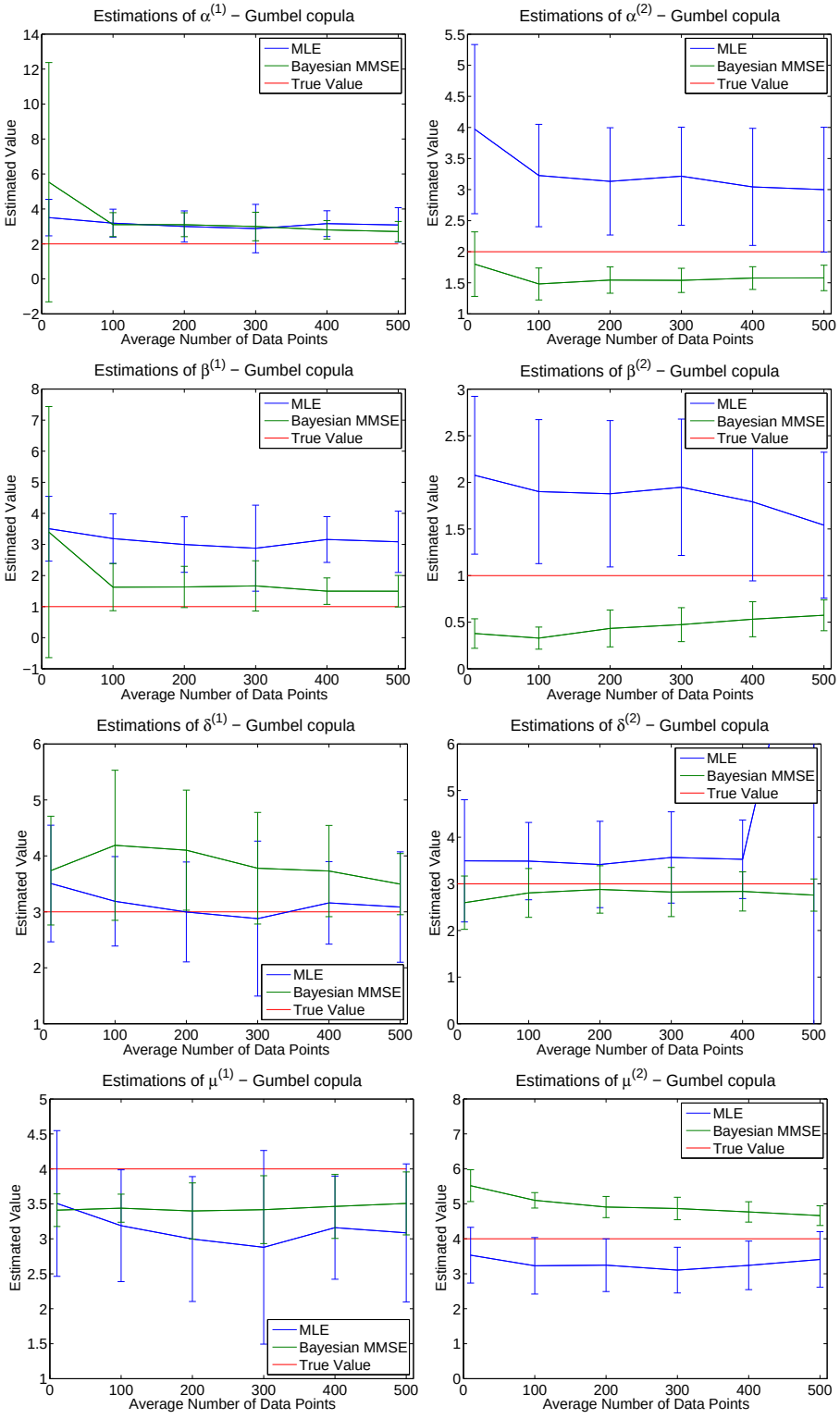


Figure 5.9: NIG parameter estimates using the 2-step Bayesian approach

Estimation of Copula Parameters

The estimates of the both the Gumbel and Gaussian dependence parameter are shown below.

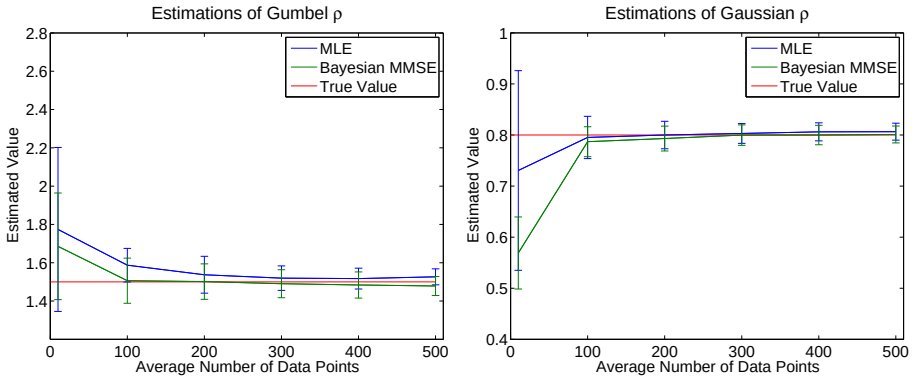


Figure 5.10: Copula parameter estimates using the 2-step Bayesian approach with NIG marginals

Trace Plots

Trace plots of the sampled estimates for the α and β parameters of the NIG distribution are shown for the 2 risk factors in both the Gumbel and Gaussian cases, along with the trace plot of the dependence parameter. The plots are for the estimation of one particular set of data with 500 points. Differences in behaviour can be seen between the different risk factors. Trace plots for μ and δ can be found in the Appendix and exhibit the same behaviour as the other parameters for their corresponding risk factor.

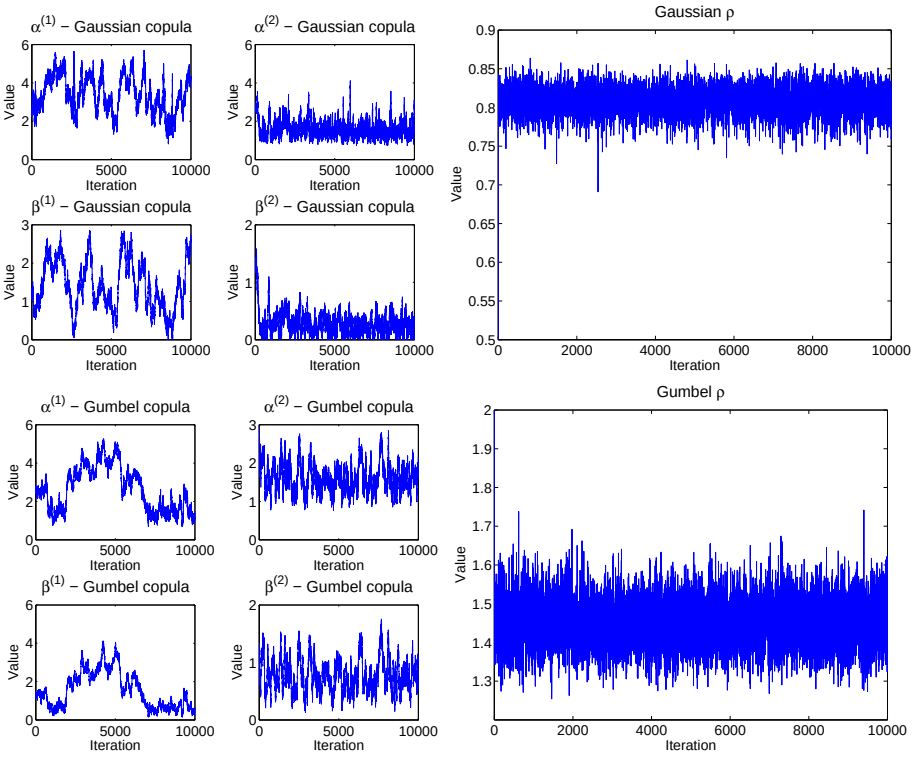


Figure 5.11: Trace plots for NIG model parameter estimates

Autocorrelation Plots

The autocorrelation functions of the sampled values of α and β are shown for both risk factors. The autocorrelation of the sampled dependence parameters are also shown.

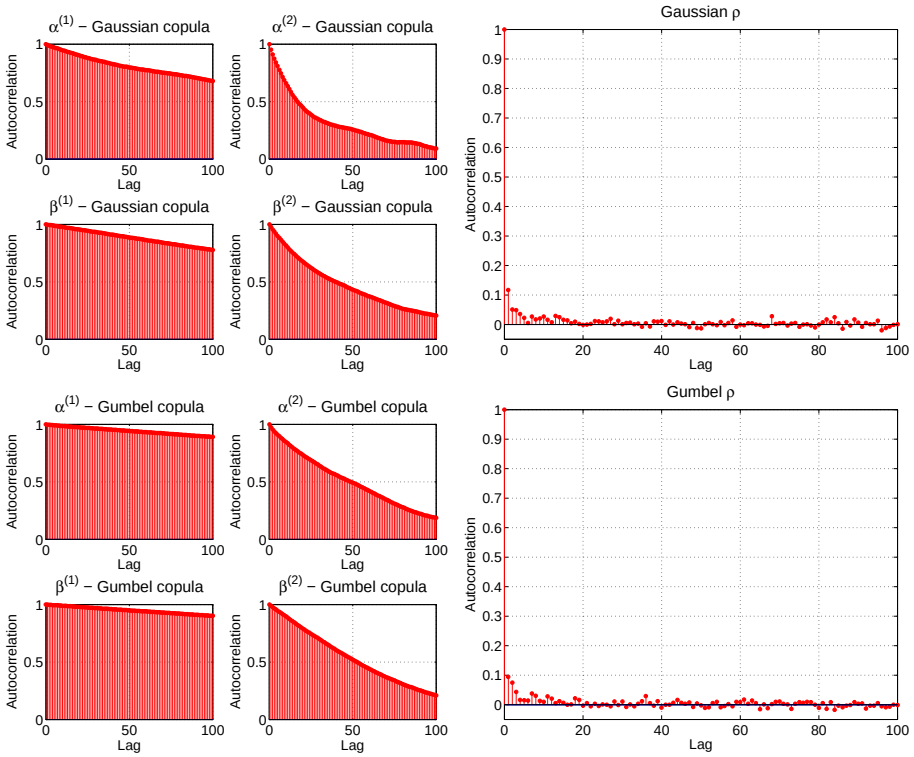


Figure 5.12: Autocorrelation plots for NIG model parameters estimates

The issue of high dimensionality has manifested here as we can see the slowness in convergence of the parameter estimation to their true values with respect to data size compared to the Lognormal severities. Furthermore, unlike in the case of the Lognormal severity model, the Bayesian estimates are clearly more accurate than the corresponding MLEs. However computational time is also much greater creating a situation where the risk practitioner must weigh the benefits and costs of accuracy versus time.

The tail heaviness and skewness parameters α and β converged slower than the scale and location parameters δ and μ as expected since they are related to the tail of the distribution. The dependence parameters for both the Gaussian and Gumbel copulas have shown good accuracy for a reasonable amount of data, converging after 100 and 300 points respectively, which is attainable for some OpRisks.

From the MCMC diagnostic plots, we can see the convergence rate of the Markov chain for the NIG model is noticeably slower than the Lognormal severities. This is direct effect of the increase in model dimension and leads to compounded computational times for model estimation under this approach. Furthermore, the convergence rate of the Markov chain also differs significantly across risk factors as well as across data sets from different dependence models, albeit the latter is not as noticeable. This suggests that the mean of the prior distributions can affect MCMC convergence rate drastically.

The trace plots and autocorrelations show that the chain for the parameters of the first risk factor requires many more iterations to reach their corresponding stationary distributions. In the case of the dataset generated using a Gaussian copula, the chain does not reach its stationary distribution after 10,000 iterations. Methods such as sub sampling, as described in (**?**), can be used to increase the convergence rate and is subject to future work in this context. However despite the difference in MCMC convergence, the corresponding MMSE estimator's convergence towards the true model parameter is similar across risk factors. From an OpRisk modelling perspective, this means that expert opinions and scenario analysis, which are used to form the prior, can help speed up the estimation time but does not necessarily improve the accuracy of the estimators.

5.2.5 Prior Sensitivities

In this section we look at the effects of vague priors on the accuracy of parameter estimates. We have already seen in the previous section that the mean of the priors can affect convergence rates of the Markov chain but did not affect the overall accuracy of the estimate.

The Lognormal severity model with Gumbel copula was used as the basis of this study and the 2-step estimation procedure was employed. We use prior distributions with variances that are 10 times the size of those in the previous section. The prior specifications are given below:

- $\mu^{(1)} \sim$ Normal(3,10),
- $\mu^{(2)} \sim$ Normal(7,10),
- $\sigma^{(1)} \sim$ Inverse-Chi-Square($8,10^{1/4}\sqrt{4/5}$),
- $\sigma^{(2)} \sim$ Inverse-Chi-Square($12,10^{1/4}\sqrt{4/5}$),
- $\rho_{Gumbel} \sim 1+\gamma$ where $\gamma \sim$ Gamma(1,10).

Results

The estimates of the parameters using the above vague priors are plotted as a function of the number of data points used. Corresponding estimates using informative priors are also shown for comparison.

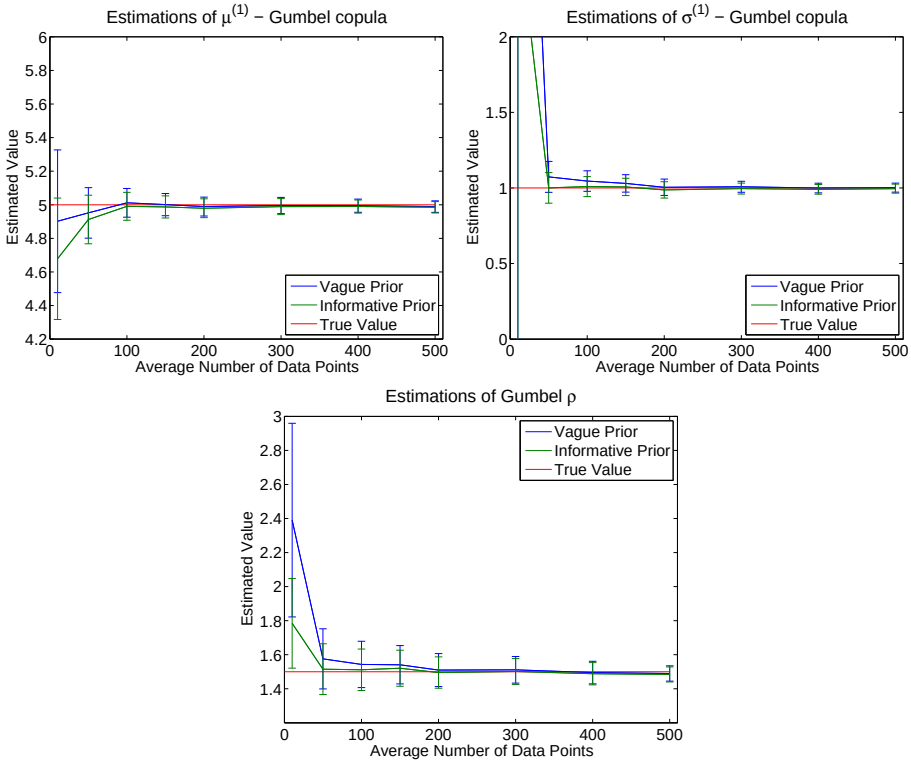


Figure 5.13: Lognormal and copula parameter estimates using vague priors

The difference between estimates are minuscule for fits obtained using vague and informative priors in all parameters. Although informative priors produced slightly more consistent estimates around the true value, the convergence rate of the two models are almost the same. This suggests that the accuracy of our model is not very sensitive to prior specification, as long as they are reasonable, and the main driver is the data itself. Although this means there is less risk of prior related errors, it also means that expert opinions cannot always play a large role in reducing data requirements.

Chapter 6

Model Selection

To determine whether the chosen model and estimated parameters appropriately model the dependence structure of a given data set, goodness-of-fit (GoF) tests are required. In the context of copula models, GoF tests are essentially akin to tests for multivariate distributions. However, the marginal distributions are often unknown which adds a layer of complication to the procedure. Currently, there is a relatively small amount of literature on GoF tests for copula models. Existing tests are outlined in (**?**) can be categorized into three broad areas:

1. Tests developed specifically for testing particular copula structures, e.g. the Gaussian copula.
2. Tests that can be applied for any family of copulas but require arbitrary parameters, such as kernels and weights, to be chosen.
3. Tests that can be used for all copula families and require no choice for their use, also known as "blanket tests".

Some currently available blanket tests as seen in (**?**) are described below.

6.1 Goodness-of-Fit Tests

General Setup

Let F_j be the marginal distribution function of the jth variable in a multivariate dataset. Its empirical distribution is then given as:

$$\hat{F}_j(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_{ij} \leq t) \tag{6.1}$$

We define $\boldsymbol{U_1}, ..., \boldsymbol{U_n}$ as a set of pseudo-observations with $\boldsymbol{U_i} = (U_{i1}, ..., U_{id})$ and $U_{ij} = \dfrac{n\hat{F}_j(X_{ij})}{n+1}$, where the factor $\dfrac{n}{n+1}$ is used to prevent potential problems on the boundary of $[0,1]^d$.

The general hypothesis that will be tested in each of the following procedures will be:

$$H_0 \ : \ C \in C_0$$
$$H_1 \ : \ C \notin C_0$$

where C is the underlying multivariate distribution of the dataset with parameter θ and C_0 is the specified parametric family of copulas. This hypothesis tests whether the population multivariate distribution is well represented by our specified copula model.

Empirical copula based tests

The empirical copula function is given by:

$$C_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, ..., U_{id} \leq u_d) \tag{6.2}$$

where $\boldsymbol{u} = (u_1, ..., u_d) \in [0,1]^d$ and $\boldsymbol{U_i} = (U_{i1}, ..., U_{id})$ are pseudo-observations. Since C_n is completely non-parametric, it is a viable benchmark for testing our desired hypothesis. A natural way to test the goodness-of-fit in this context is to examine a distance measure between C_n and an estimate of C under H_0, C_{θ_0}.

Let the empirical process $\mathbb{C}_n = \sqrt(n)(C_n - C_{\theta_n})$. Then we can use the following Cramer-von Mises (CvM) and Kolmorogov-Smirnov (KS) statistics for inference. These are given respectively as:

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(\boldsymbol{u})^2 dC_n(\boldsymbol{u}) \tag{6.3}$$

$$T_n = sup_{\boldsymbol{u}\in[0,1]^d} |\mathbb{C}_n(\boldsymbol{u})| \tag{6.4}$$

Sufficiently large values of these test statistics will result in a rejection of H_0. Corresponding p-values can be computed from their limiting distributions, however this will be dependent on the asymptotic behaviour of the process $\mathbb{C}_n$. In practice, approximate p-values can only be obtained by using specially adapted Monte Carlo methods as there are no standard limiting distributions for S_n and T_n. Tests based on these statistics have been shown to be consistent.

Kendall's transformation based tests

Consider the mapping:

$$\boldsymbol{X} \mapsto V = H(\boldsymbol{X}) = C(\mathcal{U}_1, ..., \mathcal{U}_1)$$

where the multivariate distribution of $\boldsymbol{\mathcal{U}} = (\mathcal{U}_1, ..., \mathcal{U}_d)$ is C and $\mathcal{U}_i = F_i(X_i)$. This is known as the Kendall's transform. Let K be the univariate distribution of V and $V_i = C_i(\boldsymbol{U}_i)$, then a consistent non-parametric estimator for K is given by:

$$K_n(v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(V_i \leq v),\ v \in [0,1] \tag{6.5}$$

Let the estimate of K under H_0 be K_θ, then we can test the hypothesis:

$$H_0\text{"} : K \in K_\theta$$

using an appropriate distance measure. It is important to note that in general, $H_0 \subset H_0$", so failure to reject H_0" does not necessarily mean the acceptance of H_0. An example of where these two hypotheses are equivalent is in the context of bivariate Archimedean copulas.

One process which can be used to test H_0" is the empirical process

$$\mathbb{K}_n = \sqrt{n}(K_n - K_{\theta_n})$$

The CvM and KS statistics based on this process is given respectively by:

$$S_n^{(K)} = \int_{[0,1]^d} \mathbb{K}_n(\boldsymbol{u})^2 dK_n(\boldsymbol{u}) \tag{6.6}$$

$$T_n^{(K)} = sup_{\boldsymbol{u}\in[0,1]^d} |\mathbb{K}_n(\boldsymbol{u})| \tag{6.7}$$

Sufficiently large values of these test statistics will result in a rejection of H_0". P-values can be derived

from the limiting distributions of these statistics but are dependent on the asymptotic behaviour of $\mathbb{K}_n$ and hence must be approximated via simulation. This approach is not a consistent test for H_0 as H_0 is not equivalent to H_0" in general.

Rosenblatt's transformation based tests

The Rosenblatt probability integral transform (PIT) method (**?**) converts dependent data into a new set of data that are independent and uniform on $[0, 1]$ assuming H_0 holds. This method can be seen as an inverse of simulation. Tests for independent uniformity can then be carried out to complete the GoF test.

Let $\boldsymbol{X}_i = (X_{i1}, ..., X_{id}),\ i = 1, ..., n$ be random vectors with marginal distributions $F(x_j)$ and conditional distributions $F(x_j|x_1, ..., x_{d-1})$, for $j = 1, ..., d$. Define the Rosenblatt transformation of $\boldsymbol{X}_i$ to be $\boldsymbol{T}_i = (T_{i1}, ..., T_{id})$, where:

$$\begin{aligned} T_{i1} &= F(x_{i1}) \\ T_{i2} &= F(x_{i2}|x_{i1}) \\ &\vdots \\ T_{id} &= F(x_{id}|x_{i1}, ..., x_{i,d-1}) \end{aligned} \tag{6.8}$$

For a d-dimensional copula model C we have, as seen in (**?**):

$$C_i(u_i|u_1, ..., u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, ..., u_i)}{\partial u_1 ... \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C_{i-1}(u_1, ..., u_{i-1})}{\partial u_1 ... \partial u_{i-1}} \tag{6.9}$$

where

$$\begin{aligned} C_i(u_1, ..., u_i) &= C(u_1, ..., u_i, 1, ..., 1),\ i = 2, ..., d-1, \\ C_1(u_1) &= u_1, \\ C_d(u_1, ..., u_d) &= C(u_1, ..., u_d). \end{aligned}$$

$\boldsymbol{T}_1, ..., \boldsymbol{T}_n$ are independent and uniformly distributed on $[0, 1]^d$ if H_0 is true. This can be verified by calculating $\chi_i = \sum_{j=1}^{d} (\phi^{-1}(T_{ij})^2)$ and testing whether the observed sample $\chi_1, ..., \chi_n$ comes from a chi-square distribution. Some common procedures are the Anderson-Darling (AD) test and Pearson's Chi-Square test. The empirical distribution of $\chi_1, ..., \chi_n$ is given by:

$$G_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\chi_i \leq t) \tag{6.10}$$

Then the AD statistics is given by:

$$A_n = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \Big(\log\big(G(\chi_{(i)})\big) + \log\big(1 - G(\chi_{(n+1-i)})\big) \Big) \tag{6.11}$$

where $\chi_{(1)} \leq \cdots \leq \chi_{(n)}$ are the order statistics of χ_i. A large value of these statistics will lead to the rejection of H_0. The corresponding p-value can be calculated based on the limiting distribution of the standard AD statistic.

The Pearson's Chi-Square test statistic is:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6.12}$$

where

- X^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution,
- O_i = observed frequency in the sample data,
- E_i = expected frequency under proposed distribution, in this case the χ^2 with 2 degrees-of-freedom,
- n = the number of bins used for partitioning the data.

The test statistic X^2 has an asymptotic χ^2 distribution with $n - p$ degrees-of-freedom, where p is the number of parameters in the hypothesized model plus 1. The corresponding p-value can be deduced to assess the goodness-of-fit.

Bayesian Selection

Model comparisons can be made in the Bayesian context by using a quantity known as the Bayes factor, see (**?**). Given data X, consider two hypothetical models H_0 and H_1 which we wish to fit the data to, parameterized by θ_0 and θ_1 respectively. The Bayes factor is the ratio of the likelihoods of the data under each model.

$$\begin{aligned} B_{10} &= \frac{\pi(X|H_1)}{\pi(X|H_0)} \\ &= \frac{\int \pi(\theta_0|H_0)\pi(X|\theta_0, H_0)\, d\theta_0}{\int \pi(\theta_1|H_1)\pi(X|\theta_1, H_1)\, d\theta_1} \end{aligned} \tag{6.13}$$

Bayes factor can be interpreted as evidence provided by the data in favour of a particular model over the other. One approach in making decisions regarding model selection using Bayes factor can be summarized in the follow table:

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 1/2	1 to 3.2	Almost none
1/2 to 1	3.2 to 10	Somewhat
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Table 6.1: Bayes Factor Criteria

6.2 Case Study: Probability Integral Transformation

In this section we examine how well the PIT based goodness-of-fit test performs in rejecting misspecified copula models. Dependent uniform samples were generated from known copulas and are fit to other copula models using MLE. The PIT goodness-of-fit test was then carried out to determine if the fitted model should be rejected.

Testing Procedure

1. Samples are generated from the following copula models:
   - Gaussian copula - $\rho = 0.5$,
   - Standard t copula - $\rho = 0.5,\ \nu = 3$,
   - Gumbel copula - $\rho = 2$.
2. Each sample was fit to all 3 copula models above.
3. The probability integral transform method was used to determine whether each fitted model should be rejected as the appropriate dependence structure. The Pearson's chi-square test with 10 bins was used to test whether the transformed samples are chi-square distributed.

4. Data was generated and fitted for various sample sizes 500 times and the percentage of rejection was recorded.

True Copula	**Fitted Copula**	**n = 50**	**100**	**500**	**1000**	**2500**	**5000**
Gaussian	Gaussian	5.4%	7.2%	5.8%	5.8%	6.6%	7.6%
	t	4.2%	4.8%	4.6%	3.2%	5%	4.2%
	Gumbel	3.6%	4.6%	6%	5.2%	7.4%	12.2%
t	Gaussian	7.2%	9%	43.6%	81.4%	99.4%	100%
	t	2.4%	3.6%	2.8%	4.4%	4.6%	5.2%
	Gumbel	4.4%	6.4%	25.6%	57.6%	93.6%	100%
Gumbel	Gaussian	5.4%	6.4%	16.8%	32%	74.2%	97%
	t	3.6%	3.8%	4.4%	4.8%	4.4%	6.4%
	Gumbel	3%	5.2%	4.6%	3.8%	3%	2.8%

Table 6.2: Rejection percentage of fitted models using PIT for various data sizes n

Results

Table 6.2 shows the proportion of fitted models that were rejected using the PIT goodness-of-fit test. It is clear the power of the test increases as more data is available, with rejection rates increasing with respect to the sample size. Of the three copula models tested, the PIT test performs the best for the t copula. In the case of 5,000 data points, it will always reject misspecified models. This is most likely due to the presence of tail dependence in both the upper and lower tails causing models without such prominent features to be rejected. On the other hand, the t copula is almost always accepted when fit to other copulas, showing its flexibility. As goodness-of-fit cannot be clearly established in low data settings, the t copula presents a safe choice for modelling dependences in OpRisk if high dimensionality is not an issue.

The true copula model with the lowest overall rejection rate is the Gaussian copula. This suggests that other copula models can approximate the Gaussian dependence structure well, as the Gaussian copula has no real defining features except symmetry and concentrated density around the mean. From an OpRisk point of view, this means risk practitioners have no reason to model dependence structures using a Gaussian copula aside from convenience, as other copulas can almost always fit the data just as well even if the true dependence relationship is a Gaussian copula.

Conclusion

Operational risk is one of the largest risk classes banks are exposed to and consequently financial regulators require quantitative modelling be utilized in this area to gain a better understanding of the loss process. Banks are required to hold capital to guard against losses associated with operational risk at the 99.9% quantile of the total loss distribution across all risk factors. Modelling multiple risks simultaneously is believed to lead to diversification benefits which reduce the capital a bank is required to hold. In this thesis we explored dependence modelling using the concept of copula models, which are multivariate distributions with uniform marginals that can be used to construct full joint distributions of multiple risk factors.

We modelled the annual loss process of operational risk factors as a compound distribution and studied various copula models that can be used to describe the dependence structure between the frequencies and the severities of the losses. Basic copula models that are appropriate for use in an operational risk context are the Gumbel and t copulas due to their ability to model upper tail dependence between risks. This allows associations between extreme events to be modelled which is important to capital estimation as the prescribed risk measure is the 99.9% quantile of the total loss distribution, a quantity in the far upper tail. We also looked at the properties of more advanced copula models such as mixture copulas, which are linear combinations basic copula models, and vine copulas, which are multivariate distributions decomposed into a series of bivariate copulas.

We studied the induced correlation between two annual loss processes and their joint value-at-risk when dependence was modelled at various levels of the loss distribution using Monte Carlo simulation. We concluded that applying dependence at both the frequency and severity levels allowed a significantly higher range of dependences to be accounted for. We also observed differences in the capital estimation results when different copula models were used as the dependence structure. In particular, the Gaussian copula consistently produced the lowest VaR while the t and Gumbel copulas resulted in higher VaR due to their allowance of tail dependence. In regards to diversification benefits, we can speculate from the results that positive correlation between risk factors lead to a higher value-at-risk than if the they were independent. This means positively correlated risks will lead to negative diversification benefits which will require the bank to hold more capital than if dependence was not modelled. On the other hand, capital reduction can be achieved when modelling negatively correlated risk factors.

Two methods of parameter estimation for copula models were examined: maximum likelihood and Bayesian inference. A 2-step procedure was used for maximum likelihood estimation and a parallel Monte Carlo Markov Chain method involving slice sampling was adopted for the Bayesian approach. Case studies were performed to analyze the accuracy of these procedures by fitting to data generated from known severity distributions. From the results of our testing, we can conclude that for distributions with a small number of parameters such as Lognormal, maximum likelihood estimation is preferred as it is computationally faster and achieves the same level of accuracy as the Bayesian model. Furthermore, prior distributions are not needed under the maximum likelihood approach. However for more complex distributions such as the Normal Inverse Gaussian, our Bayesian procedure consistently yielded more accurate results than the maximum likelihood. However, the maximum likelihood estimation was significantly faster so risk practitioners must consider the trade-off between time and accuracy when deciding between estimation techniques. We also saw that copula parameters and parameters of simple marginal distributions are estimated reasonably well in low data settings, which is the case for operational risks.

However, parameters of more complex marginal distributions require a substantial amount of data to estimate accurately.

Finally goodness-of-fit tests were described and one particular test, the probability integral transform, was selected for further investigation in a simulation study. We found that the test performed well for rejecting misspecified copulas if the true copula had prominent features such as upper and lower tail dependence.

Areas for future research include extending capital estimation to a larger number of risk factors and modelling using a wider range of copula models. The effectiveness of joint estimation of parameters under both the maximum likelihood and Bayesian approaches can also be studied, especially for vine copulas.

Appendix A

Mixture Copula Tail Dependence

Here we show that the tail dependence coefficients of mixture copulas are just linear combinations of the tail dependence coefficients of their component copulas.

Let C_j be the distribution function of the j$^{(th)}$ copula in the mixture with upper and lower tail dependence coefficients given by λ_{U_j} and λ_{L_j} respectively and weights w_i, for $j = 1, ..., n$. Also denote C_{mix} as the distribution function of the mixture copula with upper and lower tail dependence coefficients $\lambda_{U_{mix}}$ and $\lambda_{L_{mix}}$. Then using (2.13) and (2.14) and the fact that weights sum to 1, i.e. $\sum_{i=1}^{n} w_i = 1$, the tail dependence coefficients of the mixture copula is given as:

$$\begin{aligned}
\lambda_{L_{mix}} &= \lim_{u\to 0} \frac{C_{mix}(u,u)}{u} \\
&= \lim_{u\to 0} \frac{\sum_{i=1}^{n} w_i C_i(u,u)}{u} \\
&= \sum_{i=1}^{n} w_i \lim_{u\to 0} \frac{C_i(u,u)}{u} \\
&= \sum_{i=1}^{n} w_i \lambda_{L_j} \\
\lambda_{U_{mix}} &= \lim_{u\to 1} \frac{1-2u+C_{mix}(u,u)}{1-u} \\
&= \lim_{u\to 1} \frac{1-2u+\sum_{i=1}^{n} w_i C_i(u,u)}{1-u} \\
&= \lim_{u\to 1} \frac{\sum_{i=1}^{n} w_i \big(1-2u+C_i(u,u)\big)}{1-u} \\
&= \sum_{i=1}^{n} w_i \lim_{u\to 1} \frac{1-2u+C_i(u,u)}{1-u} \\
&= \sum_{i=1}^{n} w_i \lambda_{U_j}
\end{aligned}$$

Hence the tail dependence coefficients of the mixture copula are just linear combinations of the tail dependence coefficients of the components in the mixture.

Appendix B

Copula Simulation Algorithms

In this section we outline simulation procedures for copula models presented in Chapter 2.

Gaussian Copula

1. Simulate d independent samples $z_1, ...z_d$ from a standard Normal distribution, N(0,1).
2. Find the Cholesky decomposition $\Sigma = AA^T$, where Σ is the correlation matrix associated with the Gaussian copula.
3. Set $X = AZ$, where $Z = (z_1, ..., z_d)^T$.
4. Set $u_i = \Phi(X_i)$, where Φ is the standard Normal distribution function.
5. $(u_1, ..., u_d)$ is a d-variate sample from the Gaussian copula with correlation matrix Σ.

Bivariate Archimedean Copula

1. Generate two independent Uniform$(0, 1)$ samples u and v.
2. Set $q = K^{-1}(u)$, where $K(t) = t - \dfrac{\phi(t)}{\phi'(t)}$, for some t, is the distribution function of the copula and $\phi(t)$ is its generator function. This can be done using a numerical root-finding algorithm where necessary.
3. Set $s = \phi^{-1}[v\,\phi(q)]$ and $t = \phi^{-1}[(1 - v)\,\phi(q)]$.
4. (s, t) is a bivariate sample from the Archimedean copula.

Archimedean Copula - Laplace Transform Method

1. Simulate d independent samples $v_1, ..., v_d$ from a Uniform(0,1) distribution.
2. Simulate y from D such that the Laplace Transform of D satisfies $L(D) = \phi^{-1}$ and $D(0) = 0$.
3. Find $s_i = -(\ln v_i)/y$ for $i = 1, ..., d$.
4. Calculate $u_i = \phi^{-1}(s_i)$ for $i = 1, ..., d$.
5. $(u_1, ..., u_d)$ is a d-variate sample from the Archimedean copula.

The D distribution for the main Archimedean copulas are given as:

- Clayton: $\Gamma(\frac{1}{\rho}, 1),\ \rho \in (0, \infty)$ (Gamma).
- Gumbel: $\mathrm{S}\Big(\frac{1}{\rho}, 1, \big(\cos(\frac{\pi}{2\rho})\big)^{\rho}, 0\Big),\ \rho \in [1, \infty)$ (α-Stable).
- Frank: $\dfrac{-1}{\ln(1-\alpha)}\dfrac{\alpha^k}{k},\ \alpha = (1 - e^{-\rho})$ (Logarithmic).

Standard t-Copula

1. Simulate d independent samples $z_1, ...z_d$ from a standard Normal distribution, N(0,1).
2. Simulate S from a Chi-Square distribution with ν degrees of freedom, χ^2_ν.
3. Find the Cholesky decomposition $\Sigma = AA^T$, where Σ is the correlation matrix associated with the t-copula.
4. Set $Y = AZ$, where $Z = (z_1, ..., z_d)^T$.
5. Set $X = \sqrt{\dfrac{\nu}{S}}Y$.
6. Set $u_i = t(X_i)$, where t is the Student's-t distribution function.
7. $(u_1, ..., u_d)$ is a d-variate sample from the standard t-Copula with ν degrees of freedom and correlation matrix $\mathbf{\Sigma}$.

Mixture Copula

1. Simulate $w_i \times N$ independent samples from copula i for all copulas in the mixture.
2. The samples combined together form the mixture copula.

Canonical Vine Copula

Sample $w_1, ..., w_n$ independent uniform random numbers on $[0, 1]$
$x_1 = v_{1,1} = w_1$
For $i = 2, ..., n$
    $v_{i,1} = w_i$
    For $k = i-1, i-2, ..., 1$
        $v_{i,1} = h^{-1}(v_{i,1}, v_{k,k}, \Theta_{k,i-k})$
    End For
    $x_i = v_{i,1}$
    If $i = n$ **then**
        Stop
        $(x_1, ..., x_n)$ is a sample from the Canonical Vine.
    End If
    For $j = 1, ..., i-1$
        $v_{i,j+1} = h(v_{i,j}, v_{j,j}, \Theta_{j,i-j})$
    End For
End For

where

- $\Theta_{i,j}$ is the set of parameters of the copula density $c_{i,i+j|i+1,...,i+j-1}$.

- $h(x, v, \Theta) = F(x|v) = \dfrac{\partial C_{x,v}(x, v, \Theta)}{\partial v}$.
- $h^{-1}(x, v, \Theta)$ is the inverse of $h(x, v, \Theta)$ with respect to u.

D-Vine Copula

Sample $w_1, ..., w_n$ independent uniform random numbers on $[0, 1]$
$x_1 = v_{1,1} = w_1$
$x_2 = v_{2,1} = h^{-1}(w_2, v_{1,1}, \Theta_{1,1})$
$v_{2,2} = h(v_{1,1}, v_{2,1}, \Theta_{1,1})$
For $i = 3, ..., n$
    $v_{i,1} = w_i$
    For $k = i-1, i-2, ..., 2$
        $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,2k-2}, \Theta_{k,i-k})$
    End For
    $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,1}, \Theta_{1,i-1})$
    $x_i = v_{i,1}$
    If $i = n$ **then**
        Stop
        $(x_1, ..., x_n)$ is a sample from the D Vine.
    End If
    $v_{i,2} = h(v_{i-1,1}, v_{i,1}, \Theta_{1,i-1})$
    $v_{i,3} = h(v_{i,1}, v_{i-1,1}, \Theta_{1,i-1})$
    If $i > 3$ **then**
        For $j = 2, ..., j-2$
            $v_{i,2j} = h(v_{i-1,2j-2}, v_{i,2j-1}, \Theta_{j,i-j})$
            $v_{i,2j+1} = h(v_{i,2j-1}, v_{i-1,2j-2}, \Theta_{j,i-j})$
        End For
    End If
$v_{i,2i-2} = h(v_{i-1,2j-4}, v_{i,2j-3}, \Theta_{i-1,1})$
End For

where

- $\Theta_{i,j}$ is the set of parameters of the copula density $c_{i,i+j|i+1,...,i+j-1}$.
- $h(x, v, \Theta) = F(x|v) = \dfrac{\partial C_{x,v}(x, v, \Theta)}{\partial v}$.
- $h^{-1}(x, v, \Theta)$ is the inverse of $h(x, v, \Theta)$ with respect to u.

Appendix C

Copula Plots

Bivariate scatterplots of copula models described in Chapter 2 but not presented in Chapter 3, due to irrelevance to OpRisk, are shown here.

Clayton Copula

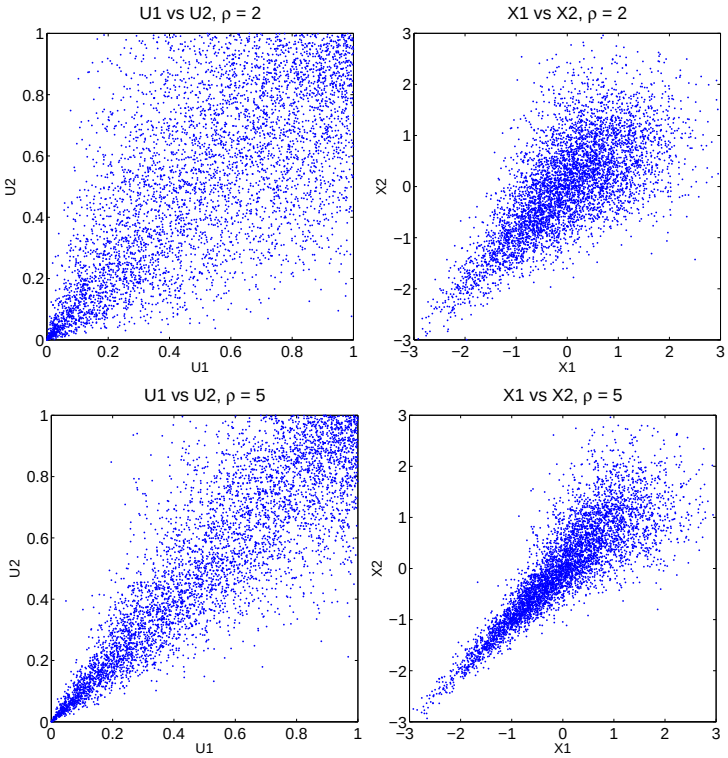


Figure C.1: Clayton copula samples

Frank Copula

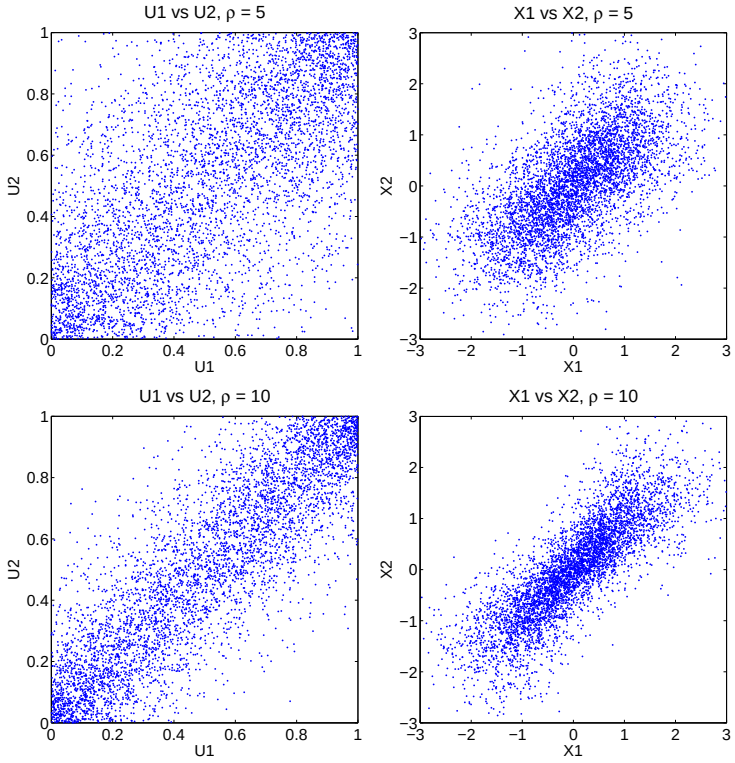


Figure C.2: Frank copula samples

Appendix D

Error Bar Calculations for Induced Correlation and VaR

The steps below outline the methodology used to calculate the simulation errors presented in Chapter3.

1. N iterations of losses were generated for each risk factor.
2. Losses were divided into batches of 5,000 iterations each.
3. The Induced Kendall's Tau, 95% and 99.9% Value-at-Risk measures were calculated for each batch.
4. The errors used in the study are the sample standard deviation of the above measures taken across all of the batches:

$$\epsilon = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \tag{D.1}$$

   where

   - $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$,
   - $n = N/50$ is the number of batches,
   - x_i is the sample measure of the ith batch.

5. The error bars were calculated as:

$$(\tilde{x} - 2\epsilon, \tilde{x} + 2\epsilon) \tag{D.2}$$

   where

   - $\tilde{x}$ is the sample measure of the entire projection.

Appendix E

Bayesian Parameter Estimates

Results of Bayesian parameter estimation that were not included in Chapter 5 are shown in this section.

Lognormal Severity Model

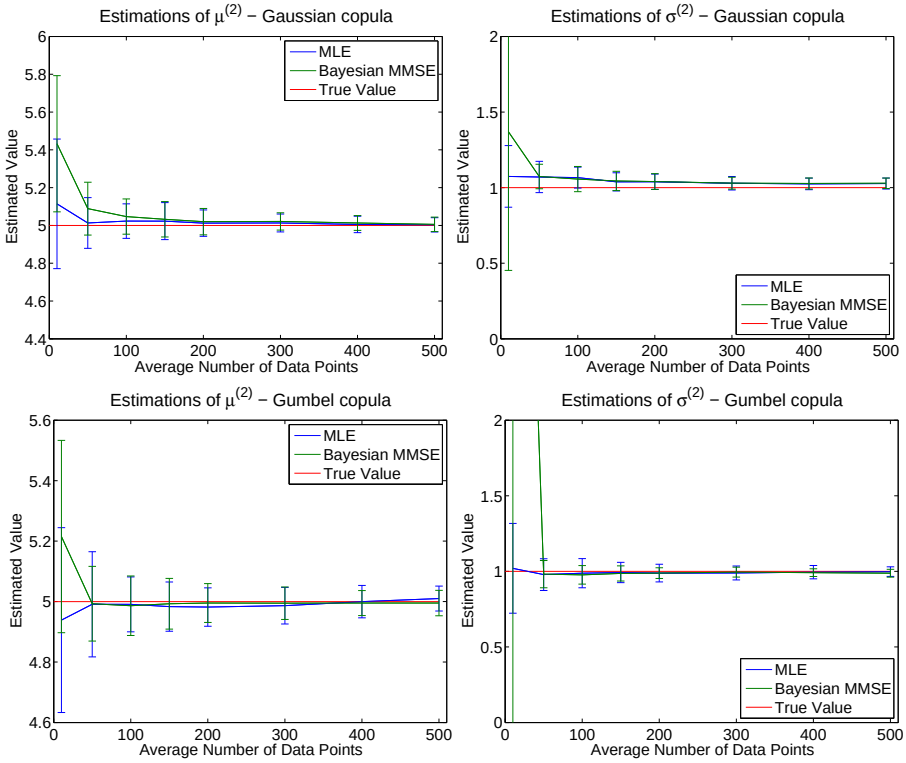


Figure E.1: Lognormal and copula parameters using the 2-step Bayesian approach

Normal Inverse Gaussian Severity Model

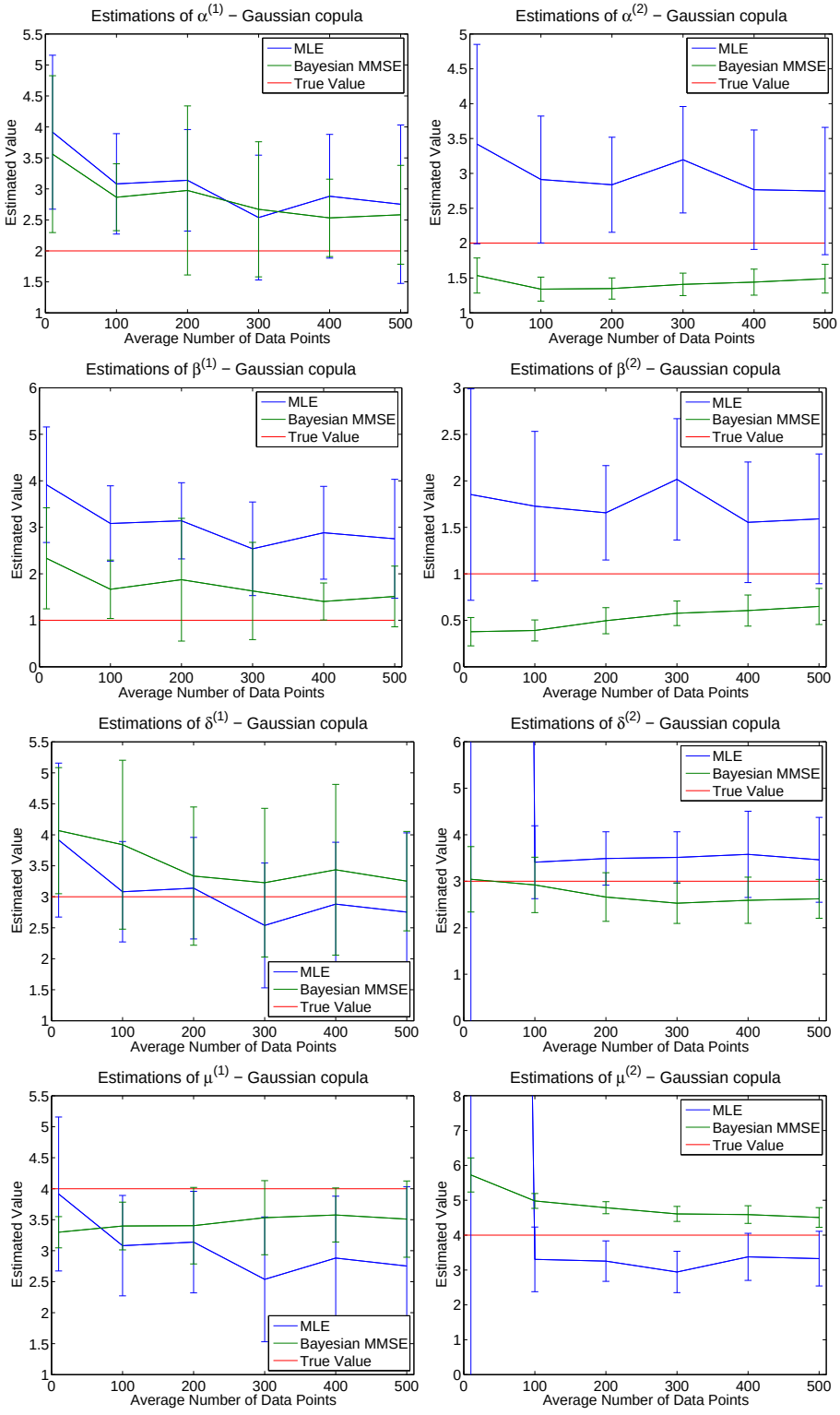


Figure E.2: NIG and copula parameters using the 2-step Bayesian approach

MCMC Diagnostics

Lognormal Severity Model

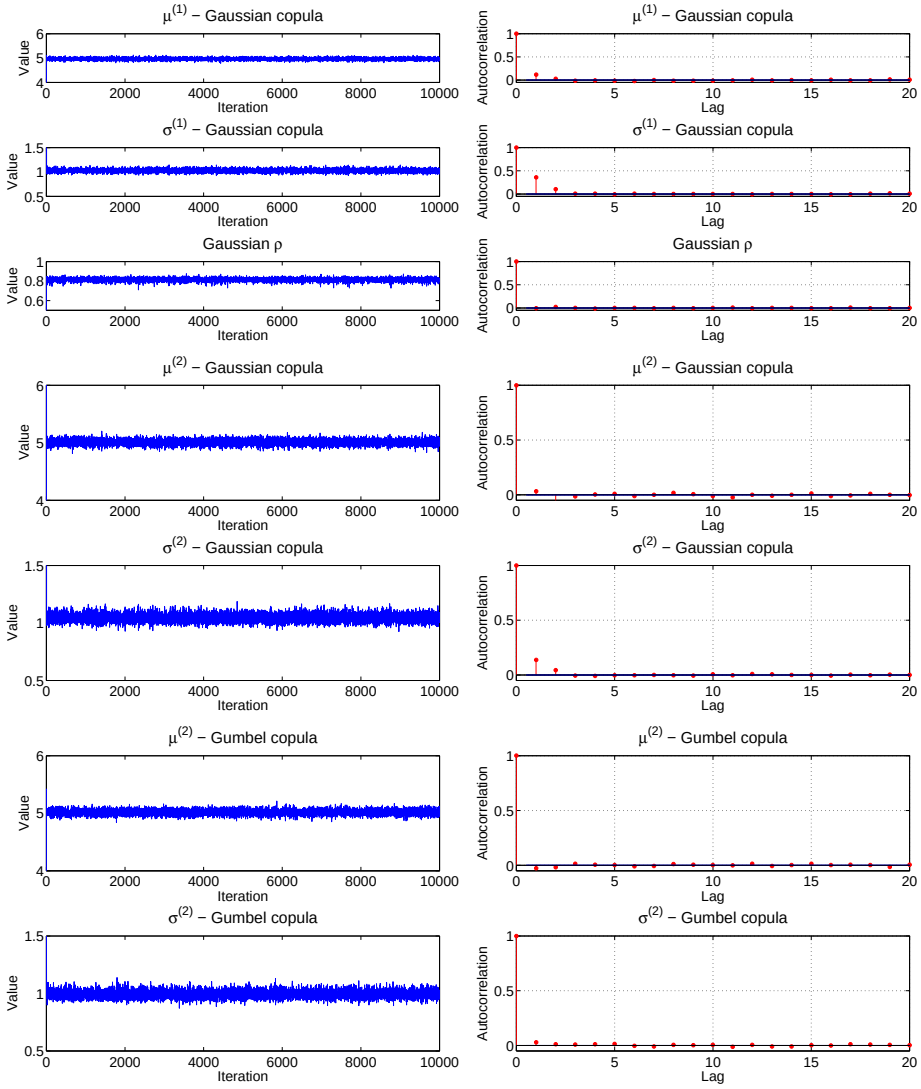


Figure E.3: MCMC diagnostic plots for the Lognormal model

Normal Inverse Gaussian Severity Model

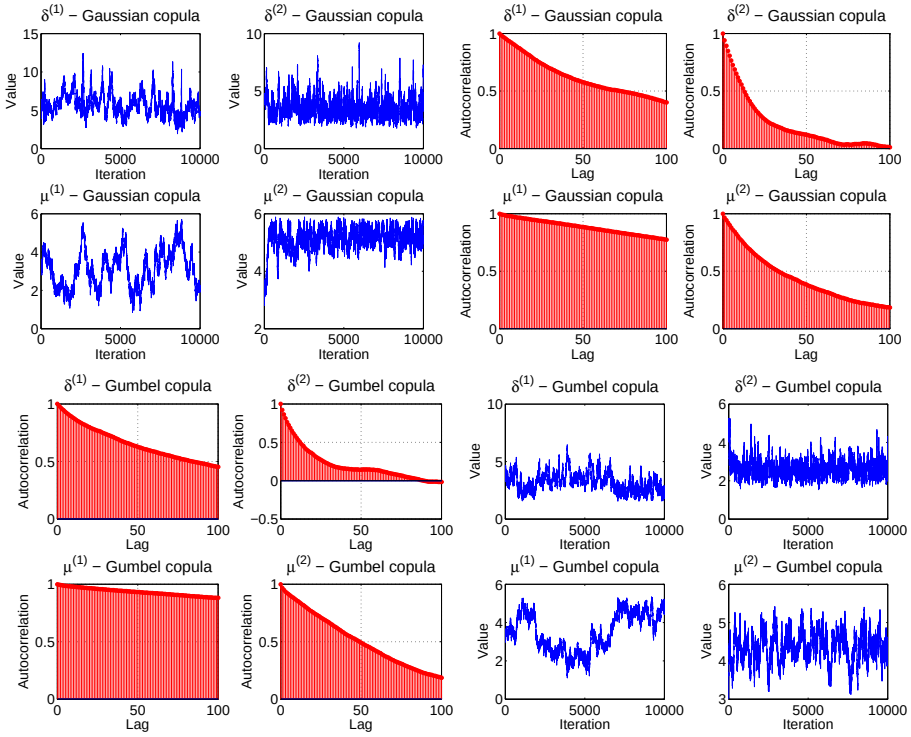


Figure E.4: NIG and copula parameters using the 2-step Bayesian approach

Lognormal Severity Model with Vague Priors

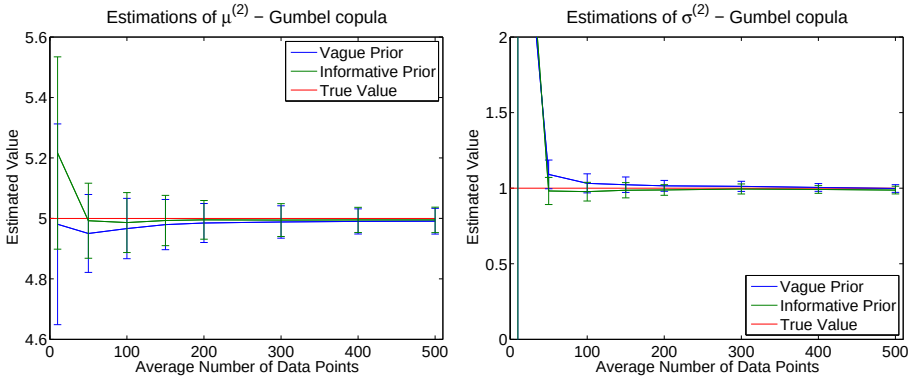


Figure E.5: Lognormal parameters using the 2-step Bayesian approach and vague priors