

A DEPENDENCE MODEL FOR UNBALANCED CROP INSURANCE INDEMNITY
AMOUNTS

By

Yi-Hsuen Lin

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics—Master of Science

2023

ABSTRACT

In this thesis, a copula model is constructed to estimate dependency and calculate the Value at Risk for insurance coverage under the dependence of the covered losses. The dependence model is illustrated in a three-dimensional setting to simplify the complex theoretical functions and provide an accessible introduction to the copula model. The modeling uses the U.S. crop insurance dataset aggregated by each county level and commodity type. The composite likelihood approach helps to simplify the computation in high-dimensional problems by approximating the negative log-likelihood using bivariate components. In this study, the majorization-minimization principle is employed to estimate the parameters of the normal copula by minimizing the composite likelihood iteratively. To avoid overfitting and result in a valid correlation matrix, the L1 penalty is applied to induce sparsity and shrink irrelevant parameters toward zero. The optimal tuning parameter is selected based on the BIC score to generate a positive semi-definite correlation matrix for the result. In the Appendix of the thesis, the dependence model is extended to a high dimension. The Value at Risk computed for the fictional insurance contract in the data analysis results in a higher value when considering dependence between variables.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Gee Lee, for his guidance and offering me the opportunity to participate in this research. This thesis would not have reached its completion without his constant support throughout the research process. His belief in my capabilities has always motivated me, driving me to achieve more than I thought possible. Words cannot adequately express how lucky I am to have learned and grown under his mentorship.

I extend my deep appreciation to my thesis committee members, Dr. Haiyan Liu and Dr. Haolei Weng, who provide their invaluable time and constructive suggestions that have enriched the content and quality of my work.

Most importantly, none of this would have been possible without my family and friends. I am profoundly grateful to my parents, San-Ching Lin and Mei-Ling Chen, for their unconditional love and support. Being your daughter is an immeasurable blessing I deeply treasure. Your constant encouragement and understanding have given me the freedom to explore my passions, pursue my dreams without hesitation, and evolve into the person I am today. I am also thankful to my lovely furry friend, Bella, who accompanied me during late-night paper writing and through frustrated moments, bringing joy even on the most challenging days.

This thesis is a culmination of the support and encouragement I've received from each one of you. Thank you all for being an integral part of this remarkable journey!

TABLE OF CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vi
INTRODUCTION	1
CHAPTER 1: COPULA MODEL	4
1.1 Assumptions	4
1.2 Estimation of Copula Parameters	5
1.3 The Distribution of Correlation Matrices	13
CHAPTER 2: SIMULATION	15
2.1 Influence of Spurious Correlations	15
2.2 C and IC Analysis	17
2.3 Model Performance	18
CHAPTER 3: RESULTS	21
3.1 Data Description	21
3.2 Marginal Models	21
3.3 Dependence Model	23
3.4 Value at Risk (VaR)	24
CHAPTER 4: CONCLUSION	27
BIBLIOGRAPHY	28
APPENDIX	29

LIST OF FIGURES

Figure 1.1: Graph of ρ_{jk} as a function of ψ_{jk}	6
Figure 2.1: Influence of spurious correlations	16
Figure 3.1: Scatter plot of liabilities and indemnities at the county level	23
Figure 3.2: Solution path and optimal tuning parameter for selected commodities	24
Figure 3.3: Simulated density of indemnities for a fictional insurance coverage	26

LIST OF TABLES

Table 2.1: C and IC analysis result	18
Table 3.1: Dataset of liabilities and indemnities	22
Table 3.2: Coefficient estimates for the marginal gamma model	23
Table 3.3: Correlation of selected commodities	23

INTRODUCTION

Dependency refers to the relationship between variables, and the analysis of dependency is crucial since it helps building accurate models and managing risk. Suppose we are given n sample points for a p -dimensional multivariate response, where each of the p marginal response has a different distribution. In this case, the copula approach can be employed to model the dependence between random variables. The classic textbook Nelsen (1999) provides an introduction to the copula model. The literature on copula modeling has developed significantly in recent years and provides numerous extended introduction to the copula models, such as Hofert *et al.* (2018) and Czado (2019) including examples in the R programming language, or Frees and Valdez (1998) exploring the modeling in the actuarial science context. For the use of the copula approach for count data, the reader may see Nikoloulopoulos and Karlis (2010), or Shi and Emiliano (2014).

Sklar's Theorem allows us to derive the multivariate distribution function of a random vector through the use of copula functions, with its corresponding marginal distributions. If we attempt to estimate the normal copula parameters with a p -dimensional correlation matrix, where the number of dependence parameters exceeds the samples size n , ($\binom{p}{2} > n$), it will be considered a high-dimensional problem. In this situation, using maximum likelihood to estimate the parameters may arise several problems. One problem is the potential overfitting with spurious correlations, it happens when all off-diagonal entries of the correlation matrix are non-zero, and the model may be fitting the noise that is not truly meaningful. Another problem is the difficulty in ensuring the estimated correlation matrix satisfies the positive semi-definite property.

Due to these problems, we hope to automatically set some correlation entries to zero during the estimate process, so that we could introduce sparsity that simplifies the model, avoid overfitting, and further result in a valid correlation matrix. There are various ways to achieve this, such as the lasso and ridge regression approach. In this paper, the lasso penalty (L1 penalty) is used to select and shrink some irrelevant correlation parameters toward zero. Nevertheless, while reducing the number of non-zero entries, the process cannot ensure that the resulting matrix is positive semi-definite; it can only increase the likelihood of obtaining a valid correlation matrix. More

relevant details will be discussed in the following chapter.

The lasso is a variable selection method that solves the optimization problem with a constraint, and it was initially introduced in the statistical lecture by Tibshirani (1996). In the thesis, the lasso approach is used to add a penalty term to the negative likelihood function and adjust the tuning parameter to control the severity of the penalty. Due to the accessibility of efficient algorithms for solving the estimation problem, the lasso method has become increasingly popular and developed several extensions, such as shrinking group variables in Yuan and Lin (2006), applying to non-normal responses in Yang and Zou (2015), and the Tweedie compound distribution in Qian *et al.* (2016). The Tweedie compound distribution is commonly used in actuarial science, insurance, and financial fields to model data with both continuous and discrete characteristics. The distribution can be applied in frequency-severity regression as a distribution for modeling insurance losses, which is the severity component of the regression.

Frequency-severity regression is commonly used in actuarial science to analyze the insurance claim occurrence (frequency) and claim amounts (severity). One can refer to Frees *et al.* (2011), Frees (2014) to gain the introduction and applications of frequency-severity regression. In the thesis, the approach used for indemnity modeling has a close connection with the frequency-severity regression approach, both modeling responses consisting of continuous and discrete components. Moreover, the gamma distribution is mainly used to model the crop insurance indemnities in the research.

This research demonstrates the use of the lasso technique on the multivariate normal copula, which helps to simplify the copula parameters by inducing sparsity. The approach takes advantage of the composite likelihood approach for estimation. The composite likelihood approach, explained in Zhao and Joe (2005), Song *et al.* (2005), and Joe (2014), constructs an approximation to the log-likelihood function using bivariate components. It is used for estimating the dependence parameters in the copula model since the log-likelihood may be hard to compute. It is a straightforward method that may be adaptable to other copula families.

The main intent of this research is to explore the dependence model and underlying theoretical

functions that enable the selection and estimation of the normal copula parameters in a three-dimensional setting. By concentrating on three dimensions, we can simplify the explanation of the core concepts and provide a more approachable introduction to the copula model. Although we restrict our focus to three dimensions, the approach explained in this thesis can be extended to higher dimensions. And extensions to other copula families might be possible using the methodologies and insights gained from the research.

The rest of the thesis is organized into four parts. Chapter 1 illustrates the assumptions and approach utilized in the dependence model in three-dimensional setting. The section also introduces the distribution of correlation matrices that is useful while generating valid correlation matrices and may be used in future work. Chapter 2 conducts some simulation studies using the copula estimation routine to evaluate the performance and properties of the model. Chapter 3 shows the results of the estimation using the copula model and the aggregated U.S. crop insurance dataset and calculates Value at Risk for the insurance portfolio losses. Moreover, the Appendix shows the extensions of the normal copula model in Chapter 1 to high dimensions.

CHAPTER 1: COPULA MODEL

1.1 Assumptions

Suppose we are given 3-dimensional multivariate responses of n samples, denoted by (y_{i1}, y_{i2}, y_{i3}) for $i = 1, \dots, n$, and the marginal distributions $F_{i1}(y_{i1})$, $F_{i2}(y_{i2})$, $F_{i3}(y_{i3})$ are known. Then, the task is to use these to derive the joint density among the variables, $f(y_{i1}, y_{i2}, y_{i3})$. Various copula models could fulfill the task, and an introduction to the copulas is explained in Nelsen (1999). Here, we assume using a multivariate normal copula C that links the univariate marginals to their joint distribution and can be written as

$$\begin{aligned} C(u_1, u_2, u_3; \Sigma) &= Pr(U_1 < u_1, U_2 < u_2, U_3 < u_3) \\ &= \Phi_3(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3); \Sigma), \end{aligned} \tag{1.1}$$

where U_i is uniformly distributed on the interval $[0, 1]$, Φ_3 is the cumulative distribution function of a 3-dimensional multivariate normal distribution with zero means and correlation matrix Σ , and Φ^{-1} is the inverse cumulative distribution function of standard normal distribution. In this case, the density of the copula C is

$$c(u_1, u_2, u_3; \Sigma) = \frac{\partial^3}{\partial u_1 \partial u_2 \partial u_3} C(u_1, u_2, u_3; \Sigma). \tag{1.2}$$

Using the probability integral transform $(u_1, u_2, u_3) = (F_1(y_1), F_2(y_2), F_3(y_3))$, the normal copula C can also be written as

$$C(F_{i1}(y_{i1}), F_{i2}(y_{i2}), F_{i3}(y_{i3}); \Sigma) = F(y_{i1}, y_{i2}, y_{i3}). \tag{1.3}$$

By the chain rule, the joint density of the multivariate response is

$$\begin{aligned} f(y_{i1}, y_{i2}, y_{i3}) &= \frac{\partial^3}{\partial y_{i1} \partial y_{i2} \partial y_{i3}} C(F_{i1}(y_{i1}), F_{i2}(y_{i2}), F_{i3}(y_{i3})) \\ &= c(F_{i1}(y_{i1}), F_{i2}(y_{i2}), F_{i3}(y_{i3}); \Sigma) \cdot f_{i1}(y_{i1}) \cdot f_{i2}(y_{i2}) \cdot f_{i3}(y_{i3}). \end{aligned} \tag{1.4}$$

1.2 Estimation of Copula Parameters

The correlation matrix Σ is symmetric and displays the dependence parameter between the i -th and j -th variables in (i, j) entry. In the three-dimensional setting,

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}. \quad (1.5)$$

In the research, the pairwise composite likelihood method is used to estimate the dependence parameters ρ_{ij} . The method is well explained in a multivariate setting by Zhao and Joe (2005), and it is helpful to lower the computational complexity by reducing the dimensionality of the data, especially in high-dimensional problems with complex dependencies among the variables. Assuming all of the three-dimensional multivariate responses for each sample are fully defined, the composite negative log-likelihood function will be

$$\begin{aligned} \ell(\Sigma) = - \sum_{i=1}^n \frac{1}{3-1} & \left[\log c(F_{i1}(y_{i1}), F_{i2}(y_{i2}); \rho_{12}) + \log f_1(y_{i1}) + \log f_2(y_{i2}) \right. \\ & + \log c(F_{i1}(y_{i1}), F_{i3}(y_{i3}); \rho_{13}) + \log f_1(y_{i1}) + \log f_3(y_{i3}) \\ & \left. + \log c(F_{i2}(y_{i2}), F_{i3}(y_{i3}); \rho_{23}) + \log f_2(y_{i2}) + \log f_3(y_{i3}) \right], \end{aligned} \quad (1.6)$$

where $1/(3-1)$ is the weight to match the log-likelihood when independent case and c is a density of bivariate normal copula. However, in general, one or more responses for some of the samples may be undefined, then the weight will become $1/(m_i - 1)$ where m_i is the number of defined responses in observation i . Note that we know the marginal distributions $F_{i1}(y_{i1})$, $F_{i2}(y_{i2})$, $F_{i3}(y_{i3})$, as well as the densities $f_{i1}(y_{i1})$, $f_{i2}(y_{i2})$, $f_{i3}(y_{i3})$ for $i = 1, 2, \dots, n$, the negative log-likelihood $\ell(\Sigma)$ is a function in the dependence parameters only. The optimal values of the parameters ρ_{jk} that minimize the negative log-likelihood function $\ell(\Sigma)$ would fall between $-\infty$

and ∞ . Nonetheless, in order to ensure the dependence parameters satisfy $-1 < \rho_{jk} < 1$, we set

$$\rho_{jk} = \frac{\tan^{-1}(\psi_{jk})}{\pi/2}, \quad -\infty < \psi_{jk} < \infty, \quad (1.7)$$

Figure 1.1 shows that the function is a one-to-one mapping, and for any ψ_{jk} , the corresponding dependence parameters ρ_{jk} will fall within the range of -1 and 1.

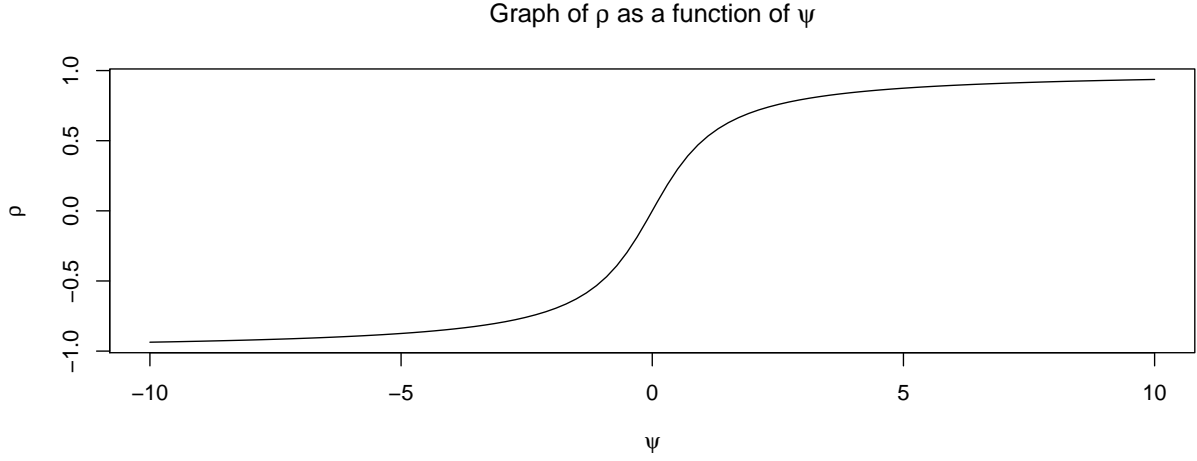


Figure 1.1: Graph of ρ_{jk} as a function of ψ_{jk}

Then, we use the L1 penalty method to reduce the number of non-zero parameters and avoid overfitting, which is helpful in high-dimensional problems which have a large number of parameters to estimate. Now, the penalized negative log-likelihood is

$$\ell_P(\Psi) = \ell(\Sigma) + \lambda(|\psi_{12}| + |\psi_{13}| + |\psi_{23}|) \quad (1.8)$$

where λ is tuning parameter that control the severity of the penalty and

$$\Psi = \begin{bmatrix} 1 & \psi_{12} & \psi_{13} \\ \psi_{21} & 1 & \psi_{23} \\ \psi_{31} & \psi_{32} & 1 \end{bmatrix}. \quad (1.9)$$

Then, the problem now is to estimate the alternative parameters ψ_{ij} by minimizing the penalized

negative log-likelihood function $\ell_P(\Psi)$. Some of the estimates may end up shrunk to zero due to the lasso penalty. The tuning parameter λ controls the number of zero entries in the estimated matrix Ψ . When λ increases, more and more coefficients are set to zero and eliminated. There are different ways to determine λ , such as cross-validation and minimizing AIC (or BIC). In the thesis, the tuning parameter is optimized by minimizing the BIC score. Since we assume that c is the density of a bivariate normal copula, the negative log-likelihood function can be written as

$$\begin{aligned} \ell(\Sigma) = & - \sum_{i=1}^n \frac{1}{3-1} \left[\log \left(\frac{1}{\sqrt{1-\rho_{12}^2}} \right) + \log \left(\frac{1}{\sqrt{1-\rho_{13}^2}} \right) + \log \left(\frac{1}{\sqrt{1-\rho_{23}^2}} \right) \right. \\ & - \frac{(a_i^2 + b_i^2)\rho_{12}^2 - 2a_ib_i\rho_{12}}{2(1-\rho_{12})^2} - \frac{(a_i^2 + c_i^2)\rho_{13}^2 - 2a_ic_i\rho_{13}}{2(1-\rho_{13})^2} - \frac{(b_i^2 + c_i^2)\rho_{23}^2 - 2b_ic_i\rho_{23}}{2(1-\rho_{23})^2} \\ & \left. + \log f_1(y_{i1}) + \log f_2(y_{i2}) + \log f_1(y_{i1}) + \log f_3(y_{i3}) + \log f_2(y_{i2}) + \log f_3(y_{i3}) \right], \end{aligned} \quad (1.10)$$

where

$$\begin{aligned} a_i &= \sqrt{2} \operatorname{erf}^{-1}(2u_{i1} - 1), \quad u_{i1} = F_{i1}(y_{i1}) \\ b_i &= \sqrt{2} \operatorname{erf}^{-1}(2u_{i2} - 1), \quad u_{i2} = F_{i2}(y_{i2}) \\ c_i &= \sqrt{2} \operatorname{erf}^{-1}(2u_{i3} - 1), \quad u_{i3} = F_{i3}(y_{i3}), \end{aligned} \quad (1.11)$$

with

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2) dt. \quad (1.12)$$

In order to estimate the dependent parameters, let

$$\begin{aligned}
\Phi_{i12} &= -\frac{1}{3-1} \left[\log \left(\frac{1}{\sqrt{1-\rho_{12}^2}} \right) - \frac{(a_i^2 + b_i^2)\rho_{12}^2 - 2a_i b_i \rho_{12}}{2(1-\rho_{12}^2)^2} + \log f_{i1}(y_{i1}) + \log f_{i2}(y_{i2}) \right] \\
\Phi_{i13} &= -\frac{1}{3-1} \left[\log \left(\frac{1}{\sqrt{1-\rho_{13}^2}} \right) - \frac{(a_i^2 + c_i^2)\rho_{13}^2 - 2a_i c_i \rho_{13}}{2(1-\rho_{13}^2)^2} + \log f_{i1}(y_{i1}) + \log f_{i3}(y_{i3}) \right] \\
\Phi_{i23} &= -\frac{1}{3-1} \left[\log \left(\frac{1}{\sqrt{1-\rho_{23}^2}} \right) - \frac{(b_i^2 + c_i^2)\rho_{23}^2 - 2b_i c_i \rho_{23}}{2(1-\rho_{23}^2)^2} + \log f_{i2}(y_{i2}) + \log f_{i3}(y_{i3}) \right]
\end{aligned} \tag{1.13}$$

so that

$$\ell(\Sigma) = \sum_{i=1}^n \Phi_{i12} + \Phi_{i13} + \Phi_{i23}. \tag{1.14}$$

Remember that we map ρ_{jk} to ψ_{jk} by $\rho_{jk} = 2 \tan^{-1}(\psi_{jk})/\pi$ for ensuring the dependence parameters satisfy the condition of between -1 and 1. To obtain the Hessian and gradient of log-likelihood $\ell(\Sigma)$, the log-likelihood $\ell(\Sigma)$ should be differentiated in terms of the parameters ψ_{jk} using the chain rule. Then consider

$$\frac{\partial \Phi_{i12}}{\partial \psi_{12}} = \frac{\partial \Phi_{i12}}{\partial \rho_{12}} \frac{\partial \rho_{12}}{\partial \psi_{12}}, \quad \frac{\partial \Phi_{i13}}{\partial \psi_{13}} = \frac{\partial \Phi_{i13}}{\partial \rho_{13}} \frac{\partial \rho_{13}}{\partial \psi_{13}}, \quad \frac{\partial \Phi_{i23}}{\partial \psi_{23}} = \frac{\partial \Phi_{i23}}{\partial \rho_{23}} \frac{\partial \rho_{23}}{\partial \psi_{23}}, \tag{1.15}$$

where

$$\begin{aligned}
\frac{\partial \Phi_{i12}}{\partial \rho_{12}} &= -\frac{1}{3-1} \left[\frac{\rho_{12}}{1-\rho_{12}^2} + \frac{a_i b_i \rho_{12}^2 - (a_i^2 + b_i^2)\rho_{12} + a_i b_i}{(1-\rho_{12}^2)^2} \right] \\
\frac{\partial \Phi_{i13}}{\partial \rho_{13}} &= -\frac{1}{3-1} \left[\frac{\rho_{13}}{1-\rho_{13}^2} + \frac{a_i c_i \rho_{13}^2 - (a_i^2 + c_i^2)\rho_{13} + a_i c_i}{(1-\rho_{13}^2)^2} \right] \\
\frac{\partial \Phi_{i23}}{\partial \rho_{23}} &= -\frac{1}{3-1} \left[\frac{\rho_{23}}{1-\rho_{23}^2} + \frac{b_i c_i \rho_{23}^2 - (b_i^2 + c_i^2)\rho_{23} + b_i c_i}{(1-\rho_{23}^2)^2} \right],
\end{aligned} \tag{1.16}$$

and

$$\frac{\partial \rho_{12}}{\partial \psi_{12}} = \frac{2}{\pi(1 + \psi_{12}^2)}, \quad \frac{\partial \rho_{13}}{\partial \psi_{13}} = \frac{2}{\pi(1 + \psi_{13}^2)}, \quad \frac{\partial \rho_{23}}{\partial \psi_{23}} = \frac{2}{\pi(1 + \psi_{23}^2)}. \quad (1.17)$$

For the Hessian, we have

$$\begin{aligned} \frac{\partial^2 \Phi_{i12}}{\partial \psi_{12}^2} &= \frac{\partial^2 \Phi_{i12}}{\partial \rho_{12}^2} \left(\frac{\partial \rho_{12}}{\partial \psi_{12}} \right)^2 + \frac{\partial \Phi_{i12}}{\partial \rho_{12}} \left(\frac{\partial^2 \rho_{12}}{\partial \psi_{12}^2} \right) \\ \frac{\partial^2 \Phi_{i13}}{\partial \psi_{13}^2} &= \frac{\partial^2 \Phi_{i13}}{\partial \rho_{13}^2} \left(\frac{\partial \rho_{13}}{\partial \psi_{13}} \right)^2 + \frac{\partial \Phi_{i13}}{\partial \rho_{13}} \left(\frac{\partial^2 \rho_{13}}{\partial \psi_{13}^2} \right) \\ \frac{\partial^2 \Phi_{i23}}{\partial \psi_{23}^2} &= \frac{\partial^2 \Phi_{i23}}{\partial \rho_{23}^2} \left(\frac{\partial \rho_{23}}{\partial \psi_{23}} \right)^2 + \frac{\partial \Phi_{i23}}{\partial \rho_{23}} \left(\frac{\partial^2 \rho_{23}}{\partial \psi_{23}^2} \right), \end{aligned} \quad (1.18)$$

where

$$\begin{aligned} \frac{\partial^2 \Phi_{i12}}{\partial \rho_{12}^2} &= -\frac{1}{2} \left[\frac{1 + \rho_{12}^2}{(1 - \rho_{12}^2)^2} + \frac{2a_i b_i \rho_{12} - (a_i^2 + b_i^2)}{(1 - \rho_{12}^2)^2} + \frac{4 \{a_i b_i \rho_{12}^2 - (a_i^2 + b_i^2) \rho_{12} + a_i b_i\} (1 - \rho_{12}^2) \rho_{12}}{(1 - \rho_{12}^2)^4} \right] \\ \frac{\partial^2 \Phi_{i13}}{\partial \rho_{13}^2} &= -\frac{1}{2} \left[\frac{1 + \rho_{13}^2}{(1 - \rho_{13}^2)^2} + \frac{2a_i c_i \rho_{13} - (a_i^2 + c_i^2)}{(1 - \rho_{13}^2)^2} + \frac{4 \{a_i c_i \rho_{13}^2 - (a_i^2 + c_i^2) \rho_{13} + a_i c_i\} (1 - \rho_{13}^2) \rho_{13}}{(1 - \rho_{13}^2)^4} \right] \\ \frac{\partial^2 \Phi_{i23}}{\partial \rho_{23}^2} &= -\frac{1}{2} \left[\frac{1 + \rho_{23}^2}{(1 - \rho_{23}^2)^2} + \frac{2b_i c_i \rho_{23} - (b_i^2 + c_i^2)}{(1 - \rho_{23}^2)^2} + \frac{4 \{b_i c_i \rho_{23}^2 - (b_i^2 + c_i^2) \rho_{23} + b_i c_i\} (1 - \rho_{23}^2) \rho_{23}}{(1 - \rho_{23}^2)^4} \right], \end{aligned} \quad (1.19)$$

and

$$\frac{\partial^2 \rho_{12}}{\partial \psi_{12}^2} = -\frac{4\psi_{12}}{\pi(1 + \psi_{12}^2)^2}, \quad \frac{\partial^2 \rho_{13}}{\partial \psi_{13}^2} = -\frac{4\psi_{13}}{\pi(1 + \psi_{13}^2)^2}, \quad \frac{\partial^2 \rho_{23}}{\partial \psi_{23}^2} = -\frac{4\psi_{23}}{\pi(1 + \psi_{23}^2)^2}. \quad (1.20)$$

The $\binom{3}{2}$ dimensional vector of parameters is

$$\boldsymbol{\psi} = (\psi_{12}, \psi_{13}, \psi_{23})^T. \quad (1.21)$$

The $\binom{3}{2} \times \binom{3}{2}$ Hessian matrix \mathbf{H}_ψ is diagonal and the off-diagonal elements are zero

$$\mathbf{H}_\psi = \begin{bmatrix} \mathbf{H}_\psi(1, 2) & 0 & 0 \\ 0 & \mathbf{H}_\psi(1, 3) & 0 \\ 0 & 0 & \mathbf{H}_\psi(2, 3) \end{bmatrix}. \quad (1.22)$$

where

$$\mathbf{H}_\psi(1, 2) = \frac{\partial^2}{\partial \psi_{12}^2} \ell(\boldsymbol{\Sigma}) = \sum_{i=1}^n \frac{\partial^2}{\partial \psi_{12}^2} [\Phi_{i12} + \Phi_{i13} + \Phi_{i23}] = \sum_{i=1}^n \frac{\partial^2 \Phi_{i12}}{\partial \psi_{12}^2} \quad (1.23)$$

and similarly

$$\mathbf{H}_\psi(1, 3) = \sum_{i=1}^n \frac{\partial^2 \Phi_{i13}}{\partial \psi_{13}^2}, \quad \mathbf{H}_\psi(2, 3) = \sum_{i=1}^n \frac{\partial^2 \Phi_{i23}}{\partial \psi_{23}^2}. \quad (1.24)$$

And the $\binom{3}{2}$ gradient vector \mathbf{U}_ψ is

$$\begin{aligned} \mathbf{U}_\psi &= (\mathbf{U}_\psi(1, 2), \mathbf{U}_\psi(1, 3), \mathbf{U}_\psi(2, 3))^T \\ &= \left(\sum_{i=1}^n \frac{\partial \Phi_{i12}}{\partial \psi_{12}}, \sum_{i=1}^n \frac{\partial \Phi_{i13}}{\partial \psi_{13}}, \sum_{i=1}^n \frac{\partial \Phi_{i23}}{\partial \psi_{23}} \right)^T \end{aligned} \quad (1.25)$$

Note $\tilde{\boldsymbol{\psi}}$ denotes the parameters updated in the outer loop of Algorithm 1, and $\check{\boldsymbol{\psi}}$ for the inner loop. Then, the second-order (quadratic) Taylor series approximation of the unpenalized negative log-likelihood function $\ell(\boldsymbol{\Sigma})$ around a given point $\tilde{\boldsymbol{\psi}}$ is

$$\ell_Q(\boldsymbol{\psi}) = \ell(\tilde{\boldsymbol{\psi}}) + \tilde{\mathbf{U}}_\psi^T (\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}}) + \frac{1}{2} (\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}})^T \tilde{\mathbf{H}}_\psi (\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}}), \quad (1.26)$$

where $\tilde{\mathbf{U}}_\psi$ is the gradient vector and $\tilde{\mathbf{H}}_\psi$ is the Hessian matrix obtained at the given point $\tilde{\boldsymbol{\psi}} = (\tilde{\psi}_{12}, \tilde{\psi}_{13}, \tilde{\psi}_{23})^T$. Now, the problem is to minimize the following penalized objective function:

$$P_Q(\boldsymbol{\psi}) = \ell_Q(\boldsymbol{\psi}) + \lambda (|\psi_{12}| + |\psi_{13}| + |\psi_{23}|). \quad (1.27)$$

However, the log-likelihood $\ell_Q(\boldsymbol{\psi})$ is complicated and the penalized objective function $P_Q(\boldsymbol{\Psi})$ may be non-smooth, so the problem cannot be directly handled by first and second-order conditions for optimization. Thus, the majorization-minimization algorithm is used to accomplish the task, and it is well explained in Yang and Zou (2015) and Qian *et al.* (2016). The key step of applying the majorization-minimization is constructing a surrogate function that majorizes the objective function. The algorithm estimates the parameters by iteratively minimizing a sequence of surrogate functions. In this paper, we construct the surrogate function by second-order Taylor expansion. First, define

$$\mathbf{U}_Q = \nabla \ell_Q(\boldsymbol{\psi}) \quad (1.28)$$

$$\mathbf{H}_Q = \nabla^2 \ell_Q(\boldsymbol{\psi}), \quad (1.29)$$

where the entries of \mathbf{U}_Q are given by

$$\begin{aligned} \mathbf{U}_Q(1, 2) &= \tilde{\mathbf{U}}_{\boldsymbol{\psi}}(1, 2) + \tilde{\mathbf{H}}_{\boldsymbol{\psi}}(1, 2)(\psi_{12} - \tilde{\psi}_{12}) \\ &= \left(\sum_{i=1}^n \frac{\partial \tilde{\Phi}_{i12}}{\partial \psi_{12}} \right) + \left(\sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i12}}{\partial \psi_{12}^2} \right) (\psi_{12} - \tilde{\psi}_{12}), \\ \mathbf{U}_Q(1, 3) &= \tilde{\mathbf{U}}_{\boldsymbol{\psi}}(1, 3) + \tilde{\mathbf{H}}_{\boldsymbol{\psi}}(1, 3)(\psi_{13} - \tilde{\psi}_{13}) \\ &= \left(\sum_{i=1}^n \frac{\partial \tilde{\Phi}_{i13}}{\partial \psi_{13}} \right) + \left(\sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i13}}{\partial \psi_{13}^2} \right) (\psi_{13} - \tilde{\psi}_{13}), \\ \mathbf{U}_Q(2, 3) &= \left(\sum_{i=1}^n \frac{\partial \tilde{\Phi}_{i23}}{\partial \psi_{23}} \right) + \left(\sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i23}}{\partial \psi_{23}^2} \right) (\psi_{23} - \tilde{\psi}_{23}), \end{aligned} \quad (1.30)$$

and the diagonal entries of \mathbf{H}_Q are given by

$$\begin{aligned}\mathbf{H}_Q(1, 2) &= \tilde{\mathbf{H}}_\psi(1, 2) = \sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i12}}{\partial \psi_{12}^2}, \\ \mathbf{H}_Q(1, 3) &= \tilde{\mathbf{H}}_\psi(1, 3) = \sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i13}}{\partial \psi_{13}^2}, \\ \mathbf{H}_Q(2, 3) &= \tilde{\mathbf{H}}_\psi(2, 3) = \sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{i23}}{\partial \psi_{23}^2},\end{aligned}\tag{1.31}$$

where \mathbf{H}_Q is also diagonal.

Then, to ensure the surrogate function majorizes the objective function $P_Q(\psi)$, we substitute the Hessian matrix \mathbf{H}_Q with its largest eigenvalue in the second-order Taylor expansion of the unpenalized negative log-likelihood function $\ell_Q(\psi)$. Given $\ell_Q(\psi)$, consider updating current estimates $\check{\psi}$. The new parameters $\check{\psi}^{(new)}$ are given by solving

$$\check{\psi}^{(new)} = \arg \min_{\check{\psi}} \ell_Q(\check{\psi}) + \tilde{\mathbf{U}}_Q^T(\psi - \check{\psi}) + (\psi - \check{\psi})^T \tilde{\gamma}_\psi (\psi - \check{\psi}) + \lambda(|\psi_{12}| + |\psi_{13}| + |\psi_{23}|). \tag{1.32}$$

The solution for this problem is given by

$$\check{\psi}_{jk}^{(new)} = \frac{\left(\tilde{\gamma}_\psi \check{\psi}_{jk} - \tilde{\mathbf{U}}_Q(j, k) \right) \left(1 - \frac{\lambda}{|\tilde{\gamma}_\psi \check{\psi}_{jk} - \tilde{\mathbf{U}}_Q(j, k)|} \right)}{\tilde{\gamma}_\psi}, \tag{1.33}$$

where $\tilde{\gamma}_\psi$ is the largest eigenvalue of $\tilde{\mathbf{H}}_Q$. The entire procedure is summarized in Algorithm 1.

Algorithm 1: Algorithm for penalized estimation of the copula parameters.

1. Initialize $\tilde{\psi}$.
 2. (Outer Loop) Update the penalized objective function in Equation (1.27).
 - (a) Compute \tilde{U}_Q , and \tilde{H}_Q .
 - (b) Compute $\tilde{\gamma}_\psi$, the largest eigenvalue of \tilde{H}_Q .
 - (c) (Inner Loop) Obtain the minimizer of the objective function in (1.27).
 - Initialize $\check{\psi} = \tilde{\psi}$.
 - Repeat the following until $\check{\psi}$ converges.
 - Compute $\check{\psi}^{(new)}$ by Equation (4.38).
 - Set $\check{\psi} = \check{\psi}^{(new)}$.
 - Set $\tilde{\psi} = \check{\psi}$.
 3. Repeat until $\tilde{\psi}$ converges.
-

1.3 The Distribution of Correlation Matrices

The authors Pourahmadi and Wang (2015) have shown any p dimensional correlation matrix Σ can be Cholesky decomposed into $\Sigma = \mathbf{B}\mathbf{B}^T$, where \mathbf{B} is a lower triangular matrix with entries b_{jk} , and

$$b_{jk} = \begin{cases} 1 & \text{for } j = k = 1, \\ \cos \theta_{j1} & \text{for } j = 2, \dots, p, \\ \prod_{h=1}^{k-1} \sin \theta_{jh} & \text{for } j = k > 1, \\ \cos \theta_{jk} \cdot \prod_{h=1}^{k-1} \sin \theta_{jh} & \text{for } 2 \leq k \leq j-1, \end{cases} \quad (1.34)$$

where $\theta_{jk}, j > k$ are some angles. For example, in a three dimensional setting,

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ \cos \theta_{21} & \sin \theta_{21} & 0 \\ \cos \theta_{31} & \cos \theta_{32} \sin \theta_{31} & \sin \theta_{32} \sin \theta_{31} \end{bmatrix}. \quad (1.35)$$

The matrix \mathbf{B} is unique if its diagonal entries are positive, or equivalently if the angles are restricted to the range $(0, \pi)$. Furthermore, the transformation from Σ to a $\binom{p}{2} \times \binom{p}{2}$ matrix $\Theta = (\theta_{jk})$ is one-to-one, where $\theta_{jk} = 0$ for $j \leq k$ and θ_{jk} are angles for $j > k$. Thus, this helps us to generate valid high-dimensional correlation matrices using $\binom{p}{2}$ parameters in the matrix Θ .

Now, we may wonder if the resulting matrix from the estimating procedure shown in Section 1.2 would be a valid correlation matrix that is positive semi-definite. However, it would not be an issue for practical application since the estimated correlations will be close to the true underlying correlations of the data when the sample size is large enough. Practically, the resulting correlation matrix in Section 3 is positive semi-definite since the smallest eigenvalue is positive. Figure 3.2 shows that the smallest eigenvalue decreases when the tuning parameter gets smaller, which means too many non-zero parameters may result in an invalid correlation matrix. Future work may use the $\binom{p}{2}$ parameters in matrix Θ as alternatives to correlations ρ_{jk} in the copula model, to avoid any possible theoretical problems in the algorithm. For this, the following compact expression from Pourahmadi and Wang (2015) will be useful to transform correlations to thetas:

$$\rho_{jk} = c_{j1}c_{k1} + \sum_{h=2}^{j-1} c_{jh}c_{kh} \prod_{l=1}^{h-1} s_{jl}s_{kl} + c_{kj} \prod_{l=1}^{j-1} s_{jl}s_{kl}, \quad 1 \leq j < k \leq p, \quad (1.36)$$

where $c_{jk} = \cos \theta_{jk}$ and $s_{jk} = \sin \theta_{jk}$. However, applying the expression to the copula model may increase the computational complexity, especially in the differentiation parts. In the thesis, we rely on the $\rho_{ij} = \tan^{-1}(\psi_{ij})/(\pi/2)$ mapping instead of the alternative parameterization using Θ for its simplicity.

CHAPTER 2: SIMULATION

2.1 Influence of Spurious Correlations

The purpose of this simulation is to analyze the influence of spurious correlations and further demonstrate the advantages of setting certain correlations to zero. Assume we have the following correlation matrices with their corresponding situation:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.9 & 0.1 \\ 0.9 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0.9 & -0.1 \\ 0.9 & 1 & -0.1 \\ -0.1 & -0.1 & 1 \end{bmatrix}, \quad (2.1)$$

- \mathbf{A} : Positive spurious correlations in the other off-diagonal entries
- \mathbf{B} : True underlying correlation matrix without spurious correlations
- \mathbf{C} : Negative spurious correlations in the other off-diagonal entries

The task is to analyze the variance and the 95th percentile of insurance portfolio losses with each dependence matrix from the simulation. To do so, we simulate B samples from copulas with each correlation matrix and denote the randomly generated uniform variables by $u_{A,ij}$, $u_{B,ij}$, $u_{C,ij}$, for $i = 1, \dots, B$ and $j = 1, 2, 3$. Assume the distribution of each response is known, which is a gamma distribution with shape 2 and scale 100. The realized losses can be simulated using inverse transform sampling: $y_{A,ij} = F^{-1}(u_{A,ij})$, $y_{B,ij} = F^{-1}(u_{B,ij})$, $y_{C,ij} = F^{-1}(u_{C,ij})$. Since we are interested in the simulated insurance portfolio losses with the dependence structure \mathbf{A} , \mathbf{B} , and \mathbf{C} , the losses should be aggregated for analysis

$$\begin{aligned} T_{\mathbf{A},i} &= (y_{\mathbf{A},i,1} + y_{\mathbf{A},i,2} + y_{\mathbf{A},i,3}), \quad i = 1, \dots, B \\ T_{\mathbf{B},i} &= (y_{\mathbf{B},i,1} + y_{\mathbf{B},i,2} + y_{\mathbf{B},i,3}), \quad i = 1, \dots, B \\ T_{\mathbf{C},i} &= (y_{\mathbf{C},i,1} + y_{\mathbf{C},i,2} + y_{\mathbf{C},i,3}), \quad i = 1, \dots, B, \end{aligned} \quad (2.2)$$

The variance and the 95th percentiles are plotted with increasing sample sizes B and shown in Figure 2.1.

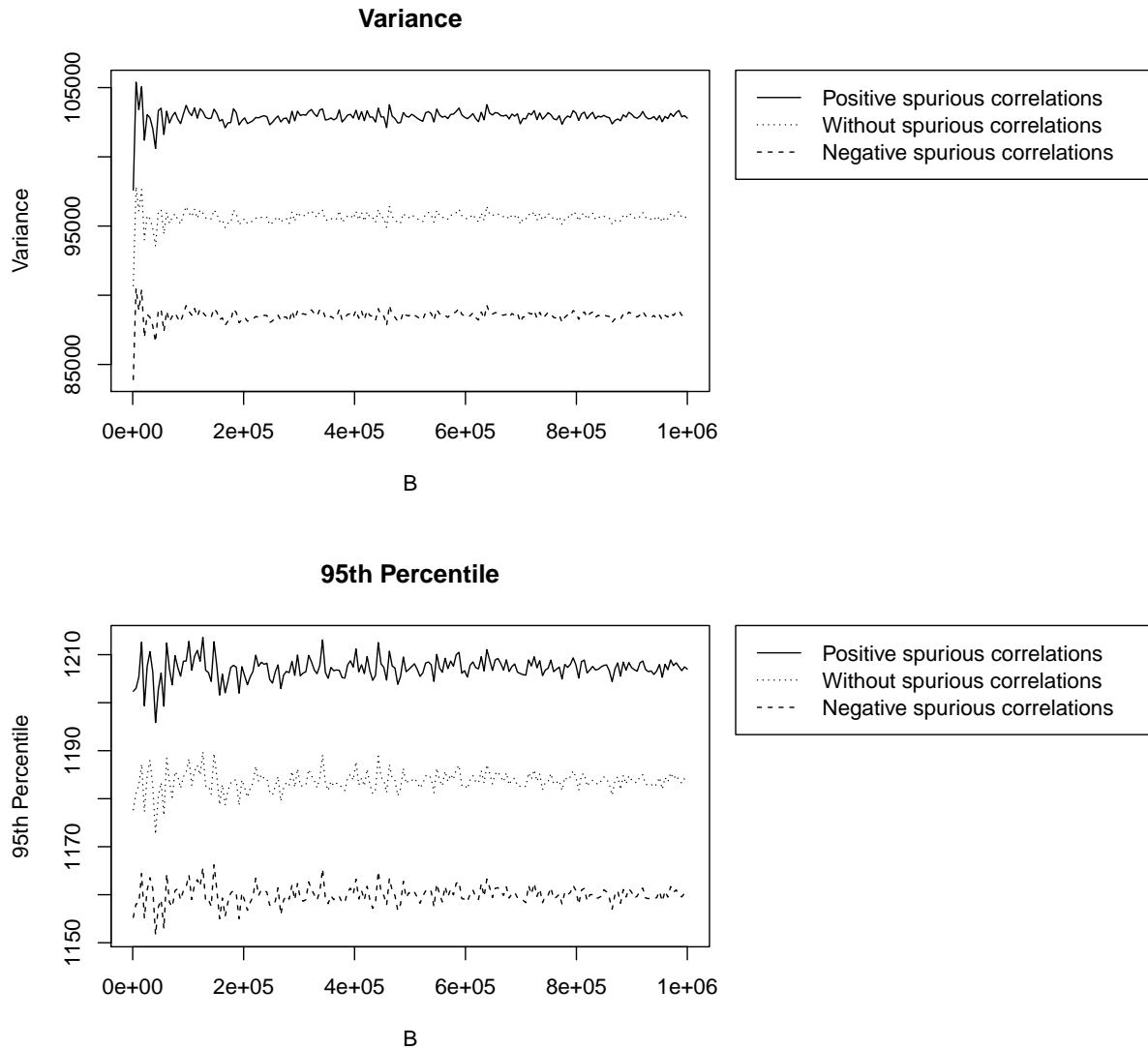


Figure 2.1: Influence of spurious correlations

In Figure 2.1, it is obvious that spurious correlations may affect the estimations of portfolio risk. Positive spurious correlations may overestimate the portfolio variance and 95th percentile, while negative spurious correlations resulted in underestimation. This indicates the usefulness of inducing sparsity into the resulting correlation matrix that may reduce the spurious correlations.

2.2 C and IC Analysis

The goal of this simulation is to know how the sample size n and the dimension p impact the algorithm on identifying zero and non-zero correlation coefficients. In order to reach the goal, we plan to randomly generate a p -dimensional correlation matrix with some zero entries and simulate n observations from a copula using this matrix. Note that the distribution of correlation matrices in Section 1.3 is useful when generating a correlation matrix with a particular dependence structure, and the inputs of the generation approach are the $\binom{p}{2}$ parameters in the lower triangular matrix Θ . Here, we set 90% of the $\binom{p}{2}$ parameters in Θ to $\pi/2$ and the rest 10% to $\pi/4$. It results in a correlation matrix with 90% of the off-diagonal entries being zero, since $b_{jk} = 0$ for $k < j$ when $\cos(\theta_{jk}) = \cos(\pi/2) = 0$. Comparing the generated matrix, which is the true correlation matrix that produces the data, and the estimated correlation matrix from the algorithm, we classify the entries of the correlation matrix into the following cases:

- CZ (Correctly identified zero): This occurs when the true coefficient is zero, and the algorithm correctly identifies it as zero.
- ICZ (Incorrectly identified zero): This occurs when the true underlying coefficient is zero, but the algorithm thinks it is non-zero.
- CNZ (Correctly identified non-zero): This occurs when the true coefficient is non-zero, and the algorithm thinks it is non-zero.
- ICNZ (Incorrectly identified non-zero): This occurs when the true coefficient is non-zero, but the algorithm identifies it as zero.

We duplicate the simulation 100 times, and the average number of CZ, ICZ, CNZ, and ICNZ cases with different sample size (n) and dimension (p) are listed in Table 2.1. Note that the Oracle is the idealized situation, the HDcop is the high-dimensional algorithm of our research, and the MLE is the maximum likelihood estimation approach.

In the first panel of Table 2.1, the number of responses is fixed to $p = 10$, the algorithm has a high CZ value for all sample sizes, indicating that it was successful at identifying the zero

coefficients. The ICZ value is higher when the sample size is small, but it decreases quickly as the sample size increases. On the contrary, the CNZ value increases as the sample size increases. On the other hand, MLE seems to be unsuccessful at identifying the true zero coefficients, but it identifies the true non-zero well. In the second panel of Table 2.1, we fix the sample size to $n = 400$, and the number of correctly identified coefficient (CZ and CNZ) increases when dimension p increases. However, the MLE seems poor at identifying any true zero and non-zero coefficients.

Table 2.1: C and IC analysis result

Sample size (n)		100	200	300	400	500	600	700	800	900	1,000
Oracle	CZ	40	40	40	40	40	40	40	40	40	40
	ICZ	0	0	0	0	0	0	0	0	0	0
	CNZ	5	5	5	5	5	5	5	5	5	5
	ICNZ	0	0	0	0	0	0	0	0	0	0
HDcop	CZ	39.90	39.54	39.48	39.49	39.44	39.49	39.37	39.46	39.38	39.40
	ICZ	3.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CNZ	1.51	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	ICNZ	0.10	0.46	0.52	0.51	0.56	0.51	0.63	0.54	0.62	0.60
MLE	CZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ICZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CNZ	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	ICNZ	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00

Dimension (p)		3	4	5	6	7	8	9	10	11	12
Oracle	CZ	2.68	5.26	8.95	13.53	18.38	24.51	31.41	39.61	47.85	57.73
	ICZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CNZ	0.32	0.74	1.05	1.47	2.62	3.49	4.59	5.39	7.15	8.27
	ICNZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HDcop	CZ	2.61	5.25	8.82	13.05	18.41	24.29	31.27	38.70	47.61	56.68
	ICZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	CNZ	0.32	0.59	1.05	1.86	2.32	3.42	4.43	5.75	6.87	8.55
	ICNZ	0.07	0.16	0.13	0.09	0.27	0.29	0.30	0.56	0.52	0.75
MLE	CZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ICZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CNZ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ICNZ	2.68	5.41	8.95	13.14	16.68	24.58	31.57	39.26	48.13	57.43

2.3 Model Performance

The goal of this simulation is to test the performance and precision of the dependence model described in Section 1.2. In other words, we want to compare the underlying correlation matrix, which produced the data, and the resulting matrix from the estimating procedure. First, according to the generation approach in Section 1.3, we randomly generate a lower triangular matrix Θ with

$\binom{5}{2}$ angles in range $(0, \pi)$ as input. For example,

$$\begin{aligned} &(\theta_{21}, \theta_{31}, \theta_{41}, \theta_{51}, \theta_{32}, \theta_{42}, \theta_{52}, \theta_{43}, \theta_{53}, \theta_{54}) = \\ &(1.570, 1.568, 0.964, 1.213, 1.561, 2.108, 1.743, 2.415, 1.690, 1.201) \end{aligned} \quad (2.3)$$

Then, by Equation 1.34, the lower triangular matrix \mathbf{B} is as follow

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.0008 & 1 & 0 & 0 & 0 \\ 0.0028 & 0.0098 & 0.9999 & 0 & 0 \\ 0.5702 & -0.4204 & -0.5275 & 0.4689 & 0 \\ 0.3502 & -0.1605 & -0.1097 & 0.3312 & 0.8543 \end{bmatrix} \quad (2.4)$$

So that, our underlying correlation matrix Σ will be

$$\Sigma = \mathbf{B}\mathbf{B}^T = \begin{bmatrix} 1 & 0.001 & 0.003 & 0.57 & 0.35 \\ 0.001 & 1 & 0.01 & -0.42 & -0.16 \\ 0.003 & 0.01 & 1 & -0.53 & -0.11 \\ 0.57 & -0.42 & -0.53 & 1 & 0.48 \\ 0.35 & -0.16 & -0.11 & 0.48 & 1 \end{bmatrix} \quad (2.5)$$

After simulating 2000 observations from copula with dependent matrix Σ and applying the copula estimation routine shown in Section 1.2, we obtain the estimated dependence parameters and show them as the following correlation matrix \mathbf{A} .

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0.5475 & 0.2977 \\ 0 & 1 & 0 & -0.3848 & -0.1583 \\ 0 & 0 & 1 & -0.4822 & -0.0705 \\ 0.5475 & -0.3848 & -0.4822 & 1 & 0.4540 \\ 0.2977 & -0.1583 & -0.0705 & 0.4540 & 1 \end{bmatrix} \quad (2.6)$$

Comparing the resulting matrix \mathbf{A} with the underlying correlation matrix Σ , all of the estimates seem close to the true value, and the underlying coefficients that are close to zero, such as 0.001, 0.003, and 0.01 in Σ , end up being shrunk to exactly zero in the resulting matrix \mathbf{A} . It indicates that the dependent model estimates the parameters well and successfully eliminates the irrelevant ones.

CHAPTER 3: RESULTS

In this section, the method in Chapter 1 is applied to estimate the dependence of indemnity amounts among the selected commodities in the U.S. federal crop insurance data. Here, the selected commodities are the top 3 prevalent crops: corn, soybean and wheat.

3.1 Data Description

The raw dataset is the U.S. crop insurance dataset on the Risk Management Agency website (RMA) ¹ which is publicly available and includes the liability amounts and indemnity amounts for each farm. The samples are composed of 368,720 farms, located in 2356 different counties in the United States. Each farm grows only one of the crop commodities and the commodity types are 188 in total. Crops grown in similar regions shared environmental conditions and may be exposed to common hazards, leading to the correlation in the indemnity amounts. The dependency among commodity types in the same region may be important for the insurance/reinsurance company interested in the Value at Risk (VaR) of specific insurance contracts covering any subset of the commodities.

For the analysis in this thesis, the crop indemnity amounts and liability amounts for each farm are further aggregated by county level, resulting in 2356 observations. Each observation (county) has the total indemnity and liability for each commodity type. In the aggregated dataset, we are given the observed indemnity Z_{ij} and liability L_{ij} amount. Table 3.1 shows the partial dataset of the aggregated liabilities amount and indemnities amount by the top 3 prevalent crops (corn, soybean, and wheat). According to the dataset, we can see that observed indemnity Z_{ij} must be zero when liability L_{ij} is zero, and Z_{ij} may be zero or positive number when L_{ij} is positive.

3.2 Marginal Models

Suppose the random variable Y_{ij} represents the realization of the indemnity amount for the j -th commodity type in county i . As the assumption of Y_{ij} is gamma distributed, we select the positive observed indemnity amounts Z_{ij} as Y_{ij} (response) with their corresponding liability amounts L_{ij}

¹ <https://www.rma.usda.gov>

Table 3.1: Dataset of liabilities and indemnities

County (i)	Liab. Corn(L_{i1})	Indem. Corn(Z_{i1})	Liab. Soy Bean(L_{i2})	Indem. Soy Bean(Z_{i2})	Liab. Wheat(L_{i3})	Indem. Wheat(Z_{i3})
1	109918	0	119156	10957	60485	15482
2	1432266	79946	1564468	297804	126290	9605
3	329826	3909	118968	0	37598	33660
4	188870	0	552515	9069	25364	11957
5	351199	0	97843	51486	0	0
6	320717	4261	43851	0	0	0
7	297396	1084	556543	66829	239153	96086
8	59827	0	0	0	0	0
9	647271	17755	1563561	153198	682013	451107
10	85054	7163	72413	31204	5738	0

(exposure) from the aggregate dataset. Then, $Y_{ij} > 0$ for each i and j allow us to fully define the density and distribution of Y_{ij} for the copula model.

The gamma distribution is commonly used in actuarial science since it is only defined for positive values, and this is appropriate for modeling the variables that represent quantities in the insurance field that cannot be negative, such as claim sizes and loss amounts. Although heavy-tail distributions are recommended for individual loss severity modeling in the recent literature, we still use the gamma distribution instead of heavy-tailed loss distributions in this research for simplicity and practical purpose. Moreover, according to the central limit theorem, the skewness in the distribution of aggregated indemnity amount will reduce while the loss amounts for more and more farms are aggregated, thus gamma distribution may be a more suitable choice for the model.

Since the indemnity amount Y_{ij} follows gamma distribution and the expected indemnity amount can be exposed using the corresponding liability amount L_{ij} , the marginal model is defined by GLM with gamma distributed dependent variable. The right panel in Figure 3.1 shows that the liability amounts and the indemnity amounts have a positive relationship at the county level after log transformation.

Therefore, we assume a log link, and the marginal model is defined as

$$\log(\mathbb{E}(Y_{ij})) = \beta_0 + \beta_1 \log(L_{ij}). \quad (3.1)$$

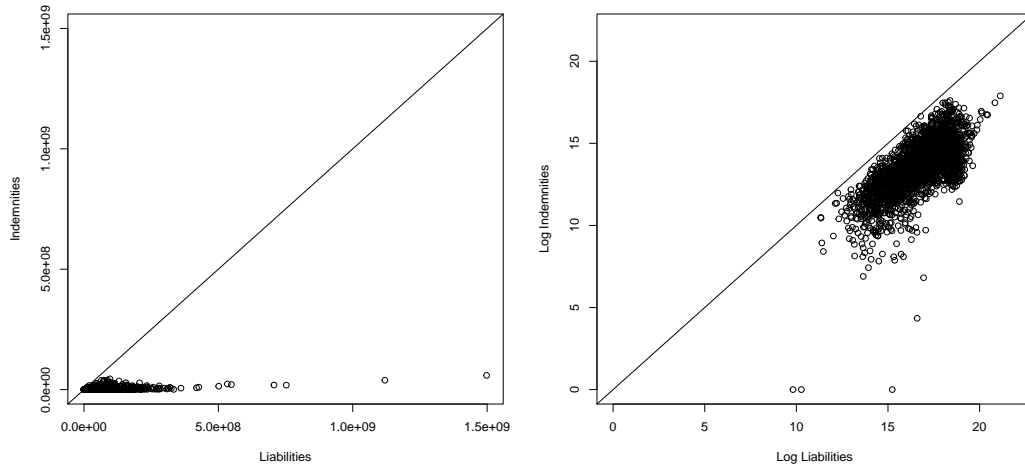


Figure 3.1: Scatter plot of liabilities and indemnities at the county level

And the estimated coefficients of the model for each commodity type are shown in Table 3.2.

Table 3.2: Coefficient estimates for the marginal gamma model

	Corn		Soybean		Wheat	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	0.8038	0.3750	2.1346	0.2531	0.1038	0.2860
LogExposurePos	0.7738	0.0235	0.6753	0.0161	0.8445	0.0213

3.3 Dependence Model

Using the marginals and the dependence model described in Chapter 1, we obtain estimates for the dependence parameters of the selected commodities shown in Table 3.3. These estimates are obtained using the tuning parameter that minimizes the BIC score, as shown in Figure 3.2.

Table 3.3: Correlation of selected commodities

	Corn	Soybean	Wheat
Corn	1		
Soybean	0.4660	1	
Wheat	0.1821	0.1923	1

In Figure 3.2, the correlation estimates are shown for different log lambda in the left penal. Although there are only 3 correlations, we can still observe that more coefficients are shrunk to zero when lambda increases. The BIC scores corresponding to different log lambda values are

displayed in the upper right panel, and the vertical line indicates the value of log lambda that minimized the BIC score. Here, $\log(\lambda) = -9.21034$ leads to the minimum BIC and will be set as the optimal tuning parameter to estimate coefficients. The smallest eigenvalue of the correlation matrix Σ with the corresponding log lambda is shown in the bottom right panel, where the panel indicates that the correlation matrix for the optimal tuning parameter is positive semi-definite since all of its eigenvalues are non-negative.

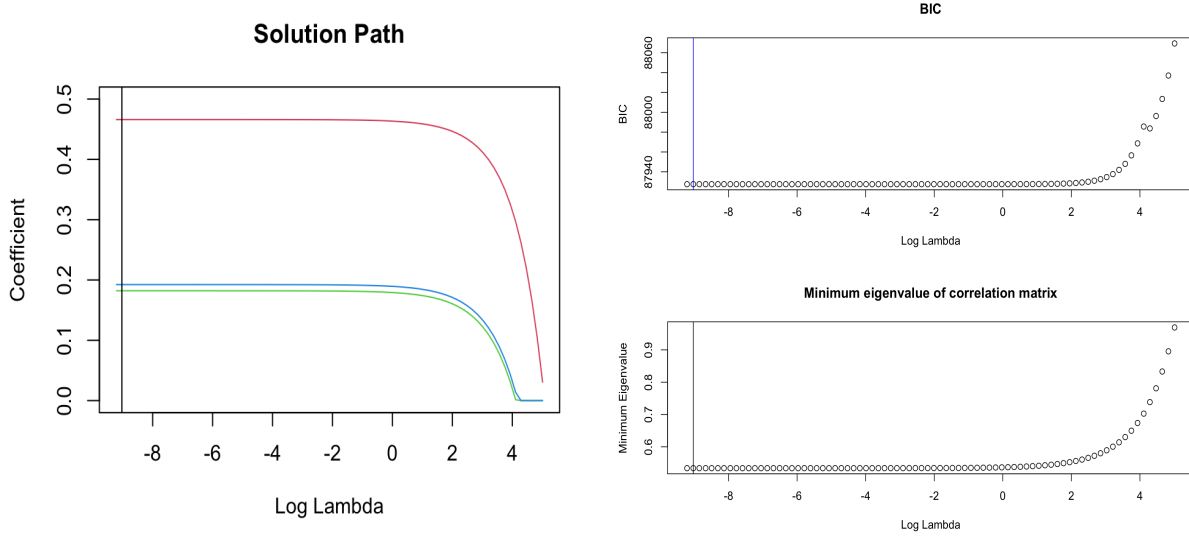


Figure 3.2: Solution path and optimal tuning parameter for selected commodities

In the 3-dimensional copula case, with only three dependent parameters, there is sufficient data to estimate the parameters accurately. Thus, the lasso penalty might not be as crucial, and the optimal tuning parameter may not shrink any coefficient to exactly zero, as depicted in Figure 3.2. However, in higher-dimensional copula cases, the model may heavily rely on regularization (L1 penalty) to deal with sparse data and prevent overfitting, leading to more coefficients being shrunk toward zero, ultimately resulting in a positive semi-definite correlation matrix for the optimal tuning parameter.

3.4 Value at Risk (VaR)

In order to demonstrate the practical application of the estimated model, let's consider a fictional insurance contract that covers a selection of commodities, including corn, soybean, and wheat.

Suppose the correlations among the commodity type are shown in Table 3.3, and the marginal models are defined as equation (3.1) with coefficients in Table 3.2. Furthermore, the log liability amounts for each commodity are 10.

Using the marginal models and the given log liability amounts, the distribution of the indemnity amount for each commodity may be known. And further, the indemnity amount can be randomly sampled using the simulation from copula with correlations shown in Table 3.3 and inverse cdf transformation. In other words, random indemnity amounts can be simulated using the copula and marginal models based on the dependence structure for the selected commodities. The simulation is repeated $B = 100,000$ times, and the indemnity amounts are aggregated each run to calculate the Value at Risk for the insurance portfolio losses. To compare the Value at Risk for the insurance coverage under the dependence and independence model, this simulation process may be applied to an independence model using an identity matrix as the correlation matrix.

Once the simulations are complete, the resulting aggregate indemnity amounts from the two models can be compared using histograms. Figure 3.3 displays the histograms for the dependence model and the independence model, and the vertical line indicates the 95th percentile for each case. It can be observed that the dependence model results in a more skewed distribution of aggregate indemnity amounts and a higher 95th percentile, indicating that the dependence among commodity types can impact the Value at Risk for insurance coverage. The 95th percentile is 42,255 for the dependent case, which corresponds to the Value at Risk at a 95% confidence level, indicating there is a 5% chance that the losses will exceed 42,255. In contrast, the 95th percentile is 37,683 under independent circumstances.

To gain a more comprehensive understanding of the insurance portfolio's risk, we calculate the Conditional Value at Risk (CVaR), also known as Tail Value at Risk. CVaR is the expected loss given that the loss exceeds the VaR level, it offers more information about the potential severity of extreme events. The CVaR at a 95% confidence level is 53,220 for the dependent case, and 46,086 for the independent case. For this fictional insurance contract, both the Conditional Value at Risk (CVaR) and Value at Risk (VaR) turn out to be larger under dependence. This observation

reminds us that when variables are dependent, we might be especially cautious about the potential for substantial losses.

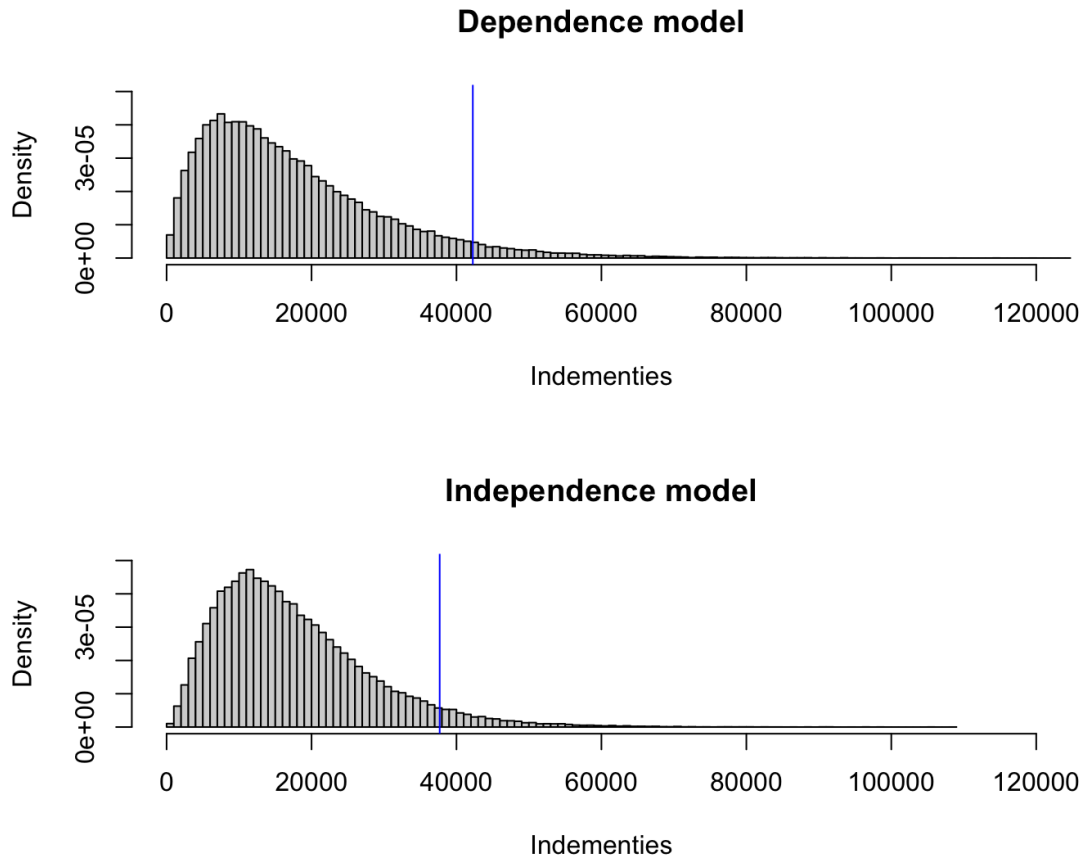


Figure 3.3: Simulated density of indemnities for a fictional insurance coverage

CHAPTER 4: CONCLUSION

In this thesis, the algorithm estimates the normal copula parameters of the composite pairwise likelihood with the L1 penalty, which is used to induce sparsity into the resulting correlation matrix and avoid overfitting. Inducing sparsity allows the dependence parameters to be estimated even in high-dimensional problems, thus it is useful when the number of dependence parameters among the p commodity types is larger than the sample size n . In the estimation approach, the majorization-minimization is used to minimize the composite likelihood function iteratively since the function with penalty has complex constraints and is difficult to optimize directly.

In practical application, the resulting dependence model can be used to calculate Value at Risk for an insurance contract covering any subset of commodities. The Value at Risk is a crucial measure for insurance companies to facilitate risk assessment and ensure sufficient capital reserves for potential losses. With the estimated dependence model, the distribution of aggregated indemnity amount for the subset of commodities can be simulated using marginal and copula models, and Value at Risk of indemnity amount covered by contract can be further computed. According to result in Section 3.4, insurance contract with positive dependencies among commodities may have higher Value at Risk than one with independence model.

The thesis focuses on using 3 dimensions as an example to illustrate the dependence model and the extension to the high-dimensional model is shown in the Appendix. With the three-dimensional setting, the complex theoretical functions and core concepts may be simplified and easily understood. There are some limitations that may be addressed in future work. One is the approach does not use the true distribution of correlation matrices in Section 1.3, thus it is possible to result in an invalid correlation matrix with non-positive semi-definiteness. Another is the approach is restricted in normal copula cases. For future research, apply the alternative parameterization using Θ to the algorithm. Moreover, the approach explained in this thesis may be extended to other copula families. Last, we only set liability amounts as exposure for modeling the expected indemnity amount in the thesis, however, if there are more explanatory variables available, including them as exposures may help improve the marginal model.

BIBLIOGRAPHY

- Nelsen, R.B. *An Introduction to Copulas*; Vol. 139, Springer Science & Business Media, 1999.
- Hofert, M.; Kojadinovic, I.; Mächler, M.; Yan, J. *Elements of Copula Modeling with R*; Springer, 2018.
- Czado, C. *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*; Springer, 2019.
- Frees, E.W.; Valdez, E.A. Understanding relationships using copulas. *North American Actuarial Journal* **1998**, 2, 1–25.
- Nikoloulopoulos, A.K.; Karlis, D. Regression in a copula model for bivariate count data. *Journal of Applied Statistics* **2010**, 37, 1555–1568.
- Shi, P.; Emiliano, V. Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal* **2014**, 2014, 159–179.
- Tibshirani, R. Regression shrinkage and selection via the lasso. *Statistics and Computing* **1996**, 58, 267–288.
- Yuan, M.; Lin, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B (Methodological)* **2006**, 68, 49–67.
- Yang, Y.; Zou, H. A fast unified algorithm for solving group-lasso penalized learning problems. *Statistics and Computing* **2015**, 25, 1129–1141.
- Qian, W.; Yang, Y.; Zou, H. Tweedie's compound poisson model with grouped elastic net. *Journal of Computational and Graphical Statistics* **2016**, 25, 606–625.
- Frees, E.W.; Gao, J.; Rosenberg, M.A. Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal* **2011**, 15, 377–392.
- Frees, E.W. Frequency and severity models. In *Predictive Modeling Applications in Actuarial Science*; Frees, E.W.; Meyers, G.; Derrig, R.A., Eds.; Cambridge University Press: Cambridge, 2014.
- Zhao, Y.; Joe, H. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* **2005**, 33, 335–356.
- Song, P.X.K.; Fan, Y.; Kalbfleisch, J.D. Maximization by parts in likelihood inference. *Journal of the American Statistical Association* **2005**, 100, 1145–1158.
- Joe, H. *Dependence Modeling with Copulas*; CRC Press, 2014.
- Pourahmadi, M.; Wang, X. Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor. *Statistics and Probability Letters* **2015**, 106, 5–12.
- Frees, E.W.; Lee, G.Y.; Yang, L. Multivariate frequency-severity regression models in insurance. *Risks* **2016**, 2016.

APPENDIX

Assumptions of General Case (p dimensions)

Consider the general problem, where we are given observations of a continuous multivariate response Z_{ij} , and liability (exposure) amounts L_{ij} . The observations are given in two matrices:

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1p} \\ L_{21} & L_{22} & \cdots & L_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \cdots & L_{np} \end{bmatrix}. \quad (4.1)$$

In practice, we may observe a matrix of explanatory variables in addition to the liabilities. Yet, for simplicity of the illustration, we assume in this paper that there are no additional explanatory variables observed. The principle may be easily extended to the case with more covariates (explanatory variables). We assume that Z_{ij} holds the value of zero when L_{ij} is zero. Yet, this is not the only case when Z_{ij} is zero. Typically, the value Z_{ij} may be either zero or a positive number with some probability, say

$$Pr(R_{ij} = 1 | L_{ij} > 0) = \mathbb{E}[R_{ij} | L_{ij} > 0], \quad (4.2)$$

where R_{ij} may be thought of as an indicator of the response being positive, given a positive liability. This is a typical situation arising in insurance claims modeling, where the data are unbalanced due to the response being undefined when the liability amount L_{ij} is zero, and given a positive L_{ij} the response is observed. In practice, the frequency of insurance claims may also be observed, resulting in the modeling of average severities, as opposed to the total severity. An example of where the frequencies are observed is the article by Frees *et al.* (2016). The method proposed in this paper can be extended to this situation as well. Yet, for now let's assume we are interested in modeling the total severity Z_{ij} directly. We hope to understand the dependence structure of the observations of the Z_{ij} response. For this, we decompose the response, so that

$$Z_{ij} = I(L_{ij} > 0) \cdot R_{ij} \cdot Y_{ij}. \quad (4.3)$$

Here, we assume that the indemnity amount Y_{ij} satisfies $Y_{ij} > 0$ for each i and j . This allows the density of Y_{ij} to be well defined. We imagine that there is a binary random variable R_{ij} for each indemnity amount Y_{ij} , that holds the value 1 when the observed indemnity amount is positive. We also assume that R_{ij} and Y_{ij} are independent. We may define the observed indemnity amount as the product of Y_{ij} and R_{ij} , so that $R_{ij} \cdot Y_{ij}$ are the observed indemnity amounts. Notice that the following relationship holds:

$$Z_{ij} = \begin{cases} Y_{ij} & \text{if } R_{ij} = 1 \text{ and } L_{ij} > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4.4)$$

The liability amount L_{ij} may also be zero or a positive number. For this reason, the observed response Z_{ij} is positive only if the liability is positive, and zero (in fact undefined) otherwise. We consider the general situation, where the marginal models are either given, or may be estimated from data separately using the Inference for Margins (IFM) approach, described in books and articles on copula modeling, such as Joe (2014). Thus, we assume that we know the parametric form and parameters of the distributions $F_{i1}(y_{i1}), F_{i2}(y_{i2}), \dots, F_{ip}(y_{ip})$, as well as the densities $f_{i1}(y_{i1}), f_{i2}(y_{i2}), \dots, f_{ip}(y_{ip})$. In other words, we are given marginal models dependent on the liability amounts L_{ij} so that

$$F_{ij}(y) = F(y; L_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, p. \quad (4.5)$$

Using these, we hope to find a joint distribution among p variables, $f(y_{i1}, y_{i2}, \dots, y_{ip})$. This task may be accomplished using a copula. Here, we make an assumption that there exists a multivariate normal copula C of p -dimension, linking the marginal distributions, so that

$$C(u_1, \dots, u_p) = \Phi_p(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p); \Sigma), \quad (4.6)$$

where Φ_p is the cumulative distribution function for a p -dimensional multivariate normal distribution

with zero means and correlation matrix Σ , and Φ^{-1} are the inverse cumulative distribution functions for standard normal distributions. The density for the copula is

$$c(u_1, \dots, u_p; \Sigma) = \frac{\partial^p}{\partial u_1 \dots \partial u_p} C(u_1, \dots, u_p; \Sigma). \quad (4.7)$$

The joint density is then given by

$$f(y_{i1}, y_{i2}, \dots, y_{ip}) = c(F_{i1}(y_{i1}), F_{i2}(y_{i2}), \dots, F_{ip}(y_{ip}); \Sigma) \cdot f_{i1}(y_{i1}) \cdot f_{i2}(y_{i2}) \dots f_{ip}(y_{ip}). \quad (4.8)$$

Note that by separating Y_{ij} from Z_{ij} by factoring out R_{ij} and L_{ij} , we may assume for every observation i , there exists such a density. Because we are assuming the marginal models are given, we may assume the input for the copula estimation routine is a matrix of uniformly distributed values u_{ij} , and indicators $r_{ij}^* = r_{ij} \cdot I(L_{ij} > 0)$ where

$$u_{ij} = \begin{cases} F_i(z_{ij}) & R_{ij} = 1, \text{ and } L_{ij} > 0 \\ 0 & R_{ij} = 0, \text{ or } L_{ij} = 0, \end{cases} \quad (4.9)$$

and

$$r_{ij}^* = \begin{cases} 1 & R_{ij} = 1, \text{ and } L_{ij} > 0 \\ 0 & R_{ij} = 0, \text{ or } L_{ij} = 0. \end{cases} \quad (4.10)$$

Shrinkage Estimation of Copula Parameters

In order to estimate the dependence parameters in the matrix Σ , we use the composite likelihood approach. Specifically, we minimize the composite negative log-likelihood function

$$\ell(\Sigma) = - \sum_{i=1}^n \sum_{\substack{j>k \\ r_{ij}^*=1 \\ r_{ik}^*=1}} \frac{1}{m_i - 1} \left[\log c(F_{ij}(y_{ij}), F_{ik}(y_{ik}); \rho_{jk}) + \log f_j(y_{ij}) + \log f_k(y_{ik}) \right], \quad (4.11)$$

where c is a bivariate normal copula, and the weight for observation i , is m_i and is obtained by

$$m_i = \sum_{j=1}^p r_{ij}^*, \quad (4.12)$$

or in other words, the number of non-zero responses for observation i . Note that given the marginal distributions F_{ij} for each $i = 1, \dots, n, j = 1, \dots, p$, the likelihood in Equation (4.11) is a function in the dependence parameters only. Note that Σ is symmetric.

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}. \quad (4.13)$$

If p is large, then the problem is high-dimensional, since the number of dependence parameters in the matrix Σ is exactly $\binom{p}{2}$. In order to automatically reduce the number of non-zero dependence parameters, we use a shrinkage approach. For this, let

$$\rho_{jk} = \frac{\tan^{-1}(\psi_{jk})}{\pi/2}, \quad -\infty < \psi_{jk} < \infty, \quad (4.14)$$

which ensures that for any ψ_{jk} , the resulting dependence parameter satisfies $-1 < \rho_{jk} < 1$. We then attempt to minimize the penalized negative log likelihood

$$\ell_P(\Psi) = \ell(\Sigma) + \lambda \sum_{\substack{j > k \\ r_{ij}^* = 1 \\ r_{ik}^* = 1}} |\psi_{jk}|, \quad (4.15)$$

where

$$\Psi = \begin{bmatrix} 1 & \psi_{12} & \cdots & \psi_{1p} \\ \psi_{21} & 1 & \cdots & \psi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{p1} & \psi_{p2} & \cdots & 1 \end{bmatrix}. \quad (4.16)$$

The problem boils down to estimating the parameters in the Ψ matrix, some of which may end up being zero due to the L_1 penalty on the likelihood. This may be done using coordinate descent. In this thesis, the tuning parameter λ is to be determined by minimizing the BIC score after solving the problem for a set of λ values. In order to estimate the dependence parameters, let

$$\Phi_{ijk} = -\frac{1}{m_i - 1} \left[\log c(F_{ij}(y_{ij}), F_{ik}(y_{ik}); \rho_{jk}) + \log f_{ij}(y_{ij}) + \log f_{ik}(y_{ik}) \right], \quad (4.17)$$

so that

$$\ell(\Sigma) = \sum_{i=1}^n \sum_{\substack{j>k \\ r_{ij}^*=1 \\ r_{ik}^*=1}} \Phi_{ijk}. \quad (4.18)$$

Here, we assume that c is the density of a bivariate normal copula. This assumption allows us to write the following:

$$\Phi_{ijk} = -\frac{1}{m_i - 1} \left[\log \left(\frac{1}{\sqrt{1 - \rho_{jk}^2}} \right) - \frac{(a_i^2 + b_i^2)\rho_{jk}^2 - 2a_i b_i \rho}{2(1 - \rho_{jk}^2)^2} + \log f_{ij}(y_{ij}) + \log f_{ik}(y_{ik}) \right], \quad (4.19)$$

where

$$\begin{aligned} a_i &= \sqrt{2} \operatorname{erf}^{-1}(2u_{ij} - 1), & u_{ij} &= F_{ij}(y_{ij}) \\ b_i &= \sqrt{2} \operatorname{erf}^{-1}(2u_{ik} - 1), & u_{ik} &= F_{ik}(y_{ik}), \end{aligned} \quad (4.20)$$

with

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2) dt. \quad (4.21)$$

Furthermore, remember that we have parametrized the dependence parameters so that we have $\rho_{jk} = 2 \tan^{-1}(\psi_{jk})/\pi$. In order to obtain the gradient and hessian of the likelihood in terms of the parameters ψ_{jk} , consider

$$\frac{\partial \Phi_{ijk}}{\partial \psi_{jk}} = \frac{\partial \Phi_{ijk}}{\partial \rho_{jk}} \frac{\partial \rho_{jk}}{\partial \psi_{jk}}. \quad (4.22)$$

The first and second terms in the product may be obtained by differentiation.

$$\frac{\partial \Phi_{ijk}}{\partial \rho_{jk}} = -\frac{1}{m_i - 1} \left[\frac{\rho_{jk}}{1 - \rho_{jk}^2} + \frac{a_i b_i \rho_{jk}^2 - (a_i^2 + b_i^2) \rho_{jk} + a_i b_i}{(1 - \rho_{jk}^2)^2} \right] \quad (4.23)$$

$$\frac{\partial \rho_{jk}}{\partial \psi_{jk}} = \frac{2}{\pi(1 + \psi_{jk}^2)}. \quad (4.24)$$

Furthermore, we have

$$\frac{\partial^2 \Phi_{ijk}}{\partial \psi_{jk}^2} = \frac{\partial^2 \Phi_{ijk}}{\partial \rho_{jk}^2} \left(\frac{\partial \rho_{jk}}{\partial \psi_{jk}} \right)^2 + \frac{\partial \Phi_{ijk}}{\partial \rho_{jk}} \left(\frac{\partial^2 \rho_{jk}}{\partial \psi_{jk}^2} \right), \quad (4.25)$$

where

$$\begin{aligned} \frac{\partial^2 \Phi_{ijk}}{\partial \rho_{jk}^2} = -\frac{1}{m_i - 1} & \left[\frac{1 + \rho_{jk}^2}{(1 - \rho_{jk}^2)^2} + \frac{\{2a_i b_i \rho_{jk} - (a_i^2 + b_i^2)\} (1 - \rho_{jk}^2)^2}{(1 - \rho_{jk}^2)^4} \right. \\ & \left. + \frac{4 \{a_i b_i \rho_{jk}^2 - (a_i^2 + b_i^2) \rho_{jk} + a_i b_i\} (1 - \rho_{jk}^2) \rho_{jk}}{(1 - \rho_{jk}^2)^4} \right], \end{aligned} \quad (4.26)$$

and

$$\frac{\partial^2 \rho_{jk}}{\partial \psi_{jk}^2} = -\frac{4\psi_{jk}}{\pi(1 + \psi_{jk}^2)^2}. \quad (4.27)$$

Denote the flattened vector of parameters by a $\binom{p}{2}$ dimensional vector

$$\boldsymbol{\psi} = (\psi_{21}, \psi_{31}, \dots, \psi_{p1}, \psi_{32}, \psi_{42}, \dots, \psi_{p2}, \dots, \psi_{p,p-1})^T. \quad (4.28)$$

The Hessian matrix has dimension $\binom{p}{2} \times \binom{p}{2}$, and is diagonal. For simplicity, we denote the entries of the Hessian matrix using j and k , the Hessian matrix \mathbf{H}_ψ has entries

$$\mathbf{H}_\psi(j, k) = \sum_{i=1}^n \frac{\partial^2 \Phi_{ijk}}{\partial \psi_{jk}^2}, \quad j > k, \quad (4.29)$$

where in the three dimensional example, there are only three possible values for the j and k indices: $(2, 1)$, $(3, 1)$, and $(3, 2)$. Denote the gradient vector by

$$\mathbf{U}_\psi(j, k) = \sum_{i=1}^n \frac{\partial \Phi_{ijk}}{\partial \psi_{jk}}, \quad j > k. \quad (4.30)$$

Remember that j and k are indices for the dependence parameter matrix, not the Hessian matrix itself. Also remember that the off diagonal elements of \mathbf{H}_ψ are zero. We take a second order Taylor series approximation to the unpenalized negative log-likelihood function in Equation (4.18) around a given point $\tilde{\boldsymbol{\psi}}$.

$$\ell_Q(\boldsymbol{\psi}) = \ell(\tilde{\boldsymbol{\psi}}) + \tilde{\mathbf{U}}_\psi^T(\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}}) + \frac{1}{2}(\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}})^T \tilde{\mathbf{H}}_\psi(\boldsymbol{\psi} - \tilde{\boldsymbol{\psi}}), \quad (4.31)$$

Then, we would like to minimize the following penalized objective function:

$$P_Q(\boldsymbol{\psi}) = \ell_Q(\boldsymbol{\psi}) + \lambda \sum_{\substack{j>k \\ R_{ij}=1 \\ R_{ik}=1}} |\psi_{jk}|. \quad (4.32)$$

For this, we use the majorization minimization principle. First, define

$$\mathbf{U}_Q = \nabla \ell_Q(\boldsymbol{\psi}) \quad (4.33)$$

$$\mathbf{H}_Q = \nabla^2 \ell_Q(\boldsymbol{\psi}), \quad (4.34)$$

where the entries of \mathbf{U}_Q are given by

$$\begin{aligned} U_Q(j, k) &= \tilde{\mathbf{U}}_{\boldsymbol{\psi}}(j, k) + \tilde{\mathbf{H}}_{\boldsymbol{\psi}}(j, k)(\psi_{jk} - \tilde{\psi}_{jk}) \\ &= \left(\sum_{i=1}^m \frac{\partial \tilde{\Phi}_{ijk}}{\partial \psi_{jk}} \right) + \left(\sum_{i=1}^m \frac{\partial^2 \tilde{\Phi}_{ijk}}{\partial \psi_{jk}^2} \right) (\psi_{jk} - \tilde{\psi}_{jk}), \end{aligned} \quad (4.35)$$

and

$$\mathbf{H}_Q(j, k) = \tilde{\mathbf{H}}_{\boldsymbol{\psi}}(j, k) = \sum_{i=1}^n \frac{\partial^2 \tilde{\Phi}_{ijk}}{\partial \psi_{jk}^2}, \quad (4.36)$$

where the simplicity is due to the fact that $\mathbf{H}_{\boldsymbol{\psi}}$ is diagonal. Notice that because of the simplicity, \mathbf{H}_Q is also diagonal. Given $\ell_Q(\boldsymbol{\psi})$, consider updating current estimates $\check{\boldsymbol{\psi}}$. The new parameters $\check{\boldsymbol{\psi}}^{(new)}$ is given by solving

$$\check{\boldsymbol{\psi}}^{(new)} = \arg \min_{\boldsymbol{\psi}} \ell_Q(\check{\boldsymbol{\psi}}) + \tilde{\mathbf{U}}_Q^T(\boldsymbol{\psi} - \check{\boldsymbol{\psi}}) + (\boldsymbol{\psi} - \check{\boldsymbol{\psi}})^T \tilde{\mathbf{H}}_Q(\boldsymbol{\psi} - \check{\boldsymbol{\psi}}) + \lambda \sum_{\substack{j>k \\ r_{ij}^*=1 \\ r_{ik}^*=1}} |\psi_{jk}|. \quad (4.37)$$

The solution for this problem is given by

$$\check{\psi}_{jk}^{(new)} = \frac{\left(\tilde{\gamma}_{\boldsymbol{\psi}} \check{\psi}_{jk} - \tilde{\mathbf{U}}_Q(j, k) \right) \left(1 - \frac{\lambda}{|\tilde{\gamma}_{\boldsymbol{\psi}} \check{\psi}_{jk} - \tilde{\mathbf{U}}_Q(j, k)|} \right)}{\tilde{\gamma}_{\boldsymbol{\psi}}}, \quad (4.38)$$

where $\tilde{\gamma}_{\boldsymbol{\psi}}$ is the largest eigenvalue of $\tilde{\mathbf{H}}_Q$.