Technische Universität München

Zentrum Mathematik

Modelling Dependence between Loss Triangles using Copula and DVine Constructions

Diplomarbeit

von

Matthias Lindinger

Themenstellerin:	Prof. Dr. Claudia Czado
Betreuer:	Vinzenz Erhardt, Stefan Antes
Abgabetermin:	01.12.2010

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 01.12.2010

Acknowledgments

Taking this opportunity I would like to thank all those people who guided and accompanied me over the last years of my studies and without whom this thesis would have never been realized in the existing form. First of all my thanks go to Prof. Claudia Czado for giving me the opportunity to write this thesis at the Chair of Mathematical Statistics and for the comfortable working atmosphere. Furthermore, I want to express my sincere gratitude to my supervisors Dr. Vinzenz Erhardt and Stefan Antes for all their helpful advice and encouragement. I am indebted to them for the many pleasant and fruitful discussions over the last six months. Last but not least I must not forget to thank my family for simply everything, especially for their unlimited support throughout my whole time as a student. My special thanks go to my friends who made university life much more enjoyable.

Contents

1 Introduction **1**

2 Gaussian Copula Model **3**

2.1 Notation . . . 3
2.2 The Chain Ladder Method . . . 4
2.3 Gaussian Copula Model for Loss Triangle Dependence . . . 6
2.4 Model Summary . . . 7
2.5 Dependence Properties . . . 8
2.6 Parameter Estimation . . . 13
2.6.1 Heuristic Parameter Estimation . . . 14
2.6.2 Maximum Likelihood Estimation . . . 16
2.6.3 Stepwise Maximum Likelihood Estimation . . . 20
2.6.4 Estimation Summary . . . 22
2.7 Loss Triangle Simulation . . . 27
2.8 Forecasting future payments . . . 28

3 Application to Real Insurance Data **30**

3.1 Data . . . 30
3.2 Shapiro-Wilk test for normality . . . 30
3.3 Data Transformation . . . 31
3.4 Simulation Study . . . 39
3.4.1 General Setup . . . 39
3.4.2 Dependence structure . . . 41
3.4.3 Discussion of the Results . . . 42
3.5 Parameter Estimation . . . 47

4 D-Vine Model for Loss Triangle Dependence **49**

4.1 Copulas . . . 49
4.2 PCC's for Multivariate Distributions . . . 50
4.2.1 Example in Dimension 5 . . . 53
4.2.2 PCC with n Variables . . . 53
4.3 Simulation from a Pair-Copulae decomposed Model . . . 53
4.3.1 Introduction of h-functions . . . 54
4.3.2 Sampling from a D-Vine . . . 54
4.4 Inference for a specified Pair-Copula Decomposition . . . 54

4.5 Model Selection 56
4.6 D-Vine Model for Loss Triangle Dependence 57
4.6.1 Data Preparation 58
4.6.2 Model Formulation 59
4.6.3 Tests for Non-Nested Model Selection 60
4.7 Application to real Insurance Data 62
4.7.1 Determination of Top-Tree Structure 62
4.7.2 Pair-Copula Estimation 63
4.8 Data Forecast and Fit 68

5 Model comparison 71
5.1 Data Fit 71
5.2 Forecast of Future Payments 76
5.2.1 Forecast based on Real Data 76
5.2.2 Forecast based on Simulated Data 80

Chapter 1

Introduction

Loss reserving is one of the most critical issues in actuarial mathematics. There is always a certain period of time between the date where a financial claim occurs and the date of it's regulation, since the claim needs to be reported to the insurance company and the claim needs to be regulated, which can, especially for large claims, take a relatively long time. While this period is quite short for most lines of business (i.e. 1 - 2 months) it can take several years in the liability insurance. The main reason for this is the fact, that, especially in cases where a person is hurt, it can take very long before the final amount of the financial claim is known, i.e. if the amount depends on a judicial decision or on the success of a long term medical treatment.
For each of these claims a so-called single-claim-reserve is calculated. But if the actual costs exceed this reserve the insurer faces a runoff-loss. The reserve for typical runoff-losses is called IBNER-reserve (IBNER = *incurred but not enough reserved*).
The importance of loss reserving mainly arises from two facts. First, the reserve is an important building block for calculating the premium for a certain line of business. Second the reserves form one of the main parts of the liability side of the balance sheet of the insurance company. Hence the reserves massively affect the monetary success of the company.
Due to the importance of the topic a lot of work has been done in order to find reasonable estimates for reserves for certain lines of business. The most common methods for forecasting reserves needed for future obligations are the chain ladder method, the grossing-up method, the Bornhuetter-Ferguson method and many more (see Schmidt [2001] and Mack [1997]).
In the recent years these methods have been criticised for the fact that they estimate reserves for each line of business without making any assumptions about possible dependence between loss triangles for different lines of business.
The study of copulas and their applications in statistics is a rather recent phenomenon. But, especially in the past 15 years (see Nelsen [1999]), the interest in copulas is growing rapidly and copula theory is applied in many areas of research, like i.e. financial returns, survival analysis and many more.
In 2009 the australian Piet de Jong published a paper where he applies basic copula theory to loss triangle data (de Jong [2009]). In this paper a Gaussian copula model is applied to data of an US insurance company. The dependence structure between different loss

triangles is modelled by a correlation matrix **R**.
This work will extend the model proposed by de Jong [2009] as displayed in chapter 2 and apply this theory to real insurance data consisting of 5 loss triangles in chapter 3. In addition a D-Vine model as described in Aas and H.Bakken [2009] or Czado and Min [2010] will be fitted to the data in chapter 4. A comparison of these models will be given in chapter 5.
The models presented in this work may improve the reserving practice of insurances since classic models for loss reserving do not adjust for possible dependencies in the data.

Chapter 2

Gaussian Copula Model

2.1 Notation

This section will provide the reader with the basic notation used throughout the following article.

- Loss triangle: the m different loss triangles are indexed by $l = 1, \ldots, m$, where each l stands for only one triangle.

- Observation period: the length of the time period in which the data has been observed is denoted by n.

- Accident year i: the year in which an accident, the company has to pay for, occurs is indexed with i , $i = 1, 2, \ldots, n$. Thus if $i = 1$ the accident year is the year when the data collection started, ..., if $i = n$ the accident year is the last year for which data is available.

- Development year j: Payments made for accidents in the year they occur are, by definition, settled in development year $j = 0$, payments made one year after the accident in development year $j = 1$,...,payments made in the kth year following the accident in development year $j = k$, and so on. Therefore the index j representing the development year runs from $j = 0, \ldots, n - 1$.

- Calendar year t: the calendar years are denoted by $t = 1, ..., n$, where $t = 1$ corresponds to the calendar year in which data collection started , ..., $t = n$ to the last calendar year of data collection.

- The triangle entries p_{lij} are the sum of all payments made for accidents occured in accident year i, $i = 1, \ldots, n$ in the jth year after those accidents in loss triangle l, l= $1, \ldots, m$.

This notation might seem difficult for readers who are not familiar with loss triangles. Therefore, I will use a simple example :
Consider a time period of $n = 5$ years, i.e, the period from 2005 to 2009, where payments for a certain business line (triangle) l have been obvserved.

These payments are collected in a triangle. Each row of this triangle represents one accident year i, $i = 1, \ldots, n = 5$ (note: $i = 1$ corresponds to accident year 2005, $i = 5$ corresponds to accident year 2009), while each column stands for one development year j, $j = 0, \ldots, n-1 = 4$.
Thus the first column includes payments made in the same calendar year as the occured accidents, the second column those payments made in the calendar year following the accident, and so on.
Payments p_{lij} for $t = i + j \leq n$ are observed in calendar year $t = i + j$, while those for $t = i + j > n$ lie in the future.
Aim of this work will be to find reasonable estimates for future payments.

Accident	Development year j				
Year i	0	1	2	3	4
1	p_{l10}	p_{l11}	p_{l12}	p_{l13}	p_{l14}
2	p_{l20}	p_{l21}	p_{l22}	p_{l23}	
3	p_{l30}	p_{l31}	p_{l32}		
4	p_{l40}	p_{l41}			
5	p_{l50}				

Table 2.1: Loss triangle structure

In the triangle displayed in Table 2.1 p_{l30} stands for the payments made in development year $j = 0$ (2007) for accidents occured in accident year $i = 3$ (2007) in triangle l, p_{l31} for payments made in development year $j = 1$ (2008) for accidents occured in accident year $i = 3$ (2007) in triangle l, and so on.
Payments p_{lij} with $i + j > n$ ($n = 5$) lie in the future and are still unknown.
In the remainder of this work many statements will rely on payments p_{lij} falling into the same calendar year t, t= $1, \ldots, n$. In general there are t different payments p_{lij} falling into calendar year t, namely those with $i + j = t$. All of them are located on a diagonal of the triangle. Note that unlike matrix notation this diagonal goes from the lower left entry of the payment rectangle to the upper right entry!
In the triangle displayed above those are:

- for t=1 (2005) : $p_{l10}(1 + 0 = 1 = t)$
- for t=2 (2006) : p_{l20} and $p_{l11}(2 + 0 = 1 + 1 = 2 = t)$
- for t=3 (2007) : p_{l30} , p_{l21} and p_{l12}
- for t=4 (2008) : p_{l40} , p_{l31} , p_{l22} and p_{l13}
- for t=5 (2009) : p_{l50} , p_{l41} , p_{l32} , p_{l23} and p_{l14}

2.2 The Chain Ladder Method

The most common method of forecasting future payments in loss triangles is the so called chain-ladder method (see Mack [1997]). It's technique is quite simple, but due to this and

a lack of alternatives it is applied by most companies.
The chain ladder method is typically applied to single loss triangles and therefore no assumptions on triangle dependence are to be made.
The main idea behind this method is the assumption that the relative change in future cumulative payments from one development year to the next is the same as for payments observed in the past.
From now on entry CPL_{ij} is interpreted as the cumulative paid loss for accident year i from development year 0 up to development year j:

$$CPL_{ij} := \sum_{k=0}^{j} p_{ij}.$$

Cumulated payments CPL_{ij} are observable for calendar years $i + j = t \leq n$, $i = 1, \dots, n$, $j = 0, 1, \dots n - 1$, $t = 1 \dots n$. Payments in calendar $i + j = t > n$ are not observable and hence need to be estimated.
The observable claims can be represented by the following triangle:

Accident year i	Development year j 0	1	$\dots$	k	$\dots$	$\dots$	n-1
1	$CPL_{1,0}$	$CPL_{1,1}$	$\dots$	$CPL_{1,k}$	$\dots$	$\dots$	$CPL_{1,n-1}$
2	$CPL_{2,0}$	$CPL_{2,1}$	$\dots$	$CPL_{2,k}$	$\dots$	$CPL_{2,m-1}$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$			
n-k	$CPL_{n-k,0}$	$CPL_{n-k,1}$	$\dots$	$CPL_{n-k,k}$			
$\vdots$	$\vdots$	$\vdots$					
n-1	$CPL_{n-1,0}$	$CPL_{n-1,1}$					
n	$CPL_{n,0}$						

For $i \in 1, \dots, n$ and $j \in 0, \dots, n - 1$, the random variable

$$DF_{i,j} := \frac{CPL_{i,j}}{CPL_{i,j-1}} \ , \ \forall j = 1, \dots, n-1 \ , \ DF_{i,0} := 1$$

is said to be the individual development factor of accident year i and development year j.
In other words $DF_{i,j}$ denotes the increase of the cumulated paid loss for accident year i from development year $j - 1$ up to calendar year j.
For $i + j \leq n$ these factors can easily be obtained from the data, those for $i + j > n$ can be estimated as follows:

$$\hat{DF}_j^{CL} := \frac{\sum_{i=0}^{n-j} CPL_{i,j}}{\sum_{i=0}^{n-j} CPL_{i,j-1}} = \sum_{i=1}^{n-j} \frac{CPL_{i,j-1}}{\sum_{h=1}^{n-j} CPL_{h,j-1}} DF_{i,j}.$$

Thus the estimated chain ladder factors $\hat{DF}_j^{CL}$ are a volume weighted average of the observed individual development factors $DF_{i,j}$, $i+j \leq n$. This means the model assumes that future procentual increases of $CPL_{\cdot j}$ in development year j only depend on the observed individual development factors and are the same for all i, where $i + j > n$.

In practice there are many slight modifications of how to estimate $\hat{DF}_j^{CL}$, $j = 0, \ldots, n-1$, i.e., an exclusion of the greatest and smallest value of $DF_{i,j}$,$i + j \leq n$, summation only over the past k years, k typically in {3, ... ,7} or a combination of these two methods.
From the estimated chain ladder factors prediction of future payments is straightforward:

$$\hat{CPL}_{i,j}^{CL} := CPL_{i,n-i} \prod_{k=n-i+1}^{j} \hat{DF}_k^{CL}.$$

Now the loss triangle can easily be filled and one obtains a loss square.

2.3 Gaussian Copula Model for Loss Triangle Dependence

In this section a Gaussian copula model for loss triangle dependence is introduced. Basics for this model can be found in de Jong [2009]. Discussion is based on payments and no longer on cumulative payments.
Let p_{lij} the payment for loss triangle l, $l = 1, \ldots, m$ with respect to accident year i, $i = 1, \ldots, n$ and development year j, $j = 0, \ldots, n-1$.
One basic assumption of the model is that the observed payments p_{lij}, $i + j \leq n$, $l = 1, \ldots, m$ are realizations of a distribution function F_{lj}, $P_{lij} \sim F_{lj}$, for some not yet further specified distribution function F_{lj}. Thus actual payments are treated as random variables. As indicated by the index F_{lj} it is assumed that payments in each column $j, j = 0, \ldots, n-1$ of each triangle $l, l = 1, \ldots, m$ follow their unique columnwise distribution F_{lj}. The uniformely transformed payments $F_{lj}(P_{lij})$ are modeled via a Gaussian copula as follows:

$$F_{lj}(P_{lij}) \overset{\mathrm{d}}{=} \Phi(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \tag{2.1}$$

$$\boldsymbol{\alpha}_t := \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \in \mathbb{R}^m \ , \ \boldsymbol{\alpha}_t \sim N(\mathbf{0}, \mathbf{R}), \ \mathbf{R} \in \mathbb{R}^{m \times m} \ , \ t = i + j \tag{2.2}$$

$$\epsilon_{lij} \sim N(0,1) \ \text{ i.i.d.} \ , \ c_l^2 + s_l^2 = 1 \quad , \ c_l, s_l \in [0,1] \tag{2.3}$$

The critical feature of this model is that the same $\alpha_{l,i+j} \in \mathbb{R}$ enters each and every entry P_{lij} of triangle l, $l = 1, \ldots, m$ falling into calendar year t, $t = i + j$. Further the α_{lt} ,$t = i + j$ are related for different l, $l = 1, \ldots, m$, as determined by the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$. Those triangle-specific numbers $\alpha_{1t}, \ldots, \alpha_{mt}$ are collected in the vector $\boldsymbol{\alpha}_t \in \mathbb{R}^m$,$t = 1, \ldots, n$.
The vectors $\boldsymbol{\alpha}_t$,$t = 1, \ldots, n$, are modeled as iid normal random vectors with mean $\mathbf{0}$

and correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$.
The correlation between the m different triangles is measured by $\mathbf{R}$. Component α_{lt} of $\boldsymbol{\alpha}_t$ manifests itself in the t different entries P_{lij} of triangle l falling in calendar year t, $t = i + j$.
The real numbers c_l and s_l are a tool to model the influence of triangle l on other triangles $k,\ \ k = 1, \ldots, l-1\ ,\ l+1, \ldots m$. Here $c_l\ , 0 \leq c_l \leq 1,$ stands for the "communality" of triangle l, while $s_l\ , 0 \leq s_l \leq 1$, is called "specificity" of triangle l, $c_l^2 + s_l^2 = 1$. If $c_l \approx 0$ ($s_l \approx 1$) then α_{lt} is given zero weight, meaning that movements in triangle l are mainly caused by the random term $\epsilon_{lij} \sim N(0, 1)$.
If $c_l = 1$ ($s_l = 0$) then α_{lt} is given full weight. In this case the values of $F_{l\cdot}(P_{l\cdot\cdot})$ are the same for all entries $P_{lij}, i = 1, \ldots, n\ , j = 0, \ldots, n-1,\ , i + j = t$ on the t-th same diagonal of triangle l and this constant level is correlated with movements in other triangles as indicated by the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$.
Equations (2.1), (2.2) and (2.3) spell out the joint behaviour of loss triangles. To complete the model requires marginal models for each payment $P_{lij}\ , i = 1, \ldots, n\ ,\ j = 0, \ldots, n-1\ ,\ l = 1, \ldots, m$.
This is accomplished by assuming that, for some triangle-specific strictly increasing transformation $\varphi_l : (0, \infty) \to \mathbb{R}\ ,\ l = 1, \ldots, m$, $\varphi(P_{lij})$ is normally distributed with means and variances that depend only on the triangle l, $l = 1, \ldots, m$ and development year j, $j = 0, \ldots, n-1$:

$$\varphi_l(P_{lij}) \sim N(\mu_{lj}, \sigma_{lj}^2), \tag{2.4}$$

where μ_{lj} and σ_{lj}^2 denote the mean and variance of the j-th column of the transformed loss triangle $\varphi_l(P_{lij}),\ \ l = 1, \ldots, m$.
The value of $F_{lj}(P_{lij})$ does not change under strictly increasing transformations. This implies

$$\begin{aligned} F_{lj}(P_{lij}) &\overset{\mathrm{d}}{=} \Phi\left(\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}}\right) \\ \overset{(2.1)}{\Rightarrow} \Phi\left(\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}}\right) &\overset{\mathrm{d}}{=} \Phi(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \\ \Leftrightarrow \frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}} &\overset{\mathrm{d}}{=} c_l \alpha_{l,i+j} + s_l \epsilon_{lij} \\ \Leftrightarrow \varphi_l(P_{lij}) &\overset{\mathrm{d}}{=} \mu_{lj} + \sigma_{lj}(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \end{aligned} \tag{2.5}$$

Note that $E[c_l \alpha_{l,i+j} + s_l \epsilon_{lij}] = 0$ and $\mathrm{var}(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) = 1$ and therefore $\varphi_l(P_{lij})$ has mean μ_{lj} and variance σ_{lj}^2.

2.4 Model Summary

Combining the basic model with the assumptions on the marginal behaviour of payments P_{lij} finally yields the following model :

$$F_{lj}(P_{lij}) \overset{\mathrm{d}}{=} \Phi(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \overset{\mathrm{d}}{=} \Phi\left(\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}}\right) \tag{2.6}$$

where

- F_{lj} is the distribution function of payments P_{lij} in the j-th column of loss triangle l
- the random vector $\boldsymbol{\alpha}_t := \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \in \mathbb{R}^m$, $\boldsymbol{\alpha}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$, $\mathbf{R} \in \mathbb{R}^{m \times m}$, $t = i + j$, is used to model the common behaviour of payments P_{lij} falling into the same calendar year $t, t = i + j$ and to measure dependence between triangles
- the random variable $\epsilon_{lij} \sim \mathbf{N}(0, 1)$ is a random noise term
- the real numbers c_l and s_l , $c_l, s_l \in [0, 1], c_l^2 + s_l^2 = 1$ moderate the effects of the systematic term α and the random term ϵ, respectively
- $\varphi_l(P_{lij})$ transforms payments P_{lij} in a way that $\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}} \sim N(0, 1)$ holds for each column $j, j = 0, \ldots, n-1$ of each triangle $l, l = 1, \ldots, m$ where μ_{lj} and σ_{lj} denote the mean and standard deviation of the j-th column of the transformed triangle $\varphi_l(P_{lij})$

2.5 Dependence Properties

This section will provide the reader with the basic assumptions on how dependence between loss triangles is modelled and how it is measured.

The correlation between triangles is modeled via the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$ of $\alpha_{\mathbf{t}} \in \mathbb{R}^m$. Because of the assumed time independence of $\boldsymbol{\alpha}_t$, the model implies that payments made in the same calendar year t, $t = 1, \ldots, n$, are related, but payments made in different calendar years t are independent.
With

$$\mathbf{P}_{lt} := \begin{pmatrix} P_{li_1j_1} \\ P_{li_2j_2} \\ \vdots \\ P_{li_tj_t} \end{pmatrix} \in \mathbb{R}^t ,$$

where $i_1 + j_1 = i_2 + j_2 = \cdots = i_t + j_t = t$, $\mathbf{P}_{lt}$ is a vector of length t containing all entries in triangle l in the same calendar year t or equivalently all entries on the t-th diagonal of triangle l.

In terms of the model notation, with $P_{lij} \sim F_{lj}$ for some distribution function F_{lj}, it follows:

$$F_{l\cdot}(\mathbf{P}_{lt}) := \begin{pmatrix} F_{lj_1}(P_{li_1j_1}) \\ F_{lj_2}(P_{li_2j_2}) \\ \vdots \\ F_{lj_t}(P_{li_tj_t}) \end{pmatrix} \overset{\mathrm{d}}{=} \begin{pmatrix} \Phi(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) \\ \Phi(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2}) \\ \vdots \\ \Phi(c_l\alpha_{lt} + s_l\epsilon_{li_tj_t}) \end{pmatrix} =: \Phi(c_l\mathbf{1}_t\alpha_{lt} + s_l\boldsymbol{\epsilon}_{lt}),$$

where $\boldsymbol{\epsilon}_{lt} := \begin{pmatrix} \epsilon_{li_1j_1} \\ \epsilon_{li_2j_2} \\ \vdots \\ \epsilon_{li_tj_t} \end{pmatrix}$, $i_1 + j_1 = i_2 + j_2 = \cdots = i_t + j_t = t$, $\mathbf{1}_t := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^t$

and both $F_{l\cdot}$ and Φ operate componentwise.

The random term $\boldsymbol{\epsilon}_{lt} \sim \mathbf{N}(\mathbf{0}, \mathbf{I_t})$, with $\mathbf{I_t}$ being the t-dimensional identity matrix, is independent of $\boldsymbol{\alpha}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$, $\mathbf{R} \in \mathbb{R}^{m \times m}$.

As shown before, for some triangle-specific strictly increasing transformation φ_l , $l = 1, \ldots, m$, where $\varphi_l(P_{lij}) \sim N(\mu_{lj}, \sigma_{lj}^2)$, $\varphi_l(P_{lij})$ can be written as :

$$\varphi_l(P_{lij}) \overset{\mathrm{d}}{=} \mu_{lj} + \sigma_{lj}(c_l\alpha_{l,i+j} + s_l\epsilon_{lij})$$

Thus the correlations between transformed payments in the $m = 5$ triangles can be summarized by the following theorem:

Theorem 2.1. *The dependence structure between payments P_{lij} , $l = 1, \ldots, m$, $i = 1, \ldots, n$, $j = 0, \ldots, n-1$, for which*

$$F_{lj}(P_{lij}) \overset{d}{=} \Phi(c_l\alpha_{l,i+j} + s_l\epsilon_{lij}) \overset{d}{=} \Phi\left(\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}}\right)$$

holds , can be summarized as follows:

(i) The correlation between two transformed payments $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$ in the same triangle l, $l = 1, \ldots, m$ falling in the same calendar year t, $t = i_1 + j_1 = i_2 + j_2$ equals the square of the communality c_l^2 of triangle l :

$$\mathrm{corr}(\varphi_l(P_{li_1j_1}), \varphi_l(P_{li_2j_2})) = c_l^2 + \mathbb{1}_{i_1=i_2} \cdot s_l^2 \ ,$$

where $\mathbb{1}$ denotes the indicator function.

(ii) The correlation between transformed payments $\varphi_l(P_{li_1j_1})$ and $\varphi_k(P_{ki_2j_2})$ in different triangles l and k , $l \neq k$, falling into the same calendar year t, $t = 1, \ldots, n$, $t = i_1 + j_1 = i_2 + j_2$ equals $c_lc_kr_{lk}$:

$$\mathrm{corr}(\varphi_l(P_{li_1j_1}), \varphi_k(P_{ki_2j_2})) = c_lc_kr_{lk}$$

(iii) There is no dependence between transformed payments $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$falling in different calendar years $t = i_1 + j_1$ and $t' = i_2 + j_2$, $t \neq t'$:

$$\mathrm{corr}(\varphi_l(P_{li_1j_1}), \varphi_l(P_{li_2j_2})) = 0 \ \ \forall \ i_1 + j_1 = t \neq t' = i_2 + j_2$$

Proof of Theorem 2.1:

(i) Let $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$ be two transformed entries of triangle l with $i_1 + j_1 = i_2 + j_2$. Then we have:

$$cov(\varphi_l(P_{li_1j_1}), \varphi_l(P_{li_2j_2})) \overset{2.5}{=}$$
$$cov(\mu_{lj_1} + \sigma_{lj_1}(c_l\alpha_{l,i_1+j_1} + s_l\epsilon_{li_1j_1}), \mu_{lj_2} + \sigma_{lj_2}(c_l\alpha_{l,i_2+j_2} + s_l\epsilon_{li_2j_2})) =$$
$$cov(\mu_{lj_1} + \sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}), \mu_{lj_2} + \sigma_{lj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2}))$$

where the last equality follows from the fact, that $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$ fall into the same calendar year t ($t = i_1 + j_1 = i_2 + j_2$).
Next one uses the well-known fact, that for two random variables X, Y the following equation holds:

$$\text{cov}(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$$

$$\begin{aligned} \text{Next define } X &:= \mu_{l\cdot j_1} + \sigma_{l\cdot j_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) \\ \text{and } \quad Y &:= \mu_{l\cdot j_2} + \sigma_{l\cdot j_2}(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2}) \end{aligned}$$

Then

$$\begin{aligned} E[XY] &= E\left[(\mu_{lj_1} + \sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1})) \cdot (\mu_{lj_2} + \sigma_{lj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2}))\right] \\ &= E[\mu_{lj_1} \cdot \mu_{lj_2} + \mu_{lj_1}\sigma_{lj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2}) + \mu_{lj_2}\sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) \\ &\quad +\sigma_{lj_1}\sigma_{lj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1})(c_l\alpha_{lt} + s_l\epsilon_{li_2j_2})] \\ &= \mu_{lj_1}\mu_{lj_2} + \sigma_{lj_1}\sigma_{lj_2}E\left[(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) \cdot (c_l\alpha_{lt} + s_l\epsilon_{li_2j_2})\right] \end{aligned}$$

because $E[c_l\alpha_{lt} + s_l\epsilon_{lij}] = 0 \;\forall\, i, j$, due to $\alpha_{lt} \sim N(0,1), \epsilon_{lij} \sim N(0,1)$, $c_l^2 + s_l^2 = 1$.

$$\begin{aligned} \Rightarrow\; E[XY] &= \mu_{lj_1}\mu_{l\cdot 2} + \sigma_{lj_1}\sigma_{lj_2}E\left[c_l^2\alpha_{lt}^2 + c_l\alpha_{lt}s_l\epsilon_{li_1j_1} + c_l\alpha_{lt}s_l\epsilon_{li_2j_2} + s_l^2\epsilon_{li_1j_1}\epsilon_{li_2j_2}\right] \\ &= \mu_{lj_1}\mu_{lj_2} + \sigma_{lj_1}\sigma_{lj_2} \cdot [c_l^2 + s_l^2 \cdot \mathbb{1}_{i_1=i_2}] \,, \end{aligned}$$

due $E\left[c_l\alpha_{lt}s_l\epsilon_{lij}\right] = 0$ following from the assumed independence of α and ϵ.

With $E[X] = \mu_{lj_1}$ and $E[Y] = \mu_{lj_2}$ it follows that:

$$E[X]E[Y] = \mu_{lj_1}\mu_{lj_2}$$

$$\Rightarrow \text{cov}(X,Y) = E[XY] - E[X]E[Y] = \sigma_{lj_1}\sigma_{lj_2} \cdot [c_l^2 + s_l^2 \cdot \mathbb{1}_{i_1=i_2}]$$

Thus the correlation between two transformed entries $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$ of triangle l , $l = 1,\ldots,m$ falling into the same calendar year t , $t = 1,\ldots,n$ equals c_l^2, the square of the communality of triangle l. As denoted by the indicator function $\mathbb{1}_{i_1=i_2} \cdot s_l^2$ the correlation of any transformed payment with itself equals $c_l^2 + s_l^2 = 1$ following from the definitions of c_l and s_l.

(ii) Let $X := \mu_{lj_1} + \sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) = \varphi_l(P_{li_1j_1})$
and $Y := \mu_{kj_2} + \sigma_{kj_2}(c_k\alpha_{kt} + s_l\epsilon_{ki_2j_2}) = \varphi_k(P_{ki_2j_2})$

$\Rightarrow$ $E[X] = \mu_{lj_1}$ and $E[Y] = \mu_{kj_2}$

$$E[XY] = E[\mu_{lj_1}\mu_{kj_2} + \mu_{lj_1}\sigma_{kj_2}(c_k\alpha_{kt} + s_k\epsilon_{ki_2j_2}) + \mu_{kj_2}\sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1}) +$$
$$+ \sigma_{lj_1}\sigma_{kj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1})(c_k\alpha_{kt} + s_k\epsilon_{ki_2j_2})]$$

- $E[\mu_{lj_1}\mu_{kj_2}] = \mu_{lj_1}\mu_{kj_2}$
- $E[\mu_{lj_1}\sigma_{kj_2}(c_k\alpha_{kt} + s_k\epsilon_{ki_2j_2}] = E[\mu_{kj_2}\sigma_{lj_1}(c_l\alpha_{lt} + s_l\epsilon_{li_1j1})] = 0$

  following from $\alpha_{lt} \sim N(0,1)$ and $\epsilon_{lij} \sim N(0,1)$, $l = 1,\ldots,m$.

- $E[\sigma_{lj_1}\sigma_{kj_2}(c_l\alpha_{lt} + s_l\epsilon_{li_1j_1})(c_k\alpha_{kt} + s_k\epsilon_{ki_2j_2})] =$

  $E[\sigma_{lj_1}\sigma_{lj_2}(c_lc_k\alpha_{lt}\alpha_{kt} + c_ls_k\alpha_{lt}\epsilon_{ki_1j_1} + c_ks_l\alpha_{kt}\epsilon_{li_1j_1} + s_ls_k\epsilon_{li_1j_1}\epsilon_{ki_2j_2})] =$

  $\sigma_{lj_1}\sigma_{kj_2}c_lc_kE[\alpha_{lt}\alpha_{kt}]$

- $E[\alpha_{lt}\alpha_{kt}] = cov(\alpha_{lt} , \alpha_{kt}) = r_{lk}$, $R = (r)_{lk}$, $E[\alpha_{lt}] = E[\alpha_{kt}] = 0$

$$\Rightarrow E[XY] = \mu_{lj_1}\mu_{kj_2} + \sigma_{lj_1}\sigma_{kj_2}c_lc_kr_{lk}$$
$$\Rightarrow cov(X,Y) = \sigma_{lj_1}\sigma_{kj_2}c_lc_kr_{lk}$$

Since $Var(X) = \sigma^2_{lj_1}$ and $Var(Y) = \sigma_{kj_2}$ it follows that

$$corr(X, Y) = c_l c_k r_{lk}$$

This means that the correlations induced by $\mathbf{R}$ are moderated by the communalities $c_l,\ l = 1, \ldots, m$. Thus high inner triangle variation in any triangle l, which is indicated by s_l being close to 1 leads to low dependence with other triangles $l',\ l' \neq l$.

(iii) The proof of Theorem 2.1 (iii) follows directly from the model definition:
Remember:

$$F_{lj}(P_{lij}) \stackrel{\mathrm{d}}{=} \Phi(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}),\ t = i + j \quad , \quad c_l^2 + s_l^2 = 1 \ ,$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R}), \qquad \epsilon_{lij} \sim N(0, 1) \ ,$$

where the vectors $\boldsymbol{\alpha}_t \in \mathbb{R}^m, t = 1, \ldots, n$ are modelled as $i.i.d$ normal random vectors with mean $\mathbf{0}$ and correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$.
This means that $\boldsymbol{\alpha}_t,\ t = i_1 + j_1$ is independent from $\alpha_{t'}$, $t' = i_2 + j_2,\ t' \neq t$.
Obviously even stronger independence assumptions hold for ϵ :

$$\epsilon_{li_1j_1} \perp \epsilon_{li_2j_2} \quad \forall \quad i_1 \neq i_2 \quad \text{or } j_1 \neq j_2 \ ,$$

indicating that each ϵ_{lij} is independent of all other $\epsilon_{li'j'}$ except from itself .

$\Rightarrow \operatorname{corr}(P_{li_1j_1}, P_{li_2j_2}) = 0 \ \ \forall \ \ i_1 + j_1 \neq i_2 + j_2$ regardless of triangle $l,\ l = 1, \ldots, m$.

□

Thus while $\mathbf{R}$ is a correlation matrix, the actual correlations between triangles are contained in the matrix $\mathbf{C}$,

$$\mathbf{C} := (\sqrt{\mathbf{I} - \mathbf{S}})\mathbf{R}(\sqrt{\mathbf{I} - \mathbf{S}}), \tag{2.7}$$

where $\mathbf{S} := \operatorname{diag}(s_1^2, \ldots, s_m^2) \ \in \mathbb{R}^{m \times m}$ and $\mathbf{I}_m$ the m-dimensional identity matrix,

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & \cdots & r_{1m} \\ r_{12} & 1 & r_{23} & \cdots & \cdots & r_{2m} \\ r_{13} & r_{23} & 1 & r_{34} & \cdots & r_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{1,m-1} & \cdots & r_{m-2,m-1} & 1 & \cdots & r_{m-1,m} \\ r_{1m} & r_{2m} & \cdots & \cdots & r_{m-1,m} & 1 \end{pmatrix} \in \mathbb{R}^{m \times m},$$

the correlation matrix of $\boldsymbol{\alpha}_t \in \mathbb{R}^m$ and $\sqrt{\cdot}$ operates componentwise.

$$\Rightarrow \mathbf{C} = \begin{pmatrix} c_1^2 & c_1 c_2 r_{12} & \cdots & c_1 c_{m-1} r_{1,m-1} & c_1 c_m r_{1,m} \\ c_1 c_2 r_{12} & c_2^2 & \cdots & c_2 c_{m-1} r_{2,m-1} & c_2 c_m r_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_1 c_{m-1} r_{1,m-1} & c_2 c_{m-1} r_{2,m-1} & \cdots & c_{m-1}^2 & c_{m-1} c_m r_{m-1,m} \\ c_1 c_m r_{1,m} & c_2 c_m r_{2,m} & \cdots & \cdots & c_m^2 \end{pmatrix}$$

From the definition of $\mathbf{C} \in \mathbb{R}^{m \times m}$ it follows that entry $(l\ ,\ k)$ of $\mathbf{C}$ equals the correlation between two payments in triangles l and k, $\ l, k = 1, \ldots, m$, falling into the same calendar year t, $\ t = 1, \ldots, n$. If $l = k$ entry $(l\ ,\ l)$ of $\mathbf{C}$ equals the correlation between two distinct payments in triangle l in the same calendar year t.
If $\mathbf{R} = \mathbf{I}$ there is zero correlation between different triangles l and k, $\ l \neq k$, and correlation c_l^2 between two transformed payments $\varphi(P_{lij})$ and $\varphi(P_{li'j'})$, $i \neq i'$ in triangle l falling into the same calendar $t = i + j = i' + j'$.
Even though the correlation matrix $\mathbf{R} \in \mathbb{R}^m$ models the dependence between triangles it is important to stress that these correlations are moderated by the communalities c_l of each triangle l, $\ l = 1 \ldots, m$. Therefore high specifities s_l indicate low dependence while high communalities c_l indicate high dependence between the m different triangles (remember $c_l^2 + s_l^2 = 1\ ,\ \forall\ \ l = 1, \ldots, m$).

2.6 Parameter Estimation

In the case of m loss triangles there are $m \cdot (m+1)/2$ parameters in order to completely specify the model.
First one needs to estimate the specifity s_l for each triangle $l = 1, \ldots, m$. Once those are specified one obtains the communalities c_l via $\ c_l^2 + s_l^2 = 1\ \ \forall\ \ l = 1, \ldots, m$.
Second the entries of the matrix $\mathbf{R} \in \mathbb{R}^m$ have to be estimated. Since $\mathbf{R}$ is a correlation matrix it has ones on the diagonal and is symmetric. Thus only $(m^2 - m)/2$ parameters have to be estimated, say the upper triangle excluding the diagonal of $\mathbf{R}$.
In this section 3 different possibilities on how to estimate the model parameters will be presented. Prior to discussing these methods in detail some theoretical framework needs to be developed.
As mentioned earlier payments in each triangle are transformed in a way that

$$\varphi_l(P_{lij})\ \sim\ N(\mu_{lj}, \sigma_{lj}^2),$$

where $\varphi_l\ ,\ \ l = 1, \ldots, m$ is a strictly increasing transformation, μ_{lj} is the column mean of the j-th column of the transformed triangle $\varphi_l(l)\ , l = 1, \ldots, m$ and σ_{lj} stands for the standard deviation of the j-th column of the transformed triangle $l\ , l = 1, \ldots, m$.
Next form

$$z_{lij} := \Phi^{-1}\left(\hat{G}\left(\frac{\varphi_l(p_{lij}) - \hat{\mu}_{lj}}{\hat{\sigma}_{lj}}\right)\right), \tag{2.8}$$

where $\hat{G}$ denotes the empirical distribution function of $\frac{\varphi(p_{lij})-\hat{\mu}_{lj}}{\hat{\sigma}_{lj}}$ calculated with respect to all i and j for each triangle l , $\hat{\mu}_{lj}$ the empirical column mean of the j-th column of transformed triangle l and $\hat{\sigma}_{lj}$ the empirical standard deviation, namely

$$\hat{\mu}_{lj} := \frac{1}{n-j}\sum_{i=1}^{n-j}\varphi_l(p_{lij})$$

and

$$\hat{\sigma}_{lj} := \sqrt{\frac{1}{n-j-1}\sum_{i=1}^{n-j}(\varphi_l(p_{lij})-\hat{\mu}_{lj})^2}$$

Comparing (2.8) to (2.6) it follows

$$z_{lij} \approx c_l\alpha_{l,i+j} + s_l\epsilon_{lij} \quad , \ i+j=t, \tag{2.9}$$

because $\hat{G}(x) \approx \Phi(x) \quad \forall x \in \mathbb{R}$.

2.6.1 Heuristic Parameter Estimation

The first approach of estimating the model parameters is a heuristic approach. Nevertheless, as will be shown later, it performs quite well. Furthermore this estimation method is easy to implement and performs quick.
Estimation is based on the so-called z-scores z_{lij} defined in (2.8):

$$z_{lij} \approx c_l\alpha_{l,i+j} + s_l\epsilon_{lij}, i+j=t$$

$$\text{Let} \ \ \overline{z}_{lt} := \frac{1}{t}\sum_{i+j=t} z_{lij} \quad \text{and} \quad \overline{\epsilon}_{lt} := \frac{1}{t}\sum_{i+j=t}\epsilon_{lij} \quad . \tag{2.10}$$

Then $\overline{z}_{lt}$ and $\overline{\epsilon}_{lt}$ are the arithmetic means of the t different z-scores z_{lij} and random terms ϵ_{lij} in one triangle l , $l = 1,\ldots,m$, falling into calendar year t, $t = i+j$.
It follows

$$z_{lij} - \overline{z}_{lt} \approx c_l\alpha_{l,i+j} + s_l\epsilon_{lij} - \frac{1}{t}\sum_{i+j=t} z_{lij} =$$

$$c_l\alpha_{l,i+j} + s_l\epsilon_{lij} - \frac{1}{t}\sum_{i+j=t}(c_l\alpha_{l,i+j} + s_l\epsilon_{lij}) =$$

$$c_l\alpha_{l,i+j} + s_l\epsilon_{lij} - c_l\alpha_{l,i+j} - \frac{1}{t}\sum_{i+j=t} s_l\epsilon_{lij},$$

where the last equality follows from the fact, that the same α_{lt} enters each and every transformed entry $\varphi(p_{lij})$ of triangle l in a given calendar year $t = i+j$. Recall, that from the definitions of t, l and j there are t different entries $\varphi(p_{lij})$, $i+j=t$ falling into

calendar year t.
Thus one obtains

$$z_{lij} - \overline{z}_{lt} \approx s_l \epsilon_{lij} - \frac{1}{t} \sum_{i+j=t} s_l \epsilon_{lij} = s_l(\epsilon_{lij} - \overline{\epsilon}_{lt}), \tag{2.11}$$

following directly from the definition of $\overline{\epsilon}_{lt}$.
Next one takes a look at the distributions of ϵ_{lij} and $\overline{\epsilon}_{lt}$, respectively:

$$\epsilon_{lij} \sim N(0,1) \Rightarrow \; \overline{\epsilon}_{lt} := \frac{1}{t} \sum_{i+j=t} \epsilon_{lij} \;\sim\; N(0, \frac{1}{t}).$$

By taking the square of both sides of (2.11) this finally leads to estimates

$$\hat{s}_l^2 = \frac{2}{n(n+1)} \sum_{t=1}^{n} \sum_{i+j=t} (z_{lij} - \overline{z}_{lt})^2 \;, \quad \hat{c}_l^2 = 1 - \hat{s}_l^2 \;, \quad \hat{\alpha}_{lt} = \frac{\overline{z}_{lt}}{\hat{c}_l},$$

where

- the estimate $\hat{s}_l^2$ is obtained by ignoring the term $(\epsilon_{lij} - \overline{\epsilon}_{lt})^2$ in the square of (2.11). This is justified by:

$$\begin{aligned}
E\left[(\epsilon_{lij} - \overline{\epsilon}_{lt})^2\right] &= E\left[\epsilon_{lij}^2\right] + E\left[\overline{\epsilon}_{lt}^2\right] - 2E\left[\epsilon_{lij}\overline{\epsilon}_{lt}\right] \\
E\left[\epsilon_{lij}^2\right] &= 1 \;,\; \text{due } \epsilon_{lij}^2 \sim \chi_1^2 \\
E\left[\overline{\epsilon}_{lt}^2\right] &= \frac{1}{t^2} E\left[\left(\sum_{i+j=t} \epsilon_{lij}\right)^2\right] = \frac{1}{t} \\
E\left[\epsilon_{lij}\overline{\epsilon}_{lt}\right] &= \frac{1}{t} \;.
\end{aligned}$$

  This is verified by using the well-known fact that

$$\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n} a_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \tag{2.12}$$

holds for each sequence $a_i,\ i \in \mathbb{N}$ of real numbers.

$$\begin{aligned}
E[\overline{\epsilon}_{lt}^2] &= E[\left(\frac{1}{t}\sum_{i+j=t}\epsilon_{lij}\right)^2] \\
&= \frac{1}{t^2}E[\left(\sum_{i+j=t}\epsilon_{lij}\right)^2] \\
&\overset{2.11}{=} \frac{1}{t^2}E[\sum_{i+j=t}\epsilon_{lij}^2 + 2\sum_{i=1}^{t-1}\sum_{i'=i+1}^{n}\epsilon_{lij}\epsilon_{li'j'}] \\
&\overset{\epsilon_{lij}\perp\epsilon_{li'j'}}{=} \frac{1}{t^2}E[\sum_{i+j=t}\epsilon_{lij}^2] + 2\sum_{i=1}^{t-1}\sum_{i'=i+1}^{n}E[\epsilon_{lij}]E[\epsilon_{li'j'}] \\
&\overset{E[\epsilon_{lij}]=0}{=} \frac{1}{t^2}E[\sum_{i+j=t}\epsilon_{lij}^2] \overset{E[\chi_t^2]=t}{=} \frac{1}{t^2}\cdot t = \frac{1}{t}\ , \\
\Rightarrow E[\epsilon_{lij}\cdot\overline{\epsilon}_{lt}] &= cov(\epsilon_{lij},\overline{\epsilon}_{lt}) + E[\epsilon_{lij}]E[\overline{\epsilon}_{lt}] \\
&= cov(\epsilon_{lij},\frac{1}{t}\sum_{i+j=t}\epsilon_{lij}) \\
&= \frac{1}{t}cov(\epsilon_{lij},\sum_{i+j=t}\epsilon_{lij}) = \frac{1}{t}\ ,
\end{aligned}$$

since $\epsilon_{lij}\perp\epsilon_{li'j'}\ \ \forall\ i \neq i'$ and $cov(\epsilon_{lij},\epsilon_{lij}) = Var(\epsilon_{lij}) = 1$.

$$\Rightarrow E\left[(\epsilon_{lij}-\overline{\epsilon}_{lt})^2\right] = 1 + \frac{1}{t} - \frac{2}{t} = \frac{t-1}{t} \approx 1 \text{ for large } t$$

- the estimate of the square of the communality of triangle l , $\hat{c}_l^2 = 1 - \hat{s}_l^2$, comes directly from the model definition (2.3).
- and the estimate $\hat{\alpha}_{lt} = \frac{\overline{z}_{lt}}{\hat{c}_l}$ follows from the definition of $\overline{z}_{lt}$:

$$\overline{z}_{lt} := \frac{1}{t}\sum_{i+j=t} z_{lij} \approx c_l\alpha_{lt} + s_l\overline{\epsilon}_{lt} \quad ,\overline{\epsilon}_{lt} \sim N(0,\frac{1}{t}).$$

Once $\hat{\alpha}_{lt}$ is estimated for each triangle l and calendar year t estimating $\hat{\mathbf{R}}$ is straight forward:
$\hat{\mathbf{R}}$ is the sample correlation matrix computed from the $\hat{\alpha}_{lt}$ or equivalently $\overline{z}_{lt}$.

2.6.2 Maximum Likelihood Estimation

MLE for Normal Data

Suppose $\mathbf{x} = (x_1,\ldots,x_n)$ is a realization of a random vector $\mathbf{X} = (X_1,\ldots,X_n) \in \mathbb{R}^n$, $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu},\mathbf{P})$, for some mean vector $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n) \in \mathbb{R}^n$ and a correlation matrix

$\mathbf{P} \in [-1,1]^{n\times n}$. The elements of $\boldsymbol{\mu}$ and $\mathbf{P}$ are unknown and collected in a vector called $\boldsymbol{\theta}$. The likelihood function of $\mathbf{x} = (x_1, \ldots, x_n)$ is defined as follows :

$$l(\mathbf{x}, \boldsymbol{\theta}) := f(x_1, \ldots, x_n|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n |\det(\mathbf{P})|}}\ exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Instead of maximizing this expression the following term is minimized:

$$-ll(\mathbf{x}, \boldsymbol{\theta}) := -\log(l(\mathbf{x}, \boldsymbol{\theta})) = -\frac{1}{2}\log((2\pi)^n \det(\mathbf{P})) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Since constant terms and factors have no influence on optimization problems, maximizing $l(\mathbf{x}, \theta)$ is equivalent to minimizing the expression:

$$\log(\det(\mathbf{P})) + (\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Application to the Model Data

From the definitions of α and ϵ

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R})\ \forall\ \ t\ \ ,\ t = 1, \ldots, n\ ,\ \text{where } \mathbf{R} \text{ is a }\ m \times m \text{ correlation matrix}$$

and $\epsilon_{lij} \sim N(0,1)$ one obtains that that the z-scores are marginally distributed $N(0,1)$:

$$\Rightarrow z_{lij} := c_l\alpha_{l,i+j} + s_l\epsilon_{lij}\ \ \sim\ \ N(0,1)\ ,\ c_l^2 + s_l^2 = 1.$$

Collecting all entries z_{lij} of one triangle in a given calendar year $t = i+j$ into a single vector yields

$$\mathbf{z_{lt}} := c_l\alpha_{lt}\mathbf{1_t} + s_l\epsilon_{lt}, \quad l = 1, \ldots, m,\ t = 1, \ldots, n,$$

$$c_l \in \mathbb{R}, s_l \in \mathbb{R}, \alpha_{lt} \in \mathbb{R},\ \mathbf{1}_t = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^t,\ \boldsymbol{\epsilon}_{lt} = \begin{pmatrix} \epsilon_{lij} \\ \vdots \\ \epsilon_{li'j'} \end{pmatrix} \in \mathbb{R}^t, i+j = \cdots = i'+j' = t.$$

Thus $\mathbf{z}_{lt} \in \mathbb{R}^t$, $\boldsymbol{\epsilon}_{lt} \sim N(0, I_t)$ and vector $\boldsymbol{\alpha}_t \in \mathbb{R}^m$ of α_{lt} has correaltion matrix $\mathbf{R} \in \mathbb{R}^{m\times m}$. In the next step all vectors $\mathbf{z}_{lt} \in \mathbb{R}^t$, $t = 1, \ldots, n$, $l = 1, \ldots, m$ are put together in one vector $\mathbf{z}_t$ for each t , $t = 1, \ldots, n$:

$$\mathbf{z}_t := \begin{pmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{mt} \end{pmatrix} \in \mathbb{R}^{m\cdot t}.$$

Definition 2.2 (Kronecker product)**.** *Horn and Johnson [2008]p. 243*
The Kronecker product of two matrices $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m\times n}$ *and* $\mathbf{B} = [b_{ij}] \in \mathbb{R}^{p\times q}$ *is denoted by* $\mathbf{A} \otimes \mathbf{B}$ *and is defined to be the block matrix*

$$\mathbf{A} \otimes \mathbf{B} \equiv \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{mp\times nq}.$$

Notice that $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$ *in general,*

Using matrix-vector-notation and definition 2.2, $\mathbf{z_t}$, $t = 1, \dots, n$ can be written as

$$\mathbf{z_t} = (\sqrt{\mathbf{I}_m - \mathbf{S}} \otimes 1_t)\boldsymbol{\alpha}_t + (\sqrt{\mathbf{S}} \otimes \mathbf{I}_t)\boldsymbol{\epsilon}_t,$$

where $\mathbf{I}_m = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{m\times m}$, $\mathbf{I_t} = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \vdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{t\times t}$,

$\mathbf{S} = \begin{pmatrix} s_1^2 & 0 & \dots & \dots & 0 \\ 0 & s_2^2 & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & s_m^2 \end{pmatrix} \in \mathbb{R}^{m\times m}$, $\mathbf{1}_t = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^t$, $\boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \in \mathbb{R}^m$,

$\boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{pmatrix} \in \mathbb{R}^{m\cdot t}$ and $\sqrt{\cdot}$ operates componentwise.

Due to the independence of payments falling into different calendar years t, $t = 1, \dots, n$, the $\mathbf{z}_t$, $t = 1, \dots, n$ are independent random vectors.
From the definitions of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\epsilon}_t$ it follows that $\boldsymbol{\alpha}_t \sim N(\mathbf{0}, \mathbf{R})$, $\mathbf{R} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I_{m\cdot t}})$. Thus the vectors $\mathbf{z}_t$, $t = 1, \dots, n$ have mean zero and covariance matrices

$$\mathbf{CM}_t := (\sqrt{\mathbf{I}_m - \mathbf{S}} \otimes \mathbf{1_t})\mathbf{R}(\sqrt{\mathbf{I}_m - \mathbf{S}} \otimes \mathbf{1_t})^T + \mathbf{S} \otimes \mathbf{I_t} = (\mathbf{C} \otimes \mathbf{1_t}\mathbf{1_t}^T) + (\mathbf{S} \otimes \mathbf{I_t}) \in \mathbb{R}^{m\cdot t\times m\cdot t},$$

where $\mathbf{C} = (\sqrt{\mathbf{I} - \mathbf{S}})\mathbf{R}(\sqrt{\mathbf{I} - \mathbf{S}})$ is defined in section 2.6 and $\sqrt{\cdot}$ operates componentwise.
From $c_l^2 + s_l^2 = 1 \;\; \forall \;\; l = 1, \dots, m$, it follows that $\mathbf{CM}_t$ actually is a correlation matrix

and has the form:

$$\mathbf{CM}_t = \begin{pmatrix} 1 & c_1^2 & \dots & c_1c_2r_{12} & \dots & c_1c_2r_{12} & \dots & c_1c_mr_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ c_1^2 & \dots & 1 & c_1c_2r_{12} & \dots & c_1c_2r_{12} & \dots & c_1c_mr_{1m} \\ c_1c_2r_{12} & \dots & c_1c_2r_{12} & 1 & c_2^2 & \dots & \dots & \vdots \\ \vdots & \ddots & \vdots & c_2^2 & \ddots & \dots & \dots & \vdots \\ c_1c_2r_{12} & \dots & c_1c_2r_{12} & \vdots & \dots & 1 & \dots & c_2c_mr_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \dots & \dots & \dots & \dots & \dots & \dots & 1 \end{pmatrix}.$$

Here the ones on the diagonal arise from $c_l^2 + s_l^2 = 1 \;\; \forall \;\; l = 1, \dots, m$.
To sum up one finally obtains that $\mathbf{z}_t \sim N(\mathbf{0}_{mt}, \mathbf{CM}_t)$, $t = 1, \dots, n$, where $\mathbf{0}_{mt}$ is a vector of $m \cdot t$ zeroes.
Finally the vectors $\mathbf{z}_t$, $t = 1, \dots, n$, are collected in a single vector $\mathbf{z}$,

$$\mathbf{z} := \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_n \end{pmatrix} \in \mathbb{R}^{m \cdot n \cdot \frac{n+1}{2}} \;,\; \mathbf{z} \;\sim\; N(\mathbf{0}, \mathbf{CM}) \;,$$

with

$$\mathbf{CM} := \begin{pmatrix} \mathbf{CM}_{t=1} & 0 & \dots & \dots & 0 \\ 0 & \mathbf{CM}_{t=2} & 0 & \dots & \vdots \\ \vdots & \dots & \ddots & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \mathbf{CM}_{t=n} \end{pmatrix} \in \; \mathbb{R}^{m \cdot n \cdot (\frac{n+1}{2}) \times m \cdot n \cdot (\frac{n+1}{2})} \tag{2.13}$$

It follows that the negative log-likelihood $-ll(\mathbf{z}, \boldsymbol{\theta})$ of $\mathbf{z}$ has the form

$$-ll(\mathbf{z}, \boldsymbol{\theta}) := \log(\det(\mathbf{CM})) + \mathbf{z}^T \mathbf{CM}^{-1} \mathbf{z}.$$

In order to simplify the numerical calculation of this expression one can make use of the block structure of **CM** and some rules for determinants and inverses of block matrices:

Theorem 2.3. *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are full-ranked matrices,* $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times m}$, *and* $\mathbf{C} := \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}$

$$\Rightarrow (i) \quad \det(\mathbf{C}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$$

$$(ii) \quad \mathbf{C}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{pmatrix}.$$

Recursive application of Theorem 2.3 to **CM** yields that the negative log-likelihood $-ll$ of $\mathbf{z}$ can be expressed as follows:

$$\overset{Theorem 2.3}{\Rightarrow} -ll(\mathbf{z},\boldsymbol{\theta}) = \sum_{t=1}^{n} \; log(\; det(\mathbf{Cm}_t)) + \mathbf{z}_t^T \mathbf{CM}_t^{-1} \mathbf{z}_t$$

This expression can be minimized with respect to choice of the $m(m+1)/2$ model parameters contained in $\boldsymbol{\theta} \in \mathbb{R}^{m(m+1)/2}$. These are in detail $s_1, \ldots, s_m$ and the upper diagonal entries r_{lk} of the correalation matrix $\mathbf{R}$, $\mathbf{R} = (r)_{lk}$, $l, k = 1, \ldots, m$ (remember: $\boldsymbol{\alpha}_t \sim N(\mathbf{0}, \mathbf{R})$).

2.6.3 Stepwise Maximum Likelihood Estimation

In the previous section maximum-likelihood estimation of the $m(m+1)/2$ model parameters is described. The critical issue using this approach is that $m \cdot n \cdot \frac{n+1}{2}$ data points are used to estimate $m(m+1)/2$ parameters. For a large number m of different loss triangles and a short observation period n this can lead to poor estimates due to a small ratio of available data and number of parameters to be estimated.
Thus this section will introduce a maximum-likelihood based estimation method making the most of the available data. The basic idea is to separately estimate the m specifities s_l of each triangle l, $l = 1\ldots, m$ and the $m \cdot \frac{m-1}{2}$ upper-diagonal entries of the correlation matrix $\mathbf{R}$, $\mathbf{R} \in \mathbb{R}^{m\times m}$.

Estimation of the Specifities s_l

Prior to estimating the entries of the correlation matrix $\mathbf{R} \in \mathbb{R}^{m\times m}$ one estimates the specifities $s_1, \ldots, s_m$. Estimation of the parameter s_l makes use of all transformed payments $\varphi_l(p_{lij})$ in triangle l, $l = 1\ldots, m$.
According to **Theorem 2.1** transformed payments $\varphi_l(P_{li_1j_1})$ and $\varphi_l(P_{li_2j_2})$ in the same triangle l, $l = 1, \ldots, m$ are correlated if they fall into the same calendar year $t = i_1 + j_1 = i_2 + j_2$, while there is zero correlation between transformed payments falling in different calendar years $t = i_1 + j_1$ and $t' = i_2 + j_2$, $t \neq t'$. These results are used to estimate the specifity parameter s_l of triangle l, $l = 1, \ldots, m$ as follows:

- Let again

$$z_{lij} := \Phi^{-1}\left(\hat{G}\left(\frac{\varphi_l(p_{lij}) - \hat{\mu}_{lj}}{\hat{\sigma}_{lj}}\right)\right) \overset{\mathrm{d}}{\approx} c_l\alpha_{l,i+j} + s_l\epsilon_{lij} \quad \sim \quad N(0,1)\ .$$

- In a first step all z-scores z_{lij} in triangle l and calendar year $t = i + j$ are collected in a single vector $\mathbf{z}_{lt}$. Since each triangle has been observed for n years one obtains n different vectors $\mathbf{z}_{lt}$, $t = 1, \ldots, n$:

$$\mathbf{z}_{l1} := (z_{l,1,0}) \in \mathbb{R},\ \mathbf{z}_{l2} := \begin{pmatrix} z_{l,2,0} \\ z_{l,1,1} \end{pmatrix} \in \mathbb{R}^2,\ \mathbf{z}_{l3} := \begin{pmatrix} z_{l,3,0} \\ z_{l,2,1} \\ z_{l,1,2} \end{pmatrix} \in \mathbb{R}^3,$$

$$\ldots,\ \mathbf{z}_{lt} := \begin{pmatrix} z_{l,t,0} \\ z_{l,t-1,1} \\ \vdots \\ z_{l,2,t-2} \\ z_{l,1,t-1} \end{pmatrix} \in \mathbb{R}^t, \ldots,\ \mathbf{z}_{nt} := \begin{pmatrix} z_{l,n,0} \\ z_{l,n-1,1} \\ \vdots \\ z_{l,2,n-2} \\ z_{l,1,n-2} \end{pmatrix} \in \mathbb{R}^n$$

Theorem 2.1 tells us that two entries of vector $\mathbf{z}_{lt} \in \mathbb{R}^t$ have correlation $c_l^2 = 1 - s_l^2$.

$$\Rightarrow \mathbf{K}_t := corr(z_{lt}) = \begin{pmatrix} 1 & c_l^2 & \ldots & c_l^2 \\ c_l^2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_l^2 \\ c_l^2 & \ldots & c_l^2 & 1 \end{pmatrix} \ \in \ \mathbb{R}^{t \times t} \tag{2.14}$$

- In the next step the vectors $\mathbf{z}_{l1}, \ldots, \mathbf{z}_{ln}$ are collected in a single vector $\mathbf{z}_l$, containing all transformed entries $\varphi(p_{lij})$ of triangle l, $l = 1, \ldots, m$ sorted by calendar year $t, t = 1, \ldots, n$:

$$\mathbf{z}_l := \begin{pmatrix} \mathbf{z}_{l1} \\ \mathbf{z}_{l2} \\ \vdots \\ \mathbf{z}_{lt} \\ \vdots \\ \mathbf{z}_{ln} \end{pmatrix} \ \in \mathbb{R}^{\frac{n(n+1)}{2}}$$

From **Theorem 2.1** we know that transformed payments falling in different calendar years are, according to the model, uncorrelated.

$$\Rightarrow \mathbf{z}_l \ \sim \ N(\mathbf{0}, cov(\mathbf{z}_l)) \ \text{ with } \ cov(\mathbf{z}_l) = \mathbf{K}(c_l) \ \in \mathbb{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}} \tag{2.15}$$

which means that covariance matrix $\mathbf{K}$ of vector $\mathbf{z}_l \in \mathbb{R}^{\frac{n(n+1)}{2}}$, $l = 1, \ldots, m$ depends only on the communality c_l of triangle l.

$$\Rightarrow \mathbf{K}(c_l) := cov(\mathbf{z}_l) = \begin{pmatrix} \mathbf{K}_1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \mathbf{K}_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{K}_t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 & \mathbf{K}_n \end{pmatrix} \ \in \ \mathbb{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}} \tag{2.16}$$

Thus the negative log-likelihood of $\mathbf{z}_l$ has the form

$$-ll(\mathbf{z}_l, c_l) = log(det(\mathbf{K})) + \mathbf{z}_l^T \ \mathbf{K}^{-1} \ \mathbf{z}_l \quad .$$

This expression can be minimized with respect to choice of c_l , $l = 1, \ldots, m$.
Application to each triangle l yields an estimate $\hat{c}_l$ of c_l.
Note that $\frac{n(n+1)}{2}$ data points have been used to estimate 1 parameter.

Estimation of R

The entries of the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$ are estimated in the same manner as in the previous section: First all z-scores z_{lij} , $i = 1, \ldots, n$, $j = 0, \ldots, n-1$, $l = 1 \ldots, m$ are collected in a single vector $\mathbf{z} \in \mathbb{R}^{mn\frac{n+1}{2}}$, $\mathbf{z} \sim N(\mathbf{0}, \mathbf{CM})$, where $\mathbf{CM}$ is defined in (2.11). Thus, like before, the negative log-likelihood of $\mathbf{z}$ has the form:

$$-ll(\mathbf{z}, \mathbf{R}) = log(det(\mathbf{CM})) + \mathbf{z_l}^T\ \mathbf{CM}^{-1}\ \mathbf{z_l}\ \ .$$

This expression can be minimized with respect to choice of $\frac{m(m-1)}{2}$ entries of the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$, $\alpha_{\mathbf{t}} \in \mathbb{R}^m \sim N(\mathbf{0}, \mathbf{R})$ i.i.d. for all t , $t = 1 \ldots, n$.
Note, that in contrast to the previous chapter, $\frac{mn(n+1)}{2}$ observations are used to estimate $\frac{m(m-1)}{2}$ parameters, while before $\frac{m(m+1)}{2}$ parameters had to be estimated with the same amount of available data.

2.6.4 Estimation Summary

Each of the three estimation methods displayed relies on the so-called z-scores

$$z_{lij} := \Phi^{-1}(\hat{G}\left(\frac{\varphi_l(p_{lij}) - \hat{\mu}_{lj}}{\hat{\sigma_{lj}}}\right)\ ,$$

where $\hat{G}$ denotes the empirical distribution function of $\frac{\varphi_l(p_{lij}) - \hat{\mu}_{lj}}{\hat{\sigma_{lj}}}$ calculated with respect to all i , $i = 1, \ldots, n$ and j , $j = 0, \ldots, n-1$. Due to the normality assumption (2.4)

$$\varphi(P_{lij}) \sim N(\mu_{lj}, \sigma_{lj}^2)$$

it follows that $\hat{G} \approx \Phi$. Together with the model definition (2.6)

$$F_{lj}(P_{lij}) \stackrel{\mathrm{d}}{=} \Phi(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \stackrel{\mathrm{d}}{=} \Phi\left(\frac{\varphi(P_{lij}) - \mu_{lj}}{\sigma_{lj}}\right)$$

it follows that

$$z_{lij} \approx c_l \alpha_{l,i+j} + s_l \epsilon_{lij}.$$

The heuristic method estimates the specifities s_l of data only in triangle l , $l = 1, \ldots, m$

via

$$\hat{s}_l^2 = \frac{2}{n(n+1)} \sum_{t=1}^{n} \sum_{i+j=t} (z_{lij} - \frac{1}{t} \sum_{i+j=t} z_{lij})^2.$$

Based on these estimates $\hat{\alpha_{lt}}$ is estimated for each triangle l and calendar year t , $t = 1, \ldots, n$:

$$\hat{\alpha_{lt}} = \frac{\frac{1}{t} \sum_{i+j=t} z_{lij}}{\sqrt{1 - \hat{s}_l^2}}$$

Estimate $\hat{\mathbf{R}}$ of the correlation matrix $\mathbf{R}$ is calculated as the sample correlation from the $\hat{\alpha}_{lt}$. In contrast to the heuristic approach the maximum-likelihood approach and the

stepwise maximum-likelihood estimation do not estimate $\hat{\alpha}_{lt}$. Instead the entries $r_{l,k}$ of $\mathbf{R}$ are estimated directly from the data.
The maximum-likelihood based estimation is based on minimizing

$$-ll(\mathbf{z},\boldsymbol{\theta}) := \log(\det(\mathbf{CM})) + \mathbf{z}^T\mathbf{CM}^{-1}\mathbf{z}$$

with respect to $s_1,\ldots,s_5$ and the upper-diagonal entries of $\mathbf{R}$ contained in $\boldsymbol{\theta}$, where

$$\mathbf{CM} := \begin{pmatrix} \mathbf{CM}_{t=1} & 0 & \ldots & \ldots & 0 \\ 0 & \mathbf{CM}_{t=2} & 0 & \ldots & \vdots \\ \vdots & \ldots & \ddots & \ldots & \vdots \\ \vdots & \ldots & \ldots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \mathbf{CM}_{t=n} \end{pmatrix} \in \mathbb{R}^{m\cdot n\cdot(\frac{n+1}{2})\times m\cdot n\cdot(\frac{n+1}{2})} ,$$

with

$$\mathbf{CM_t} = \begin{pmatrix} 1 & c_1^2 & \ldots & c_1c_2r_{12} & \ldots & c_1c_2r_{12} & \ldots & c_1c_mr_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ldots & \vdots \\ c_1^2 & \ldots & 1 & c_1c_2r_{12} & \ldots & c_1c_2r_{12} & \ldots & c_1c_mr_{1m} \\ c_1c_2r_{12} & \ldots & c_1c_2r_{12} & 1 & c_2^2 & \ldots & \ldots & \vdots \\ \vdots & \ddots & \vdots & c_2^2 & \ddots & \ldots & \ldots & \vdots \\ c_1c_2r_{12} & \ldots & c_1c_2r_{12} & \vdots & \ldots & 1 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_1c_mr_{1m} & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1 \end{pmatrix}$$

and $\mathbf{z} \in \mathbb{R}^{mn\frac{n+1}{2}}$ is the vector containing all z-scores z_{lij} , $l = 1,\ldots,m$, $i = 1,\ldots,n$, $j = 0,\ldots,n-1$ sorted by triangle l and calendar year t.

The stepwise maximum-likelihood approach estimates $s_1,\ldots,s_m$ prior to $\mathbf{R}$.
For each triangle l , $l = 1,\ldots,m$ estimate $\hat{s}_l$ of s_l is obtained by minimizing

$$-ll(\mathbf{z_l}, c_l) = log(det(\mathbf{K})) + \mathbf{z}_l^T\ \mathbf{K}^{-1}\ \mathbf{z}_l \quad ,$$

with respect to $c_l = \sqrt{1-s_l^2}$, where
$\mathbf{z}_l \in \mathbb{R}^{\frac{n(n+1)}{2}}$ is a vector containing all z-scores z_{lij} of triangle l sorted by calendar year t and

$$\Rightarrow \mathbf{K} := \begin{pmatrix} \mathbf{K}_1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \mathbf{K}_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{K}_t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 & \mathbf{K}_n \end{pmatrix} \in \mathbb{R}^{\frac{n(n+1)}{2}\times\frac{n(n+1)}{2}} ,$$

where

$$\mathbf{K}_t := \begin{pmatrix} 1 & c_l^2 & \dots & c_l^2 \\ c_l^2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_l^2 \\ c_l^2 & \dots & c_l^2 & 1 \end{pmatrix} \in \mathbb{R}^{t \times t} .$$

Given the estimates $\hat{s}_1, \dots, \hat{s}_m$ estimation of $\mathbf{R}$ is performed like described for the maximum-likelihood approach with the only difference that

$$-ll(\mathbf{z}, \mathbf{R}) := \log(\det(\mathbf{CM})) + \mathbf{z}^T \mathbf{CM}^{-1} \mathbf{z}$$

is now minimized with the respect to the $\frac{m(m-1)}{2}$ upper-diagonal entries of $\mathbf{R}$ since the specifities $s_1, \dots, s_m$ have already been estimated.
Algorithms 1, 2 and 3 give a quick overview on the three introduced estimation methods.

Algorithm 1 Heuristic approach for parameter estimation

```
for l = 1 to m do
  part 1 : Calculation of z̄_lt ∀t = 1,...,n :
  z̄_l = c()
  for t = 1 to n do
    cumsum := 0
    for j = 0 to t-1 do
      cumsum = cumsum + z_{l,t-j,j}
    end for
    z̄_lt = 1/t * cumsum
    z̄_l = c(z̄_l, z̄_lt)
  end for
  part 2 : Parameter estimation :
  varsum := 0
  for t = 1 to n do
    for j = 0 to n-1 do
      varsum = varsum + (z_{l,t-j,j} - z̄_l[t])^2
    end for
  end for
  ŝ_l^2 = 2/(n(n+1)) * varsum
  ĉ_l^2 = 1 - ŝ_l^2
  α̂_lt = z̄_l[t] / ĉ_l
end for
```

Algorithm 2 Maximum-Likelihood based parameter estimation

Part 1 : data vector

$z_{lij} := c_l \alpha_{l,i+j} + s_l \epsilon_{lij}$

collection of all z-scores z_{lij} in each triangle l , $l = 1, \ldots, m$ in a single vector $\mathbf{z} \in \mathbb{R}^{mn(n+1)/2}$ sorted by calendar year t , $t = 1, \ldots, n$, $t = i+j$ and triangle l, $l = 1, \ldots, m$:

$\mathbf{z} := c()$
for $t = 1$ to n **do**
  $\mathbf{z}_t := c()$
  for $l = 1$ to m **do**
    for $i = 1$ to t **do**
      $\mathbf{z}_t := c(\mathbf{z}_t, z_{l,i,t-i})$
    end for
  end for
  $\mathbf{z} := \begin{pmatrix} \mathbf{z} \\ \mathbf{z}_t \end{pmatrix}$
end for

Part 2 : covariance structure

$\boldsymbol{\theta} \in \mathbb{R}^{m(m+1)/2}$ vector containing starting values for optimization :
$\boldsymbol{\theta} := (s_1, \ldots, s_m, r_{1,2}, \ldots, r_{1m}, r_{2,3}, \ldots, r_{2,m}, \ldots, r_{m-1,m})^T$
$\mathbf{S} := diag(s_1^2, \ldots, s_m^2) \in \mathbb{R}^{m \times m}$, $\mathbf{I} := diag(1, \ldots, 1) \in \mathbb{R}^{m \times m}$, $\mathbf{1}_t := (1, \ldots, 1)^T \in \mathbb{R}^t$
$\mathbf{R} := matrix(1, m, m) \in \mathbb{R}^{m \times m}$
for $i = 1$ to m **do**
  for $j = i + 1$ to m **do**
    $\mathbf{R}[i, j] := r_{i,j}$
    $\mathbf{R}[j, i] := \mathbf{R}[i, j]$
  end for
end for
for $t = 1$ to n **do**
  $\mathbf{CM}_t := \left(\sqrt{\mathbf{I} - \mathbf{S}} \otimes \mathbf{1}_t\right) \mathbf{R} \left(\sqrt{\mathbf{I} - \mathbf{S}} \otimes \mathbf{1}_t\right)^T + \mathbf{S} \otimes \mathbf{I}_t$
end for
$\mathbf{CM} := diag(\mathbf{CM}_1, \mathbf{CM}_2, \ldots, \mathbf{CM}_n) \in \mathbb{R}^{m \cdot n \cdot \frac{n+1}{2} \times m \cdot n \cdot \frac{n+1}{2}}$

Part 3: optimization of the likelihood

With $\mathbf{z} \sim N(\mathbf{0}, \mathbf{CM})$
the quantity $-ll := log(det(\mathbf{CM})) + \mathbf{z}^T \mathbf{CM}^{-1} \mathbf{z}$
is minimized with respect to the $m(m + 1)/2$ parameters contained in θ.
Typically the estimates $\hat{\mathbf{s}}^H$ and $\hat{\mathbf{R}}^H$ obtained by the heuristic approach are used as starting values for $\boldsymbol{\theta}$

Algorithm 3 Stepwise maximum-likelihood parameter estimation

Part 1 : estimation of the specifities $s_1, \ldots, s_m$

$z_{lij} := c_l \alpha_{l,i+j} + s_l \epsilon_{lij}$

collection of all z-scores z_{lij} in one triangle l , $l = 1, \ldots, m$ sorted by calendar year $t = i + j$, $t = 1, \ldots, n$ in a single vector $\mathbf{z}_l \in \mathbb{R}^{\frac{n(n+1)}{2}}$:

$\mathbf{z_l} := c()$

for $t = 1$ to n **do**

$\mathbf{z}_{lt} = c()$

for $i = 1$ to t **do**

$\mathbf{z}_{lt} = c(\mathbf{z}_{lt}, z_{li,t-i})$

end for

$\mathbf{z}_l := \begin{pmatrix} \mathbf{z}_l \\ \mathbf{z}_{lt} \end{pmatrix}$

end for

With $\mathbf{K}_t := \begin{pmatrix} 1 & c_l^2 & \ldots & c_l^2 \\ c_l^2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_l^2 \\ c_l^2 & \ldots & c_l^2 & 1 \end{pmatrix} \in \mathbb{R}^{t \times t}$ it follows:

$\mathbf{z}_l \sim N(\mathbf{0}, \mathbf{K})$, where

$$\mathbf{K} := cov(\mathbf{z}_l) = \begin{pmatrix} \mathbf{K}_1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \mathbf{K}_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{K}_t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 & \mathbf{K}_n \end{pmatrix} \in \mathbb{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}}$$

minimizing $-ll := log(det(\mathbf{K})) + \mathbf{z}_l^T \ \mathbf{K}^{-1} \ \mathbf{z}_l$

with respect to parameter $c_l = \sqrt{1 - s_l^2}$ yields an estimate $\hat{s}_l^{sML}$ of s_l.

Part 2 : estimation of the correlations

$\boldsymbol{\theta} \in \mathbb{R}^{m(m-1)/2}$ vector containing starting values for optimization :

$\boldsymbol{\theta} := (r_{1,2}, \ldots, r_{1,m}, r_{2,3}, \ldots, r_{2,m}, \ldots, r_{m-1,m})^T$

$\mathbf{S} := diag(\hat{s}_1^{sML}, \ldots, \hat{s}_m^{sML})^2 \in \mathbb{R}^{m \times m}$, $\mathbf{I}_m := diag(1, \ldots, 1) \in \mathbb{R}^{m \times m}$,

$\mathbf{1}_t := (1, \ldots, 1)^T \in \mathbb{R}^t$

$\mathbf{R} := matrix(1, m, m) \in \mathbb{R}^{m \times m}$

for $i = 1$ to m **do**

for $j = i + 1$ to m **do**

$\mathbf{R}[i, j] := r_{i,j}$, $\mathbf{R}[j, i] := \mathbf{R}[i, j]$

end for

end for

for $t = 1$ to n **do**

$\mathbf{CM}_t := \left(\sqrt{\mathbf{I} - \mathbf{S}} \otimes \mathbf{1}_t\right) \mathbf{R} \left(\sqrt{\mathbf{I}_m - \mathbf{S}} \otimes \mathbf{1}_t\right)^T + \mathbf{S} \otimes \mathbf{I}_t$

end for

$\mathbf{CM} := diag(\mathbf{CM}_1, \mathbf{CM}_2, \ldots, \mathbf{CM}_n) \in \mathbb{R}^{m \cdot n \cdot \frac{n+1}{2} \times m \cdot n \cdot \frac{n+1}{2}}$

Part 3: optimization of the likelihood

$\mathbf{z} \sim N(\mathbf{0}, \mathbf{CM})$

minimize $-ll := log(det(\mathbf{CM})) + \mathbf{z}^T \mathbf{CM}^{-1} \mathbf{z}$

with respect to the $m(m-1)/2$ parameters contained in $\boldsymbol{\theta}$, i.e. the upper diagonal entries of $\mathbf{R}$.

Typically the estimates $\hat{\mathbf{R}}^H$ are used as starting values for $\boldsymbol{\theta}$.

2.7 Loss Triangle Simulation

The methodoligy of simulating Loss Triangles and henceforth reproducing data with a certain dependence structure is slightly different for ML-approaches and the heuristic approach since, in the heuristic approach, one estimates α_{lt} , $l = 1, \ldots, m$, $t = 1, \ldots, n$, directly from the data, while in the ML-approaches only $\mathbf{R} \in \mathbb{R}^{m \times m}$ is estimated and the values of α_{lt} are obtained by drawing from $\mathbf{N}(0, \mathbf{R})$. Algorithms 4 and 5 display the different methods.

Algorithm 4 Generating Loss Triangles using ML-Approaches

for $l = 1$ to m **do**
  for $j = 1$ to $n-1$ **do**
    $\hat{\mu}_{lj} = \frac{1}{n-j} \sum_{i=1}^{n-j} \varphi(p_{lij})$, $\hat{\sigma}^2_{lj} = \frac{1}{n-j} \sum_{i=1}^{n-j} (\varphi(p_{lij}) - \hat{\mu}_{lj})^2$
  end for
end for
for $t = 1$ to n **do**
  draw $a_t = \begin{pmatrix} a_{1,t} \\ \vdots \\ a_{m,t} \end{pmatrix} \in \mathbb{R}^m$ from $N(\mathbf{0}, \hat{\mathbf{R}})$
end for
for $l = 1$ to m **do**
  for $i = 1$ to n **do**
    for $j = 0$ to $n - i$ **do**
      draw e_{lij} from $N(0, 1)$
      $\hat{p}_{lij} := \varphi^{-1} \left(\hat{\mu}_{lj} + \hat{\sigma}_{lj} \left(\hat{c}_l a_{l,i+j} + \hat{s}_l e_{lij} \right) \right)$
    end for
  end for
end for

Algorithm 5 Generating Loss Triangles using the Heuristic Approach

```
for l = 1 to m do
  for j = 1 to n − 1 do
    μ̂_lj = 1/(n−j) Σ_{i=1}^{n−j} φ(p_lij) , σ̂²_lj = 1/(n−j) Σ_{i=1}^{n−j} (φ(p_lij) − μ̂_lj)²
  end for
end for
for l = 1 to m do
  for t = 1 to n do
    a_lt = ((1/t) Σ_{i+j=t} z_lij) / sqrt(1 − ŝ_l²)
  end for
end for
for l = 1 to m do
  for i = 1 to n do
    for j = 0 to n − i do
      draw e_lij from N(0, 1)
      p̂_lij := φ^{−1} (μ̂_lj + σ̂_lj (ĉ_l a_{l,i+j} + ŝ_l e_lij))
    end for
  end for
end for
```

2.8 Forecasting future payments

Forecasting of future payments is based on the parameter estimates $\hat{s}_1, \ldots, \hat{s}_m$ of $s_1, \ldots, s_m$ and $\hat{\mathbf{R}}$ of $\mathbf{R}$. Applying the Gaussian copula model for loss triangle dependence makes it possible to forecast future payments $p_{lij}, i+j = t$ for $n+1 \leq i+j \leq 2n$, $i = 1, \ldots, n$, $j = 1, \ldots, n-1$.
The upper index on j assumes that forecasts based on the model can only be made until development year $j = n-1$. This seems reasonable since for development years $j \geq n$ no data is available. Payments p_{lij} with $j \geq n$, however, can be estimated using tail estimation methods.
Thus, using the model, each loss triangle l , $l = 1, \ldots, m$ gets filled and one obtains a loss square of size $n \times n$. Note that $\frac{n(n+1)}{2}$ payments are known while $\frac{n(n-1)}{2}$ still need to be estimated in order to receive a loss square.
Forecasting future payments basically relies on

$$\varphi_l(P_{lij}) \stackrel{d}{=} \mu_{lj} \ + \ \sigma_{lj}(c_l \alpha_{l,i+j} + s_l \epsilon_{lij}) \ ,$$

$$\boldsymbol{\alpha}_t := \begin{pmatrix} \alpha_{1t} \\ \vdots \\ \alpha_{mt} \end{pmatrix} \in \mathbb{R}^m \ , \ \boldsymbol{\alpha}_t \sim N(\mathbf{0}, \mathbf{R}) \ , \ \epsilon_{lij} \sim N(0,1).$$

Given the estimates $\hat{c}_1, \ldots, \hat{c}_m, \hat{s}_1, \ldots, \hat{s}_m$ and $\hat{\mathbf{R}}$ estimation of future payments is achieved as follows:

First one draws $\mathbf{a}_t \in \mathbb{R}^m$ from $N(\mathbf{0}, \hat{\mathbf{R}})$ for $t = n+1, n+2, \ldots, 2n$.
Next $\frac{n(n-1)}{2}$ draws e_{lij} from $N(0,1)$ are made, one for each payment to be forecasted. With

$$\hat{z}_{lij} := \hat{c}_l a_{l,i+j} + \hat{s}_l e_{lij} \quad , i+j = t$$

future payment p_{lij} , $i+j=t$, $n+1 \le t \le 2n-1$ is estimated by

$$\hat{p}_{lij} := \varphi_l^{-1}(\hat{\mu}_{lj} + \hat{\sigma}_{lj}\hat{z}_{lij}). \tag{2.17}$$

Note that due to $\hat{c}_l^2 + \hat{s}_l^2 = 1$ we have that

$$\hat{z}_{lij} \ \sim \ N(0,1) \tag{2.18}$$

and therefore

$$\hat{\mu}_{lj} + \hat{\sigma}_{lj}\hat{z}_{lij} \ \sim \ N(\hat{\mu}_{lj}, \hat{\sigma}_{lj}). \tag{2.19}$$

Algorithm 6 gives a brief overview on forecasting future payments using the Gaussian Copula Model.

Algorithm 6 Forecasting future payments

```
for l = 1 to m do
  for j = 1 to n−1 do
    μ̂_lj = 1/(n−j) Σ_{i=1}^{n−j} φ(p_lij) ,  σ̂²_lj = 1/(n−j) Σ_{i=1}^{n−j} (φ(p_lij) − μ̂_lj)²
  end for
end for
for t = n+1 to 2n−1 do
  draw a_t = (a_{1,t}, ..., a_{m,t})ᵀ ∈ ℝ^m from N(0, R̂)
end for
for l = 1 to m do
  for i = 2 to n do
    for j = n−i+1 to n−1 do
      draw e_lij from N(0,1)
      p̂_lij := φ^{−1}(μ̂_lj + σ̂_lj (ĉ_l a_{l,i+j} + ŝ_l e_lij))
    end for
  end for
end for
```

Chapter 3

Application to Real Insurance Data

3.1 Data

The provided data consists of $m = 5$ different loss triangles, where each triangle represents one specific line of business. For reasons of secrecy the data as well as the lines of business have been made anonymous. Data collection started in 1993 and the last data available come from the year 2008. Thus, in terms of the model notation, we have $m = 5$ different loss triangles observed through a period of $n = 16$ years. So we have a total of $m\frac{n(n+1)}{2} = 680$ observations p_{lij} , $\frac{n(n+1)}{2} = 136$ for each line of business.

3.2 Shapiro-Wilk test for normality

In statistics, the Shapiro-Wilk test (Shapiro and Wilk [1965]) tests the null hypothesis that a sample $x_1, \ldots, x_n$ came from a normally distributed population $X_1, \ldots, X_n$. It was published in 1965 by Samuel Shapiro and Martin Wilk.
Basic assumption of this test is that $X_1, \ldots, X_n$ are independent with constant mean μ and constant variance σ^2.

$$H_0 : X_1, \ldots, X_n \text{ are i.i.d. } \sim N(\mu, \sigma^2)$$

vs.

$$H_1 : \text{not } H_0$$

The observations $x_1, \ldots, x_n$ get ordered to receive $x_{(1)}, \ldots, x_{(n)}$ where $x_{(j)}$ is the jth order statistic, i.e., the jth-smallest number in the sample.

The expected values m_i of the ordered statistics can, under the assumption that H_0 holds, be derived by the formula

$$m_i = \Phi^{-1}\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right) ,$$

where n is the total number of observations and

$$\Phi(x_{(i)}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_{(i)}-\mu)^2}{2\sigma^2}} .$$

The test statistic W for the Shapiro-Wilk test for normality is defined as

$$W := \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{(n-1)s^2} \,, \tag{3.1}$$

where

$$\begin{aligned}
(a_1, \ldots, a_n) &:= \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}} \\
m &:= (m_1, \ldots, m_n)^T \\
V &:= \begin{pmatrix} cov(m_1, m_1) & \ldots & \text{cov}(m_1, m_n) \\ \vdots & \ddots & \vdots \\ cov(m_n, m_1) & \ldots & cov(m_n, m_n) \end{pmatrix} \\
s^2 &:= \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} \\
\overline{x} &:= \frac{\sum_{i=1}^{n} x_i}{n}.
\end{aligned}$$

The value of test statistic W is compared with a critical value $W_{crit}(n, \beta)$ for a given sample size n and significance level β. These critical values are summarized in many textbooks for $3 \leq n \leq 50$, like i.e. Pearson and Hartley [1972]. If $W < W_{crit}(n, \beta)$ H_0 is rejected, while for $W \geq W_{crit}(n, \beta)$ the null hypothesis cannot be rejected.
What in context of this work makes this test superior to other well-known tests for normality, like i.e. the Kolmogorov-Smirnoff test or Pearsons chi-square test, is the fact, that the Shapiro-Wilk test performs quite well, even for small sample sizes (see Chen and Wilk [1968]).

3.3 Data Transformation

As described in the previous chapter one basic model assumption is that payments P_{lij} , $l = 1, \ldots, m$, $i = 1, \ldots, n$, $j = 0, \ldots, n-1$, in each column j of each loss triangle l are modelled as random variables with unique columnwise distribution function F_{lj}. Further it is assumed, that for some monotonic tranfsormation φ

$$\frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}} \sim N(0, 1) \tag{3.2}$$

where μ_{lj} denotes the column mean of $\varphi(P_{lij})$ and σ_{lj}^2 the column variance of $\varphi(P_{lij})$.
So, the first step in applying the Gaussian copula model for loss triangle dependence to the data is it to find a monotonic transformation φ such that the transformed payments $\varphi(P_{lij})$ are normally distributed with means and standard deviations that only depend on triangle l and development year j:

$$\varphi_l(P_{lij}) \sim N(\mu_{lj}, \sigma_{lj}) \;\; \forall\, l = 1, \ldots, m \;\;, j = 0 \ldots, n-1$$

There is a wide field of possible transformations of observed data to normally distributed data. Therefore some of the most common transformations will be introduced and applied to the data. The goodness of these different transformations will be tested by the Shapiro-Wilk test for normality and graphical tools such as normal probability plots.

Application to Data

In this Section several possible transformation φ_l , $l = 1, \dots, m$ will be applied to the observed payments p_{lij}.

Transformation	Assumption
$\varphi_1(p_{lij}) = p_{lij}$	$P_{lij} \sim N(\mu_{lj}, \sigma^2_{lj})$
$\varphi_2(p_{lij}) = log(p_{lij})$	$P_{lij} \sim LN(\mu_{lj}, \sigma^2_{lj})$
$\varphi_3(p_{lij}) = \frac{-1}{p_{lij}}$	$\frac{-1}{P_{lij}} \sim N(\mu_{lj}, \sigma^2_{lj})$
$\varphi_4(p_{lij}) = log(log(p_{lij}) + c)$	$log(log(P_{lij}) + c) \sim N(\mu_{lj}, \sigma^2_{lj})$
$\varphi_5(p_{lij}) = \sqrt{p_{lij}}$	$\sqrt{P_{lij}} \sim N(\mu_{lj}, \sigma^2_{lj})$

Table 3.1: Overview over possible normal transformations for observed payments

Table 3.1 displays 5 strictly increasing data transformations $\varphi_1, \dots, \varphi_5$, namely the identity function, the log-transformation, an inverse transformation, a log-shifted-log transformation and the squareroot-transformation.
Note that parameter $c \in \mathbb{R}$ for transformation $\varphi_4(p_{lij}) = log(log(p_{lij}) + c)$ is chosen such that

$$log(p_{lij}) + c > 0$$

holds for $l = 1 \dots, m$, $i = 1, \dots, n$ and $j = 0, \dots, n-1$.

Of course there is quite a number of other possible normal transformations and many more than those mentioned have been applied to real insurance data, but for ease of simplicity only the 5 transformations $\varphi_1, \dots, \varphi_5$ will be discussed. In order to check the normality of transformed observations $\varphi_u(p_{lij})$, $u = 1 \dots, 5$ the Shapiro-Wilks test for normality will be performed for transformed data in each column j , $j = 0, \dots, 15$ of each loss triangle. Test statistics that are higher than a given critcal value indicate that the hypothesis that the transformed data $\frac{\varphi_u(p_{lij}) - \hat{\mu}_{lj}}{\hat{\sigma}_{lj}}$ in column j is distributed standard normal can be rejected:

$$\varphi_u(\mathbf{P_{lj}}) := \begin{pmatrix} \varphi_u(P_{l,1,j}) \\ \varphi_u(P_{l,2,j}) \\ \vdots \\ \varphi_u(P_{l,n-j,j}) \end{pmatrix} \in \mathbb{R}^{n-j}$$

is the vector containing all transformed entries in triangle $l, l = 1, \dots, 5$ and development year j , $j = 0, \dots, 15$. The Shapiro-Wilk test tests the vector $\varphi_u(\mathbf{P_{lj}})$ for normality:

$$H_0 : \frac{\varphi_u(\mathbf{P_{lj}}) - \mu_{lj}}{\sigma_{lj}} \sim N(0,1)$$

vs.

$$H_1 : \frac{\varphi_u(\mathbf{P_{lj}}) - \mu_{lj}}{\sigma_{lj}} \sim F \neq \Phi$$

The obtained test statistics , W_j^{SW} for the Shapiro-Wilk test will be compared for each of the transformations φ_u , $u = 1, \ldots, 5$.
With

$$N_\beta^{SW} := \sum_{j=0}^{12} \mathbb{1}\{W_j^{SW} \leq W_{crit}(\beta, n-j)\} \tag{3.3}$$

where W_j^{SW} denotes the value of the test statistic and $W_{crit}(\beta, n-j)$ denotes the critical value depending on significance level β and number of observations per column $n-j$. Thus N_β^{SW} counts the number of columns in which H_0 is rejected at significance level β.

	triangle 1		triangle 2		triangle 3		triangle 4		triangle 5	
transformation	$N_{0.05}^{SW}$	$N_{0.1}^{SW}$	$N_{0.05}^{SW}$	$N_{0.1}^{SW}$	$N_{0.05}^{SW}$	$N_{0.1}^{SW}$	$N_{0.05}^{SW}$	$N_{0.1}^{SW}$	$N_{0.05}^{SW}$	$N_{0.1}^{SW}$
p_{lij}	4	6	9	11	8	11	8	10	7	7
$log(p_{lij})$	**1**	**0**	8	11	9	11	4	4	1	2
$-1/p_{lij}$	3	4	**2**	**0**	**1**	**0**	**1**	**0**	**0**	**0**
$log(log(p_{lij}) + c)$	3	3	7	11	5	8	1	2	1	3
$\sqrt{p_{lij}}$	2	7	8	11	6	9	5	9	1	6

Table 3.2: Shapiro Wilk Test for each triangle and each transformation $\varphi_1, \ldots, \varphi_5$. Bold numbers indicate the lowest values of N^{SW} for a given triangle

Table 3.2 displays $N_{0.05}^{SW}$ and $N_{0.1}^{SW}$, i.e. the number of columns for which the normal assumption can be rejected according to the Shapiro-Wilks Test for normality, for each triangle $l = 1, \ldots, 5$ and each of the proposed transformations.
This procedure leads to choice of transformation $\varphi_2(p_{lij}) = log(p_{lij})$ for triangle 1 and choice of transformation $\varphi_3(p_{lij}) = \frac{-1}{p_{lij}}$ for triangles 2, 3, 4 and 5.
These choices will be justified using graphical tools like histograms and normal probability plots, i.e. a plot of $\hat{G}\left(\frac{\varphi_u(p_{lij})-\hat{\mu}_{lj}}{\hat{\sigma}_{lj}}\right)$ against $\Phi\left(\frac{\varphi_u(p_{lij})-\hat{\mu}_{lj}}{\hat{\sigma}_{lj}}\right)$ for a certain choice of $\varphi_u, u = 1, \ldots, 5$. Note that

$$\hat{\mu}_{lj} := \frac{1}{n-j} \sum_{i=1}^{n-j} \varphi_u(p_{lij})$$

and

$$\hat{\sigma}_{lj} := \sqrt{\frac{1}{n-j-1} \sum_{i=1}^{n-j} (\varphi_u(p_{lij}) - \hat{\mu}_{lj})^2}.$$

Figures 3.1, ..., 3.5 show histograms of $\frac{\varphi_u(p_{lij})-\hat{\mu}_{lj}}{\hat{\sigma}_{lj}}$ for each transformation φ_u and all $i = 1, \ldots, 16$, $j = 0, \ldots, 15$ as well as the normal based probability for the chosen transformation.

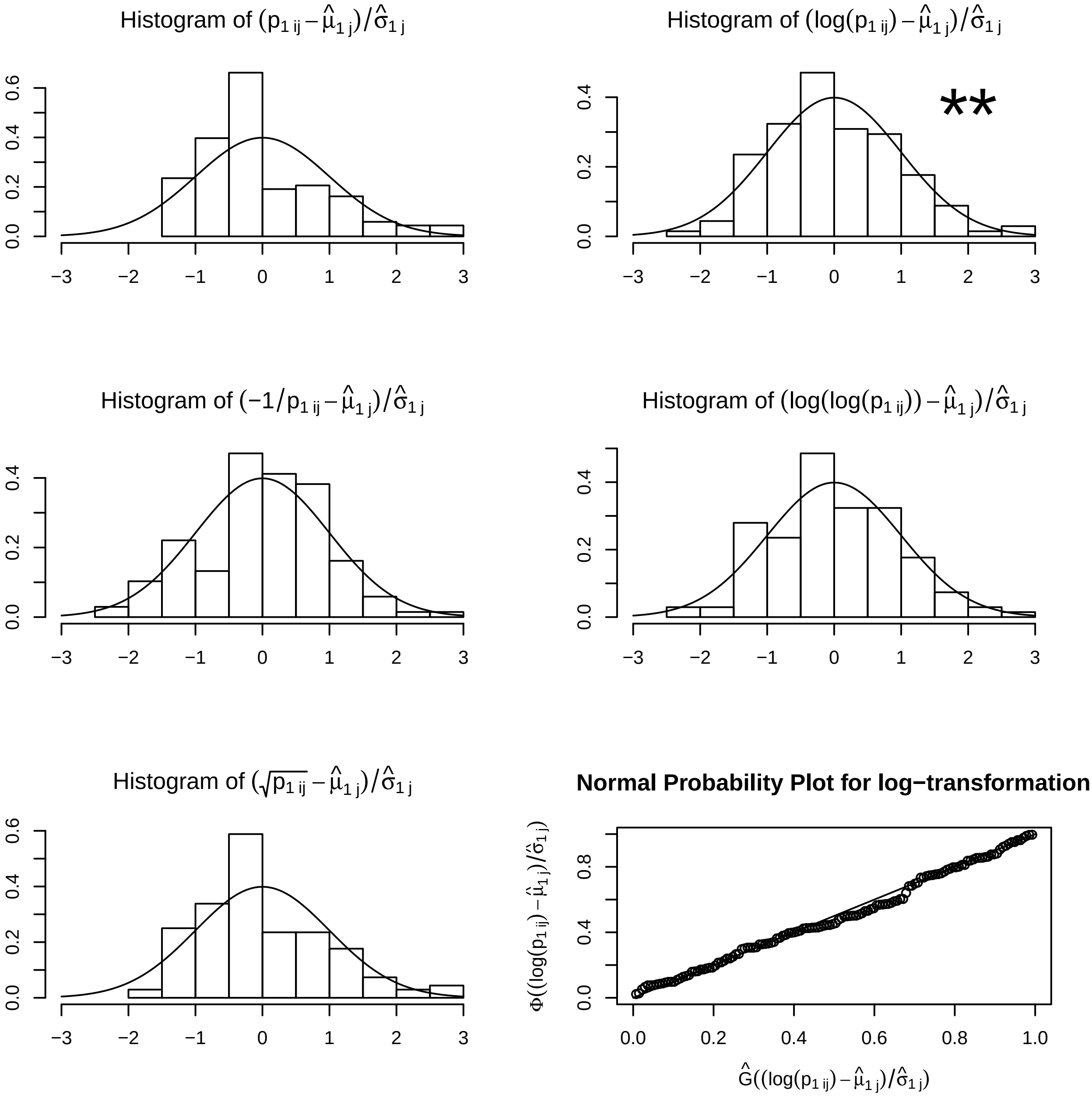

Figure 3.1: Histograms of $\frac{\varphi_u(p_{1ij})-\hat{\mu}_{1j}}{\hat{\sigma}_{1j}}$ for $u = 1, \ldots, 5$. ** displays chosen transformation. The lower right graphic displays the Normal Probability Plot for the chosen transformation.

As displayed in Figures 3.1,..., 3.5 the chosen transformations for triangles $1, \ldots, 5$ seem to fit the normal assumption well and therefore the graphical tools justify the theoretical results obtained from the Shapiro-Wilks test for normality.

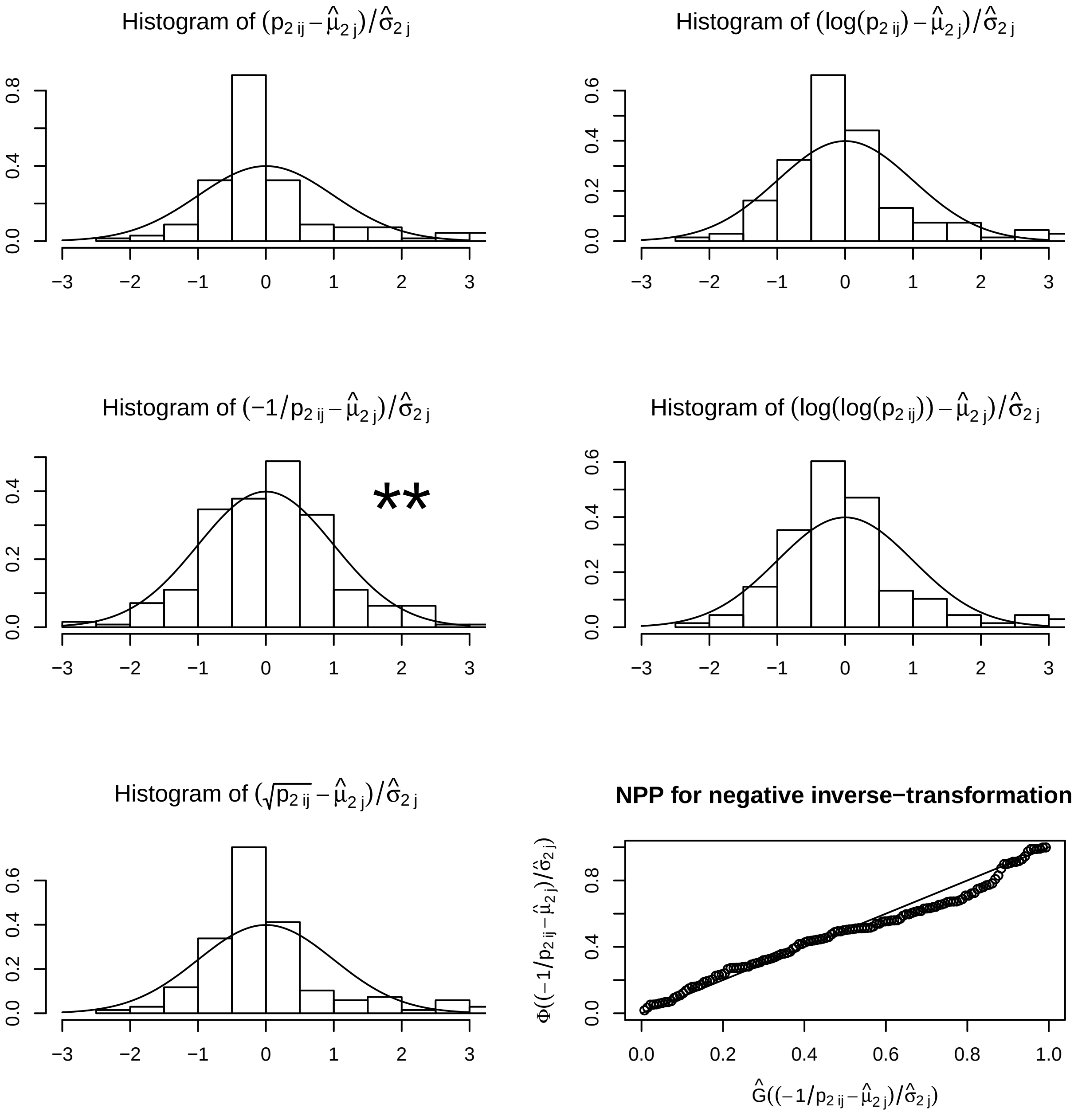

Figure 3.2: Histograms of $\frac{\varphi_u(p_{2ij})-\hat{\mu}_{2j}}{\hat{\sigma}_{2j}}$ for $u = 1, \ldots, 5$. ** displays chosen transformation. The lower right graphic displays the Normal Probability Plot for the chosen transformation.

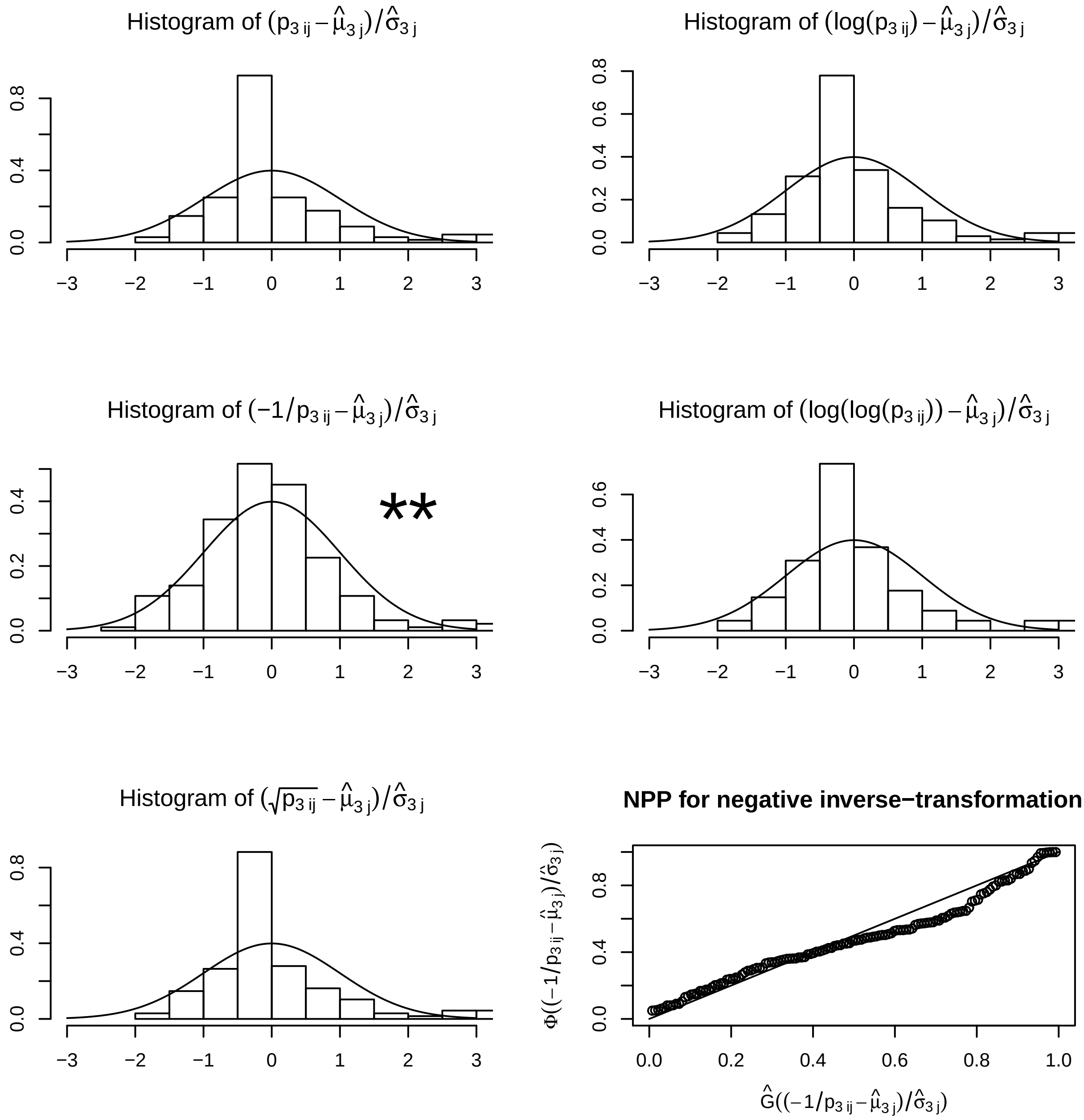

Figure 3.3: Histograms of $\frac{\varphi_u(p_{3ij}) - \hat{\mu}_{3j}}{\hat{\sigma}_{3j}}$ for $u = 1, \ldots, 5$. ** displays chosen transformation. The lower right graphic displays the Normal Probability Plot for the chosen transformation.

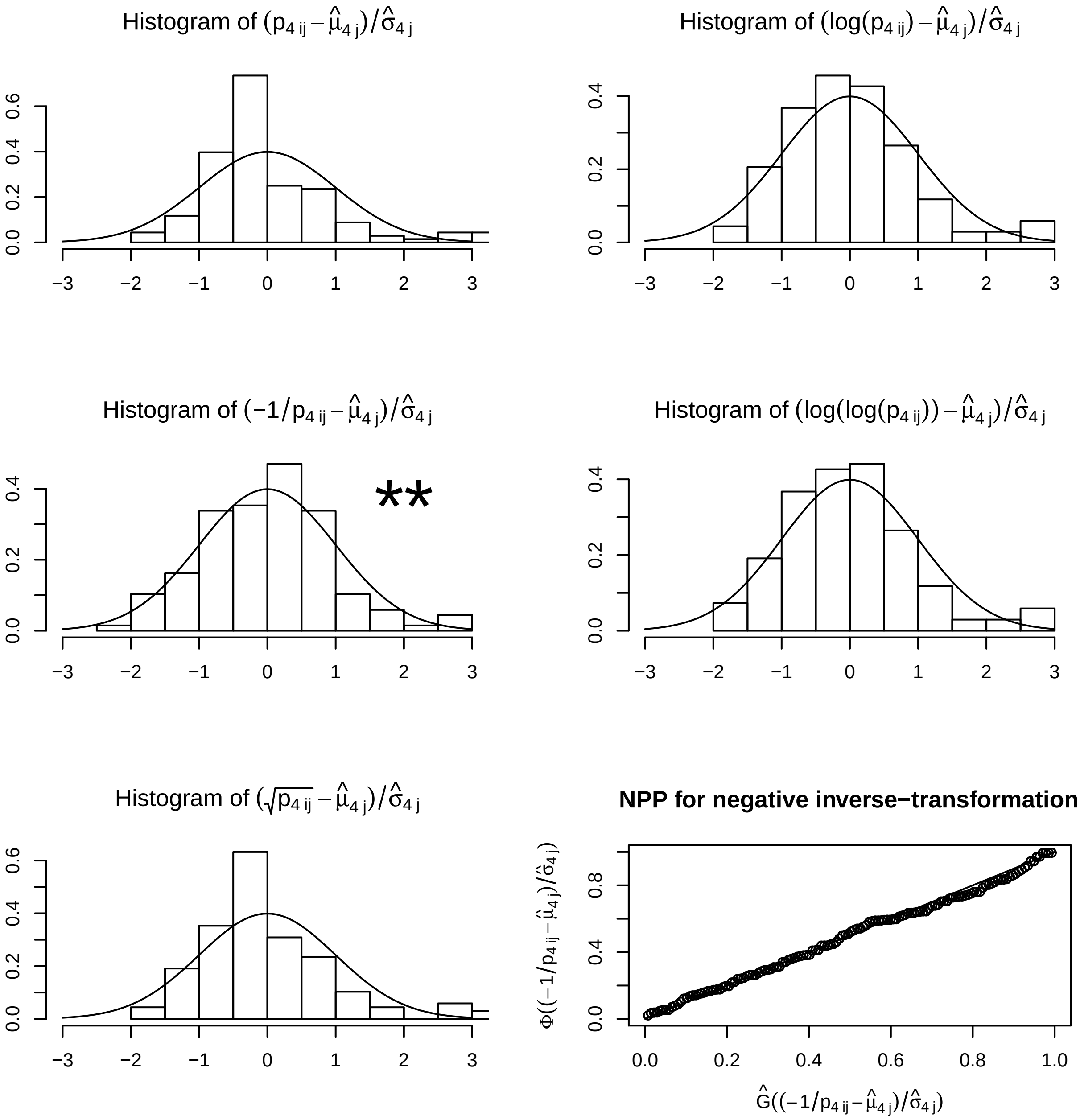

Figure 3.4: Histograms of $\frac{\varphi_u(p_{4ij})-\hat{\mu}_{4j}}{\hat{\sigma}_{4j}}$ for $u = 1, \ldots, 5$. ** displays chosen transformation. The lower right graphic displays the Normal Probability Plot for the chosen transformation.

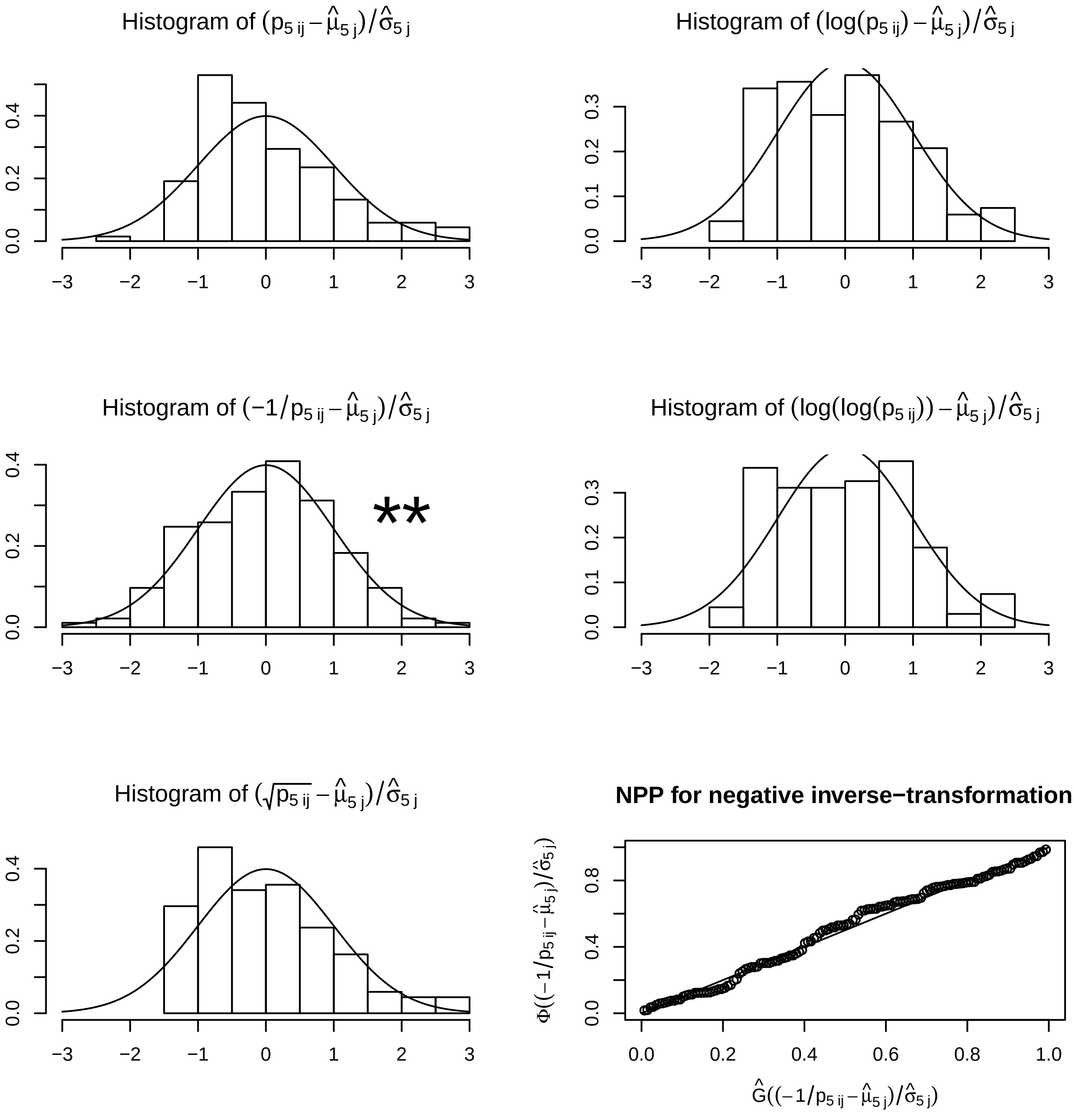

Figure 3.5: Histograms of $\frac{\varphi_u(p_{5ij})-\hat{\mu}_{5j}}{\hat{\sigma}_{5j}}$ for $u = 1, \dots, 5$. ** displays chosen transformation. The lower right graphic displays the Normal Probability Plot for the chosen transformation.

3.4 Simulation Study

3.4.1 General Setup

In this Section the 3 proposed methods of estimating the model parameters $s_1, \dots, s_m$ and $\mathbf{R}$ will be compared. These methods are the heuristic approach, the maximum-likelihood based estimation and a stepwise maximum-likelihood estimation.
In order to get insights on the accuracy of each of these methods a simulation study is performed as follows:

- For a given set of predefined parameters $s_1, \dots, s_m$, $\mathbf{R} \in \mathbb{R}^{m \times m}$, m loss triangles of length n will be simulated. Thus, for each triangle $\frac{n(n+1)}{2}$ entries are simulated. Each of the three estimation methods will be applied to the simulated data and the obtained estimates $s_1^H, \dots, s_m^H, \hat{\mathbf{R}}^H$ for the heuristic estimation, $s_1^{ML}, \dots, s_m^{ML}, \hat{R}^{ML}$ for the maximum-likelihood approach and $s_1^{SML}, \dots, s_m^{SML}, \hat{R}^{SML}$ for the stepwise maximum-likelihood estimation will be compared with the predefined set of parameters $s_1, \dots, s_m, \mathbf{R}$.

- In the performed study there will be a total of $M = 9$ different prespecified parameter configurations. These configurations will be combinations of the specifity vector $\mathbf{s} \in \mathbb{R}^m$ and the chosen correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$.
  There is a set of 3 possible settings for $\mathbf{s}$, namely a setting $\mathbf{s}_e$ with constant specifity for each triangle

$$\mathbf{s}_e := \begin{pmatrix} s_{e,1} \\ s_{e,2} \\ \vdots \\ s_{e,m} \end{pmatrix} := \begin{pmatrix} s \\ s \\ \vdots \\ s \end{pmatrix} \in \mathbb{R}^m \ , \ s \in [0,1] \ ,$$

  a setting $\mathbf{s}_g$ with growing value of s_l , $l = 1 \dots, m$

$$\mathbf{s_g} := \begin{pmatrix} s_{g,1} \\ s_{g,2} \\ \vdots \\ s_{g,m} \end{pmatrix} := \begin{pmatrix} s \\ s+a \\ s+2a \\ s+3a \\ \vdots \\ s+(m-1)a \end{pmatrix} \in \mathbb{R}^m \ , \ s \in [0,1) \ , \ a \in (0, \frac{1-s}{m-1}]$$

  and a setting $\mathbf{s}_r$ with random values from $unif(0,1)$ for $s_1, \dots, s_m$. There are 3 different settings for the correlation matrix $\mathbf{R}$ as well:

$$\mathbf{R}_e(\rho) := \begin{pmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \rho & \dots & \vdots \\ \vdots & \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho & \dots & \dots & \rho & 1 \end{pmatrix} \in \mathbb{R}^{m \times m} \ , \ \rho \ \in \ (-1,1) \ ,$$

$$\mathbf{R}_{AR(1)}(\rho) := \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{m-1} \\ \rho & 1 & \ddots & \ddots & \vdots \\ \rho^2 & \ddots & \ddots & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^{m-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix} \in \mathbb{R}^{m\times m} \ , \ \rho \ \in \ (-1,1)$$

and one random setting for the upper-diagonal entries $r_{1,2},\dots,r_{m-1,m}$ of the symmetric correlation matrix $\mathbf{R}$ with the restriction that $\mathbf{R}$ is positive definite. Based on a prespecified setting of $\mathbf{s}$ and $\mathbf{R}$ simulation of loss triangles is accomplished as described in Algortithm 7. The entries of these simulated triangles are interpreted

Algorithm 7 Loss triangle simulation

```
for t = 1 to n do
  draw a_t := (a_1t, ..., a_mt)^T ∈ R^m from N(0, R)
  for l = 1 to m do
    for i = 1 to t do
      j := t − i
      draw e_lij from N(0, 1)
      z_lij := sqrt(1 − s_l^2) a_lt + s_l e_lij
    end for
  end for
end for
```

as the so-called z-scores defined in (2.7) which are the basis for the parameter estimation described in section 2.6.

- For each simulated data set one obtains $\frac{m(m+1)}{2}$ parameter estimates for each of the estimation methods, namely m specifity estimates

$$\hat{s}_1,\dots,\hat{s}_m \ , \ \hat{s}_l \in \{\hat{s}_l^{ML},\hat{s}_l^{H},\hat{s}_l^{SML}\} \ , \ l = 1\dots,m$$

and $\frac{m(m-1)}{2}$ estimates for the upper triangle entries of the correlation matrix $\mathbf{R}$

$$\hat{r}_{1,2},\hat{r}_{1,3},\dots,\hat{r}_{m-1,m} \ , \ \hat{r}_{l,k} \in \{\hat{r}_{l,k}^{MsL},\hat{r}_{l,k}^{H},\hat{r}_{l,k}^{SML}\},$$

where $\hat{s}_l^{ML}$ denotes the maximum-likelihood estimate of s_l , $\hat{s}_l^{H}$ the heuristic estimate and $\hat{s}_l^{SML}$ the stepwise maximum-likelihood estimate of s_l , $\hat{r}_{l,k}^{ML}$ the maximum likelihood estimate of $r_{l,k}$, and so on.

- Comparison of the estimation methods is accomplished by comparing the bias and the mean squared error of the parameter estimates for each of the different parameter configurations and each of the 3 methods of estimating these parameters.

In detail the simulation study is performed for $m = 5$ different triangles of lentgh $n = 16$. Thus for each configuration $\frac{m(m+1)}{2} = 15$ parameters, namely $s_1, \ldots, s_5$ and $r_{1,2}, \ldots, r_{4,5}$, i.e. the upper-diagonal entries of $\mathbf{R} \in \mathbb{R}^{m \times m}$ need to be estimated out of $\frac{m \cdot n(n+1)}{2} = 680$ simulated data points.

configuration	choice of $\mathbf{s}$	choice of $\mathbf{R}$
1	$\mathbf{s}_e$ with $s = 0.7$	$\mathbf{R}_e(\rho)$ with $\rho = 0.2$
2	$\mathbf{s}_g$ with $s = 0.5$ and $a = 0.075$	$\mathbf{R}_e(\rho)$ with $\rho = 0.2$
3	$\mathbf{s}_e$ with $s = 0.7$	$\mathbf{R}_e(\rho)$ with $\rho = 0.7$
4	$\mathbf{s}_g$ with $s = 0.5$ and $a = 0.075$	$\mathbf{R}_e(\rho)$ with $\rho = 0.7$
5	$\mathbf{s}_e$ with $s = 0.7$	$\mathbf{R}_{AR(1)}(\rho)$ with $\rho = 0.2$
6	$\mathbf{s}_g$ with $s = 0.7$ and $a = 0.075$	$\mathbf{R}_{AR(1)}(\rho)$ with $\rho = 0.2$
7	$\mathbf{s}_e$ with $s = 0.7$	$\mathbf{R}_{AR(1)}(\rho)$ with $\rho = 0.7$
8	$\mathbf{s}_e$ with $s = 0.7$ and $a = 0.075$	$\mathbf{R}_{AR(1)}(\rho)$ with $\rho = 0.7$
9	$\mathbf{s}_r$	$\mathbf{R}_r$

Table 3.3: Overview over different parameter configurations used in the simulation study

Table 3.3 displays $M = 9$ different parameter settings. For each of these settings $m = 5$ loss triangles have been simulated $K = 100$ times, which results in a total of $N = 900$ simulations of $m = 5$ different loss triangles.

3.4.2 Dependence structure

Based on Theorem 2.1 the dependence structure between the entries z_{lij} of the $m = 5$ simulated triangles can easily be summarized. Remember that, according to the model, payments are only correlated if they fall into the same calendar year.

dependence	correlation
inner triangle	$1 - s_l^2$
between triangles	$\sqrt{1 - s_l^2}\sqrt{1 - s_k^2} r_{l,k}$

Table 3.4: Dependence structure of simulated triangles based on Theorem 2.1

Table 3.4 displays the general dependence structure between simulated payments falling into the same calendar year. This structure is different for each of the parameter configurations $1, \ldots, 9$. Since in configuration9 the entries of $\mathbf{s}$ and $\mathbf{R}$ are random, comparison is based on configuration $1, \ldots, 8$.

Table 3.5 displays the range of inner- and between-triangle correlation depending on choice of configuration, i.e. an interval with the lowest possible correlation as left end point and the highest possible correlation as right end point of the interval.

For configurations 1,3,5 and 7 the inner-triangle correlation is at the constant level 0.51 since for these configurations $\mathbf{s}_e$ with $s = 0.7$, $1 - 0.7^2 = 0.51$ has been chosen. The between-triangle correlation is constant only for configurations 1 and 3 because only for

configuration	inner-triangle correlation	between-triangle correlation
1	[0.51 ; 0.51]	[0.10 ; 0.10]
2	[0.36 ; 0.75]	[0.08 ; 0.14]
3	[0.51 ; 0.51]	[0.36 ; 0.36]
4	[0.36 ; 0.75]	[0.28 ; 0.49]
5	[0.51 ; 0.51]	[0.001 ; 0.10]
6	[0.36 ; 0.75]	[0.001 ; 0.14]
7	[0.51 ; 0.51]	[0.18 ; 0.18]
8	[0.36 ; 0.75]	[0.12 ; 0.49]

Table 3.5: Dependence structure for configurations 1,...,8 by range of correlations

these $\mathbf{s}_e$ and $\mathbf{R}_e(\rho)$ are picked. The different levels of this constant correlation arises due to the fact that in configuration 1 $\rho = 0.2$ and in configuration 3 $\rho = 0.7$ is chosen.

Inner triangle correlation varies between $0.36 = 1 - 0.8^2$ $(0.8 = s_{g,5})$ and $0.75 = 1 - 0.5^2(0.5 = s_{g,1}$ for configurations 2,4,6 and 8). Correlation between triangles is very small especially for configurations 5 and 6. For i.e. configuration 5 payments z_{1ij} and z_{5ij} in triangles 1 and 5 have correlation 0.001, since $\sqrt{1-0.7^2}\sqrt{1-0.7^2}0.2^4 = 0.00082 \approx 0.001$. Between-triangle dependence, hence, is quite strong for configurations 3,4 and 8.

3.4.3 Discussion of the Results

Comparison of the three proposed estimation methods is based on comparing the relative bias and the relative mean squared error of the estimates.

Definition 3.1 (Bias)**.** *Shao [2007] In an estimation problem, the bias of an estimator $\hat{\theta}$ of a real-valued parameter θ of the unknown population is defined to be*

$$b(\theta) := E[\hat{\theta}] - \theta.$$

An estimator $\hat{\theta}$ is said to be unbiased for θ if and only if $b(\theta) = 0$.

Definition 3.2 (Mean squared error)**.** *Shao [2007] The mean squared error (MSE) of $\hat{\theta}$ as an estimator of θ is defined to be*

$$MSE(\theta) := E[(\hat{\theta} - \theta)^2] = b(\theta)^2 + Var(\hat{\theta}).$$

Low bias b and low mean squared MSE error for a certain estimation method indicate that this method estimates the true but unknown parameters well.

According to the tables presented in Figure 3.6 and Figure 3.7 each of the estimation methods performs quite well in estimating the specifities $s_1, \dots, s_5$. In terms of relative bias the stepwise maximum likelihood approach has a slight advantage compared to the other approaches. The heuristic as well as the maximum likelihood method underestimate the true value of s in almost every configuration whereas no obvious pattern can be determined for the stepwise maximum likelihood method. Taking into account the relative

rBias	Parameter	s_1	s_2	s_3	s_4	s_5
conf. 1	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	-0.060	-0.057	-0.067	-0.067	-0.074
	ML	-0.010	-0.009	-0.017	-0.023	-0.031
	sML	**0.004**	**0.001**	**-0.006**	**-0.011**	**-0.014**
conf. 2	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	-0.060	-0.059	-0.054	-0.059	-0.061
	ML	-0.002	-0.007	-0.006	-0.012	-0.020
	sML	**0.002**	**0.002**	**0.006**	**0.001**	**-0.001**
conf. 3	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	-0.063	-0.066	-0.051	-0.064	-0.060
	ML	0.004	-0.010	**0.001**	-0.010	-0.007
	sML	**-0.001**	**-0.004**	0.009	**-0.001**	**0.003**
conf. 4	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	-0.058	-0.052	-0.065	-0.054	-0.070
	ML	**-0.002**	**0.003**	-0.009	**-0.001**	-0.025
	sML	0.004	0.009	**-0.002**	0.011	**-0.009**
conf. 5	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	-0.064	-0.066	-0.057	-0.080	-0.067
	ML	-0.020	-0.020	-0.013	-0.030	-0.019
	sML	**-0.004**	**-0.009**	**0.001**	**-0.017**	**-0.006**
conf. 6	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	-0.062	-0.059	-0.069	-0.066	-0.065
	ML	-0.006	-0.003	-0.017	-0.019	-0.021
	sML	**-0.002**	**0.003**	**-0.009**	**-0.006**	**-0.003**
conf. 7	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	-0.063	-0.074	-0.059	-0.063	-0.064
	ML	-0.009	-0.021	-0.009	-0.013	-0.014
	sML	**-0.001**	**-0.011**	**0.001**	**-0.001**	**-0.001**
conf. 8	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	-0.056	-0.070	-0.071	-0.070	-0.070
	ML	**0.002**	**-0.017**	**-0.012**	**-0.018**	**-0.023**
	sML	0.040	0.021	0.069	-0.044	-0.099
conf. 9	true value					
	Heuristic	-0.054	-0.061	-0.071	-0.065	-0.066
	ML	-0.021	-0.018	-0.019	-0.018	-0.018
	sML	**-0.011**	**-0.005**	**0.002**	**-0.005**	**0.005**

Figure 3.6: Relative Bias of the specifity estimates s_1,...,s_5 ordered by configuration and estimation method ; best values bolded

rMSE	Parameter	s_1	s_2	s_3	s_4	s_5
conf. 1	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.0054	0.0049	0.0064	0.0060	0.0063
	ML	**0.0030**	**0.0026**	**0.0037**	0.0031	0.0031
	sML	0.0031	0.0029	0.0036	**0.0030**	**0.0029**
conf. 2	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	0.0036	0.0040	0.0048	0.0054	0.0059
	ML	0.0022	**0.0021**	**0.0028**	**0.0030**	0.0030
	sML	**0.0020**	0.0023	0.0031	0.0033	**0.0028**
conf. 3	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.0051	0.0057	0.0041	0.0054	0.0054
	ML	0.0036	0.0030	0.0026	0.0030	0.0031
	sML	**0.0027**	**0.0030**	**0.0026**	**0.0027**	**0.0031**
conf. 4	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	0.0038	0.0040	0.0052	0.0051	0.0074
	ML	0.0026	0.0028	0.0032	**0.0033**	0.0039
	sML	**0.0024**	**0.0026**	**0.0029**	0.0036	**0.0038**
conf. 5	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.0057	0.0057	0.0047	0.0063	0.0051
	ML	**0.0029**	0.0034	0.0024	0.0027	0.0023
	sML	0.0030	**0.0029**	**0.0024**	**0.0023**	**0.0023**
conf. 6	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	0.0038	0.0038	0.0054	0.0058	0.0056
	ML	0.0022	0.0023	0.0026	0.0030	0.0021
	sML	**0.0022**	**0.0021**	**0.0025**	**0.0029**	**0.0020**
conf. 7	true value	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.0049	0.0069	0.0049	0.0050	0.0060
	ML	0.0024	0.0039	0.0029	0.0029	0.0034
	sML	**0.0023**	**0.0039**	**0.0027**	**0.0026**	**0.0033**
conf. 8	true value	0.5	0.575	0.65	0.725	0.8
	Heuristic	0.0036	0.0045	0.0054	0.0066	0.0068
	ML	**0.0024**	**0.0021**	**0.0026**	**0.0033**	**0.0033**
	sML	0.0086	0.0031	0.0071	0.0041	0.0094
conf. 9	true value					
	Heuristic	0.0046	0.0044	0.0058	0.0057	0.0053
	ML	0.0239	0.0214	0.0233	0.0207	0.0202
	sML	**0.0025**	**0.0016**	**0.0022**	**0.0025**	**0.0023**

Figure 3.7: Relative Mean Squared Error of the specifity estimates $s_1, ..., s_5$ ordered by configuration and estimation method ; best values bolded

	Parameter	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{1,5}$	$r_{2,3}$	$r_{2,4}$	$r_{2,5}$	$r_{3,4}$	$r_{3,5}$	$r_{4,5}$
conf. 1	true value	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	Heuristic	-0.110	**-0.155**	**0.055**	**-0.235**	**-0.050**	**0.055**	**-0.140**	**-0.100**	**-0.180**	**-0.045**
	ML	-0.860	-0.915	-0.610	-0.810	-0.695	-0.555	-0.885	-0.710	-0.695	-0.765
	sML	**0.070**	-0.165	-0.185	-0.265	0.135	0.075	0.160	0.220	-0.205	-0.080
conf. 2	true value	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	Heuristic	**-0.255**	-0.080	**0.080**	**0.070**	-0.125	**-0.180**	-0.145	**-0.125**	**-0.150**	**-0.147**
	ML	-0.910	-0.650	-0.405	-0.605	-0.750	-0.915	-0.610	-0.740	-0.995	-0.825
	sML	-0.300	**-0.025**	0.110	-0.075	**-0.035**	-0.190	**-0.105**	-0.260	-0.270	-0.250
conf. 3	true value	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Heuristic	**-0.161**	**-0.171**	-0.229	**-0.174**	**-0.219**	**-0.174**	**-0.167**	**-0.203**	**-0.176**	**-0.187**
	ML	-0.327	-0.291	-0.459	-0.271	-0.451	-0.369	-0.283	-0.440	-0.374	-0.427
	sML	-0.223	-0.210	-0.211	-0.216	-0.234	-0.263	-0.249	-0.219	-0.224	-0.244
conf. 4	true value	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Heuristic	**-0.163**	**-0.151**	**-0.167**	**-0.183**	**-0.186**	-0.223	**-0.196**	**-0.201**	-0.220	-0.256
	ML	-0.301	-0.279	-0.357	-0.294	-0.377	-0.414	-0.336	-0.369	-0.474	-0.490
	sML	-0.226	-0.190	-0.207	-0.199	-0.209	**-0.219**	-0.223	-0.216	**-0.206**	**-0.200**
conf. 5	true value	0.2	0.04	0.008	0.0016	0.2	0.04	0.008	0.2	0.04	0.2
	Heuristic	0.085	**-0.145**	**-1.625**	**3.750**	**-0.105**	**-0.100**	**1.375**	**-0.110**	**-0.322**	**-0.124**
	ML	-0.470	-0.875	-4.125	-9.875	-0.815	-0.950	-4.750	-0.745	-2.375	-0.425
	sML	**0.080**	-0.175	-2.750	4.875	0.160	-0.234	1.575	0.245	-0.443	-0.145
conf. 6	true value	0.2	0.04	0.008	0.0016	0.2	0.04	0.008	0.2	0.04	0.2
	Heuristic	**-0.045**	**0.150**	**-1.500**	**2.020**	**-0.124**	**-0.200**	**0.625**	**-0.035**	**-0.275**	**0.040**
	ML	-0.370	-0.852	-4.125	-8.250	-0.465	-1.225	-3.625	-0.675	-1.275	-0.250
	sML	-0.075	0.225	-2.250	-5.032	0.163	-0.325	-1.375	0.285	-0.910	0.070
conf. 7	true value	0.7	0.49	0.34	0.24	0.7	0.49	0.34	0.7	0.34	0.7
	Heuristic	**-0.157**	-0.216	**-0.397**	**-0.279**	**-0.143**	**-0.267**	**-0.091**	**-0.083**	**-0.282**	**-0.169**
	ML	-0.294	-0.592	-0.835	-1.175	-0.484	-0.592	-0.647	-0.443	-0.853	-0.291
	sML	-0.201	**-0.198**	-0.641	-0.608	0.213	-0.406	-0.185	-0.166	-0.424	-0.223
conf. 8	true value	0.7	0.49	0.34	0.24	0.7	0.49	0.34	0.7	0.34	0.7
	Heuristic	**-0.116**	**-0.143**	**-0.221**	**-0.354**	-0.117	**-0.288**	**-0.271**	**-0.151**	**-0.221**	-0.276
	ML	-0.233	-0.480	-0.556	-1.104	-0.376	-0.600	-0.776	-0.476	-1.015	-0.463
	sML	-0.223	-0.165	-0.394	-0.754	**0.111**	-0.376	0.271	0.279	-0.303	**-0.227**
conf. 9	true value	random	random	random	random	random	random	random	random	random	random
	Heuristic	**-0.162**	**-0.140**	**-0.150**	**-0.170**	-0.164	**-0.182**	-0.184	**-0.212**	**-0.186**	**-0.164**
	ML	-0.326	-0.470	-0.378	-0.530	-0.526	-0.588	-0.528	-0.666	-0.690	-0.648
	sML	-0.312	-0.162	-0.268	-0.262	**0.156**	-0.318	**-0.182**	-0.290	-0.206	-0.318

Figure 3.8: Relative Bias of the correlation estimates r_{lk} ordered by configuration and estimation method ; best values bolded

	Parameter	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{1,5}$	$r_{2,3}$	$r_{2,4}$	$r_{2,5}$	$r_{3,4}$	$r_{3,5}$	$r_{4,5}$
conf. 1	true value	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	Heuristic	0.340	**0.345**	**0.335**	**0.295**	0.330	0.350	0.370	**0.335**	**0.325**	**0.285**
	ML	0.675	0.675	0.590	0.595	0.570	0.550	0.665	0.560	0.495	0.540
	sML	**0.320**	0.395	0.345	0.365	**0.300**	**0.275**	**0.260**	0.340	0.330	0.295
conf. 2	true value	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	Heuristic	**0.320**	0.320	0.370	**0.275**	**0.295**	**0.325**	**0.380**	0.405	**0.340**	**0.405**
	ML	0.535	0.500	0.490	0.430	0.565	0.540	0.600	0.610	0.650	0.645
	sML	0.330	**0.310**	**0.270**	0.385	0.315	0.340	0.400	**0.300**	0.385	0.411
conf. 3	true value	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.071	0.064	0.091	0.070	0.093	**0.071**	**0.071**	0.090	**0.070**	**0.087**
	ML	0.234	0.174	0.261	0.169	0.257	0.193	0.164	0.260	0.211	0.256
	sML	**0.070**	**0.049**	**0.057**	**0.046**	**0.063**	0.087	0.081	**0.053**	0.073	0.088
conf. 4	true value	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Heuristic	0.071	0.067	0.069	0.074	0.076	0.086	0.079	0.071	0.084	0.104
	ML	0.176	0.181	0.190	0.150	0.209	0.239	0.203	0.187	0.277	0.283
	sML	**0.056**	**0.044**	**0.049**	**0.046**	**0.043**	**0.049**	**0.050**	**0.057**	**0.051**	**0.050**
conf. 5	true value	0.2	0.04	0.008	0.0016	0.2	0.04	0.008	0.2	0.04	0.2
	Heuristic	0.342	**1.153**	**2.502**	**5.622**	**0.365**	**1.425**	**2.125**	**0.355**	**1.250**	0.350
	ML	0.555	3.025	6.253	12.428	0.605	2.625	6.137	0.485	2.425	0.402
	sML	**0.295**	2.125	4.625	8.757	0.390	1.974	3.125	0.515	2.025	**0.315**
conf. 6	true value	0.2	0.04	0.008	0.0016	0.2	0.04	0.008	0.2	0.04	0.2
	Heuristic	0.305	**1.525**	**2.562**	**7.500**	**0.357**	**1.314**	**2.375**	**0.390**	**1.275**	**0.355**
	ML	0.355	1.850	7.000	15.625	0.562	3.052	5.625	0.620	2.925	0.570
	sML	**0.265**	1.600	4.375	9.375	0.424	2.025	2.983	0.525	1.924	0.370
conf. 7	true value	0.7	0.49	0.34	0.24	0.7	0.49	0.34	0.7	0.49	0.7
	Heuristic	0.050	0.147	**0.182**	**0.279**	0.084	**0.106**	**0.221**	0.119	0.197	0.076
	ML	0.151	0.388	0.518	0.833	0.291	0.373	0.488	0.297	0.547	0.154
	sML	**0.039**	**0.104**	0.212	0.329	**0.034**	0.163	0.256	**0.070**	**0.118**	**0.064**
conf. 8	true value	0.7	0.49	0.34	0.24	0.7	0.49	0.34	0.7	0.49	0.7
	Heuristic	0.059	**0.108**	**0.079**	**0.238**	**0.059**	**0.108**	**0.215**	0.120	**0.119**	**0.069**
	ML	0.107	0.294	0.253	0.679	0.206	0.343	0.415	0.309	0.635	0.296
	sML	**0.050**	0.136	0.203	0.313	0.063	0.112	0.279	**0.081**	0.182	0.073
conf. 9	true value										
	Heuristic	**0.079**	**0.070**	**0.082**	**0.063**	**0.075**	**0.089**	**0.081**	**0.073**	**0.078**	**0.072**
	ML	0.222	0.235	0.196	0.202	0.223	0.212	0.229	0.242	0.227	0.233
	sML	0.191	0.106	0.107	0.130	0.241	0.108	0.102	0.113	0.087	0.109

Figure 3.9: Relative MSE of the correlation estimates r_{lk} ordered by configuration and estimation method ; best values bolded

MSE yields similar results.
Again, the stepwise maximum likelihood estimation dominates the other methods, for configuration 1,2 and 8, however, the maximum likelihood approach performs slightly better.
The results for estimation of the entries of the correlation matrix $\mathbf{R} \in \mathbb{R}^{m \times m}$ are contained in the tables presented in Figure 3.8 and Figure 3.9.
These tables show that the maximum likelihood estimation performs very poorly. The reason for this is the fact, that according to the model definition (2.1) the same α_{lt} , $l = 1, \ldots, m$, $t = 1, \ldots, n$, enters each value of triangle l on the t-th diagonal.
Hence, in the special case of $m = 5$ triangles and $n = 16$ only $5 \cdot 16 = 80$ values of α_{lt} are available for estimating $m \cdot (m+1)/2 = 10$ correlations r_{lk} , $l, k = 1, \ldots, m$. The same drawback occurs for the stepwise maximum likelihood approach, but as mentioned in section 2.6.3 this approach estimates less parameters with the same amount of data and henceforth the obtained estimates are better in a certain sense. Even though the heuristic approach performs best in estimating the correlation matrix $\mathbf{R}$ the results are rather unsatisfactory since even this method has high relative bias and relative MSE.
But there are several possibilities to improve the results, such as regarding "longer" triangles, i.e. triangles for which more data is available, or to apply the Gaussian copula model to less than $m = 5$ triangles, since the number of different triangles m enters quadratically into the set of parameters to be estimated, which is $m(m+1)/2$.
These drawbacks and the fact that the model described in this chapter is limited to Gaussian copulas only leads to a more general approach which is described in the following chapter.

3.5 Parameter Estimation

The results of application of the heuristic, maximum-likelihood and stepwise maximum-likelihood approach to the $m = 5$ loss triangles containing real insurance data are displayed in Table 3.6. For reasons of secrecy the estimated values of α_{lt} will not be displayed in this work.

Heuristic Model	1	2	3	4	5	s_l
triangle 1	**0.25**	**0.14**	**0.10**	**0.18**	**0.20**	0.86
triangle 2	0.44	**0.42**	**0.01**	**0.00**	**0.30**	0.76
triangle 3	0.31	0.01	**0.44**	**0.31**	**0.06**	0.75
triangle 4	0.60	0.01	0.80	**0.34**	**0.13**	0.81
triangle 5	0.64	0.73	0.13	0.36	**0.39**	0.78
ML Model	1	2	3	4	5	s_l
triangle 1	**0.22**	**−0.07**	**−0.11**	**−0.06**	**−0.07**	0.88
triangle 2	-0.29	**0.27**	**0.16**	**0.11**	**0.24**	0.86
triangle 3	-0.36	0.49	**0.43**	**0.25**	**0.03**	0.76
triangle 4	-0.30	0.51	0.57	**0.17**	**0.14**	0.91
triangle 5	-0.23	0.78	0.07	0.55	**0.37**	0.80
stepwise ML Model	1	2	3	4	5	s_l
triangle 1	**0.12**	**0.16**	**0.13**	**0.17**	**0.16**	0.94
triangle 2	0.78	**0.33**	**0.20**	**0.20**	**0.29**	0.82
triangle 3	0.61	0.57	**0.36**	**0.22**	**0.08**	0.80
triangle 4	0.96	0.70	0.74	**0.24**	**0.18**	0.87
triangle 5	0.80	0.90	0.24	0.66	**0.31**	0.83

Table 3.6: Estimation results for the heuristic, ML and stepwise ML approach for the $m = 5$ loss triangles. The bolded numbers show the values of $c_l c_k r_{lk}$, the other numbers the estimates of r_{lk}. The specifity estimates s_l are presented in the right column.

As can be seen from Table 3.6 the estimated specifities are at the same level independent of the estimation approach. This is not surprising since this fact has already been indicated by the results of the simulation study presented in the previous section.
The specifities for all triangles are close to 1 which leads to low communalities c_l and low actual correlations between the triangles, since the correlation parameters r_{lk} are scaled by the communalities by Theorem 2.1.
The estimates of the correlations r_{lk} for small values of r_{lk} vary alot between the estimation methods but due to the results of the simulation study and high uncertainty for values around 0 this could be expected. For higher values of r_{lk} and $c_l c_k r_{lk}$ each estimation method yields quite similar estimates. Overall it can be stated that actual correlations $c_l c_l r_{lk}$ between the loss triangles are quite low mainly due to high specifity estimates.

Chapter 4

D-Vine Model for Loss Triangle Dependence

This chapter introduces a D-vine model for loss triangle dependence. In contrast to the model proposed by de Jong [2009] presented in Chapter 2 dependence structure will no longer be restricted to calendar year dependence. The theory about vines presented in the latter is mainly based on Aas and H.Bakken [2009] and Czado and Min [2010].

4.1 Copulas

Copulas are d-dimensional multivariate distributions with uniformly distributed marginal distributions on $[0, 1]$. They are very useful for modeling a dependence structure of multivariate data.
Sklar's theorem states that given a joint distribution function $F(x_1, ..., x_d)$ for d variables, and respective marginal distribution functions $F_i(x_i), i = 1, ..., d$, there exists a copula C such that the copula binds the margins to give the joint distribution.

Theorem 4.1. ***Sklar's theorem*** *Nelsen [1999]*
Let $\mathbf{X} = (X_1, ..., X_d)^t \in \mathbb{R}^d$ *be a d-dimensional random vector with joint distribution function* $F(x_1, ..., x_d)$ *and marginal distributions* $F_i(x_i), i = 1, ..., d$. *Then there exist a copula* $C(u_1, ..., u_d)$ *such that*

$$F(x_1, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d))$$

for every $x_1, \ldots, x_d \in \mathbb{R}$. *The copula* $C(u_1, ..., u_d)$ *is unique if the marginal distributions are continuous.*

More details can be found in the books by Joe [1997] and Nelsen [1999]. From now on we only consider absolutely continuous distributions with a joint density function $f(x_1, ..., x_d)$ and marginal density functions $f_i(x_i)$ for $i = 1, \ldots, d$.

The copula $C(u_1, ..., u_d)$ of a multivariate distribution $F(x_1, ..., x_d)$ with margins $F_i(x_i)$, $i = 1, \ldots, d$ is given by

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \ , \tag{4.1}$$

where the F_i^{-1}'s , $i = 1 \ldots, d$ are the inverse distribution functions of the marginals. The copula density received by differentiating (4.1) is given by

$$c(u_1, u_2, \ldots, u_d) = \frac{f(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d))}{f_1\left(F_1^{-1}(u_1)\right) \cdots f_d\left(F_d^{-1}(u_d)\right)} \tag{4.2}$$

where $F_i^{-1}(u_i)$ is the inverse of the margins $F_i(x_i)$ for $i = 1, ..., d$. Using **Sklar's theorem** , the multivariate density $f(x_1, ..., x_d)$ is a product of the corresponding copula density with marginal densities $f_i(x_i), i = 1, ..., d$ and is given by

$$f(x_1, ..., x_d) = c(F_1(x_1), ...F_d(x_d d)) \cdot f_1(x_1) \cdots f_d(x_d), \tag{4.3}$$

thus separating the dependence structure from the marginal structure.

4.2 PCC's for Multivariate Distributions

Using **Sklar's theorem** multivariate distributions with given margins can be easily constructed. However this general approach does not give a solution for the construction of flexible multivariate distributions which fit desired data well. In this section we give such a construction proposed first by Joe [1996], organized by Bedford [2002] and applied to Gaussian copulas only. Later Aas and H.Bakken [2009] used bivariate Gaussian, t, Gumbel and Clayton copulas as building blocks. Let $f(x_1, ..., x_d)$ be a d-dimensional density function and $c(u_1, ..., u_d)$ be the corresponding copula density function. In the sequel we denote by $f_{j|i}(x_j|\mathbf{x_i})$ the conditional density of x_j given $\mathbf{x_i} := (X_{i1}, \ldots, X_{ik})'$ for $i := (i_1, \ldots, i_k)'$. It is well known that the density $f(x_1, \ldots, x_d)$ can be factorized as

$$f(x_1, \ldots, x_d) = f_d(x_d) \cdot f_{d-1|d}(x_{d-1}|x_d) \cdot f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d) \cdot \cdots \cdot f_{1|2\cdots d}(x_1|x_2, \ldots, x_d). \tag{4.4}$$

The above factorization is a simple consequence from the definition of conditional densities. The second factor $f_{d-1|d}(x_{d-1}|x_d)$ on the right hand side of (4.4) can be represented as a product of a copula density and the marginal density $f_d(x_d)$ in the following way. Consider the bivariate density function $f_{(d-1)d}(x_{d-1}, x_d)$ with marginal densities $f_{d-1}(x_{d-1})$ and $f_d(x_d)$, respectively. Using Sklar's Theorem for $d = 2$, we have that the conditional density $f_{d-1|d}(x_{d-1}|x_d)$ is given by

$$\begin{aligned} f_{d-1|d}(x_{d-1}|x_d) &= \frac{f_{(d-1)d}(x_{d-1}, x_d)}{f_d(x_d)} \\ &= c_{(d-1)d}(F_{d-1}(x_{d-1}), F_d(x_d)) \cdot f_{d-1}(x_{d-1}) \end{aligned}$$

Similarly, the conditional density $f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d)$ is given by

$$f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d) = \frac{f_{d-2(d-1)|d}(x_{d-2}, x_{d-1}|x_d)}{f_{d-1|d}(x_{d-1}|x_d)}$$

$$= \frac{c_{(d-2)(d-1)|d}(F_{d-2|d}(x_{d-2}|x_d), F_{d-1|d}(x_{d-1}|x_d)) \cdot f_{d-2|d}(x_{d-2}|x_d) \cdot f_{d-1|d}(x_{d-1}|x_d)}{f_{d-1|d}(x_{d-1}|x_d)}$$

$$= c_{(d-2)(d-1)|d}(F_{d-2|d}(x_{d-2}|x_d), F_{d-1|d}(x_{d-1}|x_d)) \cdot c_{(d-2)d}(x_{d-2}, x_d) \cdot f_{d-2}(x_{d-2}). \quad (4.5)$$

Copula density $c_{(d-2)(d-1)|d}(\cdot,\cdot)$ is the conditional copula density corresponding to the conditional distribution $F_{(d-2)(d-1)|d}(x_{d-2}, x_{d-1}|x_d)$. Further $F_{d-i|d}(x_{d-i}|x_d)$ is the conditional distribution function of x_{d-i} given x_d for $i = 1, 2$.
Note that in general the conditional copula density $c_{(d-2)(d-1)|d}(F_{d-2|d}(x_{d-2}|x_d), F_{d-1|d}(x_{d-1}|x_d))$ depends on the given conditioning value x_d. Relation (4.5) can be generalized for a conditioning vector $\mathbf{v}$ of dimension $k(1 < k < d-1)$. Here the starting point is

$$f_{xv_j|\mathbf{v}_{-j}}(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}(F_{x|\mathbf{v}_{-j}}(x|\mathbf{v}_{-j}), F_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j}) \cdot f_{xv_j|\mathbf{v}_{-j}}(x|\mathbf{v}_{-j}) \ ,$$

where v_j is an arbitrary chosen component of $\mathbf{v}$ and the $(k-1)$-dimensional vector $\mathbf{v_{-j}}$ is the vector $\mathbf{v}$ without the component v_j . Finally we can represent each conditional density term on the right hand side of (4.4) as the product of the corresponding marginal density and copula density terms. This shows that $f(x_1, \ldots, x_d)$ is the product of marginal densities and paircopula density terms. The pair-copula density terms are unconditional copulas evaluated at marginal distribution function values or conditional copulas evaluated at univariate conditional distribution function values. The above construction was defined in Aas and H.Bakken [2009] and was called the pair copula construction (PCC) for multivariate distributions. Joe [1996] showed that the conditional distribution function $F_{u|\mathbf{v}}(u|\mathbf{v})$ appearing in the PCC are partial derivatives with respect to the second argument of the conditional copula given by

$$F_{x|\mathbf{v}}(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}. \quad (4.6)$$

Here $C_{x,v_j|\mathbf{v}_{-j}}(\cdot,\cdot)$ is a bivariate copula distribution function. It is clear that there are many pair copula constructions for a random vector $\mathbf{X}$. In order to systemize PCCs, Bedford [2002] introduced tree representations called regular vines. In this work we consider only a particular regular vines, namely D-vines (see Kurowicka [2006] or Aas and H.Bakken [2009]). For the convenience of the reader we give here the construction of D-vines for d random variables.

First of all the random variables should be labeled from 1 to d and this labeling should remain fixed. The D-vine consists of $d-1$ trees $T_i, i = 1, \ldots, d-1$. Figure 4.1 displays the tree representation of a D-vine for 5 variables on which the construction of D-vine is below illustrated. The first tree consists of the d labeled nodes. The nodes are placed along a line one after another according to the value of their labels. Further $d-1$ edges connect neighboring nodes. Now each edge of the first tree gets its label. The edge label are elements of the symmetric difference of labels of neighboring nodes this edge connects. The symmetric difference of two sets A and B is defined by $A\Delta B := (A \backslash B) \cup (B \backslash A)$. For example the symmetric difference of the sets $\{1\}$ and $\{2\}$ is the set $\{1, 2\}$ and this illustration corresponds to the nodes 1 and 2 of the tree T_1 connected with the edge 12

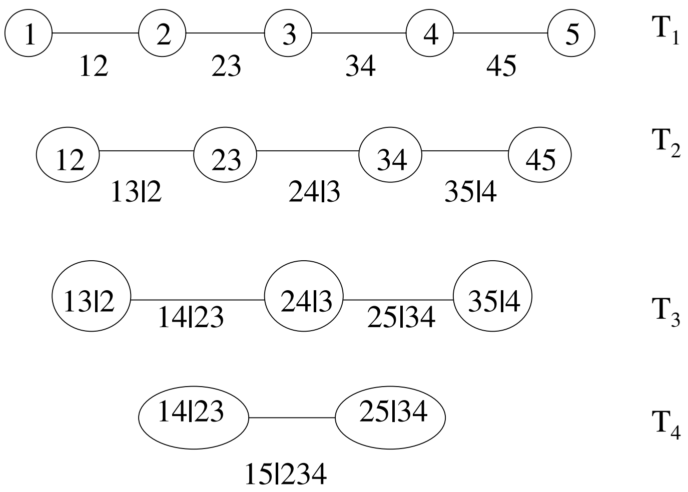

Figure 4.1: A D-vine with 5 variables, 4 trees and 10 edges. Each edge may be may be associated with a pair-copula.

from Figure 4.1. In the second tree edges of the first tree become nodes. Thus the second tree has $d-1$ nodes. The nodes in the second tree are connected with an edge if the corresponding edges in the first tree shared a node. There are altogether $d-2$ edges in the second tree. From now on, the edge labels consist of two label sets separated by a vertical line "|". The first label set before the vertical line is the symmetric difference of labels of neighboring nodes the edge connects. The second label set is made of common labels of the neighboring nodes sharing the edge. Thus the symmetric difference of the sets $\{1,2\}$ and $\{2,3\}$ is the set $\{1,3\}$ and their intersection is the set $\{2\}$. This example corresponds to the nodes 12 and 23 of the tree T_2 connected with the edge $13|2$ in Figure 4.1. In general, the ith tree T_i consists of $d+1-i$ nodes. Nodes are connected with an edge if the corresponding edges in tree T_{i-1} share a node. There are altogether $d-i$ edges. Edges are labeled according to the rule from the second tree treating the two label sets of a node as one set. Thus ignoring the vertical line the labels of the nodes in the third tree T_3 of Figure 2 are obtained as follows. The symmetric difference of the labels of the first two nodes $13|2$ and $24|3$ in Tree T_3 is the set $1,4$ and their interaction is the set $2,3$. Therefore the connecting edge has label $14|23$. The last $(d-1)$-th tree consists of two nodes with labels $1(d-1)|2,...,(d-2)$ and $2d|3,...,(d-1)$, respectively. These two nodes are connected with one edge labeled as $1d|2,...,(d-1)$. Now each edge corresponds to a pair-copula density and its label indicates the subindex of the pair-copula.

Note that the presence of the vertical line in the edge label indicates that the corresponding copula is conditional. Further the second label set after the vertical line corresponds to the set of conditioning variables, while the first label set corresponds to the two variables which will be conditioned. For a d-dimensional density $f(x_1, \ldots, x_d)$ the PCC of the D-vine is given in Aas and H.Bakken [2009] as follows

$$f(x_1, \ldots, x_d) =$$

$$\prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1} \left(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})\right). \quad (4.7)$$

For simplicity I have dropped the subindex of the conditional distribution functions of type $F(x_i|x_{i+1}, \ldots, x_{i+j-1})$. Thus, the PCC representation for D-vines given in (4.7) is the product of d marginal densities and $d(d-1)/2$ bivariate copulas.

4.2.1 Example in Dimension 5

In the special case of $d = 5$ variables the general expression for the D-Vine structure is

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5) &= f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot f(x_5) \\ &\cdot \; c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4)) \\ &\cdot \; c_{45}(F_4(x_4), F_5(x_5)) \\ &\cdot \; c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \\ &\cdot \; c_{35|4}(F(x_3|x_4), F(x_5|x_4)) \\ &\cdot \; c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot c_{25|34}(F(x_2|x_3, x_4), F(x_5|x_3, x_4)) \\ &\cdot \; c_{15|234}(F(x_1|x_2, x_3, x_4), F(x_5|x_2, x_3, x_4)) \end{aligned}$$

4.2.2 PCC with n Variables

Considering Figure (4.1) there are $n!$ choices to arrange the order in the tree T_1 for an n-dimensional D-vine,. Since we have undirected edges, i.e. $c_{ij|D} = c_{ji|D}$ for all pairs i, j and arbitrary conditioning sets for D-vines, we can reverse the order in the tree T_1 for a D-vine without changing the corresponding vine. Therefore we have only $\frac{n!}{2}$ different trees on the first level. Given such a tree T_1, the trees $T_2, T_3, \ldots, T_{n-1}$ are completely determined. This implies that the number of distinct D-vines on n nodes is given by $\frac{n!}{2}$.

4.3 Simulation from a Pair-Copulae decomposed Model

Simulation from vines is briefly discussed in Bedford [2001a], Bedford [2001b] and Kurowicka [2007]. In this section it will be shown that the simulation algorithm for D-vines is straightforward and simple to implement. This simulation algortihm is based on the definition of the so-called h-function (see Aas and H.Bakken [2009]).

4.3.1 Introduction of h-functions

The pair-copula construction involves marginal conditional distributions of the form $F(x|\mathbf{v})$. For every j, Joe [1996] showed that

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})} \tag{4.8}$$

where $C_{ij|k}$ is a bivariate copula distribution function. For the special case where v is univariate we have

$$F(x|v) = \frac{\partial C_{xv}(F_x(x), F_v(v))}{\partial F_v(v)}.$$

In the remainder of this work the function $h(x, v, \boldsymbol{\theta})$ will be used to represent this conditional distribution function when x and v are uniform, i.e. $f(x) = f(v) = 1, F(x) = x$ and $F(v) = v$. That is,

$$h(x, v, \boldsymbol{\theta}) := F(x|v) = \frac{\partial C_{x,v}(x, v, \boldsymbol{\theta})}{\partial v} \ , \tag{4.9}$$

where the second parameter of $h(\cdot)$ always corresponds to the conditioning variable and $\boldsymbol{\theta}$ denotes the set of parameters for the copula of the joint distribution function of x and v. Further, let $h^{-1}(u, v, \boldsymbol{\theta})$ be the inverse of the h-function with respect to the first variable u, or equivalently the inverse of the conditional distribution function.

4.3.2 Sampling from a D-Vine

Algorithm 8 gives the procedure for sampling from the D-vine. It also consists of one main for-loop containing one for-loop for sampling the variables and one for-loop for computing the needed conditional distribution functions. The h-function is defined by (4.9) in section 4.3.1 , but here $\boldsymbol{\theta}_{j,i}$ is the set of parameters of the copula density $c_{i,i+j|i+1,\ldots,i+j-1}(\cdot)$.

4.4 Inference for a specified Pair-Copula Decomposition

In this section it will be described how the parameters of the D-vine density given by (4.7) can be estimated by maximum likelihood. Assume that n variables at T time points are observed. Let $\mathbf{x}_i = (x_{i,1}, ..., x_{i,T}); i = 1, \ldots n$ denote the data set. Here, each random variable $X_{i,t}$ is assumed to be uniform in [0,1]. It is assumed for simplicity that the T observations of each variable are independent over time. This is not a necessary assumption, as stochastic dependencies and time series dynamics can easily be incorporated.
For the D-vine, the log-likelihood is given by

$$\sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \sum_{t=1}^{T} log(c_{i,i+j|i+1,\ldots,i+j-1}(F(x_{i,t}|x_{i+1,t}, \cdots, x_{i+j-1,t}), F(x_{i+j,t}|x_{i+1,t}, \cdots, x_{i+j-1,t}))). \tag{4.10}$$

Algorithm 8 Simulation algorithm for D-vine. Generates one sample $x_1, ...x_n$ from the vine.

```
Sample w_i; i = 1, ..., n independent uniform on [0,1] .
x_1 = v_{1,1} = w_1
x_2 = v_{2,1} = h^{-1}(w_2, v_{1,1}, θ_{1,1})
v_{2,2} = h(v_{1,1}, v_{2,1}, θ_{1,1})
for i = 3 to n do
  v_{i,1} = w_i
  for k = i − 1 to 2 do
    v_{i,1} = h^{-1}(v_{i,1}, v_{i−1,2k−2}, θ_{k,i−k})
  end for
  v_{i,1} = h^{-1}(v_{i,1}, v_{i−1,1}, θ_{1,i−1})
  x_i = v_{i,1}
  if i == n then
    Stop
  end if
  v_{i,2} = h(v_{i−1,1}, v_{i,1}, θ_{1,i−1})
  v_{i,3} = h(v_{i,1}, v_{i−1,1}, θ_{1,i−1})
  if i > 3 then
    for j = 2 to i − 2 do
      v_{i,2j} = h(v_{i−1,2j−2}, v_{i,2j−1}, θ_{j,i−j})
      v_{i,2j+1} = h(v_{i,2j−1}, v_{i−1,2j−2}, θ_{j,i−j})
    end for
  end if
  v_{i,2i−2} = h(v_{i−1,2i−4}, v_{i,2i−3}, θ_{i−1,1})
end for
```

The D-vine log-likelihood must be numerically optimised. Algortihm 9 evaluates the likelihood, where $L(\mathbf{x}, \mathbf{v}, \boldsymbol{\theta})$ is the log-likelihood of the chosen bivariate copula with parameters $\boldsymbol{\theta}$ given the data vectors $\mathbf{x}$ and $\mathbf{v}$. That is,

$$L(\mathbf{x}, \mathbf{v}, \boldsymbol{\theta}) = \sum_{t=1}^{T} log(c(x_t, v_t, \boldsymbol{\theta})) \ , \tag{4.11}$$

where $c(u, v, \boldsymbol{\theta})$ is the density of the bivariate copula with parameters $\boldsymbol{\theta}$ and $\theta_{j,j}$ is the set of parameters of copula density $c_{i,i+j|i+1,...,i+j-1}(\cdot)$.

This, at first sight, non-trivial procedure will be explained using a simple example. Consider, for example a 3-dimensional data set with $U[0, 1]$ distributed variables. Then (4.10) reduces to

$$\sum_{t=1}^{T} \left(log(c_{12}(x_{1,t}, x_{2,t}, \theta_{11})) + log(c_{23}(x_{2,t}, x_{3,t}\theta_{12})) + log(c_{13|2}(v_{1,t}, v_{2,t}, \theta_{21})) \right) ,$$

where

$$v_{1,t} = F(x_{1,t}|x_{2,t}) = h(x_{1,t}, x_{2,t}, \theta_{11})$$

and

$$v_{2,t} = F(x_{3,t}|x_{2,t}) = h(x_{3,t}, x_{2,t}, \theta_{12}).$$

The parameters to be estimated are $\boldsymbol{\theta} = (\theta_{11}, \theta_{12}, \theta_{21})$,where $\theta_{j,i}$ is the set of parameters of the corresponding copula density $c_{i,i+j|i+1,\ldots,i+j-1}(\cdot,\cdot)$. Following the procedure described above, one first estimates the parameters of the three copulae involved by a sequential procedure, and then maximises the full log-likelihood using the parameters obtained from the stepwise procedure as starting values.

Algorithm 9 Likelihood evaluation for a D-vine

log-likelihood = 0
for $i = 1$ to n **do**
  $\mathbf{v}_{0,i} = \mathbf{x}_i$
end for
for $i = 1$ to $n-1$ **do**
  log-likelihood = log-likelihood + $L(\mathbf{v}_{0,i}, \mathbf{v}_{0,i+1}, \theta_{1,i})$
end for
$\mathbf{v}_{1,1} = h(\mathbf{v}_{0,1}, \mathbf{v}_{0,2}, \theta_{1,1})$
for $k = 1$ to $n-3$ **do**
  $\mathbf{v}_{1,2k} = h(\mathbf{v}_{0,k+2}, \mathbf{v}_{0,k+1}, \theta_{1,k+1})$
  $\mathbf{v}_{1,2k+1} = h(\mathbf{v}_{0,k+1}, \mathbf{v}_{0,k+2}, \theta_{1,k+1})$
end for
$\mathbf{v}_{1,2n-4} = h(\mathbf{v}_{0,n}, \mathbf{v}_{0,n-1}, \theta_{1,n-1})$
for $j = 2$ to $n-1$ **do**
  for $i = 1$ to $n-j$ **do**
    log-likelihood = log-likelihood + $L(\mathbf{v}_{j-1,2i-1}, \mathbf{v}_{j-1,2i}, \theta_{j,i})$
  end for
  if $j == n-1$ **then**
    Stop
  end if
  $\mathbf{v}_{j,1} = h(\mathbf{v}_{j-1,1}, \mathbf{v}_{j-1,2}, \theta_{j,1})$
  if $n > 4$ **then**
    for $i = 1$ to $n-j-2$ **do**
      $\mathbf{v}_{j,2i} = h(\mathbf{v}_{j-1,2i+2}, \mathbf{v}_{j-1,2i+1}, \theta_{j,i+1})$
      $\mathbf{v}_{j,2i+1} = h(\mathbf{v}_{j-1,2i+1}, \mathbf{v}_{j-1,2i+2}, \theta_{j,i+1})$
    end for
  end if
  $\mathbf{v}_{j,2n-2j-2} = h(\mathbf{v}_{j-1,2n-2j}, \mathbf{v}_{j-1,2n-2j-1}, \theta_{j,n-j})$
end for

4.5 Model Selection

In the previous section it is described how to do inference for a specific pair-copula decomposition. However, this is only a part of the full estimation problem. Full inference for a pair-copula decomposition should in principle consider (a) the selection of a specific factorisation, (b) the choice of pair-copula types, and (c) the estimation of the copula parameters. For smaller dimensions (say 3 and 4), one may estimate the parameters of all possible decompositions using the procedure described in the previous section and compare the resulting log-likelihoods. This is in practice infeasible for higher dimensions, since the number of possible decompositions increases very rapidly with the dimension of the data set. One should instead determine which bivariate relationships are most important to model correctly, and let this determine which decomposition(s) to estimate. Given data and an assumed pair-copula decomposition, it is necessary to specify the parametric shape of each paircopula. For example, for the decomposition in the 3-dimensional case displayed above we need to decide which copula type to use for $C_{12}(\cdot,\cdot), C_{23}(\cdot,\cdot)$ and $C_{13|2}(\cdot,\cdot)$. The pair-copulae do not have to belong to the same family. The resulting multivariate distribution will be valid if one chooses for each pair of variables the parametric copula that best fits the data. If one chooses not to stay in one predefined class, one needs a way of determining which copula to use for each pair of (transformed) observations. The sequential estimation procedure for fitting a D-Vine is as follows:

(a) Fit each of the possible copula types to the data in tree 1 and estimate the copula parameters using the original data.

(b) Determine which copula type fits the data best using the Vuong Test (see Vuong [1989]) and the Clarke Test (see Clarke [2007]).

(c) Verify the test decisission by plotting the original data.

(d) Transform observations as required for tree 2, using the definition of the $h(\cdot)$ function.

(e) Determine which copula types to use in tree 2 in the same way as in tree 1.

(f) Iterate.

The observations used to select the copulae at a specific level depend on the specific pair-copulae chosen upstream in the decomposition.
Once each pair-copula of the D-Vine is specified the log-likelihood of the D-Vine given in (4.10) is maximised with respect to the copula parameters of each pair-copula. The parameters obtained from the pair-copula specification are used as starting values.

4.6 D-Vine Model for Loss Triangle Dependence

This section introduces a D-Vine based model for loss triangle dependence based on the Gaussian copula model introduced in chapter 2. In a first step assumed inner-triangle dependence will be eliminated to obtain *i.i.d.* data in each triangle. In a second step this modified loss triangle data will be modelled using a D-Vine decomposition.

4.6.1 Data Preparation

Basis for the whole model are the normalized standardized payments

$$Z_{lij} := \frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}} \sim N(0,1), \tag{4.12}$$

where φ_l , $l = 1 \dots, m$, is a triangle specific transformation, such that $\varphi_l(P_{lij}) \sim N(\mu_{lj}, \sigma_{lj}^2)$ holds for all $i = 1, \dots, n$ and $j = 0, \dots, n-1$, with μ_{lj} and σ_{lj}^2 denoting the mean and the variance of transformed payments $\varphi_l(P_{lij})$ in the j-th column of transformed triangle l. As a direct consequence of **Theorem 2.1** we have that

$$cov(Z_{li_1j_1}, Z_{li_2j_2}) = c_l^2 \quad \forall \; i_1 + j_1 = i_2 + j_2 \;, \tag{4.13}$$

where c_l^2 is the communality of triangle l defined in (2.2). With

$$\mathbf{Z}_l := \begin{pmatrix} Z_{l,1,0} \\ Z_{l,2,0} \\ Z_{l,1,1} \\ Z_{l,3,0} \\ Z_{l,2,1} \\ \vdots \\ Z_{l,n,0} \\ \vdots \\ Z_{l,2,n-2} \\ Z_{l,1,n-1} \end{pmatrix} \in \mathbb{R}^{\frac{n(n+1)}{2}} \tag{4.14}$$

being the vector containing all Z_{lij} in one triangle sorted by calendar year $t = i + j$ and development year $j = 0, \dots, n-1$ we have that

$$\mathbf{Z}_l \sim N(\mathbf{0}, \mathbf{C}(c_l)) \;, \tag{4.15}$$

with covariance matrix $\mathbf{C}(c_l)$ depending only on the communality c_l of triangle l. Note that $\mathbf{C}(c_l)$ has the form

$$\mathbf{C}(c_l) = \begin{pmatrix} 1 & & & & & & & & & & & 0 \\ & 1 & c_l^2 & & & & & & & & & \\ & c_l^2 & 1 & & & & & & & & & \\ & & & 1 & c_l^2 & c_l^2 & & & & & & \\ & & & c_l^2 & 1 & c_l^2 & & & & & & \\ & & & c_l^2 & c_l^2 & 1 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & 1 & c_l^2 & \dots & c_l^2 \\ & & & & & & & c_l^2 & 1 & \ddots & \vdots \\ & & & & & & & \vdots & \ddots & \ddots & c_l^2 \\ 0 & & & & & & & c_l^2 & \dots & c_l^2 & 1 \end{pmatrix} \in \mathbb{R}^{\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}} \;,$$

where each block has size $t \times t$, $t = i + j$, $t = 1, \ldots, n$ and entries outside these blocks are 0. Note that the block size increases by 1 from the top left corner of the matrix to the bottom right corner.
In order to obtain *i.i.d.* data this inner-triangle covariance structure needs to be eliminated. This is achieved by defining

$$\mathbf{W}_l := \mathbf{C}(c_l)^{-\frac{1}{2}} \mathbf{Z}_l \ \sim \ N(\mathbf{0}, \mathbf{I}) \ , \tag{4.16}$$

where $\mathbf{I}$ denotes the $\frac{n(n+1)}{2}$ - dimensional identity matrix.
Thus the entries of $\mathbf{W}_l$ are independent and hence triangle dependence will only occur between different triangles.

4.6.2 Model Formulation

The model presented in this section is based on the definitions of the vectors $\mathbf{W}_1, \ldots, \mathbf{W}_m$. For ease of simplicity the indexation of the entries W_{lij} of $\mathbf{W}_l$, $l = 1, \ldots, m$ will be slightly changed:

$$\mathbf{W}_l := \left(W_{l,1}, W_{l,2}, \ldots, W_{l, \frac{n(n+1)}{2} - 1}, W_{l, \frac{n(n+1)}{2}} \right) \ , \tag{4.17}$$

i.e. vector $\mathbf{W}_l$ consists of entries W_{lk} sorted first by calendar year and second by development year, $k = 1, \ldots, \frac{n(n+1)}{2}$, implying

- $k = 1$ corresponds to $i = 1$ and $j = 0$
- $k = 2$ corresponds to $i = 2$ and $j = 0$
- $k = 3$ corresponds to $i = 1$ and $j = 1$
- ...
- $k = \frac{n(n-1)}{2} + 1$ corresponds to $i = n$ and $j = 0$
- $k = \frac{n(n+1)}{2} - 1$ corresponds to $i = 2$ and $j = n - 2$

and

- $k = \frac{n(n+1)}{2}$ corresponds to $i = 1$ and $j = n - 1$.

From (4.16) we obtain that

$$W_{lk} \sim N(0,1) \ i.i.d. \ \forall \ l = 1, \ldots, m \text{ and } k = 1, \ldots, \frac{n(n+1)}{2}. \tag{4.18}$$

As a direct consequence of (4.18) we have that

$$U_{lk} := \Phi(W_{lk}) \sim \text{unif}(0,1) \ . \tag{4.19}$$

The common density of the vectors $\mathbf{U}_l := \left(U_{l,1}, \dots, U_{l,\frac{n(n+1)}{2}}\right)$, $l = 1, \dots, m$ will be modeled by a m-dimensional PCC-decomposition.
Putting these building blocks together yields the following model:

$$\begin{aligned} Z_{lij} &:= \frac{\varphi_l(P_{lij}) - \mu_{lj}}{\sigma_{lj}} \sim N(0,1) \\ \mathbf{Z}_l &:= (Z_{l,1,0}, Z_{l,2,0}, \dots, Z_{l,n,0}, \dots, Z_{l,2,n-2}, Z_{l,1,n-1})^T \sim N(\mathbf{0}, \mathbf{C}(c_l)) \\ \mathbf{W}_l &:= \mathbf{C}(c_l)^{-\frac{1}{2}} \mathbf{Z}_l \sim N(\mathbf{0}, \mathbf{I}) \\ \mathbf{W}_l &= \left(W_{l,1}, \dots, W_{l,\frac{n(n+1)}{2}}\right) \in \mathbb{R}^{\frac{n(n+1)}{2}} \\ U_{lk} &:= \Phi(W_{lk}) \sim \text{unif}(0,1). \end{aligned} \tag{4.20}$$

The joint density of $u_{1k}, \dots, u_{mk}$ is defined as

$$\begin{aligned} g(u_{1k}, \dots, u_{mk}) = \prod_{j=1}^{k-1} \prod_{i=1}^{k-j} c_{i,i+j|i+1,\dots,i+j-1}(&F(u_{i,k}|u_{i+1,k}, \dots, u_{i+j-1,k}), \\ &F(u_{i+j,k}|u_{i+1,k}, \dots, u_{i+j-1,k})). \end{aligned} \tag{4.21}$$

4.6.3 Tests for Non-Nested Model Selection

This section gives a brief overview over the two tests for model selection used in this work, namely the Vuong Test and the Clarke Test (see Vuong [1989] and Clarke [2007] for details). Both tests can be used to make a decision, whether a certain model is closer to the true, but unknown, specification than another model. As indicated by the caption of this section both tests can be used for non-nested models.

The Vuong Test

Consider two models, $M_{\boldsymbol{\xi}} = f(\mathbf{Y}_i|\boldsymbol{\xi})$ and $\tilde{M}_{\boldsymbol{\gamma}} = g(\mathbf{Y}_i|\boldsymbol{\gamma})$, where $\boldsymbol{\xi}$ and $\boldsymbol{\gamma}$ denote the set of model parameters and $\mathbf{Y}_i \in \mathbb{R}^d$, $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)^T$. The null hypothesis of the test is

$$H_0 : E_0 \left[\log \frac{f(\mathbf{Y}_i|\boldsymbol{\xi}_*)}{g(\mathbf{Y}_i|\boldsymbol{\gamma}_*)} \right] = 0, \tag{4.22}$$

where E_0 is the expectation under the true model and $\boldsymbol{\xi}_*, \boldsymbol{\gamma}_*$ are the pseudo-true values of $\boldsymbol{\xi}$ and $\boldsymbol{\gamma}$ (see White [1982]). H_0 indicates that both models are equally close to the true specification. Vuong [1989] proves under general conditions that the expected value given in the null hypothesis can be consequently estimated by $(1/n)$ time the likelihood ratio statistic LR_n,

$$\frac{1}{n} LR_n(\hat{\boldsymbol{\xi}}_n, \hat{\boldsymbol{\gamma}}_n) \overset{a.s.}{\rightarrow} E_0 \left[\log \frac{f(\mathbf{Y}_i|\boldsymbol{\xi}_*)}{g(\mathbf{Y}_i|\boldsymbol{\gamma}_*)} \right], \tag{4.23}$$

where $\hat{\boldsymbol{\xi}}_n$ and $\hat{\boldsymbol{\gamma}}_n$ are the maximum likelihood estimators of $\boldsymbol{\xi}_*$ and $\boldsymbol{\gamma}_*$. The resulting likelihood ratio statistic is asimptotically normally distributed, and the actual test is therefore

$$\text{under } H_0 : \frac{LR_n(\hat{\boldsymbol{\xi}}_n, \hat{\boldsymbol{\gamma}}_n)}{\sqrt{n}\hat{\omega}_n} \xrightarrow{\text{d}} N(0,1), \tag{4.24}$$

where the numerator is the difference in the summed log-likelihoods for the two models, $LR_n(\hat{\boldsymbol{\xi}}_n, \hat{\boldsymbol{\gamma}}_n) \equiv L_n^f(\hat{\boldsymbol{\xi}}_n) - L_n^g(\hat{\boldsymbol{\gamma}}_n)$, and $\hat{\omega}_n$ is the estimated standard deviation calculated as,

$$\hat{\omega}_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f(\mathbf{Y}_i|\hat{\boldsymbol{\xi}}_n)}{g(\mathbf{Y}_i|\hat{\boldsymbol{\gamma}}_n)}\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\frac{f(\mathbf{Y}_i|\hat{\boldsymbol{\xi}}_n)}{g(\mathbf{Y}_i|\hat{\boldsymbol{\gamma}}_n)}\right]^2. \tag{4.25}$$

The Vuong statistic is sensitive to the number of estimated coefficients in eachmodel, and therefore the test must be corrected for the model dimensionality. Vuong [1989] suggests to adjust the test statistic as

$$L\tilde{R}_n(\hat{\boldsymbol{\xi}}_n, \hat{\boldsymbol{\gamma}}_n) \equiv LR_n(\hat{\boldsymbol{\xi}}_n, \hat{\boldsymbol{\gamma}}_n) - (\frac{p}{2}\log n - \frac{q}{2}\log n), \tag{4.26}$$

where p and q are the number of estimated parameters in $M_{\boldsymbol{\xi}}$ and $\tilde{M}_{\boldsymbol{\gamma}}$.

The Clarke Test

Clarke [2007] proposes a distribution-free alternative which applies a modified paired sign test to the differences in the individual log-likelihoods from two nonnested models. The null hypothesis of this test is

$$H_0 : P_0\left(\log\frac{f(\mathbf{Y}_i|\boldsymbol{\xi}_*)}{g(\mathbf{Y}_i|\boldsymbol{\gamma}_*)} > 0\right) = 0.5. \tag{4.27}$$

Equation (4.27) states that under the null hypothesis, the log-likelihood ratios should be evenly distributed around zero. Thus, half the log-likelihood ratios should be greater than zero and half less than zero. The difference between equation (4.27) and equation (4.23) is that the expectation in equation (4.23) is replaced with the median in equation (4.27). Letting $d_i = \log f(\mathbf{Y}_i|\hat{\boldsymbol{\xi}}_n) - \log g(\mathbf{Y}_i|\hat{\boldsymbol{\gamma}}_n)$, the test statistic is

$$B = \sum_{i=1}^{n}\mathbb{1}_{(0,\infty)}(d_i), \tag{4.28}$$

where $\mathbb{1}$ is the indicator function. Equation (4.28) is the number of positive differences, and it is distributed Binomial with parameters n and $p = 0.5$. If model $M_{\boldsymbol{\xi}}$ is "better" than model $\tilde{M}_{\boldsymbol{\gamma}}$, B will be significantly larger than its expected value under the null hypothesis ($n/2$). For an upper tail test, we reject the null hypothesis of equivalence when $B \geq c_\beta$, where c_β is chosen to be the smallest integer such that

$$\sum_{c=c_\beta}^{n}\binom{n}{c}0.5^n \leq \beta.$$

For a lower tail test, the inequality is reversed, and the sum goes from $c = 0$ to $c = c_\beta$. Correction for model dimension is achieved by applying the average correction to the individual log-likelihood ratios. That is, one corrects the the individual log-likelihoods for model $M_{\boldsymbol{\xi}}$ by a factor of $(p/2n)\log n$ and the individual log-likelihoods for model $\tilde{M}_{\boldsymbol{\gamma}}$ by a factor of $(q/2n)\log n$, where $dim(\boldsymbol{\xi}) = p$ and $dim(\boldsymbol{\gamma}) = q$.

4.7 Application to real Insurance Data

The data consists of the $m = 5$ loss triangles for different lines of business mentioned in chapter 3. After transformation to $N(0,1)$ resulting in $\mathbf{Z}_l$ defined in (4.14) the inner-triangle correlation needs to be captured in order to transform the data to an *i.i.d.* sample in each triangle $l, l = 1, \ldots 5$. This is achieved by forming $\mathbf{W}_l := \mathbf{C}(c_l)^{-\frac{1}{2}}\mathbf{Z}_l$ displayed in (4.16).

triangle l	1	2	3	4	5
c_l	0.352	0.578	0.596	0.494	0.559

Table 4.1: estimated communalities of triangles $1, \ldots, 5$ obtained by the stepwise ML approach used for calculating $\mathbf{W}_l$

Table 4.1 shows the estimated communalities $\hat{c}_l^{sML}$ for the $m = 5$ triangles obtained by the stepwise maximum-likelihood estimation displayed in Table 3.6. These values are used to calculate the vectors $\mathbf{U}_l$, $l = 1, \ldots, 5$ described in section 4.6.2.

4.7.1 Determination of Top-Tree Structure

The D-Vine structure in tree 1 is determined by calculating Spearman's ρ_S and Kendall's τ (see Nelsen [1999]) for each pair $(\mathbf{U}_l, \mathbf{U}_{l'}), l \neq l'$ of data vectors, where $\mathbf{U}_l := \Phi(\mathbf{W}_l)$ and $\Phi(\cdot)$ acts componentwise.

ρ_S	$\mathbf{U}_1$	$\mathbf{U}_2$	$\mathbf{U}_3$	$\mathbf{U}_4$	$\mathbf{U}_5$
$\mathbf{U}_1$	1.000	0.049	**0.213**	0.103	**0.237**
$\mathbf{U}_2$		1.000	0.055	**0.138**	0.117
$\mathbf{U}_3$			1.000	**0.218**	0.067
$\mathbf{U}_4$				1.000	-0.024
$\mathbf{U}_5$					1.000

Table 4.2: Estimates of Spearman's ρ_S for pairs $U_l, U_{l'}, l \neq l'$ of data vectors; top 4 correlations bolded

τ	$\mathbf{U}_1$	$\mathbf{U}_2$	$\mathbf{U}_3$	$\mathbf{U}_4$	$\mathbf{U}_5$
$\mathbf{U}_1$	1.000	0.025	**0.143**	0.089	**0.162**
$\mathbf{U}_2$		1.000	0.038	**0.098**	0.081
$\mathbf{U}_3$			1.000	**0.167**	0.043
$\mathbf{U}_4$				1.000	-0.020
$\mathbf{U}_5$					1.000

Table 4.3: Estimates of Kendall's τ for pairs $U_l, U_{l'}, l \neq l'$ of data vectors; top 4 correlations bolded

Table 4.2 displays the estimates of ρ_S and Table 4.3 shows the estimates of τ for all pairs of variables. The strongest correlations between loss triangles indicated by these

measures of dependence are used to determine the Vine structure in tree 1. High dependence is measured for the pairs $(\mathbf{U}_1, \mathbf{U}_3)$,$(\mathbf{U}_3, \mathbf{U}_4)$,$(\mathbf{U}_1, \mathbf{U}_5)$ and $(\mathbf{U}_2, \mathbf{U}_4)$. Typically the pairs with high dependence are used to determine the structure of the top-tree.
This leads to the pair-copula decomposition $\mathbf{5} - \mathbf{1} - \mathbf{3} - \mathbf{4} - \mathbf{2}$ in the top-tree of the Vine. This Vine has the structure of a D-Vine.

4.7.2 Pair-Copula Estimation

In the top tree of the D-Vine four pair-copulas need to be estimated. For each pair $(\mathbf{U}_5, \mathbf{U}_1)$,$(\mathbf{U}_1, \mathbf{U}_3)$,$(\mathbf{U}_3, \mathbf{U}_4)$ and $(\mathbf{U}_4, \mathbf{U}_2)$ a t-, Gumbel-, Clayton and Gaussian copula is fitted.

	$(\mathbf{U}_5, \mathbf{U}_1)$	$(\mathbf{U}_1, \mathbf{U}_3)$	$(\mathbf{U}_3, \mathbf{U}_4)$	$(\mathbf{U}_4, \mathbf{U}_2)$
ρ_t	0.264	0.224	0.283	0.144
df	12.99	6.101	1.546	2.771
ρ_N	0.271	0.193	0.205	0.109
λ_G	1.206	1.182	1.242	1.143
λ_C	0.266	0.136	0.250	0.084

Table 4.4: Estimated copula parameters for t-, Gumbel-, Clayton and Gaussian copula for each pair-copula in the top tree of the D-Vine

Table 4.4 shows the parameter estimates for each of the fitted copulas for each pair $(\mathbf{U}_l, \mathbf{U}_k)$ in the top-tree of the D-Vine. Here ρ_t and df denote the parameter set estimated for the t-copula, ρ_N the parameter for the Gaussian copula, λ_G the parameter for the Gumbel copula and λ_C the parameter for the Clayton copula.
Fitting four different copula-types to a certain pair $(\mathbf{U}_l, \mathbf{U}_k)$ of data yields four marginal models for the dependence structure between $\mathbf{U}_l$ and $\mathbf{U}_k$. These marginal models are compared using the Vuong Test and the Clarke Test.
The log-likelihoods needed to compute the test statistics for both tests are obtained as follows:

- For a certain pair $(\mathbf{u}_l, \mathbf{u}_{l'})$ of observations define

$$\mathbf{u}_{ll'} := \begin{pmatrix} u_{l,1} & u_{l',1} \\ \vdots & \vdots \\ u_{l,\frac{n(n+1)}{2}} & u_{l',\frac{n(n+1)}{2}} \end{pmatrix} := \begin{pmatrix} \tilde{\mathbf{u}}_{ll',1} \\ \vdots \\ \tilde{\mathbf{u}}_{ll',\frac{n(n+1)}{2}} \end{pmatrix},$$

  where $\mathbf{u}_{ll'} \in \mathbb{R}^{\frac{n(n+1)}{2} \times 2}$ and $\tilde{\mathbf{u}}_{ll',k} \in \mathbb{R}^2$, $k = 1, \ldots, \frac{n(n+1)}{2}$.

- Next one computes a vector of individual log-likelihoods of each fitted pair copula (PC) for each pair of observations, namely

$$\mathbf{V}_{PC} := \begin{pmatrix} \log c_{ll',\theta}(\tilde{\mathbf{u}}_{ll',1}, \theta) \\ \vdots \\ \log c_{ll',\theta}(\tilde{\mathbf{u}}_{ll',\frac{n(n+1)}{2}}, \theta) \end{pmatrix} \in \mathbb{R}^{\frac{n(n+1)}{2} \times 2},$$

where θ is the set of copula parameters of the respective copula type. The elements of $\mathbf{V}_{PC}$ are now used to compute the test statistics for the Vuong Test and the Clarke Test.

The test decisions are used to determine which of the four copula types mentioned fits the dependence structure of a certain pair of observations in the top-tree best.

Pair	$\mathbf{U}_5 - \mathbf{U}_1$			$\mathbf{U}_1 - \mathbf{U}_3$			$\mathbf{U}_3 - \mathbf{U}_4$			$\mathbf{U}_4 - \mathbf{U}_2$		
Copula	N	G	C	N	G	C	N	G	C	N	G	C
t	N/N	G/G	-/C	N/-	G/-	-/-	t/t	-/t	t/t	-/t	-/t	-/t
N		-/G	-/N		-/G	-/N		G/G	-/-		-/G	-/-
G			-/G			-/G			-/G			-/G

Table 4.5: Test decisions based on Vuong and Clarke test for each possible pair copula in the top tree of the D-Vine; t=t-copula, N=Gauss, G=Gumbel, C=Clayton

Copula-type choice is based on a simple scoring system relying on the test decisions of the Vuong- and the Clarke test displayed in Table 4.5:
If either test prefers a certain copula-type this copula receives two points, the copula refused by either test zero points. If the test doesn't prefer either of the two copula-types both receive one point. Thus when comparing four different copula-types for each pair of loss-triangle the copula with the highest score is chosen.
For the data pairs in the top tree this leads to the scores and resulting decisions displayed in Table 4.6.

data pair	$(\mathbf{U}_5, \mathbf{U}_1)$	$(\mathbf{U}_1, \mathbf{U}_3)$	$(\mathbf{U}_3, \mathbf{U}_4)$	$(\mathbf{U}_4, \mathbf{U}_2)$
t-copula	1	4	**11**	**9**
Gauss-copula	8	7	8	7
Gumbel-copula	**10**	**9**	2	4
Clayton-copula	5	4	3	4
resulting choice	Gumbel	Gumbel	t-copula	t-copula

Table 4.6: Scores and resulting Copula decisions for the pair copulas in the top-tree of the D-Vine; highest scores bolded

The set of edges of the second tree of the D-Vine is given by $\mathbf{53|1-14|3-32|4}$. For the third tree we have $\mathbf{54|13-12|34}$ and $\mathbf{52|134}$ for the fourth tree. The procedure for determining a certain pair copula for each data pair is the same as for the top-tree with the only difference that no longer the original data is used for estimating the copula parameters. Instead one makes use of the definition of the $h(\cdot)$ function in order to compute observations for trees 2,3 and 4 as displayed in (4.9).
The test decisions used for determining a certain copula type for the second tree are displayed in Table 4.7 while those for the third and bottom tree are displayed in Table 4.8.

Pair	$\mathbf{U}_{5\|1} - \mathbf{U}_{3\|1}$			$\mathbf{U}_{1\|3} - \mathbf{U}_{4\|3}$			$\mathbf{U}_{3\|4} - \mathbf{U}_{2\|4}$		
Copula	N	G	C	N	G	C	N	G	C
t	-/-	-/-	-/-	-/-	-/-	-/-	-/t	-/t	-/-
N		-/G	-/-		-/G	-/-		-/G	-/-
G			-/G			-/-			-/-

Table 4.7: Test decisions based on Vuong and Clarke test in the second tree of the D-Vine; t=t-copula, N=Gauss, G=Gumbel, C=Clayton

Pair	$\mathbf{U}_{53\|1} - \mathbf{U}_{14\|3}$			$\mathbf{U}_{14\|3} - \mathbf{U}_{32\|4}$			$\mathbf{U}_{54\|13} - \mathbf{U}_{12\|34}$		
Copula	N	G	C	N	G	C	N	G	C
t	N/N	-/-	C/C	N/N	G/G	-/-	N/N	G/G	C/C
N		N/N	-/-		-/G	-/-		-/G	-/-
G			-/C			-/G			-/-

Table 4.8: Test decisions based on Vuong and Clarke test for each pair copula in the third and bottom tree of the D-Vine; t=t-copula, N=Gauss, G=Gumbel, C=Clayton

Applying the scoring procedure described above to the second, third and bottom tree of the D-Vine yields the copula choices displayed in Table 4.9.

data pair	copula choice	score
$\mathbf{U}_{5\|1} - \mathbf{U}_{3\|1}$	Gumbel	8
$\mathbf{U}_{1\|3} - \mathbf{U}_{4\|3}$	Gumbel	7
$\mathbf{U}_{3\|4} - \mathbf{U}_{2\|4}$	t-copula	8
$\mathbf{U}_{53\|1} - \mathbf{U}_{14\|3}$	Gauss-copula	10
$\mathbf{U}_{14\|3} - \mathbf{U}_{32\|4}$	Gumbel	10
$\mathbf{U}_{54\|13} - \mathbf{U}_{12\|34}$	Gumbel	9

Table 4.9: Copula choice for each pair copula in the second, third and bottom tree of the D-Vine with resulting score of chosen copula type

The parameters of the D-vine are estimated using Algorithm 9. For each pair-copula, the loglikelihood is computed using (4.11).

Parameter	Final Est.	Sequential Est.	Copula Type
λ_{51}	1.203	1.206	Gumbel
λ_{13}	1.174	1.181	Gumbel
ρ_{34}	0.316	0.283	t
ν_{34}	1.772	1.559	
ρ_{42}	0.184	0.143	t
ν_{42}	2.801	2.772	
$\lambda_{53\|1}$	1.049	1.051	Gumbel
$\lambda_{14\|3}$	1.049	1.047	Gumbel
$\rho_{32\|4}$	0.017	0.002	t
$\nu_{32\|4}$	2.933	3.323	
$\rho_{54\|13}$	-0.022	-0.019	Gaussian
$\lambda_{12\|34}$	1.095	1.092	Gumbel
$\lambda_{52\|134}$	1.077	1.073	Gumbel
log-likelih.	38.15	37.80	

Table 4.10: Estimated parameters for five-dimensional pair-copula decomposition compared to sequential estimates

Table 4.10 shows the starting values obtained using the sequential estimation procedure (right column) presented above, and the final parameter values together with the corresponding loglikelihood values. The final parameter values are obtained using Algorithm 9. As can be seen from the table, the likelihood slightly increases when estimating all parameters simultaneously.

In order to verify the choice based on Vuong's and Clarke's Test each data pair is plotted. Table 4.11 displays these plots. The plots support the choice of the certain pair-copulas, especially in the top tree of the D-Vine. In the two top right plots one can see a clustering around a fictious line from the bottom left to the top right of each plot, which is typical for t-copulas and in the two top left plots one can observe a clustering in the top right of the plots which supports the choice of a Gumbel-copula for these two data pairs (for details concearning typical copula shapes see Nelsen [1999]). In the lower trees of the D-Vine graphical evidence isn't as strong as for the top tree but due to a rather low data basis of 136 data points per loss triangle this isn't too surprising. The respective contour plots are shown in Table 4.12. These also support the copula choice in the top tree but don't give evidence for the copula choices in the lower trees which arises from the low data basis and weak correlations as displayed in Table 3.6.

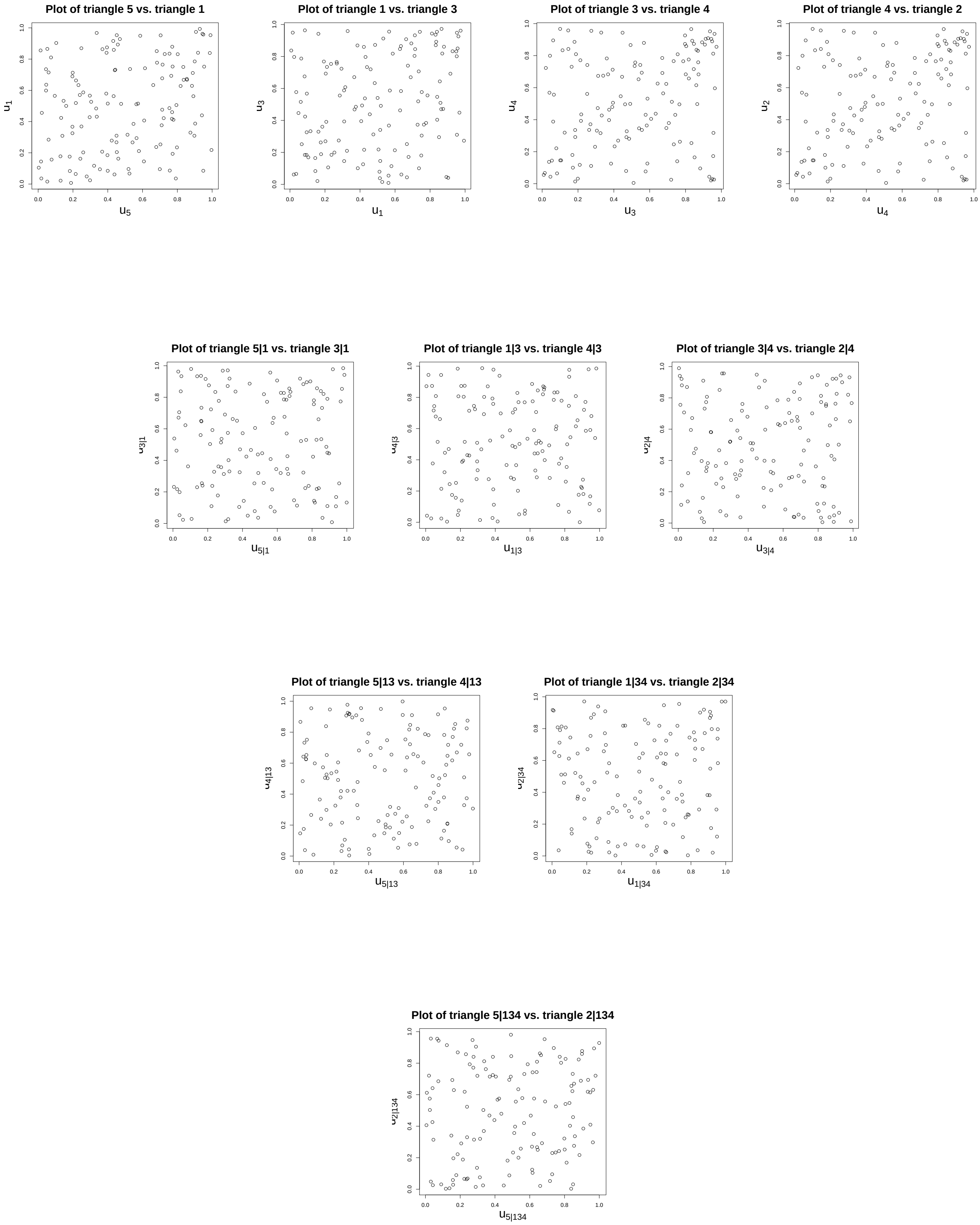

Table 4.11: Plot of the data vectors used to estimate the ten pair-copulae in the D-Vine construction

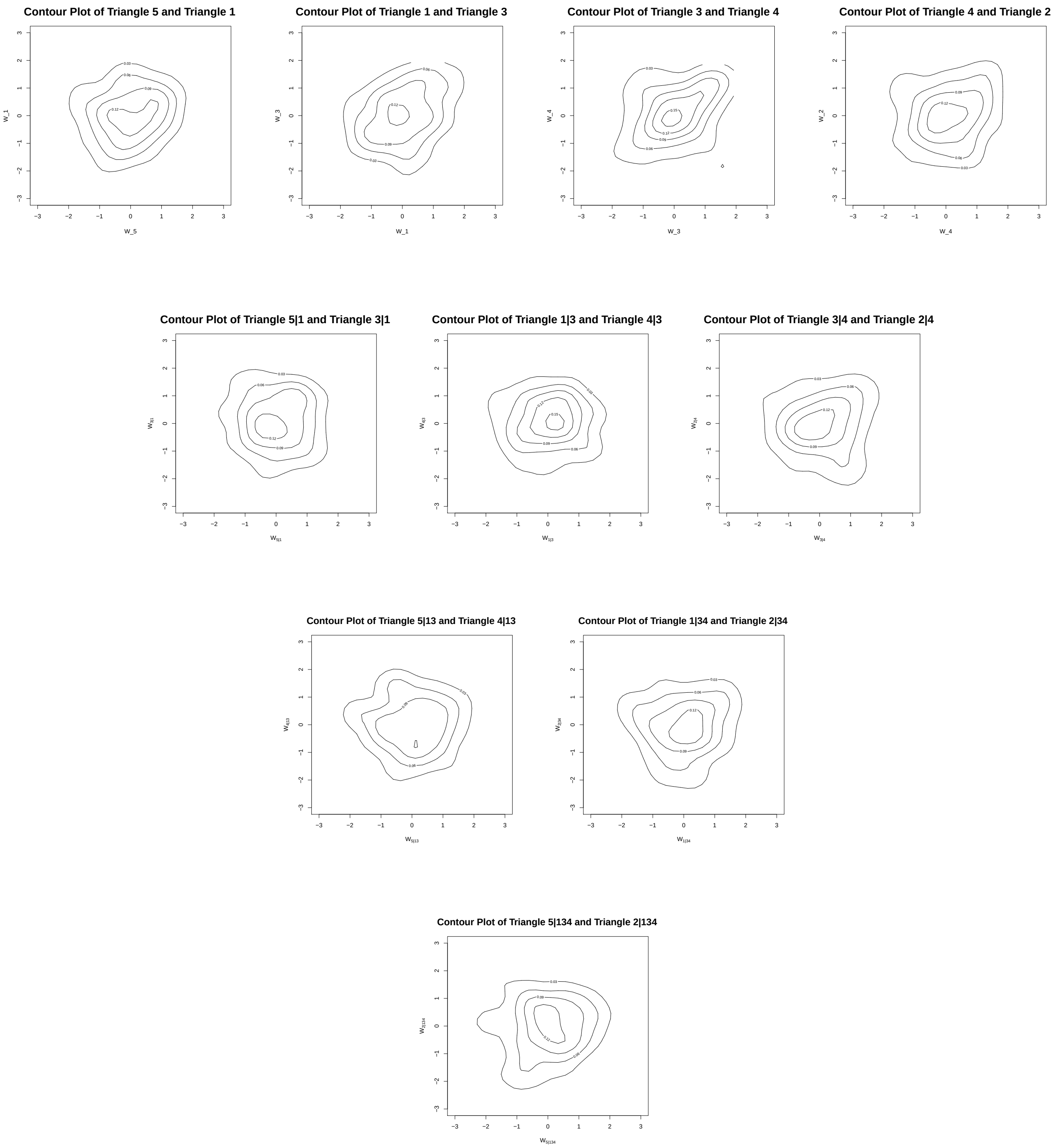

Table 4.12: Contour plots of the data vectors used to estimate the ten pair-copulae in the D-Vine construction

4.8 Data Forecast and Fit

The procedure of how to simulate loss triangles from a specified D-Vine model and how to forecast future payments will be described in this section. The general procedures of data

fit and data forecast are quite similar, but for ease of completeness each will be presented in a single algorithm. These algorithms are based on a certain D-Vine with m edges and $m-1$ nodes in the top tree and known parameters.
The simulation of m loss triangles of length n is described in Algorithm 10 and data forecast in Algorithm 11.

Algorithm 10 Data Fit D-Vine Model

draw $\mathbf{U} \in \mathbb{R}^{n(n+1)/2 \times m}$ from the D-Vine
for $l = 1$ to m **do**
  $\mathbf{U}_l = \mathbf{U}[:, l]$
end for
for $l = 1$ to m **do**
  $\mathbf{W}_l = \Phi^{-1}(\mathbf{U}_l)$,
  where Φ operates componentwise and denotes the distribution function of the $N(0,1)$ distribution
end for
for $l = 1$ to m **do**
  $\mathbf{Z}_l = \mathbf{C}(c_l)^{\frac{1}{2}}_{Fit} * \mathbf{W}_l$
  with $*$ denoting matrix- vector- multiplication
end for

Algorithm 11 Data Forecast D-Vine Model

draw $\mathbf{U} \in \mathbb{R}^{n(n-1)/2 \times m}$ from the D-Vine
for $l = 1$ to m **do**
  $\mathbf{U}_l = \mathbf{U}[:, l]$
end for
for $l = 1$ to m **do**
  $\mathbf{W}_l = \Phi^{-1}(\mathbf{U}_l)$,
  where Φ operates componentwise and denotes the distribution function of the $N(0,1)$ distribution
end for
for $l = 1$ to m **do**
  $\mathbf{Z}_l = \mathbf{C}(c_l)^{\frac{1}{2}}_{Forecast} * \mathbf{W}_l$
  with $*$ denoting matrix- vector- multiplication
end for

Note that c_l denotes the communality of triangle l and $\mathbf{C}(c_l)$ is the inner-triangle correlation matrix of triangle l defined in (2.16) in section 2.6.3. It is important to stress that for data fit $\mathbf{C}(c_l)$ has dimension $n(n+1)/2 \times n(n+1)/2$ while for forecasting future payments the dimension reduces to $n(n-1)/2 \times n(n-1)/2$ which follows directly from the structure of loss triangles. The general structure of $\mathbf{C}(c_l)$ is displayed in (4.29) for data fit and in (4.30) for data forecast.

$$\mathbf{C}(c_l)_{Fit} = \begin{pmatrix} 1 & & & & & & & & & & 0 \\ & 1 & c_l & & & & & & & & \\ & c_l & 1 & & & & & & & & \\ & & & 1 & c_l & c_l & & & & & \\ & & & c_l & 1 & c_l & & & & & \\ & & & c_l & c_l & 1 & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & & c_l & \dots & c_l \\ & & & & & & & c_l & \ddots & \ddots & \vdots \\ & & & & & & & \vdots & \ddots & \ddots & c_l \\ 0 & & & & & & & c_l & \dots & c_l & 1 \end{pmatrix} \in \mathbb{R}^{n(n+1)/2 \times n(n+1)/2} \tag{4.29}$$

$$\mathbf{C}(c_l)_{Forecast} = \begin{pmatrix} 1 & c_l & \dots & c_l & & & & & & & 0 \\ c_l & \ddots & \ddots & \vdots & & & & & & & \\ \vdots & \ddots & \ddots & \vdots & & & & & & & \\ c_l & \dots & c_l & 1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & 1 & c_l & c_l & & & \\ & & & & & c_l & 1 & c_l & & & \\ & & & & & c_l & c_l & 1 & & & \\ & & & & & & & & 1 & c_l & \\ & & & & & & & 1 & c_l & 1 & \\ 0 & & & & & & & & & & 1 \end{pmatrix} \in \mathbb{R}^{n(n-1)/2 \times n(n-1)/2} \tag{4.30}$$

Note that both matrices have block diagonal form with 0 outside these blocks. In (4.29) the block in the top left corner of the matrix has size 1×1, the one in the bottom right corner has size $n \times n$ and the block size for all other blocks increases by 1. In (4.30) it is almost vice versa with the only difference that the largest block in the top left corner has dimension $(n-1) \times (n-1)$ and the block size decreases by 1 from the top left corner to the bottom right corner of the matrix.

Once the vectors $\mathbf{Z}_l \in \mathbb{R}^{n(n+1)/2}$ which are containing the z-scores z_{lij} are obtained, payments p_{lij} can easily be obtained by inverting

$$z_{lij} = \frac{\varphi_l(p_{lij}) - \mu_{lj}}{\sigma_{lj}} \ ,$$

where μ_{lj} and σ_{lj} are calculated from the original data prescribed in section 2.6.

In order to minimize random effects when drawing from the D-Vine a reasonable data fit or forecast should rely on multiple application of either Algorithm 10 or Algorithm 11.

Chapter 5

Model comparison

In this chapter the two different models presented in this work will be compared. This is achieved by first comparing them in terms of data fit and later in terms of data forecast. In order to give a brief overview about the main differences between the models these will once again be outlined here shortly.
The Gaussian copula model assumes dependence between payments p_{lij} only if these fall into the same calendar $t = i + j$. The different approaches in this model, the heuristic, maximum likelihood and stepwise maximum likelihood approach all provide the possibility to estimate the model parameters, namely the specifities s_l and the correlation matrix $\mathbf{R}$. Main difference between them is the fact that in contrast to the maximum likelihood approaches one may estimate the calendar year effects α_{lt} directly from the data in the heuristic approach.
The D-Vine model assumes the same inner-triangle dependencies as the Gaussian copula model but makes different assumptions about between-triangle dependence. In this model it is assumed that payments between different loss triangles are correlated no matter in which calendar year t they fall. These dependencies are modelled using a D-Vine construction.

5.1 Data Fit

As a first step the different models introduced in this work will be compared in terms of how well each model fits the original data. Data fit is based on multiple application of the presented algorithms for data generation. These are in detail

- Algorithm 5 for the heuristic approach in the Gaussian copula model
- Algorithm 4 for the ML and stepwise-ML approach in the Gaussian copula model and
- Algorithm 10 for the D-Vine model.

For each triangle the original data is plotted against the fitted data obtained by applying these algorithms. Figures 5.1 to 5.5 display these plots. One thing all plots have in common is the fact that in terms of reproducing one can easily observe that each of the different

models fit the data well in a sense, that the level of payments in different development years is reproduced very good. A closer look yields that the best fit is produced by the heuristic approach. The reason for this is that in contrast to the 3 other approaches, namely the ML-, stepwise ML- and DVine- approach, this methods estimates the values of α_{lt} , $t = 1, \ldots, n$, directly from the data whereas the other methods simulate these values from a certain distribution. Thus the heuristic approach not only matches the level of payments but also reproduces outliers quite well since unusually high payments p_{lij} in one calendar year $t = i + j$ lead to a high value of α_{lt} which itself leads to high fitted values and vice versa.
In order to support these statements an analytical analysis of data fit is performed. This is done by defining the variables

$$B_l^M := \sum_{i+j \leq n} |p_{lij} - \hat{p}_{lij}^M|, \tag{5.1}$$

which denotes the sum of the absolute differences between the original data p_{lij} and the fitted data $\hat{p}_{lij}^M$ where M indicates the model being used to fit the data and

$$R_l^M := \sum_{i+j \leq n} (p_{lij} - \hat{p}_{lij}^M)^2, \tag{5.2}$$

which is the sum of the squared differences between original and fitted data. The lower the values of B_l^M and R_l^M the better the method fits the original data for loss triangle l. In order to check which method fits the data best for all triangles one compares

$$B^M := \sum_{l=1}^{m} B_l^M \tag{5.3}$$

and

$$R^M := \sum_{l=1}^{m} R_l^M. \tag{5.4}$$

The values of B_l^M are displayed in Table 5.1 and those for R_l^M in Table 5.2. As indicated by the data plots the heuristic approach fits the original data better than the other methods. The maximum likelihood and stepwise maximum likelihood approaches perform worst and both yield values of B_l^M and R_l^M at the same level. The D-Vine model fits the data better than the maximum likelihood approaches but not as good as the heuristic approach of the Gaussian copula model does.
Thus when comparing the models in terms of data fit it is recommendable to apply the heuristic approach when trying to reproduce loss triangle data.

Model	B_1^M	B_2^M	B_3^M	B_4^M	B_5^M	B^M
Heuristic	**6602**	**1590**	**2500**	**3241**	**7188**	**21121**
Maximum Likelihood	7704	2507	3514	4387	10216	28328
stepwise ML	7912	2510	3513	4383	10216	28534
D-Vine	7691	2517	3553	4857	9548	28166

Table 5.1: Values of B_l^M for the heuristic, maximum likelihood, stepwise maximum likelihood and D-Vine model for each triangle and the sum of all triangles; best values bolded

Model	R_1^M	R_2^M	R_3^M	R_4^M	R_5^M	R^M
Heuristic	**725**	**63**	**185**	**254**	**1945**	**3172**
Maximum Likelihood	1012	193	379	437	4311	6332
stepwise ML	995	193	379	437	4312	6316
D-Vine	1022	190	384	569	4111	6276

Table 5.2: Values of R_l^M for the heuristic, maximum likelihood, stepwise maximum likelihood and D-Vine model for each triangle and the sum of all triangles; best values bolded; values scaled by 10^3

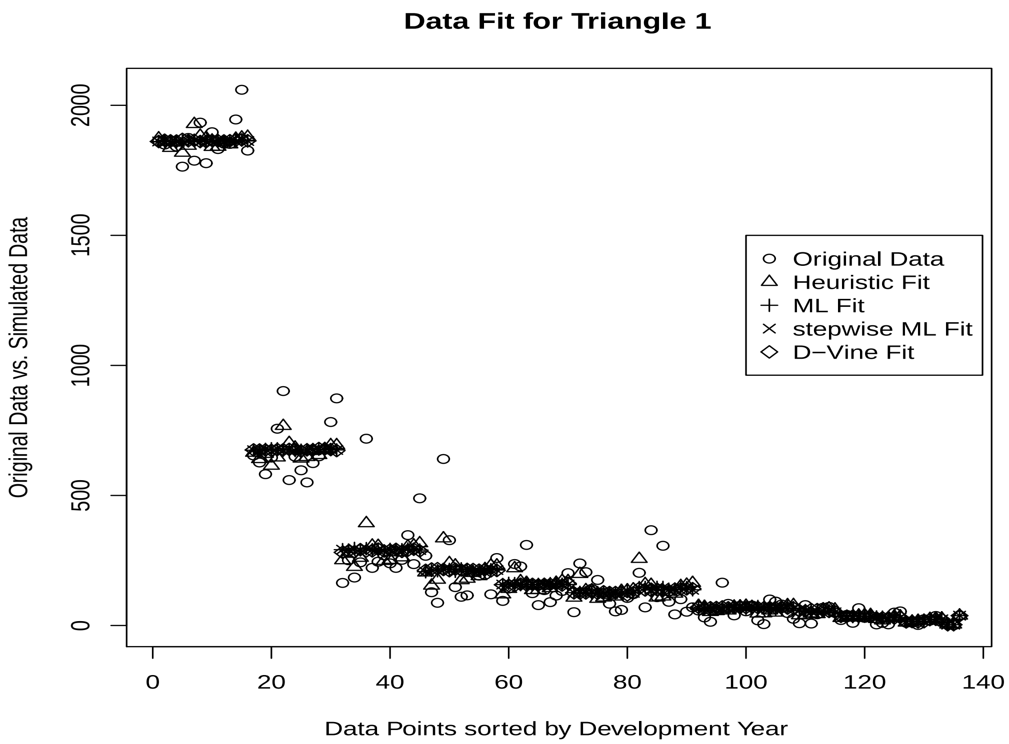

Figure 5.1: Comparison of the different models in terms of data fit for triangle 1

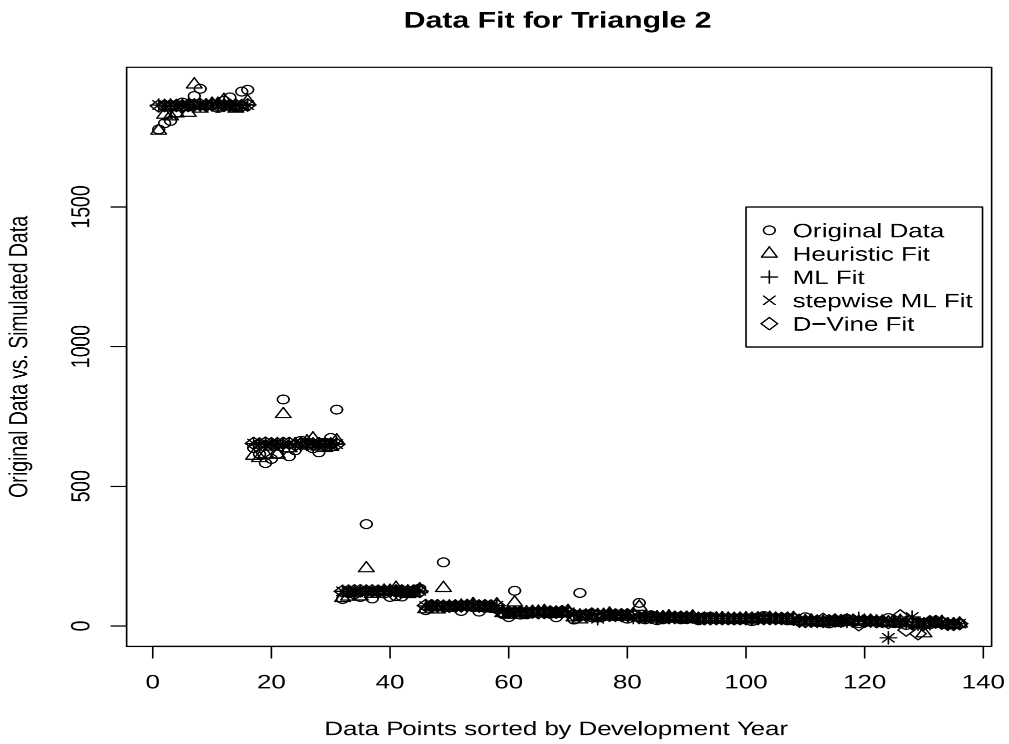

Figure 5.2: Comparison of the different models in terms of data fit for triangle 2

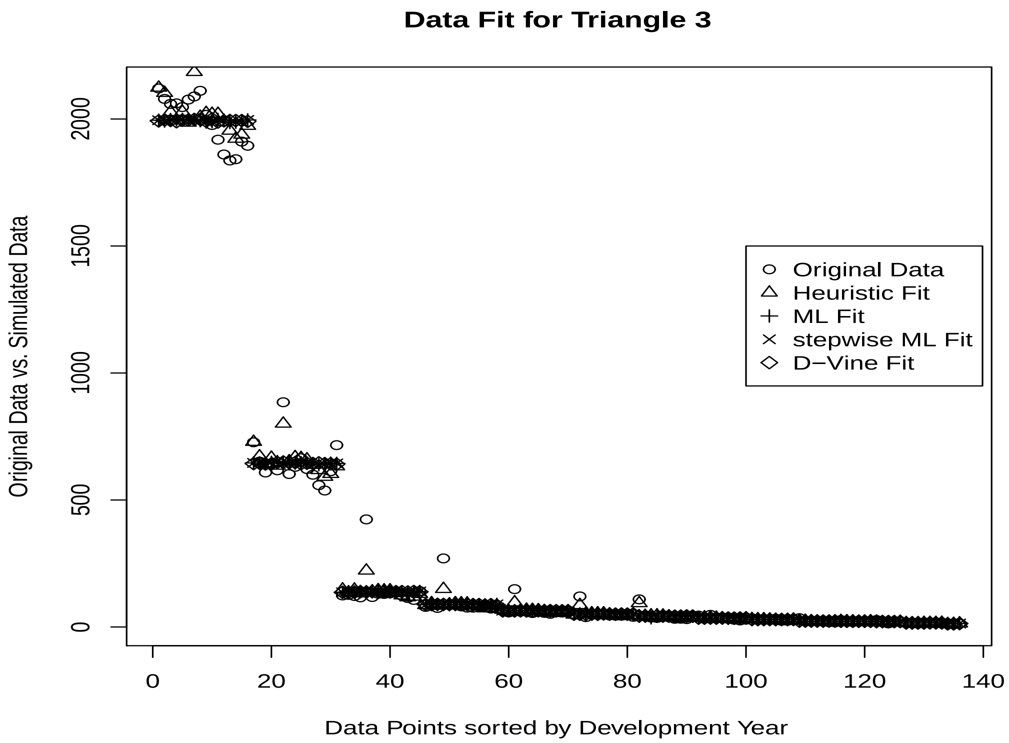

Figure 5.3: Comparison of the different models in terms of data fit for triangle 3

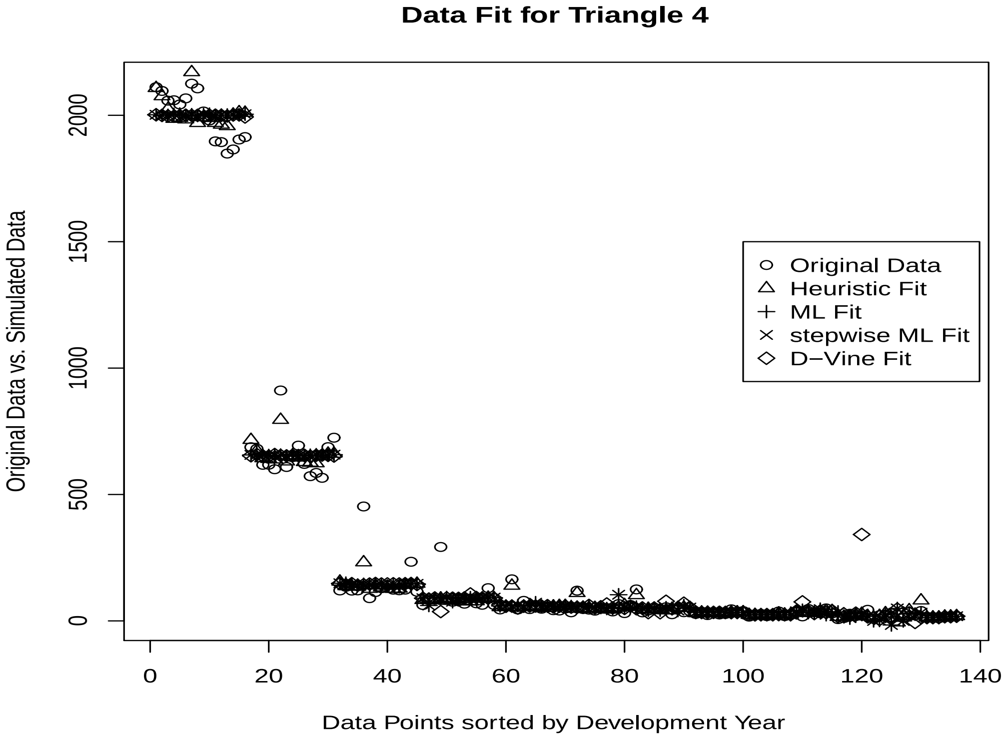

Figure 5.4: Comparison of the different models in terms of data fit for triangle 4

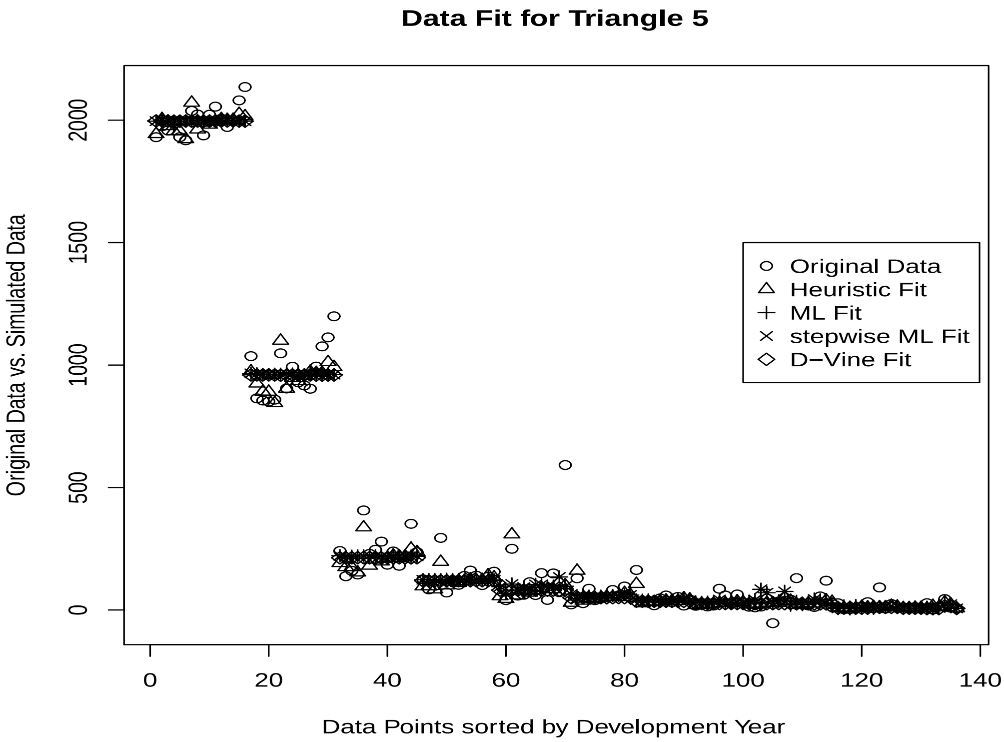

Figure 5.5: Comparison of the different models in terms of data fit for triangle 5

5.2 Forecast of Future Payments

Since the main difference between the described models and standard techniques such as the Chain-Ladder method is the fact that the Chain-Ladder method does not take into account possible dependencies in the runoff-structure of loss triangles. Hence there is no diversification effect when applying this method.
Thus in this section the Gaussian copula model, the D-Vine approach and the Chain-Ladder method will be compared based on simulated data in terms of diversification. This is achieved by first comparing the different methods in terms of forecasting real insurance data and secondly in terms of how well forecasts fit simulated data.

5.2.1 Forecast based on Real Data

In this section comparison is based on forecasting the $m = 5$ loss triangles containing real insurance data provided by a german reinsurance company.
As a first step a large number, i.e. $N = 10000$, of predictions of future payments $\hat{p}^r_{lij}$, $i + j > n$, $r = 1, \dots N$ is made. Note that the Chain-Ladder method only needs to be applied once, since it is deterministic and henceforth predictions won't change when applying this method several times. The predicted payments are then used to calculate the estimated total future obligations $\hat{TO}^r$ for all triangles.

$$\hat{TO}^r = \hat{TO}^r_1 + \cdots + \hat{TO}^r_m \ , \ r = 1, \dots, N \tag{5.5}$$

$$\hat{TO}^r_l = \sum_{i,j:i+j>n} \hat{p}^r_{lij} \ , \ l = 1, \dots, m \ , \ r = 1, \dots, N. \tag{5.6}$$

Note that $\hat{TO}^r_l$ denotes the estimated total future obligation of triangle l in simulation period r, $r = 1, \dots N$, which is the sum of all predicted payments $\hat{p}^r_{lij}$, $i + j > n$ and $\hat{TO}^r$ denotes the total future obligations for all triangles.
Calculating $\hat{TO}^r$, i.e. $N = 10000$ times, and taking the mean of these estimates leads to a Monte-Carlo estimate $\tilde{TO}$ of TO. This procedure finally yields 5 different values of $\tilde{TO}$, since future payments are simulated using 5 different methods. Namely, these different values are

- $\tilde{TO}^H$ for the Gaussian copula model using the heuristic estimates of $\mathbf{R}$ and $\mathbf{s}$ displayed in Table 3.6
- $\tilde{TO}^{ML}$ for the Gaussian copula model using the maximum likelihood estimates of $\mathbf{R}$ and $\mathbf{s}$ displayed in Table 3.6
- $\tilde{TO}^{sML}$ for the Gaussian copula model using the stepwise maximum likelihood estimates of $\mathbf{R}$ and $\mathbf{s}$ displayed in Table 3.6
- $\tilde{TO}^{D-Vine}$ for the D-Vine model using the parameters displayed in Table 4.10 and
- $\tilde{TO}^{CL}$ for the Chain-Ladder method.

Since, naturally, the amount of total future obligations is most important to reinsurance companies model choice can be influenced by the value of $\tilde{TO}$. Thus the models presented in this work will be compared in terms of the obtained value of $\tilde{TO}$. The lower the value the higher the diversification benefit is and vice versa. However, it has to be noted that it is not recommendable to base model choice only on selecting the model leading to the lowest reserves since an underestimation of the reserves leads to unforeseen payments in the future. It is important to stress that there is no diversification benefit at all using the Chain-Ladder method. Since the Chain-Ladder method is deterministic it is not possible to measure reserve uncertainty for this model and hence the tables below only present the estimated reserves obtained by applying this model but no quantiles.
The results of $N = 10000$ contemporaneous simulations of future payments $\hat{p}^r_{lij}$ in the $m = 5$ different loss triangles are presented in Tables 5.3 to 5.7.

model / results	5% quantile	median	mean	95% quantile
Heuristic	5878	7233	7315	9037
ML	5932	7243	7327	8973
stepwise ML	6112	7252	7318	8742
D-Vine	5670	7169	7301	9385
Chain Ladder			7153	

Table 5.3: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for triangle 1, $\hat{TO}^r_1$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

Table 5.3 displays the results provided by the Heuristic-, ML-, stepwise ML-, DVine- and Chain Ladder- model for loss triangle 1. The empirical mean of $\hat{TO}^r_1$ is almost the same for the stochastic models, the Chain Ladder estimate lies slightly below. Thus applying one of the stochastic models would lead to higher reserves for loss triangle 1.
From the second and fifth column one can observe that in the stepwise Maximum Likelihood model the outcomes are most centered around the mean, while those for the D-Vine model spread widely around the mean. This leads to high reserve uncertainty in the DVine-model while the reserve uncertainty for the stepwise Maximum Likelihood model is rather low for loss triangle 1.

model / results	5% quantile	median	mean	95% quantile
Heuristic	2743	3037	3055	3424
ML	2783	3055	3043	3367
stepwise ML	2762	3044	3055	3391
D-Vine	2747	3037	3054	3414
Chain Ladder			3194	

Table 5.4: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for triangle 2, $\hat{TO}^r_2$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

From Table 5.4 one can see that, in contrast to loss triangle 1, the mean of $\hat{TO}_2^r$ for the Chain Ladder model is higher than it is for the stochastic models. Again the stochastic models yield empirical means at the same level. Applying one of the stochastic models presented in this work would lower the reserves for loss triangle 2. The spread around the empirical mean is almost at the same level for each model and thus in terms of reserve uncertainty none is favoured.

model / results	5% quantile	median	mean	95% quantile
Heuristic	3571	3855	3872	4234
ML	3569	3855	3873	4236
stepwise ML	3581	3855	3869	4211
D-Vine	3560	3849	3865	4220
Chain Ladder			3619	

Table 5.5: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for triangle 3, $\hat{TO}_3^r$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

The results for loss triangle 3 presented in Table 5.5 show that application of the Chain Ladder method leads to lower reserves than applying one of the stochastic models. Again, the empirical means of $\hat{TO}_3^r$ are at one level for the Gaussian Copula models and the D-Vine model. In terms of reserve uncertainty none of the models is favoured.

model / results	5% quantile	median	mean	95% quantile
Heuristic	3674	4093	4117	4649
ML	3748	4098	4119	4553
stepwise ML	3714	4102	4121	4590
D-Vine	3662	4098	4122	4659
Chain Ladder			3960	

Table 5.6: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for triangle 4, $\hat{TO}_4^r$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

For loss triangle 4 we have that the empirical means of $\hat{TO}_4^r$ are at the same level for the 4 different stochastic models. The empirical mean yielded by the Chain Ladder model is slightly below this level.
For the Heuristic model and the D-Vine model one can observe higher reserve uncertainty than for the Maximum Likelihood model and the stepwise Maximum Likelihood model.

model / results	5% quantile	median	mean	95% quantile
Heuristic	3637	4378	4436	5452
ML	3651	4375	4441	5440
stepwise ML	3670	4438	4369	5408
D-Vine	3704	4392	4448	5397
Chain Ladder			4727	

Table 5.7: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for triangle 5, $\hat{TO}_5^r$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

Table 5.7 displays the simulated results for loss triangle 5. As it is the case for all other loss triangles the empirical means of $\hat{TO}_5^r$ barely differ for the stochastic models. The Chain Ladder model suggests higher reserves than the stochastic models.
In terms of reserve uncertainty the stepwise Maximum Likelihood model or the D-Vine model should be favoured for loss triangle 5.
Since the total amount of future obligations in all triangles is most relevant to any insurance company Table 5.8 displays the simulation results for the sum of the $m = 5$ loss triangles used in this work.

model / results	5% quantile	median	mean	95% quantile
Heuristic	20342	22674	22795	25586
ML	21101	22748	22815	24703
stepwise ML	20501	22716	22801	25351
D-Vine	20485	22656	22790	25514
Chain Ladder			22653	

Table 5.8: Comparison of the Heuristic-, ML-, stepwise ML, D-Vine and Chain Ladder model in terms of different outcomes of the total future obligations for the sum of $m = 5$ loss triangles , $\hat{TO}^r$, $r = 1, \ldots, N$, including 5% and 95% empirical quantiles, the mean and median of $N = 10000$ simulations of future payments.

According to Table 5.8 the estimated total future obligations for the sum of the 5 loss triangles does not depend on the choice of a certain stochastic model since the empirical means of $\hat{TO}^r$ are almost the same for the Heuristic, Maximum Likelihood, stepwise Maximum Likelihood and D-Vine model. The estimate obtained from the Chain Ladder model is slightly lower. Thus application of one of the stochastic models presented in this work would lead to higher reserves than application of the Chain Ladder method.
The Maximum Likelihood model leads to low reserve uncertainty since the empirical quantiles are closer to the empirical mean for this method than they are for the other methods and should therefore be favoured in terms of reserve uncertainty. But as displayed in section 3.4 the parameter estimates for this model are quite unstable.
To conclude it has to be noted that each of the models provides significant advantages as well as drawbacks. Thus it mainly depends on the reserving actuaries in charge to choose a certain model depending on the available data and goals of the company.

5.2.2 Forecast based on Simulated Data

In this section the different methods of forecasting future payments will be applied to simulated data. This is achieved as follows:

- As a first step $m \cdot K$ different loss squares are generated. In order to obtain a simulated loss square one generates a loss triangle using Algorithm 5 for the heuristic approach in the Gaussian copula model, Algorithm 4 for the maximum likelihood and stepwise maximum likelihood approach and Algorithm 10 for the D-Vine model.
- Each of these $m \cdot K$ simulated loss triangles is filled to obtain a loss square by using Algorithm 6 for the Gaussian copula model and Algorithm 11 for the D-Vine model.
- The sum of the total future obligations of triangle l in the k-th simulation $TO^O_{l,k}$ is interpreted as the observed total future obligation for loss triangle l in the k-th simulation, $k = 1, \ldots, K$.
- After saving the values of $TO^O_{l,k}$ and deleting future payments $p^k_{lij},\ i + j > n,\ k = 1, \ldots, K$, in each loss triangle one makes D forecasts of each of the $m \cdot K$ simulated loss triangles in order to obtain $\hat{TO}_{l,d,k},\ d = 1, \ldots, D$, which denotes the simulated total future obligations for loss triangle l in the d-th step of simulation.
- Taking the $(1-\beta)\cdot 100\%$ and the $\beta\cdot 100\%$ empirical quantiles of $\hat{TO}_{l,d,k},\ d = 1, \ldots, D$, yields an interval $I_{l,k}$ which covers $TO^O_{l,k}$ with probability $(1 - 2\beta) \cdot 100\%$.
- Comparison between the different models is now based on these intervals. In detail one counts the numbers of simulations C for which $TO^O_{l,k}$ is not covered in $I_{l,k}$ and compares the lengths of these intervals for the different methods.

This procedure is applied for each of the different models, namely the Gaussian copula model using the heuristic, maximum likelihood and stepwise maximum likelihood approach and the D-Vine model. The parameters used to generate the K different loss squares are those presented in Table 3.6 for the Gaussian copula model and those presented in Table 4.10 for the D-Vine model.

Model	C					average length of $I_{l,k}$				
Heuristic	14	12	**10**	9	17	317	660	658	1013	1487
Maximum Likelihood	8	**11**	**10**	9	**9**	304	**568**	647	**858**	1448
stepwise ML	14	**11**	9	12	12	**262**	609	**614**	927	**1375**
D-Vine Model	**10**	12	11	**10**	8	288	604	632	921	1402

Table 5.9: Values of C for each triangle (each column represents one triangle) and method and average length of $I_{l,k}$ for each triangle and each method for $K = 100$, $D = 1000$ and $\beta = 0.05$; values of C with closest difference to $10 = 2 \cdot 0.05 \cdot 100$ and smallest average interval lengths bolded

Table 5.9 displays the values of C for eacht triangle $l,\ l = 1, \ldots, 5$ and each of the different models as well as the average length of the intervals $I_{l,k},\ k = 1, \ldots, K$ for

$K = 100$, $D = 1000$ and $\beta = 0.05$. As can be seen from the table the maximum likelihood and stepwise maximum likelihood approaches in the Gaussian copula model produce the tightest intervals $I_{l,k}$. In terms of C being close to it's expected value 10 the maximum likelihood approach and the D-Vine model yield better results than the heuristic and the stepwise maximum likelihood approach.
Thus in terms of data forecast and low desired reserve uncertainty it is recommended to use the maximum likelihood approaches in the Gaussian copula model since the D-Vine model leads to higher reserve uncertainty than these approaches.

Discussion

The models presented in this work may overcome one point of criticism for existing loss reserving methods since they provide the possibility to adjust for dependencies between payments made in one loss triangle or between payments in different loss triangles.
The Gaussian copula model presented in chapter 2 gives explicit estimates of correlation parameters between loss triangles and within a certain loss triangle. The D-Vine model presented in chapter 4 first eliminates the inner-triangle dependence in order to receive loss triangles with independent entries. A D-Vine structure is fitted on these transformed triangles to model the dependence structure.
Even though these models are quite different they provide very similar reserve estimates as discussed in chapter 5. The models are capable of reproducing loss triangle data. In terms of this data fit it has been shown that the Gaussian copula model using the heuristic approach performs best. In terms of data forecast, however, one could see that the maximum likelihood approaches outperform the other models.
What despite the allocation of dependencies makes the models superior to standard techniques such as the Chain-Ladder method is the fact that one obtains quantiles for the triangle reserves. Thus the insurance company applying the models can easily determine a reserve which is very unlikely to be non-sufficient.
To sum up the models derived in this thesis are an alternative to the classic models for reserve estimation and may even displace those due to their ability to adjust for dependencies and reserve uncertainty.

Bibliography

C. Czado A.Frigessi Aas, K. and H.Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, pages 182–198, 2009.

Cooke R.M. Bedford, T. Monte carlo simulation of vine dependent random variables for applications in uncertainty analysis. *Proceedings of ESREL2001. Turin, Italy*, 2001a.

Cooke R.M. Bedford, T. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2001b.

Cooke R.M. Bedford, T. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2002.

Shapiro S.S. Chen, H. J. and M.B. Wilk. A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63:1343–1372, 1968.

Kevin A. Clarke. A simple distribution-free test for nonnested model selection. *Political Analysis*, 15:347–363, 2007.

C. Czado and A. Min. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics, to appear*, 2010.

Piet de Jong. Modelling dependence between loss triangles using copulas. *North American Actuarial Journal*, 2009.

Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 9th edition, 2008.

H. Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Rueschendorf, L., Schweizer B., Talyor, M.D.(Eds), Distributions with Fixed Marginals and Related Topics*, 1996.

H. Joe. *Multivariate models and dependence concepts.* Chapman and Hall, 1st edition, 1997.

R.Cooke Kurowicka, D. Uncertainty analysis with high dimensional dependence modelling. *Wiley Series in Probability and Statistics. Chichester: John Wiley and Sons Ltd*, 2006.

R.Cooke Kurowicka, D. Sampling algorithms for generating joint uniform distributions using the vine - copula method. *Computational Statistics and Data Analysis*, 51:2889–2906, 2007.

T. Mack. *Schadenversicherungsmathematik*. Verlag Versicherungswirtschaft, 2nd edition, 1997.

Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics, 2nd edition, 1999.

A. V. Pearson and H. O. Hartley. Biometrica tables for statisticians. *Cambridge University Press*, 2, 1972.

Klaus D. Schmidt. *Versicherungsmathematik*. Springer-Verlag, Heidelberg, New York, 1st edition, 2001.

Jun Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer, Madison, 2nd edition, 2007.

S.S. Shapiro and M.B. Wilk. An analysis of variance test for normality. *Biometrika*, 3: 591–611, 1965.

Quang H. Vuong. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica*, 57:307–333, 1989.

Halbert White. Maximum likelihood estimator of misspecified models. *Econometrica*, 50: 1–25, 1982.