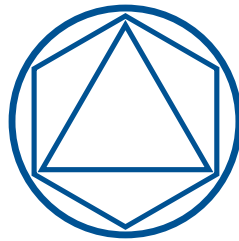




Technische Universität München

Department of Mathematics



D-Vine Copula Based Modelling and Forecasting of Exposure Limits in Reinsurance

Master's Thesis

by

Felix Hunschede

Supervisor:	Prof. Claudia Czado, PhD
Advisors:	Prof. Claudia Czado, PhD Nicole Barthel
Submission Date:	27.09.2017

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

München, 27.09.2017

Acknowledgements

First of all, I want to thank Prof. Claudia Czado for giving me the opportunity to work on this thesis under her supervision. Her help and guidance were always greatly appreciated and helped me immensely during the last months.

In the same way I am thankful to my advisor, Nicole Barthel, for her many helpful suggestions, the amount of time and the patience committed to this project.

I also want to thank Munich Re for giving me the chance to work with them during this project. The whole team of the Credit Risk department and especially my advisor, Dr. Jianing Zhang, have invested many hours and thoughts into this work and were always there to help me, for which I am very grateful.

I would like to specifically appreciate the work of the many developers of the `R` project (R Core Team, 2017) whose free software made most of this thesis feasible. In the same way, my appreciation goes out to the developers of the packages `ggplot2` (Wickham, 2009) and `tikz` (Tantau, 2015) with which all of the graphics in this thesis were created but not cited otherwise. And finally, I want to express my deepest gratitude towards my parents for their unconditional support, help and patience throughout the many years of my studies.

Abstract

We provide a D-vine copula based method to model and forecast data, which are relevant in reinsurance, e.g. for appropriate risk capital allocation. We deal with *unbalanced data*, where the number of observations increases over time. Using the *Inference for Marginals* (IFM) method, we first obtain probability integral transformed random variables by modelling the marginals via different methods. This includes elaborate regression models such as *Generalized Additive Models to account for location, scale and shape* (GAMLSS). Second, we fit D-vine models, which naturally capture the serial dependence inherent in the data. A concept for forecasting future data by shifting the fitted D-vine structure is introduced. The quality of predictions is assessed with *continuous rank probability scoring* (CRPS).

Zusammenfassung

Wir stellen eine D-Vine Copula basierte Methode vor, die zur Modellierung und Vorhersage von Daten genutzt werden kann, welche in der Rückversicherung z.B. zur Allokation von Risikokapital relevant sind. Wir arbeiten mit einem sogenannten *unbalanced setting*, in dem die Anzahl der Beobachtungen über die Zeit zunimmt. Mithilfe der *Inference for Marginals* (IFM) Methode erhalten wir zuerst sogenannte *probability integral transformed* Zufallsvariablen, indem wir die Ränder mit verschiedenen Methoden modellieren. Darunter auch komplexe Regressionsmodelle, wie etwa die *Generalized Additive Models for location, scale and shape* (GAMLSS). Als zweiten Schritt passen wir D-Vine Modelle an, welche von Natur aus die sequentielle Abhängigkeit der Daten abbilden. Anschließend wird ein Konzept zur Vorhersage von zukünftigen Daten durch das Verschieben der angepassten D-Vine Struktur eingeführt. Schlussendlich wird die Qualität der Vorhersagen durch das sogenannte *Continuous Rank Probability Scoring* (CRPS) bewertet.

Contents

1	Introduction	1
2	Theoretical background	3
2.1	Dependence modelling using copulas	3
2.2	Dependence modelling using D-vine copulas	12
2.3	Marginal modelling	19
2.4	Forecasting with D-vine copulas	23
3	Exploratory data analysis	29
3.1	Data patterns	30
3.2	Graphical analysis	33
4	Marginal model fitting	43
4.1	Marginal model specification	43
4.2	Overview of different marginal model approaches	50
4.3	Marginal models comparison	50
4.4	Comparison of u-data	55
5	Dependence modelling	59
5.1	Pairs plots	59
5.2	Vine copula estimations	64
6	Conclusion	79
A	Extended D-vine estimation results and contour plots	81
	Bibliography	103

Chapter 1

Introduction

A reinsurance works as an insurance for another (primary) insurance. Looking at a contract between these two parties, one important piece of information that the primary insurance shares with the reinsurance is a list of so-called exposure limits. These limits are of special interest when the reinsurance company wants to, for example, calculate pricings or allocate risk capital. Thus, good knowledge of these limits is important for reinsurance businesses. However, usually only data from the past are available while the reinsurance is interested in information on future time points.

In order to provide better knowledge on future time points, a statistical model is needed. When it comes to multivariate data, typical features are different marginal distributions, tail dependency between certain pairs as well as non-symmetric dependencies. Classical distributions, such as the multivariate Gaussian distribution can be used to model the data but may not be flexible enough to capture the aforementioned behaviour. Especially complex dependence structures of the data can be challenging to model correctly. The theory of copulas, based on the work of Sklar (1959), allows us to model the underlying dependence structure separately from the margins. This approach is very flexible and has thus gained a lot of popularity (see e.g. Nelsen, 2010 for an overview)

In this thesis, we will apply the copula approach by making use of so-called vine copulas, as introduced by Aas, Czado et al. (2009). The advantage of vine copulas is the application of simple so-called bivariate pair-copulas as building blocks for modelling in arbitrary dimensions. With the help of vine copulas we will fit multiple models in this thesis in order to describe the dependencies and allow forecasts of future exposure limits.

The remainder of this thesis is organized as follows. Chapter 2 lays out the necessary theoretical background for this thesis. First, the theory of copula and vines, in particular D-vines, is introduced and accompanied by the procedure of fitting a vine and diagnostic methods for them. Then, details on marginal models, including the introduction of the regression approach called Generalized Additive Models for Location, Scale and Shape (GAMLSS), are given and again accompanied by a few diagnostic tools. In the last section of this chapter, the concept of shifting a fitted model is proposed in order to forecast data and finally, a tool for assessing the forecasting quality called continuous rank probability score (CRPS) is presented.

Chapter 3 provides an exploratory data analysis of the available data. We first determine a set of data that we want to work with and give details of its properties. In the second half, we visually explore this data set with multiple plots in order to determine further characteristics.

In Chapter 4, we introduce four different ways to model the margins. The first model is based on a simple non-parametric density estimation of the limits themselves. The second model works with standardized observations whose density is estimated again non-parametrically. The third and fourth model make use of the previously mentioned GAMLSS models to gain residuals in a parametric way before estimating their densities once parametrically and once non-parametrically. The chapter also includes comparisons of the marginal models and diagnostics for the resulting data.

Chapter 5 begins with a visual analysis of the dependence structure via pairs plots. It then continues with the estimation of D-vine copulas based on four different constraints: There are two sets of copula families that are allowed and for each set the estimation procedure once contains a test for independence and once doesn't. At the end, the estimation results are compared regarding the fit and also the quality of estimation is assessed with the help of the continuous rank probability scoring (CRPS).

Finally, a short conclusion is given in Chapter 6.

Chapter 2

Theoretical background

In this chapter, we introduce the necessary theory and tools used in our subsequent work. This includes basics in copula theory and dependence measures, the concept of D-vines, a brief introduction into the GAMLSS framework, as well as some diagnostic and scoring tools. We also propose a way to forecast data in this D-vine copula framework and present the CRPS method to assess the forecast quality.

2.1 Dependence modelling using copulas

We start with an introduction to copula theory. Copulas are tools for modelling dependence of several random variables.

Definition 1 (Copula). *A d -dimensional copula is a multivariate distribution function $C : [0, 1]^d \rightarrow [0, 1]$ with uniformly distributed marginals. Let (U_1, \dots, U_d) be a random vector where $U_j \sim \mathcal{U}[0, 1]$ for $j = 1, \dots, d$, then*

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) \quad \text{for } u_1, \dots, u_d \in [0, 1].$$

A fundamental statement about copulas is Sklar's theorem (Sklar, 1959) as it provides the foundation for the application of copulas by giving us a link between multivariate distributions and their copulas. It states that for every multivariate distribution F of a random vector $\mathbf{X} = (X_1, \dots, X_d)$ with marginal distributions F_1, \dots, F_d , there exists a copula C such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

and in the case that F is absolutely continuous (which we will assume throughout this thesis), this decomposition is unique.

A similar relationship can be given for the density function

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d), \quad (2.1)$$

where f_1, \dots, f_d are the marginal densities and c is the copula density defined as

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

If we are only interested in the dependence structure of the random vector \mathbf{X} , we can apply the probability integral transform (PIT) to its known marginals to obtain the so-called u-data:

$$U_j = F_j(X_j) \sim \mathcal{U}[0, 1], \quad j = 1, \dots, d. \quad (2.2)$$

And similarly

$$X_j = F_j^{-1}(U_j) \sim F_j, \quad j = 1, \dots, d.$$

So the copula C corresponding to the cumulative distribution function F of \mathbf{X} is the distribution function of $F_1(X_1), \dots, F_d(X_d)$ and we obtain for $(U_1, \dots, U_d) \sim C$:

$$(F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)) \sim F.$$

Thus, based on Sklar's theorem, we can use the following 2-stage process: First, estimate the marginal distributions to obtain u-data of which the margins are approximately uniformly distributed on $[0, 1]$ and then estimate a copula describing the dependence structure for the estimated u-data which inherits the dependence structure of the original data, see Joe and Xu (2016) and Joe (2005) for details on the 2-stage estimation process, called inference for marginals (IFM).

Dependence measures

As copulas describe a dependence structure, we now want to introduce measures to quantify this dependence.

First, we introduce the rank-based global dependence measure called Kendall's rank correlation coefficient, also called Kendall's τ , developed by Kendall (1938):

Definition 2 (Kendall's τ). *For the i.i.d. (independent and identically distributed) random vectors $(X, Y)'$ and $(\tilde{X}, \tilde{Y})'$, Kendall's τ is defined as*

$$\tau(X, Y) = \mathbb{P}\left(\left(X - \tilde{X}\right)\left(Y - \tilde{Y}\right) > 0\right) - \mathbb{P}\left(\left(X - \tilde{X}\right)\left(Y - \tilde{Y}\right) < 0\right).$$

We list some properties of Kendall's τ from Embrechts, Lindskog et al. (2001):

Proposition 1 (Properties of Kendall's τ). *Let $(X, Y)'$ be a vector of continuous random variables with copula C . Then*

1. $\tau(X, Y) = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1.$
2. $\tau(X, Y) \in [-1, 1].$
3. *If X and Y are independent, then $\tau(X, Y) = 0.$*
4. *If X and Y are perfectly positively dependent, then $\tau(X, Y) = 1.$*
5. *If X and Y are perfectly negatively dependent, then $\tau(X, Y) = -1.$*

As the first property shows, Kendall's τ can be expressed using only copulas and thus is independent of the marginal distributions. Hence, it allows us to completely describe the dependence captured by the copula.

Next, we describe the concept of tail dependence, which is relevant for the study of dependence between extreme values as it relates to the dependence in the upper-right- or lower-left-quadrant tail of a bivariate distribution (from Embrechts, Lindskog et al. (2001)):

Definition 3 (Tail dependence coefficients). Let $(X, Y)'$ be a vector of continuous random variables with marginal distribution functions F and G . The coefficient of upper tail dependence of $(X, Y)'$ is

$$\lim_{u \nearrow 1} \mathbb{P}(Y > G^{-1}(u) | X > F^{-1}(u)) =: \lambda_U(X, Y)$$

provided that the limit $\lambda_U \in [0, 1]$ exists.

If $\lambda_U(X, Y) \in (0, 1]$, X and Y are said to be asymptotically dependent in the upper tail and if $\lambda_U(X, Y) = 0$, X and Y are said to be asymptotically independent in the upper tail. Similarly,

$$\lim_{u \searrow 0} \mathbb{P}(Y < G^{-1}(u) | X < F^{-1}(u)) =: \lambda_L(X, Y)$$

is the coefficient of lower tail dependence.

Further, Embrechts, Lindskog et al. (2001) show that the concept of tail dependence is a copula property and give the equivalent definition:

Definition 4 (Copula tail dependence coefficients). If a bivariate copula C is such that

$$\lim_{u \nearrow 1} (1 - 2u + C(u, u)) / (1 - u) =: \lambda_U(X, Y)$$

exists, then C has upper tail dependence if $\lambda_U(X, Y) \in (0, 1]$ and upper tail independence if $\lambda_U(X, Y) = 0$. Similarly, if

$$\lim_{u \searrow 0} C(u, u)/u =: \lambda_L(X, Y)$$

exists, then C has lower tail dependence if $\lambda_L(X, Y) \in (0, 1]$ and upper tail independence if $\lambda_L(X, Y) = 0$.

Bivariate copula families

In this section, we will list a number of commonly used bivariate copula families (descriptions are mostly from Barthel (2015)).

The first one is the non-parametric independence copula (Embrechts, McNeil et al., 2002)

Example 1 (Independence copula). Let $U_1, U_2 \stackrel{i.i.d.}{\sim} \mathcal{U}[0, 1]$. Then the independence copula is given by

$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) = u_1 u_2 .$$

Next, we introduce the Gaussian and t-copula, which belong to the so-called class of elliptical copulas, named for the elliptical bivariate distributions they are based on. They are both symmetric with respect to the diagonal as well as with respect to the counterdiagonal (Embrechts, McNeil et al., 2002).

Example 2 (Gaussian copula). Let Φ_ρ denote the cumulative distribution function of a bivariate standard normal distribution (zero mean, unit variance) and correlation $\rho \in (-1, 1)$. Let Φ denote the cumulative distribution function of a standard

univariate normal distribution. Then the bivariate Gaussian copula with parameter ρ is given by

$$C(u_1, u_2; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) .$$

Example 3 (Student t-copula). Let $t_{\nu, \rho}$ denote the cumulative distribution function of a bivariate t-distribution with zero mean, correlation $\rho \in (-1, 1)$ and ν degrees of freedom. Let t_ν^{-1} denote the inverse of a cumulative distribution function of a standard univariate t-distribution with ν degrees of freedom. Then the bivariate t-copula (also called Student t-copula) with parameters ν and ρ is given by

$$C(u_1, u_2; \nu, \rho) = t_{\nu, \rho}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) .$$

For the class of so-called Archimedean copulas, we introduce the Clayton, Gumbel, Frank and Joe copula (Clayton (1978), Gumbel (1960), Frank (1979) and Joe (1993)). Archimedean copulas can be constructed with an appropriate generator function φ (Nelsen, 2010):

Definition 5 (Bivariate Archimedean copulas). Let $\varphi : [0, 1] \rightarrow [0, \infty]$ be a continuous, strictly monotone decreasing and convex function such that $\varphi(1) = 0$ and let the pseudo-inverse of φ be the function $\varphi^{[-1]} : [0, \infty] \rightarrow [0, 1]$ given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) , & 0 \leq t \leq \varphi(0) , \\ 0 , & \varphi(0) \leq t \leq \infty , \end{cases}$$

with $\varphi^{[-1]}(t) = \varphi^{-1}(t)$ for $\varphi(0) = \infty$. Then

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$$

is a bivariate Archimedean copula with generator φ .

Example 4 (Clayton copula). Choosing

$$\varphi(t) = \frac{1}{\theta} (t^{-\theta} - 1)$$

with dependence parameter $\theta \in (0, \infty)$, the bivariate Clayton copula is given by

$$C(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-1/\theta} .$$

Example 5 (Gumbel copula). Choosing

$$\varphi(t) = (-\log(t))^\theta$$

with dependence parameter $\theta \in [1, \infty)$, the bivariate Gumbel copula is given by

$$C(u_1, u_2; \theta) = \exp \left[- \left((-\log(u_1))^\theta + (-\log(u_2))^\theta \right)^{1/\theta} \right] .$$

Example 6 (Frank copula). Choosing

$$\varphi(t) = -\log \left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1} \right)$$

with dependence parameter $\theta \in \mathbb{R} \setminus \{0\}$, the bivariate Frank copula is given by

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{(\exp(-\theta) - 1)} \right).$$

Example 7 (Joe copula). *Choosing*

$$\varphi(t) = -\log \left(1 - (1 - t)^\theta \right)$$

with dependence parameter $\theta \in [1, \infty)$, the bivariate Joe copula is given by

$$C(u_1, u_2; \theta) = 1 - \left((1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta \right)^{1/\theta}.$$

Similarly to Archimedian copulas, extreme value copulas can be constructed with the help of the Pickands' representation:

$$C(u, v; A) = \exp \left[\log(uv) A \left(\frac{\log(u)}{\log(uv)} \right) \right], \quad (2.3)$$

where $A : [0, 1] \rightarrow [0, 1]$ is the so-called Pickands' dependence function (Pickands, 1981). Note that for $u = v = 1$, the dependence function A is set to 0. One example of extreme value copulas is the Tawn copula (Tawn, 1988):

Example 8 (Tawn copula). *Choosing the Pickands' dependence function as*

$$A(t) = 1 - \beta + (\beta - \alpha)t + \left[(\alpha(1 - t))^\theta + (\beta t)^\theta \right]^{1/\theta},$$

for $t \in [0, 1]$ with $\alpha, \beta \in [0, 1]$ and $\theta \in [1, \infty)$, and inserting it into equation (2.3), we obtain the Tawn copula with three parameters. The special cases where $\beta = 1$ or $\alpha = 1$ are called Tawn Type 1 and Tawn Type 2 copula respectively (which have two parameters).

For each of the above mentioned copula families, Table 2.1 shows the Kendall's τ values as well as upper and lower tail dependence coefficients in terms of the copula parameter(s). Figures 2.1 and 2.2 show scatter plots of simulated copula data as well as normal marginal contour plots of the underlying copula densities for each of the copula families, i.e. the contour plots consider the transformation to a joint distribution with normal $\mathcal{N}(0, 1)$ margins.

Family	Kendall's τ	τ range	Lower tail dependence	Upper tail dependence
Gaussian	$\tau = \frac{2}{\pi} \arcsin(\rho)$	$\tau \in [-1, 1]$	-	-
Student t	$\tau = \frac{2}{\pi} \arcsin(\rho)$	$\tau \in [-1, 1]$	$2t_{\nu+1} \left(-\sqrt{\nu+1} \sqrt{\frac{1-\rho}{1+\rho}} \right)$	$2t_{\nu+1} \left(-\sqrt{\nu+1} \sqrt{\frac{1-\rho}{1+\rho}} \right)$
Clayton	$\tau = \frac{\theta}{\theta+2}$	$\tau \in [0, 1]$	$2^{-1/\theta}$	-
Gumbel	$\tau = 1 - \frac{1}{\theta}$	$\tau \in [0, 1]$	-	$2 - 2^{1/\theta}$
Frank	$\tau = 1 - \frac{4}{\theta} + 4 \frac{D_1(\theta)}{\theta}$ with $D_1(\theta) = \int_0^\theta \frac{x/\theta}{\exp(x)-1} dx$ (Debye function)	$\tau \in [-1, 1]$	-	-
Joe	$\tau = 1 + \frac{4}{\theta^2} \int_0^1 x \log(x) (1-x)^{2(1-\theta)/\theta} dx$	$\tau \in [0, 1]$	-	$2 - 2^{1/\theta}$
Tawn Type 1	$\tau = \int_0^1 \frac{t(1-t)A''(t)}{A(t)} dt$ where $A(t) = (1-\delta)t + [(\delta(1-t))^\theta + (t)^\theta]^{1/\theta}$	$\tau \in [0, 1]$	-	$\delta + 1 - (\delta^\theta + 1)^{1/\theta}$
Tawn Type 2	$\tau = \int_0^1 \frac{t(1-t)A''(t)}{A(t)} dt$ where $A(t) = 1 - \delta + (\delta - 1)t + [(1-t)^\theta + (\delta t)^\theta]^{1/\theta}$	$\tau \in [0, 1]$	-	$\delta + 1 - (\delta^\theta + 1)^{1/\theta}$

TABLE 2.1: Kendall's τ values as well as upper and lower tail dependence coefficients (where available) in terms of the copula parameter(s) for the families given in Examples 2 to 8 (from Schepsmeier, Stoeber et al. (2017b)).

All of the above copulas can also be rotated by 90, 180 and 270 degrees. This is useful as some of them only have a limited dependence range or can only capture lower or upper tail dependence (see Table 2.1). Rotating can be defined as follows (Joe, 1993):

Definition 6. For a copula function $C(\cdot, \cdot)$ with density $c(\cdot, \cdot)$, its counter-clockwise rotations are given by

- 90 degrees: $C^{90}(u_1, u_2) := u_2 - C(1 - u_1, u_2)$
 $c^{90}(u_1, u_2) := c(1 - u_1, u_2)$
- 180 degrees: $C^{180}(u_1, u_2) := u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$
 $c^{180}(u_1, u_2) := c(1 - u_1, 1 - u_2)$
- 270 degrees: $C^{270}(u_1, u_2) := u_1 - C(u_1, 1 - u_2)$
 $c^{270}(u_1, u_2) := c(u_1, 1 - u_2)$

Remark that the 180 degree rotation of a copula is often called the survival copula.

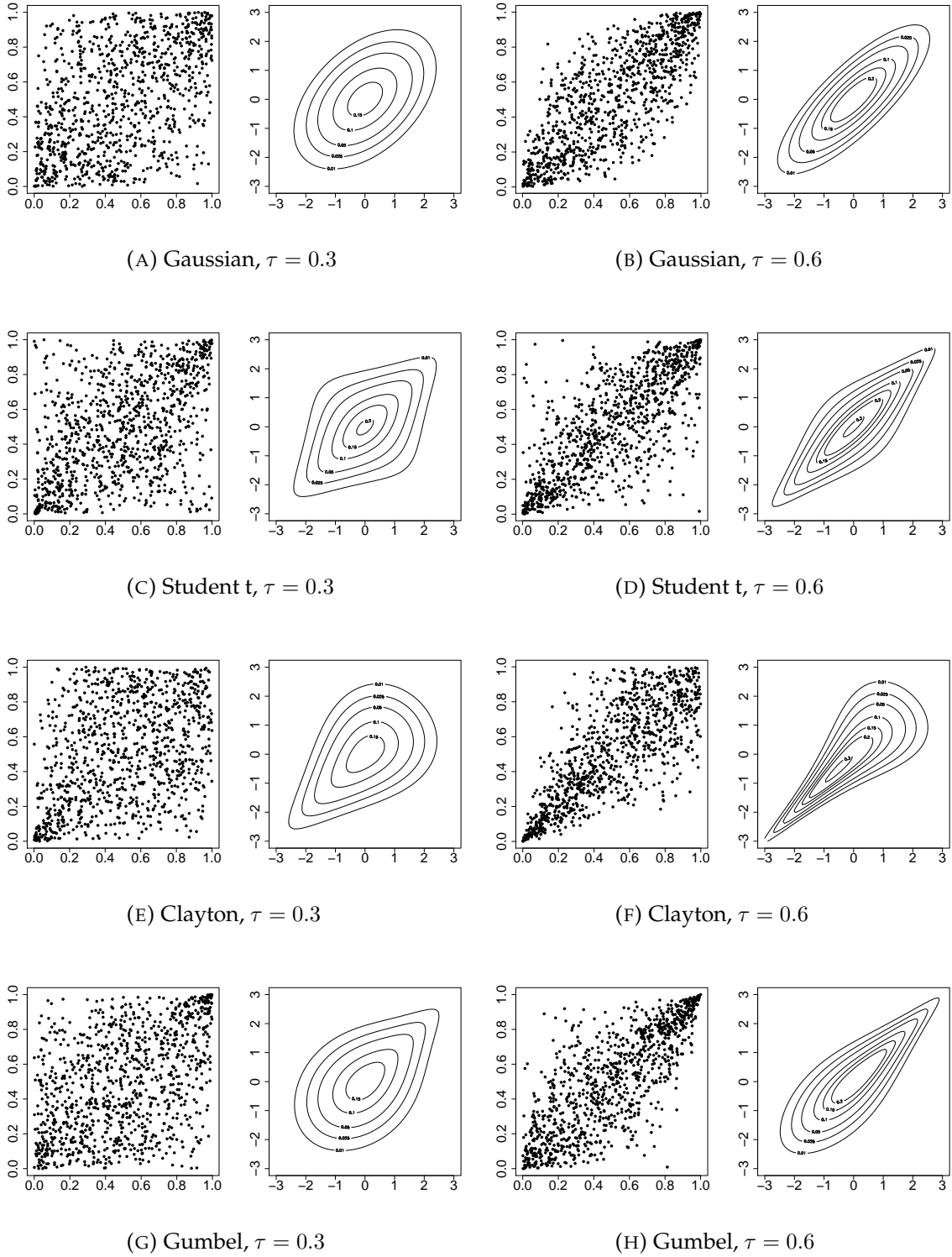


FIGURE 2.1: Scatter plots of simulated copula data for 1000 simulations as well as normal contour plots of the underlying copula density with weaker ($\tau = 0.3$) and stronger ($\tau = 0.6$) dependence for the families given in Examples 2 to 5.

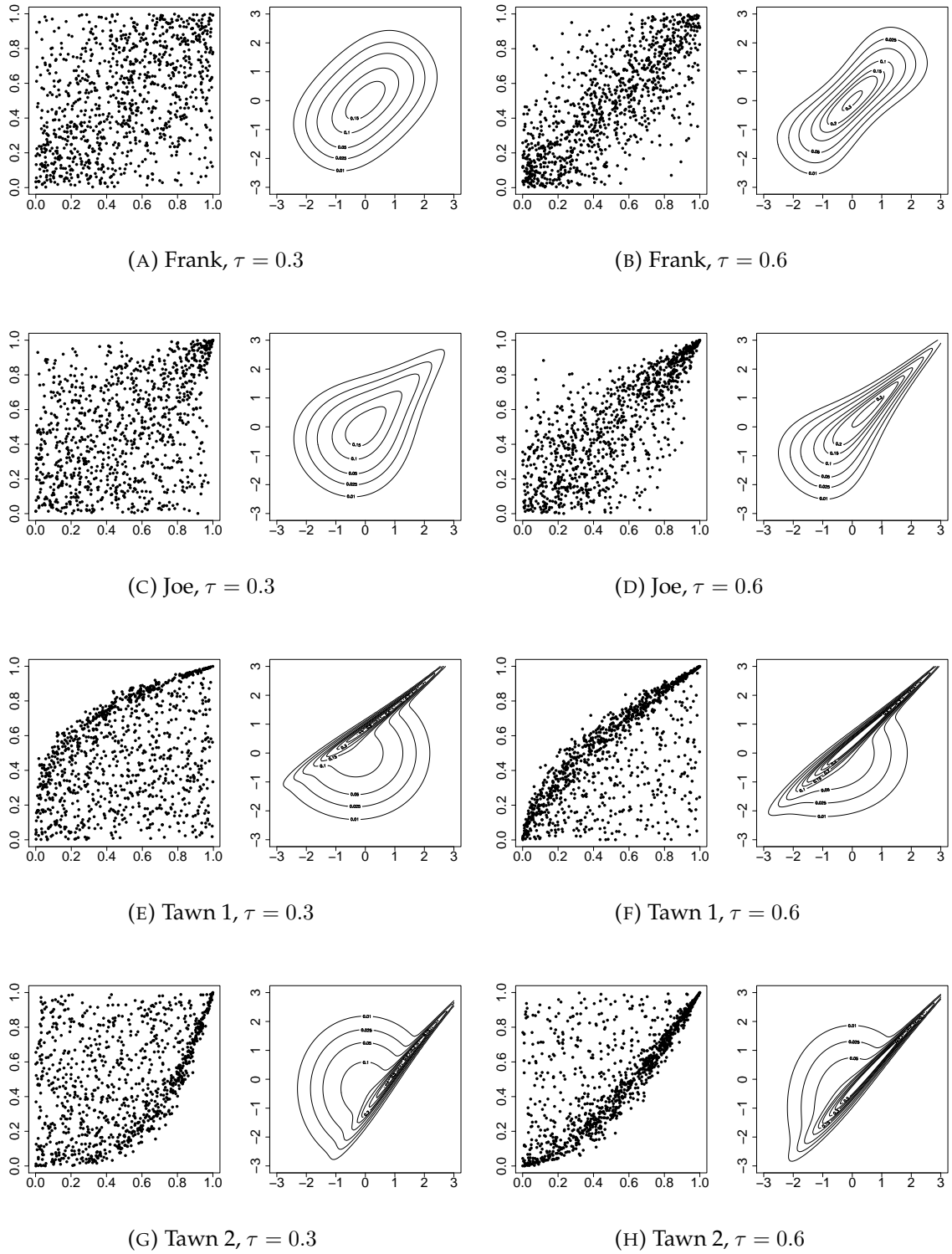


FIGURE 2.2: Scatter plots of simulated copula data for 1000 simulations as well as normal contour plots of the underlying copula density with weaker ($\tau = 0.3$) and stronger ($\tau = 0.6$) dependence for the families given in Examples 6 to 8.

2.2 Dependence modelling using D-vine copulas

At the end of the last section, we gave examples of bivariate copulas (such as Gaussian, Gumbel or Tawn copula). They are of special interest as they provide the foundation for so-called vine copulas, which we will discuss in this section (as introduced, for example, in Bedford and Cooke (2001)). A vine is a graphical tool that allows for the ordering of decomposed densities via recursive factorisation into pairs. (see Diestel (2000) for an extensive introduction to graph theory, of which we will use some basics). This can be of advantage as there are many bivariate parametric pair-copulas available to choose from, making this a very flexible model. In contrast, only few copula families exist for higher dimensions.

A vine \mathcal{V} is a nested set of connected trees, where the edges in the first tree are the nodes of the second tree, the edges in the second tree are the nodes of the third tree and so on.

A regular vine (or R-vine) is a vine in which two edges in a tree are joined by an edge in the next tree only if these edges share a common node. This is called the proximity condition.

In this thesis, we will focus on so-called ordered D-vines. A D-vine is characterized by assigning no more than two edges to each node in the first tree. This way, a D-vine is uniquely determined by its first tree and we call it an ordered D-vine if the nodes are ordered in an increasing way, i.e. 1-2-3-4-... and so on. As our data set consists of observations at successive time points, this serial dependence naturally leads to the use of ordered D-vines.

In the context of copulas, a (D- or R-) vine copula consists of the vine structure as described above where each edge is associated with a (conditional) bivariate copula. Figure 2.3 shows a graphical representation of a 4-dimensional D-vine structure where we can also see the division into trees (denoted by T_1, T_2 and T_3). Here, a 4-dimensional copula density is factorised into pair-copulas as labelled on the edges, e.g. 1, 2 refers to the copula connecting nodes 1 and 2 while 1, 4; 2, 3 refers to the copula connecting nodes 1 and 4 conditioned on 2 and 3.

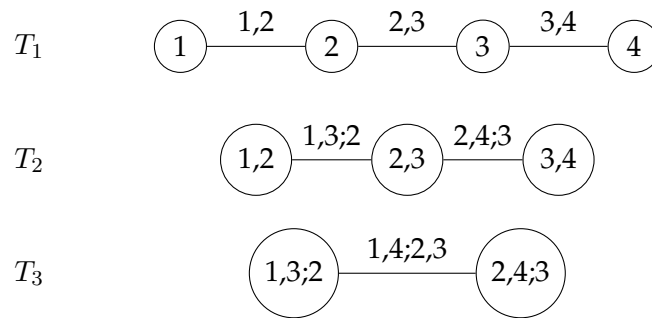


FIGURE 2.3: A 4-dimensional ordered D-vine structure. The connecting edges are labelled with the corresponding pair-copulas.

We will now introduce some notation (taken from Kraus and Czado (2017)) to help us simplify the following parts.

Given a subset $D \subset \{1, \dots, d\}$ and $i, j \in \{1, \dots, d\} \setminus D$ with $i < j$, we denote by $C_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$ the copula corresponding to the conditional

distribution of (X_i, X_j) given $\mathbf{X}_D = \mathbf{x}_D$, where $\mathbf{x}_D = \{x_u : u \in D\}$, and by $c_{ij;D}(\cdot, \cdot; \mathbf{x}_D)$ the associated copula density.

In the same way, $F_{i|D}(\cdot | \mathbf{x}_D)$ denotes the conditional distribution function of X_i given $\mathbf{X}_D = \mathbf{x}_D$.

And finally, we denote by $C_{i|D}(\cdot; \mathbf{u}_D)$ the conditional distribution of the PIT random variable U_i given $\mathbf{U}_D = \mathbf{u}_D$.

We call D the conditioning set and $\{i, j\}$ the conditioned set. Now recall the ascending order of nodes in an ordered D-vine and the fact that each edge is assigned a bivariate copula. Thus, in the case of ordered D-vines, for a given conditioned set $\{i, j\}$, the conditioning set is always given by $D = \{i + 1, \dots, j - 1\}$.

A common assumption when working with D-vine copulas (and R-vine copulas in general) is to assume that the copulas $c_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$ actually do not depend on the specific observations of the conditioning vector \mathbf{x}_D , i.e. $c_{i,j;D}(\cdot, \cdot; \mathbf{x}_D) \equiv c_{i,j;D}(\cdot, \cdot)$. This is the so-called simplifying assumption (discussed, for example, in Stoeber, Joe et al. (2013)) which we will use from now on.

In case that a copula is parametric and we want to refer to the parameters, they will be denoted by the parameter vector $\boldsymbol{\theta}_{i,j;D}$ and we write $c_{i,j;D}(\cdot, \cdot; \boldsymbol{\theta}_{i,j;D})$.

We now want to derive a pair-copula representation for multivariate densities. Thus, recall Sklar's theorem applied to densities from equation (2.1), stating that the cumulative density function $f(x_1, \dots, x_d)$ can be written as the product of the marginal densities and the d -dimensional copula density, i.e.

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d)$$

and that we can, for example, recursively decompose

$$f(x_1, \dots, x_d) = f_d(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, \dots, x_d). \quad (2.4)$$

Following Czado (2010), we can see in the 2-dimensional case that

$$f(x_1, x_2) = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2)$$

for the associated pair-copula density $c_{1,2}(\cdot, \cdot)$ and it follows for the conditional density

$$f(x_1|x_2) = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \quad (2.5)$$

In the 3-dimensional case, we have for example

$$f(x_1|x_2, x_3) = c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot f(x_1|x_2), \quad (2.6)$$

where $c_{1,3;2}(\cdot, \cdot)$ is the corresponding pair-copula density for the conditional distribution of (X_1, X_3) given X_2 . By inserting equation (2.5) into (2.6) we get

$$f(x_1|x_2, x_3) = c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1).$$

In this way, we can decompose (2.4) into the marginals and bivariate copula densities (Czado, 2010):

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f_k(x_k) \prod_{i=1}^{d-1} \prod_{j=i+1}^d c_{i,j;D} (F_{i|D}(x_i|x_D), F_{j|D}(x_j|x_D)) ,$$

with $D = \{i+1, \dots, j-1\}$.

This is called an ordered D-vine density with order $X_1 - X_2 - \dots - X_d$. If all margins are uniform (as is the case when PIT variables are used), we call this an ordered D-vine copula.

As shown in Joe (1997), it is possible to evaluate arguments of these copula densities (which are univariate conditional distribution functions) using only the pair-copulas from lower trees by recursively applying so-called h-functions:

Definition 7 (h-functions). For a pair-copula $C_{i,j;D}(\cdot, \cdot)$, a set $D \subset \{1, \dots, d\}$ and $i, j \in \{1, \dots, d\} \setminus D$ with $i < j$,

$$h_{i|j;D}(\tilde{u}_i|\tilde{u}_j) := \frac{\partial}{\partial \tilde{u}_j} C_{i,j;D}(\tilde{u}_i, \tilde{u}_j) \text{ and}$$

$$h_{j|i;D}(\tilde{u}_i|\tilde{u}_j) := \frac{\partial}{\partial \tilde{u}_i} C_{i,j;D}(\tilde{u}_i, \tilde{u}_j)$$

are the h-functions corresponding to the pair-copula $C_{i,j;D}(\cdot, \cdot)$.

Now, let $j \in D$ and $D_{-j} := D \setminus j$ to get (see Joe (1997))

$$F_{i|D}(x_i|x_D) = h_{i|j;D_{-j}} \left(F_{i|D_{-j}}(x_i|x_{D_{-j}}) \middle| F_{j|D_{-j}}(x_j|x_{D_{-j}}) \right) , \quad (2.7)$$

where the argument is again an h-function, illustrating the recursive character.

We will illustrate this with the 4-dimensional example from Figure 2.3 and start at the third tree (T_3) copula $c_{1,4;2,3}(\cdot, \cdot)$. Its arguments are:

$$F_{1|2,3}(x_1|x_2, x_3) = h_{1|3;2} \left(F_{1|2}(x_1|x_2) \middle| F_{3|2}(x_3|x_2) \right) ,$$

$$F_{4|2,3}(x_4|x_2, x_3) = h_{4|2;3} \left(F_{4|3}(x_4|x_3) \middle| F_{2|3}(x_2|x_3) \right) .$$

We can replace the new arguments again using h-functions:

$$F_{1|2}(x_1|x_2) = h_{1|2}(F_1(x_1)|F_2(x_2)) ,$$

$$F_{3|2}(x_3|x_2) = h_{3|2}(F_3(x_3)|F_2(x_2)) ,$$

$$F_{4|3}(x_4|x_3) = h_{4|3}(F_4(x_4)|F_3(x_3)) ,$$

$$F_{2|3}(x_2|x_3) = h_{2|3}(F_2(x_2)|F_3(x_3)) .$$

Thus, we can write the arguments of $c_{1,4;2,3}(\cdot, \cdot)$ as

$$F_{1|2,3}(x_1|x_2, x_3) = h_{1|3;2} \left(h_{1|2}(F_1(x_1)|F_2(x_2)) \middle| h_{3|2}(F_3(x_3)|F_2(x_2)) \right)$$

and

$$F_{4|2,3}(x_4|x_2, x_3) = h_{4|2;3} \left(h_{4|3}(F_4(x_4)|F_3(x_3)) \middle| h_{2|3}(F_2(x_2)|F_3(x_3)) \right) ,$$

which now only depend on the pair-copulas and their parameters from the lower trees T_1 and T_2 as well as the marginal distributions F_1, F_2, F_3, F_4 .

Next, we also want to introduce the concept of truncated vines and the special case of a Markov tree dependence structure:

Definition 8 (Truncated D-vine). *A d -dimensional ℓ -truncated D-vine with $\ell \in \{1, \dots, d-1\}$ is a D-vine where all pair-copulas after ℓ trees are replaced by the independence copula.*

When $\ell = 1$, the result is a Markov tree dependence structure, i.e. two variables not connected by an edge are conditionally independent given the variables of the conditioning set (Brechmann and Joe, 2014). Truncation is often used in order to reduce the complexity of a model where the number of parameters increases exponentially in larger dimensions.

Vine fitting

Assuming we have estimated the marginals in the first step, we can perform the second step of estimating the vine copula, which can generally be broken down into three tasks:

- Selection of the R-vine structure.
- Selection of bivariate copula families for each pair in the structure.
- Estimation of the corresponding parameter(s) for each bivariate copula pair.

The first task (selection of the structure) is most commonly done by a sequential method based on Kendall's τ (Dissmann, Brechmann et al. (2013) for details) but we will restrict ourselves to an ordered D-vine structure, as introduced before, because the specific pair structure in the first tree is particularly suited for the serial dependence in our data.

In the next steps, copula families are selected for each pair by exploiting the tree structure of the vine: We start by estimating all unconditional pair-copulas in the first tree. The estimation of pair-copulas is achieved by first fitting all considered copulas using maximum likelihood estimation and then computing an information criterion (AIC or BIC, chosen beforehand) to find the family with the minimal value. We then compute the variables needed in the second tree by applying the corresponding h-functions (from Definition 7) and estimate the conditional copulas in the second tree and so on.

This way, we can completely fit a vine copula model with only bivariate estimations, which is quite efficient. This so-called sequential approach is also implemented in the R package `VineCopula` which we will use in this thesis (Schepsmeier, Stoeber et al., 2017a).

Alternatively, it is possible to perform a joint maximum likelihood estimation for the whole vine copula. However, this method can be computationally heavy. In fact, the sequential estimation approach was originally introduced to find good starting values for the joint estimation and thus reduces the computation time. But as the sequential approach proved to already provide excellent estimation results, it is now widely used in practice (see Hobæk Haff et al. (2013) for asymptotic behaviour of these sequential estimates).

Vine diagnostics

We now want to introduce several methods to compare different ordered D-vine models to each other. As we will later fit ordered D-vines with different families that are used for forecasting, we want to be able to compare these models with respect to how well the fitted dependence structure matches our data. As a basis, we first introduce the log-likelihood function for a D-vine copula:

Proposition 2 (Ordered D-vine copula log-likelihood). *Consider an ordered d -dimensional D-vine copula with a set of marginal parameters α as well as a set of copula parameters θ and let N be the number of observed individuals, i.e. the complete data set is given by $\mathbf{x}' = (\mathbf{x}'_1, \dots, \mathbf{x}'_N)$ with $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,d})' \in [0, 1]^d$ for $i = 1, \dots, N$. Then, the log-likelihood of the ordered D-vine copula is given by*

$$L(\alpha, \theta | \mathbf{x}) = \sum_{i=1}^N \sum_{j=1}^d \log [f_j(x_{i,j} | \alpha_j)] \\ + \sum_{i=1}^N \sum_{j=1}^{d-1} \sum_{k=j+1}^d \log [c_{j,k;D} (F_{j|D}(x_{i,j} | x_{i,j+1}, \dots, x_{i,k-1}), \\ F_{k|D}(x_{i,k} | x_{i,j+1}, \dots, x_{i,k-1}); \theta_{j,k;D})],$$

where $\theta_{j,k;D}$ is the vector of parameters for the copula $c_{j,k;D}(\cdot, \cdot; \theta_{j,k;D})$.

However, we will later see that our data consist of individuals with different dimension, i.e. the lengths of the \mathbf{x}_i are not identical. This is often called an unbalanced setting. In this case, a single vine is not enough as individuals with less than d dimensions are missing nodes in the d -dimensional vine. Thus, there are actually multiple vines of different dimensions, corresponding to the dimensions of the observed individuals. See Table 2.2 for an example where individuals have different dimensions.

	Time points				
	1	2	3	4	
x_1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	4-dimensional D-vine
x_2		$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	
x_3		$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	3-dimensional D-vine
x_4			$x_{4,3}$	$x_{4,4}$	2-dimensional D-vine
x_5				$x_{5,4}$	

TABLE 2.2: Example of individual observations with different dimensions. Only the first individual x_1 has enough observations to be used for the 4-dimensional vine copula (blue). Observations from individuals x_1, x_2 and x_3 starting at time point 2 are used for the 3-dimensional vine copula (red). All individuals except x_5 are used from time point 3 onwards for the 2-dimensional vine copula (green).

The log-likelihood function can thus be given in a more general setting:

Proposition 3 (Ordered D-vine copulas log-likelihood with different dimensions). *Let N be the number of observed individuals $\mathbf{x}' = (\mathbf{x}'_1, \dots, \mathbf{x}'_N)$ with*

different dimensions, where the maximum dimension is d . Each \mathbf{x}_i of dimension $d_i \in \{2, \dots, d\}$ is given by $\mathbf{x}_i = (x_{i,1+d-d_i}, \dots, x_{i,d})' \in [0, 1]^{d_i}$ for $i = 1, \dots, N$. Consider a set of D-vine copulas with maximum dimension d , a set of marginal parameters α and a set of copula parameters θ . Then, the log-likelihood of the D-vine copulas is given by

$$L(\alpha, \theta | \mathbf{x}) = \sum_{i=1}^N \sum_{j=1+d-d_i}^d \log [f_j(x_{i,j} | \alpha_j)] \\ + \sum_{i=1}^N \sum_{j=1+d-d_i}^{d-1} \sum_{k=j+1}^d \log \left[c_{j,k;D}^{d_i} \left(F_{j|D}(x_{i,j} | x_{i,j+1}, \dots, x_{i,k-1}), \right. \right. \\ \left. \left. F_{k|D}(x_{i,k} | x_{i,j+1}, \dots, x_{i,k-1}); \theta_{j,k;D}^{d_i} \right) \right],$$

where $\theta_{j,k;D}^{d_i}$ is the vector of parameters for the copula $c_{j,k;D}^{d_i}(\cdot, \cdot; \theta_{j,k;D}^{d_i})$ which has dimension d_i .

Note that, as mentioned before, we will apply the method of inference for marginals (IFM) in order to estimate parameters. This means that the estimation of marginal parameters α and copula parameters θ will not be done simultaneously. Thus, from now on, we will denote by $L(\theta | \mathbf{x})$ the log likelihood function of D-vine copulas with fixed marginal parameters.

Also note that there are at most $d - 1$ different vines to consider since the individuals' dimensions d_i range from 2 to d . For the rest of this thesis, we will assume that for a given pair $\{j, k\}$, we have

$$c_{j,k;D}^{d_i}(\cdot, \cdot; \theta_{j,k;D}^{d_i}) = c_{j,k;D}(\cdot, \cdot; \theta_{j,k;D})$$

for all $d_i \in \{2, \dots, d\}$, i.e. we assume that the copula families and their parameters are identical for the same pairs, regardless of the underlying observed individuals' dimensions. This means the dependence structure is independent of the number of available observations. See Figure 2.4 for an illustration in the 3-dimensional case.

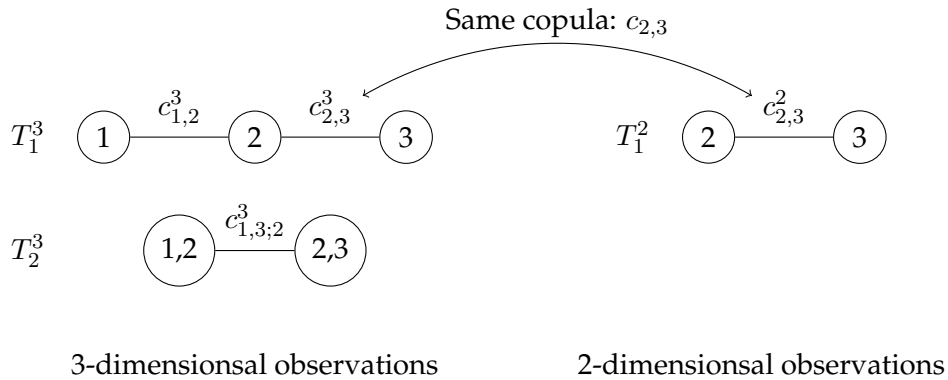


FIGURE 2.4: Example of ordered D-vine copula for 3-dimensional and 2-dimensional observations. The copula for the pair $\{2, 3\}$ is the same regardless of the dimension of the vine.

With this, we can also define the well-known Akaike and Bayesian Information criteria (AIC and BIC) for D-vine copulas which allow us to compare different vine estimations (Akaike, 1998 and Schwarz et al., 1978):

Definition 9 (AIC and BIC). *Let m be the number of parameters and θ the parameter set of a D-vine copula model with observations $\mathbf{x}' = (\mathbf{x}'_1, \dots, \mathbf{x}'_N)$ and log-likelihood function $L(\theta|\mathbf{x})$. Then the Akaike Information Criterion (AIC) is defined as*

$$AIC := -2L(\theta|\mathbf{x}) + 2m$$

and we call $2m$ the penalty term.

Similarly, the Bayesian Information Criterion (BIC) is defined as

$$BIC := -2L(\theta|\mathbf{x}) + \log(N)m,$$

where $\log(N)m$ is the penalty term.

The penalty term of the BIC depends on the number of observed individuals N . However, in our case, the number of individuals changes over time (recall Table 2.2) and it is not clear what that number should be for the BIC. Thus, we also work with an alternative of the BIC proposed by Killisches and Czado (2017) where each parameter is weighted according to the number of observed individuals that contribute to its estimation:

Definition 10 (BIC for D-vines). *For a d -dimensional D-vine model, let m_i be the number of parameters of the D-vine model restricted to the observations from i to d for $i = 1, \dots, d$ and define $\Delta m_i := m_i - m_{i+1}$ for $i = 1, \dots, d-1$ and $\Delta m_d := m_d$. Further denote by N_i the number of individuals observed at time i . Then the BIC of the D-vine model is given by*

$$BIC_D := -2L(\theta|\mathbf{x}) + \sum_{i=1}^d \Delta m_i \log(N_i)$$

Additionally to the AIC and BIC, we want to note the so-called Vuong test (Vuong, 1989). It can be used to directly compare two non-nested models and tells us the preferred model at a given significance level:

Definition 11 (Vuong test). *Let c_1 and c_2 be two vine copulas in terms of their densities with estimated parameter sets $\hat{\theta}_1$ and $\hat{\theta}_2$. We then compute the test statistic ν as the standardized sum of the log differences of their pointwise likelihoods $m_i := \log \frac{c_1(u_i; \hat{\theta}_1)}{c_2(u_i; \hat{\theta}_2)}$ for observations $u_i \in [0, 1]$ with $i = 1, \dots, N$, i.e.*

$$statistic := \nu = \frac{\frac{1}{N} \sum_{i=1}^N m_i}{\sqrt{\sum_{i=1}^N (m_i - \bar{m})^2}}.$$

As ν is asymptotically standard normal, according to the null hypothesis

$$H_0 : \mathbb{E}[m_i] = 0, \quad \text{for all } i = 1, \dots, N,$$

the vine model with density c_1 is preferred at level α if

$$\nu > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right),$$

while the other model is preferred if $\nu < -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

The statistic can also be corrected for the number of parameters used in the models, either by the Akaike or the Schwarz correction, which correspond to the penalty terms in the AIC and BIC.

With these, we have a selection of tools to help us compare different vine models.

2.3 Marginal modelling

Recall that Sklar's theorem allows us to use a 2-step process of handling the marginal distributions and the copulas separately. We now want to introduce the tools needed for the first step, i.e. fitting the marginals.

Kernel density estimation

As described before, in order to gain the approximately uniformly distributed data for the copulas, we can use the probability integral transformation (PIT) if we have the cumulative distribution functions (or a sufficiently good fit) of each marginal. Estimation of the distribution functions can be done either parametrically or non-parametrically. We choose a kernel density estimation (KDE) to non-parametrically obtain a continuous cdf, which is described in detail in Nagler (2017a). This so-called continuous convolution KDE allows us to gain a continuous cdf from discrete samples by adding random noise to the observations.

Given N observations $x_i, i = 1, \dots, N$ with unknown density f , and a random vector $\mathbf{E} \in \mathbb{R}^n$ with i.i.d components E_i (see Nagler (2017a) for classes of appropriate distributions), the continuous density estimator of f is given by

$$\hat{f}(x) = \frac{1}{Nb_N} \sum_{i=1}^N K\left(\frac{(x_i + E_i) - x}{b_N}\right),$$

where $b_n > 0$ and K a symmetric density function.

Once we have estimated corresponding cdfs for each marginal, we only need to apply it to the data to obtain approximately uniformly distributed u-data, similarly to equation (2.2) in the beginning (where we considered known marginals instead of estimated ones).

Regression

Instead of applying a KDE to the data to obtain an estimated cdf, we can also first apply a regression with explanatory variables. We then work with the resulting residuals in order to obtain data that have a possible effect of explanatory variables removed while still inheriting the dependence structure.

Our data have a few important characteristics, which we will describe in more detail in later chapters. These characteristics include large outliers, a left truncation as well as changing behaviours of mean and standard deviation over time that also depend on explanatory variables. Thus, we tried different regression methods for our data, from a simple linear regression over more advanced models such as Generalized Linear Models (GLM) to

the complex Generalized Additive Models for Location, Scale and Shape (GAMLSS).

Neither the simple regressions nor the GLMs were able to capture the behaviour of our data sufficiently. Hence, we decided to work within the GAMLSS framework, which can be seen as an extension to Generalized Linear Models and which provides a good fit in our case. While GLMs are already quite flexible compared to ordinary linear regression as they allow for response variables that have error distribution models other than a normal distribution (specifically, exponential family distributions), GAMLSS allow for an even wider range of distributions, including highly skew and/or kurtotic continuous and discrete distributions. The following definition of GAMLSS is taken from Stasinopoulos, Rigby et al. (2006):

Definition 12 (GAMLSS). A GAMLSS assumes independent observations x_i for $i = 1, 2, \dots, N$ with probability density function $f(x_i | \Psi^i)$ conditional on Ψ^i where $\Psi^i = (\Psi_1^i, \Psi_2^i, \dots, \Psi_p^i)$ is a vector of p parameters, each of which is related to the explanatory variables. Let $p = 4$ and denote the parameters as $(\mu_i, \sigma_i, \nu_i, \tau_i)$, corresponding to location, scale and two shape parameters.

Let $\mathbf{x}' = (x_1, x_2, \dots, x_N)$ be the N length vector of the response variable. Also for $k = 1, 2, 3, 4$, let $g_k(\cdot)$ be known monotonic link functions relating the k -th parameter $\Psi_k = (\Psi_k^1, \dots, \Psi_k^N)'$ to explanatory variables by semi-parametric additive models given by

$$g_k(\Psi_k) = \eta_k = \mathbf{Z}_k \beta_k + \sum_{j=1}^{J_k} h_{j,k}(\mathbf{z}_{j,k}), \quad (2.8)$$

where Ψ_k, η_k and $\mathbf{z}_{j,k}$ for $j = 1, 2, \dots, J_k$ and $k = 1, 2, 3, 4$ are vectors of length N .

The function $h_{j,k}$ is a non-parametric additive function of the explanatory variable $\mathbf{Z}_{j,k}$ evaluated at $\mathbf{z}_{j,k}$. The explanatory vectors $\mathbf{z}_{j,k}$ are assumed fixed and known. Also \mathbf{Z}_k , for $k = 1, 2, 3, 4$ are fixed design matrices while β_k are the parameter vectors.

One distribution family that the GAMLSS framework allows us to work with is the Generalized Gamma (GG) distribution which we will focus on for our model fittings. The Generalized Gamma distribution has the following density:

$$f(x | \mu, \sigma, \nu) = \frac{\delta^\delta z^\delta |\nu| e^{-\delta z}}{\Gamma(\delta) x},$$

where $z = (x/\mu)^\nu$, $\delta = 1/(\sigma^2 |\nu|^2)$ for $x > 0$, $\mu > 0$, $\sigma > 0$ and $-\infty < \nu < \infty$. Thus, we will have to estimate the three parameters μ, σ and ν for a GG regression. Choosing $\nu = 1$, this distribution becomes the Gamma distribution. When choosing $\nu = 0$, the distribution becomes the log-normal distribution. See Figure 2.5 for some examples of the Generalized Gamma density with different parameters (chosen similar to those appearing in our data).

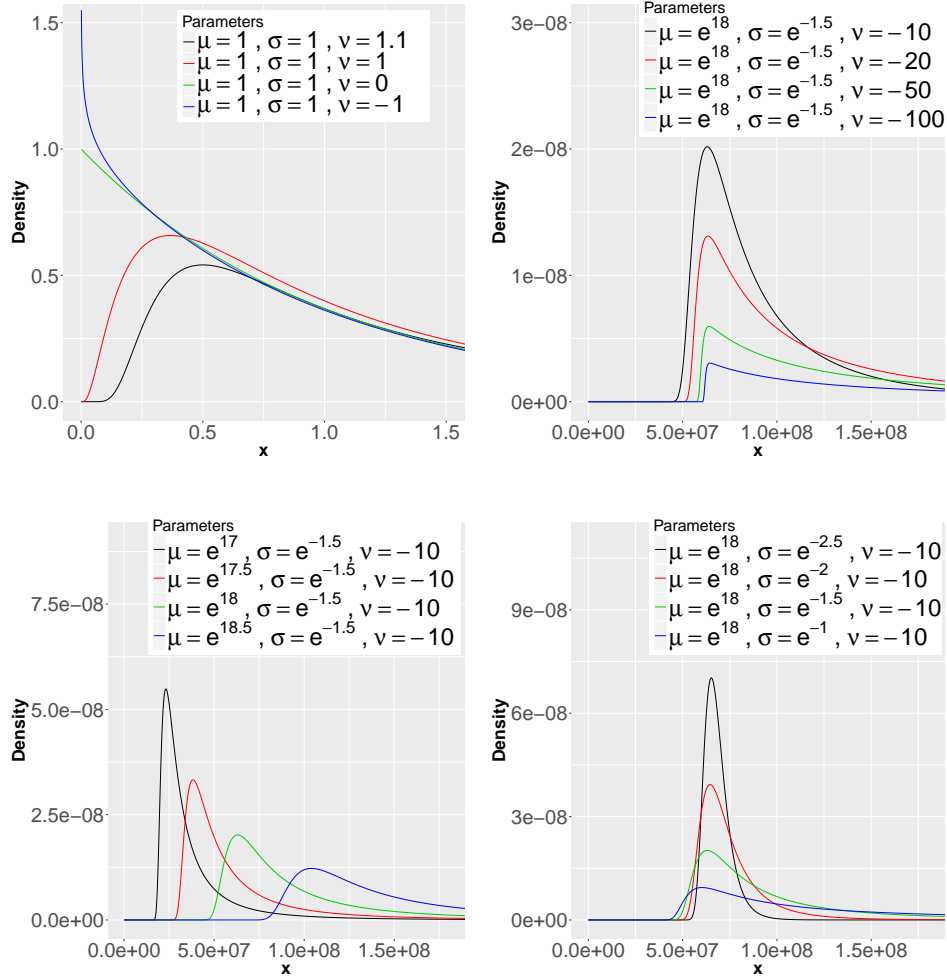


FIGURE 2.5: Example densities of Generalized Gamma distributions for different parametrizations.

Residual diagnostics

When working with a regression as described above, we need to make sure that it fits the data well enough. As we are interested in the residuals, we have to confirm that they behave as imposed by the underlying model to make sure the model is appropriate. Thus, we first need to define the residuals used in the GAMLSS case:

Definition 13 (Normalized quantile residuals). *For each time $t \in \{1, \dots, d\}$ and $i = 1, \dots, N_t$, let $x_{i,t}$ be the observed values and $\hat{F}(\cdot | \hat{\Psi}^{i,t})$ the corresponding fitted cumulative distribution function from a GAMLSS regression with fitted parameters $\hat{\Psi}^{i,t}$. Denote by Φ^{-1} the inverse cumulative distribution function of a standard normal random variable. Then*

$$r_{i,t} := \Phi^{-1} \left(\hat{F} \left(x_{i,t} | \hat{\Psi}^{i,t} \right) \right)$$

are the corresponding normalized quantile residuals.

These residuals should follow a normal distribution if the model is a good fit. See also Dunn and Smyth (1996) for details on normalized quantile residuals.

Thus, to get approximately uniformly distributed data, we can simply apply the cdf of a standard normal distribution (denoted by Φ) to these residuals, i.e.

$$\hat{u}_{i,t} := \Phi(r_{i,t}) .$$

Now that we have defined the residuals, we want a way to check them for goodness of fit. One way to do so is by looking at the so-called Quantile-Quantile plots (Q-Q plots) where the empirical quantiles of residuals are plotted against the theoretical quantiles of a standard normal distribution. If the values lie along a line, the distribution of residuals has the same shape as the standard normal distribution (up to location and scale).

Figure 2.6 shows examples of Q-Q plots with a line that passes through the first and third quartiles to identify deviations from a straight line. We can see the (average) behaviour in the case of heavy and light tailed distributions as well as right skewed ones (a left skewed distribution would show quantiles mirrored to the ones of the right skewed example). Note that when looking at 'real' data, patterns may be harder to make out due to the randomness in the data (especially with small sample sizes).

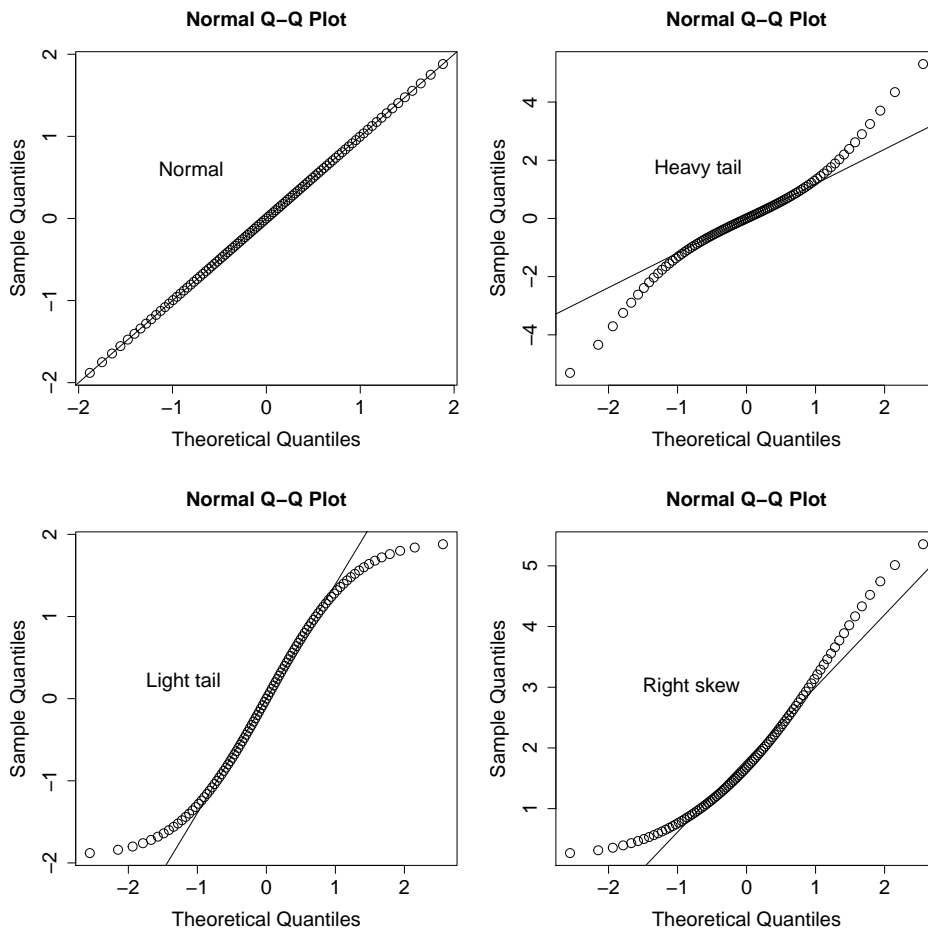


FIGURE 2.6: Examples of Q-Q plots for different sample distributions showing different behaviours. Top left: Normal distribution. Top right: t-distribution. Bottom left: Uniform distribution. Bottom right: Gamma distribution.

U-data diagnostics

As mentioned above, the obtained u-data are always just approximately uniformly distributed. In order to check whether the data are consistently estimated to be used in the second estimation step, we can perform the so-called Kolmogorov-Smirnov (KS) test. This test compares the empirical and a given theoretical cumulative distribution function and uses the maximum difference as a test statistic.

Denote by $\hat{\mathbb{P}}$ the empirical distribution and by \mathbb{P} the given theoretical distribution. Then the null hypothesis H_0 and alternative H_1 are given by

$$\begin{aligned} H_0 : \hat{\mathbb{P}} &= \mathbb{P}_0 , \\ H_1 : \hat{\mathbb{P}} &\neq \mathbb{P}_0 . \end{aligned}$$

For a given set of n observations x_1, \dots, x_n , let \hat{F} be their empirical distribution function and F be the theoretical uniform distribution function. Then the KS-test statistic D is given by

$$D = \sup_{x \in [0,1]} \left| \hat{F}(x) - F(x) \right| .$$

Now let $K_{1-\alpha}$ be the $1 - \alpha$ quantile of the Kolmogorov distribution (Wang, Tsang et al., 2003), then the null hypothesis H_0 is rejected at level α if

$$\sqrt{n}D > K_{1-\alpha} .$$

As the test assumes that the data are drawn from a uniform distribution, a rejection of H_0 (i.e. a low p-value) implies non-uniform distribution.

2.4 Forecasting with D-vine copulas

The following sections will first propose a method of shifting copulas in order to be able to forecast data from a fitted D-vine copula and then introduce the continuous rank probability score which is used for assessing the quality of forecasts.

Shifting copulas for forecasting

In order to forecast (i.e. simulate) data from a fitted ordered D-vine copula, we only need to know the marginal distributions and the dependence structure. However, as we only have fitted marginals and the fitted dependence structure until the current time point, we need to decide how to proceed to get these information for the next time point in order to obtain forecasts.

We first deal with the problem of the unknown dependence structure and propose a concept of shifting the already fitted copula structure to include the new point.

Thus, let \mathcal{C} be the d -dimensional fitted ordered D-vine copula with bivariate pair-copulas $c_{i,j;D}(\cdot, \cdot; \theta_{i,j;D})$ for $i, j \in \{1, \dots, d\}$ with $i < j$ and order $1 - 2 - \dots - d$. Then the new d -dimensional ordered D-vine copula \mathcal{C}^P used for predictions consists of the nodes $\{2, \dots, d+1\}$ with order $2 - 3 - \dots - d+1$ and the corresponding bivariate pair-copulas $c_{k,l;k+1,\dots,l-1}^P(\cdot, \cdot; \theta_{k,l;k+1,\dots,l-1}^P)$

are defined as

$$c_{k,l;k+1,\dots,l-1}^P(\cdot, \cdot; \theta_{k,l;k+1,\dots,l-1}^P) := c_{(k-1)(l-1);k,\dots,l-2}(\cdot, \cdot; \theta_{k-1,l-1;k,\dots,l-2}) ,$$

for $k, l \in \{2, \dots, d+1\}$ with $k < l$. Figure 2.7 illustrates the procedure in the 4-dimensional case.

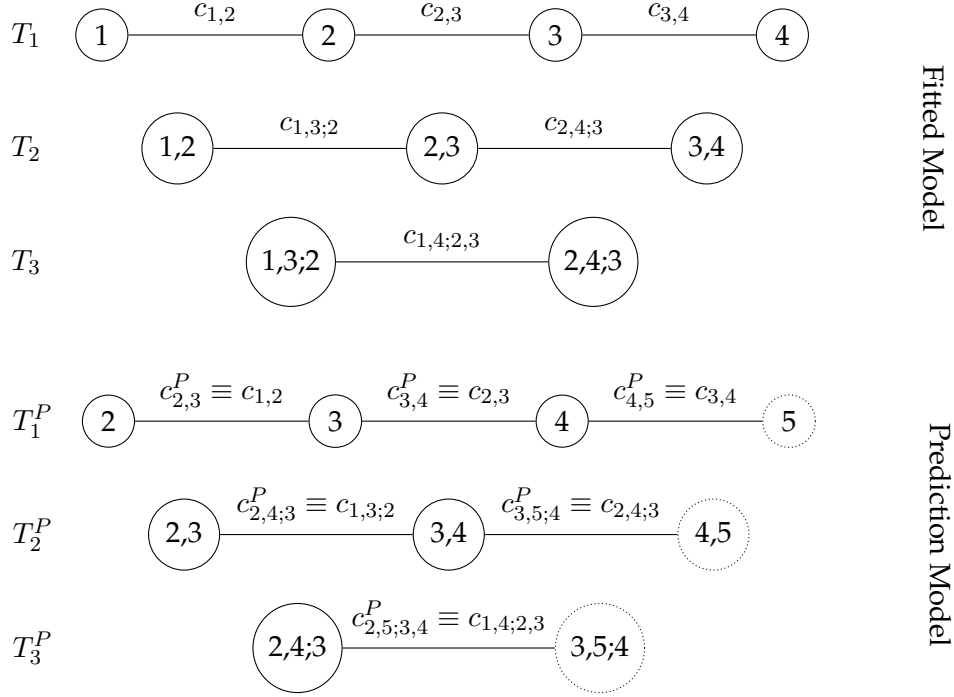


FIGURE 2.7: Example of the shifting of copulas from the fitted (top) to the prediction (bottom) vine in the case of a 4-dimensional D-vine where we add the fifth node. The connecting edges are labelled with the corresponding pair-copulas. Note that the pair-copulas in the bottom vine are identical to the ones from the top vine for different pairs.

Given this new D-vine copula and observed \mathbf{u} -data $(\mathbf{u}_2, \dots, \mathbf{u}_d)$, we can forecast the data points \mathbf{u}_{d+1} . Recall however, that in our case the \mathbf{u} -data have different dimensions. Thus, the forecasted data \mathbf{u}_{d+1} must also have the lowest dimension of the \mathbf{u} -data that is used for the forecasting. Note that this approach of course assumes that the dependence structure does not vary too much over time in order to be appropriate, which we will later see to be the case for our data set.

The next question is what marginal distribution to use to transform the forecasted \mathbf{u} -data back into the original scale. If the marginal distributions were estimated non-parametrically, we propose to also shift the marginal distribution from the last time point, i.e. $F_{d+1}(\cdot) := F_d(\cdot)$. Again, this assumes that also the marginal distribution does not vary too much over time, which will be appropriate for our data.

In case the marginal distribution was estimated parametrically, we can also try to predict the parameters of the next time point, e.g. by regression.

Continuous rank probability score

Once we have forecasted data, we need a method to assess the quality of our forecasts. For this, we can use the so-called continuous rank probability score (CRPS) that assigns a numerical score based on the forecasted value and the actually realized value (see Matheson and Winkler (1976)):

Definition 14 (Continuous rank probability score (CRPS)). *Let \mathcal{P} consist of the Borel probability measures on \mathbb{R} and identify a probabilistic forecast, i.e. a member of the class \mathcal{P} , with its cumulative distribution function F . Denote by y an actually observed value. Then the continuous rank probability score is defined as*

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(t) - \mathbb{1}_{\{y \leq t\}})^2 dt ,$$

where $\mathbb{1}_{\{y \leq t\}}$ represents the indicator function of the event $\{y \leq t\}$.

An important characteristic of the CRPS is that it not only punishes forecasts with expected values farther away from the observations but also considers their variation. Figure 2.8 illustrates this behaviour for two example forecasts where both have the same mean but Forecast 2 (red) has a larger standard deviation. The second image shows the difference between the cumulative distribution functions of the forecasts and the actual observation 2 (which is simply an indicator function for the set $\{2 \leq t\}$). Due to the higher standard deviation, the absolute difference in the second image is larger for Forecast 2 and thus the squared values are also larger or equal than those of Forecast 1 in the third image. As the CRPS takes the integral over these squared differences, Forecast 2 receives a larger CRPS value than Forecast 1.

The CRPS can also be written as (see e.g. Gneiting and Raftery (2007)):

$$\text{CRPS}(F, y) = \mathbb{E}[|X - y|] - \frac{1}{2} \mathbb{E}[|X - \tilde{X}|] ,$$

where $\mathbb{E}[\cdot]$ represents the expected value and X, \tilde{X} are independent copies of a random variable with distribution function F and finite first moment.

A feature of the CRPS is that it penalizes predictions less when their probabilities are close to the actual value, and more when their probabilities are farther away from it. A lower score is preferred to a higher one.

In practice, we can use an approximation for the CRPS. For $m \in \mathbb{N}^+$ let

$$q_X\left(\frac{i}{m}\right) := F^{-1}\left(\frac{i}{m}\right) , \quad i = 1, \dots, m-1 ,$$

denote the $\frac{i}{m}$ -quantile of the random variable X with cdf F (we leave out the 0% and 100% quantiles for numerical reasons). Then the CRPS can be approximated by

$$\begin{aligned} \text{CRPS}(F, y) &\approx \\ &\frac{1}{m-1} \sum_{i=1}^{m-1} \left| q_X\left(\frac{i}{m}\right) - y \right| - \frac{1}{2(m-1)^2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \left| q_X\left(\frac{i}{m}\right) - q_X\left(\frac{j}{m}\right) \right| . \end{aligned}$$

We can apply this scoring method to our copula framework by using the conditional quantile function as described in Kraus and Czado (2017) (also

recall the notations from section 2.2 for D-vine copulas): Assume we have $d + 1$ variables X_1, \dots, X_{d+1} , $d \geq 1$ with $X_j \sim F_j$ and $j = 1, \dots, d + 1$. Then the conditional quantile function of X_{d+1} given $(X_1, \dots, X_d) = (x_1, \dots, x_d)$ for $\alpha \in (0, 1)$ is given by

$$q_{d+1|1,\dots,d}(\alpha|x_1, \dots, x_d) := F_{d+1|1,\dots,d}^{-1}(\alpha|x_1, \dots, x_d) .$$

Applying the probability integral transformation, we get $U_j = F_j(X_j)$ with corresponding values $u_j = F_j(x_j)$. Then

$$F_{d+1|1,\dots,d}(y|x_1, \dots, x_d) = C_{d+1|1,\dots,d}(u_{d+1}|u_1, \dots, u_d)$$

and thus

$$F_{d+1|1,\dots,d}^{-1}(\alpha|x_1, \dots, x_d) = F_{d+1}^{-1} \left(C_{d+1|1,\dots,d}^{-1}(\alpha|u_1, \dots, u_d) \right) ,$$

which gives us a representation of the conditional quantile function in terms of the marginal distribution of X_{d+1} and the conditional copula quantile function $C_{d+1|1,\dots,d}^{-1}$ conditioned on the PIT values of x_1, \dots, x_d .

An estimation of the conditional quantile function can be obtained by first estimating the marginals F_j for $j = 1, \dots, d + 1$ as well as the copula $C_{1,\dots,d+1}$ and then setting

$$\hat{q}_{d+1|1,\dots,d}(\alpha|x_1, \dots, x_d) := \hat{F}_{d+1}^{-1} \left(\hat{C}_{d+1|1,\dots,d}^{-1}(\alpha|\hat{u}_1, \dots, \hat{u}_d) \right) ,$$

where $\hat{u}_j := \hat{F}_j(x_j)$ for $j = 1, \dots, d$.

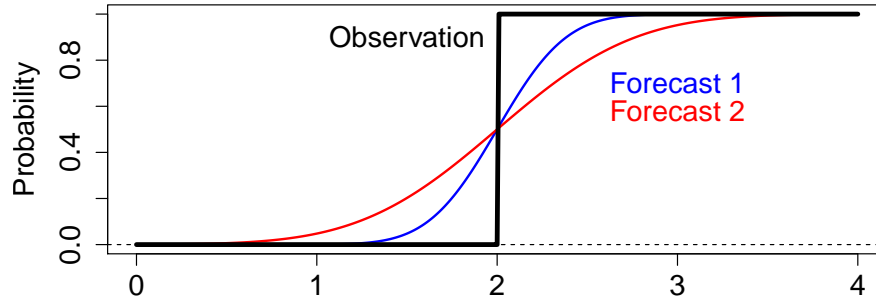
Recall from before that we can express the cdf $C_{d+1|1,\dots,d}(u_{d+1}|u_1, \dots, u_d)$ and its inverse with h-functions and inverse h-functions. Thus, we can apply the approximated CRPS from before to a copula model by applying the conditional quantile function.

Assuming we have a Markov tree dependence structure (as we will later see to be the case for our data set), the quantile function only needs to be conditioned on the last time point d , thus we obtain the CRPS of each individual observation $x_{i,d}$ as

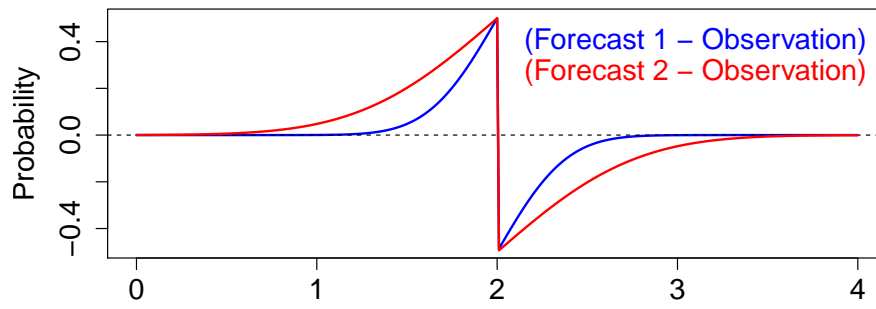
$$\begin{aligned} \text{CRPS}_i & (F_{d+1|d}(x_{i,d+1}|x_{i,d}), x_{i,d+1}) \\ &= \frac{1}{m-1} \sum_{i=1}^{m-1} \left| F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{i}{m} \middle| u_d \right) \right) - x_{i,d+1} \right| \\ & - \frac{1}{2(m-1)^2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \left| F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{i}{m} \middle| u_d \right) \right) - F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{j}{m} \middle| u_d \right) \right) \right| . \end{aligned} \tag{2.9}$$

Since we have multiple observations at the last time point d , we assign the forecasting method a score by averaging the scoring rule over all observations at time d . Thus, for a given forecasting method, i.e. a given set of margins and copulas, the overall score of the method is given by

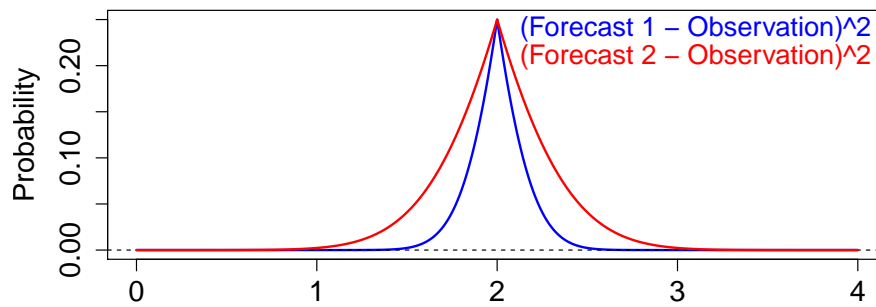
$$\text{CRPS} = \frac{1}{N_d} \sum_{i=1}^{N_d} \text{CRPS}_i (F_{d+1|d}(x_{i,d+1}|x_{i,d}), x_{i,d+1}) .$$



(A) The cumulative distribution functions for the deterministic observation (black) and two forecasts (blue and red) with the same mean 2 but different standard deviations.



(B) The difference between the forecasts' and the observation's cdf, i.e. $F(t) - \mathbb{1}_{\{2 \leq t\}}$ for $t \in \{0, \dots, 4\}$.



(C) The square of the differences, i.e. $(F(t) - \mathbb{1}_{\{2 \leq t\}})^2$ for $t \in \{0, \dots, 4\}$.

FIGURE 2.8: Illustration of the different CRPS values resulting from two forecasts with the same mean but different standard deviations. The higher standard deviation (red forecast) is punished more.

Chapter 3

Exploratory data analysis

In this chapter, we want to introduce the data set at hand and analyse its characteristics. We later want to model the marginal distributions from this data for our 2-step approach.

Our data consist of so called exposure limits. When a primary insurance enters a reinsurance contract, they will provide information about their own portfolio of insured risks. One of those information is called the exposure limit: The primary insurance sets these individual limits for each risk defining how much exposure that risk can contribute at most, i.e. when a company is insured at the primary insurance, this is the maximum amount of damage covered by the insurance. However, at any given time, the company may have less exposure than the limit (e.g. a company producing less goods for some reason) or more than the limit (in this case, the additional exposure is not covered). In fact, for large companies, the exact amount changes constantly. We only know the exposure limits (which are fixed for a given period) and have no knowledge of the underlying 'real' exposure.

It is important to quantify these exposure limits in order to model future losses. However, since the information on exposure limits is only known up to and including the present time point and we want to predict future events, we want a model to predict the evolution of these limits.

A note on truncation

In practice, the primary insurance will only deliver exposure limits that are above a certain threshold (identical for all companies) in order to reduce the amount of data to only relevant information. Thus, a data point for a company only shows actual exposure if its amount is above this threshold. In reality, the primary insurance may change the threshold of reporting as time progresses. This means a data point can show zero exposure for a company not only when there is no exposure but also when the threshold is set above the company's exposure (note that we do not know what the actual reason is). The zero values and the fact that we only have data available larger than the threshold is called left-truncation and it causes a bias in the estimation of the data's density. While there are ways to deal with this bias, we do not use them for our analysis as the effects can be neglected in our case. However, we will still notice this behaviour in many of the following plots.

3.1 Data patterns

We first want to give a short overview of the data set. We have exposure limits for 22 time points beginning in 2011 Q1 and ending in 2016 Q2 (one data point per quarter) from a total of 2073 companies. Considering the above mentioned truncation and the fact that not all companies have continuous contracts since 2011, there are a total of 10,156 non-zero data points. For each company, we also know the country where it is based. Additionally, we have data points from 2016 Q3 available for each observation from 2016 Q2. We do not use this information for modelling purposes but will later come back to it when assessing the quality of forecasting.

When looking at the exposure limit development over time, it is useful to look at the data in the form of patterns to get a better overview. A pattern describes at which points in time a company has active exposures.

We observe 486 unique patterns. See Figure 3.1 for a visualization of those patterns.

Since we want to predict future exposure limits, we will only concentrate on those patterns that have an active exposure at the last (22nd) time point (i.e. 2016 Q2) which correspond to the first 168 patterns in Figure 3.1.

As we can see, many patterns have several discontinuities and only sparse data available, especially in the early quarters. The reason for this in most cases is either that the company was not insured (i.e. an exposure of zero) or, as mentioned before, that the threshold of reporting was set above the exposure.

One can deal with these gaps in several ways:

- Shifting the data to remove gaps. For example, if quarters 14 and 15 were unavailable, we could move the data of quarters 1-13 up 2 quarters, i.e. they become data of quarters 3-15 and the first 2 quarters are unavailable. However, this would distort the actual information of the development of the limits.
- Estimating the missing data; data imputation.
- Disconnecting the continuous sets. Again for example, quarters 14 and 15 are missing. We could handle the two disconnected sets (1-13 and 16-22) as two separate patterns, one that has only the first 13 points available and one that has no data for the first 15 points.

For now, we will use the last method and exclude all data points in our first block that are not continuously connected to the last quarter. This results in 4264 data points remaining, or about 42% of the complete data set. The corresponding unique patterns are displayed in Figure 3.2 for the reduced first block. The number of occurrences displayed show that longer patterns (i.e. those that start early) are present only a few times. An exception is the third pattern, starting at 2012 Q1, which appears more often. Most patterns however start in 2014 or later.

When we look at the data column-wise instead, we see how many observations per time point are available, regardless of the pattern they belong to. This is summarized in Table 3.1. As we can see, some early time points have only few observations.

Recall that additionally to the amount of exposure limit, we also know the country where each company is based in. We observe a total of 85 different countries in the selected block of data. We can group those countries into 5 regions: Africa, Americas, Asia, Europe and Oceania. See Table 3.2 for an overview of how many observations per region are available and the earliest time point of data for each region.

We decide to use only data from 2014 Q2 onwards for the next steps, as this set of data contains observations from all regions at all time points. Table 3.3 lists the number of observations for each region and each time step from 2014 Q2 onwards.

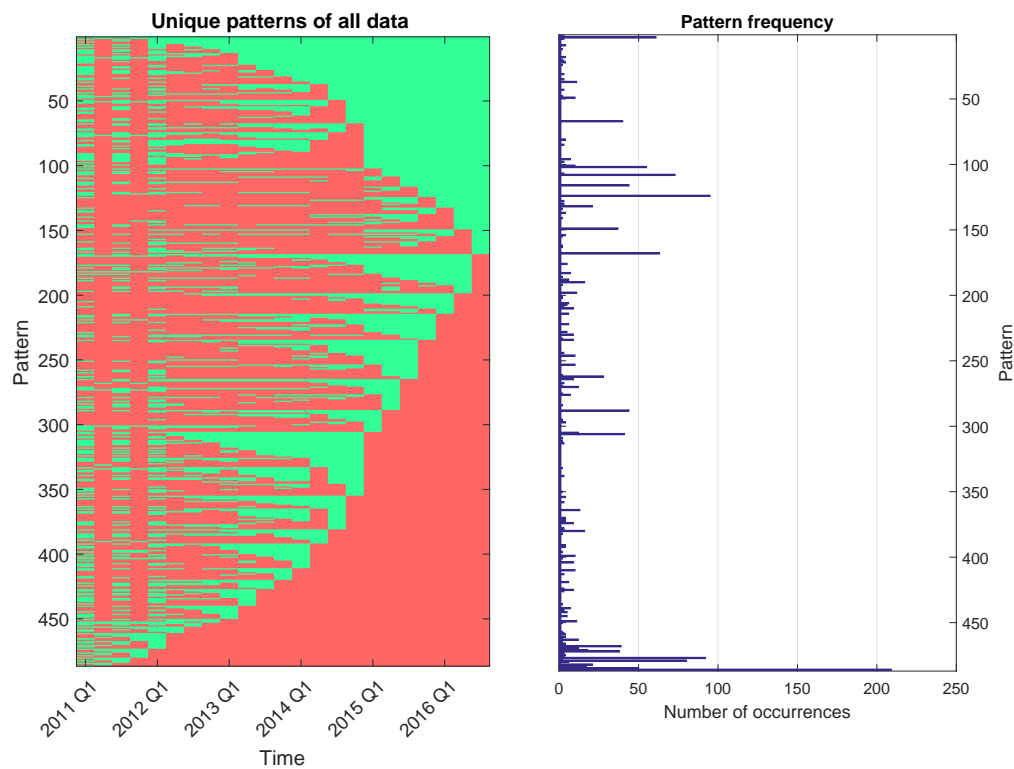


FIGURE 3.1: Unique pattern occurrences for our data. The right figure shows the frequency of each pattern.

Time point	2011 Q1	2011 Q2	2011 Q3	2011 Q4	2012 Q1	2012 Q2
Observations	3	3	3	4	70	70

Time point	2012 Q3	2012 Q4	2013 Q1	2013 Q2	2013 Q3	2013 Q4
Observations	72	77	81	99	104	111

Time point	2014 Q1	2014 Q2	2014 Q3	2014 Q4	2015 Q1	2015 Q2
Observations	117	129	153	210	321	401

Time point	2015 Q3	2015 Q4	2016 Q1	2016 Q2
Observations	452	554	586	644

TABLE 3.1: Observations per time point in the investigated block as displayed in Figure 3.2.

Region	Asia	Americas	Europe	Africa	Oceania
Total observations	1843	1302	1046	128	83
Earliest data point	2012 Q1	2011 Q1	2012 Q1	2014 Q2	2012 Q1

TABLE 3.2: Observations split up by region and the time of the earliest observation for each region.

Time point	Asia	Americas	Europe	Africa	Oceania
2014 Q2	43	51	29	3	3
2014 Q3	57	55	32	5	4
2014 Q4	82	69	46	7	6
2015 Q1	115	106	84	9	7
2015 Q2	156	121	104	13	7
2015 Q3	179	131	118	16	8
2015 Q4	242	142	138	23	9
2016 Q1	259	149	145	24	9
2016 Q2	288	157	159	28	12

TABLE 3.3: Observations per time point and per region from 2014 Q2 onwards.

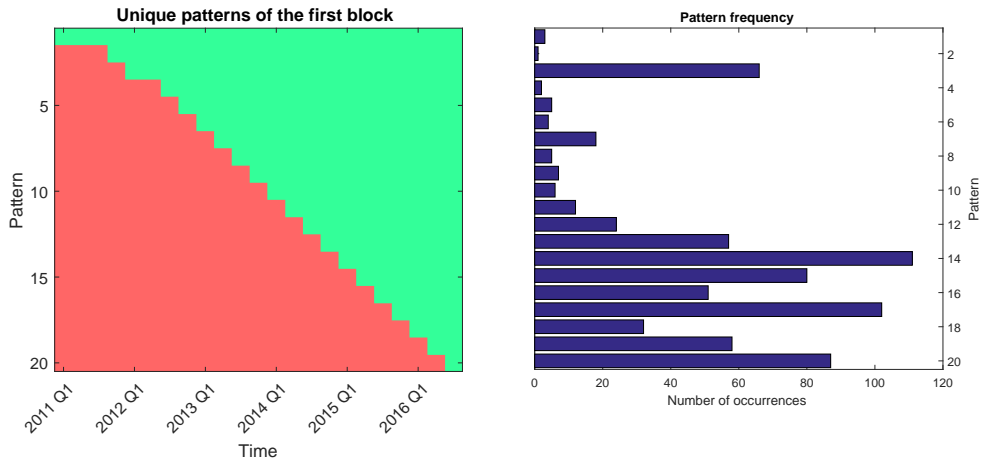


FIGURE 3.2: Unique pattern occurrences for the investigated block with only data connected to 2016 Q2. The right figure shows the frequency of each pattern.

3.2 Graphical analysis

In the last section, we decided to focus on the patterns that have connected data points until the last quarter, because we are interested in the development from one quarter to the next. We have also seen that data in earlier quarters are rare in general and that not all regions contain data from the beginning, i.e. we are dealing with an unbalanced setting.

As we use only data from 2014 Q2 onwards for the next steps, this means we have

- $d = 9$ time points resulting in 9 different patterns.
- For each time point $t \in \{1, \dots, 9\}$, we have $N_t \in \mathbb{N}^+$ observations available with $N_1 \leq \dots \leq N_d$.
- The total number of companies at the last time point is $N_d = 644$.
- For each company $i \in N_d$, we know its associated region $R_i \in \{\text{Europe, Americas, Asia, Africa, Oceania}\}$.

Thus, we have a data matrix $\mathbf{X} = (x_{i,t})$, where $x_{i,t} \in \mathbb{R}^+$ for $t = 1, \dots, d$ and $i = 1, \dots, N_t$ and $x_{i,t}$ is empty otherwise as we observe no data:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d-1} & x_{1,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N_1,1} & \vdots & \ddots & \vdots & \vdots \\ & \vdots & \ddots & \vdots & \vdots \\ & & \ddots & \vdots & \vdots \\ & x_{N_2,2} & \ddots & \vdots & \vdots \\ & & & \vdots & \vdots \\ & & & & \vdots \\ & & & x_{N_{d-1},d-1} & \vdots \\ & & & & \vdots \\ & & & & \vdots \\ & & & & x_{N_d,d} \end{bmatrix}$$

We have 4181 entries with observed values while 3129 are missing. In the following we will take a closer look at this set of data to determine which methods could be used on it to model the marginals.

In a first step, we look at the limits evolution over time, i.e. for each company i we plot $x_{i,t}$ for $t = 1, \dots, 9$ as a line, seen in Figure 3.3. Besides a few exceptions, most limits are too close to each other to see details here. We notice though that the majority of companies' limits are below a certain limit and only a few companies have higher limits.

Thus, we take a look at the mean limits for each region in Figure 3.4: For each time point t and each region R , let $N_{R,t}$ denote the number of observations available at time t for region R . Then the mean of region R at time t is given by

$$\bar{\mu}_t^R := \frac{1}{N_{R,t}} \sum_{i=1}^{N_t} x_{i,t} \mathbb{1}_{\{R_i=R\}}. \quad (3.1)$$

Notable are the different behaviours between the regions Americas, Asia, Europe and the two regions Africa, Oceania. Additionally we notice a changing behaviour in the first 3 regions' limits after the year 2014.

Figure 3.5 shows the same mean limits over time with added confidence bands of limits for each region. The confidence bands are based on the 90% and 10% quantiles of the limit data. We notice that these confidence bands don't show a lot of differences between regions from the year 2015 onwards.

This can also be seen in Figure 3.6 where we plot the standard deviation of each region dataset over time, which is given by

$$\bar{\sigma}_t^R := \sqrt{\frac{\sum_{i=1}^{N_t} \mathbb{1}_{\{R_i=R\}} (x_{i,t} - \bar{\mu}_t^R)^2}{N_{R,t} - 1}}. \quad (3.2)$$

While in the first 3 time steps all regions behave very differently, afterwards the 3 regions Americas, Asia, Europe show a similar downward trend while Africa and Oceania are increasing in variance.

This pattern is also illustrated in Figure 3.7 where we plot the mean

limits with confidence bands based on the standard deviation. Later points in time are more similar while earlier points show a bigger difference.

Figure 3.8 shows the development of the relationship between mean and standard deviation over time, i.e. we compute $\bar{\mu}_t^R / \bar{\sigma}_t^R$. We observe that Africa and Oceania have the most variation over time, which can be attributed to the fact that these regions contain only few observations. The other 3 regions show a more constant behaviour but still have some variation, especially in early time points.

Next we look at a boxplot graphic of the limit data in Figure 3.9. We can identify a small linear downward trend over time on average but see many larger outliers. To remedy this, we look at the logarithms of the limits in Figure 3.10. This helps to reduce the amount of outliers and we can see the linear downward trend more clearly.

As done before, we group the data into the 5 regions Asia, Americas, Europe, Africa and Oceania. This way, we can investigate different developments from geographically separated companies and still have enough data points per region. Figure 3.11 shows the boxplots of logarithmic limits of the whole time for each region. While the average is roughly the same for all regions, there is a lot more variation for the three regions Asia, Americas and Europe than the other two, as we could also observe in previous plots.

To be able to differentiate between the different regions over time, we look at the boxplots of the logarithmic limits for each region separately in Figure 3.12: We observe that the downward trend holds true for Americas, Asia and Europe. In each case, we can also observe a more drastic decrease at the end of 2014. Africa shows very little variation and seems to remain rather stable. Only data from Oceania shows a linear upward trend. These two regions also have less outliers, however they are also the ones with the smallest number of total observations.

Finally, to combat the outliers even more, we look at a Box-Cox transformation of the data, i.e. for observations $x_{i,t}$ and parameter $\lambda \in \mathbb{R}$, we consider the transformed values

$$\tilde{x}_{i,t} = \frac{x_{i,t}^\lambda - 1}{\lambda}.$$

Figure 3.13 shows the boxplots for a lambda of -0.6, derived in R by using `BoxCox.lambda(Limit, method="loglik")`. Indeed, the transformed data has very few outliers now while the general trends observed from the log transformation mostly remain and can be seen even more clearly. It should be noted that the ranges are more compressed now and the absolute differences have become small.

Thus, to summarize, we have seen evidence that for marginal modelling purposes, we should try approaches that can model changing variance and that it might be useful to combine regions that behave similar into groups.

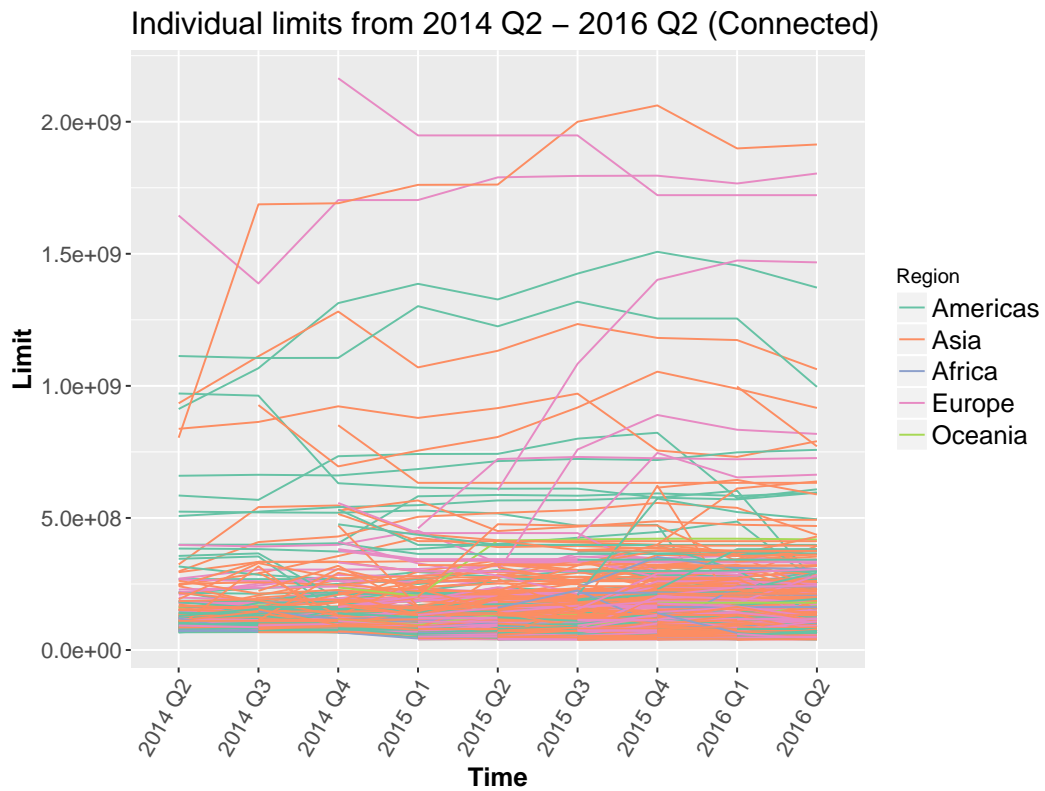


FIGURE 3.3: The connected limit development of all companies over time. Each region is colour coded.

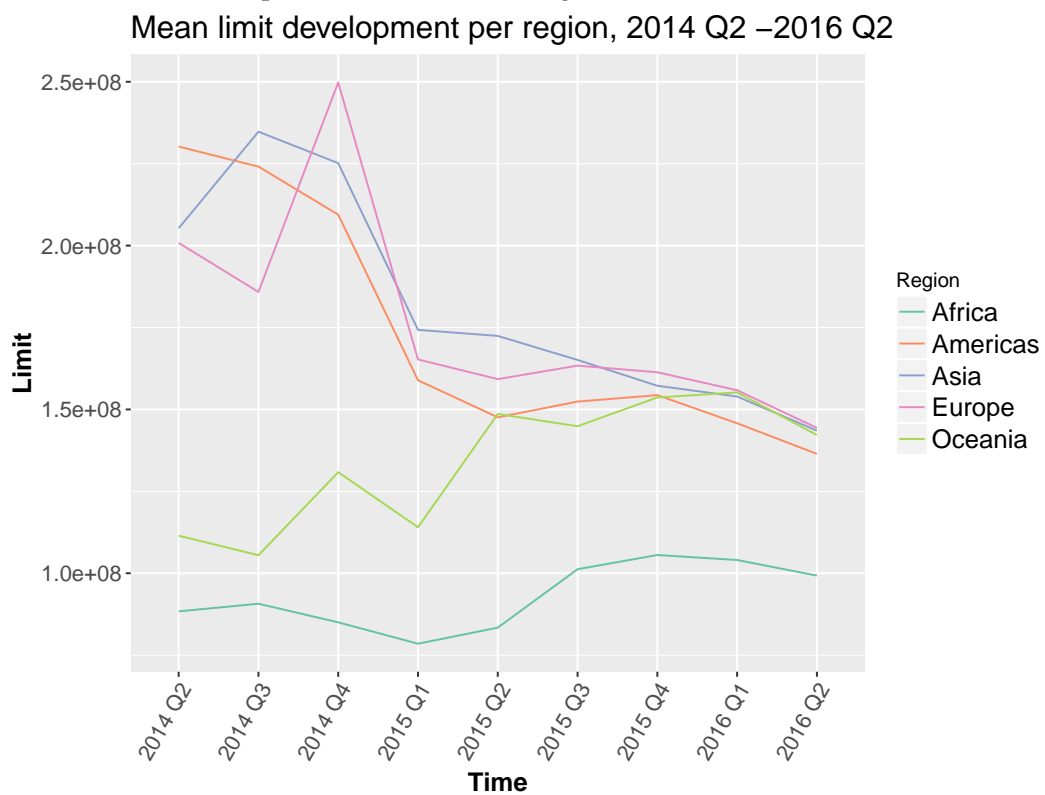


FIGURE 3.4: The average exposure limit per time and region. Each region is colour coded.

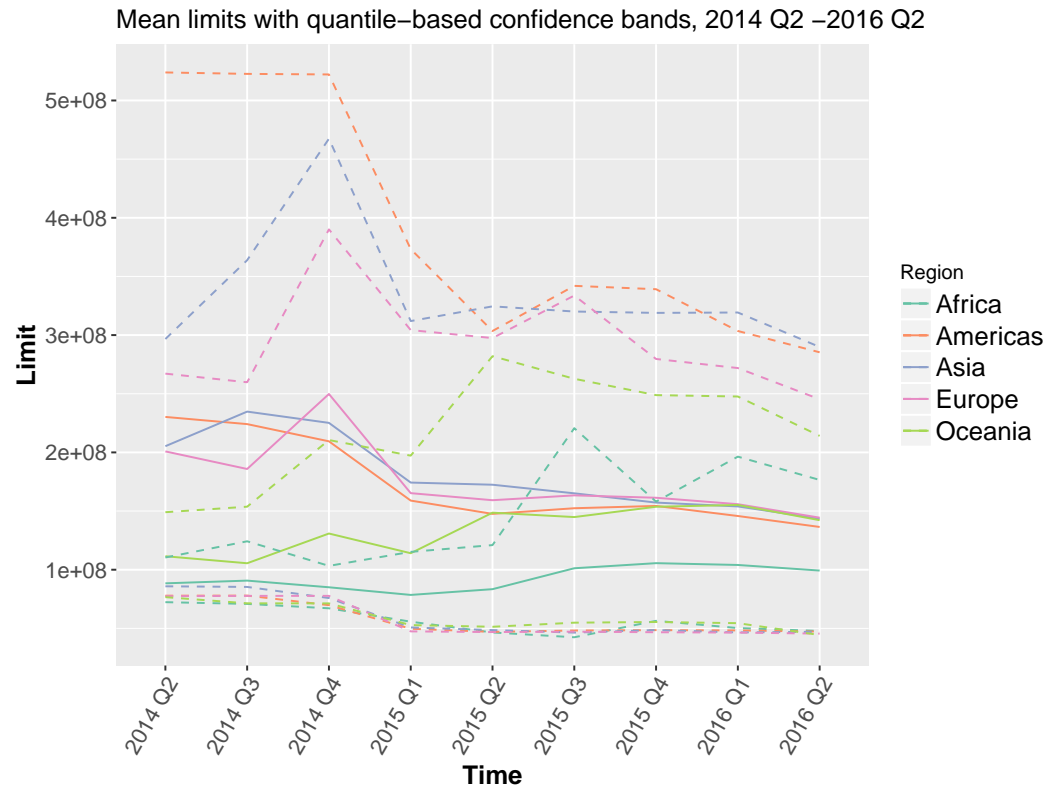


FIGURE 3.5: Mean Limits with confidence bands based on quantiles. The upper/lower band limit at time t is given by the 90% and 10% quantile respectively. Each region is colour coded.

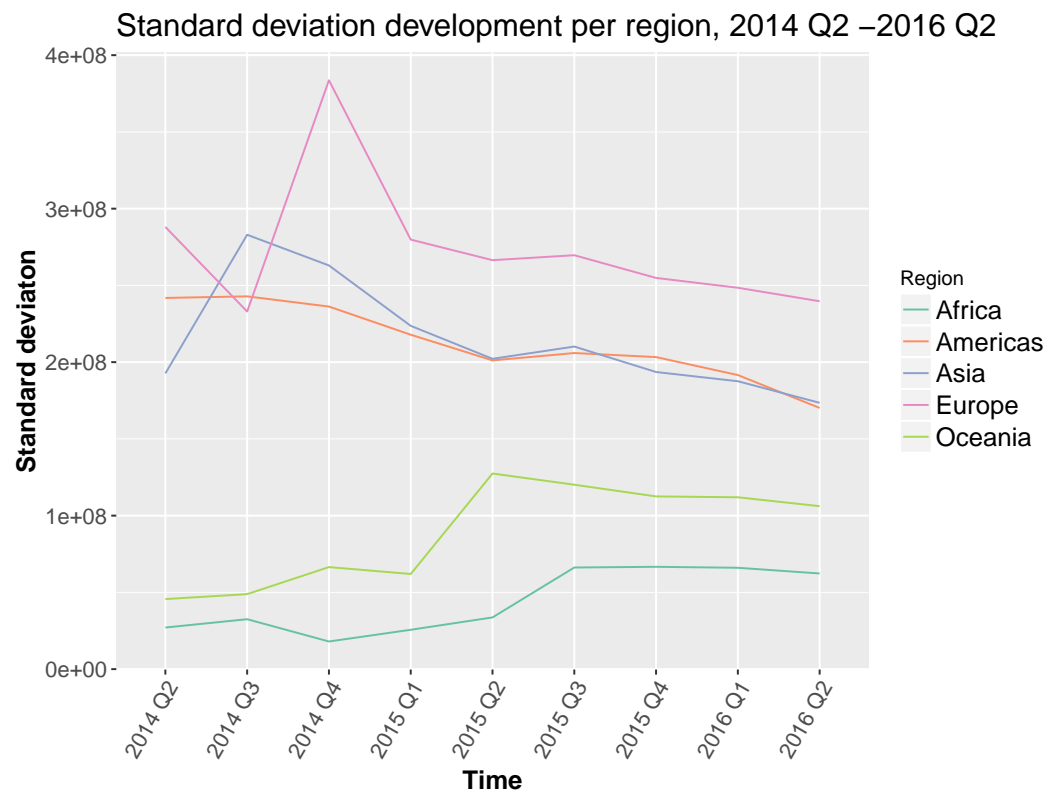


FIGURE 3.6: Standard deviation of the data set per time and region. Each region is colour coded.

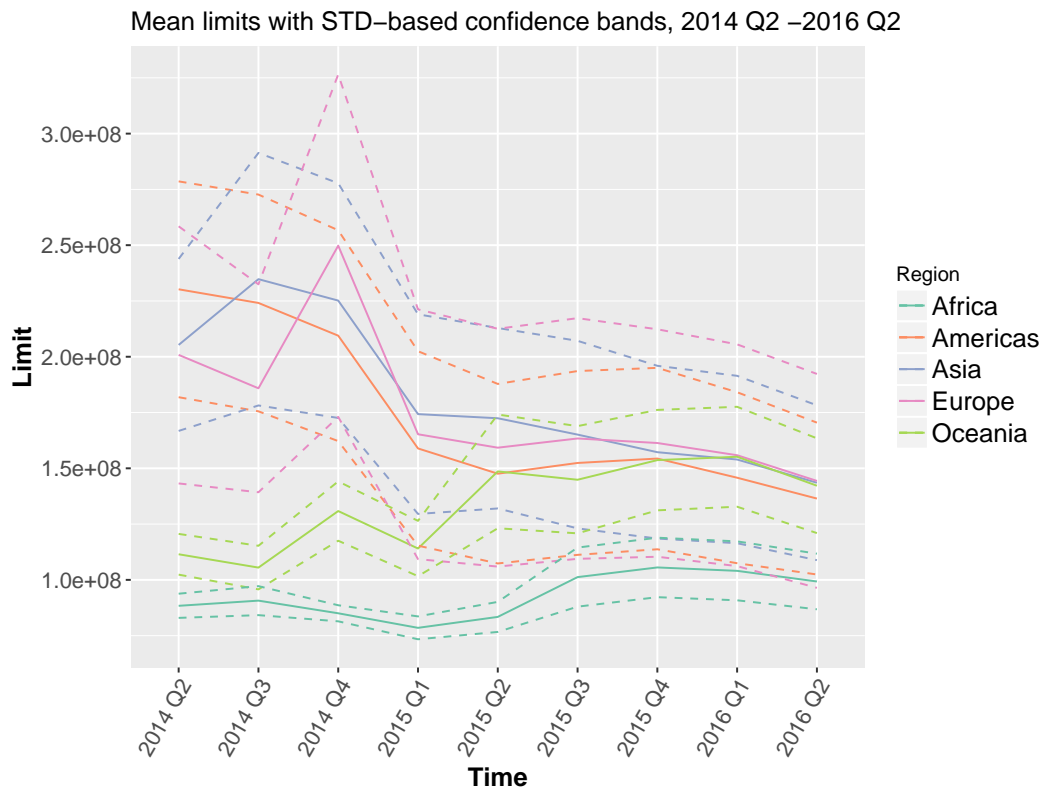


FIGURE 3.7: Mean Limits with confidence bands based on standard deviation. The upper/lower band limit at time t is given by $\text{Mean}_t \pm 0.2 * SD_t$ (note that 0.2 is an arbitrary scale to improve readability). Each region is colour coded.

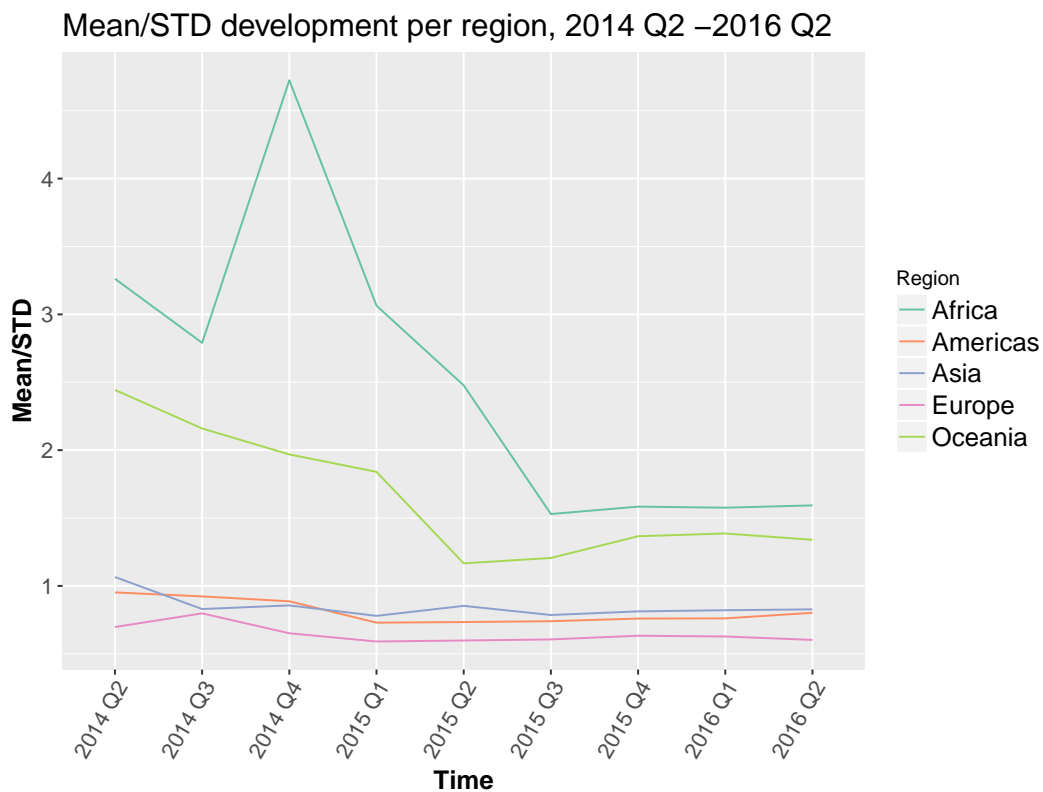


FIGURE 3.8: Mean divided by standard deviation over time per Region. Each region is colour coded.

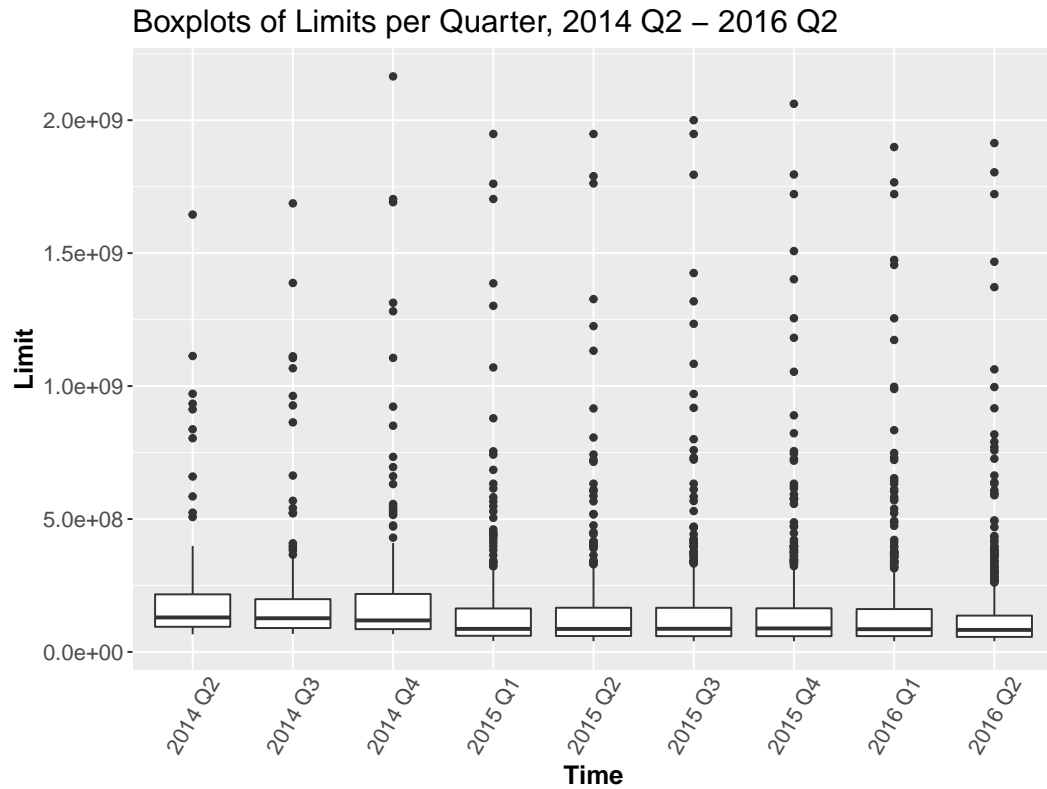


FIGURE 3.9: Boxplots for all limits over time.

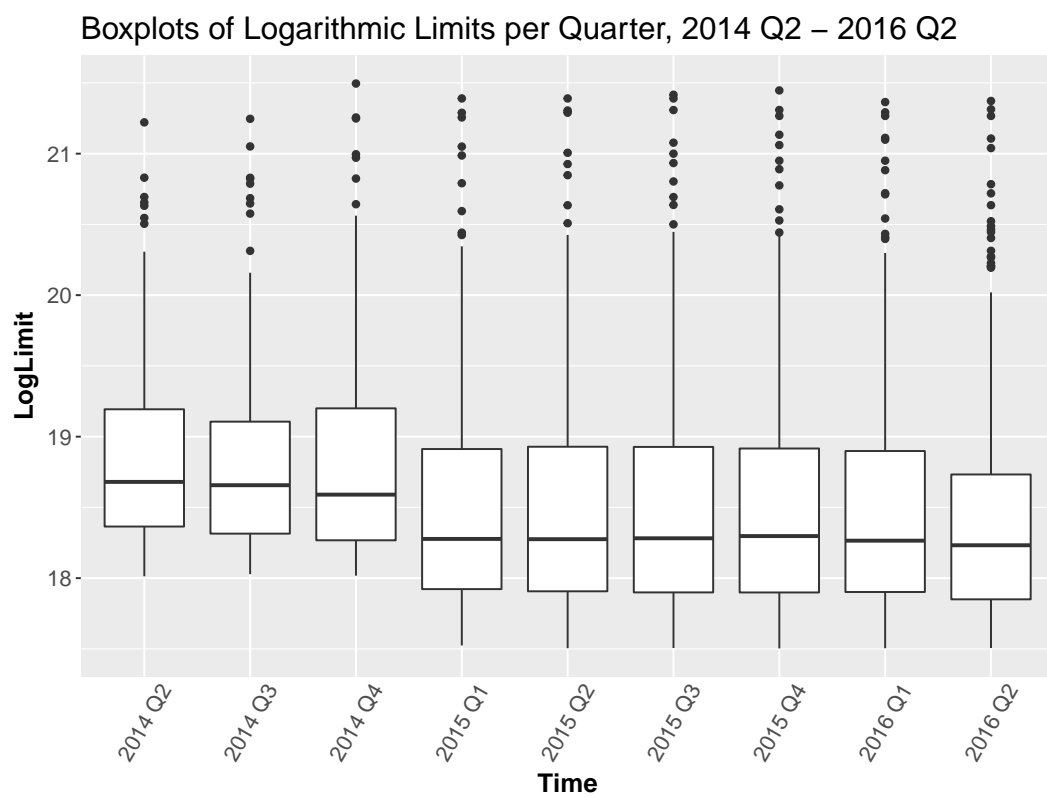


FIGURE 3.10: Boxplots for the logarithm of all limits over time.

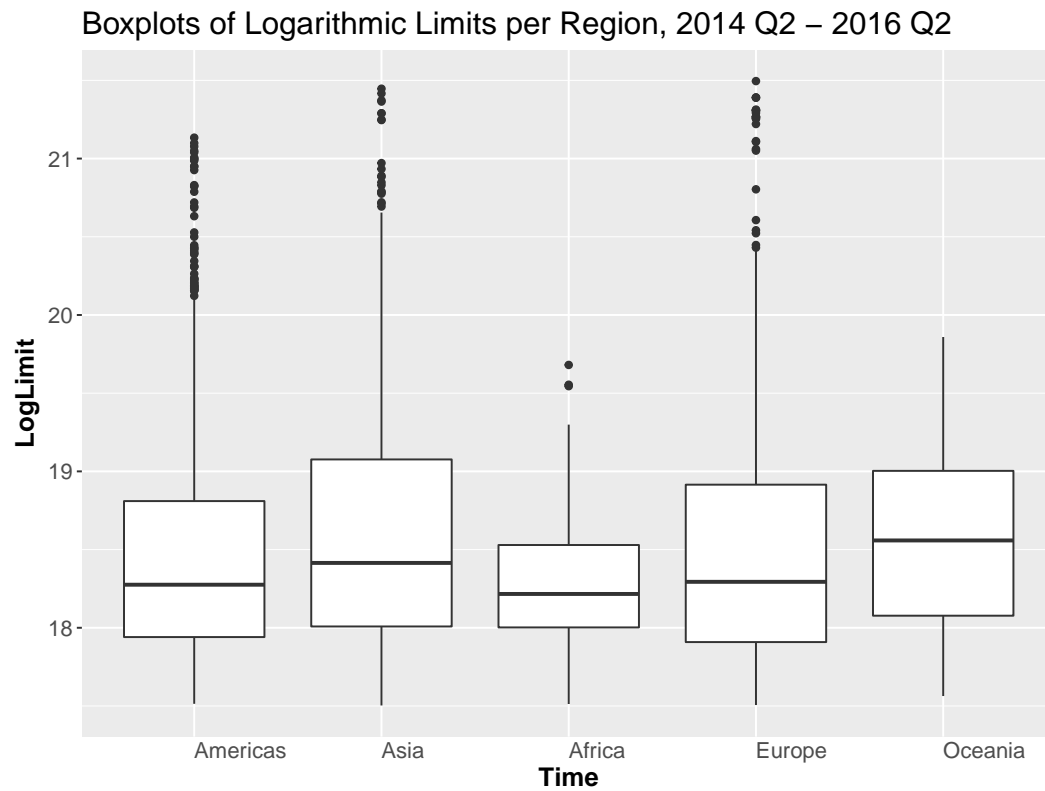


FIGURE 3.11: Boxplots for the logarithm of all limits per region.

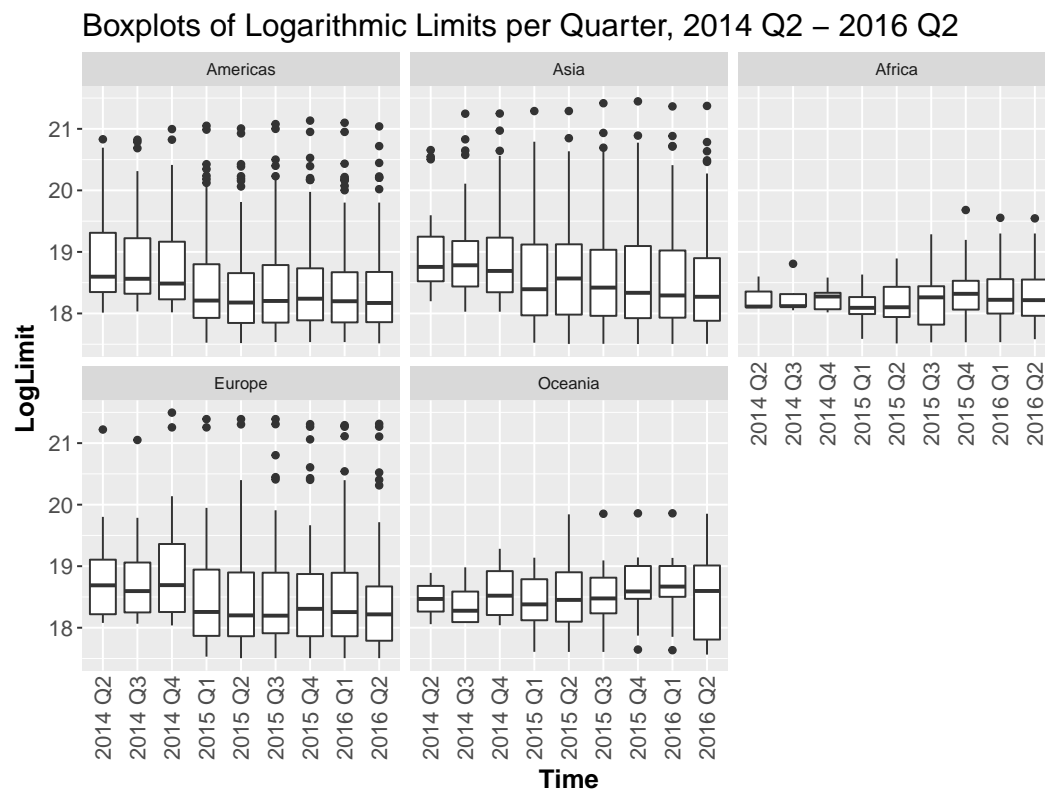


FIGURE 3.12: Boxplots for the logarithm of all limits over time split up by region.

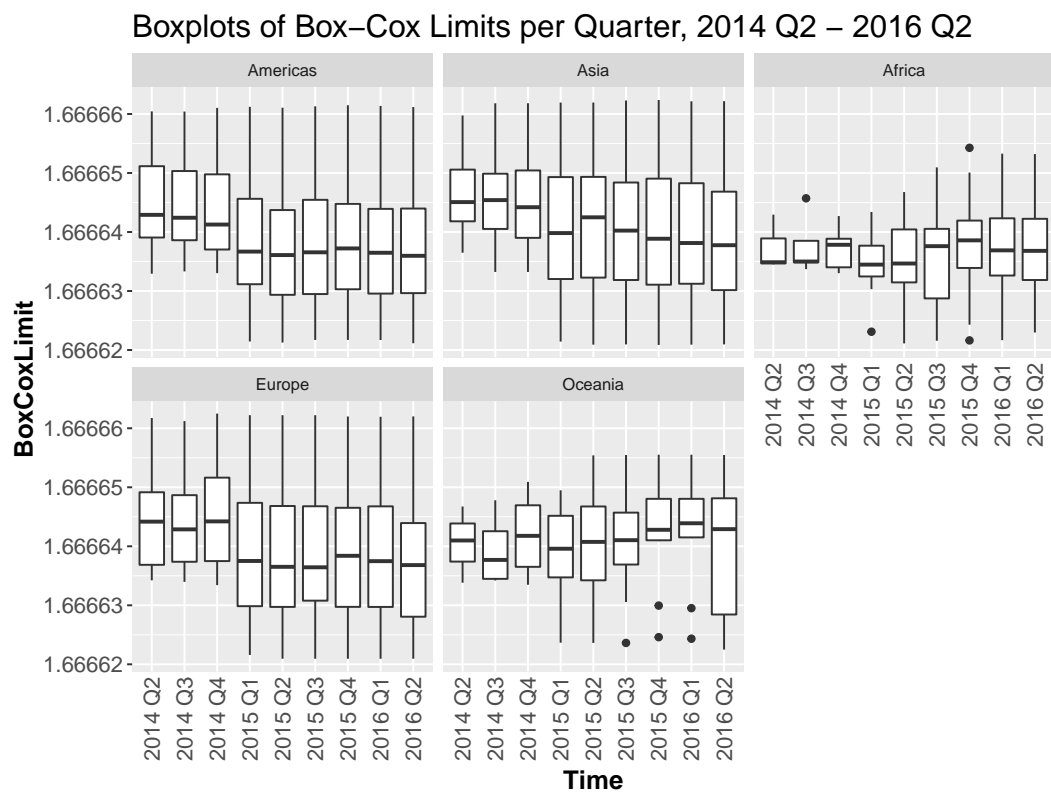


FIGURE 3.13: Boxplots for the Box-Cox transformation of all limits over time split up by region. The Box-Cox parameter is chosen as $\lambda = -0.6$.

Chapter 4

Marginal model fitting

In Chapter 3, we discussed in detail our available data set of exposure limits. In this chapter, we will apply different models introduced in Chapter 2 that we use to work with this limit data. Recall from Chapter 2 that the goal of each model is to estimate the distributions of marginals in order to gain approximately uniformly distributed data that we can then use as a basis for our copula model. Overall we apply four different models. Each section contains a short description of the methods used and how we arrive at the uniformly distributed data sets. The last part of this chapter uses diagnostics to compare the estimations.

4.1 Marginal model specification

Model 1: Non-parametric probability integral of limits

The first model we introduce uses a non-parametric kernel density estimation of the limit data at each time step. With this, we also obtain estimated cumulative distribution functions for each point in time that we apply to the limit data to gain uniformly distributed data (i.e. we use the probability integral transformation).

Thus, this model does not differentiate between regions and estimates only one density for all exposure limits at a given time step. In R, this is achieved by applying the function `kde1d` from the package `kdevine` to our data (Nagler, 2017b). The resulting estimated densities are displayed in Figure 4.1. We can observe that the densities show similar behaviours in general, while the first three quarters are slightly shifted to the right. This captures the behaviour seen in the previous Chapter 3, where we noticed higher limits for these three quarters.

We also note the effect of the left truncation which results in the density being set to zero for limits below a certain threshold. On the other hand, densities for large limits tend to have a noticeable variation. While it is possible to smooth the density by adjusting the bandwidth used in the estimation, this would also cause the density for lower limits to increase (as the whole curve widens) which in turn is not a good fit for our truncated data. Thus, we decide to choose the bandwidth best balancing both difficulties in this case.

The function `pkde1d` from the same R package allows us to get the corresponding estimated cumulative distribution functions $\hat{F}_t^{X,NPIT}$ for each time step t . Let denote as before N_t the number of total observations at time

t and the limit data by $x_{i,t}$ for $i = 1, \dots, N_t$ to obtain

$$\hat{u}_{i,t}^{X, NPIT} := \hat{F}_t^{X, NPIT}(x_{i,t}),$$

where for each t and $i = 1, \dots, N_t$, the $\hat{u}_{i,t}^{X, NPIT}$ are now approximately uniformly distributed in the interval $[0, 1]$.

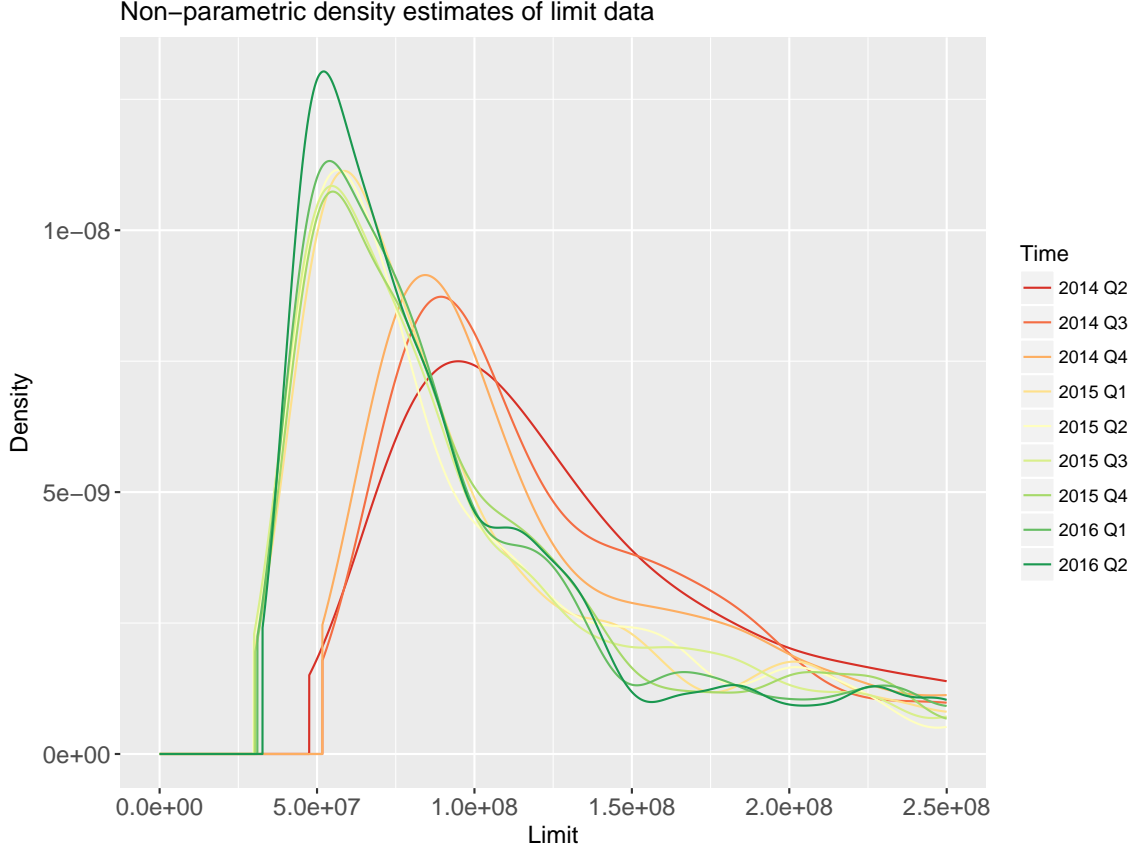


FIGURE 4.1: Overlaid non-parametrically estimated densities of limits for each time point between 2014 Q2 and 2016 Q2.

Model 2: Non-parametric probability integral of standardized observations

In this model, we take a form of standardized observations by detrending and normalizing the limit data at each time step. An important difference to the previous model is that we do so for each region separately. This means we assume that the region has an effect on the exposure limit. As we have seen that the three regions Americas, Europe and Asia behave similar while the remaining two regions Oceania and Africa show a different behaviour, we can combine these regions into a "Main" region (America, Europe, Asia) and "Other" region (Oceania, Africa). See Figure 4.3 for kernel density estimates for these combined regions at each point in time. We can observe that the densities indeed behave differently.

For the next step, we will however keep the distinction between all 5 regions. We will keep the notation from before and additionally we denote by R_i the region where company i is located.

Thus, we take the means and standard deviations of regions as defined in equations (3.1) and (3.2) to get the simplified residuals for each time t by setting

$$\tilde{x}_{i,t} := \frac{x_{i,t} - \bar{\mu}_t^{R_i}}{\bar{\sigma}_t^{R_i}}, \quad i = 1, \dots, N_t.$$

Similarly to the first model, we estimate the density of these adjusted limits non-parametrically. Figure 4.2 shows the resulting densities for each time step. We can observe that all densities have similar shapes and that there are more large positive outliers than smaller ones, as we could also see in the EDA plots from the previous chapter and which is again mostly explained by the left truncation. For this KDE, we increased the bandwidth in order to smooth out the right parts of the curves. The bandwidth was increased until the point where the resulting u-data was still approximately uniformly distributed based on the KS-test at the 0.05 confidence level. Again, this gives us a balance between smooth densities and an appropriate fit for our data.

Again, we arrive at the uniformly distributed data by applying the corresponding estimated cumulative distribution functions $\hat{F}_t^{STDX, NPIT}$:

$$\hat{u}_{i,t}^{STDX, NPIT} := \hat{F}_t^{STDX, NPIT}(\tilde{x}_{i,t}), \quad i = 1, \dots, N_t.$$

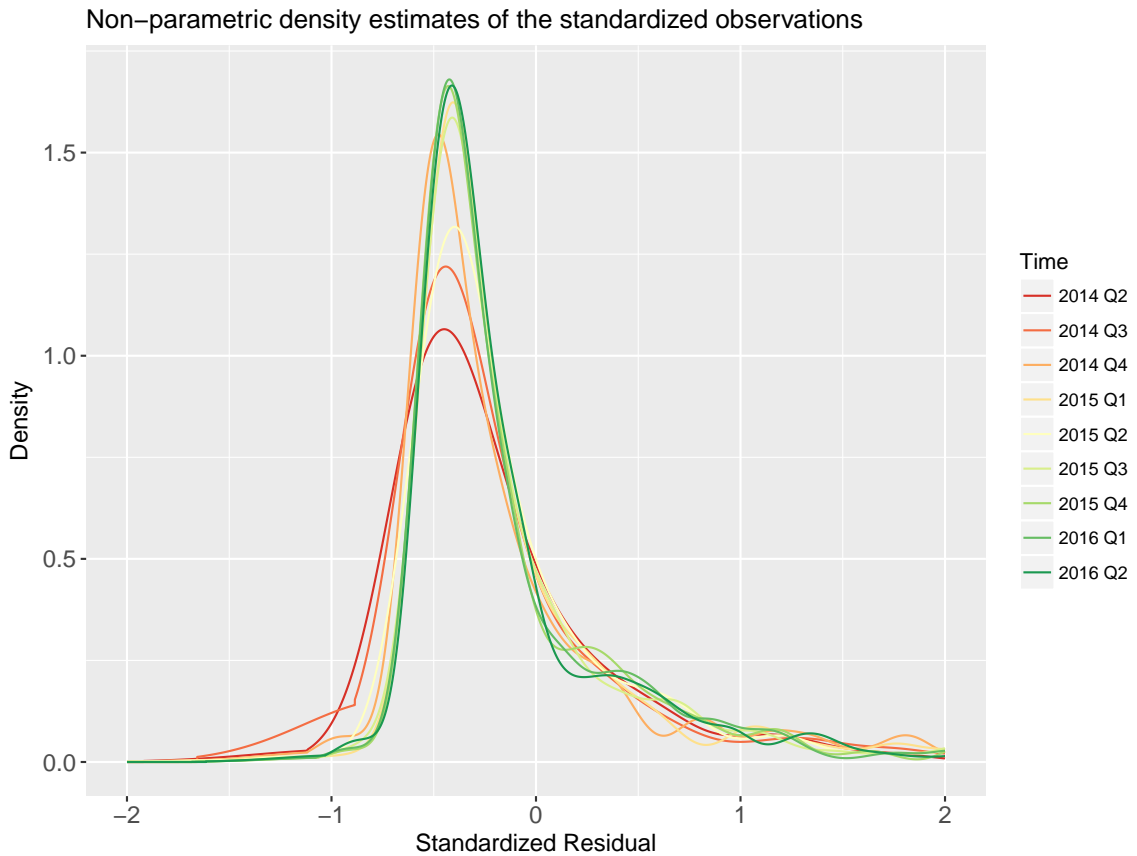


FIGURE 4.2: Overlaid non-parametrically estimated densities of standardized observations (i.e. subtracted the means and divided by standard deviations) for each time point between 2014 Q2 and 2016 Q2.

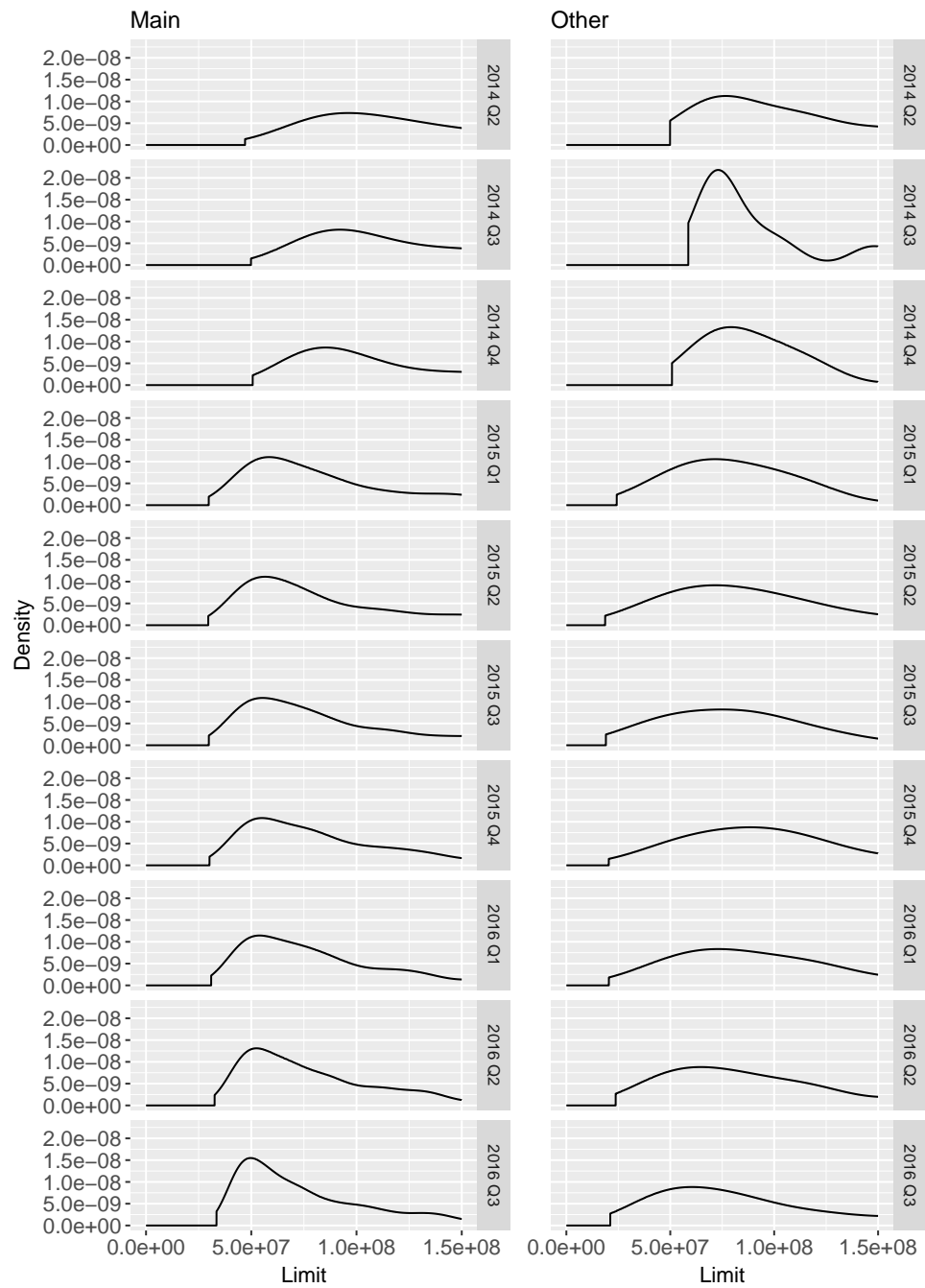


FIGURE 4.3: Kernel density estimates of limit data per time for the combined regions "Main" and "Other".

Model 3: Parametric probability integral of regression residuals

For our third model, we will not work with the limits directly and instead fit a regression model to the data first. From this regression model, we gain residuals for each time point which we will transform into approximately uniformly distributed data by using a probability integral transformation.

Thus, we perform Generalized Gamma regressions (see Chapter 2) on the limit data at each time point t . As we have seen evidence in Chapter 3 that both the mean and variance depend on the region, we use the region as the explaining variable for both μ and σ in these regressions.

This time, we use the combined regions "Main" and "Other" as introduced before. As link functions we choose the logarithm for both μ and σ while ν has the identity function as its link.

Thus, for each time t we estimate the parameters with the following three equations (see equation 2.8 from the definition):

$$\begin{aligned}\log(\hat{\mu}_R) &= \hat{\beta}_1^0 + \hat{\beta}_1^1 \times \mathbb{1}_{(R=\text{Other region})}(R) , \\ \log(\hat{\sigma}_R) &= \hat{\beta}_2^0 + \hat{\beta}_2^1 \times \mathbb{1}_{(R=\text{Other region})}(R) , \\ \hat{\nu} &= \hat{\beta}_3^0 ,\end{aligned}$$

where $R \in \{\text{Main region}, \text{Other region}\}$ is the region variable.

Using the R package `gamlss` (Rigby and Stasinopoulos, 2005) we fit our data accordingly. Table 4.1 shows the estimated parameters for each time step. Figures 4.4 and 4.5 display the resulting densities (once for the "Main" region and once for "Other"). As before, we can note the slightly different behaviour of the first three time points. Especially for these, the region has a notable effect on the fit.

	Main		Other		
Time	$\log(\hat{\mu})$	$\log(\hat{\sigma})$	$\log(\hat{\mu})$	$\log(\hat{\sigma})$	$\hat{\nu}$
2014 Q2	18.30	-1.28	18.16	-1.71	-9.71
2014 Q3	18.23	-1.49	18.12	-1.98	-15.25
2014 Q4	18.07	-2.27	18.04	-2.65	-80.46
2015 Q1	17.84	-1.16	17.82	-1.35	-8.84
2015 Q2	17.78	-1.24	17.73	-1.34	-10.65
2015 Q3	17.72	-1.41	17.67	-1.47	-15.50
2015 Q4	17.84	-1.13	17.86	-1.20	-8.19
2016 Q1	17.74	-1.35	17.75	-1.38	-12.89
2016 Q2	17.72	-1.44	17.72	-1.48	-14.92

TABLE 4.1: Estimated parameters for each time point. The first two columns refer to the estimators of the logarithm of location and scale parameters μ and σ for the "Main" region, the next two columns to the "Other" region. The last column contains the estimators for the shape parameter ν which is estimated independently of the region.

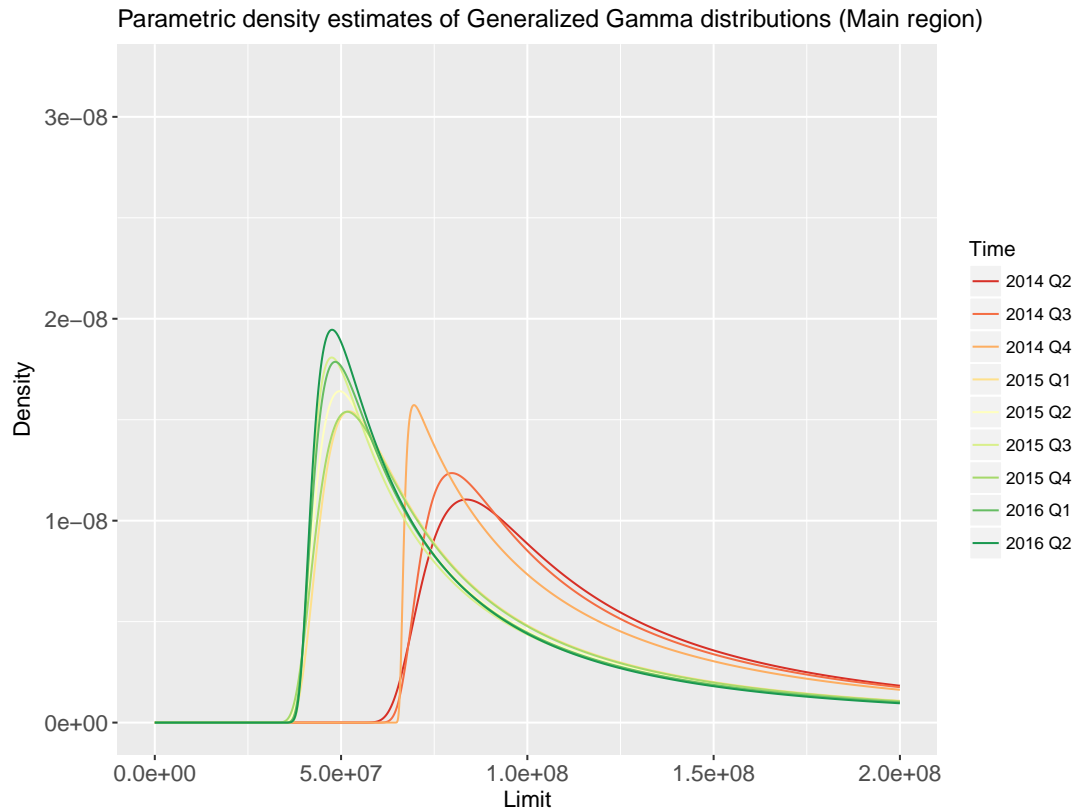


FIGURE 4.4: Overlaid parametrically estimated densities of Generalized Gamma distributions for the "Main" region.

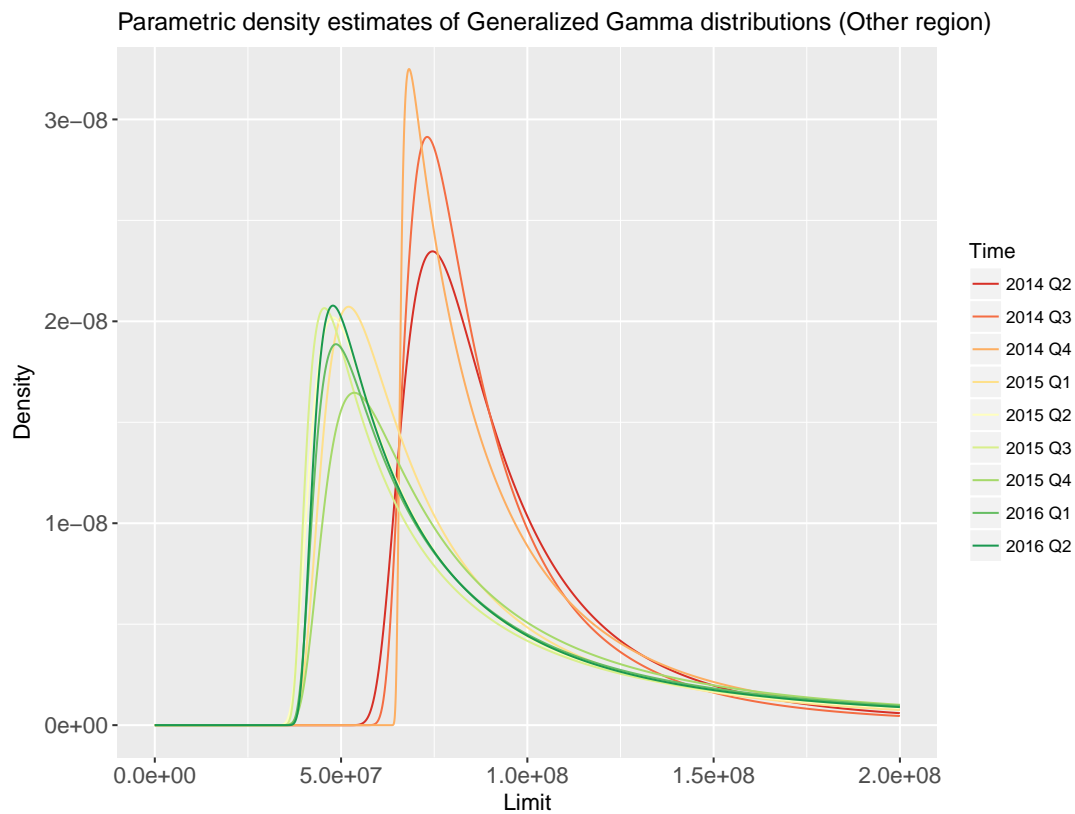


FIGURE 4.5: Overlaid parametrically estimated densities of Generalized Gamma distributions for the "Other" region.

With the fitted models, we can get the corresponding residuals $r_{i,t}$ for each time step t and $i = 1, \dots, N_t$. As shown before, to get approximately uniformly distributed data, we can simply apply the cumulative distribution function (cdf) of a standard normal to these residuals, i.e.

$$\hat{u}_{i,t}^{GAMLSS,PPIT} := \Phi(r_{i,t}), \quad i = 1, \dots, N_t.$$

Model 4: Non-parametric probability integral of regression residuals

The fourth and last model is similar to the third one. We take the same GAMLSS regression and its residuals as a starting point. However, instead of using the probability integral transformation from the standard normal distribution, we estimate the densities of residuals non-parametrically.

Let $r_{i,t}$ be the normalized quantile residual of company i from the regression performed at time t , as defined before. We then use the R package `kdevine` and its function `kde1d` to estimate a density of residuals at each time point. Figure 4.6 shows the resulting densities. Similarly to previous models, we apply the corresponding estimated cdfs $\hat{F}_t^{GAMLSS,NPIT}$ to the residuals to obtain

$$\hat{u}_{i,t}^{GAMLSS,NPIT} := \hat{F}_t^{GAMLSS,NPIT}(r_{i,t}).$$

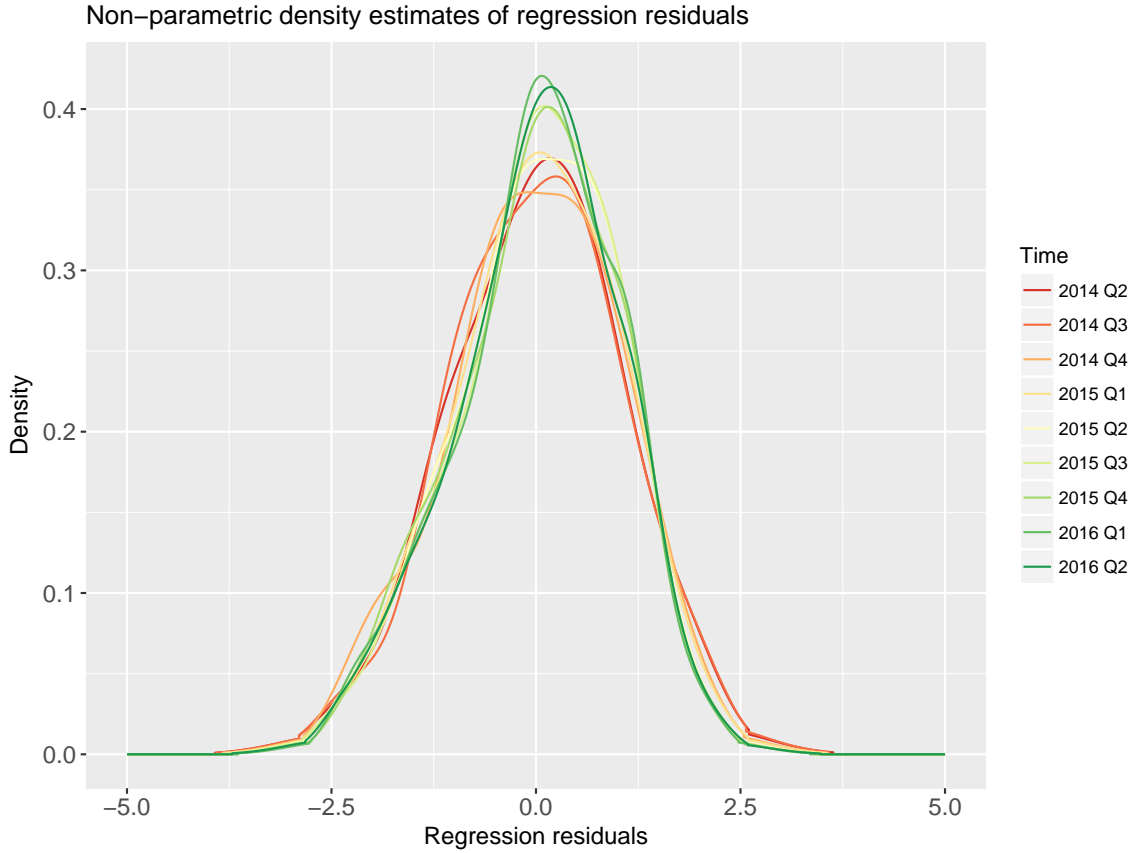


FIGURE 4.6: Overlaid non-parametrically estimated densities of regression residuals for each time point between 2014 Q2 and 2016 Q2.

4.2 Overview of different marginal model approaches

In the previous sections, we introduced 4 different models with the goal to gain approximately uniformly distributed data from our original limit data. Figure 4.7 shows an overview of the procedures for each model. Summarized, the important aspects of each model are:

- Model X,NPIT (KDE of limit data): We assume identical distributions of exposure limits for all regions. No parametric assumptions are imposed on the data.
- Model STDX,NPIT (KDE of standardized observations): A region effect is removed in a simple approach. Detrending and normalization of limits based on region result in standardized observations in a non-parametric way.
- Model GAMLSS,PPIT and model GAMLSS,NPIT (GAMLSS residuals): A region effect is removed based on complex regression models to obtain residuals parametrically. We also group together regions into two aggregated groups "Main" and "Other".

The next section will analyse the resulting u-data from each model and compare them.

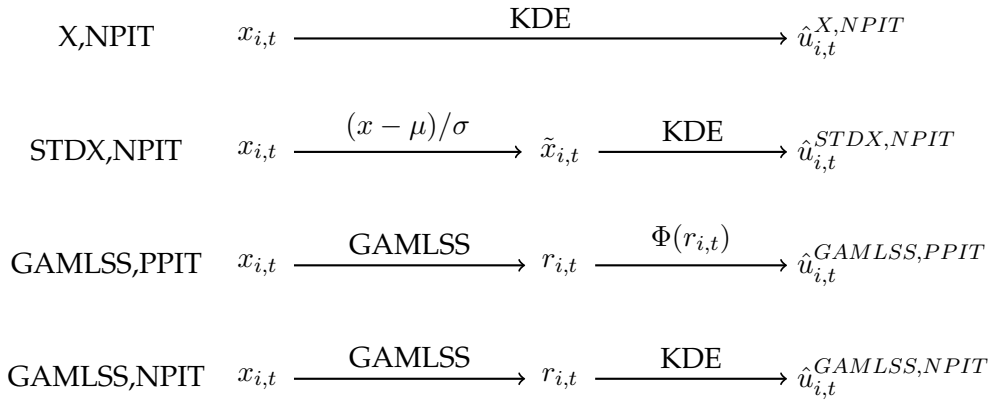


FIGURE 4.7: Simplified overview of the four introduced marginal models.

4.3 Marginal models comparison

In the previous section, we introduced the four marginal models that we work with and showed their associated fits. This chapter will compare these models to each other and discuss the results. Specifically, we start with regression diagnostics for the Generalized Gamma regression and then compare the u-data that we gained from each model with each other

Regression diagnostics

Model GAMLSS,PPIT and model GAMLSS,NPIT each started with Generalized Gamma (GG) regressions to obtain residuals. We will take a closer look at the regressions to determine whether it is appropriate to use the results.

Recall that we fit a separate GG model to the data at each time step since we want to estimate the marginals, and that the only covariate is the region of the companies (combined into two regions called "Main region" and "Other region"). Thus, as we have 9 time points, we have 9 regressions to analyse. In the following we will look at the regression output from R as well as histograms and Q-Q plots.

R output

At each point in time t , we call the `gamlss` function from the `gamlss` package with the following parameters:

```
gamlss(formula = Limit ~NewRegion, sigma.formula =
~NewRegion, family = GG, data = na.omit(Limits[,t]),
method = mixed(20, 60)).
```

Table 4.2 shows an extract of the summary for each model. It lists the estimates as well as standard deviations and p-values for the hypothesis test with the null hypothesis and alternative

$$H_0 : \beta = 0 ,$$

$$H_1 : \beta \neq 0$$

for each parameter β . We can see that some p-values are rather high, specifically those associated with estimations for the "Other region" covariate. This means that the effect of this variable is not very significant. However, in our case this is not a problem as we do not want to use the models to predict values directly. Instead we are interested in the residuals and their behaviour. Figure 4.8 visualizes the evolution of the parameters over time. We can once again observe a different behaviour during the first three quarters. Noticeable is also the estimated value of the third parameter ν and its large standard deviation at 2014 Q4.

		2014 Q2	2014 Q3	2014 Q4	2015 Q1	2015 Q2
$\log \mu$	β_1^0	18.3(0.08)	18.23(0.07)	18.07(0.03)	17.84(0.06)	17.78(0.06)
	p-value	<2E-16	<2E-16	<2E-16	<2E-16	<2E-16
	β_1^1	-0.14(0.09)	-0.11(0.07)	-0.02(0.02)	-0.02(0.07)	-0.05(0.06)
	p-value	0.12	0.12	0.34	0.77	0.45
$\log \sigma$	β_2^0	-1.28(0.21)	-1.49(0.23)	-2.27(0.39)	-1.16(0.15)	-1.24(0.16)
	p-value	9.16E-09	1.24E-09	1.57E-08	2.46E-14	2.52E-13
	β_2^1	-0.43(0.23)	-0.49(0.19)	-0.38(0.15)	-0.19(0.14)	-0.1(0.12)
	p-value	0.07	0.01	0.01	0.18	0.39
ν	β_3^0	-9.71(4.53)	-15.25(7.54)	-80.46(62.09)	-8.84(2.85)	-10.65(3.8)
	p-value	0.03	0.04	0.20	0.00	0.01

		2015 Q3	2015 Q4	2016 Q1	2016 Q2
$\log \mu$	β_1^0	17.72(0.06)	17.84(0.06)	17.74(0.07)	17.72(0.05)
	p-value	<2E-16	<2E-16	<2E-16	<2E-16
	β_1^1	-0.05(0.05)	0.03(0.06)	0(0.05)	0(0.04)
	p-value	0.36	0.62	0.97	0.95
$\log \sigma$	β_2^0	-1.41(0.2)	-1.13(0.14)	-1.35(0.2)	-1.44(0.17)
	p-value	5.49E-12	3.00E-14	7.74E-11	3.66E-16
	β_2^1	-0.05(0.11)	-0.07(0.1)	-0.04(0.09)	-0.04(0.08)
	p-value	0.63	0.51	0.71	0.62
ν	β_3^0	-15.5(6.59)	-8.19(2.68)	-12.89(5.67)	-14.92(5.54)
	p-value	0.02	0.00	0.02	0.01

TABLE 4.2: R-output for the Generalized Gamma regressions. The values in brackets behind estimations refer to the standard deviations.

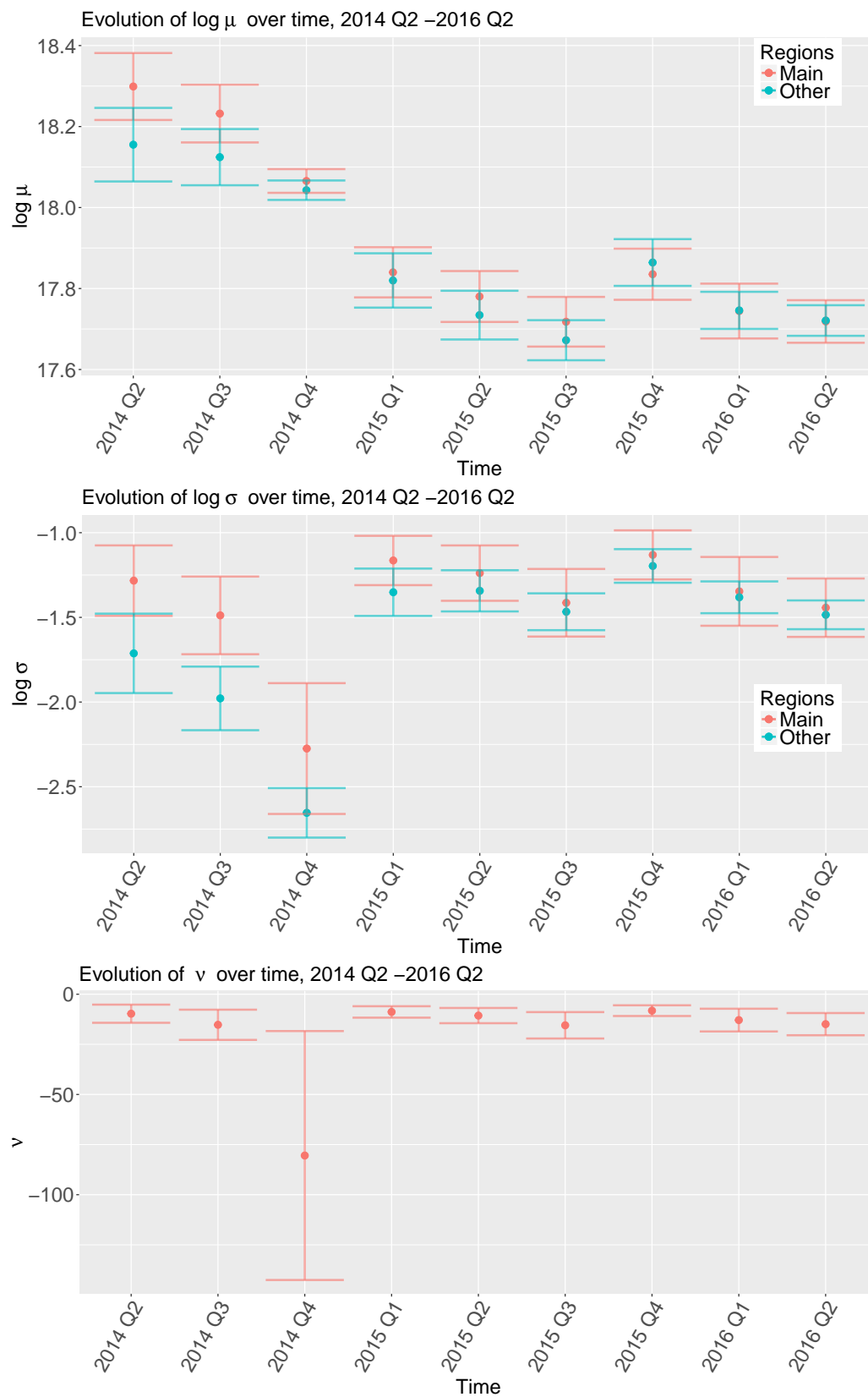


FIGURE 4.8: Evolution of the GAMLSS regression parameters over time. The estimated values are visualized by the points and the standard deviations are shown as error bars.

Residuals plots

In order to determine the goodness of fit, we take a closer look at the residuals. We begin by showing histograms of residuals for each time step in Figure 4.9. We want the residuals to be approximately normally distributed (see Chapter 2 for an explanation). Comparing the values to the overlaid red density of a standard normal distribution, we see that this holds true more or less for most time points, although larger negative values seem to be slightly underrepresented in our data.

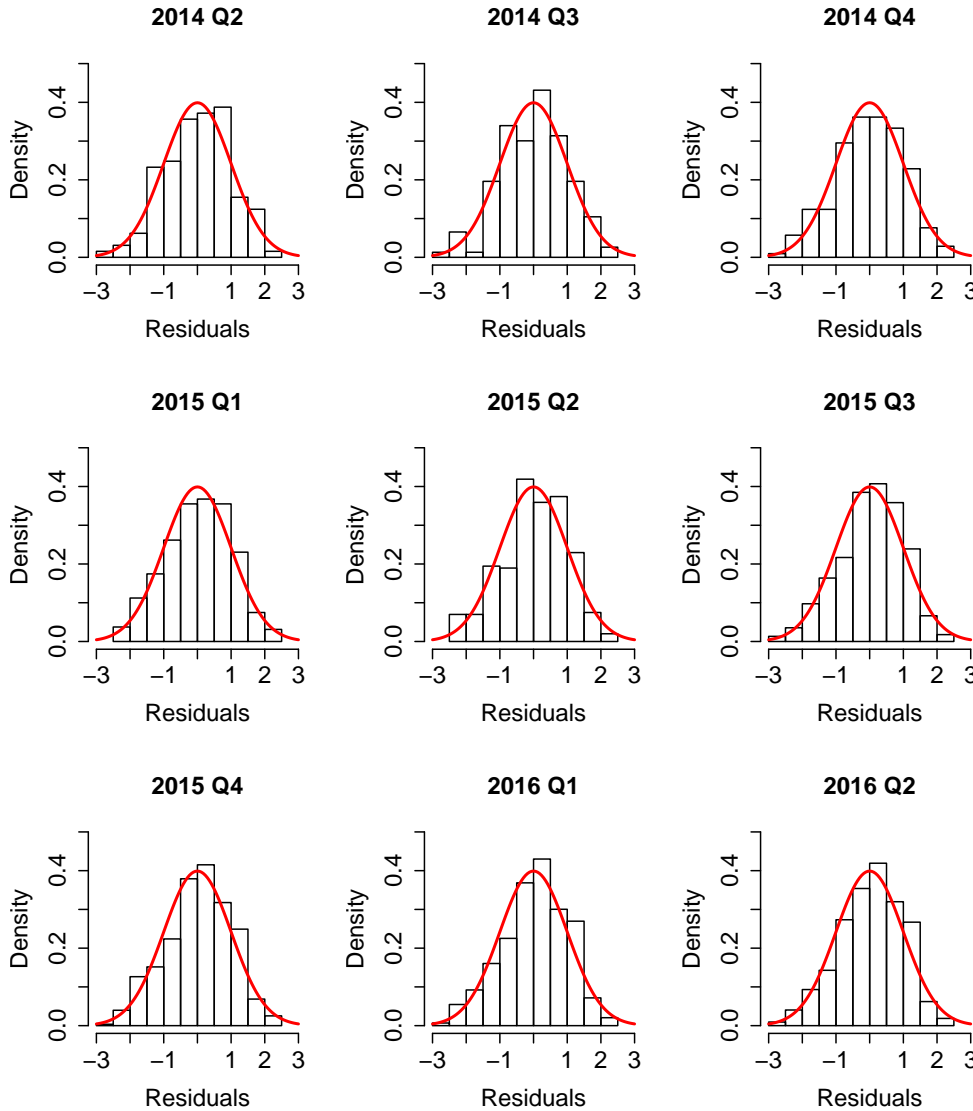


FIGURE 4.9: Histograms of residuals from the Generalized Gamma regressions for each time step. The red overlaid curve represents a standard normal density.

Next, we look at the Q-Q plots of residuals. In Figure 4.10 we can see that the quantiles show a generally good behaviour except for larger negative values, which deviate upwards from the line, implying there are too few large negative values (as we could also observe in the histograms before).

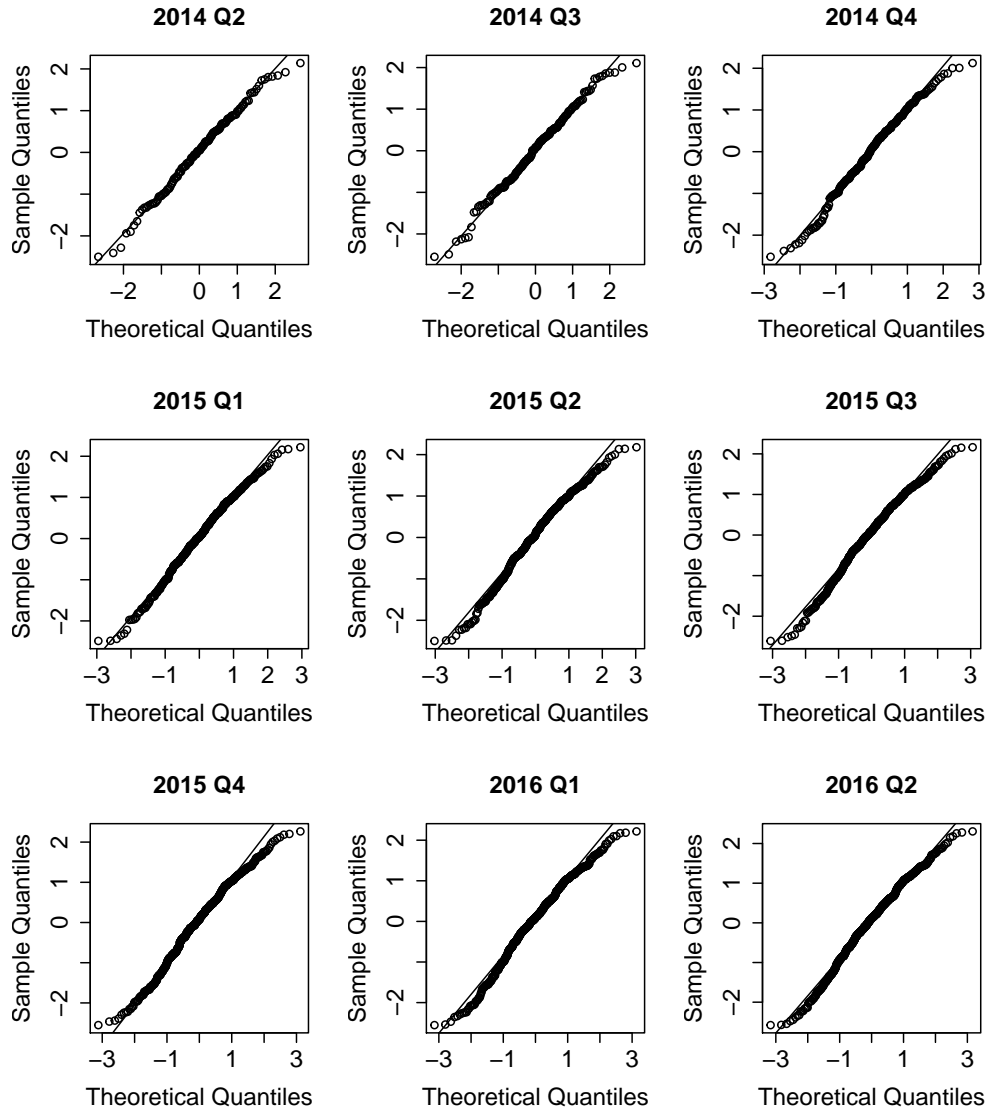


FIGURE 4.10: Q-Q plots of residuals from the Generalized Gamma regressions for each time step. The theoretical quantiles refer to a standard normal distribution.

To summarize, both the histograms and the Q-Q plots show that our Generalized Gamma regressions have well behaving residuals.

4.4 Comparison of u-data

As outlined when we introduced our models, we try to obtain approximately uniformly distributed data sets as inputs for a copula vine model. In this section, we will examine if the resulting data from our models satisfy this assumption.

For this, we look at histograms and perform the Kolmogorov-Smirnov (KS) test (see Chapter 2 for details). Figure 4.11 and 4.12 show the histograms of u-data for each model and time point. Additionally, the p-values from the KS-test are listed below each histogram. Recall that we smoothed the density of the second model STD_X,NPIT as much as possible until the point where the p-value was at 0.05. We can see that only the last two time

points for model GAMLSS,PPIT have a smaller p-value than 0.05. While the last time point comes very close (0.046), the second to last one has a p-value of only 0.03. This is problematic for the forecasting method applied later as it heavily relies on these two time points. However, we keep using these data for the next parts and will note the resulting differences later. Other than that, we conclude that in general, all the data sets are consistently estimated to be used in the second estimation step of our approach.

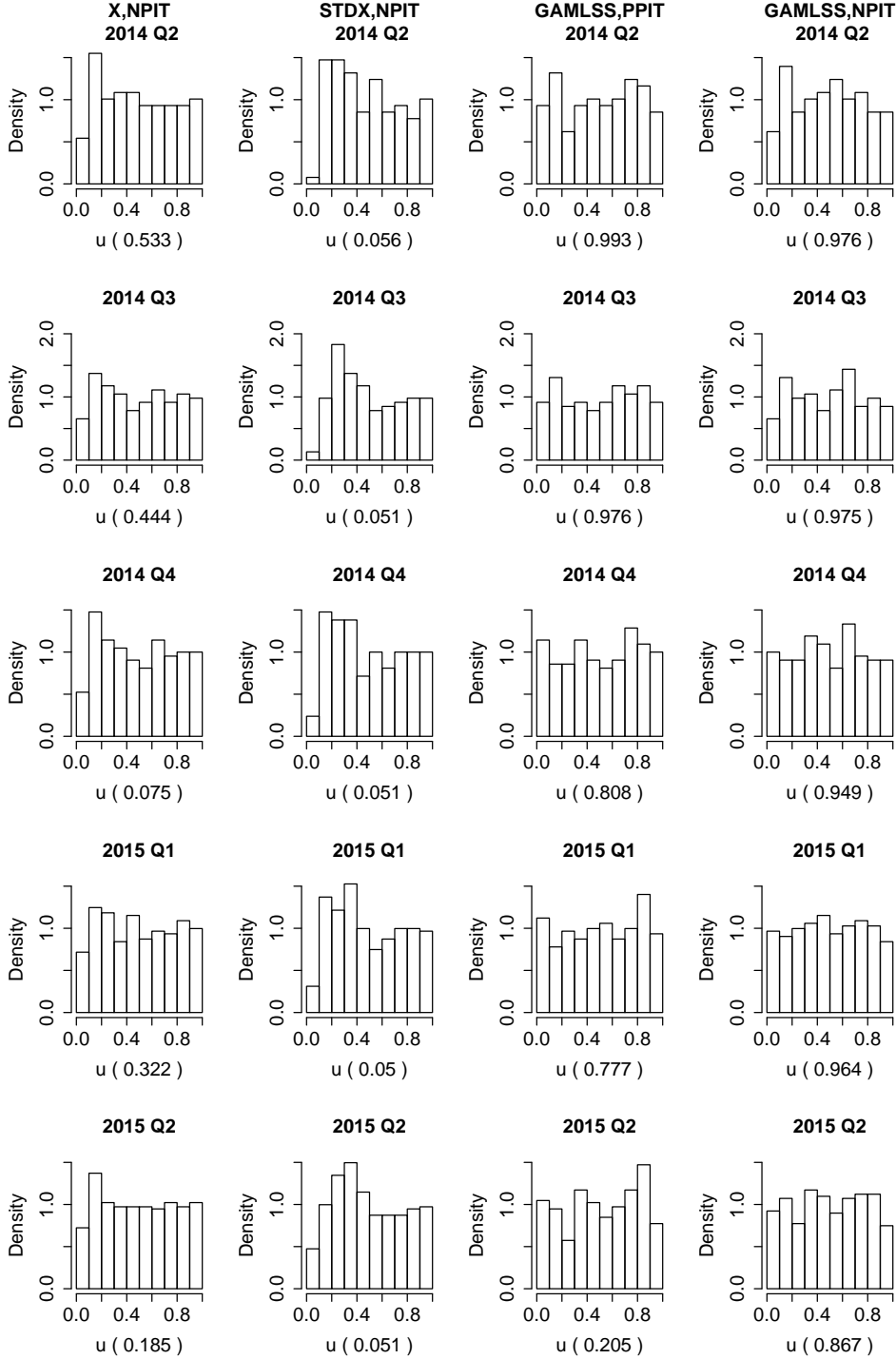


FIGURE 4.11: Histograms of u-data for each model at the first 5 time steps. Below each plot, the associated p-value from the KS-test is listed in brackets.

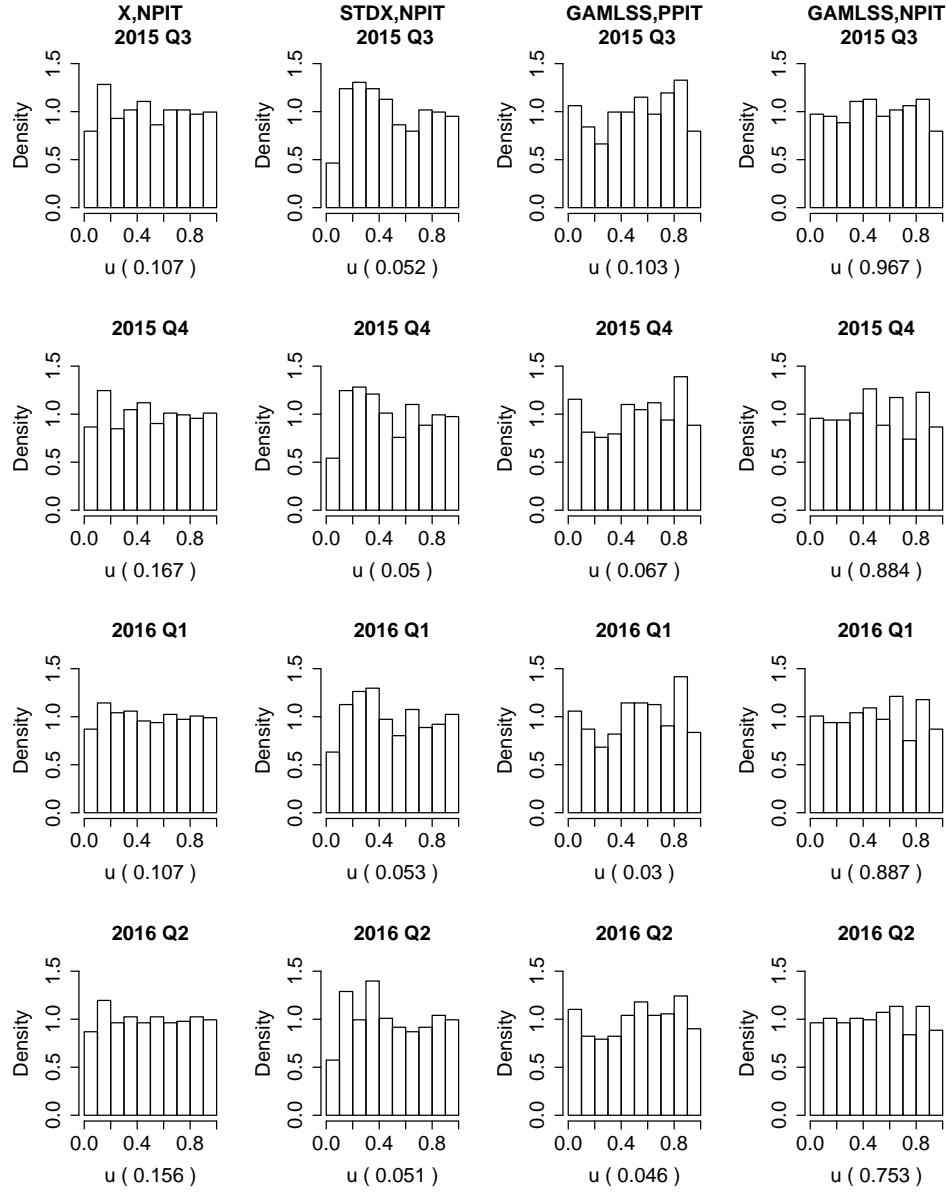


FIGURE 4.12: Histograms of u-data for each model at the last 4 time steps. Below each plot, the associated p-value from the KS-test is listed in brackets.

Chapter 5

Dependence modelling

Until now, we dealt with the models for marginal estimations and their results. In the following sections, we turn to the dependency modelling. In a first step, we explore some pairs plots before we deal with the estimation of D-vines in the main part. At the end of the chapter, we compare the different fits and take a look at the forecasting capabilities with the help of continuous rank probability scores.

5.1 Pairs plots

First, we take a look at pairs plots for each model. The plots in Figures 5.1 to 5.4 show the histograms of our u-data on the diagonal, corresponding pairwise scatter plots and the associated Kendall's τ values on the upper right as well as pairwise density normal contour plots on the lower left.

In each case, we can observe that the dependency is higher for pairs that are immediately next to each other while it decreases for pairs farther away. This makes sense, as the exposure limit for a company would not change drastically from one time point to another and is more dependant on limits from time points closer to it. This observation also confirms our focus on ordered D-vines as its structure with pairs directly next to each other in the first tree implies higher dependency for these pairs. We can also see similar contour plots for all models which tend to exhibit upper tail dependencies.

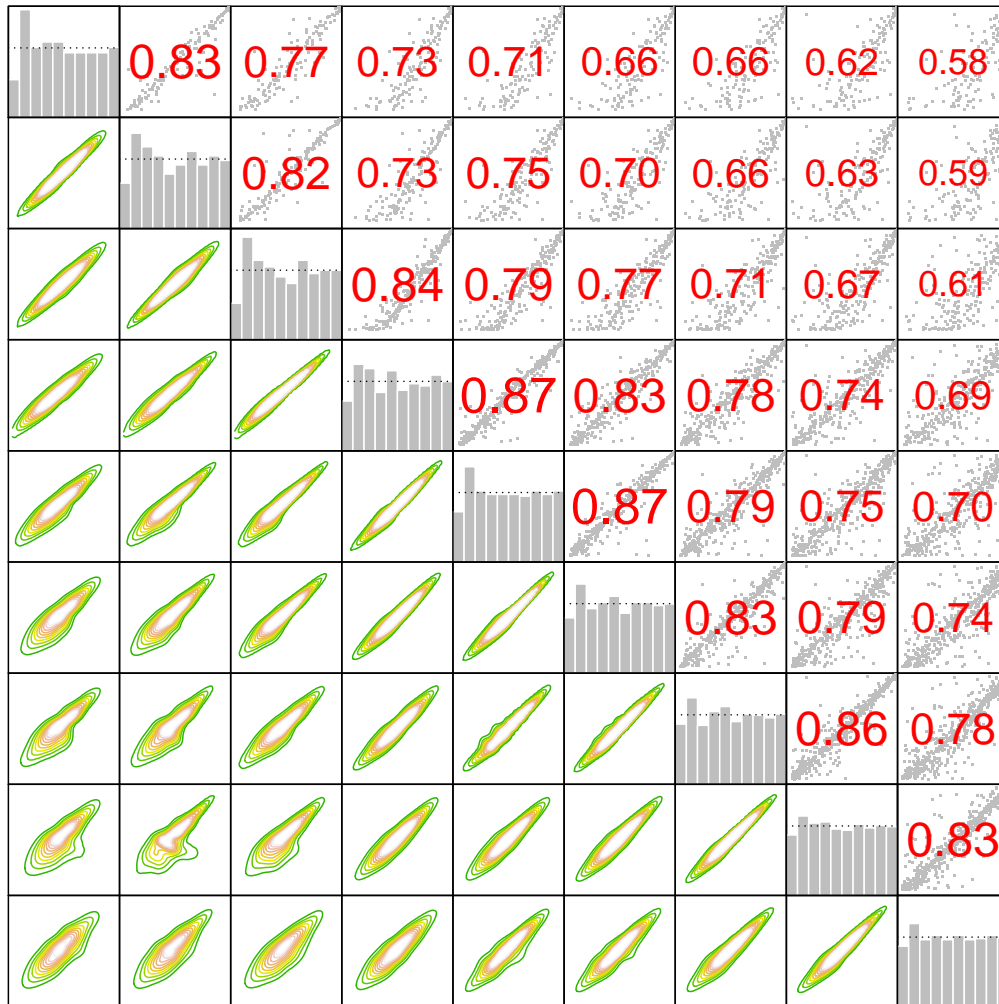
Model X_t NPIT

FIGURE 5.1: Pairs plots for the u -data from Model X_t NPIT where the limit's density is estimated non-parametrically. The data in the top left correspond to the first, the data in the bottom right to the last time point.

Model STDX,NPIT

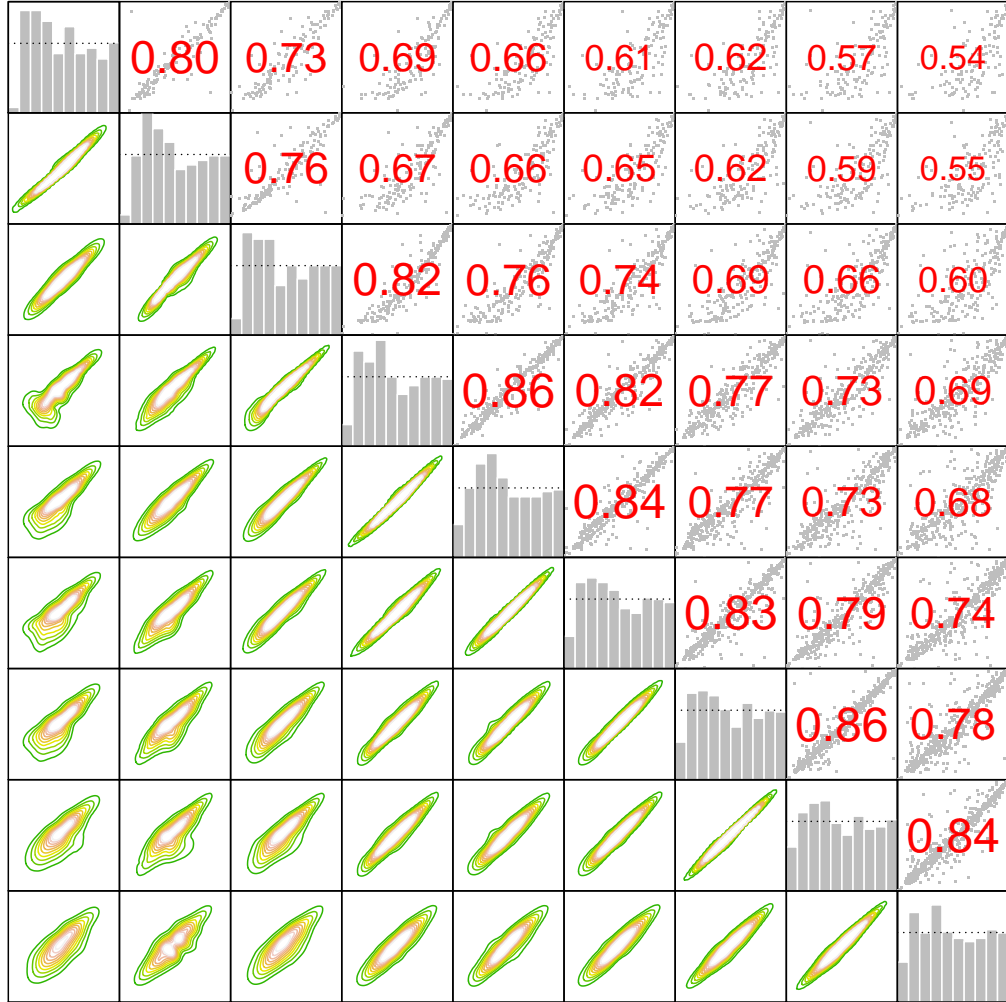


FIGURE 5.2: Pairs plots for the u-data from Model STDX,NPIT where the limit data is first standardized before its density is estimated non-parametrically. The data in the top left correspond to the first, the data in the bottom right to the last time point.

Model GAMLSS,PPIT

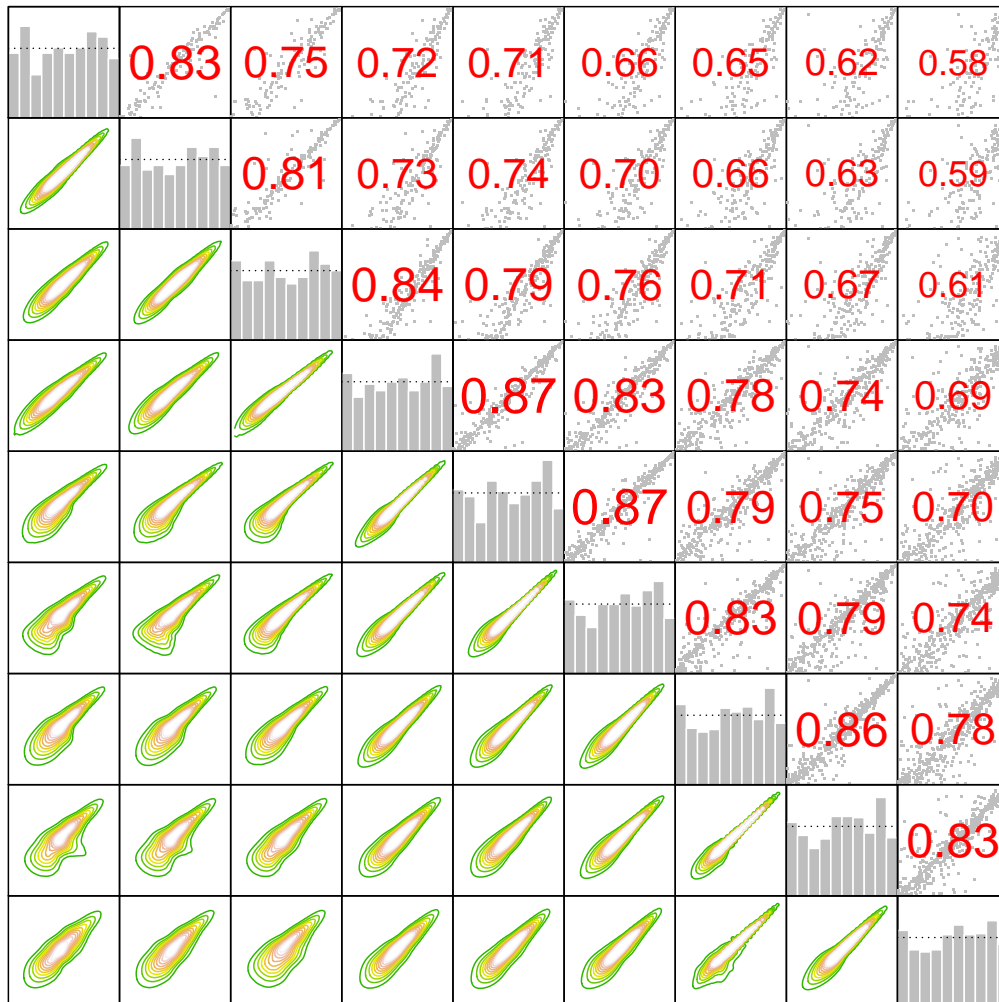


FIGURE 5.3: Pairs plots for the u-data from Model GAMLSS,PPIT where the parametric PITs of residuals from the Generalized Gamma regressions are used. The data in the top left correspond to the first, the data in the bottom right to the last time point.

Model GAMLSS,NPIT

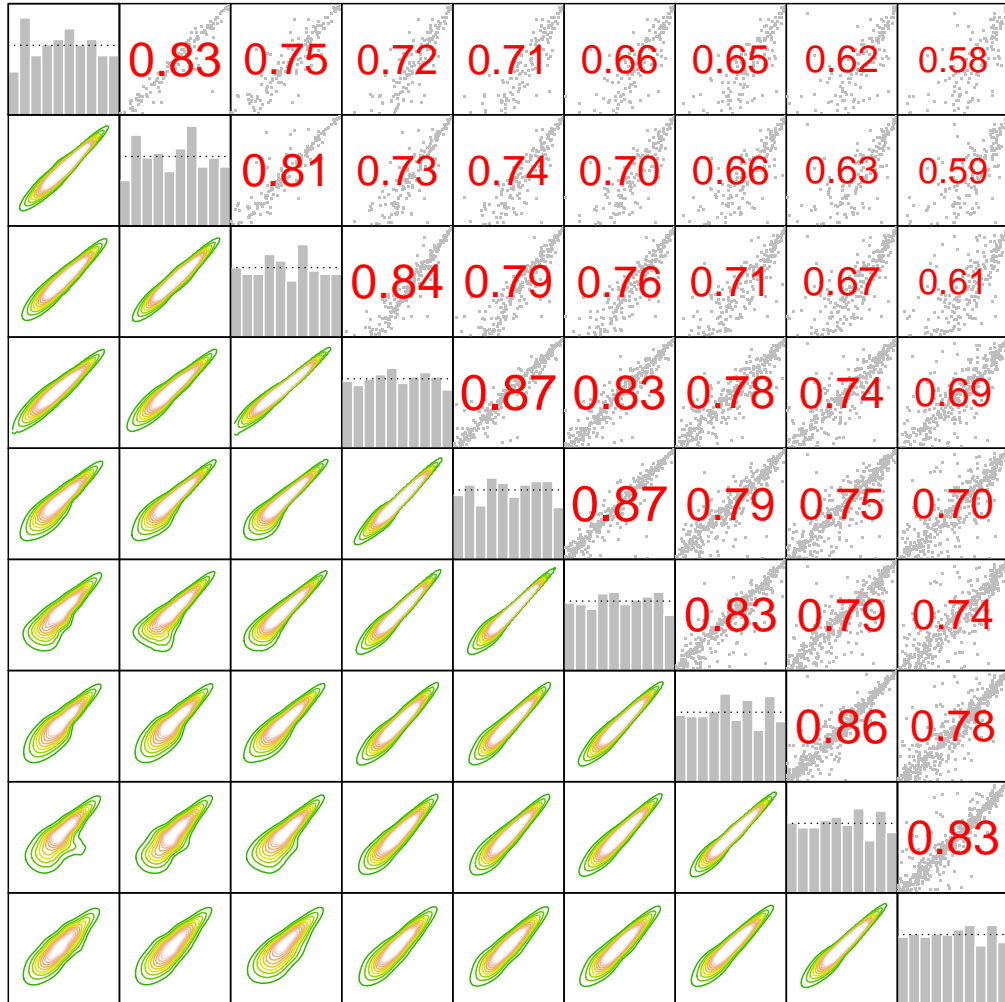


FIGURE 5.4: Pairs plots for the u-data from Model GAMLSS,NPIT where the non-parametric PITs of residuals from the Generalized Gamma regressions are used. The data in the top left correspond to the first, the data in the bottom right to the last time point.

5.2 Vine copula estimations

We now want to fit vine copula models to the different marginal u-data sets from the previous sections. As a first step, we show different estimations and then compare the results visually as well as with Vuong tests and an application of the continuous rank probability score.

Comparison of different D-vine fittings

In order to fit our data, we make use of the `VineCopula` package in R and apply the function `RVineCopSelect`. Note that we always use the BIC selection criterion in the following parts which penalizes copula families with more parameters more than the AIC criterion.

We now fit a D-vine copula model with different optimization constraints. We start with a fit that allows for all families implemented in the package, which are (including possible rotations):

- Independence,
- Gaussian,
- Student t,
- Clayton,
- Gumbel,
- Frank,
- Joe,
- BB1, BB6, BB7, BB8,
- Tawn type 1 and Tawn type 2.

Also recall the examples of copula families given in Chapter 2 for most of these. We then reduce the possible families to more ‘simple’ ones (i.e. families with only one parameter) which are Gaussian, Clayton, Gumbel, Frank and Joe copula as well as the 90, 180 and 270 degrees rotations of these and also the Independence copula. Additionally, we will run the function for these scenarios with the option to perform a hypothesis test for independence before the bivariate copula selection.

We present the results in Tables 5.1 to 5.4, where we also include information about AIC, BIC, D-vine based BIC_D and log-likelihood values as well as the number of estimated copula parameters (i.e. excluding marginal parameters). Table 5.5 shows the same information grouped by marginal model for better comparability. Note that for the BIC, the number of observations N is chosen as the number of individuals at the last time step. According to the AIC, the D-vine model with all families and no independence test is preferred for each marginal model, while the BIC prefers it only for the first two marginal models and gives a lower (better) score to the D-vine model with all families and an independence test performed in the case of both GAMLSS marginal models. The conclusion from the D-vine based BIC_D differs from the original BIC only in the case of the GAMLSS, PPIT model where it favours the case of no independence test.

The corresponding Kendall's τ values for each bivariate pair-copula are listed in brackets behind the families. For details on the estimated parameters of each copula see the tables in Appendix A.

There, we can see that all dependent copulas in trees 2 to 9 are either the independence copula or have very low Kendall's τ values. This behaviour becomes especially apparent in the two estimations where an independence test at significance level 0.05 was performed for each copula and nearly all copulas on higher trees are selected as Independence.

Thus, we truncate the vines after the first tree and assume a Markov tree dependence structure. We also note that the first tree does not change when performing an independence test, as no copula there is selected as the Independence copula, indicating a strong dependence for subsequent time points. This means that from now on we only need to consider two of the vine models: One where all families were included in the fitting and one where the list of families was reduced as detailed above.

Figures 5.5 and 5.6 show the corresponding D-vine models visualized as pairwise normal contour plots for the first tree. Contour plots for all trees can be found in Appendix A which again underline the observation that the dependent copulas in higher trees all show independence.

For the first tree, we observe two important behaviours: First, same pairs show very similar contours regardless of the underlying marginal model. And second, the dependence strength as well as the copula families don't change a lot over time. Especially the second observation is important for our approach of shifting the D-vine structure when it comes to forecasting. Because the dependence between two subsequent points stays similar from one pair to the next one, the shifting approach is appropriate in this case.

Tree	Pairs	$X, NPIT$	$STD X, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	t (0.84)	t (0.85)	Tawn T1 (0.83)	Tawn T2 (0.80)
	3,2	Tawn T2 (0.81)	t (0.80)	Tawn T2 (0.77)	Tawn T2 (0.78)
	4,3	Tawn T2 (0.62)	Tawn T2 (0.63)	Tawn T2 (0.58)	Tawn T2 (0.61)
	5,4	t (0.88)	t (0.87)	Tawn T1 (0.87)	Tawn T2 (0.84)
	6,5	t (0.90)	t (0.88)	Tawn T1 (0.89)	Tawn T1 (0.89)
	7,6	t (0.87)	t (0.87)	Tawn T1 (0.84)	Tawn T1 (0.85)
	8,7	Tawn T2 (0.85)	Tawn T2 (0.86)	Tawn T2 (0.84)	Tawn T2 (0.84)
	9,8	Tawn T2 (0.79)	Tawn T2 (0.81)	Tawn T2 (0.79)	Tawn T2 (0.79)
2	3,1 2	Ind.	t (0.03)	Ind.	Ind.
	4,2 3	Ind.	t (-0.11)	Joe 270° (-0.07)	Joe 270° (-0.06)
	5,3 4	Ind.	Ind.	BB8 90° (-0.22)	Ind.
	6,4 5	Ind.	t (0.11)	Gumbel 90° (-0.10)	Clayton 270° (-0.10)
	7,5 6	Ind.	t (-0.13)	Joe (0.04)	t (-0.07)
	8,6 7	Ind.	Ind.	BB8 (0.11)	Ind.
	9,7 8	Ind.	Gumbel (0.04)	Ind.	Ind.
3	4,1 3,2	Frank (-0.19)	Ind.	t (-0.19)	Ind.
	5,2 4,3	Gumbel (0.11)	t (0.03)	Ind.	Frank (0.21)
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	t (-0.01)	t (0.00)	Gumbel 180° (0.03)	Ind.
	8,5 7,6	Ind.	Ind.	Joe 90° (-0.05)	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Tawn T2 180° (0.08)	Ind.	Joe 270° (-0.03)
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Ind.	Joe 270° (-0.01)
	9,5 8,7,6	Ind.	Gumbel 90° (-0.10)	Ind.	Ind.
5	6,1 5,4,3,2	Ind.	t (-0.15)	Tawn T2 90° (0.00)	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Joe (0.12)	Ind.
	8,3 7,6,5,4	Ind.	Ind.	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Joe 90° (-0.17)	Clayton 90° (-0.08)	Ind.	Ind.
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	AIC	-3509.7	-3361.0	-3450.1	-3447.6
	BIC	-3413.2	-3204.6	-3316.1	-3342.0
	BIC _D	-3433.3	-3342.6	-3363.6	
	Log-likelihood	1775.8	1715.5	1755.1	1746.8
	# of Parameters	21	35	30	23

TABLE 5.1: Estimation results of D-vines with no constraints (i.e. all families included and no independence test) for each marginal model. Kendall's τ values are listed in brackets.

Tree	Pairs	$X, NPIT$	$STD_X, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	t (0.84)	t (0.85)	Tawn T1 (0.83)	Tawn T2 (0.80)
	3,2	Tawn T2 (0.81)	t (0.80)	Tawn T2 (0.77)	Tawn T2 (0.78)
	4,3	Tawn T2 (0.62)	Tawn T2 (0.63)	Tawn T2 (0.58)	Tawn T2 (0.61)
	5,4	t (0.88)	t (0.87)	Tawn T1 (0.87)	Tawn T2 (0.84)
	6,5	t (0.90)	t (0.88)	Tawn T1 (0.89)	Tawn T1 (0.89)
	7,6	t (0.87)	t (0.87)	Tawn T1 (0.84)	Tawn T1 (0.85)
	8,7	Tawn T2 (0.85)	Tawn T2 (0.86)	Tawn T2 (0.84)	Tawn T2 (0.84)
	9,8	Tawn T2 (0.79)	Tawn T2 (0.81)	Tawn T2 (0.79)	Tawn T2 (0.79)
2	3,1 2	Ind.	Ind.	Ind.	Ind.
	4,2 3	Ind.	Ind.	Joe 270° (-0.07)	Ind.
	5,3 4	Ind.	Ind.	Ind.	Ind.
	6,4 5	Ind.	t (0.11)	Ind.	Ind.
	7,5 6	Ind.	t (-0.13)	Ind.	Ind.
	8,6 7	Ind.	Ind.	BB8 (0.11)	Ind.
	9,7 8	Ind.	Ind. (0.04)	Ind.	Ind.
3	4,1 3,2	Ind.	Ind.	Ind.	Ind.
	5,2 4,3	Gumbel (0.11)	Ind.	Ind.	Frank (0.21)
	6,3 5,4	Ind.	Ind.	Frank (-0.14)	Ind.
	7,4 6,5	Ind.	Ind.	Gumbel 180° (0.03)	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Ind.	Ind.	Ind.
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Gumbel 90° (-0.1)	Ind.
	9,5 8,7,6	Ind.	Gumbel 90° (-0.10)	Ind.	Ind.
5	6,1 5,4,3,2	Ind.	t (-0.15)	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Joe (0.11)	Ind.
	8,3 7,6,5,4	Ind.	Ind.	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Ind.	Ind.	Ind.	Ind.
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
AIC		-3483.2	-3302.4	-3424.7	-3439.6
BIC		-3405.1	-3199.7	-3322.0	-3361.5
BIC _D		-3420.6	-3217.6	-3339.9	-3377.0
Log-likelihood		1758.6	1674.2	1735.4	1736.8
# of Parameters		17	23	23	17

TABLE 5.2: Estimation results of D-vines for all marginal models with all families included during estimation and a hypothesis test for independence applied before selecting a copula (Independence copula is chosen if null hypothesis of independence can't be rejected at significance level of 0.05).

Kendall's τ values are listed in brackets.

Tree	Pairs	$X, NPIT$	$STD\bar{X}, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	Gumbel (0.82)	Gumbel (0.82)	Gumbel (0.82)	Gumbel (0.83)
	3,2	Gumbel (0.82)	Gumbel (0.77)	Gumbel (0.79)	Gumbel (0.81)
	4,3	Clayton 180° (0.61)	Gumbel (0.68)	Clayton 180° (0.60)	Clayton 180° (0.63)
	5,4	Gumbel (0.87)	Gumbel (0.87)	Joe (0.84)	Gumbel (0.87)
	6,5	Gumbel (0.88)	Gumbel (0.87)	Joe (0.86)	Joe (0.86)
	7,6	Gumbel (0.84)	Gumbel (0.84)	Gumbel (0.83)	Gumbel (0.83)
	8,7	Gumbel (0.87)	Gumbel (0.88)	Joe (0.85)	Joe (0.85)
	9,8	Gumbel (0.83)	Gumbel (0.84)	Joe (0.80)	Gumbel (0.83)
2	3,1 2	Ind.	Gumbel (0.12)	Ind.	Ind.
	4,2 3	Ind.	Ind.	Gumbel 270° (-0.11)	Frank (-0.14)
	5,3 4	Ind.	Ind.	Frank (-0.14)	Ind.
	6,4 5	Ind.	Joe (0.08)	Joe 90° (-0.1)	Joe 90° (-0.11)
	7,5 6	Ind.	Frank (-0.23)	Joe (0.03)	Frank (-0.14)
	8,6 7	Joe 180° (0.01)	Joe 180° (0.01)	Joe 180° (0.01)	Joe 180° (0.01)
	9,7 8	Frank (-0.10)	Frank (-0.10)	Ind.	Frank (-0.13)
3	4,1 3,2	Frank (-0.16)	Frank (-0.16)	Ind.	Ind.
	5,2 4,3	Gumbel (0.06)	Joe(0.04)	Gumbel (0.05)	Gaussian (0.17)
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	Ind.	Ind.	Ind.	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Ind.	Joe 270° (-0.02)	Gumbel 270° (-0.06)
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Joe 90° (-0.15)	Clayton 270°(-0.11)	Ind.	Joe 90° (-0.18)
	8,4 7,6,5	Ind.	Ind.	Gumbel 270° (-0.02)	Gumbel 270° (-0.02)
	9,5 8,7,6	Ind.	Frank (-0.14)	Gaussian (-0.08)	Gumbel 90° (-0.06)
5	6,1 5,4,3,2	Gumbel 270° (-0.05)	Gumbel 90° (-0.11)	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Gumbel 90° (-0.13)	Ind.	Ind.
	8,3 7,6,5,4	Clayton 90° (-0.30)	Gumbel 270° (-0.22)	Ind.	Frank (-0.29)
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Gumbel 90° (-0.12)	Clayton 90° (-0.17)	Frank (-0.30)	Gumbel 270° (-0.22)
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Frank (-0.27)	Frank (-0.16)	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	AIC	-3205.6	-3014.5	-3285.2	-3233.7
	BIC	-3127.5	-2916.2	-3204.7	-3141.8
	BIC _D	-3151.9	-2936.8	-3220.0	-3160.9
	Log-likelihood	1619.8	1529.3	1660.6	1636.9
	# of Parameters	17	22	18	20

TABLE 5.3: Estimation results of D-vines for all marginal models with only selected families (see beginning of this section) included for estimation and no independence test.

Kendall's τ values are listed in brackets.

Tree	Pairs	$X, NPIT$	$STD\mathbf{X}, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	Gumbel (0.82)	Gumbel (0.82)	Gumbel (0.82)	Gumbel (0.83)
	3,2	Gumbel (0.82)	Gumbel (0.77)	Gumbel (0.79)	Gumbel (0.81)
	4,3	Clayton 180° (0.61)	Gumbel (0.68)	Clayton 180° (0.60)	Clayton 180° (0.63)
	5,4	Gumbel (0.87)	Gumbel (0.87)	Joe (0.84)	Gumbel (0.87)
	6,5	Gumbel (0.88)	Gumbel (0.87)	Joe (0.86)	Joe (0.86)
	7,6	Gumbel (0.84)	Gumbel (0.84)	Gumbel (0.83)	Gumbel (0.83)
	8,7	Gumbel (0.87)	Gumbel (0.88)	Joe (0.85)	Joe (0.85)
	9,8	Gumbel (0.83)	Gumbel (0.84)	Joe (0.80)	Gumbel (0.83)
2	3,1 2	Ind.	Ind.	Ind.	Ind.
	4,2 3	Ind.	Ind.	Ind.	Ind.
	5,3 4	Ind.	Ind.	Ind.	Ind.
	6,4 5	Ind.	Joe (0.08)	Ind.	Ind.
	7,5 6	Ind.	Frank (-0.23)	Ind.	Frank (-0.14)
	8,6 7	Ind.	Ind.	Joe 180° (0.01)	Ind.
	9,7 8	Ind.	Ind.	Ind.	Ind.
3	4,1 3,2	Ind.	Ind.	Ind.	Frank (-0.18)
	5,2 4,3	Gumbel (0.06)	Ind.	Ind.	Gumbel (0.10)
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	Ind.	Ind.	Ind.	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Ind.	Ind.	Ind.
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Ind.	Ind.
	9,5 8,7,6	Ind.	Frank (-0.12)	Ind.	Ind.
5	6,1 5,4,3,2	Ind.	Ind.	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Ind.	Ind.
	8,3 7,6,5,4	Ind.	Gumbel 270° (-0.23)	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Ind.	Ind.	Ind.	Ind.
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
AIC		-3148.3	-2936.0	-3246.2	-3195.5
BIC		-3107.0	-2882.4	-3206.0	-3145.0
BIC _D		-3115.5	-2891.1	-3212.3	-3155.8
Log-likelihood		1583.2	1480.0	1632.1	1608.7
# of Parameters		9	12	9	11

TABLE 5.4: Estimation results of D-vines for all marginal models with only selected families (see beginning of this section) included for estimation and a hypothesis test for independence applied before selecting a copula (Independence copula is chosen if null hypothesis of independence can't be rejected at significance level of 0.05). Kendall's τ values are listed in brackets.

X,NPIT				
	All families no ind. test	All families with ind. test	Selected families no ind. test	Selected families with ind. test
AIC	-3509.7	-3483.2	-3205.6	-3148.3
BIC	-3413.2	-3405.1	-3127.5	-3107.0
BIC _D	-3433.3	-3420.6	-3151.9	-3115.5
LL	1775.8	1758.6	1619.8	1583.2
Para.	21	17	17	9

STDN,NPIT				
	All families no ind. test	All families with ind. test	Selected families no ind. test	Selected families with ind. test
AIC	-3361.0	-3302.4	-3014.5	-2936.0
BIC	-3204.6	-3199.7	-2916.2	-2882.4
BIC _D	-3237.4	-3217.6	-2936.8	-2891.1
LL	1715.5	1674.2	1529.3	1480.0
Para.	35	23	22	12

GAMLSS,PPIT				
	All families no ind. test	All families with ind. test	Selected families no ind. test	Selected families with ind. test
AIC	-3450.1	-3424.7	-3285.2	-3246.2
BIC	-3316.1	-3322.0	-3204.7	-3206.0
BIC _D	-3342.6	-3339.9	-3220.0	-3212.3
LL	1755.1	1735.4	1660.6	1632.1
Para.	30	23	18	9

GAMLSS,NPIT				
	All families no ind. test	All families with ind. test	Selected families no ind. test	Selected families with ind. test
AIC	-3447.6	-3439.6	-3233.7	-3195.5
BIC	-3342.0	-3361.5	-3141.8	-3145.0
BIC _D	-3363.6	-3377.0	-3160.9	-3155.8
LL	1746.8	1736.8	1636.9	1608.7
Para.	23	17	20	11

TABLE 5.5: Tables showing the AIC, BIC, D-vine based BIC_D, log-likelihood (LL) and number of copula parameters (Para.) for each marginal model and D-vine model.

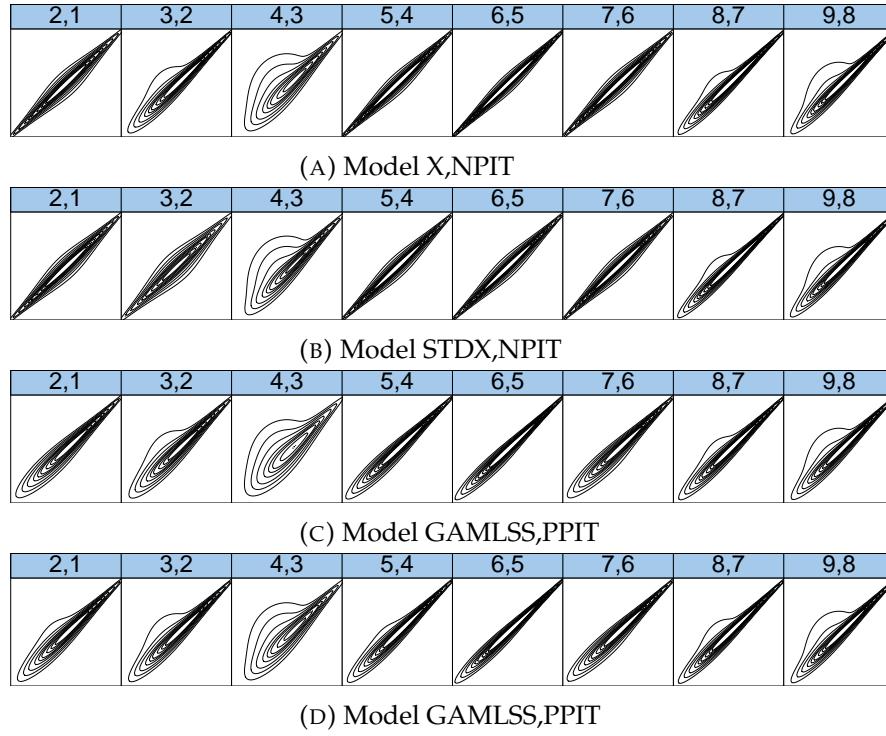


FIGURE 5.5: Normal contour plots of the first trees for each marginal model with all families allowed. The numbers correspond to the pairs in the first tree.

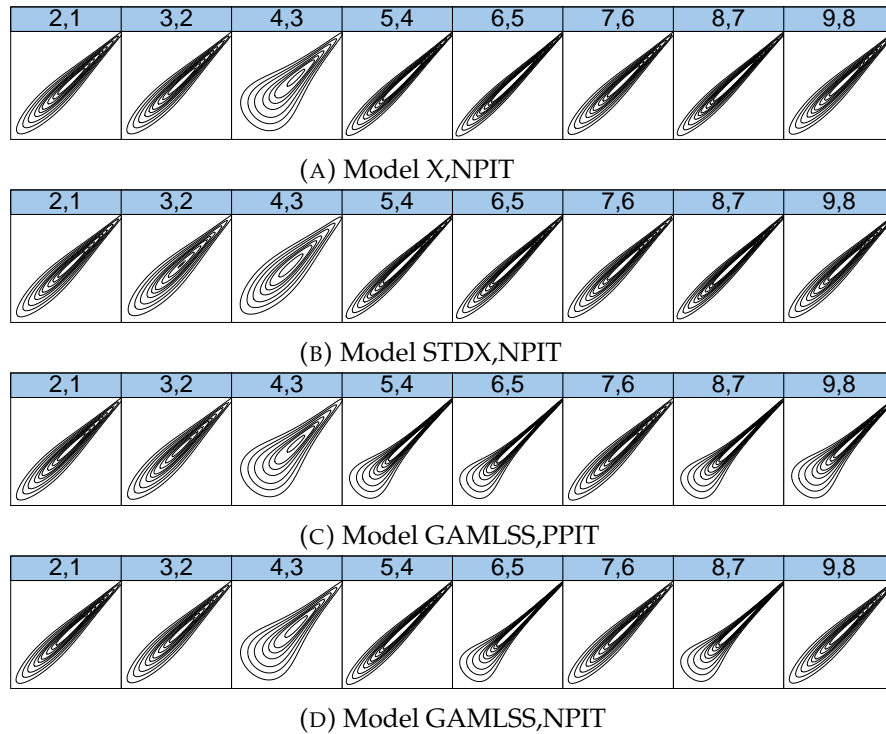


FIGURE 5.6: Normal contour plots of the first trees for each marginal model with only selected families allowed. The numbers correspond to the pairs in the first tree.

Vuong tests to compare D-vine copula fits

After we have seen the different vine model fittings and compared them visually in the last section, we now want to perform the so-called Vuong test (see Chapter 2) to determine which model is preferable. For this, we apply the `RVineVuongTest` function from the `VineCopula` package. For each marginal model, we compare the D-vine model without independence test to the one with independence test, once for the case of all families allowed and once for the case with only selected families.

The results are shown in Tables 5.6 and 5.7 where we always compare the case of no independence test to the one with independence test. We use the Schwarz test statistic to adjust for model complexity and find that at the 0.05 significance level, there is no difference between the D-vine models that were estimated with the independence tests and the ones without.

	X _i NPIT	STD _X NPIT	GAMLSS PPIT	GAMLSS NPIT
Statistic	3.51	2.61	2.53	1.25
p-value	0.00	0.01	0.01	0.21
Statistic Akaike	2.70	1.85	1.63	0.50
p-value Akaike	0.01	0.06	0.10	0.62
Statistic Schwarz	1.53	0.77	0.35	-0.57
p-value Schwarz	0.13	0.44	0.73	0.57

TABLE 5.6: Results of the Vuong tests of the vine models with all families allowed, for each marginal model. The vine models without an independence test performed are compared to the ones with an independence test.

	X _i NPIT	STD _X NPIT	GAMLSS PPIT	GAMLSS NPIT
Statistic	2.93	2.45	2.50	3.39
p-value	0.00	0.01	0.01	0.00
Statistic Akaike	2.91	1.95	1.71	2.31
p-value Akaike	0.02	0.05	0.09	0.02
Statistic Schwarz	1.38	1.24	0.58	0.75
p-value Schwarz	0.17	0.22	0.56	0.45

TABLE 5.7: Results of the Vuong tests of the vine models with only selected families allowed, for each marginal model. The vine models without an independence test performed are compared to the ones with an independence test.

Thus, we test the D-vine models with all families allowed to the ones without next. Table 5.8 shows the results. Regardless of the test statistic, the tests indicate a significantly better fit for the model with all families allowed at the 0.05 level, despite the use of more parameters.

Note that, as the Vuong test needs observations for all copulas that are part of the vine, we can only use the limited number of complete observations from the first to the last time point.

	X,NPIT	STDX,NPIT	GAMLSS PPIT	GAMLSS NPIT
Statistic	5.27	4.47	3.34	4.08
p-value	0.00	0.00	0.00	0.00
Statistic Akaike	5.03	4.22	2.89	3.89
p-value Akaike	0.00	0.00	0.00	0.00
Statistic Schwarz	4.69	3.85	2.24	3.62
p-value Schwarz	0.00	0.00	0.03	0.00

TABLE 5.8: Results of the Vuong tests of the vine models, for each marginal model. The vine models with all families allowed are compared to the ones with only selected families.

Continuous rank probability scores of predictions

In order to compare the forecasting capabilities, we now apply the continuous rank probability score (CRPS) to our methods. We do so for both of the D-vine fittings and each underlying marginal model.

Recall for the following that we apply a shifting approach, as detailed at the end of Chapter 2, to obtain a dependence structure between the current and future time point and that we also shift the marginal distribution from the current to the next time point.

Also note that we use the truncated versions of vines and thus only have to include copula families of the first tree.

Thus, we now compute Equation 2.9 which is given as

$$\begin{aligned}
 \text{CRPS}_i & (F_{d+1|d}(x_{i,d+1}|x_{i,d}), x_{i,d+1}) \\
 &= \frac{1}{m-1} \sum_{i=1}^{m-1} \left| F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{i}{m} \middle| u_d \right) \right) - x_{i,d+1} \right| \\
 &- \frac{1}{2(m-1)^2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \left| F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{i}{m} \middle| u_d \right) \right) - F_{d+1}^{-1} \left(C_{d+1|d}^{-1} \left(\frac{j}{m} \middle| u_d \right) \right) \right|,
 \end{aligned}$$

for an individual observation $x_{i,d}$ with $i = 1, \dots, N_d$ at time point d .

In our case, the last known time point is 2016 Q2, i.e. $d = 9$. The conditional bivariate pair-copula needed for $C_{d+1|d}^{-1}(\cdot|\cdot)$ is the conditional pair-copula $C_{d-1|d}(\cdot|\cdot)$ from one time step before and the marginal distribution F_{d+1} is also given as the marginal distribution of time d , i.e. F_d . Recall that we also have the true values of exposure limits available for the new time point 2016 Q3, i.e. $x_{i,d+1}$ for all $i = 1, \dots, N_t$.

We set $m = 100$ to use 99 conditional quantiles for each individual CRPS value and obtain the overall CRPS by averaging over each individual CRPS.

The resulting values are listed in Table 5.9. The best (i.e. lowest) score is attained by the ordered D-vine copula based on the limited list of families and the u-data from the X,NPIT marginal model while the STDX,NPIT marginal model has the worst scores for both D-vine models. Recall that the underlying u-data from the second marginal model were also problematic regarding the uniform distribution assumption, which could explain the worse performance.

	X,NPIT	STDX,NPIT	GAMLSS PPIT	GAMLSS NPIT
All Families	13,210,221	39,813,604	14,706,204	13,534,497
Limited Families	12,465,080	36,491,709	12,725,194	12,831,170

TABLE 5.9: Continuous rank probability scores for both considered D-vine specifications and each marginal model.

In Figure 5.7 (and Figure 5.8 in more detail without the extreme outliers), the boxplots of CRPS values are displayed for each model. We observe that the difference in scores is mostly driven by larger outliers, while the medians are very similar, except for the STDX,NPIT marginal model. Figure 5.9 and 5.10 again show boxplots of the CRPS values, this time one for values from the "Main" group and one for the "Other" group. We can see that medians remain similar between groups but most outliers come from values in the "Main" group, implying a better forecast quality for data from the "Other" group. Note, however, that the number of observations from the second group is also much smaller than in the first one.

As an alternative way to display the differences in CRPS values, Figure 5.11 shows the histograms of CRPS values for each model. Again, we can see that, except for STDX,NPIT, all models have a very similar distribution of values.

While these results show us which model is preferred, the CRPS only gives us relative values to compare models to each other. In order to check whether our models are actually useful for predictions, we compare them to two benchmark cases: First, we use the Independence copula instead of our fitted copula families to compute CRPS values and second, we apply a Gauss copula. For the Gauss copula, the parameters used are estimated the same way as before by allowing only the Gauss family as a possible copula in the estimation process. The results are listed in table 5.10. We can see that all values are larger than the lowest score from our models (12,465,080), implying that our best model is actually a better fit for forecasting. However, we also note that our previous models with the STDX,NPIT marginal model, which are the worst performers according to the CRPS values, also have higher scores than the Gauss model.

Thus, the dependence structure for our data is neither independent nor shows a Gaussian behaviour and we conclude that a more complex dependence modelling approach, such as our copula models, is indeed necessary.

	X,NPIT	STDX,NPIT	GAMLSS PPIT	GAMLSS NPIT
Independence	64,799,020	70,891,560	64,941,686	64,820,170
Gauss	15,178,996	33,965,212	19,069,157	16,714,786

TABLE 5.10: Continuous rank probability scores for the benchmark cases with the Independence copula and a Gauss copula for each marginal model.

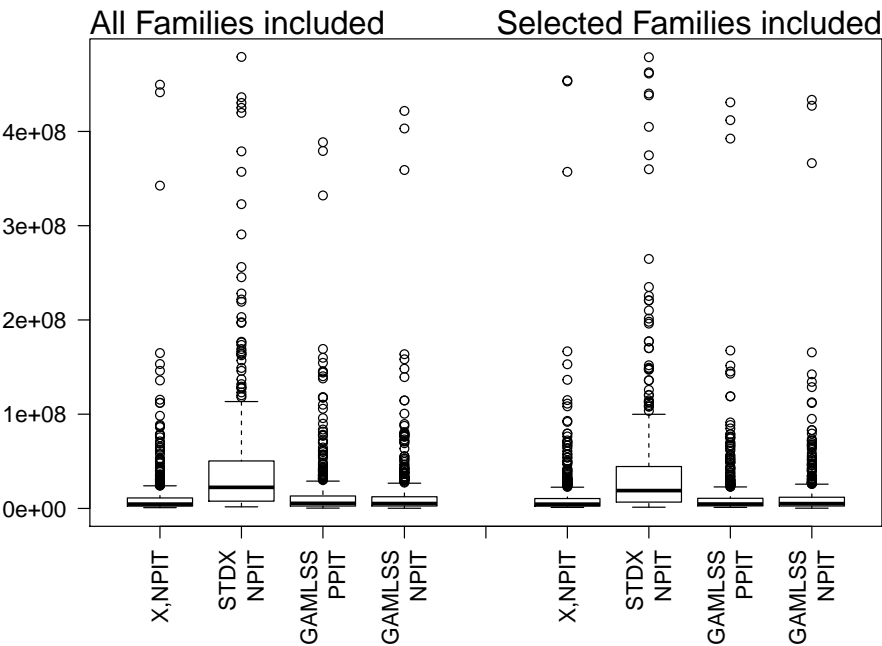


FIGURE 5.7: Boxplot of CRPS values for each marginal and vine model.

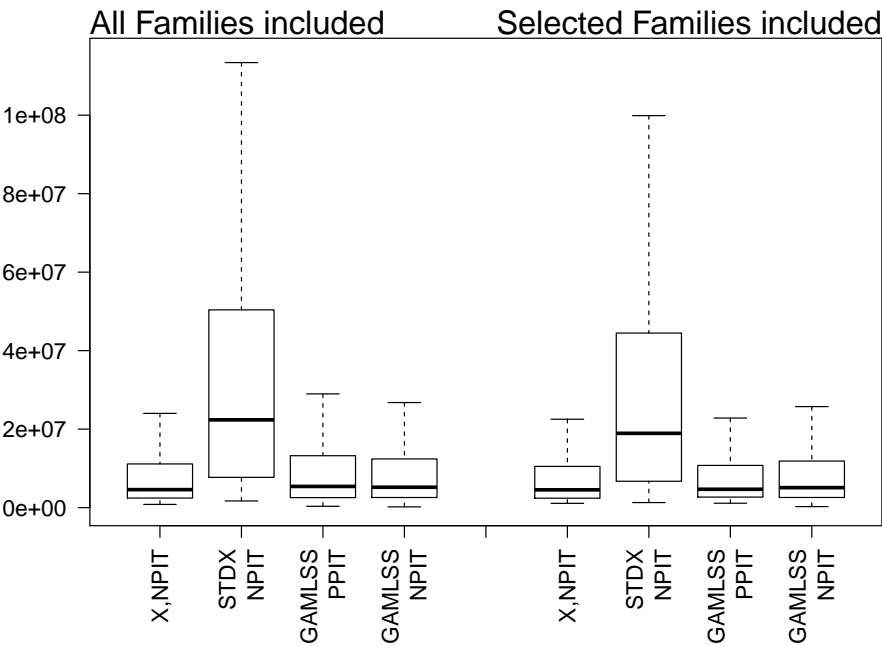


FIGURE 5.8: Boxplot of CRPS values for each marginal and vine model with outliers removed for better comparability.

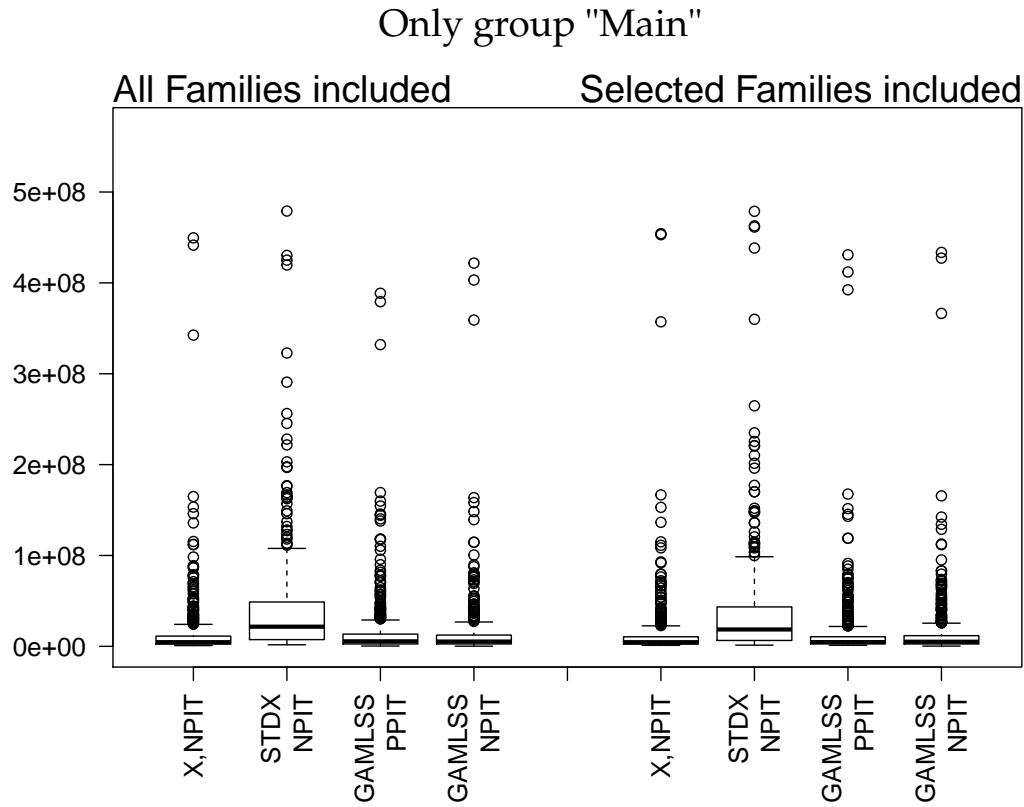


FIGURE 5.9: Boxplot of CRPS values for each marginal and vine model of data from group "Main".

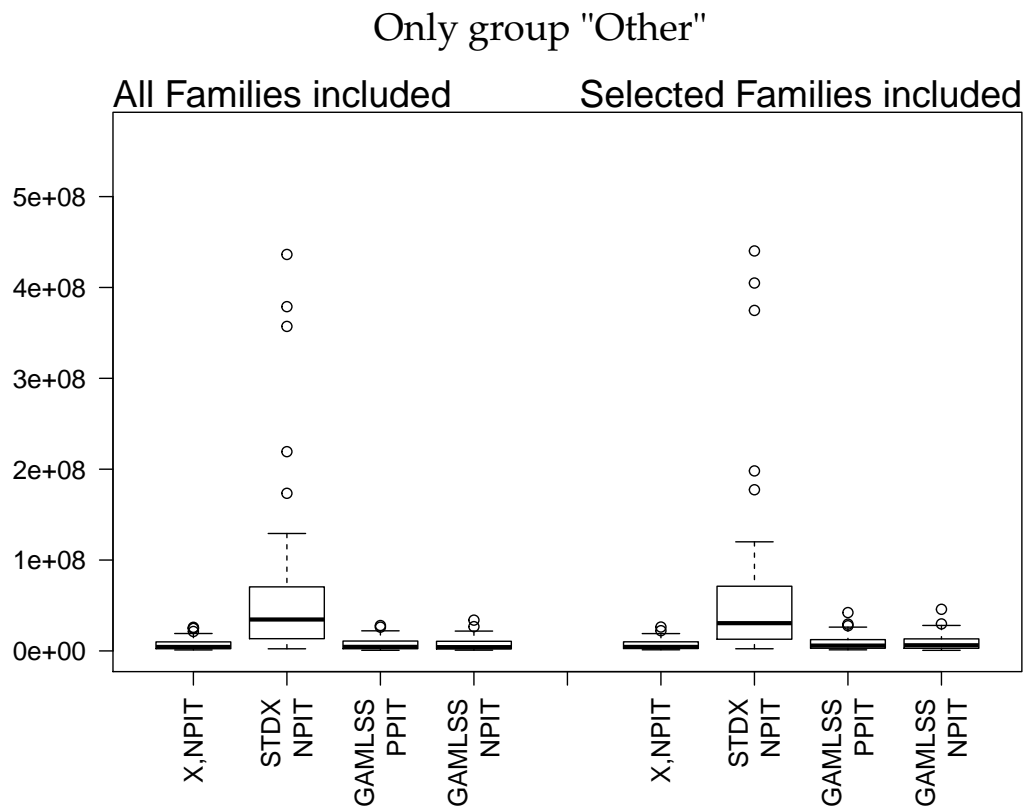


FIGURE 5.10: Boxplot of CRPS values for each marginal and vine model of data from group "Other".

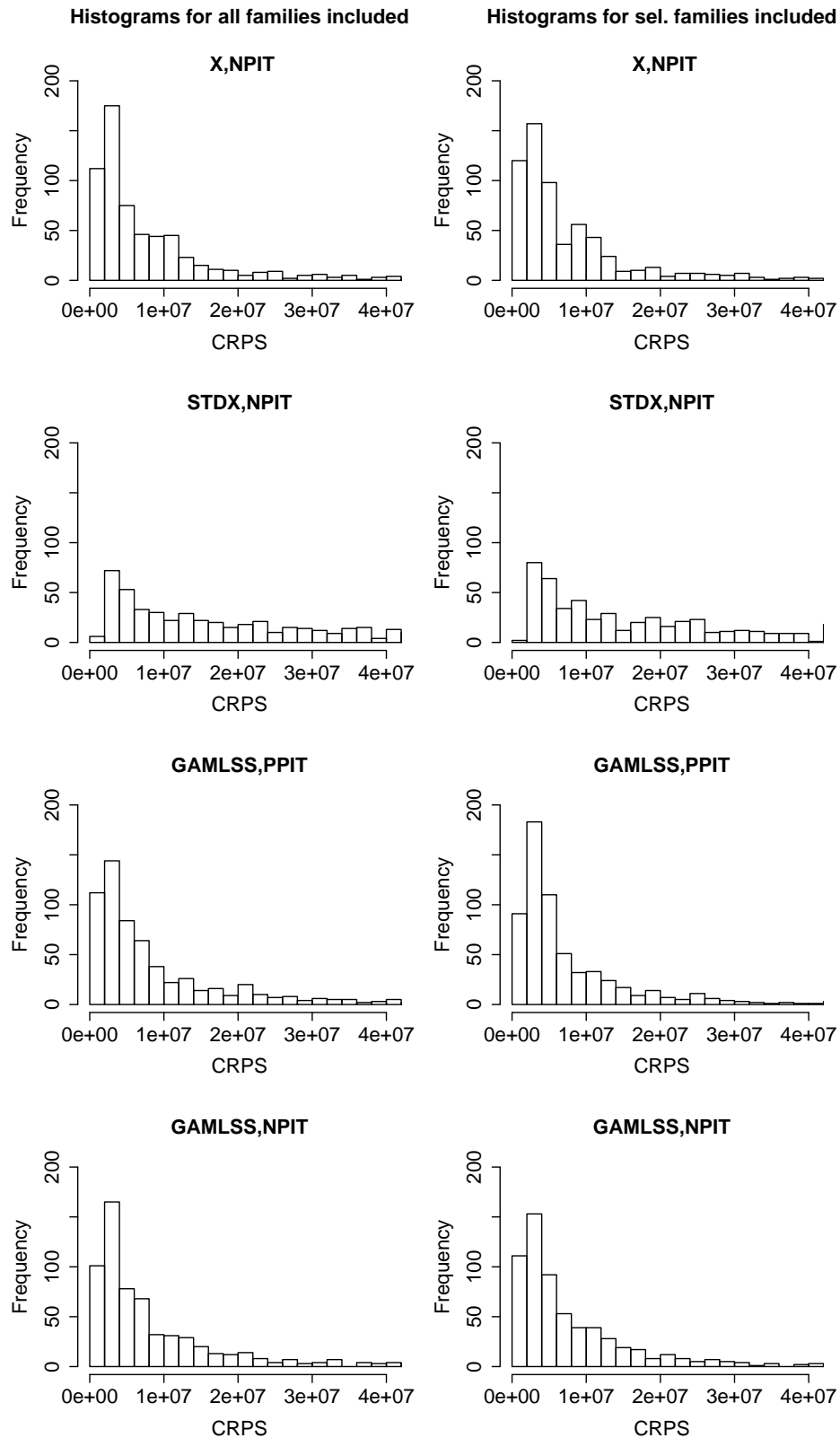


FIGURE 5.11: Histograms of CRPS values for each marginal and vine model cut off at 4×10^7 .

Chapter 6

Conclusion

In this thesis, we concerned ourselves with the task of modelling and forecasting exposure limit data with the help of D-vine copulas.

We discussed the necessary theory and applied different methods and models to our data, regarding both the modelling of marginals as well as the dependence structure. In the end, we compared the models to benchmarks and also among each other. While not all marginal models turned out to work well, most of them outperformed the benchmarks of Independence and Gaussian copula models. In particular, we found that the modelling of margins using just a KDE of the data combined with a D-vine structure based on simple, single parameter bivariate copula families is the preferred model. This also implies that a different behaviour between regions is not very relevant. Aside from the important observation that the dependence structure was not a Gaussian one and thus showed the necessity of a more complex approach such as our D-vine copula models, we also observed that the dependence structure in the data exhibits a Markov property, i.e. the conditioned dependences between time points farther away is negligible.

When working with this kind of exposure data, we thus showed that a rather easy to implement model can forecast the data very well when applying our proposed approach to shift the structure, despite the difficulties of an unbalanced setting and truncated data.

The results can thus be used to get a good estimation of future exposure limits which is an important parameter for tasks such as risk capital allocation and pricing of reinsurance contracts.

Appendix A

Extended D-vine estimation results and contour plots

In this Appendix, Tables A.1 to A.4 show the complete estimation results for the D-vine fittings of Section 5.2, including estimated parameters for each pair-copula. Figures A.1 to A.16 show the corresponding contour plots for all pair-copulas.

Tree	Pairs	$X, NPIT$	$STD\bar{X}, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	t (0.84) par = 0.97, par2 = 2	t (0.85) par = 0.97, par2 = 2	Tawn T1 (0.83) par = 6.33, par2 = 0.99	Tawn T2 (0.80) par = 6.75, par2 = 0.93
	3,2	Tawn T2 (0.81) par = 7.76, par2 = 0.92	t (0.80) par = 0.95, par2 = 2	Tawn T2 (0.77) par = 6.39, par2 = 0.9	Tawn T2 (0.78) par = 7.18, par2 = 0.9
	4,3	Tawn T2 (0.62) par = 4.96, par2 = 0.74	Tawn T2 (0.63) par = 6, par2 = 0.72	Tawn T2 (0.58) par = 3.97, par2 = 0.74	Tawn T2 (0.61) par = 4.86, par2 = 0.74
	5,4	t (0.88) par = 0.98, par2 = 2	t (0.87) par = 0.98, par2 = 2	Tawn T1 (0.87) par = 8.2, par2 = 0.99	Tawn T2 (0.84) par = 8.28, par2 = 0.95
	6,5	t (0.90) par = 0.99, par2 = 2	t (0.88) par = 0.98, par2 = 2	Tawn T1 (0.89) par = 10.19, par2 = 0.98	Tawn T1 (0.89) par = 10.5, par2 = 0.98
	7,6	t (0.87) par = 0.98, par2 = 2	t (0.87) par = 0.98, par2 = 2	Tawn T1 (0.84) par = 7.21, par2 = 0.97	Tawn T1 (0.85) par = 7.52, par2 = 0.98
	8,7	Tawn T2 (0.85) par = 10.26, par2 = 0.94	Tawn T2 (0.86) par = 10.92, par2 = 0.94	Tawn T2 (0.84) par = 8.99, par2 = 0.94	Tawn T2 (0.84) par = 9.91, par2 = 0.93
	9,8	Tawn T2 (0.79) par = 9.3, par2 = 0.88	Tawn T2 (0.81) par = 9.48, par2 = 0.89	Tawn T2 (0.79) par = 8.48, par2 = 0.88	Tawn T2 (0.79) par = 8.88, par2 = 0.88
2	3,1 2	Ind.	t (0.03) par = 0.05, par2 = 3.04	Ind.	Ind.
	4,2 3	Ind.	t (-0.11) par = -0.17, par2 = 3.85	Joe 270° (-0.07) par = -1.14	Joe 270° (-0.06) par = -1.1
	5,3 4	Ind.	Ind.	BB8 90° (-0.22) par = -1.84, par2 = -0.89	Ind.
	6,4 5	Ind.	t (0.11) par = 0.17, par2 = 3.96	Gumbel 90° (-0.10) par = -1.11	Clayton 270° (-0.10) par = -0.21
	7,5 6	Ind.	t (-0.13) par = -0.21, par2 = 4.05	Joe (0.04) par = 1.07	t (-0.07) par = -0.11, par2 = 6.77
	8,6 7	Ind.	Ind.	BB8 (0.11) par = 1.31, par2 = 0.95	Ind.
	9,7 8	Ind.	Gumbel (0.04) par = 1.04	Ind.	Ind.
3	4,1 3,2	Frank (-0.19) par = -1.79	Ind.	t (-0.19) par = -0.3, par2 = 30	Ind.
	5,2 4,3	Gumbel (0.11) par = 1.13	t (0.03) par = 0.05, par2 = 3.04	Ind.	Frank (0.21) par = 2
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	t (-0.01) par = -0.01, par2 = 5.57	t (0.00) par = 0, par2 = 4.26	Gumbel 180° (0.03) par = 1.03	Ind.
	8,5 7,6	Ind.	Ind.	Joe 90° (-0.05) par = -1.09	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Tawn T2 180° (0.08) par = 4.93, par2 = 0.08	Ind.	Joe 270° (-0.03) par = -1.06
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Ind.	Joe 270° (-0.01) par = -1.01
	9,5 8,7,6	Ind.	Gumbel 90° (-0.10) par = -1.12	Ind.	Ind.
5	6,1 5,4,3,2	Ind.	t (-0.15) par = -0.23, par2 = 4.17	Tawn T2 90° (0.00) par = -17.94, par2 = 0	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Joe (0.12) par = 1.25	Ind.
	8,3 7,6,5,4	Ind.	Ind.	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Joe 90° (-0.17) par = -1.37	Clayton 90° (-0.08) par = -0.17	Ind.	Ind.
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.

TABLE A.1: Complete results of D-vines with all families included and no independence test.

Tree	Pairs	$X, NPIT$	$STD\bar{X}, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	t (0.84) par = 0.97, par2 = 2	t (0.85) par = 0.97, par2 = 2	Tawn T1 (0.83) par = 6.33, par2 = 0.99	Tawn T2 (0.80) par = 6.75, par2 = 0.93
	3,2	Tawn T2 (0.81) par = 7.76, par2 = 0.92	t (0.80) par = 0.95, par2 = 2	Tawn T2 (0.77) par = 6.39, par2 = 0.9	Tawn T2 (0.78) par = 7.18, par2 = 0.9
	4,3	Tawn T2 (0.62) par = 4.96, par2 = 0.74	Tawn T2 (0.63) par = 6, par2 = 0.72	Tawn T2 (0.58) par = 3.97, par2 = 0.74	Tawn T2 (0.61) par = 4.86, par2 = 0.74
	5,4	t (0.88) par = 0.98, par2 = 2	t (0.87) par = 0.98, par2 = 2	Tawn T1 (0.87) par = 8.2, par2 = 0.99	Tawn T2 (0.84) par = 8.28, par2 = 0.95
	6,5	t (0.90) par = 0.99, par2 = 2	t (0.88) par = 0.98, par2 = 2	Tawn T1 (0.89) par = 10.19, par2 = 0.98	Tawn T1 (0.89) par = 10.5, par2 = 0.98
	7,6	t (0.87) par = 0.98, par2 = 2	t (0.87) par = 0.98, par2 = 2	Tawn T1 (0.84) par = 7.21, par2 = 0.97	Tawn T1 (0.85) par = 7.52, par2 = 0.98
	8,7	Tawn T2 (0.85) par = 10.26, par2 = 0.94	Tawn T2 (0.86) par = 10.92, par2 = 0.94	Tawn T2 (0.84) par = 8.99, par2 = 0.94	Tawn T2 (0.84) par = 9.91, par2 = 0.93
	9,8	Tawn T2 (0.79) par = 9.3, par2 = 0.88	Tawn T2 (0.81) par = 9.48, par2 = 0.89	Tawn T2 (0.79) par = 8.48, par2 = 0.88	Tawn T2 (0.79) par = 8.88, par2 = 0.88
	3,1 2	Ind.	Ind.	Ind.	Ind.
	4,2 3	Ind.	Ind.	Joe 270° (-0.07) par = -1.14	Ind.
2	5,3 4	Ind.	Ind.	Ind.	Ind.
	6,4 5	Ind.	t (0.11) par = 0.17, par2 = 3.96	Ind.	Ind.
	7,5 6	Ind.	t (-0.13) par = -0.21, par2 = 4.05	Ind.	Ind.
	8,6 7	Ind.	Ind.	BB8 (0.11) par = 1.31, par2 = 0.95	Ind.
	9,7 8	Ind.	Ind.	Ind.	Ind.
	4,1 3,2	Ind.	Ind.	Ind.	Ind.
	5,2 4,3	Gumbel (0.11) par = 1.13	Ind.	Ind.	Frank (0.21) par = 1.94
3	6,3 5,4	Ind.	Ind.	Frank (-0.14) par = -1.27	Ind.
	7,4 6,5	Ind.	Ind.	Gumbel 180° (0.03) par = 1.03	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
	5,1 4,3,2	Ind.	Ind.	Ind.	Ind.
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
4	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Gumbel 90° (-0.1) par = -1.11	Ind.
	9,5 8,7,6	Ind.	Gumbel 90° (-0.10) par = -1.12	Ind.	Ind.
	6,1 5,4,3,2	Ind.	t (-0.15) par = -0.23, par2 = 4.17	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Joe (0.11) par = 1.23	Ind.
5	8,3 7,6,5,4	Ind.	Ind.	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
6	9,3 8,7,6,5,4	Ind.	Ind.	Ind.	Ind.
	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
7	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
8					

TABLE A.2: Complete results of D-vines with all families included and independence test performed.

Tree	Pairs	$X, NPIT$	$STD X, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	Gumbel (0.82) par = 5.53	Gumbel (0.82) par = 5.49	Gumbel (0.82) par = 5.65	Gumbel (0.83) par = 6.03
	3,2	Gumbel (0.82) par = 5.62	Gumbel (0.77) par = 4.32	Gumbel (0.79) par = 4.87	Gumbel (0.81) par = 5.37
	4,3	Clayton 180° (0.61) ar = 3.06	Gumbel (0.68) par = 3.1	Clayton 180° (0.60) par = 3.06	Clayton 180° (0.63) par = 3.36
	5,4	Gumbel (0.87) par = 7.67	Gumbel (0.87) par = 7.59	Joe (0.84) par = 11.16	Gumbel (0.87) par = 7.48
	6,5	Gumbel (0.88) par = 8.59	Gumbel (0.87) par = 7.7	Joe (0.86) par = 13.16	Joe (0.86) par = 13.39
	7,6	Gumbel (0.84) par = 6.24	Gumbel (0.84) par = 6.26	Gumbel (0.83) par = 5.79	Gumbel (0.83) par = 6.02
	8,7	Gumbel (0.87) par = 7.96	Gumbel (0.88) par = 8.35	Joe (0.85) par = 11.81	Joe (0.85) par = 12.02
	9,8	Gumbel (0.83) par = 5.98	Gumbel (0.84) par = 6.37	Joe (0.80) par = 8.64	Gumbel (0.83) par = 5.79
2	3,1 2	Ind.	Gumbel (0.12) par = 1.13	Ind.	Ind.
	4,2 3	Ind.	Ind.	Gumbel 270° (-0.11) par = -1.13	Frank (-0.14) par = -1.31
	5,3 4	Ind.	Ind.	Frank (-0.14) par = -1.31	Ind.
	6,4 5	Ind.	Joe (0.08) par = 1.14	Joe 90° (-0.1) par = -1.2	Joe 90° (-0.11) par = -1.23
	7,5 6	Ind.	Frank (-0.23) par = -2.16	Joe (0.03) par = 1.06	Frank (-0.14) par = -1.25
	8,6 7	Joe 180° (0.01) par = 1.01	Joe 180° (0.01) par = 1.01	Joe 180° (0.01) par = 1.01	Joe 180° (0.01) par = 1.01
	9,7 8	Frank (-0.10) par = -0.93	Frank (-0.10) par = -0.87	Ind.	Frank (-0.13) par = -1.21
3	4,1 3,2	Frank (-0.16) par = -1.49	Frank (-0.16) par = -1.47	Ind.	Ind.
	5,2 4,3	Gumbel (0.06) par = 1.07	Joe (0.04) par = 1.08	Gumbel (0.05) par = 1.05	Gaussian (0.17) par = 0.27
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	Ind.	Ind.	Ind.	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Ind.	Joe 270° (-0.02) par = -1.04	Gumbel 270° (-0.06) par = -1.07
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Joe 90° (-0.15) par = -1.32	Clayton 270° (-0.11) par = -0.24	Ind.	Joe 90° (-0.18) par = -1.4
	8,4 7,6,5	Ind.	Ind.	Gumbel 270° (-0.02) par = -1.02	Gumbel 270° (-0.02) par = -1.02
	9,5 8,7,6	Ind.	Frank (-0.14) par = -1.26	Gaussian (-0.08) par = -0.12	Gumbel 90° (-0.06) par = -1.06
5	6,1 5,4,3,2	Gumbel 270° (-0.05) par = -1.05	Gumbel 90° (-0.11) par = -1.12	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Gumbel 90° (-0.13) par = -1.15	Ind.	Ind.
	8,3 7,6,5,4	Clayton 90° (-0.30) par = -0.86	Gumbel 270° (-0.22) par = -1.27	Ind.	Frank (-0.29) par = -2.8
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Gumbel 90° (-0.12) par = -1.14	Clayton 90° (-0.17) par = -0.41	Frank (-0.30) par = -2.89	Gumbel 270° (-0.22) par = -1.28
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Frank (-0.27) par = -2.54	Frank (-0.16) par = -1.44	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.

TABLE A.3: Complete results of D-vines with selected families included and no independence test.

Tree	Pairs	$X, NPIT$	$STD X, NPIT$	$GAMLSS, PPIT$	$GAMLSS, NPIT$
1	2,1	Gumbel (0.82) par = 5.53	Gumbel (0.82) par = 5.49	Gumbel (0.82) par = 5.65	Gumbel (0.83) par = 6.03
	3,2	Gumbel (0.82) par = 5.62	Gumbel (0.77) par = 4.32	Gumbel (0.79) par = 4.87	Gumbel (0.81) par = 5.37
	4,3	Clayton 180° (0.61) ar = 3.06	Gumbel (0.68) par = 3.1	Clayton 180° (0.60) par = 3.06	Clayton 180° (0.63) par = 3.36
	5,4	Gumbel (0.87) par = 7.67	Gumbel (0.87) par = 7.59	Joe (0.84) par = 11.16	Gumbel (0.87) par = 7.48
	6,5	Gumbel (0.88) par = 8.59	Gumbel (0.87) par = 7.7	Joe (0.86) par = 13.16	Joe (0.86) par = 13.39
	7,6	Gumbel (0.84) par = 6.24	Gumbel (0.84) par = 6.26	Gumbel (0.83) par = 5.79	Gumbel (0.83) par = 6.02
	8,7	Gumbel (0.87) par = 7.96	Gumbel (0.88) par = 8.35	Joe (0.85) par = 11.81	Joe (0.85) par = 12.02
	9,8	Gumbel (0.83) par = 5.98	Gumbel (0.84) par = 6.37	Joe (0.80) par = 8.64	Gumbel (0.83) par = 5.79
2	3,1 2	Ind.	Ind.	Ind.	Ind.
	4,2 3	Ind.	Ind.	Ind.	Ind.
	5,3 4	Ind.	Ind.	Ind.	Ind.
	6,4 5	Ind.	Joe (0.08) par = 1.14	Ind.	Ind.
	7,5 6	Ind.	Frank (-0.23) par = -2.16	Ind.	Frank (-0.14) par = -1.25
	8,6 7	Ind.	Ind.	Joe 180° (0.01) par = 1.01	Ind.
	9,7 8	Ind.	Ind.	Ind.	Ind.
3	4,1 3,2	Ind.	Ind.	Ind.	Frank (-0.18) par = -1.68
	5,2 4,3	Gumbel (0.06) par = 1.07	Ind.	Ind.	Gumbel (0.10) par = 1.11
	6,3 5,4	Ind.	Ind.	Ind.	Ind.
	7,4 6,5	Ind.	Ind.	Ind.	Ind.
	8,5 7,6	Ind.	Ind.	Ind.	Ind.
	9,6 8,7	Ind.	Ind.	Ind.	Ind.
4	5,1 4,3,2	Ind.	Ind.	Ind.	Ind.
	6,2 5,4,3	Ind.	Ind.	Ind.	Ind.
	7,3 6,5,4	Ind.	Ind.	Ind.	Ind.
	8,4 7,6,5	Ind.	Ind.	Ind.	Ind.
	9,5 8,7,6	Ind.	Frank (-0.12) par = -1.26	Ind.	Ind.
5	6,1 5,4,3,2	Ind.	Ind.	Ind.	Ind.
	7,2 6,5,4,3	Ind.	Ind.	Ind.	Ind.
	8,3 7,6,5,4	Ind.	Gumbel 270° (-0.23) par = -1.29	Ind.	Ind.
	9,4 8,7,6,5	Ind.	Ind.	Ind.	Ind.
6	7,1 6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	8,2 7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
	9,3 8,7,6,5,4	Ind.	Ind.	Ind.	Ind.
7	8,1 7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.
	9,2 8,7,6,5,4,3	Ind.	Ind.	Ind.	Ind.
8	9,1 8,7,6,5,4,3,2	Ind.	Ind.	Ind.	Ind.

TABLE A.4: Complete results of D-vines with selected families included and independence test performed.

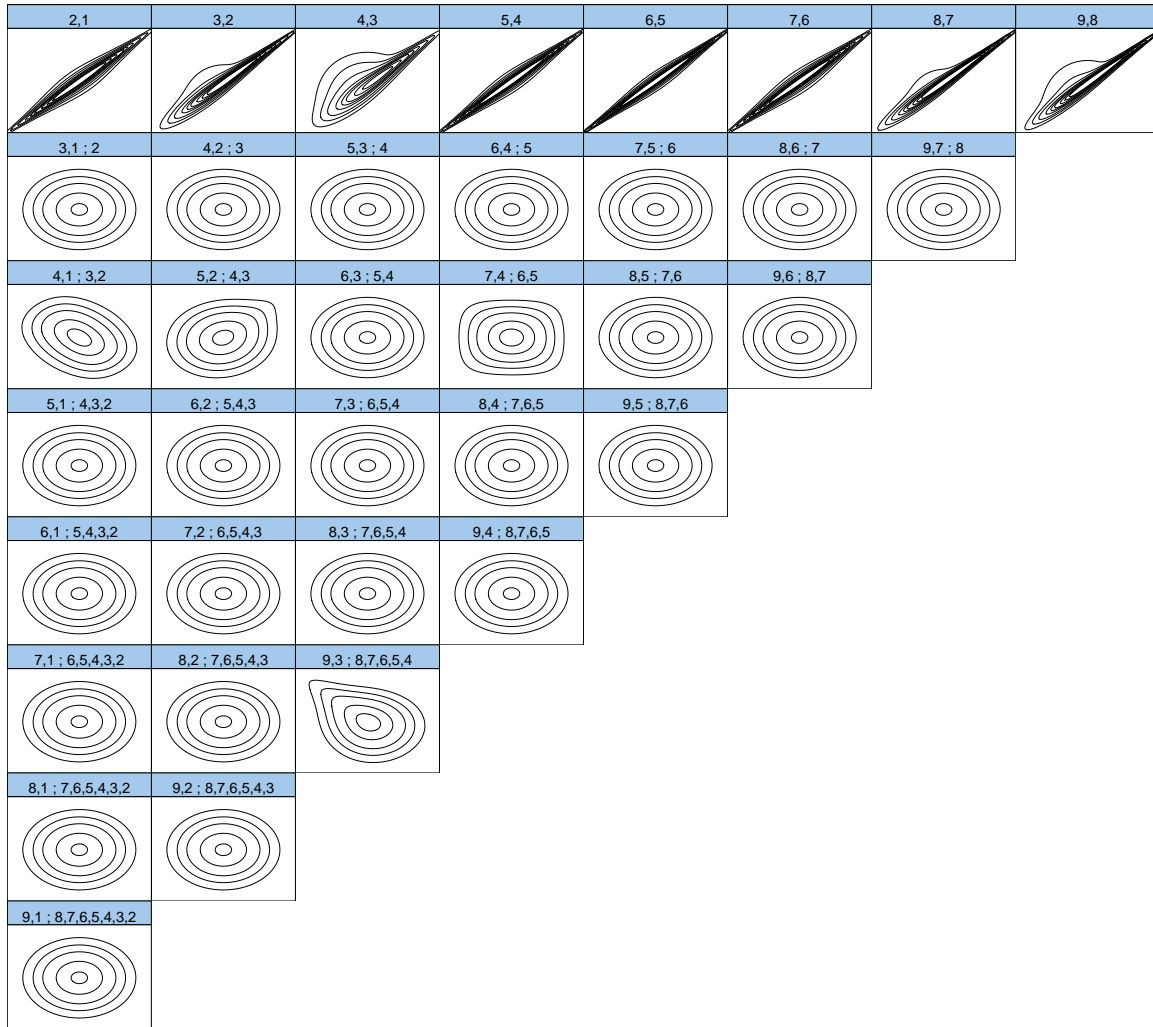


FIGURE A.1: Contour plots of all pair copulas from the D-vine with all families allowed, no independence test performed and the marginal Model X,NPIT.

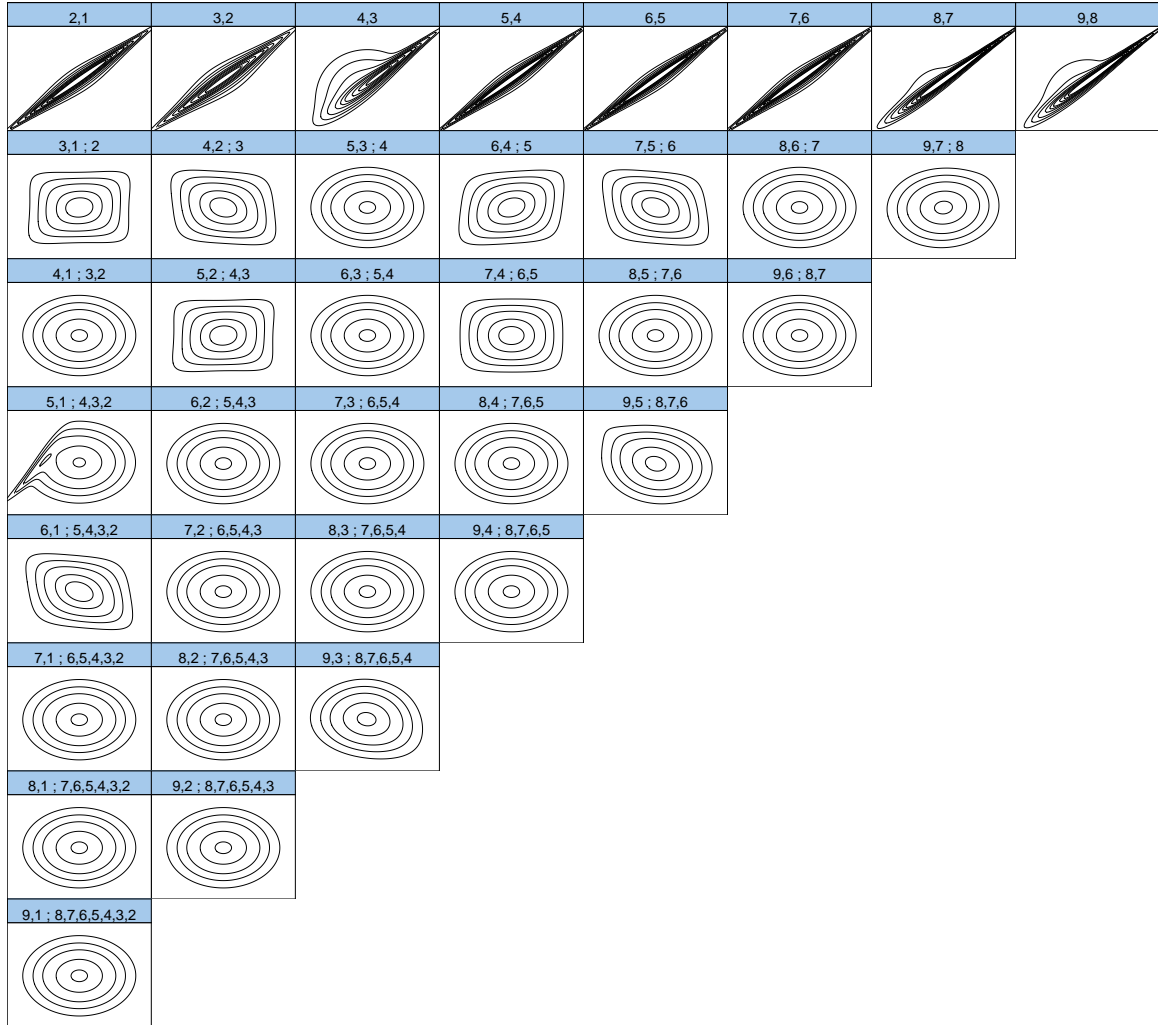


FIGURE A.2: Contour plots of all pair copulas from the D-vine with all families allowed, no independence test performed and the marginal Model STD_X,NPIT.

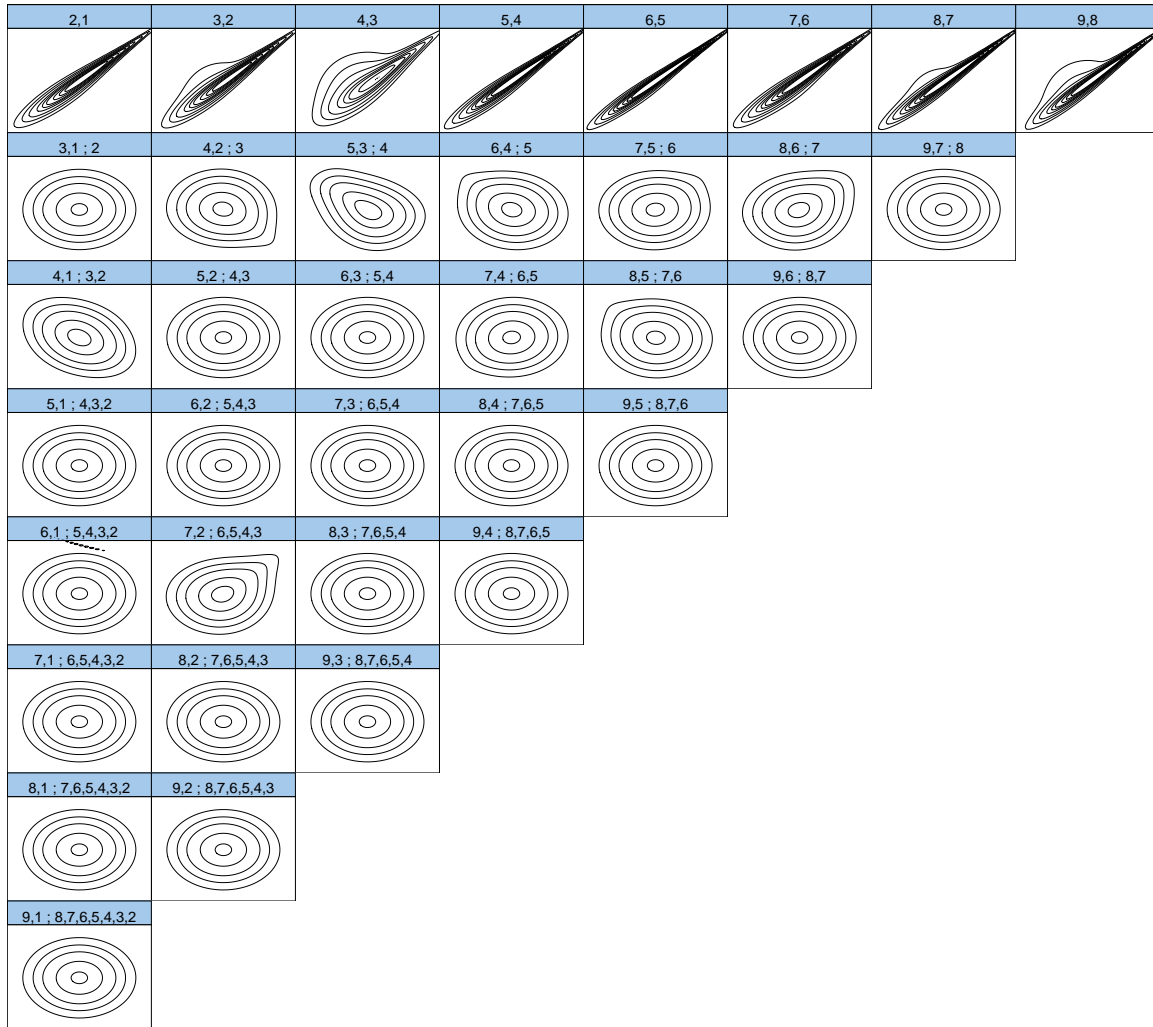


FIGURE A.3: Contour plots of all pair copulas from the D-vine with all families allowed, no independence test performed and the marginal Model GAMLSS, PPIT.

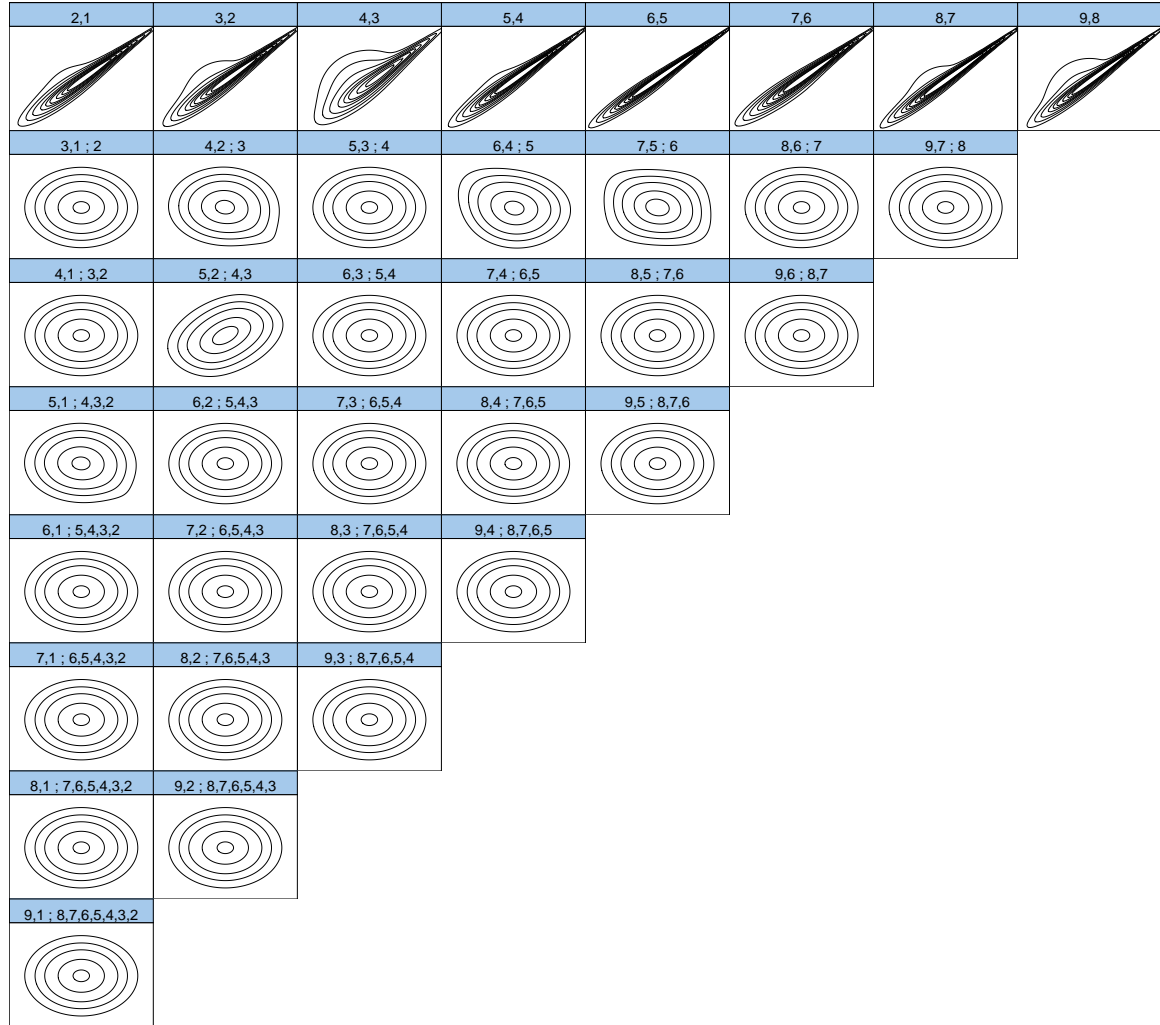


FIGURE A.4: Contour plots of all pair copulas from the D-vine with all families allowed, no independence test performed and the marginal Model GAMLSS, NPIT.

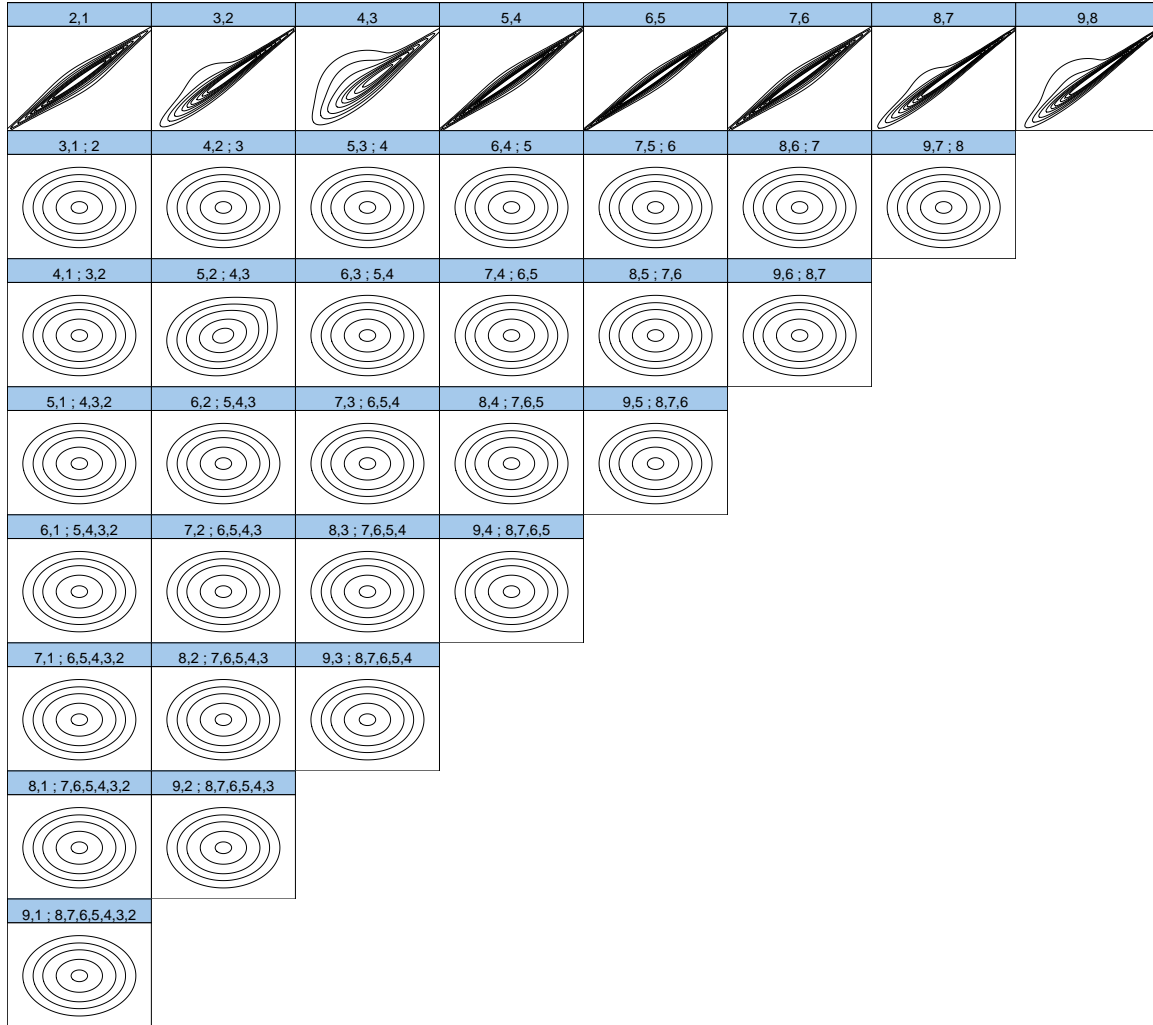


FIGURE A.5: Contour plots of all pair copulas from the D-vine with all families allowed, independence test performed and the marginal Model X,NPIT.

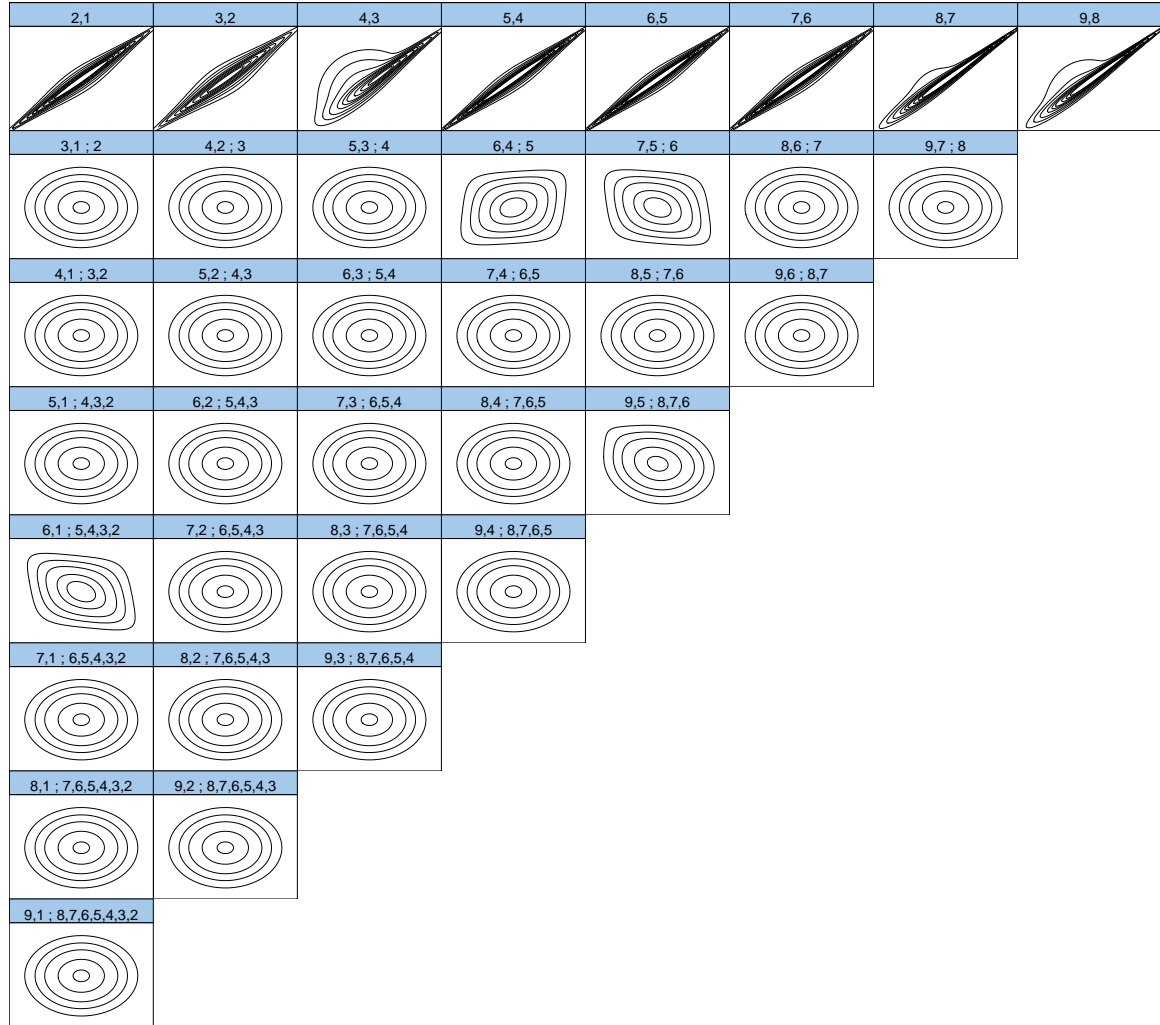


FIGURE A.6: Contour plots of all pair copulas from the D-vine with all families allowed, independence test performed and the marginal Model STD_X,NPIT.

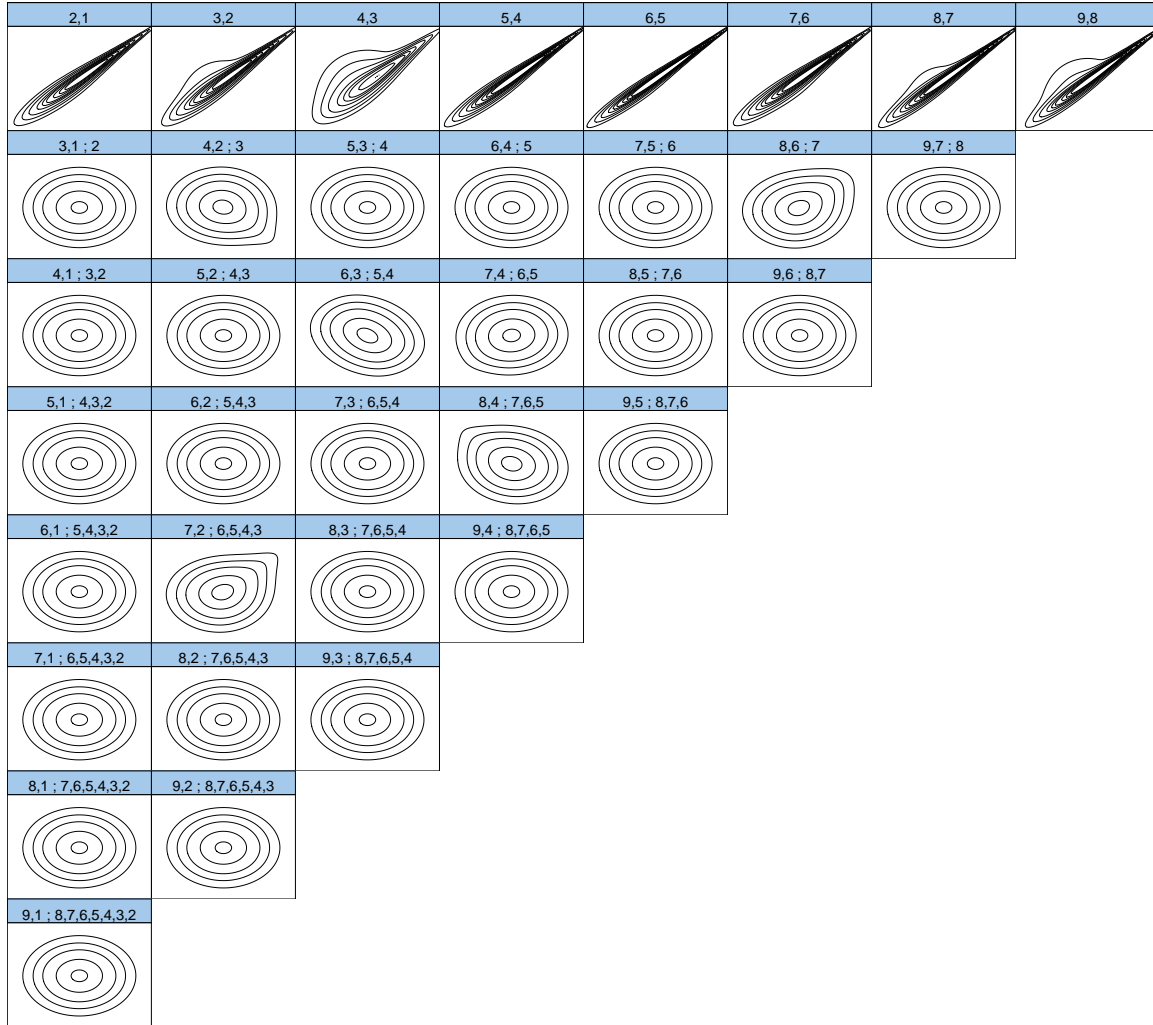


FIGURE A.7: Contour plots of all pair copulas from the D-vine with all families allowed, independence test performed and the marginal Model GAMLSS, PPIT.

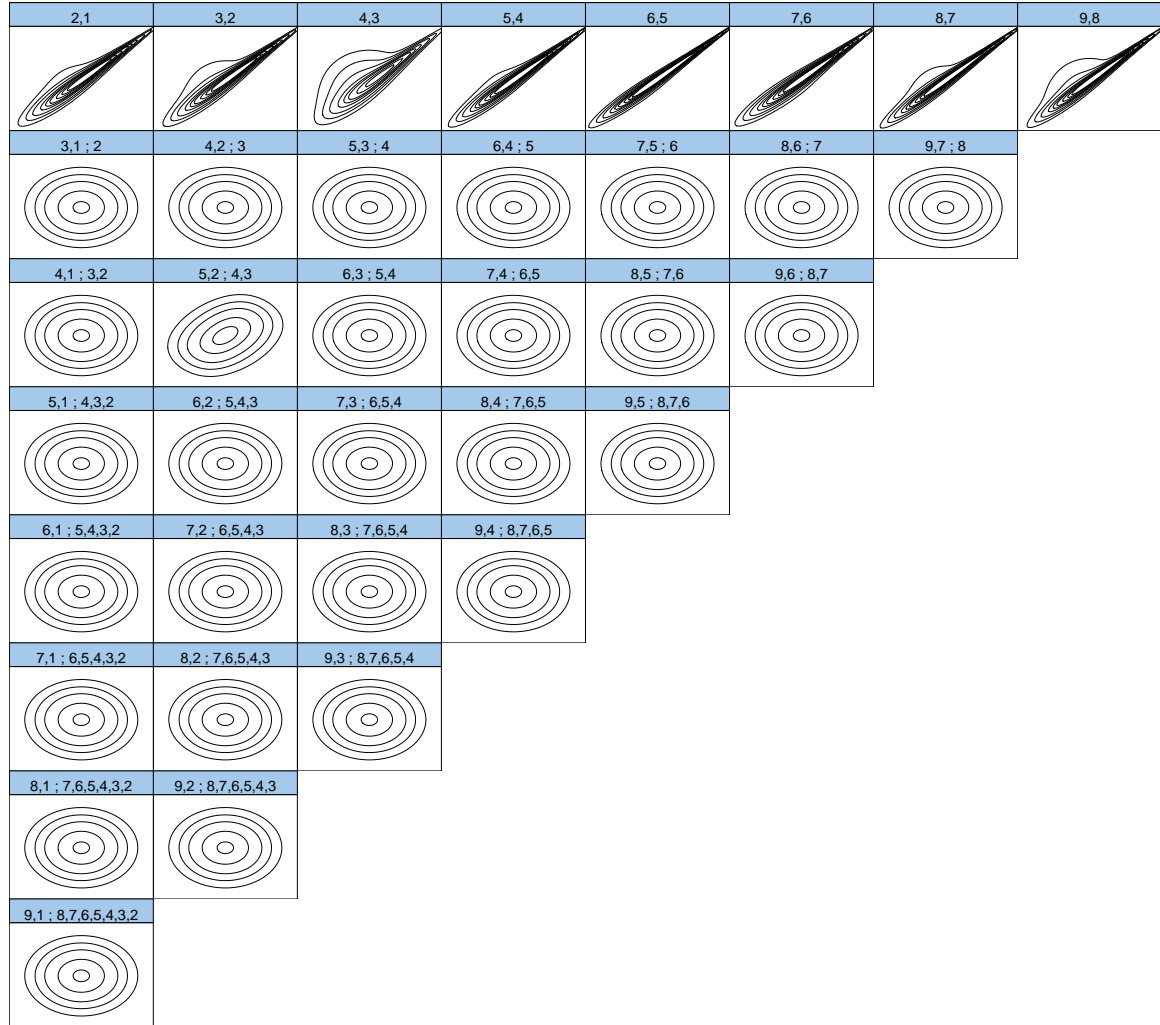


FIGURE A.8: Contour plots of all pair copulas from the D-vine with all families allowed, independence test performed and the marginal Model GAMLSS,NPIT.

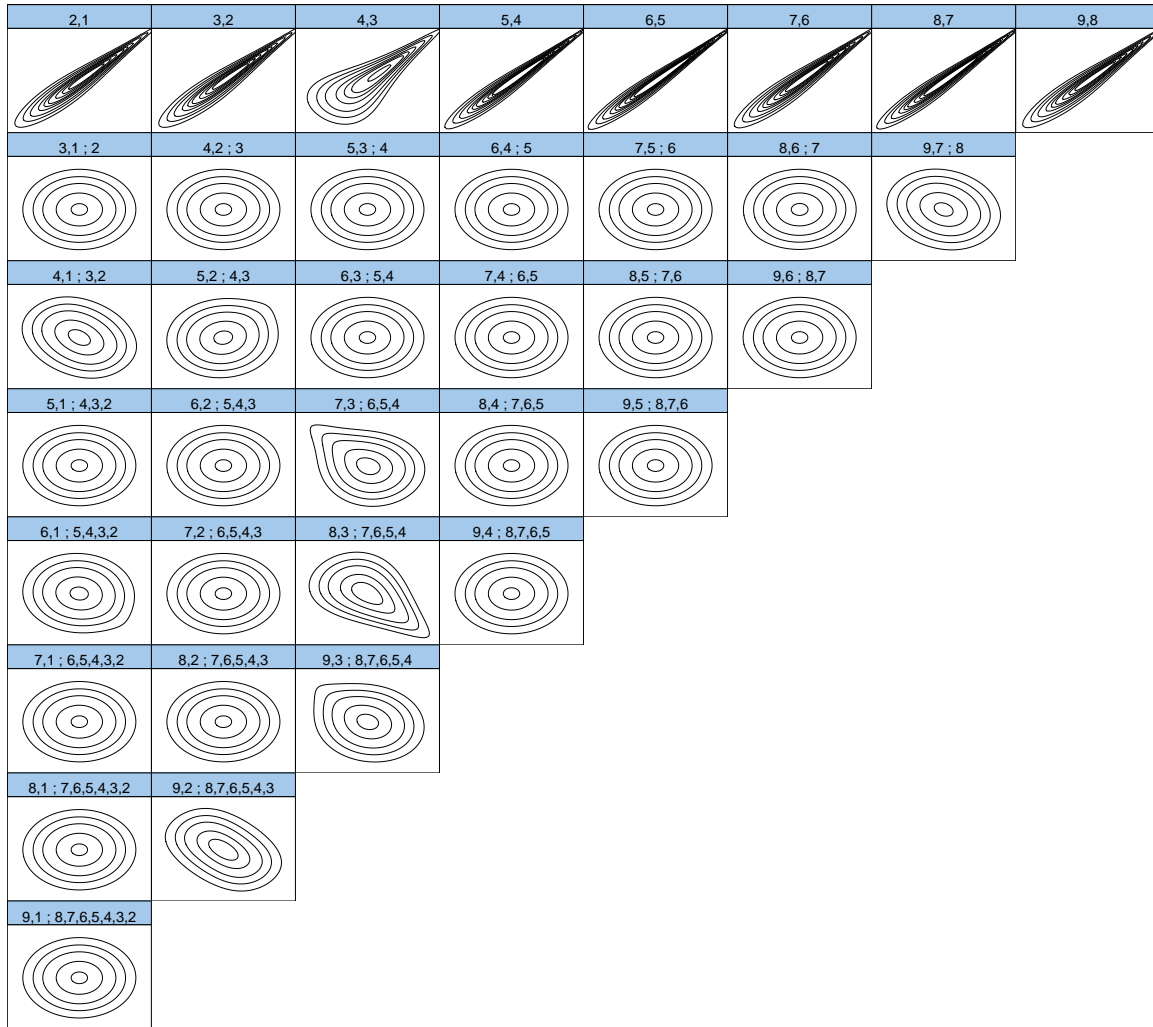


FIGURE A.9: Contour plots of all pair copulas from the D-vine with selected families allowed, no independence test performed and the marginal Model X_iNPIT.

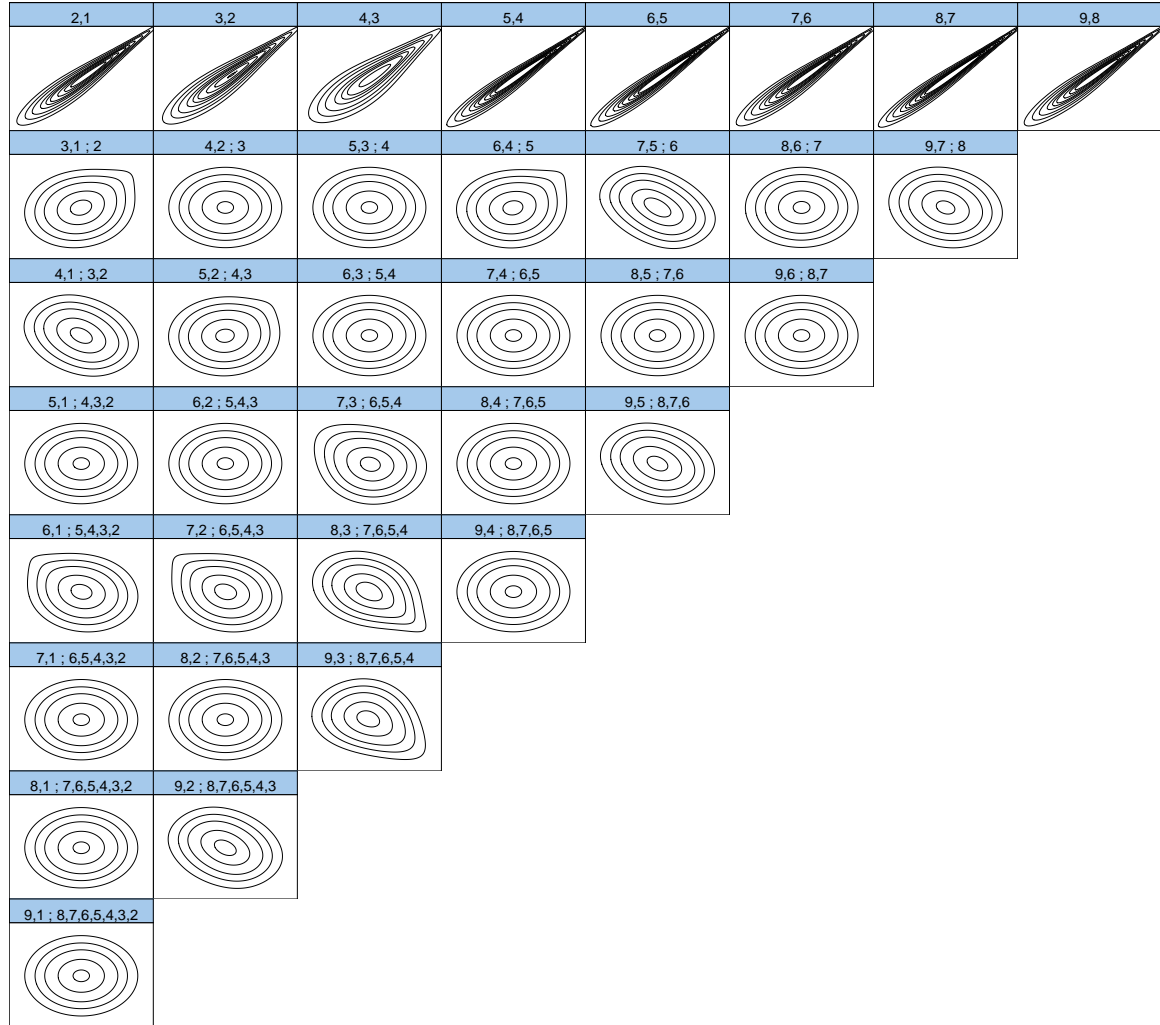


FIGURE A.10: Contour plots of all pair copulas from the D-vine with selected families allowed, no independence test performed and the marginal Model STDX,NPIT.

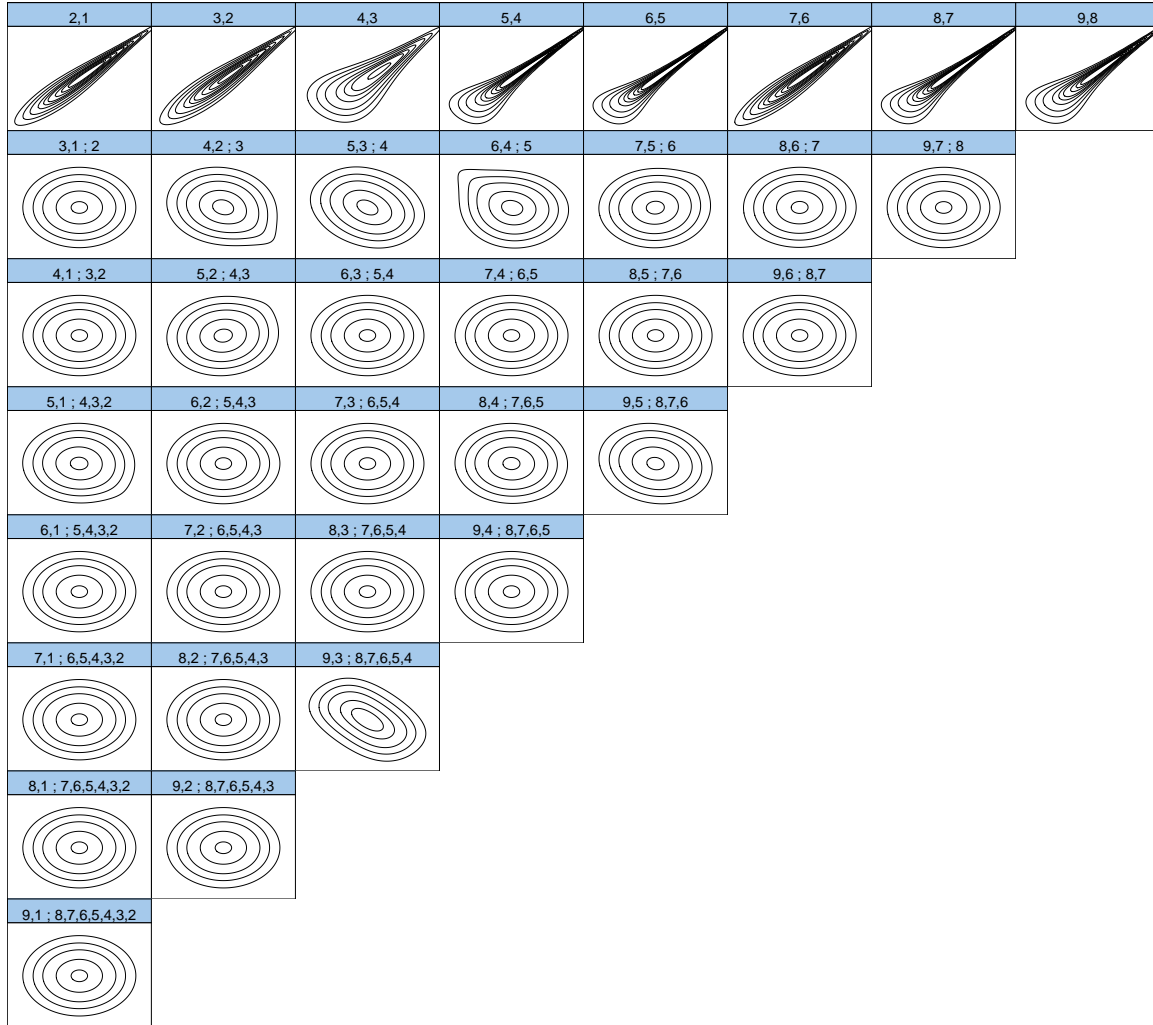


FIGURE A.11: Contour plots of all pair copulas from the D-vine with selected families allowed, no independence test performed and the marginal Model GAMLSS,PPIT.

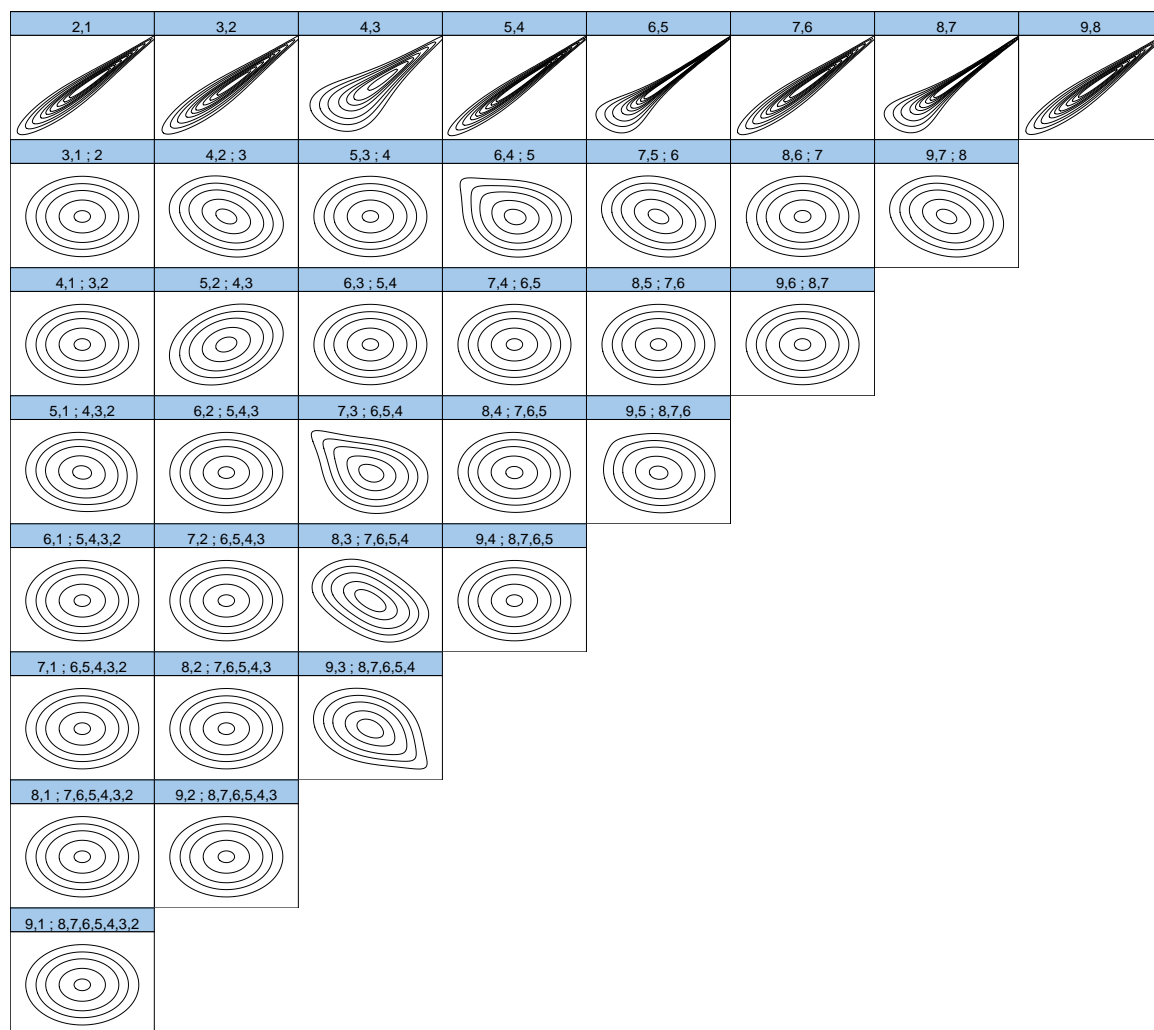


FIGURE A.12: Contour plots of all pair copulas from the D-vine with selected families allowed, no independence test performed and the marginal Model GAMLSS,NPIT.

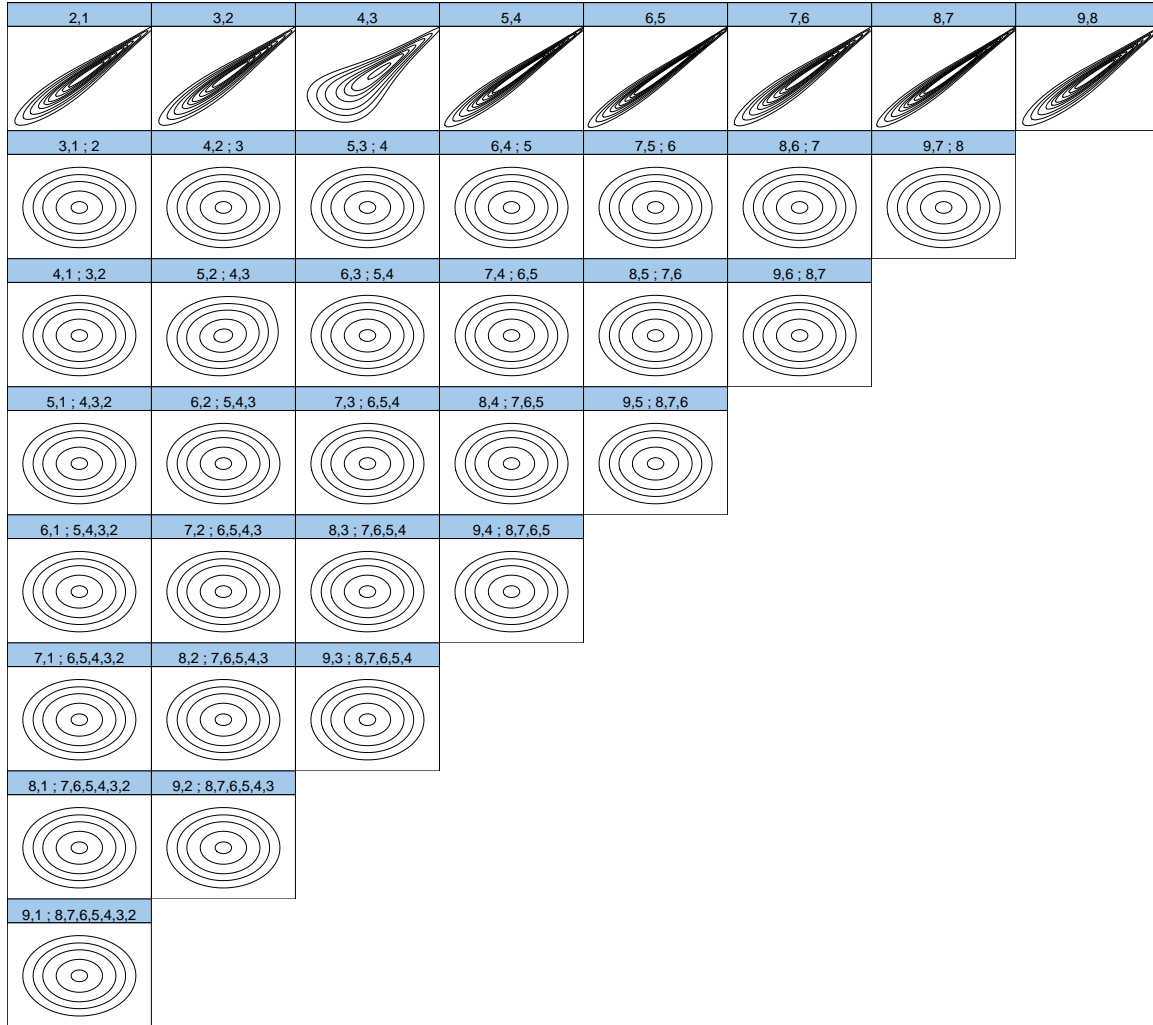


FIGURE A.13: Contour plots of all pair copulas from the D-vine with selected families allowed, independence test performed and the marginal Model X,NPIT.

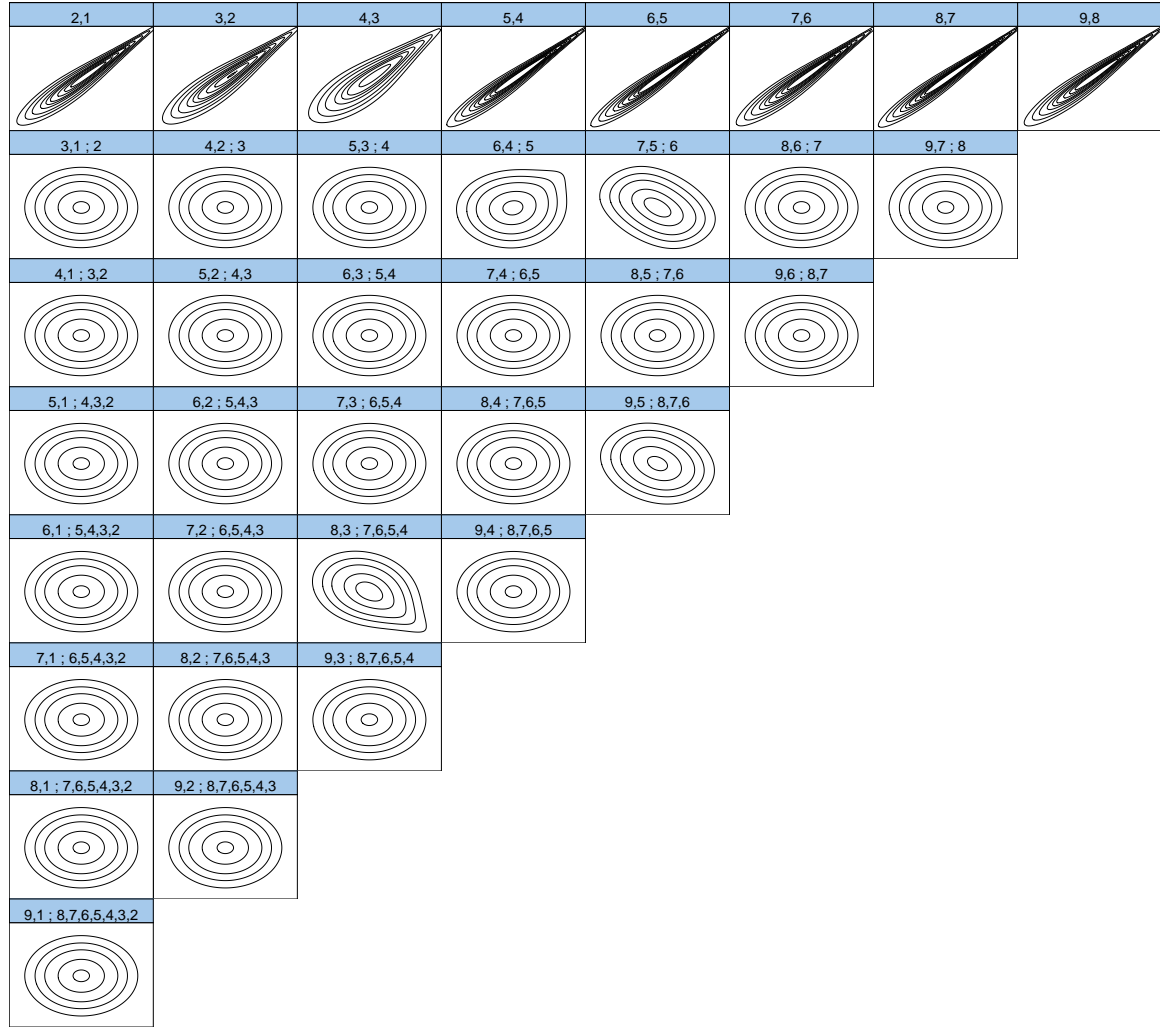


FIGURE A.14: Contour plots of all pair copulas from the D-vine with selected families allowed, independence test performed and the marginal Model STDX,NPIT.

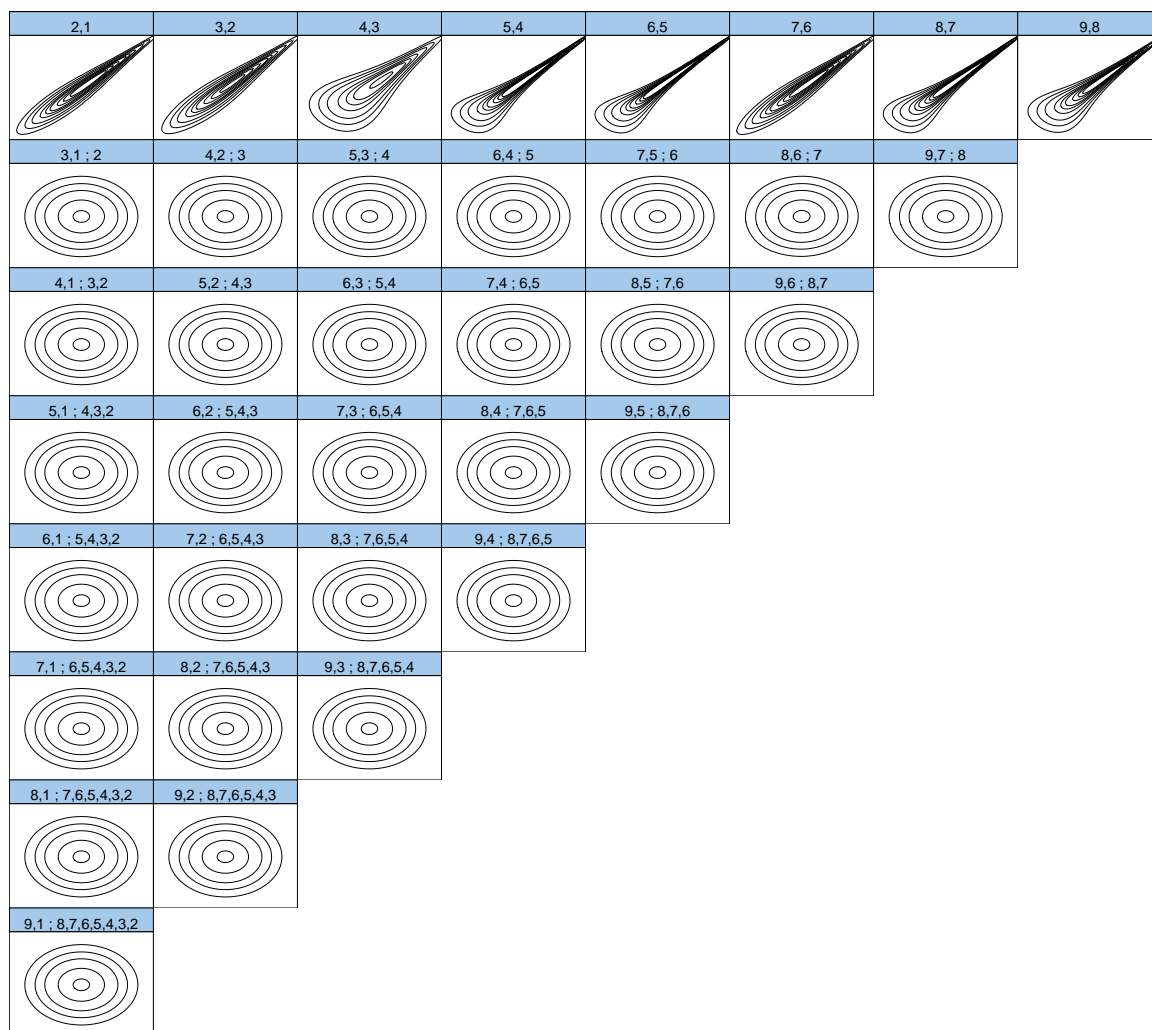


FIGURE A.15: Contour plots of all pair copulas from the D-vine with selected families allowed, independence test performed and the marginal Model GAMLSS, PPIT.

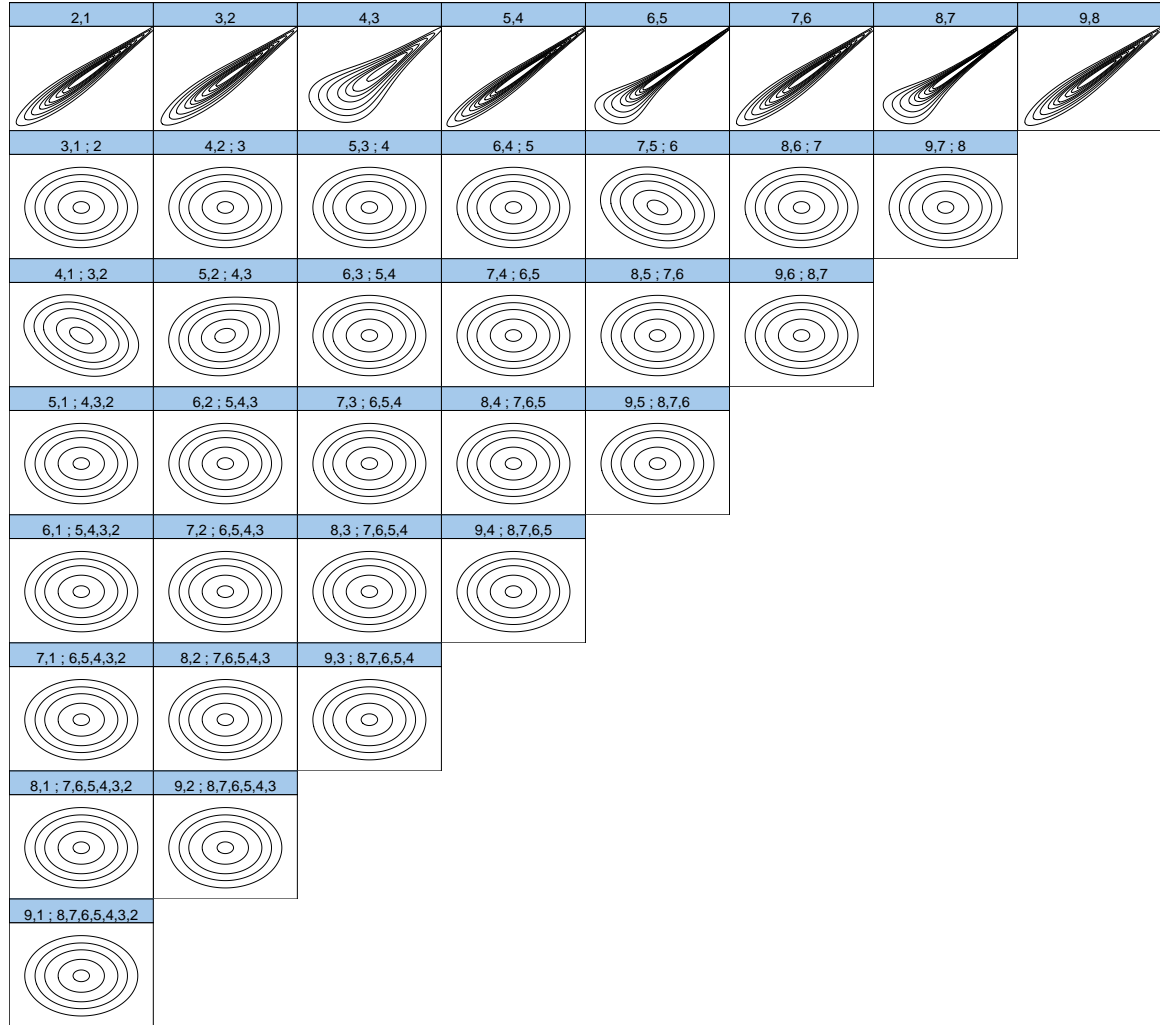


FIGURE A.16: Contour plots of all pair copulas from the D-vine with selected families allowed, independence test performed and the marginal Model GAMLSS,NPIT.

Bibliography

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi and Henrik Bakken (2009). "Pair-copula constructions of multiple dependence". In: *Insurance: Mathematics and economics* 44.2, pp. 182–198.
- Akaike, Hirotugu (1998). "Information theory and an extension of the maximum likelihood principle". In: *Selected Papers of Hirotugu Akaike*. Springer, pp. 199–213.
- Barthel, Nicole (2015). "Multivariate Survival Analysis using Vine-Copulas". MA thesis. Technische Universität München.
- Bedford, Tim and Roger M Cooke (2001). "Probability density decomposition for conditionally dependent random variables modeled by vines". In: *Annals of Mathematics and Artificial intelligence* 32.1, pp. 245–268.
- Brechmann, Eike Christian and Harry Joe (2014). "Parsimonious parameterization of correlation matrices using truncated vines and factor analysis". In: *Computational Statistics & Data Analysis* 77, pp. 233–251.
- Clayton, David G (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence". In: *Biometrika*, pp. 141–151.
- Czado, Claudia (2010). "Pair-copula constructions of multivariate copulas". In: *Copula theory and its applications*. Springer, pp. 93–109.
- Diestel, Reinhard (2000). *Graph theory*. Springer.
- Dissmann, Jeffrey, Eike C Brechmann, Claudia Czado and Dorota Kurowicka (2013). "Selecting and estimating regular vine copulae and application to financial returns". In: *Computational Statistics & Data Analysis* 59, pp. 52–69.
- Dunn, Peter K and Gordon K Smyth (1996). "Randomized quantile residuals". In: *Journal of Computational and Graphical Statistics* 5.3, pp. 236–244.
- Embrechts, Paul, Filip Lindskog and Alexander McNeil (2001). "Modelling dependence with copulas". In: *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.
- Embrechts, Paul, Alexander McNeil and Daniel Straumann (2002). "Correlation and dependence in risk management: properties and pitfalls". In: *Risk management: value at risk and beyond*, pp. 176–223.
- Frank, Maurice J (1979). "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$ ". In: *Aequationes mathematicae* 19.1, pp. 194–226.
- Gneiting, Tilmann and Adrian E Raftery (2007). "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American Statistical Association* 102.477, pp. 359–378.
- Gumbel, Emil Julius (1960). "Distributions des valeurs extrêmes en plusieurs dimensions". In: *Publ. Inst. Statist. Univ. Paris* 9, pp. 171–173.
- Hobæk Haff, Ingrid et al. (2013). "Parameter estimation for pair-copula constructions". In: *Bernoulli* 19.2, pp. 462–491.

- Joe, Harry (1993). "Parametric families of multivariate distributions with given margins". In: *Journal of multivariate analysis* 46.2, pp. 262–282.
- (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.
- (2005). "Asymptotic efficiency of the two-stage estimation method for copula-based models". In: *Journal of Multivariate Analysis* 94.2, pp. 401–419.
- Joe, Harry and James Jianmeng Xu (2016). "The estimation method of inference functions for margins for multivariate models". In:
- Kendall, Maurice G (1938). "A new measure of rank correlation". In: *Biometrika* 30.1/2, pp. 81–93.
- Killiches, Matthias and Claudia Czado (2017). "A D-vine copula based model for repeated measurements extending linear mixed models with homogeneous correlation structure". In: *arXiv preprint arXiv:1705.06261*.
- Kraus, Daniel and Claudia Czado (2017). "D-vine copula based quantile regression". In: *Computational Statistics & Data Analysis*, pp. 1–18.
- Matheson, James E and Robert L Winkler (1976). "Scoring rules for continuous probability distributions". In: *Management science* 22.10, pp. 1087–1096.
- Nagler, Thomas (2017a). "A generic approach to nonparametric function estimation with mixed data". In: *arXiv preprint arXiv:1704.07457*.
- (2017b). *kdevine: Multivariate Kernel Density Estimation with Vine Copulas*. R package version 0.4.1. URL: <https://CRAN.R-project.org/package=kdevine>.
- Nelsen, Roger B (2010). "An Introduction to Copulas". In:
- Pickands, James (1981). "Multivariate extreme value distributions". In: *Proceedings 43rd Session International Statistical Institute*. Vol. 2, pp. 859–878.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Rigby, R. A. and D. M. Stasinopoulos (2005). "Generalized additive models for location, scale and shape,(with discussion)". In: *Applied Statistics* 54, pp. 507–554.
- Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler and Tobias Erhardt (2017a). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.1.3. URL: <https://CRAN.R-project.org/package=VineCopula>.
- Schepsmeier, Ulf, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler, Tobias Erhardt, Carlos Almeida, Aleksey Min, Claudia Czado, Mathias Hofmann et al. (2017b). "Package 'VineCopula'". In: *VineCopula Reference Manual* (see also cran.r-project.org/web/packages/VineCopula).
- Schwarz, Gideon et al. (1978). "Estimating the dimension of a model". In: *The annals of statistics* 6.2, pp. 461–464.
- Sklar, M (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- Stasinopoulos, DM, RA Rigby and C Akantziliotou (2006). "Instructions on how to use the GAMLSS package in R". In: *Accompanying documentation in the current GAMLSS help files* (see also www.gamlss.org).

- Stoeber, Jakob, Harry Joe and Claudia Czado (2013). "Simplified pair copula constructions—limitations and extensions". In: *Journal of Multivariate Analysis* 119, pp. 101–118.
- Tantau, Till (2015). *The TikZ and PGF Packages. Manual for version 3.0.0*. URL: <http://sourceforge.net/projects/pgf/>.
- Tawn, Jonathan A (1988). "Bivariate extreme value theory: models and estimation". In: *Biometrika*, pp. 397–415.
- Vuong, Quang H (1989). "Likelihood ratio tests for model selection and non-nested hypotheses". In: *Econometrica: Journal of the Econometric Society*, pp. 307–333.
- Wang, Jingbo, Wai Wan Tsang and George Marsaglia (2003). "Evaluating Kolmogorov's distribution". In: *Journal of Statistical Software* 8.18.
- Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.