Technische Universität München

Department of Mathematics

Assessing system relevance of financial institutions using pair-copula constructions for modeling

Master's Thesis
by
Otto Kähm

Supervisor: Professor Claudia Czado, Ph.D.
Advisor: Professor Claudia Czado, Ph.D.
Submission date: August 4, 2014

I hereby confirm that I have written this Master's thesis independently and with help only from the specified resources and support.

Munich, August 4, 2014

List of Abbreviations and Symbols

AIC	Akaike information criterion
BIC	Bayesian information criterion
CDS	Credit default swap
LLH	Log-likelihood
$F(\cdot)$	Distribution function
$f(\cdot)$	Density function
$F_1(\cdot), ..., F_d(\cdot)$	Marginal distribution functions
$f_1(\cdot), ..., f_d(\cdot)$	Marginal density functions
$F(\cdot|\cdot)$	Conditional distribution functions
$f(\cdot|\cdot)$	Conditional density function
$C(\cdot)$	Copula distribution function
$c(\cdot)$	Copula density function
θ	Parameter of the copula $C(\cdot)$
ρ	Pearson's correlation coefficient
$\hat{\rho}$	Empirical version of Pearson's correlation coefficient
τ	Kendall's tau
$\hat{\tau}$	Empirical version of Kendall's tau
$E[\cdot]$	Expectation
$Var[\cdot]$	Variance
$\mathrm{Diag}(1)$	$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$
$f(\mathbf{x}) \propto g(\mathbf{x})$	$f(\mathbf{x}) = const * g(\mathbf{x})$
$\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$	d- dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ.
$\mathcal{N}(\mu, \sigma)$	1- dimensional normal distribution with mean μ and standard deviation σ.
MH	Metropolis Hastings
ESS	Effective sample size
Unif[0,1]	Uniform distribution on [0,1]
Φ	distribution function of standard normal distribution
$logit(\mathbf{x})$	$\left(\ln\left(\frac{x_i}{1-x_i} \right) \right)_{\{i=1,...,d\}}$
$\mathbf{D} = \{i_1, \ldots, i_I\} \subset \{1, \ldots, n\}$	Conditional set
$-\mathbf{D}$	$\{1, \ldots, n\} \backslash \mathbf{D}$
$\mathbf{u}$	$(\underset{\in [0,1]^{d-I}}{\mathbf{u_{-D}}}, \underset{\in [0,1]^{I}}{\mathbf{u_D}}) \in [0,1]^d$
U-Level	Variables taking values in $[0,1]^d$
$\lVert \cdot \rVert_{\mathrm{TV}}$	Total variation
L-Level	logit transformed U-Level variables
Z-Level	Φ transformed U-Level variables
$\mathbb{R}$	Real numbers
$\mathbb{N}$	Natural numbers
$\mathbf{X} = (X_1, ..., X_d)^t$	Vector of random variables
$\mathbf{x} = (x_1, ..., x_d)^t$	Vector of observation of random variables $X_1, ..., X_d$

Contents

List of Figures

List of Tables

List of Algorithms

1 Introduction

During the financial crises, the interdependent nature of financial institutions was one key reason for the severe crises (Bluhm and Krahnen, 2011; Drehmann and Tarashev, 2013). Therefore, the evaluation of the dependence structure and its effect on the performance of financial markets, in particular banks, became a major task of the financial sector.

One key driver of scientific research in this area is the stress test being applied by the European Central Bank (ECB, 2013), which has aroused a lot of public interest in this topic worldwide.

Because of the significance of this issue, this Master's thesis was aimed to develop methods to evaluate the behavior of financial institutions in situations of financial crises. On this basis, one key question was to elicit which financial institutions are of systemic relevance (Risk and Soundness, 2009). Since many academic models failed to evaluate the systemic risk of financial institutions adequately (see Risk and Soundness, 2009), investigating this issue became increasingly important.

Most of the currently used models apply multivariate Gaussian distributions, which determine the interdependencies of financial markets by correlation analysis. These data have, by definition, a Gaussian copula. This approach, however, constitutes a great weakness, as Gaussian copulas are unable to model effects like tail dependencies. For systemic risks as described in the work of Risk and Soundness, 2009, it is not sufficient to measure the correlation, but the more extreme effects, like tail dependencies, have to be measured. In this study, this was achieved by including a wide range of dependence structures between random variables which can cover these effects. Therefore, the approach of modeling dependencies with **R-vine** models will be introduced, which are much more flexible than previous models.

In order to answer the thesis question, stress tests to the financial sector will be applied. For this purpose, we use the market data on credit default swaps (CDS) of 38 financial institutions, which were kindly made available by Hendrich, 2012. With this data set, stress situations with increasing stress level, caused by a number of systemic banks located in different geographical regions, were simulated. Given these stress situations, it was possible to analyze the probability distributions of these financial institutions.

The following is a short overview of the chapter's contents: In **Chapter 2** and **3**, the derivation of some theoretical concepts was carried out, which are fundamental for the understanding of simulations. The algorithms for these simulations and the theoretic background will be developed in **Chapter 4** and **5**. In **Chapter 6** extensive simulations

with simulated data will be performed to validate the accuracy and the reliability of the methods used. In **Chapter 7** the empirical data will be analyzed descriptively using regression analyses to initially examine the relationships between the financial institutions. The final simulations with the empirical data set will be conducted in **Chapter 8**, where the results are also presented and analyzed.

In **Chapter 9** a new approach of regression for financial institutions with C-vine dependence structures, located in different geographic regions, will be presented. This idea was inspired by binning analyses originally used for geographical data, which will be used here to develop similarity-based copulas. It will be shown that this method can be used for general vines and results in good fitting models, despite its low amount of variables.

2 Preliminaries

Throughout this thesis several basic tools will be applied repeatedly. The most important of these recurring concepts will be presented briefly in the current chapter.

In the first section, different scaling levels will be introduced, which will be needed for the simulations carried out in this thesis. In the second section, basic results for a conditional density are presented.

It should be noted here, that vectors will be denoted by **bold** letters in this thesis.

2.1 Different scaling levels

In this thesis three different scaling levels for the random variables are used. These levels and their different areas of application are explained in the following.

The first level is the **U-Level**:

$$\mathbf{U} = (U_1, \ldots, U_d)^t \in [0,1]^d \tag{2.1}$$

$$\mathbf{U} \sim \pi_U(\cdot) \tag{2.2}$$

This is the level of the R-vine copula realizations. Its density is denoted by $\pi_U(\cdot)$.

In some cases, we will need to work with random variables taking values in $\mathbb{R}^d$. One reason is that for certain algorithms (see **Chapter 5.4**) an unbounded domain will be needed. To do this, the **U** variable needs to be transformed to $\mathbb{R}^d$. One possible choice is the logit transformation:

Definition 2.1 (Logit Transformation)**.**
logit: $(0,1)^d \to \mathbb{R}^d$ *is defined as*

$$\boldsymbol{\ell} = logit(\mathbf{u}) = \left(\log \left(\frac{u_1}{1-u_1} \right), \ldots, \log \left(\frac{u_d}{1-u_d} \right) \right)^t$$

The inversion of the logit is:

$$logit^{-1}(\boldsymbol{\ell}) = \left(\left(\frac{e^{l_1}}{1+e^{l_1}} \right), \ldots, \left(\frac{e^{l_d}}{1+e^{l_d}} \right) \right)^t$$

The advantage of *logit* is that the function and its inversion are both known in closed form. That leads us to the following definition:

The **L-Level** is the level of the *logit* transformed **U** variable.

$$\mathbf{L} = logit(\mathbf{U}) = \left(\log \left(\frac{U_1}{1 - U_1} \right), \ldots, \log \left(\frac{U_d}{1 - U_d} \right) \right)^t \sim \pi_L(\cdot) \qquad (2.3)$$

This level will be used for algorithms on $\mathbb{R}^d$ as the logit transformation has an easy closed form, and hence is fast to evaluate numerically. Its density is denoted by $\pi_L(\cdot)$.

The last level which will be used is the **Z-Level**, which is obtained by applying the inverted cumulative distribution function of the normal distribution Φ^{-1} on the **U** variable.

$$\mathbf{Z} = \left(\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_d) \right)^t \sim \pi_Z(\cdot) \qquad (2.4)$$

This level is mainly used for visualization purposes, as the shapes of many pair-copulas on the **Z-Level** are well known and used for visualization as in Kurrowicka and Joe, 2011. The problem of the **Z-Level** is that it does not have a closed form representation and requires numerical integration, which is slow to evaluate (Conrad, 2005).
Comparing the **L-Level** and the **Z-Level**, we see that the **L-Level** values take more extreme values than the **Z-Level** values. This behavior is shown in **Figure 1**.
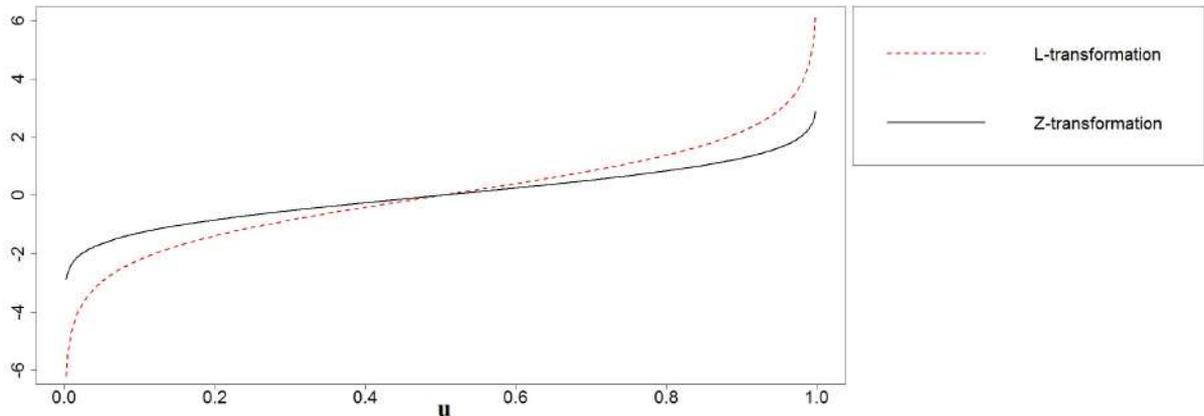
Figure 1: Comparison of the transformation on the **L-Level** and on the **Z-Level**

If no specific level is selected, the real valued random variable is denoted by **X**. The density is denoted by π, if no specific level was selected.

2.2 Proportionality of conditional densities

Closed form solutions for the simulation of conditional copulas were investigated by Hendrich, 2012 and Brechmann et al., 2013. However, these solutions only work when conditioning on one variable. Therefore, one aim of this thesis is to develop more general methods, which will be done in **Chapter 5**.

As conditional random variables play a crucial role in this thesis, the notation for conditional sets is introduced.

In general, we denote a d-dimensional random variable as:

$$\mathbf{X} = (X_1, \ldots, X_d)^t$$

The set of indices wanted to condition on is denoted by $\mathbf{D} = \{i_1, \ldots, i_I\} \subset \{1, \ldots, d\}$. The remaining set is denoted by $-\mathbf{D} = \{1, \ldots, d\}\backslash\mathbf{D}$.

The question to be solved is how a random variable can be simulated, for which the vector $\mathbf{X_D} := (X_i)_{i \in \mathbf{D}}$ with $|\mathbf{D}| = I$ is set to fixed values. The vector $(X_i)_{i \in -\mathbf{D}}$ is denoted by $\mathbf{X_{-D}} := (X_i)_{i \in -\mathbf{D}}$. Thus, the density function becomes a conditional density function. The conditional density, as seen in the calculation below, is proportional to the full density:

$$\pi(\mathbf{x_{-D}}|\mathbf{X_D} = \mathbf{u_D}) = \frac{\pi(\mathbf{x_{-D}}, \mathbf{x_D})}{\displaystyle\int_{\mathbb{R}^{d-I}} \pi(\mathbf{x_{-D}}, \mathbf{x_D})d\mathbf{x_{-D}}} \tag{2.5}$$

$$\propto \pi(\mathbf{x_{-D}}, \mathbf{x_D}) \tag{2.6}$$

The fact that the density of the conditional random variable up to a proportional constant can be evaluated, will be used in the following chapters to simulate realizations from this random variable with the **Metropolis Hastings algorithm**, which will be introduced in **Chapter 5**, as described in Chib and Greenberg, 1995.

3 R-vines

As we want to evaluate the connectedness of financial institutions we need to introduce a statistical method to measure dependence of random variables. As we have stated in the **Introduction**, most of todays models use multivariate normal distributions or multivariate Student distributions. These models, however, only measure the dependence among two companies by their Pearson correlation ρ. These models are clearly very limited, and it was mentioned by Huang et al., 2009, that these limited models, which describe the reality badly, were one reason of the financial crises.

Therefore, in this chapter we will introduce the **regular-vine copulas** (R-vine copulas), which are a flexible method to model the dependence among stochastic variables, as used by Bedford and Cooke, 2002.

We will introduce the bivariate pair-copulas, which describe the dependence of two random variables. We will further derive methods for the estimation of these pair-copulas and methods to analyze their fit.

Additionally we will explain the pair-copula construction in order to analyze higher dimensional vectors in this chapter.

3.1 Copulas

At first we will introduce the general definition of a copula which describes the dependence of a multivariate random variable with uniform margins.

Definition 3.1 (Copula).
A d-dimensional copula $C : [0,1]^d \to [0,1]$ is a joint d-dimensional distribution function for a random vector $\mathbf{U}$ with uniform margins on the set $[0,1]$:

$$C(u_1, u_2, ..., u_d) := \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, ..., U_d \leq u_d]$$

To illustrate the definition, we will show an example of a copula:

Example 3.1 (Independence copula)**.**
Let $U_1, ..., U_d$ be i.i.d. random variables with uniform margins on $[0, 1]$, the **copula** of
$\mathbf{U} = (U_1, ..., U_d)$ has the form

$$
\begin{aligned}
C(u_1, u_2, ..., u_d) :=& \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, ..., U_d \leq u_d] \\
=& \mathbb{P}[U_1 \leq u_1] \cdot \mathbb{P}[U_2 \leq u_2] \cdot ... \cdot \mathbb{P}[U_d \leq u_d] \\
=& u_1 \cdot u_2 \cdot ... \cdot u_d \\
=& \prod_{i=1}^{d} u_i
\end{aligned}
$$

As can be seen from **Definition 3.1**, copulas are defined as cumulative distribution functions with uniform margins. This raises the question, how one can describe dependencies of random variables witch arbitrary marginal distributions. This question is solved by Sklar's theorem.

Theorem 3.1 (Sklar's Theorem)**.**
For each distribution function F, of a d-dimensional real-valued random variable, with marginal distribution functions $F_1, ..., F_d$, there exists a d-dimensional copula $C : [0, 1]^d \rightarrow [0, 1]$ such that for all $\mathbf{x} \in \mathbb{R}^d$

$$
F(\mathbf{x}) = C(F_1(x_1), ..., F_d(x_d))
$$

If all marginal distribution functions $F_1, ..., F_d$ are continuous, then C is unique.

Sklar's theorem gives an easy way to construct multidimensional distributions, using the wide area of estimating one dimensional distributions. This theorem shows that the construction of multidimensional distributions, can be decomposed into two steps:

- Estimating the appropriate marginal distributions

- Estimating the copula for the transformed multivariate random variable.

The choice of the marginal distributions was already carried out in the thesis of Hendrich, 2012, which we are kindly allowed to use. We thus can focus on the task of choosing the appropriate copula. For this task we will introduce the R-vine copulas, which represent a flexible and good to estimate copula family. To do this, we have to analyze 2 dimensional copulas, the so called **pair-copulas**.

3.2 Pair-copulas

At first we will introduce some two dimensional copulas, on which we will focus.

Definition 3.2 (Gaussian copula).
The copula of a two dimensional normal vector (X_1, X_2) with correlation matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ and mean $\boldsymbol{\mu} \in \mathbb{R}^2$ is given by:

$$C_\rho(u_1, u_2) = \Phi_{\boldsymbol{\mu},\Sigma}(\Phi_{\mu_1,\sigma_1}^{-1}(u_1), \Phi_{\mu_2,\sigma_2}^{-1}(u_2))$$

Where $\Phi_{\mu,\sigma}$ denotes the cumulative distribution function of a $\mathcal{N}_{\mu,\sigma}$ distributed random variable and $\Phi_{\boldsymbol{\mu},\Sigma}$ the cumulative distribution function of a two dimensional Gaussian random variable.
The density of the copula is given by:

$$c(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2((\Phi^{-1}(u_1))^2 + (\Phi^{-1}(u_2))^2) - 2\rho\Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^2)}\right)$$

Another important pair-copula which we will need is the Student copula:

Definition 3.3 (Student copula).
The Student copula is defined as the copula of the multivariate random variable (X_1, X_2) with multivariate Student distribution and parameters ν, $\boldsymbol{\mu}$ and Σ :

$$C_{\rho,\nu}(u_1, u_2) = t_{\nu,\boldsymbol{\mu},\Sigma}(t_{\nu,\mu_1,\sigma_1}^{-1}(u_1), t_{\nu,\mu_2,\sigma_2}^{-1}(u_2))$$

It has the following density:

$$c(u_1, u_2)_{\rho,\nu} = \frac{f_{\nu,\Sigma}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))}{f_\nu(t_\nu^{-1}(u_1)) \cdot f_\nu(t_\nu^{-1}(u_2))}$$

Where $f_{\nu,\Sigma}$ is the density of the 2 dimensional Student distribution, and t_ν is the distribution function of the one dimensional Student distribution.

An important observation is that the Gaussian copula has the parameter ρ and the Student copula the parameters ρ and ν. The Gaussian copula and the Student copula belong to the class of **elliptical copulas**.

Another important copula, is the Frank copula, which is an example of **archimedean copulas**:

Definition 3.4 (Frank copula)**.**
The Frank copula with parameter θ is defined as:

$$C(u_1, u_2) = -\frac{1}{\theta} log \left(1 + \frac{(e^{-\theta u_1} - 1) \cdot (e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$$

We will now derive the Kendall's τ which describes the strength of dependency among two random variables. It will thus be useful to describe the extent of dependency of a pair-copula and make the strength of dependencies of different pair-copulas better to compare.

Definition 3.5 (Kendall's τ)**.**
Let (X, Y) be a vector of two random variables and $(X^\star, Y^\star)$, an independent random variable with the same distribution. Then **Kendall's τ** *is defined as:*

$$\tau(X, Y) = \mathbb{P}[(X - X^\star) \cdot (Y - Y^\star) > 0] - \mathbb{P}[(X - X^\star) \cdot (Y - Y^\star) < 0]$$

As this definition is a theoretical value, but we will work with empirical data, we also need an empirical version of Kendall's τ, which is given below:

Definition 3.6 (Empirical Kendall's τ)**.**
 The empirical estimate of Kendall's τ, $\hat{\tau}(\mathbf{x}, \mathbf{y})$, for $\mathbf{x} = (x_i)_{\{i=1,\ldots,n\}}, \mathbf{y} = (y_i)_{\{i=1,\ldots,n\}}$ is defined as:

$$C := \#\{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\} | x_i < x_j \text{ and } y_i < y_j\}$$
$$D := \#\{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\} | x_i < x_j \text{ and } y_i > y_j\}$$
$$T_x := \#\{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\} | x_i = x_j \text{ and } y_i \neq y_j\}$$
$$T_y := \#\{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\} | x_i \neq x_j \text{ and } y_i = y_j\}$$
$$\hat{\tau}(\mathbf{x}, \mathbf{y}) := \frac{C - D}{\sqrt{(C + D + T_x) \cdot (C + D + T_y)}}$$

Kendall's τ measures the strength of the co-movement of two random variables as the Pearson correlation ρ does. The relation of ρ and τ for Student copulas is answered by Demarta and McNeil, 2005:

Lemma 3.1 (Pearson correlation and Kendall's τ of Student copula)**.**
The Pearson correlation is a function of τ given by:

$$\rho(\tau) = \sin\left(\tau \cdot \frac{\pi}{2}\right)$$

One advantage of the definition of Kendall's τ compared to the classical Pearson correlation is, that its value is not changed by strictly monotone increasing transformations:

Lemma 3.2 (Invariance of Kendall's τ to strictly monotone increasing transformations)**.**
Let t be a strictly monotone increasing function. If we apply the transformation to X, Y, the value of τ does not change.

Proof. As t is a strictly monotone increasing function, we have:

$$t(x) > t(y) \Leftrightarrow x > y \tag{3.1}$$

$$
\begin{aligned}
\tau(t(X), t(Y)) =& \mathbb{P}[(t(X) - t(X^\star)) \cdot (t(Y) - t(Y^\star)) > 0] \\
& - \mathbb{P}[(t(X) - t(X^\star)) \cdot (t(Y) - t(Y^\star)) < 0] \\
=& \mathbb{P}[(t(X) > t(X^\star), t(Y) > t(Y^\star)) \cup (t(X) < t(X^\star), t(Y) < t(Y^\star))] \\
& - \mathbb{P}[(t(X) > t(X^\star), t(Y) > t(Y^\star)) \cup (t(X) < t(X^\star), t(Y) < t(Y^\star))] \\
\overset{(3.1)}{=}& \mathbb{P}[(X > X^\star, Y > Y^\star) \cup (X < X^\star, Y < Y^\star)] \\
& - \mathbb{P}[(X > X^\star, Y > Y^\star) \cup (X < X^\star, Y < Y^\star)] \\
=& \tau(X, Y)
\end{aligned}
$$

$\square$

One important conclusion of **Lemma 3.2** is that the variables on the **U-Level** and the original variables with arbitrary margins have the same Kendall's τ values, as the distribution functions are always monotonously increasing.

Furthermore, **Lemma 3.2** shows that the Kendall's τ value of $\mathbf{U}_1, \mathbf{U}_2$ and their Z-transformations $\mathbf{Z}_1, \mathbf{Z}_2$ are the same, as the Z-transformation is strictly increasing. Thus, we will use the Kendall's τ value to measure association of two random variables, as we can now switch between the **U-Level** and the **Z-Level**, without changes in the Kendall's τ value.

The reasons above show that the Kendall's τ values are appropriately used in order to describe the dependency among two random variables in our setting, as monotone increasing transformations are frequently applied.

A definition, which is important for the analysis of pair-copulas, is the symmetry:

Definition 3.7 (Symmetric pair-copulas)**.**
We call pair-copulas symmetric, if they are symmetric in their parameters:

$$C(u,v) = C(v,u)$$

While Kendall's τ measures the average dependence between two random variables, we also want to define a measure for the dependence in the "extremes". Therefore, we introduce the upper and lower tail dependence.

Definition 3.8 (Upper and Lower Tail dependence)**.**
Let U_1, U_2 be two random variables with uniform marginal distributions. We define the upper tail dependence as:

$$\lambda^{upper} = \lim_{t \nearrow 1} \mathbb{P}[U_1 > t | U_2 > t]$$

We define the lower tail dependence as:

$$\lambda^{lower} = \lim_{t \searrow 0} \mathbb{P}[U_1 \leq t | U_2 \leq t]$$

For symmetric copulas, as we use them in this thesis, the value of the lower and the upper tail dependence does not change if we interchange the order of U_1 and U_2. Gaussian, Student and Frank copulas are important examples for symmetric copulas.
In our setting the upper-tail dependence is of interest, as we will simulate random variables with positive dependencies and conditioning variables set to high quantiles.
One can interpret the upper-tail dependence as a measure of the extremeness of the pair-copula, as it denotes the limit of the conditional probability of U_1 lying over a high quantile, given that U_2 is over that high quantile. In other words, given that U_2 is very extreme, how high is the probability that U_1 is also very extreme.
We will now discuss the tail dependence of the copulas introduced in this chapter:

Lemma 3.3 (Tail dependence of Gaussian copula and Frank copula)**.**
Random variables with Gaussian copula or Frank copula have an upper and lower tail dependence of 0.

Thus, **Lemma 3.3** shows, that Gaussian and Frank copulas do not cover extreme dependencies. However, the Student copula has a tail dependence and thus covers extremer dependencies.

Lemma 3.4 (Tail dependence of Student copula)**.** *The upper and lower tail dependence for a Student copula with parameters* ρ, ν *is given by:*

$$\lambda = 2t_{\nu+1}\left(-(\sqrt{\nu+1}) \cdot \sqrt{\frac{1-\rho}{1+\rho}}\right)$$

As the Student copula has two parameters, we want to study both of the parameters in order to be able to interpret the resulting copula better.

If we analyse, for a fixed value of ρ, the tail dependence of a Student copula, we get the following result:

Lemma 3.5 (Tail dependence of Student copula)**.** *For fixed values of* $\rho \in (-1, 1)$, *the upper and lower tail dependence is given by:*

$$\lambda = 2t_{\nu+1}\left(-(\sqrt{\nu+1}) \cdot \sqrt{\frac{1-\rho}{1+\rho}}\right)$$

is a strictly ***decreasing*** *function in* ν.

Proof. As it has been shown in Demarta and McNeil, 2005, in Theorem 4.3, the tail dependence of two random variables with Student copula can be written as:

$$\lambda(\nu) = \frac{\int_{(\pi/2-arcsin(\rho_{i,j}))/2}^{\pi/2} cos^{\nu}(t)dt}{\int_0^{\pi/2} cos^{\nu}(t)dt} \tag{3.2}$$

where $\psi := (\pi/2 - arcsin(\rho_{i,j}))/2 \in [0, \pi/2]$.
We now want to show that λ is a monotonously decreasing function in ν. This is equivalent to

$$\lambda(\nu) > \lambda(\nu + \epsilon) \quad \forall \epsilon > 0, \nu > 2 \tag{3.3}$$

We will proof inequality (3.3) using equation (3.2).

$$\lambda(\nu) > \lambda(\nu + \epsilon)$$

$$\Leftrightarrow \frac{\int_\psi^{\pi/2} cos^\nu(t)dt}{\int_0^{\pi/2} cos^\nu(t)dt} > \frac{\int_\psi^{\pi/2} cos^{\nu+\epsilon}(t)dt}{\int_0^{\pi/2} cos^{\nu+\epsilon}(t)dt}$$

$$\Leftrightarrow \frac{\int_0^{\pi/2} cos^{\nu+\epsilon}(t)dt}{\int_\psi^{\pi/2} cos^{\nu+\epsilon}(t)dt} > \frac{\int_0^{\pi/2} cos^\nu(t)dt}{\int_\psi^{\pi/2} cos^\nu(t)dt}$$

$$\Leftrightarrow 1 + \frac{\int_0^{\psi} cos^{\nu+\epsilon}(t)dt}{\int_\psi^{\pi/2} cos^{\nu+\epsilon}(t)dt} > 1 + \frac{\int_0^{\psi} cos^\nu(t)dt}{\int_\psi^{\pi/2} cos^\nu(t)dt}$$

$$\Leftrightarrow \int_0^{\psi} cos^{\nu+\epsilon}(t)dt \cdot \int_\psi^{\pi/2} cos^\nu(t)dt > \int_0^{\psi} cos^\nu(t)dt \cdot \int_\psi^{\pi/2} cos^{\nu+\epsilon}(t)dt \qquad (3.4)$$

It is thus sufficient to proof inequality (3.4).
To proof this inequality we note:

$$cos^\epsilon(\psi) = min_{t\in[0,\psi]}(cos^\epsilon(t)) \qquad (3.5)$$

$$cos^\epsilon(\psi) = max_{t\in[\psi,\pi/2]}(cos^\epsilon(t)) \qquad (3.6)$$

Thus, we can proof inequality (3.4) as follows:

$$\int_0^{\psi} cos^{\nu+\epsilon}(t)dt \cdot \int_\psi^{\pi/2} cos^\nu(t)dt = \int_0^{\psi} cos^\nu(t) \cdot cos^\epsilon(t)dt \cdot \int_\psi^{\pi/2} cos^\nu(t)dt$$

$$\overset{(3.5)}{>} \int_0^{\psi} cos^\nu(t) \cdot cos^\epsilon(\psi)dt \cdot \int_\psi^{\pi/2} cos^\nu(t)dt$$

$$= \int_0^{\psi} cos^\nu(t)dt \cdot \int_\psi^{\pi/2} cos^\nu(t) \cdot cos^\epsilon(\psi)dt$$

$$\overset{(3.6)}{>} \int_0^{\psi} cos^\nu(t)dt \cdot \int_\psi^{\pi/2} cos^\nu(t) \cdot cos^\epsilon(t)dt$$

$$= \int_0^{\psi} cos^\nu(t)dt \cdot \int_\psi^{\pi/2} cos^{\nu+\epsilon}(t)dt$$

$$\square$$

This theorem gives us an interpretation of the role of the parameter ν in the Student copula: The lower the value of ν, the stronger the tail dependence and thus extreme co-movements occur more often. The result of **Lemma 3.5** will be used in **Section 9.4**

to analyse the results.

The complete list of pair-copulas used in this thesis is given in **Table 1**:

Name	Number of parameters
Gaussian	1
Student t	2
Clayton (with 90°,180°,270° rotations)	1
Gumbel (with 90°,180°,270° rotations)	1
Joe (with 90°,180°,270° rotations)	1
Frank	1
Clayton-Gumbel (with 90°,180°,270° rotations)	2
Joe-Gumbel (with 90°,180°,270° rotations)	2
Joe-Clayton (with 90°,180°,270° rotations)	2
Joe-Frank (with 90°,180°,270° rotations)	2
BB1 (with 90°,180°,270° rotations)	2
BB7 (with 90°,180°,270° rotations)	2
BB8 (with 90°,180°,270° rotations)	2

Table 1: List of pair-copulas used in this thesis

The total number of copulas used is 36.

3.3   Pair-Copula Constructions

In this section we will show a method to construct multidimensional copulas out of pair-copulas. For more details we refer to Bedford and Cooke, 2002, Aas et al., 2009 and Kurrowicka and Joe, 2011.

It is a well known fact that in case of **independence** of a d-dimensional vector we can write its density as the product of its marginal densities:

$$\pi(\mathbf{x}) = \prod_{k=1}^{d} \pi_k(x_k) \tag{3.7}$$

As we have seen in **Theorem 3.1** and **Example 3.1** the independence of variables and the associated independence copula is only a special form of the general copula setting.

Thus, in this chapter we will show a generalized method to decompose densities in to products similar to equation (3.7).

At first we observe that we can rewrite the density as follows:

$$\pi(\mathbf{u}) = \frac{\pi(u_1, ..., u_d)}{\pi(u_1, ..., u_{d-1})} \cdot \pi(u_1, ..., u_{d-1})$$

$$= \pi(u_d | u_1, ..., u_{d-1}) \cdot \pi(u_1, ..., u_{d-1})$$

If one iterates this procedure, the following general decomposition can be obtained:

$$\pi(u_1, ..., u_d) = \left[\prod_{r=2}^{d} \pi(u_r | u_1, ...u_{r-1}) \right] \cdot \pi(u_1) \tag{3.8}$$

One can note, that the decomposition in (3.8) is not unique, but other orders of conditioning are possible, too.

As a second component to construct general R-vines, we need to apply **Sklar's Theorem**. As the cumulative distribution function of uniform variables is the identity function, we get:

$$\pi(u_1, u_2) = c_{1,2}(u_1, u_2)$$

and thus we get the following result:

$$\pi(u_2 | u_1) = \frac{c_{1,2}(u_1, u_2)}{\pi(u_1)} = c_{1,2}(u_1, u_2)$$

In three dimensions we get:

$$\begin{aligned}
\pi(u_3 | u_1, u_2) &= \frac{\pi(u_1, u_2, u_3)}{\pi(u_1, u_2)} \\
&= \frac{\pi(u_2, u_3 | u_1)}{\pi(u_2 | u_1)} \\
&= \frac{c_{2,3;1}(F_{2|1}(u_2|u_1), F_{3|1}(u_3|u_1))\pi(u_2|u_1)\pi(u_3|u_1)}{\pi(u_2|u_1)} \\
&= c_{2,3;1}(F_{2|1}(u_2|u_1), F_{3|1}(u_3|u_1))c_{1,2}(u_1, u_2)
\end{aligned}$$

The general form of conditional densities is thus:

$$\pi(v|\mathbf{u}) = c_{v,u_j;\mathbf{u}_{-j}}(F(v|\mathbf{u}_{-j}), F(u_j|\mathbf{u}_{-j})) \cdot \pi(v|\mathbf{u}_{-j}) \tag{3.9}$$

Equations (3.9) and (3.8) lead to the topic of **pair-copula constructions**. As this theory is very complicated, we cannot develop the whole theory in this chapter. However, we state the definitions and the construction of regular vines, as we will need this in the following chapters.

At first we give the formal definition of a regular vine which we will use in this thesis.

Definition 3.9 (Regular Vine Tree Sequence)**.** *A regular vine tree sequence is a set of trees $T_1, ..., T_{d-1}$ on d elements with nodes N_i and edges E_i for $i = 1, ..., d-1$ with:*

1. *T_1 has nodes $N_1 = \{1, ..., d\}$ and edges E_1*

2. *For $i = 2, ..., d-1$ the trees T_i has nodes $N_i = E_{i-1}$ and edge set E_i*

3. *(proximity condition) For $i = 2, ..., d-1$ If two edges in tree T_i are to be joined by an edge in tree T_{i+1} they must share a common node:*
   $\{a, b\} \in E_i$ with $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ it must hold that $\#(a \cap b) = 1$

Definition 3.10 (Regular Vine Copula)**.** *A regular vine copula $RV = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \boldsymbol{\theta}(\mathcal{B}(\mathcal{V})))$ on d variables is a copula with uniform margins, such that:*

1. *$\mathcal{V}$ is a regular vine on d elements as defined in **Definition 3.9***

2. *$\mathcal{B}(\mathcal{V}) = \{C_{i(e),j(e)|D(e)}|e \in E_m, m = 1, ..., d-1\}$ is a set of $d(d-1)/2$ copula families identifying the conditional distribution of $U_{i(e)}, U_{j(e)}|\mathbf{U}_{D(e)}$*

3. *$\boldsymbol{\theta}(\mathcal{B}(\mathcal{V})) = \{\boldsymbol{\theta}_{i(e),j(e)|D(e)}|e \in E_m, m = 1, ..., d-1\}$ is the set of parameter vectors, corresponding to the copulas in $\mathcal{B}(\mathcal{V})$.*

The index i of the specific tree in the tree sequence $(T_i)_{i=1,...,d-1}$ will be called the **levels** in the R-vine.

The triplet $RV = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \boldsymbol{\theta}(\mathcal{B}(\mathcal{V})))$ specifies the regular vine copula (**R-vine**). We thus need to focus on the selection and estimation of each element of the triplet.

To derive the vine-copula model properly we further need the definition of h-functions:

Definition 3.11 (h-functions)**.**
The h-functions are defined as:

$$h(u|\mathbf{v}, \boldsymbol{\theta}) := F(u|\mathbf{v}) = \frac{\partial C_{uv_j|\mathbf{v}_{-j}}(F(u|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})|\boldsymbol{\theta})}{\partial F(v_j|\mathbf{v}_{-j})}$$

With this definitions we can now write the general form of the decomposition of multivariate densities on the U-Level:

$$\pi(\mathbf{u}) = \prod_{j=1}^{d-1} \prod_{e \in E_j} c_{a_e,b_e|D_e}(F_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e}), F_{b_e|D_e}(u_{b_e}|\mathbf{u}_{D_e})) \quad (3.10)$$

A special case of the R-vine is the **C-vine**:

Definition 3.12 (C-vine).
As C-vines we define the set of R-vines whose trees are stars (for definition of stars see Tutte, 1984). Its density is given by:

$$\pi(\mathbf{x}) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1:(j-1)}(F(u_j|u_1,...,u_{j-1}), F(u_{j+i}|u_1,...,u_{j-1})|\boldsymbol{\theta}_{j,i+j|1:(j-1)})$$

An important conclusion from this decomposition is that the density of a copula is uniquely determined by its pair-copulas. We can thus construct much more flexible copulas by pair-copula constructions.

3.4 Model selection

In this section we will develop methods to construct R-vine copulas.
One theoretic possibility to choose the R-vine copula would be to apply the maximum likelihood estimation directly, as the density is available in a closed form solution. Thus, one could choose the R-vine copula.
This is, however, very slow in higher dimensions.
Therefore, we can divide the R-vine copula specification into 3 parts, which we will cover in this section:

- Selecting of the R-vine structure

- Choosing the bivariate copula families for each pair in the R-vine structure

- Estimating the corresponding parameters for each copula

3.5   Parameter estimation

To specify the pair-copulas, we need to choose the pair-copula families and their parameters.

At first we show how the parameters for a pair-copula is estimated. We will use two different methods:

Definition 3.13 (Maximum likelihood estimation)**.**
The Maximum likelihood estimation of the parameter $\boldsymbol{\theta}$ of a pair-copula, given two vectors $\mathbf{u} = (u_i)_{i=1,\dots,d}, \mathbf{v} = (v_i)_{i=1,\dots,d}$, is defined as:

$$\hat{\boldsymbol{\theta}} := argmax_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{d} \log(c(u_i, v_i|\boldsymbol{\theta}))$$

Definition 3.14 (Inversion of Kendall's τ)**.**
The estimation of $\boldsymbol{\theta}$ by inversion of Kendall's τ of a pair-copula, for two vectors $\mathbf{u}$ and $\mathbf{v}$ is only well defined, if $\boldsymbol{\theta}$ can be written as a function of τ. It is defined as:

$$\hat{\boldsymbol{\theta}} := \boldsymbol{\theta}(\hat{\tau}(\mathbf{u}, \mathbf{v}))$$

The advantage of the inversion of Kendall's τ is that this method is much faster, as it does not require a numerical optimization and the values of Kendall's τ are already given by the Dissmann algorithm.
Given the parameters $\hat{\boldsymbol{\theta}}$ we can now evaluate the AIC for each of the parametric pair-copula families as explained in Dissmann et al., 2013. Following Akaike, 1998, we define the AIC for a pair-copula as:

Definition 3.15 (AIC criterion for pair-copulas)**.**
Given the two vectors $\mathbf{u}, \mathbf{v}$ and their copula $c(\cdot, \cdot|\boldsymbol{\theta})$, we define the AIC as:

$$AIC = -2\sum_{i=1}^{d} ln[c(u_i, v_i|\boldsymbol{\theta})] + 2|\boldsymbol{\theta}|$$

According to the AIC criterion, the pair-copula with the lowest value is chosen.

3.6   Sequential estimates

We have so far focused on the estimation of pair-copulas. We will now estimate entire R-vines.
The direct maximum likelihood estimation of the R-vine copula is possible, however the vector of parameters has too many parameters to be easily numerically calculable. Thus,

we apply tree-wise algorithms which work successively on the levels of the R-vine: On each level the tree is chosen first. The second step is to choose the pair-copulas as described above. After this, the algorithm proceeds with the transformed variables satisfying the conditions in **Definition 3.9**.

A common choice for the procedure to choose the tree sequence is **Algorithm 1**, the Dissmann algorithm as proposed in Dissmann et al., 2013. This algorithm requires a functional measure to quantify the strength of dependence between two random variables. The usual function which is used in this context is the Kendall's τ value.

Algorithm 1 Dissmann algorithm

Choose an R-vine model structure based on Kendall's τ

Require: Set of vectors$\{\mathbf{u}_i, \mathbf{u}_j\}, 1 \leq i < j \leq d$ with $\mathbf{u}_i \in \mathbb{R}^n$ uniformly distributed, set of proposed pair-copulas $\mathcal{C}$

1: **for** $i \leftarrow 1$ **to** d **do**
2:    **for** $j \leftarrow 1$ **to** d **do**
3:       Calculate the empirical Kendall's $\hat{\tau}_{i,j} = \tau(\mathbf{u}_i, \mathbf{u}_j)$
4:    **end for**
5: **end for**
6: Select the spanning tree that maximizes the sum of absolute empirical Kendall's τ:

$$T_k = argmax \sum_{e=\{i,j\}\in T_k} |\hat{\tau}_{i,j}|$$

7: For all pairs of neighbouring variables in the tree, estimate the parameters for the proposed pair-copulas and select a copula according to the AIC criterion.
8: Transform the observations using the h-functions of the chosen pair-copula and parameter from Step 7 to obtain the transformed values.
9: Use these transformed observations to calculate the empirical Kendall's τ for all pairs fulfilling the proximity condition.
10: Proceed with Step 6 until the R-vine is fully specified.

For more details about the algorithm see Dissmann et al., 2013.

3.7   Interpretation of pair plots

As we will use different plots for visualization and for validation of our algorithms, we will show examples of contour plots on the **Z-Level** in this section, such that the results can be interpreted more easily, as it can be compared with these plots.
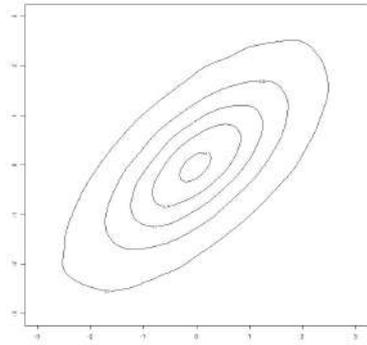
(a1) Z-Level contour plot: Gaussian copula ($\tau = 0.5$)
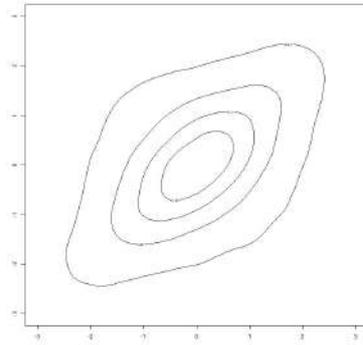
(a2) Z-Level contour plot: Student copula ($\tau = 0.5$, df=4)
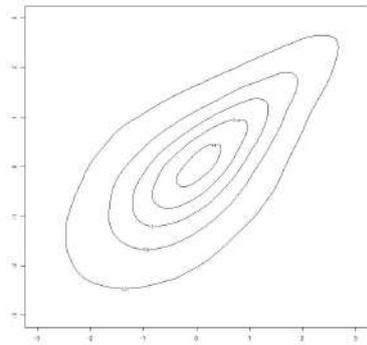
(b1) Z-Level contour plot: Gumbel copula ($\tau = 0.5$)
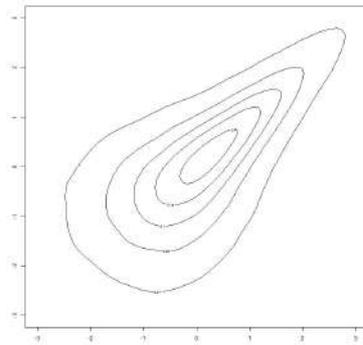
(b2) Z-Level contour plot: Joe copula ($\tau = 0.5$)
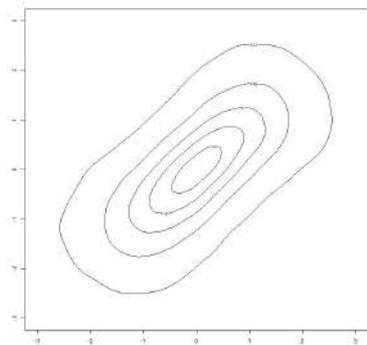
(c1) Z-Level contour plot: Frank copula ($\tau = 0.5$)
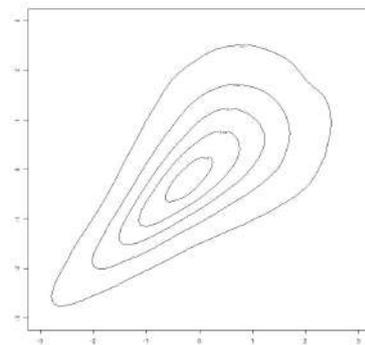
(c2) Z-Level contour plot: Clayton copula ($\tau = 0.5$)

Figure 2: Summary of Z-Level contour plots for different pair-copulas

One can see in **Figure 2** that very different shapes can result from the choices of different copulas with the same value of Kendall's τ.

The Clayton copula will be used to show the rotations of copulas. These rotations allow us to fit a bigger scale of different dependencies, as shown in **Figure 3**:
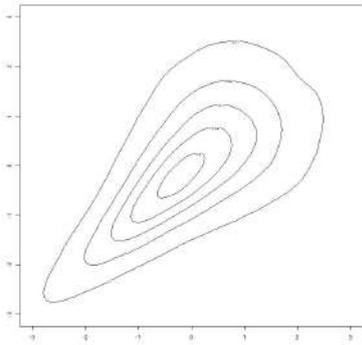
(a1) Z-Level contour plot
Clayton copula ($\tau = 0.5$)
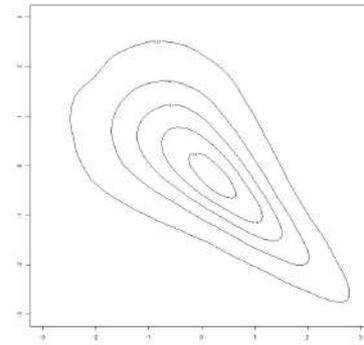
(a2) Z-Level contour plot: 90° rotated
Clayton copula ($\tau = -0.5$)
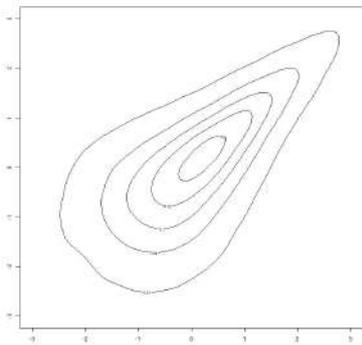
(b1) Z-Level contour plot: 180° rotated
Clayton copula ($\tau = 0.5$)
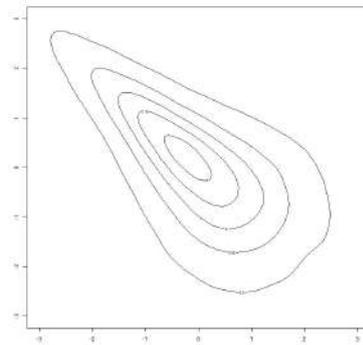
(b2) Z-Level contour plot: 270° rotated
Clayton copula ($\tau = -0.5$)

Figure 3: Summary of Z-Level contour plots for rotated Clayton copulas

4 Monte Carlo methods

In this chapter, we will introduce Monte Carlo methods, a widely used class of stochastic algorithms. These algorithms approximate functions by stochastic simulations, often with Markov chains, which we will therefore introduce in the first section of this chapter, before we cover Monte Carlo methods in the second part.

The main area of application of Monte Carlo methods are problems, where no closed form solution is available, as described in Gilks et al., 1998. For this reason, we will apply Monte Carlo methods to perform the stress tests, as closed form solutions will be not available here.

For further details we refer to Roberts et al., 2004.

4.1 Markov Chains

We first want to define discrete Markov Chains as follows:

Definition 4.1. *(Time discrete Markov chain)*
A stochastic process $(\mathbf{X}_r)_{r \in \mathbb{N}}$ with values in $\mathbb{R}^d$ is called a (time discrete) Markov chain, if for all (Lebesgue) measurable sets $A \subset \mathbb{R}^d$ and all $r \in \mathbb{N}$ the following equation holds:

$$\mathbb{P}[\mathbf{X}_{r+1} \in A | \mathbf{X}_0, ..., \mathbf{X}_r] = \mathbb{P}[\mathbf{X}_{r+1} \in A | \mathbf{X}_r]$$

This equation denotes that for a Markov chain, given the present, the future and the past of the process are conditionally independent.

Definition 4.2. *(Irreducible stationary distribution)*
A Markov chain $(\mathbf{X}_n)_{n \geq 1}$ has a stationary distribution π if for all $r \geq 1$:

$$\mathbf{X}_r \sim \pi \Rightarrow \mathbf{X}_{r+1} \sim \pi$$

The Markov chain is further irreducible w.r.t. π if:

$$\mathbb{P}[\exists n \in \mathbb{N} : \mathbf{X}_n \in A | X_0 = x] > 0 \text{ for all } x \in \mathbb{R}^d$$

For all $A \subset \mathbb{R}^d$ such that $\pi(A) > 0$

A further rather technical definition is the aperiodicity of Markov chains:

Definition 4.3. *(Aperiodic Markov chains)*
A Markov chain $(\mathbf{X}_n)_{n\geq 1}$ is aperiodic, if for all sequences $\{E_0, ..., E_{m-1}\}$, such that
for all $j = i + 1 \mod m$ with $\mathbb{P}[\mathbf{X}_{i+1} \in E_{i+1}|\mathbf{X}_0 = x] = 1 \; \forall x \in E_i$ it holds:
The length of the sequence m is 1.

A problem that we will face in **Chapter 5** is that the simulation of random variables is not directly possible, but we can construct a Markov chain, which converges in distribution towards a process with the desired distribution. To quantify the difference between the desired distribution and the distribution of the realizations, we need to introduce a distance on the space of distributions:

Definition 4.4. *(Total variation)*
Let μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The total variation is defined as:

$$||\mu||_{TV} = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \mu(A)$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel algebra on $\mathbb{R}^d$

The following theorem denotes the limiting properties of a great class of Markov chains:

Theorem 4.1. *(Convergence of Markov chains)*
For an irreducible and aperiodic Markov chain $(\mathbf{X}_n)_{n\geq 1}$ with stationary distribution π and values in $\mathbb{R}^d$, it holds that:

$$\lim_{n\to\infty} ||\pi(\cdot) - \mathbb{P}[\mathbf{X}_n \in \cdot|\mathbf{X}_0 = x]||_{TV} = 0$$

for π- almost all x.

Theorem 4.1 implies, that: $\lim_{n\to\infty} \mathbb{P}[\mathbf{X}_n \in A|\mathbf{X}_0 = x] = \pi(A)$ for all measurable sets $A \subset \mathbb{R}^d$.
We will thus approximate the distribution π as follows:

$$\pi(A) \approx \mathbb{P}[\mathbf{X}_n \in A|\mathbf{X}_0 = x]$$

As the convergence is only valid as a limit, we need a sufficiently big values of n.

4.2   Monte Carlo methods

Monte Carlo methods, which make use of Markov chains, approximate expectation values, that are not possible to evaluate in closed form, in the following way:

Definition 4.5. *(Monte Carlo method)*
Let $\mathbf{X}_1, ..., \mathbf{X}_R$ be random variables in $\mathbb{R}^d$ with distribution π, and let $h : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. We approximate $\mathbb{E}_\pi[h(\mathbf{X})]$ as follows:

$$\frac{1}{R} \sum_{r=1}^{R} h(\mathbf{X}_r) = \overline{h}_R \approx \mathbb{E}_\pi[h(\mathbf{X})]$$

One example is the estimation of a distribution function by:

$$\widehat{F}_R(x) = \frac{1}{R} \sum_{r=1}^{R} \mathbb{I}_{\{\mathbf{X}_r \leq x\}}$$

If we assume, that the random variables $(\mathbf{X}_r)_{r \in \mathbb{N}}$ are stochastically independent, the convergence of the average follows from the strong law of large numbers. However, in our framework the random variables are usually not independent, but are realizations of a Markov chain (see Gilks et al., 1998). In this case, proofs of the convergence can be found for a wide class of Markov chains, as described in Atchad et al., 2009.

Furthermore, the Markov chain is not distributed according to π, but converges towards it. If the distribution of the first B realizations of the Markov chain is too far away from the stationary distribution, we delete this values from the sample for which we apply the Monte Carlo approximation. This procedure is called Markov chain Monte Carlo method with **burn-in**:

Definition 4.6. *(Markov chain Monte Carlo method with burn-in)*
Let $\mathbf{X}_1, ..., \mathbf{X}_R$ be realizations of a Markov chain with stationary distribution π with values in $\mathbb{R}^d$. Let $h : \mathbb{R}^d \to \mathbb{R}$ be a measurable function and burn-in $B \in \{0, ..., R-1\}$. We approximate $\mathbb{E}_\pi[h(\mathbf{X})]$ as follows:

$$\frac{1}{R-B} \sum_{r=1+B}^{R} h(\mathbf{X}_r) \approx \mathbb{E}_\pi[h(\mathbf{X})]$$

Higher values of B and $R - B$ improve the convergence, but it has to be decided which values should be chosen.

Two essential questions arise from this approach:

- How long does the burn-in B have to be, such that the Markov chain is sufficiently close to the limiting distribution?

- How long does the length $R - B$ of the simulations after burn-in have to be to achieve a sufficient convergence of the estimate?

These questions are not solved in general, as we will see in **Section 5.3**. It is not possible to proof this question in general, however some theoretical results which can be used will be derived. Because of the lack of theoretical proofs of the convergence, other methods were developed to evaluate the convergence by statistical tests, which we describe in the following section.

4.3 Full convergence diagnostics

In this section, we will summarize some methods to validate the convergence of the Monte Carlo algorithm, as in our case, it is not possible to prove the convergence theoretically.

As a statistical test for the length of the burn-in for the Markov chain we apply the **Heidelberger and Welch stationarity test** (for further details see Geyer, 2012). For the test we require the definition of stationarity:

Definition 4.7 (Stationary processes)**.**
A process $(\mathbf{X}_r)_{r \in \mathbb{N}}$ with values in $\mathbb{R}^d$ is strictly stationary, if for all $r, t, s \in \mathbb{N}$, $(\mathbf{x_1}, ..., \mathbf{x_{t+1}})$ with $\mathbf{x}_i \in \mathbb{R}^d$:

$$F_{\mathbf{X}_{r,...,\mathbf{X}_{r+t}}}(\mathbf{x}_1, ..., \mathbf{x}_{t+1}) = F_{\mathbf{X}_{r+s,...,\mathbf{X}_{r+s+t}}}(\mathbf{x}_1, ..., \mathbf{x}_{t+1})$$

The null hypothesis of the test is that the Markov chain is a realization of a stationary process, which would be the case if the Markov chain would have converged. If the null hypothesis is rejected, the first 1% of the chain are deleted as a burn-in and the test is repeated. We perform the Heidelberger- Welch test, with a confidence level of 99% (see Cowles and Carlin, 1996).
This test will be applied to all margins and the burn-in is taken as the maximum of all proposed burn-in values. This leads to multiple testing, which is a conservative approach, as it leads to higher burn-in values.

Another problem, which we will cover in this chapter, is the possibility of very high correlation within the simulated Markov chain. To measure these effects, we introduce the

following approach:

Assuming that we want to estimate a function h of the first R elements in our process $(\mathbf{X}_r)_{1 \leq r \leq R}$, we can use the ergodic average.

$$\overline{h}_R = \frac{1}{R} \sum_{r=1}^{R} h(\mathbf{X}_r)$$

To analyze the ergodic average, we define the autocorrelation of a process at different time points.

Definition 4.8 (Autocorrelation).
The autocorrelation of a process $(\mathbf{X}_r)_{r \geq 1}$ with values in $\mathbb{R}^d$ is measured by
$\rho : \mathbb{N} \times \mathbb{N} \to [-1, 1]^d$:

$$\rho(t, s) = \frac{\mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu}_t) \cdot (\mathbf{X}_s - \boldsymbol{\mu}_s)]}{\sqrt{\mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu}_t)^2] \cdot \mathbb{E}[(\mathbf{X}_s - \boldsymbol{\mu}_s)^2]}}$$

Definition 4.9 (White noise).
A stochastic process is called white noise if its autocorrelation function satisfies:

$$\rho(t, s) = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

As the autocorrelation measures the inter time dependence of a process, one sees that the white noise has no time correlation.

To measure the convergence of the ergodic average, we want to get a formula for the variance of $\overline{h}_R$. We follow the approach of Burke, 2012, who uses the assumption, that the Markov chain $\mathbf{X}_r$ is an AR(1) process :

Definition 4.10 (AR(1) processes).
An AR(1) process is a stochastic process $(\mathbf{X}_r)_{r \in \mathbb{N}}$ in $\mathbb{R}^d$ with constants
$c \in \mathbb{R}$, $\phi \in (-1, 1)$:

$$\mathbf{X}_r = c + \phi \mathbf{X}_{r-1} + \boldsymbol{\epsilon}_r$$

Where $\boldsymbol{\epsilon}_r$ is a white noise process with constant variance.

This assumption is not generally true for Markov chains, as our increments are not white noise. However, this can be used as an approximation.

In the case of a one dimensional stationary AR(1) process, the autocorrelation has the following easy form:

Theorem 4.2. *The auto covariance of a one dimensional stationary AR(1) process has the following form:*

$$\rho(t, s) = \rho^{|t-s|}$$

with $\rho \in (-1, 1)$

This theorem shows that the autocorrelation of a one dimensional stationary AR(1) process drops exponentially with t for fixed values of s as shown in **Figure 4**. One can see that for higher values of ρ the cross-correlation decreases more slowly.
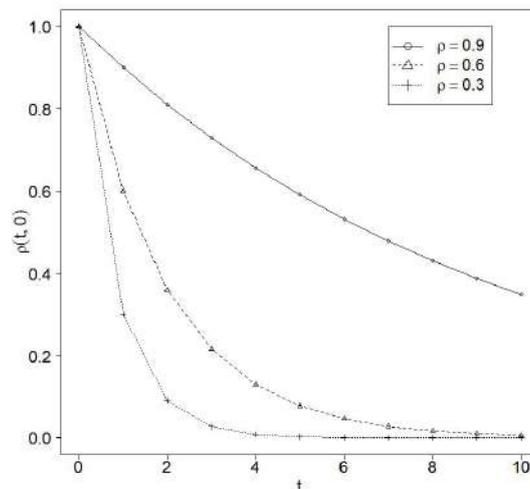
Figure 4: Visualization of $\rho(t, 0)$ for different values of ρ

If we assume our Markov chain to be a stationary AR(1) process, the variance of our estimator satisfies the following equation.

$$Var\left(\overline{h}_R\right) = \frac{1+\rho}{1-\rho}\frac{1}{R}Var(h(\mathbf{X}_1)) \tag{4.1}$$

Formula (4.1) leads to the definition of effective sample size, as described in Burke, 2012. If we divide $Var(h(\mathbf{X}_1))$ by $Var\left(\overline{h}_R\right)$, we get the ratio that measures by how much the

individual variance is larger than the variance of the ergodic average:

$$\frac{Var(h(\mathbf{X}_1))}{Var\left(\overline{h}_R\right)} = R \cdot \frac{1-\rho}{1+\rho}$$

This leads to the definition of the effective sample size:

Definition 4.11 (Effective sample size).
The effective sample size of a sample $(X_r)_{r\in\{1,...,R\}}$ *is defined as*

$$ESS = R \cdot \frac{1-\rho}{1+\rho}$$

where ρ denotes the autocorrelation coefficient in the AR(1) process.

The effective sample size of a sample denotes the number of independent random variables with the same distribution, which contain the same amount of information as the sample has.
High correlation in the sample generated can reduce the amount of information which is contained in the chains.

As an estimate for ESS, we use the empirical version:

Definition 4.12 (Empirical effective sample size).
The empirical effective sample size is defined as:

$$\widehat{ESS} = R \cdot \frac{1-\hat{\rho}}{1+\hat{\rho}}$$

Where $\hat{\rho}$ is the empirical estimate of ρ.

For an upper bound of the confidence interval, we apply Chebyshev's inequality:

Theorem 4.3 (Chebyshev's inequality).
For a random variable $\mathbf{X} \in \mathbb{R}^d$ *with mean* $\boldsymbol{\mu}$ *and finite variance* σ^2 *and all* $\epsilon > 0$ *it holds:*

$$\mathbb{P}\left[|\mathbf{X} - \boldsymbol{\mu}| \geq \epsilon\right] \leq \frac{\sigma^2}{\epsilon^2}$$

Applying Chebyshev's inequality and assuming, that the process is approximately an

$AR(1)$ process, we get the following inequality:

$$\mathbb{P}\left[|\overline{h}_R - \boldsymbol{\mu}| \geq \epsilon\right] \leq \frac{Var(\overline{h}_R)}{\epsilon^2}$$
$$= \frac{1+\rho}{1-\rho}\frac{1}{R}Var(h(\mathbf{X}_1)) \cdot \frac{1}{\epsilon^2}$$
$$= \frac{1}{ESS}Var(h(\mathbf{X}_1)) \cdot \frac{1}{\epsilon^2}$$

Thus, we get the following equation for a significance level α:

$$\frac{1}{ESS}Var(h(\mathbf{X}_1)) \cdot \frac{1}{\epsilon^2} = \alpha$$
$$\Leftrightarrow \epsilon = \sqrt{\frac{1}{ESS}Var(h(\mathbf{X}_1)) \cdot \frac{1}{\alpha}}$$

If we plug in the empirical values, we get an approximate confidence interval to the significance level α:

$$\left[\overline{h}_R + \hat{\sigma}\sqrt{\frac{1}{\widehat{ESS}} \cdot \frac{1}{\alpha}} \ ; \ \overline{h}_R - \hat{\sigma}\sqrt{\frac{1}{\widehat{ESS}} \cdot \frac{1}{\alpha}}\right] \tag{4.2}$$

Equation (4.2) shows that high values of $\widehat{ESS}$ lead to smaller confidence intervals. Thus, $\widehat{ESS}$ can be used as a measure of the information contained in a sample.

If we regard R as fixed, $\widehat{ESS}$ is a decreasing function of $\hat{\rho}$ and can be interpreted as follows: A high autocorrelation effects the effective sample size negatively, as the earlier entries have a strong effect upon the later entries and thus the amount of newly gained information is highly limited.

Thus, a big sample which has a small $\widehat{ESS}$ is not useful, as it requires a lot of storage space, but only gives a limited amount of information. The $\widehat{ESS}$ will therefore be stated for the specific simulations.

5   Metropolis Hastings algorithm

In the last chapter we introduced Monte Carlo methods, which work with Markov chains. In this chapter we will introduce the Metropolis Hastings algorithm, which generates specific Markov chains, which we will use in this thesis.

A common situation when applying Monte Carlo methods is that direct simulation of a d-dimensional random variable is not possible and its density $\pi(\cdot)$ is only known up to a normalizing constant. The most important case of such a situation is the conditional distribution, as explained in **Section 2.2**.

The Metropolis Hastings algorithm (abbreviation MH) is a commonly used algorithm to get samples from random variables, with a density only known up to a multiplicative constant, as described in Chib and Greenberg, 1995.
In this thesis, we will focus on the simulation of random variables with an R-vine dependence structure.

Our specific aim is to sample conditional realizations of a random variable with an R-vine dependence structure, where some entries are set to a fixed value. As we derived in **Chapter 3**, the density of random variables with R-vine dependence structures can be evaluated in closed form, thus their conditional density is known up to a multiplicative constant as shown in equation (2.5). Because direct simulation of these conditional random variables is not possible, we will apply the Metropolis Hastings algorithm.

The key idea of this MH algorithm is to construct a Markov chain $(\mathbf{U}_r)_{r\geq 1}$, the distribution of which converges to its stationary distribution with density $\pi_U(\cdot)$. The realizations of this Markov chain are then chosen as random samples of the stationary distribution with density $\pi_U(\cdot)$.

Example 5.1.
In this example we derive a conditional distribution for which direct simulation is not possible. In **Example 5.3** we will show that the methods which we derive in this chapter enable us to carry out simulations using this distribution.
As a copula, we use a mixture of two copulas with the following C-vine structures:

$$C(u) = 0.5 \cdot C^*(u) + 0.5 \cdot C^{'}(u)$$

	Family	τ		Family	τ		Family	τ
			Level 1					
$C^*_{1,2}$	Student $df = 5$	0.3	$C^*_{1,3}$	Student $df = 5$	0.1	$C^*_{1,4}$	Student $df = 5$	0.5
			Level 2					
$C^*_{2,3|1}$	Student $df = 5$	0.3	$C^*_{2,4|1}$	Student $df = 5$	0.5			
			Level 3					
$C^*_{3,4|1,2}$	Student $df = 5$	0.3						

List of pair-copulas of C^*

	Family	τ		Family	τ		Family	τ
			Tree 1					
$C'_{1,2}$	Student $df = 5$	0	$C'_{1,3}$	Student $df = 5$	0	$C'_{1,4}$	Student $df = 5$	-0.9
			Tree 2					
$C'_{2,3|1}$	Student $df = 5$	0	$C'_{2,4|1}$	Student $df = 5$	0			
			Tree 3					
$C'_{3,4|1,2}$	Student $df = 5$	0						

List of pair-copulas of C'

Table 2: List of pair-copulas of C^* and C'

The density becomes a mixture of π^* and π':

$$\pi_U(u) = 0.5 \cdot \pi^*(u) + 0.5 \cdot \pi'(u)$$

We denote the density of the Student pair-copula by st, the density of the Student distribution by f and the cumulative distribution function of the Student distribution by t. This density π_U has no closed form solution as one can see from the following calculation:

$$\pi_U(u_4|u_1 = 0.1, u_2 = 0.1, u_3 = 0.1) = \frac{\pi_U(u_4, 0.1, 0.1, 0.1)}{\int_{[0,1]} \pi_U(u_4, 0.1, 0.1, 0.1) du_4}$$

To evaluate this expression, we need to evaluate the integral in the denominator:

$$\int_{[0,1]} \pi_U(u_4, 0.1, 0.1, 0.1) du_4$$

$$= 0.5 \cdot \int_{[0,1]} \pi^*(u_4, 0.1, 0.1, 0.1) du_4 + 0.5 \cdot \int_{[0,1]} \pi^{'}(u_4, 0.1, 0.1, 0.1) du_4$$

$$= 0.5 \cdot \int_{[0,1]} st_{\nu=5,\ \tau=0.3}(u, 0.1) \cdot st_{\nu=5,\ \tau=0.1}(u, 0.1) \cdot st_{\nu=5,\ \tau=0.3}(u, 0.1) du \cdot const$$

$$+ 0.5 \cdot \int_{[0,1]} st_{\nu=5,\ \tau=-0.9}(u, 0.1) du$$

$$= 0.5 \cdot \int_{[0,1]} \left[\frac{f_{\nu=5,\tau=0.3}(t_{\nu=5}^{-1}(u), t_{\nu=5}^{-1}(0.1))}{f_{\nu=5}(t_{\nu=5}^{-1}(u)) \cdot f_{\nu=5}(t_{\nu=5}^{-1}(0.1))} \cdot \frac{f_{\nu=5,\tau=0.1}(t_{\nu=5}^{-1}(u), t_{\nu=5}^{-1}(0.1))}{f_{\nu=5}(t_{\nu=5}^{-1}(u)) \cdot f_{\nu=5}(t_{\nu=5}^{-1}(0.1))} \right.$$

$$\left. \cdot \frac{f_{\nu=5,\tau=0.3}(t_{\nu=5}^{-1}(u), t_{\nu=5}^{-1}(0.1))}{f_{\nu=5}(t_{\nu=5}^{-1}(u)) \cdot f_{\nu=5}(t_{\nu=5}^{-1}(0.1))} \right] du \cdot const$$

$$+ 0.5 \cdot \int_{[0,1]} \frac{f_{\nu=5,\tau=-0.9}(t_{\nu=5}^{-1}(u), t_{\nu=5}^{-1}(0.1))}{f_{\nu=5}(t_{\nu=5}^{-1}(u)) \cdot f_{\nu=5}(t_{\nu=5}^{-1}(0.1))} du$$

This integral has no closed form solution. As it is one dimensional, one could integrate it numerically, however this would not be possible in higher dimensions.

As a consequence, there is no closed form expression for the distribution function and the density function. Thus, standard simulation procedures are not applicable and the MH algorithm has to be applied.

Summed up ,the conditional density $\pi_U(u_4|u_1, u_2, u_3)$ is not known in closed form. It is also not possible to simulate this random variable in closed form.

5.1   Nonadaptive Metropolis Hastings algorithm

One specific version of the MH algorithm, as described in Chib and Greenberg, 1995, is the nonadaptive MH algorithm. This algorithm has the advantages of being easy to understand and easy to implement.

This algorithm requires a d-dimensional transition density of the proposal kernel $q_U(\cdot, \cdot) : [0,1]^d \times [0,1]^d \to \mathbb{R}^+$. According to the last state $\mathbf{u}_{r-1}$, a proposed update $\mathbf{u}_r^*$ is sampled in each time step from the conditional proposal distribution with density $q_U(\mathbf{u}_{r-1}, \cdot)$. The acceptance probability $\alpha(\mathbf{u}_{r-1}, \mathbf{u}_r^*)$ of the proposal is given as:

$$\alpha(\mathbf{u}_{r-1}, \mathbf{u}_r^*) = \max\left(1, \frac{\pi_U(\mathbf{u}_r^*)}{\pi_U(\mathbf{u}_{r-1})} \frac{q_U(\mathbf{u}_r^*, \mathbf{u}_{r-1})}{q_U(\mathbf{u}_{r-1}, \mathbf{u}_r^*)}\right).$$

Algorithm 2 Nonadaptive Metropolis Hastings algorithm

Simulate $\mathbf{u}_{1 \leq r \leq R}$, $\mathbf{u}_r \in [0,1]^d$

with $\pi_U(\cdot)$ as the approximate density of $\mathbf{u}_r$

Require: Starting value $\mathbf{u}_0$, d-dimensional transition density of the proposal kernel
$q_U(\cdot, \cdot) : [0,1]^d \times [0,1]^d \to \mathbb{R}_+$, number of runs R

1: **for** $r \leftarrow 1$ **to** R **do**

2:    $\mathbf{u}_r^* \sim q_U(\mathbf{u}_{r-1}, \cdot)$

3:    $\alpha = \dfrac{\pi_U(\mathbf{u}_r^*)}{\pi_U(\mathbf{u}_{r-1})} \dfrac{q_U(\mathbf{u}_r^*, \mathbf{u}_{r-1})}{q_U(\mathbf{u}_{r-1}, \mathbf{u}_r^*)}$

4:    $v \sim \text{Unif}[0,1]$

5:    **if** $v < \alpha$ **then**

6:        $\mathbf{u}_r \leftarrow \mathbf{u}_r^*$

7:    **else**

8:        $\mathbf{u}_r \leftarrow \mathbf{u}_{r-1}$

9:    **end if**

10: **end for**

Example 5.2. (Nonadaptive Metropolis Hastings algorithm with independence proposal) The easiest way to implement the MH algorithm, explained in Chib and Greenberg, 1995, is to use a proposal density independent of $\mathbf{u}_{r-1}$, called the independence proposal. This is illustrated in **Figure 5**:
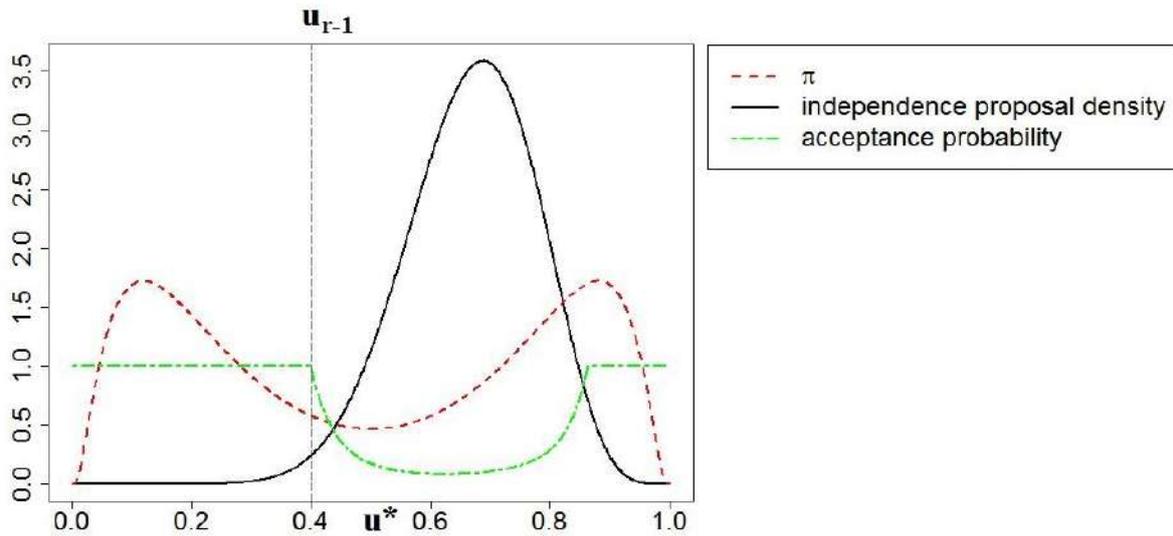
Figure 5: Visualization of an update proposal in one dimension, with independence proposal

Figure 5 illustrates the acceptance rule of **Algorithm 2** with a proposal density $q_U(\mathbf{u}_{r-1}, \mathbf{u}_r^*)$, independent of $\mathbf{u}_{r-1}$. Hence, the value of α becomes $\dfrac{\pi_U(\mathbf{u}_r^*)}{\pi_U(\mathbf{u}_{r-1})} \dfrac{q_U(\mathbf{u}_{r-1})}{q_U(\mathbf{u}_r^*)}$.

For the next iteration, the algorithm works the following way:

A proposal u_r^* from the conditional proposal density is drawn and accepted if $v < \alpha$, which happens with probability $\max \left(1, \dfrac{\pi_U(\mathbf{u}_r^*)}{\pi_U(\mathbf{u}_{r-1})} \dfrac{q_U(\mathbf{u}_{r-1})}{q_U(\mathbf{u}_r^*)} \right)$.

In this formula, we can see that there are two sources which influence the acceptance probability. The first part of the fraction $\dfrac{\pi_U(\mathbf{u}_r^*)}{\pi_U(\mathbf{u}_{r-1})}$ implies, that proposal values with a higher likelihood under the limiting distribution are more likely to be accepted. The second part $\dfrac{q_U(\mathbf{u}_{r-1})}{q_U(\mathbf{u}_r^*)}$ of the proposal implies, that proposal values with lower likelihood under the proposal distribution are more likely accepted.

In this example, the proposal distribution is badly chosen, as its density is not similar to the limiting distribution. This will be discussed in more detail in the following section.

5.2 Metropolis Hastings algorithm with symmetric random-walk increments

The classical approach for the proposal distribution of the MH algorithm, as it is introduced by Atchad et al., 2009, is the symmetric-increment random walk method (SRWM). This approach is applied to simulate a random variable in $\mathbb{R}^d$.

The proposal realization $\mathbf{X}_r^*$ is simulated by adding a random walk increment $\boldsymbol{\epsilon}_r \in \mathbb{R}^n$ with density f, to the current random variable $\mathbf{X}_r$.

$$\mathbf{X}_r^* = \mathbf{X}_r + \boldsymbol{\epsilon}_r \tag{5.1}$$

The distribution of the increments $\boldsymbol{\epsilon}_r$ is chosen to be symmetric around zero. This means that the density f evaluated at $\boldsymbol{\epsilon}$ and $-\boldsymbol{\epsilon}$ has the same value:

$$f(\boldsymbol{\epsilon}) = f(-\boldsymbol{\epsilon}) \tag{5.2}$$

With the use of the indicator function $\mathbb{I}$ we see that its mean is zero:

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{\epsilon}) &= \mathbb{E}(\epsilon_i)_{i=1,\dots,d} \\
&= \mathbb{E}(\epsilon_i \cdot \mathbb{I}_{\epsilon_i>0})_{i=1,\dots,d} + \mathbb{E}(\epsilon_i \cdot \mathbb{I}_{\epsilon_i<0})_{i=1,\dots,d} \\
&= \mathbb{E}(\epsilon_i \cdot \mathbb{I}_{\epsilon_i>0})_{i=1,\dots,d} - \mathbb{E}(\epsilon_i \cdot \mathbb{I}_{\epsilon_i>0})_{i=1,\dots,d} \\
&= 0
\end{aligned}
$$

Thus, the conditional mean of $\mathbf{X}_r^*$ given $\mathbf{X}_r$ is:

$$
\mathbb{E}(\mathbf{X}_r^*|\mathbf{X}_r) = \mathbf{X}_r
$$

Note that this implies that the transition density of the proposal kernel is a function of $\boldsymbol{\epsilon}_r = \mathbf{x}_r^* - \mathbf{x}_r$:

$$
q(\mathbf{x}_r, \mathbf{x}_r^*) = f(\mathbf{x}_r^* - \mathbf{x}_r) \tag{5.3}
$$

The advantage of this algorithm is that the acceptance probability has an easy evaluable form, as it does not depend on the specific density of the increments (see Atchad et al., 2009). With equations (5.2) and (5.3) the acceptance probability becomes:

$$
\begin{aligned}
\alpha &= \frac{\pi_x(\mathbf{x}_r^*)}{\pi_x(\mathbf{x}_r)} \frac{q(\mathbf{x}_r^*, \mathbf{x}_r)}{q(\mathbf{x}_r, \mathbf{x}_r^*)} \\
&= \frac{\pi_x(\mathbf{x}_r^*)}{\pi_x(\mathbf{x}_r)} \frac{f(\mathbf{x}_r^* - \mathbf{x}_r)}{f(-(\mathbf{x}_r - \mathbf{x}_r^*))} \\
&= \frac{\pi_x(\mathbf{x}_r^*)}{\pi_x(\mathbf{x}_r)}
\end{aligned}
$$

Essential for a fast convergence of the algorithm is the choice of distribution for the increments $\boldsymbol{\epsilon}_r$, as we will show in the following section.

If the standard deviance of the increments $\boldsymbol{\epsilon}_r$ is too small, the chain moves too slowly. If the standard deviance is too large, to many proposals are rejected. In Rosenthal, 2010, it is shown that both of the two cases must be avoided. As wrong choices lead to very slow convergence, and finding a good proposal distribution manually can take a lot of time, the class of adapted MH algorithms were developed. These algorithms are explained in the next chapter.

5.3 Convergence of the Metropolis Hastings algorithm under the assumption of lighter-than-exponential tails

One major question is the choice of the d-dimensional proposal density $q_U(\cdot, \cdot)$. Different choices of the proposal density can result in very different convergence speeds of the simulation.

The theoretical explanation for the rates of convergence of this approach are given in Holden, 1998.

Holden, 1998 derives an upper bound for the speed of convergence of the random variable, which is simulated by R runs of **Algorithm 2** towards its limiting distribution. This result requires specific conditions, which are explained in this chapter.

To study the speed of convergence, we introduce some new notation. We denote the density of $\mathbf{U}_r$ in **Algorithm 2** by p^r.

In the algorithm the transition kernel can change in each time step. We denote the d-dimensional transition density of the proposal kernel in step r by $q_U^r(\cdot, \cdot)$.

To measure the speed of convergence, we introduce a measure for the distance between the distribution of the simulated random variables and the desired distribution, which we call the supreme relative error.

Definition 5.1 (Supreme relative error).
Let $\pi_U(\mathbf{u}) > 0$ for $\mathbf{u} \in [0,1]^d$

$$R_M^r := \sup_{\mathbf{u} \in (0,1)^d} \left\{ \left| \frac{p^r(\mathbf{u})}{\pi_U(\mathbf{u})} - 1 \right| \right\}$$

Holden (1998) states, that if a constant c_r exists, such that in iteration r

$$q_U^r(\mathbf{u}, \mathbf{u}^*) \geq c_r \cdot \pi_U(\mathbf{u}^*) \qquad \forall \mathbf{u}, \ \mathbf{u}^* \in (0,1)^d \tag{5.4}$$

an upper bound for R_M^{r+1} is given by:

$$R_M^{r+1} \leq (1 - c_r) R_M^r \tag{5.5}$$

Property (5.4) is a strong condition, as it says that $q_U^r(\mathbf{u}, \cdot)/c_r$ is an envelope for $\pi_U(\cdot)$.

If we iterate inequality (5.5), we get the following upper bound for R_M^{r+1}:

$$R_M^{r+1} \leq \left[\prod_{i=1}^{r} (1 - c_i) \right] R_M^0 \tag{5.6}$$

Hence it shows that, if $\prod_{i=1}^{r} (1 - c_i)$ converges to 0 as $r \to \infty$ and R_M^0 is finite, then R_M^r converges to 0 as $r \to \infty$.

If c_i is bounded from below by a positive constant c, we get geometrical convergence:

$$R_M^{r+1} \leq \left[\prod_{i=1}^{r} (1 - c_i) \right] R_M^0 \leq (1 - c)^r R_M^0$$

So the order of convergence depends on the coefficients c_r, which depend on q_U^r. We want to maximize c_r as this speeds up the convergence of our algorithm. For this, we need to find densities q_U^r which are envelopes for π_U and lead to values of c_r that are close to 1.

A specific example, which can easily be evaluated in closed form, is obtained when the proposal distribution is chosen as the uniform distribution. The conditional density becomes $q_U^r(\mathbf{u}, \cdot) \equiv 1$ and $c_r = c$. Hence, to satisfy (5.4), $\pi_U(\mathbf{u}^*)$ has to be bounded by the constant $1/c$. This shows that all bounded densities can be simulated fast by using the uniform density as a proposal.

As a consequence, the result suggests the use of different transition densities in each step: As R_M^r is non increasing and bounded from above by R_M^0, it will converge to 0 as long as the coefficients c_r are bounded by a positive coefficient from below.

The convergence of this algorithms does not follow from the convergence of the usual version of the MH algorithm.
The convergence of this special case follows from Bai et al., 2011b.

The proof of the convergence can be shown under different conditions. One of these conditions is explained as follows:

Definition 5.2 (Lighter-than-exponentially tailed densities)**.**
The density $\pi_L(\cdot)$ on $\mathbb{R}^n$ is lighter-than-exponentially tailed if it is positive and has a continuous first derivative

$$\limsup_{|\mathbf{x}| \to \infty} \left\langle \frac{\mathbf{x}}{||\mathbf{x}||}, \nabla \ln(\pi_L(\mathbf{x})) \right\rangle = -\infty \tag{5.7}$$

We cite an important result from Bai et al., 2011b:

Theorem 5.1.
If the density $\pi_L(\cdot)$ on $\mathbb{R}^n$ is lighter-than-exponentially tailed, the MH algorithm converges towards its stationary distribution.

The assumption for $\pi_L(\cdot)$ to be lighter-than-exponentially tailed is a strong assumption, which is difficult to prove theoretically.
Thus, we will have to use methods as explained in **Section 4.3** indicating whether our algorithm has converged.

5.4   Adaptive Metropolis Hastings algorithm for the L-Level

Influenced by equation (5.6) in **Chapter 5.3**, we want to introduce an algorithm working on the **L-Level**, which changes its transition probability in each step in order to speed up the convergence of the algorithm.

The class of adapted MH algorithms modifies its proposal density q^r according to the first $r - 1$ elements of the Markov chain. Hence, the algorithm does not depend upon the right guess for the proposal distribution, but chooses an appropriate distribution on its own. Their major advantage is the good convergence properties and their flexibility.

In general, we define adaptive Markov chains as in Rosenthal, 2010:

We have a set of time homogeneous transition kernels $\{P_\gamma : \gamma \in \mathbb{R}^{d \times d}\}$, each having the same distribution $\pi(\cdot)$ as its stationary distribution.
At time step 0 we begin with starting values $\mathbf{x}_0, \boldsymbol{\gamma}_0$, then $\mathbf{X}_1$ is simulated from $P_{\boldsymbol{\gamma}_0}(\mathbf{x}_0, \cdot)$ and $\boldsymbol{\Gamma}_1$ is calculated from $\mathbf{X}_1$ and $\boldsymbol{\gamma}_0$.

Then at each step $r \geq 1$, we perform the following two steps:

1. Simulate $\mathbf{X}_{r+1}$ from $P_{\boldsymbol{\Gamma}_r}(\mathbf{X}_r, \cdot)$

2. Calculate $\boldsymbol{\Gamma}_{r+1}$ deterministically from $\boldsymbol{\Gamma}_r$ and $\mathbf{X}_{r+1}$, and thus change the transition probability.

With the filtration $(\mathcal{G}_r)_{r \geq 1} = (\sigma(\mathbf{X}_1, \ldots, \mathbf{X}_r, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_r))_{r \geq 1}$, which is generated by the process $\mathbf{X}_r$ and the process $\boldsymbol{\Gamma}_r$, we can now define adaptive Markov chains in general:

Definition 5.3 (adaptive Markov chain)**.**
An adaptive Markov chain $\mathbf{Z} = \{(\mathbf{X}_r, \boldsymbol{\Gamma}_r) : r \geq 1\}$ *with its natural filtration*
$(\mathcal{G}_r)_{r \geq 1} = (\sigma(\mathbf{X}_1, \ldots, \mathbf{X}_r, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_r))_{r \geq 1}$ *satisfies the following equation:*

$$\mathbb{P}_{\mathbf{x}_1, \gamma_1}(\mathbf{X}_{r+1} \in A | \mathcal{G}_r) = \mathbb{P}_{\mathbf{x}_1, \gamma_1}(\mathbf{X}_{r+1} \in A | \mathbf{X}_r, \boldsymbol{\Gamma}_r) = P_{\boldsymbol{\Gamma}_r}(\mathbf{X}_r, A), \tag{5.8}$$

Where $\boldsymbol{\Gamma}_r$ *is a* $\sigma(\mathbf{X}_{r-1}, \boldsymbol{\Gamma}_{r-1})$ *measurable random variable.*

The name "adapted" comes from the property, that the transition kernel adapts to the past of the chain. In our case, this can be seen as a learning procedure: The Markov chain learns from its past by changing the transition kernel.

We want to derive some further properties of adapted Markov chains.
A special case of adapted Markov chains are the ones which are independent of $\boldsymbol{\Gamma}_r$:

$$\mathbb{P}_{\mathbf{x}_1, \gamma_1}(\mathbf{X}_{r+1} \in A | \mathbf{X}_r, \boldsymbol{\Gamma}_r) = P(\mathbf{X}_r, A)$$

A further property of adapted Markov chains is that random mixes of adapted Markov chains stay adapted Markov chains.

Definition 5.4 (Random mixture of two adapted Markov chains)**.**
Let $\{P_\gamma : \gamma \in \mathcal{Y}\}$ *and* $\{P_\zeta^* : \zeta \in \mathcal{Z}\}$ *be two sets of time homogeneous transition kernels with the same stationary distribution* $\pi(\cdot)$. *We define a random mixture with mixture parameter* $\beta \in [0, 1]$ *and filtration* $(\mathcal{G}_r^{mix})_{r \geq 1} = (\sigma(\mathbf{X}_1, \ldots, \mathbf{X}_r, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_r, \mathbf{Z}_1, \ldots, \mathbf{Z}_r))_{r \geq 1}$ *as Markov chain satisfying the following equation:*

$$\begin{aligned}
\mathbb{P}_{\mathbf{x}_1, \gamma_1, \zeta_1}(\mathbf{X}_{r+1} \in A | \mathcal{G}_r^{mix}) =& \beta \cdot \mathbb{P}_{\mathbf{x}_1, \gamma_1}(\mathbf{X}_{r+1} \in A | \mathbf{X}_r, \boldsymbol{\Gamma}_r) + (1 - \beta) \cdot \mathbb{P}_{\mathbf{x}_1, \zeta_1}(\mathbf{X}_{r+1} \in A | \mathbf{X}_r, \mathbf{Z}_r) \\
=& \beta \cdot P_{\boldsymbol{\Gamma}_r}(\mathbf{X}_r, A) + (1 - \beta) \cdot P_{\mathbf{Z}_r}^*(\mathbf{X}_r, A)
\end{aligned}$$

The following theorem will be needed in the next chapter.

Theorem 5.2. *A random mixture of two adapted Markov chains is again an adapted Markov chain.*

Proof. Set $\mathcal{Y}' = \mathcal{Y} \times \mathcal{Z}$, $\mathbf{\Gamma}'_r = (\mathbf{\Gamma}_r, \mathbf{Z}_r)$ and $\gamma'_1 = (\gamma_1, \zeta_1)$.
We can now see that the new Markov chain is again adapted:

$$\mathbb{P}_{\mathbf{x}_1, \gamma'_1}(\mathbf{X}_{r+1} \in A | \mathcal{G}^{mix}_r) = \beta \cdot \mathbb{P}_{\mathbf{x}_1, \gamma_1}(\mathbf{X}_{r+1} \in A | \mathcal{G}^{mix}_r) + (1 - \beta) \cdot \mathbb{P}_{\mathbf{x}_1, \zeta_1}(\mathbf{X}_{r+1} \in A | \mathcal{G}^{mix}_r)$$

$\square$

5.5 Adaptive Metropolis Hastings algorithm with Gaussian increments

After we studied adapted Markov chains in general, we now focus on adapted MH algorithms, which generate adapted Markov chains.
A special case of adapted Markov chains is the adapted random walk with Gaussian increments.
We work on the **L − Level** with the random variables $(\boldsymbol{\ell}_r)_{r \geq 0}$, because this algorithm requires values in $\mathbb{R}^d$.
The $r + 1$th proposal has the same form as in equation (5.1):

$$\boldsymbol{\ell}^*_{r+1} = \boldsymbol{\ell}_r + \boldsymbol{\epsilon}_r$$

Where $\boldsymbol{\epsilon}_r \sim \mathcal{N}_d(\mathbf{0}, \Sigma_r)$. The transition proposal density fulfils property (5.2), which is required for the algorithm.

The recursive definition of the covariance matrix Σ_r and the vector $\boldsymbol{\mu}_r$, which is only needed for the normalization for the proposal density, and is proposed in Haario et al., 2001, are given by:

$$\boldsymbol{\mu}_{r+1} = \boldsymbol{\mu}_r + \frac{1}{r+1}(\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r) \tag{5.9}$$

$$\Sigma_{r+1} = \Sigma_r + \frac{1}{r+1}\left((\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r) \cdot (\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)^t - \Sigma_r\right) \tag{5.10}$$

with the initial values $\boldsymbol{\mu}_1 = (0, \ldots, 0)$, and $\Sigma_1 = \operatorname{diag}(1)$.

In this case $\Gamma_r = (\Gamma_r^*, \Gamma_r')$ with $\Gamma_r^* \in \mathbb{R}^d$, $\Gamma_r' \in \mathbb{R}^{d \times d}$ is given by:

$$\Gamma_r = (\Gamma_r^*, \Gamma_r') = (\boldsymbol{\mu}_r, \Sigma_r)$$

The recursive formula for Γ_{r+1} is in our case given by:

$$\begin{aligned} \Gamma_{r+1} &= (\Gamma_{r+1}^*, \Gamma_{r+1}') \\ &= (\Gamma_r^* + \frac{1}{r+1}(\boldsymbol{\ell}_{r+1} - \Gamma_r^*), \Gamma_r' + \frac{1}{r+1}\left((\boldsymbol{\ell}_{r+1} - \Gamma_r^*) \cdot (\boldsymbol{\ell}_{r+1} - \Gamma_r^*)^t - \Gamma_r'\right)) \end{aligned}$$

Algorithm 3 Joint adaptive Metropolis Hastings algorithm with Gaussian increments

---

Simulate $\mathbf{u}_{1 \le r \le R}$, $\mathbf{u}_r \in [0,1]^d$

with $\pi_U(\cdot)$ as the approximate density of $\mathbf{u}_r$

Require: Starting value $\boldsymbol{\ell}_1$, initial values $\boldsymbol{\mu}_1$, Σ_1, number of runs R

1: **for** $r \leftarrow 1$ **to** R **do**
2:    $\boldsymbol{\epsilon}_r \sim \mathcal{N}_d(\mathbf{0}_r, \Sigma_r)$
3:    $\boldsymbol{\ell}_{r+1}^* = \boldsymbol{\ell}_r + \boldsymbol{\epsilon}_r$
4:    $\alpha = \dfrac{\pi_L(\boldsymbol{\ell}_{r+1}^*)}{\pi_L(\boldsymbol{\ell}_r)}$
5:    $v \sim \text{Unif}[0,1]$
6:    **if** $v < \alpha$ **then**
7:      $\boldsymbol{\ell}_{r+1} \leftarrow \boldsymbol{\ell}_{r+1}^*$
8:    **else**
9:      $\boldsymbol{\ell}_{r+1} \leftarrow \boldsymbol{\ell}_r$
10:   **end if**
11:   $\boldsymbol{\mu}_{r+1} = \boldsymbol{\mu}_r + \dfrac{1}{r+1}(\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)$
12:   $\Sigma_{r+1} = \Sigma_r + \dfrac{1}{r+1}((\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r) \cdot (\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)^t - \Sigma_r)$
13: **end for**

Our new Markov chain is the process $\tilde{\boldsymbol{\ell}}_r = (\boldsymbol{\ell}_r, \Sigma_r, \boldsymbol{\mu}_r)$, as $\boldsymbol{\ell}_r$ is no longer a Markov chain. The reason is that the realizations of $(\boldsymbol{\ell}_k)_{1 \le k \le (r-1)}$ influence the value of Σ_r and $\boldsymbol{\mu}_r$ and thus the transition probabilities, which makes the process no longer independent of the past.

If we rewrite equations (5.9) and (5.10), we can see that **Algorithm 3** is a special version

of the **Covariance Matrix Adaption Evolution Strategy**. This method was originally developed for the stochastic maximization of real valued functions as described in Mueller and Sbalzarini, 2010, and in Hoshimura, 2007. Both methods are similar, as they search for the area with the highest values. The difference is that optimization only requires a single point, the global maximum, while the MH algorithm needs the whole area with high likelihood, as most random variables have to be drawn from this area.

The more general form of equations (5.9) and (5.10) is a linear combination of the current value and the preceding value:

$$\boldsymbol{\mu}_{r+1} = (1 - \gamma_r)\,\boldsymbol{\mu}_r + \gamma_r \boldsymbol{\ell}_{r+1}$$

$$\Sigma_{r+1} = (1 - \gamma_r)\,\Sigma_r + \gamma_r(\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)(\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)^t$$

$$\text{With } \gamma_r := \frac{1}{r+1}$$

In this context, other choices of γ_r are also possible. The choice of γ_r weights the influence of the new values on the distribution. Other choices can result in other convergence properties of the algorithm. If we choose $\gamma_r = \dfrac{1}{r^2+1}$, the values of $\boldsymbol{\mu}_{r+1}$ and Σ_{r+1} are much less influenced by the current value $\boldsymbol{\ell}_r$. If we choose $\gamma_r = \dfrac{1}{\sqrt{r+1}}$, the influence of the value of $\boldsymbol{\ell}_r$ on $\boldsymbol{\mu}_{r+1}$ and Σ_{r+1} is much stronger (see Mueller and Sbalzarini, 2010).

One other possibility to reduce the computation time would be to perform the update only every k'th step and to apply the non adaptive MH algorithm from this step on.

One shortfall of the adaptive Gaussian increments is that if the initial distribution is far away from the target density, the convergence of the adapted MH is slow. This can be caused by a starting value far away from the stationary distribution, which strongly influences the convergence of Σ_r and $\boldsymbol{\mu}_r$.

In **Figure 6** we see a trajectory with a starting value which is simulated from a random distribution, which is different from the stationary distribution. This starting value has a low likelihood in the stationary distribution. Additionally, we see a trajectory with a starting value simulated from the stationary distribution.

All increments of the random walk, starting in its stationary distribution, are therefore realizations of the stationary distribution.

However, the random walk not starting in its stationary distribution moves fast towards the area with a high likelihood under its stationary distribution. This, however, slows down the convergence of Σ_r and $\boldsymbol{\mu}_r$, as the direction and shape of the first random walk

increments are different from the increments in the stationary case. Thus, the first steps of the Markov chain $\tilde{\mathbf{L}}_r$ slow down the convergence of $\mathbf{L}_r$.
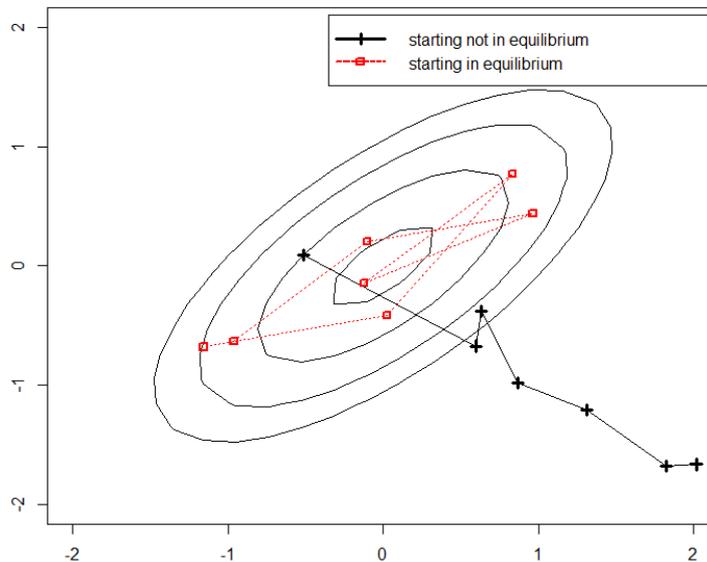

Figure 6: Trajectories generated by the Metropolis Hastings algorithm starting in equilibrium and not in equilibrium.

Thus, to avoid the effect of badly chosen starting values, we restart the adapted MH algorithm after a certain number of steps (e.g. 5000) at its current value by setting $r = 1$. This improves the convergence of $\boldsymbol{\mu}_r$ and Σ_r, because the first simulations can be far away from the stationary distribution and thus slow down the convergence.

As in most simulations high burn-in times are required, this procedure has the further advantage that it can reduce the burn-in.

5.6   Metropolis Hastings with univariate Beta distributions

One big theoretical problem which arises from the use of Gaussian increments is that the proposal realizations are not heavy tailed. If the target density has more than one local maximum, the Gaussian Markov chain will likely concentrate around one local maximum and rarely move to another local maximum. A second shortfall of the use of adapted Gaussian increments is the slow convergence at the boundaries of $[0, 1]^d$. Therefore, $\mathbf{L}_k$ has to converge to ∞, which takes a lot of time and makes later jumps to other regions extremely unlikely.

To overcome these problems and to speed up the convergence, we propose a second proposal rule which works on the **U-Level**.

Our approach is to define a one dimensional transition probability, which leads to a faster convergence of the Markov chain. For this we choose the beta distribution defined on [0,1].

Definition 5.5 (Beta distribution)**.**
The beta distribution is defined for $0 < y < 1$, $a > 0, b > 0$ through the following density

$$f(y|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \tag{5.11}$$

We want to investigate the convergence behavior of the density at the limit of its definition range. Depending on its parameters, the density has the following limiting values:

$$\lim_{y \to 0} f(y|a,b) = \begin{cases} b & \text{if } a = 1 \\ \infty & \text{if } a < 1 \\ 0 & \text{if } a > 1 \end{cases}$$

$$\lim_{y \to 1} f(y|a,b) = \begin{cases} a & \text{if } b = 1 \\ \infty & \text{if } b < 1 \\ 0 & \text{if } b > 1 \end{cases}$$

If we neglect the cases in which either b or a is equal to 1, we have 4 different cases which are illustrated in Figure 7:
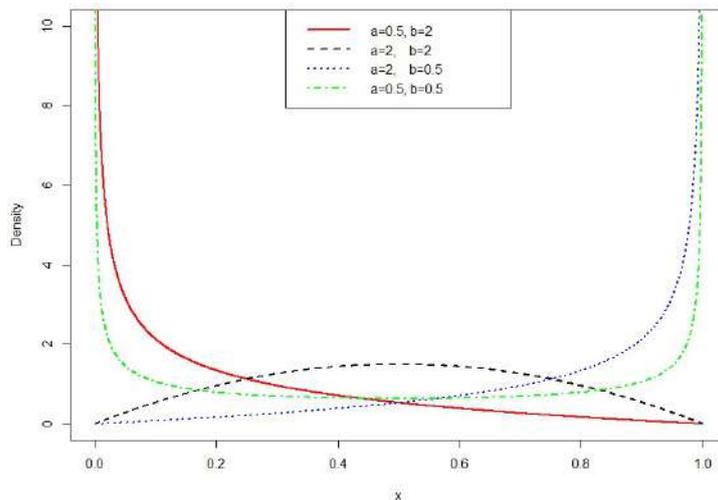

Figure 7: Densities of beta distributions with different parameters

Since we want the one dimensional distribution to be centered around the current value

u_i, this leads to equations (5.12) and (5.13).

$$\mathbb{E}[Y] = u_i = \frac{a}{a+b} \tag{5.12}$$

$$\Leftrightarrow b = a\frac{1-u_i}{u_i} \tag{5.13}$$

We can clearly exclude the case when $a < 1$, $b < 1$, as in this case u has a high probability of either being close to 0 or close to 1. This is rarely observed in our case and thus the convergence would possibly be slowed down.

In the case that our fixed values are quite high (e.g. 0.99) and an upper tail dependence or lower tail dependence is present, it is reasonable to choose b or a smaller than 1, respectively.

Our proposal is to choose a random distribution for a. The value of b is then defined by equation (5.12).

The **first case** which we want to look at is where the density $f(y)$ is 0 for $y \in \{0; 1\}$, which results in the following inequalities:

$$a > 1 \text{ and } b = a \cdot \frac{1-u_i}{u_i} > 1$$

$$\Rightarrow a > \frac{u_i}{1-u_i}$$

$$\Rightarrow a > \max(1; \frac{u_i}{1-u_i})$$

The **second case** is where the density f converges to 0 at one limit, and to ∞ at the other. In the case, where $\lim_{y \to 1} f(y) = \infty$ and $\lim_{y \to 0} f(y) = 0$, we have to choose the constants in the following way:

$$a > 1 \text{ and } b = a \cdot \frac{1-u_i}{u_i} < 1$$

$$\Rightarrow a < \frac{u_i}{1-u_i}$$

$$\Rightarrow \frac{u_i}{1-u_i} > a > 1$$

This is only applicable if:

$$\frac{u_i}{1-u_i} > 1$$

$$\Leftrightarrow u_i > 0.5$$

This satisfies $\lim_{y \to 1} f(y) = \infty$ for the proposal density, which is a natural assumption for large values of u_i.

In the case where $\lim_{y \to 1} f(y) = 0$ and $\lim_{y \to 0} f(y) = \infty$, we have to choose the constants in the following way:

$$a < 1 \text{ and } b = a \cdot \frac{1 - u_i}{u_i} > 1$$

$$\Rightarrow a > \frac{u_i}{1 - u_i}$$

$$\Rightarrow 1 > a > \frac{u}{1 - u_i}$$

This case only applies if:

$$1 > \frac{u_i}{1 - u_i}$$

$$\Leftrightarrow u_i < 0.5$$

This again satisfies our assumption where $\lim_{y \to 0} f(y) = \infty$, as u_i has a low value.

As we want both cases to happen often and at the same time, the simulation should be fast. We propose the following distribution for a:

We choose a large constant $c \in \mathbb{R}_+$ to ensure, that the upper limit l_{upper} is bounded. Then we define:

$$l_{upper} = \max\left(1, \min\left(\frac{u_i}{1 - u_i}, c\right)\right)$$

$$l_{lower} = \min\left(1, \frac{u_i}{1 - u_i}\right)$$

$$a \sim \mathrm{Unif}\,[l_{lower}, l_{upper}]$$

To get the proposal density, we have to integrate over the set of possible values of a.

$$\kappa(u_i, u_i^*) = \frac{1}{l_{upper} - l_{lower}} \int_{l_{lower}}^{l_{upper}} \frac{\Gamma(a \cdot (1 + \frac{1 - u_i}{u_i}))}{\Gamma(a)\Gamma(a \cdot \frac{1 - u_i}{u_i})} u_i^{*a-1}(1 - u_i^*)^{a \cdot \frac{1 - u_i}{u_i} - 1} \mathrm{d}a$$

$$\kappa(u_i^*, u_i) = \frac{1}{l_{upper} - l_{lower}} \int\limits_{l_{lower}}^{l_{upper}} \frac{\Gamma(a \cdot (1 + \frac{1-u_i^*}{u_i^*}))}{\Gamma(a)\Gamma(a \cdot \frac{1-u_i^*}{u_i^*})} u_i^{a-1} (1 - u_i)^{a \cdot \frac{1-u_i^*}{u_i^*} - 1} da$$

These integrals can only be evaluated numerically, which is one of the major reasons for the long computation time needed in each update step.

The acceptance probability thus becomes:

$$\alpha = \frac{\pi_U(u^*)}{\pi_U(u)} \prod_{i \in \mathbf{D}} \frac{I(u_i^*, u_i)}{I(u_i, u_i^*)}$$

5.7   Mixture of Metropolis Hastings algorithms

In this chapter we have shown that both, the adapted Gaussian approach on the **L-Level** and the beta approach on the **U-Level**, have strengths and weaknesses. Hence, we propose a mixture off both algorithms to overcome their weaknesses.

The major advantage of **Algorithm 3** is its high convergence speed when the area with high likelihood is reached. The reason for this is that due to its adeptness, the algorithm uses the past values to "learn" the shape of the stationary density.

However, it works on the **L-Level** and Gaussian increments are not heavy tailed. If it starts far away from the stationary distribution, it can take some time to reach the area with high likelihood. It also converges badly for distributions with multi-modal densities.

Algorithm 4 overcomes these problems, as it works on the **U-Level** and allows for big jumps. Distributions with bimodal densities, for example, are not a big problem for this algorithm.

However, the algorithm does not change its transition probabilities according to the past. This fact together with the large number of big jumps makes the acceptance rate rather low.

The property of the algorithm not to change its transition probabilities according to the past for big jumps is a good property, because the aim of big jumps is to move to areas which have not been reached yet. As these areas have not been reached, an algorithm which changes its transition probabilities according to the past would not be appropriate.

To overcome these problems, we propose a mixture of both algorithms as follows:

We choose a value of $\beta \in (0, 1)$, where β is the probability with which an update with univariate Beta distributions, like in **Section 5.6**, is applied. With probability $1 - \beta$ an update with adaptive Gaussian increments, like in **Section 5.5**, is applied.

To illustrate a case where our new algorithm has better convergence, we conducted a simulation in **Example 5.3**.

Notice that in the algorithm, the **MH algorithm with adapted Gaussian increments** works on the **L-Level** and the **MH algorithm with univariate Beta distributions** works on the **U-Level**.

Algorithm 4 Metropolis Hastings algorithm with mixture of adapted Gaussian increments and univariate Beta distributions

---

Simulate $\mathbf{u}_{1\leq r\leq R}$, $\mathbf{u}_r \in [0,1]^d$
with $\pi_U(\cdot)$ as the approximate density of $\mathbf{u}_r$

Require: Starting value $\boldsymbol{\ell}_1$, initial values $\boldsymbol{\mu}_1$, Σ_1, probability of beta updates β, number of runs R, big constant c $\in \mathbb{R}_+$

1: **for** $r \leftarrow 1$ **to** R **do**

2: $\quad w \sim \text{Unif}[0,1]$

3: $\quad$ **if** $w < \beta$ **then**

4: $\quad\quad \mathbf{u} = logit^{-1}(\mathbf{l_r})$

5: $\quad\quad \alpha = 1$

6: $\quad\quad$ **for** $i \leftarrow 1$ **to** d **do**

7: $\quad\quad\quad l_{upper} = \max\left(1, \min\left(\frac{u_i}{1-u_i}, c\right)\right)$

8: $\quad\quad\quad l_{lower} = \min\left(1, \frac{u_i}{1-u_i}\right)$

9: $\quad\quad\quad a \sim \text{Unif}[l_{lower}, l_{upper}]$

10: $\quad\quad\quad b = a \cdot \frac{1-u_i}{u_i}$

11: $\quad\quad\quad u_i \sim beta(a,b)$

12: $\quad\quad\quad \kappa(u_i, u_i^*) = \frac{1}{l_{upper}-l_{lower}} \cdot \int\limits_{l_{lower}}^{l_{upper}} \frac{\Gamma(a\cdot(1+\frac{1-u_i}{u_i}))}{\Gamma(a)\Gamma(a\cdot\frac{1-u_i}{u_i})}u_i^{*a-1}\left(1-u_i^*\right)^{a\cdot\frac{1-u_i}{u_i}-1}\mathrm{d}a$

13: $\quad\quad\quad \kappa(u_i^*, u_i) = \frac{1}{l_{upper}-l_{lower}} \cdot \int\limits_{l_{lower}}^{l_{upper}} \frac{\Gamma(a\cdot(1+\frac{1-u_i^*}{u_i^*}))}{\Gamma(a)\Gamma(a\cdot\frac{1-u_i^*}{u_i^*})}u_i^{a-1}\left(1-u_i\right)^{a\cdot\frac{1-u_i^*}{u_i^*}-1}\mathrm{d}a$

14: $\quad\quad\quad \alpha = \alpha * \frac{I(u_i^*,u_i)}{I(u_i,u_i^*)}$

15: $\quad\quad$ **end for**

16: $\quad\quad v \sim \text{Unif}[0,1]$

17: $\quad\quad$ **if** $v < \alpha$ **then**

18: $\quad\quad\quad \boldsymbol{\ell}_{r+1} \leftarrow logit(\mathbf{u}^*)$

19: $\quad\quad$ **else**

20: $\quad\quad\quad \boldsymbol{\ell}_{r+1} \leftarrow \boldsymbol{\ell}_r$

21: $\quad\quad$ **end if**

22: $\quad$ **else**

23: $\quad\quad \boldsymbol{\epsilon}_r \sim \mathcal{N}_d(\mathbf{0}, \Sigma_r)$

24: $\quad\quad \boldsymbol{\ell}_{r+1}^* = \boldsymbol{\ell}_r + \boldsymbol{\epsilon}_r$

25: $\quad\quad \alpha = \dfrac{\pi_L(logit(\boldsymbol{\ell}_{r+1}^*))}{\pi_L(logit(\boldsymbol{\ell}_r))}$

26: $\quad\quad v \sim \text{Unif}[0,1]$

27: $\quad\quad$ **if** $v < \alpha$ **then**

28: $\quad\quad\quad \boldsymbol{\ell}_{r+1} \leftarrow \boldsymbol{\ell}_{r+1}^*$

29: $\quad\quad$ **else**

30: $\quad\quad\quad \boldsymbol{\ell}_{r+1} \leftarrow \boldsymbol{\ell}_r$

31: $\quad\quad$ **end if**

32: $\quad\quad \boldsymbol{\mu}_{r+1} = \boldsymbol{\mu}_r + \dfrac{1}{r+1}(\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)$

33: $\quad\quad \Sigma_{r+1} = \Sigma_r + \dfrac{1}{r+1}((\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r) \cdot (\boldsymbol{\ell}_{r+1} - \boldsymbol{\mu}_r)^t - \Sigma_r)$

34: $\quad$ **end if**

35: **end for**

A theoretical question which we still have to solve is whether **Algorithm 4** is a valid adaptive MH algorithm. This is proven in **Theorem 5.2** for which we have to ensure that both of our algorithms are adapted. The **Gaussian increments Algorithm** is adapted and the **beta increments Algorithm** is adapted, as our family of transition kernels is chosen to be constant $P_\gamma \equiv P$.

To show that the new mixed **Algorithm 4** has better convergence properties for bimodal distributions, we compare **Algorithm 4**, with $\beta = 0.5$, with **Algorithm 3** in the following example.

Example 5.3.

As an example where **Algorithm 3** converges very slowly, we introduce a conditional R-vine copula density, which has a bimodal density.

We used a C-vine, where we conditioned on 3 variables, and simulated the 4th variable. Analog to the similarity-based approach in **Chapter 9**, we apply a mixture of two different pair-copula families. For both pair-copula families we choose Student copulas, however with different values of Kendall's τ. The density has its maximum at 0.1, but it has a second local maximum at 0.9.

The Markov chain simulation is generated by **Algorithm 3**, which moves fast towards the area of the maximum at 0.1. In this case, however, the Markov chain does not move to the area at 0.9 where the density has a local maximum, as the Gaussian increments are not heavy tailed enough to cross the area with low likelihood between these two areas. We will discuss this effect in more detail:

Moving from the area around 0.1 towards the are of 0.9 is theoretically very unlikely, as the area between these two points has very low likelihood. The reason is that moving from a point with high likelihood to a point with much lower likelihood is an unlikely event due to the construction of the chain via the MH algorithm (Compare **Figure 5**). As Gaussian increments are not heavy tailed, several steps through the area with low likelihood have to be made in order to reach the local maximum.

However, **Algorithm 4** allows for big jumps. Thus, it is not surprising that in this case it leads to much faster convergence.
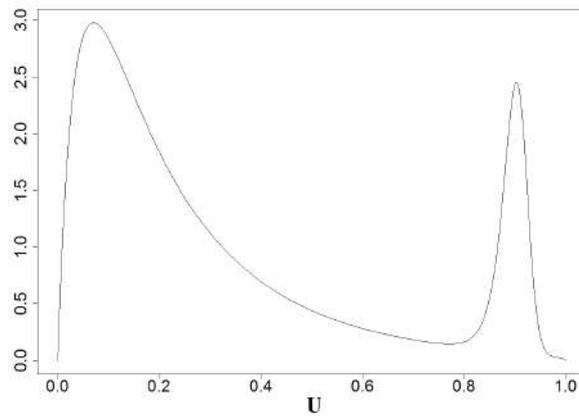
The distribution we used for the example, is a **linear mixture of copulas**, which will appear in **Chapter 9**. The copula we use is a mixture of two copulas with the following **C-vine** structure:
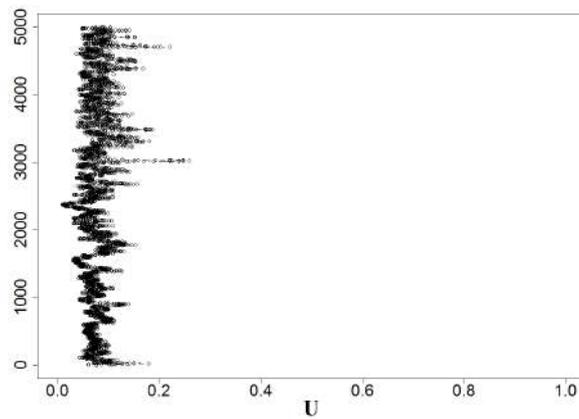
$$C(u) = 0.5 \cdot C^*(u) + 0.5 \cdot C'(u)$$

The density thus becomes a mixture of π^* and π':

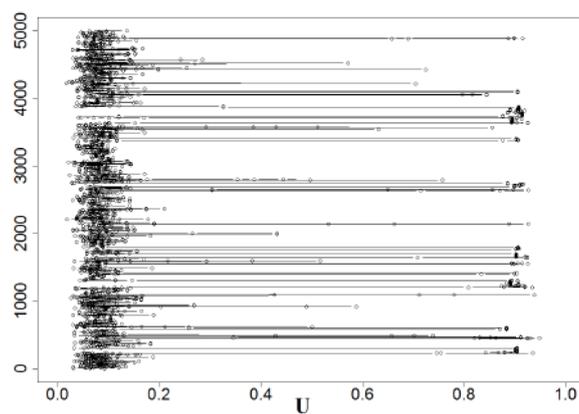$$\pi_U(u) = 0.5 \cdot \pi^*(u) + 0.5 \cdot \pi'(u)$$

We simulate the conditional density $\pi_U(u_4|u_1, u_2, u_3)$, where we set $(u_1, u_2, u_3) = (0.1, 0.1, 0.1)$.

(a) The stationary density $\pi_U(u_4|u_1, u_2, u_3)$, where we set $(u_1, u_2, u_3) = (0.1, 0.1, 0.1)$

.

(b) The trajectory generated by **Algorithm 3**

(c) The trajectory generated by **Algorithm 4** with $\beta = 0.5$

Figure 8: Summary of plots validating better convergence of **Algorithm 4** in example 5.3

5.8   Further possible methods to improve speed of convergence

Given that the MH algorithms with Gaussian increments for distributions with a bimodal density may not converge (as seen in Example 5.3), there are other methods proposed to solve this problem.

One possible solution can be to adopt a **regional adaptive algorithm with online recursion (RAPTOR)**, as described in Bai et al., 2011a. This method, however, is not a straight forward method, as it requires knowledge of the regions where the different modes are located. Afterwards, the set $[0,1]^d$ has to be divided into disjoint subsets manually with only one mode in each subset. Since we want our algorithm to run fully automatically for different situations, this represents no sensible method in our case.

Another approach to overcome this problems is to use mixtures of adaptive MH algorithms with Gaussian increments (see Luengo and Martino, 2013). This procedure has been proven to converge in simulations of multi modal densities. This, however, does not overcome the problem of Gaussian increments not having big jumps.

6 Simulation study with simulated data

In this chapter we will carry out simulation studies in two settings in order to validate the convergence of the MH algorithm specified in **Algorithm 4** for the purposes of R-vines.

At first we will perform simulation studies in a low number of dimensions, where we simulate from a 4-dimensional C-vine and condition on 2 random variables. One advantage in this case is that the random variable is 2-dimensional and thus we can graphically validate its distribution.
For this low-dimensional setting we will perform 8 different scenarios, which are explained in more detail in **Section 6.2**. We will also compare the distribution of the simulated random variables with the theoretical distribution in this part of the chapter. Additionally we will analyze the pair-copula of the conditional random variable.

To validate the convergence in higher dimension, we will perform a high-dimensional example in the second part of the chapter. We will use a 10 dimensional C-vine, where we condition on 5 random variables. A graphical validation of the resulting 5-dimensional distribution is not possible, therefore we evaluate the marginal distributions.
We will perform 4 different scenarios in this high-dimensional setting, which are described in **Section 6.4**.

Before we get to those topics described above, we will first explain the methods, which we will further use in an introductory example in **Section 6.1**. This example will exemplify, how we evaluate the theoretical distributions later on and which plots we use to validate the convergence.

6.1 Introductory example

We simulate the conditional value of a 4 dimensional R-vine. All 4 dimensional R-vines are either equivalent to a C-vine or a D-vine (see Kurrowicka and Joe, 2011 p. 142). In our case, we choose the C-vine (see **Definition 3.12**).

We conducted several scenarios (see **Table 5**), of which we want to explain the first scenario in more detail to illustrate the methods applied to validate the convergence.

For the low-dimensional simulations we consider the following families of pair-copulas: Gaussian, Student, Gumbel (and rotations). In this introductory example we only considered the first two of these. For the high-dimensional simulations we will additionally

consider Frank and Joe pair-copulas.

$C_{1,2}$	Student
	$df = 5$
	$\tau = -0.8$
$C_{1,3}$	Student
	$df = 8$
	$\tau = 0.4$
$C_{1,4}$	Student
	$\tau = 0.3$
$C_{2,4|1}$	Gaussian
	$\tau = 0.5$
$C_{2,3|1}$	Gaussian
	$\tau = 0.5$
$C_{3,4|1,2}$	Gaussian
	$\tau = 0.7$

Table 3: Summary of pair-copulas used in the introductory example

In this example, we condition on the random variables u_1 and u_2, which we set to $(u_1, u_2) = (0.95, 0.95)$.

We simulated 10000 runs with a burn-in of 2000 simulations. The burn-in test by Heidelberger and Welch with confidence level of 0.99 indicated no need for a further burn-in.

Figure 9 below shows that the random variables are moderately autocorrelated, whereas the autocorrelation function looks similar to an AR(1) process, which encourages the use of $\widehat{ESS}$ (compare **Section 4.3**).
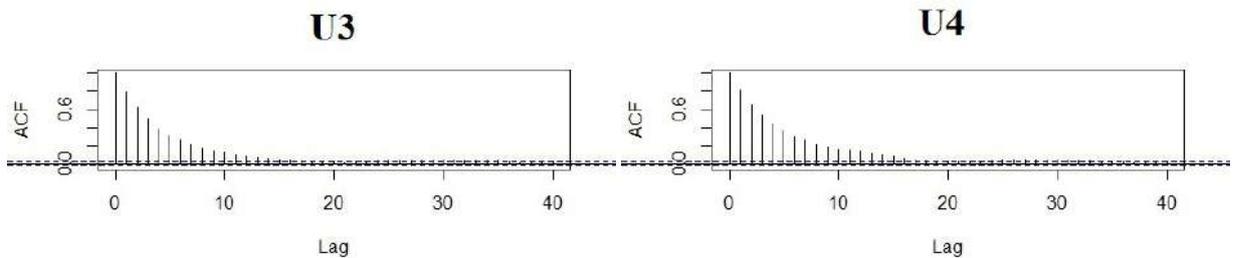

Figure 9: Sample autocorrelation of the Markov chain in the example

The effective sample size $\widehat{ESS}$ of the margins of the Markov chain are 1015 and 1132. These values are regarded as large enough to approximate the theoretical distribution

adequately.

As our random variable is 2 dimensional, it is possible to plot its whole theoretic distribution, and compare it with the realizations of the MH algorithm.
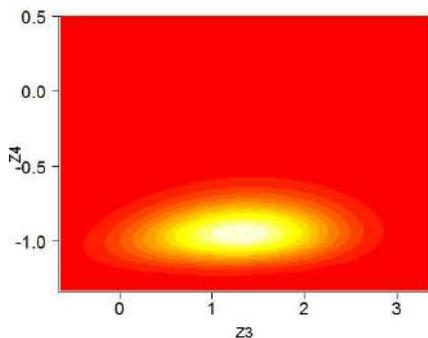We evaluated the density on an equidistant grid numerically, to draw the heat-map figures. These theoretical values are compared to the empirical scatter plot of the simulated values.
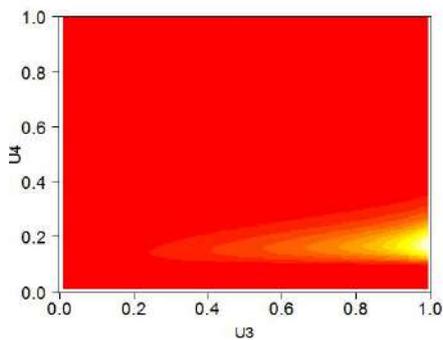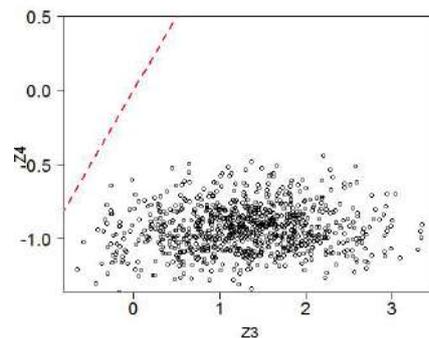

(a1) Heat map plot of (u_3, u_4) (U-Level)   (a2) Simulated pair-plot of (u_3, u_4) (U-Level)
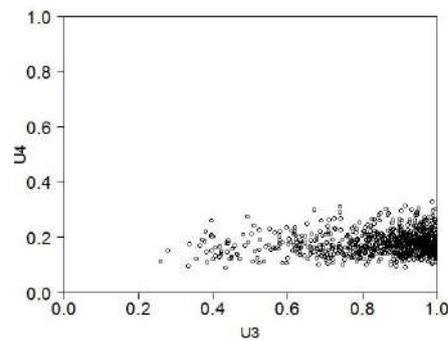
(b1) Heat map plot of (u_3, u_4) (Z-Level)   (b2) Simulated pair-plot of (u_3, u_4) (Z-Level)
Figure 10: Example of a heat-maps for the density

To quantify how good the convergence is we evaluate the plots in **Figure 10**. The plots on the left side ((a1) and (b1)) show the density heat map of the **theoretical** distribution. In this picture regions with bright color indicate high values of the density, whereas areas with low density are indicated by dark red color. These plots are evaluated by numerical integration of the density. The plots on the right side ((a2) and (b2)) show the results of the simulation study. In (b2) the bisector is drawn as well, to give a better orientation in the plot.
To evaluate whether the algorithm has converged, we can compare the pairs of plots. If the plots on the left and the right side look similar, meaning that the points on the right side are located in the same area where the density has high values, then the algorithm

has converged.

Thus, the plots in **Figure 10** would indicate good convergence, as the plots have the same shapes.

We additionally plotted the theoretical and the empirical marginal density functions to make it easier to see the regions of bad convergence.

To analyze the marginal densities, we integrate numerically the density evaluated on an equidistant grid, using the trapezoidal rule. These marginal densities are again integrated numerically to receive the cumulative distribution functions.
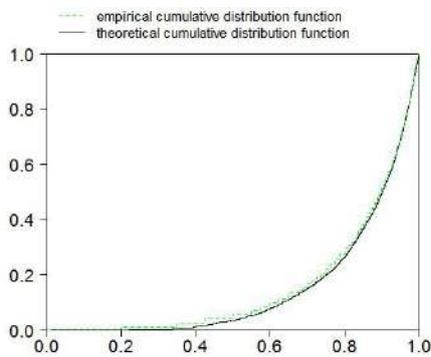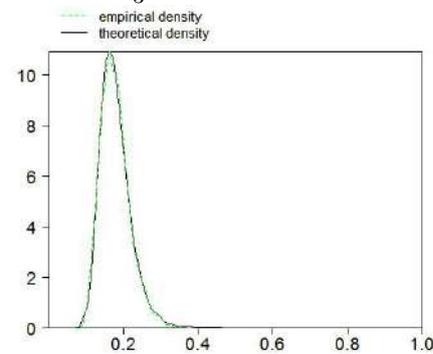

(a1) Empirical and theoretical cumulative distribution function of u_3
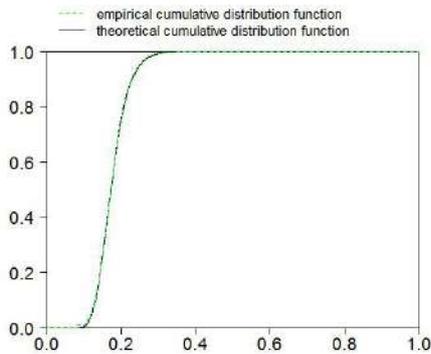
(a2) Empirical and theoretical density function of u_3

(b1) Empirical and theoretical cumulative distribution function of u_4
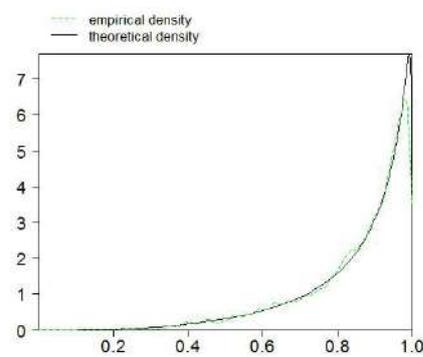
(b2) Empirical and theoretical density function of u_4

Figure 11: Example of a summary for marginal distribution

To evaluate whether the algorithm has converged we further evaluate the marginal distributions, shown in **Figure 11**. We compare the empirical cumulative distribution function and the empirical density function with the numerically evaluated theoretical cumulative distribution function and the theoretical density function. If these functions have the same shape, this indicates that the algorithm has converged, as is the case in this example.

6.2   Low-dimensional simulation study

To validate the convergence of **Algorithm 4**, which we introduced in **Section 5.7**, we first conducted a low-dimensional simulation study. The results of this study are summarized in this section.

We simulated conditional random variables of a 4 dimensional C-vine copula, where we conditioned on 2 variables. The advantage of this approach is that the random variables are 2 dimensional.

Furthermore, we choose β, the proportion of multidimensional beta updates, to be 10%. Thus, the probability of updates via adapted Gaussian increments was 90%.

To validate the algorithm in different settings, 8 different scenarios were applied, which are summarized in **Table 5**. The results of the different scenarios can be seen in **Figures 12-20**.

|	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**	**Scenario 7**	**Scenario 8**
$C_{1,2}$	Student $df = 5$ $\tau = -0.8$	Student $df = 5$ $\tau = 0.5$	Student $df = 5$ $\tau = -0.8$	Student $df = 5$ $\tau = 0.5$	Student $df = 3$ $\tau = -0.8$	Student $df = 3$ $\tau = 0.5$	Student $df = 3$ $\tau = -0.8$	Student $df = 3$ $\tau = 0.5$
$C_{1,3}$	Student $df = 8$ $\tau = 0.4$	Student $df = 8$ $\tau = 0.1$	Student $df = 8$ $\tau = 0.4$	Student $df = 8$ $\tau = 0.1$	Gumbel $\tau = 0.4$	Gumbel $\tau = 0.1$	Gumbel $\tau = 0.4$	Gumbel $\tau = 0.1$
$C_{1,4}$	Student $df = 3$ $\tau = 0.3$	Student $df = 3$ $\tau = -0.8$	Student $df = 3$ $\tau = 0.3$	Student $df = 3$ $\tau = -0.8$	180° rotated Gumbel $\tau = 0.3$	90° rotated Gumbel $\tau = -0.8$	180° rotated Gumbel $\tau = 0.3$	90° rotated Gumbel $\tau = -0.8$
$C_{2,3\|1}$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0.1$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0.1$	Student $df = 5$ $\tau = 0.7$	Student $df = 5$ $\tau = 0.1$	Student $df = 5$ $\tau = 0.7$	Student $df = 5$ $\tau = 0.1$
$C_{2,4\|1}$	Gaussian $\tau = 0.5$	Gaussian $\tau = 0.2$	Gaussian $\tau = 0.5$	Gaussian $\tau = 0.2$	Gumbel $\tau = 0.5$	Gumbel $\tau = 0.2$	Gumbel $\tau = 0.5$	Gumbel $\tau = 0.2$
$C_{3,4\|1,2}$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0$	Gaussian $\tau = 0.7$	Gaussian $\tau = 0$
Conditional values	$(u_1, u_2) = (0.95, 0.95)$	$(u_1, u_2) = (0.95, 0.95)$	$(u_1, u_3) = (0.95, 0.95)$	$(u_1, u_3) = (0.95, 0.95)$	$(u_1, u_2) = (0.95, 0.95)$	$(u_1, u_2) = (0.95, 0.95)$	$(u_1, u_3) = (0.95, 0.95)$	$(u_1, u_3) = (0.95, 0.95)$
Dependence	high	low	high	low	high	low	high	low
Symmetry	symmetric	symmetric	symmetric	symmetric	asymmetric	asymmetric	asymmetric	asymmetric
Burn-in	2000	2000	2000	2000	2000	2000	2000	2000
Simulation size (after burn-in)	10000	10000	10000	10000	10000	10000	10000	10000
$\widehat{ESS}$ of random variables (min,max)	(1015,1132)	(630,731)	(642, 834)	(369,995)	(378,390)	(631,747)	(463,722)	(915,968)

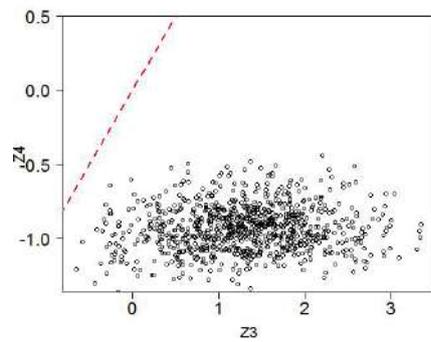Table 4: Overview of scenarios in the low-dimensional simulation study
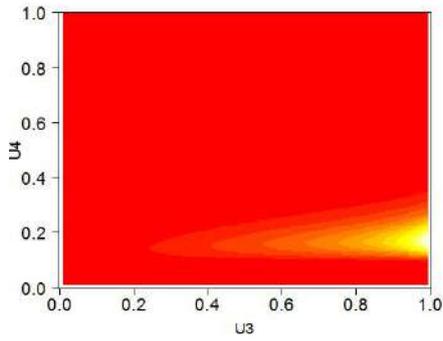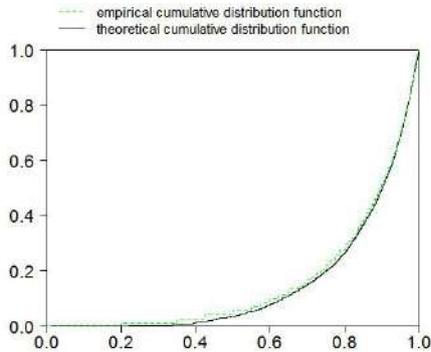
(a1) Heat map plot of (u_3, u_4) (U-Level)   (a2) Simulated pair-plot of (u_3, u_4) (U-Level)
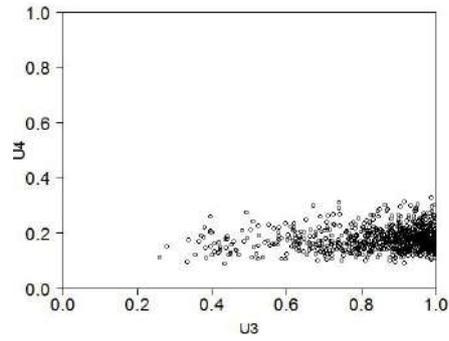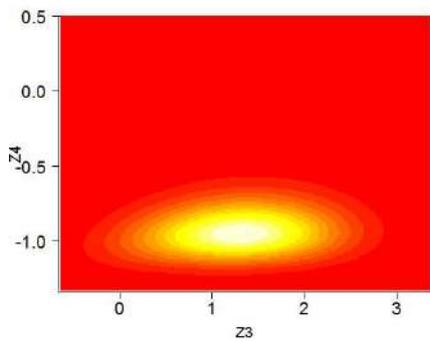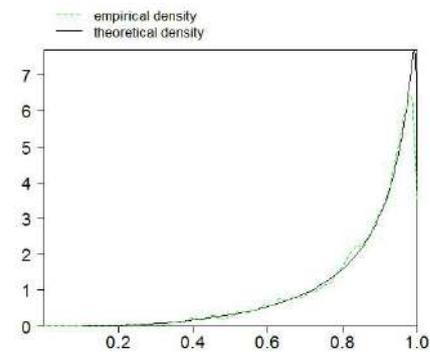
(b1) Heat map plot of (u_3, u_4) (Z-Level)   (b2) Simulated pair-plot of (u_3, u_4) (Z-Level)

(c1) Empirical and theoretical cumulative distribution function of u_3 (c2) Empirical and theoretical density function of u_3

(d1) Empirical and theoretical cumulative distribution function of u_4 (d2) Empirical and theoretical density function of u_4

Figure 12: Summary of low-dimensional scenario 1

(a1) Heat map plot of (u_3, u_4) (U-Level)

(a2) Simulated pair-plot of (u_3, u_4) (U-Level)

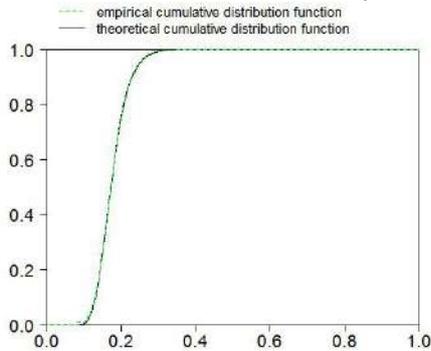(b1) Heat map plot of (u_3, u_4) (Z-Level)

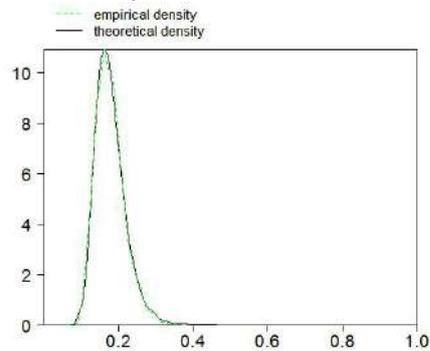(b2) Simulated pair-plot of (u_3, u_4) (Z-Level)

(c1) Empirical and theoretical cumulative distribution function of u_3

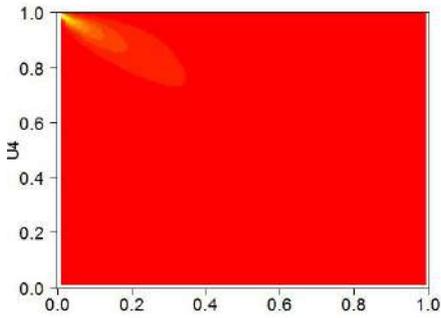(c2) Empirical and theoretical density function of u_3

(d1) Empirical and theoretical cumulative distribution function of u_4

(d2) Empirical and theoretical density function of u_4

Figure 13: Summary of low-dimensional scenario 2

(a2) Heat map plot of (u_2, u_4) (U-Level)

(a2) Simulated pair-plot of (u_2, u_4) (U-Level)

(b1) Heat map plot of (u_2, u_4) (Z-Level)

(b2) Simulated pair-plot of (u_2, u_4) (Z-Level)

(c1) Empirical and theoretical cumulative distribution function of u_2

(c2) Empirical and theoretical density function of u_2

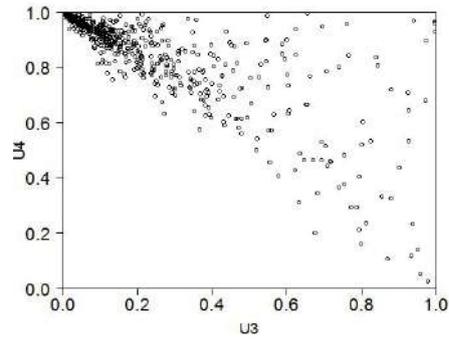(d1) Empirical and theoretical cumulative distribution function of u_4

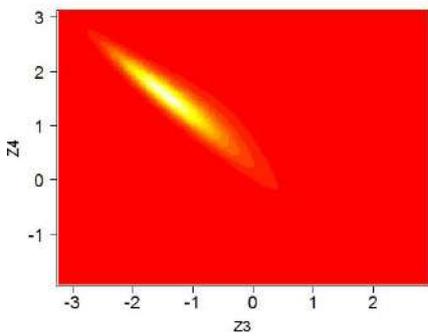(d2) Empirical and theoretical density function of u_4

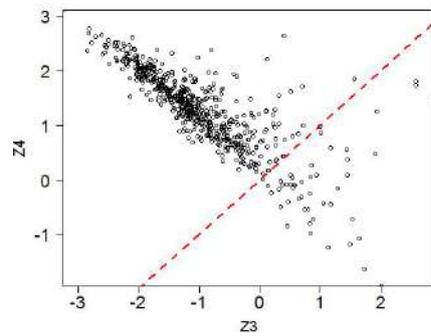Figure 14: Summary of low-dimensional scenario 3

(a1) Heat map plot of (u_2, u_4) (U-Level)

(a2) Simulated pair-plot of (u_2, u_4) (U-Level)

(b1) Heat map plot of (u_2, u_4) (Z-Level)
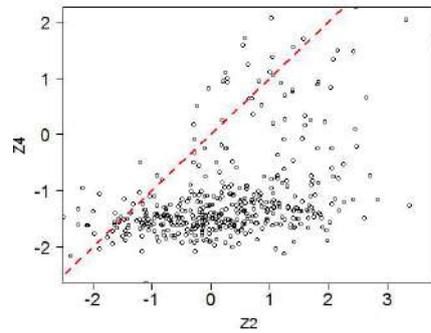
(b2) Simulated pair-plot of (u_2, u_4) (Z-Level)
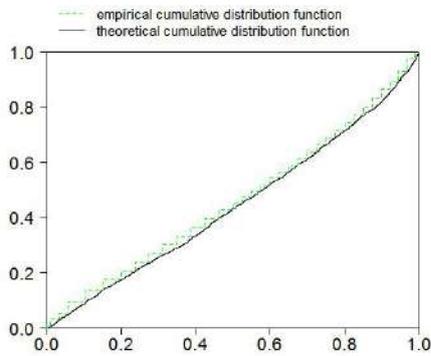
(c1) Empirical and theoretical cumulative distribution function of u_2

(c2) Empirical and theoretical density function of u_3

(d1) Empirical and theoretical cumulativedistribution function of u_4

(d2) Empirical and theoretical density function of u_4

Figure 15: Summary of low-dimensional scenario 4
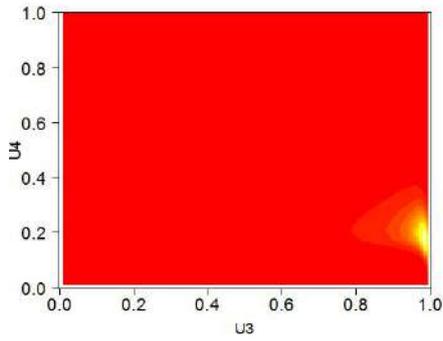
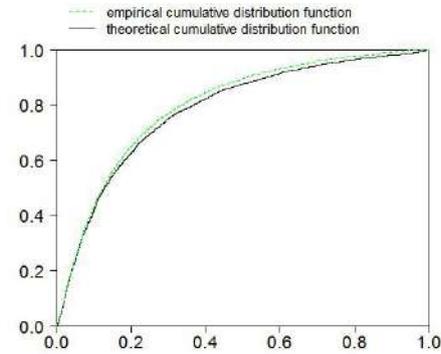(a1) Heat map plot of (u_3, u_4) (U-Level)
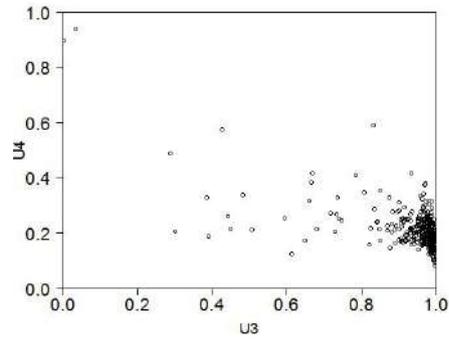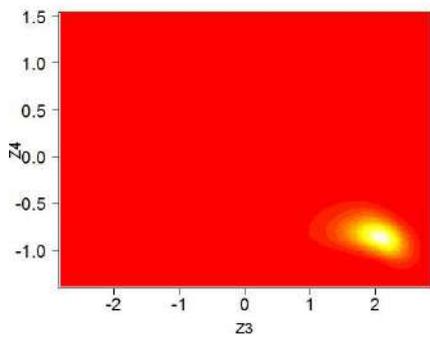
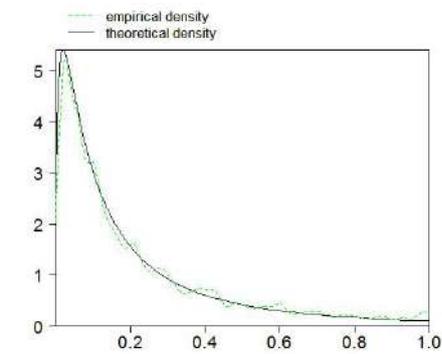(a2) Simulated pair-plot of (u_3, u_4) (U-Level)

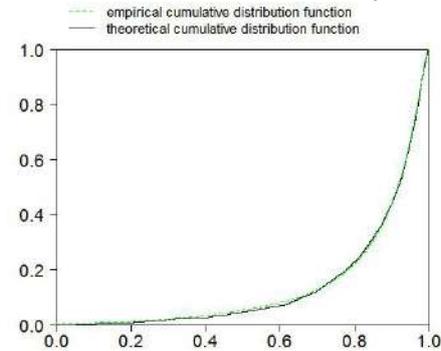(b1) Heat map plot of (u_3, u_4) (Z-Level)
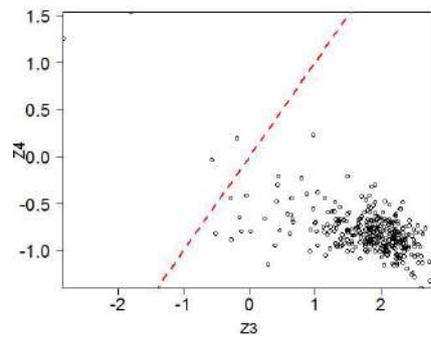
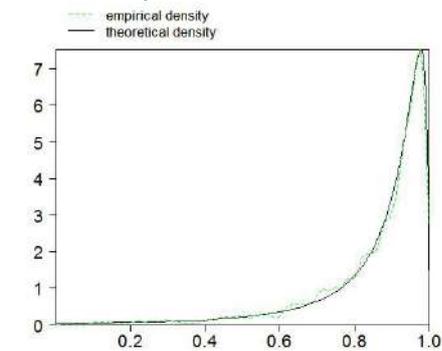(b2) Simulated pair-plot of (u_3, u_4) (Z-Level)
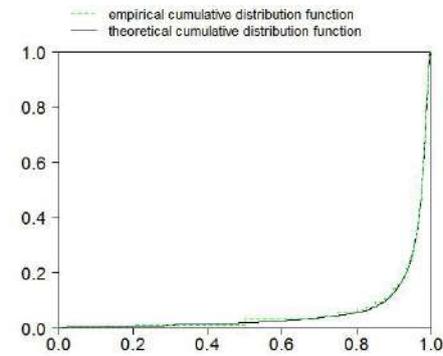
(c1) Empirical and theoretical cumulative distribution function of u_3
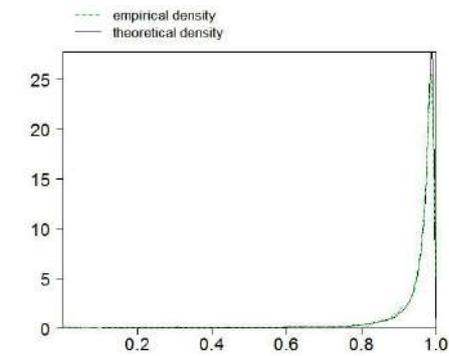
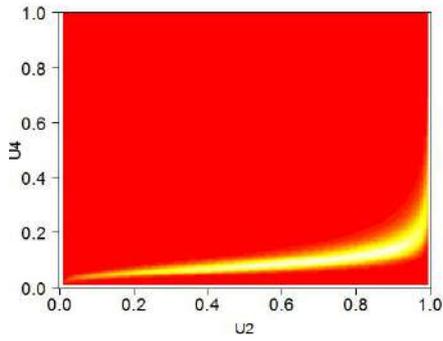(c2) Empirical and theoretical density function of u_3
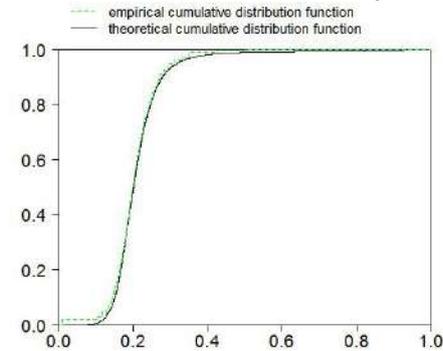
(d1) Empirical and theoretical cumulative distribution function of u_4
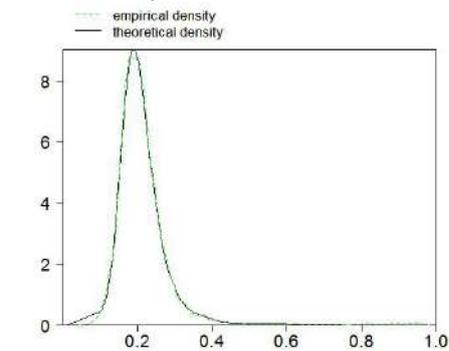
(d2) Empirical and theoretical density function of u_4

Figure 16: Summary of low-dimensional scenario 5
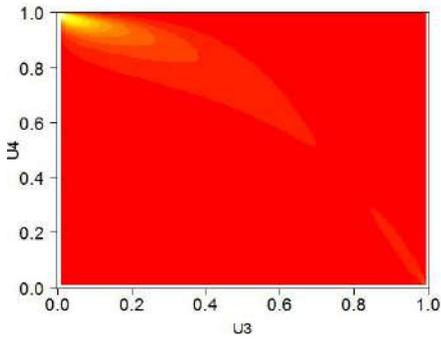
(a1) Heat map plot of (u_3, u_4) (U-Level) (a2) Simulated pair-plot of (u_3, u_4) (U-Level)
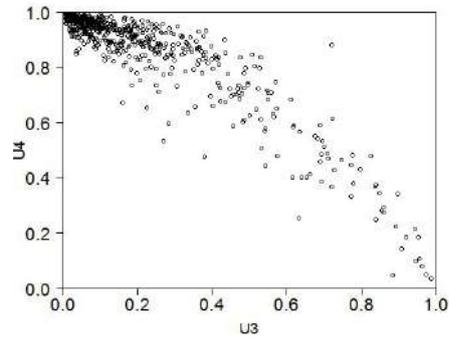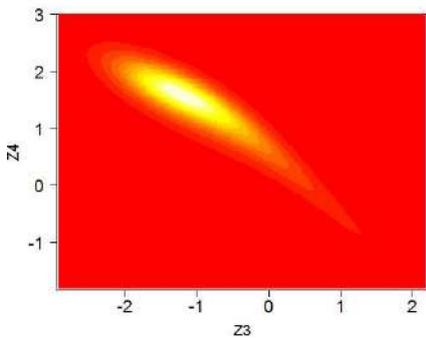
(b1) Heat map plot of (u_3, u_4) (Z-Level) (b2) Simulated pair-plot of (u_3, u_4) (Z-Level)
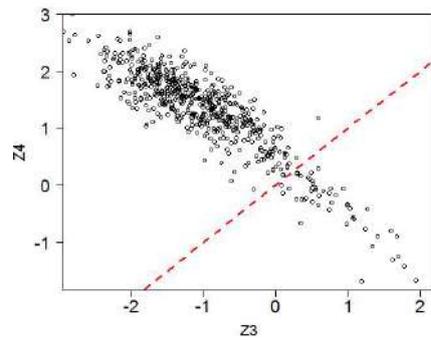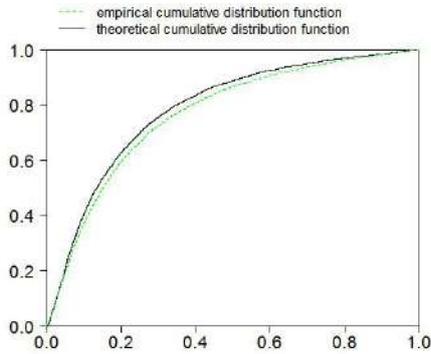
(c1) Empirical and theoretical cumulative distribution function of u_3 (c2) Empirical and theoretical density function of u_3
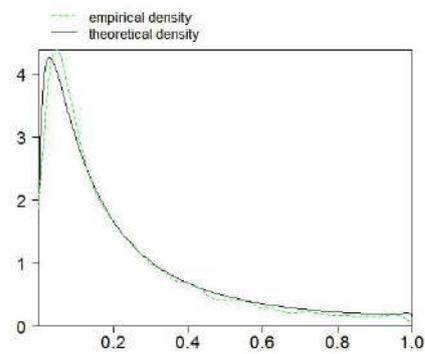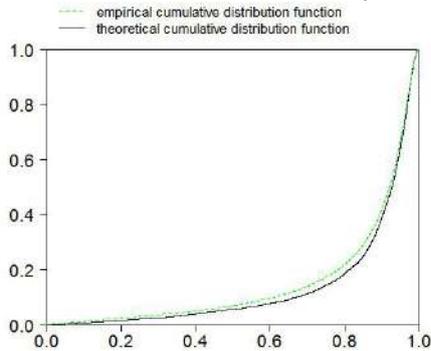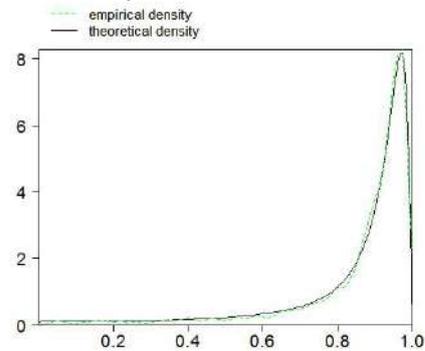
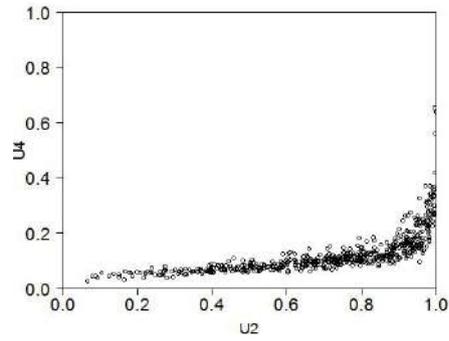(d1) Empirical and theoretical cumulative distribution function of u_4 (d2) Empirical and theoretical density function of u_4
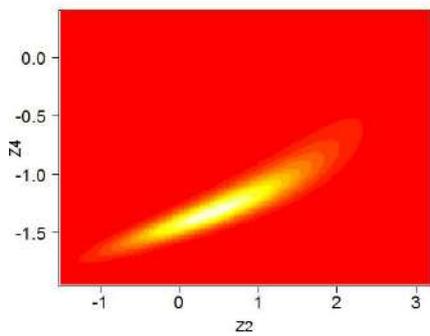
Figure 17: Summary of low-dimensional scenario 6
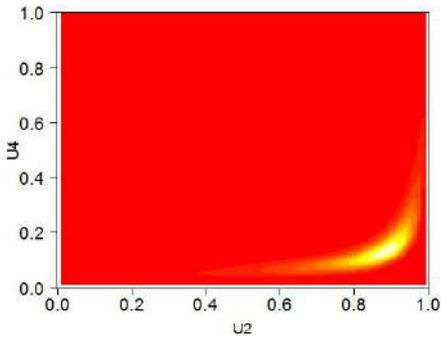
(a1) Heat map plot of (u_2, u_4) (U-Level)
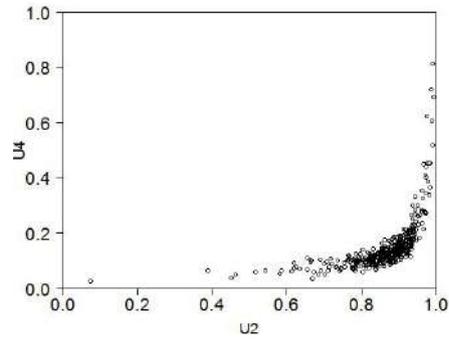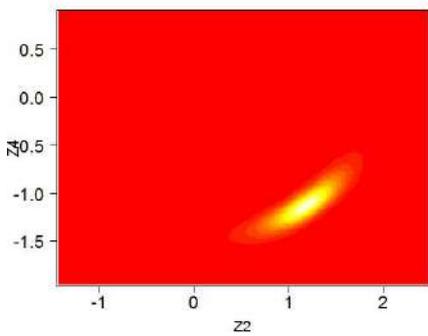
(a2) Simulated pair-plot of (u_2, u_4) (U-Level)

(b1) Heat map plot of (u_2, u_4) (Z-Level)
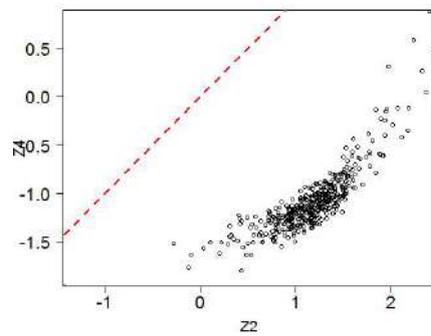
(b2) Simulated pair-plot of (u_2, u_4) (Z-Level)
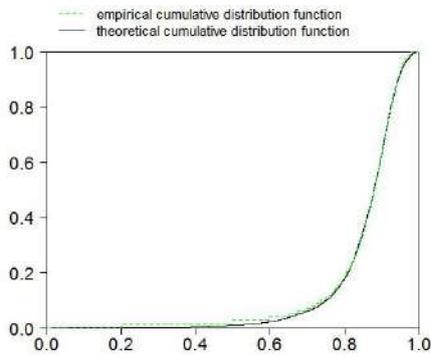
(c1) Empirical and theoretical cumulative distribution function of u_2
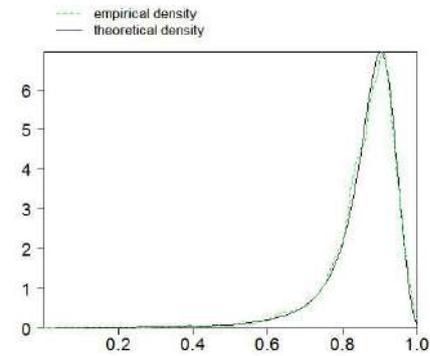
(c2) Empirical and theoretical density function of u_2
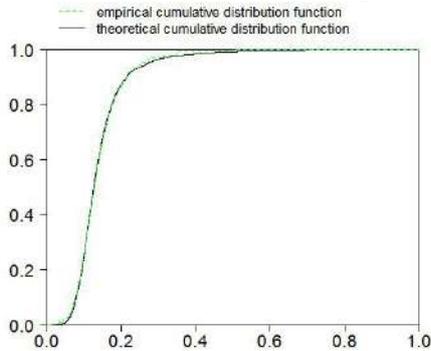
(d1) Empirical and theoretical cumulative distribution function of u_4
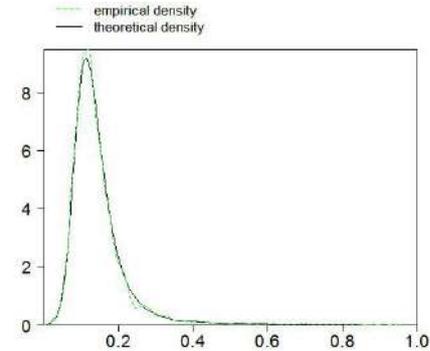
(d2) Empirical and theoretical density function of u_4
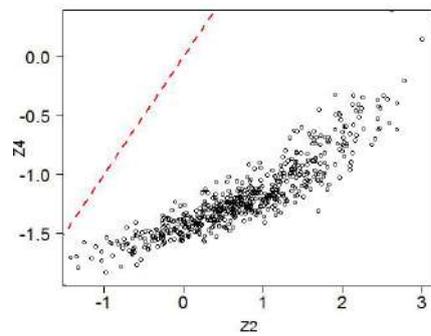
Figure 18: Summary of low-dimensional scenario 7
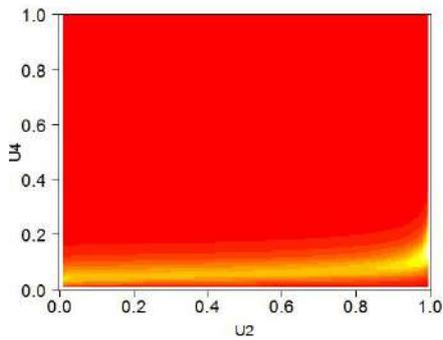
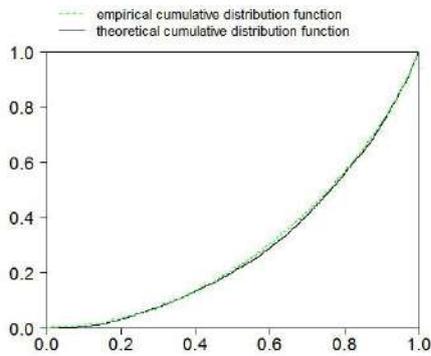(a1) Heat map plot of (u_2, u_4) (U-Level)
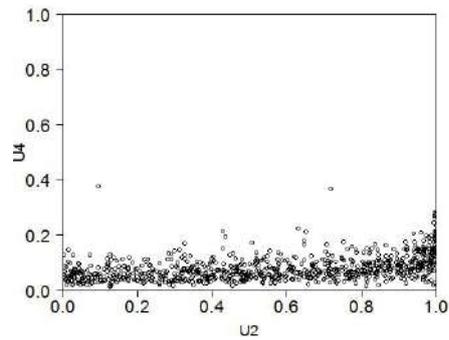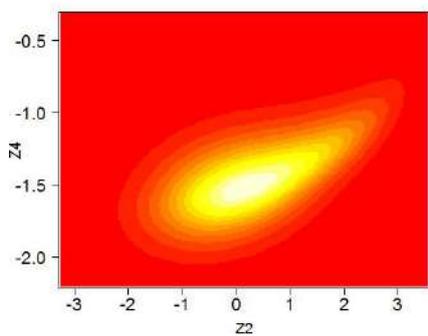
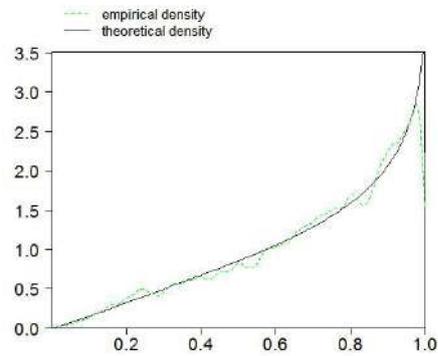(a2) Simulated pair-plot of (u_2, u_4) (U-Level)

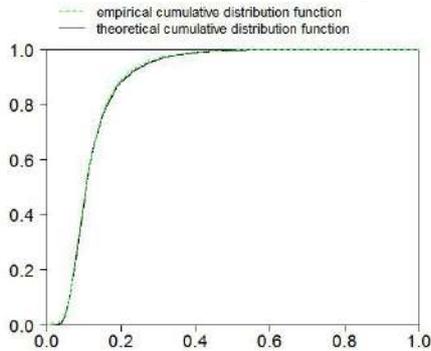(b1) Heat map plot of (u_2, u_4) (Z-Level)
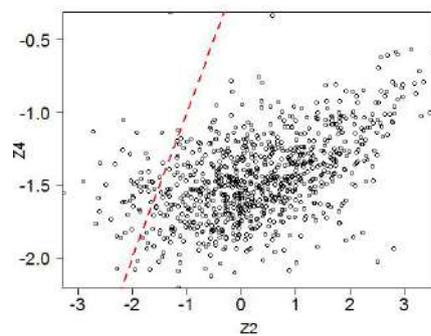
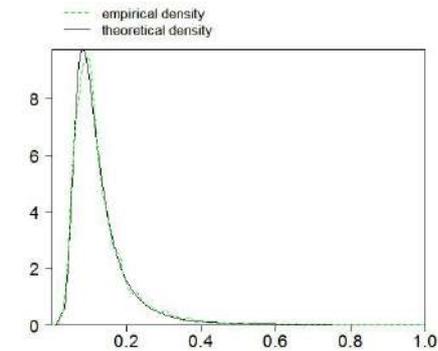(b2) Simulated pair-plot of (u_2, u_4) (Z-Level)
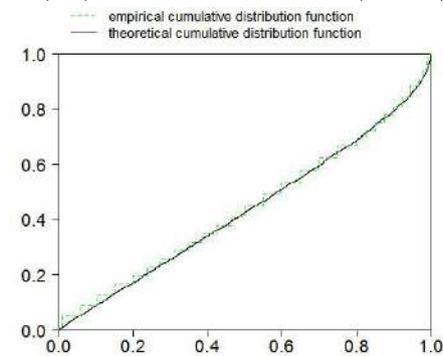
(c1) Empirical and theoretical cumulative distribution function of u_2
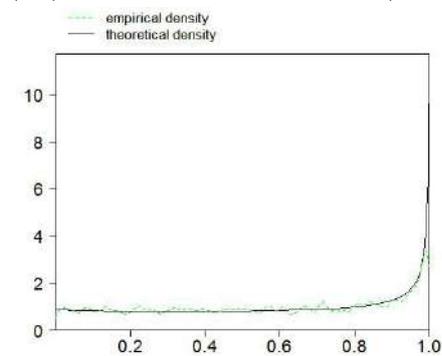
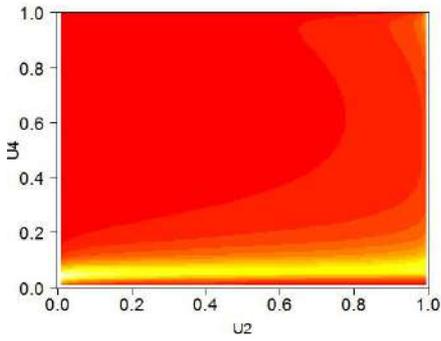(c2) Empirical and theoretical density function of u_2
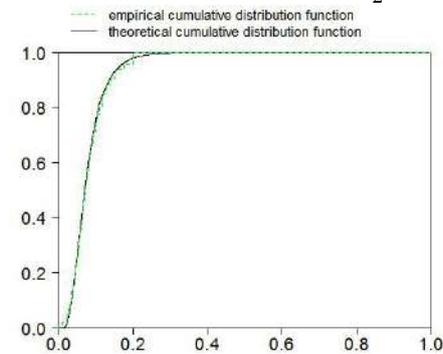
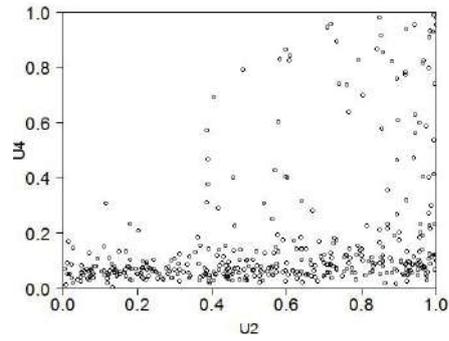(d1) Empirical and theoretical cumulative distribution function of u_4
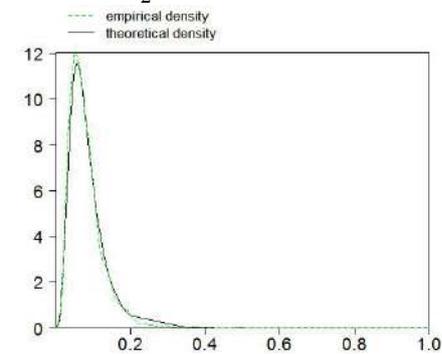
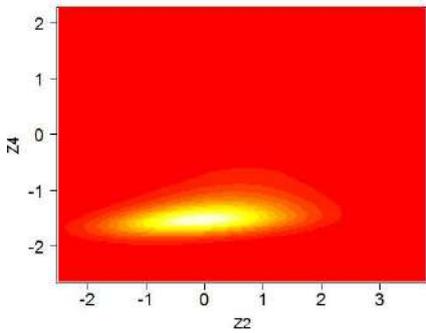(d2) Empirical and theoretical density function of u_4

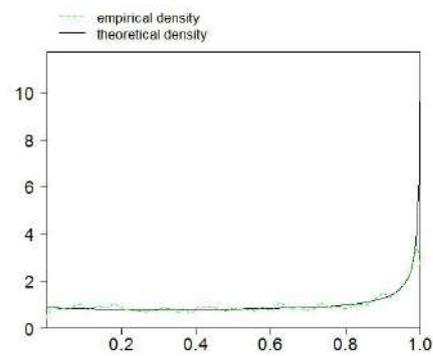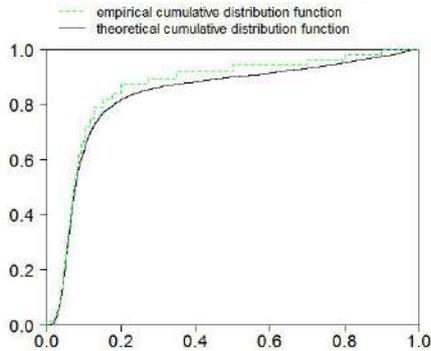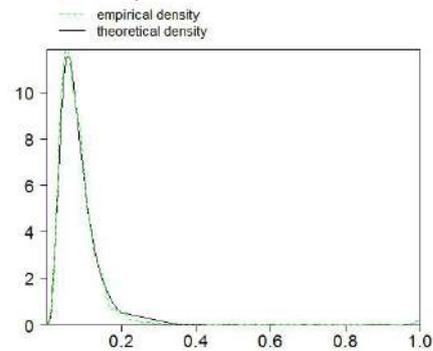Figure 19: Summary of low-dimensional scenario 8

6.3   Analysis of the resulting copulas

In this section we will analyze the pair-copulas of the resulting two dimensional vectors, given by the 2 conventional variables generated in **Section 6.2**.

To analyze the resulting dependence we performed an R-vine analysis of the resulting two dimensional vector.

At first we need to transform the entries of the vector so that the margins are uniformly distributed. For this we introduce several stochastic distributions.

At first we introduce a distribution which has as its support the interval $[0, 1]$ in which our simulations are (**U-Level**).

Definition 6.1 (Generalized Beta distribution)**.** *The generalized beta distribution with parameters (shape1= a, shape2= b,shape3= c) is defined as a random distribution with density:*

$$f(x) = \begin{cases} \dfrac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot \dfrac{c \cdot (x^{ac}) \cdot (1-x^c)^{(b-1)}}{x} & \text{if } x \in [0,1] \\ 0 & \text{else} \end{cases}$$

A further distribution which we fit to the reciprocal variable $1/U_i$ the Gamma distribution:

Definition 6.2 (Gamma distribution)**.** *The Gamma distribution with parameters (shape= a, rate= r) is defined as the distribution with density:*

$$f(x) = \begin{cases} \dfrac{r^a}{\Gamma(a)} x^{a-1} e^{-x \cdot r} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

Additionally we fitted the non-standardized Student distribution to the **L-Level** variables.

Definition 6.3 (Non-standardized Student distribution)**.** *The scale family Student distribution with parameters* (m, s, ν) *is defined as the distribution of* X:

$$X = m + Y \cdot s$$

where Y *has Student distribution with* ν *degrees of freedom, meaning that its density has the form:*

$$f_Y(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \cdot \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where Γ *is the Gamma function.*

For each variable the adequate distribution was chosen so that the transformed random variable had uniform margins.

The variables of the fitted distributions and the chosen pair-copulas for the transformed variables, are summarized in **Table 5**. In **Figure 20** we additionally evaluate the **Z-Level** plots of the transformed variables.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
Pair-copula family	Clayton	Student	Gaussian	Gumbel	90° rotated BB7	270° rotated BB7	Gaussian	BB8
Kendall's τ implied by copula parameter	.0647	-0.6981	-0.7622	0.3777	-0.3361	-0.6495	-0.6933	0.3275
Second parameter of copula	-	2.837	-	-	-0.7392	-1.3934		0.9688
Marginal transformation of u_2 resp. u_3	generalized beta (shape1=2.3535, shape2=0.8345, shape3=2.0625)	non-standardized Student on logit scale (m=-1.8665, s=1.4237, df=6.6715)	generalized beta (shape1=2.8258, shape2=0.8696, shape3=0.7358)	generalized beta (shape1=0.2867, shape2=0.5679, shape3=3.3163)	non-standardized Student on logit scale (m=3.6293, s=0.922, μ=3.9012)	non-standardized Student on logit scale (m=-1.8155, s=1.3424, μ=8.4499)	non-standardized Student on logit scale (m=1.9809, s=0.5888, μ=4.9847)	generalized beta (shape1=0.2405, shape2=0.5845, shape3=4.0227)
Marginal transformation of u_4	generalized beta (shape1=14.2489, shape2=82.4604, shape3=1.1091)	non-standardized Student on logit scale (m=2.3787, s=1.2758, μ=12.503)	reciprocal gamma (shape=4.5632, rate=0.4535)	non-standardized Student on logit scale (m=-2.5686, s=0.4951, μ=4.8203)	non-standardized Student on logit scale (m=-1.3781, s=0.2429, μ=3.0707)	non-standardized Student on logit scale (m=2.5079, s=1.0592, μ=4.832)	reciprocal gamma (shape=6.0525, rate=0.7251)	generalized beta (shape1=8.8175, shape2=27.8739, shape3=0.5513)
Symmetry of original copula	symmetric	symmetric	symmetric	symmetric	asymmetric	asymmetric	asymmetric	asymmetric
Dependence of original copula	high	low	high	low	high	low	high	low
$\widehat{ESS}$ of random variables (min,max)	(1015,1132)	(630,731)	(642, 834)	(369,995)	(378,390)	(631,747)	(463,722)	(915,968)

Table 5: Overview of scenarios in the low-dimensional simulation study
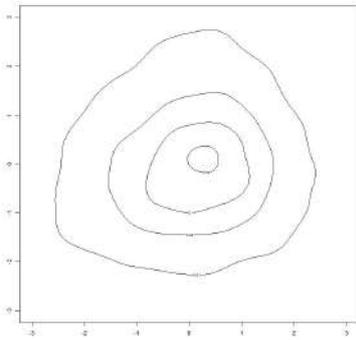
(a1) Z-Level contour plot of transformed (u_3, u_4) (Scenario 1)
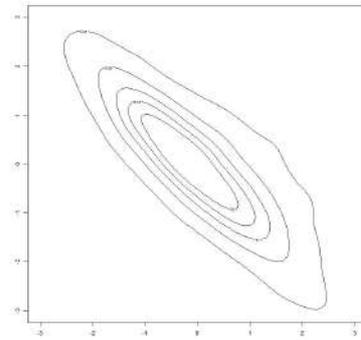
(a2) Z-Level contour plot of transformed (u_3, u_4) (Scenario 2)
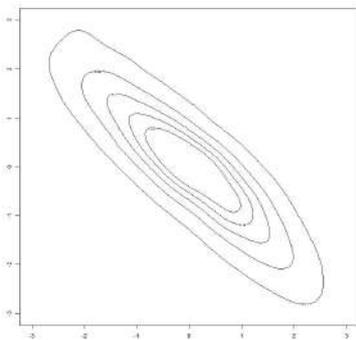
(b1) Z-Level contour plot of transformed (u_2, u_4) (Scenario 3)
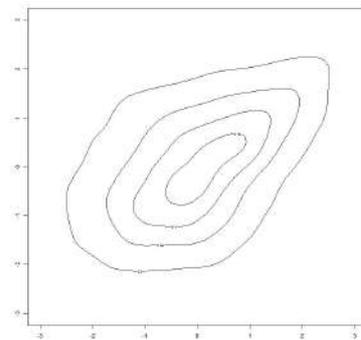
(b2) Z-Level contour plot of transformed (u_2, u_4) (Scenario 4)
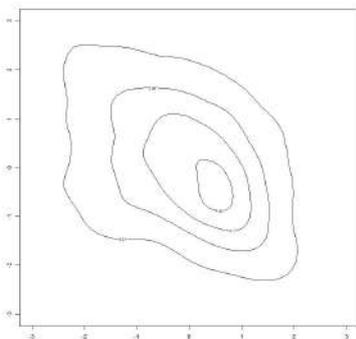
(c1) Z-Level contour plot of transformed (u_3, u_4) (Scenario 5)
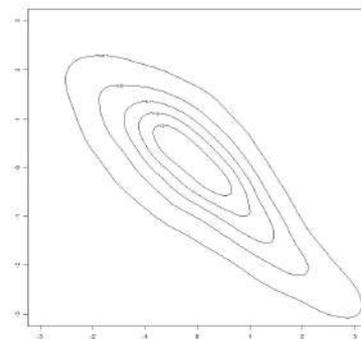
(c2) Z-Level contour plot of transformed (u_3, u_4) (Scenario 6)
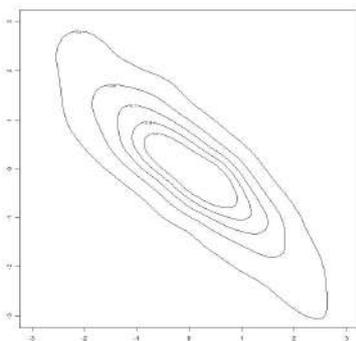
(d1) Z-Level contour plot of transformed (u_2, u_4) (Scenario 7)
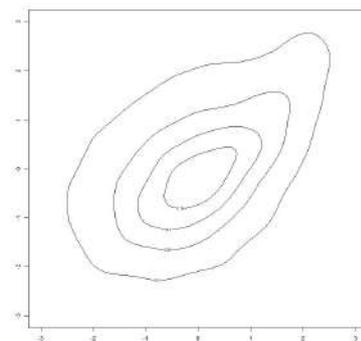
(d2) Z-Level contour plot of transformed (u_2, u_4) (Scenario 8)

Figure 20: Z-Level contour plots for the low-dimensional simulation studies

These results from **Figure 20** underline the need to evaluate the resulting distributions in more detail. Even copulas with low dependence can result in dependent copulas with high correlation. Thus, predicting the resulting dependent copula by only observing the underling copulas is not possible. This is one reason why the analysis of these resulting copulas is of great interest.

6.4   High-dimensional simulation study

The second simulation study, that we conducted is implemented in higher dimensions. In contrast to the low-dimensional case, it is no longer possible to visualize the whole distribution by a single plot. We thus focus on the marginal distribution of the random variables.

In our example we will condition in a 10-dimensional C-vine on the variable $(u_1, u_2, u_3, u_4, u_5)$. This random variable has again a C-vine structure, thus we can numerical evaluate the density of the conditional random variable $(u_6, u_7, u_8, u_9, u_{10} | u_1, u_2, u_3, u_4, u_5)$.

We evaluated the marginal density $(u_6 | u_1, u_2, u_3, u_4, u_5)$ in closed form, as it is given by the C-vine construction. Thus for this density integration was not needed. For the remaining set of variables e.g., $(u_{10} | u_1, u_2, u_3, u_4, u_5)$, the density of $(u_6, u_7, u_8, u_9, u_{10} | u_1, u_2, u_3, u_4, u_5)$ was numerically integrated over the variable (u_6, u_7, u_8, u_9) using the adaptive integration implemented R package *cubature* (see Berntsen et al., 1991). To ensure that the integration converged, we applied successive integration, in which the number of iterations was gradually increased. If the total difference of the successive values for the integration differed more than 5%, indicating not enough iterations, the integration was repeated. In the following integration, the number of iterations was increased by the factor 5 until a limit of 1000000 iterations was reached, as otherwise the computational effort would be to high. This limit was reached in several cases, which is one major reason for the insufficient convergence indicated by jagged curves as seen in the high-dimensional densities (e2) and (f2) in **Figures 21-24**. From the density functions we received the cumulative distribution function by numerical integration with the trapezoidal rule.

The integration took a long time and showed that even in 5 dimension evaluating densities numerically is no longer easily possible. The results thus underlined the need to apply MH algorithms, as these needed much less computation time.

We analyzed 4 different scenarios with several pair-copulas used in the pair-copula construction which are summarized in **Table 6**.

As the detailed list of used pair-copulas is too long, we give a more detailed list in front of each of the evaluated scenarios.

Scenario	1	2	3	4
Used pair-copulas	Student	Student	Student, Clayton, Gumbel, Frank, Joe	Student, Clayton, Gumbel, Frank ,Joe (and rotations of all copulas)
Set of values for τ	$\{0.4\}$	$\{0.4, -0.4\}$	$\{0.4\}$	$\{0.4, -0.4\}$
Conditional set	$(u_1, u_2, u_3, u_4, u_5) =$ $(0.9, 0.9, 0.9, 0.9, 0.9)$	$(u_1, u_2, u_3, u_4, u_5) =$ $(0.9, 0.9, 0.9, 0.9, 0.9)$	$(u_1, u_2, u_3, u_4, u_5) =$ $(0.9, 0.9, 0.9, 0.9, 0.9)$	$(u_1, u_2, u_3, u_4, u_5) =$ $(0.9, 0.9, 0.9, 0.9, 0.9)$
Effective sample size $\widehat{ESS}$ (min,max)	$(4305, 4588)$	$(3251, 4303)$	$(300, 756)$	$(356, 1095)$
Burn-in	2000	2000	2000	2000
Simulation size (after burn-in)	100000	100000	100000	100000

Table 6: Overview of high-dimensional scenarios

If one compares the effective sample sizes, one can see that scenarios 1 and 2 have much higher values than scenarios 3 and 4, which indicates a higher number of contained information and thus a better convergence. The explanation for this is that these scenarios have only elliptical copulas. This observation appears to be plausible, as our adaptive algorithm also applies elliptical copulas as proposal distributions, as the proposal variables are multivariate normal distributed and thus have an elliptical copula. As in **Section 5.3**, it was shown, that proposal distribution close to the distribution from which simulations we want to simulate from lead to faster convergence, this indicates that the conditional distributions seem to be elliptical, too. Thus, we can also assume in **Chapter 8** to receive better convergence when many pair-copulas in the R-vine are elliptical.

To visualize the resulting distribution, we plotted the resulting joint distribution on the U-Level and on the Z-Level. For the proof of the convergence, we plotted the marginal distribution functions and the marginal densities as describe above.

To visualize the joint distribution we will apply pair-plots on the **U-Level** and **Z-Level**. For the **Z-Level** plot we additionally plot the marginal histograms on the diagonal of the pair-plot, and the Kendall's τ on the upper triangular.

	family	τ		family	τ		family	τ
				Level 1				
$C_{1,2}$	Student $df = 5$	0.4	$C_{1,3}$	Student $df = 5$	0.4	$C_{1,4}$	Student $df = 5$	0.4
$C_{1,5}$	Student $df = 5$	0.4	$C_{1,6}$	Student $df = 5$	0.4	$C_{1,7}$	Student $df = 5$	0.4
$C_{1,8}$	Student $df = 5$	0.4	$C_{1,9}$	Student $df = 5$	0.4	$C_{1,10}$	Student $df = 5$	0.4
				Level 2				
$C_{2,3|1}$	Student $df = 5$	0.4	$C_{2,4|1}$	Student $df = 5$	0.4	$C_{2,5|1}$	Student $df = 5$	0.4
$C_{2,6|1}$	Student $df = 5$	0.4	$C_{2,7|1}$	Student $df = 5$	0.4	$C_{2,8|1}$	Student $df = 5$	0.4
$C_{2,9|1}$	Student $df = 5$	0.4	$C_{2,10|1}$	Student $df = 5$	0.4			
				Level 3				
$C_{3,4|1,2}$	Student $df = 5$	0.4	$C_{3,5|1,2}$	Student $df = 5$	0.4	$C_{3,6|1,2}$	Student $df = 5$	0.4
$C_{3,7|1,2}$	Student $df = 5$	0.4	$C_{3,8|1,2}$	Student $df = 5$	0.4	$C_{3,9|1,2}$	Student $df = 5$	0.4
$C_{3,10|1,2}$	Student $df = 5$	0.4						
				Level 4				
$C_{4,5|1,2,3}$	Student $df = 5$	0.4	$C_{4,6|1,2,3}$	Student $df = 5$	0.4	$C_{4,7|1,2,3}$	Student $df = 5$	0.4
$C_{4,8|1,2,3}$	Student $df = 5$	0.4	$C_{4,9|1,2,3}$	Student $df = 5$	0.4	$C_{4,10|1,2,3}$	Student $df = 5$	0.4
				Level 5				
$C_{5,6|1,2,3,4}$	Student $df = 5$	0.4	$C_{5,7|1,2,3,4}$	Student $df = 5$	0.4	$C_{5,8|1,2,3,4}$	Student $df = 5$	0.4
$C_{5,9|1,2,3,4}$	Student $df = 5$	0.4	$C_{5,10|1,2,3,4}$	Student $df = 5$	0.4			
				Level 6				
$C_{6,7|1,2,3,4,5}$	Student $df = 5$	0.4	$C_{6,8|1,2,3,4,5}$	Student $df = 5$	0.4	$C_{6,9|1,2,3,4,5}$	Student $df = 5$	0.4
$C_{6,10|1,2,3,4,5}$	Student $df = 5$	0.4						
				Level 7				
$C_{7,8|1,2,3,4,5,6}$	Student $df = 5$	0.4	$C_{7,9|1,2,3,4,5,6}$	Student $df = 5$	0.4	$C_{7,10|1,2,3,4,5,6}$	Student $df = 5$	0.4
				Level 8				
$C_{8,9|1,2,3,4,5,6,7}$	Student $df = 5$	0.4	$C_{8,10|1,2,3,4,5,6,7}$	Student $df = 5$	0.4			
				Level 9				
$C_{9,10|1,2,3,4,5,6,7,8}$	Student $df = 5$	0.4						

Table 7: Summary of the tree-wise chosen pair-copulas (high-dimensional scenario 1)
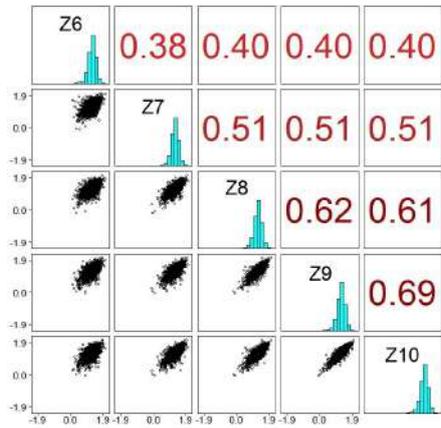
(a1) Pair-plot with Kendall's τ values of $(z_6, z_7, z_8, z_9, z_{10})$ (Z-Level)
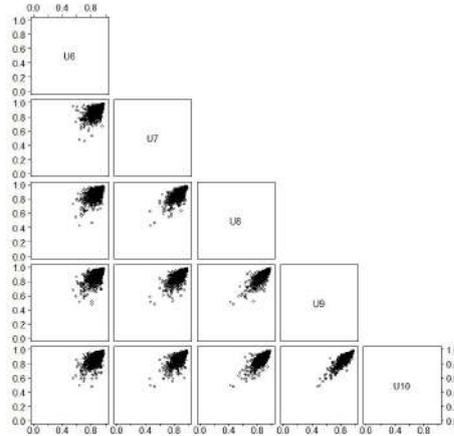
(a2) Pair-plot of $(u_6, u_7, u_8, u_9, u_{10})$ (U-Level)
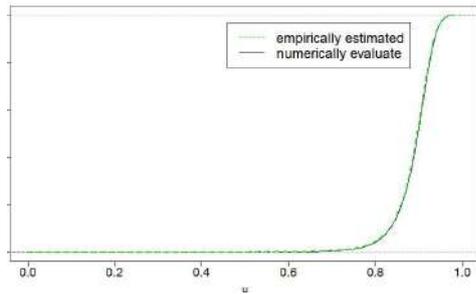
(b1) Empirical and theoretical cumulative distribution function of u_6
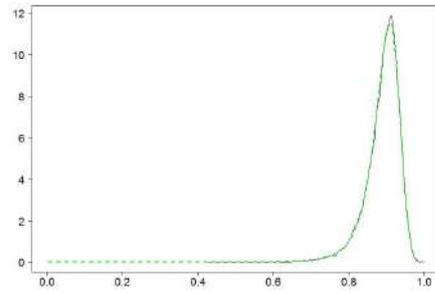
(b2) Empirical and theoretical density function of u_6

(c1) Empirical and theoretical cumulative distribution function of u_7

(c2) Empirical and theoretical density function of u_7

Figure 21: Summary of high-dimensional scenario 1

(cont.)(d1) Empirical and theoretical cumulative distribution function of u_8

(cont.)(d2) Empirical and theoretical density function of u_8

(cont.)(e1) Empirical and theoretical cumulative distribution function of u_9

(cont.)(e2) Empirical and theoretical density function of u_9

(cont.)(f1) Empirical and theoretical cumulative distribution function of u_{10}

(cont.)(f2) Empirical and theoretical density function of u_{10}

Figure 21 (cont.): Summary of high-dimensional scenario 1

	Family	τ		Family	τ		Family	τ
				Level 1				
$C_{1,2}$	Student $df = 5$	0.4	$C_{1,3}$	Student $df = 5$	0.4	$C_{1,4}$	Student $df = 5$	0.4
$C_{1,5}$	Student $df = 5$	0.4	$C_{1,6}$	Student $df = 5$	-0.4	$C_{1,7}$	Student $df = 5$	-0.4
$C_{1,8}$	Student $df = 5$	0.4	$C_{1,9}$	Student $df = 5$	0.4	$C_{1,10}$	Student $df = 5$	0.4
				Level 2				
$C_{2,3|1}$	Student $df = 5$	0.4	$C_{2,4|1}$	Student $df = 5$	0.4	$C_{2,5|1}$	Student $df = 5$	0.4
$C_{2,6|1}$	Student $df = 5$	-0.4	$C_{2,7|1}$	Student $df = 5$	-0.4	$C_{2,8|1}$	Student $df = 5$	0.4
$C_{2,9|1}$	Student $df = 5$	0.4	$C_{2,10|1}$	Student $df = 5$	0.4			
				Level 3				
$C_{3,4|1,2}$	Student $df = 5$	0.4	$C_{3,5|1,2}$	Student $df = 5$	0.4	$C_{3,6|1,2}$	Student $df = 5$	-0.4
$C_{3,7|1,2}$	Student $df = 5$	-0.4	$C_{3,8|1,2}$	Student $df = 5$	0.4	$C_{3,9|1,2}$	Student $df = 5$	0.4
$C_{3,10|1,2}$	Student $df = 5$	0.4						
				Level 4				
$C_{4,5|1,2,3}$	Student $df = 5$	0.4	$C_{4,6|1,2,3}$	Student $df = 5$	-0.4	$C_{4,7|1,2,3}$	Student $df = 5$	-0.4
$C_{4,8|1,2,3}$	Student $df = 5$	0.4	$C_{4,9|1,2,3}$	Student $df = 5$	0.4	$C_{4,10|1,2,3}$	Student $df = 5$	0.4
				Level 5				
$C_{5,6|1,2,3,4}$	Student $df = 5$	-0.4	$C_{5,7|1,2,3,4}$	Student $df = 5$	-0.4	$C_{5,8|1,2,3,4}$	Student $df = 5$	0.4
$C_{5,9|1,2,3,4}$	Student $df = 5$	0.4	$C_{5,10|1,2,3,4}$	Student $df = 5$	0.4			
				Level 6				
$C_{6,7|1,2,3,4,5}$	Student $df = 5$	-0.4	$C_{6,8|1,2,3,4,5}$	Student $df = 5$	0.4	$C_{6,9|1,2,3,4,5}$	Student $df = 5$	0.4
$C_{6,10|1,2,3,4,5}$	Student $df = 5$	0.4						
				Level 7				
$C_{7,8|1,2,3,4,5,6}$	Student $df = 5$	0.4	$C_{7,9|1,2,3,4,5,6}$	Student $df = 5$	0.4	$C_{7,10|1,2,3,4,5,6}$	Student $df = 5$	0.4
				Level 8				
$C_{8,9|1,2,3,4,5,6,7}$	Student $df = 5$	0.4	$C_{8,10|1,2,3,4,5,6,7}$	Student $df = 5$	0.4			
				Level 9				
$C_{9,10|1,2,3,4,5,6,7,8}$	Student $df = 5$	0.4						

Table 8: Summary of the tree-wise chosen pair-copulas (high-dimensional scenario 2)

(a1) Pair-plot with Kendall's τ values of $(z_6, z_7, z_8, z_9, z_{10})$ (Z-Level)

(a2) Pair-plot of $(u_6, u_7, u_8, u_9, u_{10})$ (U-Level)
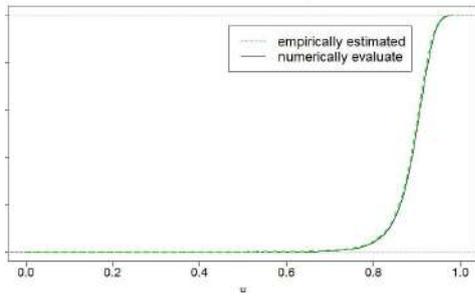
(b1) Empirical and theoretical cumulative distribution function of u_6
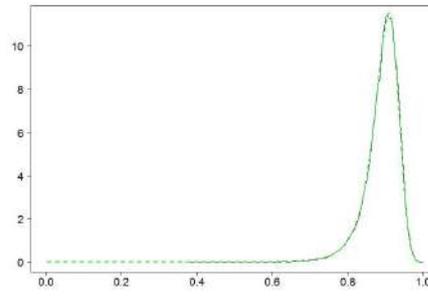
(b2) Empirical and theoretical density function of u_6
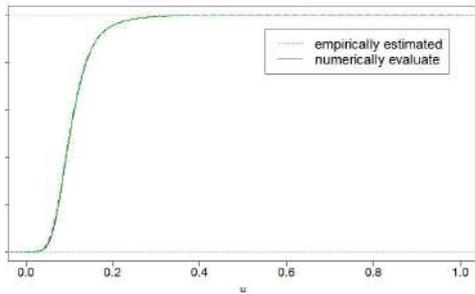
(c1) Empirical and theoretical cumulative distribution function of u_7
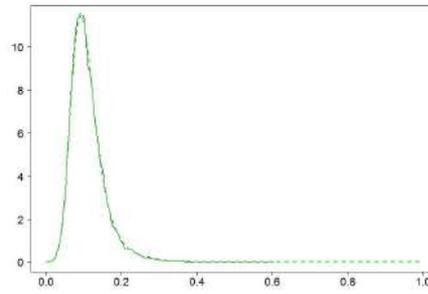
(c2) Empirical and theoretical density function of u_7

Figure 22: Summary of high-dimensional scenario 2
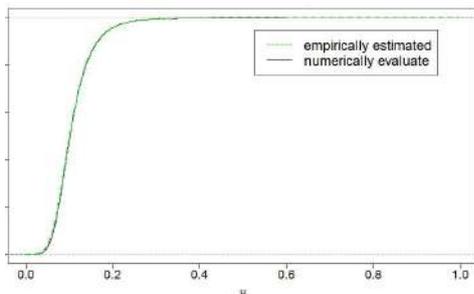
(cont.)(d1) Empirical and theoretical cumulative distribution function of u_8
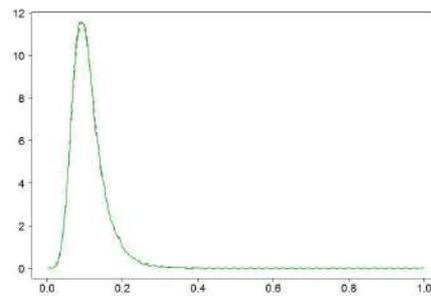
(cont.)(d2) Empirical and theoretical density function of u_8
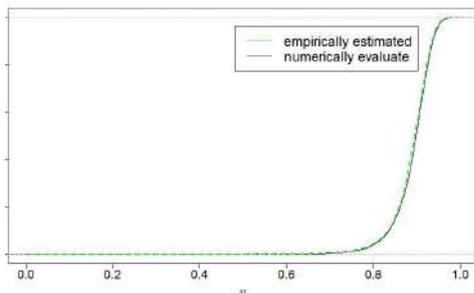
(cont.)(e1) Empirical and theoretical cumulative distribution function of u_9
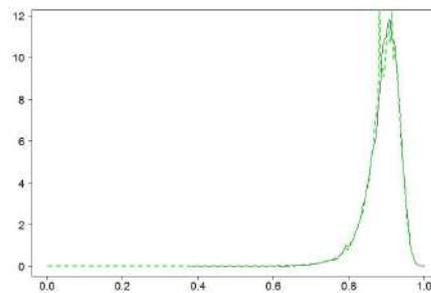
(cont.)(e2) Empirical and theoretical density function of u_9
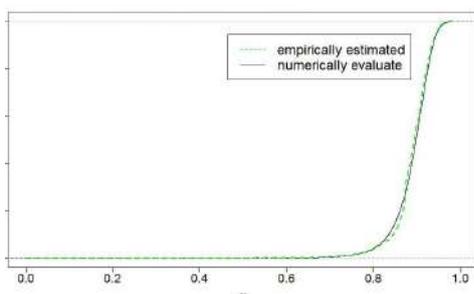
(cont.)(f1) Empirical and theoretical cumulative distribution function of u_{10}
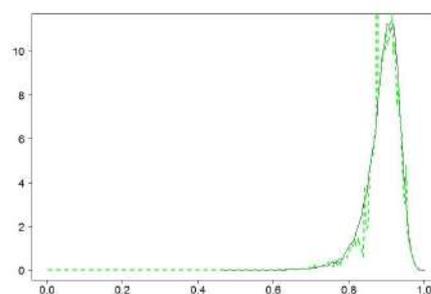
(cont.)(f2) Empirical and theoretical density function of u_{10}

Figure 22 (cont.): Summary of high-dimensional scenario 2

	Family	τ		Family	τ		Family	τ
			Level 1					
$C_{1,2}$	Student $df=5$	0.4	$C_{1,3}$	Student $df=5$	0.4	$C_{1,4}$	Student $df=5$	0.4
$C_{1,5}$	Clayton	0.4	$C_{1,6}$	Frank	0.4	$C_{1,7}$	Student $df=5$	0.4
$C_{1,8}$	Student $df=5$	0.4	$C_{1,9}$	Gumbel	0.4	$C_{1,10}$	Clayton	0.4
			Level 2					
$C_{2,3|1}$	Student $df=5$	0.4	$C_{2,4|1}$	Student $df=5$	0.4	$C_{2,5|1}$	Clayton	0.4
$C_{2,6|1}$	Frank	0.4	$C_{2,7|1}$	Student $df=5$	0.4	$C_{2,8|1}$	Student $df=5$	0.4
$C_{2,9|1}$	Gumbel	0.4	$C_{2,10|1}$	Clayton	0.4			
			Level 3					
$C_{3,4|1,2}$	Student $df=5$	0.4	$C_{3,5|1,2}$	Clayton	0.4	$C_{3,6|1,2}$	Frank	0.4
$C_{3,7|1,2}$	Student $df=5$	0.4	$C_{3,8|1,2}$	Student $df=5$	0.4	$C_{3,9|1,2}$	Gumbel	0.4
$C_{3,10|1,2}$	Clayton	0.4						
			Level 4					
$C_{4,5|1,2,3}$	Clayton	0.4	$C_{4,6|1,2,3}$	Frank	0.4	$C_{4,7|1,2,3}$	Student $df=5$	0.4
$C_{4,8|1,2,3}$	Student $df=5$	0.4	$C_{4,9|1,2,3}$	Gumbel	0.4	$C_{4,10|1,2,3}$	Clayton	0.4
			Level 5					
$C_{5,6|1,2,3,4}$	Frank	0.4	$C_{5,7|1,2,3,4}$	Student $df=5$	0.4	$C_{5,8|1,2,3,4}$	Student $df=5$	0.4
$C_{5,9|1,2,3,4}$	Gumbel	0.4	$C_{5,10|1,2,3,4}$	Clayton	0.4			
			Level 6					
$C_{6,7|1,2,3,4,5}$	Student $df=5$	0.4	$C_{6,8|1,2,3,4,5}$	Student $df=5$	0.4	$C_{6,9|1,2,3,4,5}$	Gumbel	0.4
$C_{6,10|1,2,3,4,5}$	Clayton	0.4						
			Level 7					
$C_{7,8|1,2,3,4,5,6}$	Student $df=5$	0.4	$C_{7,9|1,2,3,4,5,6}$	Gumbel	0.4	$C_{7,10|1,2,3,4,5,6}$	Clayton	0.4
			Level 8					
$C_{8,9|1,2,3,4,5,6,7}$	Gumbel	0.4	$C_{8,10|1,2,3,4,5,6,7}$	Clayton	0.4			
			Level 9					
$C_{9,10|1,2,3,4,5,6,7,8}$	Clayton	0.4						

Table 9: Summary of the tree-wise chosen pair-copulas (high-dimensional scenario 3)
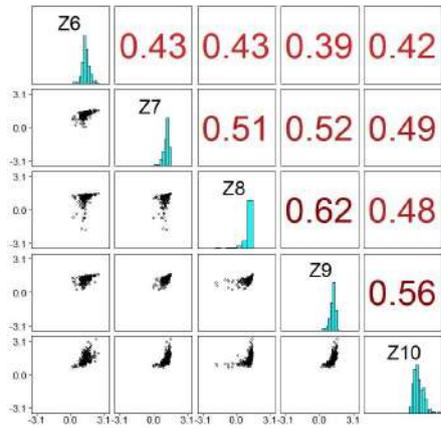
(a1) Pair-plot with Kendall's τ values of $(z_6, z_7, z_8, z_9, z_{10})$ (Z-Level)
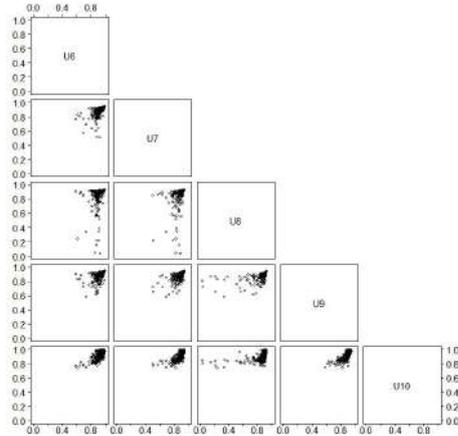
(a2) Pair-plot of $(u_6, u_7, u_8, u_9, u_{10})$ (U-Level)
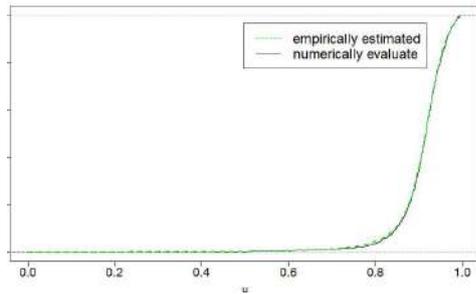
(b1) Empirical and theoretical cumulative distribution function of u_6
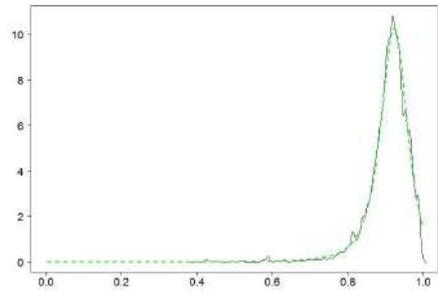
(b2) Empirical and theoretical density function of u_6
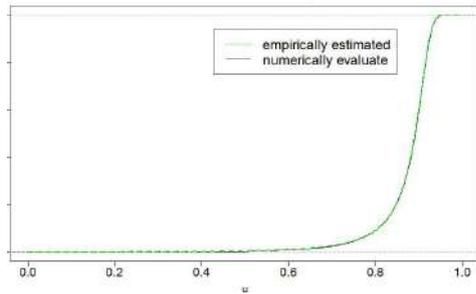
(c1) Empirical and theoretical cumulative distribution function of u_7
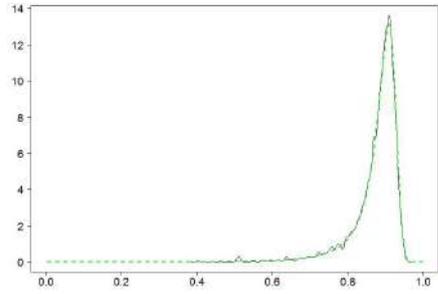
(c2) Empirical and theoretical density function of u_7

Figure 23: Summary of high-dimensional scenario 3
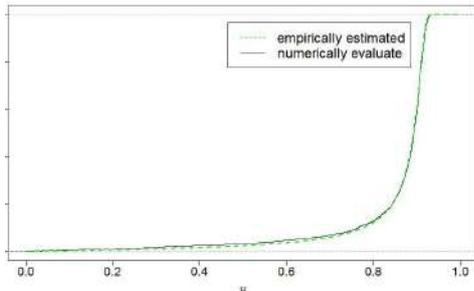
(cont.)(d1) Empirical and theoretical cumulative distribution function of u_8
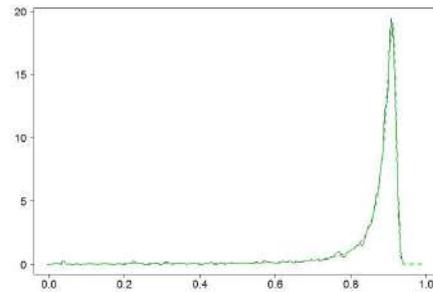
(cont.)(d2) Empirical and theoretical density function of u_8
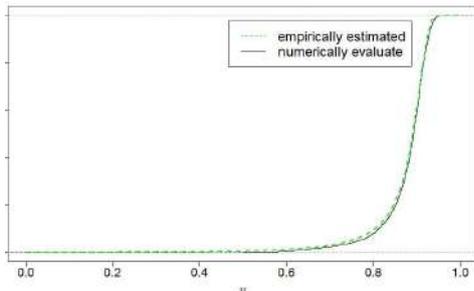
(cont.)(e1) Empirical and theoretical cumulative distribution function of u_9
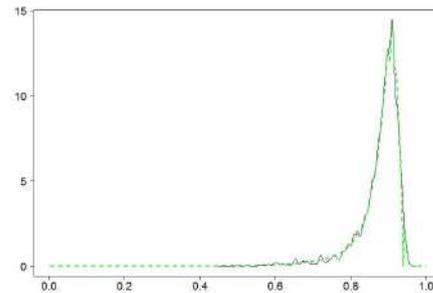
(cont.)(e2) Empirical and theoretical density function of u_9
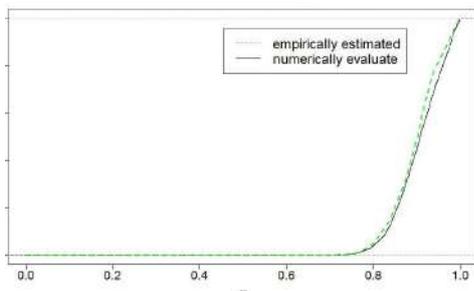
(cont.)(f1) Empirical and theoretical cumulative distribution function of u_{10}
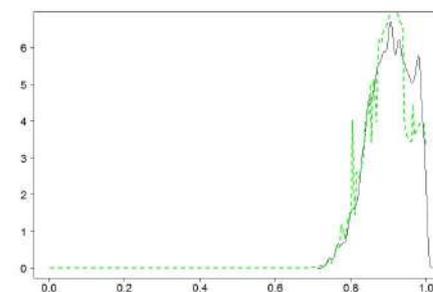
(cont.)(f2) Empirical and theoretical density function of u_{10}

Figure 23 (cont.): Summary of high-dimensional scenario 3

	Family	τ		Family	τ		Family	τ
				Level 1				
$C_{1,2}$	Student $df=5$	0.4	$C_{1,3}$	Student $df=5$	0.4	$C_{1,4}$	Student $df=5$	0.4
$C_{1,5}$	Clayton	0.4	$C_{1,6}$	Frank	-0.4	$C_{1,7}$	90° rotated Gumbel	-0.4
$C_{1,8}$	90° rotated Joe	0.4	$C_{1,9}$	Gumbel	0.4	$C_{1,10}$	Clayton	0.4
				Level 2				
$C_{2,3|1}$	Student $df=5$	0.4	$C_{2,4|1}$	Student $df=5$	0.4	$C_{2,5|1}$	Clayton	0.4
$C_{2,6|1}$	Frank	-0.4	$C_{2,7|1}$	90° rotated Gumbel	-0.4	$C_{2,8|1}$	90° rotated Joe	0.4
$C_{2,9|1}$	Gumbel	0.4	$C_{2,10|1}$	Clayton	0.4			
				Level 3				
$C_{3,4|1,2}$	Student $df=5$	0.4	$C_{3,5|1,2}$	Clayton	0.4	$C_{3,6|1,2}$	Frank	-0.4
$C_{3,7|1,2}$	90° rotated Gumbel	-0.4	$C_{3,8|1,2}$	90° rotated Joe	0.4	$C_{3,9|1,2}$	Gumbel	0.4
$C_{3,10|1,2}$	Clayton	0.4						
				Level 4				
$C_{4,5|1,2,3}$	Clayton	0.4	$C_{4,6|1,2,3}$	Frank	0.4	$C_{4,7|1,2,3}$	90° rotated Gumbel	0.4
$C_{4,8|1,2,3}$	90° rotated Joe	0.4	$C_{4,9|1,2,3}$	Gumbel	0.4	$C_{4,10|1,2,3}$	Clayton	0.4
				Level 5				
$C_{5,6|1,2,3,4}$	Frank	-0.4	$C_{5,7|1,2,3,4}$	90° rotated Gumbel	-0.4	$C_{5,8|1,2,3,4}$	90° rotated Joe	0.4
$C_{5,9|1,2,3,4}$	Gumbel	0.4	$C_{5,10|1,2,3,4}$	Clayton	0.4			
				Level 6				
$C_{6,7|1,2,3,4,5}$	90° rotated Gumbel	0.4	$C_{6,8|1,2,3,4,5}$	90° rotated Joe	0.4	$C_{6,9|1,2,3,4,5}$	Gumbel	0.4
$C_{6,10|1,2,3,4,5}$	Clayton	-0.4						
				Level 7				
$C_{7,8|1,2,3,4,5,6}$	90° rotated Joe	0.4	$C_{7,9|1,2,3,4,5,6}$	Gumbel	0.4	$C_{7,10|1,2,3,4,5,6}$	Clayton	0.4
				Level 8				
$C_{8,9|1,2,3,4,5,6,7}$	Gumbel	0.4	$C_{8,10|1,2,3,4,5,6,7}$	Clayton	0.4			
				Level 9				
$C_{9,10|1,2,3,4,5,6,7,8}$	Clayton	0.4						

Table 10: Summary of the tree-wise chosen pair-copulas (high-dimensional scenario 4)
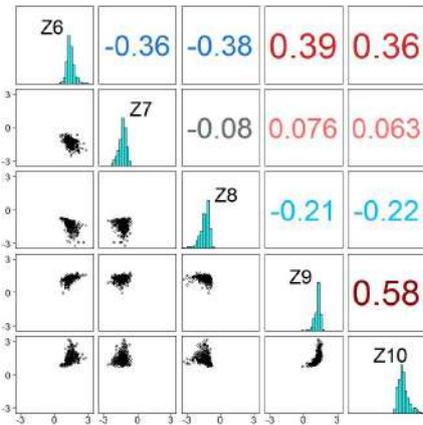
(a1) Pair-plot with Kendall's τ values of $(z_6, z_7, z_8, z_9, z_{10})$ (Z-Level)
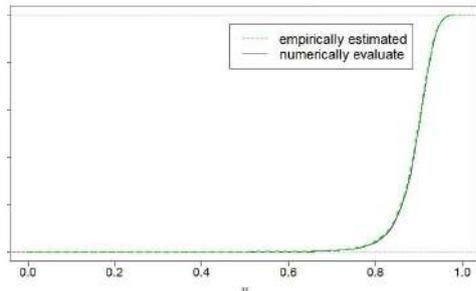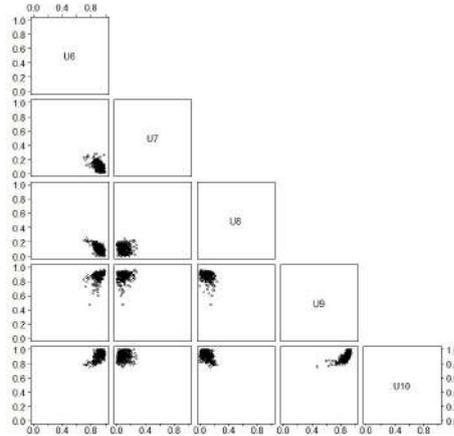

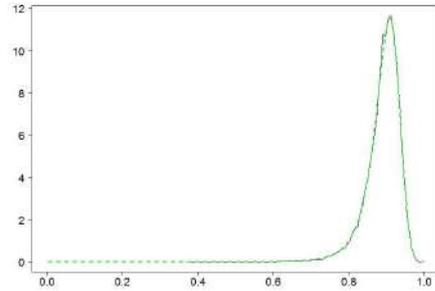(a2) Pair-plot of $(u_6, u_7, u_8, u_9, u_{10})$ (U-Level)
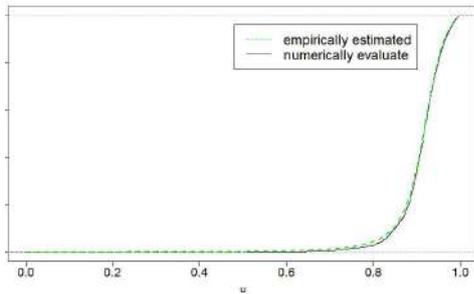
(b1) Empirical and theoretical cumulative distribution function of u_6
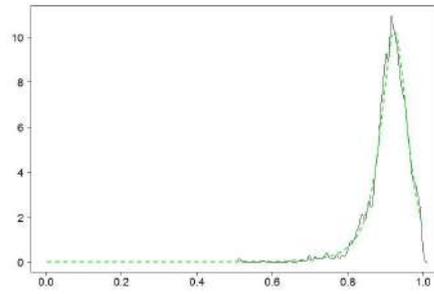
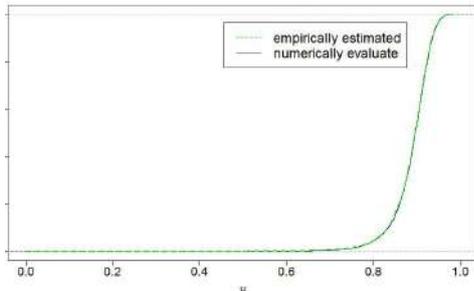
(b2) Empirical and theoretical density function of u_6
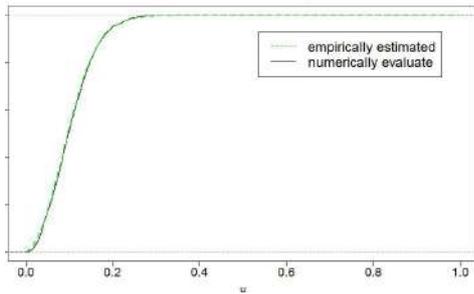
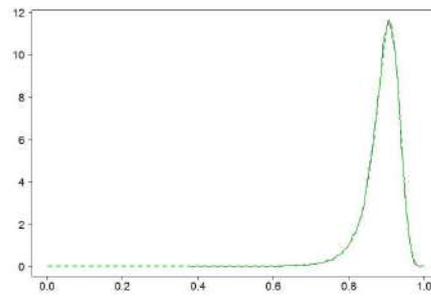(c1) Empirical and theoretical cumulative distribution function of u_7
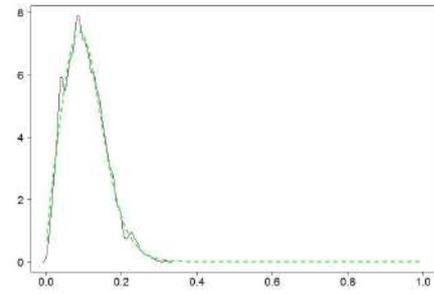
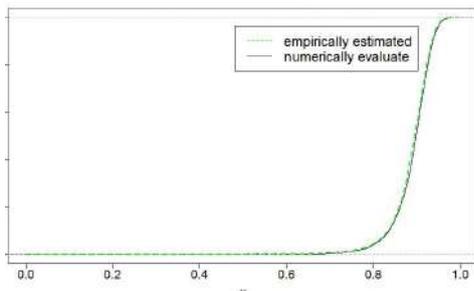(c2) Empirical and theoretical density function of u_7
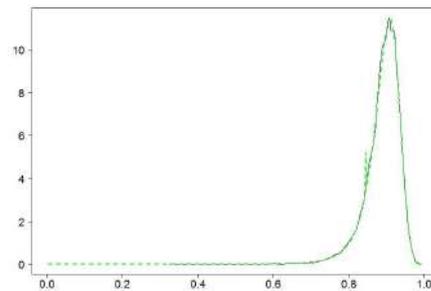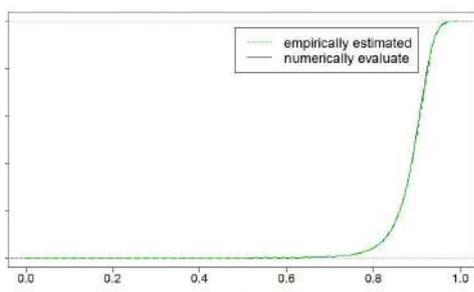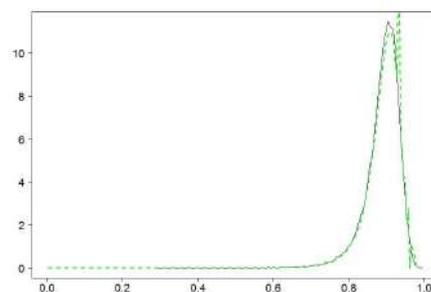
Figure 24: Summary of high-dimensional scenario 4
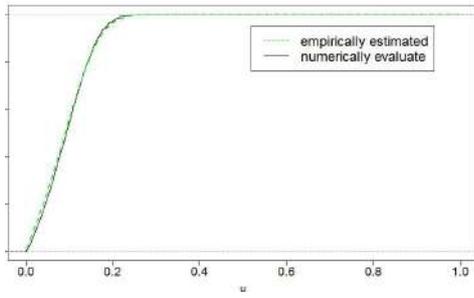
(cont.)(d1) Empirical and theoretical cumulative distribution function of u_8
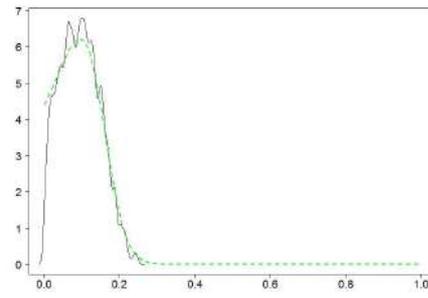
(cont.)(d2) Empirical and theoretical density function of u_8
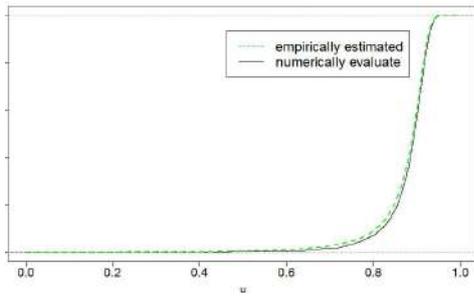
(cont.)(e1) Empirical and theoretical cumulative distribution function of u_9
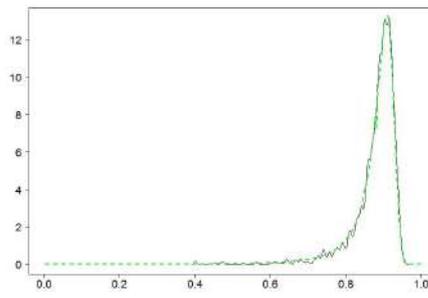
(cont.)(e2) Empirical and theoretical density function of u_9

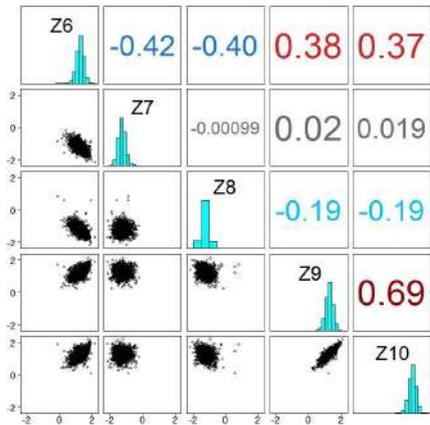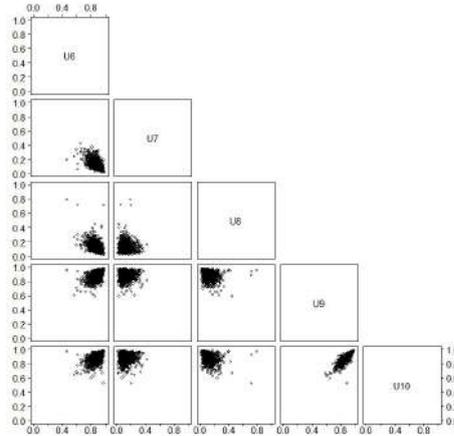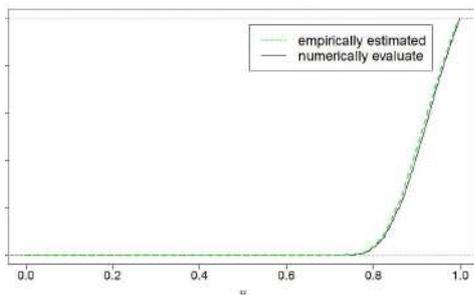(cont.)(f1) Empirical and theoretical cumulative distribution function of u_{10}
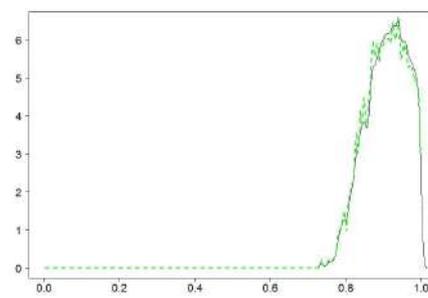
(cont.)(f2) Empirical and theoretical density function of u_{10}

Figure 24 (cont.): Summary of high-dimensional scenario 4

6.5 Conclusions of the implications on this work

Summarizing the results of the simulation studies applied, we can say that the MH algorithm can be regarded as having converged to the stationary distribution if the test by Welch and Heidelberger does not suggest further need for a burn-in and $\widehat{ESS} \geq 100$. However,bigger values are recommended. If the values of $\widehat{ESS}$ are too small, one should be careful when considering the interpretation of the results, as they can contain a high uncertainty.

One observation during the simulation was that, small scale simulations require less iterations to get certain levels of $\widehat{ESS}$. This is caused by the fact, that smaller dimensioned vectors also have a smaller number of dependencies within the vector which have to be matched by the simulated random variables.

A further observation is, that the simulation of R-vines with elliptical copulas also results in higher values of $\widehat{ESS}$. As the proposal distribution of our adaptive MH algorithm is also elliptical, this fits with the theoretical findings derived in **Chapter 4**. As we showed in **Section 5.3**, the convergence is improved when the proposal distribution is close to the stationary distribution.

7   Descriptive analysis of the empirical data set

So far, we have examined the feasibility and validity of the methods used. In this chapter, we will thoroughly describe the empirical data set which we have used for the current work. As we use market data on credit default swap (CDS) spreads, we will therefore first give a comprehensive insight into the meaning of CDS spreads and explain possible benefits of this approach. In the second part, a detailed overview of the institutions contained in the data set will be provided, followed by a descriptive analysis of these data along with regression analysis.

7.1   CDS spreads

Using balance sheets gives an economic point of view which is based on their capability to predict the future of a company. However, the approach to analyze CDS spreads is more of an actuarial/financial-mathematical point of view based on statistical methods. In this approach, the predictions are based on statistical methods.

A CDS is a bilateral over-the-counter credit derivative contract that enables trading the default risks of an underlying corporation.

The protection buyer pays a fixed annual premium, called a CDS spread, for a fixed period of time to the protection seller, who in turn compensates the protection buyer in case of a credit event of the reference entity. Thus, since a CDS enables to trade the default risk of companies, it is comparable to an insurance contract. The CDS spread thereby represents the price of this protection granted. Thus, institutions with higher CDS spreads but same maturity indicate higher default probability.

Most of current financial stress tests, which can be found in published literature, are using balance sheets of financial institutions, which are subject to legal restrictions and results are therefore limited.
As CDSs are traded credit derivatives, one can consider them to be fair priced and therefore appropriate to measure the risk of financial institutions.
It has been shown by Kou and Varotto, 2005, that CDS spreads can even predict decisions of S&P and Moody's. This fact emphasizes that statistical analyses using CDS spreads have the potential to result in meaningful results considering the default risk of companies. Another study by Cont et al., 2010, used network methods to evaluate the systemic risk of financial institutions. It was mentioned that the inter-similarity of financial institutions can increases the probability of contagion of financial distress. It has been shown in this

study, that balance sheet sizes alone are not capable of indicating systemic importance. In addition, Huang et al., 2009, explained, that only measuring correlation of CDS spreads is a very restricted approach and thus other methods, as explained in this thesis, have to be developed.

It is denoted in Chu et al., 2010, that until now, mainly Gaussian and Student models are used in the financial context, which represents a great limitation.
Even if modeled by multivariate Brownian motion, it is implied that their joined copula is normal (see Cont and Kan, 2011). Therefore, we will apply **R-vines** to the transformed CDS spread data set to fully exploit the potential of the data.

On the other hand, one possible weakness of CDS spreads as suggested by Markose et al., 2009, are the strong self-reflexive properties. Increasing CDS spreads can accelerate the default event as ratings-downgrade follow. This observation represents a drawback of CDS spreads. However as time series model were applied to the data set, these effects were eliminated from the data set.

7.2 Overview of the data set

For our analysis, daily changes of senior CDS spreads were collected from Bloomberg with a maturity of five years for the period from January 4th, 2006 to October 25th, 2011. This data set had to be transformed to the **U-Level** using estimated cumulative distribution functions. Additional time series had to be applied to receive independent random variables. For further information of this procedure see Hendrich, 2012.

The companies included in the investigation are summarized in **Table 11**. For each of the companies, their short name, their financial sector, their country the geographical region and whether the company was quoted as systemically important by Risk and Soundness, 2009, is given in the table. For further details we refer to Hendrich, 2012.

Company name	Short name	Sector	Geographical region	Country	Systemic bank
Citigroup Inc.	Citi	Banking	North America	United States	Yes
Goldman Sachs Inc.	GS	Banking	North America	United States	Yes
JP Morgan Chase & Co. Corp.	JPM	Banking	North America	United States	Yes
Banco Bilbao Vizcaya Argentaria S.A.	BBVA	Banking	Europe	Spain	No
Banco Santander S.A.	BS	Banking	Europe	Spain	Yes
Barclays plc	Barclays	Banking	Europe	United Kingdom	Yes
BNP Paribas S.A.	BNP	Banking	Europe	France	Yes
Deutsche Bank AG	DB	Banking	Europe	Germany	Yes
Intesa Sanpaolo S.p.A.	Intesa	Banking	Europe	Italy	No
Royal Bank of Scotland Ltd.	RBS	Banking	Europe	United Kingdom	Yes
Société Générale S.A.	SG	Banking	Europe	France	Yes
Standard Chartered plc	StanCha	Banking	Europe	United Kingdom	No
UBS AG	UBS	Banking	Europe	Switzerland	Yes
Unicredit S.p.A.	Unicredit	Banking	Europe	Italy	Yes
Bank of China Ltd.	BoC	Banking	Asia-Pacific	China	Yes
Kookmin Bank	Kookmin	Banking	Asia-Pacific	South Korea	No
Sumitomo Mitsui Financial Group K.K.	Sumitomo	Banking	Asia-Pacific	Japan	Yes
Westpac Banking Corp.	Westpac	Banking	Asia-Pacific	Australia	No

Table 11: Overview of the banks selected for further analyses including geographical and sectoral clustering

Company name	Short name	Sector	Geographical region	Country	Systemic bank
ACE Ltd.	ACE	Insurance	North America	Bermuda	/
Allstate Corp.	Allstate	Insurance	North America	United States	/
American International Group Inc.	AIG	Insurance	North America	United States	/
Chubb Corp.	Chubb	Insurance	North America	United States	/
Hartford Financial Services Group Inc.	Hartford	Insurance	North America	United States	/
XL Group plc	XLG	Insurance	North America	Bermuda	/
Aegon N.V.	Aegon	Insurance	Europe	Netherlands	/
Allianz SE	Allianz	Insurance	Europe	Germany	/
Assicurazioni Generali S.p.A.	AssGen	Insurance	Europe	Italy	/
Aviva plc	Aviva	Insurance	Europe	United Kingdom	/
AXA S.A.	AXA	Insurance	Europe	France	/
Hannover Ruck AG	HannRe	Insurance	Europe	Germany	/
Legal & General Group plc	LG	Insurance	Europe	United Kingdom	/
Munich Re AG	MR	Insurance	Europe	Germany	/
Prudential plc	Prudential	Insurance	Europe	United Kingdom	/
SCOR SE	SCOR	Insurance	Europe	France	/
Swiss Re AG	SwissRe	Insurance	Europe	Switzerland	/
Zurich Insurance Group AG	Zurich	Insurance	Europe	Switzerland	/
QBE Insurance Group Ltd.	QBE	Insurance	Asia-Pacific	Australia	/
Tokio Marine Holdings K.K.	TM	Insurance	Asia-Pacific	Japan	/

Table 11 (cont.): Overview of the insurances selected for further analyses including geographical and sectoral clustering

As there are 3 North American banks included in the model, we will in the following chapter compare these institutions to 3 major European banks by asset size as it is stated in Tor and Sarfraz, 2013.

We will analyze the dependencies among different financial institutions in this section. For this we will use the empirical Kendall's τ as given in **Definition 3.6**. The values of $(\hat{\tau}_{i,j})_{i<j}$ are calculated for institutions i and j. The empirical distribution of these is investigated in the following figures.
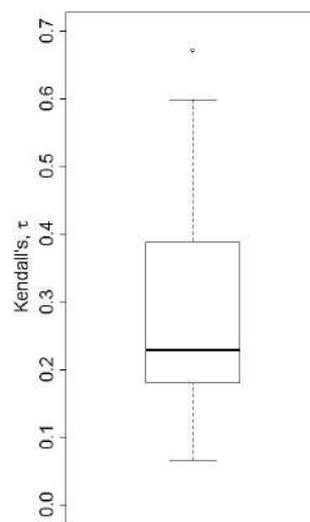
Figure 25: Box plot of the empirical Kendall's τ values

In **Figure 25** we can see that the empirical Kendall's τ values vary in a great range of variables. The median value is at 0.228, however also a lot of higher values of Kendall's τ do exist.

The box plot indicates one outlier, which is the Kendall's τ of the company *Banco Bilbao Vizcaya Argentaria S.A. (BBVA)* and *Banco Santander S.A. (BS)*. These companies are both Spanish, which is one explanation for the strong dependence. One question which arises from this observation is, whether these institutions will have a high Kendall's τ in the simulations again. This question will be answered in **Section 8.12**.

To investigate the distribution of the Kendall's τ values in more detail we plot the histogram in **Figure 26**. We also plot the distribution of the different geographical regions of the pair of companies associated with the Kendall's τ values indicated by different coloring in this histogram.
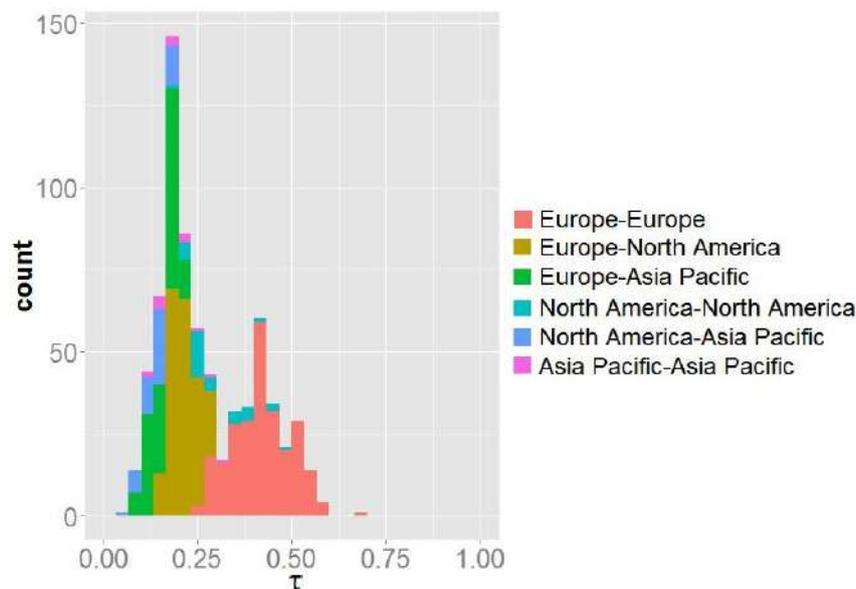
Figure 26: Histogram of the empirical Kendall's τ values with different colors for companies from different geographical regions

In **Figure 26** we can see that the highest values of Kendall's τ can be observed within the European market, followed by the North American market. The lowest values can be found between Asian Pacific institutions on one side and European or North American institutions on the other side.

7.3   Regression model

We want to apply a regression model using the empirical Kendall's τ values to describe the data set. The advantage of using the Kendall's τ values compared to the $\hat{\theta}_{i,j}$ of the R-vine model is that these values can be compared among all families of copulas, to summaries the strength of the dependence in the pair-copula.

With these results we will try to predict the results of the stress test in **Chapter 8**.

To analyze the dependencies in the data set model, we want to apply a regression with a linear model. For this we have to transform the values of $\hat{\tau}_{i,j}$ to the whole real axis. This can be done by the Fisher transformation.

Definition 7.1 (Fisher transformation).
The Fisher transformation is defined as:

$$z(\tau) = \frac{1}{2} \ln \left(\frac{1+\tau}{1-\tau} \right)$$

Thus, for all values of $\hat{\tau}_{i,j}$ the associated z values are:

$$z_{i,j} := z(\hat{\tau}_{i,j}) = \frac{1}{2} \ln \left(\frac{1 + \hat{\tau}_{i,j}}{1 - \hat{\tau}_{i,j}} \right)$$

In our statistical model we include the following characteristics, which we will use as the set of explaining variables:

An institution can fit to the following 3 geographical locations:

1. North America

2. Europe

3. Asia Pacific

The institution is also part of one of the following core businesses:

1. bank

2. insurance

We first examine a full regression model. In a second step we analyze a smaller regression model in order to compare both models.

If we assume, that the Kendall's τ of two companies, dependence on the characteristics of both companies, the coefficients for the following matrix of combinations has to be estimated.

	Bank & North America	Bank & Europe	Bank & Asia Pacific	Insurance & North America	Insurance & Europe	Insurance & Asia Pacific
Bank & North America	β_1	β_2	β_4	β_7	β_{11}	β_{16}
Bank & Europe	β_2	β_3	β_5	β_8	β_{12}	β_{17}
Bank & Asia Pacific	β_4	β_5	β_6	β_9	β_{13}	β_{18}
Insurance & North America	β_7	β_8	β_9	β_{10}	β_{14}	β_{19}
Insurance & Europe	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{20}
Insurance & Asia Pacific	β_{16}	β_{17}	β_{18}	β_{19}	β_{20}	β_{21}

Table 12: First proposed set of regression coefficients for the analysis of empirical Kendall's τ values (full regression model)

The problem of this model is the high number of values that have to be estimated. Thus, we propose a second model with a lower number of variables. Assume the 6 possible properties of a company (Bank & North America, Bank & Europe, Bank & Asia Pacific, Insurance & North America, Insurance & Europe, Insurance & Asia Pacific) are numbered from 1 to 6 with:

$$z_{i,j} = \beta_1 + \beta_2 \cdot (I_{i\ Bank} + I_{j\ Bank}) + \beta_3 \cdot (I_{i\ NorthAmerica} + I_{j\ NorthAmerica})$$
$$+ \beta_4 \cdot (I_{i\ Europe} + I_{j\ Europe}) + \sum_{k=1}^{6} \beta_{k+4} I_{i,\ j\ \text{share combination k}}$$

This leads to the following symmetric coefficient matrix:

	Bank & North America	Bank & Europe	Bank & Asia Pacific	Insurance & North America	Insurance & Europe	Insurance & Asia Pacific
Bank & North America	$\beta_1 + 2\beta_2 + 2\beta_3 + \beta_5$	$\beta_1 + 2\beta_2 + \beta_3 + \beta_4$	$\beta_1 + 2\beta_2 + \beta_3$	$\beta_1 + \beta_2 + 2\beta_3$	$\beta_1 + \beta_2 + \beta_3 + \beta_4$	$\beta_1 + \beta_2 + \beta_3$
Bank & Europe	$\beta_1 + 2\beta_2 + \beta_3 + \beta_4$	$\beta_1 + 2\beta_2 + 2 \cdot \beta_4 + \beta_6$	$\beta_1 + 2\beta_2 + \beta_4$	$\beta_1 + \beta_2 + \beta_3 + \beta_4$	$\beta_1 + \beta_2 + 2\beta_4$	$\beta_1 + \beta_2 + \beta_4$
Bank & Asia Pacific	$\beta_1 + 2\beta_2 + \beta_3$	$\beta_1 + 2\beta_2 + \beta_4$	$\beta_1 + 2\beta_2 + \beta_7$	$\beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_2 + \beta_4$	$\beta_1 + \beta_2$
Insurance & North America	$\beta_1 + \beta_2 + 2\beta_3$	$\beta_1 + \beta_2 + \beta_3 + \beta_4$	$\beta_1 + \beta_2 + \beta_3$	$\beta_1 + 2\beta_3 + \beta_8$	$\beta_1 + \beta_3 + \beta_4$	$\beta_1 + \beta_3$
Insurance & Europe	$\beta_1 + \beta_2 + \beta_3 + \beta_4$	$\beta_1 + \beta_2 + 2\beta_4$	$\beta_1 + \beta_2 + \beta_4$	$\beta_1 + \beta_3 + \beta_4$	$\beta_1 + 2\beta_4 + \beta_9$	$\beta_1 + \beta_4$
Insurance & Asia Pacific	$\beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_2 + \beta_4$	$\beta_1 + \beta_2$	$\beta_1 + \beta_3$	$\beta_1 + \beta_4$	$\beta_1 + \beta_{10}$

Table 13: Second proposed set of regression coefficients for the analysis of empirical Kendall's τ values (smaller regression model)

The resulting values of the smaller regression are summarized in **Table 14**:

parameter of the model	Estimate values	p-value
β_1 (Intercept)	0.0181	0.0328
β_2 (Bank)	0.0169	0.00000415
β_3 (North America)	0.0559	$< 10^{-16}$
β_4 (Europe)	0.1617	$< 10^{-16}$
β_5 (Both companies are: Bank & North America)	0.3336	$< 10^{-16}$
β_6 (Both companies are: Bank & Europe)	0.1362	$< 10^{-16}$
β_7 (Both companies are: Bank & Asia Pacific)	0.148	$2.22 \cdot 10^{-11}$
β_8 (Both companies are: Insurance & North America)	0.2147	$< 10^{-16}$
β_9 (Both companies are: Insurance & Europe)	0.1506	$< 10^{-16}$
β_{10} (Both companies are: Insurance & Asia Pacific)	0.1226	0.0166

Table 14: Parameters of second proposed set of coefficients for the analysis of the data set (smaller regression model)

To compare the models with coefficients as described in **Table 12** and in **Table 13**, we performed a Cox-test for non-nested models (see H.Greene, 2003). This test analysis, if the variables of one of the two compared models, if added to the other model, add further information. This would indicate that the model is to small.

The set of variables of the full model, adds further information to the second model (p-value< 0.000001). The second model can not add further information to the first model, as this model is already the full model. Thus, the full model is significantly better.

We calculate and analyze the full model described in **Table 12** and receive the empirical estimated values as listed in **Table 15**:

	Bank & North America	Bank & Europe	Bank & Asia Pacific	Insurance & North America	Insurance & Europe	Insurance & Asia Pacific
Bank & North America	0.497	0.262	0.135	0.248	0.252	0.117
Bank & Europe	0.262	0.511	0.159	0.195	0.403	0.131
Bank & Asia Pacific	0.135	0.159	0.200	0.155	0.174	0.200
Insurance & North America	0.248	0.195	0.155	0.344	0.201	0.137
Insurance & Europe	0.252	0.403	0.174	0.201	0.492	0.155
Insurance & Asia Pacific	0.117	0.131	0.200	0.137	0.155	0.140

Table 15: Estimated values of the regression coefficients for the analysis of empirical Kendall's τ values with heat-map coloring (full regression)

One can see in **Table 15** that the strongest dependencies exist among European banks (0.511), among North American banks (0.497), among European insurances (0.492) and between European insurances and European banks (0.403). On the other hand, the dependency of Asian companies with companies from other regions is very low. We will compare these results with the results of the simulation study in **Chapter 8**.

From the values in **Table 15** one can see that the dependence within the North American banking market is the highest parameter, which can be seen as one possible reason for the severe impact of the financial crises since 2007.

Based on these empirical Kendall's τ values, we can make predictions of the results of the stress tests in **Chapter 8**.

An obvious prediction would be, that a stress situation induced by European banks would effect other European banks the most. These effects should be the strongest effects, among all dependencies, as within the European banking sector the strongest dependencies exist. Also, one would assume that institutions from the Asia-Pacific region to be less effected by stress tests.

Summarizing the results, the geographical region has a more important role than the financial sector in the stress tests.

8   Applied stress tests based on the MH algorithm

In this chapter we will perform stress tests to empirically fitted R-vine models, based on the data explained in **Chapter 7**. This will be done for 4 different market models, one of which is the full model and 3 are sub-models. For these models we will at first choose the R-vine structure based on maximum likelihood estimation, the Dissmann algorithm and the AIC criterion as explained in **Section 3.6** and then perform different stress tests to this models.

In the stress test we will simulate conditional random variables, derived by the 4 R-vine models, with **Algorithm 4** which was developed in **Chapter 5** and has been validated in **Chapter 6**. This procedure will produce conditional random variables on the **U-Level**. Thus we will work on the **quantile level**, meaning that e.g. 0.99 is the 99% quantile. We will evaluate the effect of the stress situations upon the different institutions, by comparing the distributions of these conditional random variables.

In this chapter we compare 4 different financial models which are given below for which we investigate the effect of a financial crises:

- Model 1: World market

- Model 2: European market

- Model 3: North American market

- Model 4: Banking market

In each model, we conduct up to 3 kinds of stress test. More specifically, we either stress the European systemic banks, the North American systemic banks or the 3 major European systemic banks which are *Barclays pls*, *BNP Paribas S.A* and *Deutsche Bank AG*. For the North American banks and the major European systemic banks we set the values to either the 0.9 quantile, the 0.95 quantile or the 0.99 quantile and simulate the conditional random variable of the remaining variables, given these conditional values. As we have 8 systemic European banks, given in **Table 11**, and since high quantiles result in longer burn-in periods and slower convergence, we limit the quantile levels to 0.9, 0.95 for the set of European systemic banks in order to ensure reliable convergence of the data.
We summarize the conducted stress tests in the **Table 25**.

To estimate the effect of shocks to the financial institutions, we analyzed the distribution of their conditional values by using different stress tests. For this we will compare the median of the random variables. As we work on the **U-Level**, the unconditional

random variables have a median value of 0.5.

As we evaluate CDS spreads, increased values denote an increase in the default probability. Thus, we will evaluate how much the median values of the conditional random variables have been increased compared to this value.

8.1 Copula model selection

We implemented vine copula models with all pair-copulas listed in **Table 1**. These copulas are the pair-copulas which are implemented in the R-package *VineCopula* (see **Table 1** in **Section 3.2**). We compared 4 different vine-structures for all 4 models:

- full R-vine

- R-vine with pairwise independence tests

- full C-vine

- C-vine with pairwise independence tests

The pairwise independence test was performed as described in Genest and Favre, 2007, with a confidence level of $\alpha = 0.05$.

Our aim was to find the model, which is closest to the true model. To compare the different models we therefore performed Vuong tests as described in Vuong, 1989. If the test does not show, that one of the compared models is significantly closer to the actual model than all other models, we choose the model according to the Akaike information criterion (AIC), meaning that lower AIC values indicate a better model.

8.2   World market model

As we can see from **Table 16**, the Vuong test favors the full R-vine over the R-vine model with pairwise independence tests and the full C-vine over the C-vine model with pairwise independence tests. However, the test does not show a significant difference between the full C-vine and the full R-vine model, thus we choose the vine model according to the AIC criterion. From **Table 17**, it can be concluded that we select the **full C-vine** model. We additionally list the values of the log-likelihood (LLH) and BIC in **Table 17**.

Comparison		Vuong test	Vuong test with AIC correction	Vuong test with BIC correction
full R-vine	test statistic	1.3904	1.7432	2.6648
to full C-vine	p-value	0.164	0.081	0.007
R-vine with independence tests	test statistic	-0.2483	1.0602	4.4776
to C-vine with independence tests	p-value	0.803	0.289	0.000
full C-vine	test statistic	*12.742*	4.1398	-18.3262
to C-vine with independence tests	p-value	*0.000*	0.000	0.000
full R-vine	test statistic	*14.393*	4.8077	-20.2255
to R-vine with independence tests	p-value	*0.000*	0.000	0.000

Table 16: Statistics and p-values of the Vuong test comparing different copula models (World market)

	Number of parameters	LLH	AIC	BIC
full R-vine	940	21237.82	-39984.41	-35685.74
R-vine with independence tests	447	20497.54	-40101.08	-37766.27
full C-vine	959	21162.96	*-40407.91*	-35398.77
C-vine with independence tests	519	20511.21	-39984.41	-37273.52

Table 17: Comparison of different copula models (World market)

8.3   European market model

The same procedure as described above was undertaken with the second model. As we can see from **Table 18**, the Vuong test favors the full C-vine over the C-vine model with independence tests, but the Vuong test does not show a significant difference between the full vine models and the vine models with independence tests. Therefore, we choose the vine model according to the AIC criterion. From **Table 19**, it can be concluded that we select the **full R-vine** model.

Comparison		Vuong test	Vuong test with AIC correction	Vuong test with BIC correction
full R-vine	test statistic	1.254	1.117	0.7591
to full C-vine	p-value	0.209	0.263	0.447
R-vine with independence tests	test statistic	-1.0591	-0.2582	1.8333
to C-vine with independence tests	p-value	0.289	0.796	0.066
full C-vine	test statistic	*7.0596*	2.8298	-8.2168
to C-vine with independence tests	p-value	*0.000*	0.004	0.000
full R-vine	test statistic	-1.0591	-0.2582	1.8333
to R-vine with independence tests	p-value	0.289	0.796	0.066

Table 18: Statistics and p-values of the Vuong test comparing different copula models (European market)

	Number of parameters	LLH	AIC	BIC
full R-vine	225	16089.26	*-31384.52*	-29310.88
R-vine with independence tests	145	15831.86	-31139.73	-29771.22
full C-vine	228	16034.35	-31286.69	-29244.38
C-vine with independence tests	160	15880.8	-31163.59	-29601.82

Table 19: Comparison of different copula models (European market)

8.4 North American market model

As we can see from **Table 20**, the Vuong test favors the full R-vine over the R-vine model with independence tests and the full C-vine over the C-vine model with independence tests. However, the test does not show a significant difference between the full C-vine and the full R-vine model, thus we choose the vine model according to the AIC criterion. From **Table 21**, it can be concluded, that we select the **full R-vine** model.

Comparison		Vuong test	Vuong test with AIC correction	Vuong test with BIC correction
full R-vine	test statistic	1.6829	1.8338	2.2281
to full C-vine	p-value	0.092	0.066	0.025
R-vine with independence tests	test statistic	1.8517	1.7769	1.5816
to C-vine with independence tests	p-value	0.064	0.075	0.113
full C-vine	test statistic	*2.2203*	0.9591	-2.3344
to C-vine with independence tests	p-value	*0.026*	0.337	0.019
full R-vine	test statistic	*2.6702*	1.4065	-1.8939
to R-vine with independence tests	p-value	*0.007*	0.159	0.058

Table 20: Statistics and p-values of the Vuong test comparing different copula models (North American market)

	Number of parameters	LLH	AIC	BIC
full R-vine	63	3141.22	*-6156.45*	-5827.385
R-vine with independence tests	55	3124.32	-6138.64	-5851.364
full C-vine	65	3118.92	-6107.85	-5768.342
C-vine with independence tests	54	3099.56	-6091.12	-5809.065

Table 21: Comparison of different copula models (North American market)

8.5 Banking market model

As we can see from **Table 22**, the Vuong test favors the full R-vine over the R-vine model with independence tests and the full C-vine over the C-vine model with independence tests. However, the test does not show a significant difference between the full C-vine and the full R-vine model, thus we choose the vine model according to the AIC criterion. From **Table 23**, it can be concluded, that we select the **full R-vine** model.

Comparison		Vuong test	Vuong test with AIC correction	Vuong test with BIC correction
full R-vine to full C-vine	test statistic	0.0476	0.1558	0.4384
	p-value	0.962	0.876	0.661
R-vine with independence tests to C-vine with independence tests	test statistic	-0.5645	-0.0479	1.301
	p-value	0.572	0.961	0.193
full C-vine to C-vine with independence tests	test statistic	*6.7856*	2.867	-7.367
	p-value	*0.000*	0.004	0.000
full R-vine to R-vine with independence tests	test statistic	*6.9518*	2.8463	-7.8756
	p-value	*0.000*	0.004	0.000

Table 22: Statistics and p-values of the Vuong test comparing different copula models (Banking market)

	Number of parameters	LLH	AIC	BIC
full R-vine	225	9029.78	*-17609.57*	-16434.33
R-vine with independence tests	145	8894.31	-17498.64	-16741.26
full C-vine	228	9028.46	-17600.93	-16410.02
C-vine with independence tests	160	8910.71	-17501.43	-16665.7

Table 23: Comparison of different copula models (Banking market)

We want to compare the R-vine structures estimated for the 4 models to evaluate the expected speed of convergence.

	World market model	European market model	North American market model	Banking market model
Gaussian	41 (5.83%)	4 (1.58%)	1 (2.78%)	9 (5.88%)
Student	198 (28.17%)	126 (49.8%)	23 (63.89%)	55 (35.95%)
Clayton	45 (6.4%)	7 (2.77%)	0 (0%)	11 (7.19%)
Gumbel	27 (3.84%)	7 (2.77%)	1 (2.78%)	9 (5.88%)
Frank	128 (18.21%)	40 (15.81%)	3 (8.33%)	22 (14.38%)
Joe	13 (1.85%)	3 (1.9%)	1 (2.78%)	2 (1.31%)
BB1	4 (0.57%)	0 (0%)	0 (0%)	0 (0%)
BB8	27 (3.84%)	8 (3.16%)	1 (2.78%)	8 (5.23%)
180° rotated Clayton	40 (5.69%)	11 (4.35%)	0 (0%)	8 (5.23%)
180° rotated Gumbel	38 (5.41%)	12 (4.74%)	2 (5.56%)	6 (3.92%)
180° rotated Joe	28 (3.98%)	8 (3.16%)	0 (0%)	6 (3.92%)
180° rotated BB1	1 (0.14%)	0 (0%)	0 (0%)	1 (0.65%)
180° rotated BB7	0 (0%)	0 (0%)	1 (2.78%)	0 (0%)
180° rotated BB8	21 (2.99%)	9 (3.56%)	2 (5.56%)	6 (3.92%)
90° rotated Clayton	24 (3.41%)	3 (1.19%)	0 (0%)	3 (1.96%)
90° rotated Gumbel	10 (1.42%)	0 (0.4%)	0 (0%)	0 (0%)
90° rotated Joe	9 (1.28%)	4 (1.58%)	0 (0%)	2 (1.31%)
90° rotated BB8	3 (0.43%)	1 (0.4%)	0 (0%)	1 (0.65%)
270° rotated Clayton	28 (3.98%)	6 (2.37%)	1 (2.78%)	1 (0.65%)
270° rotated Gumbel	3 (0.43%)	0 (0%)	0 (0%)	1 (0.65%)
270° rotated Joe	13 (1.85%)	3 (1.19%)	0 (0%)	1 (0.65%)
270° rotated BB8	2 (0.84%)	0 (0%)	0 (0%)	1 (0.65%)

Table 24: Distribution of pair-copula families for each vine model

In **Table 24** one can see a summary of the pair-copulas chosen in each model. One can see that overall Student pair-copulas are chosen most often. The second most often chosen pair-copula is the Frank copula.

However, one can see great differences among the models, while in the World market model only 28% of the pair-copulas are Student copulas, in the North American market model 63% of the pair-copulas are Student copulas. Considering the conclusions of **Section 6.5**, we would thus assume the World market to have worse convergence properties. Because of this reason, we will apply longer running MH algorithms for the World market model.

8.6   Overview of simulated stress tests

For the fitted R-vine and C-vine structures we conditioned on different sets of variables, which are summarized in **Table 25**. We simulated realizations of these conditional random variables by applying **Algorithm 4**, the MH algorithm with a proportion of $\beta = 0.1$ univariate Beta distributed proposed updates and a proportion of $1 - \beta = 0.9$ adapted Gaussian proposed updates.

Based on the dimension of the models and a medium size simulation with 20000 samples, we choose different values for the number of the runs of the MH algorithm. To ensure convergence, we choose a minimum burn-in length of 30000 and a simulation size after initial burn-in between 30000 and 280000, depending on the results of the first medium size simulation.

To investigate if further burn-in was required the test by Welch and Heidelberger as described in **Section 4.3**, was applied to the output of the algorithm. According to the result of the test, the burn-in was increased if required.

	Model 1								Model 2				
Test No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Market	World market								European market				
Stress-inducing group	European systemic banks		North American systemic banks			major European banks			European systemic banks		major European banks		
Stress level (moderate/ severe/ extreme)	mode-rate (0.9)	severe (0.95)	mode-rate (0.9)	severe (0.95)	ex-treme (0.99)	mode-rate (0.9)	severe (0.95)	ex-treme (0.99)	mode-rate (0.9)	severe (0.95)	mode-rate (0.9)	severe (0.95)	ex-treme (0.99)
Chosen vine model	full C-vine								full R-vine				
Burn-in	55000	80000	70000	110000	130000	30000	55000	105000	30000	30000	30000	35000	40000
Simulation size (after burn-in)	225000	200000	160000	120000	150000	260000	225000	175000	40000	40000	40000	40000	40000
Effective sample size $\widehat{ESS}$ (minimum,maximum)	(431, 664)	(478, 542)	(226, 411)	(171, 264)	(176, 286)	(425, 703)	(282, 537)	(94, 525)	(463, 624)	(382, 602)	(347, 575)	(229, 386)	(198, 501)

Table 25: Overview of stress tests performed with the empirical data set

	Model 3			Model 4				
Test No.	14	15	16	17	18	19	20	21
Market	North American market			Banking market				
Stress-inducing group	North American systemic banks			European systemic banks		North American systemic banks		
Stress level (moderate/ severe/ extreme)	moderate (0.9)	severe (0.95)	extreme (0.99)	moderate (0.9)	severe (0.95)	moderate (0.9)	severe (0.95)	extreme (0.99)
Chosen vine model	full R-vine			full R-vine				
Burn-in	30000	30000	30000	33000	30000	30000	33000	36000
Simulation size	30000	30000	30000	30000	40000	40000	27000	24000
Effective sample size $\widehat{ESS}$ (minimum,maximum)	(1211, 1307)	(1115, 1354)	(1299, 1376)	(305, 423)	(911, 1100)	(864, 1150)	(275, 431)	(279, 383)

Table 25 (cont.): Overview of stress tests performed with the empirical data set

In the following **Sections 8.7- 8.10**, we will show the results of different stress tests listed in **Table 25**. In **Section 8.11** and **8.12**, we then discuss the possible interpretation of these stress tests.

The first figure in each of the following sections shows the performance of the respective market in stress situations with different stress levels. The companies printed in ***italic and bold*** letters represent the institutions inducing the current stress situation. We therefore use the **U-Level** transformed CDS spreads, which can be seen as a measure for the default probability as described in **Section 7.1**. As higher values of the **U-Level** variables would also result in higher values in the original non-transformed variables, these would indicate a higher default probability in the specific stress situation. The dotted colored lines depict the overall medians observed in the specific scenarios. For each institution, it is thus possible to evaluate whether the reaction is stronger than the overall median.

The second figure in each section shows the differences in the performance of the respective markets in increasingly extreme stress situations. The represented values indicate by how much the effect of induced stress upon the institutions is influenced by the severity of the crisis, from a moderate over a severe to an extreme financial crisis.

For improved visualization, we assigned a separate color to each region: Red for North America, blue for Europe and green for Asia Pacific. Additionally, bright colors indicate banks, whereas light colors indicate insurances. We therefore have 6 combinations, each of which is represented by a characteristic color:

- North American banks

- North American insurances

- European banks

- European insurances

- Asian Pacific banks

- Asian Pacific insurances

These colorings will make it easer to compare the different performances in the stress situations.

8.7   Performance of the world market model
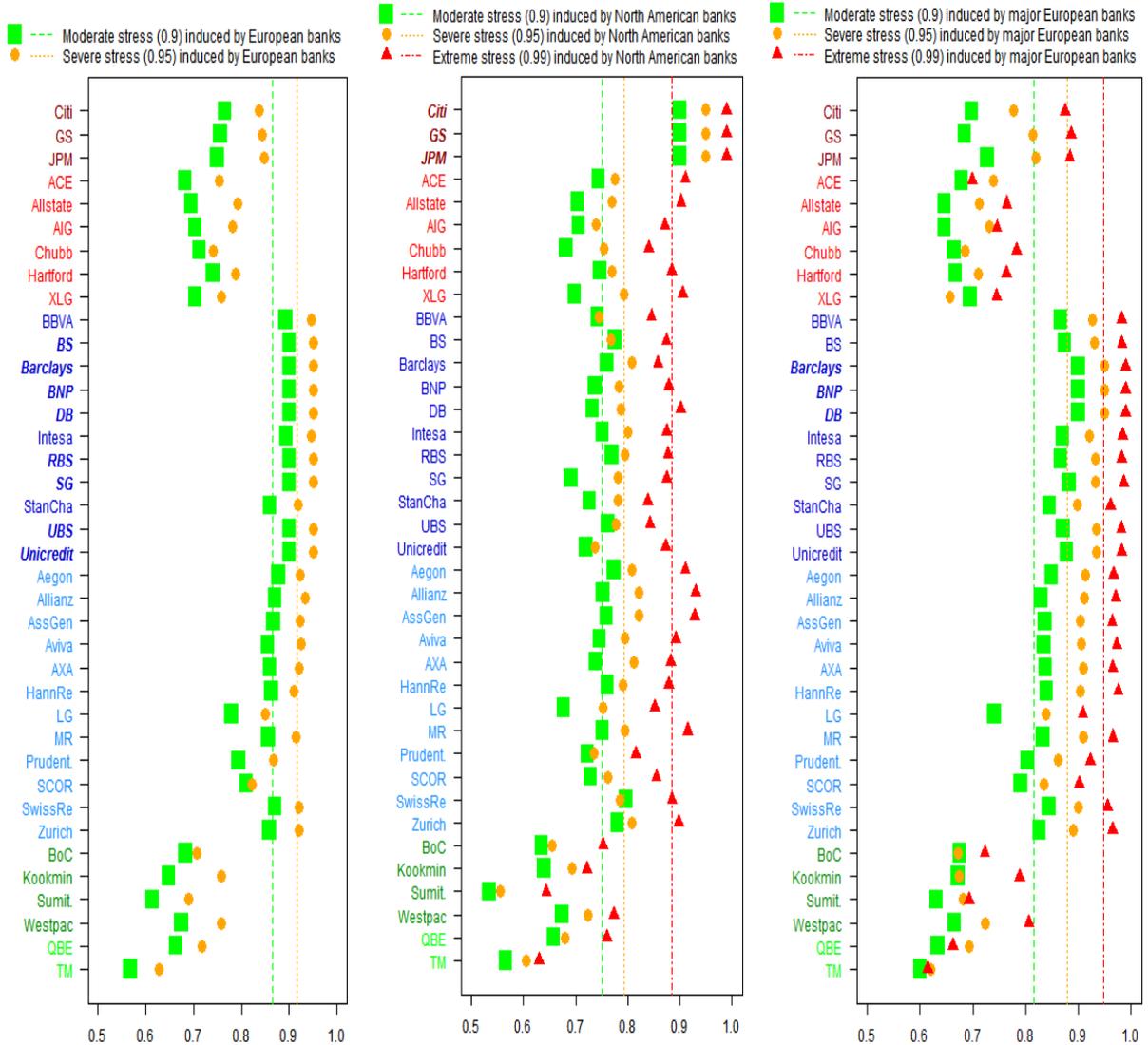
(a) Performance of the world market in stress situations induced by European systemic banks[1]

(b) Performance of the world market in stress situations induced by North American systemic banks[1]

(c) Performance of the world market in stress situations induced by major European banks

Figure 27: Performance of the world market in stress situations - Summary of median values
[1] Symbols depict the different institutions, whereas the dotted horizontal line represents the overall median value
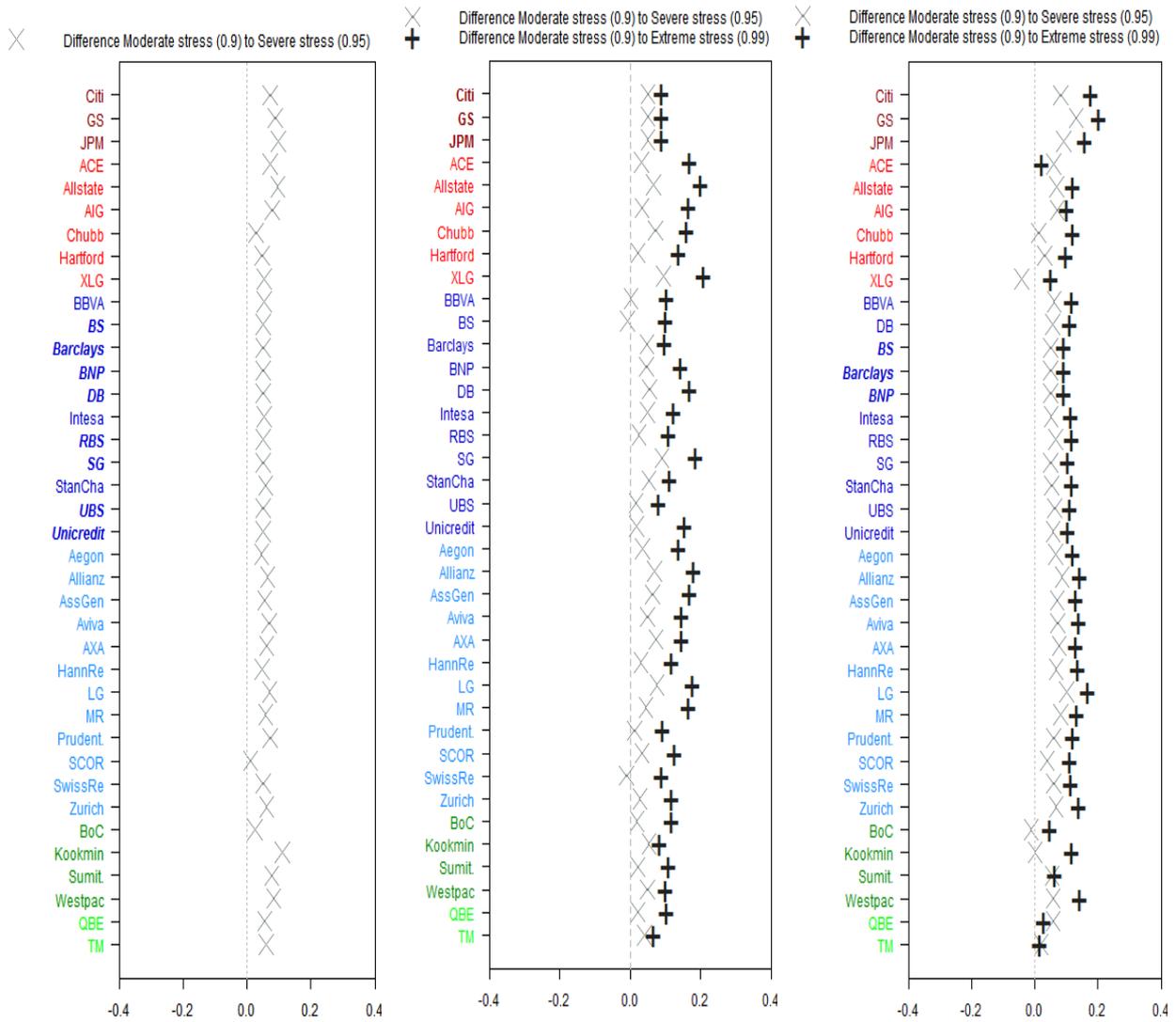
(a) Differences in the performance of the world market in stress situations induced by European systemic banks

(b) Differences in the performance of the world market in stress situations induced by North American systemic banks

(c) Differences in the performance of the world market in stress situations induced by major European banks

Figure 28: Differences in the performance of the world market in stress situations - Summary of differences in median values

8.8   Performance of the European market model
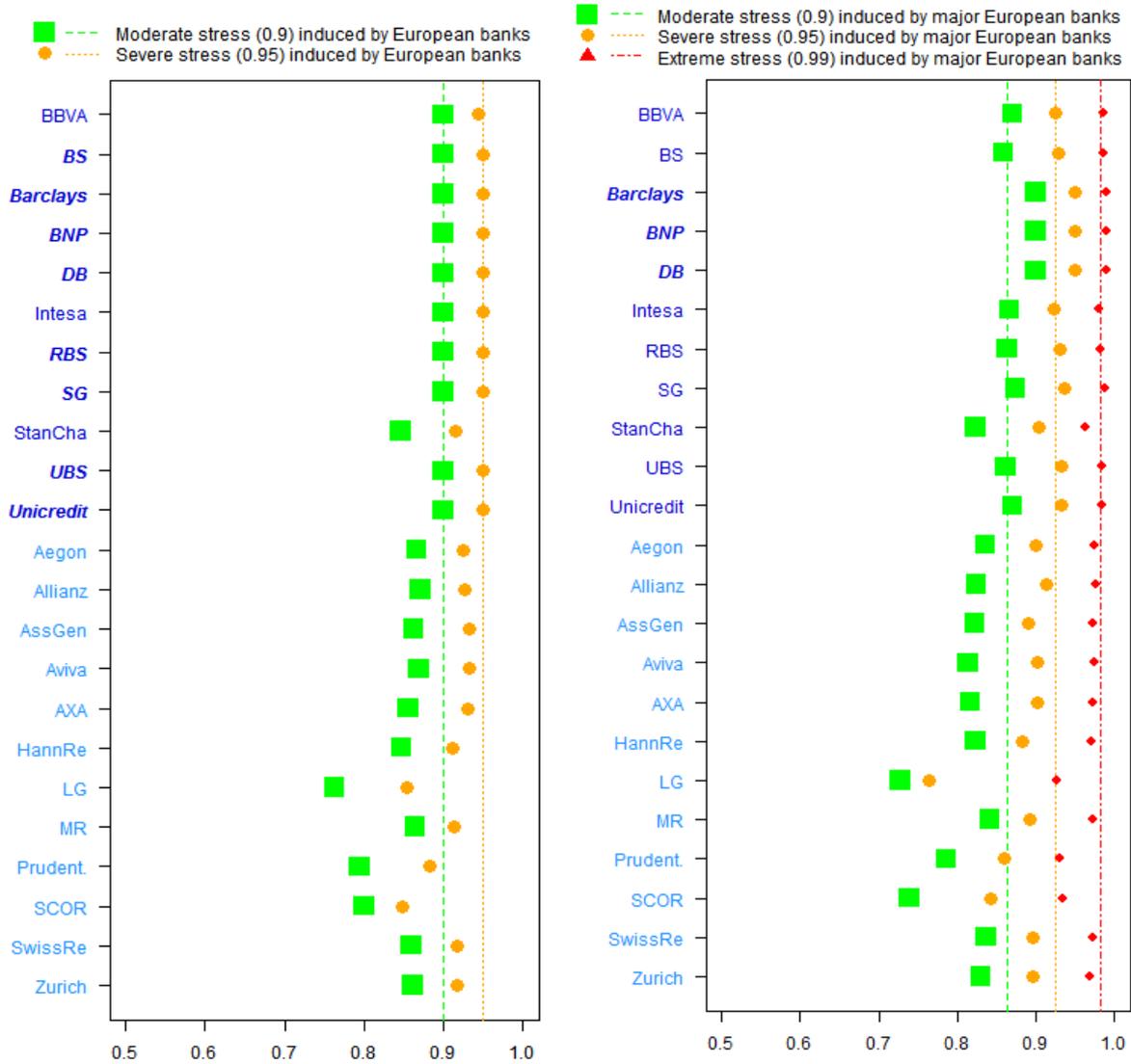
(a) Performance of the European market in stress situations induced by European systemic banks[1]

(b) Performance of the European market in stress situations induced by major European banks[1]

Figure 29: Performance of the European market in stress situations - Summary of median values

[1] Symbols depict the different institutions, whereas the dotted horizontal line represents the overall median value
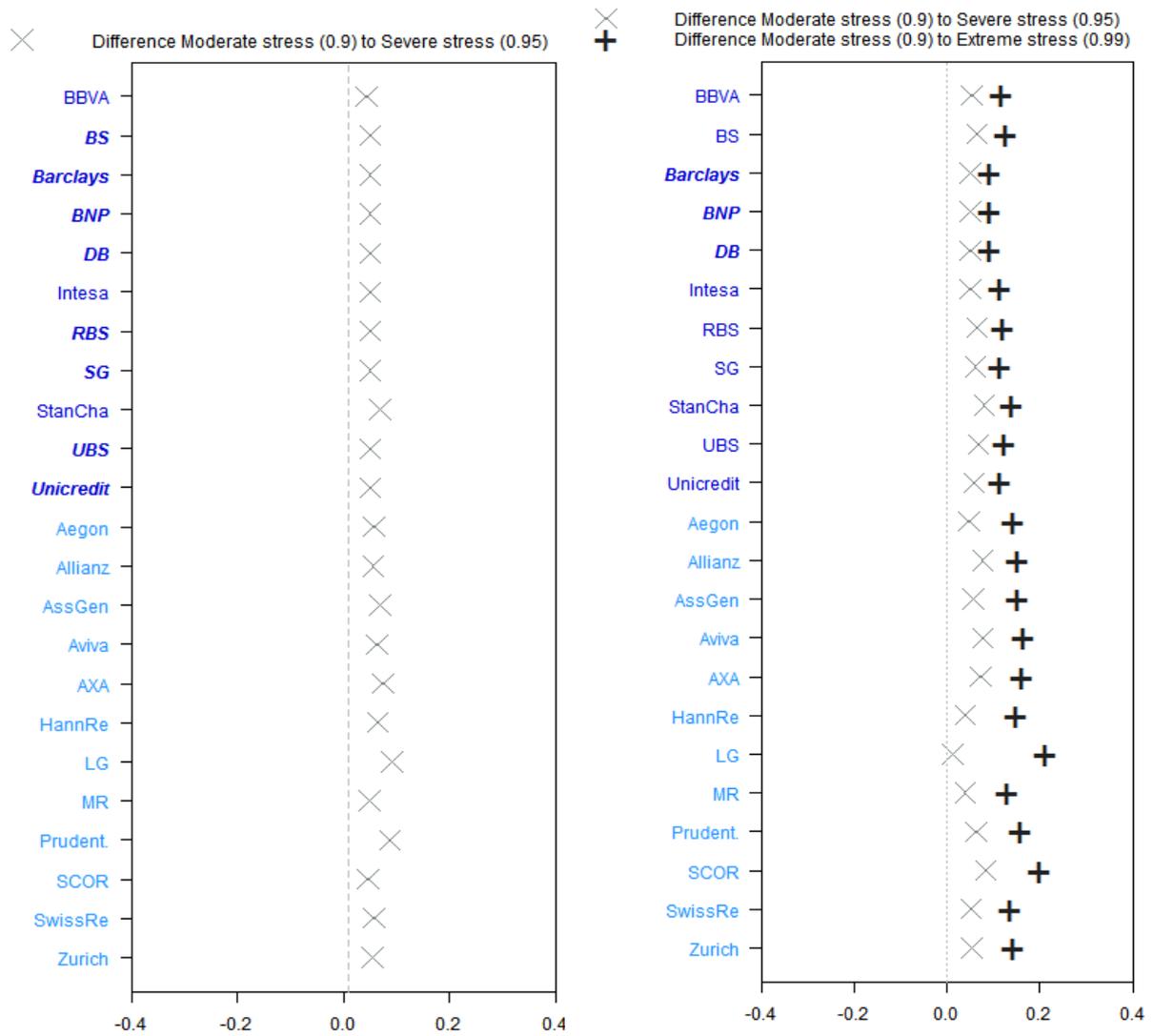
(a) Differences in the performance of the European market in stress situations induced by European systemic banks

(b) Differences in the performance of the European market in stress situations induced by major European banks

Figure 30: Differences in the performance of the European market in stress situations - Summary of differences in median values

8.9   Performance of the North American market model
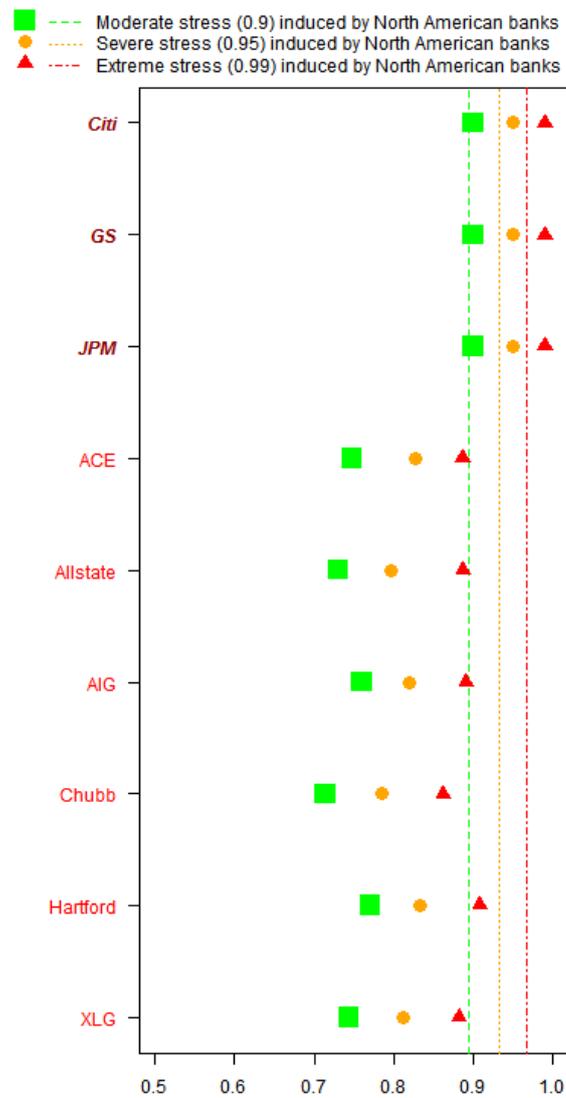
Figure 31: Performance of the North American market in stress situations induced by North American systemic banks - Summary of median values[1]

[1] Symbols depict the different institutions, whereas the dotted horizontal line represents the overall median value
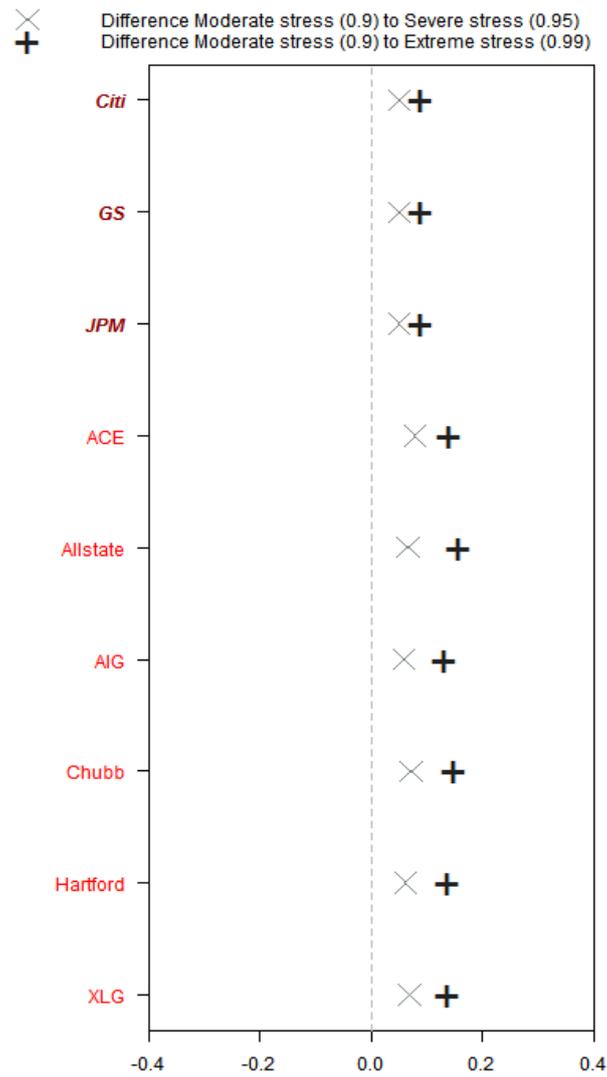
Figure 32: Performance of the North American market in stress situations induced by North American systemic banks - Summary of median values

8.10   Performance of the banking market model
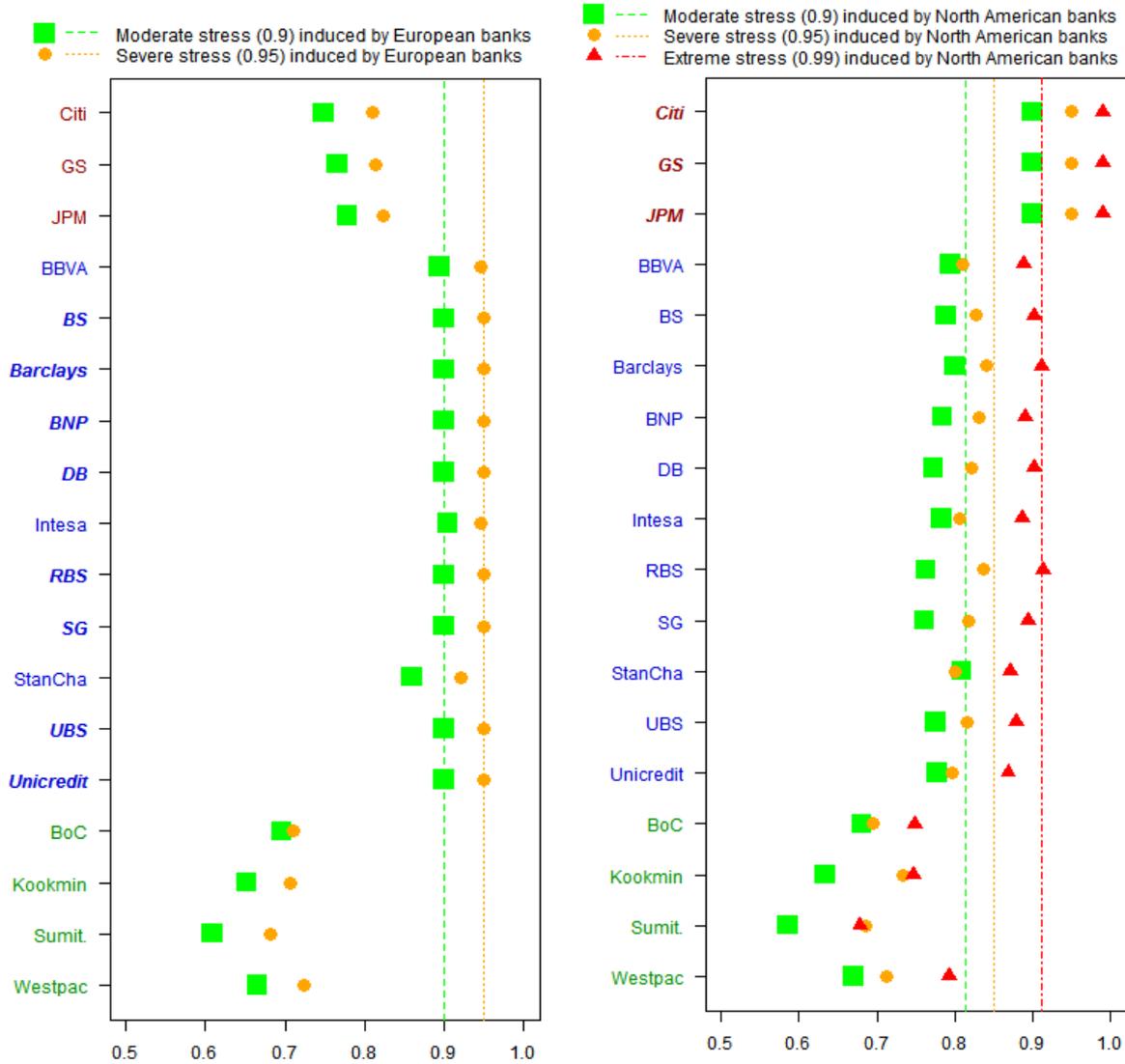
(a) Performance of the banking market in stress situations induced by European systemic banks[1]

(b) Performance of the banking market in stress situations induced by North American systemic banks[1]

Figure 33: Performance of the banking market in stress situations - Summary of median values
[1] Symbols depict the different institutions, whereas the dotted horizontal line represents the overall median value
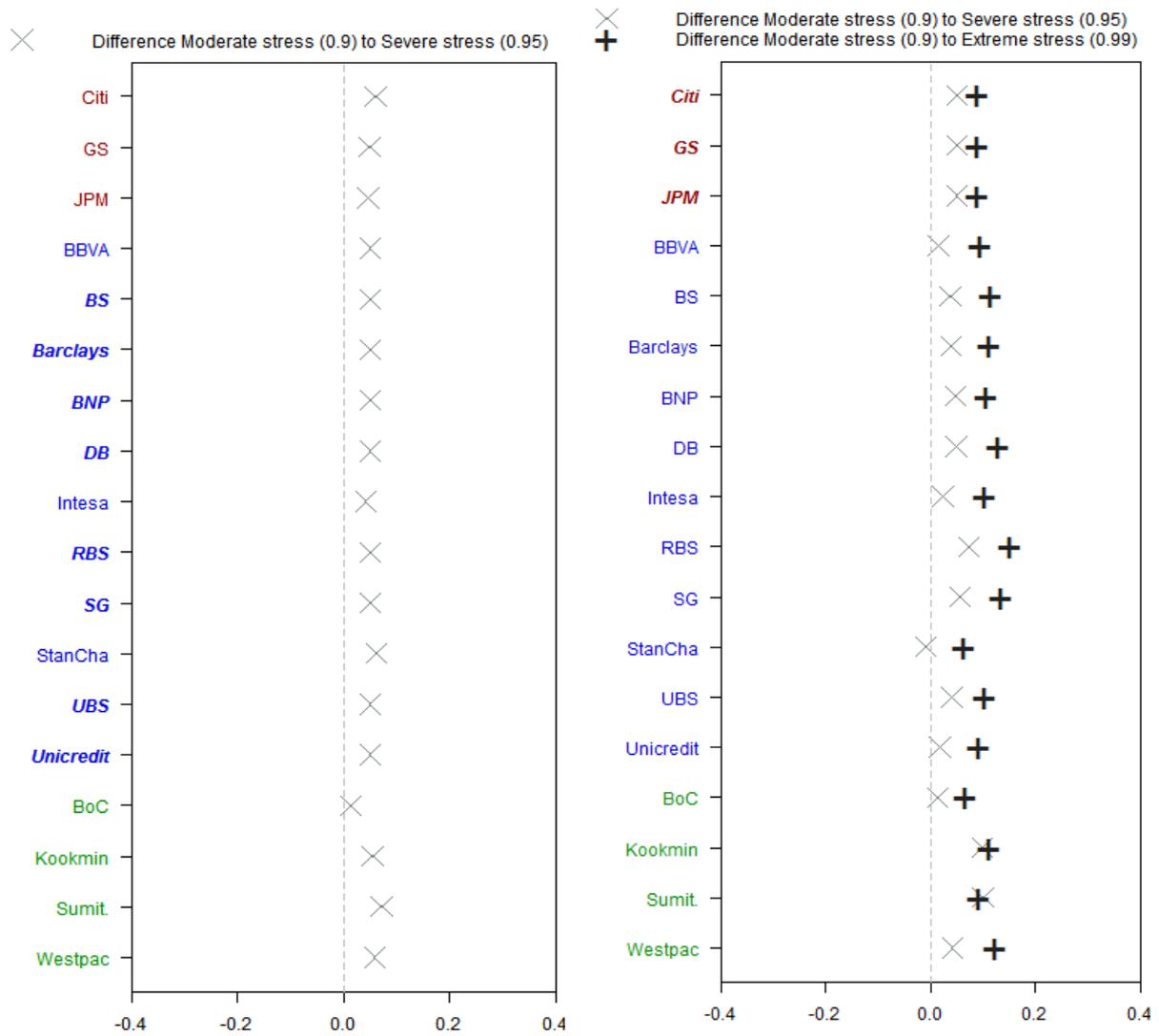
(a) Differences in the performance of the banking market in stress situations induced by European systemic banks

(b) Differences in the performance of the banking market in stress situations induced by North American systemic banks

Figure 34: Differences in the performance of the banking market in stress situations - Summary of differences in median values

8.11   Interpretation of Z-Level plots

So far, we have only visualized the median values of the stress tests. However, we also want to give a comprehensive overview of the whole situation on financial markets, given a stress situation. For this we developed a graphic method of providing a great range of information.

The new form of **Z-Level** plot can visualize the whole joint distribution of the markets under stress. One small example for the **Z-Level** plot is shown in **Figure 35**.
On the **upper triangular**, the empirical Kendall's τ values are plotted and colored to indicate the strength of dependence of the companies in the stress situation ranging from dark blue for low dependencies, over gray for medium dependencies to dark red for strong dependencies.
On the **diagonal**, the empirical histograms of the conditional random variables are plotted. Based on the fact, that the variables are transformed to the **Z-Level** the unconditional random variables are $\mathcal{N}_{0,1}$ distributed, as the **U-Level** variables are uniformly distributed on $[0, 1]$. The unconditional random variable transformed to the **Z-Level** has median zero, marked as a gray dotted line in the figures. In the stress situations the values will usually be shifted to the right, indicating higher default probabilities. The histogram can moreover visualize the whole distribution of the conditional random variable, which provides further information besides the median values. The different colors of the histograms indicate North American banks, North American insurances, European banks, European insurances, Asian Pacific banks and Asian Pacific insurances (see **Appendix**).
On the **lower triangular**, there are the contour plots of the bivariate distributions. As the empirical Kendall's τ values denote the strength of dependence only by a single number, we use these images to give a better indication of the specific type of co-movement of the values. The derived shape plots can therefore be compared to the plots given in **Section 3.7** in order to evaluate the type of dependence.

The **Z-Level** plots for all 21 stress tests are given in **Appendix A**.
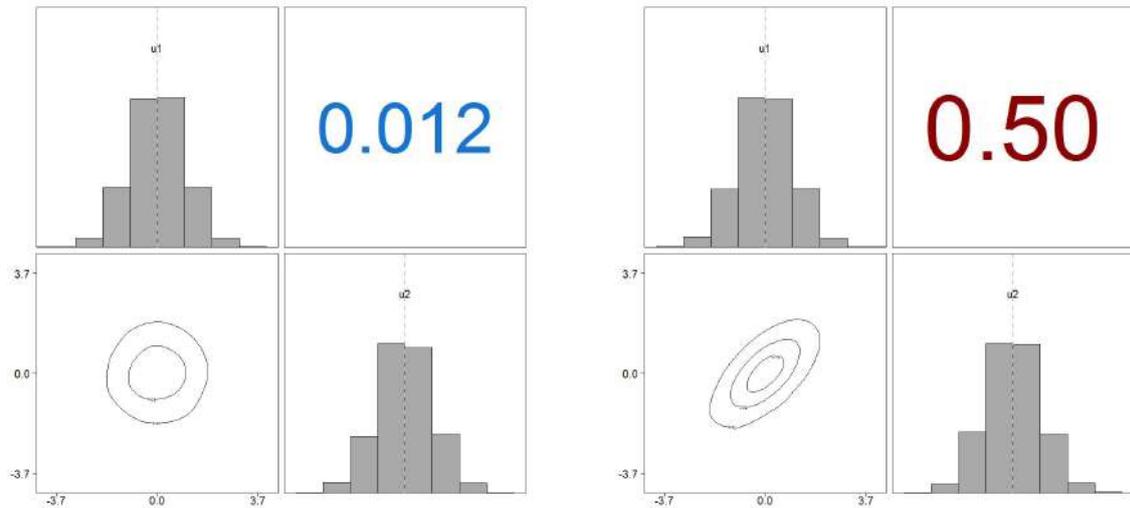
(a1) Independence copula with uniform margins

(a2) Gaussian copula with $\tau = 0.5$ and uniform margins
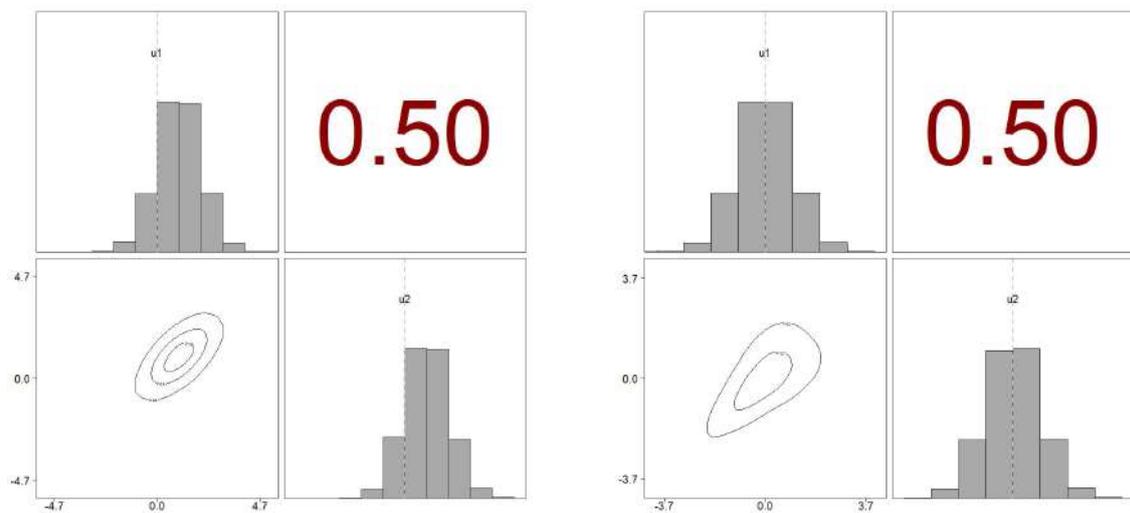
(b1) Gaussian copula with $\tau = 0.5$ and margins shifted to the right

(b2) Clayton copula with $\tau = 0.5$ and uniform margins

Figure 35: Example illustrating Z-Level plot and Kendall's τ values

As mentioned above **Figure 35** is an example for a **Z-Level** plot, generated by 5000 random samples. In (a1), (a2) and (b2) the random variables both have uniform margins, as it can be seen in the histograms. Thus, these random variables have the same uniform distribution as the unconditional random variables. Furthermore, by considering the histograms, it is shown in (b1) that the random variables are shifted to the right, which is usually the case in stress situations. In (a1) and (a2) one can compare random variables with the same marginal distribution, but with different Kendall's τ values. It is also recognizable how the shape of the contour plot changes. When comparing (a2) and (b2) one can see the marginal distributions and the Kendall's τ values are the same, but the

contour plots look differently, as the kind of co-movement is different. This can lead to different interpretations of the dependence.

8.12   Interpretation of the results of the simulated stress situations

In this section we will at first analyze the results of the median values for all 4 markets. We will hereby analyze which institutions are effected the most in the conducted stress tests. The stress test is therefore conducted with strictly monotonously transformed daily CDS spread changes, thus high increases represent an increase in the default probability. This in return means, that institutions with higher values are stronger effected than institutions with lower values.

In the second part of the interpretation we will focus on the **Z-Level** plots, introduced in **Section 8.11**.

First, we will interpret the median values: In all market models, it can be observed that with increasing stress level the overall median value also increases. This observation is reasonable from an economical point of view, as it implies that in more extreme stress situations the average default probability increases. This observation therefore supports the validation of the stress tests applied.

A further validating observation becomes apparent when comparing the results of the world market and the European market for the stress induced by either a larger group of 8 European banks or 3 major European banks.

In the case of the larger group of 8 systemic European banks, the overall median is higher than in the stress test situation induced by the smaller group. This can also be seen as economically plausible, as a bigger proportion of the market is under stress, and thus the stress situation is more extreme. However, in **Figure 29**, it can be seen that the overall median is only moderately increased. This shows that *Barclays*, *BNP* and *DB* representing the major 3 institutions in Europe, whereas other institutions do not contribute much to the increase in stress.

It is therefore initially important to investigate the results of the European banks across all stress tests in the following paragraph. *Standard Chartered plc (Stancha)* is less effected by stress induced by European banks than the other European banks. This indicates that this company is less interconnected to the European banking market.

It can be noted that the European banks are far more effected by moderate and severe stress situations induced by North American banks than the North American insurances.

More precisely, the overall medians of the European banks in moderate and severe stress are 0.743 and 0.784, while the overall medians of the North American insurances are 0.714 and 0.768. This observation changes in the extreme stress test situation where the overall median of the North American insurances is higher (0.887) than the median value of the European banks (0.867). This observation hence demonstrates that the geographical region is not always the most influential factor.

Last but not least, the European insurances which are effected the least by stress situations induced by one of the three stress inducing groups in the world market and the European market are *Legal and General Group pls (LG)*, *Prudential plc (Prudent)* and *Score SE (SCOR)*. These institutions are therefore relatively less effected by financial crises.

Regarding the world market and the North American market, the North American insurance which is least influenced by stress induced by North American banks, is *Chubb Corp. (Chubb)*.

In view of the Asian Pacific market, the institutions here are effected little by stress induced by either European or North American institutions. *Tokio Marine Holdings K. K. (TM)* is the least effected institution in most of the stress tests. The least effected bank in most of the stress tests is *Sumitomo Mitsui Financial Group K. K. (Sumit.)*. Both institutions, *TM* and *Sumit.* are Japanese institutions. This indicates that the Japanese financial market would be least effected by crises induced by North American or European banks, considering all financial markets included in this study.

Thus the probability that an European or North American crises will cross-over to the Asian Pacific financial sector can be regarded as rather small.

In the second part we will now focus on the Z-scale plots, introduced in **Section 8.11** and shown in detail in the **Appendix**.

As a full analysis of all scenarios would take too long, some key results will be summarized.

It can be observed that institutions from the same geographical region form a cluster in the heat-map, as institutions within these groups have high Kendall's τ values.

In stress tests 1 and 2 (**Figure A.1, A.2**) and also in stress tests 9 and 10 (**Figure A.9, A.10**) we see that the European banks, especially *BBVA* and *Intesa*, are effected more than the European insurances are, as their distribution is more increased than the distribution of the insurances. However, surprisingly, the Kendall's τ values within the

group of banks and between the group of banks and the group of insurances is very low, whereas the Kendall's τ values in the group of insurances are much higher. This has therefore an interesting interpretation, as: *BBVA* and *Intesa* are effected the most by an European induced crises, however they do not effect other institutions.

In the stress tests 1-2, and 6-8 (**Figure A.1, A.2, A.6, A.7, A.8**) one can see that the North American companies are not as strongly effected as the European institutions, but the dependencies within the North American institutions are still high.

A still open question is the comparison of the stress induced by European and North American banks. Just comparing the overall median values in the world market model in **Figure 27** is not appropriate, as the high number of European institutions, which are more effected by stress induced by European banks makes the comparison biased.
We therefore propose two procedures to compare the effects:
At first, the effect of both groups upon the Asian-Pacific institutions is measured by comparing the overall median of the Asian Pacific institutions in these stress situations.
As a second component the comparison of the effect of North American banks upon the major European banks and vice versa was chosen. Thus, in **Figure 36** the effects of the analysis can be compared more easily.
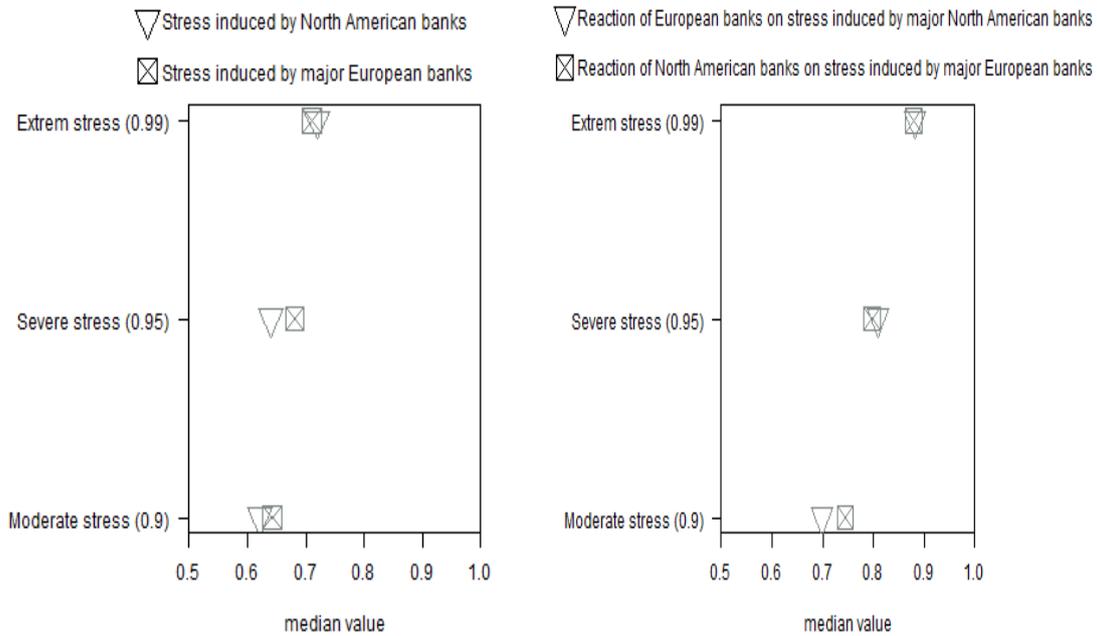
(a) Comparison of the median values of: -The effect of North American banks upon Asian Pacific institutions -The effect of major European banks upon Asian Pacific institutions

(b) Comparison of the median values of: -The effect of North American banks upon major European banks -The effect of major European banks upon North American banks

Figure 36: Comparison of the effect of stress induced by North American banks and major European banks

The comparison of the effect of stress induced by major European and major North American institutions upon Asian institutions is shown in **Figure 36 (a)**. In **Figure 36 (b)** we compare how the major European and major North American institutions are effecting each other.

It can be seen that in moderate stress situations, the effect of European institutions is stronger than the effect of North American institutions. However, this effect is no longer present in more severe stress situations. Thus, only in moderate stress situations the effect of European institutions is more important than the effect of North American institutions.

9 Similarity-based pair-copula constructions

Influenced by the work of Gräler and Pebesma, 2011, where an approach was applied to geographical distances between observation points, we developed a new approach which we describe in the following pages.

This method will strongly differ from other methods, as we will derive procedures, which can be used for regression models on C-vines.

In this respect, we will introduce the concept of **similarity** of financial institutions, which corresponds to the role of geographical distances as reported by Gräler and Pebesma, 2011, in a way that high values of **similarity** can be interpreted as small distances between points and thus as a high dependence.

9.1 Similarity-based pair-copulas

We want to define a similarity measure which evaluates the strength of the association of two vectors with uniform margins. This measure should indicate the strength of connectedness of two institutions, which we want to evaluate in this thesis.

The general definition of a similarity measure is as follows:

Definition 9.1 (Similarity measure)**.**
*A **similarity measure** $s : [0,1]^n \times [0,1]^n \times \{1,...,d\} \times \{1,...,d\} \to \mathbb{R}_+$
assigns two random vectors $\mathbf{u}_i, \mathbf{u}_j$ with values in $[0,1]^n$ associated with two institutions (i,j), with a positive real similarity value. This similarity measure s defines a **similarity matrix** for a set of d vectors $(\mathbf{u}_i)_{i=1,...,d}$ defined as:*

$$\mathcal{S} = (s_{i,j})_{i,j=1,...,d} = (s(\mathbf{u}_i, \mathbf{u}_j, (i,j)))_{i,j=1,...,d}$$

Higher values of $s(\mathbf{u}_i, \mathbf{u}_j, (i,j))$ represent more similarity between $\mathbf{u}_i$ and $\mathbf{u}_j$, while lower values indicate less similarity. One example for such a similarity measure, where s is a function of $\mathbf{u}_i$ and $\mathbf{u}_j$, is the **Pearson correlation**. In this case, the matrix $\mathcal{S}$ becomes a correlation matrix.

Another example for a similarity measure containing the indices i and j is the **similarity of institutions**, which will be used in **Section 9.4**:

Definition 9.2 (Similarity of institutions)**.**
The similarity between the institutions i and j, with measurements $\mathbf{u}_i \in [0,1]^n$ and $\mathbf{u}_j \in [0,1]^n$ is given by:

$$s(\mathbf{u}_i, \mathbf{u}_j, i, j) := |\hat{\tau}(\mathbf{u}_i, \mathbf{u}_j)| + geographical\_closeness(i,j)$$

The values for the closeness among the geographical regions, *geographical_closeness(i,j)*, can be chosen arbitrarily, but should be symmetric in i and j.
Higher values for the geographical closeness indicate a high inter-similarity of these geographical regions, which can be interpreted as small geographical distances between these institutions.
For the analysis, the following values listed in **Table 27** were chosen:

	North America	Europe	Asia Pacific
North America	1	$\dfrac{0.002}{1.002}$	$\dfrac{0.002}{1.002}$
Europe	$\dfrac{0.002}{1.002}$	1	$\dfrac{0.001}{1.001}$
Asia Pacific	$\dfrac{0.002}{1.002}$	$\dfrac{0.001}{1.001}$	1

Table 26: Chosen values for *geographical_closeness*

The advantage of including (i,j) in the function s is that further properties of the financial institutions can be included, as for example geographical regions.

In this example the following values would be derived to calculate the similarity measure: As an example, the similarity between the *Deutsche Bank (DB)* and 4 different institutions including the *Citigroup Inc (Citi), Standard Chartered plc (StanCha), Legal and General Group plc (LG)* and *QBE Insurance Group Ltd. (QBE)* was assessed. We therefore receive the following Kendall's τ values:

$$\tau_{DB,Citi} = 0.24173$$
$$\tau_{DB,StanCha} = 0.38681$$
$$\tau_{DB,LG} = 0.29171$$
$$\tau_{DB,QBE} = 0.15498$$

The corresponding similarity values, which are received by adding the *geographical_closeness*, are:

$$s_{DB,Citi} = 0.24173 + 0.00199 = 0.24372$$
$$s_{DB,StanCha} = 0.38681 + 1 = 1.38681$$
$$s_{DB,LG} = 0.29171 + 1 = 1.29171$$
$$s_{DB,QBE} = 0.15498 + 0.00099 = 0.15597$$

It can be seen, how the *geographical_closeness* increases some values a lot. These effects will be further investigated in **Section 9.4**.

The function s is crucial for the definition of the newly developed approach of **similarity-based pair-copulas**, which we will present in this chapter. The construction of our R-vine structure is thereby based on the triplet $RV = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \boldsymbol{\theta}(\mathcal{B}(\mathcal{V})))$. In this chapter we will introduce a new approach to construct this triplet RV based on the values of the similarity measure s.

The crucial idea in our method is to construct pair-copulas out of **linear mixtures of two different pair-copulas**. The proportion and the families of these combinations will depend on the values of s.

For this, we want to partition the set of the $\binom{d}{2}$ values of $(s_{i,j})_{i>j}$ into $B+1$ sets, which correspond to B boundaries for these sets. In other words, we define B quantiles, $(q_{\alpha_i})_{i=1,\ldots,B}$ with the quantile levels $(\alpha_i)_{i=1,\ldots,B}$, and divide the set into $B+1$ sub-intervals, that are the intervals $[0, q_{\alpha_1}]$, $(q_{\alpha_1}, q_{\alpha_2}], \ldots, (q_{\alpha_{B-1}}, q_{\alpha_B}], (q_{\alpha_B}, \infty)$.

We want to choose the quantile levels such that the difference between neighboring values is constant. By denoting the difference of the level of two neighboring quantiles as Δ and the following calculation, we get the values of the quantile levels:

$$\alpha_{i+1} - \alpha_i = \Delta$$
$$\alpha_1 = \Delta/2$$
$$1 - \alpha_B = \Delta/2$$

We set the lowest quantile level as $\dfrac{1}{2B}$ and the highest quantile level as $1 - \dfrac{1}{2B}$.

Written in form of a telescoping sum, we thus get:

$$\Rightarrow 1 = \alpha_1 + \sum_{i=1}^{B-1}(\alpha_{i+1} - \alpha_i) + 1 - \alpha_B$$

$$= \Delta/2 + (B-1)\Delta + \Delta/2$$

$$= B \cdot \Delta$$

$$\Leftrightarrow \Delta = \frac{1}{B}$$

As values of α_i, we get:

$$\alpha_i = \frac{1}{2B} + \frac{i-1}{B} \tag{9.1}$$

As we want to work with linear combinations of copulas we need to define these combinations:

Definition 9.3 (Linear mixtures of two copulas)**.**
Given two pair-copulas C_1 and C_2 we can now define a **linear mixture of two copulas** *as:*

$$C_s(u,v) = \lambda(s) \cdot C_1(u,v) + (1 - \lambda(s)) \cdot C_2(u,v,) \tag{9.2}$$

This combination is still a copula, as linear combinations of cumulative distribution functions stay cumulative distribution functions.

The influence of s is illustrated in **Figure 37**. One can see how the linear combination changes depending on the value of s.
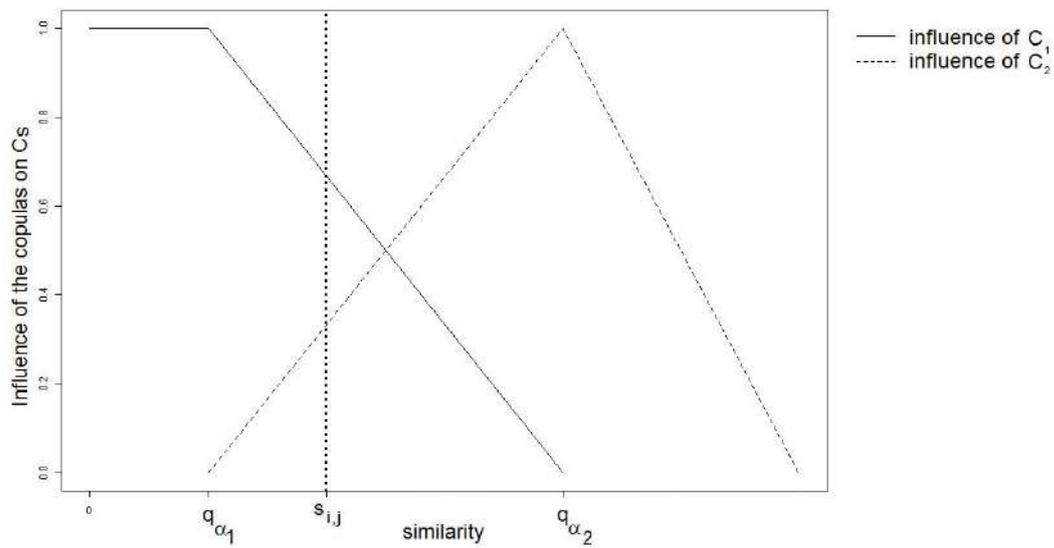
Figure 37: Illustration of the contribution of pair-copulas to the linear combinations based on similarity values (see equation (9.2))

In **Figure 38** and **Figure 39** the contour plots for a range of different copulas are shown. One can see that a great range of new contours are possible. This enables this method to fit certain copulas even better than the traditional approach does.
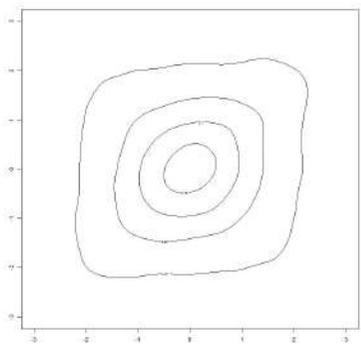
(a1) C_1=Frank, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.1$ (a2) C_1=Frank, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.5$ (a3) C_1=Frank, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.9$
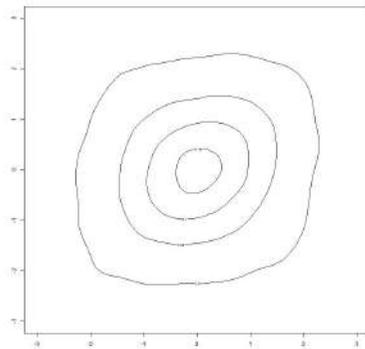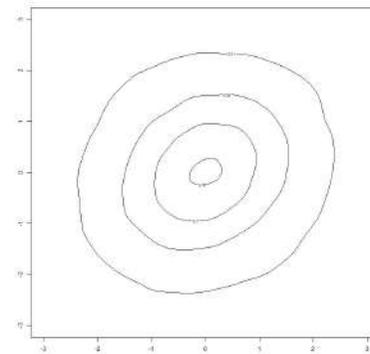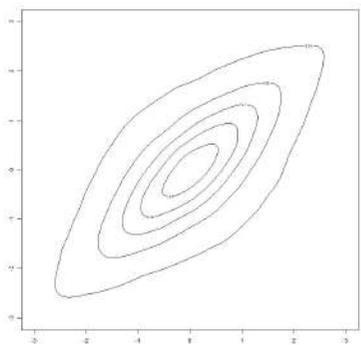
(b1) C_1=Frank, $C_2 = t_3$, $\tau = 0.5$, $\lambda = 0.1$ (b2) C_1=Frank, $C_2 = t_3$, $\tau = 0.5$, $\lambda = 0.5$ (b3) C_1=Frank, $C_2 = t_3$, $\tau = 0.5$, $\lambda = 0.9$
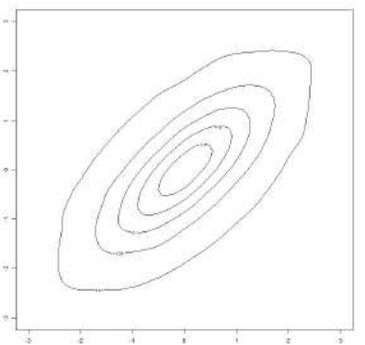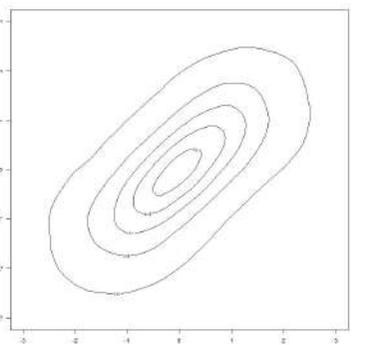
Figure 38: Summary of Z-Level contour plots for $\tau = 0.1$ (first row) and $\tau = 0.5$ (second row) (see equation (9.2))
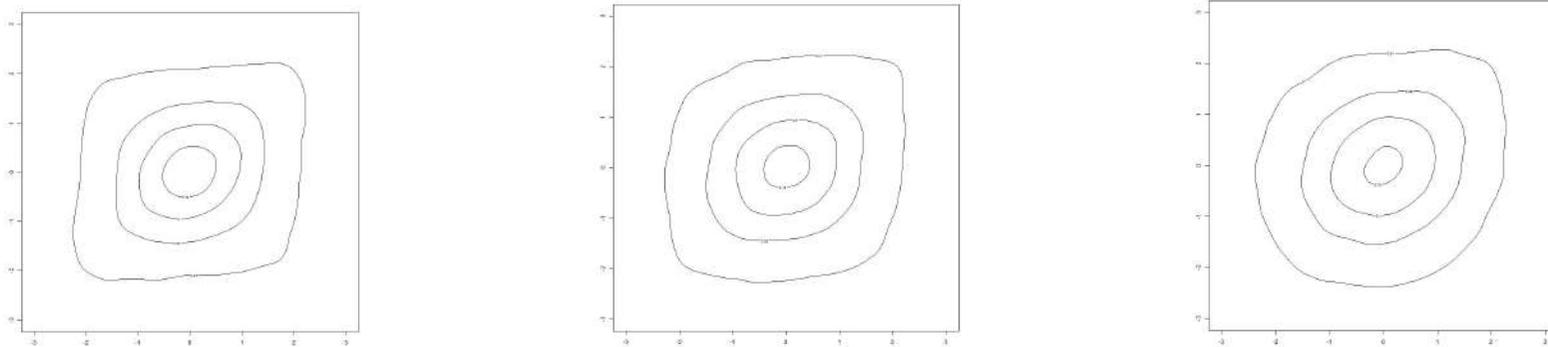
(a1) C_1=Gumbel, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.1$ (a2) C_1=Gumbel, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.5$ (a3) C_1=Gumbel, $C_2 = t_3$, $\tau = 0.1$, $\lambda = 0.9$
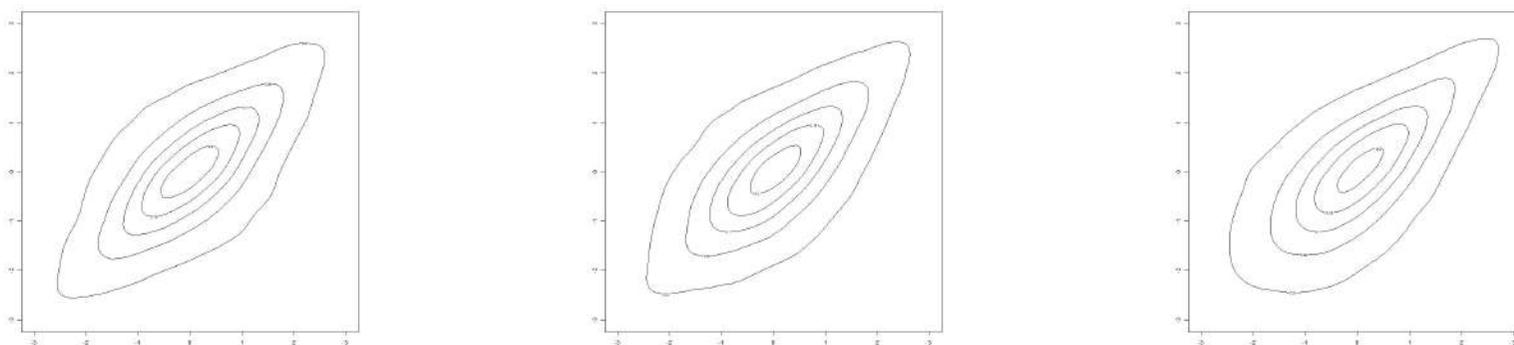
(b1) C_1=Gumbel, $C_2 = t_3$,$\tau = 0.5$, $\lambda = 0.1$ (b2) C_1=Gumbel, $C_2 = t_3$, $\tau = 0.5$,$\lambda = 0.5$ (b3) C_1=Gumbel, $C_2 = t_3$,$\tau = 0.5$,$\lambda = 0.9$

Figure 39: Summary of Z-Level contour plots for $\tau = 0.1$ (first row) and $\tau = 0.5$ (second row) (see equation (9.2))

To define the pair-copulas similarly to Gräler and Pebesma, 2011, we can now define the λ values based on similarity values given the quantiles $(q_{\alpha_i})_{i=1,...,B}$:

Definition 9.4 (Weights)**.**
*For each value of $s_{i,j} \in \mathcal{S}$, we define $\lambda(s_{i,j})$ as the proportion of the **similarity** to the nearest higher quantile q_{α_i} relative to the size of the interval. If $s_{i,j}$ is smaller than q_{α_1}, we set $\lambda(s_{i,j}) = 1$, if $s_{i,j}$ is larger than q_{α_B}, we set $\lambda(s_{i,j}) = 0$:*

$$\lambda(s) := \begin{cases} 1 & \text{if } s \leq q_{\alpha_1} \\ \frac{q_{\alpha_b}-s}{q_{\alpha_B}-q_{\alpha_{b-1}}} & \text{if } s \in (q_{\alpha_{b-1}}, q_{\alpha_b}] \\ 0 & \text{if } s > q_{\alpha_B} \end{cases}$$

According to this approach, we will associate the set of quantiles $\mathsf{q} = (q_{\alpha_1}, ..., q_{\alpha_B})$ with a set of pair-copulas $\mathsf{C} = (C_1(\cdot, \cdot, \theta_1), ..., C_B(\cdot, \cdot, \theta_B))$.

In **Section 9.4** we will use that the similarity measure includes the values of Kendall's τ. This has the advantage, that the Dissmann algorithm (**Algorithm 1**) also requires the calculation of all Kendall's τ values. As we will therefore have to calculate all empirical values of Kendall's τ ($\hat{\tau}(\mathbf{u}_i, \mathbf{u}_j)$), we use these values to estimate the values of θ by the inversion of Kendall's τ, as explained in **Definition 3.14**.

This is why the **set of proposed pair-copulas** for the estimation has to be defined:

Definition 9.5 (Set of proposed pair-copulas)**.**
The set of proposed pair-copulas, from which the pair-copulas in C are chosen from is defined as:

$$\mathcal{C} = \{C_{i,\theta_i}\}_{i=1,...,m}$$

The set only contains copulas whose value of θ_i is a function of τ.

By only including copulas whose value of θ_i is a function of τ, we can reduce the computational effort, as we have to calculate the values of Kendall's τ only once.

Against this background and the definition of $\lambda(s)$, we can now define the **similarity-based pair-copulas**.

Definition 9.6 (Similarity-based pair-copulas)**.**
Given the quantiles $\mathsf{q} = (q_{\alpha_1}, ..., q_{\alpha_B})$ *and the set of associated pair-copulas*
$\mathsf{C} = (C_1(\cdot, \cdot, \theta_1), ..., C_B(\cdot, \cdot, \theta_B))$, *we define the pair-copula* C_s *as:*

$$C_s(u, v; \tau, s, \mathsf{C}, \mathsf{q}) := \begin{cases} C_1(u, v, \theta_1(\tau)) & \text{if } s \leq q_{\alpha_1} \\ \lambda(s)C_{b-1}(u, v, \theta_b(\tau)) + (1 - \lambda(s))C_b(u, v, \theta_b(\tau)) & \text{if } s \in (q_{\alpha_{b-1}} \; q_{\alpha_b}] \\ C_B(u, v, \theta_B(\tau)) & \text{if } s > q_{\alpha_B} \end{cases}$$

With this background, it becomes more clear why the quantile levels were chosen in this way. We choose the values such that each of the intervals contains the same number of pairs, while the two outer intervals, $[0, q_{\alpha_1}]$ and (q_{α_B}, ∞), contain half the number of the observations of the inner intervals. This can be explained as follows: All inner copulas as for example C_2 determine approximately half of the dependence in interval $(q_{\alpha_1}, q_{\alpha_2}]$ and half of the interval $(q_{\alpha_2}, q_{\alpha_3}]$, as it appears only as one component of the pair-copula C_s for these intervals. As each of these intervals contains a share of Δ of the data set, this can be interpreted as determining a proportion of $\Delta/2 + \Delta/2$ of the total dependency. The outer copulas as for example C_1 also determines approximately half of the dependence in the interval $(q_{\alpha_1}, q_{\alpha_2}]$, however it determines the whole dependence in the interval $[0, q_{\alpha_1}]$. As a result, the copula again determines approximately a share of $\Delta/2 + \Delta/2$ of the dependency. Thus, this choice of the quantile levels makes all copulas in the set $\mathsf{C} = (C_1(\cdot, \cdot, \theta_1), ..., C_B(\cdot, \cdot, \theta_B))$ approximately equally important and better to interpret.

One important function for the construction of C-vines is the h-function, as defined in **Definition 3.11**:

$$h(u, v|\theta) = \frac{\partial C_{u,v}(u, v, \theta)}{\partial v}$$

By using the linearity of the differential operator $\frac{\partial}{\partial v}$, we can derive a formula for the h-function of $C_s(u, v; \tau, s, \mathsf{C}, \mathsf{q})$, which therefore is again a linear combination. We get the following definition:

Definition 9.7 (h-functions for similarity-based pair-copulas).
Given $\tau, s, \mathsf{C}, \mathsf{q}$ we define the h function of a similarity-based pair-copulas:

$$
h_s(u,v;\tau,s,\mathsf{C},\mathsf{q}) := \begin{cases} h_1(u,v|\theta_1(\tau)) & \text{if } s \le q_{\alpha_1} \\ \lambda(s)h_{b-1}(u,v|\theta_b(\tau)) + (1-\lambda(s))h_b(u,v|\theta_b(\tau)) & \text{if } s \in (q_{\alpha_{b-1}}, q_{\alpha_b}] \\ h_B(u,v|\theta_B(\tau)) & \text{if } s > q_{\alpha_B} \end{cases}
$$

This formula shows, that the construction of the C-vine copula is a straight forward calculation, if the pair-copulas $\mathsf{C} = (C_1(\cdot,\cdot,\theta_1),...,C_B(\cdot,\cdot,\theta_B))$ are chosen. We therefore have to analyze how we should choose these functions by applying a procedure similar to the maximum likelihood estimation. For this, we have to investigate the density induced by C_s.

Corollary 9.1.
The density of the pair-copula C_s has the following form:

$$
c_s(u,v;\tau,s,\mathsf{C},\mathsf{q}) := \begin{cases} c_1(u,v,\theta_1(\tau)) & \text{if } s \le q_{\alpha_1} \\ \lambda(s)c_{b-1}(u,v,\theta_b(\tau)) + (1-\lambda(s))c_b(u,v,\theta_b(\tau)) & \text{if } s \in (q_{\alpha_{b-1}}, q_{\alpha_b}] \\ c_B(u,v,\theta_B(\tau)) & \text{if } s > q_{\alpha_B} \end{cases}
$$

9.2   C-Vine with similarity-based pair-copulas for jointly chosen pair-copulas

The result derived above can now be used to calculate the density of the copula. The general formula for a C-vine density, as described in **Chapter 3**, is given as follows:

$$
c(\mathbf{u}) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+i|1,...,j-1}(F(u_j|u_1,...,u_{j-1}), F(u_{j+i}|u_1,...,u_{j-1}))
$$
$$
\text{for } \mathbf{u} = (u_1,...,u_d)^t
$$

As mentioned in **Section 3.4**, it is not easy to apply a maximization on this copula density directly, as it leads to very high computational effort. Thus, we have to propose another faster method to calculate the best fitting copulas.

The approach to estimate the suitable copulas sequentially, as described in **Section 3.6**, is not appropriate, as the families of pair-copulas are not chosen independently on each vine level but on all vine levels simultaneously. We thus implemented a **composite likelihood**

approach (see Varin et al., 2011).

We therefore first calculate the logarithmic likelihood of the product of the $\binom{d}{2}$ pairs of companies, given the values $\hat{\boldsymbol{\tau}} = (\hat{\tau}_{i,j})_{i,j=1,\ldots,d}$ and $\mathcal{S} = (s_{i,j})_{i,j=1,\ldots,d}$.

The aim of this procedure is to choose the copulas $\mathsf{C} = (C_1(\cdot,\cdot,\theta_1),\ldots,C_B(\cdot,\cdot,\theta_B))$ from $\mathcal{C}$, the set of proposed copulas, as defined in **Definition 9.5**:

$$
\begin{aligned}
\mathcal{L}_s(\mathbf{u}_1,\ldots,\mathbf{u}_d;\hat{\boldsymbol{\tau}},\mathcal{S},\mathsf{C},\mathsf{q}) &= log\left(\prod_{k=1}^{n}\prod_{i=1}^{d}\prod_{j=i+1}^{d} c_s(u_{i,k},u_{j,k};\hat{\tau}_{i,j},s_{i,j},\mathsf{C},\mathsf{q})\right) \\
&= \sum_{k=1}^{n}\sum_{i=1}^{d}\sum_{j=i+1}^{d} log(c_s(u_{i,k},u_{j,k};\hat{\tau}_{i,j},s_{i,j},\mathsf{C},\mathsf{q}))) \qquad (9.3) \\
&= \sum_{(i,j):j>i,s_{i,j}\leq bin_1}\sum_{k=1}^{n} log(c_1((u_{i,k},u_{j,k},\theta_1(\hat{\tau}_{i,j}))) \\
&\quad + \sum_{b=2}^{B-1}\sum_{(i,j):j>i,s_{i,j}\in(q_{\alpha_{b-1}},q_{\alpha_B}]}\sum_{k=1}^{n}\Bigg[log(\lambda(s_{i,j})c_{b-1}(u_{i,k},u_{j,k},\theta_{b-1}(\hat{\tau}_{i,j})) \\
&\qquad\qquad\qquad\qquad\qquad + (1-\lambda(s_{i,j}))c_b(u_{i,k},u_{j,k},\theta_b(\hat{\tau}_{i,j})))\Bigg] \\
&\quad + \sum_{(i,j):j>i,\hat{s}_{i,j}>q_{\alpha_B}}\sum_{k=1}^{n} log(c_B(u_{i,k},u_{j,k},\theta_B(\hat{\tau}_{i,j})))
\end{aligned}
$$

We can see above that the expression c_1 influences the first two summands. In the same way each of the functions influences two summands, as it can be seen from **Figure 37**. Thus, the choice of the the copula families maximizing the function $\mathcal{L}_s$ is not independently possible.

The sum in equation (9.3) can be seen as a optimizing problem of the following form:

$$
(f_i^\star)_{i=1,\ldots,B} = argmax_{f_i\in\{f_1,\ldots,f_n\}}\sum_j log(f_1(a_{j,1})+\ldots+f_B(a_{j,B}))
$$

In our case these functions f_i are the density function, which we want to choose. We approximate the solution by optimizing over each f_i:

$$
f_i' = argmax_{f=f_1,\ldots,f_n}\sum_j log(f(a_{j,i}))
$$

Using these chosen values we can thus approximate our solution as follows:

$$
(f'_i)_{i=1,\ldots,B} \approx (f^\star_i)_{i=1,\ldots,B} \qquad (9.4)
$$

As we used a **composite likelihood** approach, which is an approximation itself, we considered it to be appropriate to use the approximation as described in equation (9.4). With this assumption, we receive the following **discrete maximization**:

$$c_1 := argmax_{\{c \in \mathcal{C}\}} \left[\sum_{(i,j):j>i, s_{i,j} \leq bin_1} \sum_{k=1}^{n} log(c(u_{i,k}, u_{j,k}, \theta(\hat{\tau}_{i,j}))) \right.$$
$$\left. + \sum_{(i,j):j>i, s_{i,j} \in (q_{\alpha_1}, q_{\alpha_2}]} \sum_{k=1}^{n} log(\lambda(s_{i,j})c(u_{i,k}, u_{j,k}, \theta(\hat{\tau}_{i,j}))) \right]$$

Following this approach, we get:

$$c_b := argmax_{\{c \in \mathcal{C}\}} \left[\sum_{(i,j):j>i, s_{i,j} \in (q_{\alpha_{b-1}}, q_{\alpha_b}]} \sum_{k=1}^{n} log((1 - \lambda(s_{i,j}))c(u_{i,k}, u_{j,k}, \theta(\hat{\tau}_{i,j}))) \right.$$
$$\left. + \sum_{(i,j):j>i, s_{i,j} \in (q_{\alpha_b}, q_{\alpha_{b+1}}]} \sum_{k=1}^{n} log(\lambda(s_{i,j})c(u_{i,k}, u_{j,k}, \theta(\hat{\tau}_{i,j}))) \right]$$

Analogously, we get:

$$c_B := argmax_{\{c \in \mathcal{C}\}} \left[\sum_{(i,j):j>i, s_{i,j} \in (q_{\alpha_{B-1}}, q_{\alpha_B}]} \sum_{k=1}^{n} log((1 - \lambda(s_{i,j}))c(u_{i,k}, u_{j,k}, \theta_B(\hat{\tau}_{i,j}))) \right.$$
$$\left. + \sum_{(i,j):j>i, s_{i,j} \geq q_{\alpha_B}} \sum_{k=1}^{n} log(c(u_{i,k}, u_{j,k}, \theta_B(\hat{\tau}_{i,j}))) \right]$$

In this way, we have derived a fast numerical procedure to approximate the best fitting copulas. From this step on the pair-copula construction can proceed as usual:

1. The vine structure, which is in our case restricted to be a C-vine, is chosen by the Dissmann algorithm. As we apply a C-vine, this is equivalent to choosing the inner node i_1.

2. The pair-copulas $C_{j,i_1}(\cdot,\cdot) = C_s(\cdot,\cdot,\tau_{j,i_1}, s_{j,i_1}, \mathsf{C}, \mathsf{q})$ required for the first level of the C-vine, are chosen.

3. The values of $\mathbf{u}_j$ are transformed using the h-function associated with the pair-copulas C_{j,i_1}. And thus the values $\mathbf{u}_{j|i_1}$ are calculated.

4. For this transformed values, the procedure again starts from step 1.

This procedure will be repeated until the whole C-vine is specified.

Summarized, the developed procedure works as follows:

At first the empirical values of Kendall's $\hat{\boldsymbol{\tau}} = \{\hat{\tau}_{i,j}\}_{i \in \{1,\ldots,d-1\},\ j \in \{i+1,\ldots,d\}}$ are calculated.
Second, the values of $\mathcal{S}$ are calculated, in this example, this requires the values of $\hat{\boldsymbol{\tau}}$. Based on these values the quantiles q are calculated.
With all these values the set of copulas $\mathsf{C} = (C_1, \ldots, C_B)$ as described in **Section 9.2** is chosen with a composite likelihood approach. This requires a discrete optimization, where each pair-copula C_i is chosen from the set of proposed copulas $\mathcal{C}$.
Based on the empirical Kendall's τ values $\hat{\boldsymbol{\tau}}$ the Dissmann algorithm is applied to choose the root node for the next tree level. As we restrict the choice to the class of C-vines the Dissmann algorithm is equivalent to choosing the inner node i_1 by the following maximization:

$$i_1 = argmax_{i=1,\ldots,d} \sum_{j \neq i} |\tau_{i,j}|$$

Based on the values of $\hat{\boldsymbol{\tau}}$, $\mathcal{S}$ and q the pair-copulas $C_{i_1,j}$ are chosen, based on the interval in which $s_{i_1,j}$ lies and the value of $\hat{\tau}_{i_1,j}$ (see **Definition 9.6** and **Figure 37**).
With these values, the variables are transformed and thus the values $\mathbf{u}_{j|i_1}$ are calculated.

On the **second level** of the C-vine this procedure changes. The values of $\hat{\boldsymbol{\tau}}$, $\mathcal{S}$ are calculated based on the pairs of transformed variables $\mathbf{u}_{j|i_1}$. The value of i_2 is again calculated by maximization. However, the set C is not newly estimated. Thus, the pair-copulas $C_{i_2,j;i_1}$ are chosen directly based on the values of $s_{i_2,j}$ and $\hat{\tau}_{i_2,j;i_1}$.
The same procedure, as for the second level, is applied on the **following levels** of the C-vine.

This procedure is summarized and visualized in **Figure 40**.
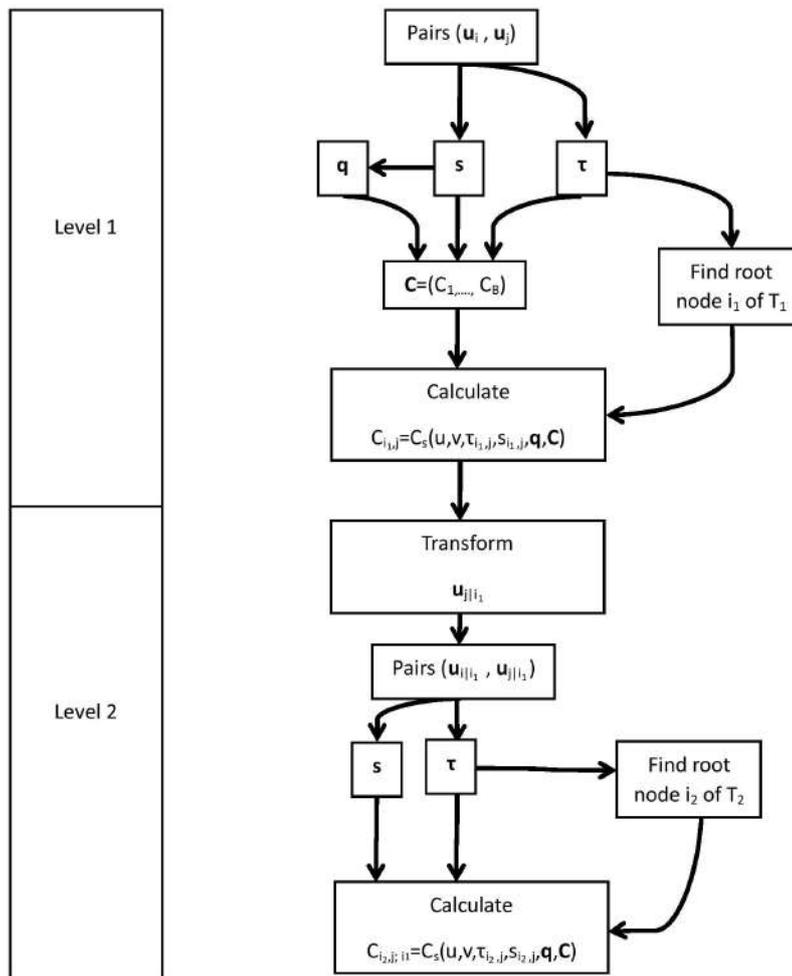
Figure 40: Summary of the similarity-based C-vine construction

9.3   C-vine with similarity-based pair-copulas chosen adaptively up to level K

The developed approach can successively be applied to the values on the different C-vine levels. For this, the until now fixed values of C and q are newly chosen on the consecutive tree levels. And thus the copulas C_s is updated on consecutive tree levels. The major reason for this procedure is to improve the fit of the similarity-based model by increasing its flexibility.

The new choice of the C_s is repeated until a fixed level K. From this C-vine level on C_s is not changed and the algorithm proceeds as described in **Section 9.2**.

This procedure will be called **Renewing the choice of similarity-based copulas up to level K**.

The question which arises is until which level K which number of repetitions should be chosen. To solve this problem we introduce the AIC criterion for similarity-based C-vines:

Definition 9.8 (AIC for similarity-based C-vines)**.**
Given a similarity-based C-vine, and the empirically estimated value $\hat{\boldsymbol{\theta}}$ we define the AIC as:

$$AIC = 2 \cdot B \cdot K - 2 \cdot \textit{log-likelihood}(\hat{\boldsymbol{\theta}})$$

Where $B \cdot K$ stands for the number of variables in the model which is equal to the product of B, the number of families per level, and K the level up to which similarity-based copulas are chosen.

We will choose the model with the lowest values of AIC.

This newly changed procedure will be visualized, as in the preceding section, with **renewing the choice of similarity-based copulas up to level** 2.
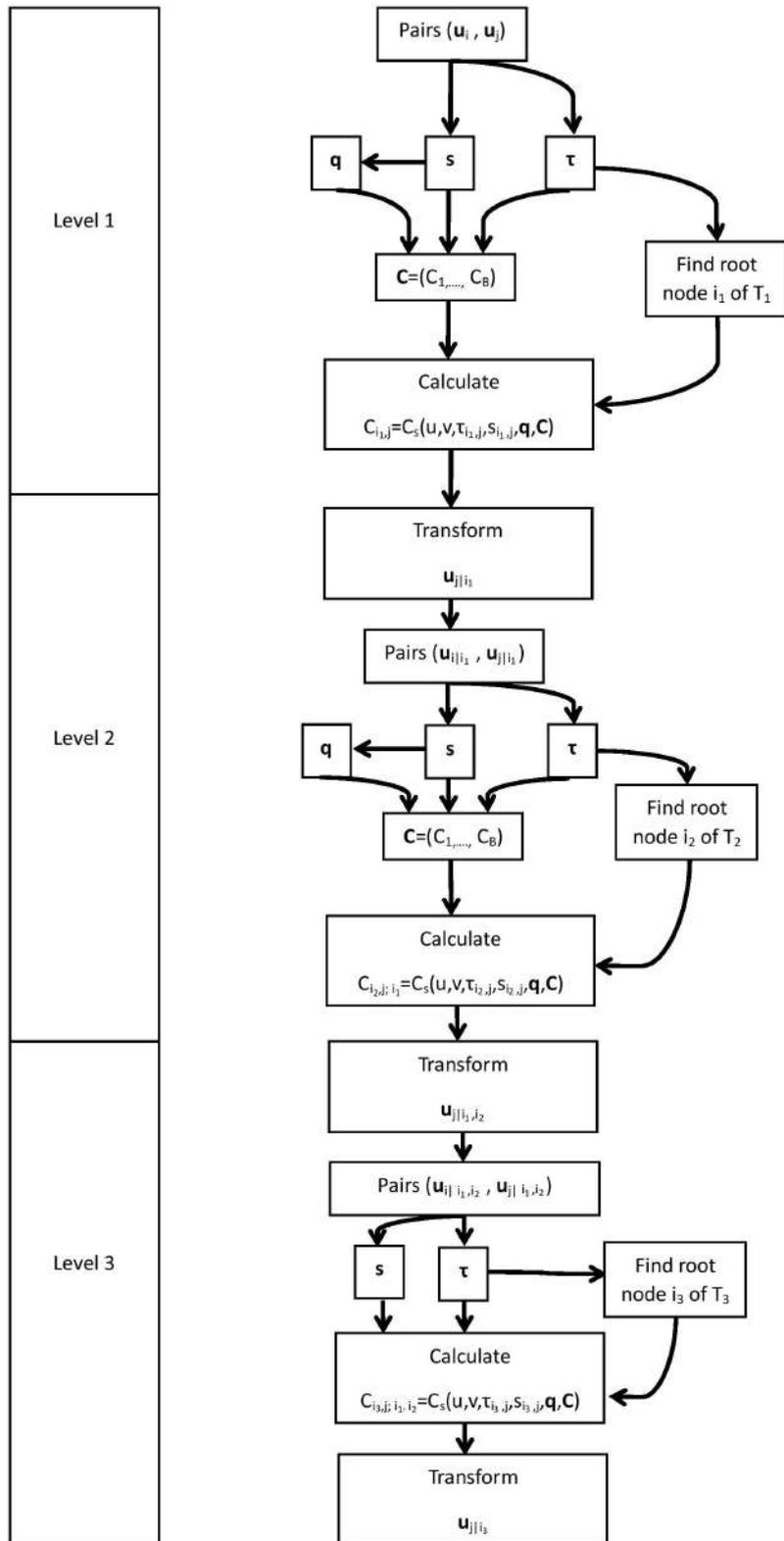
Figure 41: Summary of the similarity-based C-vine construction with renewing up to level $K = 2$

On the following pages we will provide the algorithms derived in this chapter to be able to construct similarity-based C-vines.

Algorithm 5 Choosing pair-copula families

Calculating the similarity of the institutions and choosing the pair-copula families

Require: Quantile levels $\alpha_1, ..., \alpha_B$, $(\mathbf{u}_i)_{i=1,...,d}$ with $\mathbf{u}_i \in [0,1]^n$, set of proposed (continuous) copulas $\mathcal{C}$ with parameter θ only depending on τ, similarity function s

1: **for** $i \leftarrow 1$ **to** d, $j \leftarrow i+1$ **to** d **do**

2:     $s_{i,j} = s(\mathbf{u}_i, \mathbf{u}_j, i, j)$

3: **end for**

4: **for** $b \leftarrow 1$ **to** B **do**

5:     Calculate the quantile q_{α_b} with level α_b of $(s_{i,j})_{1 \leq i < j \leq d}$

6: **end for**

7: Define function λ and calculate $(\lambda_{i,j})_{1 \leq i < j \leq d} = (\lambda(s_{i,j}))_{1 \leq i < j \leq d}$ given $(q_{\alpha_b})_{b=1,...,B}$

8: $c_{1,\theta_1} = argmax_{c_1 \in \mathcal{C}} \left[\sum_{\substack{j>i: \\ s_{i,j} \leq q_{\alpha_1}}} \sum_{k=1}^{n} log(c_{1,\theta_1(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right.$

9: $\left. + \sum_{\substack{j>i: \\ s_{i,j} \in (q_{\alpha_1}, q_{\alpha_2}]}} \sum_{k=1}^{n} log(\lambda_{i,j} c_{1,\theta_1(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right]$

10: **for** $b \leftarrow 2$ **to** B-1 **do**

11:     $c_b = argmax_{c_b \in \mathcal{C}} \left[\sum_{\substack{j>i: \\ s_{i,j} \in (q_{\alpha_{b-1}}, q_{\alpha_b}]}} \sum_{k=1}^{n} log((1-\lambda_{i,j}) c_{b,\theta_b(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right.$

12:     $\left. + \sum_{\substack{j>i: \\ s_{i,j} \in (q_{\alpha_b}, q_{\alpha_{b+1}}]}} \sum_{k=1}^{n} log(\lambda_{i,j} c_{b,\theta_b(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right]$

13: **end for**

14: $c_B = argmax_{c_B \in \mathcal{C}} \left[\sum_{\substack{j>i: \\ s_{i,j} > q_{\alpha_B}}} \sum_{k=1}^{n} log(c_{B,\theta_B(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right.$

15: $\left. + \sum_{\substack{j>i: \\ s_{i,j} \in (q_{\alpha_{B-1}}, s_{\alpha_B}]}} \sum_{k=1}^{n} log((1-\lambda_{i,j}) c_{B,\theta_B(\hat{\tau}_{i,j})}(u_{i,k}, u_{j,k})) \right]$

Algorithm 6 Pair-copulas and h-functions as part of the similarity-based copula approach

Derive h-functions and densities for a R-vine structure using the similarity-based copula approach

---

Require: Quantiles $q_1, ..., q_B$, $(\mathbf{u}_i)_{i=1,...,d}$ with $\mathbf{u}_i \in [0,1]^n$, set of densities $\{c_{1,\theta(\tau)}, ..., c_{B,\theta(\tau)}\}$ with parameter θ only depending on τ,

matrix $\mathcal{S} = (s_{i,j})_{i,j=1,...,d}$, matrix $(\hat{\tau}_{i,j})_{i,j=1,...,d}$

1: Function$(c_s)(u, v, \hat{\tau}_{i,j}, s_{i,j})$

2: **if** $s_{i,j} \leq q_{\alpha_1}$ **then**

3: $\quad c_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = c_{1,\theta_1(\hat{\tau}_{i,j})}(u, v)$

4: **end if**

5: **for** $b \leftarrow 2$ **to** B **do**

6: $\quad$ **if** $s_{i,j} \in (q_{\alpha_{b-1}}, q_{\alpha_b}]$ **then**

7: $\quad\quad c_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = \lambda_{i,j} c_{b-1,\theta_b(\hat{\tau}_{i,j})}(u, v) + (1 - \lambda_{i,j}) c_{b,\theta_b(\hat{\tau}_{i,j})}(u, v)$

8: $\quad$ **end if**

9: **end for**

10: **if** $s_{i,j} > q_{\alpha_B}$ **then**

11: $\quad c_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = c_{B,\theta_B(\hat{\tau}_{i,j})}(u, v)$

12: **end if**

13: EndFunction

14: Function$(h_s)(u, v, \hat{\tau}_{i,j}, s_{i,j})$

15: **if** $s_{i,j} \leq q_{\alpha_1}$ **then**

16: $\quad h_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = h_1(u, v | \theta_1(\hat{\tau}_{i,j}))$

17: **end if**

18: **for** $b \leftarrow 2$ **to** B **do**

19: $\quad$ **if** $s_{i,j} \in (q_{\alpha_{b-1}}, q_{\alpha_b}]$ **then**

20: $\quad\quad h_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = \lambda_{i,j} h_{b-1}(u, v | \theta_b(\hat{\tau}_{i,j})) + (1 - \lambda_{i,j}) h_b(u, v | \theta_b(\hat{\tau}_{i,j}))$

21: $\quad$ **end if**

22: **end for**

23: **if** $s_{i,j} > q_{\alpha_B}$ **then**

24: $\quad h_s(u, v, \hat{\tau}_{i,j}, s_{i,j}) = h_B(u, v | \theta_B(\hat{\tau}_{i,j}))$

25: **end if**

26: EndFunction

The h-function and the copula functions, given by **Algorithm 6** are used to implement the C-vine procedure as described in **Section 3.3**.

To compare the fit of this procedure with the individually chosen R-vine copula, we calculate the log-likelihood of the observations $(\mathbf{u}_k)_{k=1,\dots,n}$ as it is needed for the AIC values.
The general algorithm to calculate the log-likelihood for C-vines is given by Aas et al., 2009, and is described in **Algorithm 7** which requires the h-function and the vine density given by **Algorithm 6**.

Algorithm 7 Log-likelihood evaluation for a similarity-based C-vine
---
Calculate the log-likelihood for a similarity-based C-vine

Require: Observations $(\mathbf{u_i})_{\mathbf{i=1},\dots,\mathbf{d}}$ with $u_i \in \mathbb{R}^n$, h-function h_s and density c_s given by
　　Algorithm 6, matrix $\mathcal{S} = (s_{i,j})_{i,j=1,\dots,d}$, matrix $(\hat{\tau}_{i,j})_{i,j=1,\dots,d}$

1: log-likelihood=0

2: $i = 1$

3: **for** $i \leftarrow 1$ **to** d **do**

4: 　　$\mathbf{v_{0,i}} = \mathbf{u_i}$

5: **end for**

6: **for** $j \leftarrow 1$ **to** $d-1$ **do**

7: 　　**for** $i \leftarrow 1$ **to** $d-j$ **do**

8: 　　　　**for** $k \leftarrow 1$ **to** n **do**

9: 　　　　　　log-likelihood=log-likelihood+$log(c_s(v_{j-1,1,k}, v_{j-1,i+1,k}, \hat{\tau}_{j,i}, s_{j,i}))$

10: 　　　　**end for**

11: 　　**end for**

12: 　　**if** $j == d-1$ **then**

13: 　　　　STOP ALGORITHM

14: 　　**end if**

15: 　　**for** $i \leftarrow 1$ **to** $d-j$ **do**

16: 　　　　**for** $k \leftarrow 1$ **to** n **do**

17: 　　　　　　$v_{j,i,k} = h_s(v_{j-1,i+1,k}, v_{j-1,1,k}, \hat{\tau}_{j,i}, s_{j,i})$

18: 　　　　**end for**

19: 　　**end for**

20: **end for**

9.4   Application of C-vine regression analysis with similarity-based pair-copulas

In this section we will continue with the example for similarity measures, the **similarity of institutions**, which we introduced in **Section 9.1**. We will then apply the concept of similarity-based pair-copulas to the data set described in **Chapter 7**.

Definition 9.9 (Similarity of institutions).
The similarity between the institutions i and j, with measurements $\mathbf{u}_i \in [0,1]^n$ and $\mathbf{u}_j \in [0,1]^n$, is given by:

$$s(\mathbf{u}_i, \mathbf{u}_j, i, j) := |\hat{\tau}(\mathbf{u}_i, \mathbf{u}_j)| + geographical\_closeness(i,j)$$

The values for the closeness among the geographical regions, $geographical\_closeness(i,j)$, can be chosen arbitrarily, but should be symmetric in i and j. These values can be seen as regression parameters, as one possibility could be finding the set of values $geographical\_closeness(i,j)$ which maximizes the likelihood. The comparison of the goodness of fit of these values can be achieved by the AIC criterion, as introduced in **Definition 9.8**.
Higher values for the geographical closeness indicate a high inter-similarity of these geographical regions, which can be interpreted as small geographical distances between these institutions.
For the analysis the following values listed in **Table 27** were chosen:

	North America	Europe	Asia Pacific
North America	1	$\dfrac{0.002}{1.002}$	$\dfrac{0.002}{1.002}$
Europe	$\dfrac{0.002}{1.002}$	1	$\dfrac{0.001}{1.001}$
Asia Pacific	$\dfrac{0.002}{1.002}$	$\dfrac{0.001}{1.001}$	1

Table 27: Chosen values for geographical closeness

Knowing that the value of $geographical\_closeness(i,j)$ has an upper bound of 1 and a lower bound of $\frac{0.001}{1.001}$, and that the value of $|\hat{\tau}|$ lies in $[0,1]$, we thereby get the following

bounds for the value of $s(\mathbf{u}_i, \mathbf{u}_j, i, j)$:

$$\frac{0.001}{1.001} \leq s(\mathbf{u}_i, \mathbf{u}_j, i, j) \leq 2$$

With that done, we can now calculate the similarity for each pair of institutions. As we have 38 institutions, this leads to $\binom{38}{2} = 703$ pairs, each of which is related to a specific similarity.

By choosing $\alpha_i = \frac{1}{2B} + \frac{i-1}{2B}$ we receive the following quantile levels α_i:
$\alpha_1 = \frac{1}{18}, \alpha_2 = \frac{3}{18}, \alpha_3 = \frac{5}{18}, \alpha_4 = \frac{7}{18}, \alpha_5 = \frac{9}{18}, \alpha_6 = \frac{11}{18}, \alpha_7 = \frac{13}{18}, \alpha_8 = \frac{15}{18}, \alpha_9 = \frac{17}{18}$.

In our example, the quantiles are located at the following positions:
$q_{\alpha_1} = 0.1389$; $q_{\alpha_2} = 0.1837$; $q_{\alpha_3} = 0.2033$; $q_{\alpha_4} = 0.223$; $q_{\alpha_5} = 0.2643$; $q_{\alpha_6} = 1.259$; $q_{\alpha_7} = 1.372$; $q_{\alpha_8} = 1.424$; $q_{\alpha_9} = 1.5129$.
These quantiles are shown as vertical blue lines in **Figure 42**, together with an interpretation of the strength of the dependencies within the different regions:
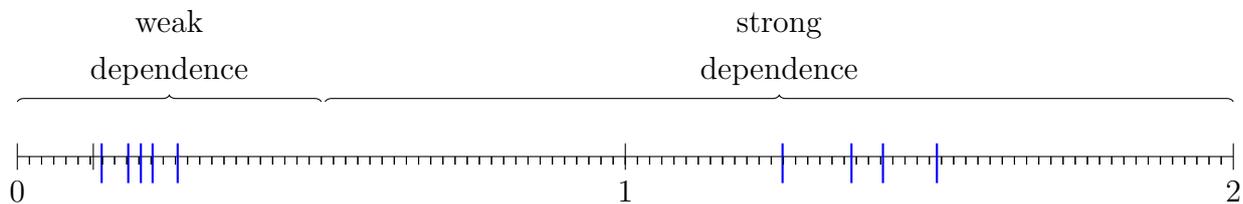

Figure 42: Location of the quantiles in the example, within the interval [0,2]

In this section we apply the algorithms summarized in **Algorithm 5**, **Algorithm 6** and in **Algorithm 7** to estimate and construct the similarity-based C-vine.

As the set of pair-copula families we choose: Gaussian, Gumbel, Frank and Student copulas with $\nu \in \{3, 4, 5, 6, 7, 10, 15, 20, 25\}$.

To improve the fit of the model we will use the concept of **Renewing of the choice of similarity-based copulas up to level** K, as described in **Section 9.3**. This can improve the fit, as the model becomes more flexible, as the choice of the location of the quantiles and the pair-copula families can vary.

Table 28 summarizes the AIC values, the log-likelihood values and the number of parameters used in the models depending on the vine tree level, up to which the choice of similarity-based copulas is renewed. To compare these models with the standard approach, we add the full C-vine model chosen in **Chapter 8**.

As a decision rule for the choice of the model, we choose the AIC value. According to this rule, it is optimal to renew the choice of similarity-based copulas up to level 4.

C-vine level K up to which the choice of similarity-based pair-copula are renewed	Log-likelihood	AIC values	Number of parameters
1	-41205.06	72316	7
2	20041.21	-40207.21	14
3	19389.93	-39234.68	21
4	*20232.17*	*-40446.05*	*28*
5	19725.36	-39273.75	35
6	20134.33	-40172.15	42
7	20071.86	-40169.63	49
8	20166.32	-40325.76	56
9	19838.66	-40201.5	63
10	19952.03	-39724.05	70
full C-vine	21162.96	-40407.91	959

Table 28: Comparing models with different choices of the value K up to which the choice of similarity-based pair-copula is renewed

The total log-likelihood of the chosen similarity-based C-vine is 20232.17, whereas the log-likelihood of the full C-vine model with free copula choice was 211162.96. This indicates that the similarity-based model describes the data worse. This is reasonable, as the similarity-based model is much less flexible.

However, comparing the resulting AIC with the AIC of the full C-vine model shows a different result, as the full model has an AIC of -40407.91 but the similarity-based model has an even lower value of -40446.05. This better result is due to the low number of parameters required in the similarity-based model.

Thus compared to the small number of parameters included in the model, it has a surprisingly good fit.

In **Table 29** the detailed values of the quantiles and the pair-copulas are summarized.

Level	q_{α_1}	q_{α_2}	q_{α_3}	q_{α_4}	q_{α_5}	q_{α_6}	q_{α_7}	q_{α_8}	q_{α_9}
1	0.1389	0.1836	0.2033	0.223	0.2643	1.2586	1.371	1.4239	1.5129
2	0.0559	0.0811	0.0942	0.1071	0.1283	1.1162	1.1545	1.192	1.3191
3	0.0341	0.0511	0.0649	0.0779	0.0961	1.051	1.0934	1.1437	1.2217
4	0.0245	0.04	0.0511	0.0651	0.0866	1.0193	1.0648	1.1251	1.2019

(a) Summary of the positions of the quantiles

Level	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
1	t_5	t_{10}	t_{10}	t_{10}	t_3	t_{10}	t_6	t_6	t_4
2	t_{25}	Frank	t_{25}	Frank	t_{25}	t_{15}	t_{10}	t_{10}	t_6
3	t_{15}	t_{25}	t_{15}	t_{25}	t_{25}	Frank	t_{15}	t_{15}	t_7
4	Frank	t_{25}	Frank	t_{25}	t_{25}	t_{25}	t_{20}	t_{15}	t_7

(b) Summary of the chosen pair-copula families

Table 29: Summary of the C-vine structures of the similarity-based copula model

Level	1	2	3	4	5	6	7	8	9	10
Similarity-based C-vine	Allianz	BNP	Zurich	BS	LG	StanCha	Aegon	Intesa	HannRe	Barclays
Traditional full C-vine	Allianz	XLG	BNP	Zurich	Hartford	BBVA	LG	JPM	Aegon	QBE
Level	11	12	13	14	15	16	17	18	19	20
Similarity-based C-vine	AXA	SCOR	SG	ACE	AssGen	Hartford	BBVA	Prudential	JPM	DB
Traditional full C-vine	Barclays	ACE	AssGen	Kookmin	HannRe	DB	Chubb	TM	Unicredit	SCOR
Level	21	22	23	24	25	26	27	28	29	30
Similarity-based C-vine	Westpac	Chubb	SwissRe	BoC	XLG	RBS	TM	Citi	Unicredit	QBE
Traditional full C-vine	Citi	SwissRe	RBS	Prudential	SG	AXA	BoC	Sumitomo	StanCha	Allstate
Level	31	32	33	34	35	36	37	38		
Similarity-based C-vine	Allstate	Aviva	AIG	Kookmin	MR	UBS	GS	Sumitomo		
Traditional full C-vine	BS	AIG	MR	Intesa	UBS	Westpac	GS	Aviva		

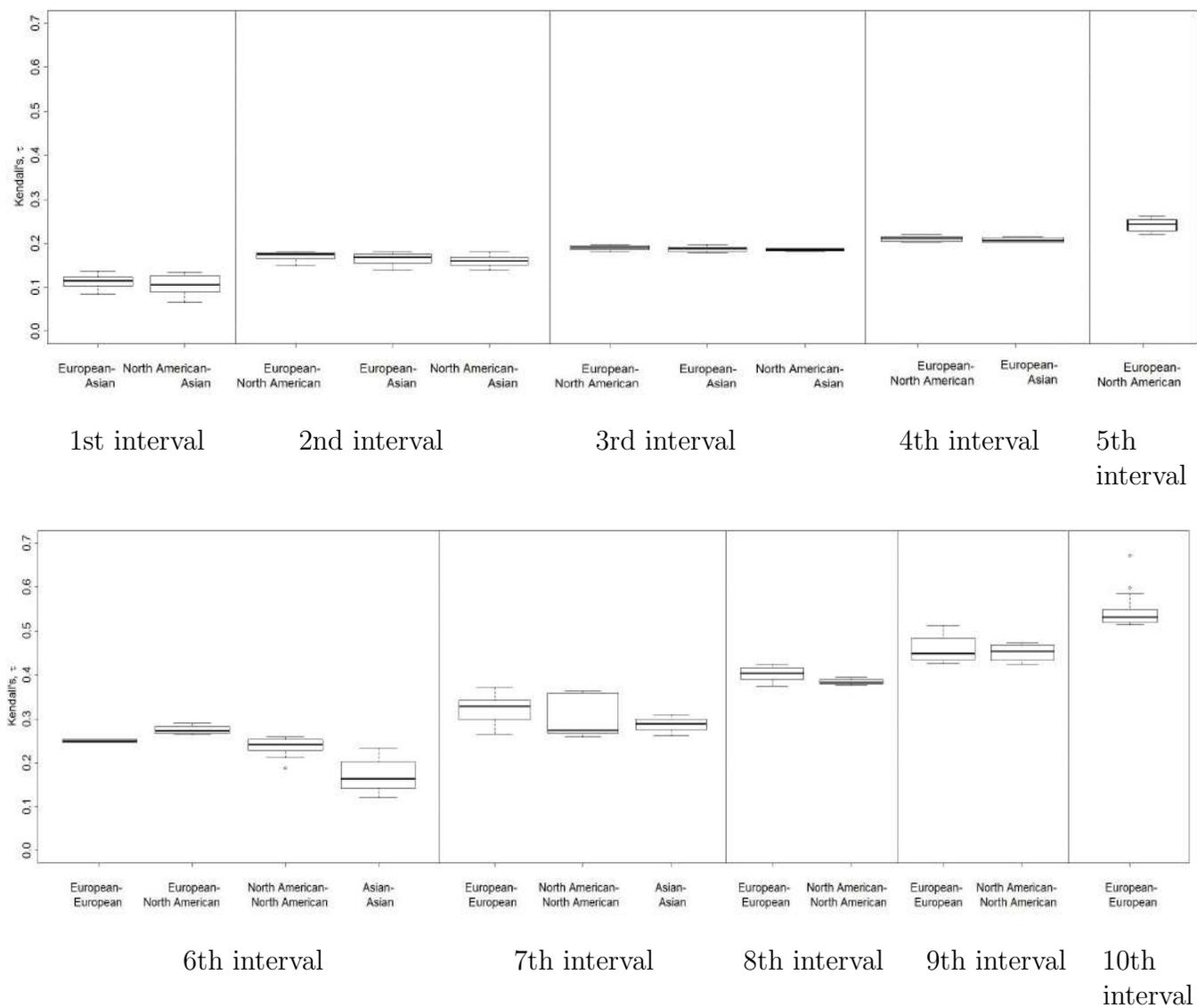Table 30: Summary of the inner nodes of the C-vine tree structures

Figure 43: Box plots of the Kendall's τ values for the different intervals calculated by model 1 and different geographical regions

In **Table 29** one can apply the interpretation of the tail dependence of the pair-copulas. As we saw in **Lemma 3.3**, the tail dependence of the Frank copula is zero. In **Lemma 3.5** we saw that the tail dependence of Student copulas is decreasing in ν.

In most of the cases the degrees of freedom parameter is decreasing from level to level and thus the tail dependence decreases. This indicates that the strongest dependencies are observed in lower levels.

Interestingly, C_5 has a high tail dependence on Level 1 as $\nu = 3$. C_5 effects the interval $(q_{\alpha_4}, q_{\alpha_5}]$ and $(q_{\alpha_5}, q_{\alpha_6}]$. The interval $(q_{\alpha_5}, q_{\alpha_6}]$, however, contains a big proportion of similarity values between European and North American institutions as one can see from **Figure 43**. These have very high Kendall's τ values, as they have to compensate for their low values of geographical closeness, to lie in the interval $(q_{\alpha_5}, q_{\alpha_6}]$.

In **Table 30** one can see the full C-vine specification of the traditional full C-vine and the similarity-based C-vine. As both models are C-vines, it is sufficient to list the inner nodes of the models.
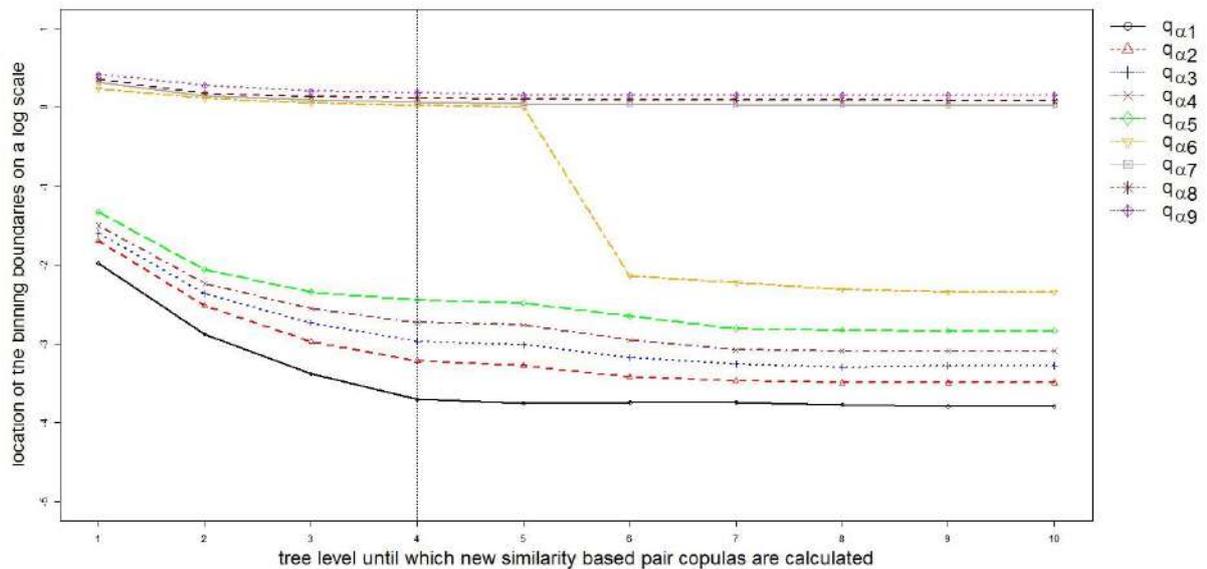

Figure 44: Locations of the bins on different tree levels

It is illustrated in **Figure 44**, that the quantile values are decreasing from level to level. Another observation, that is one explanation for the bad fit of model 1 in **Table 28**, is the sharp decrease in values of the quantiles after the first level. These changing values result in a badly fitting model as shown in **Table 28**.

One can further see a sharp decrease in the value of q_{α_6} between level 5 and 6. The reason for this is that the Dissmann algorithm chooses the companies with the highest sum of Kendall's τ values, these, however, have also higher similarity values. Thus, after several

repetitions of the algorithm many of these values have been chosen as the interior node of the tree of the C-vine, and were therefore not included in the following tree levels, thus the quantile drops.

In **Figure 45** we plot the histograms for the Kendall's τ values and the similarity values for each pair of institutions. The different geographical regions are indicated by different colors. In **Figure 45 (b)** the effect of the *geographical_closeness* is illustrated. One can see how, compared to the set of Kendall's τ values, the set of values is divided into two groups. This is the effect of the *geographical_closeness*: Depending on whether both companies come from the same region or not, the similarity value is high or low.
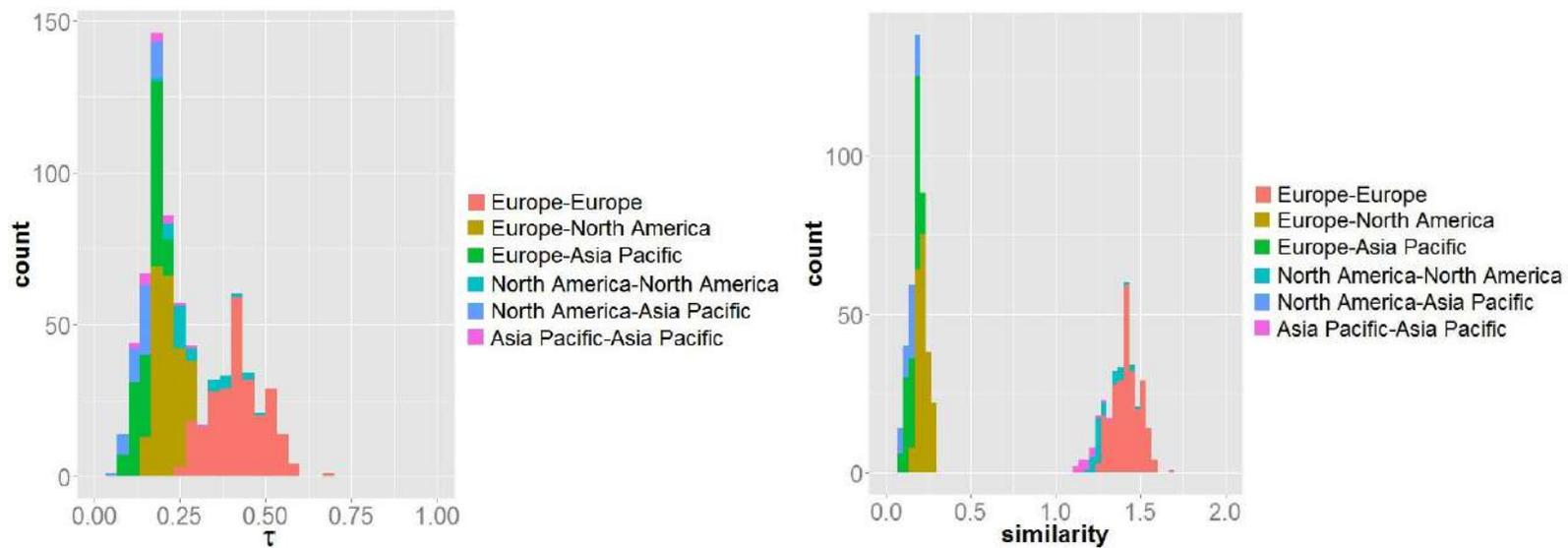
(a) Histogram of Kendall's τ values

(b) Histogram of similarity values

Figure 45: Histogram of Kendall's τ values and similarity values for all pairs of companies

10 Conclusions and Outlook

In this thesis a method was developed to simulate conditional random variables with arbitrary continuous R-vine dependency structures using the MH algorithm. Therefore, it is possible with this method to condition on arbitrary sets of variables with arbitrary values.

The proposed method was then validated and successfully applied to a data set of transformed CDS spreads as these were considered to be capable of measuring default probabilities. Stress tests were thereby conducted as simulations of conditional random variables: Specially, this allows us to asses the reactions of financial markets and institutions.

The methodology used has several advantages, which will be listed in the following.

The advantage of using R-vine models for stress tests is that they are capable to model dependencies much more flexible than usual approaches using the multidimensional Gaussian or Student distributions.

Another contribution of this thesis is that the stress tests carried out do not use balance sheets in contrast to most of the currently used methods. An advantage of the approach described in this thesis is that CDS spreads are not influenced by political decisions. A further advantage is that no new sensitive data have to be collected to implement the stress tests.

The results of the stress tests are consistent with the study carried out by Hendrich, 2012, as we confirmed that geographical regions are of key importance in stress testing. However, this thesis reveals many new findings about stress testing as a measure of the sensitivity of the financial sector to systemic risk indicators. Some of the key results can be summarized as follows:

- One of the most explaining variables for the interconnectedness of financial institutions is their corresponding geographical region.

- Even within same geographical regions strong differences in the reaction of financial institutions upon stress situations were present.

- Major European banks seem to be at least as systemically important as major North American banks.

- North American banks have a stronger effect upon European banks than upon North American insurances.

- Asian Pacific financial institutions are not strongly effected by stress situations induced by North American banks or European banks.

- The major European banks are in the case of a crises, the key driver of financial stress within Europe.

A further improvement in the speed of convergence of the adaptive MH algorithm could be achieved by including other adapted proposal distributions. Particularly other adapted copulas could be used for better simulations from non-elliptical distributions, as until now Gaussian copulas and uniform Beta distributions were used predominantly.

In addition, to monitor world-wide market developments, the approach of this thesis could be extended by including further financial institutions. These extended models could improve the comprehensive analysis of world-wide financial markets as more complex dependence structures can be taken into account. This issue, however, was not included in the scope of this thesis, as such models require high-performance computing, exceeding the already large computation time in this work.

In **Chapter 9**, additionally to the stress tests conducted, we established a new method of similarity-based pair-copula construction for C-vines which is capable to describe the empirical data well, despite of the low number of parameters estimated. Such an approach has only been used for geographical data so far, e.g. rain fall statistics. Taking advantage of this method, we analogously included the geographical region of financial institutions as explanatory variables.

Further parameters, especially the balance sheet sizes of the institutions, could be used to further define the similarity of institutions. In this way, the two major topics of this thesis could be linked more closely.
Another point to mention is, that in this thesis the dependence was assumed to be constant over time. The similarity-based approach, however, could be extended to use time varying similarity measures, and thus analyze the dependency during historical financial crises.

A potential overall problem which could evolve with the implementation of these stress tests, is that the self-reflexiveness of CDS spreads (see Markose et al., 2009) could even increase. The influence of trading CDS on the stress test results of certain financial institutions may therefore be taken into consideration. However, the CDS market has a high trading volume (see Oehmke and Zawadowski, 2013) and thus this effect would probably be not as strong.

A   Additional figures - Results of stress tests

For an overview of the performed stress tests, see **Table 25**. In the appendix the plots as described in **Section 8.11** are listed.
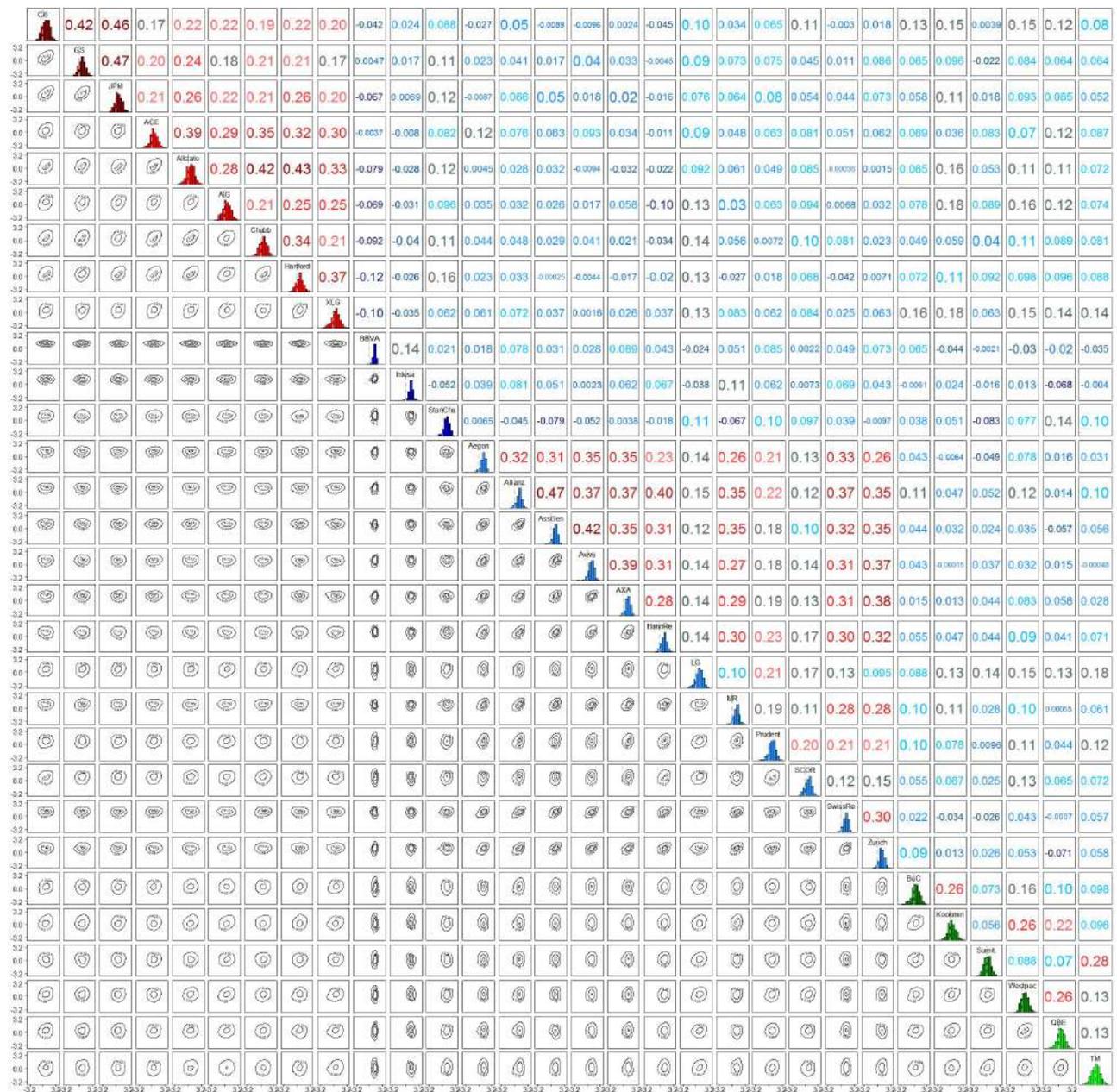
Stress test 1

Figure A.1: Z-Level plot and Kendall's τ values: Performance of the world market in a moderate stress situation (0.9) induced by European systemic banks
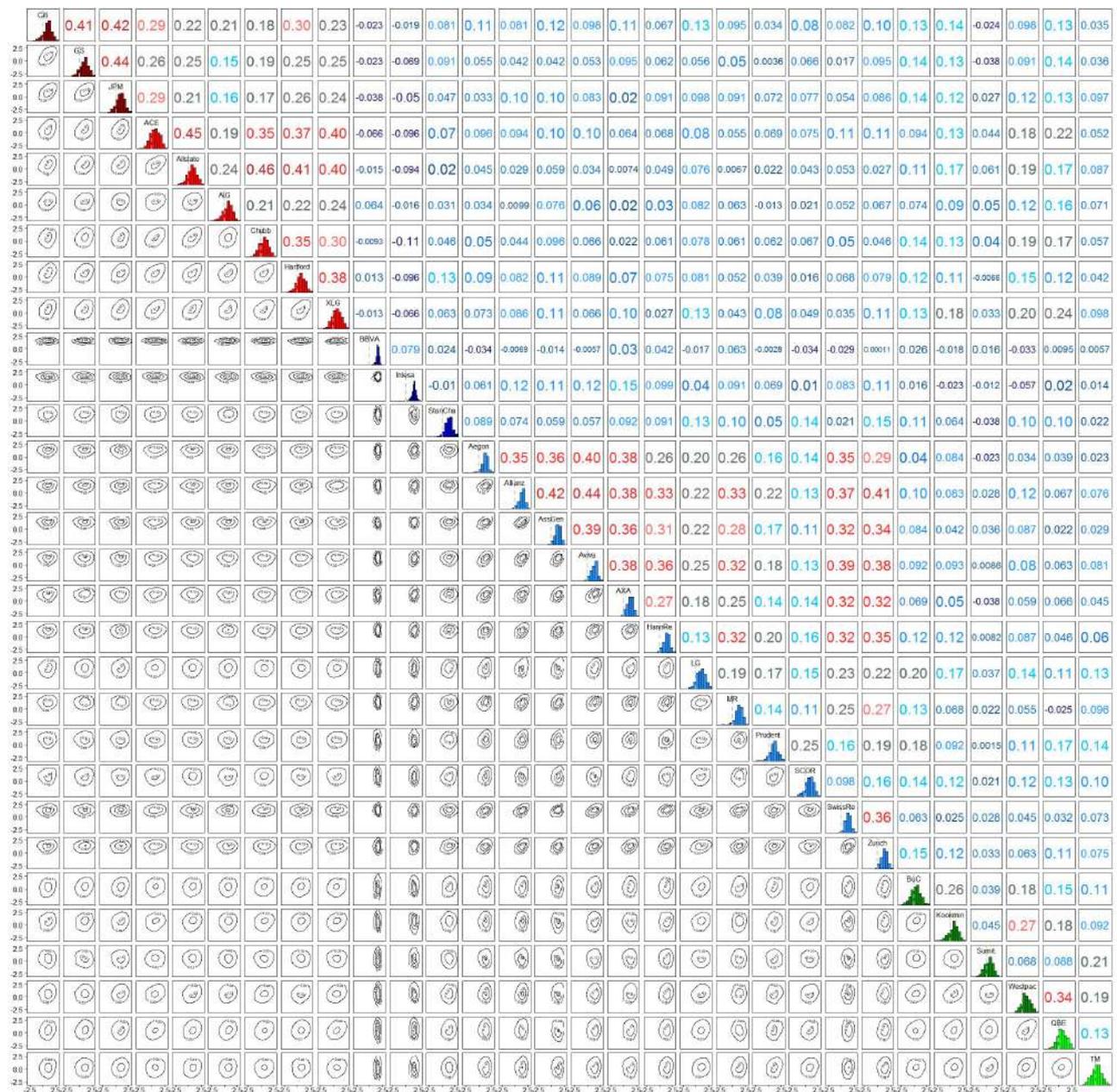
Stress test 2

Figure A.2: Z-Level plot and Kendall's τ values: Performance of the world market in a severe stress situation (0.95) induced by European systemic banks
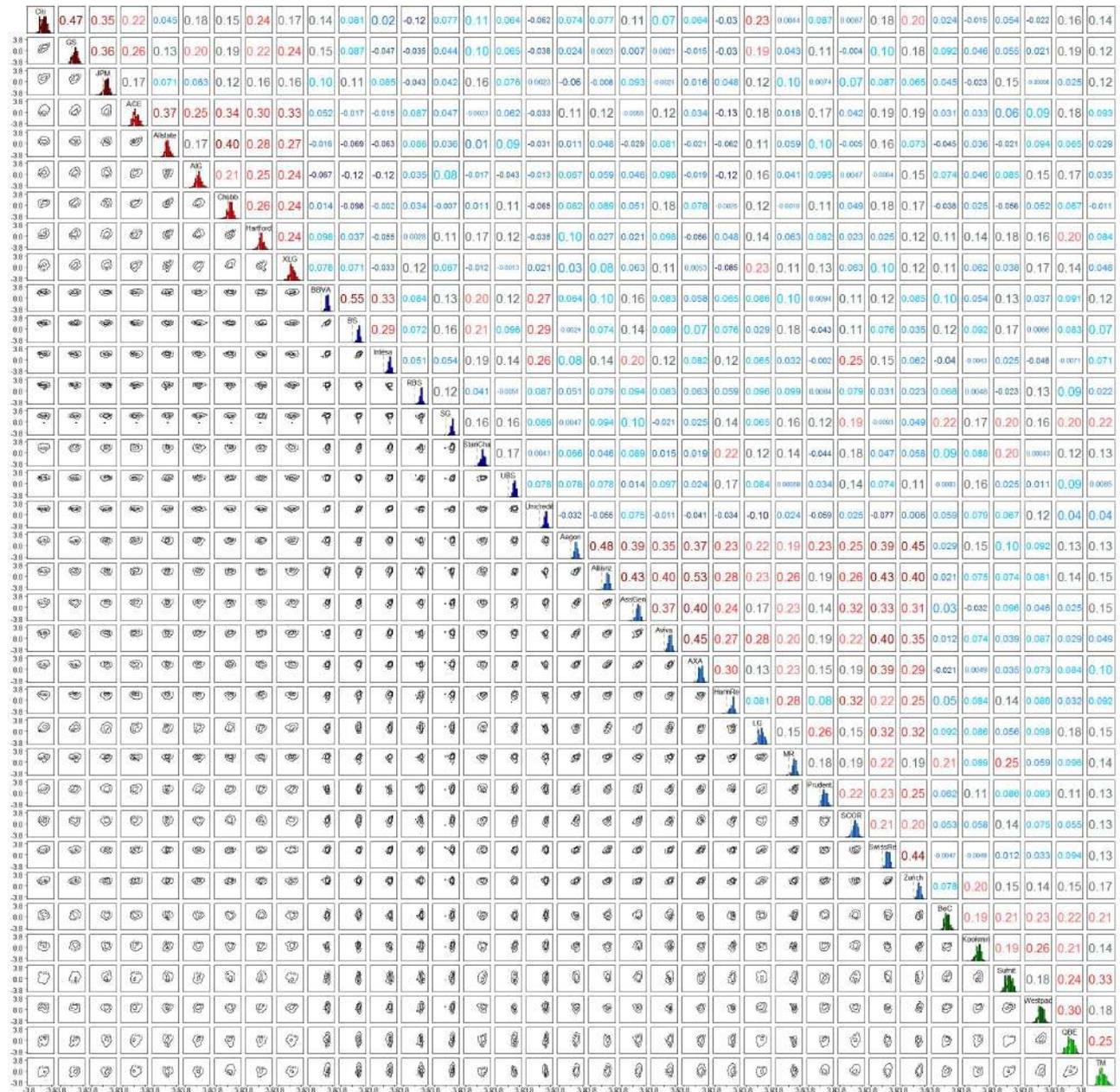
Stress test 3

Figure A.3: Z-Level plot and Kendall's τ values: Performance of the world market in a moderate stress situation (0.9) induced by North American systemic banks

Stress test 4

Figure A.4: Z-Level plot and Kendall's τ values: Performance of the world market in a severe stress situation (0.95) induced by North American systemic banks

Stress test 5

Figure A.5: Z-Level plot and Kendall's τ values: Performance of the world market in an extreme stress situation (0.99) induced by North American systemic banks

Stress test 6

Figure A.6: Z-Level plot and Kendall's τ values: Performance of the world market in a moderate stress situation (0.9) induced by major European banks
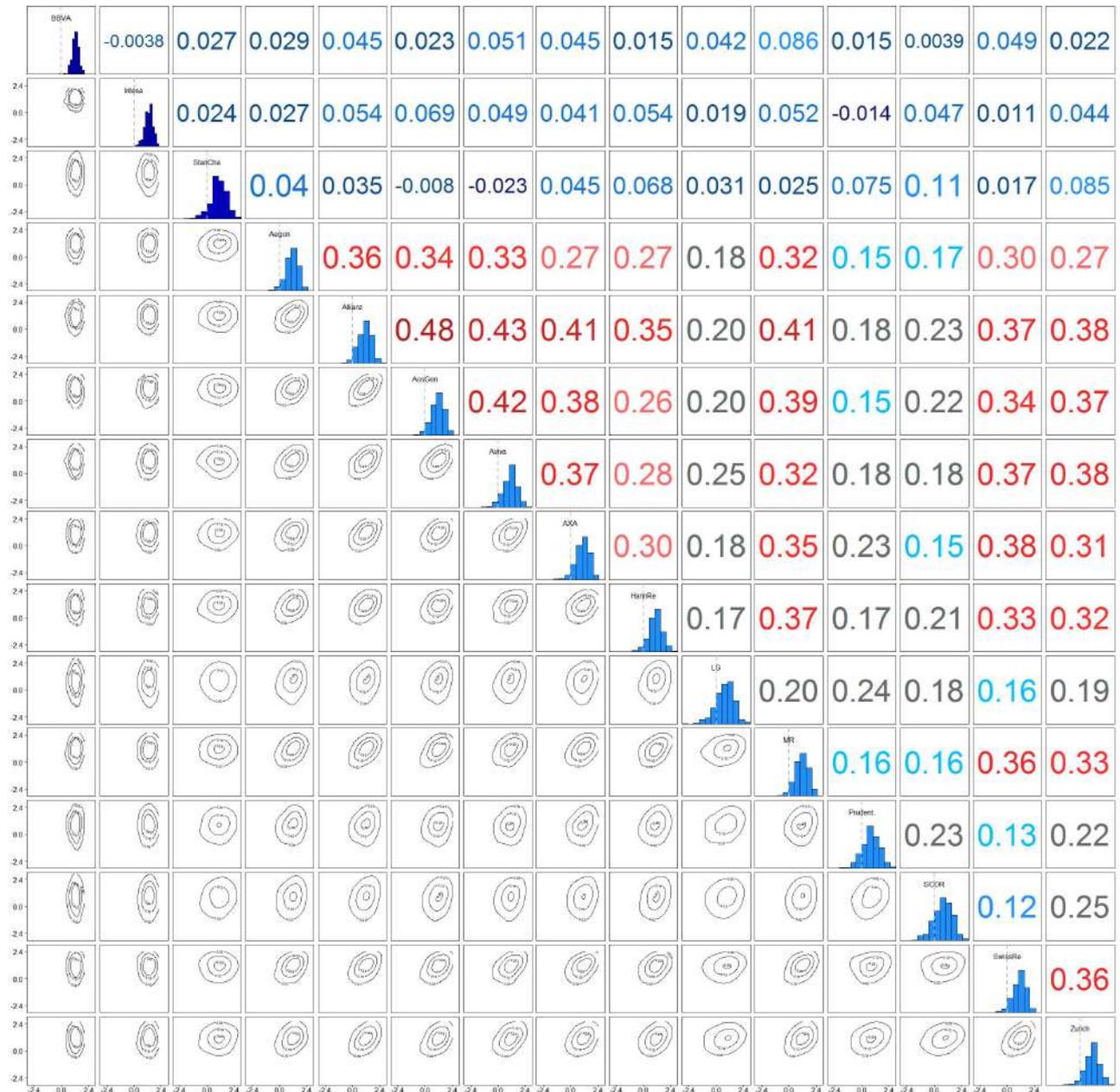
Stress test 7

Figure A.7: Z-Level plot and Kendall's τ values: Performance of the world market in a severe stress situation (0.95) induced by major European banks

Stress test 8

Figure A.8: Z-Level plot and Kendall's τ values: Performance of the world market in an extreme stress situation (0.99) induced by major European banks

Stress test 9

Figure A.9: Z-Level plot and Kendall's τ values: Performance of the European market in a moderate stress situation (0.9) induced by European systemic banks
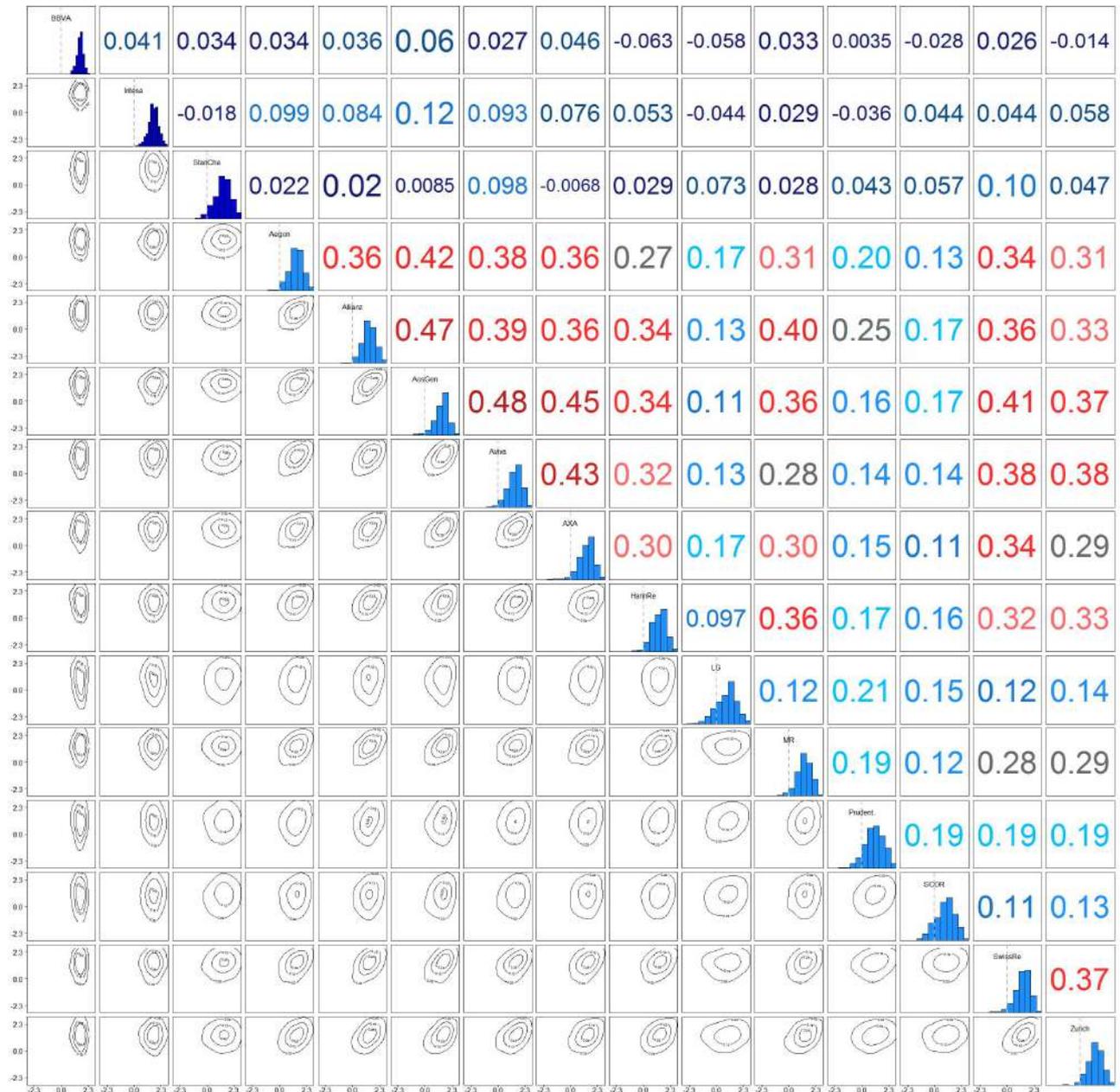
Stress test 10

Figure A.10: Z-Level plot and Kendall's τ values: Performance of the European market in a severe stress situation (0.95) induced by European systemic banks
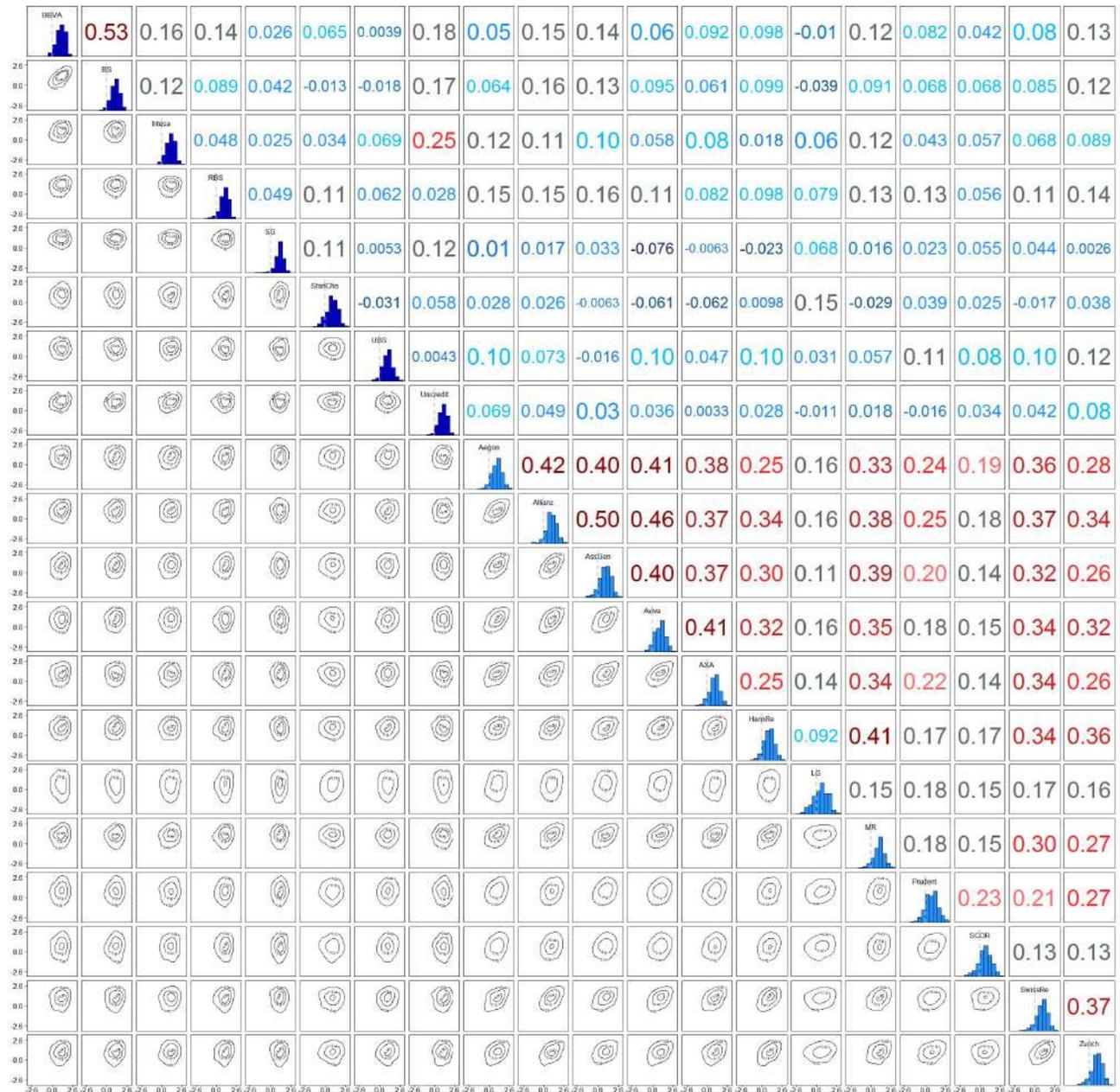
Stress test 11

Figure A.11: Z-Level plot and Kendall's τ values: Performance of the European market in a moderate situation (0.9) induced by biggest European banks
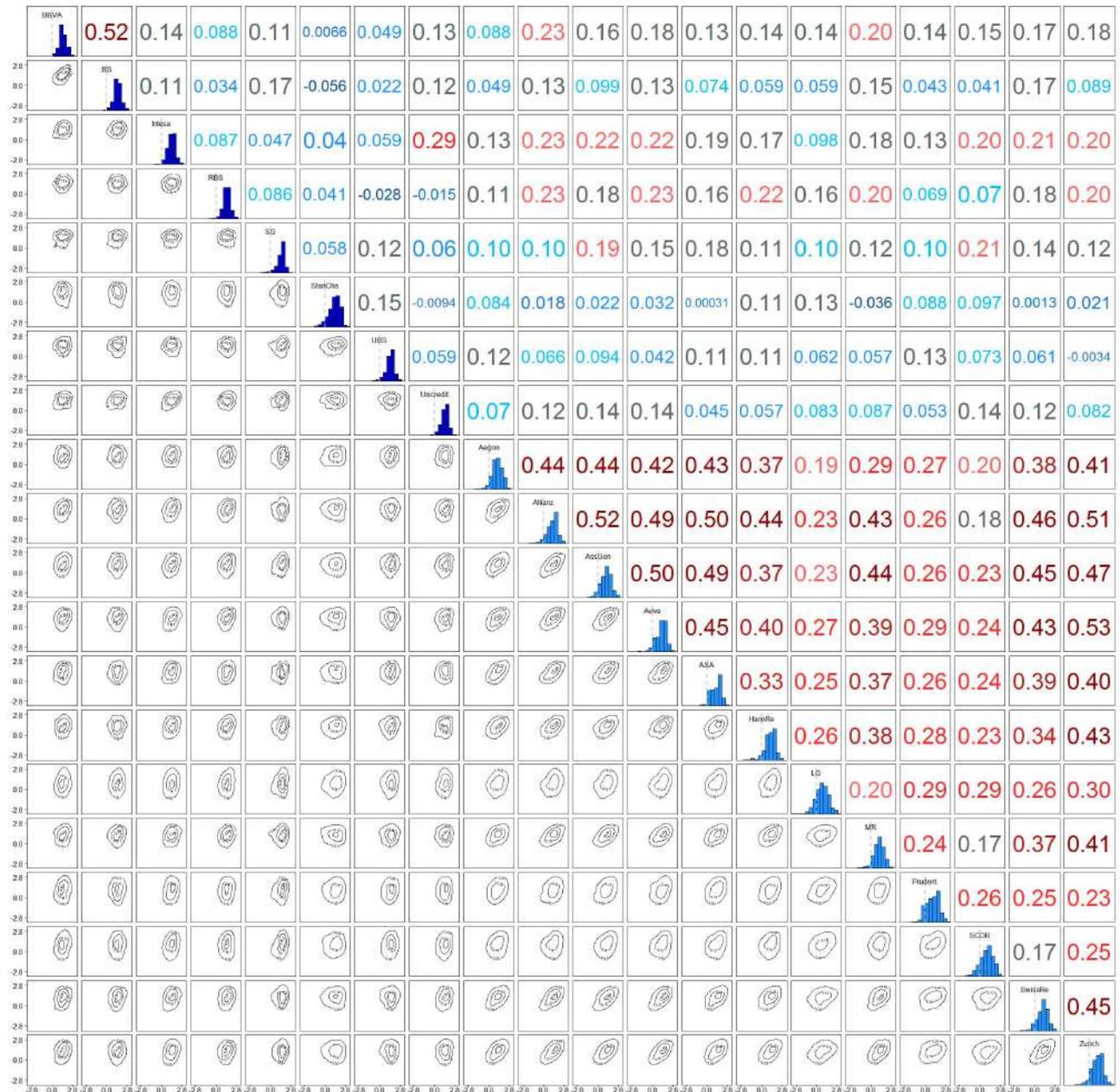
Stress test 12

Figure A.12: Z-Level plot and Kendall's τ values: Performance of the European market in a severe stress situation (0.95) induced by biggest European banks
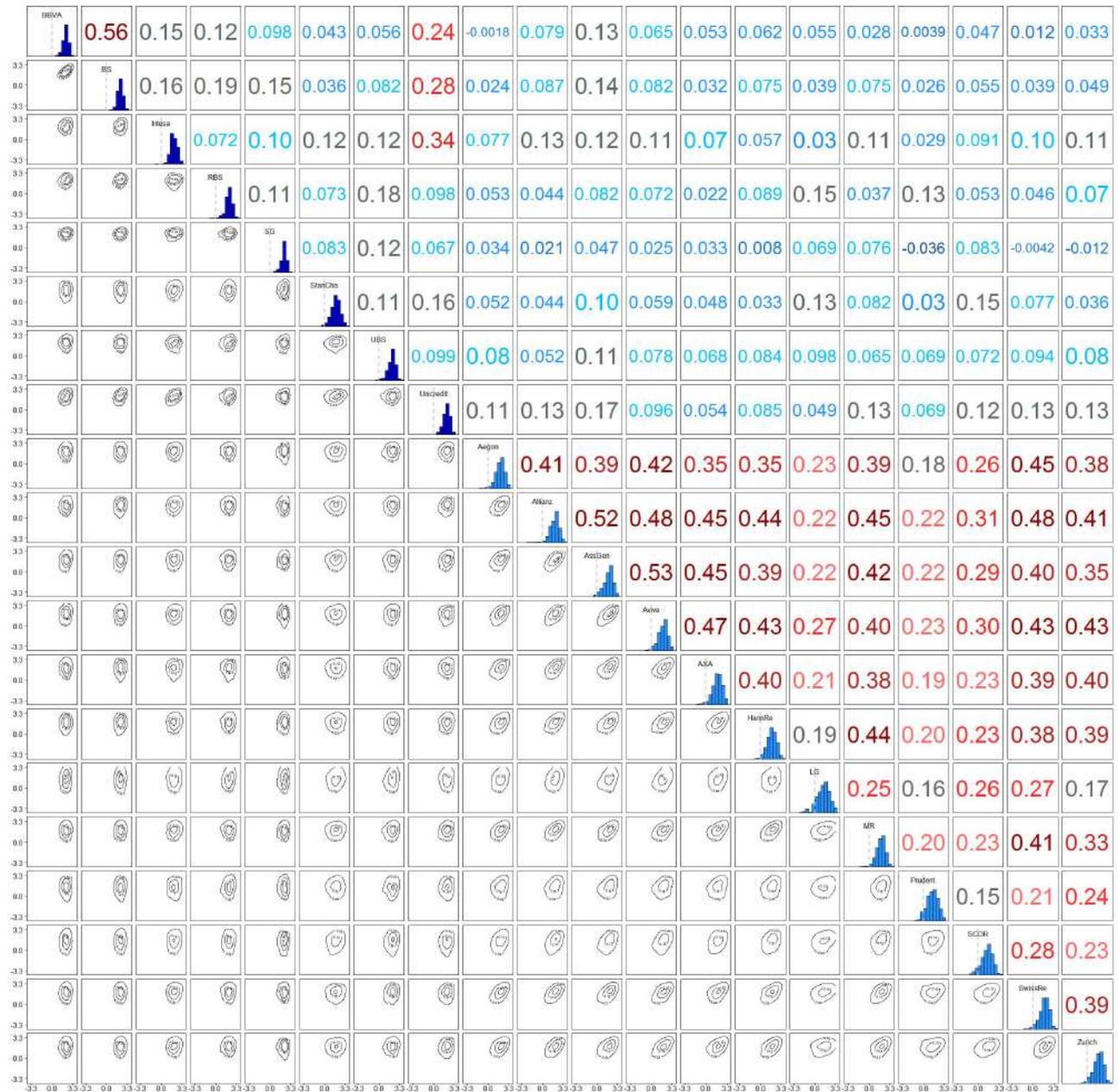
Stress test 13

Figure A.13: Z-Level plot and Kendall's τ values: Performance of the European market in an extreme stress situation (0.99) induced by biggest European banks
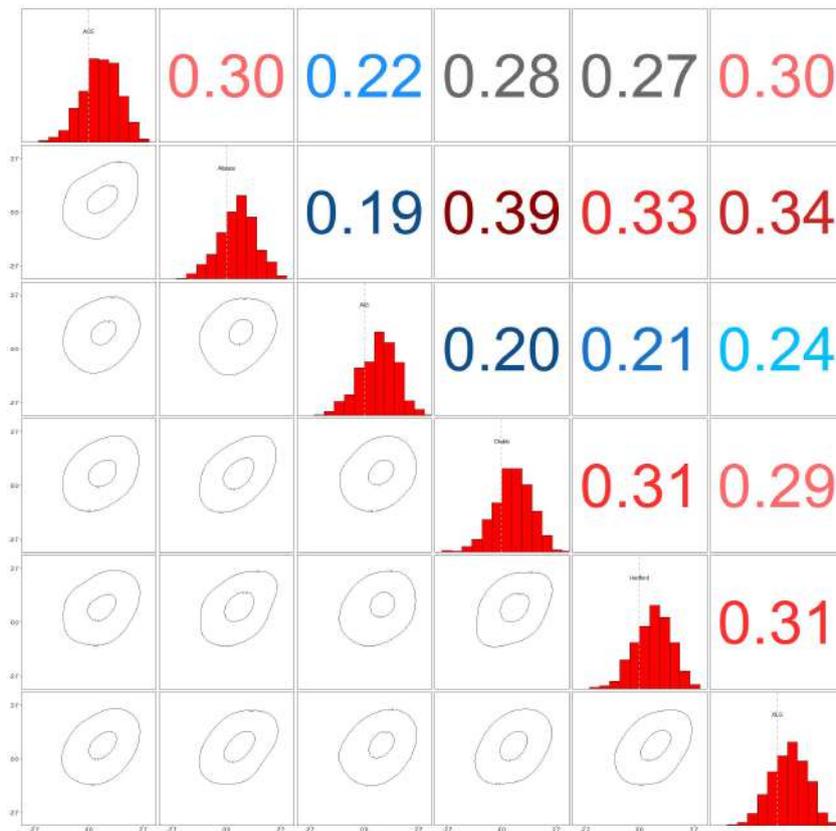
Stress test 14

Figure A.14: Z-Level plot and Kendall's τ values: Performance of the North American market in a moderate stress situation (0.9) induced by North American systemic banks

Stress test 15
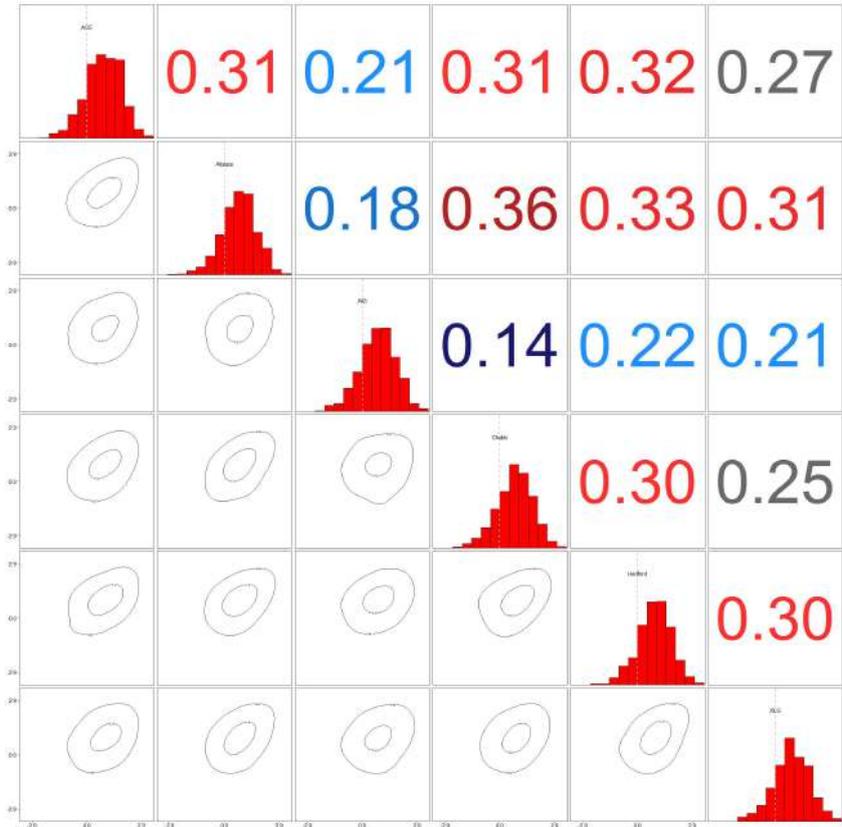

Figure A.15: Z-Level plot and Kendall's τ values: Performance of the North American market in a severe stress situation (0.95) induced by North American systemic banks
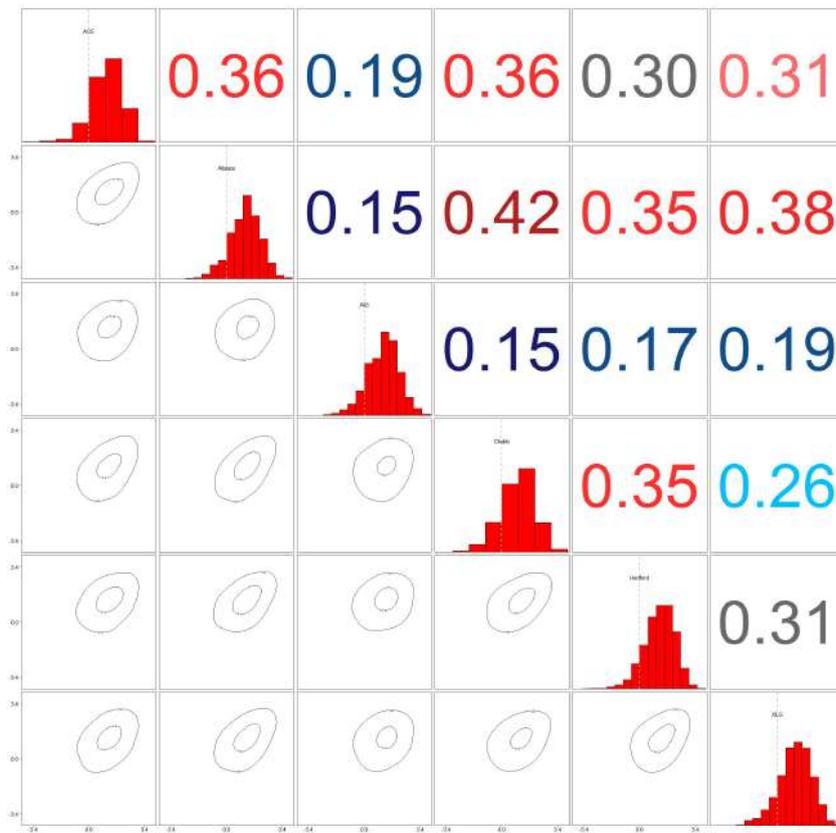
Stress test 16

Figure A.16: Z-Level plot and Kendall's τ values: Performance of the North American market in an extreme stress situation (0.99) induced by North American systemic banks
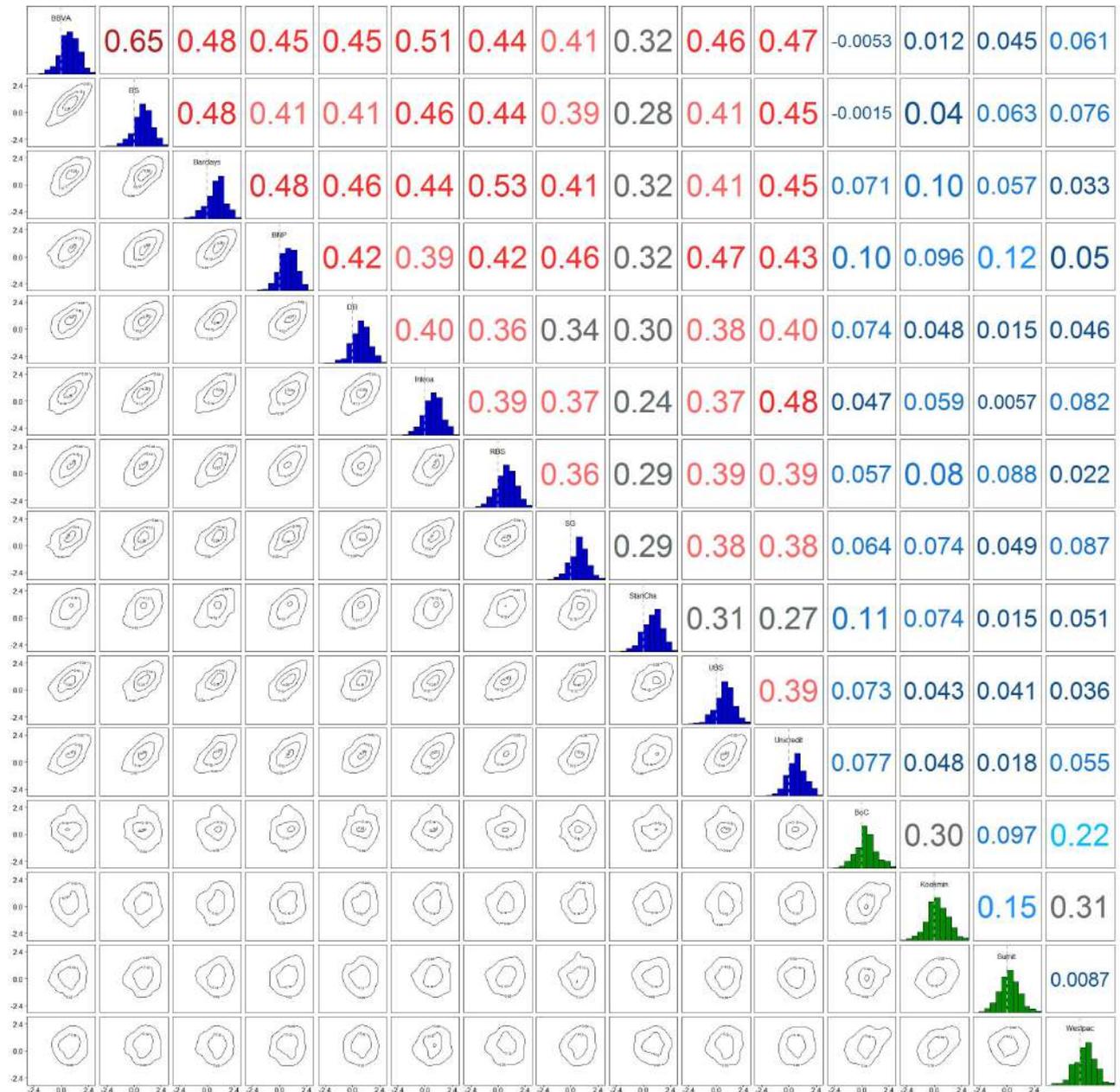
Stress test 17

Figure A.17: Z-Level plot and Kendall's τ values: Performance of the banking market in a moderate stress situation (0.9) induced by North American systemic banks

Stress test 18

Figure A.18: Z-Level plot and Kendall's τ values: Performance of the banking market in a severe stress situation (0.95) induced by North American systemic banks

Stress test 19

Figure A.19: Z-Level plot and Kendall's τ values: Performance of the banking market in an extreme stress situation (0.99) induced by North American systemic banks

Stress test 20

Figure A.20: Z-Level plot and Kendall's τ values: Performance of the banking market in a moderate stress situation (0.9) induced by European systemic banks
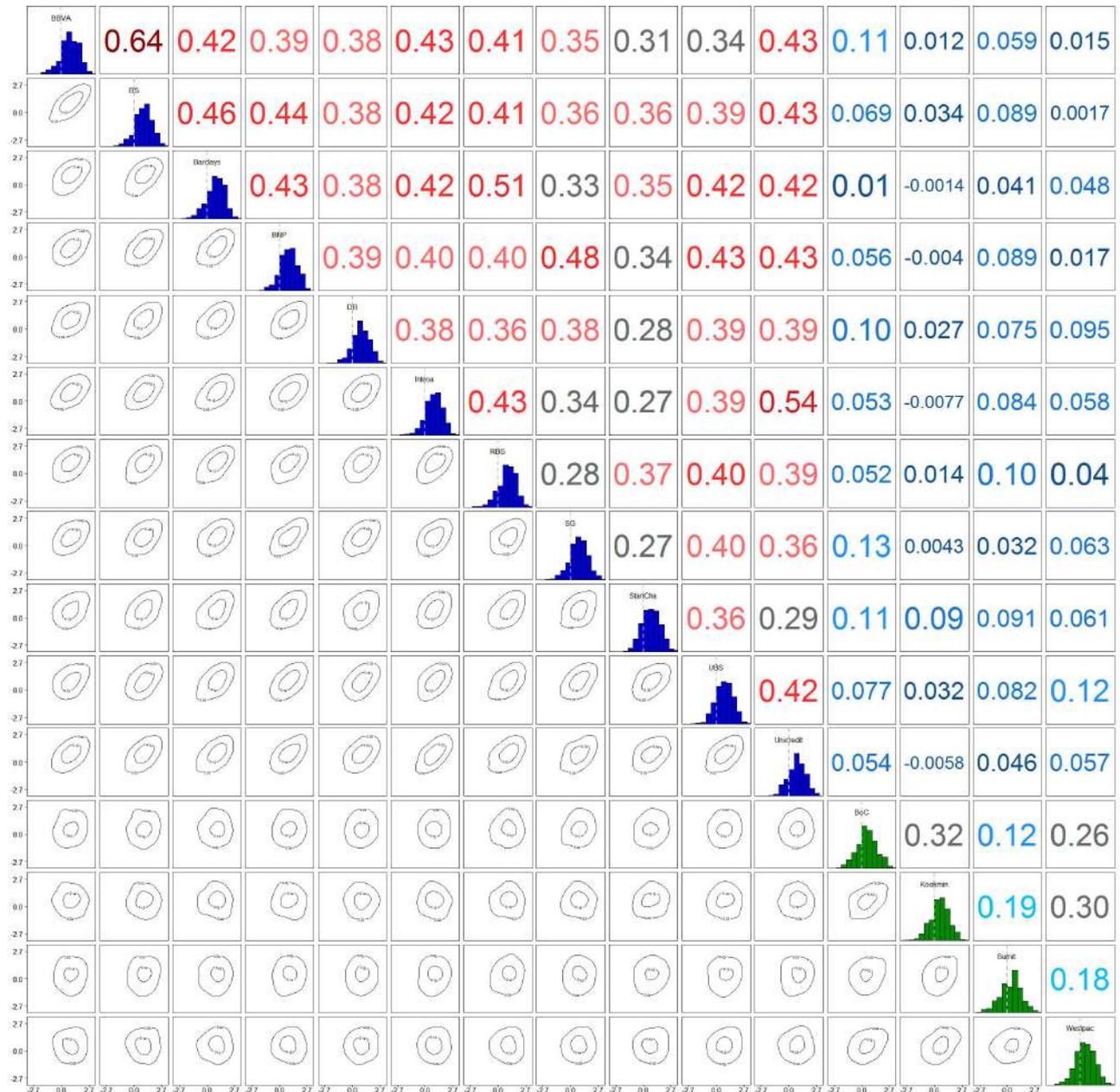
Stress test 21
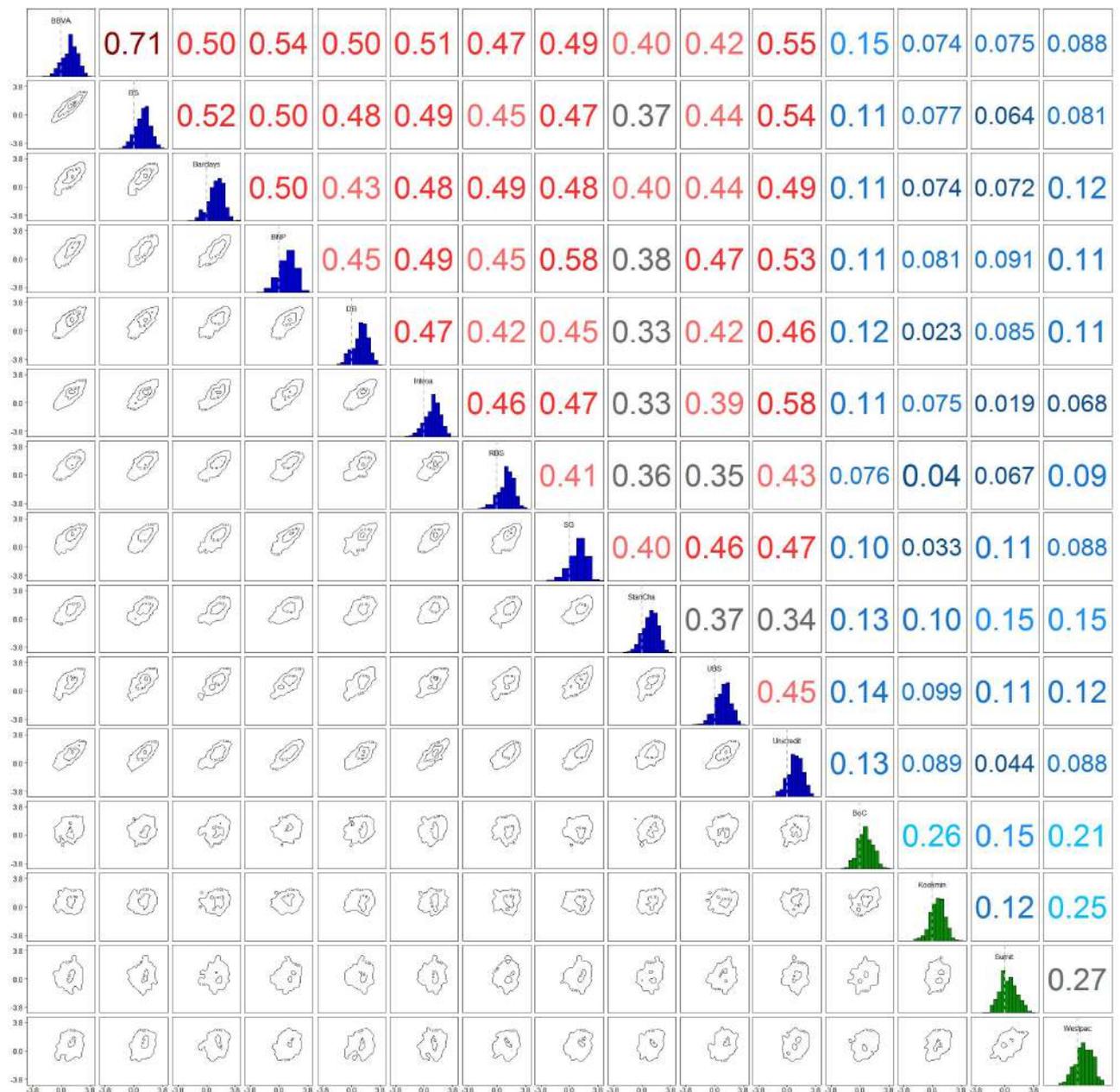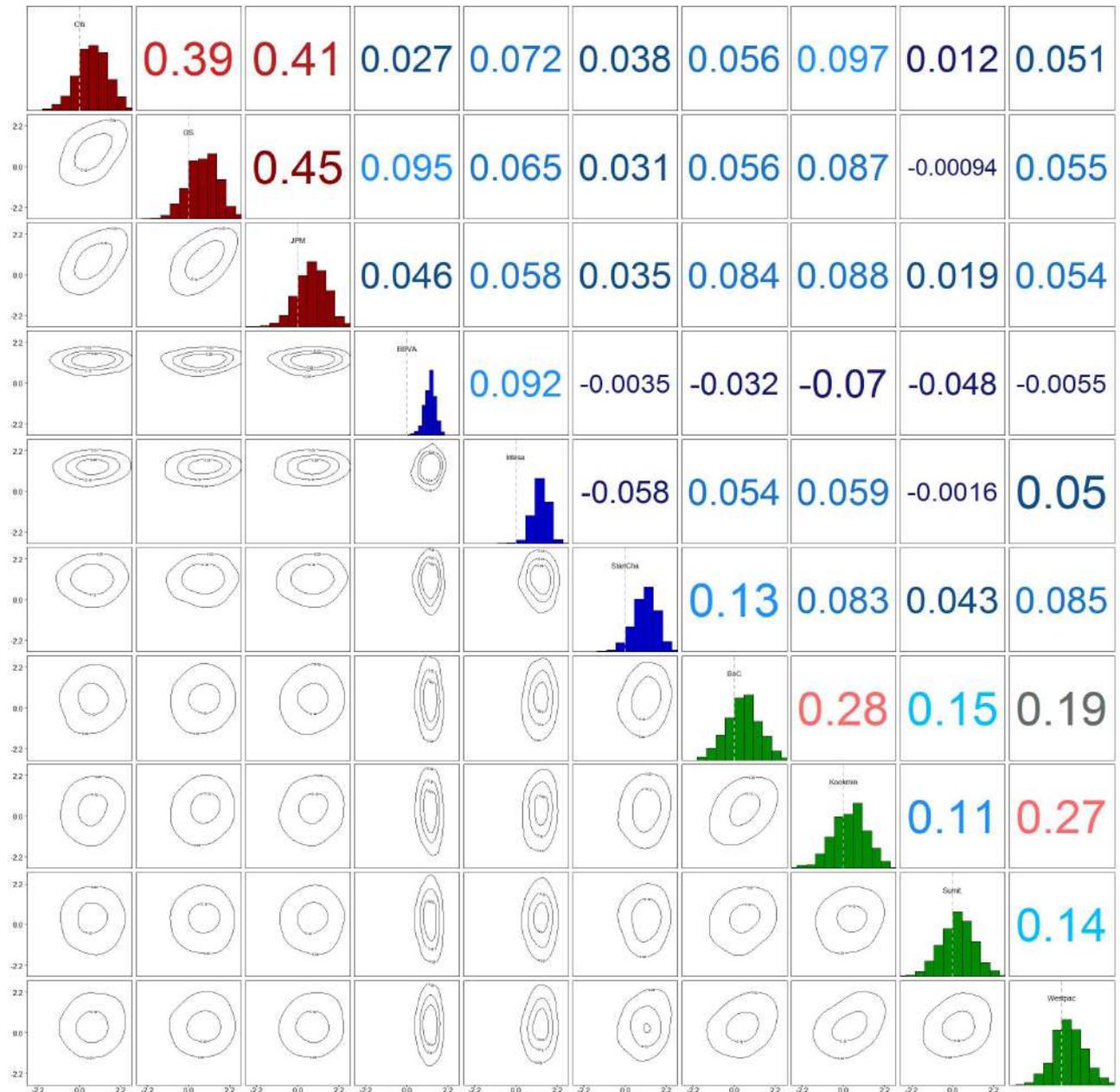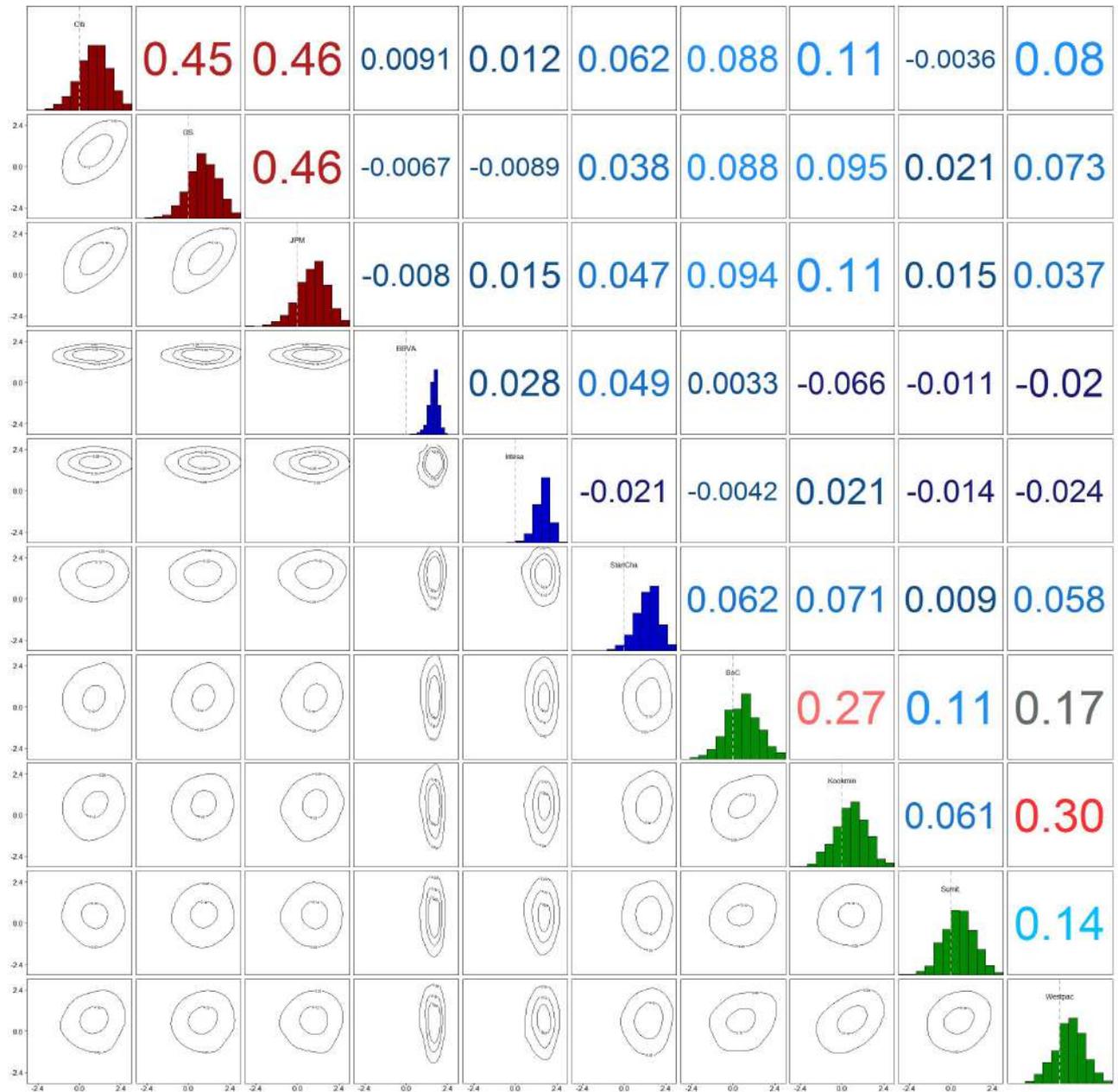

Figure A.21: Z-Level plot and Kendall's τ values: Performance of the banking market in a severe stress situation (0.95) induced by European systemic banks

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Andrieu, C. and Robert, C. P. (2001). Controlled mcmc for optimal sampling. Technical report.

Atchad, Y., Fort, G., Moulines, E., and Priouret, P. (2009). Adaptive markov chain monte carlo: Theory and methods.

Bai, Y., Craiu, R., and Narzo, A. F. D. (2011a). Divide and conquer: A mixture-based approach to regional adaptation for mcmc.

Bai, Y., Roberts, G. O., and Rosenthal, J. S. (2011b). On the containment condition for adaptive markov chain monte carlo algorithms. *Adv. Appl. Stat.*, 21:1–54.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068.

Berntsen, J., Espelid, T. O., and Genz, A. (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software (TOMS)*, 17(4):437–451.

Bindel, D. and Goodman, J. (2009). Principles of scientific computing.

Bluhm, M. and Krahnen, J. P. (2011). Default risk in an interconnected banking system with endogeneous asset markets. Technical report, CFS working paper.

Board, F. S. (2009). Guidance to assess the systemic importance of financial institutions, markets and instruments: initial considerations. *Report to G20 Finance Ministers and Governors.*

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Brechmann, E. C., Hendrich, K., and Czado, C. (2013). Conditional copula simulation for systemic risk stress testing. *Insurance: Mathematics and Economics*, 53(3):722–732.

Burke, O. (2012). Statistical methods: Autocorrelation: Mcmc output analysis. Technical report, University of Oxford.

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

Chu, Y.-S., Constantinou, N., and O'Hara, J. (2010). An analysis of the determinants of the itraxx cds spreads using the skewed student'st ar-garch model. *University of Essex-Centre for Computational Finance and Aconomic Agents Working Paper Series*, 40.

Colander, D., Haas, A., Goldberg, M., Kirman, A., Lux, T., and Sloth, B. The financial crisis and the systemic failure of academic economics. *Lessons from the Financial Crisis: Causes, Consequences, and Our Economic Future*, pages 427–436.

Conrad, B. (2005). Impossibility theorems for elementary integration. In *Academy Colloquium Series. Clay Mathematics Institute, Cambridge, MA*.

Cont, R. and Kan, Y. H. G. (2011). Statistical modeling of credit default swap portfolios. *Available at SSRN 1771862*.

Cont, R., Moussa, A., et al. (2010). Network structure and systemic risk in banking systems. *Edson Bastos e, Network Structure and Systemic Risk in Banking Systems (December 1, 2010)*.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904.

Czado, C., Brechmann, E., and Gruber, L. (2013). Selection of vine copulas. In Jaworski, P., Durante, F., and Härdle, W. K., editors, *Copulae in mathematical and quantitative finance*, Lecture Notes in Statistics, pages 17–37. Springer Berlin Heidelberg.

Czado, C., Schepsmeier, U., and Min, A. (2012). Maximum likelihood estimation of mixed c-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255.

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

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Drehmann, M. and Tarashev, N. (2013). Measuring the systemic importance of interconnected banks. *Journal of Financial Intermediation*, 22(4):586–607.

ECB (2013). A macro stress-testing framework for bank solvency analysis. Technical report, ECB.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368.

Geyer, C. J. (2012). *Stationary Stochastic Processes*, chapter Stationary Processes.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1998). *Markov chain Monte Carlo in practice*, volume 17. Chapman and Hall.

Gräler, B. and Pebesma, E. (2011). The pair-copula construction for spatial data: a new approach to model spatial dependency. *Procedia Environmental Sciences*, 7:206–211.

Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hendrich, K. (2012). Copula-based analysis of interdependence among companies in the banking and insurance sector. Master's thesis, TUM.

H.Greene, W. (2003). *Econometric Analysis*. New York University.

Holden, L. (1998). Geometric convergence of the metropolis-hastings simulation algorithm. *Statistics & probability letters*, 39(4):371–377.

Hoshimura, K. (2007). Covariance matrix adaptation evolution strategy for constrained optimization problem. Master's thesis, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University.

Huang, X., Zhou, H., and Zhu, H. (2009). A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance*, 33(11):2036–2049.

Kim, D., Kim, J.-M., Liao, S.-M., and Jung, Y.-S. (2013). Mixture of d-vine copulas for modeling dependence. *Computational Statistics & Data Analysis*, 64:1–19.

Kou, J. and Varotto, S. (2005). Predicting agency rating migrations with spread implied ratings. Technical report, Henley Business School, Reading University.

Kurrowicka, D. and Joe, H., editors (2011). *Dependence modeling: Vine copula handbook*. World Scientific.

Luengo, D. and Martino, L. (2013). Fully adaptive gaussian mixture metropolis-hastings algorithm. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6148–6152. IEEE.

Markose, S., Giansante, S., Gatkowski, M., and Shaghaghi, A. R. (2009). Too interconnected to fail: Financial contagion and systemic risk in network model of cds and other credit enhancement obligations of us banks. *analysis*.

Morales-Napoles, O., C. R. K. D. (2010). About the number of vines and regular vines on n nodes. *Discrete Applied Mathematics*.

Mueller, C. and Sbalzarini, I. (2010). Gaussian adaptation as a unifying framework for continuous black-box optimization and adaptive monte carlo sampling. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–8.

Oehmke, M. and Zawadowski, A. (2013). The anatomy of the cds market. *Columbia University and Boston University Working Paper*.

Risk, M. S. and Soundness, R. F. (2009). Global financial stability report.

Roberts, G. O., Rosenthal, J. S., et al. (2004). General state space markov chains and mcmc algorithms. *Probability Surveys*, 1:20–71.

Rosenthal, J. S. (2010). Optimal proposal distributions and adaptive mcmc.

Sahlin, K. (2011). Estimating convergence of markov chain monte carlo simulations. *Stockholm University, Master Thesis*.

Schmidl, D., Czado, C., Hug, S., Theis, F. J., et al. (2013). A vine-copula based adaptive mcmc sampler for efficient inference of dynamical systems. *Bayesian Analysis*, 8(1):1–22.

Tor, M. and Sarfraz, S. (2013). Largest 100 banks in the world. *SNL*.

Tutte, W. T. (1984). Graph theory, volume 21 of encyclopedia of mathematics and its applications.

Varin, C., Reid, N. M., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.