Technische Universität München

Zentrum Mathematik

Bayesian Analysis of Multivariate Time Series Models based on Pair-Copula Construction

Diplomarbeit

von

Florian Gärtner

Themenstellerin:	Prof. Dr. Claudia Czado
Betreuer:	Dr. Aleksey Min
Abgabetermin:	23. September 2008

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

Garching, den 23. September 2008

Acknowledgments

First, let me thank Prof. Dr. Claudia Czado for her excellent mentoring and valuable support throughout the whole developing process of this diploma thesis.

In addition, I owe thanks to Dr. Aleksey Min who did not spare any efforts to answer my often tricky questions and to help me with his expert guidance. He was always willing to listen to me and help me solve my encountered problems.

Special thanks goes to my friend Linda who helped me inestimably with all questions concerning the English language.

Furthermore, I would like to thank my parents who unremittingly supported me during my years of study. They made this work possible. Especially, I want to mention my mother Renate who supported me with the necessary accuracy in correcting produced typing errors.

Finally, I also want to thank my fiancée Katrin for her encouragement, understanding, patience and love during my whole studies and particularly during the difficult periods of this work.

Zusammenfassung

Diese Diplomarbeit handelt von der gemeinsamen bayesianischen Schätzung von marginalen Zeitreihen- und Copula-Parametern eines D-Vines, der aus bivariaten t-Copulas besteht.

Zuerst wird die notwendige Theorie für diese Arbeit erläutert. Dies geschieht in Form von grundlegenden Sätzen der mathematischen Statistik. Dann werden die verwendeten ein- und mehrdimensionalen Wahrscheinlichkeitsverteilungen und die Abhängigkeitsmaße definiert. Nach der Einführung der Grundlagen der Zeitreihentheorie werden Copulas und die damit eng verknüpften Pair-Copula Zerlegungen und D-Vines erläutert. Auch die Erzeugung von Zufallsvariablen bzw. die Bestimmung des Likelihoods wird gezeigt. Am Ende des zweiten Kapitels werden noch die Markov Chain Monte Carlo Verfahren und die verwendeten Schätzmethoden vorgestellt.

Im dritten Kapitel steht eine kurze Einführung in die Theorie der australischen Strom- und Ladungsmärkte, bevor die konkret beobachteten Daten beschrieben und analysiert werden. Es handelt sich um je 1142 Beobachtung der durchschnittlichen täglichen Ladungsmenge in den australischen Bundesstaaten Queensland, New South Wales, Victoria und South Australia. Anschließend werden mögliche Vorbearbeitungsmöglichkeiten gezeigt und auf die beobachteten Daten angewendet. Die beste dieser Methode wird durchgeführt, um eine Stationarität der Beobachtungen und die Modellierung durch einen AR(1)-Prozess zu errreichen.

Das vierte Kapitel enthält die Entwicklung des Modells für den Likelihood bzw. für die A-posteriori-Verteilung der einzelnen zu schätzenden Parameter. Diese Ergebnisse werden in einer Simulationsstudie getestet, ehe alle möglichen Pärchen der beobachteten vierdimensionalen und vorbearbeiteten Daten geschätzt werden. Hierbei zeigt sich, dass die Simulationsergebnisse sehr nahe bei den entsprechenden Werten eines Maximum Likelihood Schätzers liegen.

Im fünften Kapitel wird das Modell auf vier Dimensionen erweitert, wobei eine Pair-Copula Zerlegung mit t-Copulas verwendet wird. Die Simulationsstudie zeigt sehr gute Schätzungen, die jedoch auch eng bei den ML-Schätzern liegen. Nun wenden wir unseren Algorithmus auf die beobachteten vierdimensionalen australischen Ladungsdaten an. Hierbei tritt beim Abhängigkeitsparameter der AR(1)-Zeitreihe der Fall auf, dass für

New South Wales, Victoria und South Australia der ML-Wert nicht mehr im Bayes'schen Konfidenzintervall liegt. Weiterhin vergleichen wir die gemeinsame Schätzung mit einer Schätzung in zwei Schritten, wobei zuerst die marginalen Zeitreihenparameter per MLE und dann die Copula-Parameter via MCMC geschätzt werden. Die Ergebnisse hierbei sind sehr eng beieinander. Wir validieren innerhalb vier verschiedener Schätzmethoden die Anpassung an die beobachteten Daten mit einem mehrdimensionalen Cramér-Test und kommen zu dem Ergebnis, dass ein reduziertes D-Vine-Modell mit t-Copulas am besten passt. Schließlich schätzen wir noch bayesianisch einen D-Vine aus t-Copulas, einen D-Vine aus Normal-Copulas und eine vierdimensionale t-Copula jeweils mit AR(1)-Marginalien in einem kompletten und in einem reduzierten Modell und berechnen die A-posteriori-Modell-Wahrscheinlichkeiten nach Congdon (2006). Hierbei ergibt sich, dass das reduzierte Modell einer vierdimensionalen t-Copula mit AR(1)-Marginalien die höchste Wahrscheinlichkeit hat.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Modeling multivariate dependency is an important task in statistics. However, the observed data often follows any kind of additional marginal structure. Therefore, the modeling is very complicated. With the introduction of the concept of copulas by Sklar (1959) based on the works of Hoeffding and Fréchet in the 1940s and the early 1950s, it is possible to model the marginal structure and the dependency separately. A good overview of copulas can be found in Nelsen (2006). Instead of using a higher-dimensional copula which is often very complicated to fit, Bedford and Cooke (2001, 2002) introduced a possible decomposition into bivariate building blocks, the pair copulas, and its margins. Hereby, you decompose the joint density function into a cascade of bivariate copulas of the original data and their conditional and unconditional distribution functions. This pair-copula decomposition is a very flexible way of constructing multivariate dependency concepts. The graphical concept of vines, introduced by Bedford and Cooke (2001, 2002), helps with the modeling and the determination of the necessary copulas. Aas et al. (2007) give a good overview of the theory of pair-copula decomposition and show the application of this theory by denoting the sampling algorithms and the algorithm for determining the joint likelihood. Very important for the application to real data is the multivariate Student t-copula, e.g. described in Demarta and McNeil (2005). A Bayesian analysis of the copula approach is done by Huard et al. (2006) and Dalla Valle (2007).

We want to model our marginals with an autoregressive structure. There is a great variety of literature dealing with fitting autoregressions, both frequentist methods, e.g. Box and Jenkins (1976), as well as Bayesian methods, e.g. Zellner (1971, Chapter 7), Box and Jenkins (1976, p.250), Monahan (1983), Marriott and Smith (1992) and Barnett et al. (1996).

If the marginal structure as well as the dependency is considered, this is often done by applying a copula approach since the marginal densities can be defined separately from the dependency structure. However, most of the proposed methods are two-step approaches whereby they do not account for the parameter uncertainty of time series and copula parameters simultaneously. Pitt et al. (2006) and Silva and Lopes (2008) consider a copula approach with marginal modeling in a Bayesian setting, but Silva and Lopes (2008) for example fix the parameter for the degrees of freedom for the t-copula and consider only

one-parametric families. Ausin and Lopes (2008) perform a joint Bayesian estimation for a copula GARCH model.

Our approach is now to estimate jointly the marginal time series and the copula parameters with an Markov Chain Monte Carlo method. We use the Metropolis-Hastings algorithm with independence proposal density (cf. Tierney (1994)). Therefore, we develop our model at first in two dimensions. We perform a simulation study and try to fit the copula AR(1) model to real data. We choose Australian load data, namely daily observations for the states of Queensland, New South Wales, Victoria and South Australia. After having preprocessed these time series to reach stationarity, we fit marginal AR(1) models and a t-copula to each possible pair out of this set. Then we generalize our algorithm to four dimensions using the pair copula decomposition with a D-vine of t-copulas. In this setting, we perform a simulation study before we jointly estimate the time series and D-vine pair copula parameters for the real data, the Australian load data. Since a lot of methods described in literature take a two-step approach by estimating the marginal parameters first and the copula parameters afterwards, we want to compare this two-step method (marginal parameters estimated first by MLE, then MCMC for the time series parameters) to the joint MCMC estimation of all parameters. Another method widely used in econometrics is the vector autoregressive model. Since we encounter big difficulties in comparing multivariate time series models to marginal time series and copula models for the dependency, we try an internal validation of the estimation methods by using a multivariate Cramér test. We also test this for the joint MCMC of the marginal times series parameters D-vine pair-copula decomposition parameters, for a reduced joint MCMC where the parameters of the conditional copulas in the D-vine are set to $\nu = 100$ and $\rho = 0$ and for a marginal AR(1) model and a four-dimensional t-copula for the dependence structure. Finally, we perform a Bayesian model selection using the method of Congdon (2006). The marginal time series parameters in all cases are estimated in the same way. The different methods vary only in the modeling of the dependency. We compare a D-vine of bivariate t-copulas and its reduced model, a D-vine of bivariate normal copulas and its reduced model and a four-dimensional t-copula and its reduced model. We estimate in each case with a MCMC and compare all these methods on the basis of posterior model probabilities.

This diploma thesis is organized as follows:
In Chapter 2, we introduce some basic theorems of mathematical statistics, the used distributions and dependency measures as well as the theory of time series, copulas, pair-copula decomposition and D-vine and Markov Chain Monte Carlo method.
In Chapter 3, we describe the Australian electricity market and we perform an exploratory data analysis of the observed load data. Then we show different methods for preprocessing the data to get stationarity and we choose the best preprocessing for fitting afterwards a marginal AR(1) model.
Chapter 4 contains the explanation of the two-dimensional modeling, especially the procedure for the marginal time series. Then we perform a simulation study and show the estimation of real two-dimensional data.
In Chapter 5, we illustrate the basic procedures for the four-dimensional joint estimation

of the time series and pair-copula parameters. We implement a simulation study and estimate the parameters for the observed load data. Then we compare the joint Bayesian estimation to a two-step estimation and we perform an internal validation of different estimation methods. Finally, we run a Bayesian model selection for six different models. Chapter 6 contains a short overview and an outlook.

Chapter 2

Definitions and Basic Properties

2.1 Basic Ideas of Mathematical Statistics and Bayesian Inference

The basic idea of Bayesian Statistics consists of the point of view that not only the data X but also the parameters θ are random and follow an unknown distribution. So starting with the *likelihood* of the observed data as a function of θ, $l(\theta) = f(x|\theta)$, and a first view of the distribution of the parameters, the *prior distribution* $\pi(\theta)$, we want to calculate a more exactly distribution of θ, the *posterior distribution* $f(\theta|x)$, using the additional knowledge gained by the observed data. Very useful in this context is the *theorem of Bayes*, which also gave the name for this whole discipline of statistics.

Theorem 2.1 (Bayes theorem)

$$f(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)}$$

where $f(x) = \int f(x|\theta)\pi(\theta)\mathrm{d}\theta$.

This can be proved easily by applying the laws of conditional densities.
As the left-hand-side is a function of θ, we can shorten the formula as the data x as well as $f(x)$ are constant with regard to θ and so no essential part for determining the density of θ. So it remains

$$f(\theta|x) \propto l(\theta)\pi(\theta).$$

Very useful to generate random variables with a given distribution function is the *probability integral transform* introduced by Rosenblatt (cf. Rosenblatt (1952)).

Theorem 2.2 (Probability Integral Transform) *Let X be a real valued random variable and F(x) its continuous cumulative distribution function. Let U denote a random variable that is uniformly distributed on* $(0, 1)$. *Then Y=F(X) is uniformly distributed on* $(0, 1)$. *Furthermore,* $Z = F^{-1}(U)$ *has cdf F.*

Proof: cf. Angus (1994) □

The subsequent theorems will help to find the joint distribution of several functions of a continuous random vector. We will need this for the implementation of the Bayesian Inference. The presentation of this topic corresponds to the one in Bickel and Doksum (2001).
Let $\boldsymbol{h} = (h_1, \ldots, h_k)'$, where each h_i is a real-valued function on $\mathbb{R}^k$. Thus, $\boldsymbol{h}$ is a transformation from $\mathbb{R}^k$ to $\mathbb{R}^k$. Recall that the *Jacobian* $J_{\boldsymbol{h}}(\boldsymbol{t})$ of $\boldsymbol{h}$ evaluated at $\boldsymbol{t} = (t_1, \ldots, t_k)'$ is by definition the determinant

$$J_{\boldsymbol{h}}(\boldsymbol{t}) = \begin{vmatrix} \frac{\partial}{\partial t_1} h_1(\boldsymbol{t}) & \ldots & \frac{\partial}{\partial t_1} h_k(\boldsymbol{t}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial t_k} h_1(\boldsymbol{t}) & \vdots & \frac{\partial}{\partial t_k} h_k(\boldsymbol{t}) \end{vmatrix}.$$

A very important result of this section, the *transformation theorem* 2.4, is based on the following theorem.

Theorem 2.3 (change of variable theorem for multiple integrals) *Let $\boldsymbol{h} = (h_1, \ldots, h_k)'$ be a transformation defined on an open subset $\mathcal{B}$ of $\mathbb{R}^k$. Suppose that:*

(i) $\boldsymbol{h}$ *has continuous first partial derivatives in* $\mathcal{B}$.

(ii) $\boldsymbol{h}$ *is one-to-one on* $\mathcal{B}$.

(iii) The Jacobian of $\boldsymbol{h}$ *does not vanish on* $\mathcal{B}$.

Let f be a real-valued function (defined and measurable) on the range $\boldsymbol{h}(\mathcal{B}) = \{(h_1(\boldsymbol{t}), \ldots, h_k(\boldsymbol{t})) : \boldsymbol{t} \in \mathcal{B}\}$ of $\boldsymbol{h}$ and suppose f satisfies

$$\int_{\boldsymbol{h}(\mathcal{B})} |f(\boldsymbol{x})| \mathrm{d}\boldsymbol{x} < \infty.$$

Then for every (measurable) subset $\mathcal{K}$ of $\boldsymbol{h}(\mathcal{B})$ we have

$$\int_{\mathcal{K}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{h}^{-1}(\mathcal{K})} f(\boldsymbol{h}(\boldsymbol{t})) |J_{\boldsymbol{h}}(\boldsymbol{t})| \mathrm{d}\boldsymbol{t}.$$

Proof: cf. Apostol (1974, p. 421) □

Hereby is $\mathrm{d}\boldsymbol{x}$ an abbreviation for $\mathrm{d}x_1 \cdots \mathrm{d}x_k$ and $\boldsymbol{h}^{-1}$ denotes the inverse transformation of $\boldsymbol{h}$; that is, $\boldsymbol{h}^{-1}(\boldsymbol{x}) = \boldsymbol{t}$ if and only if $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{t})$. We also use the following calculation (cf. Apostol (1974, p. 417)),

$$J_{\boldsymbol{h}^{-1}}(\boldsymbol{t}) = \frac{1}{J_{\boldsymbol{h}}(\boldsymbol{h}^{-1}(\boldsymbol{t}))}.$$

Now we can formulate the important *transformation theorem.*

Theorem 2.4 (Transformation theorem) *Let $\boldsymbol{X}$ be continuous and let $\mathcal{S}$ be an open subset of $\mathbb{R}^k$ such that $P(\boldsymbol{X} \in \mathcal{S}) = 1$. If $\boldsymbol{g} = (g_1, \ldots, g_k)'$ is a transformation from $\mathcal{S}$ to $\mathbb{R}^k$ such that $\boldsymbol{g}$ and $\mathcal{S}$ satisfy the conditions of Theorem 2.3, then the density of $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ is given by*

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = p_{\boldsymbol{X}}(\boldsymbol{g}^{-1}(\boldsymbol{y}))|J_{\boldsymbol{g}^{-1}}(\boldsymbol{y})|$$

for $\boldsymbol{y} \in \boldsymbol{g}(\mathcal{S})$.

Proof: cf. Bickel and Doksum (2001, p. 486) □

Very useful for calculating the asymptotic distribution of a function of an asymptotic normal distributed sequence of random variables is the *Delta method*. For proofing this result we need the following theorem.

Theorem 2.5 (Slutsky's Theorem) *If $U_n \xrightarrow{\mathscr{D}} U$ and $V_n \xrightarrow{P} v_0$ (a constant), then*

(i) $U_n + V_n \xrightarrow{\mathscr{D}} U + v_0$,

(ii) $U_n V_n \xrightarrow{\mathscr{D}} v_0 U$.

Proof: cf. Bickel and Doksum (2001, p.467) □

A direct application of Slutsky's theorem yields the following corollary.

Corollary 2.6 *Suppose that a_n is a sequence of constants tending to ∞, b is a fixed number, and $a_n(Z_n - b) \xrightarrow{\mathscr{D}} X$. Let g be a function of a real variable that is differentiable and whose derivative g' is continuous at b. Then*

$$a_n \left[g(Z_n) - g(b)\right] \xrightarrow{\mathscr{D}} g'(b)X.$$

Now we can denote the one dimensional *Delta method*.

Theorem 2.7 *Let $X_1, \ldots, X_n$ be an i.i.d. real valued sequence of random variables, $h : \mathbb{R} \to \mathbb{R}$ a real valued function, $E(X_1^2) < \infty$ and h is differentiable at $\mu = E(X_1)$. Then follows for the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$*

$$\sqrt{n}\left(h(\bar{X} - h(\mu)\right) \xrightarrow{\mathscr{D}} Y$$

where

$$Y \sim N(0, [h'(\mu)]^2\sigma^2)$$

and $\sigma^2 = \mathrm{Var}(X_1)$.

Proof: follows with first order Taylor expansion and Slutsky's theorem (2.5).
cf. Bickel and Doksum (2001, p. 312) □

A quite easy, cheap (in terms of computational effort) and particularly numerically stable way to solve linear equations systems and to compute the square root of a symmetric and positive definite matrix is the *Cholesky decomposition*.

Theorem 2.8 *For each symmetric and positive definite matrix $A \in \mathbb{R}^{n\times n}$ there exists a unique lower triangular matrix $L \in \mathbb{R}^{n\times n}$ with positive diagonal elements so that $A = LL'$.*

Proof: cf. Kanzow (2005, p. 78) □

For a practical use of the Cholesky decomposition, we need to know the algorithm. Therefore at first some notation issues:

Let $A = LL'$ with a lower triangular matrix $L = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{pmatrix}$.

Now we can quote the algorithm (cf. Kanzow (2005, p. 80)).

Algorithm 2.1 Cholesky Decomposition for symmetric and positive definite matrices

1: **FOR** $j = 1 : n$ **DO**
2: $\quad l_{jj} := \sqrt{a_{jj} - \sum_{m=1}^{j-1} l_{jm}^2}$;
3: $\quad$ **FOR** $i = (j+1) : n$ **DO**
4: $\quad\quad l_{ij} := (a_{ij} - \sum_{m=1}^{j-1} l_{jm} l_{im})/l_{jj}$;
5: $\quad$ **END FOR**
6: **END FOR**

2.2 Univariate Distributions

The distributions presented in this section are used as prior distributions for the Bayesian analysis. At first we will clarify some notations: $X \sim F$ means that X is a random variable with distribution function F. $X \sim p$ says that p is the density of the random variable X.
This section is based on the presentation in Bickel and Doksum (2001).

2.2.1 Uniform distribution

The *Uniform distribution* on (a, b) is denoted by $\mathcal{U}(a, b)$. Its density function is

$$p(x) = \frac{1}{(b-a)},\ a < x < b$$

with (a, b) any pair of real numbers with $a < b$. The corresponding distribution function is given by

$$F(x) = \frac{(x-a)}{(b-a)} \text{ for } a < x < b.$$

If X has a $\mathcal{U}(a, b)$ distribution, then

$$E(X) = \frac{a+b}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}.$$

2.2.2 Normal distribution

The *Normal distribution* is denoted by $\mathcal{N}(\mu, \sigma^2)$ with location parameter μ and scale parameter σ. Its density function is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

where μ can be any real number while σ is positive.
The normal distribution with $\mu = 0$ and $\sigma = 1$ is known as the *standard normal distribution*. Its density will be denoted by $\varphi(x)$ and its distribution function by $\Phi(x)$.
If X has a $\mathcal{N}(\mu, \sigma^2)$ distribution, then

$$E(X) = \mu, \text{Var}(X) = \sigma^2.$$

2.2.3 Gamma and Inverse Gamma distribution

The *Gamma distribution* is denoted by $\Gamma(p, \lambda)$ with a shape parameter p and an inverse scale parameter λ. Its density function is

$$g_{p,\lambda}(x) = \frac{\lambda^p x^{p-1} e^{-\lambda x}}{\Gamma(p)}$$

for $x > 0$, where the parameters p and λ have to be positive. $\Gamma(p)$ denotes the *Euler Gamma function* defined by

$$\Gamma(p) := \int_0^\infty t^{p-1} e^{-t} \mathrm{d}t.$$

If X has a $\Gamma(p, \lambda)$ distribution, then

$$E(X) = \frac{p}{\lambda}, \text{Var}(X) = \frac{p}{\lambda^2}.$$

The special case $p = 1$ corresponds to the *exponential distribution* $\mathcal{E}(\lambda)$.

Furthermore, there is a connection to the χ^2 distribution. Let k be a positive integer. The *chi squared density with k degrees of freedom* corresponds to the gamma density with

$p = \frac{1}{2}k$ and $\lambda = \frac{1}{2}$ and is denoted by χ^2_k.

Very interesting for using as a prior distribution is the *inverse gamma distribution* $IG(\alpha, \beta)$ with shape parameter α and scale parameter β. Its density function

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\frac{\beta}{x}} \tag{2.1}$$

is defined over the support $x > 0$. Its cumulative distribution function is the *regularized gamma function*

$$F_{\alpha,\beta}(x) = \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}$$

where the numerator is the upper incomplete gamma function ($\Gamma(a, b) := \int_b^\infty t^{a-1}e^{-t}\mathrm{d}t$) and the denominator is the gamma function.
If X has a $IG(\alpha, \beta)$ distribution, then

$$E(X) = \frac{\beta}{\alpha - 1} \text{ for } \alpha > 1, \mathrm{Var}(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \text{ for } \alpha > 2.$$

Furthermore, $\frac{1}{X} \sim Gamma(\alpha, \frac{1}{\beta})$.

2.2.4 Beta distribution

The *Beta distribution*, denoted by $\beta(r, s)$, is closely connected to the previously defined Gamma distribution. Its probability density function with two positive shape parameters r and s is given by

$$b_{r,s}(x) = \frac{1}{B(r, s)} x^{r-1}(1 - x)^{s-1}$$

for $0 < x < 1$, where $B(r, s) := \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$ is the *Beta function*.
If X has a $\beta(r, s)$ distribution, then

$$E(X) = \frac{r}{r+s}, \mathrm{Var}(X) = \frac{rs}{(r+s)^2(r+s+1)}.$$

If Y and Z are independently $\Gamma(r, t)$ and $\Gamma(s, t)$ distributed, respectively, then $\frac{Y}{Y+Z}$ has a $\beta(r, s)$ distribution.

Interesting as a prior distribution is the linearly transformed $\beta(a, b)$ distribution on the interval $(-1, 1)$ (cf. Joe (2006)):

$$g_{a,b}(x) = \frac{1}{2}[B(a, b)]^{-1} \left(\frac{1+x}{2}\right)^{a-1} \left(\frac{1-x}{2}\right)^{b-1}, \qquad x \in (-1, 1)$$

where $B(a, b)$ is the Beta function defined above. If $a = b$, the density function simplifies to

$$g_a(x) = \frac{1}{2}[B(a, a)]^{-1} \left(\frac{1+x}{2}\right)^{a-1} \left(\frac{1-x}{2}\right)^{a-1} = 2^{-2a+1}[B(a, a)]^{-1}(1 - u^2)^{a-1}. \tag{2.2}$$

2.2.5 t distribution

The *t distribution* was published in 1908 by William Gosset under the pseudonym 'Student'. He recognized that the standardized mean of a sum of normal distributed random variables is no more normal if the variance is unknown and must be estimated with the observed sample variance.
So, more generally spoken, if Z has a standard normal distribution and is independent of V which is χ^2_ν distributed, then follows by *Cochran's Theorem* that

$$\frac{Z}{\sqrt{\frac{V}{\nu}}}$$

has a t distribution with ν degrees of freedom.

The probability density function of the t distribution with positive parameter ν, namely the degrees of freedom, is given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where $\Gamma(x)$ is the Gamma function.
If X is t distributed with ν degrees of freedom, then

$$E(X) = 0 \text{ for } \nu > 1, \mathrm{Var}(X) = \frac{\nu}{\nu - 2} \text{ for } \nu > 2.$$

2.2.6 Cauchy distribution

A generalization of a special case of the above defined t distribution is the *Cauchy distribution* as the standard Cauchy distribution corresponds to a t distribution with one degree of freedom.
The probability density function of the Cauchy distribution with location parameter x_0 and positive scale parameter γ is defined as

$$f(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}.$$

Its cumulative distribution function is given by

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}.$$

The Cauchy distribution with location parameter $x_0 = 0$ and scale parameter $\gamma = 1$ is called *standard Cauchy distribution*. It is closely related to the standard normal distribution: Let U and V be two independent random variables with a standard normal distribution. Then $\frac{U}{V}$ is standard Cauchy distributed.
The Cauchy distribution is very heavy tailed so that no moments exist, especially no mean

and no variance.

In a later chapter, we will use a *half Cauchy distribution* as a prior function, i.e. that the support is only on the positive side of x_0. So the density function of a half Cauchy distribution with parameters x_0 and γ is

$$f(x) = \frac{2}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}, \qquad x \geq x_0 \tag{2.3}$$

since the Cauchy density is symmetric with symmetry axis $v = x_0$.

2.3 Multivariate Distributions

Multivariate distributions are generalizations of univariate distributions to higher dimensions. In our case, we will mostly use them as a basis for the construction of elliptical copulas which will follow in later chapters of the diploma thesis.
This section will follow mainly the presentation in Bickel and Doksum (2001).

2.3.1 Multivariate Normal distribution

There are two possibilities to define a *multivariate normal distribution*, but both lead to the same family and are therefore equivalent.

- $\boldsymbol{U} \in \mathbb{R}^d$ has a multivariate (d-variate) normal distribution if and only if $\boldsymbol{U}$ can be written as

  $$\boldsymbol{U} = \boldsymbol{\mu} + A\boldsymbol{Z}$$

  where $\boldsymbol{\mu} \in \mathbb{R}^d, A \in \mathbb{R}^{d\times d}$ are constant and $\boldsymbol{Z} = (Z_1, \ldots, Z_d)'$ where Z_j are independent standard normal variables. Then

  $$E(\boldsymbol{U}) = \boldsymbol{\mu}, \mathrm{Var}(\boldsymbol{U}) = AA' =: \Sigma$$

- $\boldsymbol{U} \in \mathbb{R}^d$ has a multivariate normal distribution if and only if for every $\boldsymbol{a} \in \mathbb{R}^d$ non-random, $\boldsymbol{a}'\boldsymbol{U} = \sum_{j=1}^d a_j U_j$ has a univariate normal distribution.

We denote the d-variate normal distribution with mean $\boldsymbol{\mu}$ and symmetric positive definite variance-covariance matrix Σ by $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$. Its probability density function is given by

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}[\det(\Sigma)]^{d/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

We can further calculate the marginal and conditional distributions (cf. Koop (2005)). Suppose that $\boldsymbol{Y} \in \mathbb{R}^d, \boldsymbol{Y} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ is partitioned as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_{(1)} \\ \boldsymbol{Y}_{(2)} \end{pmatrix}$$

where $\boldsymbol{Y}_{(i)} \in \mathbb{R}^{d_i}, i = 1, 2$ with $d_1 + d_2 = d$ and $\boldsymbol{\mu}$ and Σ have been partitioned conformably as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma'_{(12)} & \Sigma_{(22)} \end{pmatrix}.$$

Then

- The marginal distribution of $Y_{(i)}$ is $\mathcal{N}_{d_i}(\boldsymbol{\mu}_{(i)}, \Sigma_{(ii)})$ for $i = 1, 2$.
- The conditional distribution of $Y_{(1)}$ given $Y_{(2)} = y_{(i)}$ is $\mathcal{N}_{d_1}(\boldsymbol{\mu}_{(1|2)}, \Sigma_{(1|2)})$ where

  $$\boldsymbol{\mu}_{(1|2)} = \boldsymbol{\mu}_{(1)} + \Sigma_{(12)}\Sigma_{(22)}^{-1}(\boldsymbol{y}_{(2)} - \boldsymbol{\mu}_{(2)})$$

  and

  $$\Sigma_{(1|2)} = \Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma'_{(12)}$$

  and vice versa.

If we have only two dimensions, we can further simplify the density function. Let (X, Y) have a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and let

$$\rho := \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2}$$

be the correlation coefficient where σ_1 and σ_2 denote the standard deviation of X and Y, respectively. Then, we define the variance-covariance matrix of (X, Y) as the matrix of central second moments

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

So the probability density function is given by

$$\begin{aligned} f(x, y) = & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right.\right. \\ & \left.\left. -2\rho\frac{(x-\mu_1)}{\sigma_1}\frac{(y-\mu_2)}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right\} \right]. \end{aligned}$$

2.3.2 Multivariate t distribution

A continuous d-dimensional random vector, $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$, has a *multivariate t distribution*, denoted by $\boldsymbol{Y} \sim t(\boldsymbol{\mu}, \Sigma, \nu)$, with parameters $\boldsymbol{\mu}$, the mean vector, with dispersion or scatter matrix Σ, a quadratic, symmetric positive definite $d \times d$-matrix, and with a positive scalar ν, the degrees of freedom, if its probability density function is given by

$$f(\boldsymbol{y}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d|\Sigma|}} \left(1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}{\nu} \right)^{-\frac{\nu+d}{2}}.$$

If $\boldsymbol{Y} \sim t(\boldsymbol{\mu}, \Sigma, \nu)$, then

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} \text{ for } \nu > 1, \text{Var}(\boldsymbol{Y}) = \frac{\nu}{\nu - 2}\Sigma \text{ for } \nu > 2.$$

Very useful is also the fact that we can explicitly calculate the marginals and conditionals of a multivariate t distribution (cf. DeGroot (2003) and Koop (2005)).
Let $\boldsymbol{Y} \sim t(\boldsymbol{\mu}, \Sigma, \nu), \boldsymbol{Y} \in \mathbb{R}^d, \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_{(1)} \\ \boldsymbol{Y}_{(2)} \end{pmatrix}, \boldsymbol{Y}_{(i)} \in \mathbb{R}^{d_i}, d_1 + d_2 = d$ and
$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma'_{(12)} & \Sigma_{(22)} \end{pmatrix}$. Then

- The marginal distribution of $\boldsymbol{Y}_{(i)}$ is $t(\boldsymbol{\mu}_{(i)}, \Sigma_{(ii)}, \nu)$ for $i = 1, 2$.
- The conditional distribution of $\boldsymbol{Y}_{(1)}$ given $\boldsymbol{Y}_{(2)} = \boldsymbol{y}_{(2)}$ is $t(\boldsymbol{\mu}_{(1|2)}, \Sigma_{(1|2)}, \nu + d_1)$ where

$$\begin{aligned} \boldsymbol{\mu}_{(1|2)} &= \boldsymbol{\mu}_{(1)} + \Sigma_{(12)}\Sigma_{(22)}^{-1}(\boldsymbol{y}_{(2)} - \boldsymbol{\mu}_{(2)}) \\ \Sigma_{(1|2)} &= h_{(1|2)} \left[\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma'_{(12)}\right] \end{aligned}$$

  and

$$h_{(1|2)} = \frac{1}{\nu + d_2}\left[\nu + (\boldsymbol{y}_{(2)} - \boldsymbol{\mu}_{(2)})^t\Sigma_{(22)}^{-1}(\boldsymbol{y}_{(2)} - \boldsymbol{\mu}_{(2)})\right]$$

  and vice versa.

If $\boldsymbol{Y}$ is bivariate t distributed with zero mean, we can further simplify the density function:

$$f(y_1, y_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)}}\left(1 + \frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}. \tag{2.4}$$

We will denote the bivariate distribution function by $t_{\nu,\rho}$.
Applying the cited calculations of Koop (2005) mentioned above, we can deduce in the bivariate case:

$$Z := Y_1|Y_2 \sim t(\mu_{(1|2)}, \Sigma_{(1|2)}, \nu + 1) \tag{2.5}$$

where

$$\begin{aligned} \mu_{(1|2)} &= \rho \cdot y_2 \text{ with } \Sigma_{12} = \rho \\ \Sigma_{(1|2)} &= \frac{1}{\nu + 1} \cdot (\nu + y_2^2) \cdot (1 - \rho^2) \end{aligned}$$

and

$$F(y_1|Y_2 = y_2) = t_{\nu+1}\left(\frac{y_1 - \rho y_2}{\sqrt{\frac{(\nu+y_2)\cdot(1-\rho^2)}{\nu+1}}}\right) \tag{2.6}$$

where $t_{\nu+1}(\cdot)$ is the distribution function of the univariate Student's t distribution with $\nu + 1$ degrees of freedom.

2.3.3 Elliptical Distributions

Following Embrechts, Lindskog, and McNeil (2003), we define the class of *elliptical distributions*.

Definition 2.9 *If $\boldsymbol{X}$ is a n-dimensional random vector and, for some $\boldsymbol{\mu} \in \mathbb{R}^n$ and some $n \times n$ nonnegative definite symmetric matrix Σ, the characteristic function $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{t})$ of $\boldsymbol{X} - \boldsymbol{\mu}$ is a function of the quadratic form $\boldsymbol{t}'\Sigma\boldsymbol{t}$, $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{t}) = \phi(\boldsymbol{t}'\Sigma\boldsymbol{t})$, we say that $\boldsymbol{X}$ has an elliptical distribution with parameters $\boldsymbol{\mu}, \Sigma$ and ϕ, and we write $\boldsymbol{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \phi)$.*

For $n = 1$ is the class of elliptical distributions identical to the class of one-dimensional symmetric distributions. The function ϕ is called a *characteristic generator*.

The characteristic function of a multivariate normal distributed random vector $\boldsymbol{X}$ with mean $\boldsymbol{\mu}$ and variance Σ is

$$\phi_{\boldsymbol{X}}(\boldsymbol{x}) = \exp(\imath \boldsymbol{x}'\boldsymbol{\mu}) \cdot \exp(-\frac{1}{2}\boldsymbol{x}'\Sigma\boldsymbol{x}).$$

So the multivariate normal distribution belongs to the class of elliptical distributions, namely to the subclass of Kotz type distributions (cf. Fang, Kotz, and Ng (1990)).

The characteristic function of a multivariate t distributed random vector $\boldsymbol{Y}$ following Kibria and Joarder (2006) is given by

$$\phi_{\boldsymbol{Y}}(\boldsymbol{y}) = \exp(\imath \boldsymbol{y}'\boldsymbol{\mu}) \cdot \frac{\left\|(\nu\Sigma)^{1/2}\boldsymbol{y}\right\|^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} \cdot K_{\nu/2}\left(\left\|(\nu\Sigma)^{1/2}\boldsymbol{y}\right\|\right)$$

where $\left\|\boldsymbol{t}\right\| = \sqrt{t't}$ and $K_{\nu/2}\left(\left\|(\nu\Sigma)^{1/2}\boldsymbol{y}\right\|\right)$ is the Macdonald function with order $\nu/2$ and argument $\left\|(\nu\Sigma)^{1/2}\boldsymbol{y}\right\|$. The Macdonald function can be represented by

$$K_\alpha(t) = \left(\frac{2}{t}\right)^\alpha \frac{\Gamma(\alpha+1)}{\sqrt{\pi}} \int_0^\infty (1+u^2)^{-(\alpha+1/2)} \cos(tu)\mathrm{d}u,$$

where $t > 0$ and $\alpha > -1/2$.
This characteristic function can be rewritten as

$$\phi_{\boldsymbol{Y}}(\boldsymbol{y}) = \exp(\imath \boldsymbol{y}'\boldsymbol{\mu}) \cdot \psi(\boldsymbol{y}'\Sigma\boldsymbol{y})$$

for some function $\psi(\cdot)$ and therefore the multivariate t distributions belong to the class of elliptical distributions, namely to the subclass of multivariate Pearson Type VII distributions (cf. Fang, Kotz, and Ng (1990)).

2.4 Dependence Measures

The presentation of this topic follows mainly Chapter 3.2 in Kurowicka and Cooke (2006).

Definition 2.10 (Independence) *Random variables $X_1, \ldots, X_n$ are independent if for any intervals $I_1, \ldots, I_n$,*

$$P(X_1 \in I_1, \ldots, X_n \in I_n) = \prod_{i=1}^{n} P(X_i \in I_i).$$

Definition 2.11 (Product Moment Correlation) *The product moment correlation (also linear or Pearson correlation) of random variables X, Y with finite expectations $E(X), E(Y)$ and finite variances σ_X^2, σ_Y^2, is*

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}.$$

Definition 2.12 (Rank Correlation) *The rank correlation (or Spearman's rho) of random variables X, Y with cumulative distribution functions F_X and F_Y and joint distribution function F is*

$$\rho_s(X,Y) = \rho(F_X(X), F_Y(Y)).$$

Definition 2.13 (Kendall's tau) *Let $(X_1, Y_1), (X_2, Y_2)$ be two independent pairs of random variables with joint distribution function F and marginal distributions F_X and F_Y. Kendall's tau is given by*

$$\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0).$$

Following Joe (1997), Spearman's rho and Kendall's tau can also be expressed in terms of an integral of the joint distribution function and the copula C associated with F, respectively.

$$\rho_s(X,Y) = 12 \int\int F_X(x) F_Y(y) \mathrm{d}F(x,y) - 3 = 12 \int\int C(u,v) \mathrm{d}u\mathrm{d}v - 3$$

$$\tau = 4 \int F \mathrm{d}F - 1 = 4 \int C \mathrm{d}C - 1 = 4E(C(U,V)) - 1 \qquad \text{for } U, V \sim Unif(0,1)$$

We can find a relationship between the linear correlation, Spearman's rho and Kendall's tau.

Lemma 2.14 (Pearson(1904)) *Let (X, Y) be random vectors with joint normal distribution, then*

$$\rho(X,Y) = 2\sin(\frac{\pi}{6}\rho_s(X,Y))$$

The following result holds for all elliptical distributions.

Lemma 2.15 *Let (X, Y) be random vectors with joint normal or joint Student t distribution, then*

$$\tau(X,Y) = 2\arcsin(\rho(X,Y))/\pi$$

Proof: cf. Lindskog, McNeil, and Schmock (2003) □

A partial correlation can be defined in terms of partial regression coefficients.

Definition 2.16 (Partial Correlation) *Let X_i be random variables with zero mean and standard deviation $\sigma_i = 1, i = 1, \ldots, n$. Let the numbers $b_{12;3,\ldots,n}, \ldots, b_{1n;2,\ldots,n-1}$ minimize*

$$E((X_1 - b_{12;3,\ldots,n}X_2 - \ldots - b_{1n;2,\ldots,n-1}X_n)^2).$$

Then the partial correlation is defined as

$$\rho_{12;3,\ldots,n} := sgn(b_{12;3,\ldots,n})\sqrt{b_{12;3,\ldots,n}b_{21;3,\ldots,n}}.$$

Cramér (1946, p. 306) shows the equivalence of the above defined partial correlation to

$$\rho_{12;3,\ldots,n} = -\frac{C_{12}}{\sqrt{C_{11}C_{22}}},$$

where $C_{i,j}$ denotes the (i,j)-th cofactor (also called minor) of the correlation matrix, i.e. the determinant of the submatrix obtained by removing row i and column j .
The partial correlation $\rho_{12;3,\ldots,n}$ can be interpreted as the correlation between the orthogonal projections of X_1 and X_2 on the plan orthogonal to the space spanned by $X_3, \ldots, X_n$ (cf. Kendall (1945, p. 372 f.) and Baba, Shibata, and Sibuya (2004)).
Yule and Kendall (1965) gave an recursive formula for recursively computing partial correlations:

$$\rho_{12;3,\ldots,n} = \frac{\rho_{12;3,\ldots,n-1} - \rho_{1n;3,\ldots,n-1} \cdot \rho_{2n;3,\ldots,n-1}}{\sqrt{1 - \rho^2_{1n;3,\ldots,n-1}}\sqrt{1 - \rho^2_{2n;3,\ldots,n-1}}}. \tag{2.7}$$

Definition 2.17 (Conditional Correlation) *The conditional correlation of Y and Z given X*

$$\begin{aligned}\rho_{YZ|X} &= \rho(Y|X, Z|X) \\ &= \frac{E(YZ|X) - E(Y|X)E(Z|X)}{\sigma(Y|X)\sigma(Z|X)}\end{aligned}$$

is the product moment correlation computed with the conditional distribution of Y and Z given X.

Following the results of Baba, Shibata, and Sibuya (2004) and Gatz (2007), the partial correlations and the conditional correlations coincide (i.e. $\rho_{12;3,\ldots,n} = \rho_{12|3,\ldots,n}$) if the variables belong to the class of elliptical distributions, e. g. the multivariate normal and the multivariate Student t distribution. The partial and conditional correlations of copulas have to be interpreted on the level of the distributions, not on the level of uniform data. However, zero partial correlation or zero conditional correlation do not imply conditional independence except for normal distributions.

2.5 Time Series

2.5.1 Univariate Time Series

The theory of time series as introduced in this section follows the description in Brockwell and Davis (1991).

Definition 2.18 (Stochastic process) *A stochastic process is a family of random variables* $\{X_t, t \in \mathcal{T}\}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathcal{P})$

As we look at time series, the index (or parameter) set $\mathcal{T}$ is a set of time points, i.e. $1, 2, 3, \ldots$. According to the definition of a random variable, if $t \in \mathcal{T}$ is fixed, X_t is a function $X_t(\cdot)$ on the set Ω. In contrast, if $\omega \in \Omega$ is fixed, $X_\cdot(\omega)$ is a function on $\mathcal{T}$.
The covariance matrix helps to gain insights into the dependency structure for a finite number of random variables. As we often have an infinite number of random variables, we have to extend this concept.

Definition 2.19 (Autocovariance function) *If* $\{X_t, t \in \mathcal{T}\}$ *is a process such that* $\mathrm{var}(X_t) < \infty$ *for each* $t \in \mathcal{T}$*, then the autocovariance function* $\gamma_X(\cdot, \cdot)$ *of* $\{X_t\}$ *is defined by*

$$\gamma_X(r, s) := \mathrm{Cov}(X_r, X_s) = E\left[(X_r - E(X_r))\,(X_s - E(X_s))\right], \quad r, s \in \mathcal{T}$$

Definition 2.20 (Stationarity) *The time series* $\{X_t, t \in \mathbb{Z}\}$*, with index set* $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$*, is said to be stationary if*

(i) $E\left(|X_t|^2\right) < \infty$ *for all* $t \in \mathbb{Z}$*,*

(ii) $E(X_t) = m$ *for all* $t \in \mathbb{Z}$*,*

and

(iii) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ *for all* $r, s, t \in \mathbb{Z}$*.*

If $\{X_t, t \in \mathbb{Z}\}$ is stationary then $\gamma_X(r, s) = \gamma_X(r - s, 0)$ for all $r, s \in \mathbb{Z}$. So we can redefine the autocovariance function as a function of only one variable by

$$\gamma_X(h) := \gamma_X(h, 0) = \mathrm{Cov}(X_{t+h}, X_t) \qquad \text{for all } t, h \in \mathbb{Z}.$$

So we say that $\gamma_X(h)$ is the autocovariance function at lag h. The *autocorrelation function (acf)* of $\{X_t, t \in \mathbb{Z}\}$ is defined analogously as the function whose value at lag h is

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_{t+h}, X_t) \qquad \text{for all } t, h \in \mathbb{Z}. \tag{2.8}$$

There is also another important definition of stationarity which is frequently used.

Definition 2.21 (Strict stationarity) *The time series* $\{X_t, t \in \mathbb{Z}\}$ *is said to be strictly stationary if the joint distribution of* $(X_{t_1}, \ldots, X_{t_k})'$ *and* $(X_{t_1+h}, \ldots, X_{t_k+h})'$ *are the same for all positive integers* k *and for all* $t_1, \ldots, t_k, h \in \mathbb{Z}$*.*

It follows easily that a strict stationary process with finite second moments is stationary. But the converse of this statement is not true. As a counterexample, we can look at a sequence of random variables $\{X_t\}$ such that X_t is exponentially distributed with mean one when t is odd and normally distributed with mean one and variance one when t is even. Then the time series $\{X_t, t \in \mathbb{Z}\}$ is clearly stationary, but it cannot be strictly stationary.
However, there is one important case where stationarity implies strict stationarity.

Definition 2.22 (Gaussian Time Series) *The process $\{X_t\}$ is a Gaussian time series if and only if $(X_{t_1}, \ldots, X_{t_k})'$ are multivariate Gaussian for $t_1 < t_2 < \ldots < t_k \forall k$.*

It follows easily that a stationary Gaussian time series is strictly stationary.

Now we want to define a class of time series defined in terms of linear difference equations with constant coefficients. Therefore, we define first the error process.

Definition 2.23 (White Noise) *The process $\{Z_t\}$ is said to be white noise with mean 0 and variance σ^2, written*

$$\{Z_t\} \sim WN(0, \sigma^2),$$

if and only if $\{Z_t\}$ has zero mean and the covariance function $\gamma_Z(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$

If the random variables $\{Z_t\}$ are independently and identically distributed with mean 0 and variance σ^2 then we shall write

$$\{Z_t\} \sim IID(0, \sigma^2).$$

We can generate a very wide class of stationary processes by using white noise as the forcing terms in a set of linear difference equations. This leads to the notion of an *autoregressive-moving average (ARMA) process.*

Definition 2.24 (ARMA(p, q) process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,*

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \tag{2.9}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$. We say that $\{X_t\}$ is an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

The equations (2.9) can be written symbolically in the more compact form

$$\phi(B)X_t = \theta(B)Z_t, \qquad t = 0, \pm 1, \pm 2, \ldots$$

where ϕ and θ are the p^{th} and q^{th} degree polynomials

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \tag{2.10}$$

and

$$\theta(z) = 1 - \theta_1 z - \ldots - \theta_q z^q$$

and B is the *backward shift operator* defined by

$$B^j X_t = X_{t-j}, \qquad j = 0, \pm 1, \pm 2, \ldots. \tag{2.11}$$

The polynomials ϕ and θ will be referred to as the *autoregressive* and *moving average polynomials*, respectively, of the difference equations (2.10).

As an example, we will look at the AR(1) process, i.e.

$$Y_t = \gamma Y_{t-1} + \epsilon_t \tag{2.12}$$

where $\epsilon_t \sim N(0, \sigma^2)$ iid and $|\gamma| < 1$.

First, we have to investigate the existence and uniqueness of a stationary solution of the process defined in (2.12).

$$\begin{aligned} Y_t &= \gamma Y_{t-i} + \epsilon_t \\ &= \epsilon_t + \gamma(\gamma Y_{t-2} + \epsilon_{t-1}) \\ &= \ldots = \epsilon_t + \gamma\epsilon_{t-1} + \ldots + \gamma^{k+1} Y_{t-k-1} \end{aligned}$$

Since $|\gamma| < 1$ and $\{X_t\}$ is stationary, $\|Y_t\|^2 = E(Y_t^2)$ is constant and we can conclude that

$$\|Y_t - \sum_{j=0}^{k} \gamma^j \epsilon_{t-j}\|^2 = \gamma^{2k+2} \|Y_{t-k-1}\|^2 \overset{\text{a.s.}}{\to} 0 \text{ as } k \to \infty$$

Since $\sum_{j=0}^{\infty} \gamma^j \epsilon_{t-j}$ is mean-square convergent (by the Cauchy criterion), it follows that

$$Y_t = \sum_{j=0}^{\infty} \gamma^j \epsilon_{t-j} \tag{2.13}$$

So $\{Y_t\}$ defined by (2.13) is stationary since

$$E(Y_t) = \sum_{j=0}^{\infty} \gamma^j E(\epsilon_{t-j}) = 0 \tag{2.14}$$

and

$$\begin{aligned} \text{Cov}(Y_{t+h}, Y_t) &= \lim_{n\to\infty} E\left[\left(\sum_{j=0}^{n} \gamma^j \epsilon_{t+h-j}\right)\left(\sum_{k=0}^{n} \gamma^k \epsilon_{t-k}\right)\right] \\ &= \sigma^2 \gamma^{|h|} \sum_{j=0}^{\infty} \gamma^{2j} \qquad \text{(since } E(\epsilon_i \epsilon_j) = 0 \text{ for } i \neq j \text{ and } \epsilon_t \sim N(0, \sigma^2) \text{ i.i.d.)} \\ &= \frac{\sigma^2 \gamma^{|h|}}{1 - \gamma^2} \end{aligned} \tag{2.15}$$

Furthermore, $\{Y_t\}$ as defined by (2.13) satisfies the difference equation (2.12) and is therefore the unique stationary solution.

An important technique to analyze the cyclical patterns of data is the *spectral analysis*. It decomposes a time series into a few underlying sine and cosine functions with given wavelength. We express the wavelength in terms of *frequency* ω, i.e. the number of cycles per time unit. The period T of a sinusoidal function is defined as the required time for one full cycle, therefore $T = 1/\omega$. Spectral analysis and the periodogram, its basic tool, help to understand frequency dependent variability.

Definition 2.25 *Let $\{x_1, \ldots, x_n\}$ be a vector of observations. The periodogram is defined as*

$$I_n(\omega_k) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t \exp(-\imath t\omega_k)\right|^2, \tag{2.16}$$

where $\omega_k = 2\pi k/n$ are the Fourier frequencies expressed in terms of radians per unit time, $k = 1, \ldots, [n/2]$ and $[z]$ denotes the largest integer less than or equal to z.

With a periodogram, we can measure the frequency dependent variability since the ordinate of the periodogram at Fourier frequency ω_k is proportional to the variance accounted for by that frequency component. Hence, relatively large values of $I_n(\omega_k)$ indicate a cycle of period $1/\omega_k$ (cf. Weron (2006) and Brockwell and Davis (1991)).

2.5.2 Multivariate Time Series

A generalization of this univariate theory to multivariate data is also possible.
This section follows mainly the description in Brockwell and Davis (1991, §11) Let

$$\boldsymbol{X}_t := (X_{1t}, \ldots, X_{mt})', \qquad t = 0, \pm 1, \pm 2, \ldots \tag{2.17}$$

be a m-variate time series, $EX_{it}^2 < \infty \; \forall t \; \forall i$, with mean vector

$$\boldsymbol{\mu}_t := E(\boldsymbol{X}_t) = (\mu_{1t}, \ldots, \mu_{mt})' \tag{2.18}$$

and covariance matrices

$$\Gamma(t+h, t) := E[(\boldsymbol{X}_{t+h} - \boldsymbol{\mu}_{t+h})(\boldsymbol{X}_t - \boldsymbol{\mu}_t)'] = [\gamma_{ij}(t+h, t)]_{i,j=1}^{m}. \tag{2.19}$$

Definition 2.26 (Stationary Multivariate Time Series) *The series* (2.17) *with means* (2.18) *and covariances* (2.19) *is said to be stationary, if $\boldsymbol{\mu}_t$ and $\Gamma(t+h, t), h = 0, \pm 1, \ldots,$ are independent of t.*

If $\boldsymbol{X}_t$ is stationary, we use the notation

$$\boldsymbol{\mu} := E(\boldsymbol{X}_t) = (\mu_1, \ldots, \mu_m)'$$

and

$$\Gamma(h) := E[(\boldsymbol{X}_{t+h} - \boldsymbol{\mu}_{t+h})(\boldsymbol{X}_t - \boldsymbol{\mu}_t)'] = [\gamma_{ij}(h)]_{i,j=1}^m.$$

$\boldsymbol{\mu}$ is called the mean of the series and $\Gamma(h)$ is the covariance matrix at lag h. If $\{\boldsymbol{X}_t\}$ is stationary with covariance function $\Gamma(\cdot)$, then for each i, X_{it} is stationary with covariance function $\gamma_{ii}(\cdot)$. For $i \neq j$, we call the function $\gamma_{ij}(\cdot)$ the cross-covariance function of the two series X_{it} and X_{jt}. Notice that in general $\gamma_{ij}(\cdot)$ is not the same as $\gamma_{ji}(\cdot)$. Furthermore, we define the correlation matrix function $R(\cdot)$ by

$$R(h) := \left[\frac{\gamma_{ij}(h)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}\right]_{i,j=1}^m \in \mathbb{R}^{m\times m}.$$

This function $R(\cdot)$ corresponds to the covariance matrix function of the normalized series which we obtain when we subtract $\boldsymbol{\mu}$ from $\boldsymbol{X}$ and then divide each component by its standard deviation.

Definition 2.27 (Multivariate White Noise) *The m-variate series $\{\boldsymbol{Z}_t, t = 0, \pm 1, \pm 2, \ldots\}$ is said to be white noise with mean $\mathbf{0}$ and covariance matrix $\tilde{\Sigma}$, written*

$$\{\boldsymbol{Z}_t\} \sim WN(\mathbf{0}, \tilde{\Sigma}),$$

if and only if $\{\boldsymbol{Z}_t\}$ is stationary with mean vector $\mathbf{0}$ and covariance matrix function,

$$\Gamma(h) = \begin{cases} \tilde{\Sigma} & \text{if } h = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We use further the notation

$$\{\boldsymbol{Z}_t\} \sim IID(\mathbf{0}, \tilde{\Sigma}),$$

to indicate that the random vectors $\boldsymbol{Z}_t, t = 0 \pm 1, \ldots,$ are independently and identically distributed with mean $\mathbf{0}$ and covariance matrix Γ.

Generalizing the univariate case we can define a very useful class of multivariate stationary processes $\{\boldsymbol{X}_t\}$ by requiring that $\{\boldsymbol{X}_t\}$ should satisfy a set of linear difference equations with constant coefficients.

Definition 2.28 (Multivariate ARMA(p, q) Process) *$\{\boldsymbol{X}_t, t = 0, \pm 1, \ldots\}$ is a m-variate ARMA (p, q) process if $\{\boldsymbol{X}_t\}$ is a stationary solution of the difference equations*

$$\boldsymbol{X}_t - \Phi_1 \boldsymbol{X}_{t-1} - \ldots - \Phi_{t-p}\boldsymbol{X}_{t-p} = \boldsymbol{Z}_t - \Theta_1 \boldsymbol{Z}_{t-1} - \ldots - \Theta_{t-q}\boldsymbol{Z}_{t-q}, \tag{2.20}$$

where $\Phi_1, \ldots, \Phi_{t-p}, \Theta_1, \ldots, \Theta_{t-q} \in \mathbb{R}^{m\times m}$ and $\{\boldsymbol{Z}_t\} \sim WN(\mathbf{0}, \tilde{\Sigma}$.

The equation (2.20) can be written in the more compact form

$$\Phi(B)\boldsymbol{X}_t = \Theta(B)\boldsymbol{Z}_t, \qquad \boldsymbol{Z}_t \sim WN(\boldsymbol{0}, \tilde{\Sigma}),$$

where $\Phi(z) := I - \Phi_1 z - \ldots - \Phi_p z^p$ and $\Theta(z) := I - \Theta_1 z - \ldots - \Theta_q z^q$ are matrix-valued polynomials, I is the $m \times m$- identity matrix and B denotes the backward shift operator (like in (2.11)). Each component of the matrices $\Phi(z), \Theta(z)$ is a polynomial with real coefficients and degree less than or equal to p, q respectively.

We will look at a multivariate AR(1) process as an example. This process satisfies

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t, \qquad \boldsymbol{Z}_t \sim WN(\boldsymbol{0}, \tilde{\Sigma}). \tag{2.21}$$

By exactly the same argument as in the univariate case from (2.12) to (2.13), we can express $\boldsymbol{X}_t$ as

$$\boldsymbol{X}_t = \sum_{j=0}^{\infty} \Phi^j \boldsymbol{Z}_{t-j}, \tag{2.22}$$

provided that all the eigenvalues of Φ are less than 1 in absolute value, i.e. provided

$$\det(I - z\Phi) \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ such that } |z| \leq 1. \tag{2.23}$$

If this condition is satisfied then the series (2.22) converges (componentwise) both in mean square and absolutely with probability 1. Moreover it is the unique stationary solution (2.21). The condition (2.23) is the multivariate analogue of the condition $|\phi| < 1$, required for the existence of the causal representation (2.22) in the univariate case (cf. (2.13)).

2.6 Copula

The concept of copulas is very useful. It is a good help to describe multivariate dependencies and to separate them from marginal behavior. N.I. Fisher describes this in his article in *Encyclopedia of Statistical Sciences* (Fisher (1997)):

> Copulas [are] of interest to statisticians for two main reasons: Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions, sometimes with a view to simulation.

A copula is a function which "couples" the multivariate distribution function to their univariate marginal distribution functions. The basics of this theory were found independently by Hoeffding in the 1940s (who only used the interval $[-1/2, 1/2]^2$ and not the unit square) and by Fréchet in the early 1950s. Abe Sklar was the first to use the term "copula" to describe these distributions with fixed marginals in his seminal paper "Fonctions de répartition à n dimensions et leur marges" (Sklar (1959)). He also mentioned there his famous theorem for the first time which is now known as "Sklar's theorem" (2.31). Although the term "copula" was not used very often, a lot of people made research in

this topic, frequently called "distributions with fixed/given marginals".

The presentation of this topic follows mainly Nelsen (2006).
At first some notation: the *unit square* $\mathbf{I}^2$ is the product $\mathbf{I} \times \mathbf{I}$ where $\mathbf{I} = [0, 1]$. With "copula" we always mean a 2-dimensional copula; of course, there are generalizations into higher dimensions possible. Furthermore, all distribution functions considered here are continuous unless otherwise stated.

Definition 2.29 (Copula) *A Copula is a function C from $\mathbf{I}^2$ to $\mathbf{I}$ with the following properties:*

(i) For every u, v in $\mathbf{I}$,

$$C(u, 0) = 0 = C(0, v)$$

(ii) For every u, v in $\mathbf{I}$,

$$C(u, 1) = u \text{ and } C(1, v) = v; \tag{2.24}$$

(iii) For every u_1, u_2, v_1, v_2 in $\mathbf{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \tag{2.25}$$

If a function has property (2.24), then it is called *grounded*. A function where (2.25) holds, is called *2-increasing*.

Copulas are bounded in both directions: there exists an upper bound and a lower bound.

Theorem 2.30 *Let C be a copula. Then for every (u, v) in $\mathbf{I}^2$,*

$$max(u + v - 1, 0) \leq C(u, v) \leq min(u, v).$$

Proof: Nelsen (2006, p. 11) □

The bounds themselves are also copulas. As an honor for the contributions of these scientists, they are called *Fréchet-Hoeffding lower bound* $W(u, v)$ and *Fréchet-Hoeffding upper bound* $M(u, v)$, respectively. Another important copula is the *product copula* $\Pi(u, v) = u \cdot v$.

Now we will state the famous *theorem of Sklar* which shows the unique connection between the marginal distributions functions and the joint bivariate distribution function.

Theorem 2.31 (Sklar's theorem) *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x, y in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$,*

$$H(x, y) = C(F(x), G(y)). \tag{2.26}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on the range of F times the range of G. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (2.26) is a joint distribution function with margins F and G.

Proof: cf. Nelsen (2006, pp. 18 ff) □

In fact, the copula C from (2.26) has the form

$$C(u,v) = H\left(F^{-1}(u), G^{-1}(v)\right), \qquad u,v \in \mathbf{I}.$$

We also get some properties concerning the partial derivatives of a copula. We use the word "almost" in the sense of a Lebesgue measure, i.e. the points where a property is not satisfied have Lebesgue measure zero.

Theorem 2.32 *Let C be a copula. For any v in* $\mathbf{I}$*, the partial derivative* $\frac{\partial C(u,v)}{\partial u}$ *exists for almost all u, and for such u and v,*

$$0 \leq \frac{\partial}{\partial u} C(u,v) \leq 1.$$

Similarly, for any u in $\mathbf{I}$*, the partial derivative* $\frac{\partial C(u,v)}{\partial v}$ *exists for almost all v, and for such u and v,*

$$0 \leq \frac{\partial}{\partial v} C(u,v) \leq 1.$$

Furthermore, the functions $u \mapsto \frac{\partial C(u,v)}{\partial v}$ *and* $v \mapsto \frac{\partial C(u,v)}{\partial u}$ *are defined and non-decreasing almost everywhere on* $\mathbf{I}$*.*

Proof: cf. Nelsen (2006, p. 14) □

Theorem 2.33 *Let C be a copula. If* $\frac{\partial C(u,v)}{\partial v}$ *and* $\frac{\partial^2 C(u,v)}{\partial u \partial v}$ *are continuous on* $\mathbf{I}^2$ *and* $\frac{\partial C(u,v)}{\partial u}$ *exists for all* $u \in (0,1)$ *when* $v = 0$*, then* $\frac{\partial C(u,v)}{\partial u}$ *and* $\frac{\partial^2 C(u,v)}{\partial v \partial u}$ *exist in* $(0,1)^2$ *and* $\frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{\partial^2 C(u,v)}{\partial v \partial u}$*.*

So we can define the density of a copula.

Definition 2.34 (Copula density) *Let C be a copula which is twice partial differentiable. Then the function* $c : \mathbf{I}^2 \to \mathbf{I}$ *with*

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}$$

is called the density of the copula C.

This helps to see a connection between the copula density function, the joint density function and the marginal density functions using the chain rule for differentiation and Sklar's theorem 2.31. If f and g, h denote the nonzero density and the marginal densities of (X,Y) with joint distribution function F and marginal distribution functions G and H, respectively, then

$$c(G(x), H(y)) = \frac{f(x,y)}{g(x) \cdot h(y)}$$

is the copula density of C. Hence, we can rewrite the density in terms of the copula density and the marginal densities as

$$f(x, y) = c(G(x), H(y)) \cdot g(x) \cdot h(y). \tag{2.27}$$

Very remarkable is also the fact that copulas are invariant under increasing continuous transformations of the marginals.

Theorem 2.35 *If x, y have copula C and T_1, T_2 are increasing continuous functions, then $T_1(x), T_2(y)$ also have copula C.*

Proof: cf. Embrechts, McNeil, and Straumann (2002) □

Besides the already mentioned boundary copulas and the product copula, we can define a lot of different ones. Very interesting in our context are the *elliptical copulas* - Normal and Student t-copula - which are defined with the use of distribution functions.

Definition 2.36 (Bivariate Normal Copula) *Let Φ_ρ be the cumulative distribution function of the bivariate Standard normal distribution and let Φ^{-1} be the inverse of the univariate Standard normal cdf. Then we define the Normal or Gauss Copula by using Sklar's theorem as*

$$C_\rho(u, v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$$

for all (u, v) in $\mathbf{I}^2$.

The bivariate normal copula has the density

$$c(u, v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(u^2+v^2) - 2\rho uv}{2(1-\rho^2)}\right).$$

Very important for our work is the *Student t-Copula* which we will use in the pair-copula construction (cf. section 2.7) of a D-vine (cf section 2.8).

Definition 2.37 (Bivariate t-Copula) *Let $t_{\nu,\rho}$ be the bivariate t distribution function and let t_ν^{-1} be the inverse of the standard univariate t distribution function. Then we define the Student t-Copula (or t-Copula) by using Sklar's theorem as*

$$\begin{aligned} C^t_{\nu,\rho}(u, v) &= t_{\nu,\rho}\left[t_\nu^{-1}(u), t_\nu^{-1}(v)\right] \\ &= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left\{1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right\}^{-\frac{\nu+2}{2}} \mathrm{d}t\,\mathrm{d}s \end{aligned} \tag{2.28}$$

for all (u, v) in $\mathbf{I}^2$. If $\nu > 2$, then ρ is the usual linear correlation coefficient of the corresponding bivariate t_ν-distribution.

The density of the bivariate t-Copula is given by

$$c^t_{\nu,\rho}(u,v) = \frac{1}{\sqrt{1-\rho^2}} \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})^2} \frac{\left\{1 + \frac{s^2-2\rho st+t^2}{\nu(1-\rho^2)}\right\}^{-\frac{\nu+2}{2}}}{\left(1+\frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \tag{2.29}$$

where $s = t_\nu^{-1}(u), t = t_\nu^{-1}(v)$.

If we want to model a distribution with more extremal event, it is reasonable to use a t-Copula instead of the Normal Copula as the tails of the t-Copula are much heavier than the ones of the Normal Copula. We see this easily by comparing the 3D surface and the contour plot of a t-Copula and a Normal Copula density with the same correlation coefficient (see Figures 2.1 - 2.3).

Furthermore, it is possible to create a *meta-t-Copula with standard normal margins*: We take univariate standard normal r.v., utilize the probability-integral transform (2.2) with the standard normal distribution function to get univariate r.v. and plug them into a t-Copula:

$$C^{\text{meta-t}}_{\nu,\rho}(x,y) = C^t_{\nu,\rho}(\Phi^{-1}(x), \Phi^{-1}(y))$$

where $x, y \in \mathbb{R}$ are realizations of standard normal random variables.
The 3D surface and the contour plot of this Meta-t-Copula density for different correlation parameters is shown in Figure 2.4

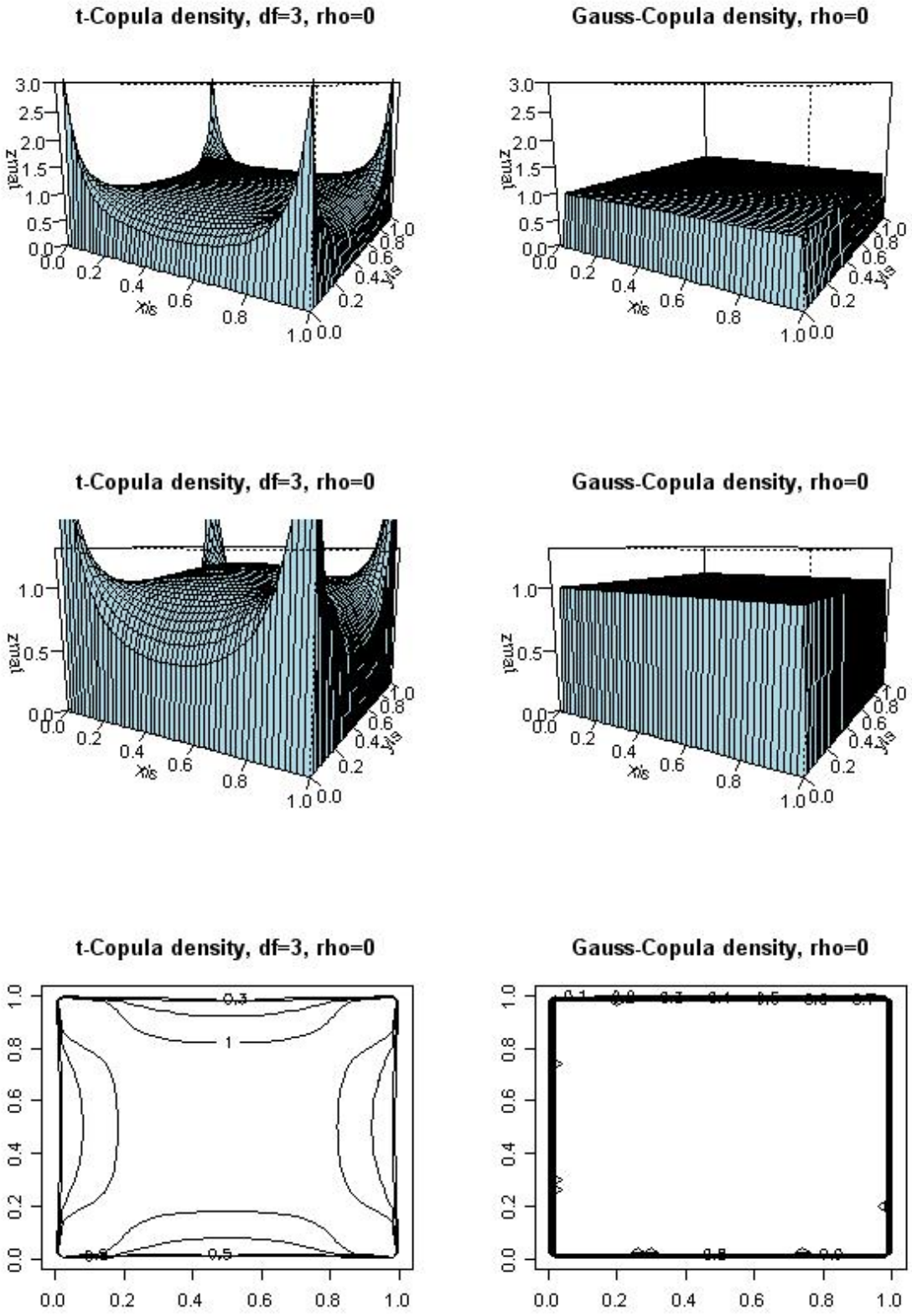

Figure 2.1: Comparison of a t-Copula and a Normal Copula density with correlation 0

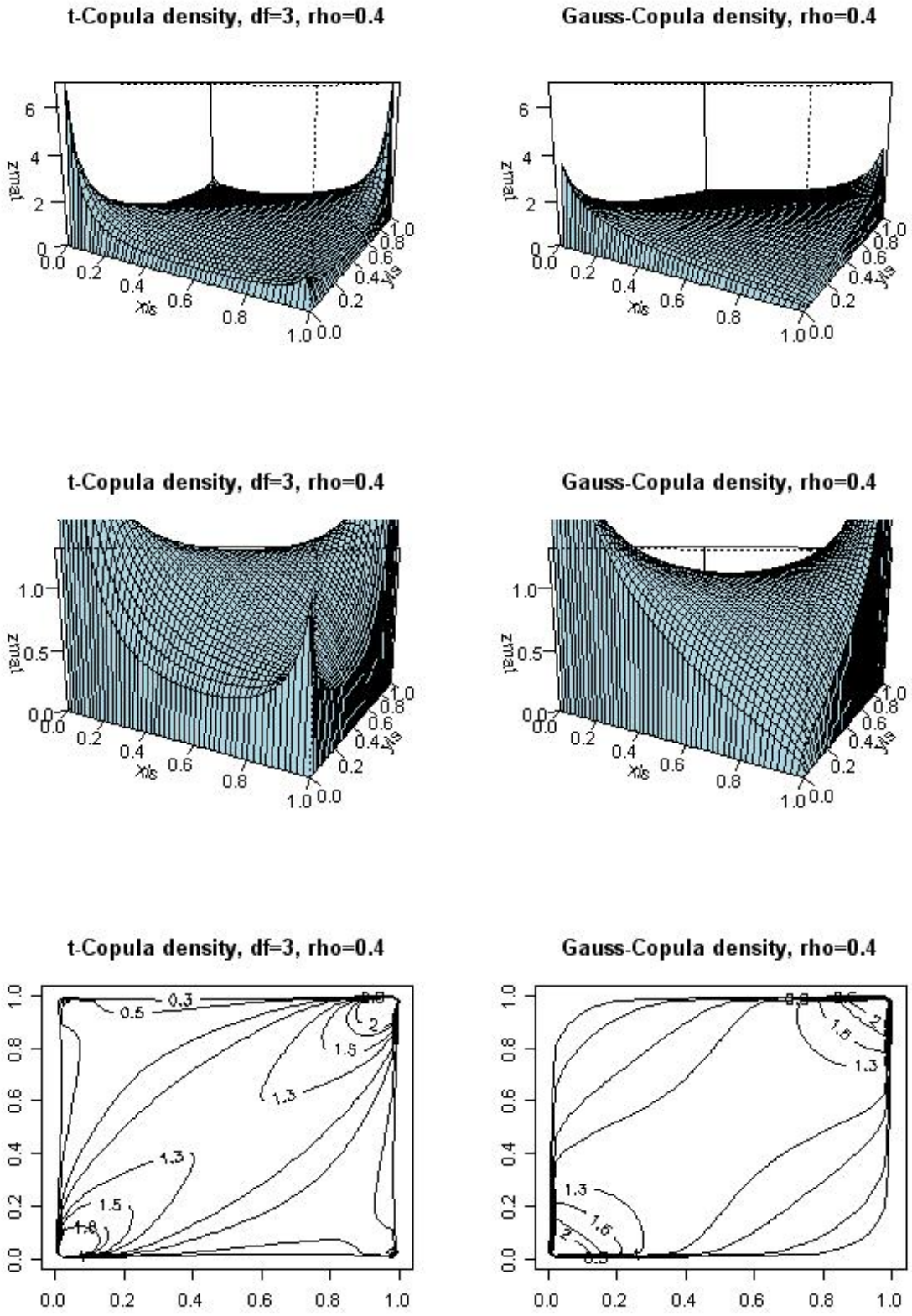

Figure 2.2: Comparison of a t-Copula and a Normal Copula density with correlation 0.4

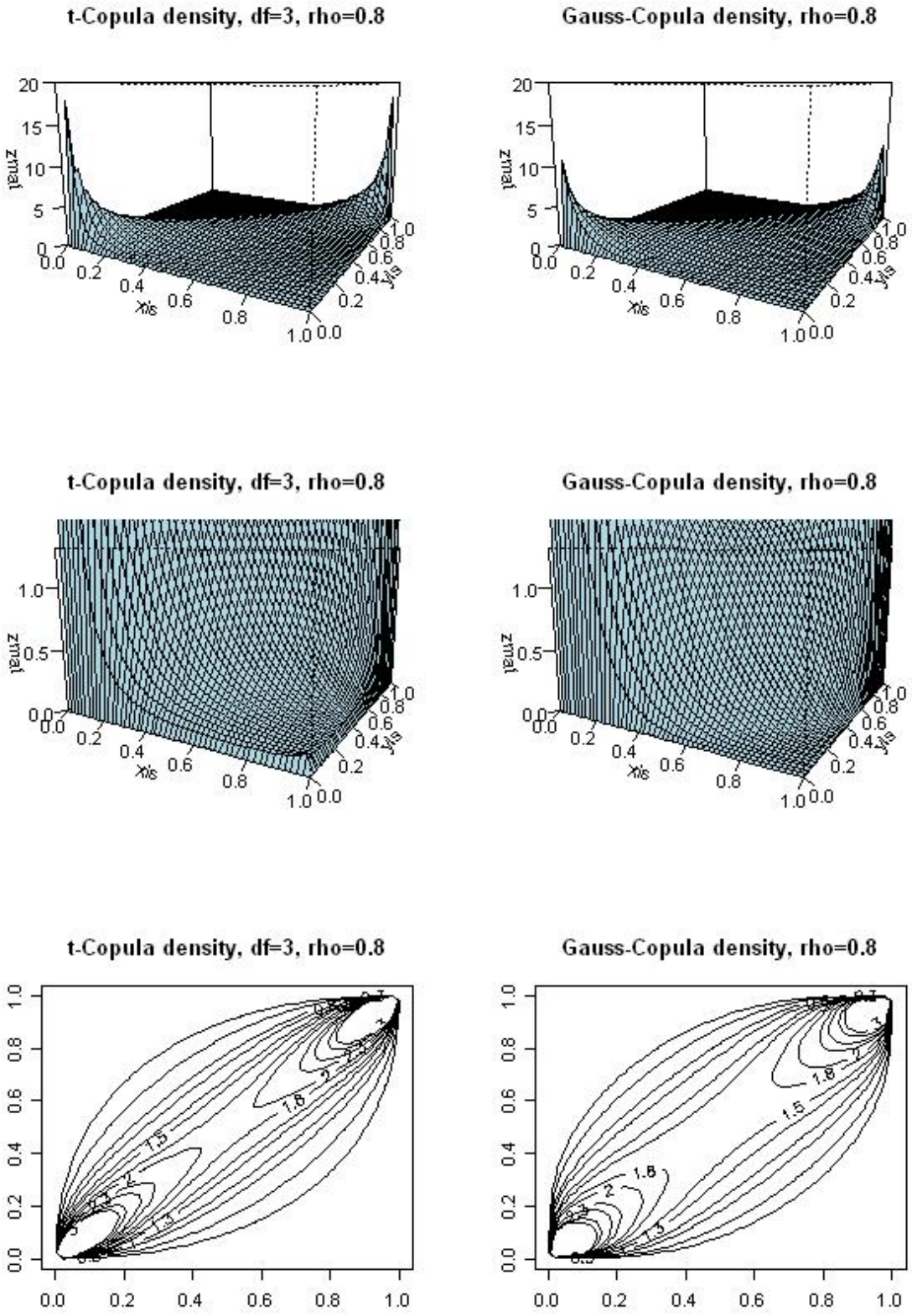

Figure 2.3: Comparison of a t-Copula and a Normal Copula density with correlation 0.8

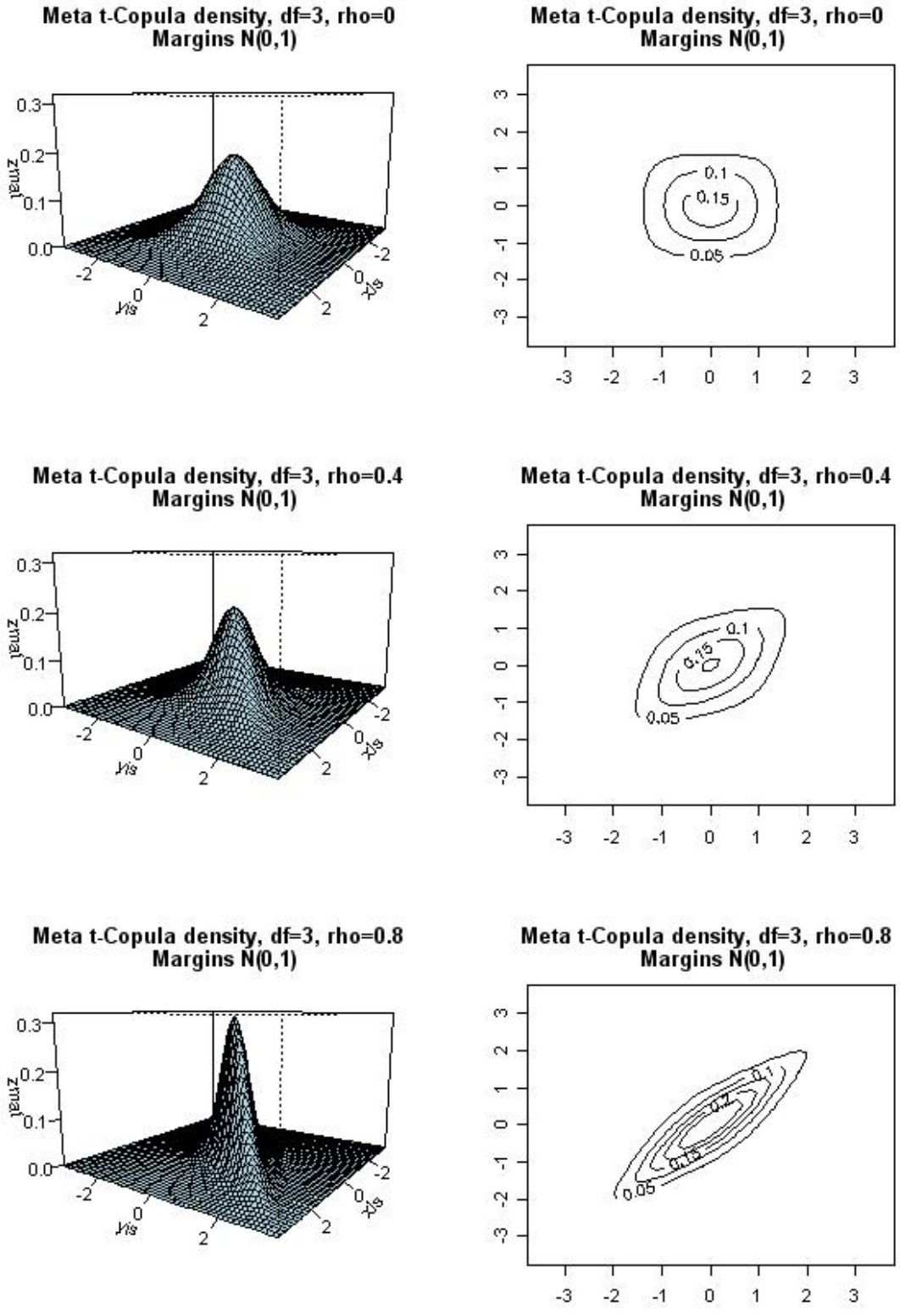

Figure 2.4: Meta-t-Copula density with standard normal margins for correlation parameters 0, 0.4 and 0.9

2.7 Pair-Copula Construction

The pair-copula construction is a useful tool to simplify the work with high dimensional multivariate densities. It was first proposed by Bedford and Cooke (2002) inspired by the work of Joe (1996). Using conditional densities, we can factorize each multivariate density with bivariate copulas, to some extent with conditional densities, and univariate marginal densities.
The presentation of this topic follows mainly the paper of Aas et al. (2007).

Suppose that the vector $\boldsymbol{X} = (X_1, \dots X_d)$ of random variables has the joint density function $f(x_1, \dots, x_d)$. Now we can decompose this density using conditional densities

$$f(x_1, \dots, x_d) = f_d(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, \dots, x_d). \tag{2.30}$$

This factorization is unique up to a re-labelling of the variables.
Using the multivariate generalization of the copula density (2.27), we get

$$f(x_1, \dots, x_d) = c_{1\cdots d}(F_1(x_1), \dots, F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d). \tag{2.31}$$

If we have the bivariate case, (2.31) can be written as

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \tag{2.32}$$

where c_{12} is the appropriate pair-copula density for the transformed variables $F_1(x_1)$ and $F_2(x_2)$.
This factorization is also possible for a conditional density, it follows easily that

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1). \tag{2.33}$$

If we have three variables X_1, X_2 and X_3, we can decompose them as

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_1, x_2). \tag{2.34}$$

We can further factorize

$$f(x_2|x_1) \overset{(2.33)}{=} c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2)$$

and

$$\begin{aligned} f(x_3|x_1, x_2) &= \frac{f(x_2, x_3|x_1)}{f(x_2|x_1)} \\ &= \frac{c_{23|1}(F(x_2|x_1), F(x_3|x_1)) f(x_2|x_1) f(x_3|x_1)}{f(x_2|x_1)} \\ &= c_{23|1}(F(x_2|x_1), F(x_3|x_1)) f(x_3|x_1) \\ &\overset{(2.33)}{=} c_{23|1}(F(x_2|x_1), F(x_3|x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3). \end{aligned}$$

So, plugging all into (2.34), we get

$$\begin{aligned} f(x_1,x_2,x_3) =& c_{23|1}(F(x_2|x_1),F(x_3|x_1)) \\ & \cdot c_{12}(F_1(x_1),F_2(x_2)) \cdot c_{13}(F_1(x_1),F_3(x_3)) \\ & \cdot f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3). \end{aligned} \tag{2.35}$$

If we use another decomposition into conditional densities in (2.34), we get an alternative pair-copula construction.

So, it is obvious that (2.30) can be decomposed under appropriate regularity conditions into any pair-copula construction times its marginal densities using iteratively the general formula

$$f(x|\boldsymbol{v}) = c_{xv|\boldsymbol{v}_{-j}}\left(F(x|\boldsymbol{v}_{-j}),F(v_j|\boldsymbol{v}_{-j})\right) \cdot f(x|\boldsymbol{v}_{-j})$$

for a n-dimensional vector $\boldsymbol{v}$. v_j means any completely arbitrary chosen element of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ is the $(n-1)$-dimensional vector consisting of $\boldsymbol{v}$ without its j-th component.
For applying this pair-copula decomposition, we need to calculate marginal conditional densities of the form $F(x|\boldsymbol{v})$. Joe (1996) showed that

$$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}\left(F(x|\boldsymbol{v}_{-j}),F(v_j|\boldsymbol{v}_{-j})\right)}{\partial F(v_j|\boldsymbol{v}_{-j})}, \tag{2.36}$$

for every j, where $C_{ij|k}$ is a bivariate copula distribution function. We can further simplify this if v is univariate:

$$F(x|v) = \frac{\partial C_{x,v}\left(F(x),F(v)\right)}{\partial F(v)}.$$

As we will need this conditional density more often in the following parts of this diploma thesis, we define it with a fixed notation:

$$h(x,v;\theta) := F(x|v) = \frac{\partial C_{x,v}\left(F(x),F(v)\right)}{\partial F(v)} \tag{2.37}$$

where θ denotes the appropriate parameters of the joint copula and the second parameter of $h(\cdot)$ is the one on which the conditional distribution function is conditioned on.

We can now easily deduce the concrete value of this $h(\cdot)$-function for the t-copula with standard normal margins. The t-Copula with standard normal margins has the cdf (cf. (2.28)):

$$C^t_{\nu,\rho}(\Phi(y_1),\Phi(y_2)) \quad = t_{\nu,\rho}\left[t_\nu^{-1}(\Phi(y_1)),t_\nu^{-1}(\Phi(y_2))\right].$$

Using the conditional distribution (2.5), its conditional cumulative density function (2.6) and the definition of the $h(\cdot)$-function (2.37) we get:

$$h(\Phi(y_1),\Phi(y_2);\nu,\rho) \quad = t_{\nu+1}\left[\frac{t_\nu^{-1}(\Phi(y_1))-\rho t_\nu^{-1}(\Phi(y_2))}{\sqrt{\frac{\left(\nu+\left(t_\nu^{-1}(\Phi(y_2))\right)^2\right)\cdot(1-\rho^2)}{\nu+1}}}\right] \tag{2.38}$$

and the inverse of the $h(\cdot)$-function

$$h^{-1}(\Phi(y_1),\Phi(y_2);\nu,\rho) \quad = t_\nu\left\{t_{\nu+1}^{-1}(\Phi(y_1))\sqrt{\frac{\left(\nu+\left(t_\nu^{-1}(\Phi(y_2))\right)^2\right)\cdot(1-\rho^2)}{\nu+1}}+\rho t_\nu^{-1}(\Phi(y_2))\right\} \quad (2.39)$$

2.8 D-Vines

For organizing the high number of possible pair-copula constructions, Bedford and Cooke (2001, 2002) introduced a graphical model called the *regular vine*. As this class is still very large, we will concentrate here on a special case, the *D-vine* introduced by Kurowicka and Cooke (2004). The presentation of this chapter will follow mainly the paper of Aas et al. (2007).

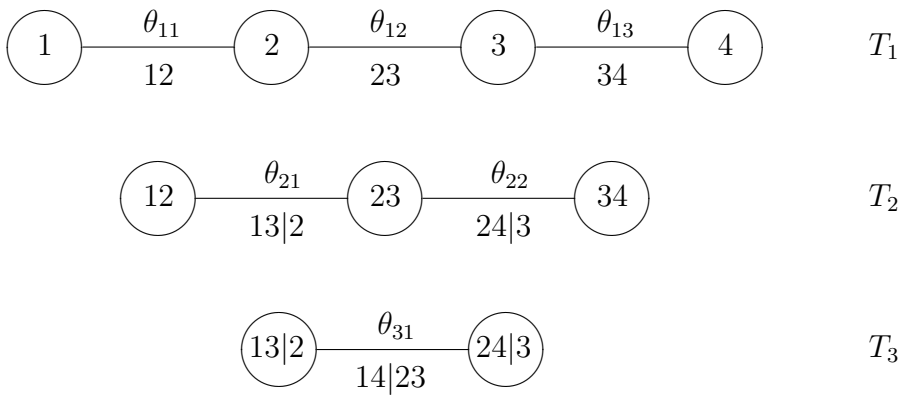

Figure 2.5: A D-vine with 4 variables, 3 trees and 6 edges. Each edge may be associated with a pair-copula. The corresponding copula parameters are denoted above the edges.

The D-vine is one special way of decomposing the joint density. It can be illustrated e.g. in the form of a nested set of trees. Figure 2.5 shows a four-dimensional D-vine. It has three trees $T_j, j = 1, 2, 3$ where tree T_j has $5 - j$ nodes and $4 - j$ edges. We can identify each edge with a pair-copula density and the label of the edge corresponds to the subscript of the pair-copula density, e.g. edge $13|2$ corresponds to the pair-copula density $c_{13|2}(\cdot)$. The corresponding parameters are denoted above the edges, e.g. θ_{21} are the corresponding copula parameters for the pair-copula density $c_{13|2}(\cdot)$, $\nu_{13|2}$ and $\rho_{13|2}$. We only need the $n(n-1)/2$ edges and the marginal densities to specify the decomposition. The nodes of one tree only help to determine the edges for the next tree. Only if two edges in T_j which become nodes in T_{j+1} share a common node, then they are joined by an edge in T_{j+1}. No node in any tree of the D-vine is connected to more than two edges. Note that the tree structure only helps to identify the necessary pair-copula decompositions, it is however not strictly necessary for applying the pair-copula methodology.

We specialize the density of a n-dimensional distribution in terms of a regular vine given by Bedford and Cooke (2001) to a D-vine. So the density $f(x_1, \ldots, x_n)$ corresponding to a D-vine can be written as

$$\prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\ldots,i+j-1} \left\{F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})\right\}, \quad (2.40)$$

where index j identifies the trees while i runs over the edges in each tree.

As we will work in 4 dimensions, we want to show as an example the 4-dimensional D-vine. Its structure is

$$\begin{aligned} f(x_1, x_2, x_3, x_4) =& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot \\ & \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4)) \cdot \\ & \cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \cdot \\ & \cdot c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)). \end{aligned} \quad (2.41)$$

We can see this by decomposing the joint density in the following way:

$$f(x_1, x_2, x_3, x_4) = f_2(x_2) \cdot f(x_3|x_2) \cdot f(x_1|x_2, x_3) \cdot f(x_4|x_1, x_2, x_3). \quad (2.42)$$

We can further factorize

$$f(x_3|x_2) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f_3(x_3)$$

and

$$\begin{aligned} f(x_1|x_2, x_3) &= \frac{f(x_1, x_3|x_2)}{f(x_3|x_2)} \\ &= \frac{c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot f(x_1|x_2) \cdot f(x_3|x_2)}{f(x_3|x_2)} \\ &= c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \end{aligned}$$

and

$$\begin{aligned} f(x_4|x_1, x_2, x_3) &= \frac{f(x_1, x_4|x_2, x_3)}{f(x_1|x_2, x_3)} \\ &= \frac{c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot f(x_1|x_2, x_3) \cdot f(x_4|x_2, x_3)}{f(x_1|x_2, x_3)} \\ &= c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot \frac{f(x_2, x_4|x_3)}{f(x_2|x_3)} \\ &= c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot \\ &\quad \cdot \frac{c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \cdot f(x_2|x_3) \cdot f(x_4|x_3)}{f(x_2|x_3)} \\ &= c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \cdot \\ &\quad \cdot c_{34}(F_3(x_3), F_4(x_4)) \cdot f_4(x_4). \end{aligned}$$

So plugging all in we get

$$\begin{aligned} f(x_1,x_2,x_3,x_4) =& f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4)\cdot \\ & \cdot c_{12}(F_1(x_1),F_2(x_2))\cdot c_{23}(F_2(x_2),F_3(x_3))\cdot c_{34}(F_3(x_3),F_4(x_4))\cdot \\ & \cdot c_{13|2}(F(x_1|x_2),F(x_3|x_2))\cdot c_{24|3}(F(x_2|x_3),F(x_4|x_3))\cdot \\ & \cdot c_{14|23}(F(x_1|x_2,x_3),F(x_4|x_2,x_3)). \end{aligned}$$

We calculate easily that there are in total 12 different four dimensional D-vine decompositions. If we generalize this result, we can prove that number of distinct D-vines on n nodes is given by $n!/2$. We can order the variables in the first tree on $n!$ possible way. Since the edges are undirected - the order of the conditioned values does not matter - the reversed first tree does not change the corresponding vine. So it remains $n!/2$ different first trees. Since the first tree totally determines the whole D-vine, there are $n!/2$ different decompositions.

Using the definition of the $h(\cdot)$-function (2.37) and the relationship in (2.36), we can further simplify the calculations for the four dimensional D-vine, in detail the conditional distribution functions. To simplify the notation in this example, we omit the corresponding parameters of the copula.

$$\begin{aligned} F(x_1|x_2,x_3) &= \int_{-\infty}^{x_1} c_{13|2}(F(u_1|x_2),F(x_3|x_2))\cdot f(u_1|x_2)\mathrm{d}u_1 \\ &= \int_{-\infty}^{x_1} \frac{\partial^2 C_{13|2}(F(u_1|x_2),F(x_3|x_2))}{\partial F(u_1|x_2)\partial F(x_3|x_2)}\cdot\frac{\partial F(u_1|x_2)}{\partial u_1}\mathrm{d}u_1 \\ &= \frac{\partial}{\partial F(x_3|x_2)}\int_{-\infty}^{x_1} \frac{\partial C_{13|2}(F(u_1|x_2),F(x_3|x_2))}{\partial F(u_1|x_2)}\cdot\frac{\partial F(u_1|x_2)}{\partial u_1}\mathrm{d}u_1 \\ &= \frac{\partial}{\partial F(x_3|x_2)}\int_{-\infty}^{x_1} \frac{\partial C_{13|2}(F(u_1|x_2),F(x_3|x_2))}{\partial u_1}\mathrm{d}u_1 \\ &= \frac{\partial}{\partial F(x_3|x_2)} C_{13|2}(F(x_1|x_2),F(x_3|x_2)) \\ &= \frac{\partial}{\partial \eta} C_{13|2}(F(x_1|x_2),\eta)|_{\eta=F(x_3|x_2)} \\ &\overset{(2.37)}{=} h(F(x_1|x_2),F(x_3|x_2)) \end{aligned}$$

and recursively

$$= h(h(F(x_1),F(x_2)),h(F(x_3),F(x_2))). \tag{2.43}$$

So we get for example

$$\begin{aligned} c_{14|23}(F(x_1|x_2,x_3),F(x_4|x_2,x_3)) = c_{14|23}(&h(h(F(x_1),F(x_2)),h(F(x_3),F(x_2))),\\ &h(h(F(x_4),F(x_3)),h(F(x_2),F(x_3)))). \end{aligned}$$

2.9 Sampling and Inference from a D-Vine

This section will follow the exposition of Aas et al. (2007) very closely. At first, we will denote the general algorithm for sampling n dependent variables from a D-vine with a marginal standard normal distribution using probability integral transform 2.2. We sample first $w_1, \dots, w_n$ independent uniform variables on $[0, 1]$.Then, we have to set

$$\begin{aligned} a_1 &= w_1 \\ a_2 &= F^{-1}(w_2|a_1) \\ a_3 &= F^{-1}(w_3|a_2, a_1) \\ \dots &= \dots \\ a_n &= F^{-1}(w_n|a_{n-1}, \dots, a_1). \end{aligned}$$

To get the marginal standard normal distribution, we have to calculate

$$y_i = \Phi^{-1}(a_i), \qquad i = 1, \dots, n$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function.
To calculate the conditional distribution function $F(x_j|x_1, \dots, x_{j-1})$ for each j, we use the definition of the $h(\cdot)$-function in (2.37) and the relationship in (2.36) recursively. We always choose for the D-vine

$$F(x_j|x_1, \dots, x_{j-1}) = \frac{\partial C_{j,1|2,\dots,j-1}(F(x_j|x_2, \dots, x_{j-1}), F(x_1|x_2, \dots, x_{j-1}))}{\partial F(x_1|x_2, \dots, x_{j-1})}.$$

In total, we have to compute $(n-2)^2$ conditional distribution functions for simulating n variables of a D-vine. $\theta_{j,i}$ denotes the set of parameters of the copula density $c_{i,i+j|i+1,\dots,i+j-1}(\cdot, \cdot)$ (cf. Figure 2.5).

The simulation algorithm for a 4-dimensional D-vine with standard normal margins follows the D-vine decomposition described in (2.42):

$$f(x_1, x_2, x_3, x_4) = f_2(x_2) \cdot f(x_3|x_2) \cdot f(x_1|x_2, x_3) \cdot f(x_4|x_1, x_2, x_3).$$

So we can simulate the four random variables in the following way using the probability integral transform 2.2 and the iterative application of the $h(\cdot)$-function like in (2.43):
Sample four random variables $w_1, \dots, w_4$ independent uniform on $[0, 1]$. Then set

$$v_{0,2} := w_2$$

and it follows

$$\begin{aligned} F(v_{0,3}|v_{0,2}) &= h(v_{0,3}, v_{0,2}; \underbrace{\nu_{23}, \rho_{23}}_{=:\theta_{12}}) \\ \Rightarrow v_{0,3} &:= h^{-1}(w_3, v_{0,2}; \theta_{12}) \end{aligned}$$

and

$$
\begin{aligned}
F(v_{0,1}|v_{0,2},v_{0,3}) &= h\big[h(v_{0,1},v_{0,2};\underbrace{\nu_{12},\rho_{12}}_{=:\theta_{11}}),\underbrace{h(v_{0,3},v_{0,2};\theta_{12})}_{=:v_{1,2}},\underbrace{\nu_{13|2},\rho_{13|2}}_{=:\theta_{21}}\big] \\
\Rightarrow h(v_{0,1},v_{0,2};\theta_{11}) &= \underbrace{h^{-1}(w_1,v_{1,2};\theta_{21})}_{=:v_{0,1}^{\text{help}}} \\
&\Rightarrow v_{0,1} := h^{-1}(v_{0,1}^{\text{help}},v_{0,2};\theta_{11})
\end{aligned}
$$

and

$$
\begin{aligned}
F(v_{0,4}|v_{0,1},v_{0,2},v_{0,3}) &= h(F(v_{0,4}|v_{0,2},v_{0,3}),F(v_{0,1}|v_{0,2},v_{0,3});\underbrace{\nu_{14|23},\rho_{14|23}}_{=:\theta_{31}}) = \\
&= h\Big\{h\big[h(v_{0,4},v_{0,3};\underbrace{\nu_{34},\rho_{34}}_{\theta_{13}}),\underbrace{h(v_{0,2},v_{0,3};\theta_{12})}_{=:v_{1,3}};\underbrace{\nu_{24|3},\rho_{24|3}}_{\theta_{22}}\big], \\
&\qquad h\big[\underbrace{h(v_{0,1},v_{0,2};\theta_{11})}_{=:v_{1,1}},\underbrace{h(v_{0,3};v_{0,2},\theta_{12})}_{=v_{1,2}};\theta_{21}\big];\theta_{31}\Big\} \\
\Rightarrow h\big[h(v_{0,4},v_{0,3};\theta_{13}),v_{1,3};\theta_{22}\big] &= h^{-1}\big[w_4,\underbrace{h(v_{1,1},v_{1,2};\theta_{21})}_{=:v_{2,1}};\theta_{31}\big] \\
&\Rightarrow v_{0,4} = h^{-1}\Big\{h^{-1}\big[\underbrace{h^{-1}(w_4,v_{2,1};\theta_{31})}_{=:v_{0,4}^{\text{help}_1}},v_{1,3};\theta_{22}\big],v_{0,3};\theta_{13}\Big\} \\
&= h^{-1}\big[\underbrace{h^{-1}(v_{0,4}^{\text{help}_1},v_{1,3};\theta_{22})}_{=:v_{0,4}^{\text{help}_2}},v_{0,3};\theta_{13}\big] \\
&= h^{-1}(v_{0,4}^{\text{help}_2},v_{0,3};\theta_{13})
\end{aligned}
$$

and

$$
x_i = \Phi^{-1}(v_{0,i}), \qquad i = 1,\dots,4.
$$

Now we can denote the sampling algorithm 2.2.

We are particularly interested in doing *inference* for the D-vine as we want to estimate the parameters. We assume that we observe n variables at T time points. We denote the dataset by $\boldsymbol{x}_i = (x_{i1},\dots,x_{iT}), i = 1,\dots n$. Then, we assume for simplicity, without loss of generality, that the T observations of each variable are independent over time. In the following chapters we will also apply univariate marginal time-series models and estimate all parameters in one step. Furthermore we suppose, that all margins are standard normal.

Algorithm 2.2 Simulation algorithm from four dimensional D-vine with standard normal margins. Generates one sample $x_1, \ldots, x_4$ from the D-vine.

Sample $w_1, \ldots, w_4$ independent uniform on $[0, 1]$.
$v_{0,2} = w_2$
$v_{0,3} = h^{-1}(w_3, v_{0,2}; \theta_{1,2})$
$v_{1,2} = h(v_{0,3}, v_{0,2}; \theta_{1,2}$
$v_{0,1}^{\text{help}} = h^{-1}(w_1, v_{1,2}; \theta_{2,1})$
$v_{0,1} = h^{-1}\left(v_{0,1}^{\text{help}}, v_{0,2}; \theta_{1,1}\right)$
$v_{1,1} = h(v_{0,1}, v_{0,2}; \theta_{1,1})$
$v_{1,3} = h(v_{0,2}, v_{0,3}; \theta_{1,2})$
$v_{2,1} = h(v_{1,1}, v_{1,2}; \theta_{2,1})$
$v_{0,4}^{\text{help}_1} = h^{-1}(w_4, v_{2,1}; \theta_{3,1})$
$v_{0,4}^{\text{help}_2} = h^{-1}(v_{0,4}^{\text{help}_1}, v_{1,3}; \theta_{2,2})$
$v_{0,4} = h^{-1}\left(v_{0,4}^{\text{help}_2}, v_{0,3}; \theta_{1,3}\right)$
$x_i = \Phi^{-1}(v_{0,i}), i = 1, \ldots, 4.$

The log-likelihood is given by

$$ll = \sum_{j=1}^{n}\sum_{t=1}^{T}\log[\varphi(x_{jt})] + \\ + \sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\sum_{t=1}^{T}\log\left[c_{i,i+j|i+1,\ldots,i+j-1}\left\{F(x_{i,t}|x_{i+1,t},\ldots x_{i+j-1,t}), F(x_{i+j,t}|x_{i+1,t},\ldots x_{i+j-1,t})\right\}\right].$$

where $\varphi(\cdot)$ is the probability density function of the standard normal distribution.

When we apply the $h(\cdot)$-function or a cumulative distribution function or a probability density function to a vector (i.e. a bold-typed symbol) in the following algorithm, it is meant that this is done component by component.
$\theta_{j,i}$ denotes the set of parameters of the copula density $c_{i,i+j|i+1,\ldots,i+j-1}(\cdot,\cdot)$ (cf. Figure 2.5).

So we can denote the log-likelihood for the four dimensional D-vine with standard normal margins:

$$\begin{aligned} ll =& \log(\varphi(\boldsymbol{x}_1)) + \log(\varphi(\boldsymbol{x}_2)) + \log(\varphi(\boldsymbol{x}_3)) + \log(\varphi(\boldsymbol{x}_4)) + \\ &+ \log(c_{12}(\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2))) + \log(c_{23}(\Phi(\boldsymbol{x}_2), \Phi(\boldsymbol{x}_3))) + \log(c_{34}(\Phi(\boldsymbol{x}_3), \Phi(\boldsymbol{x}_4))) + \\ &+ \log(c_{13|2}(h(\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2); \theta_{11}), h(\Phi(\boldsymbol{x}_3), \Phi(\boldsymbol{x}_2); \theta_{12}))) + \\ &+ \log(c_{24|3}(h(\Phi(\boldsymbol{x}_2), \Phi(\boldsymbol{x}_3); \theta_{12}), h(\Phi(\boldsymbol{x}_4), \Phi(\boldsymbol{x}_3); \theta_{13}))) + \\ &+ \log(c_{14|23}(h(h(\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2); \theta_{11}), h(\Phi(\boldsymbol{x}_3), \Phi(\boldsymbol{x}_2); \theta_{12}); \theta_{21}), \\ &\qquad h(h(\Phi(\boldsymbol{x}_4), \Phi(\boldsymbol{x}_3); \theta_{13}), h(\Phi(\boldsymbol{x}_2), \Phi(\boldsymbol{x}_3); \theta_{12}); \theta_{31}))). \end{aligned} \tag{2.44}$$

To simplify the notation, we denote

$$L(\boldsymbol{y}, \boldsymbol{z}, \theta) = \sum_{t=1}^{T} \log[c(y_t, z_t, \theta)]$$

where $c(u, v, \theta)$ is the density of the bivariate copula with parameters θ.
This yields the following algorithm 2.3.

Algorithm 2.3 Calculation of Log-Likelihood for 4-dim. D-Vine

1: $ll = 0$
2: **FOR** $i = 1, \ldots, 4$ **DO**
3: $\quad ll = ll + \varphi(\boldsymbol{x_i})$ {Log-Likelihood of the marginal densities}
4: $\quad \boldsymbol{v_{0,i}} = \Phi(\boldsymbol{x_i})$ {Preparation of the values for the D-vine}
5: **END FOR**
6: **FOR** $i = 1, \ldots, 3$ **DO**
7: $\quad ll = ll + L(\boldsymbol{v_{0,i}}, \boldsymbol{v_{0,i+1}}, \theta_{1i})$ {Log-Likelihood of the first tree}
8: **END FOR**
9: $\boldsymbol{v_{1,1}} = h(\boldsymbol{v_{0,1}}, \boldsymbol{v_{0,2}}; \theta_{1,1})$ {Preparing the data for the second tree}
10: $\boldsymbol{v_{1,2}} = h(\boldsymbol{v_{0,3}}, \boldsymbol{v_{0,2}}; \theta_{1,2})$
11: $\boldsymbol{v_{1,3}} = h(\boldsymbol{v_{0,2}}, \boldsymbol{v_{0,3}}; \theta_{1,2})$
12: $\boldsymbol{v_{1,4}} = h(\boldsymbol{v_{0,4}}, \boldsymbol{v_{0,3}}; \theta_{1,3})$
13: **FOR** $i = 1, 2$ **DO**
14: $\quad ll = ll + L(\boldsymbol{v_{1,2i-1}}, \boldsymbol{v_{1,2i}}, \theta_{2i})$ {Log-Likelihood of the second tree}
15: **END FOR**
16: $\boldsymbol{v_{2,1}} = h(\boldsymbol{v_{1,1}}, \boldsymbol{v_{1,2}}; \theta_{2,1})$ {Preparing the data for the third tree}
17: $\boldsymbol{v_{2,2}} = h(\boldsymbol{v_{1,4}}, \boldsymbol{v_{1,3}}; \theta_{2,2})$
18: $ll = ll + L(\boldsymbol{v_{2,1}}, \boldsymbol{v_{2,2}}, \theta_{31})$ {Log-Likelihood of the third tree}

For inference, the D-vine log-likelihood must be numerically optimized. Starting values needed in this maximization may be determined as follows (here for 4 dimensions):

(a) Estimate the parameters of the copulas in tree 1 from the original data.

(b) Compute observations (i.e. conditional distribution functions) for tree 2 using the copula parameters from tree 1 and the $h(\cdot)$-function (cf. Algorithmus 2.3 lines 9 - 12).

(c) Estimate the parameters of the copulas in tree 2 using the observations from (b).

(d) Compute observations for tree 3 using the copula parameters at level 2 and the $h(\cdot)$-function (cf. Algorithmus 2.3 lines 16/17).

(e) Estimate the parameters of the copulas in tree 3 using the observations from (d).

We can generalize this easily for higher dimensions. Each estimation is easy to perform since the data set is only two dimensional.

2.10 Markov Chain Monte Carlo methods

Markov Chain Monte Carlo (MCMC) methodology is a very useful tool to simplify difficult calculations. It was often very hard to determine the value of high-dimensional integrals and we had to develop special software. MCMC helps to solve a big range of complex problems with a quite simple approach.
MCMC uses mainly Monte Carlo integration with Markov Chains. As we will see in the following, *Monte Carlo* integration calculates complicated integrals by simply averaging over samples drawn from the required distribution. Following the law of large numbers, this average converges to the real value. *Markov Chain* Monte Carlo uses a long running cleverly constructed Markov chain for drawing these samples.
All algorithms of this technique with all its different characteristics, including the Gibbs sampler, are based on the seminal works of Metropolis et al. (1953) and Hastings (1970).

The presentation in this chapter follows mainly Gilks, Richardson, and Spiegelhalter (1996) and Gschlößl (2006).

As we have seen in Theorem 2.1, we can denote the posterior distribution of an unknown parameter θ by

$$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

where x is the observed data and $\pi(\cdot)$ is the prior distribution of θ.
While doing Bayesian inference, we are interested in special features of the posterior distribution like moments or quantiles. We can express all these quantities in terms of posterior expectations of functions of θ. We can calculate this expectation of $f(\theta)$ as

$$E\left[f(\theta)|x\right] = \frac{\int f(\theta)\pi(\theta)p(x|\theta)\mathrm{d}\theta}{\int p(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

but the normalizing constant in the denominator is often unknown and cannot be easily evaluated.

In this context, we can apply *Monte Carlo* integration. We want to calculate the expectation

$$E\left[f(X)\right] = \frac{\int f(x)\pi(x)\mathrm{d}x}{\int \pi(x)\mathrm{d}x} \tag{2.45}$$

where $f(\cdot)$ is the function of interest, X is a vector of random variables with distribution $\pi(\cdot)$. Here it is also possible that the distribution of X is known only up to a normalizing constant, i.e. $\int \pi(x)\mathrm{d}x$ is unknown.
We can evaluate (2.45) by drawing samples $\{X_t, t = 1, \ldots, n\}$ from $\pi(\cdot)$ and then approximating

$$E\left[f(X)\right] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t).$$

We can apply the law of large numbers when the samples $\{X_t\}$ are independent and so we ensure the convergence of the approximation by increasing the sample size n which is manually chosen.
But in general, it is often impossible to draw samples $\{X_t\}$ independently from $\pi(\cdot)$ since $\pi(\cdot)$ can be quite non-standard.

In this setting we can make use of the *Markov Chain* Monte Carlo technique. We construct a Markov chain having $\pi(\cdot)$ as its stationary distribution. After a sufficiently long *burn-in* period we can suppose that the sample is drawn from the stationary distribution. So we get the estimator

$$\bar{f} = \frac{1}{n-m} \sum_{t=m+1}^{n} f(X_t) \tag{2.46}$$

where the first m samples are discarded as burn-in. This estimator is called an *ergodic average* and the necessary convergence is ensured by the ergodic theorem.

The construction of such a Markov chain with stationary distribution $\pi(\cdot)$ was first introduced by Hastings (1970) generalizing the method of Metropolis et al. (1953). The so-called *Metropolis-Hastings* algorithm works in the following way (after the description of Gilks, Richardson, and Spiegelhalter (1996)): At each time t, the next state of the Markov chain X_{t+1} is determined by first sampling a candidate point Y from a proposal density $q(\cdot|X_t)$. Y is then accepted with probability $\alpha(X_t, Y)$ where

$$\alpha(X, Y) = min\left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)}\right). \tag{2.47}$$

If the candidate is rejected, the chain stays in its former state, i.e. $X_{t+1} = X_t$. Otherwise, if Y is accepted, the candidate becomes the next state, i.e. $X_{t+1} = Y$.

Algorithm 2.4 Metropolis Hastings-Algorithm with N steps

```
Initialize X_0
FOR t = 0, ..., N DO
    Sample a point Y from q(·|X_t)
    Sample a Uniform(0,1) random variable U
    IF U ≤ α(X_t, Y) THEN
        Set X_{t+1} = Y
    ELSE
        Set X_{t+1} = X_t
    END IF
END FOR
```

It is very remarkable that it does not matter which form the proposal density $q(\cdot|\cdot)$ has. The stationary distribution of the chain will always be $\pi(\cdot)$. For formal proof of this statement and the convergence of the Markov chain, see Gilks, Richardson, and Spiegelhalter (1996, p. 7f).

Furthermore, it is also possible to do *single-component Metropolis-Hastings* steps. We do not have to update all components of X en bloc but we can update each component in a single step while keeping the others constant (regardless if they are already updated or not). So, if X is d-dimensional, one complete Metropolis-Hastings step consists of d separate single-components Metropolis-Hastings steps.

The proposal density is very important for a fast convergence of the MCMC. One possibility to chose a good proposal density is the *random-walk Metropolis* algorithm for which $q(Y|X) = q(|X - Y|)$. So, the acceptance probability (2.47) reduces to

$$\alpha(X, Y) = min\left(1, \frac{\pi(Y)}{\pi(X)}\right).$$

It is a special case of the *Metropolis algorithm* introduced by Metropolis et al. (1953). In this case you often choose a normal distribution centered around the current state X. You have to tune its variance to achieve a reasonable acceptance rate.

Another possibility is the *independence sampler* introduced by Tierney in 1994 (cf. Tierney (1994)). Its proposal density $q(Y|X) = q(Y)$ does not depend on X. Therefore, the acceptance probability (2.47) can be written as

$$\alpha(X, Y) = min\left(1, \frac{w(Y)}{w(X)}\right), \tag{2.48}$$

where

$$w(X) = \frac{\pi(X)}{q(X)}.$$

For an effective operation of the independence sampler, the proposal $q(\cdot)$ should be a good approximation to the target density with slightly heavier tails. An independence proposal which is often chosen and which we will choose in our simulation studies in the following chapters is a normal distribution centered around the mode of $\pi(\cdot)$ and with its covariance matrix somewhat greater than the inverse Hessian matrix

$$\left[-\frac{d^2 \log \pi(x)}{dx'dx}\right]^{-1}$$

evaluated at the mode (for calculations cf. Gamerman and Lopes (2006, p. 83)). The mode has to be calculated in each step using a numerical optimization routine like Newton-Raphson.
So we get the following algorithm for the independence MH-Sampler. For simplicity, it is written down for only one component, but we can generalize it easily.

Algorithm 2.5 Independence Metropolis-Hastings sampler for N steps

Initialize X_0
FOR $t = 0, \ldots, N$ **DO**
    Calculate the mode θ_{mode} and the inverse Hessian at the mode $-H(\theta_{mode})^{-1}$ of the target distribution $\pi(\cdot)$, where $H(\theta) := \frac{\partial^2 \log \pi(\theta)}{\partial \theta^2}$
    Sample a candidate value $Y \sim \mathcal{N}(\theta_{mode}, -H(\theta_{mode})^{-1})$
    Sample a Uniform(0,1) random variable U
    IF $U \leq \alpha(X_t, Y) = min\left(1, \frac{\pi(Y)q(X_t)}{q(Y)\pi(X_t)}\right)$ **THEN**
        Set $X_{t+1} = Y$
    ELSE
        Set $X_{t+1} = X_t$
    END IF
END FOR

After having performed a simulation, we are interested in its results. At first, we look at the trace plot, i.e. a plot with each value of the chain connected to the next one, and at the autocorrelation function of the chain. If the trace plot switches regularly around the estimated value and the autocorrelation function tails off, the chain has a good mixing. Then we subtract the burn-in period. To further increase the independence between the single values of the Markov chain for one parameter, we rare the rest of the chain by taking only every 20th value. We can evaluate the created Markov chain by calculating the mean and the mode for each parameter. Furthermore, we can determine posterior intervals by taking the respective value of the order statistic. We care about the 2.5%, the 5%, the 50%, the 95% and the 97.5% quantile of the posterior density. Hence, we can take as an estimator the appropriate value of the order statistics of the simulated Markov chain. If, for example, we have simulated 10 000 steps, we assume a burn-in period of 1000 steps. So we disregard the first 1000 values of the chain. Then we rare the chain by taking only every 20th value. So the remaining statistics consits of 450 values. The estimator for the 2.5% quantile is the 11th value of the ordered remaining values. We also estimate the density of each parameter with a Gaussian kernel density estimator and generate a trace plot of the rared chain. Additionally, we plot the histogramm of the observed loglikelihood for each step of the chain and we estimate its density.

2.11 Estimation Procedures

Since we use a lot of different methods for estimating the unknown parameters, we denote and clarify these different procedures in the following table. Furthermore, we define the corresponding abbreviations for a later use in the analysis and comparison of our results. A parametric transformation means that we use the estimated time series parameter to transform the observed time series to the unit cube using the transformation (4.1) and the standard normal cumulative distribution function $\Phi(\cdot)$ (cf. Section 2.2.2). In contrast, a nonparametric transformation to the unit cube is done by applying the empirical cumulative distribution function on the residuals, i.e. the difference of the observed to

the fitted values.

method	estimation of the marginal parameters	estimation of the copula parameters
joint MCMC (JMCMC)	jointly estimated	jointly estimated
two-step parametric MLE Notation: TSP-MLE	MLE Notation: marg. MLE	• parametric transformation of the residuals to unit cube • copula parameters estimated by MLE using starting values determined by the algorithm denoted at the end of Section 2.9 Notation: start := starting values, C-MLE := copula MLE estimation
marginal MLE, nonparametric MCMC copula Notation: MMLE-CMCMC	MLE Notation: marg. MLE	• nonparametric transformation of the residuals to unit cube • MCMC estimation of copula parameters Notation: C-MCMC: copula MCMC estimation
two-step nonparametric MLE Notation: TSNP-MLE	MLE Notation: marg. MLE	• nonparametric transformation of the residuals to unit cube • copula parameters estimated by MLE using starting values start (C-MLE) Notation: start := starting values, C-MLE := copula MLE estimation

Table 2.1: Overview of the different applied estimation methods

The algorithm for the C-MCMC method was implemented by Aleksey Min whereas the C-MLE is calculated with the R-package 'copulaGOF' of Daniel Berg and Henrik Bakken. The marginal MLE is evaluated with the function 'arima' implemented in the 'stats'-package of R.

Chapter 3

Electricity Load in Australia

3.1 Liberalization in Electricity Markets

The presentation of the theory of electricity markets follows mainly Weron (2006).
For a long time, electricity markets were seen as natural monopolies where no markets could be established. Electricity was regarded as a vertically integrated monopoly structure consisting of generation, transport and distribution. But during the last two decades, there has been a change in this way of thinking. According to the successful liberalization in other industrial branches, different countries have started to liberalize their electricity market and to establish electricity market exchanges. This has been possible due to the technical evolution, especially in generation technique and transmission. These countries hope to gain more efficiency, to stimulate the technical innovation and to lead to efficient investment by opening their electricity markets. This process was started by Chile in 1982 and other countries like Great Britain and the Nordic European countries soon followed. In Australia, power markets in Victoria and New South Wales opened in 1994 and the Australian National Electricity Market (NEM) began operating in 1998.

The liberalization of the electricity markets is very controversial. One benefit often noted is the decrease in electricity prices and a more efficient use of electricity assets. But due to the reversal fact that many states increased their taxes on electricity, this trend is not apparent to households and small-sized firms. In former times, during the monopoly, the prices paid by these little customers were often subsidized. But during the monopoly, there was a tendency to create substantial over-capacities. This trend has in fact now been reduced.
Very controversial is also the question if the liberalization of the electricity markets creates enough incentives to invest substantially into the development of new generation and transmission capacities and techniques. Although long-term capital-intensive investments with low marginal costs would be advisable, many market participants invest only in generation plants realizable in short terms. This comes from the fact that investment decisions are no longer centrally planed, but are an outcome of the market participants. If the companies expect low electricity prices, they can further postpone necessary investments for supporting the infrastructure. So, it is questionable if these markets should include capacity payments (payment for the reliability of the generator, i.e. its availabil-

ity) like in some South American countries and in Spain, if they should organize capacity markets (like in the northeastern part of the United States) or if its enough to keep these "energy-only" markets (like in Australia or New Zealand).
In these energy-only or one-price-only markets, the compensation for both variable and fix costs is included in the wholesale electricity price. A consequence of this market theory are price spikes, i.e. abrupt and generally unanticipated large changes in spot prices which should send signals to the companies that new generation facilities are needed. But if these spikes are too rare, it is possible that further incentives like capacity payments are necessary as incentives for the companies to keep the electricity generation and transmission in a good state.

3.2 Electricity Market in Australia

In former times, the Australian electricity supply was organized as a vertically integrated monopoly with almost no trade or connection between different states. The liberalization started with the opening of the National Electricity Market (NEM) in December 1998. The first members were Victoria, Queensland, New South Wales, the Australian Capital Territory and then South Australia and Tasmania joined in 2005. It is a wholesale market for electricity to supply retailers and end-users.
The connection between the electricity producers and electricity consumers is facilitated by the establishment of the National Electricity Market Management Company (NEMMCO). This company manages a pool where the output of all generators is aggregated and scheduled to meet the forecasted demand.
Wholesale trading is done as a real-time market where supply and demand are instantaneously matched through a centrally dispatched process. As the offers are submitted by the generators every five minutes, NEMMCO determines the necessary plants and they are dispatched into production. So the market clearing price is determined every five minutes and is averaged for each trading interval (30 minutes).
Since the Australian electricity market is an energy-only market, there are a lot of price spikes. But the experience shows that these price spikes are enough incentives for the electricity companies to build new generation plants, for example as seen in South Australia in the period 1998 - 2003 where the generation capacities increased heavily after a series of price spikes to meet these peak demands.

3.3 Modeling Daily Load Data

Electricity demand clearly shows seasonal fluctuations, mostly due to changing climate conditions like temperature or the number of daylight hours.
We follow the classical technique of seasonal decomposition by thinking of a trend component T_t, a seasonal component S_t and the remaining stochastic component Y_t. This method is known as *Census I method*. So we can write the observed daily load data $\{x_1, \ldots, x_n\}$ as

$$x_t = T_t + S_t + y_t, \qquad t = 1, \ldots, n.$$

We can model this data by a broad range of different approaches (following Weron (2006) who uses techniques of Brockwell and Davis (1991)).

3.3.1 Differencing

Differencing helps to remove seasonal components and trends. It eliminates some of the autocorrelations and helps to make the time series stationary. The simple idea behind it is to consider the differences between successive pairs of observations with appropriate legs.
Since we have a clearly weekly seasonal component in daily load data, we remove it by generating the transformed series

$$z_t = x_t - x_{t-7} = (1 - B^7)x_t, \qquad t = 8, \ldots, n, \tag{3.1}$$

where B is the backward shift operator (2.11). All seasonal components of period 7 are eliminated by this transformation, but it is possible that a mean-reverting relationship (i.e. a negative autocorrelation at lag 7) can be observed. However, this can be an artifact of differencing and therefore spurious since a white noise differenced at lag 7 exhibits a similar ACF.

Another possibility is the use of differencing combined with a moving average-type smoothing:

$$z_t = x_t - \left(\frac{1}{N} \sum_{i=1}^{N} x_{t-i\cdot 7} + \frac{1}{7} \sum_{j=1}^{7} x_{t-j} - \frac{1}{7N} \sum_{i=1}^{N} \sum_{j=1}^{7} x_{t-i\cdot 7-j} \right). \tag{3.2}$$

3.3.2 Median Week

The idea is to calculate an 'average' week of the data which is then subtracted. Therefore, we calculate the median over all observations from one weekday $\hat{w}_k$, i.e. the median over all observations on Mondays, the median over all observations on Tuesdays, etc.:

$$\hat{w}_k = Median(\{x_k, x_{k+7}, x_{k+14}, \ldots, x_{k+[n-k]}\}), \qquad k = 1, \ldots, 7$$

where $[\cdot]$ is the floor function (aka as integral part) of the number and $\hat{w}_k = \hat{w}_{k-7}$ for $k > 7$. We get the 'average' week which we subtract from the observed data, i.e. the 'average' Monday value is subtracted from a Monday observation etc.:

$$z_t = x_t - \hat{w}_t$$

So we remove the seasonal component and get the stochastic component.

3.3.3 Moving Average Technique

The moving average technique also helps to remove the weekly seasonality. To eliminate the weekly component and to dampen the noise, we first estimate the trend by applying a moving average filter:

$$\hat{m}_t := \frac{1}{7}(x_{t-3} + \ldots + x_{t+3}), \qquad t = 4, \ldots, n-3.$$

Now, we can estimate the seasonal component. Hence, we calculate the average w_k of the deviations $\{(x_{k+7j} - \hat{m}_{k+7j}), 3 < k + 7j \leq n - 3\}$ for each $k = 1, \dots 7$,

$$w_k := \frac{1}{[(n-3-k)/7] + 1 - i} \sum_{j=i}^{[(n-3-k)/7]} (x_{k+7j} - \hat{m}_{k+7j}), k = 1, \dots, 7; i := \begin{cases} 1 \text{ if } k \leq 3 \\ 0 \text{ if } k > 3 \end{cases}$$

where $[\cdot]$ denotes the floor function (aka as integral part) of the number. Then we estimate the seasonal component s_k as:

$$\hat{s}_k := w_k - \frac{1}{7} \sum_{i=1}^{7} w_i, \qquad k = 1, \dots, 7,$$

where $\hat{s}_k = \hat{s}_{k-7}$ for $k > 7$.
The deseasonalized data is then defined as $y_t = x_t - \hat{m}_t - \hat{s}_t$ for $t = 3, \dots, n - 3$.

3.3.4 Sinusoid Fitting

One possibility to remove the annual seasonality and a linear trend is the fitting of a sinusoid of an one-year period and of a sinusoid with a weekly period to the data:

$$S_t = A_1 \sin\left(\frac{2\pi}{365}(t + B_1)\right) + A_2 \sin\left(\frac{2\pi}{7}(t + B_2)\right) + C \cdot t.$$

The denominator 365 stands for the number of observations in one year whereas we have 7 observations per week. This technique comes from the spectral (or Fourier) decomposition of a signal.

3.3.5 Rolling Volatility Technique

The rolling volatility technique was introduced by Weron as an alternative to estimate the annual seasonality. First, we have to calculate a vector $\{r_1, \dots, r_n\}$ of returns, i.e. log-changes of the observed time series, of length $n = m \cdot T$ being a multiple of the annual period $T = 365$ days. Then we have to do the following steps:

(i) calculate a 25-day rolling volatility:

$$v_t = \sqrt{\frac{1}{25} \sum_{i=0}^{24} (R_{t+i} - \bar{R}_t)^2}, \qquad \text{where } \bar{R}_t = \frac{1}{25} \sum_{i=0}^{24} R_{t+i},$$

for $t = 1, \dots, n$ and a vector of returns $\{R_t\}$ such that $R_1 = \dots = R_{12} = r_1, R_{12+t} = r_t$ for $t = 1, \dots, n,$ and $R_{n+12} = \dots = R_{n+24} = r_n$;

(ii) rescale the returns by dividing them by the 25-day rolling volatility,

(iii) subtract the estimated mean.

3.4 Observed Data

The used load data consists of four time series of daily observations dating from May 16th, 2005 to June 30th, 2008, in total 1142 data points per time series. It describes the average daily load demand in Gigawatt (GW) for the regions Queensland, New South Wales, Victoria and South Australia calculated by averaging the half-hourly observed data for one day. This data is available at www.nemmco.com.au.

At first, we have a look at the time series plots.

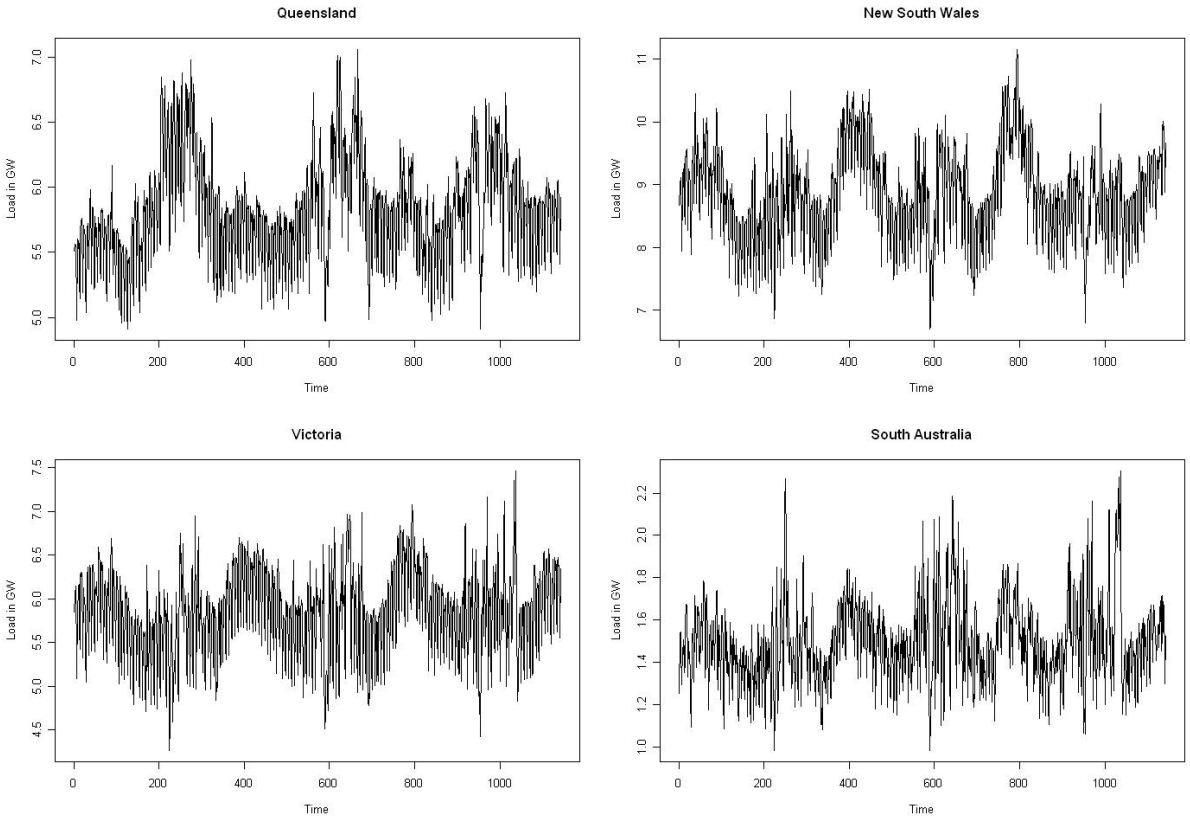

Figure 3.1: Time series plots of the observed data in the four different states

Very interesting for remarking trends or seasonal behavior are plots of the sample autocorrelation function, of the sample partial autocorrelation function and of the periodogram (cf. (2.16)). The sample autocorrelation function (ACF) is defined analogously to the autocorrelation function (2.8) as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

where $\hat{\gamma}(\cdot)$ is the sample autocovariance function given by

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(y_{t+h} - \bar{y})(y_t - \bar{y}),$$

and $\bar{y} = 1/n \sum_{t=1}^{n} y_t$ is the sample mean. The sample partial autocorrelation function (PACF) is the function $\alpha(\cdot)$ defined by

$$\hat{\alpha}(0) = 1 \qquad \text{and} \qquad \hat{\alpha}(h) = \hat{\phi}_{hh}, h \geq 1,$$

where ϕ_{hh} is the last component of

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{h}} = \hat{\Gamma}_h^{-1} \hat{\boldsymbol{\gamma}}_{\boldsymbol{h}},$$

$\hat{\Gamma}_h := [\hat{\gamma}(i-j)]_{i,j=1}^{h}$ is the covariance matrix, and $\hat{\boldsymbol{\gamma}}_{\boldsymbol{h}} := (\hat{\gamma}(1), \ldots, \hat{\gamma}(h))'$.
We can remark dependencies by plotting the sample autocorrelation function and the sample partial autocorrelation function against the time lags $h = 0, 1, \ldots n-1$.

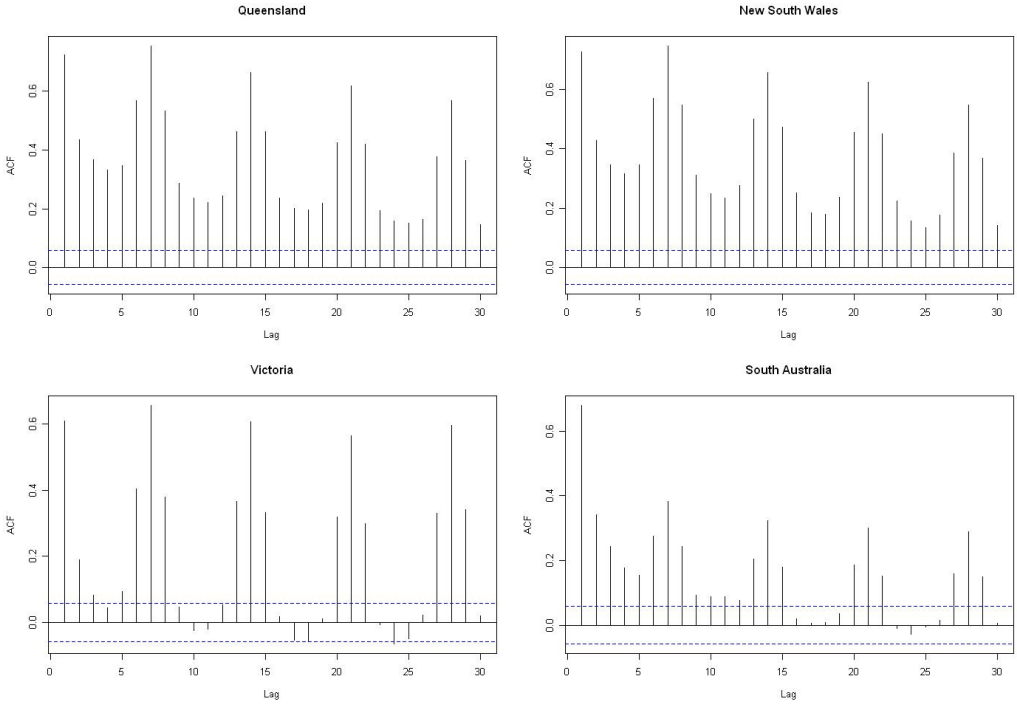

Figure 3.2: Sample autocorrelation function for the four different states

We can clearly see in the plot of the sample autocorrelation function (Figure 3.2) that there is a large dependency at lag 7 and its multiples (14, 21 etc). Additionally, we remark in the plot of the sample partial autocorrelation function (Figure 3.3) that there is a large dependency of lag one (besides other dependencies).

The plot of the periodogram (Figure 3.4) supports the theory of a weekly cycle since there is an obvious peak at the frequency 0.14 which stands for a cycle of 7 days ($T = 1/0.1428 \approx 7$). The smaller spikes at 0.28 and 0.43 indicate periods of 7/2 and 7/3, respectively, and they are called the *harmonics* (i.e. multiples of the 7 day frequency). They are a hint that the 7 day periodicity is not sinusoidal (following Weron (2006)).
Furthermore, we can see a large spike at about 0.0027 which is a sign for a yearly cycle ($T = 1/0.0027 \approx 365$). This peak is the maximum in the plots of the periodogram for

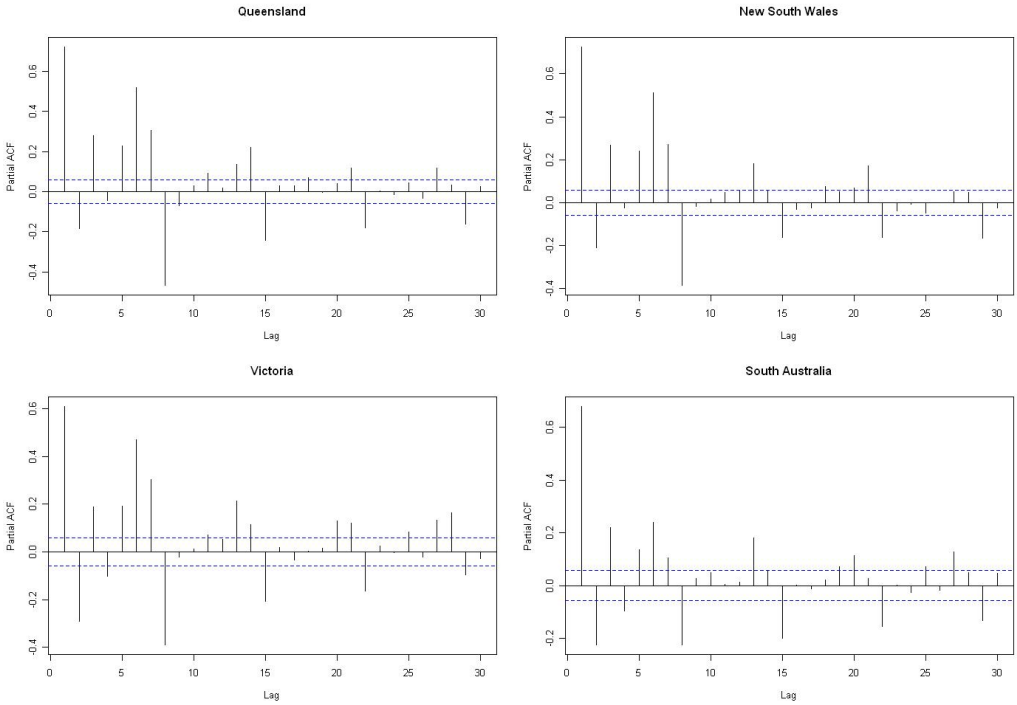

Figure 3.3: Sample partial autocorrelation function for the four different states

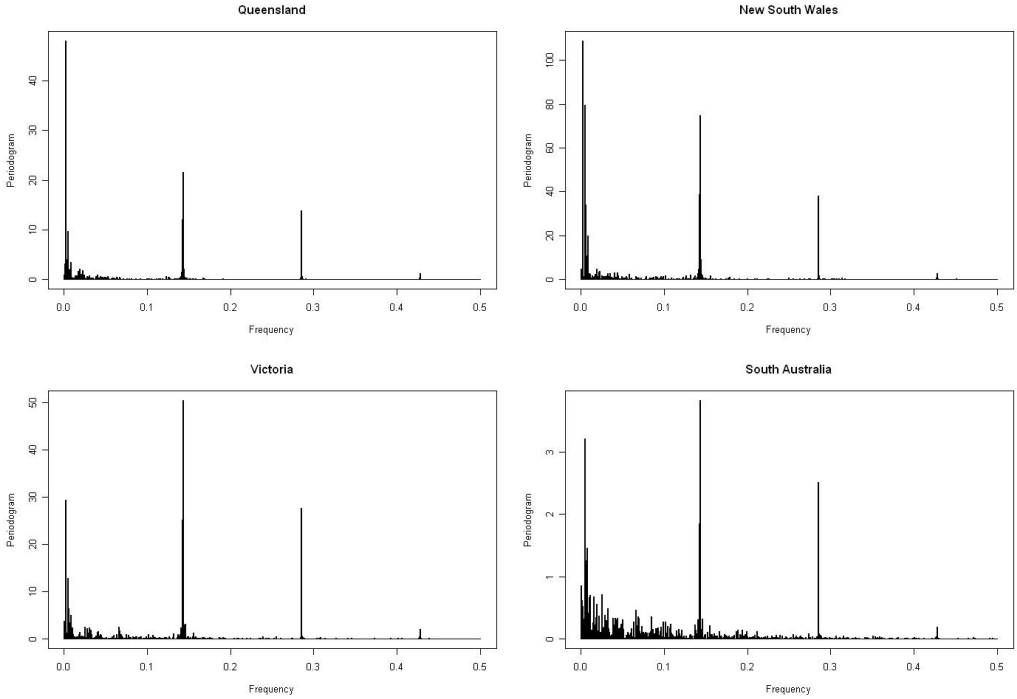

Figure 3.4: Periodogram for the four different states

Queensland and New South Wales whereas the maximum for Victoria and South Australia is at the frequency 0.14.

3.5 Preprocessing of the Observed Data

Now we want to 'preprocess' the data to remove the trend and the weekly and annual seasonality. Therefore, we use the methods described in Section 3.3.
Since we want to fit an AR(p), MA(q) or ARMA(p,q)-process to the preprocessed data, we have to know how the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of these processes behave. Following the calculations in Brockwell and Davis (1991), we get the Table 3.1.

	AR(p)	MA(q)	ARMA(p,q), $p,q>0$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Table 3.1: General Behavior of the ACF and PACF for ARMA-models

As an example, we show in Figure 3.5 a plot of the time series, of the sample autocorrelation function, of the sample partial autocorrelation function and of the periodogram of a simulated AR(1) time series with 1135 values where the autoregressive coefficient is 0.65 and the variance of the Gaussian white noise is 0.5.

We try the following methods to preprocess the data:

Nr.	Description of the Method
1	Difference the data at lag 7 (cf. (3.1))
2	Differencing at lag 7 with moving average-type smoothing (cf. (3.2))
3	Subtracting the median week (cf. Section 3.3.2)
4	Applying the moving average technique (cf. Section 3.3.3)
5	Fitting a Sinusoid for the yearly and weekly cycle with a linear trend (cf. Section 3.3.4)
6	Applying the rolling volatility technique (cf. Section 3.3.5)
7	Applying first the rolling volatility technique and afterwards fitting a sinusoid
8	Applying first the rolling volatility technique and afterwards fitting a MA(7) where only the seventh moving average coefficient is not equal to zero, i.e. $X_t = \epsilon_t + \theta_7\epsilon_{t-7}$, $\theta_7 \neq 0, \{\epsilon_t\} \sim WN(0,\sigma^2)$.

Table 3.2: 8 Methods for Preprocessing the Data

After performing all these 8 methods for preprocessing, we get the plots of the residual time series (cf. Figures 3.6 and 3.7), of the sample autocorrelation function (cf. Figures

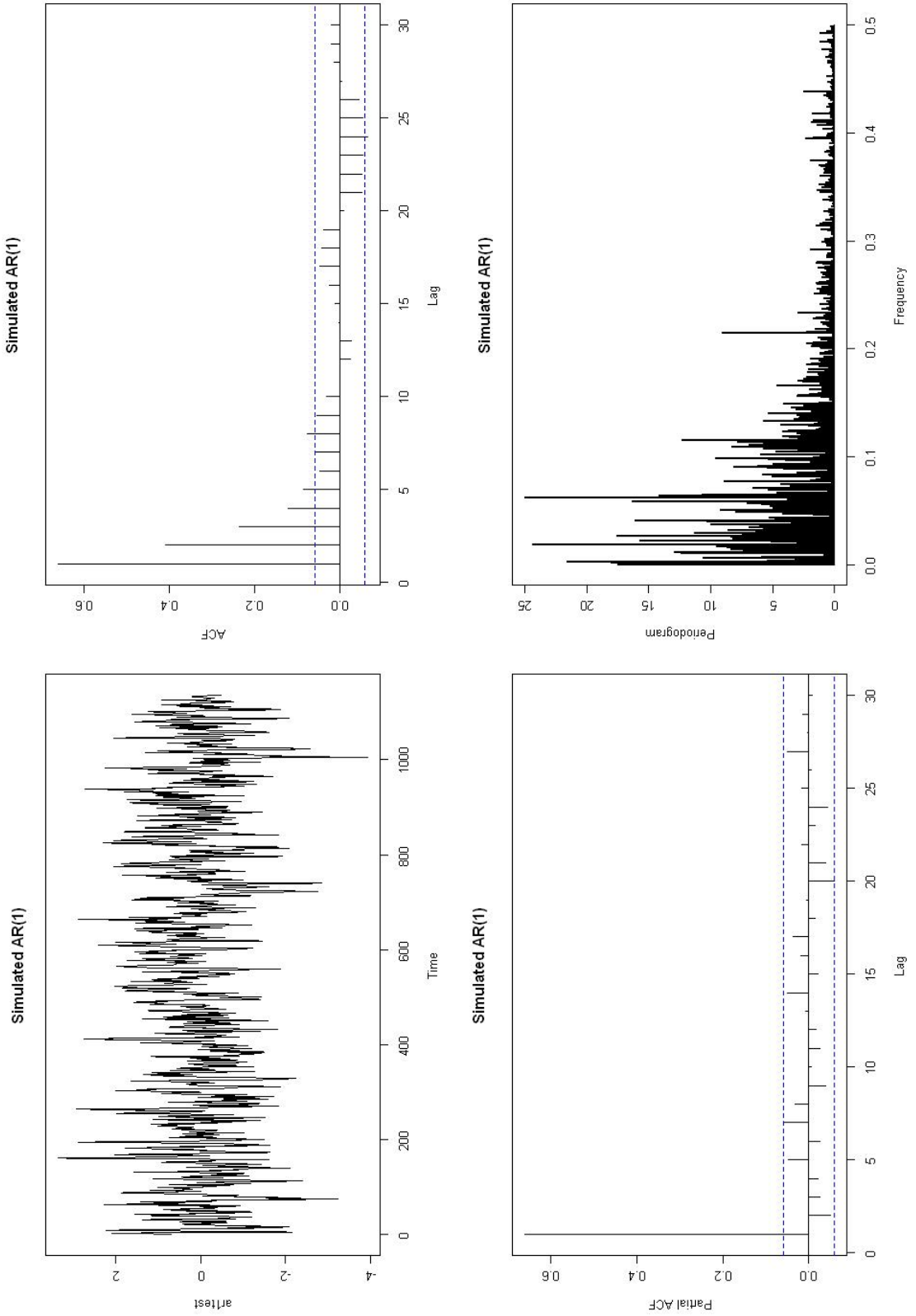

Figure 3.5: Plot of the time series, of the sample ACF, of the sample PACF and of the periodogram of a simulated AR(1)

3.8 and 3.9), of the sample partial autocorrelation function (cf. Figures 3.10 and 3.11) and of the periodogram (cf. Figures 3.12 and 3.13) of the residuals.

The results of the first two methods look quite good. However, the plotted time series has some clusters with bigger volatility and can therefore not be modeled with an AR(1) process. The ACF of the third method hardly tails off and this method will hence be disregarded. In the periodogram of the fourth method, we remark a heavy spike at 0.28 which is a clear hint for a hidden cycle of 3.5 days. The ACF of the fifth method does not tail off as well as its PACF does not cut off after the first lag. The time series plots of the sixth and the seventh method look good, but the ACF does not tail off and the PACF does not cut off. Comparing the plots of the simulated AR(1) (cf. Figure 3.5), we see that the plots of the eighth method fit best. The ACF tails off in the right way and the PACF cuts off after the first lag. The periodogram also looks very well.
So finally, we choose the preprocessing with the eighth method.

Furthermore, we want to test this data for stationarity or if there is a unit root. We use the augmented Dickey-Fuller test (ADF) with lag order 1, a Phillips-Perron test (PP) (both test for a unit root) and the KPSS test for stationarity (for details cf. Banerjee et al. (1993) and Kwiatkowski et al. (1992)). The KPSS test has the null hypothesis 'K0: The time series is stationary.' versus the alternative 'K1: The time series is not stationary.' and ADF and PP test have the null hypothesis 'H0: The time series has a unit root, i.e. the autoregressive coefficient has an absolute value of 1.' against the alternative 'H1: The absolute value of the autoregressive coefficient is smaller than 1.' The tests show the results described in Table 3.3.

	QLD		NSW		VIC		SA	
	stat.	p-value	stat.	p-value	stat.	p-value	stat.	p-value
KPSS	0.0671	> 0.1	0.126	> 0.1	0.099	> 0.1	0.065	> 0.1
	accept K0		accept K0		accept K0		accept K0	
ADF	-12.74	< 0.01	-13.70	< 0.01	-15.11	< 0.01	-15.04	< 0.01
	reject H0		reject H0		reject H0		reject H0	
PP	-334.31	< 0.01	-347.94	< 0.01	-413.37	< 0.01	-421.71	< 0.01
	reject H0		reject H0		reject H0		reject H0	

Table 3.3: Tests for Stationarity and Unit Root for the preprocessed time series using Method 8

For each of the preprocessed time series, we cannot reject the KPSS test for stationarity. In addition, we have to reject the ADF and PP test for unit roots. So, we assume stationarity for all four time series. Therefore, we will model them marginally with an AR(1) process.

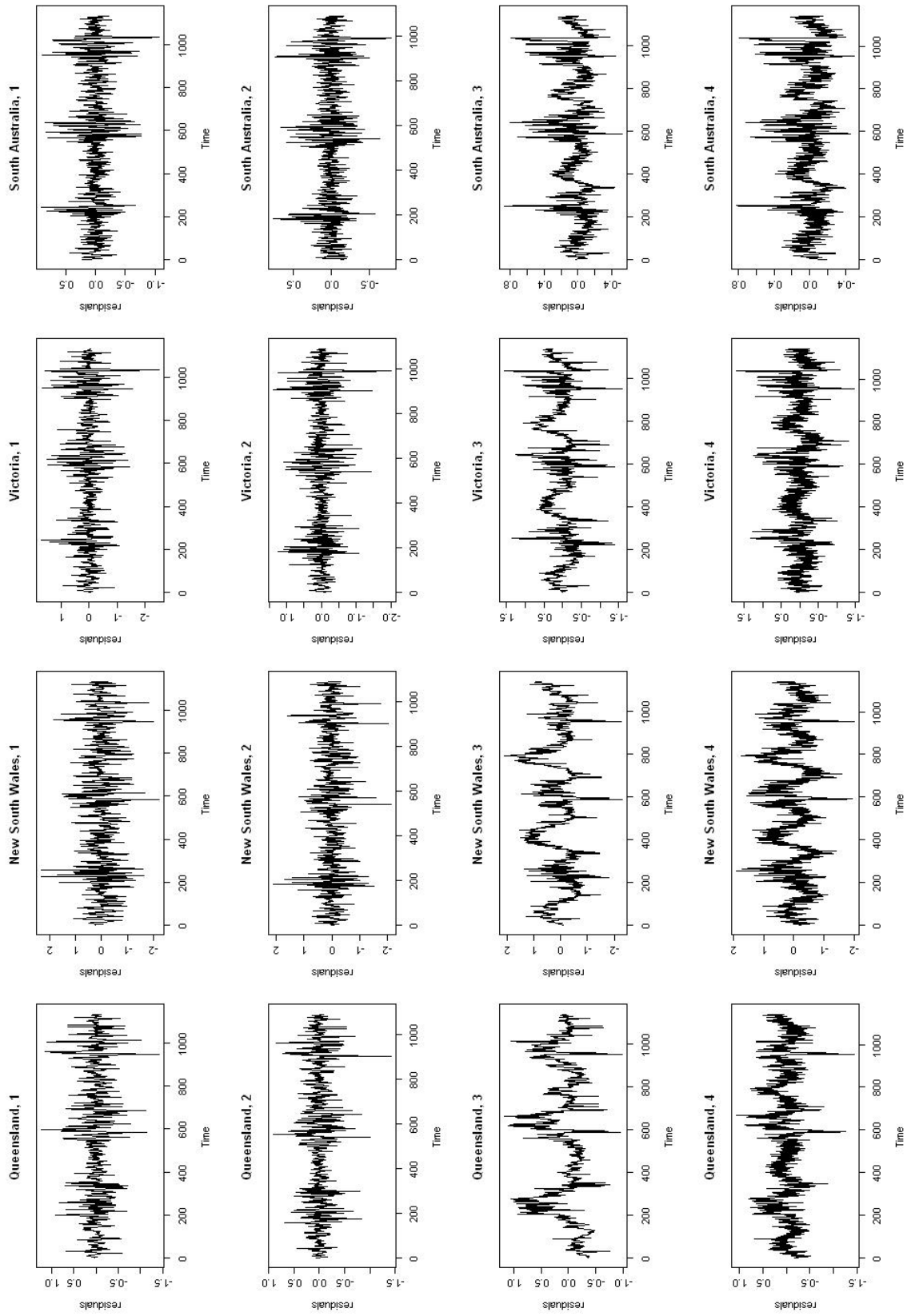

Figure 3.6: Time series plots of the residuals for the methods 1 - 4

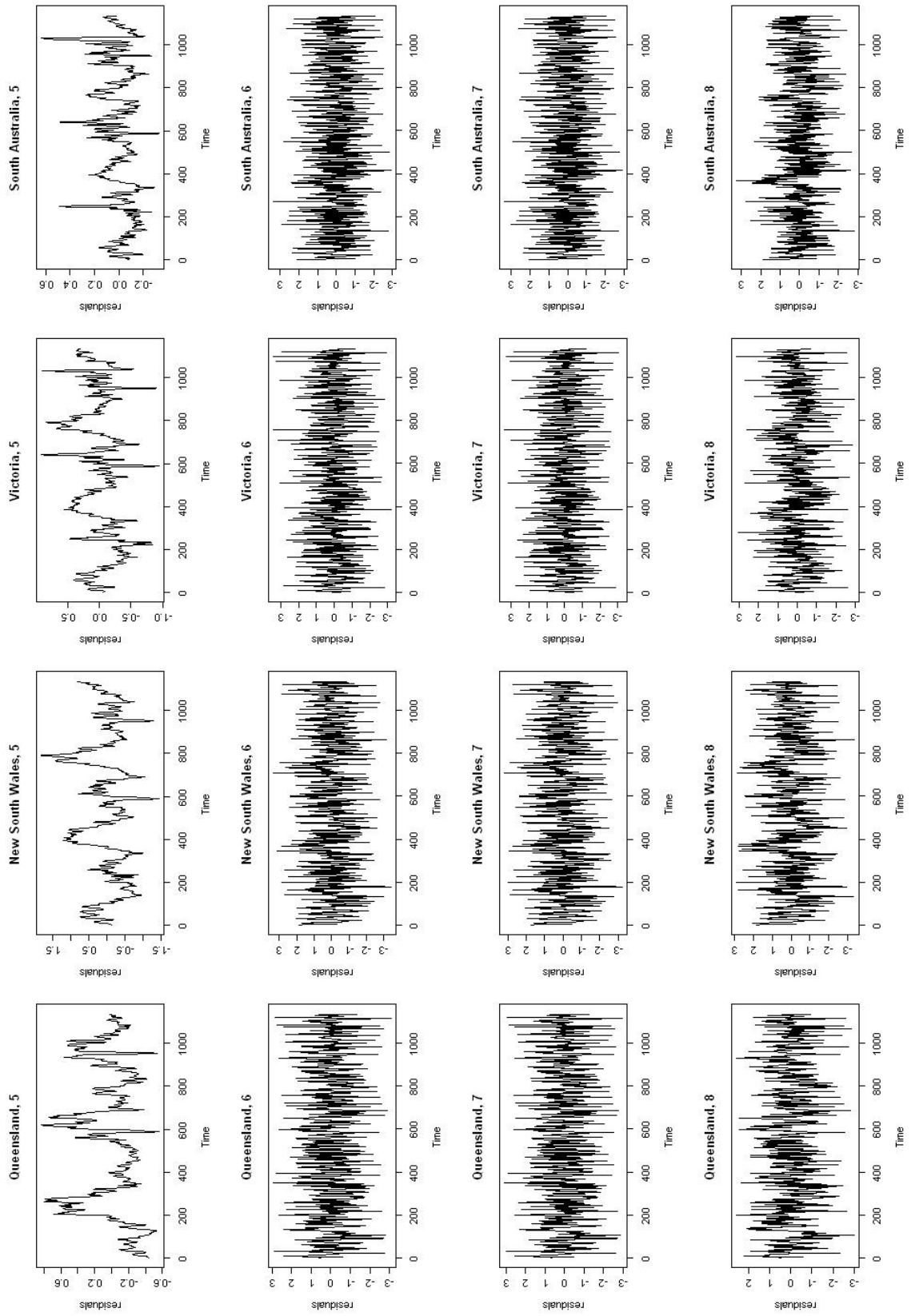

Figure 3.7: Time series plots of the residuals for the methods 5 - 8

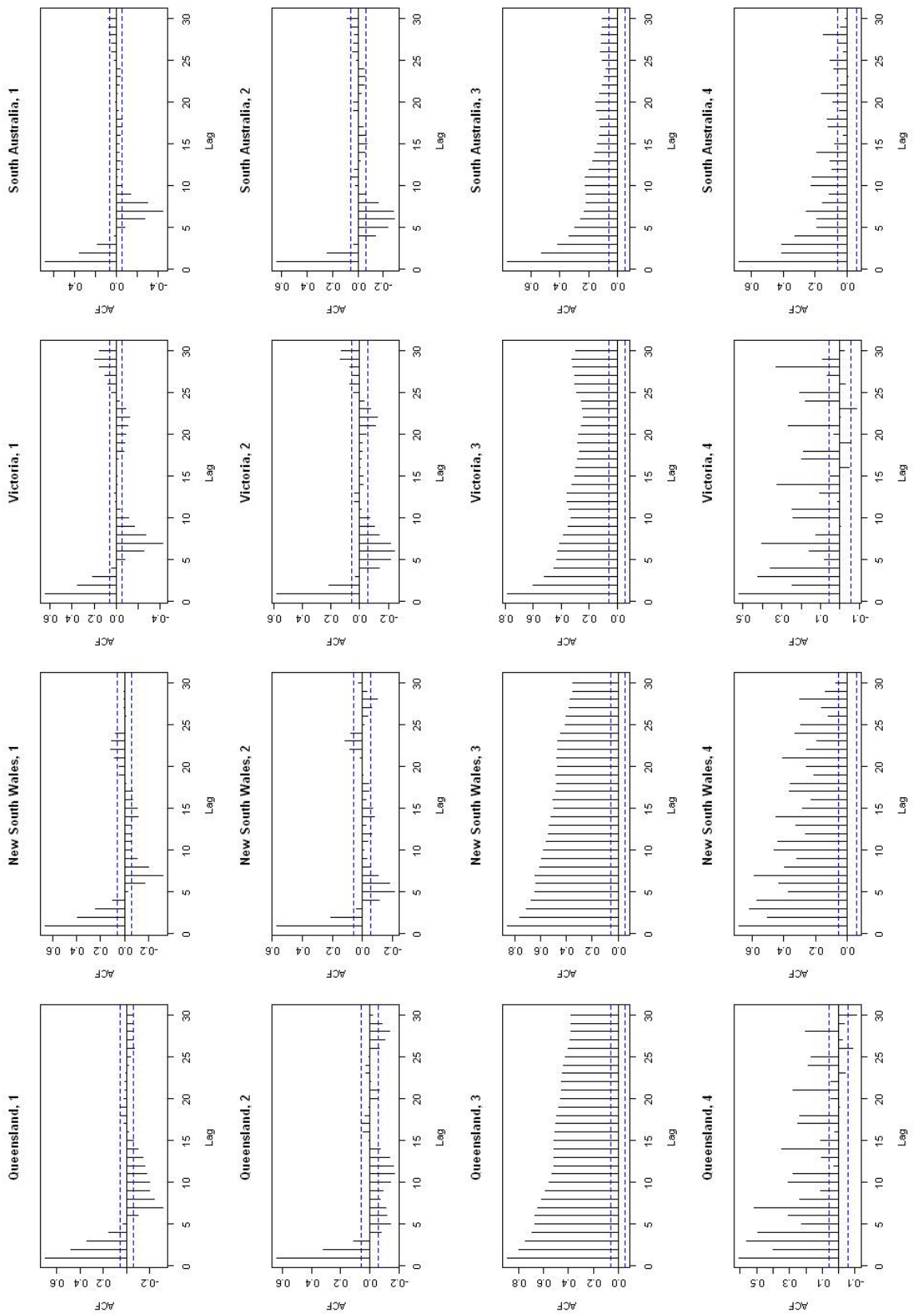

Figure 3.8: Plots of the sample ACF of the residuals for the methods 1 - 4

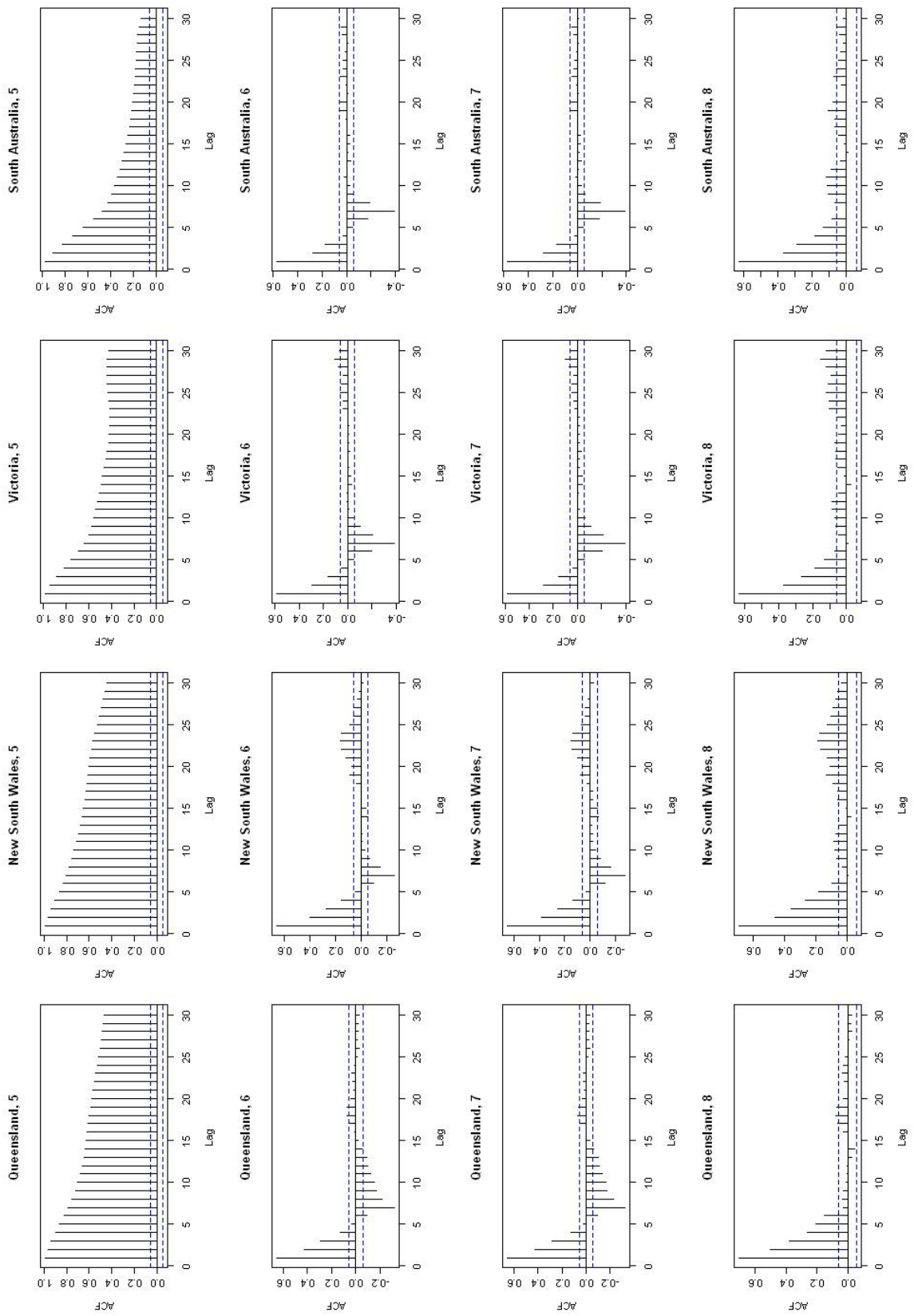

Figure 3.9: Plots of the sample ACF of the residuals for the methods 5 - 8

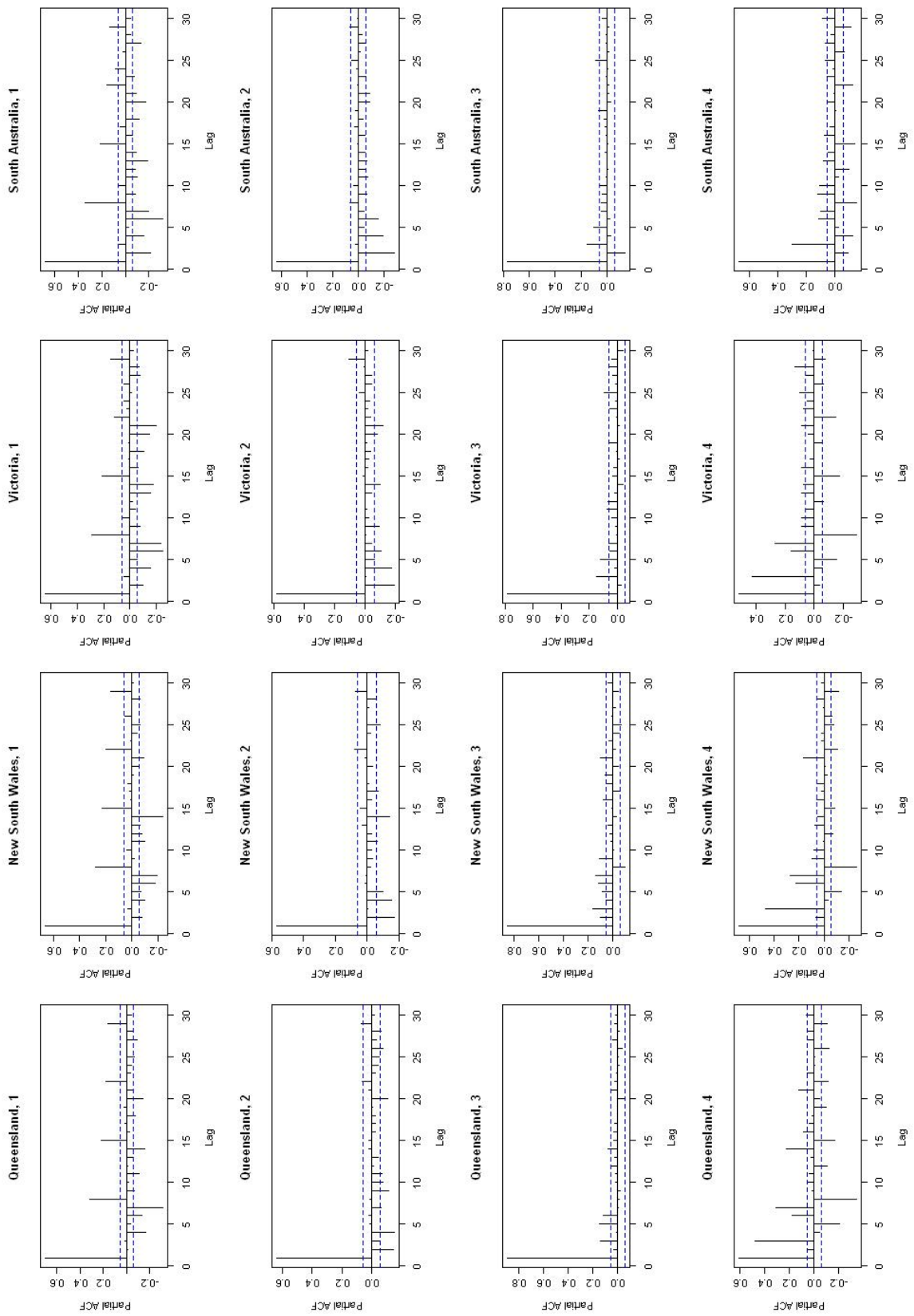

Figure 3.10: Plots of the sample PACF of the residuals for the methods 1 - 4

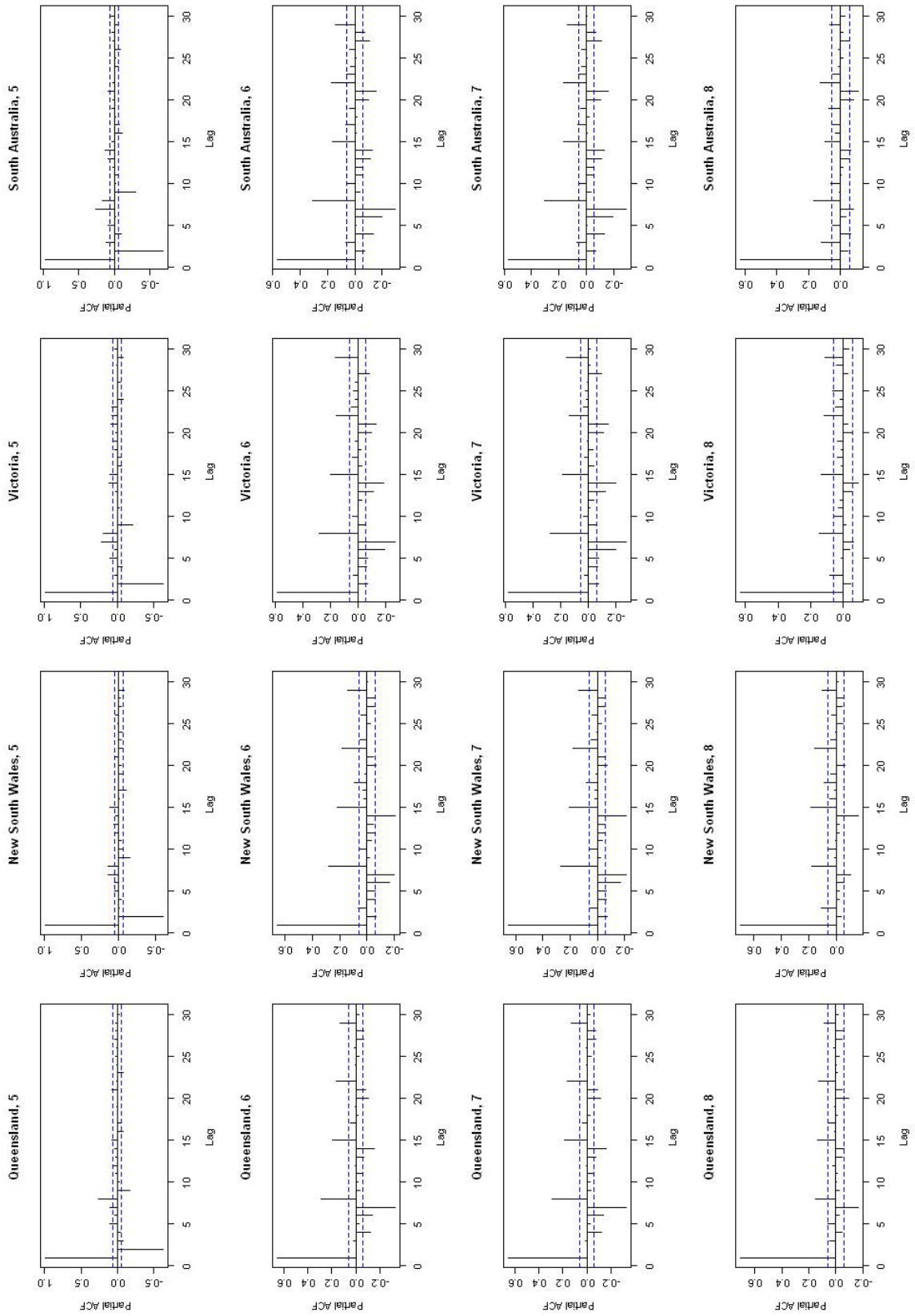

Figure 3.11: Plots of the sample PACF of the residuals for the methods 5 - 8

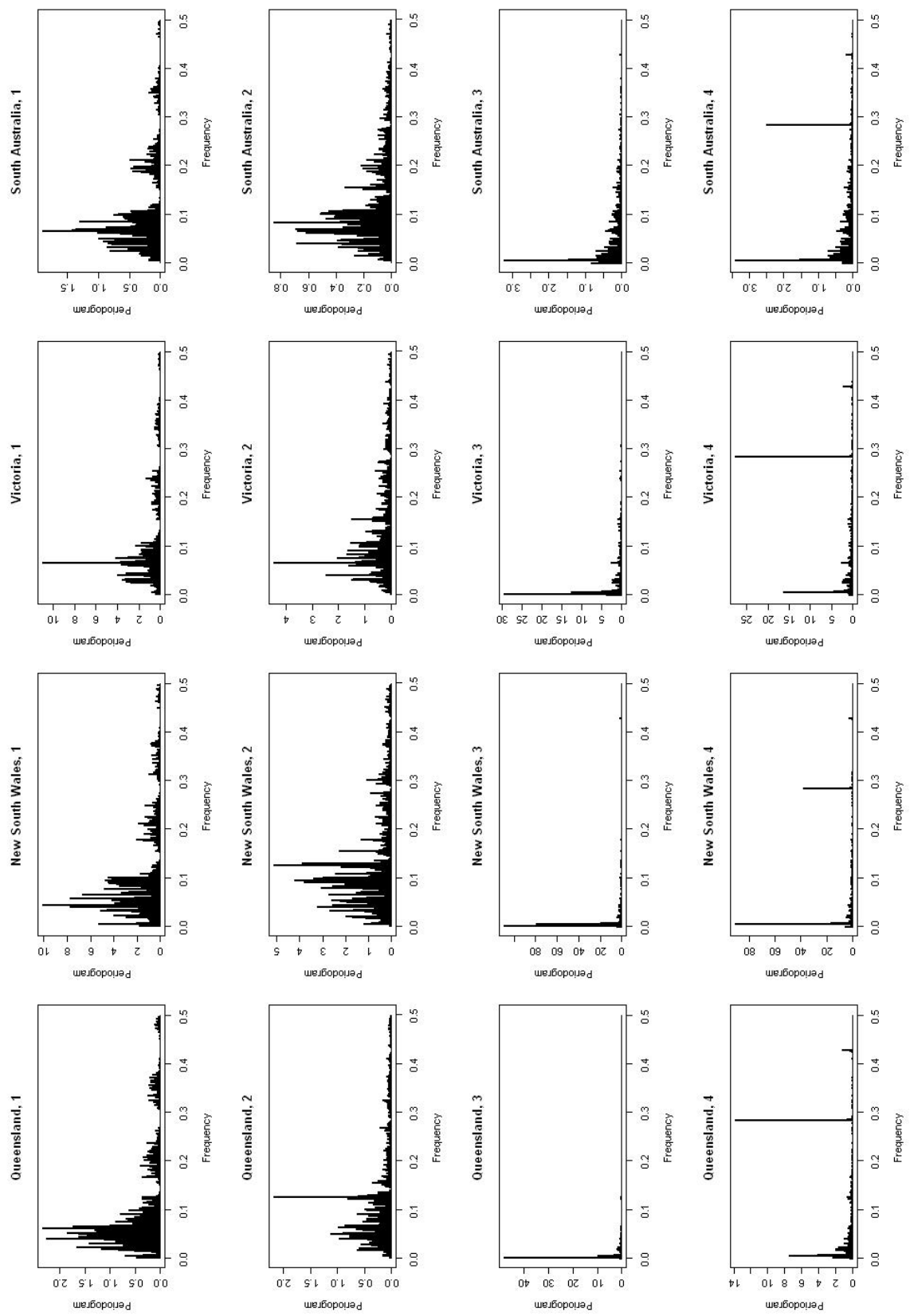

Figure 3.12: Plots of the Periodogram of the residuals for the methods 1 - 4

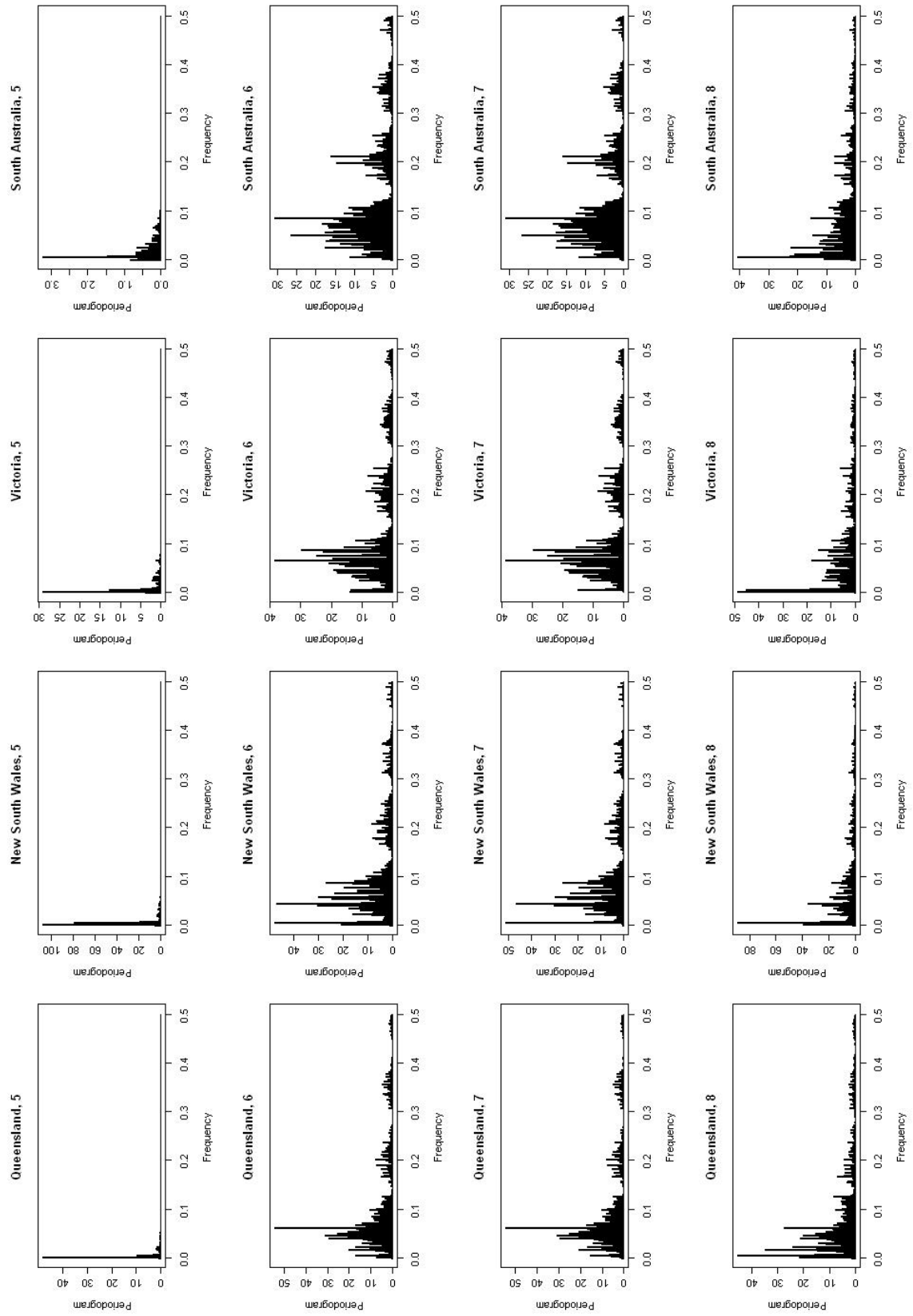

Figure 3.13: Plots of the Periodogram of the residuals for the methods 5 - 8

Chapter 4

Two-Dimensional Modeling

4.1 Basic Procedure

In this first step, we want to concentrate on only two dimensions. We estimate jointly the marginal time series and the copula parameters in one MCMC procedure. Since we have seen that there is a good fit after the preprocessing in Section 3.4, we want to model the marginal time series with an AR(1)-model (cf. (2.12)). Additionally, we describe the dependency with a t-Copula. So, we want to estimate jointly 6 parameters, two for each AR(1)-model and two for the t-Copula.

Let us first denote the marginal time series model. We have two AR(1)-processes with a Gaussian white noise error term denoted by

$$\begin{aligned} &Y_{1t} = \gamma_1 \cdot Y_{1(t-1)} + \epsilon_{1t} \quad t = 1, \dots, T, \quad \epsilon_{1t} \sim \mathcal{N}(0, \sigma_1^2) \; i.i.d. \\ &Y_{10} \sim \mathcal{N}\left(0, \frac{\sigma_1^2}{1-\gamma_1^2}\right), \left|\gamma_1\right| < 1 \end{aligned}$$

and

$$\begin{aligned} &Y_{2t} = \gamma_2 \cdot Y_{2(t-1)} + \epsilon_{2t} \quad t = 1, \dots, T, \quad \epsilon_{2t} \sim \mathcal{N}(0, \sigma_2^2) \; i.i.d. \\ &Y_{20} \sim \mathcal{N}\left(0, \frac{\sigma_2^2}{1-\gamma_2^2}\right), \left|\gamma_2\right| < 1. \end{aligned}$$

The condition $|\gamma_i| < 1, i = 1, 2$ ensures the stationarity of the time series which has a stationary Gaussian distribution with mean 0 and variance $\frac{\sigma_i^2}{1-\gamma_i^2}, i = 1, 2$ (cf. (2.14) and (2.15)). So, we have chosen the distribution of the starting values $Y_{i0}, i = 1, 2$ in this way.

To simplify the notation, we define:

$$\boldsymbol{Y^t} := \begin{pmatrix} Y_{1t} & Y_{2t} \end{pmatrix}, t = 1, \dots, T \qquad \boldsymbol{Y_i} := \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, i = 1, 2.$$

Now we can denote the joint distribution of the marginal time series using the calculations in (2.14) and (2.15):

$$\boldsymbol{Y_i} \sim \mathcal{N}_T(\mathbf{0}, \Sigma_i) \text{ with } \Sigma_i = \frac{\sigma_i^2}{1-\gamma_i^2} \begin{pmatrix} 1 & \gamma_i & \gamma_i^2 & \dots & \gamma_i^{T-1} \\ \gamma_i & 1 & \gamma_i & & \gamma_i^{T-2} \\ \gamma_i^2 & \gamma_i & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \gamma_i \\ \gamma_i^{T-1} & \dots & & \gamma_i & 1 \end{pmatrix}, i = 1, 2.$$

The different marginal time series $\boldsymbol{Y_i}$ are dependent. For getting marginal independent observations, say $\boldsymbol{Z_i} = (Z_{i1}, \dots, Z_{iT})'$, we transform $\boldsymbol{Y_i}$ by multiplying with $\Sigma_i^{-1/2}$. Therefore

$$\boldsymbol{Z_i} := \Sigma_i^{-1/2} \boldsymbol{Y_i} \sim \mathcal{N}_T(\mathbf{0}, I_T),$$

where I_T is the identity matrix of size T.

Wise (1955) shows that Σ_i^{-1} is a tridiagonal matrix of the following form

$$\Sigma_i^{-1} = \sigma_i^{-2} \begin{pmatrix} 1 & -\gamma_i & 0 & \dots & \dots & 0 \\ -\gamma_i & 1+\gamma_i^2 & -\gamma_i & 0 & & \\ 0 & -\gamma_i & 1+\gamma_i^2 & -\gamma_i & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -\gamma_i & 1+\gamma_i^2 & -\gamma_i \\ 0 & \dots & \dots & 0 & -\gamma_i & 1 \end{pmatrix}$$

and

$$\det \Sigma_i^{-1} = \sigma_i^{-2T}(1-\gamma_i^2).$$

Now we have to find the square root $\Sigma_i^{-1/2}$. We can do this by applying the Cholesky Decomposition for positive definite matrices (cf. Algorithm 2.1) yielding

$$\Sigma_i^{-1/2} = \sigma_i^{-1} \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -\gamma_i & 1 & 0 & & \vdots \\ 0 & -\gamma_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\gamma_i & \sqrt{1-\gamma_i^2} \end{pmatrix}$$

and

$$\det \Sigma_i^{-1/2} = \sigma_i^{-T} \sqrt{1-\gamma_i^2}.$$

So we have all necessary utilities for the transformation of $\boldsymbol{Y_i}$ to $\boldsymbol{Z_i}$ and vice versa using the transformation theorem 2.4:

$$\begin{aligned} \boldsymbol{Y_i} &= g(\boldsymbol{Z_i}) = \Sigma_i^{1/2}\boldsymbol{Z_i} \\ \Rightarrow g^{-1}(\boldsymbol{Y_i}) &= \boldsymbol{Z_i} = \Sigma_i^{-1/2}\boldsymbol{Y_i} \\ Dg^{-1}(\boldsymbol{Y_i}) &= \Sigma_i^{-1/2} \end{aligned} \tag{4.1}$$

where Df is the Jacobian of f and

$$|\det Dg^{-1}(\boldsymbol{Y_i})| = |\det \Sigma_i^{-1/2}| = \sigma_i^{-T}\sqrt{1-\gamma_i^2}.$$

We can denote the relationship between z_{it} and y_{it} for each single t:

$$\begin{aligned} z_{i1} &= \frac{1}{\sigma_i} y_{i1} \\ z_{i2} &= \frac{1}{\sigma_i}(-\gamma_i y_{i1} + y_{i2}) \\ z_{i3} &= \frac{1}{\sigma_i}(-\gamma_i y_{i2} + y_{i3}) \\ &\vdots \\ z_{i(T-1)} &= \frac{1}{\sigma_i}(-\gamma_i y_{i(T-2)} + y_{i(T-1)}) \\ z_{iT} &= \frac{1}{\sigma_i}(-\gamma_i y_{i(T-1)} + \sqrt{1-\gamma_i^2} y_{iT}). \end{aligned}$$

Our next step is the calculation of the joint likelihood of $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$. Since each of these time series is marginally dependent on t, we start with $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ and use the marginal independence of these values to simplify the likelihood:

$$\begin{aligned} &f(\boldsymbol{z_1}, \boldsymbol{z_2}|\sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2, \rho, \nu) = \\ &\overset{\boldsymbol{z^t} \text{ ind.}}{=} \prod_{t=1}^{T} f_t(z_{1t}, z_{2t}|\sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2, \rho, \nu) = \prod_{t=1}^{T} c_{\nu,\rho}^t(\Phi(z_{1t}), \Phi(z_{2t})) \cdot \varphi(z_{1t}) \cdot \varphi(z_{2t}). \end{aligned}$$

Hereby denotes Φ the cumulative distribution function and φ is the probability density function of the standard normal distribution. The dependency in-between both time series is modeled using the t-Copula (2.37).

We now transform these values to the observed ones by the transformation mentioned

above (cf. (4.1)):

$$
\begin{aligned}
&f(\boldsymbol{y_1}, \boldsymbol{y_2}|\sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2, \rho, \nu) = \qquad (4.2)\\
&\overset{(4.1)}{=} c^t_{\nu,\rho}(\Phi(\frac{1}{\sigma_1}y_{11}), \Phi(\frac{1}{\sigma_2}y_{21})) \cdot \varphi(\frac{1}{\sigma_1}y_{11}) \cdot \varphi(\frac{1}{\sigma_2}y_{21}) \cdot\\
&\cdot c^t_{\nu,\rho}(\Phi(\frac{1}{\sigma_1}(y_{12} - \gamma_1 y_{11})), \Phi(\frac{1}{\sigma_2}(y_{22} - \gamma_2 y_{21}))) \cdot \varphi(\frac{1}{\sigma_1}(y_{12} - \gamma_1 y_{11})) \cdot \varphi(\frac{1}{\sigma_2}(y_{22} - \gamma_2 y_{21})) \cdot\\
&\cdot c^t_{\nu,\rho}(\Phi(\frac{1}{\sigma_1}(y_{13} - \gamma_1 y_{12})), \Phi(\frac{1}{\sigma_2}(y_{23} - \gamma_2 y_{22}))) \cdot \varphi(\frac{1}{\sigma_1}(y_{13} - \gamma_1 y_{12})) \cdot \varphi(\frac{1}{\sigma_2}(y_{23} - \gamma_2 y_{22})) \cdot\\
&\cdot \ldots \cdot\\
&\cdot c^t_{\nu,\rho}(\Phi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T} - \gamma_1 y_{1(T-1)})), \Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T} - \gamma_2 y_{2(T-1)}))) \cdot\\
&\cdot \varphi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T} - \gamma_1 y_{1(T-1)})) \cdot \varphi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T} - \gamma_2 y_{2(T-1)})) \cdot\\
&\cdot \sigma_1^{-T}\sqrt{1-\gamma_1^2} \cdot \sigma_2^{-T}\sqrt{1-\gamma_2^2}.
\end{aligned}
$$

For calculating the posterior distributions of the parameters (up to constants), we need their prior distributions since

$$f(\theta|x) \propto f(x|\theta)\pi(\theta).$$

We want to do the update of the parameters during one MH-step sequentially. Therefore we will use the independence proposal MH-algorithm (cf. (2.48) and Algorithm 2.5): we will sample the values from a normal distribution with mean θ_{mode} of the posterior distribution and with the inverse Hessian matrix $[-\frac{\partial^2 log f(x)}{\partial x^2}]^{-1}$ evaluated at the mode, where $f(\theta)$ is the posterior distribution of the unknown parameter. We will determine these values by an optimization routine by maximizing the posterior density.

To ease the simulation and the optimization for finding the mode of the time series parameters, we transform them to $(-\infty, \infty)$ by applying the logarithm transformation for σ_i^2 and the Fisher-z-transformation for γ_i and ρ using the transformation theorem 2.4, respectively.

$$
\begin{aligned}
s_i &:= \log(\sigma_i^2) = 2\log\sigma_i \qquad (4.3)\\
\Rightarrow \sigma_i^2 &= \exp(s_i) \qquad \sigma_i = \sqrt{\exp(s_i)} = \exp\left(\frac{1}{2}s_i\right)\\
\frac{\partial\sigma_i^2}{\partial s_i} &= \exp(s_i)
\end{aligned}
$$

and

$$
\begin{aligned}
a_i &:= \frac{1}{2}\log\left(\frac{1+\gamma_i}{1-\gamma_i}\right) \qquad (4.4)\\
\Rightarrow \gamma_i &= \tanh(a_i)\\
\frac{\partial\gamma_i}{\partial a_i} &= \frac{1}{\cosh^2(a_i)}
\end{aligned}
$$

and

$$\begin{aligned} r &:= \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) \\ \Rightarrow \rho &= \tanh(r) \\ \frac{\partial\rho}{\partial r} &= \frac{1}{\cosh^2(r)}. \end{aligned} \tag{4.5}$$

All in all, we develop the posterior density (up to constants) first on the untransformed level. The optimization of the posterior distribution of ρ, $\sigma_i, i = 1, 2$ and $\gamma_i, i = 1, 2$ for finding the mode and the Hessian is performed on the transformed scale of $(-\infty, \infty)$. The optimization for μ is done on the original scale. Then, if necessary, we transform back to the original scale. Now, we know the mode and the Hessian and we can run the MH-algorithm on the untransformed level.

Using the results of Lewandowski, Kurowicka, and Joe (2007) and the calculations of Joe (2006), if we suppose a $Beta(1, 1)$ distribution on (-1,1) (cf. (2.2)) for ρ, the resulting matrix of partial correlation is uniformly distributed over the space of correlation matrices. Regarding ν, we suppose a half Cauchy (1,2)-prior (cf (2.3) and Ausin and Lopes (2008)) since Bauwens and Lubrano (1998) showed that the usual flat prior would lead to an improper posterior distribution. For σ_i^2 we assume an Inverse Gamma prior with shape parameter 1 and scale parameter 0.001 (cf. Congdon (2003) and (2.1)). We transform the loglikelihood of γ_i first to a_i using (4.4) and then, we assume a Normal(0,10) prior on $(-\infty, \infty)$. So:

- Prior for ρ: Beta(1, 1) on (-1,1) = Unif(1,1)
  $\pi(\rho) \propto 1$
  then Fisher-z-transformation (4.5) of the posterior distribution to r on $(-\infty, \infty)$
- Prior for ν: half Cauchy(1,2)
  $\pi(\nu) \propto \frac{1}{1+(\frac{\nu-1}{2})^2}, \qquad \nu \in (1, \infty)$
- Prior for γ_i: Fisher-z-transformation (4.4) of the loglikelihood to a_i on $(-\infty, \infty)$, then $\mathcal{N}(0, 10)$
  $\pi(a_i) \propto \exp(-\frac{a_i^2}{200})$
- Prior for σ_i^2: Inverse Gamma(1, 0.001)
  then Logarithm-transformation (4.3) of the posterior distribution to s_i on $(-\infty, \infty)$

$$\begin{aligned} \pi(u) &= \frac{0.001}{\Gamma(1)} u^{-2} \exp\left(-\frac{0.001}{u}\right) \qquad \text{(Congdon (2003, p. 15))} \\ \pi(s_i) &\overset{(4.3)}{=} \frac{0.001}{\Gamma(1)} \exp(-2s_i) \exp\left(-\frac{0.001}{\exp(s_i)}\right) \exp(s_i) \\ &= \frac{0.001}{\Gamma(1)} \exp(-s_i) \exp(-0.001 \exp(-s_i)) \\ \log \pi(s_i) &\propto -s_i - 0.001 \exp(-s_i) \end{aligned}$$

These are all necessary prerequisites for applying the independence MH-algorithm. We just have to calculate the mode θ_{mode} and the inverse of the Hessian $[-\frac{\partial^2 log f(x)}{\partial x^2}]^{-1}$ for each posterior density using

$$f(\theta|x) \propto f(x|\theta)\pi(\theta). \tag{4.6}$$

So, we get the following posterior densities on the transformed level up to proportionality constants (included the corresponding Jacobians of the transformations) with the likelihood $f(\boldsymbol{y_1}, \boldsymbol{y_2}|\nu, \rho, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2)$ denoted in (4.2):

$$f(\nu|\boldsymbol{y_1}, \boldsymbol{y_2}, \rho, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2) \propto f(\boldsymbol{y_1}, \boldsymbol{y_2}|\nu, \rho, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2) \cdot \frac{1}{1 + (\frac{\nu-1}{2})^2} \tag{4.7}$$

$$f(r|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2) \propto f(\boldsymbol{y_1}, \boldsymbol{y_2}|\nu, \tanh(r), \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2) \cdot 1 \cdot \frac{1}{\cosh^2(r)} \tag{4.8}$$

$$\begin{aligned} f(s_i|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \rho, \gamma_i, \sigma_j^2, \gamma_j) \; &\propto f(\boldsymbol{y_1}, \boldsymbol{y_2}|\nu, \rho, \exp(s_i), \gamma_i, \sigma_j^2, \gamma_j) \cdot \exp(-s_i) \cdot \\ &\cdot \exp(-0.001 \cdot \exp(-s_i)) \cdot \exp(s_i) \end{aligned} \tag{4.9}$$

$$\begin{aligned} f(a_i|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \rho, \sigma_i^2, \sigma_j^2, \gamma_j) &\propto f(\boldsymbol{y_1}, \boldsymbol{y_2}|\nu, \rho, \sigma_i^2, \tanh(a_i), \sigma_j^2, \gamma_j) \cdot \\ &\cdot \exp(-\frac{a_i^2}{200}) \cdot \frac{1}{\cosh^2(a_i)} \end{aligned} \tag{4.10}$$

We optimize the right-hand side of the above formula because the knowledge of the posterior density up to constants is enough for finding the necessary mode and Hessian. However, we have to notice another important point. Since we optimize the posterior densities for ρ, γ_i and σ_i^2 on a transformed scale, we have to retransform the determined mode and Hessian to the original scale using the Delta method (cf. Theorem 2.7). This procedure holds only for univariate parameters and its estimator.
Following the central limit theorem, it holds for the univariate mode m and its estimator $\hat{m}$ of the posterior distribution that

$$\sqrt{n}(\hat{m} - m) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \frac{1}{h}\right)$$

where h is the value of the Hessian evaluated at the mode. Using the univariate Delta method, we get:

$$\begin{aligned} \sqrt{n}(\hat{a}_i - a_i) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \frac{1}{h}\right) \\ \sqrt{n}(g(\hat{a}_i) - g(a_i)) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, g'(\hat{a}_i)^2 \frac{1}{h}\right) \\ \sqrt{n}(\tanh(\hat{a}_i) - \tanh(a_i)) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \left(\frac{1}{\cosh^2(\hat{a}_i)}\right)^2 \frac{1}{h}\right) \\ \sqrt{n}(\hat{\gamma}_i - \gamma_i) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \left(\frac{1}{\cosh^2(\hat{a}_i)}\right)^2 \frac{1}{h}\right) \end{aligned} \tag{4.11}$$

and

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \left(\frac{1}{\cosh^2(\hat{r})}\right)^2 \frac{1}{h}\right) \tag{4.12}$$

and

$$\begin{aligned}
\sqrt{n}(\hat{s_i} - s_i) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \frac{1}{h}\right) \\
\sqrt{n}(g(\hat{s_i}) - g(s_i)) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, g'(\hat{s_i})^2 \frac{1}{h}\right) \\
\sqrt{n}(\exp(\hat{s_i}) - \exp(s_i)) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, (\exp(\hat{s_i}))^2 \frac{1}{h}\right) \\
\sqrt{n}(\hat{\sigma_i^2} - \sigma_i^2) &\xrightarrow{\mathscr{D}} \mathcal{N}\left(0, (\exp(\hat{s_i}))^2 \frac{1}{h}\right).
\end{aligned} \tag{4.13}$$

So, we have to multiply the calculated variance on the transformed scale by $\frac{1}{\cosh^2(\hat{a_i})}$, $\frac{1}{\cosh^2(\hat{r})}$ and $\exp(\hat{s_i})$, respectively, to return to the original scale and to get the mode and the Hessian there which we need for the independence proposal MH-algorithm.

At first, we denote the algorithms for one step for one single parameter (cf. Algorithms 4.1 - 4.4).

Algorithm 4.1 One step of independence MH-Algorithm for ν

1: Optimize logarithm of the posterior density $f(\nu|\boldsymbol{y_1}, \boldsymbol{y_2}, \rho, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2)$ (cf. (4.7)) of ν for finding mode ν_m and Hessian $h(\nu_m)$
2: Draw a proposal value ν_{prop} from $\mathcal{N}(\nu_m, 1/h(\nu_m))$
3: Calculate acceptance probability $\alpha(\nu_{old}, \nu_{prop})$ (cf. (2.48))
4: Draw a Uniform (0,1) random variable u_1
5: **IF** $u_1 < \alpha(\nu_{old}, \nu_{prop})$ **THEN**
6: $\quad \nu_{new} := \nu_{prop}$
7: **ELSE**
8: $\quad \nu_{new} := \nu_{old}$
9: **END IF**

Algorithm 4.2 One step of independence MH-Algorithm for ρ

1: Transform ρ to r
2: Optimize the logarithm of the posterior density $f(r|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \sigma_1^2, \gamma_1, \sigma_2^2, \gamma_2)$ (cf. (4.8)) of r for finding mode r_m and Hessian $h(r_m)$
3: Transform mode and Hessian back to ρ_m and $h(\rho_m)$ using (4.12)
4: Draw a proposal value ρ_{prop} from $\mathcal{N}(\rho_m, 1/h(\rho_m))$
5: Calculate acceptance probability $\alpha(\rho_{old}, \rho_{prop})$ (cf. (2.48))
6: Draw a Uniform (0,1) random variable u_2
7: **IF** $u_2 < \alpha(\rho_{old}, \rho_{prop})$ **THEN**
8: $\quad \rho_{new} := \rho_{prop}$
9: **ELSE**
10: $\quad \rho_{new} := \rho_{old}$
11: **END IF**

Algorithm 4.3 One step of independence MH-Algorithm for $\sigma_i^2, i = 1, 2$

1: Transform σ_i^2 to s_i
2: Optimize the logarithm of the posterior density $f(s_i|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \rho, \gamma_i, \sigma_j^2, \gamma_j)$ (cf. (4.9)) of s_i for finding mode $s_{i,m}$ and Hessian $h(s_{i,m})$
3: Transform mode and Hessian back to $\sigma_{i,m}^2$ and $h(\sigma_{i,m}^2)$ using (4.13)
4: Draw a proposal value $\sigma_{i,prop}^2$ from $\mathcal{N}(\sigma_{i,m}^2, 1/h(\sigma_{i,m}^2))$
5: Calculate acceptance probability $\alpha(\sigma_{i,old}^2, \sigma_{i,prop}^2)$ (cf. (2.48))
6: Draw a Uniform (0,1) random variable u_3
7: **IF** $u_3 < \alpha(\sigma_{i,old}^2, \sigma_{i,prop}^2)$ **THEN**
8: $\quad \sigma_{i,new}^2 := \sigma_{i,prop}^2$
9: **ELSE**
10: $\quad \sigma_{i,new}^2 := \sigma_{i,old}^2$
11: **END IF**

Algorithm 4.4 One step of independence MH-Algorithm for $\gamma_i, i = 1, 2$

1: Transform γ_i to a_i
2: Optimize the logarithm of the posterior density $f(a_i|\boldsymbol{y_1}, \boldsymbol{y_2}, \nu, \rho, \sigma_i^2, \sigma_j^2, \gamma_j)$ (cf. (4.10)) of a_i for finding mode $a_{i,m}$ and Hessian $h(a_{i,m})$
3: Transform mode and Hessian back to $\gamma_{i,m}$ and $h(\gamma_{i,m})$ using (4.11)
4: Draw a proposal value $\gamma_{i,prop}$ from $\mathcal{N}(\gamma_{i,m}, 1/h(\gamma_{i,m}))$
5: Calculate acceptance probability $\alpha(\gamma_{i,old}, \gamma_{i,prop})$ (cf. (2.48))
6: Draw a Uniform (0,1) random variable u_4
7: **IF** $u_4 < \alpha(\gamma_{i,old}, \gamma_{i,prop})$ **THEN**
8: $\quad \gamma_{i,new} := \gamma_{i,prop}$
9: **ELSE**
10: $\quad \gamma_{i,new} := \gamma_{i,old}$
11: **END IF**

Then we can denote the algorithm for the joint estimation of time series and copula parameters using the independence Metropolis-Hastings sampler (cf. Algorithm 4.5).

Algorithm 4.5 Algorithm for the Joint Estimation of Time Series and Copula Parameters with N Steps in 2 Dimensions

1: **FOR** $t = 1, \ldots, N$ **DO**
2: Calculate new value for ν using Algorithm 4.1
3: Calculate new value for ρ using Algorithm 4.2
4: Calculate new value for σ_1^2 using Algorithm 4.3
5: Calculate new value for γ_1 using Algorithm 4.4
6: Calculate new value for σ_2^2 using Algorithm 4.3
7: Calculate new value for γ_2 using Algorithm 4.4
8: **END FOR**

4.2 Simulation Study

To assess the power of the implemented algorithm, we perform a simulation study. We choose 12 scenarios, create the corresponding data and estimate the parameters using marginal MLE and C-MLE (TSP-MLE, cf. Table 2.1) and using our algorithm 4.5.
Therefore, we draw the data on the z-level from a t-Copula with standard normal margins and transform them to the y-Level using the transformation (4.1). Each sample consists of 2000 simulated values, 1000 for each marginal time series. Then we estimate the time series parameters first by a Maximum Likelihood estimation (marginal MLE). We transform the observed values parametricly to the unit cube and perform a Maximum Likelihood estimation of the copula parameters (TSP-MLE). We use these values as starting values for our MCMC estimation with the independence MH-algorithm (JMCMC).
We repeat all 12 scenarios four times and calculate then the mean and the standard deviation over all four replications for each parameter.

We choose the possible values 3 and 10 for ν and the possible values 0.4 and 0.9 for ρ. We keep the variance $\frac{\sigma_i^2}{1-\gamma_i^2}$ of the stationary distribution of the marginal time series constant to 5 and we calculate corresponding parameter values for γ_i and σ_i^2. This yields the following 12 scenarios:

	ν	ρ	γ_1	σ_1^2	γ_2	σ_2^2
1	3	0.40	0.90	0.95	0.90	0.95
2	3	0.40	0.40	4.20	0.40	4.20
3	3	0.40	0.40	4.20	0.90	0.95
4	3	0.90	0.90	0.95	0.90	0.95
5	3	0.90	0.40	4.20	0.40	4.20
6	3	0.90	0.40	4.20	0.90	0.95
7	10	0.40	0.90	0.95	0.90	0.95
8	10	0.40	0.40	4.20	0.40	4.20
9	10	0.40	0.40	4.20	0.90	0.95
10	10	0.90	0.90	0.95	0.90	0.95
11	10	0.90	0.40	4.20	0.40	4.20
12	10	0.90	0.40	4.20	0.90	0.95

Table 4.1: Overview of the 12 investigated simulation scenarios

The simulation study shows the results for the average values and the standard deviations based on 4 replications printed in Tables 4.2 - 4.5.
For the low degrees of freedom of 3, we can see that the mean determined by the MCMC mostly estimates better than the value determined by TSP-MLE while the standard deviation has a tendency to be smaller than that of the TSP-MLE values. For $\nu = 10$, TSP-MLE is considerably better than the posterior mean only in the case that the correlation is high with a high variance for both marginal time series. In all other cases, the posterior mean performs as well as the TSP-MLE value with an often smaller standard deviation. The correlation parameter ρ is very well estimated by the posterior mean as well as by the TSP-MLE values. The standard deviation of the posterior mean appears to be slightly smaller. Only the estimated correlation in the seventh scenario has a bigger difference to the real value for both estimators. The values of σ^2 and γ are overall estimated well with no clear advantage for any of the estimators.
So generally, both estimation methods are very close to each other.

			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
1	ν	value	3.000	2.245	2.320	2.947	3.810	3.940	2.987	2.873	2.935
		sd		0.255	0.283	0.416	0.619	0.637	0.422	0.403	0.430
	ρ	value	0.400	0.353	0.362	0.415	0.465	0.475	0.415	0.425	0.417
		sd		0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.025
	σ_1^2	value	0.950	0.873	0.885	0.950	1.025	1.040	0.953	0.950	0.948
		sd		0.056	0.059	0.065	0.064	0.065	0.061	0.065	0.067
	γ_1	value	0.900	0.880	0.880	0.903	0.920	0.922	0.903	0.903	0.903
		sd		0.018	0.018	0.015	0.012	0.015	0.015	0.015	0.015
	σ_2^2	value	0.950	0.843	0.853	0.922	0.990	1.005	0.922	0.917	0.915
		sd		0.031	0.031	0.031	0.036	0.040	0.031	0.030	0.033
	γ_2	value	0.900	0.883	0.887	0.903	0.923	0.927	0.903	0.903	0.910
		sd		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.014
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
2	ν	value	3.000	2.223	2.302	2.880	3.715	3.958	2.938	2.810	2.877
		sd		0.250	0.259	0.406	0.696	0.817	0.448	0.373	0.422
	ρ	value	0.400	0.320	0.330	0.390	0.443	0.450	0.388	0.393	0.388
		sd		0.037	0.033	0.033	0.033	0.037	0.029	0.033	0.029
	σ_1^2	value	4.200	3.893	3.940	4.230	4.553	4.620	4.235	4.225	4.210
		sd		0.158	0.161	0.174	0.193	0.197	0.177	0.171	0.203
	γ_1	value	0.400	0.362	0.370	0.410	0.450	0.458	0.410	0.410	0.417
		sd		0.019	0.014	0.014	0.014	0.017	0.014	0.014	0.033
	σ_2^2	value	4.200	3.922	3.975	4.272	4.600	4.668	4.282	4.255	4.225
		sd		0.162	0.169	0.182	0.188	0.203	0.186	0.187	0.201
	γ_2	value	0.400	0.348	0.357	0.395	0.435	0.443	0.395	0.395	0.405
		sd		0.017	0.017	0.019	0.019	0.015	0.019	0.019	0.010
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
3	ν	value	3.000	2.308	2.417	3.005	3.812	4.010	3.042	2.950	3.027
		sd		0.335	0.369	0.519	0.757	0.823	0.530	0.502	0.519
	ρ	value	0.400	0.315	0.328	0.380	0.432	0.438	0.380	0.380	0.383
		sd		0.010	0.013	0.014	0.017	0.013	0.014	0.014	0.017
	σ_1^2	value	4.200	3.990	4.045	4.370	4.688	4.758	4.367	4.367	4.298
		sd		0.130	0.136	0.161	0.181	0.190	0.152	0.164	0.174
	γ_1	value	0.400	0.375	0.383	0.420	0.465	0.470	0.420	0.422	0.410
		sd		0.021	0.021	0.016	0.021	0.016	0.016	0.017	0.024
	σ_2^2	value	0.950	0.875	0.883	0.953	1.025	1.040	0.953	0.950	0.945
		sd		0.026	0.022	0.026	0.031	0.027	0.026	0.029	0.024
	γ_2	value	0.900	0.880	0.883	0.900	0.920	0.925	0.900	0.900	0.893
		sd		0.000	0.005	0.000	0.000	0.006	0.000	0.000	0.010

Table 4.2: Estimated posterior mean, mode and quantiles of JMCMC as well as TSP-MLE values for all parameters in scenario 1 - 3 averaged over 4 replications with its empirical standard deviation

			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
4	ν	value	3.000	2.175	2.243	2.917	3.788	3.958	2.958	2.862	3.037
		sd		0.511	0.554	0.828	1.191	1.235	0.849	0.812	0.917
	ρ	value	0.900	0.885	0.885	0.900	0.913	0.915	0.900	0.900	0.900
		sd		0.010	0.010	0.008	0.005	0.010	0.008	0.008	0.008
	σ_1^2	value	0.950	0.900	0.907	0.970	1.045	1.058	0.973	0.970	0.945
		sd		0.022	0.029	0.029	0.029	0.029	0.030	0.029	0.021
	γ_1	value	0.900	0.887	0.887	0.903	0.917	0.920	0.903	0.905	0.900
		sd		0.010	0.010	0.010	0.010	0.012	0.010	0.013	0.008
	σ_2^2	value	0.950	0.910	0.920	0.980	1.058	1.073	0.988	0.980	0.955
		sd		0.014	0.022	0.022	0.022	0.021	0.024	0.022	0.017
	γ_2	value	0.900	0.887	0.890	0.905	0.917	0.920	0.907	0.903	0.900
		sd		0.005	0.000	0.006	0.005	0.008	0.005	0.005	0.008
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
5	ν	value	3.000	2.212	2.317	2.980	3.810	3.995	3.015	2.945	3.045
		sd		0.147	0.171	0.256	0.436	0.482	0.275	0.258	0.307
	ρ	value	0.900	0.880	0.885	0.900	0.910	0.913	0.900	0.900	0.897
		sd		0.008	0.013	0.008	0.008	0.010	0.008	0.008	0.010
	σ_1^2	value	4.200	3.905	3.955	4.223	4.527	4.588	4.228	4.220	4.085
		sd		0.155	0.168	0.207	0.223	0.216	0.202	0.225	0.201
	γ_1	value	0.400	0.380	0.385	0.415	0.450	0.455	0.415	0.415	0.420
		sd		0.027	0.024	0.024	0.027	0.024	0.024	0.024	0.029
	σ_2^2	value	4.200	3.917	3.960	4.225	4.530	4.598	4.232	4.225	4.088
		sd		0.187	0.191	0.218	0.233	0.226	0.214	0.225	0.165
	γ_2	value	0.400	0.385	0.390	0.422	0.455	0.460	0.422	0.422	0.422
		sd		0.019	0.022	0.017	0.019	0.022	0.017	0.017	0.036
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
6	ν	value	3.000	2.175	2.315	2.965	3.845	4.003	3.018	2.862	3.100
		sd		0.455	0.521	0.712	1.025	1.089	0.728	0.656	0.715
	ρ	value	0.900	0.877	0.877	0.897	0.910	0.910	0.897	0.897	0.893
		sd		0.013	0.013	0.013	0.008	0.008	0.013	0.013	0.010
	σ_1^2	value	4.200	3.955	3.998	4.237	4.570	4.627	4.255	4.210	4.157
		sd		0.247	0.251	0.213	0.234	0.227	0.223	0.199	0.212
	γ_1	value	0.400	0.388	0.388	0.412	0.435	0.438	0.412	0.412	0.420
		sd		0.019	0.019	0.017	0.017	0.013	0.017	0.017	0.008
	σ_2^2	value	0.950	0.897	0.912	0.958	1.035	1.050	0.963	0.953	0.948
		sd		0.068	0.068	0.050	0.058	0.059	0.055	0.056	0.085
	γ_2	value	0.900	0.895	0.897	0.907	0.917	0.920	0.907	0.907	0.900
		sd		0.006	0.005	0.005	0.005	0.000	0.005	0.005	0.014

Table 4.3: Estimated posterior mean, mode and quantiles of JMCMC as well as TSP-MLE values for all parameters in scenario 4 - 6 averaged over 4 replications with its empirical standard deviation

			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
7	ν	value	10.000	5.745	6.180	9.992	18.265	18.855	10.875	9.025	10.990
		sd		0.640	0.784	1.882	3.723	3.637	2.047	1.577	2.671
	ρ	value	0.400	0.305	0.318	0.367	0.412	0.422	0.367	0.367	0.370
		sd		0.017	0.019	0.019	0.015	0.017	0.019	0.019	0.014
	σ_1^2	value	0.950	0.865	0.880	0.948	1.015	1.025	0.948	0.945	0.940
		sd		0.017	0.020	0.019	0.017	0.017	0.019	0.017	0.027
	γ_1	value	0.900	0.867	0.873	0.893	0.915	0.917	0.893	0.893	0.883
		sd		0.029	0.024	0.024	0.019	0.022	0.024	0.024	0.039
	σ_2^2	value	0.950	0.840	0.853	0.922	0.992	1.008	0.925	0.922	0.917
		sd		0.014	0.015	0.019	0.022	0.021	0.019	0.019	0.021
	γ_2	value	0.900	0.873	0.875	0.897	0.917	0.920	0.897	0.897	0.893
		sd		0.015	0.017	0.015	0.015	0.012	0.015	0.015	0.021
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
8	ν	value	10.000	5.135	5.468	8.848	16.008	16.803	9.558	7.970	9.367
		sd		0.501	0.500	1.149	2.923	2.614	1.516	0.915	1.265
	ρ	value	0.400	0.338	0.343	0.390	0.440	0.450	0.390	0.392	0.392
		sd		0.022	0.021	0.018	0.018	0.018	0.018	0.021	0.021
	σ_1^2	value	4.200	3.912	3.953	4.245	4.595	4.657	4.252	4.235	4.207
		sd		0.086	0.087	0.097	0.076	0.056	0.088	0.103	0.091
	γ_1	value	0.400	0.352	0.362	0.407	0.453	0.458	0.407	0.407	0.407
		sd		0.024	0.024	0.022	0.024	0.022	0.022	0.022	0.017
	σ_2^2	value	4.200	3.900	3.955	4.263	4.610	4.672	4.265	4.258	4.205
		sd		0.113	0.114	0.116	0.118	0.128	0.115	0.114	0.106
	γ_2	value	0.400	0.357	0.360	0.410	0.453	0.463	0.410	0.410	0.410
		sd		0.022	0.018	0.018	0.015	0.021	0.018	0.018	0.029
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
9	ν	value	10.000	5.635	5.978	9.060	15.545	17.303	9.710	8.380	10.172
		sd		0.932	0.982	1.968	4.839	5.851	2.303	1.761	2.564
	ρ	value	0.400	0.360	0.367	0.417	0.465	0.470	0.417	0.422	0.422
		sd		0.048	0.043	0.043	0.039	0.041	0.043	0.046	0.047
	σ_1^2	value	4.200	3.917	3.968	4.285	4.598	4.655	4.282	4.293	4.242
		sd		0.177	0.181	0.197	0.206	0.216	0.197	0.187	0.177
	γ_1	value	0.400	0.338	0.348	0.390	0.432	0.443	0.390	0.390	0.385
		sd		0.013	0.010	0.008	0.005	0.010	0.008	0.008	0.006
	σ_2^2	value	0.950	0.897	0.907	0.972	1.045	1.058	0.972	0.970	0.978
		sd		0.019	0.022	0.024	0.026	0.029	0.024	0.022	0.030
	γ_2	value	0.900	0.865	0.870	0.890	0.910	0.915	0.887	0.890	0.893
		sd		0.010	0.008	0.008	0.008	0.010	0.005	0.008	0.010

Table 4.4: Estimated posterior mean, mode and quantiles of JMCMC as well as TSP-MLE values for all parameters in scenarios 7 - 9 averaged over 4 replications with its empirical standard deviation

			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
10	ν	value	10.000	5.578	6.150	9.595	16.430	18.803	10.200	8.835	9.438
		sd		0.579	0.921	1.803	4.082	4.273	2.033	1.581	0.651
	ρ	value	0.900	0.890	0.893	0.903	0.915	0.915	0.903	0.903	0.905
		sd		0.012	0.010	0.010	0.006	0.006	0.010	0.010	0.013
	σ_1^2	value	0.950	0.885	0.897	0.963	1.035	1.048	0.963	0.958	0.953
		sd		0.042	0.044	0.050	0.048	0.044	0.050	0.056	0.043
	γ_1	value	0.900	0.877	0.883	0.903	0.913	0.923	0.903	0.903	0.900
		sd		0.010	0.005	0.005	0.005	0.005	0.005	0.005	0.008
	σ_2^2	value	0.950	0.880	0.895	0.955	1.030	1.042	0.958	0.955	0.955
		sd		0.055	0.060	0.062	0.065	0.063	0.061	0.062	0.057
	γ_2	value	0.900	0.883	0.885	0.903	0.923	0.923	0.903	0.903	0.905
		sd		0.005	0.006	0.005	0.005	0.005	0.005	0.005	0.006
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
11	ν	value	10.000	4.918	5.210	8.175	13.877	14.807	8.785	7.495	10.182
		sd		0.759	0.927	2.725	7.057	6.537	3.381	2.075	2.804
	ρ	value	0.900	0.877	0.883	0.897	0.907	0.907	0.897	0.897	0.897
		sd		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
	σ_1^2	value	4.200	3.920	3.978	4.272	4.593	4.658	4.277	4.258	4.160
		sd		0.065	0.067	0.061	0.046	0.066	0.064	0.056	0.060
	γ_1	value	0.400	0.338	0.345	0.385	0.420	0.425	0.385	0.380	0.388
		sd		0.022	0.024	0.024	0.024	0.024	0.024	0.018	0.022
	σ_2^2	value	4.200	3.853	3.902	4.192	4.502	4.572	4.195	4.190	4.088
		sd		0.132	0.133	0.116	0.130	0.121	0.121	0.116	0.132
	γ_2	value	0.400	0.343	0.353	0.390	0.425	0.430	0.390	0.390	0.388
		sd		0.022	0.022	0.022	0.021	0.022	0.022	0.022	0.015
			true	2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
12	ν	value	10.000	6.280	6.713	10.520	19.372	20.440	11.320	9.555	12.110
		sd		1.150	1.499	2.861	8.386	8.455	3.373	2.394	4.425
	ρ	value	0.900	0.893	0.895	0.905	0.915	0.917	0.905	0.907	0.903
		sd		0.005	0.006	0.006	0.006	0.005	0.006	0.005	0.005
	σ_1^2	value	4.200	4.062	4.125	4.447	4.785	4.837	4.452	4.438	4.338
		sd		0.190	0.187	0.199	0.219	0.226	0.201	0.191	0.207
	γ_1	value	0.400	0.360	0.365	0.390	0.412	0.420	0.390	0.390	0.375
		sd		0.023	0.024	0.023	0.026	0.023	0.023	0.023	0.035
	σ_2^2	value	0.950	0.907	0.922	0.995	1.068	1.083	0.995	0.995	0.972
		sd		0.010	0.015	0.013	0.015	0.017	0.013	0.013	0.017
	γ_2	value	0.900	0.880	0.883	0.893	0.905	0.907	0.893	0.895	0.890
		sd		0.008	0.005	0.005	0.010	0.010	0.005	0.010	0.008

Table 4.5: Estimated posterior mean, mode and quantiles of JMCMC as well as TSP-MLE values for all parameters in scenarios 10 - 12 averaged over 4 replications with its empirical standard deviation

4.3 Estimation of Real Data

Now we want to apply our algorithm for estimating the parameters of real data in a bivariate setting. We choose the Australian load data (cf. Section 3.4) which we have preprocessed as described there. The resulting time series consist all of 1135 observations and we have these time series for the states Queensland (QLD), New South Wales (NSW), Victoria (VIC) and South Australia (SA). We form all possible pairs of these four time series and get six different scenarios.
The estimation of the parameters of these pairs yield the results shown in Table 4.6.

We can see that the posterior mode, the posterior mean, the posterior median and the TSP-MLE value are very close together, especially for the correlation and the time series parameters. Since we are interested in the four-dimensional modeling using a D-vine to describe the dependency, we have to determine a certain order of the time series. The geographical position as well as the Australian infrastructure which is concentrated on the territories next to the coast suggests the order 'Queensland, New South Wales, Victoria and South Australia'. This proposal is supported by the results of our estimation since the estimated correlations between these time series are the highest and the estimated degrees of freedom for these copulas are the lowest (i.e. between Queensland and New South Wales, between New South Wales and Victoria and between Victoria and South Australia). So, we choose this order for the parameter estimation of the real data (cf. Section 5.3).

		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
QLD NSW	ν	4.92	5.08	7.03	9.89	10.48	7.16	6.80	7.03
	ρ	0.27	0.28	0.33	0.38	0.39	0.33	0.33	0.33
	σ_1^2	0.41	0.42	0.44	0.48	0.48	0.45	0.44	0.44
	γ_1	0.67	0.68	0.71	0.74	0.75	0.71	0.71	0.71
	σ_2^2	0.47	0.48	0.51	0.55	0.55	0.51	0.51	0.52
	γ_2	0.64	0.64	0.68	0.71	0.71	0.68	0.68	0.70
		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
QLD VIC	ν	5.39	5.66	7.62	11.43	12.19	7.95	7.36	7.75
	ρ	0.08	0.09	0.15	0.21	0.22	0.15	0.16	0.16
	σ_1^2	0.41	0.41	0.44	0.47	0.47	0.44	0.44	0.44
	γ_1	0.67	0.67	0.71	0.74	0.74	0.71	0.71	0.71
	σ_2^2	0.45	0.45	0.49	0.52	0.53	0.49	0.49	0.49
	γ_2	0.60	0.60	0.64	0.68	0.69	0.64	0.64	0.63
		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
QLD SA	ν	6.83	7.26	10.54	18.91	20.80	11.38	9.67	11.41
	ρ	0.05	0.07	0.12	0.18	0.19	0.12	0.13	0.13
	σ_1^2	0.41	0.42	0.45	0.48	0.48	0.45	0.44	0.44
	γ_1	0.67	0.67	0.71	0.74	0.75	0.71	0.71	0.71
	σ_2^2	0.46	0.46	0.50	0.53	0.54	0.50	0.50	0.50
	γ_2	0.58	0.59	0.63	0.67	0.68	0.63	0.63	0.63
		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
NSW VIC	ν	3.61	3.91	5.28	7.26	7.51	5.38	5.11	5.16
	ρ	0.33	0.34	0.39	0.45	0.46	0.39	0.40	0.38
	σ_1^2	0.48	0.49	0.52	0.55	0.56	0.52	0.52	0.52
	γ_1	0.63	0.64	0.67	0.70	0.71	0.67	0.67	0.70
	σ_2^2	0.46	0.46	0.50	0.53	0.54	0.50	0.50	0.49
	γ_2	0.53	0.53	0.57	0.60	0.61	0.57	0.57	0.63
		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
NSW SA	ν	3.78	3.98	5.21	7.14	7.75	5.35	5.04	5.23
	ρ	0.18	0.18	0.25	0.30	0.31	0.24	0.25	0.24
	σ_1^2	0.48	0.49	0.52	0.56	0.57	0.52	0.52	0.52
	γ_1	0.64	0.65	0.68	0.71	0.72	0.68	0.68	0.70
	σ_2^2	0.45	0.46	0.50	0.53	0.54	0.50	0.50	0.50
	γ_2	0.55	0.55	0.59	0.63	0.64	0.59	0.59	0.63
		2.5%	5%	50%	95%	97.5%	mean	mode	TSP-MLE
VIC SA	ν	6.07	6.61	9.89	15.65	16.18	10.26	9.34	9.94
	ρ	0.56	0.56	0.60	0.64	0.64	0.60	0.60	0.59
	σ_1^2	0.62	0.63	0.68	0.73	0.74	0.68	0.68	0.67
	γ_1	0.53	0.54	0.57	0.61	0.62	0.57	0.57	0.59
	σ_2^2	0.65	0.66	0.71	0.76	0.78	0.71	0.71	0.70
	γ_2	0.44	0.45	0.48	0.52	0.53	0.48	0.48	0.57

Table 4.6: Estimated Parameters with JMCMC and TSP-MLE for all possible pairs of the four states Queensland, New South Wales, Victoria and South Australia

Chapter 5

Four-Dimensional Modeling

5.1 Basic Procedure

Now we will look at four dimensional data. The marginal model stays the same, we choose again an AR(1)-process for all four margins. The dependency in this case is modeled via a D-vine consisting of bivariate t-Copulas as building blocks (cf. Section 2.8). This is new compared to the last section since in two dimensions, the D-vine is identical to a 'normal' t-Copula.

So we have 4 marginal AR(1) time series

$$Y_{it} = \gamma_i \cdot Y_{i(t-1)} + \epsilon_{it} \quad t = 1, \dots, T, \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma_i^2) \; i.i.d., i = 1, \dots, 4$$
$$Y_{i0} \sim \mathcal{N}(0, \frac{\sigma_i^2}{1-\gamma_i^2}), \left|\gamma_i\right| < 1.$$

Again, we define the notation

$$\boldsymbol{Y^t} := \begin{pmatrix} Y_{1t} & Y_{2t} & Y_{3t} & Y_{4t} \end{pmatrix}, t = 1, \dots, T \qquad \boldsymbol{Y_i} := \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, i = 1, 2, 3, 4.$$

Therefore, it follows

$$\Rightarrow \boldsymbol{Y_i} \sim \mathcal{N}_T(\mathbf{0}, \Sigma_i) \text{ with } \Sigma_i = \frac{\sigma_i^2}{1-\gamma_i^2} \begin{pmatrix} 1 & \gamma_i & \gamma_i^2 & \dots & \gamma_i^{T-1} \\ \gamma_i & 1 & \gamma_i & & \gamma_i^{T-2} \\ \gamma_i^2 & \gamma_i & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \gamma_i \\ \gamma_i^{T-1} & \dots & & \gamma_i & 1 \end{pmatrix}, i = 1, 2, 3, 4.$$

We use the same transformation as in the two dimensional case (cf. (4.1)) and define

$$\boldsymbol{Z_i} := \Sigma_i^{-1/2} \boldsymbol{Y_i} \sim \mathcal{N}_T(\mathbf{0}, I_T), i = 1, \dots, 4$$

with

$$\Sigma_i^{-1/2} = \sigma_i^{-1} \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -\gamma_i & 1 & 0 & & \vdots \\ 0 & -\gamma_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\gamma_i & \sqrt{1-\gamma_i^2} \end{pmatrix}$$

and

$$\det \Sigma_i^{-1/2} = \sigma_i^{-T} \sqrt{1-\gamma_i^2}.$$

At first some notation issues (cf. Figure 2.5):

$$\begin{aligned} \boldsymbol{\nu} &:= (\nu_{12}, \nu_{23}, \nu_{34}, \nu_{13|2}, \nu_{24|3}, \nu_{14|23})' & \boldsymbol{\rho} &:= (\rho_{12}, \rho_{23}, \rho_{34}, \rho_{13|2}, \rho_{24|3}, \rho_{14|23})' \\ \boldsymbol{\theta_{11}} &:= (\nu_{12}, \rho_{12})', & \boldsymbol{\theta_{12}} &:= (\nu_{23}, \rho_{24})', & \boldsymbol{\theta_{13}} &:= (\nu_{34}, \rho_{34})', \\ \boldsymbol{\theta_{21}} &:= (\nu_{13|2}, \rho_{13|2})', & \boldsymbol{\theta_{22}} &:= (\nu_{24|3}, \rho_{24|3})', & \boldsymbol{\theta_{31}} &:= (\nu_{14|23}, \rho_{14|23})' \end{aligned}$$

We model now the dependence between the time series with a D-Vine consisting of bivariate t-Copulas as building blocks. Therefore we start on the Z-Level as the observations are marginal independent there and we can write the likelihood as a product. So the likelihood is given as:

$$f(\boldsymbol{z_1}, \boldsymbol{z_2}, \boldsymbol{z_3}, \boldsymbol{z_4}|\boldsymbol{\rho}, \boldsymbol{\nu}) \overset{\boldsymbol{z^t} \text{ ind}}{=} \prod_{t=1}^{T} f_t(z_{1t}, z_{2t}, z_{3t}, z_{4t}|\boldsymbol{\rho}, \boldsymbol{\nu}).$$

Using the calculations of the D-vine in Section 2.8 and especially (2.41), we can decompose the likelihood as

$$\begin{aligned} &f(\boldsymbol{z_1},\boldsymbol{z_2}, \boldsymbol{z_3}, \boldsymbol{z_4}|\boldsymbol{\rho}, \boldsymbol{\nu}) = \\ &= \prod_{t=1}^{T} \Big[f_1(z_{1t}) \cdot f_2(z_{2t}) \cdot f_3(z_{3t}) \cdot f_4(z_{4t}) \cdot \\ &\quad \cdot c_{12}(F_{1t}(z_{1t}), F_{2t}(z_{2t})) \cdot c_{23}(F_{2t}(z_{2t}), F_{3t}(z_{3t})) \cdot c_{34}(F_{3t}(z_{3t}), F_{4t}(z_{4t})) \cdot \\ &\quad \cdot c_{13|2}(F_t(z_{1t}|z_{2t}), F_t(z_{3t}|z_{2t})) \cdot c_{24|3}(F_t(z_{2t}|z_{3t}), F_t(z_{4t}|z_{3t})) \cdot \\ &\quad \cdot c_{14|23}(F_t(z_{1t}|z_{2t}, z_{3t}), F_t(z_{4t}|z_{2t}, z_{3t})) \Big]. \end{aligned}$$

We model the marginal distribution with a standard normal distribution. Furthermore, we use the definition of the $h(\cdot)$-function for the t-Copula (2.38) to calculate the conditional cumulative density functions and we take the logarithm to get the loglikeli-

hood. This yields (cf. (2.44))

$$
\begin{aligned}
&\log(f(\boldsymbol{z_1},\boldsymbol{z_2},\boldsymbol{z_3},\boldsymbol{z_4}|\boldsymbol{\rho},\boldsymbol{\nu})) = \\
&= \sum_{t=1}^{T} \Big[\log(\varphi(z_{1t})) + \log(\varphi(z_{2t})) + \log(\varphi(z_{3t})) + \log(\varphi(z_{4t})) + \\
&\quad + \log(c_{12}(\Phi(z_{1t}),\Phi(z_{2t}))) + \log(c_{23}(\Phi(z_{2t}),\Phi(z_{3t}))) + \log(c_{34}(\Phi(z_{3t}),\Phi(z_{4t}))) + \\
&\quad + \log(c_{13|2}(h(\Phi(z_{1t}),\Phi(z_{2t});\theta_{11}),h(\Phi(z_{3t}),\Phi(z_{2t});\theta_{12}))) + \\
&\quad + \log(c_{24|3}(h(\Phi(z_{2t}),\Phi(z_{3t});\theta_{12}),h(\Phi(z_{4t}),\Phi(z_{3t});\theta_{13}))) + \\
&\quad + \log(c_{14|23}(h(h(\Phi(z_{1t}),\Phi(z_{2t});\theta_{11}),h(\Phi(z_{3t}),\Phi(z_{2t});\theta_{12});\theta_{21}), \\
&\qquad\qquad h(h(\Phi(z_{4t}),\Phi(z_{3t});\theta_{13}),h(\Phi(z_{2t}),\Phi(z_{3t});\theta_{12});\theta_{31}))).
\end{aligned}
$$

As mentioned above, we also use the transformation (4.1) to the observed data like in the two dimensional case :

$$
\begin{aligned}
&\log(f(\boldsymbol{y_1},\boldsymbol{y_2},\boldsymbol{y_3},\boldsymbol{y_4}|\sigma_i^2,\gamma_i,i=1,\ldots,4,\boldsymbol{\rho},\boldsymbol{\nu})) = \\
&\qquad = \log(\varphi(\frac{1}{\sigma_1}y_{11})) + \log(\varphi(\frac{1}{\sigma_2}y_{21})) + + \log(\varphi(\frac{1}{\sigma_3}y_{31})) + \log(\varphi(\frac{1}{\sigma_4}y_{41})) + \\
&\qquad + \log(c_{12}(\Phi(\frac{1}{\sigma_1}y_{11}),\Phi(\frac{1}{\sigma_2}y_{21}))) + \log(c_{23}(\Phi(\frac{1}{\sigma_2}y_{21}),\Phi(\frac{1}{\sigma_3}y_{31}))) + \\
&\qquad + \log(c_{34}(\Phi(\frac{1}{\sigma_3}y_{31}),\Phi(\frac{1}{\sigma_4}y_{41}))) + \\
&\qquad + \log(c_{13|2}(h(\Phi(\frac{1}{\sigma_1}y_{11}),\Phi(\frac{1}{\sigma_2}y_{21});\theta_{11}),h(\Phi(\frac{1}{\sigma_3}y_{31}),\Phi(\frac{1}{\sigma_2}y_{21});\theta_{12}))) + \\
&\qquad + \log(c_{24|3}(h(\Phi(\frac{1}{\sigma_2}y_{21}),\Phi(\frac{1}{\sigma_3}y_{31});\theta_{12}),h(\Phi(\frac{1}{\sigma_4}y_{41}),\Phi(\frac{1}{\sigma_3}y_{31});\theta_{13}))) + \\
&\qquad + \log(c_{14|23}(h(h(\Phi(\frac{1}{\sigma_1}y_{11}),\Phi(\frac{1}{\sigma_2}y_{21});\theta_{11}),h(\Phi(\frac{1}{\sigma_3}y_{31}),\Phi(\frac{1}{\sigma_2}y_{21});\theta_{12});\theta_{21}), \\
&\qquad\qquad h(h(\Phi(\frac{1}{\sigma_4}y_{41}),\Phi(\frac{1}{\sigma_3}y_{31});\theta_{13}),h(\Phi(\frac{1}{\sigma_2}y_{21}),\Phi(\frac{1}{\sigma_3}y_{31});\theta_{12});\theta_{31}))) + \\
&\qquad + \sum_{t=2}^{T-1} \Big[\log(\varphi(\frac{1}{\sigma_1}(y_{1t}-\gamma_1 y_{1(t-1)}))) + \log(\varphi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)}))) + \\
&\qquad\quad + \log(\varphi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)}))) + \log(\varphi(\frac{1}{\sigma_4}(y_{4t}-\gamma_4 y_{4(t-1)}))) + \\
&\qquad\quad + \log(c_{12}(\Phi(\frac{1}{\sigma_1}(y_{1t}-\gamma_1 y_{1(t-1)})),\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)})))) + \\
&\qquad\quad + \log(c_{23}(\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)})),\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)})))) + \\
&\qquad\quad + \log(c_{34}(\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)})),\Phi(\frac{1}{\sigma_4}(y_{4t}-\gamma_4 y_{4(t-1)})))) +
\end{aligned}
$$

$$
\begin{aligned}
&+\log(c_{13|2}(h(\Phi(\frac{1}{\sigma_1}(y_{1t}-\gamma_1 y_{1(t-1)})),\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)}));\theta_{11}),\\
&\qquad h(\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)})),\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)}));\theta_{12})))+\\
&+\log(c_{24|3}(h(\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)})),\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)}));\theta_{12}),\\
&\qquad h(\Phi(\frac{1}{\sigma_4}(y_{4t}-\gamma_4 y_{4(t-1)})),\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)}));\theta_{13})))+\\
&+\log(c_{14|23}(h(h(\Phi(\frac{1}{\sigma_1}(y_{1t}-\gamma_1 y_{1(t-1)})),\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)}));\theta_{11}),\\
&\qquad h(\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)})),\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)}));\theta_{12});\theta_{21}),\\
&\qquad h(h(\Phi(\frac{1}{\sigma_4}(y_{4t}-\gamma_4 y_{4(t-1)})),\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)}));\theta_{13}),\\
&\qquad h(\Phi(\frac{1}{\sigma_2}(y_{2t}-\gamma_2 y_{2(t-1)})),\Phi(\frac{1}{\sigma_3}(y_{3t}-\gamma_3 y_{3(t-1)}));\theta_{12});\theta_{31})))\Big]+\\
&+\log(\varphi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T}-\gamma_1 y_{1(T-1)})))+\log(\varphi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)})))+\\
&+\log(\varphi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)})))+\log(\varphi(\frac{1}{\sigma_4}(\sqrt{1-\gamma_4^2}y_{4T}-\gamma_4 y_{4(T-1)})))+\\
&+\log(c_{12}(\Phi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T}-\gamma_1 y_{1(T-1)})),\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)}))))+\\
&+\log(c_{23}(\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)})),\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)}))))+\\
&+\log(c_{34}(\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)})),\Phi(\frac{1}{\sigma_4}(\sqrt{1-\gamma_4^2}y_{4T}-\gamma_4 y_{4(T-1)}))))+\\
&+\log(c_{13|2}(h(\Phi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T}-\gamma_1 y_{1(T-1)})),\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)}));\theta_{11}),\\
&\qquad h(\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)})),\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)}));\theta_{12})))+\\
&+\log(c_{24|3}(h(\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)})),\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)}));\theta_{12}),\\
&\qquad h(\Phi(\frac{1}{\sigma_4}(\sqrt{1-\gamma_4^2}y_{4T}-\gamma_4 y_{4(T-1)})),\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)}));\theta_{13})))+\\
&+\log(c_{14|23}(h(h(\Phi(\frac{1}{\sigma_1}(\sqrt{1-\gamma_1^2}y_{1T}-\gamma_1 y_{1(T-1)})),\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)}));\theta_{11}),\\
&\qquad h(\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)})),\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)}));\theta_{12});\theta_{21}),\\
&\qquad h(h(\Phi(\frac{1}{\sigma_4}(\sqrt{1-\gamma_4^2}y_{4T}-\gamma_4 y_{4(T-1)})),\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)}));\theta_{13}),\\
&\qquad h(\Phi(\frac{1}{\sigma_2}(\sqrt{1-\gamma_2^2}y_{2T}-\gamma_2 y_{2(T-1)})),\Phi(\frac{1}{\sigma_3}(\sqrt{1-\gamma_3^2}y_{3T}-\gamma_3 y_{3(T-1)}));\theta_{12});\theta_{31})))+\\
&+\sum_{i=1}^{4}\left[\log(\sigma_i^{-T})+\log(\sqrt{1-\gamma_i^2})\right]. \qquad (5.1)
\end{aligned}
$$

For the calculation of the posterior distribution (up to constants), we need to know the priors. Using the results of Lewandowski, Kurowicka, and Joe (2007) and the calculations of Joe (2006), we suppose a $Beta((4-k)/2,(4-k)/2)$ distribution on $(-1,1)$ for ρ_j, where k is the cardinality of the set of conditioning variables. The resulting correlation matrix of partial correlations is uniformly distributed over the space of correlation matrices. So we choose:

- Prior for ρ_{12}, ρ_{23} and ρ_{34}: $Beta(2,2)$ on $(-1,1)$
  $\pi(\rho_j) \propto (1-\rho_j^2), j = 12, 23, 34$
- Prior for $\rho_{13|2}$ and $\rho_{24|3}$: $Beta(3/2,3/2)$ on $(-1,1)$
  $\pi(\rho_j) \propto (1-\rho_j^2)^{1/2}, j = 13|2, 24|3$
- Prior for $\rho_{14|23}$: $Beta(1,1)$ on $(-1,1)$ = Unif(1,1)
  $\pi(\rho_{14|23}) \propto 1$.
- Prior for each ν: half Cauchy (1,2) (cf. (2.3))
  $\pi(\nu) \propto \frac{1}{1+(\frac{\nu-1}{2})^2}, \qquad \nu \in (1,\infty)$
- Prior for $\sigma_i^2, i = 1,\ldots,4$: Inverse Gamma(1,0.001) (cf. (2.1))
  $\pi(u) = \frac{0.001}{\Gamma(1)} u^{-2} \exp\left(-\frac{0.001}{u}\right)$ (Congdon (2003, p. 15))
- Prior for a_i: Normal (0,10) on transformed scale $(-\infty,\infty)$
  $\pi(a_i) \propto \exp(-\frac{a_i^2}{200})$.

Additionally, we use again Metropolis-Hastings algorithm with independence proposal (cf. Algorithm 2.5). We develop the posterior density (up to constants) first on the untransformed level. For finding the necessary mode and the Hessian, we transform the posterior density for ρ_j, $\sigma_i^2, i = 1,\ldots,4$ and $\gamma_i, i = 1,\ldots,4$ to $(-\infty,\infty)$ using Fisher-z-transformation (4.5) to r_j, logarithm transformation (4.3) to s_i and Fisher-z-transformation (4.4) to a_i, respectively. For γ_i, we choose the prior on the transformed scale. After having determined the mode and the Hessian we transform back to the original scale, if necessary, to perform the MH-algorithm using again the Delta method for ρ like in (4.12), for σ_i^2 like in (4.13) and for γ_i like in (4.11). The only modification in the algorithm compared to the two dimensional case (cf. Algorithms 4.1 - 4.4) is that we just optimize in every 20th step to find the mode and the Hessian in order to reduce the calculation time of the computer. During the other steps, we use the most recently determined mode and Hessian to draw a proposal value.
We perform the updates in the following order:

$$\nu_{12} \to \rho_{12} \to \sigma_1^2 \to \gamma_1 \to \sigma_2^2 \to \gamma_2 \to \nu_{23} \to \rho_{23} \to \sigma_3^2 \to \gamma_3 \to \nu_{34} \to$$
$$\to \rho_{34} \to \sigma_4^2 \to \gamma_4 \to \nu_{13|2} \to \rho_{13|2} \to \nu_{24|3} \to \rho_{24|3} \to \nu_{14|23} \to \rho_{14|23}.$$

5.2 Simulation Study

For the four-dimensional case, we also performed a simulation study to assess the power of the algorithm. We created three scenarios and repeated the data creation and the

parameter estimation seven times in each scenario. Then we calculated the average for each parameter over these seven replications and we compute the standard deviation over these replications. We chose the three scenarios described in Table 5.1.

1	ν_{12}	ν_{23}	ν_{34}	$\nu_{13\|2}$	$\nu_{24\|3}$	$\nu_{14\|23}$	σ_1^2	γ_1	σ_2^2	γ_2
	7	5.5	13	10	25	10	0.15	0.65	0.05	0.7
	ρ_{12}	ρ_{23}	ρ_{34}	$\rho_{13\|2}$	$\rho_{24\|3}$	$\rho_{14\|23}$	σ_3^2	γ_3	σ_4^2	γ_4
	0.4	0.2	0.7	0.3	-0.05	-0.1	0.1	0.65	0.05	0.7
2	ν_{12}	ν_{23}	ν_{34}	$\nu_{13\|2}$	$\nu_{24\|3}$	$\nu_{14\|23}$	σ_1^2	γ_1	σ_2^2	γ_2
	4	4	4	3	3	6	0.8	0.2	0.4	0.4
	ρ_{12}	ρ_{23}	ρ_{34}	$\rho_{13\|2}$	$\rho_{24\|3}$	$\rho_{14\|23}$	σ_3^2	γ_3	σ_4^2	γ_4
	0	0.8	0.4	0.8	0.6	0.4	1.5	0.6	1	0.1
3	ν_{12}	ν_{23}	ν_{34}	$\nu_{13\|2}$	$\nu_{24\|3}$	$\nu_{14\|23}$	σ_1^2	γ_1	σ_2^2	γ_2
	10	3	7	6	4	3	2	0.4	1	0.7
	ρ_{12}	ρ_{23}	ρ_{34}	$\rho_{13\|2}$	$\rho_{24\|3}$	$\rho_{14\|23}$	σ_3^2	γ_3	σ_4^2	γ_4
	0.8	0.1	0.5	0.8	0.2	0.8	0.7	0.8	0.2	0.2

Table 5.1: Simulation Scenarios

We compare the results of the Bayesian estimation to TSP-MLE values (cf. Table 2.1). The marginal time series parameters are determined by using the calculation routine 'arima' univariately with order (1,0,0) which is already implemented in the statistic program 'R'. It is a Maximum Likelihood Estimation. Then we transform the data parametricly to the unit cube using the inverse transformation to the described one in (4.1) and the cumulative distribution function of the standard normal distribution in order to estimate the copula parameters. The first estimation values of the copula parameters, denoted by 'start' in the results, are calculated by using the algorithm for starting values described at the end of Section 2.9 with estimating the parameters by a separate Maximum Likelihood Estimation for each copula. The second method, denoted by 'C-MLE' in the results, is a Maximum Likelihood estimation of all D-vine t-copula parameters. For this second estimation method for the copula parameters, we used the implementation in 'R' of Daniel Berg and Henrik Bakken.

Analyzing the results, we can say that in general all estimation methods are close to each other. We get the impression that the Bayes estimation values have a smaller standard deviation than the marginal MLE / C-MLE and the start values. The marginal time series parameters are estimated more exactly than the copula parameters. Considering the MCMC values, for some parameters the posterior mode is closer to the true values, for others the posterior mean or the posterior median.

		true	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	value	7.000	4.529	4.800	6.889	11.411	12.929	7.346	6.460	7.094	7.101
	sd		0.624	0.627	1.256	3.049	3.882	1.443	1.079	1.327	1.304
ν_{23}	value	5.500	3.673	3.890	5.333	8.503	9.791	5.676	5.074	5.339	5.474
	sd		0.634	0.707	1.444	3.939	5.505	1.818	1.269	1.848	1.584
ν_{34}	value	13.000	7.150	7.839	14.657	36.530	42.706	17.366	12.770	19.890	19.253
	sd		1.685	1.925	6.473	19.604	23.122	7.932	4.905	15.147	15.027
$\nu_{13\|2}$	value	10.000	5.513	5.847	9.221	18.816	21.797	10.419	8.401	9.919	9.596
	sd		0.961	1.105	2.131	7.239	8.135	2.908	1.770	2.692	2.457
$\nu_{24\|3}$	value	25.000	6.337	7.299	23.963	76.896	87.754	30.400	17.933	30.627	59.414
	sd		1.336	1.470	8.980	16.914	10.649	8.626	5.407	31.952	106.541
$\nu_{14\|23}$	value	10.000	5.841	6.297	10.589	24.116	27.854	12.266	9.427	11.737	11.693
	sd		0.964	1.099	2.610	9.169	10.308	3.540	2.067	3.658	3.642
ρ_{12}	value	0.400	0.311	0.323	0.370	0.419	0.431	0.371	0.370	0.373	0.374
	sd		0.030	0.030	0.028	0.028	0.029	0.030	0.030	0.028	0.030
ρ_{23}	value	0.200	0.129	0.139	0.194	0.250	0.260	0.196	0.194	0.196	0.199
	sd		0.043	0.046	0.044	0.043	0.043	0.046	0.046	0.046	0.046
ρ_{34}	value	0.700	0.667	0.673	0.704	0.729	0.734	0.704	0.703	0.707	0.707
	sd		0.021	0.021	0.019	0.018	0.019	0.019	0.017	0.021	0.021
$\rho_{13\|2}$	value	0.300	0.236	0.244	0.301	0.349	0.357	0.301	0.301	0.301	0.303
	sd		0.024	0.023	0.023	0.025	0.023	0.023	0.023	0.025	0.023
$\rho_{24\|3}$	value	−0.050	−0.117	−0.109	−0.053	0.004	0.010	−0.051	−0.053	−0.051	−0.054
	sd		0.037	0.036	0.042	0.041	0.043	0.038	0.042	0.043	0.043
$\rho_{14\|23}$	value	−0.100	−0.157	−0.149	−0.096	−0.039	−0.030	−0.096	−0.096	−0.094	−0.094
	sd		0.024	0.023	0.026	0.025	0.023	0.026	0.026	0.025	0.025

		true	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	value	0.150	0.139	0.139	0.150	0.160	0.164	0.150	0.149	0.149
	sd		0.007	0.007	0.006	0.006	0.008	0.006	0.007	0.007
γ_1	value	0.650	0.600	0.607	0.643	0.676	0.683	0.643	0.643	0.641
	sd		0.019	0.020	0.021	0.023	0.019	0.021	0.019	0.018
σ_2^2	value	0.050	0.049	0.049	0.050	0.056	0.057	0.050	0.050	0.050
	sd		0.004	0.004	0.000	0.005	0.005	0.000	0.000	0.000
γ_2	value	0.700	0.657	0.664	0.699	0.731	0.739	0.696	0.699	0.694
	sd		0.028	0.026	0.030	0.023	0.024	0.027	0.027	0.025
σ_3^2	value	0.100	0.091	0.093	0.101	0.111	0.111	0.101	0.101	0.101
	sd		0.004	0.005	0.004	0.004	0.004	0.004	0.004	0.004
γ_3	value	0.650	0.623	0.627	0.657	0.684	0.690	0.657	0.657	0.654
	sd		0.028	0.027	0.027	0.026	0.026	0.027	0.027	0.036
σ_4^2	value	0.050	0.047	0.047	0.051	0.054	0.056	0.051	0.051	0.053
	sd		0.005	0.005	0.004	0.005	0.005	0.004	0.004	0.005
γ_4	value	0.700	0.669	0.674	0.701	0.730	0.733	0.701	0.701	0.690
	sd		0.023	0.023	0.024	0.022	0.020	0.024	0.024	0.028

Table 5.2: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE, starting values and TSP-MLE for all parameters in scenario 1 averaged over 7 replications with its empirical standard deviation

		true	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	value	4.000	3.094	3.191	3.770	4.533	4.707	3.809	3.720	3.814	3.744
	sd		0.286	0.279	0.390	0.522	0.575	0.394	0.374	0.447	0.339
ν_{23}	value	4.000	3.381	3.481	4.087	4.859	5.024	4.114	4.053	4.454	4.101
	sd		0.320	0.340	0.395	0.501	0.536	0.405	0.369	0.967	0.379
ν_{34}	value	4.000	3.066	3.174	3.874	4.849	5.137	3.926	3.820	3.840	3.899
	sd		0.220	0.236	0.337	0.498	0.588	0.348	0.336	0.410	0.350
$\nu_{13\|2}$	value	3.000	2.273	2.357	2.901	3.650	3.814	2.940	2.856	3.006	2.981
	sd		0.367	0.389	0.516	0.725	0.775	0.529	0.502	0.397	0.516
$\nu_{24\|3}$	value	3.000	2.041	2.203	3.557	5.647	6.304	3.704	3.433	3.007	3.041
	sd		0.305	0.336	0.316	0.810	0.875	0.344	0.298	0.413	0.426
$\nu_{14\|23}$	value	6.000	3.901	4.111	5.797	9.369	10.509	6.164	5.494	6.159	5.983
	sd		0.724	0.772	1.401	3.581	4.220	1.672	1.233	1.543	1.504
ρ_{12}	value	0.000	−0.069	−0.061	−0.011	0.043	0.054	−0.011	−0.013	−0.016	−0.001
	sd		0.028	0.030	0.030	0.029	0.028	0.030	0.029	0.030	0.032
ρ_{23}	value	0.800	0.776	0.781	0.799	0.819	0.820	0.799	0.799	0.800	0.804
	sd		0.018	0.020	0.017	0.017	0.015	0.017	0.017	0.021	0.018
ρ_{34}	value	0.400	0.327	0.336	0.381	0.427	0.436	0.383	0.381	0.387	0.389
	sd		0.030	0.030	0.034	0.030	0.033	0.033	0.034	0.048	0.035
$\rho_{13\|2}$	value	0.800	0.754	0.760	0.784	0.807	0.813	0.784	0.784	0.787	0.786
	sd		0.022	0.019	0.017	0.015	0.016	0.017	0.017	0.016	0.020
$\rho_{24\|3}$	value	0.600	0.539	0.551	0.599	0.639	0.647	0.597	0.600	0.593	0.594
	sd		0.045	0.040	0.034	0.031	0.030	0.034	0.033	0.035	0.035
$\rho_{14\|23}$	value	0.400	0.351	0.363	0.411	0.459	0.469	0.410	0.413	0.401	0.407
	sd		0.032	0.030	0.026	0.027	0.027	0.027	0.028	0.030	0.024
		true	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE	
σ_1^2	value	0.800	0.749	0.757	0.809	0.867	0.877	0.810	0.806	0.803	
	sd		0.020	0.022	0.021	0.022	0.025	0.022	0.020	0.025	
γ_1	value	0.200	0.166	0.173	0.200	0.223	0.229	0.200	0.199	0.209	
	sd		0.017	0.018	0.019	0.018	0.020	0.019	0.018	0.030	
σ_2^2	value	0.400	0.377	0.379	0.401	0.430	0.433	0.401	0.400	0.404	
	sd		0.014	0.012	0.016	0.014	0.015	0.016	0.014	0.015	
γ_2	value	0.400	0.377	0.377	0.397	0.416	0.420	0.397	0.397	0.396	
	sd		0.025	0.025	0.025	0.022	0.024	0.025	0.025	0.043	
σ_3^2	value	1.500	1.414	1.429	1.519	1.623	1.643	1.521	1.516	1.531	
	sd		0.042	0.045	0.047	0.050	0.052	0.047	0.049	0.044	
γ_3	value	0.600	0.579	0.581	0.599	0.614	0.619	0.599	0.599	0.597	
	sd		0.017	0.018	0.017	0.016	0.013	0.017	0.017	0.026	
σ_4^2	value	1.000	0.919	0.927	0.994	1.067	1.081	0.994	0.994	0.989	
	sd		0.037	0.034	0.039	0.039	0.038	0.039	0.039	0.034	
γ_4	value	0.100	0.054	0.060	0.091	0.120	0.127	0.091	0.093	0.083	
	sd		0.035	0.037	0.038	0.037	0.035	0.038	0.035	0.043	

Table 5.3: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE, starting values and TSP-MLE for all parameters in scenario 2 averaged over 7 replications with its empirical standard deviation

		true	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	value	10.000	8.073	8.499	11.083	16.013	17.224	11.489	10.684	16.766	10.851
	sd		1.400	1.545	2.444	5.656	6.845	2.849	2.148	10.862	2.181
ν_{23}	value	3.000	2.707	2.769	3.153	3.617	3.696	3.169	3.129	3.306	3.217
	sd		0.157	0.173	0.212	0.255	0.259	0.215	0.212	0.283	0.204
ν_{34}	value	7.000	4.971	5.257	6.923	9.921	10.726	7.207	6.681	8.363	7.267
	sd		0.506	0.545	0.888	1.805	2.156	0.984	0.833	2.124	1.008
$\nu_{13\|2}$	value	6.000	4.323	4.524	5.880	8.034	8.536	6.040	5.693	6.771	6.261
	sd		0.398	0.442	0.636	1.144	1.203	0.686	0.581	1.526	0.782
$\nu_{24\|3}$	value	4.000	2.744	3.016	5.144	8.593	9.759	5.404	4.879	4.969	4.219
	sd		0.359	0.373	0.427	0.737	1.558	0.367	0.463	1.819	0.391
$\nu_{14\|23}$	value	3.000	2.493	2.607	3.391	4.567	4.920	3.460	3.313	3.640	3.527
	sd		0.392	0.417	0.661	1.015	1.108	0.674	0.625	0.780	0.722
ρ_{12}	value	0.800	0.779	0.780	0.799	0.814	0.816	0.797	0.799	0.800	0.800
	sd		0.009	0.008	0.007	0.008	0.005	0.008	0.007	0.008	0.008
ρ_{23}	value	0.100	0.024	0.033	0.084	0.134	0.140	0.084	0.084	0.089	0.096
	sd		0.032	0.029	0.028	0.032	0.033	0.028	0.028	0.034	0.034
ρ_{34}	value	0.500	0.444	0.451	0.493	0.530	0.540	0.491	0.493	0.506	0.500
	sd		0.022	0.022	0.022	0.020	0.020	0.021	0.022	0.029	0.024
$\rho_{13\|2}$	value	0.800	0.773	0.776	0.799	0.819	0.819	0.799	0.799	0.800	0.800
	sd		0.020	0.019	0.016	0.016	0.016	0.016	0.016	0.017	0.018
$\rho_{24\|3}$	value	0.200	0.144	0.151	0.203	0.253	0.263	0.204	0.203	0.206	0.209
	sd		0.021	0.020	0.023	0.023	0.025	0.020	0.023	0.018	0.024
$\rho_{14\|23}$	value	0.800	0.770	0.779	0.804	0.823	0.827	0.803	0.804	0.800	0.801
	sd		0.017	0.017	0.013	0.014	0.010	0.014	0.013	0.015	0.013
		true	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE	
σ_1^2	value	2.000	1.869	1.884	1.999	2.120	2.147	2.001	1.993	2.014	
	sd		0.077	0.077	0.084	0.084	0.077	0.083	0.085	0.066	
γ_1	value	0.400	0.381	0.386	0.400	0.417	0.419	0.400	0.401	0.397	
	sd		0.011	0.013	0.010	0.013	0.012	0.010	0.011	0.031	
σ_2^2	value	1.000	0.930	0.940	1.001	1.069	1.084	1.001	1.001	0.997	
	sd		0.048	0.051	0.051	0.055	0.056	0.051	0.051	0.051	
γ_2	value	0.700	0.684	0.686	0.699	0.716	0.719	0.699	0.699	0.691	
	sd		0.011	0.010	0.007	0.010	0.007	0.007	0.007	0.030	
σ_3^2	value	0.700	0.646	0.654	0.700	0.746	0.754	0.700	0.699	0.703	
	sd		0.023	0.024	0.024	0.028	0.026	0.024	0.023	0.022	
γ_3	value	0.800	0.781	0.787	0.801	0.817	0.819	0.801	0.801	0.800	
	sd		0.004	0.008	0.004	0.005	0.007	0.004	0.004	0.021	
σ_4^2	value	0.200	0.187	0.189	0.200	0.216	0.217	0.200	0.200	0.200	
	sd		0.013	0.012	0.012	0.014	0.013	0.012	0.012	0.012	
γ_4	value	0.200	0.167	0.171	0.199	0.220	0.226	0.199	0.197	0.203	
	sd		0.018	0.016	0.017	0.017	0.015	0.017	0.018	0.036	

Table 5.4: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE, starting values and TSP-MLE for all parameters in scenario 3 averaged over 7 replications with its empirical standard deviation

5.3 Estimation of Real Data

As in the two-dimensional case, we also want to perform an estimation for the observed Australian load data. We use the data preprocessed for stationarity with method 8 as described in Section 3.5. Therefore, we have the four time series for Queensland, New South Wales, Victoria and South Australia, each consisting of 1134 observations. We chose this order because of geographical reasons as already described in Section 4.3. These states are neighbored and connected with all its infrastructure in this way along the Eastern Coast of Australia beginning with Queensland in the North, then New South Wales, then Victoria and South Australia following in the South West. We show again the time series plot in Figure 5.1.

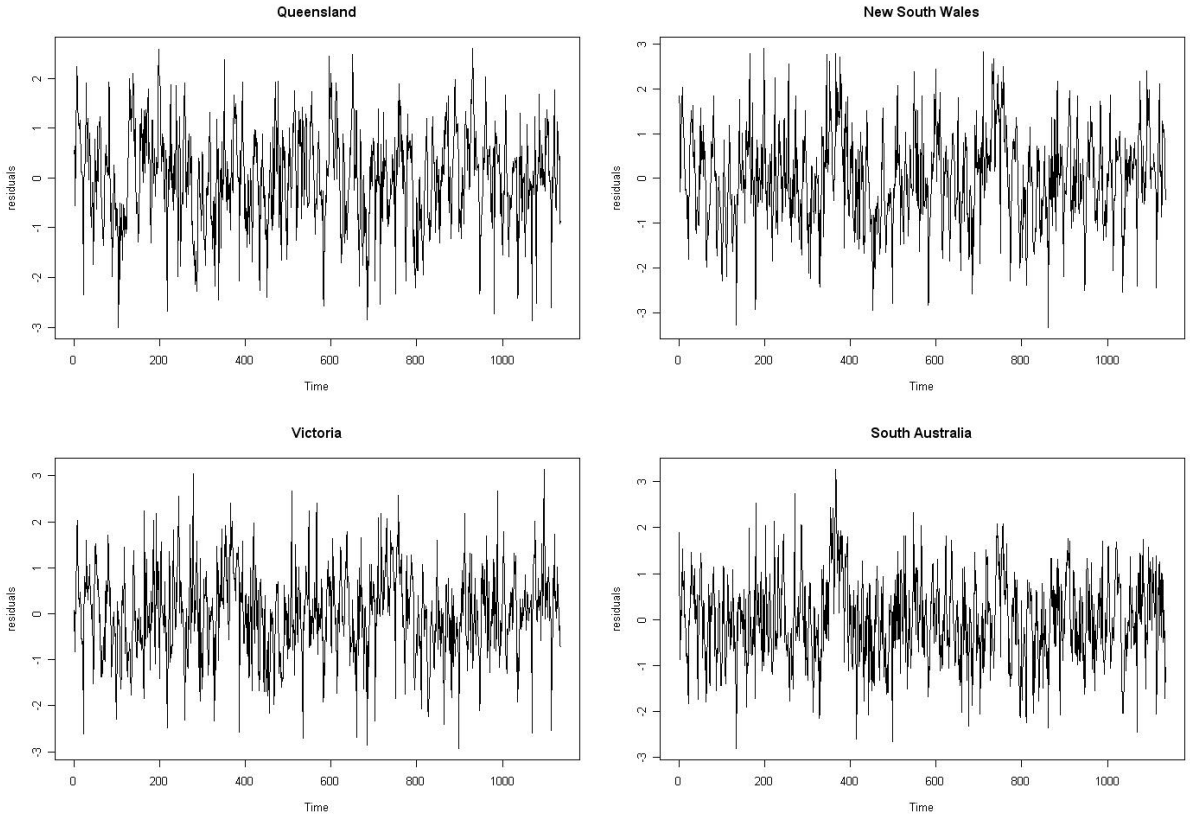

Figure 5.1: Australian Load Data preprocessed for Stationarity with Method 8

For this estimation of the real data, we want to show the whole analysis that we have performed for all performed estimation procedures. At first, we have a look at the trace plots of the estimation, i.e. the plot of all states of the Markov chain with each value connected to the next. The solid line marks the estimated posterior mode of JMCMC whereas the dashed line shows the C-MLE value and the marginal MLE for the time series parameters, respectively (cf. comment after Table 5.1).

We can see that the chain has a good mixing for each parameter and it oscillates nicely and fairly concentrated around the estimated mode. Only for the degrees of freedom for

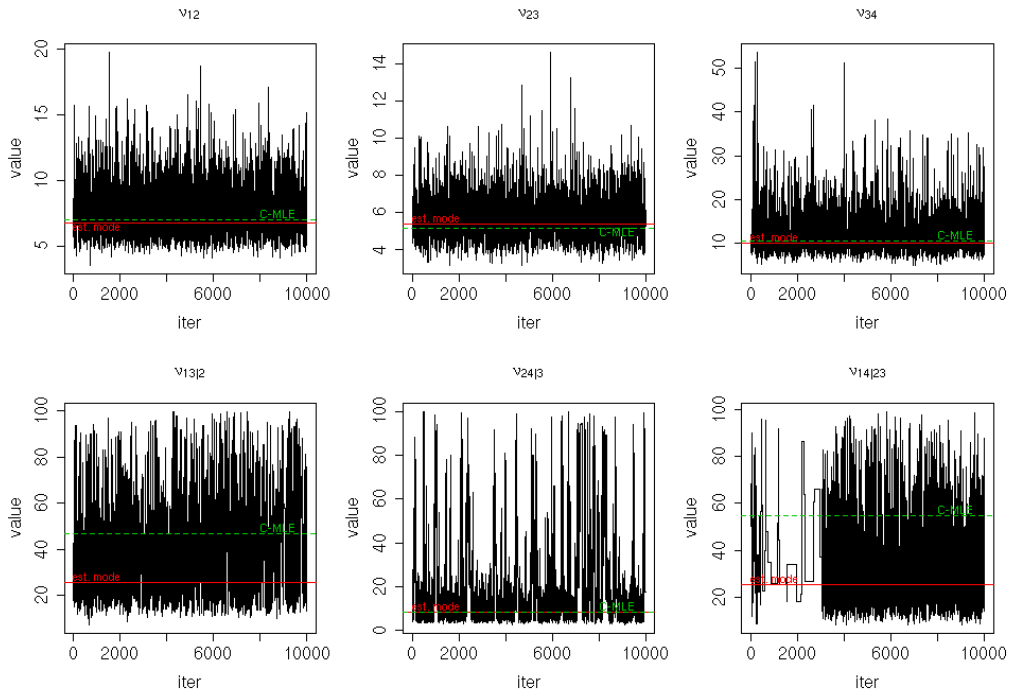

Figure 5.2: Trace Plots for all ν estimated of the observed Real Data

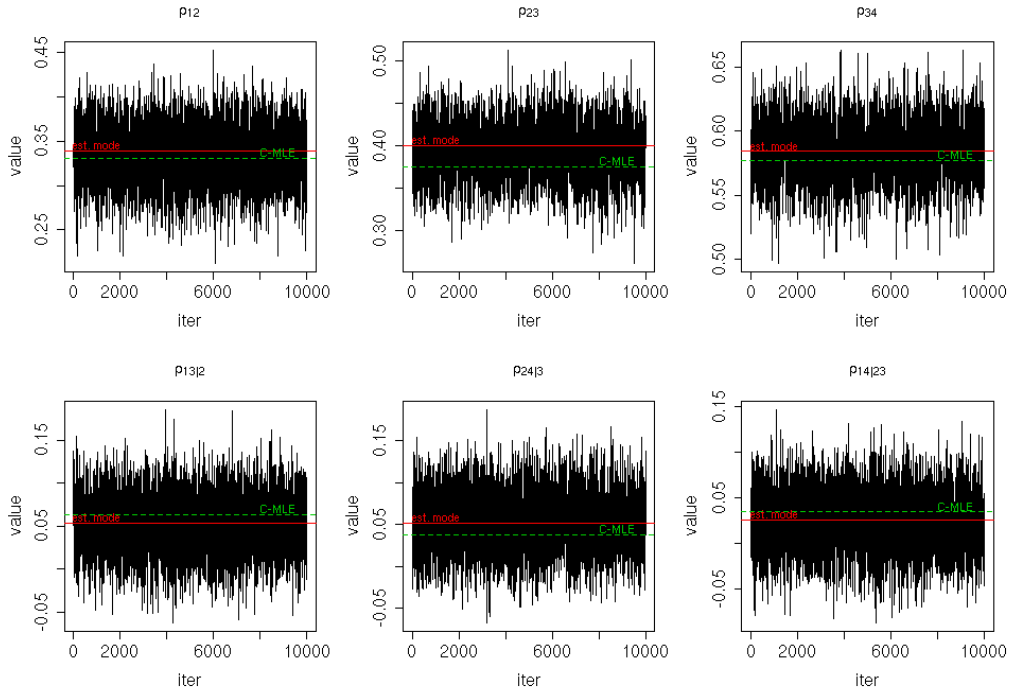

Figure 5.3: Trace Plots for all ρ estimated of the observed Real Data

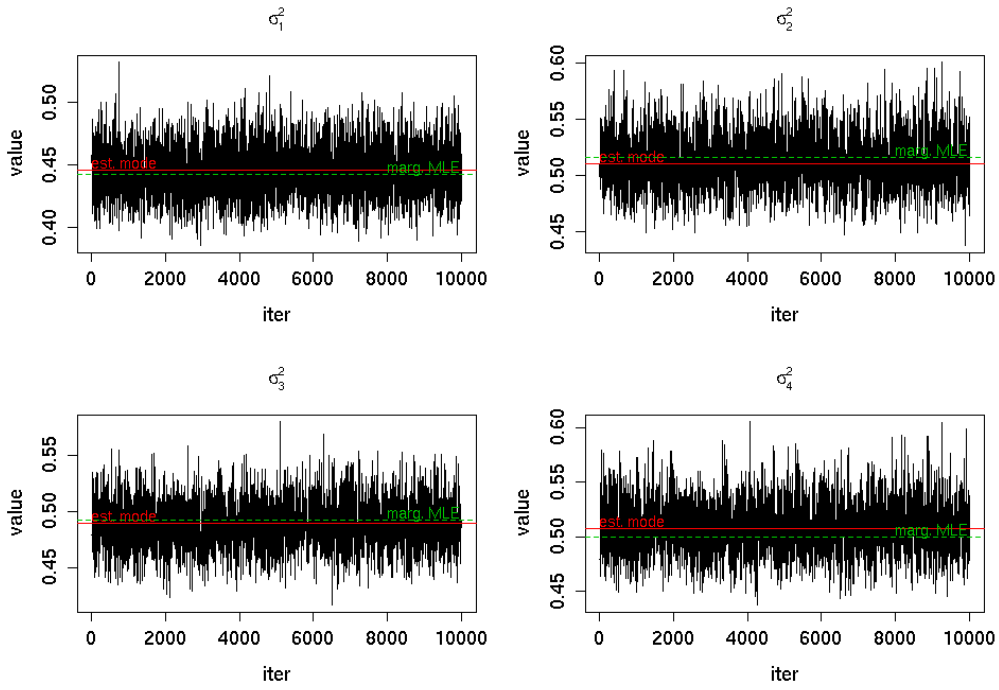

Figure 5.4: Trace Plots for all σ^2 estimated of the observed Real Data

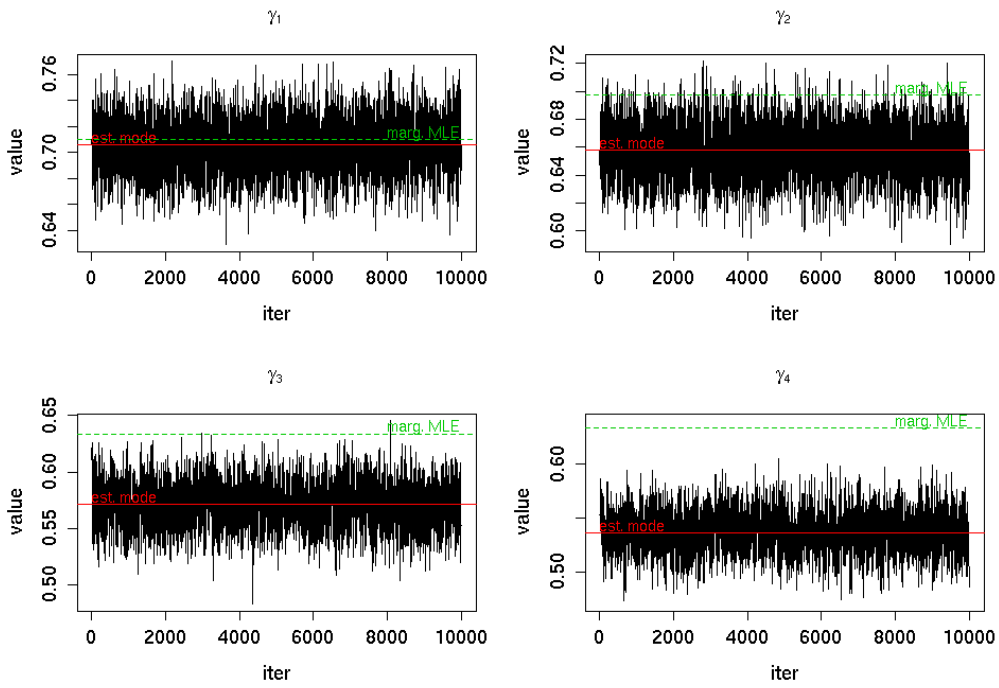

Figure 5.5: Trace Plots for all γ estimated of the observed Real Data

the conditional copulas $\nu_{13|2}$, $\nu_{24|3}$ and $\nu_{14|23}$, we have values on the whole admissible range between 1 and 100. For $\nu_{14|23}$, we remark some problems of the mixing in the first 3000 iterations. So, we increase the burn-in period for $\nu_{14|23}$ up to 3000. For all the other parameters, we choose a burn-in period of 1000 which we will disregard after having looked at the plots of the autocorrelation functions. Especially for γ_2, γ_3 and γ_4, we remark a different phenomena: the estimated posterior mode and all states of the Markov chain are completely different to the parameter values estimated marginally by MLE.

The autocorrelation functions of all parameters decay very quickly (cf. Figures 5.6 - 5.9). Only for the conditional degrees of freedom $\nu_{24|3}$ and $\nu_{14|23}$, the autocorrelation function tails off more slowly. So, if we thin out the chain by taking only every 20th value, we can suppose independence between the different sampled values.

After having discarded the burn-in period of the first 1000 values (for $\nu_{14|23}$ the first 3000 values) and having thinned out the chain by taking every 20th value, the remaining sample for each parameter consist of 500 values (350 values for $\nu_{14|23}$, respectively). Then we have a look at the remaining trace plots (cf. Figures 5.10 - 5.13). We see that the trace plots for the thinned out sample looks good. It fluctuates regularly around the estimated posterior mode and the Markov chain does not get stuck. Furthermore, we estimate the kernel density of the remaining sample with a Gaussian kernel (cf. Figures 5.14 - 5.17). We remark there, consistently to the observations in the trace plots of the whole chain, that the parameter values for γ_2, γ_3 and γ_4 are located where, according to the kernel estimation of the MCMC sample, no or nearly no probability mass is located. For all other parameter estimates, the C-MLE and the MCMC value are located closely together. When we have a look at the plot of the estimated kernel density for degrees of freedom of the conditional copulas $\nu_{13|2}$, $\nu_{24|3}$ and $\nu_{14|23}$, we remark that they have quite heavy tails on the right hand versus 100. The other estimated kernel densities of the parameters are quite concentrated.

Finally, we show the results of the estimation in a table (cf. Table 5.5). We notice that the marginal MLE values for the parameters γ_2, γ_3 and γ_4 lie even outside of the estimated posterior 95% credibility interval. The other time series and the correlation parameter estimates with MH algorithm and with marginal MLE / C-MLE are very closely together. The estimators for the degrees of freedom for the unconditioned copulas are also adjacent. Only the estimates for the degrees of freedom of the conditional copulas $\nu_{13|2}$, $\nu_{24|3}$ and $\nu_{14|23}$ have a larger difference and their absolute value is quite high.
We further remark that for the corresponding correlation parameters $\rho_{13|2}$, $\rho_{24|3}$ and $\rho_{14|23}$, zero belongs to the 90% credibility interval and it therefore cannot be rejected that these parameters are zero. Since it is possible that there is no dependency for the conditional copulas, we try to fit a normal copula or as an approximation, we fix the degrees of freedom for the conditional copulas to 100. So, to recapitulate, we try to estimate all parameters with another MCMC algorithm where we keep $\rho_{13|2}$, $\rho_{24|3}$ and $\rho_{14|23}$ fixed to 0 and $\nu_{13|2}$, $\nu_{24|3}$ and $\nu_{14|23}$ fixed at 100. For the C-MLE values, we estimate the conditional copulas as a normal copula with correlation 0. The results of the C-MLE and the MCMC are shown in Table 5.6

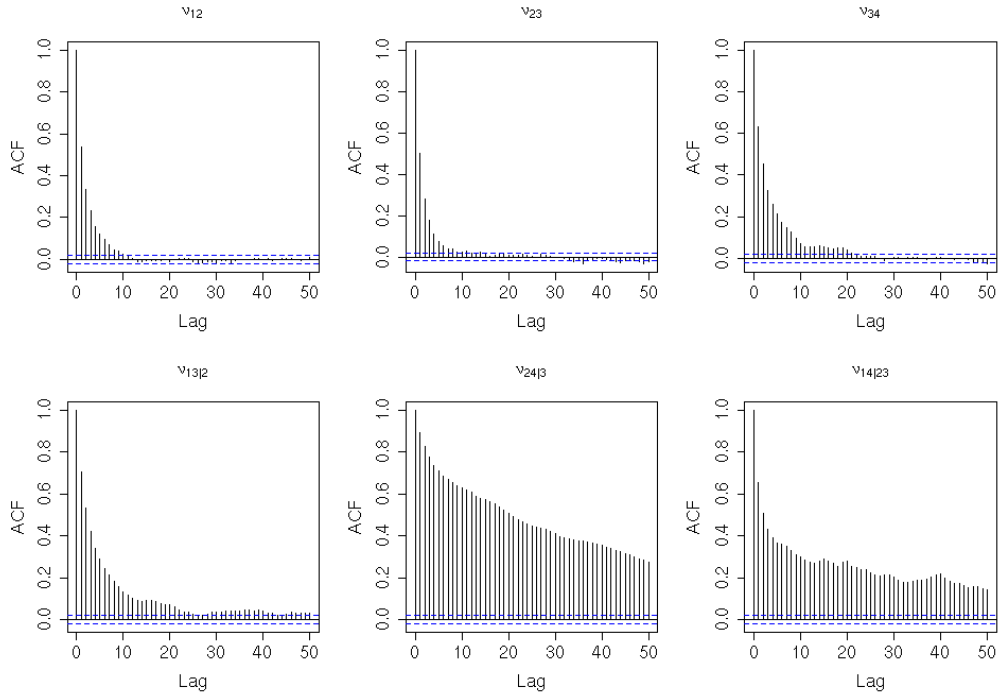

Figure 5.6: Plots of the empirical autocorrelation for MCMC iterations of ν estimated of the observed real data based on all iterations

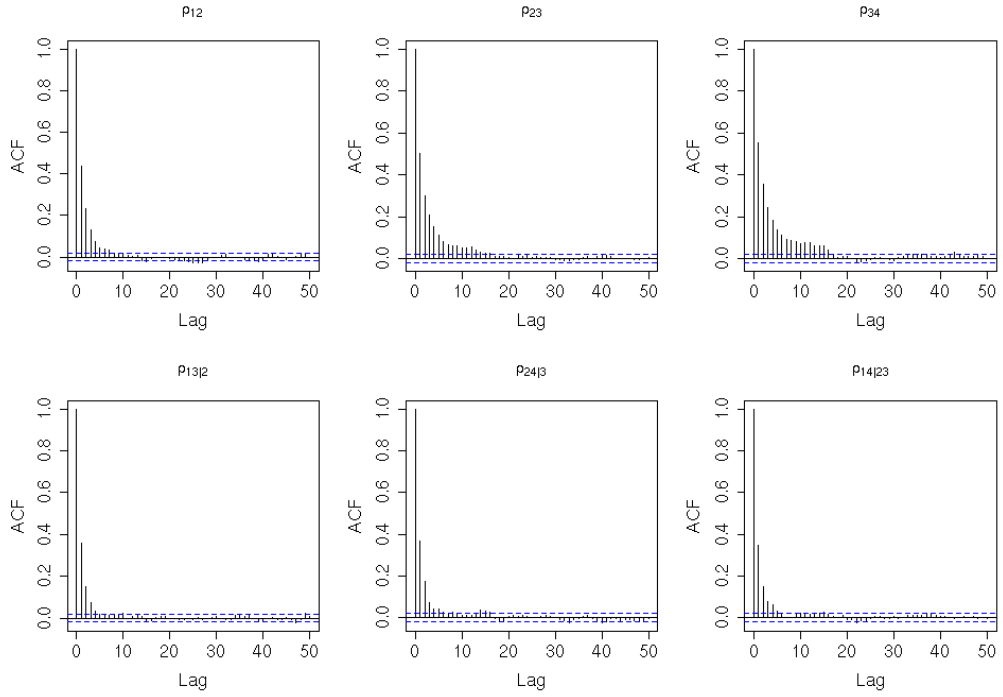

Figure 5.7: Plots of the empirical autocorrelation for MCMC iterations of ρ estimated of the observed real data based on all iterations

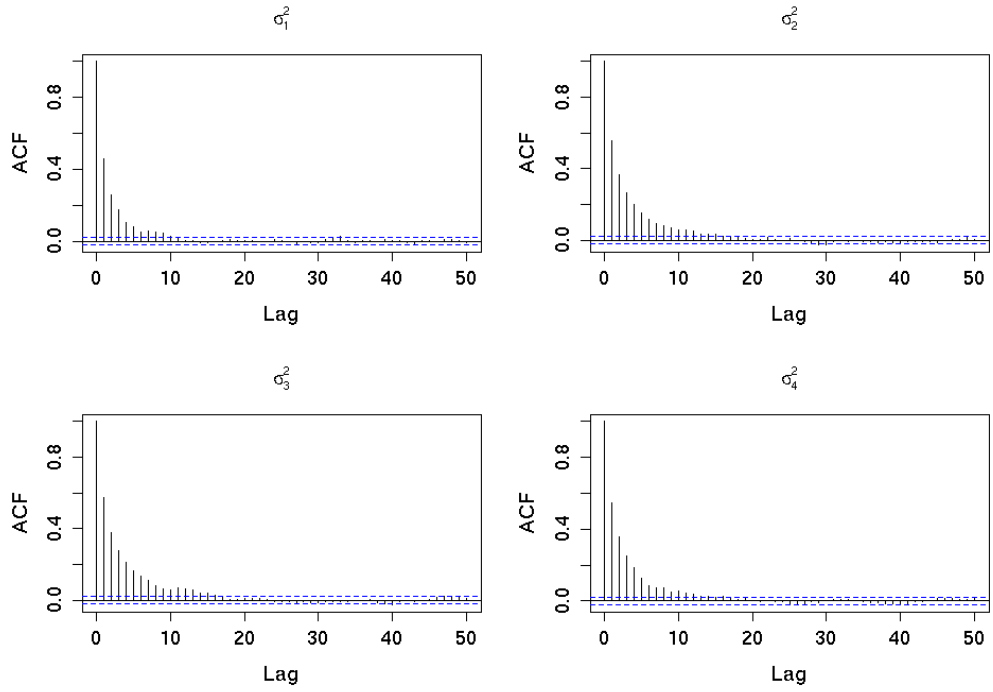

Figure 5.8: Plots of the empirical autocorrelation for MCMC iterations of σ^2 estimated of the observed real data based on all iterations

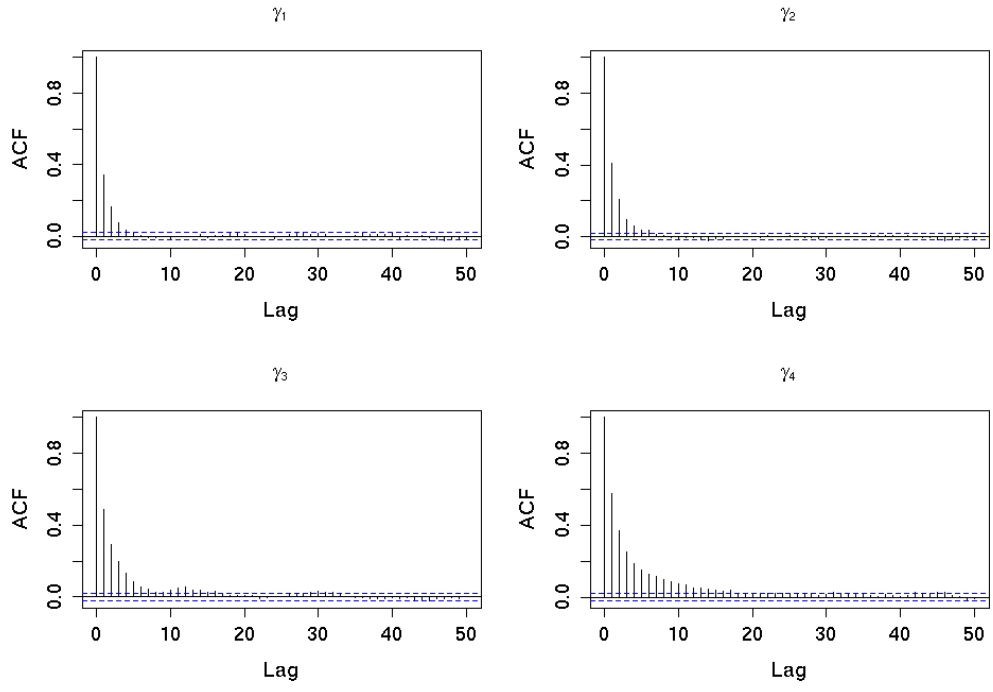

Figure 5.9: Plots of the empirical autocorrelation for MCMC iterations of γ estimated of the observed real data based on all iterations

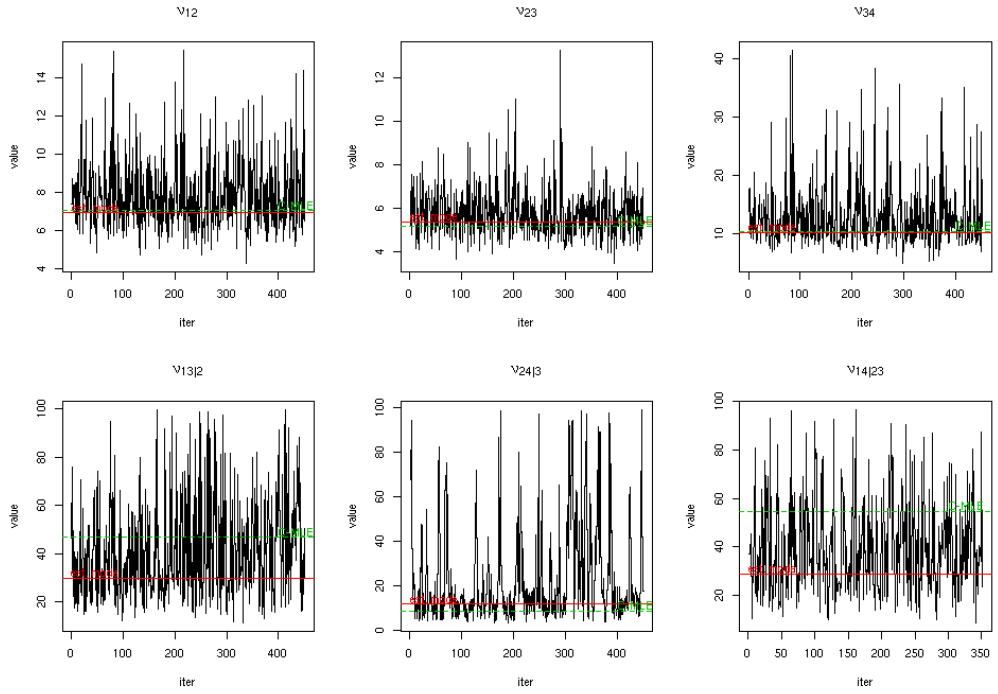

Figure 5.10: Trace Plots of the thinned out Chain of ν estimated of the observed real data

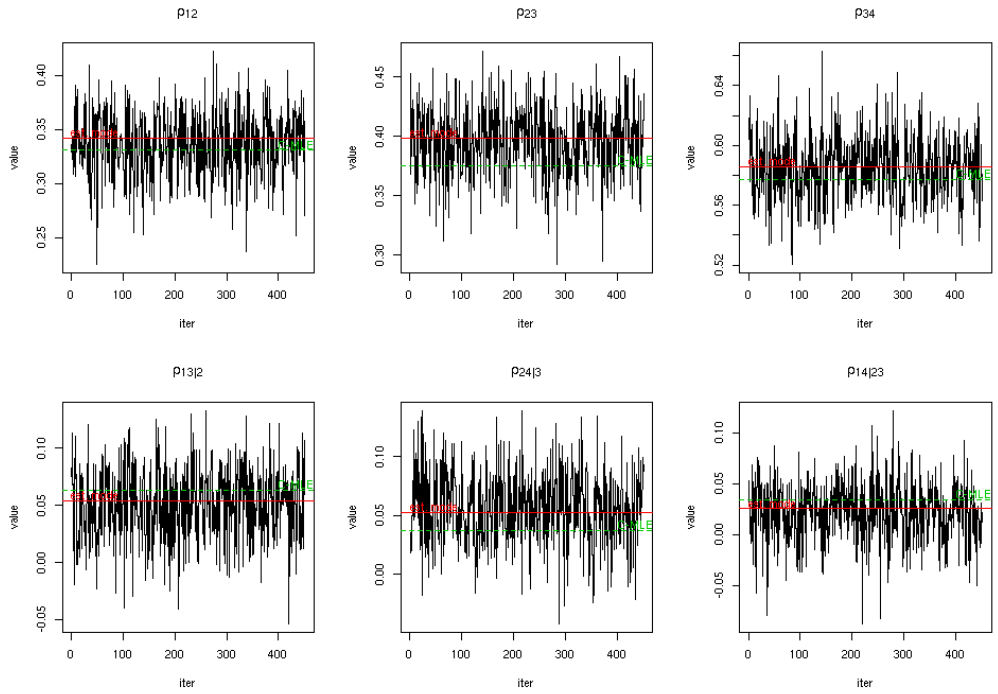

Figure 5.11: Trace Plots of the thinned out Chain of ρ estimated of the observed real data

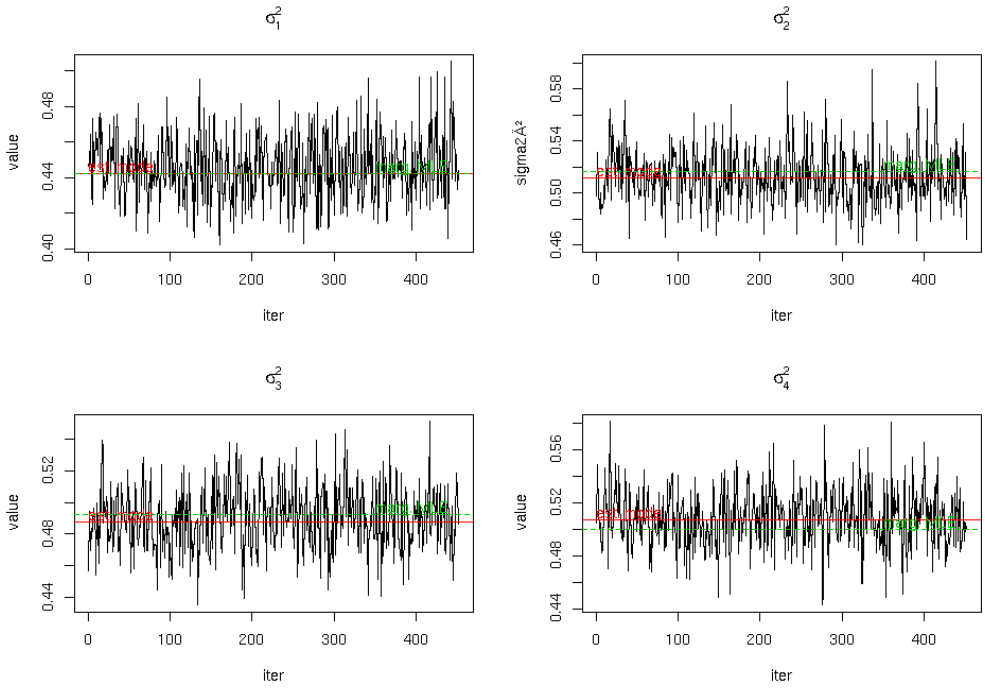

Figure 5.12: Trace Plots of the thinned out Chain of σ^2 estimated of the observed real data

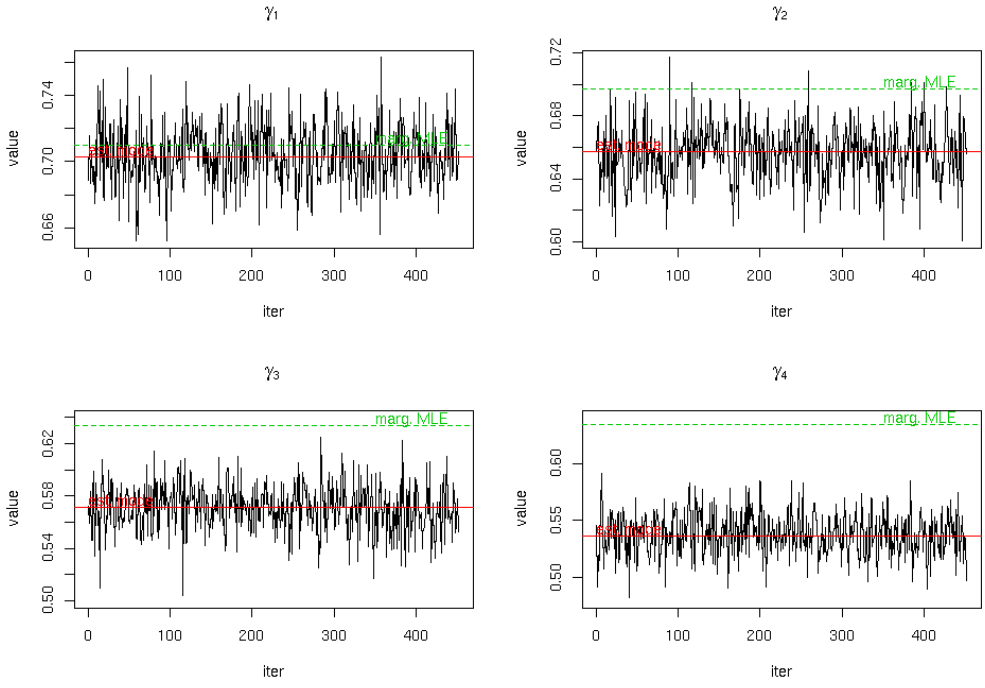

Figure 5.13: Trace Plots of the thinned out Chain of γ estimated of the observed real data

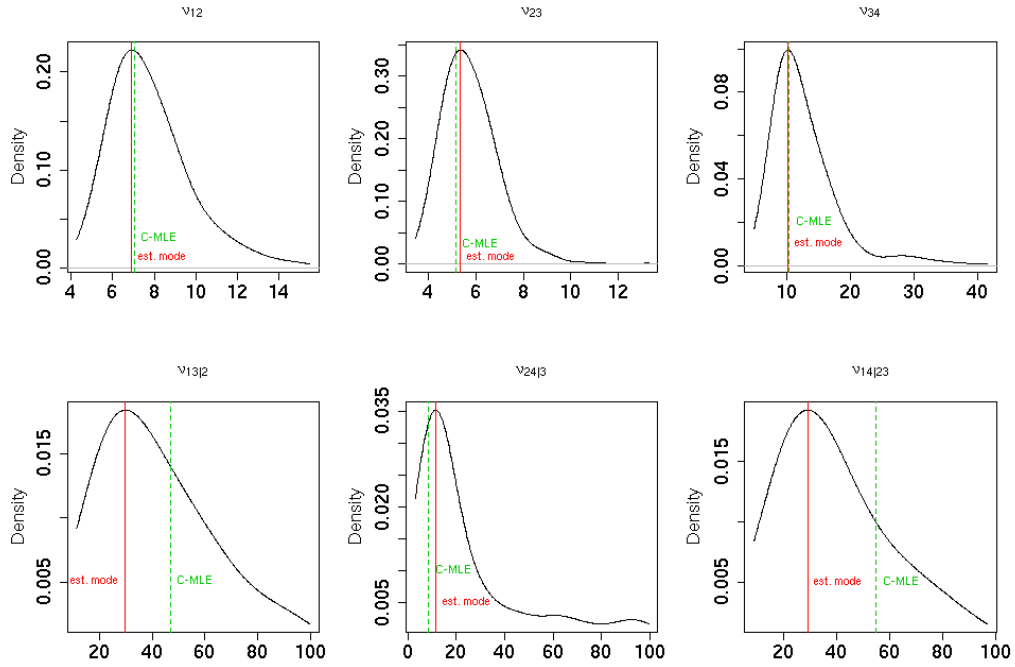

Figure 5.14: Plots of the estimated kernel density of ν estimated of the observed real data based on the thinned out MCMC chain

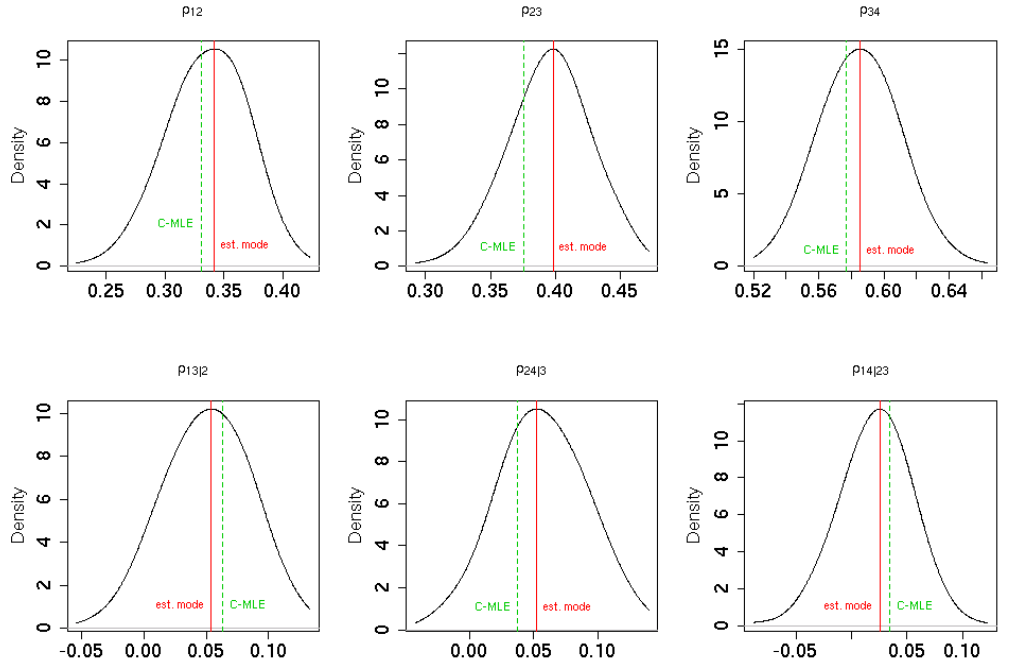

Figure 5.15: Plots of the estimated kernel density of ρ estimated of the observed real data based on the thinned out MCMC chain

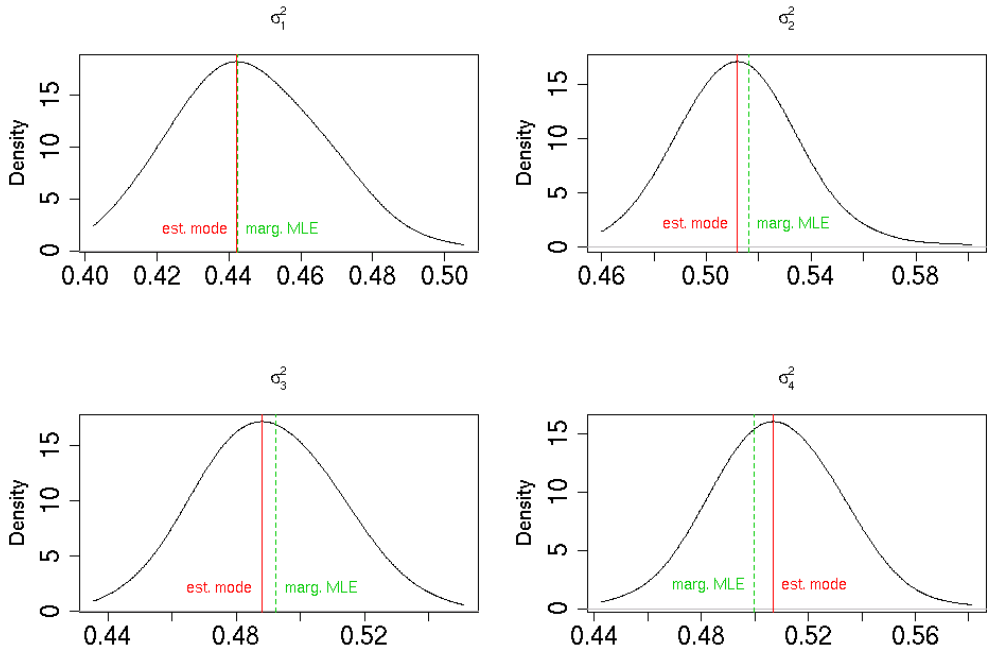

Figure 5.16: Plots of the estimated kernel density of σ^2 estimated of the observed Real Data based on the thinned out MCMC chain

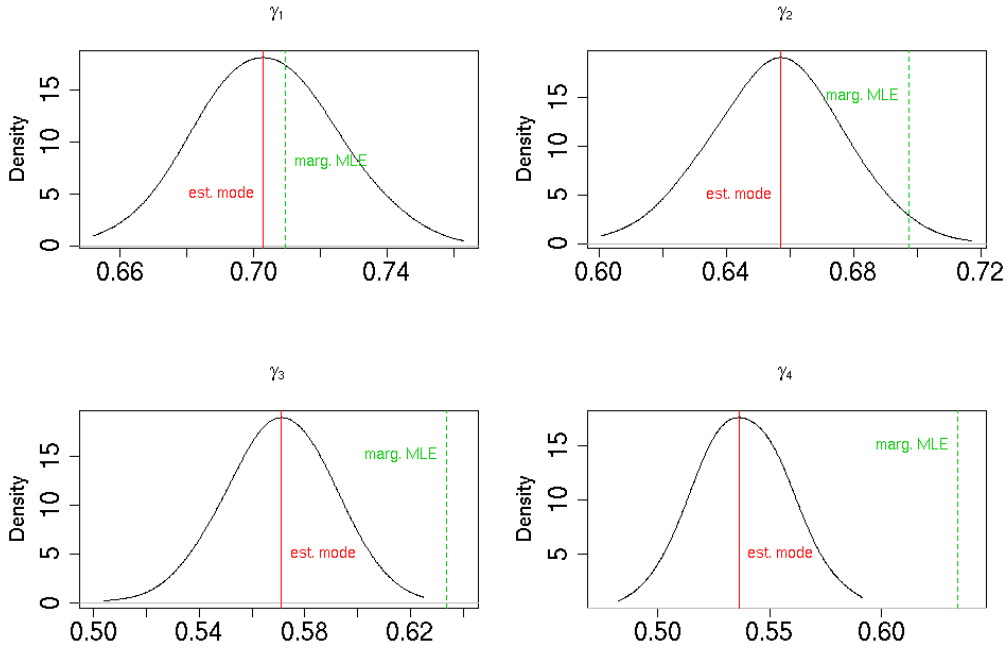

Figure 5.17: Plots of the estimated kernel density of γ estimated of the observed real data based on the thinned out MCMC chain

	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	5.17	5.45	7.37	11.72	12.71	7.80	6.92	7.03	7.32
ν_{23}	4.11	4.26	5.57	7.65	8.60	5.76	5.36	5.16	5.52
ν_{34}	7.08	7.69	11.45	24.29	29.68	12.89	10.23	10.47	11.17
$\nu_{13\|2}$	15.01	16.28	36.89	84.56	91.77	41.44	29.80	46.77	38.55
$\nu_{24\|3}$	4.20	4.73	14.25	78.22	93.44	22.84	11.60	8.56	8.34
$\nu_{14\|23}$	12.32	14.58	34.43	78.22	86.81	38.94	29.00	54.71	58.07
ρ_{12}	0.27	0.28	0.34	0.38	0.39	0.34	0.34	0.33	0.33
ρ_{23}	0.33	0.35	0.40	0.45	0.45	0.40	0.40	0.38	0.39
ρ_{34}	0.54	0.55	0.59	0.62	0.63	0.59	0.59	0.58	0.57
$\rho_{13\|2}$	−0.01	−0.00	0.05	0.10	0.11	0.05	0.05	0.06	0.06
$\rho_{24\|3}$	−0.01	0.00	0.05	0.11	0.12	0.06	0.05	0.04	0.04
$\rho_{14\|23}$	−0.04	−0.03	0.02	0.07	0.08	0.02	0.03	0.03	0.03
	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE	
σ_1^2	0.41	0.41	0.44	0.48	0.48	0.45	0.44	0.44	
γ_1	0.67	0.67	0.70	0.74	0.74	0.70	0.70	0.71	
σ_2^2	0.47	0.48	0.51	0.55	0.56	0.51	0.51	0.52	
γ_2	0.62	0.62	0.66	0.69	0.69	0.66	0.66	0.70	
σ_3^2	0.45	0.46	0.49	0.52	0.53	0.49	0.49	0.49	
γ_3	0.53	0.54	0.57	0.60	0.61	0.57	0.57	0.63	
σ_4^2	0.47	0.47	0.51	0.54	0.55	0.51	0.51	0.50	
γ_4	0.50	0.51	0.54	0.57	0.58	0.54	0.54	0.63	

Table 5.5: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE, starting values and C-MLE for the Observed Australian Load Data preprocessed for Stationarity with Method 8 (cf. Section 3.5)

	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	5.02	5.29	7.32	10.87	11.65	7.56	6.90	7.03	7.19
ν_{23}	3.87	4.07	5.34	7.28	7.86	5.47	5.17	5.16	5.26
ν_{34}	6.91	7.38	11.34	20.23	22.48	12.11	10.51	10.47	10.61
ρ_{12}	0.27	0.27	0.33	0.38	0.39	0.33	0.34	0.33	0.33
ρ_{23}	0.32	0.34	0.39	0.44	0.45	0.39	0.39	0.38	0.38
ρ_{34}	0.54	0.55	0.59	0.62	0.63	0.59	0.59	0.58	0.58
	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE	
σ_1^2	0.41	0.41	0.44	0.47	0.48	0.44	0.44	0.44	
γ_1	0.67	0.67	0.71	0.74	0.74	0.71	0.71	0.71	
σ_2^2	0.47	0.47	0.51	0.54	0.55	0.51	0.51	0.52	
γ_2	0.62	0.62	0.66	0.69	0.69	0.66	0.66	0.70	
σ_3^2	0.45	0.46	0.49	0.53	0.53	0.49	0.49	0.49	
γ_3	0.53	0.53	0.56	0.60	0.61	0.57	0.56	0.63	
σ_4^2	0.47	0.48	0.51	0.55	0.55	0.51	0.50	0.50	
γ_4	0.50	0.50	0.54	0.57	0.58	0.54	0.54	0.63	

Table 5.6: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE and C-MLE in the reduced model (all conditional correlation parameters are fixed to 0 and the corresponding degrees of freedom fixed to 100) for the Observed Australian Load Data preprocessed for Stationarity with Method 8 (cf. Section 3.5)

5.4 Comparison to Two-Step Estimation Simulation

We are also interested in the differences between the joint Bayesian estimation of all copula and time series parameters and the two step estimation consisting of a marginal Maximum Likelihood estimation of the time series parameters, an enclosing non-parametric transformation of the residuals using the empirical cumulative distribution function and the following Bayesian estimation of the D-vine parameters (MMLE-CMCMC, cf. Table 2.1). In both MH algorithms, we use the same prior densities for ν_j and ρ_j as described in this Chapter after equation (5.1).
For this comparison study, we chose a scenario adopted to the with marginal MLE and C-MLE estimated parameter values of the observed Australian load data. So, we take:

ν_{12}	ν_{23}	ν_{34}	$\nu_{13\|2}$	$\nu_{24\|3}$	$\nu_{14\|23}$	σ_1^2	γ_1	σ_2^2	γ_2
7.3	5.5	11	30	8.3	50	0.44	0.71	0.52	0.70
ρ_{12}	ρ_{23}	ρ_{34}	$\rho_{13\|2}$	$\rho_{24\|3}$	$\rho_{14\|23}$	σ_3^2	γ_3	σ_4^2	γ_4
0.33	0.38	0.57	0.06	0.04	0.03	0.49	0.63	0.50	0.63

We draw a sample from a D-vine with marginal AR(1) model with the above denoted parameters. Then we estimate the parameters using JMCMC, MMLE-CMCMC, TSP-MLE and TSNP-MLE (cf. Table 2.1). We redo these steps until we have four replications. We can interpret the results with regard to the marginal time series parameters that the Bayesian estimation is better than the marginal MLE while the standard deviation is quite equal. The D-vine copula parameters are quite similar with no obvious advantage for any of the two estimation methods.
An additional and very interesting comparison is to check the joint Bayesian estimation against a completely Bayesian two step estimation where the marginal time series parameters are also estimated using a MCMC method. Perhaps, it is possible to take the posterior density calculated in the first step Bayes estimation as a starting point for the determination of a prior distribution for the second MCMC algorithm.

		true	2.5%	5%	50%	95%	97.5%	mean	mode	start	C-MLE
ν_{12}	value	7.300	4.468	4.777	6.920	12.040	13.372	7.485	6.463	7.160	7.120
	sd		0.350	0.440	0.913	2.717	3.135	1.173	0.846	0.839	0.878
ν_{23}	value	5.500	3.743	3.922	5.348	8.178	8.982	5.610	5.178	5.422	5.513
	sd		0.357	0.378	0.514	1.664	1.932	0.689	0.429	0.593	0.577
ν_{34}	value	11.000	7.317	8.365	16.348	38.847	46.040	18.965	13.963	22.820	22.745
	sd		2.386	3.352	9.405	25.522	30.689	11.252	7.223	19.887	20.229
$\nu_{13\|2}$	value	30.000	10.572	11.820	28.968	72.155	78.565	34.258	23.032	79.250	217.925
	sd		2.421	2.728	10.190	23.661	23.224	11.538	7.240	41.500	135.653
$\nu_{24\|3}$	value	8.300	4.138	4.735	15.367	60.807	73.445	22.200	12.062	8.578	8.630
	sd		0.999	1.065	4.286	27.365	27.753	7.313	3.032	2.944	3.045
$\nu_{14\|23}$	value	50.000	10.235	11.078	28.793	70.442	76.175	34.500	22.082	62.822	129.328
	sd		2.745	2.874	16.240	25.509	22.979	17.725	9.561	42.929	131.982
ρ_{12}	value	0.330	0.270	0.280	0.333	0.383	0.390	0.333	0.335	0.338	0.338
	sd		0.022	0.022	0.022	0.019	0.014	0.022	0.019	0.017	0.017
ρ_{23}	value	0.380	0.315	0.323	0.375	0.425	0.435	0.375	0.378	0.380	0.380
	sd		0.024	0.026	0.024	0.024	0.024	0.024	0.022	0.027	0.027
ρ_{34}	value	0.570	0.520	0.527	0.568	0.605	0.607	0.565	0.568	0.568	0.570
	sd		0.032	0.033	0.033	0.026	0.025	0.029	0.033	0.025	0.029
$\rho_{13\|2}$	value	0.060	0.018	0.028	0.085	0.140	0.148	0.087	0.085	0.087	0.087
	sd		0.017	0.017	0.021	0.018	0.017	0.017	0.021	0.017	0.017
$\rho_{24\|3}$	value	0.040	−0.037	−0.030	0.028	0.083	0.092	0.028	0.028	0.030	0.030
	sd		0.049	0.045	0.050	0.046	0.046	0.050	0.050	0.045	0.045
$\rho_{14\|23}$	value	0.030	−0.020	−0.007	0.048	0.100	0.110	0.048	0.048	0.048	0.048
	sd		0.018	0.017	0.015	0.018	0.018	0.015	0.015	0.017	0.017
		true	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE	
σ_1^2	value	0.440	0.395	0.402	0.438	0.468	0.478	0.438	0.438	0.427	
	sd		0.034	0.030	0.035	0.035	0.035	0.035	0.035	0.035	
γ_1	value	0.710	0.670	0.675	0.712	0.740	0.748	0.710	0.712	0.705	
	sd		0.016	0.017	0.017	0.016	0.021	0.016	0.017	0.024	
σ_2^2	value	0.520	0.468	0.473	0.512	0.550	0.557	0.512	0.510	0.505	
	sd		0.021	0.021	0.021	0.023	0.021	0.021	0.023	0.024	
γ_2	value	0.700	0.645	0.653	0.682	0.712	0.720	0.682	0.682	0.680	
	sd		0.010	0.010	0.010	0.010	0.008	0.010	0.010	0.000	
σ_3^2	value	0.490	0.435	0.443	0.475	0.512	0.520	0.475	0.470	0.470	
	sd		0.013	0.015	0.013	0.015	0.014	0.013	0.014	0.014	
γ_3	value	0.630	0.568	0.573	0.605	0.637	0.645	0.605	0.605	0.598	
	sd		0.022	0.017	0.021	0.022	0.021	0.021	0.021	0.030	
σ_4^2	value	0.500	0.455	0.458	0.495	0.530	0.540	0.495	0.495	0.487	
	sd		0.021	0.022	0.021	0.026	0.026	0.021	0.021	0.025	
γ_4	value	0.630	0.580	0.585	0.617	0.653	0.657	0.617	0.617	0.620	
	sd		0.014	0.010	0.015	0.013	0.015	0.015	0.015	0.014	

Table 5.7: Estimated posterior mean, mode and quantiles of JMCMC as well as marginal MLE, starting values and C-MLE (TSP-MLE) for the scenario described in Section 5.4 averaged over 4 replications with its empirical standard deviation

		true	2.5%	5%	50%	95%	97.5%	mean	mode	C-MLE
ν_{12}	value	7.300	4.362	4.630	6.565	11.842	14.147	7.205	6.060	6.822
	sd		0.410	0.451	0.930	3.176	4.897	1.260	0.785	1.048
ν_{23}	value	5.500	3.825	4.067	5.418	8.140	8.812	5.633	5.133	5.450
	sd		0.307	0.431	0.665	1.695	1.870	0.778	0.545	0.644
ν_{34}	value	11.000	8.075	8.720	17.023	50.345	58.475	21.070	13.260	24.497
	sd		2.614	2.874	9.074	31.998	35.701	11.114	5.866	22.844
$\nu_{13\|2}$	value	30.000	10.915	12.100	29.688	73.138	80.500	34.517	20.415	186.493
	sd		2.586	3.046	11.321	23.122	21.123	11.693	6.254	139.359
$\nu_{24\|3}$	value	8.300	4.878	5.240	7.935	14.912	17.387	8.720	6.975	8.158
	sd		0.685	0.869	2.072	6.312	7.201	2.516	1.323	2.407
$\nu_{14\|23}$	value	50.000	10.143	11.183	25.828	76.392	84.508	31.947	17.922	100.532
	sd		2.125	2.501	6.469	4.044	4.044	5.246	4.238	133.786
ρ_{12}	value	0.330	0.270	0.283	0.330	0.380	0.388	0.330	0.333	0.338
	sd		0.026	0.021	0.022	0.014	0.010	0.022	0.021	0.017
ρ_{23}	value	0.380	0.320	0.330	0.378	0.422	0.430	0.378	0.380	0.383
	sd		0.027	0.027	0.029	0.026	0.028	0.029	0.022	0.026
ρ_{34}	value	0.570	0.525	0.532	0.568	0.597	0.605	0.568	0.568	0.570
	sd		0.034	0.030	0.025	0.025	0.026	0.025	0.025	0.029
$\rho_{13\|2}$	value	0.060	0.022	0.033	0.090	0.138	0.148	0.090	0.088	0.087
	sd		0.017	0.017	0.018	0.022	0.017	0.018	0.015	0.022
$\rho_{24\|3}$	value	0.040	−0.035	−0.028	0.028	0.087	0.098	0.030	0.030	0.028
	sd		0.047	0.046	0.042	0.042	0.042	0.045	0.045	0.042
$\rho_{14\|23}$	value	0.030	−0.020	−0.007	0.045	0.103	0.117	0.048	0.048	0.048
	sd		0.018	0.015	0.013	0.013	0.010	0.017	0.017	0.017
		true	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	value	0.440								0.427
	sd									0.035
γ_1	value	0.710								0.705
	sd									0.024
σ_2^2	value	0.520								0.505
	sd									0.024
γ_2	value	0.700								0.680
	sd									0.000
σ_3^2	value	0.490								0.470
	sd									0.014
γ_3	value	0.630								0.598
	sd									0.030
σ_4^2	value	0.500								0.487
	sd									0.014
γ_4	value	0.630								0.620
	sd									0.014

Table 5.8: Estimated posterior mean, mode and quantiles of MMLE-CMCMC as well as marginal MLE and C-MLE (TSNP-MLE) for the scenario described in Section 5.4 averaged over 4 replications with its empirical standard deviation

5.5 Internal Validation of Different Models

We have already seen that we can fit a model to the observed Australian load data by using a D-vine pair copula construction with t-copulas and marginal AR(1) models. Furthermore, we have fitted a reduced model where the correlation parameters of the conditional copulas of the D-vine $\rho_{13|2}$, $\rho_{24|3}$ and $\rho_{14|23}$ are fixed to zero and the corresponding degrees of freedom $\nu_{13|2}$, $\nu_{24|3}$ and $\nu_{14|23}$ are fixed to 100. For these two cases, we take the estimated posterior mode as the estimator for the parameters.
Now we want to suggest two further models to fit to the data. Firstly, we try to fit with Maximum Likelihood estimation a four-dimensional t-copula with a single parameter for the degrees of freedom and marginal AR(1) models. Therefore, we use the marginally estimated time series parameters with the transformation (4.1) and the cdf of the standard normal distribution to transform the observed data parametricly to uniformly distributed data. Then we try to fit the four-dimensional t-copula and to estimate the copula parameters. On the other hand, we include a vector autoregressive model of order 1, $\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t, \boldsymbol{Z}_t \sim WN(\boldsymbol{0}, \tilde{\Sigma}), \Phi \in \mathbb{R}^{m \times m}$ (cf. Equation (2.21)). We fit this model to the data using stepwise least-square estimation and calculate the residuals. Then we transform these residuals non-parametricly by its empirical cumulative distribution function to uniformly distributed data.
So, we consider the following four models:

Nr.	Description of the Estimation Method	Nr. of Parameters
1	marginal AR(1) and four dimensional t-Copula	$6+1+8=15$
2	Vector Autoregressive Process of Order 1	$16+16+4=36$
3	Joint Bayesian Estimation with marginal AR(1) and D-Vine of t-Copulas (JMCMC)	$6+6+8=20$
4	Joint Bayesian Estimation of the reduced model: marginal AR(1), unconditional Copulas as t-Copulas, conditional Copulas with correlation 0 and degrees of freedom = 100	$3+3+8=14$

We get the following results for the four methods:

(1)

marginal MLE:	σ_1^2	γ_1	σ_2^2	γ_2	σ_3^2	γ_3	σ_4^2	γ_4
	0.71	0.44	0.70	0.52	0.63	0.49	0.63	0.50
4-dim t-copula parameter:		ν	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}
		9.59	0.30	0.11	0.09	0.34	0.21	0.58

(2) $$\Phi = \begin{pmatrix} 0.66 & 0.14 & -0.09 & -0.00 \\ -0.02 & 0.54 & 0.19 & 0.16 \\ 0.00 & -0.05 & 0.50 & 0.25 \\ -0.03 & 0.01 & -0.05 & 0.66 \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} 0.43 & 0.18 & 0.10 & 0.08 \\ 0.18 & 0.46 & 0.18 & 0.13 \\ 0.10 & 0.18 & 0.47 & 0.28 \\ 0.08 & 0.13 & 0.28 & 0.50 \end{pmatrix}$$

(3)

ν_{12}	ν_{23}	ν_{34}	$\nu_{13\|2}$	$\nu_{24\|3}$	$\nu_{14\|23}$	σ_1^2	γ_1	σ_2^2	γ_2
6.92	5.36	10.23	29.80	11.60	29.00	0.44	0.70	0.51	0.66
ρ_{12}	ρ_{23}	ρ_{34}	$\rho_{13\|2}$	$\rho_{24\|3}$	$\rho_{14\|23}$	σ_3^2	γ_3	σ_4^2	γ_4
0.34	0.40	0.59	0.05	0.05	0.03	0.49	0.57	0.51	0.54

(cf. estimated posterior mode in Table 5.5)

(4)

ν_{12}	ν_{23}	ν_{34}	ρ_{12}	ρ_{23}	ρ_{34}		
6.90	5.17	10.51	0.34	0.39	0.59		
σ_1^2	γ_1	σ_2^2	γ_2	σ_3^2	γ_3	σ_4^2	γ_4
0.44	0.71	0.51	0.66	0.49	0.56	0.50	0.54

(while the df parameters of the conditional copulas are fixed at 100 and the corresponding correlation parameters are fixed at 0, cf. estimated posterior mode in Table 5.6)

Hereby we have to remark that the correlation parameters determined with method 1 are 'standard' product moment correlations and not the conditional correlation parameters as determined with the D-vine pair copula construction. To have a quick check, we want to transform them to the corresponding conditional correlations using the relation (2.7) and the fact that for the elliptical distributions, partial correlations and conditional correlation coincide (cf. the end of Section 2.4). Then we want to compare them to the results of method 3, the JMCMC. Normally, we had to compare them to the confidence intervals, but since we do not know them, we take the comparison to the credible intervals of method 3 (described in Table 5.5) as an approximation. So we get:

$$\rho_{13|2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2}} = 0.0053$$

$$\rho_{24|3} = \frac{\rho_{24} - \rho_{23}\rho_{34}}{\sqrt{1-\rho_{23}^2}\sqrt{1-\rho_{34}^2}} = 0.0183$$

$$\rho_{14|23} = \frac{\rho_{14|2} - \rho_{13|2}\rho_{34|2}}{\sqrt{1-\rho_{13|2}^2}\sqrt{1-\rho_{34|2}^2}} = 0.0292$$

Analyzing the $h(\cdot)$-function, we recognize that the degrees of freedom increase by one for the conditional copula in the second tree and by 2 in the third tree (cf. Figure 2.5). So we can deduce that $\nu_{12} = \nu_{23} = \nu_{34} = 9.59, \nu_{13|2} = \nu_{24|3} = 10.59$ and $\nu_{14|23} = 11.59$. Checking these values with the credible intervals resulting from the JMCMC, we remark that the calculated values for $\rho_{13|2}$, $\rho_{24|3}$, $\rho_{14|23}$ and ν_{12}, ν_{34} and $\nu_{24|3}$ are inside these bounds. However, the calculated values for ν_{23}, $\nu_{13|2}$ and $\nu_{14|23}$ lie outside the corresponding 95% credible interval of JMCMC. This is a clear hint that the fit of the four-dimensional copula is not as good as the fit of the D-vine with t-copulas as building blocks.

For creating a meaningful comparison of the four different methods, we encountered a lot of difficulties. How can you compare a vector autoregressive model with a combined autoregressive and copula model? Which statistic can you look at? Is there any test for this multivariate data? How can you compare these models with a test when you also have to consider the order of the values?

Therefore, we have decided to perform an internal validation of the different estimation procedures. We have chosen a multivariate Cramér test introduced by Baringhaus and Franz (2004). This test is implemented in R with the package 'cramer'. Its two-sample test statistic is

$$T_{m,n} = \frac{mn}{m+n}\left(\frac{2}{mn}\sum_{s,t=1}^{m,n}\frac{\|\boldsymbol{X^s}-\boldsymbol{Y^t}\|}{2} - \frac{1}{m^2}\sum_{s,t=1}^{m}\frac{\|\boldsymbol{X^s}-\boldsymbol{X^t}\|}{2} - \frac{1}{n^2}\sum_{s,t=1}^{n}\frac{\|\boldsymbol{Y^s}-\boldsymbol{Y^t}\|}{2}\right),$$

where $\boldsymbol{X^s}, s = 1, \dots m$ and $\boldsymbol{Y^t}, t = 1, \dots n$ are the two multivariate samples to be compared. The critical value is determined by the bootstrap method with 1000 replicates and a confidence level of 90%. We repeat the test in each method 100 times.

For the first method, the four-dimensional t-copula with AR(1) margins, we use the estimated time series parameters of the marginal MLE, the transformation (4.1) and the cdf of the standard normal distribution to transform the observed data parametricly to the unit cube. This is our basis point for the comparison. Then we estimate the copula parameters (one parameter for the degrees of freedom and 6 parameters for the correlations) out of these 1135 values. Afterwards, we simulate from the four-dimensional t-copula with the just estimated parameter values 1135 data points. We compare this sample to our basis point with the Cramér test.
The second method consists of the vector autoregressive process of order 1. Our basis points for the internal validation in this case are the residuals of the observed time series and the fitted values which we transform non-parametricly with the empirical cdf to the unit cube. We simulate now 1135 values from the estimated vector autoregressive para-meters and we calculate the residuals of the observed data and our simulated values. Then we transform them non-parametricly to the unit cube and apply the Cramér test with our basis points.
For the internal validation of the JMCMC, we transform the observed data parametricly to the unit cube using the estimated posterior mode of the marginal time series parameters, the transformation 4.1 and the cdf of the standard normal distribution. This constitutes our point of comparison for this third method. We simulate then 1135 data points from a D-vine consisting of pairs of t-copulas with the estimated posterior modes of the JMCMC and test them against our point of comparison with the Cramér test.
For the fourth method, we transform the observed values to the unit cube like in the previous method and this is also our basis point. Then we sample 1135 data points from a D-vine with the estimated parameters where the df parameters of the conditional copulas are fixed to 100 and the corresponding correlation parameters are fixed to 0. We perform then the Cramér test against the basis point.

After 100 replications, we get the following results:

Nr.	Description of the Estimation Method	Nr. of rejections
1	marginal AR(1) and four dimensional t-Copula	58
2	Vector Autoregressive Process of Order 1	64
3	Joint Bayesian Estimation with marginal AR(1) and D-Vine of t-Copulas (JMCMC)	43
4	Joint Bayesian Estimation of the reduced model: marginal AR(1), unconditional Copulas as t-Copulas, conditional Copulas with correlation 0 and degrees of freedom = 100	34

This indicates that the fit of the D-vine with bivariate t-copulas as building blocks is better than the one of the four-dimensional t-copula and of the vector autoregressive process of order 1. This observation is in line with the results of the quick check denoted above. However, we remark here that the reduced model has a better validation and it seems that this reduced model is enough for a good fit to the data.

5.6 Bayesian Model Selection

Another very interesting aspect is the Bayesian model selection. We have fitted 6 different models $M_1, \ldots, M_6$ with a MCMC method where model M_k has parameters $\boldsymbol{\theta}_k$. Now, we want to compare them on the basis of posterior model probabilities which are given by

$$P(\text{Model } M_k|\text{data}), \qquad k = 1, \ldots, 6.$$

Since we have already fitted all models in a Bayesian setup, we will use the method of Congdon (2006). Our approach follows the presentation of Czado and Min (2007). Congdon makes the assumptions that the distribution of the data is independent of $\{\boldsymbol{\theta}_{j\neq k}\}$ given M_k. Furthermore, he assumes independence among all $\boldsymbol{\theta}_k$ given Model M. Then he shows that the posterior distributions are independent and can be sampled individually. He uses the relation

$$P(M = M_k|\text{data}, \boldsymbol{\theta}) \propto P(\text{data}|\boldsymbol{\theta}, M = M_k)P(\boldsymbol{\theta}|M = M_k)P(M = M_k). \tag{5.2}$$

So if we assume that the 6 independent MCMC runs result in

$$\begin{array}{lcl} M_1 : \boldsymbol{\theta}_1^{(t)}, r = 1, \ldots R & & p(\boldsymbol{\theta}_1|\text{data}) \\ \vdots & \text{which approximate} & \vdots \\ M_6 : \boldsymbol{\theta}_6^{(t)}, r = 1, \ldots R & & p(\boldsymbol{\theta}_6|\text{data}) \end{array}$$

we get an estimator. We use $\{\boldsymbol{\theta}^{(r)} := (\boldsymbol{\theta}_1^{(r)}, \ldots, \boldsymbol{\theta}_6^{(r)}), r = 1, \ldots R\}$ and hence, we can approximate

$$P(M|\text{data}) = \int P(M|\boldsymbol{\theta}, \text{data})p(\boldsymbol{\theta}|\text{data})\mathrm{d}\boldsymbol{\theta}$$

by

$$P(M|\hat{\text{data}}) := \frac{1}{R}\sum_{r=1}^{R} P(M|\boldsymbol{\theta}^{(r)}, \text{data}).$$

Using the equation (5.2), we can estimate $P(M = M_k|\text{data}, \boldsymbol{\theta}^{(r)})$ by

$$w_k^{(r)} := \frac{G_k^{(r)}}{\sum_{j=1}^{6} G_k^{(r)}},$$

where

$$\begin{aligned} G_k^{(r)} &:= \exp(L_k^{(r)} - L_{\max}^{(r)}) \\ L_k^{(r)} &:= \log(P(\text{data}|\boldsymbol{\theta}^{(r)}, M = M_k)P(\boldsymbol{\theta}^{(r)}|M = M_k)P(M = M_k)) \\ L_{\max}^{(r)} &:= \max_{k=1,\ldots,6} L_k^{(r)}. \end{aligned}$$

Therefore, we get

$$\hat{T}_k := \frac{1}{R}\sum_{r=1}^{R} w_k^{(r)}$$

as the estimator for $P(M = M_k|\text{data})$.

We have performed a Bayesian estimation for the following six methods:

Nr.	Description of the Estimation Method
1	Joint Bayesian Estimation with marginal AR(1) and D-Vine of t-Copulas (JMCMC)
2	Joint Bayesian Estimation of the reduced model of 1: marginal AR(1), unconditional copulas as t-copulas, conditional copulas with correlation 0 and degrees of freedom = 100
3	Joint Bayesian Estimation with marginal AR(1) and D-Vine of normal copulas (approximated by t-copula with df=100)
4	Joint Bayesian Estimation of the reduced model of 4: marginal AR(1), unconditional Copulas as gauss copulas (approximated by t-copula df=100), conditional copulas with correlation 0
5	marginal AR(1) and four-dimensional t-Copula
6	marginal AR(1) and four-dimensional t-Copula with the conditional correlation parameters fixed to 0

After the MCMC simulation, we got the following results:

(1) cf. Table 5.5

(2) cf. Table 5.6

	2.5%	5%	50%	95%	97.5%	mean	mode	C-MLE
ρ_{12}	0.30	0.31	0.35	0.40	0.41	0.35	0.35	0.35
ρ_{23}	0.36	0.36	0.41	0.45	0.46	0.41	0.41	0.39
ρ_{34}	0.55	0.56	0.59	0.62	0.63	0.59	0.59	0.57
$\rho_{13\|2}$	0.01	0.02	0.07	0.12	0.13	0.07	0.07	0.09
$\rho_{24\|3}$	−0.01	0.01	0.06	0.10	0.11	0.05	0.05	0.04
$\rho_{14\|23}$	−0.03	−0.03	0.03	0.08	0.09	0.03	0.03	0.04

(3)

	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	0.41	0.41	0.44	0.47	0.48	0.44	0.44	0.44
γ_1	0.68	0.68	0.71	0.74	0.75	0.71	0.71	0.71
σ_2^2	0.47	0.48	0.51	0.55	0.56	0.51	0.51	0.52
γ_2	0.63	0.63	0.66	0.69	0.70	0.66	0.66	0.70
σ_3^2	0.45	0.46	0.49	0.53	0.54	0.49	0.49	0.49
γ_3	0.52	0.52	0.56	0.59	0.59	0.56	0.56	0.63
σ_4^2	0.47	0.47	0.51	0.55	0.55	0.51	0.51	0.50
γ_4	0.48	0.48	0.52	0.55	0.56	0.52	0.52	0.63

Estimated posterior mean, mode and quantiles of the joint MCMC as well as marginal MLE and C-MLE (TSP-MLE) for marginal AR(1) and D-Vine of normal copulas (approximated by t-copula with df=100)

	2.5%	5%	50%	95%	97.5%	mean	mode	C-MLE
ρ_{12}	0.30	0.31	0.35	0.39	0.40	0.35	0.35	0.35
ρ_{23}	0.36	0.37	0.41	0.45	0.46	0.41	0.41	0.39
ρ_{34}	0.55	0.55	0.59	0.62	0.63	0.59	0.59	0.57

(4)

	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	0.40	0.41	0.44	0.47	0.48	0.44	0.44	0.44
γ_1	0.67	0.68	0.71	0.74	0.75	0.71	0.72	0.71
σ_2^2	0.48	0.48	0.51	0.55	0.56	0.51	0.51	0.52
γ_2	0.62	0.63	0.66	0.69	0.70	0.66	0.66	0.70
σ_3^2	0.45	0.46	0.49	0.53	0.54	0.49	0.49	0.49
γ_3	0.51	0.52	0.55	0.58	0.59	0.55	0.55	0.63
σ_4^2	0.47	0.48	0.51	0.55	0.56	0.51	0.50	0.50
γ_4	0.49	0.49	0.52	0.55	0.56	0.52	0.52	0.63

Estimated posterior mean, mode and quantiles of the joint MCMC as well as marginal MLE and C-MLE (TSP-MLE) of the reduced model of 4: marginal AR(1), unconditional Copulas as gauss copulas (approximated by t-copula df=100), conditional copulas with correlation 0

	2.5%	5%	50%	95%	97.5%	mean	mode	C-MLE
ν	8.17	8.42	10.25	12.64	13.35	10.39	10.03	9.59
ρ_{12}	0.28	0.29	0.33	0.38	0.39	0.33	0.33	0.30
ρ_{23}	0.34	0.35	0.40	0.45	0.45	0.40	0.39	0.34
ρ_{34}	0.54	0.55	0.58	0.62	0.62	0.58	0.59	0.58
$\rho_{13\|2}$	−0.01	0.00	0.05	0.10	0.12	0.05	0.05	0.01
$\rho_{24\|3}$	−0.02	−0.01	0.05	0.10	0.11	0.05	0.05	0.02
$\rho_{14\|23}$	−0.05	−0.03	0.02	0.07	0.08	0.02	0.02	0.03
	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	0.41	0.41	0.44	0.48	0.49	0.45	0.44	0.44
γ_1	0.67	0.67	0.71	0.74	0.74	0.70	0.71	0.71
σ_2^2	0.46	0.47	0.50	0.54	0.55	0.50	0.50	0.52
γ_2	0.62	0.63	0.66	0.69	0.70	0.66	0.66	0.70
σ_3^2	0.45	0.45	0.49	0.52	0.53	0.49	0.49	0.49
γ_3	0.53	0.53	0.57	0.60	0.60	0.57	0.57	0.63
σ_4^2	0.47	0.47	0.51	0.55	0.55	0.51	0.51	0.50
γ_4	0.49	0.50	0.54	0.57	0.58	0.54	0.54	0.63

(5) Estimated posterior mean, mode and quantiles of the joint MCMC as well as marginal MLE and C-MLE (TSP-MLE) for marginal AR(1) and four-dimensional t-Copula

	2.5%	5%	50%	95%	97.5%	mean	mode	C-MLE
ν	5.80	6.08	7.44	9.30	9.79	7.54	7.32	9.59
ρ_{12}	0.27	0.28	0.33	0.38	0.39	0.33	0.33	0.30
ρ_{23}	0.34	0.35	0.40	0.45	0.46	0.40	0.40	0.34
ρ_{34}	0.54	0.55	0.59	0.62	0.63	0.59	0.59	0.58
	2.5%	5%	50%	95%	97.5%	mean	mode	marg. MLE
σ_1^2	0.40	0.41	0.44	0.48	0.48	0.44	0.44	0.44
γ_1	0.67	0.67	0.71	0.74	0.74	0.71	0.71	0.71
σ_2^2	0.46	0.47	0.50	0.54	0.54	0.50	0.50	0.52
γ_2	0.62	0.63	0.66	0.69	0.70	0.66	0.66	0.70
σ_3^2	0.45	0.46	0.49	0.52	0.53	0.49	0.49	0.49
γ_3	0.52	0.53	0.56	0.59	0.60	0.56	0.56	0.63
σ_4^2	0.47	0.48	0.51	0.55	0.56	0.51	0.51	0.50
γ_4	0.50	0.50	0.54	0.57	0.57	0.54	0.54	0.63

(6) Estimated posterior mean, mode and quantiles of the joint MCMC as well as marginal MLE and C-MLE (TSP-MLE) for marginal AR(1) and four-dimensional t-Copula with the conditional correlation parameters fixed to 0

Now we can calculate the posterior model probabilities calculated with the method of Congdon as described above. This yields the following results:

Nr.	Description of the Estimation Method	$\hat{T}_k$
1	Joint Bayesian Estimation with marginal AR(1) and D-Vine of t-Copulas (JMCMC)	$4.2689 \cdot 10^{-06}$
2	Joint Bayesian Estimation of the reduced model of 1: marginal AR(1), unconditional copulas as t-copulas, conditional copulas with correlation 0 and degrees of freedom = 100	0.0351
3	Joint Bayesian Estimation with marginal AR(1) and D-Vine of normal copulas (approximated by t-copula with df=100)	$4.5999 \cdot 10^{-13}$
4	Joint Bayesian Estimation of the reduced model of 4: marginal AR(1), unconditional Copulas as gauss copulas (approximated by t-copula df=100), conditional copulas with correlation 0	$3.2536 \cdot 10^{-14}$
5	marginal AR(1) and four-dimensional t-Copula	0.3245
6	marginal AR(1) and four-dimensional t-Copula with the conditional correlation parameters fixed to 0	0.6404

These results indicate clearly that the model with a marginal AR(1) structure and a four-dimensional t-copula where the conditional correlation parameters are fixed to 0 gives the best fit to the observed data. This is quite in contrast to the results of the internal validation of the different estimation methods where the JMCMC showed the lowest rejection rate.

Chapter 6

Conclusion

We have seen in the simulation studies in the previous chapters that the joint Bayesian estimation of marginal AR(1) time series parameters and D-vine t-copula parameters works very well in two dimensions as well as in four dimensions, especially for the time series and correlation parameters. We used a Metropolis-Hastings algorithm with independence proposal density. Therefore, we had to optimize the posterior density (up to constants) to find the mode and the Hessian at the mode. This took us most of the calculation time so that the operation of this algorithm required a long calculation period. Additionally, the ML estimators are very close to these estimated values so that you cannot recognize a significant advantage of the MCMC method with regard to the preciseness of the estimates. Only when we have estimated real Australian load data in the four-dimensional case, the autoregressive dependency parameters differed substantially. The MLE values even lie outside the 95 % credible interval.

Furthermore, we performed a comparison study between the joint Bayesian estimation of all parameters and a two-step approach. We hereby estimated the time series parameters first and transformed then the observed time series with these estimates and the cdf of the standard normal distribution to the unit cube to estimate the copula parameters. Our study shows that both estimation methods are very close to each other. To get an even better comparison, it is desirable that the marginal time series parameters are also estimated in a Bayesian setting. Perhaps, you could use the posterior estimates of the time series parameters to calculate a prior distribution for the copula parameters. There is still further research to do.

We also performed an internal validation of different models. At first, we encountered big difficulties in comparing the different models. How can you compare a vector autoregressive model to a four-dimensional t-copula with AR(1) margins to a D-vine of bivariate t-copulas with AR(1) margins. How can we give consideration to the sequence of the observations since this is an important point for time series models. Finally, we decided to use a multivariate two-sample test on the transformed data on the unit cube to assess the fit of the model. In this area, further proceedings are possible. A test for comparing these different models with considering the marginal situation is very desirable.

Finally, we conducted a Bayesian Model choice between 6 different models. Hereby, we used the method of Congdon (2006). It yields the result that the reduced model of a four-dimensional t-copula with all conditional correlation parameters fixed to zero has the best fit of all considered methods. This result is quite in contrast to the prior ones. Since we do not only estimate the dependency parameters, but also the marginal time series parameters, the power of this method of Congdon is not assured. There remains a lot of work to do with regard to a good model selection criteria for frequentist and Bayesian models.

Our implemented algorithm only deals with marginal AR(1) models. It would be very desirable to extend this work to general ARMA and/or GARCH models to be able to model a wide variety of different data.

Bibliography

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2007). Pair-Copula Constructions of Multiple Dependence. *Insurance: Mathematics and Economics*. doi: 10.1016/j.insmatheco.2007.02.001.

Angus, J. E. (1994, December). The Probability Integral Transform and Related Results. *SIAM Review 36*(4), 652–654.

Apostol, T. M. (1974). *Mathematical Analysis* (2nd ed.). Reading, MA: Addison-Wessley.

Ausin, M. C. and H. F. Lopes (2008, July). Time-varying joint distribution through copulas. Presentationat Conference 'Forecasting in Rio', `http://epge.fgv.br/finrio/myreview/FILES/p18.pdf`.

Baba, K., R. Shibata, and M. Sibuya (2004). Partial Correlation and Conditional Correlation as Measures of Conditional Independence. *Australian and New Zealand Journal of Statistics 46*(4), 657–664.

Banerjee, A., J. J. Dolado, J. W. Galbraith, and D. F. Hendry (1993). *Conintegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.

Baringhaus, L. and C. Franz (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis 88*, 190 – 206.

Barnett, G., R. Kohn, and S. Sheather (1996). Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics 74*, 237 – 254.

Bauwens, L. and M. Lubrano (1998). Bayesian Inference on GARCH models using the Gibbs sampler. *Econometrics Journal 1*(1), 23 – 46.

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

Bedford, T. and R. M. Cooke (2002). Vines - a New Graphical Model for Dependent Random Variables. *The Annals of Statistics 30*(4), 1031–1068.

Bickel, P. J. and K. A. Doksum (2001). *Mathematical Statistics: Basic Ideas and Selected Topics* (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Box, G. E. P. and G. M. Jenkins (1976). *Time series: Analysis, forecasting and control*. Holden Day.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). Springer Series in Statistics. New York: Springer.

Congdon, P. (2003). *Applied Bayesian Modelling*. Wiley & Sons.

Congdon, P. (2006). Bayesian model choice based on Monte Carlo estimates of posterior model probabilities. *Computational Statistics & Data Analysis 50*, 346 – 357.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Czado, C. and A. Min (2007, November 19). Bayesian Inference for Pair-Copula Construtions based on Multiple Dependence. Presentation Vine Copula Workshop TU Delft, `http://dutiosc.twi.tudelft.nl/~risk/Work2007files/czado-min.pdf`.

Dalla Valle, L. (2007). Bayesian Copulae Distributions, with Applications to Operational Risk Management. *Methodology and Computing in Applied Probability*. doi: 10.1007/s11009-007-9067-x.

DeGroot, M. H. (2003). *Probability and Statistics*. Wiley & Sons.

Demarta, S. and A. J. McNeil (2005). The t Copula and Related Copulas. *International Statistical Review 73*(1), 111 – 129.

Embrechts, P., F. Lindskog, and A. J. McNeil (2003). Modelling Dependence with Copulas and Applications to Risk Management. In S. T. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, pp. 329–348. New York: Elsevier.

Embrechts, P., A. J. McNeil, and D. Straumann (2002). Correlation and Dependency in Risk Management: Properties and Pitfalls. In M. A. H. Dempster and H. K. Moffat (Eds.), *Risk Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

Fang, K.-T., S. Kotz, and K. W. Ng (1990). *Symmetric Multivariate and Related Distributions*. Chapman Hall.

Fisher, N. I. (1997). Copulas. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences*, Volume Update Vol. 1, pp. 159 – 163. Wiley & Sons.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). CRC Press.

Gatz, J. H. (2007). Properties and Applications of the Student T Copula. Master's thesis, TU Delft. `http://dutiosc.twi.tudelft.nl/~risk/index.php?option=com\_docman\&task=doc\_download\&gid=153\&\&Itemid=13`.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996). Introducing Markov Chain Monte Carlo. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman Hall.

Gschlößl, S. (2006). *Hierarchical Bayesian Spatial Regression Models with Applications to Non-Life Insurance*. Ph. D. thesis, Mathematische Fakultät, Technische Universität München.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika 57*(1), 97 – 109.

Huard, D., G. Evin, and A.-C. Favre (2006). Bayesian Copula Selection. *Computational Statistics & Data Analysis 51*, 809 – 822.

Joe, H. (1996). Families of m-variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with Fixed Marginals and Related Topics*, Volume 28 of *Lecture Notes - Monograph Series*, pp. 120 – 141. Hayward, CA: Institute of Mathematical Statistics.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman Hall.

Joe, H. (2006). Generating Random Correlation Matrices Based on Partial Correlations. *Journal of Multivariate Analysis 97*, 2177–2189.

Kanzow, C. (2005). *Numerik linearer Gleichungssysteme: Direkte und iterative Verfahren*. Berlin: Springer.

Kendall, M. G. (1945). *The Advanced Theory of Statistics* (2nd ed.), Volume 1. Charles Griffin & Company.

Kibria, B. M. G. and A. H. Joarder (2006). On the Multivariate T-Distribution and Some of its Applications. Technical Report 359, King Fahd University of Petroleum & Minerals, Department of Mathematical Sciences, Saudi Arabia. `http://www.kfupm.edu.sa/math/TechReports\_Data/359.pdf`.

Koop, G. (2005). *Bayesian Econometrics*. Wiley & Sons.

Kurowicka, D. and R. M. Cooke (2004). Distribution-free continuous bayesian belief nets. In *Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice*, Santa Fe, New Mexico.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley & Sons.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics 54*, 159 – 178.

Lewandowski, D., D. Kurowicka, and H. Joe (2007). Generating Random Correlation Matrices Based on Vines and Extended Onion Method. *Journal of Multivariate Analysis*, submitted.

Lindskog, F., A. J. McNeil, and U. Schmock (2003). Kendall's Tau for Elliptical Distributions. In G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk - Measurement, Evaluation and Management*, Proceedings of the 8th Econometric Workshop, held in Karlsruhe, Germany from March 13-15, 2002, Heidelberg. Physica-Verlag.

Marriott, J. and A. F. M. Smith (1992). Reparametrization aspects of numerical Bayesian methods for autoregressive moving average models. *Journal of Time Series Analysis 13*, 327 – 343.

Metropolis, N. C., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics 21*(6), 1087 – 1092.

Monahan, J. (1983). Fully Bayesian analyis of ARMA time series models. *Journal of Econometrics 21*, 307 – 331.

Nelsen, R. B. (2006). *An Introduction to Copulas* (2nd ed.). Springer Series in Statistics. Springer.

Pitt, M., D. Chan, and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika 93*(3), 537 – 554.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics 23*, 470–472.

Silva, R. d. S. and H. F. Lopes (2008). Copula, marginal distributions and model selection: a Bayesian note. *Statistics and Computing 18*(3), 313 – 320.

Sklar, A. (1959). Fonctions de répartitions à n dimensions et leur marges. *Publications de l'Institut de Statistique de l'Université de Paris 8*, 229 – 231.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 22*(4), 1701 – 1762.

Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices*. Chichester, GB: Wiley & Sons.

Wise, J. (1955). The Autocorrelation Function and the Spectral Density Function. *Biometrika 42*(1/2), 151 – 159.

Yule, G. U. and M. G. Kendall (1965). *An Introduction to the Theory of Statistics*. Charles Griffin & Company.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. Wiley & Sons.