Technische Universität München
Lehrstuhl für Mathematische Statistik

Regular vine copulas with the simplifying assumption, time-variation, and mixed discrete and continuous margins

Jakob Michael Stöber

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Rudi Zagst
Prüfer der Dissertation:	
1.	Univ.-Prof. Claudia Czado, Ph.D.
2.	Prof. Roger M. Cooke, Ph.D. TU Delft, Niederlande
3.	Prof. Andrew J. Patton, Ph.D. Duke University, USA (nur schriftliche Beurteilung)

Die Dissertation wurde am 11.03.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 12.06.2013 angenommen.

Zusammenfassung

Die vorliegende Arbeit ergänzt die Theorie zu Pair-Copula-Konstruktionen (PCCs) basierend auf regulären Vines (R-vines) in einigen Aspekten. PCCs für Modelle mit sowohl diskreten als auch kontinuierlichen eindimensionalen Randverteilungen werden entwickelt, und es wird gezeigt wie Likelihood-Funktion, Score-Funktion und die beobachtete Fisher-Informationsmatrix berechnet werden können. Damit PCCs greifbar für statistische Schlussfolgerungen und Modellwahl bleiben, trifft man üblicherweise eine vereinfachende Annahme für die Copulas der bedingten Verteilungen: Es wird angenommen, dass diese Copulas nicht von den Werten der Variablen, auf die bedingt wird, abhängen. In dieser Arbeit werden alle Archimedischen Copulas vom vereinfachten Typ identifiziert, gezeigt, dass die student'sche t-Verteilung die einzige Scale Mixture von Normalverteilungen ist für welche die Annahme gilt, und eine Technik demonstriert um den benötigten Stichprobenumfang zur Unterscheidung einer multivariaten Verteilung von einer nahen vereinfachten PCC zu bestimmen. Da einige Anwendungen auf Finanzzeitreihen die Modellierung von zeitlichen Veränderungen in Abhängigkeiten notwendig machen, wird weiter ein Markov-Switching Modell entwickelt, welches es erlaubt solche Veränderungen zu studieren. Dieses Modell wird auf einen Datensatz mit US-Wechselkursen angewendet, für den Stressphasen an den Märkten, welche mit Veränderungen in Abhängig-keiten einhergehen, bestimmt werden. Um die Vorteile von PCCs für Daten mit sowohl diskreten als auch kontinuierlichen Zufallsvariablen aufzuzeigen, wird eine weitere Anwendung auf Daten von der "Second Longitudinal Study of Aging" (LSOA II) betrachtet. In dieser wird die Prävalenz von chronischen Krankenheiten untersucht, um das Verständnis von Komorbiditäten zu erweitern.

Abstract

This thesis extends the theory of regular vine (R-vine) pair copula constructions (PCCs) in several aspects. We develop PCCs for models with both discrete and continuous univariate marginal distributions and illustrate how the likelihood function, score function, and observed information matrix can be computed. To keep PCCs tractable for inference and model selection one usually makes a simplifying assumption for the copulas of conditional distributions: we assume that these copulas do not depend on the values of the variables which are conditioned on. We find all Archimedean copulas which are of the simplified type, show that Student's t distribution is the only scale mixture of Normals for which the assumption holds and demonstrate a technique to assess the sample size required to distinguish a multivariate copula from a nearby simplified PCC. Since some applications in finance require to model variations in dependence over time we develop a Markov switching R-vine model which allows to study these changes. This model is applied to a data set of US exchange rates, for which we determine periods of market stress associated with changes in dependence. To illustrate the advantages of PCCs for data with both discrete and continuous variables we consider a further application to data from the Second Longitudinal Study of Aging (LSOA II) where we analyze the prevalence of chronic conditions to broaden our understanding of comorbidities.

Acknowledgements

First of all, I am greatly indepted to my advisor Claudia Czado, who guided me through the years of writing this thesis. I want to thank her for providing intensive supervision and always being available for fruitful discussions. Likewise, I am very grateful for her encouragement to attend international conferences and I want to thank her, Prof. Robert Wolpert and the Department of Statistical Science at Duke University for giving me the opportunity to spend an inspiring semester at Duke.

It is a particular pleasure for me to thank Ulf Schepsmeier, Harry Joe, Hyokyoung Hong and Pulak Ghosh for the fruitful collaborations and inspiring exchange of ideas. Also I would like to thank Prof. Andrew Patton and Prof. Roger Cooke for acting as referees of this thesis. I want to thank all my colleagues at Technische Universität München: Without you the three years I spent writing this thesis would have been a much less pleasurable experience. Financial support through a research stipend provided by Allianz Deutschland AG and through the Elite Network of Bavaria and the TopMath Graduate Center of TUM Graduate School at Technische Universität München is gratefully acknowledged. I am grateful to Denise Lichtig and Dr. Carl-Friedrich Kreiner who are doing great work coordinating the TopMath program.

Contents

Introduction **1**

1 Preliminaries **5**

1.1 Notation . . . 5

1.2 Copulas and multivariate dependence . . . 6

1.2.1 Measures of dependence . . . 7

1.2.2 Inference . . . 9

1.3 Scoring rules and model selection criteria . . . 10

1.3.1 Log predictive score . . . 10

1.3.2 AIC/BIC . . . 11

1.4 Graph theory: regular vine tree sequence . . . 11

1.5 Decomposing a three dimensional distribution . . . 15

1.5.1 Continuous case . . . 16

1.5.2 Mixed discrete and continuous case . . . 17

1.5.3 Margins with discrete and continuous components . . . 18

1.5.4 Corresponding R-vine . . . 19

1.6 Software . . . 19

2 Pair copula constructions and regular vines **21**

2.1 PCCs for discrete and continuous margins . . . 23

2.2 Likelihood, score function and observed information . . . 28

2.2.1 Computation of the likelihood function . . . 29

2.2.2 Computation of the score function . . . 32

2.2.3 Computation of the observed information . . . 37

3 The simplifying assumption **39**

3.1 Archimedean copulas . . . 39

3.2 Elliptical copulas . . . 44

3.3 Effects of the simplifying assumption . . . 52

3.3.1 Trivariate extension of the Farlie-Gumbel-Morgenstern copula . . . 52

3.3.2 Multivariate 1-factor model . . . 54

3.3.3 Non-simplified PCCs . . . 58

4 Regime switching models **63**
4.1 Regime switching copulas . . . 63
4.1.1 Markov switching copula models . . . 64
4.1.2 Inference for Markov switching models . . . 65
4.1.3 Simulation study . . . 69
4.1.4 R-vine copula model selection . . . 71
4.2 Regime switching copulas and marginal distributions . . . 72
4.2.1 Model setup . . . 73
4.2.2 Inference and EM algorithm for the joint model . . . 74
4.2.3 Simulation study . . . 75

5 Application 1: Multivariate regime switching model of US exchange rates **81**
5.1 Markov Switching US exchange rate dependence . . . 82
5.1.1 Data description . . . 83
5.1.2 R-Vine copula with switching parameters . . . 83
5.1.3 Identifying crisis regimes . . . 86
5.1.4 Empirical findings . . . 87
5.1.5 Model comparison . . . 88
5.2 Regime switching marginal time series . . . 91

6 Application 2: Comorbidity in the second longitudinal study of aging (LSOA II) **95**
6.1 Introduction and data description . . . 95
6.2 Multivariate model . . . 97
6.2.1 R-vine copula selection . . . 98
6.2.2 Selected joint model . . . 100
6.3 Results . . . 101
6.4 Conclusions . . . 106

A Calculation of the observed information in regular vine copula models **109**

B Additional results from the simulation study in Section 4.1.3 **115**

C Conditional copula of the Student's t distribution **117**

D Application 1: additional material **119**
D.1 Exchange rate data: selected R-vines . . . 119

D.1.1 Model (1) 119
D.1.2 Models (2) 121
D.1.3 Model (3) 123
D.1.4 Model (4) 124
D.2 Regime switching marginal distributions 126

E Application 2 (LSOA II model): parameter estimates 133

Introduction

While many problems in mathematics are hundreds of years old (Boyer and Merzbach 2011) most statistical methods being used by practitioners of various fields today have been developed rather recently. With the upcoming of ever more powerful computers and information technology, the amount of data being collected has increased drastically, pushing demand for new statistical methods. Being not limited to methods where results can be obtained through calculations with pen and paper, but using sophisticated numerical optimization and Monte Carlo techniques, allows for ever more complex and comprehensive models to be applied to real world problems.

With more data becoming accessible, one area which has received considerable attention in the last years is the modeling of multivariate dependencies between observations or random variables. The mathematical "language" which has been developed to describe dependence structures is the theory of copulas. Despite attracting criticism both from a theoretical point of view (Mikosch 2006) and for its (inappropriate) use in some areas of finance (Salmon 2009, "The formula that killed Wall Street") the concept of copulas is still on the rise. While computational capabilities allow for more sophisticated applications, the main mathematical merit of the copula concept is that it allows to develop a systematic understanding of dependence beyond linear correlations (Embrechts et al. 2002). Today, many classes of copulas are available, each having its strengths in specific applications. A comprehensive overview in the context of economic forecasting has been provided by Patton (2012).

In this thesis, we will focus on a particular class of copulas, so-called regular vine (R-vine) pair copula constructions (PCCs), which are popular because they break the complex problem of specifying a d-dimensional dependence model down into choosing a decomposition and then specifying bivariate copulas. Under some assumptions, this class provides a highly flexible and tractable model in moderate dimensions (typical applications are in dimensions 3 to 50, Heinen and Valdesogo (2009) propose a vine-based model in dimension $d = 100$). We will extend the existing theory of PCCs in several aspects. While previous research focused either on the "standard" case of purely continuous random variables (Aas et al. 2009) or on multivariate discrete models (Panagiotelis et al. 2012), we present a systematic approach to PCCs where some margins are discrete and some are continuous. We will show how the likelihood, score function and observed information can be computed for R-vine PCCs and answer theoretical questions regarding the so-called "simplifying assumption". In particular, we will investigate how restrictive it is to assume that the copulas of conditional

distributions do not depend on the values of variables that are conditioned on, and which well-known multivariate models are simplified PCCs of this kind. Since some problems in finance require the modeling of dependence structures varying over time, we will then study a Markov switching (MS) R-vine copula model in which the dependence structure depends on a latent state variable. We will apply this model to a data set of US exchange rates where we identify crisis periods. To illustrate the capabilities of the R-vine PCC model for mixed discrete and continuous variables, we will study a data set from the Second Longitudinal Study of Aging (LSOA II). The general outline and contribution of each chapter is as follows.

Chapter 1 presents mathematical notation and concepts which will be required throughout this thesis. In particular, the concepts of copulas and multivariate measures of dependence are introduced, and we provide the necessary graph theoretic background for the R-vine tree structures which will be used. To motivate the concept of PCCs, a three dimensional example is discussed in detail.

Chapter 2 introduces R-vine PCCs for discrete and continuous marginal distributions and is based on material from Stöber et al. (2012). This copula model, which generalizes those of Panagiotelis et al. (2012) for discrete data and Aas et al. (2009) for continuous data has two important advantages over existing models. First, R-vine PCCs result in highly flexible dependence structures since different and (possibly) asymmetric bivariate copulas can be combined in one distribution. This allows to model asymmetries and in particular asymmetric tail dependencies (Joe et al. 2010) which are beyond the scope of e.g. the symmetric multivariate normal and Student's t distributions. Second, in the presence of discrete marginal distributions, the approach presented here has significant computational advantages since it requires only the evaluation of *bivariate* instead of higher dimensional copula functions. Furthermore, the number of required evaluations of copula functions to calculate the probability mass function (pmf) grows only *quadratically* with the number of discrete variables. This is a significant improvement compared to calculating the pmf for a given high dimensional copula by taking differences where the number of function evaluations increases *exponentially* (Panagiotelis et al. 2012). This makes parameter inference in a maximum likelihood framework feasible where competing models have to be fitted in a computationally more intense Markov chain Monte Carlo (MCMC) framework. We will show how the likelihood computations for R-vine PCCs can be implemented in computer code and develop algorithms to evaluate also the score function and observed information matrix. For the purely continuous setup, the algorithms have been made available in the R-package `VineCopula` (Schepsmeier

et al. 2012).

In **Chapter 3**, which is taken from Stöber et al. (2013), we proceed by reviewing the simplifying assumption for PCCs, i.e. that the copulas of conditional distributions do not depend on the values of variables that are conditioned on. This assumption has also been studied by Hobæk Haff et al. (2010) and Acar et al. (2012a). Our main results are to show that the only multivariate Archimedean copulas of the simplified type are those in the MTCJ copula family and that the only multivariate scale mixture of Normals fulfilling the simplifying assumption is the Student's t distribution. To investigate how severe violations of the simplifying assumption are in practice, we introduce a technique based on the Kullback-Leibler divergence which allows to assess the necessary sample size to distinguish a multivariate distribution from a nearby distribution which satisfies the simplifying assumption.

Chapter 4, which is based on material from Stöber and Czado (2013), considers a general Markov switching model based on R-vine copulas (MS-RV model) for which we develop efficient inference techniques. Using the full flexibility of R-vine PCCs, we go beyond the copula model considered by Chollete et al. (2009). For fast parameter estimation and scalability to high dimensions, we develop an approximative Expectation - Maximization (EM) algorithm based on the sequential estimation technique of Aas et al. (2009) and Hobæk Haff (2013). We do also consider inference in a Bayesian setup, where we develop an MCMC algorithm which also allows to compute credible intervals for the probability of being in a certain regime at a given point of time. This allows us to quantify the uncertainty in the time-variability of dependence for given data which is not possible for most existing models. Since in many contexts we will not have prior knowledge about the data on which the selection of appropriate R-vine structures and corresponding pair-copulas can be based we will also introduce a simple model selection heuristic based on the work of Dißmann et al. (2013) for time-constant R-vine copulas. While most of this thesis is concerned with modeling only the dependence structure or the copula, assuming that the marginal distributions are known or pre-specified, we will also consider a model where both the marginal time series and the dependence structure are subject to changes in regime in this section. We will extend the EM procedure to estimate parameters for this joint model.

In Chapters 5 and 6 we apply the developed models to two real-world data sets. In **Chapter 5**, which is based again on Stöber and Czado (2013), we consider an application of the MS models developed in Chapter 4 to a data set of US exchange rates. After the 2007/2008 financial crisis, regulators worldwide have recognized the fact that financial time series exhibit different behavior in situations of market stress and introduced new requirements for risk models of financial institutions addressing this issue. In addition to Value at Risk (VaR),

European banks are now required to report a stressed VaR (European Banking Authority 2012, SVaR), for which the risk model has to be calibrated using a period of significant stress to the banks portfolio. To detect appropriate "crisis" periods, we will present an extensive investigation of MS models for the dependence structure among the exchange rates in our data set. Since most practitioners will not only be interested in the dependence structure, we will further apply the joint model in which both the marginal time series and the copula are switching jointly. Using out-of-sample predictive logscores, we illustrate that the considered MS models produce more accurate forecasts than constant models during times of market stress.

Chapter 6 is taken from Stöber et al. (2012). Here, we analyze data from the second longitudinal study of aging (LSOA II), containing information on the absence/presence (discrete outcome) of five chronic conditions (arthritis, hypertension, heart disease, stroke and diabetes) and the body mass index (BMI, continuous outcome), requiring both discrete and continuous marginal distributions. As the population is aging in the US and most of the developed world, improved conditions and medical methods have contributed to substantially higher life expectancies. Nevertheless, as a consequence of improved survival rates for previously fatal diseases, also the proportion of adults afflicted with chronic conditions has increased substantially. The chronic conditions on which we have information in our data set are often studied in an isolated setting and clinical practice guidelines are based on such studies. However, the elderly are likely to develop comorbid conditions, i.e. suffer from more than one chronic condition at the same time. Here, we consider a joint modeling framework to improve our understanding of comorbidities.

Chapter 1

Preliminaries

In this chapter, we will state necessary preliminaries for what we will consider in the remainder of this thesis.

1.1 Notation

A d-dimensional random vector is denoted by $\mathbf{X}_{1:d} = (X_1, \ldots, X_d)$, for observations from this distribution we write $\mathbf{x}_{1:d} = (x_1, \ldots, x_d)$. We do further write $\mathbf{X}_I = \{X_i | i \in I\}$ for a set I of indices, in particular, $\mathbf{X}_{(1:d)\setminus h}$, for $h \in 1, \ldots, d$, refers to the vector where the variable with index h has been removed from $\mathbf{X}_{1:d}$. Here, $1 : d = \{1, \ldots, d\}$.
If nothing else is mentioned, we shall denote the corresponding cumulative distribution function (cdf) by $F_{1:d}(\cdot)$ and the density, which we will usually assume to exist, by $f_{1:d}(\cdot)$. The conditional distribution function of $\mathbf{X}_I$ given $\mathbf{X}_J = \mathbf{x}_J$, where I, J are non-overlapping subsets of $\{1, \ldots, d\}$, is $F_{I|J}(\cdot|\mathbf{x}_J)$ with corresponding density $f_{I|J}(\cdot|\mathbf{x}_J)$. If I has more than one element, then the copula corresponding to this conditional distribution is $C_{I;J}(\cdot|\mathbf{x}_J)$, with density $c_{I;J}(\cdot|\mathbf{x}_J)$.
For a discrete variable X_j, $j \in 1, \ldots, d$, we do also write $F_j(x_{j,0}) := F_j(x_j)$, and $F_j(x_{j,1})$ for the left-hand limit of F_j at x_j to simplify notation. In the case of $X_j \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the integers, this corresponds to $F_j(x_{j,1}) = F_j(x_j - 1)$. The corresponding probability mass function (pmf) is denoted by p_j.
For a differentiable function $f : \mathbb{R}^d \mapsto \mathbb{R}$, ($\mathbb{R}$ is the field of real numbers, the field of rational numbers is denoted by $\mathbb{Q}$) we denote the i-th partial derivative by $\partial_i f$, i.e.

$$\partial_i f(x_1, \ldots, x_i, \ldots, x_d) = \frac{\partial f(x_1, \ldots, z_i, \ldots, x_d)}{\partial z_i}\Big|_{z_i = x_i}.$$

For $g : \mathbb{R} \mapsto \mathbb{R}$, we will write g' for the first derivative and $g^{(k)}$ for the kth derivative.
For a set S, $\mathcal{P}(S)$ is the power set (i.e. the set of all subsets) and $\#S$ is the number of elements s in S. We denote the indicator function on the set by $\mathbb{1}_S$, it is defined by

$$\mathbb{1}_S(s) = \begin{cases} 1 & s \in S \\ 0 & s \notin S \end{cases}.$$

Closely related is the signum function $sgn(\cdot)$ for a real number x, which is defined as

$$sgn(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}.$$

To express logical statements we will use $\wedge$ (and), $\forall$ (for all) and $\exists$ (exists).
For a sample $(x_{i,t})_{i=1,\dots,d;t=1,\dots,T}$, the empirical distribution functions are given by $\hat{F}_i(y) = \sum_{t=1}^{T} \mathbb{1}_{\{z|z\leq y\}}(x_{i,t})/T$, and the standardized ranks are $\hat{u}_{i,t} = T\hat{F}_i(x_{i,t})/(T+1)$.
Finally, $Ran(F)$ denotes the range of a function F, i.e. for $F : A \mapsto B$, $Ran(F) = \{F(a)|a \in A\}$. If F is invertible, then the inverse is F^{-1}. If F is an increasing but not invertible function on the real line then $F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}$, $y \in \mathbb{R}$ is its generalized inverse.

1.2 Copulas and multivariate dependence

The central mathematical objects in this thesis are copulas.

Definition 1.2.1. *A d-dimensional* ***copula*** *is a cumulative distribution function in d-dimensions such that the univariate marginal distributions are uniform on* $[0; 1]$.

Let $F_{1:d}$ be the cdf of a d-dimensional probability distribution and denote the marginal distribution functions by $F_1, \dots, F_d$. Then the following theorem allows to specify $F_{1:d}$ in terms of the univariate margins and a copula $C_{1:d}$.

Theorem 1.2.2 (Sklar (1959))**.** *For every distribution function $F_{1:d}$ with univariate marginal distributions $F_1, \dots, F_d$ there exists a copula $C_{1:d}$, such that*

$$F_{1:d}(x_1, \dots, x_d) = C_{1:d}(F_1(x_1), \dots, F_d(x_d)). \tag{1.1}$$

If $F_1, \dots, F_d$ are continuous, then $C_{1:d}$ is unique. Otherwise, C is uniquely determined on $Ran(F_i) \times \cdots \times Ran(F_d)$.

If the distribution $F_{1:d}$ is absolutely continuous with respect to the standard Lebesgue measure on $\mathbb{R}^d$ and has density $f_{1:d}$ and marginal densities $f_1, \dots, f_d$, then (1.1) can be rewritten in terms of the densities as

$$f_{1:d}(x_1, \dots, x_d) = c_{1:d}(F_1(x_1), \dots, F_d(x_d)) \cdot f_1(x_1) \cdot \ldots \cdot f_d(x_d). \tag{1.2}$$

While Definition 1.2.1 and Theorem 1.2.2 are given for the (unconditional) distribution of a random vector $\mathbf{X}$, they can be extended to the conditional distribution of $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ (Patton

2006). In this case, Sklar's theorem guarantees the existence of a conditional copula $C_{\mathbf{X};\mathbf{Y}}(\cdot|\mathbf{y})$, depending on $\mathbf{y}$, such that the conditional distribution function of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ is given by

$$F_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = C_{\mathbf{X};\mathbf{Y}}\left(F_{1|\mathbf{Y}}(x_1|\mathbf{y}), \dots, F_{d|\mathbf{Y}}(x_d|\mathbf{y})\middle|\mathbf{y}\right). \tag{1.3}$$

Here, we write $C_{\mathbf{X};\mathbf{Y}}$ for the copula corresponding to a conditional distribution to clearly distinguish it from the conditional distributions of a multivariate copula. To further clarify the notation which we will use throughout this thesis, let us consider the special case of a bivariate conditional distribution of variables (X_1, X_2) given $\mathbf{X}_{3:d} = \mathbf{x}_{3:d}$. Here, the joint distribution function is

$$F_{1,2|3:d}(x_1, x_2|\mathbf{x}_{3:d}) = C_{1,2;3:d}\left(F_{1|3:d}(x_1|\mathbf{x}_{3:d}), F_{2|3:d}(x_2|\mathbf{x}_{3:d})\middle|\mathbf{x}_{3:d}\right), \tag{1.4}$$

with corresponding density function

$$\begin{aligned} f_{1,2|3:d}(x_1, x_2|\mathbf{x}_{3:d}) = c_{1,2;3:d}&\left(F_{1|3:d}(x_1|\mathbf{x}_{3:d}), F_{2|3:d}(x_2|\mathbf{x}_{3:d})\middle|\mathbf{x}_{3:d}\right) \\ &\cdot f_{1|3:d}(x_1|\mathbf{x}_{3:d}) \cdot f_{2|3:d}(x_2|\mathbf{x}_{3:d}). \end{aligned} \tag{1.5}$$

Similarly, the conditional density of X_1 given $X_2 = x_2$ and $\mathbf{X}_{3:d} = \mathbf{x}_{3:d}$ is given by

$$f_{1|2:d}(x_1|\mathbf{x}_{2:d}) = c_{1,2;3:d}\left(F_{1|3:d}(x_1|\mathbf{x}_{3:d}), F_{2|3:d}(x_2|\mathbf{x}_{3:d})\middle|\mathbf{x}_{3:d}\right) \cdot f_{1|3:d}(x_1|\mathbf{x}_{3:d}). \tag{1.6}$$

In particular in the special case of two dimensions, copulas have been studied by many authors and a variety of literature is available (see e.g. the books by Joe (1997) and Nelsen (2006)). An overview of many bivariate copula families and their derivatives, which will be important for algorithms developed later in this thesis, is given by Schepsmeier and Stöber (2012). The families we will consider in this thesis include the Gaussian/Normal, Student's t, Gumbel, Clayton/MTCJ and Frank family. Their parameterizations will be chosen as in Schepsmeier and Stöber (2012).

1.2.1 Measures of dependence

While it is necessary to specify the copula to fully determine the dependence structure of a random vector, we will often use dependence measures which do only depend on the copula, and not on the marginal distributions, to summarize observations. The two measures we will mostly use in this thesis are Kendall's τ rank correlation and the notion of tail dependence. However, many more multivariate measures of association are available and a comprehensive overview, from which also the following definitions are taken, is given in Joe (1997). We will always assume that the limits in the definitions exist and refer to the given literature for a discussion of when this applies.

Definition 1.2.3. *Let $F_{1,2}$ be a continuous bivariate distribution function and let (X_1, X_2), (X_1', X_2') be independent random vectors with distribution $F_{1,2}$. Then* ***Kendall's τ*** *is*

$$\tau = 4 \int_{\mathbb{R}^2} F_{1,2}(x_1, x_2) dF_{1,2}(x_1, x_2) - 1. \tag{1.7}$$

This is equivalent to the probability of observing two concordant pairs minus the probability of observing two discordant pairs

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

Expression (1.7) is invariant under strictly increasing transformations of the marginal variables. Hence, Kendall's τ can be written in terms of the corresponding copula $C_{1,2}$,

$$\tau = 4 \int_{[0;1]^2} C_{1,2}(u_1, u_2) dC_{1,2}(u_1, u_2) - 1.$$

For a sample of observations $\mathbf{x}_t = (x_{1,t}, x_{2,t})$, $t = 1, \ldots, T$, with positive weights ω_t, an estimator of Kendall's τ is

$$\hat{\tau}_{\omega_{1:T}}(\mathbf{x}_{1:T}) = \frac{T(T-1)/2}{\sqrt{\left(\frac{T(T-1)}{2} - n^1\right)\left(\frac{T(T-1)}{2} - n^2\right)}} \frac{1}{\sum_{i=1}^{T-1}\sum_{j=i+1}^{T} \omega_i \omega_j} \sum_{i=1}^{T-1}\sum_{j=i+1}^{T} d_{i,j}^1 d_{i,j}^2 \omega_i \omega_j,$$

where $d_{i,j}^k = sgn(x_{k,i} - x_{k,j})$, $k = 1, 2$, and n^k is the number of tied pairs with $d_{i,j}^k = 0$ (see Pozzi et al. (2012) for further references and computer code for this estimator). For equal weights $\omega_t = \omega_{t'}$, $\forall t, t'$, this is the usual estimator with adjustment for ties:

$$\hat{\tau}(\mathbf{x}_{1:T}) = \frac{1}{\sqrt{\left(\frac{T(T-1)}{2} - n^1\right)\left(\frac{T(T-1)}{2} - n^2\right)}} \sum_{i=1}^{T-1}\sum_{j=i+1}^{T} d_{i,j}^1 d_{i,j}^2.$$

Definition 1.2.4. *Let $\boldsymbol{X} = (X_1, X_2)$ be a random vector with marginal distributions F_1 and F_2.* ***X*** *has* ***upper tail dependence*** *if*

$$\lambda_U = \lim_{u \to 1^-} P(X_1 > F_1^{-1}(u) | X_2 > F_2^{-1}(u)) > 0,$$

and no upper tail dependence if $\lambda_U = 0$. Similarly, ***X*** *has* ***lower tail dependence*** *if*

$$\lambda_L = \lim_{u \to 0^+} P(X_1 \leq F_1^{-1}(u) | X_2 \leq F_2^{-1}(u)) > 0,$$

and no lower tail dependence if $\lambda_L = 0$.

Again, this definition is invariant under strictly increasing transformations and we have

$$\lambda_L = \lim_{u \to 0^+} \frac{C_{1,2}(u, u)}{u}, \qquad \lambda_U = \lim_{u \to 1^-} = \frac{1 - 2u + C_{1,2}(u, u)}{(1 - u)}.$$

For a more general multivariate notion of tail dependence we refer to Joe et al. (2010).

1.2.2 Inference

We will always consider the case where a d-variate distribution is specified in terms of d marginal densities $f_i(\cdot|\boldsymbol{\theta}_i)$ with parameters $\boldsymbol{\theta}_i$ and a copula density $c_{1:d}(\cdot|\boldsymbol{\theta}_{cop})$ with parameters $\boldsymbol{\theta}_{cop}$. Hence, for observations $(\mathbf{x}_t)_{t=1,\dots,T}$ of independent and identically distributed (i.i.d.) random variables with this distribution, the likelihood function ℓ is given by

$$\begin{aligned}&\ell(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_d,\boldsymbol{\theta}_{cop}|\mathbf{x}_{1:T})\\&\quad=\prod_{t=1}^{T}\left[c_{1:d}\left(F_1(x_{1,t}|\boldsymbol{\theta}_1),\dots,F_d(x_{d,t}|\boldsymbol{\theta}_d)\middle|\boldsymbol{\theta}_{cop}\right)\cdot f_1(x_{1,t}|\boldsymbol{\theta}_1)\cdot\ldots\cdot f_d(x_{d,t}|\boldsymbol{\theta}_d)\right],\end{aligned}$$

and the natural way to perform inference is to either apply the maximum likelihood principle or to specify a prior distribution on the parameters $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_d,\boldsymbol{\theta}_{cop}$ and then proceed under the Bayesian paradigm. Since this requires a joint maximization with respect to all parameters or the calculation of the joint posterior distribution, respectively, the necessary computations can become very demanding in high dimensions. To circumvent these challenges, other estimation methods have been developed in the literature - a comprehensive overview is given by Patton (2012). In the following, we quickly summarize two methods which separate the estimation of marginal parameters from the estimation of the copula parameters.

1.2.2.1 Inference functions for margins

If the marginal distributions and the copula do not share parameters, i.e. $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_d$ and $\boldsymbol{\theta}_{cop}$ can all be specified independently, Joe and Xu (1996) propose to perform estimation in two stages. On the first stage, the marginal dependence parameters $\boldsymbol{\theta}_i$ are estimated by maximizing the marginal likelihood functions

$$\ell_i(\boldsymbol{\theta}_i|\mathbf{x}_{1:T})=\prod_{t=1}^{T}f_i(x_{i,t}|\boldsymbol{\theta}_i),\qquad i=1,\dots,d.$$

Subsequently, $\boldsymbol{\theta}_{cop}$ is estimated by maximizing the pseudo likelihood function

$$\ell_{cop}(\boldsymbol{\theta}_{cop}|\mathbf{x}_{1:T},\hat{\boldsymbol{\theta}}_1,\dots,\hat{\boldsymbol{\theta}}_d)=\prod_{t=1}^{T}c_{1:d}\left(F_1(x_{1,t}|\hat{\boldsymbol{\theta}}_1),\dots,F_d(x_{d,t}|\hat{\boldsymbol{\theta}}_d)\middle|\boldsymbol{\theta}_{cop}\right),$$

where $\hat{\boldsymbol{\theta}}_1,\dots,\hat{\boldsymbol{\theta}}_d$ are the estimated parameters obtained on the first stage. Also if $\mathbf{x}_{1:T}$ are not observations of i.i.d. random variables but have a time series structure, this two step procedure can be applied analogously as long as the parameters of the marginal time series models and the dependence model are fully independent. While this multiple step estimation procedure is obviously asymptotically less efficient than full maximum likelihood estimation,

it is asymptotically consistent and allows to fully disentangle the dependence structure from the marginal distributions also for parameter estimation. Note that this inference method is known more generally also as multi-stage maximum likelihood estimation (see for example White (1996)).

1.2.2.2 Maximum pseudo likelihood estimation

If sufficient data is available and only the dependence structure is of interest, Genest et al. (1995) propose to go one step further and estimate the marginal distributions non parametrically. In the case of i.i.d observations, an estimator for the copula parameter is obtained by maximizing the pseudo likelihood function

$$\ell_{cop}(\boldsymbol{\theta}_{cop}|\hat{\mathbf{u}}_{1:T}, \hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_d) = \prod_{t=1}^{T} c_{1:d}\left(\hat{u}_{1,t}, \ldots, \hat{u}_{d,t} \middle| \boldsymbol{\theta}_{cop}\right),$$

where $\hat{u}_{i,t}$, $i = 1, \ldots, d$ are standardized ranks (see Section 1.1). In general, this is called a semi-parametric estimation procedure. In particular, semi parametric estimation can help to avoid estimation bias which can be induced in parametric estimation of copula parameters if marginal distributions are severely misspecified (Kim et al. 2007). Similar methods in the presence of a time series structure have been considered by Chen and Fan (2006). In both cases, the obtained estimates for the copula parameters are asymptotically consistent. In particular, while we will work with fully parametric models within the scope of this thesis, this illustrates that multi step estimation is generally feasible in the context of copula models.

1.3 Scoring rules and model selection criteria

In this section, criteria which we will apply for model comparison are introduced.

1.3.1 Log predictive score

Scoring rules, as considered by Gneiting and Raftery (2007), "assess the quality of probabilistic forecasts, by assigning a numerical score based on the predictive distribution and on the event or value that materializes". In particular, the logarithmic score or log predictive score is popular. For a predictive density $f(\cdot)$ and an observed value of $\mathbf{x}$ it is defined as

$$LogS(f, \mathbf{x}) = \ln(f(\mathbf{x})). \tag{1.8}$$

We will use this scoring rule to analyze the out-of-sample performance of our models. A comprehensive overview of scoring rules is given by Gneiting and Raftery (2007).

1.3.2 AIC/BIC

Two of the most commonly used information criteria for model selection are AIC and BIC (see e.g. Burnham and Anderson (2004) for an extended discussion). The Akaike information criterion (AIC) (Akaike 1974) is an in-sample measure of goodness-of-fit. Similar to the log predictive score it is based on the log-likelihood of a statistical model. However, since it is an in-sample measure, an additional punishment term is introduced to prevent overfitting. For a statistical model with density $f(\cdot;\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$ contains k parameters, and independent observations $\mathbf{x}_t$, $t = 1, \ldots, T$, the AIC is given by

$$AIC(f, \mathbf{x}_{1:T}) = 2k - 2\sum_{t=1}^{T} \ln(f(\mathbf{x}_t; \hat{\boldsymbol{\theta}})),$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ in Θ. In some settings, we will want to weight observations according to their relative relevance. In this case, for positive weights ω_t, $t = 1, \ldots, T$, which sum to one, a weighted AIC can be calculated as

$$AIC_{\omega_{1:T}}(f, \mathbf{x}_{1:T}) = 2k - 2\sum_{t=1}^{T} \omega_t \ln(f(\mathbf{x}_t; \hat{\boldsymbol{\theta}})),$$

where $\hat{\boldsymbol{\theta}}$ is minimizing the weighted AIC. While the penalty term for the number of parameters in the AIC does not depend on the sample size, the penalty term in the Bayesian information criterion (BIC) (Schwarz 1978) does. It is given by

$$BIC(f, \mathbf{x}_{1:T}) = \ln(T)k - 2\sum_{t=1}^{T} \ln(f(\mathbf{x}_t; \hat{\boldsymbol{\theta}})),$$

which puts a stronger penalty on the number of parameters. When used as a model selection criterion, this leads to more parsimonious models.

1.4 Graph theory: regular vine tree sequence

The common tool used to organize the PCCs, which are at the core of this thesis, is the graph theoretic construct of an R-vine tree sequence. Therefore, we will consider some basic notions of graph theory, which are required throughout the manuscript, in this section and explain how R-vine tree sequences can be described in a convenient matrix notation. The matrix notation as we consider it here has been developed by Dißmann et al. (2013) and Dißmann (2010) building on earlier work by Morales-Nápoles (2011). We will only give the most important definitions which are required in this thesis. Readers interested in graph theory as such or in a more detailed introduction to graph theoretical concepts and algorithms

should consult the classical work of König (1936) or Thulasiraman and Swamy (1992) for a more recent overview. The definitions presented here follow Dißmann et al. (2013) and Bedford and Cooke (2002) and our notation is close to Czado (2010).

The basic object in graph theory is a graph consisting of nodes and edges.

Definition 1.4.1. *A **graph** is an ordered pair $G = (N, E)$ comprising a set of nodes N and a set of edges $E \subseteq \mathcal{P}(N)$. Here, each element $e \in E$ is a two-element subset of N, i.e. $e \subseteq N$, $\#e = 2$.*

A path in G is a sequence $(e_1, \dots, e_n)$, $e_i \in E$, $e_i \neq e_j$, $i, j \in 1, \dots, n$, such that $\forall (e_i, e_{i+1}), i \in 1, \dots, n-1 : \exists n_i \in N, (n_i \in e_i) \wedge (n_i \in e_{i+1})$, i.e. a sequence of edges in which consecutive edges share a common node. If for each each pair of nodes $(n_a, n_b), n_a \neq n_b$ there exists a path $(e_1, \dots, e_n)$, such that $n_a \in e_1$ and $n_b \in e_n$, the graph is called connected. It is called acyclic if such a path does not exist for $n_a = n_b$.

*A **tree** is connected, acyclic graph.*

The graph theoretical object which is central in this thesis is a set of trees (a forest) which are connected by the following conditions.

Definition 1.4.2. *A **regular vine (R-vine) tree sequence** on d elements is an ordered set of trees $\mathcal{V} = (T_1, \dots, T_{d-1})$, $T_i = (N_i, E_i)$, $i \in 1, \dots, d-1$, such that*

(i) $N_1 = \{1, \dots, d\}$ (Tree 1 has nodes $1, \dots, d$).

(ii) For $i \in 2, \dots, d-1$, $N_i = E_{i-1}$ (The edges of tree i become the nodes in tree $i+1$).

(iii) For $i \in 2, \dots, d-1$, $\forall e = \{a, b\} \in E_i : \#(a \cap b) = 1$ (If two nodes in tree i are joined by an edge, the corresponding edges in tree $i-1$ must have a common node (proximity condition)).

The edges of the R-vine tree structure $\mathcal{V}$ will later correspond to conditional distribution functions and their corresponding copula. To clarify this, we require some further notation.

Definition 1.4.3. *For an edge $e = \{j, k\} \in E_i$ in an R-vine tree sequence, the complete union A_e is defined as*

$$A_e := \Big\{ j \in N_1 | \exists\ e_1 \in E_1, \dots, e_{i-1} \in E_{i-1} : j \in e_1 \in \dots \in e_{i-1} \in e, \Big\}.$$

*The **conditioning set** of e is $D_e := A_j \cap A_k$ and the **conditioned sets** associated with e are given by $C_{e,j} := A_j \setminus D_e$ and $C_{e,k} := A_k \setminus D_e$. We will often write $j(e) := C_{e,j}$ and $k(e) := C_{e,k}$. The constraint set $\mathcal{CV}$ of an R-vine $\mathcal{V}$ is defined as*

$$\mathcal{CV} = \Big\{ \{\{j(e), k(e)\}, D_e\} | i \in 1, \dots, d-1, e \in E_i, e = \{j, k\} \Big\}.$$

Since the constraint set contains all information required to reconstruct the R-vine $\mathcal{V}$, it is common practice to use the constraint set to denote edges of $\mathcal{V}$. By induction, the conditioned sets are singletons, i.e. $\#j(e) = \#k(e) = 1$ (Bedford and Cooke 2002). To shorten notation, the elements of the constraint set are usually given as follows: the numbers in $j(e)$ and $k(e)$ are separated by a comma and given to the left of a "|" sign, with the numbers in D_e appearing on its right.

An example of an R-vine tree sequence is drawn in Figure 1.1. Here, $2,3|1$ represents the edge $\{\{1,2\},\{1,3\}\}$ in set notation. The complete union of this edge is $\{1,2,3\}$, since $1 \in \{1,2\} \in \{\{1,2\},\{1,3\}\}$, $2 \in \{1,2\} \in \{\{1,2\},\{1,3\}\}$ and $3 \in \{1,3\} \in \{\{1,2\},\{1,3\}\}$. Also, since $\{1,2\} \cap \{1,3\} = \{1\}$, $\{1\}$ is the conditioning set, while $\{2\}$, $\{3\}$ represent the conditioned sets.

This tree sequence can be represented by a 8×8 lower triangular matrix M as follows: We choose the member of one of the conditioned sets of the edge in T_7, e.g. 8, and write it down as the $(1,1)$ element of a matrix M. Now, we go through the edges where 8 is in a conditioned set from tree T_7 to tree T_1 and write down the elements which are in the other conditioned set in the first column of the matrix. The edge in T_7 is $(7,8|1,2,3,4,5,6)$, so 7 becomes the $(2,1)$ element of M, 2 becomes the $(3,1)$ element since $(2,8|1,3,4,5,6) \in E_6$, et cetera. We obtain

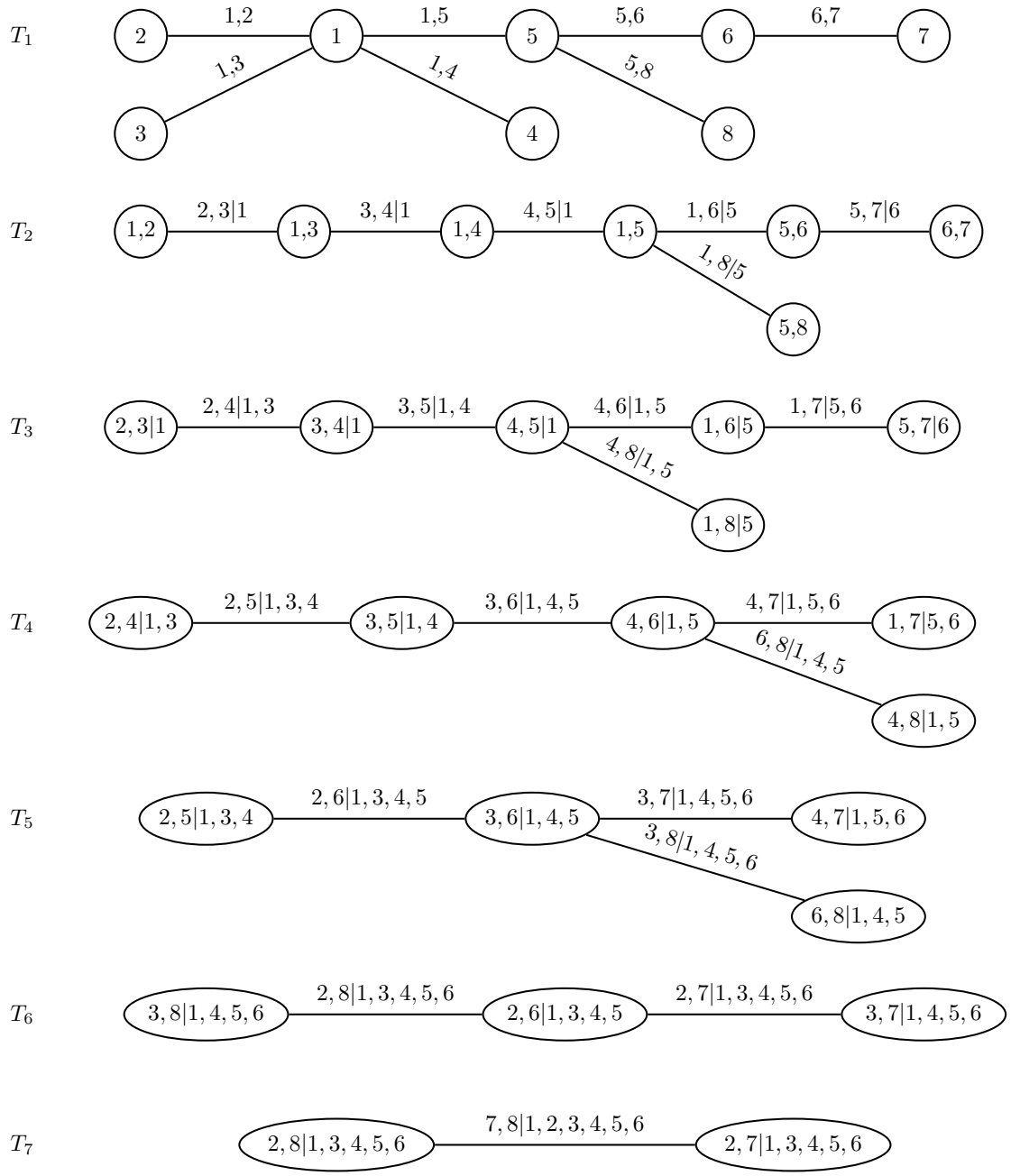
$$\begin{pmatrix} 8 & & & & & & & \\ 7 & & & & & & & \\ 2 & & & & & & & \\ 3 & & & & & & & \\ 6 & & & & & & & \\ 4 & & & & & & & \\ 1 & & & & & & & \\ 5 & & & & & & & \end{pmatrix}, \tag{1.9}$$

and delete all the edges where 8 is in a conditioned set from the vine. This leads to a reduced vine only containing variables $1,\dots,7$ with which we can proceed analogously until we end up with a matrix

$$M = \begin{pmatrix} 8 & & & & & & & \\ 7 & 7 & & & & & & \\ 2 & 2 & 6 & & & & & \\ 3 & 3 & 2 & 5 & & & & \\ 6 & 4 & 3 & 2 & 4 & & & \\ 4 & 1 & 4 & 3 & 2 & 3 & & \\ 1 & 5 & 1 & 4 & 3 & 2 & 2 & \\ 5 & 6 & 5 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \tag{1.10}$$

and have deleted all edges from the original tree sequence. The matrix M is called R-vine

Figure 1.1: An R-vine tree sequence in eight dimensions with corresponding elements of the constraint set as edge indices.



matrix, and by construction it fulfills the properties specified in Definition 1.4.4.

Definition 1.4.4. *A lower triangular matrix* $M = (m_{i,j})_{j=1,\dots,d, i \geq j}$ *where* $m_{i,j} \in \{1, \dots, d\}$ *is called* ***R-vine matrix*** *if*

(i) $\{m_{i,i}, \dots, m_{d,i}\} \subset \{m_{j,j}, \dots, m_{d,j}\}$ *for* $1 \leq j < i \leq d$ *(The entries of a selected column*

are also contained in all columns to the left of that column).

(ii) $m_{i,i} \notin \{m_{i+1,i+1}, \ldots, m_{d,i+1}\}$ *(The diagonal entry of a column is not contained in any column further to the right).*

(iii) For all $i = d-2, \ldots, 1$ *and* $k = i+1, \ldots, d$ *there exist (j,l) with* $j > i$ *and* $l > j$ *such that*

$$\Big\{m_{k,i}, \{m_{k+1,i}, \ldots, m_{d,i}\}\Big\} = \Big\{m_{j,j}, \{m_{l,j}, m_{l+1,j}, \ldots, m_{d,j}\}\Big\} \quad \textbf{\textit{or}}$$
$$\Big\{m_{k,i}, \{m_{k+1,i}, \ldots, m_{d,i}\}\Big\} = \Big\{m_{l,j}, \{m_{l+1,j}, \ldots, m_{d,j}, m_{j,j}\}\Big\}.$$

(This corresponds to the proximity condition).

To reconstruct the R-vine tree sequence in Figure 1.1 from the R-vine Matrix M in (1.10), we start with adding nodes $1, \ldots, 8$ to tree T_1. The edges connecting these nodes are determined from the diagonal elements and row 8 of the matrix: In column 7, 2 is on the diagonal and 1 in the last row, we add an edge $(1, 2)$. For column 6 of the matrix edge $(1, 3)$ is added, while $(1, 4)$ is added for column 5 and so on. Having reconstructed tree T_1, we proceed in a similar fashion with the diagonal elements and row 7 of the matrix, to determine the conditioned sets of edges in T_2. The edges in T_2 must also have one element in the conditioning set which is determined by row 8 to fulfill the proximity condition. For column 1, we add an edge $(8, 1|5)$ to tree T_2. By iterating through the remaining columns and rows and determining constraint sets analogously the complete tree sequence is reconstructed.

Since the number of possible R-vines in d dimensions is huge ($d!/2 \cdot 2^{\binom{d-2}{2}}$) (Morales-Nápoles 2011), many authors consider the subclasses of drawable vines (D-vines, left panel of Figure 1.2) or canonical vines (C-vines, right panel of Figure 1.2).[1]

Definition 1.4.5. *An R-vine tree sequence* $\mathcal{V} = (T_1, \ldots, T_{d-1})$ *is called*

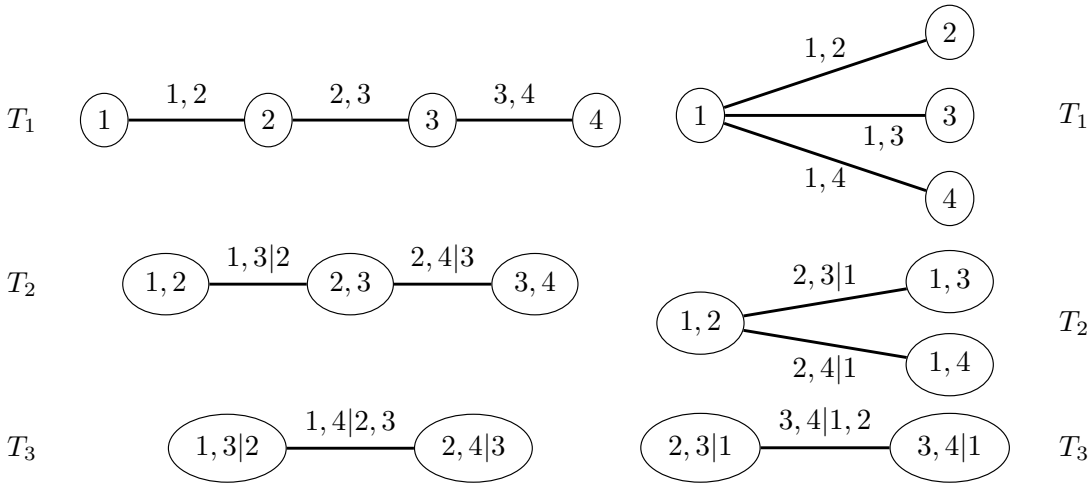
- ***D-vine*** *tree sequence if* $\forall i = 1, \ldots, d-1, n \in N_i : \#\{e \in E_i | n \in e\} \leq 2$.
- ***C-vine*** *tree sequence if* $\forall i = 1, \ldots, d-1 : \exists n \in N_i : \#\{e \in E_i | n \in e\} = d - i$.

1.5 Decomposing a three dimensional distribution

To illustrate the general principle of PCCs, let us consider a three dimensional example with variables $\mathbf{X}_{1:3} = (X_1, X_2, X_3) \in \mathbb{R}^3$.

1. The C-vines are called canonical since they are the easiest to sample, while the drawable vine probably owes its name to the fact that its tree structure has the closest resemblance to a picture of wine grapes on a vine (Kurowicka and Cooke 2006).

Figure 1.2: A D-vine (left panel) and C-vine (right panel) tree sequence in four dimensions.



1.5.1 Continuous case

If $\mathbf{X}_{1:3}$ has joint density $f_{1:3}$ with marginal densities f_1, f_2, f_3, we obtain the following decomposition by conditioning:

$$f_{1:3}(x_1, x_2, x_3) = f_{1|2,3}(x_1|x_2, x_3) \cdot f_{2|3}(x_2|x_3) \cdot f_3(x_3). \tag{1.11}$$

Using Sklar's theorem (Theorem 1.2.2), we can subsequently decompose the conditional densities in (1.11). Let us first consider the distribution of X_1 and X_3 given $X_2 = x_2$ for some $x_2 \in \mathbb{R}$. To simplify the following calculations, as well as inference and model selection procedures which we will introduce later, we will assume that the conditional copula $C_{13;2}$ does not depend on x_2. (This is called *simplifying assumption*, for a discussion see Remark 2.1.1 and Chapter 3.) For the conditional densities in (1.11), this means that

$$\begin{aligned} f_{1|2,3}(x_1|x_2, x_3) &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot f_{1|2}(x_1|x_2), \\ f_{2|3}(x_2|x_3) &= c_{2,3}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2), \end{aligned} \tag{1.12}$$

and similarly, $f_{1|2}$ is given by

$$f_{1|2}(x_1|x_2) = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1).$$

Using the copulas $C_{1,2}$ and $C_{2,3}$, the conditional distribution functions $F_{1|2}$ and $F_{3|2}$, which are required to evaluate the densities in (1.12), can be expressed as

$$\begin{aligned} F_{1|2}(x_1|x_2) &= \frac{\partial F_{1,2}(x_1, x_2)}{\partial x_2} \Big/ f_2(x_2) = \frac{\partial C_{1,2}(F_1(x_1), F_2(x_2))}{\partial x_2} \Big/ f_2(x_2) \\ &= \partial_2 C_{1,2}(F_1(x_1), F_2(x_2)), \qquad \text{and} \\ F_{3|2}(x_3|x_2) &= \partial_1 C_{2,3}(F_2(x_2), F_3(x_3)). \end{aligned} \tag{1.13}$$

The joint density is therefore given in terms of bivariate copulas and marginal distributions:

$$\begin{aligned} f_{1:3}(x_1, x_2, x_3) =& c_{13;2}\Big(\partial_2 C_{1,2}(F_1(x_1), F_2(x_2)), \partial_1 C_{2,3}(F_2(x_2), F_3(x_3))\Big) \\ & \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_3(x_3) \cdot f_2(x_2) \cdot f_1(x_1). \end{aligned}$$

1.5.2 Mixed discrete and continuous case

The same principle also applies if X_1, X_2, X_3 are not all continuous. For example, let us consider two continuous variables $\tilde{X}_1 \in \mathbb{R}$, $\tilde{X}_3 \in \mathbb{R}$ with densities $\tilde{f}_1$, $\tilde{f}_3$ and one discrete variable $\tilde{X}_2 \in \mathbb{Z}$ with pmf $\tilde{p}_2$. For the decomposition into bivariate building blocks, we start with the (generalized)[2] joint density of $\tilde{\mathbf{X}}_{1:3} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$. Given the cumulative distribution function $\tilde{F}_{1:3}$ of $\tilde{\mathbf{X}}_{1:3}$, it is given by

$$\tilde{f}_{1:3}(x_1, x_2, x_3) = \frac{\partial^2}{\partial x_1 \partial x_3} \left(\tilde{F}_{1:3}(x_1, x_2, x_3) - \tilde{F}_{1:3}(x_1, x_2 - 1, x_3) \right),$$

while the generalized density $\tilde{f}_2$ of $\tilde{X}_2$ is its pmf $\tilde{f}_2(.) = \tilde{p}_2(.)$. By conditioning, the joint density can be decomposed similarly as in the continuous case, we obtain

$$\tilde{f}_{1:3}(x_1, x_2, x_3) = \tilde{f}_{1|2,3}(x_1|x_2, x_3) \cdot \tilde{f}_{2|3}(x_2|x_3) \cdot \tilde{f}_3(x_3). \tag{1.14}$$

Using Sklar's theorem, the terms in (1.14) can be decomposed similarly as for $f_{1:3}$:

$$\begin{aligned} \tilde{f}_{1|2,3}(\tilde{x}_1|\tilde{x}_2, \tilde{x}_3) &= \tilde{c}_{1,3;2}(\tilde{F}_{1|2}(\tilde{x}_1|\tilde{x}_2), \tilde{F}_{3|2}(\tilde{x}_3|\tilde{x}_2)) \cdot \tilde{f}_{1|2}(\tilde{x}_1|\tilde{x}_2), \\ \tilde{f}_{2|3}(\tilde{x}_2|\tilde{x}_3) &= \tilde{F}_{2|3}(\tilde{x}_2|\tilde{x}_3) - \tilde{F}_{2|3}(\tilde{x}_2 - 1|\tilde{x}_3) \\ &= \partial_2 \tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2), \tilde{F}_3(\tilde{x}_3)) - \partial_2 \tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2 - 1), \tilde{F}_3(\tilde{x}_3)), \end{aligned}$$

$$\begin{aligned} \tilde{f}_{1|2}(\tilde{x}_1|\tilde{x}_2) = \frac{\partial}{\partial \tilde{x}_1} \tilde{F}_{1|2}(\tilde{x}_1|\tilde{x}_2) &= \frac{\partial}{\partial \tilde{x}_1} \left(\frac{\tilde{F}_{1,2}(\tilde{x}_1, \tilde{x}_2) - \tilde{F}_{1,2}(\tilde{x}_1, \tilde{x}_2 - 1)}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)} \right) \\ &= \frac{\partial}{\partial \tilde{x}_1} \left(\frac{\tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2)) - \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2 - 1))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)} \right) \\ &= \left(\frac{\partial_1 \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2)) - \partial_1 \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2 - 1))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)} \right) \cdot \tilde{f}_1(\tilde{x}_1). \end{aligned}$$

where we write $\tilde{C}_{ij}$ for the copula corresponding to $\tilde{F}_{ij}$. We see that $\tilde{F}_{1|2}(\tilde{x}_1|\tilde{x}_2)$ is given by

$$\tilde{F}_{1|2}(\tilde{x}_1|\tilde{x}_2) = \frac{\tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2)) - \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2 - 1))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)}, \tag{1.15}$$

2. With *generalized* density, we mean the density of $\tilde{\mathbf{X}}_{1:3}$ w.r.t. the product measure on the respective supports of the marginal variables. For discrete margins with values in $\mathbb{R}$ this is the counting measure on the set of possible outcomes, for continuous margins we consider the Lebesgue measure in $\mathbb{R}$.

and the expression for $\tilde{F}_{3|2}$ follows analogously. Thus, $\tilde{f}_{1:3}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ can again be expressed in terms of only the marginal distributions and the bivariate copulas $\tilde{C}_{1,2}$, $\tilde{C}_{2,3}$ and $\tilde{C}_{1,3;2}$:

$$\begin{aligned}\tilde{f}_{1:3}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) =& \tilde{c}_{1,3;2}\left(\frac{\tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2)) - \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2 - 1))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)},\right. \\ &\left.\frac{\tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2), \tilde{F}_3(\tilde{x}_3)) - \tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2 - 1), \tilde{F}_3(\tilde{x}_3))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)}\right) \\ &\cdot \left(\partial_2 \tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2), \tilde{F}_3(\tilde{x}_3)) - \partial_2 \tilde{C}_{2,3}(\tilde{F}_2(\tilde{x}_2 - 1), \tilde{F}_3(\tilde{x}_3))\right) \\ &\cdot \left(\frac{\partial_1 \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2)) - \partial_1 \tilde{C}_{1,2}(\tilde{F}_1(\tilde{x}_1), \tilde{F}_2(\tilde{x}_2 - 1))}{\tilde{F}_2(\tilde{x}_2) - \tilde{F}_2(\tilde{x}_2 - 1)}\right) \cdot \tilde{f}_3(\tilde{x}_3) \cdot \tilde{f}_1(\tilde{x}_1).\end{aligned}$$

1.5.3 Margins with discrete and continuous components

As a last example, let us now consider the case where $\bar{X}_1$, $\bar{X}_3$ are continuous and the distribution of $\bar{X}_2$ is a mixture between a continuous density g on $\mathbb{R}$ and point masses on $\mathbb{Q}$. In this case, the generalized density of $\bar{X}_2$ can be represented as

$$\bar{f}_2(\bar{x}_2) = p \cdot g(\bar{x}_2) + (1 - p) \sum_{q \in \mathbb{Q}} \left(\mathbb{1}_{\{q\}}(\bar{x}_2) \cdot \omega_q\right),$$

where $\omega_q \in [0; 1]$ is the point mass on $q \in \mathbb{Q}$, $p \in [0; 1]$, and $\sum_{q \in \mathbb{Q}} \omega_q = 1$.

While the conditional density $\bar{f}_{1|2,3}$ is again decomposed as

$$\bar{f}_{1|2,3}(\bar{x}_1|\bar{x}_2, \bar{x}_3) = \bar{c}_{1,3;2}(\bar{F}_{1|2}(\bar{x}_1|\bar{x}_2), \bar{F}_{3|2}(\bar{x}_3|\bar{x}_2)) \cdot \bar{f}_{1|2}(\bar{x}_1|\bar{x}_2),$$

we distinguish two cases for $\bar{f}_{2|3}$ and $\bar{f}_{1|2}$. Denoting the left hand limit of the distribution function $\bar{F}_2$ in $\bar{x}_2$ by $\bar{F}_2(\bar{x}_{2,1})$, we obtain

$$\bar{f}_{2|3}(\bar{x}_2|\bar{x}_3) = \begin{cases} \bar{c}_{2,3}(\bar{F}_2(\bar{x}_2), \bar{F}_3(\bar{x}_3)) \cdot \bar{f}_2(\bar{x}_2) & \bar{x}_2 \notin \mathbb{Q} \\ \partial_2 \bar{C}_{2,3}(\bar{F}_2(\bar{x}_2), \bar{F}_3(\bar{x}_3)) - \partial_2 \bar{C}_{2,3}(\bar{F}_2(\bar{x}_{2,1}), \bar{F}_3(\bar{x}_3)) & \bar{x}_2 \in \mathbb{Q} \end{cases},$$

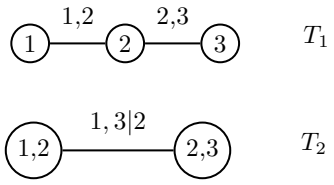
$$\bar{f}_{1|2}(\bar{x}_2|\bar{x}_2) = \begin{cases} \bar{c}_{1,2}(\bar{F}_1(\bar{x}_1), \bar{F}_2(\bar{x}_2)) \cdot \bar{f}_1(\bar{x}_1) & \bar{x}_2 \notin \mathbb{Q} \\ \left(\frac{\partial_1 \bar{C}_{1,2}(\bar{F}_1(\bar{x}_1), \bar{F}_2(\bar{x}_2)) - \partial_1 \bar{C}_{1,2}(\bar{F}_1(\bar{x}_1), \bar{F}_2(\bar{x}_{2,1}))}{\bar{F}_2(\bar{x}_2) - \bar{F}_2(\bar{x}_{2,1})}\right) \cdot \bar{f}_1(\bar{x}_1) & \bar{x}_2 \in \mathbb{Q} \end{cases},$$

and the corresponding expressions for $\bar{F}_{2|3}$ and $\bar{F}_{1|3}$ follow similarly by treating $\bar{X}_2$ as a continuous variable for $\bar{x}_2 \notin \mathbb{Q}$ and as a discrete variable for $\bar{x}_2 \in \mathbb{Q}$. This means, that the expression for the joint density $\bar{f}_{1:3}$ will be the same as for $f_{1:3}$ if $\bar{x}_2 \notin \mathbb{Q}$ and the same as for $\tilde{f}_{1:3}$ if $\bar{x}_2 \in \mathbb{Q}$, replacing $\tilde{F}_2(\tilde{x}_2 - 1)$ by the more general expression $\tilde{F}(\tilde{x}_{2,1})$ for the left-hand limit.

1.5.4 Corresponding R-vine

The R-vine tree sequence corresponding to this example is given in Figure 1.3. The first tree has the marginal variables as nodes and edges between 1 and 2 and between 2 and 3 to represent the copula functions $C_{1,2}$ and $C_{2,3}$. The second tree contains the edges from the first tree as nodes and an edge between them to represent the conditional copula $C_{1,3;2}$.

Figure 1.3: The trees representing the three dimensional example.



1.6 Software

Most computations in this dissertation have been performed using the statistical software package R (R Development Core Team 2011). In particular, the computations in Section 5.1 have been performed using R 2.8.1 on Linux and Mac OS X, while the newer versions R 2.12 (Windows) and 2.15 (Mac) have been used for the results in Sections 6 and 5.2 as well as Model (4) in 5.1.

The basic methods for R-vine copula models, such as likelihood computation and model selection, are publicly available in the R package `VineCopula` (Schepsmeier et al. 2012) which is available on CRAN. The computations for Section 5.2 (and Model (4) in 5.1) have been performed using VineCopula 1.1, while a prerelease version was used for Section 5.1. The modifications to the VineCopula package which are required to handle the data set with discrete and continuous variables in Section 6 are not publicly available at the time of writing this thesis.

Because of continuous developments in the software packages on CRAN and the basic R software, using different R versions and/or future versions of R packages might possibly yield different results.

Chapter 2

Pair copula constructions and regular vines

Since pair copula constructions, although often considered as a very recent development in statistical modeling, have more than 15 years of history with contributions by many authors over the last years, we start this chapter with a brief historical and literature overview.

The idea of constructing a multivariate dependence model from bivariate copulas as building blocks goes back to a groundbreaking paper of Joe (1996), developing "a class of m-variate distributions with given [univariate] margins and m(m-1)/2 dependence parameters". This is motivated as constructing a distribution with the same number of parameters, and thus a similar flexibility in realizable dependence structures, as the multivariate normal distribution, but with "properties that the family of multivariate normal distributions does not have", such as tail-dependence. The particular construction he describes is what would later be called a D-vine structure. While focussing on dependence properties such as the tail dependence of bivariate marginal distributions (which has been generalized in Joe et al. (2010)), Joe also discusses simulations (see Stöber and Czado (2012) for a recent overview) and derives several basic properties. Of particular importance for the inference methods which have been developed later is the property he calls "densities without integrals": given the distribution functions, conditional distribution functions and densities of bivariate copulas in the PCC, the multivariate likelihood can be expressed in terms of those functions only, without any integrals. This applies for all R-vine copulas. This property is the main reason for the computational attractiveness of R-vine models as compared to other graphical ways of specifying a dependence structure such as non-Gaussian directed acyclic graphs (DAGs) (Bauer et al. 2012). While this makes PCCs not only a very flexible but also a highly tractable model he admits that these features are paid for by the lack of permutation symmetry and the fact that not all marginal distributions are in the same family (bivariate marginal distributions can in fact be very non-standard, see Stöber and Czado (2012)). By outlining all important features of PCCs, Joe (1996) provided the basis for further developments in the field, and many details which are mentioned shortly have regained interest in recent years.

While the theoretical foundations go back to Joe (1996), the graphical structure of regular

vines which have become the "face" of PCCs has been introduced by Bedford and Cooke (2001); Bedford and Cooke (2002). In their papers, Bedford and Cooke develop the R-vine as a set of connected trees which can be used to specify distributions in a framework of expert judgement. In particular, they derive a general expression for the density of a PCC specified on an R-vine and extensively discuss the special case of a multivariate normal distribution. In this case, choosing a specific R-vine corresponds to choosing a set of partial correlations in $[-1; 1]$. As they show, these fully specify the correlation structure while having the appealing property that there are no restrictions on these partial correlations to ensure that the correlation matrix is positive semi-definite (this allows to use vines to specify distributions on the set of correlation matrices, see Lewandowski et al. (2009)).

This discussion has been extended in the book by Kurowicka and Cooke (2006), where the chapter on PCCs is centered on their relationship to correlation and rank correlation matrices. While not all rank correlation matrices can be realized using the dependence structure of a multivariate normal distribution, Kurowicka and Cooke show that all rank correlation specifications on an R-vine can in fact be realized in this way. Expressing a correlation matrix in terms of partial correlations on a vine, they also consider completing correlation matrices where some entries are missing and "repairing" misspecified, i.e. not positive semidefinite, correlation matrices. Perhaps even more importantly, they also discuss a first heuristic approach to infer a rank correlation vine from a given multivariate data set and extensively discuss sampling from an R-vine distribution.

The seminal paper of Aas et al. (2009) builds on the publications by Bedford, Cooke, Kurowicka and Joe but presents what the authors call a more "practical" approach to the theory. They illustrate their methods with an application to financial data, preparing the ground for many publications inferring PCCs from real world data.

Since then, many applications have been considered and four workshops on theory and applications of PCCs have been held in Delft, Oslo and Munich and a collection of papers from these has been published in Kurowicka and Joe (2011). Among the contributions in this book, we want to highlight in particular the chapter by Morales-Nápoles (2011) on counting the number of possible vine structures. The matrix notation for R-vines which is developed here has provided the basis for the algorithmic likelihood calculation and simulation of general R-vine distributions in Dißmann et al. (2013). While authors had usually restricted the class of R-vines to the subclass of C-Vines or D-Vines before, the algorithms and model selection heuristics developed by Dißmann (2010) made the whole class of R-vine copulas accessible. The computational techniques presented by him have been developed further and have been made available to practitioners and researchers in the R-package `VineCopula` (Schepsmeier

et al. 2012).

While the aforementioned literature considers PCCs for distributions where all univariate marginal distributions are continuous, the principle is more general. Panagiotelis et al. (2012) provide a pioneering discussion of D-vine based PCCs for discrete data, which we will build upon for the presentation of PCCs in this thesis.

In the remainder of this chapter, we consider PCCs for distributions where some univariate marginal distributions are continuous and some of the marginal variables are discrete (Subsection 2.1). This discussion is taken from Stöber et al. (2012). Having introduced the pair copula decompostion, we will develop algorithms for the calculation of the likelihood, score function and observed information in Subsection 2.2. These algorithms are taken from Stöber and Schepsmeier (2013), and are developed only for the case of R-vine copulas with continuous margins for notational simplicity. They are available also for the mixed discrete and continuous margins case, and we will hint at necessary modifications in the text.

2.1 PCCs for discrete and continuous margins

There are two basic approaches to pair copula constructions in d dimensions: Joe et al. (2010) starts from the marginal distributions, and proceeds by subsequently "glueing" them together using copula functions. By making appropriate choices to ensure that the distributions combined at each level of the construction are compatible, this leads to a joint distribution of d variables. Starting with marginal distribution functions $F_1, \dots, F_d$, his constructions proceeds roughly as follows. First, the distributions are combined to bivariate distribution functions by assigning copulas $C_{i,i+1}$, $i = 1, \dots, d$.

$$\begin{aligned} F_{1,2}(x_1, x_2) &:= C_{1,2}\left(F_1(x_1), F_2(x_2)\right) \\ F_{2,3}(x_2, x_3) &:= C_{2,3}\left(F_2(x_2), F_3(x_3)\right) \\ F_{3,4}(x_3, x_4) &:= C_{3,4}\left(F_3(x_3), F_4(x_4)\right) \\ &\vdots \end{aligned} \tag{2.1}$$

These bivariate distributions are combined to three-dimensional distributions by assigning a copula $C_{i,i+2;i+1}(\cdot, \cdot; x_{i+1})$, $i = 1, \dots, d-2$, to the conditional distribution of (X_i, X_{i+2}) given $X_{i+1} = x_{i+1}$. By integration, we obtain the three dimensional cdfs $F_{i,i+1,i+2}$.

$$\begin{aligned} F_{1,2,3}(x_1, x_2, x_3) &:= \int_{-\infty}^{x_2} C_{1,3;2}(F_{1|2}(x_1|z_2), F_{3|2}(x_3|z_2); z_2)\ dF_2(z_2) \\ F_{2,3,4}(x_2, x_3, x_4) &:= \int_{-\infty}^{x_3} C_{2,4;3}(F_{2|3}(x_2|z_3), F_{4|3}(x_4|z_3); z_3)\ dF_3(z_3) \\ &\vdots \end{aligned} \tag{2.2}$$

By choosing copulas $C_{i,i+1}$, $C_{i,i+2;i+1}$, $C_{i,i+3;i+1,i+2}$, ..., this ensures that the marginal distributions of the k dimensional distributions which are combined to a $k+1$ dimensional distribution at each step are compatible. In particular, the F_{23} margins of the cdfs F_{123} and F_{234} obtained in Equation (2.2) are identical. Since this construction is based on (conditional) distribution functions, it is obvious that it works for general distributions and does not require absolute continuity of the distributions with respect to the Lebesgue measure. In particular, the marginal distributions can correspond to discrete random variables and have point masses.

Remark 2.1.1. *Also note that in general the copula functions $C_{i,i+k;i+1,\dots,i+k-1}$ can depend on the values of the variables $X_{i+1},\dots,X_{i+k-1}$. To simplify notation in the following chapters, we will suppress this dependence of the conditional copulas on the values of the variables which are conditioned on in notation. In fact, it is a common assumption made for inference purposes that conditional copulas $C_{i,j;D}$ do not depend on the values of $\{X_k|k\in D\}$, i.e. that conditional copulas are constant. This assumption will be discussed in Chapter 3.*

For inference purposes, we will also require the densities of PCCs. To derive these, it is illustrative to follow the approach taken by Aas et al. (2009) and start with a d-dimensional random vector $\mathbf{X}_{1:d} = (X_1,\dots,X_d)$ with density $f_{1:d}(\mathbf{x}) = f_{1:d}(x_1,\dots,x_d)$ which we will decompose into an expression involving only bivariate copulas and the marginal distributions.

Here, we allow for the marginal distributions of X_j, $j=1,\dots,d$ to be either discrete or continuous. Therefore, the corresponding density $f_{1:d}$ is not with respect to the standard Lebesgue measure, but with respect to the product measure of measures on the respective support of the marginal variables. For discrete variables, this measure is the counting measure on the set of respective outcomes. For example, if the discrete variables take values in $\mathbb{Q}$ we consider the counting measure on $\mathbb{Q}$. For continuous margins, the respective measure is the Lebesgue measure on $\mathbb{R}$. This restriction to purely discrete and purely continuous marginal distributions is made only for notational convenience. A similar decomposition is possible if marginal distributions are mixtures of discrete and continuous distributions (e.g. zero-adjusted Gamma). In this case, when decomposing the density function at $\mathbf{x}_{1:d}$, a variable X_j is to be treated as a discrete variable if $P(X_j = x_j) > 0$ and as a continuous variable otherwise, as illustrated in the example in Section 1.5.3.

By subsequent conditioning, we can factorize $f_{1:d}$ into conditional densities,

$$f_{1:d}(\mathbf{x}_{1:d}) = f_{1|2:d}(x_1|\mathbf{x}_{2:d})\cdot f_{2|3:d}(x_2|\mathbf{x}_{3:d})\cdot\ldots\cdot f_d(x_d). \tag{2.3}$$

We will now further decompose the densities $f_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d})$ in this expression. For this, we choose a variable X_h, with $j<h\le d$.

X_j continuous, X_h continuous

In the case where X_j and X_h are both continuous, we obtain the relationship

$$f_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) = c_{j,h;(j+1):d\setminus h}\left(F_{j|(j+1):d\setminus h}(x_j|\mathbf{x}_{(j+1):d\setminus h}, F_{h|(j+1):d\setminus h}(x_h|\mathbf{x}_{(j+1):d\setminus h})\right) \cdot f_{j|(j+1):d\setminus h}(x_j|\mathbf{x}_{(j+1):d\setminus h}), \tag{2.4}$$

from the conditional version of Sklar's theorem for densities (Equation (1.6)).

X_j discrete, X_h discrete

If X_j and X_h are both discrete, as it is the case in Panagiotelis et al. (2012), we have

$$\begin{aligned}
&f_{j|(j+1):d}(x_j|x_{j+1},\dots,x_d) = P(X_j = x_j|\mathbf{X}_{(j+1):d} = \mathbf{x}_{(j+1):d})\\
&= \frac{P(X_j = x_j, X_h = x_h|\mathbf{X}_{(j+1):d\setminus h} = \mathbf{x}_{(j+1):d\setminus h})}{P(X_h = x_h|\mathbf{X}_{(j+1):d\setminus h} = \mathbf{x}_{(j+1):d\setminus h})}\\
&= \frac{\sum_{i_j=0,1}\sum_{i_h=0,1}(-1)^{i_j+i_h}P(X_j \le x_{j,i_j}, X_h \le x_{h,i_h}|\mathbf{X}_{(j+1):d\setminus h} = \mathbf{x}_{(j+1):d\setminus h})}{P(X_h = x_h|\mathbf{X}_{(j+1):d\setminus h} = \mathbf{x}_{(j+1):d\setminus h})}\\
&= \sum_{i_j,i_h=0}^{1}(-1)^{i_j+i_h}\frac{C_{j,h;(j+1):d\setminus h}\left(F_{j|(j+1):d\setminus h}(x_{j,i_j}|\mathbf{x}_{(j+1):d\setminus h}), F_{h|(j+1):d\setminus h}(x_{h,i_h}|\mathbf{x}_{(j+1):d\setminus h})\right)}{f_{h|(j+1):d\setminus h}(x_h|\mathbf{x}_{(j+1):d\setminus h})}\\
&= c_{j,h|(j+1):d\setminus h}\cdot f_{j|(j+1):d\setminus h}(x_j|\mathbf{x}_{(j+1):d\setminus h}),
\end{aligned} \tag{2.5}$$

where we have shortened the notation by defining

$$\begin{aligned}
&c_{j,h;(j+1):d\setminus h} :=\\
&= \sum_{i_j,i_h=0}^{1}(-1)^{i_j+i_h}\frac{C_{j,h;(j+1):d\setminus h}\left(F_{j|(j+1):d\setminus h}(x_{j,i_j}|\mathbf{x}_{(j+1):d\setminus h}), F_{h|(j+1):d\setminus h}(x_{h,i_h}|\mathbf{x}_{(j+1):d\setminus h})\right)}{f_{h|(j+1):d\setminus h}(x_h|\mathbf{x}_{(j+1):d\setminus h})f_{j|(j+1):d\setminus h}(x_j|\mathbf{x}_{(j+1):d\setminus h})}.
\end{aligned} \tag{2.6}$$

This is the discrete equivalent of the copula density in the continuous case. We will now derive similar expressions for the case where either X_j or X_h is discrete.

X_j discrete, X_h continuous

If X_j is discrete and X_h is continuous, we can write the conditional pmf as

$$\begin{aligned} f_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) &= P(X_j = x_j|\mathbf{X}_{(j+1):d} = \mathbf{x}_{(j+1):d}) \\ &= \frac{\partial}{\partial x_h} F_{j,h|(j+1):d\backslash h}(x_{j,0}, x_h|\mathbf{x}_{(j+1):d\backslash h}) - \frac{\partial}{\partial x_h} F_{j,h|(j+1):d\backslash h}(x_{j,1}, x_h|\mathbf{x}_{(j+1):d\backslash h}) \\ &= \sum_{i_j=0}^{1} (-1)^{i_j} \partial_2 C_{j,h;(j+1):d\backslash h}(F_{j|(j+1):d\backslash h}(x_{j,i_j}|\mathbf{x}_{(j+1):d\backslash h}), F_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})) \\ &= c_{j,h;(j+1):d\backslash h} \cdot f_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}), \end{aligned} \tag{2.7}$$

where we define as in the discrete case

$$\begin{aligned} &c_{j,h;(j+1):d\backslash h} := \\ &= \sum_{i_j=0}^{1} (-1)^{i_j} \frac{\partial_2 C_{j,h;(j+1):d\backslash h}(F_{j|(j+1):d\backslash h}(x_{j,i_j}|\mathbf{x}_{(j+1):d\backslash h}), F_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h}))}{f_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h})}. \end{aligned} \tag{2.8}$$

X_j continuous, X_h discrete

If X_j is continuous and X_h discrete, we obtain

$$\begin{aligned} f_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) &= \frac{\partial}{\partial x_j} F_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) \\ &= \frac{\partial}{\partial x_j} \left[\frac{P(X_j \leq x_j, X_h = x_h|\mathbf{X}_{(j+1):d\backslash h} = \mathbf{x}_{(j+1):d\backslash h})}{P(X_h = x_h|\mathbf{X}_{(j+1):d\backslash h} = \mathbf{x}_{(j+1):d\backslash h})} \right] \\ &= \sum_{i_h=0}^{1} (-1)^{i_h} \frac{F_{j,h|(j+1):d\backslash h}(x_j, x_{h,i_h}|\mathbf{x}_{(j+1):d\backslash h})}{f_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})} \\ &= \sum_{i_h=0}^{1} (-1)^{i_h} \frac{\partial_1 C_{j,h;(j+1):d\backslash h}(F_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}), F_{h|(j+1):d\backslash h}(x_{h,i_h}|\mathbf{x}_{(j+1):d\backslash h}))}{f_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})} \\ &\quad \cdot f_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}) = c_{j,h|(j+1):d\backslash h} \cdot f_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}), \end{aligned} \tag{2.9}$$

where we write

$$\begin{aligned} &c_{j,h;(j+1):d\backslash h} := \\ &= \sum_{i_h=0}^{1} (-1)^{i_h} \frac{\partial_1 C_{j,h;(j+1):d\backslash h}(F_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}), F_{h|(j+1):d\backslash h}(x_{h,i_h}|\mathbf{x}_{(j+1):d\backslash h}))}{f_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})}. \end{aligned} \tag{2.10}$$

Using Expressions (2.4) - (2.9), the joint density in (2.3) can be rewritten as a term involving the copula densities, or density equivalents, $c_{j,h|(j+1):d\backslash h}$ and conditional densities $f_{j|(j+1):d\backslash h}$. Subsequently, the conditional densities can be decomposed again until we end up with an expression involving only bivariate copula densities, or their equivalents, and densities of the

marginal variables. In the arguments of the copula densities however we require conditional distribution functions $F_{j|(j+1):d\backslash h}$ and $F_{h|(j+1):d\backslash h}$. The conditional distribution functions $F_{j|(j+1):d}$ can be expressed using copulas in a similar way as the conditional densities above. If X_h is continuous, we have

$$\begin{aligned} &F_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) \\ &\qquad = \partial_2 C_{j,h;(j+1):d\backslash h}\left(F_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}, F_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})\right), \end{aligned} \tag{2.11}$$

cf. (1.13), and for discrete X_h it follows from Equation (1.15) that

$$\begin{aligned} &F_{j|(j+1):d}(x_j|\mathbf{x}_{(j+1):d}) \\ &= \sum_{i_h=0}^{1}(-1)^{i_h}\frac{C_{j,h;(j+1):d\backslash h}\left(F_{j|(j+1):d\backslash h}(x_j|\mathbf{x}_{(j+1):d\backslash h}), F_{h|(j+1):d\backslash h}(x_{h,i_h}|\mathbf{x}_{(j+1):d\backslash h})\right)}{f_{h|(j+1):d\backslash h}(x_h|\mathbf{x}_{(j+1):d\backslash h})}. \end{aligned} \tag{2.12}$$

Expressions (2.11) and (2.12) apply for both discrete and continuous X_j. Similar relationships do of course apply for the conditional cdfs $F_{j|(j+1):d\backslash h}$ and $F_{h|(j+1):d\backslash h}$.

While the construction in (2.1) and (2.2) is also valid for other choices of copulas than $C_{i,i+1}$, $C_{i,i+2;i+1}$, ..., and the decomposition in (2.4) - (2.12) works for all permutations of variables and choices of h at each level, some restrictions do apply: In particular, for the construction of Joe (1996) we must ensure that the marginal distributions are compatible when combining k dimensional distributions to a $k+1$ dimensional distribution. Also, we want to preserve the "densities without integrals" property. This means that when combining distributions $F_{j:d\backslash h}$ and $F_{(j+1):d}$ with a copula $C_{j,h;(j+1):d\backslash h}$ we need to make sure that copulas $C_{j,k;(j+1):d\backslash\{h,k\}}$ and $C_{h,l;(j+1):d\backslash\{h,l\}}$, $k,l \in (j+1):d\backslash h$ are available such that the conditional cdfs $F_{j|(j+1):d\backslash h}$ and $F_{h|(j+1):d\backslash h}$ can be determined by recursions as in (2.11) and (2.12) without integrations.

Equivalently, we can completely decompose the multivariate distribution using (2.4) - (2.12) and then ask whether the "densities without integrals" property holds if we only know the parametric forms of the copula functions appearing in the decomposition and the marginal distributions. In general, this will not be true if we do not make an intelligent choice for the indices h in (2.4) - (2.9). As shown by Bedford and Cooke (2001); Bedford and Cooke (2002) making such choices is equivalent to requiring the copula indices above to correspond to edge indices in an R-vine tree sequence as defined in Section 1.4. For an R-vine tree sequence $\mathcal{V} = (T_1, \ldots, T_{d-1})$ with edge sets $E_1, \ldots, E_{d-1}$, the density $f_{1:d}$ is then given by

$$\begin{aligned} &f_{1,\ldots,d}(x_1,\ldots,x_d) \\ &= \prod_{i=1}^{d} f_i(x_i) \prod_{i=1}^{d-1}\prod_{e\in E_i} c_{j(e),k(e);D(e)}(F_{j(e)|D(e)}(x_{j(e)}|\boldsymbol{x}_{D(e)}), F_{k(e)|D(e)}(x_{k(e)}|\boldsymbol{x}_{D(e)})). \end{aligned} \tag{2.13}$$

The proximity condition for the tree sequence will guarantee that the required copulas to calculate the conditional distribution functions, and ultimately the density, without integrations are available. We will illustrate this in the next section where we state algorithmic procedures to calculate the likelihood function as well as the score function and observed information for a PCC defined on an R-vine tree sequence.

2.2 Likelihood, score function and observed information

This section is mainly taken from Stöber and Schepsmeier (2013). Since PCCs contain bivariate copulas as their building blocks, the likelihood functions, (conditional) distribution functions and their derivatives will be required for the bivariate copulas on the R-vine in order to apply the algorithms. A comprehensive overview of bivariate copulas, also considering all required derivatives, is given by Schepsmeier and Stöber (2012). The algorithms are publicly available in the R-package `VineCopula` (Schepsmeier et al. 2012). Being interested mainly in the copula structure, we assume that all marginal distributions are uniform and denote an observation on the unit hypercube $[0;1]^d$ by $\mathbf{u}_{1:d}$. Extending the algorithms, which are presented here, to the case of both discrete and continuous margins and calculating derivatives also with respect to parameters of marginal distributions is relatively straightforward by replacing all relations with their discrete-continuous equivalents. However, the notation will be unnecessarily complex which is why only the simplest case is presented here.

We use the matrix notation developed in Section 1.4. In addition to storing the edges of the R-vine tree sequence, we will also store the corresponding pair copula families (we call this set $\mathcal{B}$) and their parameters $\boldsymbol{\theta}$. For illustration purposes, we consider the R-vine in Figure 1.1 with corresponding R-vine matrix

$$M = \begin{pmatrix} m_{1,1} & & & & & & & \\ m_{2,1} & m_{2,2} & & & & & & \\ m_{3,1} & m_{3,2} & m_{3,3} & & & & & \\ m_{4,1} & m_{4,2} & m_{4,3} & m_{4,4} & & & & \\ m_{5,1} & m_{5,2} & m_{5,3} & m_{5,4} & m_{5,5} & & & \\ m_{6,1} & m_{6,2} & m_{6,3} & m_{6,4} & m_{6,5} & m_{6,6} & & \\ m_{7,1} & m_{7,2} & m_{7,3} & m_{7,4} & m_{7,5} & m_{7,6} & m_{7,7} & \\ m_{8,1} & m_{8,2} & m_{8,3} & m_{8,4} & m_{8,5} & m_{8,6} & m_{8,7} & m_{8,8} \end{pmatrix} = \begin{pmatrix} 8 & & & & & & & \\ 7 & 7 & & & & & & \\ 2 & 2 & 6 & & & & & \\ 3 & 3 & 2 & 5 & & & & \\ 6 & 4 & 3 & 2 & 4 & & & \\ 4 & 1 & 4 & 3 & 2 & 3 & & \\ 1 & 5 & 1 & 4 & 3 & 2 & 2 & \\ 5 & 6 & 5 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{8\times 8}. \quad (2.14)$$

For this R-vine, the corresponding parameters and copulas are stored in matrices

$$\boldsymbol{\theta} = \begin{pmatrix} \dots & & & & \\ \dots & \theta_{4,m_{6,5}|m_{7,5},m_{8,5}} & & & \\ \dots & \theta_{4,m_{7,5}|m_{8,5}} & \theta_{3,m_{7,6}|m_{8,6}} & & \\ \dots & \theta_{4,m_{8,5}} & \theta_{3,m_{8,6}} & \theta_{2,m_{8,7}} & \end{pmatrix} = \begin{pmatrix} \dots & & & & \\ \dots & \theta_{4,2|1,3} & & & \\ \dots & \theta_{4,3|1} & \theta_{3,2|1} & & \\ \dots & \theta_{4,1} & \theta_{3,1} & \theta_{2,1} & \end{pmatrix} \in \mathbb{R}^{8\times 8},$$

$$\mathcal{B} = \begin{pmatrix} \dots & & & & \\ \dots & \mathcal{B}_{4,m_{6,5}|m_{7,5},m_{8,5}} & & & \\ \dots & \mathcal{B}_{4,m_{7,5}|m_{8,5}} & \mathcal{B}_{3,m_{7,6}|m_{8,6}} & & \\ \dots & \mathcal{B}_{4,m_{8,5}} & \mathcal{B}_{3,m_{8,6}} & \mathcal{B}_{2,m_{8,7}} & \end{pmatrix} = \begin{pmatrix} \dots & & & & \\ \dots & \mathcal{B}_{4,2|1,3} & & & \\ \dots & \mathcal{B}_{4,3|1} & \mathcal{B}_{3,2|1} & & \\ \dots & \mathcal{B}_{4,1} & \mathcal{B}_{3,1} & \mathcal{B}_{2,1} & \end{pmatrix} \in \mathbb{R}^{8\times 8}.$$

Here, $\theta_{1,2}$ and $\mathcal{B}_{1,2}$ are the $(8,7)$ elements and the 8th column is left empty for notational convenience. Note that the diagonal of M is sorted in descending order which can always be achieved by relabeling the nodes. From now on, we will assume that all matrices are "normalized" in this way as this allows to simplify notation. Therefore we have $m_{i,i} = d-i+1$. Also, we use the 8-dimensional example throughout the remainder of this chapter.

2.2.1 Computation of the likelihood function

To evaluate the (log-) likelihood function (see Equation (2.13)) of an R-vine model, we require the conditional distiributions $F_{j(e)|D(e)}$ and $F_{k(e)|D(e)}$, evaluated at a d-dimensional vector of observations $(u_1, \dots, u_d)$, as arguments of the copula density $c_{j(e),k(e);D(e)}$ corresponding to edge e. To develop a programmable algorithm, we will also store these values in two matrices: V^{direct} and $V^{indirect}$. In particular, we calculate

$$V^{direct} = \begin{pmatrix} \dots & & & & \\ \dots & F_{4|m_{6,5},m_{7,5},m_{8,5}} & & & \\ \dots & F_{4|m_{7,5},m_{8,5}} & F_{3|m_{7,6},m_{8,6}} & & \\ \dots & F_{4|m_{8,5}} & F_{3|m_{8,6}} & F_{2|m_{8,7}} & \\ \dots & u_4 & u_3 & u_2 & u_1 \end{pmatrix} \in \mathbb{R}^{8\times 8}, \tag{2.15}$$

$$V^{indirect} = \begin{pmatrix} \dots & & & & \\ \dots & F_{m_{6,5}|m_{7,5},m_{8,5},4} & & & \\ \dots & F_{m_{7,5}|m_{8,5},4} & F_{m_{7,6}|m_{8,6},3} & & \\ \dots & F_{m_{8,5}|4} & F_{m_{8,6}|3} & F_{m_{8,7}|2} & \\ \dots & & & & \end{pmatrix} \in \mathbb{R}^{8\times 8}. \tag{2.16}$$

Here, $F_{i|D} := F_{i|D}(u_i|\mathbf{u}_D)$ and the last row and column of $V^{indirect}$ are left empty for notational convenience. Note that, for each pair-copula term $c_{j(e),k(e);D(e)}$ in (2.13), the

corresponding terms of V^{direct} and $V^{indirect}$ involving $F_{j(e)|D(e)}$ and $F_{k(e)|D(e)}$ can be easily determined by applying (2.11). When being able to evaluate

$$c_{4,m_{6,5};m_{7,5},m_{8,5}}(F_{4|m_{7,5},m_{8,5}}(u_4|u_{m_{7,5}},u_{m_{8,5}}),F_{m_{6,5}|m_{7,5},m_{8,5}}(u_{m_{6,5}}|u_{m_{7,5}},u_{m_{8,5}}))$$

we do also obtain

$$\begin{aligned}
&F_{4|m_{6,5},m_{7,5},m_{8,5}}(u_4|u_{m_{6,5}},u_{m_{7,5}},u_{m_{8,5}}) = \\
&= (\partial_1 C)_{4,m_{6,5};m_{7,5},m_{8,5}}(F_{4|m_{7,5},m_{8,5}}(u_4|u_{m_{7,5}},u_{m_{8,5}}),F_{m_{6,5}|m_{7,5},m_{8,5}}(u_{m_{6,5}}|u_{m_{7,5}},u_{m_{8,5}})) \quad \text{and} \\
&F_{m_{6,5}|4,m_{7,5},m_{8,5}}(u_{m_{6,5}}|u_4,u_{m_{7,5}},u_{m_{8,5}}) = \\
&= (\partial_2 C)_{4,m_{6,5};m_{7,5},m_{8,5}}(F_{4|m_{7,5},m_{8,5}}(u_4|u_{m_{7,5}},u_{m_{8,5}}),F_{m_{6,5}|m_{7,5},m_{8,5}}(u_{m_{6,5}}|u_{m_{7,5}},u_{m_{8,5}})).
\end{aligned}$$

With all such conditional distribution functions being available, the copula terms in (2.13) corresponding to the next tree T_4 can be evaluated. Following the notation in Aas et al. (2009), we write $h(\cdot,\cdot|\mathcal{B}^{k,i},\theta^{k,i})$ for the conditional distribution function corresponding to a parametric family $\mathcal{B}^{k,i}$ with parameter $\theta^{k,i}$, where $\mathcal{B}^{k,i}$ and $\theta^{k,i}$ denote the (k,i)th element of the matrices $\mathcal{B}$ and $\boldsymbol{\theta}$, respectively. For this, we assume that all copulas are exchangeable such that we do not have to distinguish between conditioning on the first and second variable,

$$h(u_1,u_2|\mathcal{B}^{k,i},\theta^{k,i}) = \partial_1 C(u_2,u_1|\mathcal{B}^{k,i},\theta^{k,i}) = \partial_2 C(u_1,u_2|\mathcal{B}^{k,i},\theta^{k,i}).$$

In this notation, we obtain for example that

$$\begin{aligned}
F_{4|2,1,3}(u_4|u_2,u_1,u_3) &= h(F_{4|1,3}(u_4|u_1,u_3),F_{2|1,3}(u_2|u_1,u_3)|\mathcal{B}_{4,2|1,3},\theta_{4,2|1,3}) \\
&= h(v_{6,5}^{direct},v_{6,6}^{indirect}|\mathcal{B}^{6,5},\theta^{6,5}).
\end{aligned}$$

To write an algorithm for evaluating the likelihood function, we must further decide whether the arguments in each step (i.e. $F_{4|3,1}(u_4|u_3,u_1), F_{2|3,1}(u_2|u_3,u_1)$ in the example) have to be picked from the matrix V^{direct} or $V^{indirect}$. For this, we exploit the descending order of the diagonal of M. From the structure of V^{direct}, we see that the first argument of the copula term with family $\mathcal{B}^{k,i}$ and parameter $\theta^{k,i}$ is stored as the (k,i)th element $v_{k,i}^{\text{direct}}$ of V^{direct}. To locate the second entry, let us denote $\tilde{M} = (\tilde{m}_{k,i}|i=1,\dots,d;k=i,\dots,d)$, where $\tilde{m}_{k,i} := \max\{m_{k,i},\dots,m_{d,i}\}$ for all $i=1,\dots,d$ and $k=i,\dots,d$. The second argument, which is $F_{m_{k,i}|m_{k+1,i},\dots,m_{d,i}}(u_{m_{k,i}}|u_{m_{k+1,i}},\dots,u_{m_{d,i}})$ must be in column $(d-\tilde{m}_{k,i}+1)$ of V^{direct} or $V^{indirect}$ by the ordering of variables. If $\tilde{m}_{k,i} = m_{k,i}$, the conditioned variable $u_{m_{k,i}}$ has the biggest index and thus the entry we are looking for must be in V^{direct}. Similarly, if $\tilde{m}_{k,i} > m_{k,i}$, the variable with the biggest index is in the conditioning set and we must choose from $V^{indirect}$.

Example 2.2.1 (Selection of arguments for $c_{4,2;1,3}$)**.** *As an example for how this procedure selects the correct arguments for copula terms in the R-vine distribution let us consider the*

copula $c_{2,4;1,3}$ *in our example. The corresponding parameter* $\theta_{4,2|1,3}$ *is stored as* $\theta^{6,5}$*, thus we are in the case where* $i = 5$ *and* $k = 6$*. Since* $\tilde{m}_{6,5} = \max\{m_{6,5}, m_{7,5}, m_{8,5}\} = \max\{2, 3, 1\} = 3$ *and* $\tilde{m}_{6,5} = 3 > 2 = m_{6,5}$ *we select as second argument the entry* $v^{indirect}_{k,(d-\tilde{m}_{k,i}+1)} = v^{indirect}_{6,6} = F_{2|1,3}(u_2|u_1, u_3)$*. This and* $v^{direct}_{6,5} = F_{4|1,3}(u_4|u_1, u_3)$ *which we have already selected are the required arguments.*

These sequential selections and calculations are performed in Algorithm 2.2.1, which was developed in Dißmann (2010) and Dißmann et al. (2013). It iterates over the edges in the R-vine tree sequence $\mathcal{V}$, subsequently calculating the likelihood and the corresponding conditional distribution functions for each bivariate copula associated with the vine.

Algorithm 2.2.1 Log-likelihood of an R-vine specification.

Require: d-dimensional R-vine specification in matrix form, i.e., M, $\mathcal{B}$, $\boldsymbol{\theta}$, set of observations $(u_1, \dots, u_d)$.

1: Set $L = 0$.
2: Set $(v^{\text{direct}}_{d,1}, v^{\text{direct}}_{d,2}, \dots, v^{\text{direct}}_{d,d}) = (u_d, u_{d-1}, \dots u_1)$.
3: Let $\tilde{M} = (\tilde{m}_{k,i} | i = 1, \dots, d; k = i, \dots, d)$ where $\tilde{m}_{k,i} = \max\{m_{k,i}, \dots, m_{d,i}\}$ for all $i = 1, \dots, d$ and $k = i, \dots, d$.
4: **for** $i = d-1, \dots, 1$ **do** {Iteration over the columns of M}
5: **for** $k = d, \dots, i+1$ **do** {Iteration over the rows of M}
6: Set $z_1 = v^{\text{direct}}_{k,i}$
7: **if** $\tilde{m}_{k,i} = m_{k,i}$ **then**
8: Set $z_2 = v^{\text{direct}}_{k,(d-\tilde{m}_{k,i}+1)}$.
9: **else**
10: Set $z_2 = v^{\text{indirect}}_{k,(d-\tilde{m}_{k,i}+1)}$.
11: **end if**
12: Set $L = L + \ln(c(z_1, z_2 | \mathcal{B}^{k,i}, \theta^{k,i}))$.
13: Set $v^{\text{direct}}_{k-1,i} = h(z_1, z_2 | \mathcal{B}^{k,i}, \theta^{k,i})$ and $v^{\text{indirect}}_{k-1,i} = h(z_2, z_1 | \mathcal{B}^{k,i}, \theta^{k,i})$.
14: **end for**
15: **end for**
16: **return** L

Remark 2.2.2. *By modifying step* 14 *of Algorithm 2.2.1 to maximize the likelihood of the copula* $c(\cdot, \cdot | \mathcal{B}^{k,i}, \theta^{k,i})$ *with respect to* $\theta^{k,i}$*, we obtain what is called the* **stepwise parameter estimation** *algorithm for R-vine copulas. This type of step-wise estimation has first been considered by Aas et al. (2009) and a discussion of its asymptotic properties is given by*

Hobæk Haff (2013). In particular, it is shown that this procedure is asymptotically consistent and asymptotic efficiency can be achieved in certain special cases. While in general maximum likelihood estimation of R-vine copula parameters results in a high-dimensional numerical maximization problem, this procedure allows to estimate R-vine parameters by solving only one dimensional maximization problems. Simulation studies show that the loss in efficiency is not big for most applications (see e.g. Hobæk Haff (2012)).

2.2.2 Computation of the score function

In this section we develop an algorithm to calculate the derivatives of the R-vine log-likelihood with respect to copula parameters and thus the score function of the model. Throughout the remainder of this chapter, we will assume that all occurring copula densities are continuously differentiable with respect to their arguments and parameters. Further, we assume that the copula parameters are all in $\mathbb{R}$, the extension to two or higher dimensional parameter spaces is straightforward but makes the notation unnecessarily complex.

To determine the log-likelihood derivatives, we will again exploit the hierarchical structure of the R-vine copula model and proceed similarly as for the likelihood calculation. The first challenge which we must overcome to develop an algorithm for the score function is to determine which of the copula terms in Expression (2.13) depend on which parameter directly or indirectly through one of their arguments. Following the steps of the log-likelihood computation and exploiting the structure of the R-vine structure matrix M, this is decided in Algorithm 2.2.2.

For example, for parameter $\theta_{4,2|1,3}$, we have $k = 6$, $i = 5$, and $g = \{1, 2, 3, 4\}$. Therefore, the copula terms in Expression (2.13) which depend on $\theta_{4,2|1,3}$ are those involving the copulas $c_{5,2;1,3,4}$, $c_{6,2;1,3,4,5}$ $c_{7,2;1,3,4,5,6}$, $c_{8,2;1,3,4,5,6}$ and $c_{8,7;1,2,3,4,5,6}$. In particular, for $c_{8,2;1,3,4,5,6}$, we have $h = \{1, 2, 3, 4, 5, 6, 8\}$ in Algorithm 2.2.2, such that $g \subset h$, which means that the corresponding likelihood term depends on $\theta_{4,2|1,3}$. In contrast, for $c_{7,3|1,4,5,6}$, we determine that $h = \{1, 3, 4, 5, 6, 7\}$, and therefore $\#(g \cap h) \neq \#g$ since $2 \notin h$. This means that the likelihood term in Expression (2.13) corresponding to $c_{7,3;1,4,5,6}$ cannot depend on $\theta_{4,2|1,3}$ by the "densities without integrals" property since the X_2 margin in any conditional cdf calculated using $c_{4,2;1,3}$ would have to be integrated out.

Knowing how a specific copula term depends on a given parameter, we can proceed with calculating the corresponding derivatives. Before we explain the derivatives in detail let us start with an example where two of the three possible cases of dependence on a given parameter are illustrated.

Algorithm 2.2.2 Determine copula terms which depend on a specific parameter.

The input of the algorithm is a d-dimensional R-vine matrix M with elements $(m_{l,j})_{l,j=1,\dots,d}$ and the row number k and column number i corresponding to the position of the parameter of interest in the parameter matrix $\boldsymbol{\theta}$. The output will be a matrix C (with elements $(c_{l,j})_{l,j=1,\dots,d}$) of zeros and ones, a one indicating that the copula term corresponding to this position in M will depend on the parameter under consideration.

Set $g := \{m_{i,i}, m_{k,i}, m_{k+1,i}, \dots, m_{d,i}\}$
Set $c_{l,j} := 0 \quad l, j = 1, \dots, d$
for $a = i, \dots, 1$ **do**
  for $b = k, \dots, a+1$ **do**
    Set $h := \{m_{a,a}, m_{b,a}, m_{b+1,a}, \dots, m_{d,a}\}$
    if $\#(g \cap h) == \#g$ **then**
      Set $c_{b,a} := 1$
    end if
  end for
end for
return C

Example 2.2.3 (3-dim)**.** *Let $x_1 \sim F_1, x_2 \sim F_2, x_3 \sim F_3$ and $u_1 = F_1(x_1), u_2 = F_2(x_2)$, $u_3 = F_3(x_3)$, then the joint density can be decomposed as*

$$\begin{aligned} f_{123}(x_1, x_2, x_3) = f_1(x_1) f_2(x_2) f_3(x_3) \cdot c_{1,2}(u_1, u_2|\theta_{1,2}) \cdot c_{2,3}(u_2, u_3|\theta_{2,3}) \\ \cdot c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2}) \end{aligned}$$

The first derivatives of $\ln f_{123}$ *with respect to the copula parameters $\theta_{1,2}$, $\theta_{2,3}$ and $\theta_{1,3|2}$ are*

$$\begin{aligned} \frac{\partial(\ln f_{123}(x_1, x_2, x_3))}{\partial \theta_{1,2}} &= \frac{\partial_{\theta_{1,2}} c_{1,2}(u_1, u_2|\theta_{1,2})}{c_{1,2}(u_1, u_2|\theta_{1,2})} \\ &\quad + \frac{\partial_1 c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})}{c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})} \cdot \partial_{\theta_{1,2}} h_{1,2}(u_1, u_2|\theta_{1,2}), \\ \frac{\partial(\ln f_{123}(x_1, x_2, x_3))}{\partial \theta_{2,3}} &= \frac{\partial_{\theta_{2,3}} c_{2,3}(u_2, u_3|\theta_{2,3})}{c_{2,3}(u_2, u_3|\theta_{2,3})} \\ &\quad + \frac{\partial_2 c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})}{c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})} \cdot \partial_{\theta_{2,3}} h_{2,3}(u_2, u_3|\theta_{2,3}) \qquad \textit{and} \\ \frac{\partial(\ln f_{123}(x_1, x_2, x_3))}{\partial \theta_{1,3|2}} &= \frac{\partial_{\theta_{1,3|2}} c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})}{c_{1,3;2}(h_{1,2}(u_1, u_2|\theta_{1,2}), h_{2,3}(u_3, u_2|\theta_{2,3})|\theta_{1,3|2})}, \end{aligned}$$

respectively.

The first case, which occurs in our example, is that the copula densities $c_{1,2}$ and $c_{2,3}$ depend on their respective parameters directly. For a general term involving a copula $c_{U,V;\mathbf{Z}}$ with parameter θ,

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta\right)\right) &= \frac{\frac{\partial}{\partial\theta}\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta\right)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta\right)}\\ &= \frac{\partial_\theta c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta\right)}.\end{aligned} \tag{2.17}$$

Further, like for $c_{1,3;2}$, a $c_{U,V;\mathbf{Z}}$ term can depend on a parameter θ through one of its arguments, say $F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)$:

$$\begin{aligned}&\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\big(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\big)\right) =\\ &\qquad= \frac{\frac{\partial c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}{\partial F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)}}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}\cdot\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\\ &\qquad= \frac{\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}\cdot\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta).\end{aligned} \tag{2.18}$$

Finally, in dimension $d \geq 4$, both arguments of a $c_{U,V;\mathbf{Z}}$ copula term can depend on a parameter θ. In this case,

$$\begin{aligned}&\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\big(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\big)\right)\\ &\qquad=\frac{\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right)}\cdot\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\\ &\qquad+\frac{\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right)}\cdot\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta).\end{aligned} \tag{2.19}$$

We see that the derivatives of copula terms corresponding to tree T_i in the vine will involve derivatives of conditional distribution functions which are determined by tree T_{i-1}. Thus, it will be convenient to store their derivatives in matrices $S1^{direct,\theta}$ and $S1^{indirect,\theta}$ related to the matrices V^{direct} and $V^{indirect}$ which have been determined during the calculation of the log-likelihood together with the terms

$$\ln\Big(c_{j(e),k(e);D(e)}(F_{j(e)|D(e)}(u_{j(e)}|\boldsymbol{u}_{D(e)}),F_{k(e)|D(e)}(u_{k(e)}|\boldsymbol{u}_{D(e)}))\Big) =: \varrho_{j(e),k(e)|D(e)},$$

for each edge e in the R-vine $\mathcal{V}$, which can also be stored in a matrix V^{values}:

$$V^{values} = \begin{pmatrix} \cdots & & & \\ \cdots & \varrho_{4,m_{6,5}|m_{7,5},m_{8,5}} & & \\ \cdots & \varrho_{4,m_{7,5}|m_{8,5}} & \varrho_{3,m_{7,6}|m_{8,6}} & \\ \cdots & \varrho_{4,m_{8,5}} & \varrho_{3,m_{8,6}} & \varrho_{2,1} \\ \cdots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.20}$$

In particular, we will determine the following matrices:

$$S1^{direct,\theta} = \begin{pmatrix} \cdots & & & \\ \cdots & \frac{\partial}{\partial\theta}F_{4|m_{6,5},m_{7,5},m_{8,5}} & & \\ \cdots & \frac{\partial}{\partial\theta}F_{4|m_{7,5},m_{8,5}} & \frac{\partial}{\partial\theta}F_{3|m_{7,6},m_{8,6}} & \\ \cdots & \frac{\partial}{\partial\theta}F_{4|m_{8,5}} & \frac{\partial}{\partial\theta}F_{3|m_{8,6}} & \frac{\partial}{\partial\theta}F_{2|1} \\ \cdots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.21}$$

$$S1^{indirect,\theta} = \begin{pmatrix} \cdots & & & \\ \cdots & \frac{\partial}{\partial\theta}F_{m_{6,5}|m_{7,5},m_{8,5},4} & & \\ \cdots & \frac{\partial}{\partial\theta}F_{m_{7,5}|m_{8,5},4} & \frac{\partial}{\partial\theta}F_{m_{7,6}|m_{8,6},3} & \\ \cdots & \frac{\partial}{\partial\theta}F_{m_{8,5}|4} & \frac{\partial}{\partial\theta}F_{m_{8,6}|3} & \frac{\partial}{\partial\theta}F_{1|2} \\ \cdots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.22}$$

$$S1^{values,\theta} = \begin{pmatrix} \cdots & & & \\ \cdots & \frac{\partial}{\partial\theta}\varrho_{4,m_{6,5}|m_{7,5},m_{8,5}} & & \\ \cdots & \frac{\partial}{\partial\theta}\varrho_{4,m_{7,5}|m_{8,5}} & \frac{\partial}{\partial\theta}\varrho_{3,m_{7,6}|m_{8,6}} & \\ \cdots & \frac{\partial}{\partial\theta}\varrho_{4,m_{8,5}} & \frac{\partial}{\partial\theta}\varrho_{3,m_{8,6}} & \frac{\partial}{\partial\theta}\varrho_{2,1} \\ \cdots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.23}$$

Here, $\frac{\partial}{\partial\theta}F_{i|D} := \frac{\partial}{\partial\theta}F_{i|D}(u_i|\mathbf{u}_D)$ and the last rows are left empty for notational convenience. The terms in $S1^{direct,\theta}$ and $S1^{indirect,\theta}$ can be determined by differentiating (2.11) similarly as we did for the copula terms in (2.17) - (2.19). For instance, we have

$$\begin{aligned}\frac{\partial}{\partial\theta}F_{U|V,\mathbf{Z}}(u|v,\mathbf{z},\theta) &= \frac{\partial}{\partial\theta}\left(h_{U,V|\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\right)\\ &= \partial_1 h_{U,V|\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\cdot\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\\ &= c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\cdot\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),\end{aligned}$$

if the copula term depends on θ through the first argument and

$$\begin{aligned}&\frac{\partial}{\partial\theta}h_{U,V|\mathbf{Z}}\big(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\big) =\\ &\qquad = \partial_2 h_{U,V|\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right)\cdot\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta),\end{aligned}$$

if the copula term depends on θ through the second argument. The complete calculations required to obtain the derivative of the log-likelihood with respect to one copula parameter θ are performed in Algorithm 2.2.3.

Algorithm 2.2.3 Log-likelihood derivative with respect to the parameter $\theta^{\tilde{k},\tilde{i}}$.

The input of the algorithm is a d-dimensional R-vine matrix M with maximum matrix $\tilde{M}$ and parameter matrix $\boldsymbol{\theta}$, and a matrix C determined using Algorithm 2.2.2 for a parameter

$\theta^{\tilde{k},\tilde{i}}$ positioned at row $\tilde{k}$ and $\tilde{i}$ in the R-vine parameter matrix $\boldsymbol{\theta}$. Further, we assume the matrices V^{direct}, $V^{indirect}$ and V^{values} corresponding to one observation from the R-vine copula distribution, which have been determined during the calculation of the log-likelihood, to be given. The output will be the value of the first derivative of the copula log-likelihood for the given observation with respect to the parameter $\theta^{\tilde{k},\tilde{i}}$.

Set $z_1 = v^{direct}_{\tilde{k},\tilde{i}}$
Set $s1^{direct}_{k,i} := 0$, $s1^{indirect}_{k,i} := 0$, $s1^{values}_{k,i} := 0$, $i = 1, \dots, d; k = i, \dots, d$
if $m_{\tilde{k},\tilde{i}} == \tilde{m}_{\tilde{k},\tilde{i}}$ **then**
    Set $z_2 = v^{direct}_{\tilde{k},d-\tilde{m}_{\tilde{k},\tilde{i}}+1}$
else
    Set $z_2 = v^{indirect}_{\tilde{k},d-\tilde{m}_{\tilde{k},\tilde{i}}+1}$
end if
Set $s1^{direct}_{\tilde{k}-1,\tilde{i}} = \partial_{\theta^{\tilde{k},\tilde{i}}} h(z_1, z_2|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}})$
Set $s1^{indirect}_{\tilde{k}-1,\tilde{i}} = \partial_{\theta^{\tilde{k},\tilde{i}}} h(z_2, z_1|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}})$
Set $s1^{values}_{\tilde{k},\tilde{i}} = \frac{\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})}$
for $i = \tilde{i}, \dots, 1$ **do**
    for $k = \tilde{k} - 1, \dots, i + 1$ **do**
        if $c_{k,i} == 1$ **then**
            Set $z_1 = v^{direct}_{k,i}$, $\tilde{z}_1 = s1^{direct}_{k,i}$
            if $m_{k,i} == \tilde{m}_{k,i}$ **then**
                Set $z_2 = v^{direct}_{k,d-\tilde{m}_{k,i}+1}$, $\tilde{z}_2 = s1^{direct}_{k,d-\tilde{m}_{k,i}+1}$
            else
                Set $z_2 = v^{indirect}_{k,d-\tilde{m}_{k,i}+1}$, $\tilde{z}_2 = s1^{indirect}_{k,d-\tilde{m}_{k,i}+1}$
            end if
            if $c_{k+1,i} == 1$ **then**
                Set $s1^{values}_{k,i} = s1^{values}_{k,i} + \frac{\partial_1 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}_1$
                Set $s1^{direct}_{k-1,i} = s1^{direct}_{k-1,i} + \partial_1 h(z_1, z_2|\mathcal{B}^{k,i}, \theta^{k,i}) \cdot \tilde{z}_1$
                Set $s1^{indirect}_{k-1,i} = s1^{indirect}_{k-1,i} + \partial_2 h(z_2, z_1|\mathcal{B}^{k,i}, \theta^{k,i}) \cdot \tilde{z}_1$
            end if
            if $c_{k+1,d-m+1} == 1$ **then**
                Set $s1^{values}_{k,i} = s1^{values}_{k,i} + \frac{\partial_2 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}_2$
                Set $s1^{direct}_{k-1,i} = s1^{direct}_{k-1,i} + \partial_2 h(z_1, z_2|\mathcal{B}^{k,i}, \theta^{k,i}) \cdot \tilde{z}_2$
                Set $s1^{indirect}_{k-1,i} = s1^{indirect}_{k-1,i} + \partial_1 h(z_2, z_1|\mathcal{B}^{k,i}, \theta^{k,i}) \cdot \tilde{z}_2$
            end if

30: **end if**
31: **end for**
32: **end for**
33: **return** $\sum_{k,i=1,\dots,d} s1^{values}_{k,i}$

In particular, this algorithm allows to replace finite-differences based numerical maximization of R-vine likelihood functions with maximization based on the analytical gradient. In a numerical comparison study across different R-vine models in 5-8 dimensions this resulted in a decrease in computation time by a factor of 4-8.

2.2.3 Computation of the observed information

Based on the algorithm for the score function introduced in the previous section, we will now present an algorithm to determine the Hessian matrix corresponding to the R-vine log-likelihood function.

Again, the derivatives of conditional distribution functions and copula log-likelihoods will be stored in matrices. Considering a derivative with respect to bivariate copula parameters θ and γ associated with the vine, it is clear that the expressions for the derivatives of the log-densities in this case will contain second derivatives of the occurring h-functions. Thus, our algorithm will determine the following matrices:

$$S2^{direct,\theta,\gamma} = \begin{pmatrix} \dots & & & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{4|m_{6,5},m_{7,5},m_{8,5}} & & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{4|m_{7,5},m_{8,5}} & \frac{\partial}{\partial\theta\partial\gamma}F_{3|m_{7,6},m_{8,6}} & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{4|m_{8,5}} & \frac{\partial}{\partial\theta\partial\gamma}F_{3|m_{8,6}} & \frac{\partial}{\partial\theta\partial\gamma}F_{2|1} \\ \dots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.24}$$

$$S2^{indirect,\theta,\gamma} = \begin{pmatrix} \dots & & & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{m_{6,5}|m_{7,5},m_{8,5},4} & & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{m_{7,5}|m_{8,5},4} & \frac{\partial}{\partial\theta\partial\gamma}F_{m_{7,6}|m_{8,6},3} & \\ \dots & \frac{\partial}{\partial\theta\partial\gamma}F_{m_{8,5}|4} & \frac{\partial}{\partial\theta\partial\gamma}F_{m_{8,6}|3} & \frac{\partial}{\partial\theta\partial\gamma}F_{1|2} \\ \dots & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.25}$$

$$S2^{values,\theta,\gamma} = \begin{pmatrix} \cdots & & & & \\ \cdots & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{4,m_{6,5}|m_{7,5},m_{8,5}} & & & \\ \cdots & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{4,m_{7,5}|m_{8,5}} & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{3,m_{7,6}|m_{8,6}} & & \\ \cdots & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{4,m_{8,5}} & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{3,m_{8,6}} & \frac{\partial}{\partial\theta\partial\gamma}\varrho_{2,1} & \\ \cdots & & & & \end{pmatrix} \in \mathbb{R}^{8\times 8} \tag{2.26}$$

Here, $\frac{\partial}{\partial\theta\partial\gamma}F_{i|D} := \frac{\partial}{\partial\theta\partial\gamma}F_{i|D}(u_i|\mathbf{u}_D)$. Since not all entries in (2.15), (2.16) and (2.20) depend on both θ and γ, not all entries in (2.24) - (2.26) will be non-zero and required in the algorithm. Employing Algorithm 2.2.2 to obtain matrices C^θ and C^γ corresponding to the parameters θ and γ, respectively, we see that the second derivatives of all elements where the corresponding matrix entry of either C^θ or C^γ is zero clearly vanish. As before, the terms in $S2^{direct,\theta,\gamma}$ and $S2^{indirect,\theta,\gamma}$ can be determined by differentiating the conditional distribution functions (Equation (2.11)). The details of these calculations and the recursive algorithm to determine second derivatives of the log-likelihood function are given in Appendix A. A simulation study performed in Stöber and Schepsmeier (2013) illustrates that the standard errors and confidence intervals computed using the results of this algorithm are appropriate.

With the algorithms presented in this chapter the required calculations for an application of PCCs to real world data sets can be performed in statistical software. Before we develop and apply concrete statistical models, however, we will provide an in-depth discussion of the simplifying assumption for PCCs in the following chapter.

Chapter 3

The simplifying assumption

As we have noted in the introduction to PCCs (see Remark 2.1.1), a commonly made assumption to keep PCCs tractable for inference is that conditional copulas do not depend on the values of variables that are conditioned on. This will be discussed on more detail in this chapter, which is taken from Stöber et al. (2013).

The simplifying assumption was previously studied by Hobæk Haff et al. (2010), providing first illustrative examples, and by Acar et al. (2012a), demonstrating for a particular data set and choice of parametric families that the simplifying assumption can be restrictive. The question for which classes of multivariate models the assumption is applicable however has remained unsolved. In this chapter, we fill this gap by studying the conditional copulas of the general classes of Archimedean (Section 3.1) and elliptical (Section 3.2) copulas. Here, we will show that the only Archimedean copulas in dimension $d \geq 3$ which are of the simplified type are those based on the Gamma Laplace transform or its extension, while Student's t copulas are the only ones arising from a scale mixture of Normals. These results illustrate that from a theoretical perspective, the simplifying assumption can be very restrictive. To assess its relevance in practical applications, we will develop a technique to assess the distance of a multivariate distribution from a nearby distribution that satisfies the simplifying assumption in Section 3.3. We will also discuss the increased flexibility by keeping the simplifying assumption only for parts of the distribution and introducing for example time-varying parameters.

3.1 Archimedean copulas

In this section, we characterize the Archimedean copulas that are simplified PCCs. A d-dimensional Archimedean copula is given by

$$C_{1:d}(u_1, \ldots, u_d) = \varphi\Big(\sum_{j=1}^{d} \varphi^{-1}(u_j) \Big), \tag{3.1}$$

for an Archimedean generator function $\varphi \in \mathcal{L}_d$. Here, $\mathcal{L}_d$ is the class of d-monotone functions $\varphi : [0, \infty) \mapsto [0, 1]$, which are strictly decreasing on $[0, \inf\{s, \varphi(s) = 0\}]$ (with $\varphi(0) = 1$, $\varphi(\infty) = 0$) and differentiable on $[0, \infty)$ up to order $d - 2$, such that $(-1)^j \varphi^{(j)} \geq 0, j =$

$1, \ldots, d-2$, and where further $(-1)^{d-2}\varphi^{(d-2)}$ is non-increasing and convex (McNeil and Nešlehová 2009; Joe 1997; Müller and Scarsini 2005). These copulas have the appealing property that the conditional distribution function is given in a simple analytical form which facilitates the study of their properties. Let us consider a Gamma random variable with shape parameter p, rate parameter b, density

$$f_\Gamma(x; b, p) = \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx},$$

mean p/b and variance p/b^2. For $\theta = 1/b = 1/p$, the Laplace transform (LT) is

$$\varphi_\Gamma(s; \theta) = \int_0^\infty e^{-sx} f_\Gamma(x; 1/\theta, 1/\theta) dx = (1+\theta s)^{-1/\theta}, \quad \theta \geq 0.$$

The Archimedean copula corresponding to this LT as generator function is called MTCJ[1] copula. From the form of the conditional distributions derived in Takahasi (1965) it is obvious that the copulas corresponding to bivariate conditional margins of the MTCJ copula are again MTCJ copulas. In particular, the MTCJ copula is the copula of the multivariate Burr's distribution which is closed under conditioning. The above Gamma LT family extends to Archimedean generators in $\mathcal{L}_d$ (for the *generalized* MTCJ family)

$$\varphi_\Gamma(s; \theta) = (1+\theta s)_+^{-1/\theta}, \quad \theta \geq -\frac{1}{d-1}, \text{ where } (x)_+ = \max\{0, x\}. \tag{3.2}$$

Mesfioui and Quessy (2008) show that the conditional generator when conditioning on m variables becomes

$$\varphi_m(s; \theta) = \left(1 + \frac{s\theta}{m\theta + 1}\right)_+^{-(m\theta+1)/\theta}. \tag{3.3}$$

For $\theta \geq -1/(d-1)$, we have $\theta/(m\theta+1) \geq -1/(d-1-m)$, so the parameter range is consistent, and the (generalized) MTCJ copula is a simplified PCC where all bivariate building blocks are MTCJ with corresponding choice of parameters. In fact, it is the only multivariate Archimedean copula that constitutes a simplified PCC.

Theorem 3.1.1. *A d-dimensional Archimedean copula is a simplified PCC if and only if its generator is in the family (3.2).*

1. This copula is also called Clayton copula due to its appearance in (Clayton 1978). It is the copula of the multivariate Pareto distribution (Mardia 1962) and of the multivariate Burr distribution (Takahasi 1965). It was first mentioned as a multivariate copula in Cook and Johnson (1981) and as a bivariate copula in Kimeldorf and Sampson (1975). The extension to negative dependence was given in Genest and MacKay (1986) for the bivariate case and in Joe (1997, pp. 157-158) for the multivariate case. Since many properties were discovered studying the corresponding distribution function and Cook and Johnson (1981) mentioned its general form we call it MTCJ copula.

Proof. Because the proof is more intuitive, we first outline the case where φ is the LT of a positive random variable. In this case, there is a representation of the copula as a mixture of powers. The mixture distribution from which the Archimedean copula arises can be written as (cf. Joe (1997, p. 86)):

$$F_{1:d}(\mathbf{x}_{1:d}) = \int_0^\infty \prod_{j=1}^d [G(x_j)]^\alpha \, dF_A(\alpha) = \int_0^\infty \prod_{j=1}^d [G(x_j)]^\alpha \, f_A(\alpha) d\alpha,$$

where F_A is the cdf of a positive random variable A, with corresponding density f_A, and

$$F(x) = \int_0^\infty [G(x)]^\alpha dF_A(\alpha) = \varphi_A(-\ln G(x))$$

is the common univariate cdf with φ_A being the LT of A. Without loss of generality we can assume that $F(x) = x$ on $[0,1]$. Then, $F_{1:d}(\mathbf{x}_{1:d})$, $x_j \in [0,1]$, $j = 1, \ldots, d$ is a copula. Also $G(x) = \exp\{-\varphi_A^{-1}(x)\}$ on $[0,1]$ is differentiable. With $g = G'$, the marginal cdf of the last k variables and its density are:

$$\begin{aligned}
F_{(d-k+1):d}(\mathbf{x}_{(d-k+1):d}) &= \int_0^\infty \left[\prod_{i=d-k+1}^d G(x_i) \right]^\alpha f_A(\alpha) d\alpha, \\
f_{(d-k+1):d}(\mathbf{x}_{(d-k+1):d}) &= \frac{\partial^k}{\partial x_{d-k+1} \cdots \partial x_d} F_{(d-k+1):d}(\mathbf{x}_{(d-k+1):d}) \\
&= \prod_{j=d-k+1}^d \left[\frac{g(x_j)}{G(x_j)} \right] \int_0^\infty \alpha^k \left[\prod_{i=d-k+1}^d G(x_i) \right]^\alpha f_A(\alpha) d\alpha,
\end{aligned}$$

and the conditional cdf of the first $d-k$ variables given the last k is:

$$\begin{aligned}
F_{1:(d-k)|(d-k+1):d}(\mathbf{x}_{1:(d-k)}|\mathbf{x}_{(d-k+1):d}) &= \frac{\partial^k F(x_1, \ldots, x_d)}{\partial x_{d-k+1} \cdots \partial x_d} \Bigg/ f_{(d-k+1):d}(\mathbf{x}_{(d-k+1):d}) \\
&= \frac{\int_0^\infty \prod_{j=1}^{d-k} [G(x_j)]^\alpha \cdot \alpha^k \prod_{i=d-k+1}^d [G(x_i)]^\alpha \, f_A(\alpha) d\alpha}{\int_0^\infty \alpha^k \prod_{i=d-k+1}^d [G(x_i)]^\alpha \, f_A(\alpha) d\alpha} \\
&= \int_0^\infty \prod_{j=1}^{d-k} [G(x_j)]^\alpha \cdot f_{A^\star}(\alpha) d\alpha,
\end{aligned}$$

where

$$f_{A^\star}(\alpha) = \frac{\alpha^k \prod_{i=d-k+1}^d [G(x_i)]^\alpha \, f_A(\alpha)}{\int_0^\infty \beta^k \prod_{i=d-k+1}^d [G(x_i)]^\beta \, f_A(\beta) d\beta} \propto \alpha^k e^{\alpha \sum_{i=d-k+1}^d \ln(G(x_i))} f_A(\alpha).$$

In this case, the density of $A^\star$ has the same parametric form as the density of A if the density of A has parameters θ and η and can be expressed as

$$f_A(\alpha; \eta, \theta) = e^{-\alpha\eta} \alpha^{\theta-1} h(\alpha) / C(\eta, \theta), \tag{3.4}$$

where h is a positive-valued function (it is not absorbed in the $\alpha^{\theta-1}$ term only if it is a non-power function) and $C(\eta, \theta)$ is a (finite) normalizing constant. From above, $A^\star$ has the same parametric form with parameters $(\eta - \sum_{i=d-k+1}^{d} \ln G(x_i), \theta + k)$. The conditional copula does not depend on $x_{d-k+1}, \ldots, x_d$ only if η is a rate (or reciprocal scale) parameter, since Archimedean copulas are invariant to scale changes of the mixing distribution (Mai and Scherer 2012, p. 60). For η to be a rate or inverse scale parameter of (3.4) only, $h(\alpha)$ must be a power of α. Hence f_A is a Gamma density, and $F_{1:d}(\mathbf{x}_{1:d})$ is a MTCJ copula.

For the general case, where $\varphi \in \mathcal{L}_d$ is not necessarily a LT, we prove the result by construction of a functional equation: Let

$$C_{1:d}(u_1, \ldots, u_d) = \varphi\Big(\sum_{j=1}^{d} \varphi^{-1}(u_j)\Big)$$

be an Archimedean copula. Let $(U_1, \ldots, U_d)$ be a random vector associated with this distribution. Suppose φ has support $[0, s_0)$ where $s_0 = \inf\{s, \varphi(s) = 0\}$ is infinite for a Laplace transform, but could be finite for $\varphi \in \mathcal{L}_d$ which is not a Laplace transform. The case of finite support implies that $\varphi(s) = 0$ for $s \geq s_0$. Let $F_{1\cdots d-1|d}(u_1, \ldots, u_{d-1}|u_d) = \partial C(u_1, \ldots, u_d)/\partial u_d$ be the conditional distribution given $U_d = u_d$. We will show now that the copula for this is another Archimedean copula, say based on ψ, where $\psi \in \mathcal{L}_{d-1}$. By differentiation, with $h = -\varphi'$ and $a = \varphi^{-1}(u_d) \in [0, s_0)$ with $0 < u_d \leq 1$,

$$F_{1:(d-1)|d}(u_1, \ldots, u_{d-1}|u_d) = h\Big(\sum_{j=1}^{d} \varphi^{-1}(u_j)\Big) \Big/ h(a),$$

with jth $(1 \leq j \leq d-1)$ margin $F_{j|d}(u_j|u_d) = h(\varphi^{-1}(u_j) + a)/h(a) =: v_j$. Note that h is monotonically decreasing, continuous, and convex by definition of $\mathcal{L}_d$. Hence $\varphi^{-1}(u_j) = h^{-1}(v_j h(a)) - a$ for $1 \leq j \leq d-1$ and the copula of the conditional distribution of $U_1, \ldots, U_{d-1}$ given $U_d = u_d$ is:

$$C_{1:(d-1);d}(v_1, \ldots, v_{d-1}; a) = h\Big(\sum_{j=1}^{d-1} h^{-1}(v_j h(a)) - (d-2)a\Big). \Big/ h(a) \tag{3.5}$$

Defining $s := \psi^{-1}(v; a) = h^{-1}(vh(a)) - a$ and $v := \psi(s; a) = h(s + a)/h(a)$, this is a Archimedean copula

$$\psi(\psi^{-1}(v_1; a) + \cdots + \psi^{-1}(v_{d-1}; a); a)$$

with generator function $\psi(\cdot, a)$. As $u_d \to 1$, $a = \varphi^{-1}(u_d) \to 0$, and $h(0) = -\varphi'(0)$ can be positive or infinite but not 0, by the definition of φ. Also, h is differentiable except at a countable number of points with h' continuous and increasing except at the countable subset,

and right derivatives exist for all points where h is finite (Rockafellar 1970, Theorems 23.1 and 25.3).

Consider first the case where $h(0)$ is finite and positive, and $h'(0)$ is finite. If h is not differentiable everywhere, $h'(0)$ and all derivatives in the following can be considered as right derivatives. The copula of the conditional distribution does not depend on u_d or a if and only if there is a continuous differentiable scale function $\gamma(a) > 0$ such that $\gamma(0) = 1$ and

$$\psi(s; a) = h(s + a)/h(a) = \psi(s\gamma(a); 0) = h(s\gamma(a))/h(0); \quad 0 \le s < s_0.$$

Writing the above functional equation in h as $h(s+a)\,h(0) = h(a)\,h(s\gamma(a))$ and differentiating with respect to a yields

$$h'(s + a)\,h(0) = h'(a)\,h(s\gamma(a)) + h(a)\,h'(s\gamma(a))\,s\gamma'(a).$$

With $a = 0$ if follows that

$$h'(s)\,h(0) = h'(0)\,h(s) + h(0)\,h'(s)\,s\gamma'(0).$$

Rewriting this equation as

$$h'(s) = \frac{h'(0)}{h(0)[1 - s\gamma'(0)]}h(s),$$

we conclude that the right derivative h' is continuous and thus h is differentiable. The above differential equation has solution $h(s) = h(0)[1-s\gamma'(0)]^{\alpha}$ where $\alpha = -h'(0)/[h(0)\gamma'(0)]$. Since $h = -\varphi'$ must be decreasing, there are 2 possibilities

(i) $s_0 = \infty$, $\gamma'(0) < 0$, $\alpha < 0$, or

(ii) $\gamma'(0) > 0$, $s_0 = 1/\gamma'(0)$, $\alpha > 0$.

By integrating h over s we obtain $\varphi(s) = (1-s\gamma'(0))_+^{1+\alpha}$, since $\psi(0) = 0$. In case (i), we must have $\alpha < -1$ and in case (ii), $1 + \alpha \ge d - 1$ in order for $\psi \in \mathcal{L}_d$ (see also Joe (1997, pp. 157-158)). In case (i) the obtained generating function has support on $[0, \infty)$ and corresponds to the "standard" MTCJ copula, whereas case (ii) with bounded φ yields the generalized MTCJ copula.

If $h'(0)$ or $h(0)$ is infinite, the above is modified as follows. Let $0 < \epsilon < s_0$. There is a continuous differentiable scale function $\gamma(a) > 0$ such that $\gamma(\epsilon) = 1$ and

$$\psi(s; a) = h(a + s)/h(a) = \psi(s\gamma(a); \epsilon) = h(s\gamma(a) + \epsilon)/h(\epsilon).$$

Cross-multiplying and differentiating the above with respect to a, and then setting a to ϵ yields

$$h'(\epsilon)\,h(s + \epsilon) = h(\epsilon)[1 - s\gamma'(\epsilon)]h'(s + \epsilon), \quad 0 < s < s_0 - \epsilon.$$

This has solution $h(s+\epsilon) = h(\epsilon)[1 - s\gamma'(\epsilon)]^\alpha$ where $\alpha = -h'(\epsilon)/[h(\epsilon)\gamma'(\epsilon)]$ so that $\psi(s;\epsilon) = h(s+\epsilon)/h(\epsilon) = [1 - s\gamma'(\epsilon)]^\alpha$. The conclusion is the same as above, because after integrating h to get φ, one would conclude that this leads to $\varphi'(0)$ and $\varphi''(0)$ being finite. □

This adds a further aspect making the MTCJ copula unique[2] among Archimedean copulas. Using different parameters than those obtained using Equation (3.3) in a PCC setup, i.e. PCCs with MTCJ copulas with freely chosen parameters as building blocks, we obtain a natural extension of the MTCJ copula, permitting different dependence for different bivariate marginals.

3.2 Elliptical copulas

In this section, we characterize the elliptical copulas that have all conditional distributions in location-scale families. The location scale family of a random vector $\mathbf{X} \in \mathbb{R}^d$ is the set $\left\{a\mathbf{X} + \mathbf{b} | a \in \mathbb{R}_0^+, \mathbf{b} \in \mathbb{R}^d\right\}$. We show that not all elliptical copulas are simplified PCCs and characterize the scale mixtures of Normals which are of the simplified type. By referring to a d-dimensional elliptical copula we mean a copula arising from an elliptical distribution such as the multivariate Normal or Student's t distribution. Following Cambanis et al. (1981), a multivariate distribution is elliptical if its characteristic function has the form

$$\varphi_{\mathbf{X}}(\mathbf{t}; \boldsymbol{\mu}, \Sigma) = \Psi(\mathbf{t}'\Sigma\mathbf{t})e^{i\mathbf{t}'\boldsymbol{\mu}},$$

for $\Psi : \mathbb{R}_0^+ \mapsto \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}^d$, and positive definite $\Sigma \in \mathbb{R}^{d\times d}$. If the distribution has a density, this implies that it is given by

$$f_{\mathbf{X}}(\mathbf{x}) = |\Sigma|^{-1/2} g\Big((\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\Big),$$

for a generator function $g : \mathbb{R}_0^+ \mapsto \mathbb{R}_0^+$, which can be uniquely determined from Ψ. For the two examples we mentioned, the generator functions have the form

$$g_{\text{Gauss},d}(t) = \frac{1}{(2\pi)^{d/2}} e^{-t/2}, \quad g_{\text{Student's}-t,d,\nu}(t) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{d/2}} \cdot \left(1 + \frac{t}{\nu}\right)^{-(\nu+d)/2},$$

and lead to simplified PCCs.

Theorem 3.2.1. *The multivariate Gaussian distribution and the multivariate Student's t distribution are PCCs of the simplified form.*

2. The MTCJ copula also is the only Archimedean copula invariant under truncation, in the sense that for a rv $\mathbf{U} \sim C$, C also is the copula of $\mathbf{U}|\mathbf{U} \leq \mathbf{a}$, $\mathbf{a} \in [0,1]^{\dim(\mathbf{U})}$ (Ahmadi Javid 2009).

For the Gaussian distribution this is quite obvious since all conditional distributions are again Gaussian and do depend on the values of the variables that they are conditioned on only through their mean, i.e., they all are in the same location family. For Student's t distribution the covariances also depend on these values, but only through a scaling factor such that all conditional distributions remain in the same location-scale family. Since changes of location and scale do not affect the copula of a multivariate distribution, this implies that Student's t distribution is a simplified PCC. For clarification, we derive the explicit form of the copulas corresponding to bivariate conditional distributions in Appendix C. Just as for the MTCJ copula, changing the degrees of freedom in the bivariate building blocks of the PCC leads to a natural extension of the multivariate Student's t distribution in which different bivariate conditional margins can have different degrees of freedom.

While the copulas corresponding to these distributions are the most popular examples they also have a unique position within the class of elliptical distributions as the following theorems show.

Theorem 3.2.2. *Let us assume that the generator function $g(\cdot)$ of the density of a d-dimensional elliptical distribution is differentiable. Then, the conditional distributions remain within the same location-scale family for all values of the variables that are conditioned on, if and only if*

(a) the support of g is $\mathbb{R}$ and the distribution is the multivariate Student's t (or Pearson type VII) distribution or in its limiting case the multivariate Normal distribution or

(b) g has compact support and $g(t;\zeta) = (1-t)_+^{\zeta}$, for $\zeta > 1$ up to rescaling (the Pearson type II distribution which is sometimes called "inverted t distribution" (Dickey 1967)).

Proof. The proof is similar to that for Archimedean copulas in that the same functional equation can be obtained. Without loss of generality, let us consider the case of a d-dimensional elliptical distribution with zero means and zero correlations. In this case, for some generator functions g_d and g_k, the density of the distribution is given by

$$f_{1:d}(\mathbf{x}_{1:d}) = g_d(x_1^2 + \ldots + x_d^2), \quad \text{with marginal density}$$
$$f_{(d-k+1):d}(\mathbf{x}_{(d-k+1):d}) = g_k(x_{d-k+1}^2 + \ldots + x_d^2).$$

For $f_{1:(d-k)|(d-k+1):d}(\mathbf{x}_{1:(d-k)}|\mathbf{x}_{(d-k+1):d})$ we obtain

$$f_{1:(d-k)|(d-k+1):d}(\mathbf{x}_{1:(d-k)}|\mathbf{x}_{(d-k+1):d}) = \frac{g_d(x_1^2 + \ldots + x_d^2)}{g_k(x_{d-k+1}^2 + \ldots + x_d^2)}.$$

For this distribution to be in the same location-scale family irrespective of the values of $x_{d-k+1}, \ldots, x_d$, we must have that for given $\sum_{i=d-k+1}^{d} x_i^2 \neq \sum_{i=d-k+1}^{d} x_i^{\star 2}$ there exists $\gamma(\mathbf{x}_{(d-k+1):d}, \mathbf{x}^{\star}_{(d-k+1):d})$ such that

$$f_{1:(d-k)|(d-k+1):d}(\gamma(\mathbf{x}_{(d-k+1):d}, \mathbf{x}^{\star}_{(d-k+1):d}) \cdot \mathbf{x}_{1:(d-k)} | \mathbf{x}^{\star}_{(d-k+1):d}) \\ \propto f_{1:(d-k)|(d-k+1):d}(\mathbf{x}_{1:(d-k)} | \mathbf{x}_{(d-k+1):d}).$$

With $\mathbf{x}^{\star}_{(d-k+1):d} = \mathbf{0}$, $a = \sum_{i=d-k+1}^{d} x_i^2$, $t = \sum_{i=1}^{d-k} x_i^2$ and $\delta(a) = 1/\gamma(\mathbf{x}_{(d-k+1):d}, \mathbf{0})$ this implies that

$$g_d(t+a) = \xi(a) \cdot g_d\left(\frac{t}{\delta(a)}\right), \tag{3.6}$$

where $\xi(a)$ equals $g_k(a)$ times a constant depending on a, and $\xi(\cdot)$, $\delta(\cdot)$ are differentiable scale functions. Since $g_d(0) = 0$ if and only if $g_d(a) = 0$ for all values of a, we must have $g_d(0) > 0$. Using $t = 0$ in Equation 3.6, we conclude that $g_d(0)$ is finite. Thus, we can define $h(t) := \frac{g_d(t)}{g_d(0)}$ and obtain from Equation 3.6 that

$$h(t+a) = h(a) \cdot h\left(\frac{t}{\delta(a)}\right).$$

Using that $\delta(0) = 1$ by the definition of δ, differentiation with respect to a yields

$$h'(t+a) = h'(a) \cdot h\left(\frac{t}{\delta(a)}\right) + h(a) \cdot h'\left(\frac{t}{\delta(a)}\right) \cdot \left(\frac{-t\delta'(a)}{\delta(a)^2}\right),$$

and for $a = 0$ this implies

$$h'(t) = h'(0) \cdot h(t) + h'(t) \cdot (-t \cdot \delta'(0)).$$

In other words, the function h must fulfill the differential equation

$$(1 + \beta t) h'(t) = \alpha h(t),$$

where $\alpha = h'(0)$, $\beta = \delta'(0)$. From this, we obtain for $\beta > 0$ that $h(t) = (1 + \beta t)^{\alpha/\beta}$ which corresponds to the elliptical generator of a Pearson type VII (scaled Student's t) distribution. For h to yield a well defined density in d dimensions, $\frac{\alpha}{\beta}$ must be given in the form $\frac{\alpha}{\beta} = -(\nu + d)/2$, $\nu > 0$, to ensure integrability with respect to $t^{d/2-1}$.

For $\beta < 0$, $h(t) = (1 - \beta t)_+^{\alpha/\beta}$, which leads to a well-defined density for $\frac{\alpha}{\beta} > -1$ and is differentiable and thus a valid solution for $\frac{\alpha}{\beta} > 1$.

By integration, we obtain that the generator function for lower-dimensional margins $g_{d-l}(t)$ is proportional to $(1 - \beta t)_+^{\alpha/\beta + l/2}$. Therefore, also the conditional distributions of the lower-dimensional margins remain within the same location-scale family for all values of the conditioning variables.

□

Here, case (b) can be seen as an analogue to the extension of the MTCJ to negative dependence. The proof that Student's t distribution is a simplified PCC relied on the fact that in this case, conditioning only affects the location and scale of the distribution. Theorem 3.2.2 shows that the t-distribution is the only elliptical distribution where this proof strategy is successful. Just as the MTCJ copula can be constructed using a Gamma mixture, also Student's t distribution arises from the multivariate Normal distribution using a Gamma mixture for the square of the inverse scale parameter. For the distribution function of the MTCJ copula, we have

$$C_{MTCJ}(\mathbf{u}_{1:d};\theta) = \int_0^\infty \prod_{i=1}^d \left[G_{MTCJ}(u_i;\theta)\right]^\alpha f_\Gamma(\alpha;1,1/\theta)d\alpha,$$

where $G_{MTCJ}(\cdot;\theta) = \exp\left\{-\left(\varphi_\Gamma\right\}^{-1}(\cdot;\theta)\right)$ is a cdf on $[0,1]$, cf. Joe (1997, p. 86), while

$$f_{t,d}(\mathbf{x}_{1:d};R,\nu) = (2\pi)^{-d/2}|R|^{-1/2}\int_0^\infty w^{d/2}\exp\{-\tfrac{1}{2}w\mathbf{x}'R^{-1}\mathbf{x}\}f_\Gamma(w;\nu/2,2/\nu)\,dw,$$

holds for the density of a multivariate Student's t distribution (Cornish 1954). When only considering scale mixtures of Normals, i.e. mixtures of random variables of the form $a\mathbf{X}$, where $a \in \mathbb{R}_0^+$ and $\mathbf{X}$ has a multivariate normal distribution, we obtain the following theorem.

Theorem 3.2.3. *Consider a d-dimensional scale mixture of Normals with correlation matrix Σ which is a simplified PCC*

(a) in $d \geq 4$ or

(b) for all positive definite correlation matrices Σ,

then the mixing distribution is the Gamma distribution.

Proof. A general scale mixture of Normals in dimension d can be written as $(X_1,\ldots,X_d) = (Z_1,\ldots,Z_d)/\sqrt{W}$ where W is a random variable on $(0,\infty)$ with density f_W, and $(Z_1,\ldots,Z_d)$ is multivariate Gaussian with zero mean vector and covariance matrix Σ. Without loss of generality, we can assume that all diagonal entries of Σ are 1, i.e., Σ is a correlation matrix. This implies for the d-variate generator g_d of the distribution of $(X_1,\ldots,X_d)$ that

$$|\Sigma|^{-1/2}g_d(\mathbf{x}'\Sigma^{-1}\mathbf{x}) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\int_0^\infty w^{d/2}\exp\{-\tfrac{1}{2}w\mathbf{x}'\Sigma^{-1}\mathbf{x}\}f_W(w)\,dw.$$

Similarly, if Σ_k is the leading $k\times k$ matrix of Σ, and $\mathbf{x}_k = (x_1,\ldots,x_k)'$, then by marginalizing out the last $d-k$ components we get

$$|\Sigma_k|^{-1/2}g_k(\mathbf{x}_k'\Sigma_k^{-1}\mathbf{x}_k) = (2\pi)^{-k/2}|\Sigma_k|^{-1/2}\int_0^\infty w^{k/2}\exp\{-\tfrac{1}{2}w\mathbf{x}_k'\Sigma_k^{-1}\mathbf{x}_k\}f_W(w)\,dw.$$

Hence, the univariate margin is

$$f_1(x_1) = g_1(x_1^2) = (2\pi)^{-1/2} \int_0^\infty w^{1/2} \exp\{-\tfrac{1}{2} w x_1^2\} f_W(w)\, dw.$$

In the equation above, the density f_W could be replaced with dF_W in a Stieltjes integral. However, this would not affect the remainder of this proof; we omit it for notational convenience. Note that

$$g_1(0) = (2\pi)^{-1/2} \mathrm{E}\,[W^{1/2}], \quad g_d(0) = (2\pi)^{-d/2} \mathrm{E}\,[W^{d/2}].$$

Let G be the cdf corresponding to the marginal generator g_1. Then, the copula density corresponding to g_d is

$$c_{1:d}(u_1, \ldots, u_d; g_d, \Sigma) = |\Sigma|^{-1/2} \frac{g_d(\mathbf{x}'\Sigma^{-1}\mathbf{x})}{g_1([G^{-1}(u_1)]^2) \cdots g_1([G^{-1}(u_d)]^2)} \tag{3.7}$$

with $x_j = G^{-1}(u_j)$. From this general form of the copula density, we will obtain two equations for the moments of the mixing variable W which will lead to necessary conditions on the conditional distributions of simplified PCCs in the elliptical class. Directly from (3.7) we get that

$$c_{1:d}(0.5, \ldots, 0.5; g_d) = |\Sigma|^{-1/2} \frac{g_d(0)}{g_1^d(0)} = |\Sigma|^{-1/2} \frac{\mathrm{E}\,[W^{d/2}]}{\mathrm{E}^{\,d}[W^{1/2}]}. \tag{3.8}$$

This means that if W_1, W_2 are two different mixing variables, then the copula densities with fixed Σ are different unless the necessary condition of

$$\frac{\mathrm{E}\,[W_1^{d/2}]}{\mathrm{E}^{\,d}[W_1^{1/2}]} = \frac{\mathrm{E}\,[W_2^{d/2}]}{\mathrm{E}^{\,d}[W_2^{1/2}]}$$

holds. To derive a second equation, let us consider

$$c_{1:d}(0.5, \ldots, 0.5, u; g_d) = |\Sigma|^{-1/2} \frac{g_d(\alpha x^2)}{g_1^{d-1}(0) g_1(x^2)} \tag{3.9}$$

where $\alpha \geq 1$ is the (d,d) element of Σ^{-1} and $x = G^{-1}(u)$. Note that α is not a scaling factor of the marginal distribution but a function of the correlation matrix. In particular, we can obtain all $\alpha \geq 1$ from correlation matrices of the form

$$\Sigma_\alpha = \begin{pmatrix} \mathbb{I}_{d-2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \sqrt{1 - \frac{1}{\alpha}} \\ \mathbf{0} & \sqrt{1 - \frac{1}{\alpha}} & 1 \end{pmatrix},$$

where $\mathbb{I}_{d-2}$ is the $(d-2) \times (d-2)$ identity matrix. Using $\partial x / \partial u = 1/g_1(x^2)$, we determine the first derivative of (3.9) with respect to u as

$$|\Sigma|^{-1/2} \left[\frac{\alpha g_d'(\alpha x^2)}{g_1^{d-1}(0) g_1(x^2)} - \frac{g_d(\alpha x^2) g_1'(x^2)}{g_1^{d-1}(0) g_1^2(x^2)} \right] \cdot \frac{2x}{g_1(x^2)}.$$

After taking the second derivative with respect to u and then setting $x = 0$ ($u = 0.5$), all terms are 0 except

$$\begin{aligned}
&\lim_{x\to 0}|\Sigma|^{-1/2}\left[\frac{ag_d'(\alpha x^2)}{g_1^{d-1}(0)g_1(x^2)} - \frac{g_d(\alpha x^2)g_1'(x^2)}{g_1^{d-1}(0)g_1^2(x^2)}\right]\cdot\frac{2}{g_1^2(x^2)} \\
&\quad = 2|\Sigma|^{-1/2}\left[\frac{\alpha g_d'(0)}{g_1^{d+2}(0)} - \frac{g_d(0)g_1'(0)}{g_1^{d+3}(0)}\right] \\
&\quad = (2\pi)^{-1}|\Sigma|^{-1/2}\left[-\frac{\alpha \mathrm{E}\,(W^{(d+2)/2})}{\mathrm{E}^{\,d+2}(W^{1/2})} + \frac{\mathrm{E}\,(W^{d/2})\mathrm{E}\,(W^{3/2})}{\mathrm{E}^{\,d+3}(W^{1/2})}\right].
\end{aligned} \tag{3.10}$$

Let us now consider the analog for conditional densities. Let $f_{1\cdots d}(\mathbf{x})$ be the density of $(X_1,\ldots,X_d)$ and let f_1 be the density of X_1 or any X_j. Let

$$f_{1:(d-1)|d}(x_1,\ldots,x_{d-1}|x_d) = f_{1:d}(\mathbf{x})/f_1(x_d).$$

For this, we decompose $\mathbf{x}'\Sigma^{-1}\mathbf{x}$ as $(\mathbf{x}^*)'\Sigma_{11\cdot 2}^{-1}\mathbf{x}^* + x_d^2$ where $\mathbf{x}^* = (x_1 - \sigma_{1d}x_d,\ldots,x_{d-1} - \sigma_{d-1,d}x_d)'$ and $\Sigma_{11\cdot 2}$ is the conditional covariance matrix of $(Z_1,\ldots,Z_{d-1})$ given Z_d. Writing the conditional densities in mixture form,

$$\begin{aligned}
f_{1:d}(\mathbf{x}) &= \int_0^\infty w^{d/2}\phi_d(\mathbf{x}w;R)\,f_W(w)\,dw \\
&= (2\pi)^{-d/2}|\Sigma|^{-1/2}\int_0^\infty w^{d/2}\exp\{-\tfrac{1}{2}w\mathbf{x}'\Sigma^{-1}\mathbf{x}\}f_W(w)\,dw \\
&= (2\pi)^{-d/2}|\Sigma_{11\cdot 2}|^{-1/2}\int_0^\infty w^{d/2}\exp\{-\tfrac{1}{2}w\mathbf{x}^{*\prime}\Sigma_{11\cdot 2}^{-1}\mathbf{x}^*\}\exp\{-\tfrac{1}{2}wx_d^2\}f_W(w)\,dw,
\end{aligned}$$

and

$$f_1(x_d) = (2\pi)^{-1/2}\int_0^\infty w^{1/2}\exp\{-\tfrac{1}{2}wx_d^2\}f_W(w)\,dw,$$

so that

$$\begin{aligned}
&f_{1:(d-1)|d}(x_1,\ldots,x_{d-1}|x_d) = \\
&a(x_d)(2\pi)^{-(d-1)/2}|\Sigma_{11\cdot 2}|^{-1/2}\int_0^\infty w^{(d-1)/2}\exp\{-\tfrac{1}{2}w\mathbf{x}^{*\prime}\Sigma_{11\cdot 2}^{-1}\mathbf{x}^*\}f_{W^*}(w)\,dw
\end{aligned} \tag{3.11}$$

is a scale mixture with mixing density $f_{W^*}(w;x_d) = w^{1/2}\exp\{-\frac{1}{2}wx_d^2\}f_W(w)/a(x_d)$, where $a(x_d)$ is a normalizing constant. We denote the random variable with this density by $W^*(x_d)$.

For $d \geq 3$, Equations (3.8) and (3.11) imply that a necessary condition for the copula corresponding to the distribution of $X_1,\ldots,X_{d-1}$ given $X_d = x_d$ to be independent of x_d is that

$$\begin{aligned}
&\frac{\int_0^\infty w^{d/2}\exp\{-\frac{1}{2}wx_d^2\}f_W(w)\,dw\cdot\left(\int_0^\infty w^{1/2}\exp\{-\frac{1}{2}wx_d^2\}f_W(w)\,dw\right)^{d-2}}{\left(\int_0^\infty w\exp\{-\frac{1}{2}wx_d^2\}f_W(w)\,dw\right)^{d-1}} \\
&\qquad = \frac{\mathrm{E}\,[\{W^*(x_d)\}^{(d-1)/2}]}{\mathrm{E}^{\,d-1}[W^{*1/2}(x_d)]}
\end{aligned} \tag{3.12}$$

is a constant over x_d. Similarly, we obtain from (3.10) that

$$|\Sigma|^{-1/2}\left[-\frac{\alpha \mathrm{E}\,(W^*(x_d)^{(d+1)/2})}{\mathrm{E}^{\,d+1}(W^*(x_d)^{1/2})}+\frac{\mathrm{E}\,(W^*(x_d)^{(d-1)/2})\mathrm{E}\,(W^*(x_d)^{3/2})}{\mathrm{E}^{\,d+2}(W^*(x_d)^{1/2})}\right], \tag{3.13}$$

must be equal to a constant $\beta(\alpha)$. To rewrite these equations, let $t=\frac{1}{2}x_d^2\geq 0$ and let V be a random variable with density $f_V(v)$ proportional to $v^{1/2}f_W(v)$. Let V_t be a random variable with density proportional to $e^{-tv}f_V(v)$; this is a Laplace transform tilt of the density of V with normalizing constant $\varphi_V(t)$, the LT of V at t. Note that V_t has finite positive integer moments for $t>0$ and $V_0=V$.

Then, (3.12) can be rewritten as

$$\frac{\int_0^\infty v^{(d-1)/2}\exp\{-vt\}f_V(v)\,dv\cdot\Big(\int_0^\infty \exp\{-vt\}f_V(v)\,dv\Big)^{d-2}}{\Big(\int_0^\infty v^{1/2}\exp\{-vt\}f_V(v)\,dv\Big)^{d-1}}=\frac{\mathrm{E}\,[V_t^{(d-1)/2}]}{\mathrm{E}^{\,d-1}[V_t^{1/2}]}, \tag{3.14}$$

which must be constant over $t\geq 0$, while (3.13) leads to

$$|\Sigma|^{-1/2}\left[-\frac{\alpha \mathrm{E}\,(V_t^{(d+1)/2})}{\mathrm{E}^{\,d+1}(V_t^{1/2})}+\frac{\mathrm{E}\,(V_t^{(d-1)/2})\mathrm{E}\,(V_t^{3/2})}{\mathrm{E}^{\,d+2}(V_t^{1/2})}\right], \tag{3.15}$$

which, for all $t\geq 0$, must be equal to a constant $\beta(\alpha)$. This implies the following recursive relationship for the moments of V_t: if we know that (3.14) is constant for $d=k$ and $d=4$, then it is also constant for $d=k+2$. Thus, it is sufficient to show that (3.14) is constant for $d=3$ and $d=4$, or $d=3$ and $d=5$.

Let us now consider case a), where the copula C is a simplified PCC in $d\geq 4$. From the three-dimensional marginal distributions, we obtain that (3.14) is constant for $d=3$. By conditioning on $X_4=x_4, X_3=x_3$, we conclude with a similar calculation as for (3.11) that

$$\frac{E[V_t^2]\cdot E^2[V_t^{1/2}]}{E^3[V_t]}=const.$$

This, together with (3.14) being constant for $d=3$ implies that (3.14) is constant for $d=5$ and thus for all $d\geq 3$.

In case b), where the copula is a simplified PCC for all Σ, we know that for a three-dimensional marginal distribution, (3.15) holds for all $\alpha\geq 0$:

$$-\frac{\alpha \mathrm{E}\,[\{W^*(x_3)\}^{4/2}]}{\mathrm{E}^{\,4}[W^{*1/2}(x_3)]}+\frac{\mathrm{E}\,[\{W^*(x_3)\}^{2/2}]\mathrm{E}\,[\{W^*(x_3)\}^{3/2}]}{\mathrm{E}^{\,5}[W^{*1/2}(x_3)]}=\beta(\alpha)|\Sigma|^{1/2}.$$

Thus, $\mathrm{E}\,[V_t^2]/\mathrm{E}^{\,4}[V_t^{1/2}]$ and $\mathrm{E}\,[V_t]\mathrm{E}\,[V_t^{3/2}]/\mathrm{E}^{\,5}[V_t^{1/2}]$ must be constants over t. Together with (3.14) being constant in t for $d=3$, this means that (3.14) is constant for $d=4$ and thus for

all $d \geq 3$. Note that, more precisely, we only require two different values of α in (3.15) for the argument above.

Summing up, we obtain that with $m = (d-1)/2$ the moments of V_t are connected to the moments of V via

$$\mathrm{E}\,[V_t^m] = a^m(t)\mathrm{E}\,[V^m]$$

for all $t > 0$, $m = 1, 1.5, 2, 2, 5, \ldots$, where $a(t) = \mathrm{E}^{\,2}[V_t^{1/2}]/\mathrm{E}^{\,2}[V^{1/2}]$.

With all of the positive integer moments of V and V_t existing, the Laplace transforms of V and V_t, for $0 \leq s \leq s_t$, where the constant s_t may depend on t, can be written as

$$\varphi_V(s) = 1 + \sum_{i=1}^{\infty}(-1)^i\mathrm{E}\,[V^i]s^i/i!,$$

$$\varphi_{V_t}(s) = 1 + \sum_{i=1}^{\infty}(-1)^i\mathrm{E}\,[V_t^i]s^i/i! = 1 + \sum_{i=1}^{\infty}(-1)^i\mathrm{E}\,[V^i]s^i a^i(t)/i!.$$

Hence $\varphi_{V_t}(s) = \varphi_V(sa(t))$ in a neighborhood of 0 for the Taylor series expansion of the LTs about 0. By Feller (1971, Section VII.6), the Taylor series in a positive neighborhood of 0 uniquely determines the distribution. Hence $V_t = a(t)V$ for $t > 0$. That Laplace transform tilting of the density leads to a scale-changed random variable, implies that V has a Gamma density (Marshall and Olkin 2007, p. 576, Theorem 18.B.6). Hence, also W is Gamma distributed, and the corresponding scale mixture is the multivariate t-distribution. □

In particular, for other scale mixtures of Normals the generator function of conditional distributions will depend on the value of variables that they are conditioned on. A simple example for such a distribution would be a two point mixture of Normal distributions having the same correlation matrix, but where $\Sigma_1 = \gamma\Sigma_2$ for a positive constant γ. This contradicts the claim made in Example 4.1 of Hobæk Haff et al. (2010) that all elliptical distributions are simplified PCCs.

It turns out that this deviation from the simplifying assumption cannot be detected by looking at conditional correlations or the popular Kendall's τ measure only. Cambanis et al. (1981) showed that the conditional correlation coefficient is equal to the partial correlation for elliptical distributions, and Lindskog et al. (2003) demonstrated that the relationship between the correlation coefficient ρ and Kendall's τ for the Normal distribution holds for all atom-free elliptical distributions.

$$\tau = \frac{2}{\pi}\arcsin(\rho).$$

This means that, for elliptical distributions, the values of Kendall's τ corresponding to bivariate conditional distributions are independent of the variables that are conditioned on.

3.3 Effects of the simplifying assumption

In this section we consider practical implications of the theoretical limitations of simplified PCCs. We illustrate how simplified PCCs approximate general distributions and discuss how they can be extended when the simplifying assumption is inappropriate.

3.3.1 Trivariate extension of the Farlie-Gumbel-Morgenstern copula

Even though it is not a plausible model for data, we start by studying the trivariate extension of the Farlie-Gumbel-Morgenstern (FGM) copula, which is given by

$$C_{1:3}(u_1, u_2, u_3; \theta) = u_1 u_2 u_3[1 + \theta(1 - u_1)(1 - u_2)(1 - u_3)],$$

for $\theta \in [-1; 1]$. This example is illustrative because here results can be obtained analytically. For $u_j = 1$, $j \in \{1, 2, 3\}$, the term in brackets is equal to 1, thus all bivariate marginal distributions are independence copulas. This is a classical example of a distribution where bivariate measures of dependence fail to detect the dependence. The density function of this copula family is

$$c_{1:3}(u_1, u_2, u_3; \theta) = 1 + \theta(1 - 2u_1)(1 - 2u_2)(1 - 2u_3).$$

Since the bivariate marginals are independence copulas, this is also the form of the conditional distribution of (U_1, U_2) given $U_3 = u_3$ and the corresponding copula, i.e.

$$f_{1,2|3}(u_1, u_2|u_3; \theta) = c_{1,2;3}(u_1, u_2|u_3; \theta) = c_{1:3}(u_1, u_2, u_3; \theta).$$

This is a bivariate FGM copula with parameter $\eta(u_3) = \theta(1 - 2u_3)$. To approximate this 3-dimensional copula using a simplified PCC, we follow the suggestions by Hobæk Haff et al. (2010). In a first step, we match the bivariate marginal distributions, i.e. C_{12} and C_{23} are independence copulas. Subsequently, the conditional copula $C_{12;3}$ will be approximated by a copula of the same parametric family

$$c_{FGM}(u_1, u_2|\eta) = 1 + \eta(1 - 2u_1)(1 - 2u_2),$$

but with constant parameter η minimizing the expected Kullback-Leibler (KL) distance from the true distribution:

$$\hat{\eta} = \text{argmin}_\eta KL_\theta(\eta), \text{ where}$$

$$\begin{aligned} KL_\theta(\eta) &= \int_{[0;1]^3} c_{1,2;3}(u_1, u_2|u_3) \ln\left(\frac{c_{1,2;3}(u_1, u_2|u_3)}{c_{FGM}(u_1, u_2|\eta)}\right) du_1 du_2 du_3 \\ &= \int_{[0;1]^3} c_{1,2;3}(u_1, u_2|u_3) \ln\left(\frac{1 + \theta(1 - 2u_1)(1 - 2u_2)(1 - 2u_3)}{1 + \eta(1 - 2u_1)(1 - 2u_2)}\right) du_1 du_2 du_3. \end{aligned}$$

. Evaluating the partial derivatives with respect to η,

$$\partial_\eta KL_\theta(\eta) = -\int_{[0;1]^3} \left(1+\theta\prod_{i=1}^{3}(1-2u_i)\right)\frac{(1-2u_1)(1-2u_2)}{1+\eta(1-2u_1)(1-2u_2)}du_1du_2du_3,$$

$$\partial_\eta^2 KL_\theta(\eta) = \int_{[0;1]^3} \left(1+\theta\prod_{i=1}^{3}(1-2u_i)\right)\frac{(1-2u_1)^2(1-2u_2)^2}{(1+\eta(1-2u_1)(1-2u_2))^2}du_1du_2du_3,$$

we obtain that $\partial_\eta KL_\theta(\eta)|_{\eta=0} = 0$ and $\partial_\eta^2 KL_\theta(\eta) > 0$, i.e. the minimum of the KL distance is attained for $\eta = 0$. This means that, in this case, the closest approximation is the 3-dimensional independence copula. The average conditional copula $C^\star_{1,2;3}$, $C^\star_{1,2;3}(u_1,u_2) = \int_0^1 C_{1,2;3}(u_1,u_2|u_3)du_3$ is also the independence copula. The quality of the simplified approximation can be assessed in terms of the KL distance. To convert the KL distance between two copulas into a more interpretable measure, we can consider the sample size needed to distinguish between the two models using a likelihood ratio statistic at confidence level $1-\alpha = 0.95$. For an approximating copula $c^\star_{1:3}$ the KL distance to the true model $c_{1:3}$ is

$$\Delta(c^\star_{1:3}; c_{1:3}) = \int_{[0;1]^3} c_{1,2,3}(u_1,u_2,u_3)\ln\left(\frac{c_{1,2,3}(u_1,u_2,u_3)}{c^\star_{1,2,3}(u_1,u_2,u_3)}\right)du_1du_2du_3 \geq 0.$$

Assuming data $(\mathbf{u}_i)_{i=1,\dots,N}$ from C_{123} to be given, an estimator for $\Delta(c^\star_{1:3}; c_{1:3})$ is

$$\hat{\Delta}_N(c^\star_{1:3}; c_{1:3}) := \frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{c_{1,2,3}(u_{1i},u_{2i},u_{3i})}{c^\star_{1,2,3}(u_{1i},u_{2i},u_{3i})}\right).$$

$\sqrt{N}(\hat{\Delta}_N(c^\star_{1:3}; c_{1:3}) - \Delta(c^\star_{1:3}; c_{1:3}))$ is asymptotically normal with variance $\sigma(c^\star_{1:3}; c_{1:3})^2$,

$$\sigma(c^\star_{1:3}; c_{1:3})^2 = \int_{[0;1]^3} c_{1,2,3}(u_1,u_2,u_3)\ln\left(\frac{c_{1,2,3}(u_1,u_2,u_3)}{c^\star_{1,2,3}(u_1,u_2,u_3)}\right)^2 du_1du_2du_3,$$

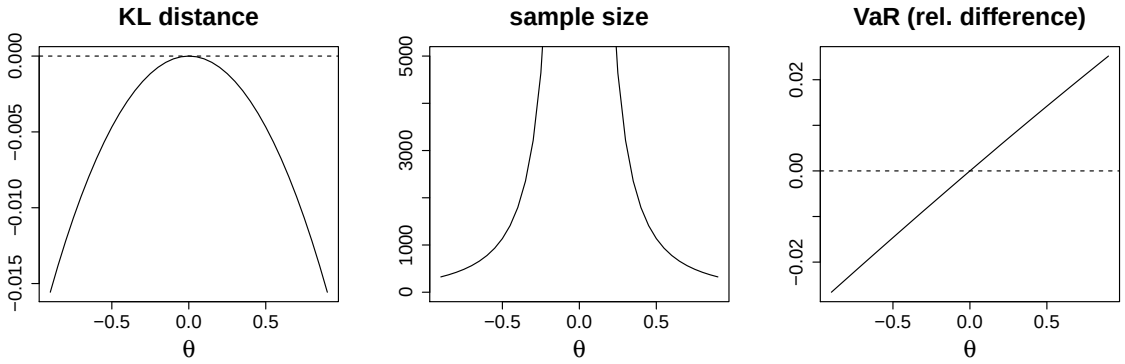
assuming that $\sigma(c^\star_{1:3}; c_{1:3})^2$ is finite (Central Limit Theorem). If C_{123} is the data generating process, we expect $\hat{\Delta}_N(c^\star_{1:3}; c_{1:3})$ to be positive for a finite sample. Thus, the required sample size to achieve a probability of $P(\hat{\Delta}_N(c^\star_{1:3}; c_{1:3}) > 0) \geq 1-\alpha$, $\alpha \in (0,1)$, i.e.

$$\begin{aligned} P(\hat{\Delta}_N(c^\star_{1:3}; c_{1:3}) > 0) &= P\left(\frac{\sqrt{N}(\hat{\Delta}_N(c^\star_{1:3}; c_{1:3}) - \Delta(c^\star_{1:3}; c_{1:3}))}{\sigma(c^\star_{1:3}; c_{1:3})} > \frac{-\sqrt{N}\Delta(c^\star_{1:3}; c_{1:3})}{\sigma(c^\star_{1:3}; c_{1:3})}\right) \\ &\approx \Phi\left(\sqrt{N}\Delta(c^\star_{1:3}; c_{1:3})/\sigma(c^\star_{1:3}; c_{1:3})\right) \geq 1-\alpha, \end{aligned}$$

is $N_\alpha(c^\star_{1:3}; c_{1:3}) > \sigma(c^\star_{1:3}; c_{1:3})^2\Phi^{-1}(1-\alpha)/\Delta(c^\star_{1:3}; c_{1:3})^2$.

For the trivariate extension of the FGM copula, this sample size N needed to distinguish between the true and the approximating model is given in Figure 3.1. The dependence is strongest for $\theta = \pm 1$, and this is where the KL distance of the simplified PCC approximation

Figure 3.1: KL distance Δ (left), sample size N needed to distinguish the models in a likelihood ratio test (middle) and relative difference in Value-at-Risk (right), for the trivariate FGM extension and the approximation by the independence copula.



from the true distribution is largest. The KL sample size needed to distinguish between the two distributions at confidence level 0.95 is 253 for $\theta = \pm 1$; for $|\theta| < 1$ several thousand observations can be necessary.

Another important measure for the quality of an approximation are the implications for the relative difference in portfolio Value-at-Risk (VaR). For a trivariate meta-normal distribution with standard normal margins and the FGM copula, the relative difference due to the simplifying approximation is illustrated in the right panel of Figure 3.1.

3.3.2 Multivariate 1-factor model

The simplifying assumption corresponds to the extremal case where conditional dependence does not vary as the values of the conditioning variables change. For the other extreme, we believe that a general class of models to consider are factor models constructed from sums of non-negative random variables. Factor models and their implied copulas have been proposed and studied by several authors in the literature including McNeil et al. (2005); Li (2000); Hull and White (2004, 2006) and Oh and Patton (2012). The 1-factor case generically has the form:

$$X_j = Z_0 + Z_j, \quad j = 1, \ldots, d, \tag{3.16}$$

where $Z_0, Z_1, \ldots, Z_d$ are independent random variables. Consider a conditional distribution such as $X_1, \ldots, X_k | X_{k+1} = x_{k+1}, \ldots, X_d = x_d$ and suppose that the Z's are all non-negative. As $x_{k+1}, \ldots, x_d$ all go to 0, the conditional dependence in $X_1, \ldots, X_k$ is close to independence because the conditioning event implies that Z_0 is close to 0. As $x_{k+1}, \ldots, x_d$ all go to ∞, the conditional dependence in $X_1, \ldots, X_k$ is strong because the conditioning event implies that

Z_0 is large with high probability. Note that the positivity of the Z's is a major reason for this. If the Z's can take positive and negative values it is not clear whether the simplifying assumption will be violated, it holds when the Z's are Gaussian.

Trivariate Gamma 1-factor model

As a concrete example, let us consider the trivariate Gamma factor model

$$X_j = Z_0 + Z_j, \quad j = 1, \ldots, 3,$$

where $Z_0, \ldots, Z_3$ are independent Gamma random variables with respective shape parameter p_j (and rate $b = 1$). Marginally, X_j is Gamma$(\eta_j, 1)$ with $\eta_j = p_0 + p_j$. Further, Z_0/X_j and X_j are independent random variables, and $Z_0/X_j \sim \text{Beta}(p_0, p_j)$ (Lukacs 1955). Given $X_3 = x_3$, this implies that a stochastic representation for (X_1, X_2) is $(Z_0^\star + Z_1, Z_0^\star + Z_2)$ where $Z_0^\star/x_3$ is an independent Beta(p_0, p_3) random variable. The Beta distribution has density $f_\beta(x; p_0, p_j) = B(p_0, p_j)^{-1} x^{p_0 - 1}(1-x)^{p_j - 1}$, where $B(p_0, p_j)$ is a normalizing constant (the beta function), mean $E[Z_0/X_j] = (p_0 + p_j)/p_j$ and variance $Var[Z_0/X_j] = \frac{p_0 p_j}{(p_0+p_j)^2(p_0+p_j+1)}$. Hence,

$$\begin{aligned}
\text{Var}[X_1|X_3 = x_3] &= \frac{x_3^2 p_0 p_3}{(p_0+p_3)^2(1+p_0+p_3)} + p_1, \\
\text{Var}[X_2|X_3 = x_3] &= \frac{x_3^2 p_0 p_3}{(p_0+p_3)^2(1+p_0+p_3)} + p_2, \\
\text{Cov}[X_1, X_2|X_3 = x_3] &= \frac{x_3^2 p_0 p_3}{(p_0+p_3)^2(1+p_0+p_3)}, \\
\text{Cor}[X_1, X_2|X_3 = x_3] &= \frac{1}{\sqrt{\left(1 + \frac{p_2(p_0+p_3)^2(1+p_0+p_3)}{x_3^2 p_0 p_3}\right)\left(1 + \frac{p_1(p_0+p_3)^2(1+p_0+p_3)}{x_3^2 p_0 p_3}\right)}}.
\end{aligned} \tag{3.17}$$

The conditional correlation is increasing in x_3 from 0 to 1 (Figure 3.2), which means that no conditional distribution is close to satisfying the simplifying assumption. In this model,

$$\begin{aligned}
f_{j3}(x_j, x_3) &= \int_0^{\min(x_j, x_3)} f_\Gamma(x_j - z; 1, p_j) f_\Gamma(x_3 - z; 1, p_3) f_\Gamma(z; 1, p_0) dz, \quad j = 1, 2, \\
f_{1:3}(\mathbf{x}_{1:3}) &= \int_0^{\min(\mathbf{x}_{1:3})} f_\Gamma(x_1 - z; 1, p_1) f_\Gamma(x_2 - z; 1, p_2) f_\Gamma(x_3 - z; 1, p_3) f_\Gamma(z; 1, p_0) dz,
\end{aligned}$$

which can be evaluated using adaptive integration. Thus, the density $c_{1:3}$,

$$c_{1:3}(u_1, u_2, u_3) = \frac{f_{1:3}(F_1^{-1}(u_1), F_2^{-1}(u_2), F_3^{-1}(u_3))}{\prod_{i=1}^3 f_i(F_i^{-1}(u_i))},$$

can be evaluated using only 1-dimensional numerical integrations. From the stochastic representation of (X_1, X_2) as $(Z_0^\star + Z_1, Z_0^\star + Z_2)$ we obtain the densities and distribution functions

by convolution as follows:

$$\begin{aligned}
f_{j|3}(y_j|x_3) &= \int_0^{\min(\frac{y_j}{x_3},1)} f_\Gamma(y_j - x_3 b; 1, p_j) f_\beta(b; p_0, p_3) db, \quad j = 1, 2, \\
F_{j|3}(y_j|x_3) &= \int_0^{\min(\frac{y_j}{x_3},1)} F_\Gamma(y_j - x_3 b; 1, p_j) f_\beta(b; p_0, p_3) db, \quad j = 1, 2, \\
f_{12|3}(y_1, y_2|x_3) &= \int_0^{\min(\frac{y_1}{x_3},\frac{y_2}{x_3},1)} f_\Gamma(y_1 - x_3 b; 1, p_1) f_\Gamma(y_2 - x_3 b; 1, p_2) f_\beta(b; p_0, p_3) db, \\
F_{12|3}(y_1, y_2|x_3) &= \int_0^{\min(\frac{y_1}{x_3},\frac{y_2}{x_3},1)} F_\Gamma(y_1 - x_3 b; 1, p_1) F_\Gamma(y_2 - x_3 b; 1, p_2) f_\beta(b; p_0, p_3) db.
\end{aligned}$$

We will now approximate the copula density $c_{1:3}$ with a simplified PCC of the form

$$\begin{aligned}
c^\star_{1:3}(u_1, u_2, u_3) &= c_{13}(u_1, u_3)\, c_{23}(u_2, u_3)\, c^\star_{12;3}(C_{1|3}(u_1|u_3), C_{2|3}(u_2|u_3)), \\
c^\star_{1:3}(u_1, u_2, u_3) &= c_{12}(u_1, u_2)\, c_{23}(u_2, u_3)\, c^\star_{13;2}(C_{1|2}(u_1|u_2), C_{3|2}(u_3|u_2)) \quad \text{or} \\
c^\star_{1:3}(u_1, u_2, u_3) &= c_{12}(u_1, u_2)\, c_{13}(u_1, u_3)\, c^\star_{23;1}(C_{2|1}(u_2|u_1), C_{3|1}(u_3|u_1)),
\end{aligned}$$

where $C_{j|3}(u_j|u_3) = F_{j|3}(F_j^{-1}(u_j)|F_3^{-1}(u_3))$, $j = 1, 2$. Here, the average conditional copula of X_1, X_2 given $X_3 = x_3$, which is also the copula one tries to estimate from data under the simplifying assumption, is

$$c^\star_{1,2;3}(u_1, u_2) = \int_0^1 c_{1,2;3}(u_1, u_2|u_3) du_3,$$

where

$$c_{1,2;3}(u_1, u_2|u_3) = \frac{f_{1,2|3}\left(F_{1|3}^{-1}\left(u_1|F_3^{-1}(u_3)\right), F_{2|3}^{-1}\left(u_2|F_3^{-1}(u_3)\right) \middle| F_3^{-1}(u_3)\right)}{f_{1|3}\left(F_{1|3}^{-1}\left(u_1|F_3^{-1}(u_3)\right) \middle| F_3^{-1}(u_3)\right) f_{2|3}\left(F_{2|3}^{-1}\left(u_2|F_3^{-1}(u_3)\right) \middle| F_3^{-1}(u_3)\right)}.$$

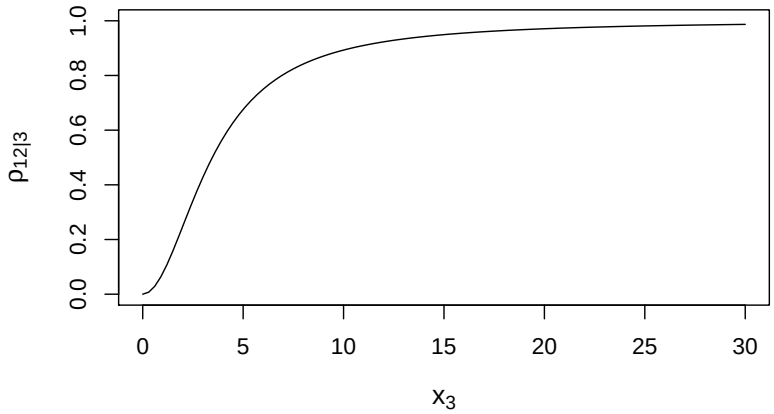
After transformation back to Gamma margins, this integral can be computed numerically using Gauss-Legendre quadrature. Similar relationships apply for $c_{13;2}$ and $c_{23;1}$, and $c^\star_{1,3;2}(u_1, u_3)$ and $c^\star_{2,3;1}(u_2, u_3)$ are obtained by permutation of indices. For some representative parameter values in the 1-factor Gamma convolution model, the sample sizes N corresponding to the distance of the approximate models of the true model are given in Table 3.1. They have been obtained by evaluating Δ and σ^2 via Monte Carlo integration.

The approach here for comparing a multivariate distribution to a nearby simplified PCC approximation also applies to higher-dimensional multivariate distributions including the Gamma factor model, but the calculations would be much more time-consuming. For dimensions $d \geq 4$, because there are many different PCCs and permutations of variables for which the simplifying assumption can be applied, the order of magnitude of sample sizes needed

Table 3.1: KL sample size N needed to differentiate the Gamma factor model from a PCC approximation with 95% confidence using a likelihood ratio test for some representative parameter vectors. KL sample sizes N are computed via Monte Carlo approximation of the integrals, and we report them to 2 significant digits. $\bar{\rho}$ is the average conditional correlation, $\bar{\rho}_{ij;k} = \int_0^\infty \rho_{ij;k}(x_k)\, f_k(x_k)\, dx_k$, and SD its expected variation over the range of the conditioning variable.

p_0	p_1	p_2	p_3	PCC approximation for					
				$C_{12;3}$		$C_{13;2}$		$C_{23;1}$	
				$\bar{\rho}_{12;3}$(SD)	N	$\bar{\rho}_{13;2}$(SD)	N	$\bar{\rho}_{23;1}$(SD)	N
1	1	1	1	0.244(0.205)	140	0.244(0.205)	140	0.244(0.205)	140
1	4	4	4	0.151(0.105)	830	0.151(0.105)	830	0.151(0.105)	830
1	1	1	4	0.376(0.183)	200	0.150(0.146)	300	0.150(0.146)	300
3	1	2	3	0.440(0.177)	160	0.337(0.171)	210	0.201(0.141)	290
4	1	2	3	0.477(0.169)	160	0.366(0.164)	220	0.216(0.135)	310
5	1	2	3	0.503(0.161)	180	0.386(0.158)	220	0.226(0.129)	310

Figure 3.2: Conditional correlation in the trivariate Gamma 1-factor model for parameters $p_1 = 1$, $p_2 = 1$, $p_3 = 1$ and $p_0 = 1$.



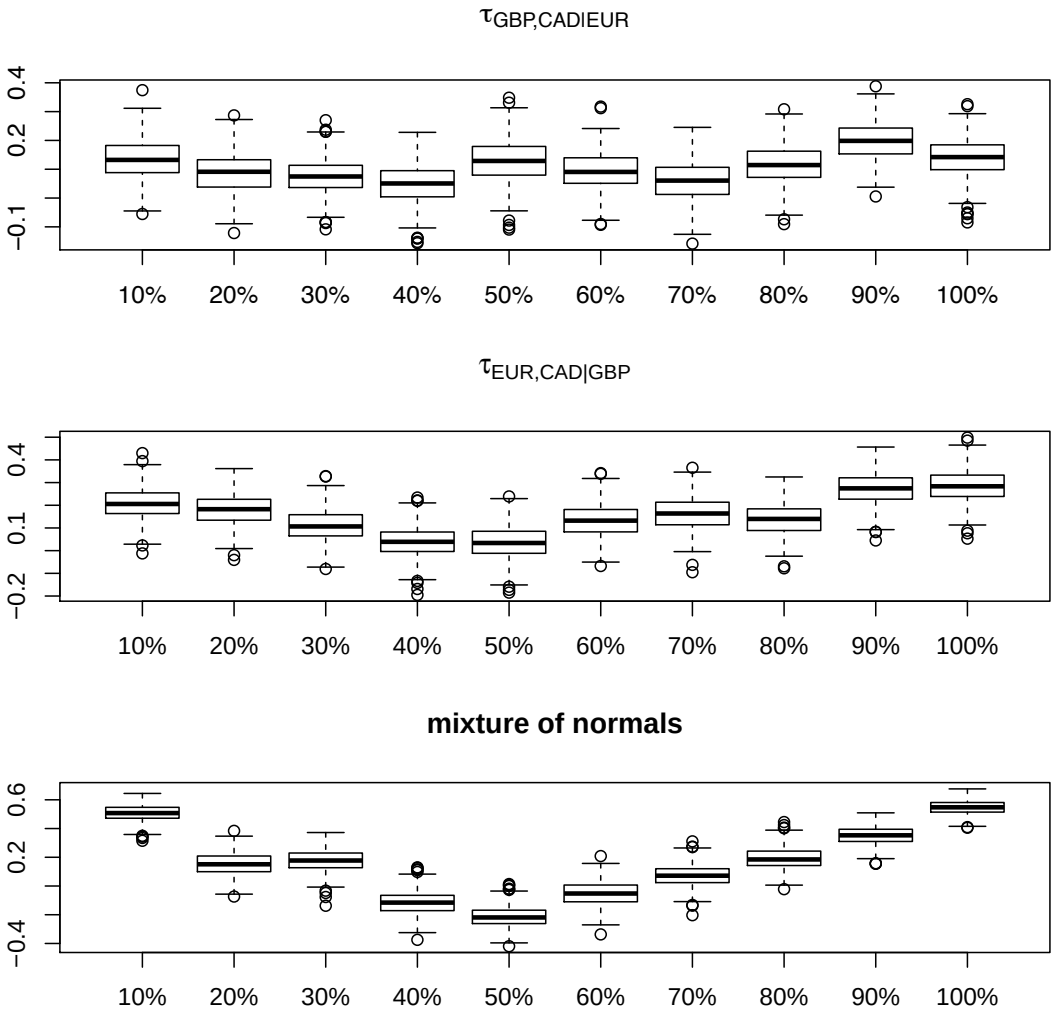
to distinguish a multivariate distribution and a "best-fitting" PCC need not be smaller than those for the trivariate Gamma factor model. Algorithms, such as in Dißmann et al. (2013) (see also Sections 4.1.4 and 6.2.1) for fitting high-dimensional R-vine PCCs, find good-fitting models based on sequential optimization criteria.

3.3.3 Non-simplified PCCs

The results presented in this chapter illustrate that the simplifying assumption for PCCs cannot be made in practical applications without some loss of generality. While for some models, the closest approximating PCC of the simplified type is virtually indistinguishable from the true model for typical data sizes, the difference can be significant for other models. In particular, the impact of a simplified model on the interpretation of results should be evaluated in empirical work and test for the simplifying assumption will need to be developed. For a given order of variables in the PCC and parametric bivariate copulas, such tests can be based on likelihood ratios as in Acar et al. (2012b).

To illustrate a simple but non-rigorous diagnostic check for the applicability of the simplifying assumption for the dependence structure of a given data set, let us consider the exchange rate data which will be analyzed in Chapter 5.1. A detailed description of this data set can be found in Section 5.1.1. We calculate the rank correlation between pairs of variables, conditioning on the corresponding values of a third variable being in a certain quantile of its distribution. To obtain a measure of uncertainty for these conditional values of Kendall's τ, we use a non-parametric bootstrap, where the data is resampled with replacement 500 times. The resulting boxplots are shown in Figure 3.3. The upper panel shows the variability in τs between the GBP/USD and CAD/USD exchange rates for different deciles of the EUR/USD exchange rate returns. One would probably attribute it to random variations, and with the given data size also the the Kendall's τ between the EUR/USD and CAD/USD rate for different deciles of GBP/USD returns presented in the mid panel does not give a clear indication for a violation of the simplifying assumption. However, the conditional τ values in the mid panel appear to depend quadratically on the deciles of the conditioning variable. Quadratic patterns like this are characteristic for mixture models as we illustrate in the lower panel of Figure 3.3 where a similar exercise was carried out for a simulated data set from a mixture of two trivariate normal distributions with correlations 0.9 and -0.3 respectively, which represents an extreme case of a mixture distribution with different strengths of dependence in both components.

Figure 3.3: upper panel: conditional Kendall's τ rank correlation between the returns of the USD/EUR and USD/CAD exchange rate, where we condition on the returns of the USD/GBP being in a given decile of the distribution.
middle panel: conditional Kendall's τ rank correlation between the returns of the USD/EUR and USD/CAD exchange rate, where we condition on the returns of the USD/GBP being in a given decile of the distribution.
lower panel: conditional Kendalls's τ for 1000 observations from a 50%:50% mixture of two trivariate normal distributions with correlations 0.9 and -0.3 respectively.
The boxplots are obtained using a non-parametric bootstrap.



If simplified PCCs are found to be inappropriate, the dependence on the values of the variables that are conditioned on can be modeled for the parameters of bivariate copulas in the PCC. When there is no a priori knowledge for how the dependence parameter should be influenced, non-parametric methods as in Acar et al. (2012a) can be applied. While this is fairly straightforward when conditioning on just one variable, it raises the question of how interactions should be included when conditioning on multiple variables. For the elliptical and Archimedean distributions which we studied in Sections 3 and 4, we observe the following:

For elliptical distributions, the Kendall's τ of conditional distributions does not depend on the values which are conditioned on. The conditional distribution however can depend on these variables. Assuming zero means and correlations, the conditional distributions of $X_1, \dots, X_{k-1})$ given $X_k = x_k, \dots, X_d = x_d$ will only depend on $a = \sum_{i=k}^{d} x_i^2$. If for an elliptical family the generator function $g_k(\cdot)$ in dimension k is monotonically decreasing (as for the multivariate Normal or Student's t distribution, and generally if $k \geq 2$, see Joe 1997, Section 4.9), this implies in particular that the conditional distribution only depends on $g_k(\mathbf{x}'_{k:d}\Sigma^{-1}_{k:d}\mathbf{x}_{k:d})$. For this reason, one might consider to make the dependence parameter of conditional copulas depend on the likelihood of observations that are conditioned on, for data where the distribution appears to be close to the elliptical family. However, since the values of Kendall's τ must not be affected - which are closely related to parameter values for most bivariate parametric families - keeping the simplifying assumption will always be a close approximation in these cases.

For an Archimedean copula with generator function φ, we have observed that when conditioning on realizations $X_k = x_k, \dots, X_d = x_d$ the conditional distribution will only depend on $a = \sum_{i=k}^{d} \varphi^{-1}(x_i)$ (see Equation (3.5)). In particular, this implies that it only depends on $\varphi(a)$, which is the cdf of $X_k, \dots, X_d$, evaluated at $X_k = x_k, \dots, X_d = x_d$. Thus, for data exhibiting dependence behavior close to that of the Archimedean class, we recommend to consider analyzing dependence of the conditional copulas on this joint probability.

While a mixture of different (simplified) copulas for different values of the conditioning variables yields the general non-simplified case, other mixture type models can be more intuitive in certain applications. This includes models where the parameters of a simplified copula are not constant over the whole dataset, but are varying over time in the context of financial time series. Examples for this were studied by Patton (2006) and Bartram et al. (2007), who consider models where the dependence parameter of the conditional copula depends on previous realizations of the time series, or Almeida and Czado (2011), who consider dependence of the parameter on an underlying AR(1) process. A discussion of how time-varying parameters can affect dependence measures is given by Manner and Segers (2011). A mixture

model of this kind, where simplified PCCs are mixed according to a latent Markov process, is introduced in the next chapter and applied in Chapter 5.1. This model implies that the copula of the multivariate time series at each point in time is given by a discrete mixture of simplified PCC's, which will usually be of non-simplified form as for the mixture of Normals.

Chapter 4

Regime switching models

In this chapter, which is based on material from Stöber and Czado (2013), we will consider a general regime switching model using R-vine copulas to describe the dependence structure.

Markov switching (MS), also called regime switching or hidden Markov models (Hamilton 1989), are time series models which allow for two or more distinct regimes. These regimes can be thought of as different states of the business cycle, the monetary policy, the economy, or more generally the world. Which regime is present at a particular point of time is governed by an underlying hidden Markov process. While there is a huge range of univariate or low dimensional applications of Markov switching models (Pelletier 2006; Garcia and Tsafack 2011; Chauvet and Hamilton 2006; Cerra and Saxena 2005; Hamilton 2005), their full potential for the modeling of a multidimensional set of underlyings and indices as it is required e.g. in the risk management of financial institutions has not yet been explored.

Here, we develop a model which can appropriately describe characteristics of financial data such as asymmetric dependence and tail dependence (Longin and Solnik 2001; Ang and Chen 2002). These are outside the world of the non tail-dependent Gaussian and symmetric Student's t distributions. In particular, we go beyond the Markov switching copula model of Chollete et al. (2009) in that we use the full flexibility of R-vine models. This superior flexibility will allow us to use truncation techniques (see Brechmann et al. (2012) and references therein), leading to a parsimonious parametrization of the model.

The chapter is divided in two sections. In the first section, we consider a regime switching model for the copula structure, assuming that we know the marginal distributions. Subsequently, we will consider a model where both the marginal distributions and the copula depend on the latent state in the second part.

4.1 Regime switching copulas

We start by considering a Markov switching R-vine (MS-RV) copula model. In this section, we will develop an approximative Expectation - Maximization (EM) type procedure in the Maximum Likelihood (ML) framework which allows for fast parameter estimation and is scalable to high dimensional applications. The algorithm is based on the sequential estimation procedure developed by Aas et al. (2009), which has been shown to be asymptotically

consistent by Hobæk Haff (2013). To address the issue of quantifying uncertainty, we further consider parameter inference for a prespecified MS-RV model in a Bayesian setup. For this, the algorithm of Min and Czado (2010), who consider Bayesian inference for a structural subclass of R-vine copulas, is generalized and extended to incorporate inference about the underlying Markov structure. In particular, we do also compute credible intervals for the probability of being in a given regime at a given point of time. While most existing models for time-varying dependence do not allow to quantify the uncertainty in the time variability of parameters, our Bayesian estimation procedure enables us to do so. The applicability and performance of our procedures for parameter estimation will be demonstrated in a simulation study. An application to exchange rate data is considered in Section 5.

The remainder of this section is structured as follows: We start with a general introduction to MS copula models in Subsection 4.1.1. Subsection 4.1.2 illustrates the calculation of the likelihood function in the presence of latent state variables and we consider parameter inference in the frequentist and the Bayesian setup in Subsections 4.1.2.1 and 4.1.2.2, respectively. The simulation study in Subsection 4.1.3 demonstrates that the estimation procedures work for simulated data. Subsection 4.1.4 presents a heuristic to select appropriate R-vine copula structures.

4.1.1 Markov switching copula models

We focus on multivariate statistical models based on general MS models introduced by Hamilton (1989). Since for now we are interested only in the copula structure, we want a model for the multivariate financial time series $\{\mathbf{U}_t = (U_{1,t}, \dots, U_{d,t}), t = 1, \dots, T\}$, where $\{U_{i,t}, t = 1, \dots T\}$ is i.i.d. uniform, using an MS approach. In this context, the dependency among $\mathbf{U}_t$ depends on a hidden latent state variable S_t, which takes on only finitely many values $k = 1, \dots, p$. These are called regimes and represent the different states of the economy. As is usual in the MS approach we assume that S_t, $t = 1, \dots, T$ is a homogeneous Markov chain (MC) in discrete time. For simplicity, we restrict to a first order MC, which is characterized by its transition matrix P with elements $P_{k,k'} := P(S_t = k' | S_{t-1} = k)$.

We use an R-vine copula to model the dependency of $\mathbf{U}_t$ in regime k ($S_t = k$). The Markov switching R-vine (MS-RV) copula for $\mathbf{u}_t$ is now fully characterized by specifying conditional densities as follows:

$$c(\mathbf{u}_t | (\mathcal{V}, \mathbf{B}, \boldsymbol{\theta})_{1,\dots,p}, S_t) = \sum_{k=1}^{p} 1_{\{k\}}(S_t) \cdot c(\mathbf{u}_t | (\mathcal{V}, \mathbf{B}, \boldsymbol{\theta})_k). \tag{4.1}$$

The complete MS-RV copula model is thus given in terms of p R-vine copula specifications and the transition matrix P which contains the parameters of the underlying Markov chain.

For inference, we will always assume the R-vine structures $\mathcal{V}_k$ and corresponding sets of copulas $\mathbf{B}_k$, $k = 1, \dots, p$, to be given and thus suppress them in the following notation. The MS R-vine copula is then described by its parameters

$$\boldsymbol{\theta}' = (\boldsymbol{\theta}'_{cop}, \boldsymbol{\theta}'_{MC}) = ((\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_p), \boldsymbol{\theta}'_{MC}),$$

where the subscript "cop" stands for copula parameters and "MC" for parameters needed for the transition matrix P. While this model does not include switching margins, the switching copula regimes induce serial dependence. Given previous realizations it will be more or less likely that the hidden variable S_t assumes a specific state. The individual marginal time series $(U_{i,t})_{t=1,\dots,T}$ however are i.i.d. uniform for $i = 1, \dots, d$.

4.1.2 Inference for Markov switching models

The first challenge in developing inference methods for MS models is that we are faced with unobserved latent variables. In order to derive an expression for the full likelihood of $\mathbf{u}_{1:T} = (\mathbf{u}_1, \dots, \mathbf{u}_T)$, we consider a decomposition of their joint density $f(\mathbf{u}_{1:T}|\boldsymbol{\theta})$ into conditional densities:

$$\begin{aligned} f(\mathbf{u}_{1:T}|\boldsymbol{\theta}) = f(\mathbf{u}_1|\boldsymbol{\theta}) \cdot \prod_{t=2}^{T} f(\mathbf{u}_t|\mathbf{u}_{1:(t-1)}, \boldsymbol{\theta}) = \left[\sum_{k=1}^{p} f(\mathbf{u}_1|S_1{=}k, \boldsymbol{\theta}_k) P(S_1{=}k|\boldsymbol{\theta}_{MC})\right] \\ \cdot \prod_{t=2}^{T} \left[\sum_{k=1}^{p} f(\mathbf{u}_t|S_t{=}k, \boldsymbol{\theta}_k) \cdot P(S_t{=}k|\mathbf{u}_{1:(t-1)}, \boldsymbol{\theta}_{MC})\right], \end{aligned} \tag{4.2}$$

where $f(\mathbf{u}_t|S_t = k, \boldsymbol{\theta}_k)$ is known from (4.1) for $t = 1, \dots, T$. The unconditional probabilities $P(S_1 = k)$ are known from the stationary distribution of the MC, which we assume to exist. To obtain the state prediction probabilities $\boldsymbol{\Omega}_{t|t-1} \in \mathbb{R}^p = \mathbb{R}^{p\times 1}$ with elements

$$\left(\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\right)_k := P(S_t{=}k|\mathbf{u}_{1:(t-1)}, \boldsymbol{\theta}) \qquad \text{for } k = 1, \dots, p$$

we can apply the filter of Hamilton (1989). Assuming $\boldsymbol{\Omega}_{t-1|t-1}$ to be given, we calculate

$$\begin{aligned} &\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) = P' \cdot \boldsymbol{\Omega}_{t-1|t-1}(\boldsymbol{\theta}) \quad \text{and} \\ &\boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta}) = \frac{\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta}) \odot (f(\mathbf{u}_t|S_t = k, \mathbf{u}_{1:(t-1)}, \boldsymbol{\theta}_k))_{k=1,\dots,p}}{\sum_{k=1}^{p} \left(\boldsymbol{\Omega}_{t|t-1}(\boldsymbol{\theta})\right)_k \odot f(\mathbf{u}_t|S_t = k, \mathbf{u}_{1:(t-1)}, \boldsymbol{\theta}_k)}, \end{aligned}$$

and obtain all probabilities which are required to evaluate the density (4.2) recursively. The operator $\odot$ denotes componentwise multiplication of two vectors. Similarly, the probability $\left(\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})\right)_{s_t} := P(S_t = s_t|\mathbf{u}_{1:T}, \boldsymbol{\theta})$, to which we will refer as the "smoothed" probability of being in state s_t at time t, can be determined by applying the following backward iterations.

$$\left(\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta})\right)_{s_t} = \left(\left(P \cdot \frac{\boldsymbol{\Omega}_{t+1|T}(\boldsymbol{\theta})}{\boldsymbol{\Omega}_{t+1|t}(\boldsymbol{\theta})}\right) \odot \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})\right)_{s_t}, \tag{4.3}$$

where also the division is to be understood componentwise.

Because of the latent state variables and the resulting dependence between parameters, direct maximization of the likelihood for given $(\mathcal{V}_k, \mathbf{B}_k)$, $k = 1, \ldots, p$, is analytically not possible and numerically difficult. In the following, we discuss a frequentist and a Bayesian approach to make inference for this kind of model tractable.

4.1.2.1 EM algorithm for MS-RV models

Hamilton (1990) proposed to overcome the problems in maximum likelihood estimation for an MS model by using an EM type (Dempster et al. 1977) algorithm. This algorithm iteratively determines parameter estimates $\boldsymbol{\theta}^l$, $l = 1, 2, \ldots$, which converge to the ML estimate for $l \to \infty$. To avoid technical difficulties when dealing with the stationary distribution of the Markov chain, we will not determine the probabilities $P(S_1{=}k)$ from the stationary distribution of the Markov chain here but include them as additional parameters of the model. Therefore, the parameter vector $\boldsymbol{\theta}^l$ is given by $(\boldsymbol{\theta}^l)' = ((\boldsymbol{\theta}^l_{cop})', (\boldsymbol{\theta}^l_{MC})')$ where $\boldsymbol{\theta}^l_{MC}$ consists of the initial state probabilities $(P(S_1 = k)^l)_{k=1,\ldots,p}$ and the transition matrix P^l with elements $P^l_{i,j} = P(S_t = j | S_{t-1} = i)^l$. This will make the inference problem computationally much more tractable and has the additional merit of allowing for the possibility of a permanent change in regimes, i.e. absorbing states of the Markov chain. Let us consider the expected pseudo likelihood function $Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ for $\boldsymbol{\theta}^{l+1}$, given observations $\mathbf{u}_{1:T}$ and the current parameter estimate $\boldsymbol{\theta}^l$, which is defined as

$$\begin{aligned} Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^l) &:= \int_{\mathbf{S}_{1:T}} \ln\Big(f(\mathbf{u}_{1:T}, \mathbf{S}_{1:T} | \boldsymbol{\theta}^{l+1})\Big) P(\mathbf{S}_{1:T} | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l) \\ &\propto \sum_{t=1}^{T} \int_{\mathbf{S}_{1:T}} \ln\Big(f(\mathbf{u}_t | S_t, \boldsymbol{\theta}^{l+1}_{cop})\Big) \cdot P(\mathbf{S}_{1:T} | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l) \\ &+ \int_{\mathbf{S}_{1:T}} \left[\sum_{t=2}^{T} \ln\Big(P(S_t | S_{t-1}, \boldsymbol{\theta}^{l+1}_{MC})\Big) + \ln(P(S_1)^{l+1})\right] \cdot P(\mathbf{S}_{1:T} | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l), \end{aligned} \tag{4.4}$$

where $\int_{\mathbf{S}_{1:T}} g(\mathbf{S}_{1:T}) := \sum_{s_1=1}^{p} \ldots \sum_{s_t=1}^{p} g(S_1 = s_1, \ldots, S_T = s_T)$ for an arbitrary function g of $\mathbf{S}_{1:T}$. The algorithm iterates the following steps:

(i) Expectation: Obtain the smoothed probabilities $\boldsymbol{\Omega}_{t|T}(\boldsymbol{\theta}^l)$ (compare Equation 4.3) of the latent states $\mathbf{S}_{1:T} = (S_1, \ldots, S_T)$ given the current parameter vector $\boldsymbol{\theta}^l$.

(ii) Maximization: Maximize $Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ with respect to $\boldsymbol{\theta}^{l+1}$.

Using the Markov property of $\mathbf{S}_{1:T}$, Kim and Nelson (2006) show that the maximum of the pseudo likelihood is attained at

$$P_{i,j}^{l+1} = \frac{\sum_{t=1}^{T} P(S_t = j, S_{t-1} = i | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)}{\sum_{t=1}^{T} P(S_{t-1} = i | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)},$$

similarly $P(S_1{=}k)^{l+1} = P(S_1 = k | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$, $k = 1, \dots, p$. In contrast to the model originally considered by Hamilton (1990), where all maximization steps could be performed analytically, this is not possible for the maximization with respect to the copula parameters $\boldsymbol{\theta}_{cop}^{l+1}$ in our case. This means that, while $\boldsymbol{\theta}_{MC}^{l+1}$ can be obtained directly, the maximization of $Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ with respect to $\boldsymbol{\theta}_{cop}^{l+1}$ has to be performed using numerical optimization methods. Since a d-dimensional R-vine copula specification, in which each pair copula has k parameters, contains $d(d-1)/2 \cdot k$ parameters, this is computationally still very challenging. To circumvent this problem, we can exchange the joint maximization with respect to $\boldsymbol{\theta}_{cop}^{l+1}$ with the stepwise maximization procedure of Aas et al. (2009) which is modified to weight each observation by $P(S_t = s_t | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$.

We call this the *stepwise EM-Algorithm*. Since tree-wise estimation of copula parameters is asymptotically consistent (Hobæk Haff 2013), this constitutes a close approximation to the "proper" EM-Algorithm. While there are theoretical results on the convergence of the EM-Algorithm shown by Wu (1983), we loose these properties with our approximation. All limit theorems however do rely on proper maximization at each step of the algorithm. This is almost impossible to guarantee in our case where we are faced with high dimensional optimization problems and have to rely on numerical techniques. While all existing models for time-varying dependence structures in high dimensions suffer from the computational burden for numerical estimation, we do only need to maximize the likelihoods of bivariate copulas in this tree-wise procedure. This reduces computation time and avoids the curse of dimensionality. The obtained estimate we denote by

$$\left(\hat{\boldsymbol{\theta}}^{\mathrm{EM}}\right)' = \left(\left(\hat{\boldsymbol{\theta}}_{cop}^{\mathrm{EM}}\right)' = \left(\left(\hat{\boldsymbol{\theta}}_1^{\mathrm{EM}}\right)', \dots, \left(\hat{\boldsymbol{\theta}}_p^{\mathrm{EM}}\right)'\right), \left(\hat{\boldsymbol{\theta}}_{MC}^{\mathrm{EM}}\right)'\right).$$

4.1.2.2 Gibbs sampling for the MS-RV model

Having derived an approximative ML procedure for our MS copula, we will now consider Bayesian estimation methods, which will enable us to quantify the uncertainty in parameter estimates. In particular, credible intervals (CIs) and posterior standard deviations are determined naturally while the uncertainty in ML parameter estimates is very hard to assess in this context. Building on ideas of Albert and Chib (1993), the Gibbs sampler which we develop consists of updates for the copula parameters, the Markov chain parameters and the

latent state vector, respectively. Iterating through all three outlined update steps will yield a sample

$$\big((\boldsymbol{\theta}^{r,\text{MCMC}})', \mathbf{S}_{1:T}^{r,\text{MCMC}}\big) = \Big(\Big(\Big((\boldsymbol{\theta}_1^{r,\text{MCMC}})', \ldots, (\boldsymbol{\theta}_p^{r,\text{MCMC}})'\Big), (\boldsymbol{\theta}_{MC}^{r,\text{MCMC}})'\Big), \mathbf{S}_{1:T}^{r,\text{MCMC}}\Big),$$

for $r = 1, \ldots, R$, where R is the number of realizations.

Update of copula parameters In order to complete the model specification in a Bayesian framework, we first have to specify prior distribution for each component of $\boldsymbol{\theta}_{cop}$. Following Min and Czado (2010), we assume uniform priors for all copula parameters in the model. For bivariate copula families where the parameter range is not compact, we restrict its support to some finite interval to avoid numerical instabilities for very small or large parameter values. If for all bivariate copulas there is a one-to-one correspondence between parameter values and Kendall's τ given in closed form, uniform priors for τ can be considered as an alternative. Furthermore, we can use a uniform prior for the correlation matrix of the model if all bivariate building blocks are Gaussian or Student's t copulas, cf. Lewandowski et al. (2009). Since the conditional distributions of the copula parameters given the remaining parameters are not available we use a Metropolis-Hastings (MH) update here. There are several choices for proposal distributions available. Min and Czado (2010) use a modification of standard random walk proposals where the normal distribution is truncated to the parameter support, while proposal variances are tuned to achieve suitable acceptance rates. This leads to poor acceptance rates in some cases with strong dependencies and to high autocorrelations in general. To overcome these problems, we consider a two point mixture of a random walk proposal with an independent normal distribution at the mode of the likelihood function for each parameter. The modes are approximated by the stepwise estimation procedure for R-vines and the standard errors are obtained from the inverse Hessian. Both distributions are assigned a weight of 0.5. Independence proposals centered around the mode have been proposed by Gamerman and Lopes (2006) and have been applied in a context similar to ours by Czado et al. (2011). While there are parameter constellations where pure random walk proposals are more favorable than independence proposals and vice versa, simulation studies showed that the chosen mixture distribution works well for all settings.

Update of Markov chain parameters For this, we will assume independent Dirichlet distributions as prior distributions for the rows of the transition matrix P, i.e.

$$(P_{k,k'})_{k'=1,\ldots,p} \sim \text{Dirichlet}\big((\alpha_{k,k'})_{k'=1,\ldots,p}\big),$$

for $k = 1, \ldots, p$. The conditional posterior distribution of the transition probabilities in P, given the other parameters, depends only on the latent state vector $\mathbf{S}_{1:T}$. Here, the likelihood function

$$l(P|\mathbf{S}_{1:T}) = \prod_{k=1}^{p} \prod_{k'=1}^{p} p_{k,k'}^{n_{k,k'}},$$

where $n_{k,k'}$ denotes the number of transitions from state k to state k' in $\mathbf{S}_{1:T}$ is multinomial. Since the Dirichlet and the multinomial distribution are conjugate distributions (see Kotz et al. (2000)), also the conditional posterior distributions are Dirichlet distributions with parameters $\alpha_{k,k'}^{posterior} = \alpha_{k,k'} + n_{k,k'}$. From these we can sample directly.

Update of the latent state vector We follow the approach by Kim and Nelson (1998) to update $\mathbf{S}_{1:T}$ jointly assuming independent non informative priors and decompose

$$P(\mathbf{S}_{1:T}|\mathbf{u}_{1:T}, \boldsymbol{\theta}_{cop}, \boldsymbol{\theta}_{MC}) = P(\mathbf{S}_{1:T}|\mathbf{u}_{1:T}) = P(S_T|\mathbf{u}_{1:T}) \cdot \prod_{i=1}^{T-1} P(S_t|S_{t+1}, \mathbf{u}_{1:T}).$$

This allows to generate S_T from $P(S_T|\mathbf{u}_{1:T})$ and S_t for $t = T-1, \ldots, 1$ from

$$P(S_t|\mathbf{u}_{1:T}, S_{t+1}) \propto P(S_{t+1}|S_t)P(S_t|\mathbf{u}_{1:t}),$$

where $P(S_t|\mathbf{u}_{1:t}) = \boldsymbol{\Omega}_{t|t}(\boldsymbol{\theta})$ can again be determined using the Hamilton filter. Note that in order to obtain the convenient conjugate prior for the elements of the transition matrix in the Gibbs update of the Markov chain parameters, the prior distribution for the initial value S_1 which is required for the Hamilton filter must be independent of of $\boldsymbol{\theta}_{cop}$, e.g. uniform. This is similar to the frequentist setup where we included the initial probabilities $P(S_1{=}k)$ as additional parameters to be estimated. Robustness studies however show that the influence of this prior distribution on the joint posterior is negligible which is why most authors ignore this (see Albert and Chib (1993) and Kim and Nelson (1998)). We will follow their suggestions and do not discuss the choice of this prior distribution any further.

4.1.3 Simulation study

This section gives the results of a simulation study which has been performed in order to demonstrate the ability of the developed Bayesian inference procedure to capture the true model in simulated data. We consider two regimes and 5 parameter setups, see Table 4.1. In all scenarios we set the Markov parameters to $P(S_t = 1|S_{t-1} = 1) = 0.95$ and $P(S_t = 2|S_{t-1} = 2) = 0.9$, the corresponding prior distributions are chosen to be uniform. For each scenario, we simulate a time series with 800 four dimensional observations. Keeping the (true)

R-vine structure and copula families we used for simulations, we obtain a posterior estimate for the parameters as follows:

(i) Starting values for the EM algorithm: Fit the copula parameters for each regime to the whole data set using the stepwise estimation procedure, and cluster the observations according to their likelihood values. Re-fit to the 400 observations which have the highest log likelihood.

(ii) Starting values for MCMC: Obtained using the stepwise EM algorithm.

(iii) Obtain 1000 independent MCMC samples from the posterior distribution of the parameters. A burn-in period is discarded and the chain is sub-sampled according to what Kass et al. (1998) call "effective sample size" (cf. Carlin and Louis (2009)).

Table 4.1: The two columns on the left define the models we consider in simulation scenarios (1) - (5). The copulas in Regime 1 (Regime 2) are chosen to be Gumbel (Gaussian) copulas and we provide the values of Kendall's τ corresponding to the chosen copula parameters. The empirical coverage probabilities on the righthand side are based on 120 data sets from each scenario.

	conditional Kendall's τ		coverage probability 90% CI		coverage probability 95% CI	
	Gumbel D-Vine (Regime 1)	Gauss C-Vine (Regime 2)	sym.	HPD	sym.	HPD
(1)	$\tau_{43\|21} = 0.4$ $\tau_{42\|1} = 0.6$, $\tau_{32\|1} = 0.6$ $\tau_{41} = 0.8$, $\tau_{31} = 0.8$, $\tau_{21} = 0.8$	$\tau_{41\|23} = 0.4$ $\tau_{42\|3} = 0.6$, $\tau_{31\|2} = 0.6$ $\tau_{43} = 0.8$, $\tau_{32} = 0.8$, $\tau_{21} = 0.8$	92%	92%	94%	94%
(2)	$\tau_{43\|21} = 0.4$ $\tau_{42\|1} = 0.6$, $\tau_{32\|1} = 0.6$ $\tau_{41} = 0.8$, $\tau_{31} = 0.8$, $\tau_{21} = 0.8$	$\tau_{41\|23} = 0.1$ $\tau_{42\|3} = 0.2$, $\tau_{31\|2} = 0.2$ $\tau_{43} = 0.3$, $\tau_{32} = 0.3$, $\tau_{21} = 0.3$	89%	89%	92%	92%
(3)	$\tau_{43\|21} = 0.1$ $\tau_{42\|1} = 0.2$, $\tau_{32\|1} = 0.2$ $\tau_{41} = 0.3$, $\tau_{31} = 0.3$, $\tau_{21} = 0.3$	$\tau_{41\|23} = 0.4$ $\tau_{42\|3} = 0.6$, $\tau_{31\|2} = 0.6$ $\tau_{43} = 0.8$, $\tau_{32} = 0.8$, $\tau_{21} = 0.8$	85%	84%	92%	93%
(4)	$\tau_{43\|21} = 0.1$ $\tau_{42\|1} = 0.2$, $\tau_{32\|1} = 0.2$ $\tau_{41} = 0.3$, $\tau_{31} = 0.3$, $\tau_{21} = 0.3$	$\tau_{41\|23} = 0.1$ $\tau_{42\|3} = 0.2$, $\tau_{31\|2} = 0.2$ $\tau_{43} = 0.3$, $\tau_{32} = 0.3$, $\tau_{21} = 0.3$	75%	75%	82%	93%
(5)	$\tau_{43\|21} = 0.3$ $\tau_{42\|1} = 0.5$, $\tau_{32\|1} = 0.3$ $\tau_{41} = 0.7$, $\tau_{31} = 0.5$ $\tau_{21} = 0.3$	$\tau_{41\|23} = 0.3$ $\tau_{42\|3} = 0.5$, $\tau_{31\|2} = 0.3$ $\tau_{43} = 0.7$, $\tau_{32} = 0.5$, $\tau_{21} = 0.3$	85%	85%	92.8%	93%

For example, if we have 5000 observations after discarding the burn-in period, and estimate an average effective sample size over the copula parameters of 1000, we take every 5th

observation to obtain an approximately independent sample of 1000 observations.

From the obtained samples, we estimate 90% and 95% symmetric (sym.) and highest posterior density (HPD) credible intervals (CIs) for the copula parameters and check whether all true copula parameters lie within these intervals. The procedure was repeated 120 times for each scenario with results reported in Table 4.1. Relative bias and mean squared error (MSE) for two selected scenarios are given in Appendix B. For all parameter setups, except Scenario 4, we observe about 90% (95%) frequentist coverage. Scenario 4 corresponds to low dependence in both regimes, thus the regimes are less distinguishable. Therefore, with clearly distinguishable regimes the outlined procedure is able to identify the true model.

4.1.4 R-vine copula model selection

In order to apply the previously discussed inference procedures, suitable R-vine copula models for all regimes must be selected in a pre-analysis. This means, that we have to identify a tree structure and subsequently select corresponding bivariate copulas.

If there is a natural order of the marginal variables as for example for the longitudinal data considered in Smith et al. (2010), this information can be incorporated in the model selection by choosing a corresponding tree structure. For a given structure, bivariate pair copula families can be selected using goodness-of-fit tests, information criteria like the Akaike information criterion (AIC) and the Bayesian information criterion (BIC, see Section 1.3.2), or exploratory data analysis using contour plots and lambda functions (see Schepsmeier (2010) for a comparison). If the variables do not have a natural ordering, we need to apply heuristics like the procedure developed by Dißmann et al. (2013).

For regime switching models, the situation is more challenging since we do not want to determine one copula structure for the whole data set but we need to select an R-vine copula for each regime. For data sets with a natural ordering of the marginal variables this information can again be incorporated in the model selection. But if there is no such ordering, an analogue of the Dissmann procedure will be required. If we have additional information about the data which allows, for each regime, to identify a period of the data set where this regime should be dominant, the R-vine copula models can be chosen by applying the algorithm of Dißmann et al. (2013) to these subsets. This will be demonstrated for exchange rates in Section 5 where we use a rolling window analysis to determine appropriate periods.

Oftentimes however, we will want to run an unsupervised algorithm which selects appropriate models automatically. For this, the model selection heuristics for R-vine copulas can be combined with the EM-algorithm for regime switching models. Instead of maximizing the pseudo likelihood function $Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^{l})$ with respect to the copula parameters, we will

select a new copula model for each possible state, weighting the observations according to the smoothed regime probabilities $P(\mathbf{S}_{1:T}|\mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ in the model selection. For a given regime k with weight vector $P(\mathbf{S}_{1:T} = k|\mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ the model selection heuristic of Dißmann et al. (2013) can be adapted as follows.

(i) For each pair of variables estimate the corresponding value of Kendall's τ from the copula data set. For weighted observations, this Kendall's τ statistic can be estimated using the algorithm of Pozzi et al. (2012), see Section 1.2.1.

(ii) Create a fully connected graph which consists of the marginal variables as nodes and where an edge is added between every pair of variables.

(iii) Associate to each edge the absolute value of the corresponding (weighted) Kendall's τ statistic as edge weight.

(iv) Determine the maximum spanning tree (MST), i.e. find a tree which maximizes the sum of edge weights using for example the algorithm of Prim (1957).

(v) For each edge in the resulting tree select a parametric bivariate copula from a catalogue of bivariate copula families and estimate its parameters. A possible selection criterion is AIC/BIC in which the observations are weighted according to $P(\mathbf{S}_{1:T} = k|\mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$, see Section 1.3.2. To achieve more parsimonious models, a bivariate independence test as in Genest and Favre (2007) can be used to pre-test for independence.

(vi) Proceed similarly until all trees and corresponding copulas are selected.

The expectation and model selection step are then iterated until the change in the resulting model likelihood in each step falls below a certain limit.

4.2 Regime switching copulas and marginal distributions

While the copula of a continuous multivariate distribution is independent of the marginal distributions by definition, it is this independent treatment of marginal distributions and the dependence structure which has attracted critics in the past (Mikosch 2006). Even more so for Markov Switching models, switches in the marginal distribution might influence the estimation of the dependence structure, and a two-step analysis might therefore be biased.

To address these issues, we will consider a model where both marginal time series and the dependence structure are subject to changes in regime. While we will have to rely partially on step-wise estimation procedures for reasons of computational tractability, we will allow

for dependence between regime changes in different marginal time series and the copula and estimate the transition probabilities jointly.

We will conduct a small simulation study comparing the results of a joint estimation to a pure two-step procedure in which marginal Markov switching models are fitted first and the copula is subsequently estimated from transformed residuals.

4.2.1 Model setup

While we have discussed a pure copula model with uniform marginal distributions in Section 4.1, both the marginal time series models and the copula will depend on the latent process $(S_t)_{t=1,\dots,T}$ here. To keep the model tractable for inference, we choose a regime switching Gaussian model for the margins. While this is a comparably simple model, it can still replicate the important features found in empirical financial data such as skewness, excess kurtosis, autocorrelation of returns and volatility clustering (Timmermann 2000). We will outline the estimation procedure for the case of a pure Gaussian model without autoregressive (AR) structure (an AR structure can be incorporated easily by using OLS regression on previous observations instead of just estimating the means in each step, cf. Hamilton (1990)). The dependence will again be described by an R-vine copula with regime switching tree structure, families and parameters. Formally, the model we are considering is given as follows. For $i = 1, \dots, d$, $t = 1, \dots, T$ and $S_t \in \{1, \dots, p\}$,

$$X_{i,t} = \mu_{i,S_t} + \sigma_{i,S_t}\epsilon_{i,t}, \qquad \epsilon_{i,t} \overset{i.i.d}{\sim} N(0,1),$$

and hence the joint distribution function in state S_t is given by

$$F_{1:d}(\mathbf{x}_t|S_t) = C\left(\Phi\left(\frac{x_{1,t}-\mu_{1,S_t}}{\sigma_{1,S_t}}\right), \dots, \Phi\left(\frac{x_{d,t}-\mu_{d,S_t}}{\sigma_{d,S_t}}\right)\middle|\mathcal{V}_{S_t}, \mathcal{B}_{S_t}, \boldsymbol{\theta}_{S_t}\right).$$

In general, this allows for p distinct parameter sets for each marginal distribution and p distinct R-vine copulas. While we expect that each marginal time series and the dependence structure can be described by a moderate number of regimes, we want to allow for regime switches to occur at different times for different margins. We therefore consider p_i regimes $S_{i,t}$ for $(X_{i,t})_{t=1,\dots,T}$ and p_c copula dependence regimes $S_{c,t}$, and

$$S'_t = (S_{1,t'}, \dots, S'_{d,t}, S'_{c,t})$$

i.e. the state space of the joint model consists of all $p = p_c \cdot \prod_{i=1}^{d} p_i$ combinations of marginal and dependence regimes. The joint distribution function in state S_t becomes

$$\begin{aligned} F_{1:d}(\mathbf{x}_t|S_t) &= C\left(F_1(x_{1,t}|\boldsymbol{\theta}_{1,S_{1,t}}), \dots, F_d(x_{d,t}\middle|\boldsymbol{\theta}_{d,S_{d,t}})|\mathcal{V}_{S_{c,t}}, \mathcal{B}_{S_{c,t}}, \boldsymbol{\theta}_{c,S_{c,t}}\right) \\ &= C\left(\Phi\left(\frac{x_{1,t}-\mu_{1,S_{1,t}}}{\sigma_{1,S_{1,t}}}\right), \dots, \Phi\left(\frac{x_{d,t}-\mu_{d,S_{d,t}}}{\sigma_{d,S_{d,t}}}\right)\middle|\mathcal{V}_{S_{c,t}}, \mathcal{B}_{S_{c,t}}, \boldsymbol{\theta}_{c,S_{c,t}}\right). \end{aligned}$$

For given R-vine tree structures $\mathcal{V}_1, \dots, \mathcal{V}_{p_c}$ and pair copulas $\mathcal{B}_1, \dots, \mathcal{B}_{p_c}$, the joint Markov switching is parametrized by

$$\boldsymbol{\theta}' = (\boldsymbol{\theta}'_{cop}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_d, \boldsymbol{\theta}'_{MC}) = ((\boldsymbol{\theta}'_{c,1}, \dots, \boldsymbol{\theta}'_{c,p_c}), (\boldsymbol{\theta}'_{1,1}, \dots, \boldsymbol{\theta}'_{1,p_1}), \dots, (\boldsymbol{\theta}'_{d,1}, \dots, \boldsymbol{\theta}'_{d,p_d}), \boldsymbol{\theta}'_{MC}),$$

where $\boldsymbol{\theta}_{cop}$ are the copula parameters, $\boldsymbol{\theta}_i$ the marginal parameters for time series $(X_{i,t})_{t=1,\dots,T}$ and $\boldsymbol{\theta}_{MC}$ consists of the independent probabilities in the transition matrix. When assuming that all marginal time series and the dependence structure switch regimes at the same time, this model and the following algorithm can be modified setting $S_{i,t} = S_{c,t} = S_t$, $i = 1, \dots, d$.

4.2.2 Inference and EM algorithm for the joint model

We will now discuss how the EM algorithm for the RV-MS model is adapted for the joint Markov Switching model. We assume the R-vine tree structures and pair copulas to be fixed and we estimate the model parameters $\boldsymbol{\theta}$. For observations $\mathbf{x}_{1:T}$, the expected pseudo log likelihood function for the step $l+1$ estimate $\boldsymbol{\theta}^{l+1}$ given $\boldsymbol{\theta}^l$ and the data (Equation (4.4)) is

$$\begin{aligned}
&Q(\boldsymbol{\theta}^{l+1}; \mathbf{x}_{1:T}, \boldsymbol{\theta}^l) := \int_{\mathbf{S}_{1:T}} \ln\Big(f(\mathbf{x}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\theta}^{l+1})\Big) P(\mathbf{S}_{1:T}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&\propto \sum_{t=1}^{T} \int_{\mathbf{S}_{1:T}} \ln\Big(f(\mathbf{x}_t|S_t, \boldsymbol{\theta}^{l+1}_{cop}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_d)\Big) \cdot P(\mathbf{S}_{1:T}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&+ \int_{\mathbf{S}_{1:T}} \left[\sum_{t=2}^{T} \ln\Big(P(S_t|S_{t-1}, \boldsymbol{\theta}^{l+1}_{MC})\Big) + \ln(P(S_1)^{l+1})\right] \cdot P(\mathbf{S}_{1:T}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&= \sum_{t=1}^{T} \sum_{s_{1,t}=1}^{p_1} \cdots \sum_{s_{d,t}=1}^{p_d} \sum_{s_{c,t}=1}^{p_c} \ln\Big(c\Big(F_1(x_{1,t}|\boldsymbol{\theta}_{1,s_{1,t}}), \dots, F_d(x_{d,t}\Big|\boldsymbol{\theta}_{d,s_{d,t}})\Big|\boldsymbol{\theta}_{c,s_{c,t}}\Big)\Big) P(S_t = s_t|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&+ \sum_{t=2}^{T} \sum_{s_t, s_{t-1}=1}^{p} \ln(P^{l+1}_{s_{t-1}, s_t}) P(S_t = s_t, S_{t-1} = s_{t-1}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&+ \sum_{s_1=1}^{p} \ln(P(S_1 = s_1)^{l+1}) P(S_1 = s_1|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) \\
&+ \sum_{t=1}^{T} \sum_{s_{1,t}=1}^{p_1} \ln\big(f_1(x_{1,t}|\boldsymbol{\theta}_{1,s_{1,t}})\big) P(S_{1,t} = s_{1,t}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l) + \dots \\
&+ \sum_{t=1}^{T} \sum_{s_{d,t}=1}^{p_d} \ln\big(f_d(x_{d,t}|\boldsymbol{\theta}_{d,s_{d,t}})\big) P(S_{d,t} = s_{d,t}|\mathbf{x}_{1:T}, \boldsymbol{\theta}^l).
\end{aligned}$$

Clearly, the parameters $\boldsymbol{\theta}^{l+1}_{MC}$ of the Markov chain (i.e. the transition matrix and the initial state probabilities) can be estimated independently from the copula and marginal parameters

and the maximum is again attained at

$$P_{i,j}^{l+1} = \frac{\sum_{t=2}^{T} P(S_t = j, S_{t-1} = i | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)}{\sum_{t=2}^{T} P(S_{t-1} = i | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)},$$

and $P(S_1 = k)^{l+1} = P(S_1 = k | \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$, k=1,...,b. The remaining step $l+1$ parameters can be estimated either by maximizing $Q(\boldsymbol{\theta}^{l+1}; \mathbf{u}_{1:T}, \boldsymbol{\theta}^l)$ jointly with respect to $\boldsymbol{\theta}_{cop}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ or by using a weighted analogue of the inference functions for margins procedure and estimating the marginal parameters in a first step before estimating the copula parameters. To keep the model computationally tractable, we propose to proceed in this multi step fashion, since it allows to compute $\hat{\boldsymbol{\theta}}_1^{l+1}, \ldots, \hat{\boldsymbol{\theta}}_d^{l+1}$ in closed form (Kim and Nelson 2006),

$$\mu_{i,k_i}^{l+1} = \frac{\sum_{t=1}^{T} x_{i,t} P(S_{i,t} = k_i | \mathbf{x}_{1:T}, \boldsymbol{\theta}^l)}{\sum_{t=1}^{T} P(S_{i,t} = k_i | \mathbf{x}_{1:T}, \boldsymbol{\theta}^l)},$$

$$\sigma_{i,k_i}^{l+1} = \sqrt{\frac{\sum_{t=1}^{T} \left(x_{i,t} - \mu_{i,k_i}^{l+1}\right)^2 P(S_{i,t} = k_i | \mathbf{x}_{1:T}, \boldsymbol{\theta}^l)}{\sum_{t=1}^{T} P(S_{i,t} = k_i | \mathbf{x}_{1:T}, \boldsymbol{\theta}^l)}}, \qquad i = 1, \ldots, d, \quad k = 1, \ldots, p_i.$$

Given the marginal parameters estimated in this first step, the estimation of $\boldsymbol{\theta}_{cop}^{l+1}$ reduces to the maximization problem considered in Section 4.1.2 and we will again employ the stepwise estimation algorithm for R-vine copulas to keep numerical computations tractable.

4.2.3 Simulation study

This section presents the result of a simulation study, comparing the estimator defined by the EM algorithm of Section 4.2.2 to the estimator defined by estimating the marginal Markov switching models using EM end then applying the algorithm of Section 4.1.2.1 to the pseudo observations

$$u_{i,t} = F(x_{i,t} | \boldsymbol{\theta}_i, \mathbf{x}_{i,1:(t-1)}),$$

$i = 1, \ldots, d$. While we expect to observe dependence between the Markov chains $(S_{i,t})_{t=1,\ldots,T}$, $(S_{c,t})_{t=1,\ldots,T}$ in real world data, joint modeling of the transition matrix significantly increases the number of parameters. To check whether the joint modeling approach improves parameter inference, we consider a three dimensional example. In total, we explore 4 simulation setups. In setups 1 and 2 the number of regimes is $p_1 = 3$ and $p_2 = p_3 = p_{cop} = 2$. Since the transition probabilities from each state must sum to one, the number of parameters of the transition matrix is $\sum_{i=1}^{3} p_i(p_i - 1) + p_{cop}(p_{cop} - 1) = 12$ (linear in the number of dimensions) under independence. It increases to $\left(\prod_{i=1}^{3} p_i \cdot p_{cop}\right)\left(\prod_{i=1}^{3} p_i \cdot p_{cop} - 1\right) = 552$ (exponential in the number of dimensions) when modeling the full transition matrix. In setups 3 and 4,

all marginal time series and the copula will share a common state variable, i.e. the number of regimes is $p = 2$.

The model parameters of our 4 simulation setups are defined in Table 4.2. Since the choice of different copula families should not influence the performance of the considered estimation procedures, we decide to conduct the simulations in a Gaussian setup. Other copula families could easily be incorporated.

To induce dependence between the Markov chains in setups 1 and 2, we build the joint transition matrix which we consider in the simulation setup by sequential conditioning. Here,

$$P_{cop} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad P_{3,n}^{S1} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad P_{3,n}^{S2} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}, \quad P_{3,s} = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix},$$

$$P_{2,n}^{S1} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \quad P_{2,n}^{S2} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}, \quad P_{2,s} = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix},$$

$$P_{1,n} = \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}, \quad P_{1,s} = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \end{pmatrix},$$

Table 4.2: Model parameters in the simulation setup. N indicates the use of a bivariate Normal copula, the copula parameters are given in terms of Kendall's τ. Here, for example $(c_{13})_1$ refers to the $1,3$ copula in regime 1, while $(c_{13})_2$ is the copula in regime 2.

			μ_1	μ_2	μ_3	σ_1	σ_2	σ_3
margins	Setup 1,2	X_1	0	0	0	0.1	0.2	0.4
		X_2	0.1	0.2		0.1	0.15	
		X_3	0	0		0.1	0.2	
margins	Setup 3,4	X_1	0	0		0.1	0.4	
		X_2	0.1	0.2		0.1	0.15	
		X_3	0	0		0.1	0.2	
			$(c_{23\|1})_1$	$(c_{13})_1$	$(c_{12})_1$	$(c_{23\|1})_2$	$(c_{13})_2$	$(c_{12})_2$
copulas	Setup 1,3		(N,0.41)	(N,0.41)	(N,0.41)	(N,0.06)	(N,0.06)	(N,0.06)
	Setup 2,4		(N,0.41)	(N,0.41)	(N,0.41)	(N,0.19)	(N,0.19)	(N,0.41)

and the joint model is defined as follows:

$$
\begin{aligned}
P(S_{cop,t+1} = j | S_{cop,t} = i) &= (P_{cop})_{i,j}\,, \\
P(S_{3,t+1} = j | S_{3,t} = i, S_{cop,t+1} = S_{cop,t}) &= \left(P_{3,n}^{S1(S2)}\right)_{i,j}\,, \\
P(S_{2,t+1} = j | S_{2,t} = i, S_{3,t+1} = S_{3,t}) &= \left(P_{2,n}^{S1(S2)}\right)_{i,j}\,, \\
P(S_{1,t+1} = j | S_{1,t} = i, S_{2,t+1} = S_{2,t}) &= (P_{1,n})_{i,j}\,, \\
P(S_{3,t+1} = j | S_{3,t} = i, S_{cop,t+1} \neq S_{cop,t}) &= (P_{3,s})_{i,j}\,, \\
P(S_{2,t+1} = j | S_{2,t} = i, S_{3,t+1} \neq S_{3,t}) &= (P_{2,s})_{i,j}\,, \\
P(S_{1,t+1} = j | S_{1,t} = i, S_{2,t+1} \neq S_{2,t}) &= (P_{1,s})_{i,j}\,,
\end{aligned}
$$

where matrices with superscript $S1$ ($S2$) correspond to simulation setup 1 (setup 2). This means, that if the dependence regime does not change ($S_{cop,t+1} = S_{cop,t}$), the probability of observing a regime change for the marginal distribution of X_3 is 0.1 (Matrix $P_{3,n}^{S1}$) for setup 1 and 0.05 (Matrix $P_{3,n}^{S2}$) for setup 2. If the dependence regime changes ($S_{cop,t+1} \neq S_{cop,t}$), the probability of observing a regime change for the marginal distribution of X_3 is 0.8 (Matrix $P_{3,s}$. Similarly, the transition probability for the marginal regime of X_2 (X_1) depends on whether we observe a change for S_3 (S_2).

In setup 3, the transition matrix for the shared state variable is the same as $P_{3,n}^{S1}$, while $P_{3,n}^{S2}$ is used for setup 4.

For all setups, the simulation was run 100 times for a typical data size of $T = 1000$. We keep the true number of regimes and R-vine structures with copula families which we have used for the simulations. Starting values for the EM algorithms are obtained by fitting a constant model for the whole data set first. Observations are then clustered according to their likelihood, and the model is re-fitted to the observations in the upper-half (for two regimes) / upper-third (for three regimes) to determine starting values for the first regime. These data points are then removed and we proceed similarly until starting values for all regimes are determined.

In setups 1 and 2 we explore the two-step approach and estimation using the joint EM algorithm where we estimate the full transition matrix. In setups 3 and 4, the joint EM algorithm is restricted to enforce $S_{1,t} = S_{2,t} = S_{3,t} = S_{c,t}$. The observed bias and root mean squared error (RMSE) for the model parameters are reported in Table 4.3 and Table 4.5 for setups 1,2 and setups 3,4, respectively.

As the parameter estimates for setups 1 and 2 reveal, the observed bias and RMSE of the joint estimation procedure are close to those of the two-step estimation approach. Also the correct classification rate (Table 4.4), which is given by the number of observations

Table 4.3: Observed bias and RMSE (in brackets) for the model parameters in setups 1 and 2 of the simulation study.

			two-step					
			$\mu_1 \cdot 10^2$	$\mu_2 \cdot 10^2$	$\mu_3 \cdot 10^2$	$\sigma_1 \cdot 10^2$	$\sigma_2 \cdot 10^2$	$\sigma_3 \cdot 10^2$
margins	Setup 1	X_1	−0.5 (2.4)	0.3 (2.0)	0.1 (2.7)	−2.3 (3.4)	−3.4 (5.1)	−1.3 (3.3)
		X_2	−0.3 (1.0)	0.6 (2.6)		−0.3 (0.9)	−0.2 (0.8)	
		X_3	−0.1 (0.9)	0.0 (0.8)		−0.1 (1.1)	0.1 (1.1)	
margins	Setup 2	X_1	0.1 (2.1)	0.1 (1.8)	0.2 (2.4)	−1.8 (2.7)	−2.7 (4.5)	−0.5 (3.5)
		X_2	−0.1 (0.9)	0.3 (1.6)		−0.1 (0.7)	−0.1 (0.7)	
		X_3	0.0 (0.7)	0.0 (0.9)		−0.0 (0.7)	0.1 (0.8)	
			$(\tau_{23\|1})_1 \cdot 10^2$	$(\tau_{13})_1 \cdot 10^2$	$(\tau_{12})_1 \cdot 10^2$	$(\tau_{23\|1})_2 \cdot 10^2$	$(\tau_{13})_2 \cdot 10^2$	$(\tau_{12})_2 \cdot 10^2$
copulas	Setup 1		−4.5 (5.7)	−3.6 (5.1)	−5.8 (6.9)	−0.7 (4.7)	−0.4 (4.1)	−0.6 (3.7)
	Setup 2		−2.4 (6.3)	−1.2 (5.6)	−3.2 (6.4)	−1.3 (4.7)	−2.5 (6.2)	−6.6 (7.7)

			joint					
			$\mu_1 \cdot 10^2$	$\mu_2 \cdot 10^2$	$\mu_3 \cdot 10^2$	$\sigma_1 \cdot 10^2$	$\sigma_2 \cdot 10^2$	$\sigma_3 \cdot 10^2$
margins	Setup 1	X_1	−0.5 (2.8)	0.3 (3.0)	0.2 (2.9)	−2.5 (3.5)	−3.0 (4.7)	0.2 (3.0)
		X_2	−0.6 (1.0)	2.5 (3.4)		−0.2 (0.7)	−0.7 (1.3)	
		X_3	−0.0 (1.0)	0.0 (1.1)		−0.3 (1.1)	1.0 (1.4)	
margins	Setup 2	X_1	0.2 (2.3)	0.3 (2.6)	0.3 (2.6)	−1.8 (2.7)	−2.1 (4.3)	0.5 (3.5)
		X_2	−0.2 (0.7)	1.1 (2.0)		−0.2 (0.6)	−0.3 (0.9)	
		X_3	0.0 (0.7)	0.1 (1.0)		−0.2 (0.7)	0.8 (1.2)	
			$(\tau_{23\|1})_1 \cdot 10^2$	$(\tau_{13})_1 \cdot 10^2$	$(\tau_{12})_1 \cdot 10^2$	$(\tau_{23\|1})_2 \cdot 10^2$	$(\tau_{13})_2 \cdot 10^2$	$(\tau_{12})_2 \cdot 10^2$
copulas	Setup 1		4.8 (7.4)	4.8 (6.9)	5.1 (7.5)	−2.3 (6.7)	−4.3 (7.6)	−1.7 (6.0)
	Setup 2		3.9 (7.9)	4.0 (6.5)	2.7 (6.5)	−0.2 (5.5)	−3.2 (6.7)	1.6 (5.5)

Table 4.4: Average correct classification rate in setups 1 and 2 of the simulation study in [%], the standard deviation is given in brackets.

	two-step				joint			
	S_1	S_2	S_3	S_{cop}	S_1	S_2	S_3	S_{cop}
Setup 1	56 (7)	71 (3)	75 (4)	78 (3)	56 (6)	73 (3)	72 (4)	75 (4)
Setup 2	62 (9)	80 (3)	83 (3)	72 (9)	65 (8)	84 (3)	82 (3)	72 (6)

for which the estimation procedure assigns the highest probability to the true regime they were simulated from, is similar. It is influenced mainly by the number of regimes (higher classification rate in the presence of two regimes than for three regimes) and the persistence of the regimes (higher persistence in simulation setup 2).

In setups 3 and 4, however, where the marginal time series and the copula all share a common state variable, the situation is different. Here, in particular the bias and RMSE of copula parameter estimates is reduced for the joint estimation algorithm (Table 4.5). Also the classification rate improves to 97% and 98% in setups 3 and 4, respectively (Table 4.6). Comparing the results for setup 3 and setup 4, we see that the higher persistence of regimes in setup 4 again improves our estimators.

Therefore, when being interested in estimating the parameters corresponding to different regimes and correctly classifying observations, the analysis can be conducted in the simpler two-step setup when the dependence between the different state variables is not too strong instead of estimating the high number of parameters in the full transition matrix. When we expect perfect dependence between the state variables, however, using the restricted joint EM algorithm for one common state variable greatly improves the estimation results.

Table 4.5: Observed bias and RMSE (in brackets) for the model parameters in setups 3 and 4 of the simulation study.

			two-step					
			$\mu_1 \cdot 10^2$	$\mu_2 \cdot 10^2$		$\sigma_1 \cdot 10^2$	$\sigma_2 \cdot 10^2$	
margins	Setup 3	X_1	0.0 (0.5)	0.2 (1.9)		0.0 (0.5)	0.0 (1.6)	
		X_2	−0.2 (0.9)	0.1 (1.4)		−0.2 (0.9)	−0.1 (1.4)	
		X_3	0.0 (0.6)	0.0 (1.1)		0.0 (0.6)	−0.1 (1.1)	
margins	Setup 4	X_1	0.0 (0.5)	0.1 (1.9)		0.1 (0.4)	0.1 (1.3)	
		X_2	−0.1 (0.7)	0.2 (1.0)		−0.1 (0.4)	−0.1 (0.6)	
		X_3	−0.1 (0.6)	0.1 (1.0)		−0.1 (0.5)	−0.2 (0.9)	
			$(\tau_{23\|1})_1 \cdot 10^2$	$(\tau_{13})_1 \cdot 10^2$	$(\tau_{12})_1 \cdot 10^2$	$(\tau_{23\|1})_2 \cdot 10^2$	$(\tau_{13})_2 \cdot 10^2$	$(\tau_{12})_2 \cdot 10^2$
copulas	Setup 3		−2.8 (4.7)	16.8 (17.0)	8.3 (8.8)	−1.0 (2.7)	−3.6 (4.5)	−2.9 (3.9)
	Setup 4		−0.5 (3.6)	12.0 (12.5)	5.4 (6.3)	−2.6 (4.0)	−6.6 (7.6)	−8.5 (9.2)

			joint					
			$\mu_1 \cdot 10^2$	$\mu_2 \cdot 10^2$		$\sigma_1 \cdot 10^2$	$\sigma_2 \cdot 10^2$	
margins	Setup 3	X_1	0.0 (0.4)	0.2 (1.9)		0.1 (0.4)	0.0 (1.4)	
		X_2	0.0 (0.5)	0.1 (0.8)		0.0 (0.3)	0.0 (0.5)	
		X_3	0.0 (0.4)	−0.1 (1.1)		0.0 (0.3)	0.1 (0.6)	
margins	Setup 4	X_1	0.0 (0.4)	0.1 (1.9)		0.1 (0.4)	0.1 (1.3)	
		X_2	0.0 (0.5)	0.1 (0.7)		0.0 (0.3)	0.0 (0.5)	
		X_3	0.0 (0.5)	0.1 (1.0)		0.0 (0.3)	0.0 (0.7)	
			$(\tau_{23\|1})_1 \cdot 10^2$	$(\tau_{13})_1 \cdot 10^2$	$(\tau_{12})_1 \cdot 10^2$	$(\tau_{23\|1})_2 \cdot 10^2$	$(\tau_{13})_2 \cdot 10^2$	$(\tau_{12})_2 \cdot 10^2$
copulas	Setup 3		0.1 (2.6)	0.5 (2.5)	0.4 (2.6)	0.4 (3.0)	0.4 (3.3)	−0.1 (2.9)
	Setup 4		0.1 (2.6)	0.4 (2.5)	0.2 (2.8)	0.4 (2.8)	0.4 (3.2)	0.1 (2.3)

Table 4.6: Average correct classification rate in setups 3 and 4 of the simulation study in [%], the standard deviation is given in brackets.

	two-step				joint
	S_1	S_2	S_3	S_{cop}	S
Setup 3	91 (2)	80 (3)	80 (3)	85 (2)	97 (1)
Setup 4	95 (1)	87 (3)	87 (3)	84 (3)	98 (1)

Chapter 5

Application 1: Multivariate regime switching model of US exchange rates

In this section, we will apply the MS model of Section 4. The section is based on Stöber and Czado (2013). Arguably, misperceptions about extremal dependencies in credit portfolios or between financial assets during economic downturns and market stress have been an important cause of the recent financial crisis. In the reverberations of the 2007/2008 banking crisis, regulating entities have turned their attention towards the fact that financial time series exhibit different behavior under market stress. This has led to new requirements for financial institutions addressing this issue. For European financial institutions, the European Banking Authority (EBA) has introduced the Stressed Value at Risk (SVaR) in addition to the standard VaR measure (European Banking Authority 2012). Here, the underlying distributions, which are assumed to calculate VaR, have to be calibrated using a period of significant stress for the banks portfolio to appropriately reflect different behavior of time series during these times.

In the literature on financial time series, different behavior during times of market stress has long been recognized (e.g. in the seminal paper of Engle (1982) which shows that variances are not constant over time). MS models are time series models in which some model parameters are state dependent. These different states of the Markov model fit nicely into the regulatory framework requiring stressed and non-stressed market conditions to be incorporated in risk management. While SVaR models are often calibrated using an "intuitive" stress period in the years 2007-2009, Markov models allow to automatically detect theses stress periods.

In this section, we will present an extensive investigation of MS models for the dependence structure in US exchange rates. We illustrate how suitable copula structures to describe "normal" and "crisis" periods can be chosen based on intuition and prior knowledge of the data and illustrate how dependencies tend to change during times of market stress. Using DIC as a Bayesian criterion of model fit, we will demonstrate that the MS models we develop provide a more accurate description of the dependence structure than a constant copula model. Since practitioners will usually not be interested in the dependence structure of

financial data alone, we will subsequently use MS models also for marginal time series and illustrate briefly that marginal regimes tend to switch at the same time as dependence regimes for US exchange rates. For the modeling of SVaR, we will be interested in determining a single crisis period to estimate marginal distributions and the dependence structure. Therefore, we will finally impose this restriction of having a single underlying regime variable. The results suggest that using a crisis period from mid 2008 to mid 2009 is appropriate for US exchange rates.

5.1 Markov Switching US exchange rate dependence

We start our application of MS models by focussing on modeling the dependence structure of multivariate data. Hence, we apply a two step estimation approach as suggested by Joe and Xu (1996). In the first step, appropriate parametric models for the marginal time series are fitted separately and used to transform the standardized residuals to approximately uniform margins. To this transformed data, we apply the copula model in the second step. Before Bayesian or frequentist parameter inference for the MS-RV model can be conducted, appropriate R-vine structures and sets of bivariate copulas for each regime need to be selected in a preanalysis. To do so, we apply the heuristic model selection techniques as outlined in Dißmann et al. (2013) and Brechmann et al. (2012) which select an R-vine tree structure sequentially to capture the most important dependencies on the first trees.

Since we will be interested in detecting a stressed and a normal period, we assume the presence of two regimes for models we consider here. Unless mentioned otherwise, the copula families we will consider are the standard Gaussian (N) copula and the tail-dependent Gumbel (G) copula. Since the Gumbel copula is not invariant with respect to rotations, we consider its standard form and rotations by 90° (G90), 180° (survival Gumbel, SG) and 270° (G270), respectively. For all models studied, we run the MCMC for 20000 iterations discarding the first 1000 as burn-in, and keep every fifth observation to reduce autocorrelations. For estimating quantiles of the posterior distribution, we further thin the output according to what Kass et al. (1998) call "effective sample size" (cf. Carlin and Louis (2009)). After this, we end up with ca. 1000 approximately i.i.d. samples.

The section is structured as follows: Subsection 5.1.1 introduces the exchange rate data we analyze, and an initial MS-RV model in which only the copula parameters are switching is fitted in Subsection 5.1.2. To gain additional information about possible regime switches also in the copula families we conduct a rolling window analysis in Subsection 5.1.2.1. Subsequently, R-vine copula structures for the "crisis" regime are selected in Subsection 5.1.3.

Subsection 5.1.4 presents findings of our analysis, and the in-sample fit of all analyzed MS-RV models is compared in Subsection 5.1.5.

5.1.1 Data description

The data set consists of 9 exchange rates against the US dollar (USD), namely the Euro (EUR), British pound (GBP), Canadian dollar (CAD), Australian dollar (AUD), Brazilian real (BRL), Japanese yen (JPY), Chinese yuan (CNY), Swiss franc (CHF) and Indian rupee (INR). The observed time period is from July 22, 2005 to July 17, 2009, resulting in 1007 daily observations and covering the default of Lehman Brothers and the 2007/2008 financial crisis. The modeling of the one dimensional margins with appropriate ARMA-GARCH models and the transformation to copula data has been performed by Czado et al. (2012). The distributions which have been chosen for the innovations of each time series and QQ plots and the results of diagnostic Ljung-Box tests on the residuals are available in Schepsmeier (2010, p. 91ff.). In total, we consider 6 models for the dependence structure of this data which will be defined as we proceed, their defining tree structures and the allowed copula families are listed in Table 5.1.

Table 5.1: R-vine models considered for the exchange rate data with tree structures $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$, copula families and parameters given in the appendix.

Model	Defined in Section	Regime 1 (no crisis)	Regime 2 (crisis)	Copulas regime 1		Copulas regime 2	Parameters
(1)	5.1.2	$\mathcal{V}_{11} = \mathcal{V}_1$	$\mathcal{V}_{12} = \mathcal{V}_1$	mixed	=	mixed	Table D.1
(2a)	5.1.3	$\mathcal{V}_{21} = \mathcal{V}_1$	$\mathcal{V}_{22} = \mathcal{V}_2$	N		SG	Table D.2 & Table D.3
(2a*)		$\mathcal{V}_{21} = \mathcal{V}_1$	$\mathcal{V}_{22} = \mathcal{V}_2$	N		SG, N	
(2b)		$\mathcal{V}_{21} = \mathcal{V}_1$	$\mathcal{V}_{22} = \mathcal{V}_2$	N		G	
(2c)		$\mathcal{V}_{21} = \mathcal{V}_1$	$\mathcal{V}_{22} = \mathcal{V}_2$	N		Student's t	
(3)	5.1.3	$\mathcal{V}_{31} = \mathcal{V}_1$	$\mathcal{V}_{32} = \mathcal{V}_3$	mixed	$\neq$	mixed	Table D.5

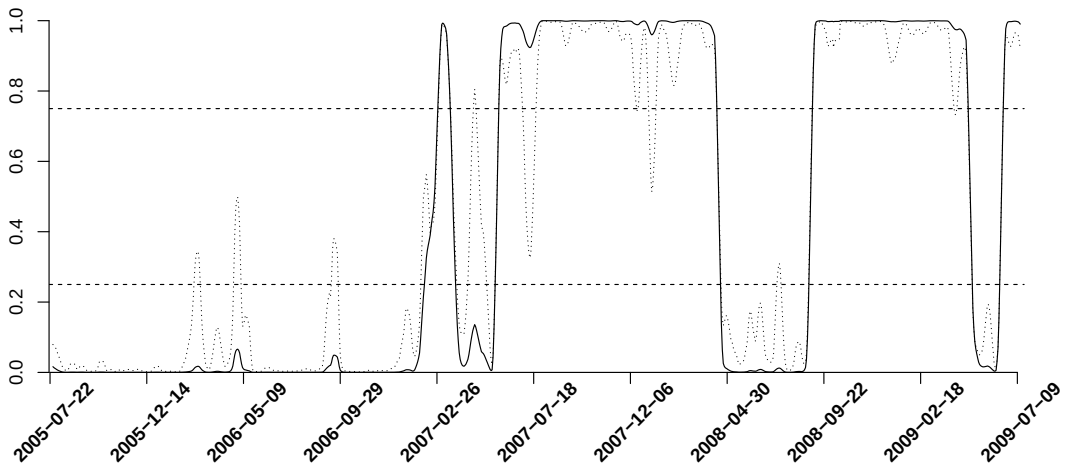
5.1.2 R-Vine copula with switching parameters

As a first model, we consider a common R-vine copula structure for the two regimes but with different parameters. To do so, we fit an R-vine with corresponding bivariate copulas to the data using the procedure of Dißmann et al. (2013). Since the estimated pair-copula parameters corresponding to higher trees indicate conditional independencies, we truncate

the R-vine copula after the second tree, i.e. we associate all edges on higher trees with independence copulas. The R-vine copula structure $\mathcal{V}_1$ resulting from this procedure is given in the appendix. We call this model Model (1).

Figure 5.1 shows the probability $P(S_t = 2|\tilde{\mathbf{u}}_T, \hat{\boldsymbol{\theta}}^{EM})$ that the hidden state variable S_t indicates the presence of regime 2 plotted against time. While regime 1 is predominant until around February 2007, regime 2 becomes more important during the later times of the financial crisis. Analyzing the parameter estimates $\hat{\boldsymbol{\theta}}_1^{EM}$ and $\hat{\boldsymbol{\theta}}_2^{EM}$ (Table D.1 in the appendix) for the two regimes, we find that regime 1 has stronger dependencies on the first tree, whereas regime 2 has stronger dependencies on the second tree. In particular, regime 2 exhibits stronger conditional negative dependencies reflected by rotated Gumbel copulas, thus creating a more asymmetric dependence structure.

Figure 5.1: Estimated probabilities of being in state 2 for Model (1). (solid: EM estimates, dotted: Bayesian estimates) The estimates where additionally smoothed using an MA(7) filter to make the differences more visible.



In order to apply our Bayesian estimation procedure, we need to distinguish both regimes to avoid model identification problems. For a detailed consideration of this issue we refer to Frühwirth-Schnatter (2001). Using our observations with regard to the strength of dependence in the two regimes identified by the EM-Algorithm, we define regime 1 to correspond to weaker conditional dependence on the second tree and regime 2 to correspond to stronger dependence on the second tree, compared by the sum of absolute values of Kendall's τ corresponding to parameters $\boldsymbol{\theta}_1^{r,\text{MCMC}}$ and $\boldsymbol{\theta}_2^{r,\text{MCMC}}$. The resulting posterior probability estimates for the state variable, i.e.

$$\hat{P}(S_t = 1|\tilde{\mathbf{u}}_T) := \frac{1}{R}\sum_{r=1}^{R} S_t^{r,\text{MCMC}},$$

for R independent MCMC samples, are plotted as dotted points in Figure 5.1. These Bayesian estimates follow those obtained from the EM algorithm closely while showing a slightly higher degree of variability.

5.1.2.1 Rolling window analysis

Having identified parameter switches in an R-vine copula model for our data set, we will now try to identify switches in the overall dependence structure. Since there is empirical evidence that dependence structures can change in times of crisis (cf. Longin and Solnik (1995), Ang and Bekaert (2002) or Garcia and Tsafack (2011)) and since tail dependencies become more important in times of extremal returns, we want to select two different R-vine copula structures. To do so, we start with a rolling window analysis, selecting and fitting R-vine models to a rolling window of 100 data points. To reduce model complexity, we decide to work again with truncated R-vines, resulting in a sufficiently flexible and parsimonious model. The copulas on the first tree were chosen to be either all Gaussian, all Gumbel or all survival Gumbel. The copulas on the second tree were set to Gaussian and the R-vines were truncated after this second tree. The resulting rolling log likelihoods are given in the left panel of Figure 5.2. For an AIC (BIC) comparison this is sufficient since the number of parameters remains the same in all models considered.

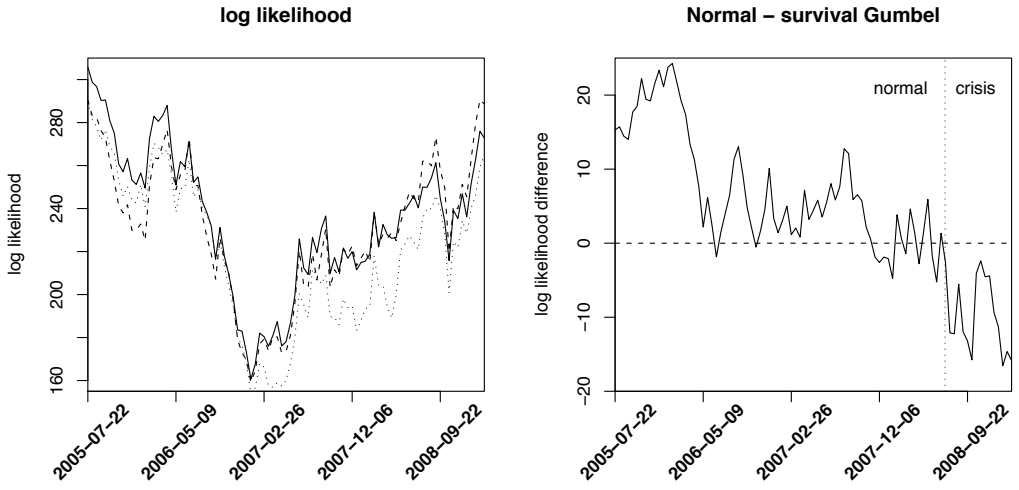
We see that, while the range of overall likelihood estimates is similar, the Gaussian model tends to give the best fit, i.e. highest log likelihood, (left panel of Figure 5.2) over the whole data set. However, the survival Gumbel model starts to outperform the Gaussian model towards the end of the observation period (right panel of Figure 5.2). Furthermore, the survival Gumbel model, in which the exchange rates taken into consideration are assumed to be lower tail dependent, tends to outperform the model with standard Gumbel copulas, corresponding to upper tail dependence. This is in accordance with the observation that the financial crisis during the observation period originated in the dollar area, quickly spreading to the world economy but with different severity e.g. to other developed countries and developing countries. Because of this, cash flows out of the dollar area, resulting in higher Foreign Currency/US dollar exchange rates, tend to be less extremely correlated than cash flows into the dollar area to settle liabilities denominating in US dollar, which results in more lower than upper tail dependence.

5.1.3 Identifying crisis regimes

Based on the observations from the rolling window analysis we will now define Models (2a) - (2c) and (3). For Models (2a) - (2c) we incorporate our knowledge about the evolution of the financial crisis into the model selection. They will be used to investigate the influence of different tail dependence structures in the modeling of the exchange rate data. Since restricting to a particular kind of tail dependence will make the MS-RV model look less favorable as compared to classical models in terms of goodness of fit, we do further consider Model (3). Here, we select tree structures in a semi-automatic way based on the rolling window analysis and allow for all bivariate copula families, i.e. the Gaussian copula and all rotations of the Gumbel copula.

For Models (2a) - (2c), we decide to select R-vine copulas as follows: For regime 1, the tree structure ($\mathcal{V}_1$) is again fitted to the whole data set, but we use only Gaussian copulas as bivariate building blocks. Since parameter estimates on the higher trees indicate weak dependence, we truncate the R-vine copula after the second tree. To determine a second structure ($\mathcal{V}_2$) we apply the procedure of Dißmann et al. 2013 to the time frame from July 10, 2008 to December 3, 2008 (first 100 days of the "crisis" period indicated in Figure 5.2). Doing so, we capture many high-impact events of the financial crisis. For the copulas on

Figure 5.2: Left-panel: log likelihood values resulting from fitting R-vines with normal (solid), survival Gumbel (dashed) and Gumbel (dotted) copulas. Right panel: difference between the values for the normal and survival Gumbel model, we indicate the period to which the structure for Model (3) is fitted. The dates are the starting observations of the rolling windows.



the first tree associated to $\mathcal{V}_2$ we consider survival Gumbel copulas (strong dependence for negative returns, Model (2a)), Gumbel copulas to capture dependencies in the upper tail (Model (2b)) and Student's t copulas to cover symmetric tail dependencies (Model (2c)). The copulas corresponding to edges on the second tree are again chosen to be Gaussian and we truncate after Tree 2. While the survival Gumbel model is preferred in the rolling window analysis, we include Models (2b) and (2c) to investigate the impact of different tail dependencies. The resulting smoothed probabilities for being in the non-Gaussian regime using Models (2a) - (2c) are given in Figure 5.3.

For Model (3), the R-vine tree structure $\mathcal{V}_3$ together with the corresponding copulas for the "crisis" regime is selected by applying the stepwise selection procedure of Dißmann et al. (2013) to the data points where the rotated Gumbel copula is outperforming the normal copula in the rolling window analysis (July 10, 2008 to July 17, 2009, annotated with "crisis" in Figure 5.2). The R-vine structure for the "normal" regime with corresponding copulas as identified from the remaining data points (July 22, 2005 to July 9, 2008) coincides with the structure for the "normal" regime in Models (2a) - (2c), $\mathcal{V}_1$. While the pair copulas corresponding to this tree structure were all chosen to be Gaussian before, we also allow for all rotations of the Gumbel family here to make use of the full flexibility the MS-RV model provides.

We employ the EM procedure for an initial fit of MS-RV models with the selected regimes. Subsequently, a Bayesian analysis using the Gibbs sampler and prior assumptions of Section 4.1.2.2 is performed.

5.1.4 Empirical findings

While the overall strength of dependence modeled in the two regimes (judging by the fitted values of Kendall's τ, Appendix D.1) is similar for all models, the results for Model (2c) with Student's t copulas are close to the results of Model (1), whereas the other two differ significantly. This was expected, since the model with t-copulas is close to a Gaussian copula model with regime switching parameters. Analyzing the estimated Kendall's τs further, we find that in Model (2c) the τ between the JPY-USD and the INR-USD exchange rate indicates negative dependence. Since the Gumbel and survival Gumbel family model positive dependence, this cannot be captured in Model (2a) or (2b), respectively. Replacing the copula for this bivariate margin by a Gauss copula (Model (2a*)), so that it captures the negative dependence, does however not significantly change the posterior estimates of the hidden state variable. This means that the observed difference in the behavior of Models (2a) and (2b) as compared to Model (2c) cannot be explained by the lack of Gumbel and Gumbel survival

copulas to allow for negative dependence. Instead, these models tend to be preferred during specific times of high impact events of the financial crisis, where the bivariate dependence structures are closer to a Gumbel copula, as indicated in the top panel of Figure 5.3 where important events are annotated.

The probability of being in a given state at a given time is a function of the observations from the multivariate time series $\tilde{\mathbf{u}}_T$ and the model parameters $(\boldsymbol{\theta}_{cop}, a, b)$, i.e.

$$p_{t,i} = P(S_t = i|\tilde{\mathbf{u}}_T, a, b, \boldsymbol{\theta}_{cop}),$$

is determined by (4.3) for given $(\boldsymbol{\theta}_{cop}, a, b)$. This means that we also obtain the posterior distribution of the state probability $p_{t,i}$, from which we can compute CIs (see Figure 5.3). The obtained pointwise 90% credible intervals (CIs) are quite narrow for the models (2a) - (2c) which shows that the time variations which are detected are in fact characteristics of the data.

Figure 5.4 shows several marginal posterior density estimates for the copula parameters in the crisis regime of Model (2a*). As we can see, the parameter value of $\tau_{INR-JPY} = 0$ which would correspond to independence is nowhere near a 90 or 95% CI, the dependence is significantly negative. For the copula between Brazil-US and China-US in contrast, the parameter values in our posterior sample are all close to 0, which means that the two time series are only weekly dependent or maybe independent.

5.1.5 Model comparison

Having discussed the stylized features of the investigated RV-MS models, we want to compare them in terms of their fit. First, we rely on in-sample methods, and use our Bayesian Gibbs sampling procedure to calculate the deviance information criterion (DIC) which has been proposed by Spiegelhalter et al. (2002). Table 5.2 shows DIC values for all models under investigation. For comparison, we also include an R-vine model without regime switches, but where the vine tree structure has not been truncated after tree 2. The first two trees of this structure correspond to Structure $\mathcal{V}_1$, it has been selected using Dißmann et al. 2013 for the full data set. To illustrate the performance of the unsupervised selection heuristic introduced in Section 4.1.4, which is based on the EM algorithm and also on the algorithm of , as compared to the manually selected regimes of this section, we further include Model (4) which has been selected by this procedure. We refer to the selected "normal" and "crisis" structures as $\mathcal{V}_{4n}$ and $\mathcal{V}_{4c}$, respectively. The first and second trees of these structures together with the chosen copula families and Kendall's τs corresponding to posterior mean estimates of the copula parameters are given in Appendix D.1.

Figure 5.3: Smoothed probabilities that the hidden state variable indicates the non-Gaussian regime. Models are from top to bottom: (2a), (2b), (2c). The solid lines correspond to Bayesian MCMC estimates, the dotted lines to 90% CIs. High impact crisis events are annotated in the upmost graph.

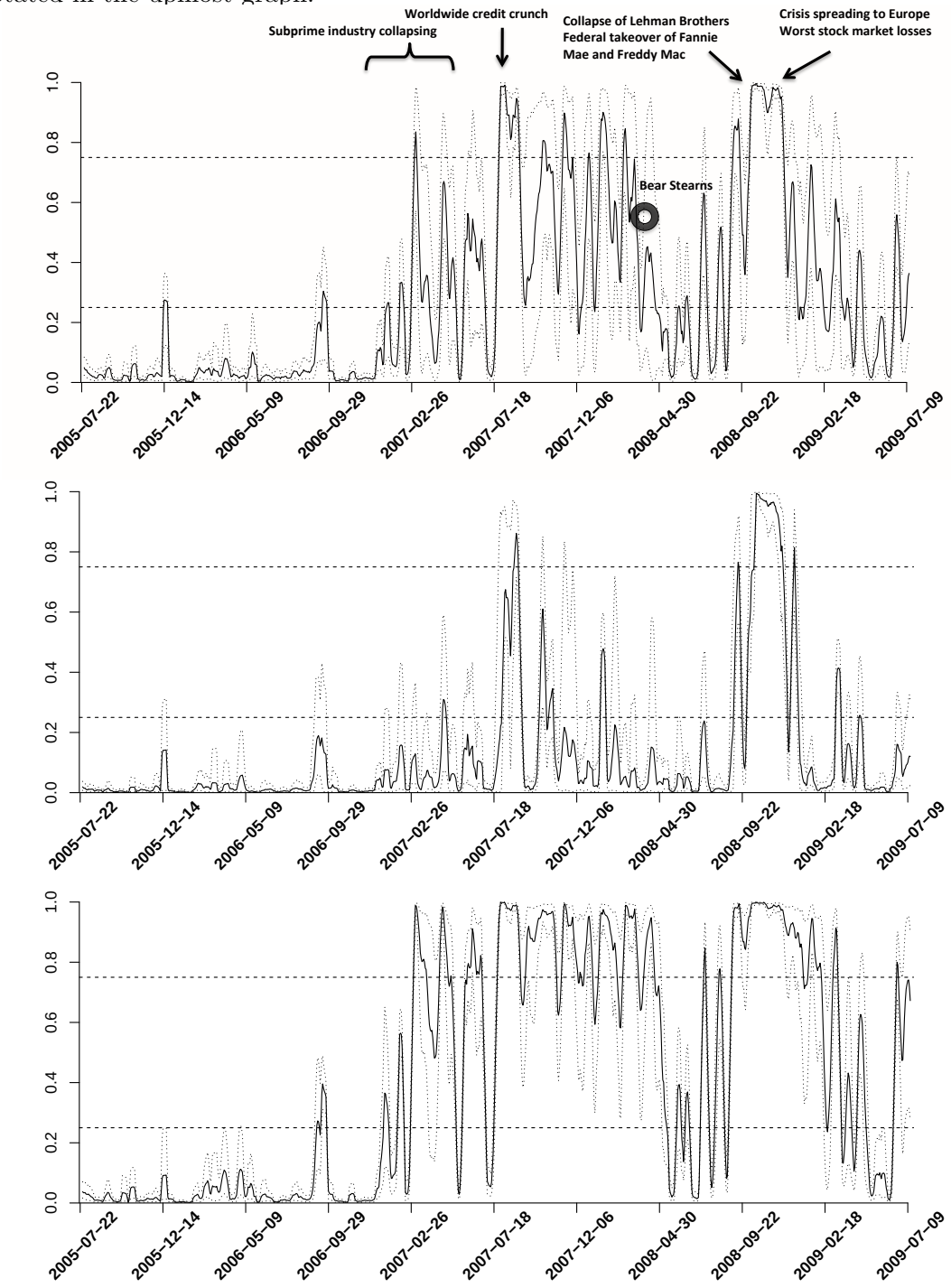


Figure 5.4: Estimated posterior densities of Kendall's τ in the crisis regime of Model (2a*) with 90% CIs. The plotted densities correspond to the parameters of the (unconditional) copulas associated to Tree 1 of the vine $\mathcal{V}_2$.

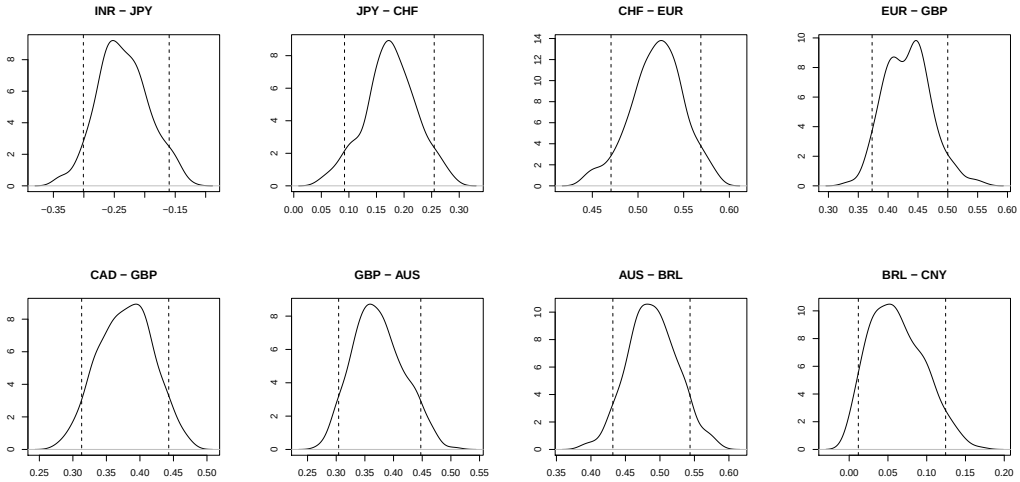


Table 5.2: DIC values for the different (MS) R-vine models that have been considered. Lower values indicate a better fit of the model to the data.

Model	(1)	(2a)	(2a*)	(2b)	(2c)	(3)	(4)	no MS
DIC	-4398	-4280	-4312	-4199	-4346	-4430	-4666	-4146

Although the full R-vine model has 36 parameters and the MS-RV models where we use truncated vines and one parametric pair copulas only have 32, even the worst MS-RV copula outperforms the model without Markov structure, which clearly supports the use of models with time-varying dependence. The DIC values further show that in terms of in-sample fit, the model with standard Gumbel copulas in the crisis regime is outperformed by the other models, which was to be expected from the rolling window analysis. Since the copulas in Model (1) were chosen maximizing pairwise AIC, it outperforms the models where we restricted the choice of copulas. The best-performing models however are Model (3), where the copula families were chosen using pairwise AIC but the R-vine structure differs between the regimes, and Model (4) which was selected by the heuristic of Section 4.1.4. While a higher DIC value for Model (4) was expected since it contains less independence copulas, the good performance of Model (3) shows that RV-MS models with switching vine structure and parameters are more suitable for this kind of data than models where only the copula parameters are switching.

It has been recognized in the past, that while MS models usually provide very good in-

sample fit and help to describe and understand time variation patterns in data, they are often not very good for forecasting (Engel 1994). However, for forecasting in a crisis period, we would expect the MS model to outperform a constant model, since it explicitly allows for structural breaks to occur and does not average over the whole past. To provide also a out-of-sample measure of model fit, we further compared the MS copula model to a constant R-vine model using one day ahead log predictive scores (see Section 1.3.1) for the period from September 19, 2008 to February 17, 2009. This period consists of 100 business days, and we compute the log predictive scores as follows: To select a suitable R-vine tree structure and associated pair-copulas, we apply the model selection heuristic presented in Section 4.1.4 to the period of July 22, 2005 to September 18, 2008. Then, we apply the EM-Algorithm to estimate the model parameters and obtain a predictive density for the September 19 return, from which we calculate the first out-of-sample log score. For comparison, we do also select a constant R-vine tree structure using the algorithm of Dißmann et al. (2013) and calculate the log score also for this model. Subsequently, the models are re-fitted to the data period including September 19, and we calculate log predictive scores for the next business day. For the 100 business days between September 19, 2008 and February 17, 2009, this results in an average log predictive score of 2.10 for the MS model and 1.67 for the constant R-vine model. This is in line with the DIC results and illustrates that at least in crisis periods MS models can also be useful for econometric forecasting.

5.2 Regime switching marginal time series

In this section, we will fit MS models both for the marginal time series of the exchange rate data and the copula. First, we will fit models using the two-step approach. Marginal time series are fitted first, and the copula model is subsequently applied to the transformed residuals. Subsequently, we will fit a joint model where all marginal time series and the dependence structure switch states at the same time. The parameter estimates for both models are given in Appendix D.2.

For the marginal exchange rate return time series, we consider simple MS Gaussian models, i.e. Markov switching means and volatilities. MS AR models have been investigated as an alternative but the likelihood improvements did not justify the additional parameters being included in the model. As expected, the MS Gaussian models outperform constant Gaussian models also in out-of-sample comparison (Table 5.3). The resulting smoothed probabilities of being in the crisis regime (which will be a high volatility regime for the marginal time series) is given in Figure 5.5 for the EUR/USD exchange rate and the copula, for the remaining

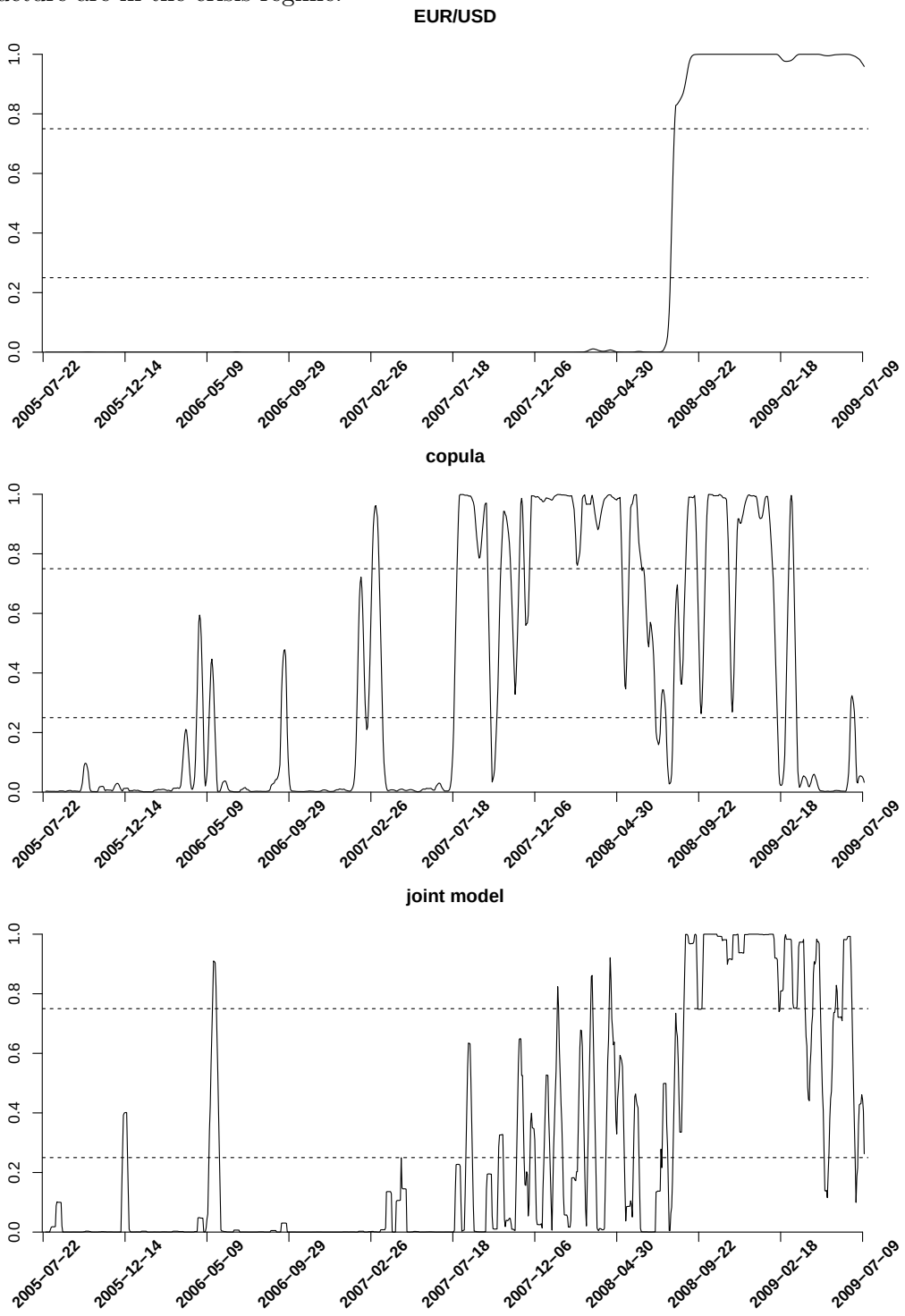
marginal time series they are illustrated in Appendix D.2. Our results show that for all major currencies the crisis periods are more or less identical, although some exchange rates move to the crisis state already in the beginning of 2008 (e.g. CHF/USD) while others follow in late 2008 (e.g. EUR/USD). Only for the currencies of countries with less developed financial markets we see additional movements in the crisis probabilities, in particular for CNY, where we see a pattern which is probably due to Chinese monetary interventions.

Fitting a copula model to the transformed residuals of the marginal time series using the model selection heuristic of Section 4.1.4, we see that the results are similar to those for the models fitted using marginal ARMA-GARCH models in Section 5.1.3. Since all marginals and the copula are in "crisis" state from mid 2008 to early 2009, and since we would ideally want to find one crisis period for the whole data, we do also fit a model with only one state variable using the algorithm of Section 4.2.2. The resulting smoothed probability of being in the crisis regime is plotted in the lower panel of Figure 5.5. While we observe more movements in the crisis state probability pre 2008 than for most marginals, the joint model confirms that mid 2008 - mid 2009 should be used as a crisis period for FX portfolios according to this analysis.

Table 5.3: Average one day ahead log predictive scores for the marginal time series in the exchange rate data (September 19, 2008 - February, 17 2009). The procedure for calculating the predictive densities is the same as for the copula model in Section 5.1.5.

	EUR	GBP	CAD	AUD	BRL	CNY	JPY	CHF	INR
MS	2.58	2.55	2.47	1.89	1.84	5.10	2.75	2.75	3.20
constant	1.54	0.76	1.11	−0.76	−0.34	4.70	2.21	2.26	2.66

Figure 5.5: Upper panel: Smoothed probability of being in the crisis regime for the EUR/USD exchange rate returns when analyzed separately of the other margins.
Middle panel: Smoothed probability that the copula between the 9 exchange rate returns is in the crisis regime.
Lower panel: Smoothed probability that all individual time series and the dependence structure are in the crisis regime.



Chapter 6

Application 2: Comorbidity in the second longitudinal study of aging (LSOA II)

In this section, which is taken from Stöber et al. (2012), we will apply the R-vine copula model for mixed discrete and continuous margins to a data set from the second longitudinal study of aging (LSOA II, data is available from `http://www.cdc.gov/nchs/lsoa/lsoa2.htm`). The section is structured as follows: First, we introduce the data set in Subsection 6.1. Our statistical model introduced in Subsection 6.2 is based on generalized linear models (GLMs), describing the distribution of the marginal variables for given covariates and an R-vine PCC to describe the dependence structure. Results are presented in Subsection 6.3, Subsection 6.4 concludes and points out some limitations of our analysis.

6.1 Introduction and data description

Most countries in the developed world, and in particular the US where the LSOA II was conducted have an aging population (Werner 2010; U.S. Census Bureau 2012). While the advances in modern medicine have prolonged life expectancies and improved the outcome of previously fatal diseases, this comes with a higher proportion of older adults living for years with chronic conditions. The occurrence of chronic conditions was recorded in the LSOA II among other data. Here, we will focus on six of the most common chronic conditions: hypertension (hyp), diabetes (dia), arthritis (art), heart disease (hd), stroke (str), and obesity/underweight via the body mass index (BMI). Information on the presence of these conditions was collected at three times using standardized telephone interviews and self-administered questionaires: the baseline interview was done in 1994-1996 (wave 1). The same subjects had two consecutive follow-up interviews between 1997 and 1998 (wave 2), and between 1999 and 2000 (wave 3). The time gaps between consecutive interviews varied by individuals, but data was collected roughly in two year intervals. For more details on how data was collected, we refer to the original study and Stöber et al. (2012). While the

aforementioned chronic conditions are all well-studied in the medical literature, proper treatment can still be difficult since many patients develop "comorbid" conditions, which refers to one or more chronic conditions occurring together with the primary condition (Table 6.1 presents an overview of the prevalence of comorbidities in our sample). However, clinical practice guidelines are generally based on one isolated disease. Little information is available considering care for multiple chronic conditions (Lugtenberg et al. 2011).

The aim of the study presented in this section is to analyze the prevalence of chronic conditions in a systematic joint modeling framework and thus broaden our understanding of comorbidities. This might help to identify promising avenues for further clinical research. Ultimately the hope in multimorbidity research is that a better understanding of interactions

Table 6.1: Percentage of subjects with each condition who have another chronic condition.

		Wave 1 No. (%) With Comorbid Condition					
Chronic condition	**No. of Subjects**	**Hyper-tension**	**Diabetes**	**Arthritis**	**Obesity**	**HD**	**Stroke**
Hypertension	893	192 (21.5)*	111 (12.4)	587 (65.7)	172 (19.3)	181 (20.3)	75 (8.4)
Diabetes	193	111 (57.5)	18 (9.3)*	140 (72.5)	51 (26.4)	59 (30.6)	19 (9.8)
Arthritis	1436	587(40.9)	140(9.7)	562(39.1)*	240(16.7)	285 (19.8)	92(6.4)
Obesity	336	172(51.2)	51(15.2)	240(71.4)	37(11)*	69(20.5)	16(4.8)
HD	434	181(41.7)	59(13.6)	285 (65.7)	69(15.9)	75 (17.3)*	47(10.8)
Stroke	139	75(54)	19(13.7)	92(66.2)	16(11.5)	47(33.8)	12(8.6)*
		Wave 2 No. (%) With Comorbid Condition					
Chronic condition	**No. of Subjects**	**Hyper-tension**	**Diabetes**	**Arthritis**	**Obesity**	**HD**	**Stroke**
Hypertension	1035	228(22)*	147 (14.2)	682 (65.9)	157 (15.2)	220 (21.3)	55 (5.3)
Diabetes	237	147 (62)	22 (9.3) *	156 (65.8)	57 (24.1)	71 (30)	17(7.2)
Arthritis	1470	682 (46.40)	156 (10.6)	523 (35.6)*	195 (13.3)	323 (22)	69 (4.7)
Obesity	282	157(55.7)	57 (20.2)	195 (69.1)	31 (11)*	45 (16)	8(2.80)
HD	441	220 (49.9)	71 (16.1)	323 (73.2)	45 (10.2)	50 (11.3)*	40 (9.1)
Stroke	92	55 (59.8)	17 (18.5)	69 (75)	8 (8.7)	40 (43.5)	5(5.4)*
		Wave 3 No. (%) With Comorbid Condition					
Chronic condition	**No. of Subjects**	**Hyper-tension**	**Diabetes**	**Arthritis**	**Obesity**	**HD**	**Stroke**
Hypertension	1115	239 (21.40)*	174(15.60)	741 (66.50)	164(14.70)	297 (26.60)	68 (6.10)
Diabetes	279	174 (62.40)	31 (11.10)*	188 (67.40)	59 (21.10)	100 (35.80)	26 (9.30)
Arthritis	1482	741 (50)	188 (12.70)	478 (32.30)*	191 (12.90)	386 (26)	81 (5.50)
Obesity	268	164 (61.20)	59. (22)	191 (71.30)	23(8.60)*	57 (21.30)	9 (3.40)
HD	545	297(54.50)	100 (18.30)	386 (70.80)	57(10.50)	70 (12.80)*	44 (8.10)
Stroke	116	68(58.60)	26 (22.40)	81 (69.80)	9(7.80)	44. (37.90)	13 (11.20)*

[1] For each pair of conditions, percentages are based on cases with no missing data on those two variables. Percentages do not add to 100 because some patients have more than two conditions.

[*] Subjects with that condition only.

and synergies between chronic diseases will help to facilitate prevention, diagnosis and treatment, lower the financial burden on the health care system, and increase patients' quality of life (Schäfer et al. 2010).

To control for the effects of known risk factors, we include several covariates in our model. Although many predictors might be potentially useful, we only concentrate on some of the most common. The incidence of chronic diseases is known to increase with age; gender affects the progression and prevalence of chronic diseases. Further, Fleischer et al. (2011) reported association of a socioeconomic gradient for education and income with the risk factor profile for chronic diseases. People coping with chronic diseases are particularly vulnerable to the hazardous health effects of tobacco use. Smoking can exacerbate and complicate symptoms of the chronic conditions. Therefore, sex, age, income, education, and smoking are included in our analysis.

6.2 Multivariate model

In this section, we introduce the joint model for the six response variables controlling for the covariates using GLMs and the copula paradigm. In a generic form, let Y_{ijt} be the response/outcome of the i-th patient for chronic disease j at observation/wave t, with $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, J$ and $t = 1, 2, \ldots, T$. The covariates we consider in our analysis for patient i, disease j and time observation t are accordingly denoted as $\mathbf{x}_{ijt}$.

6.2.0.1 Generalized linear models

For all j, t, we assume that Y_{ijt} are independent and have distribution function

$$F_j(y_{ijt}|\mu_{ijt}, \phi_{j,t}),$$

where the mean parameter $\mu_{ijt} = h_j(\mathbf{x}_{ijt}\boldsymbol{\beta}_{jt}^T)$ is a function of the covariates and ϕ_{jt} is a possible scaling parameter. In particular, for j corresponding to a continuous response variable (the BMI in the data set which we will consider later), F_j can be the inverse Gaussian distribution with distribution function

$$F_{ig}(y|\mu, \phi) = \Phi\left(\sqrt{\frac{\phi}{y}}\left(\frac{y}{\mu} - 1\right)\right) + e^{\frac{2\phi}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{y}}\left(\frac{y}{\mu} + 1\right)\right),$$

and h_j can be chosen as $h_j(.) = exp(.)$ (log-link). If j corresponds to a binary response variable indicating the presence/absence of a chronic disease, a natural choice for F_j is the

Bernoulli cdf with

$$F_b(y|\mu) = \begin{cases} 1 & y \geq 1 \\ \mu & 1 > y \geq 0 \\ 0 & 0 > y \end{cases}.$$

Here, the canonical choice for the link function h_j is $h_j = \frac{1}{1+e^{-(\cdot)}}$ (logit-link). For the selection and initial parameter estimation for marginal regression models we use the statistical software package R (R Development Core Team 2011). To select the relevant covariates and interactions from a given set of possible covariates we will apply the AIC (see Section 1.3.2). To minimize this criterion, a stepwise procedure starting with a fully saturated model (i.e. including all possible covariates and interactions) is applied, removing in each step the term with the highest possible reduction in AIC until no further reduction is possible. To fit the parameters of the GLMs, we apply iteratively reweighted least squares for maximum likelihood estimation (Green 1984).

6.2.1 R-vine copula selection

Furthermore, we assume that for any t, the marginal distributions F_j are linked with a copula function C_t. Hence, the joint distribution function for the outcome variables $(Y_{i,1,t}, \ldots, Y_{i,J,t})$ given covariates $(\mathbf{x}_{i1t}, \ldots, \mathbf{x}_{iJt})$ is given as

$$\begin{aligned} &F_t(y_{i,1,t}, y_{i,2,t}, \ldots, y_{i,J,t}|\mathbf{x}_{i1t}, \ldots, \mathbf{x}_{iJt}) \\ &\qquad = C_t(F_1(y_{i,1,t}|\mu_{i1t}, \phi_{1t}), F_2(y_{i,2,t}|\mu_{i2t}, \phi_{2t}), \ldots, F_J(y_{i,J,t}|\mu_{iJt}, \phi_{Jt})). \end{aligned} \tag{6.1}$$

The copula function C_t will result from an R-vine PCC. To select an appropriate R-vine structure and corresponding pair-copula families, the R-vine model selection procedure pioneered by Dißmann et al. (2013) which we have already used in Section 4.1.4 is modified as follows:

(i) For each pair of variables and each parametric pair copula family under consideration, calculate the corresponding value of the Akaike information criterion (AIC) from the copula data set. In a 3-dimensional example, we would calculate the AIC for the pairs $(1,3)$, $(2,3)$ and $(1,2)$.

(ii) Create a fully connected graph where the set of nodes N is the set of marginal variables (ex: $\{1,2,3\}$), and the set of edges E contains an edge between every possible pair of variables (ex: $(1,3)$, $(2,3)$ and $(1,2)$). Associate to each edge the highest AIC value which has been estimated for the corresponding variables in step (i) as edge weight.

(iii) Using the algorithm of Prim (1957) determine the maximum spanning tree corresponding to this graph, i.e. find a tree which maximizes the sum of edge weights (If the edge weights which are determined for the pairs $(1,2)$ and $(2,3)$ in our example are bigger than the weight of $(1,3)$, this is the tree T_1 in Figure 1.3, i.e. the tree containing edges $(1,2)$ and $(2,3)$).

(iv) For each edge in the resulting tree, choose the family for which we had obtained the highest AIC.

(v) For each pair of edges $i,k|D$ and $i,j|D$ sharing a common node, determine pseudo observations for the next tree by applying the conditional distribution functions $F_{k|i,D}$ and $F_{j|i,D}$ (Equations (2.11) and (2.12)) to the data. In the likelihood algorithm provided in Section 2.2, these are the values which are stored in matrices (2.15) and (2.16). Because of the proximity condition, these are all pseudo observations which might be required. (ex: $(1,2)$ and $(2,3)$ share 2, we compute $F_{1|2}$ and $F_{3|2}$) Proceed with the pseudo observations as in steps 1 to 4, while only considering edges which respect the proximity condition in step 2, until all trees together with their copula types and parameters are determined.

To decide which bivariate copula families to include in step (i) of the R-vine copula selection procedure and to demonstrate the superior predictive performance of our joint copula model compared to independent regression models, we perform a 10-fold cross-validation (see Arlot and Celisse (2010) for an overview of cross-validation procedures) as follows: The data is randomly partitioned into 10 patient sets of (almost) equal size. In each step, we leave out one of these subsets and apply the outlined variable selection procedure for the marginal models, followed by the described model selection heuristic for the R-vine copula. We consider 10 different sets of pair copula families for the R-vine copula selection as shown in Table 6.2. More details on the bivariate copula families and parametrization we use can be found in Schepsmeier and Stöber (2012). The prediction quality of the resulting models for the remaining data set is then compared using the log predictive score (see Section 1.3.1 and Gneiting and Raftery (2007) for a review of scoring rules). Table 6.2 lists the sum of log predictive scores for the 10 subsets where we subtracted the scores corresponding to the benchmark independence model.

The independence model is outperformed for all choices of family sets. Further, we see no indication of overfitting when more copula families are included, but a loss in predictive performance when some are excluded, in particular for the Gaussian and Frank copula. For this reason, we choose modelclass 9 with Gaussian, Frank, Clayton, Joe and Gumbel copulas

Table 6.2: Differences of log predictive scores to the independence models for different sets of copula families under consideration. We use the abbreviations N (Gauss), F (Frank), C (Clayton), J (Joe) and G (Gumbel). An extensive discussion of bivariate copula families and their properties can be found in Joe (1997) or Nelsen (2006).

Model class	Families	log predictive score
1	N	363.8
2	F	360.6
3	N, F	365.4
4	N, C	356.3
5	N, J	365.5
6	N, F, C	365.1
7	N, F, J	366.4
8	N, F, C, J	366.2
9	N, F, C, J, G	366.4
10	N, F, C, J, G + rotations by 90°, 180° and 270°	366.2

for our further analysis since we believe it to offer the best compromise between flexibility, predictive performance and computation tractability.

6.2.2 Selected joint model

For the whole data, the selected model is the following: For the GLMs, which we use for the modeling of marginal response variables, the covariates are given in Table E.1 in the appendix.

The dependence between these marginal models is then subsequently described using a discrete-continuous R-vine copula, with R-vine structures and associated pair copulas as shown in Figure E.1 and E.2 in the appendix. Comparison of the resulting model probabilities with the observed probabilities indicates that the model can accurately describe the observed dependence patterns.

Since the selection procedure of Section 6.2.1 selects the strongest dependencies (i.e. the dependencies where the corresponding copula terms lead to the biggest improvements in the joint likelihood) first. These are on the first trees of Figure E.1 and E.2. For the first tree T_1, copulas between BMI and diabetes, BMI and hypertension as well as heart disease and stroke are selected for all three waves of observations. This shows that these are the most important dependencies in the data. On the other hand, the copulas on higher trees correspond to

weaker conditional dependencies which might even be close to conditional independence as for the Gaussian copula on T_5 for the baseline observations ($\rho = 0.0749$, with $sd = 0.0551$, this corresponds to a p-value ≥ 0.1 and is non-significant).

In wave 2, a stronger dependence between heart disease and arthritis was observed as compared to wave 1. This information can be indicative to predict future comorbidity at the time of the baseline observation. We are not aware of any clinical literature investigating the longitudinal association between heart disease and arthritis; this dynamic association would be an interesting avenue for future research.

6.3 Results

Other than for purely continuous variables where the theoretical rank correlations and bivariate tail dependencies associated with a copula model are usually good summary statistics for the data, interpreting model results is more challenging in the presence of discrete outcomes. (Nevertheless, the theoretical Kendall's τ values corresponding to our parameter estimates in the continuous setup are included in Figures E.1 and E.2 in the appendix, for readers familiar with this parameterization.) Here, changes in strength of dependence can be expressed by different copula families being selected. In particular, the limiting dependence behavior (for large and small values of the continuous variable, respectively) of conditional distributions is different across copula families. While our inference procedure yields point estimates and standard errors for all model parameters and allows to compute p-values for the regression parameters we omit listing these estimates here (they are given in Appendix E). Instead, we compute conditional probabilities from our model to better understand the results. For example, we explored the conditional probabilities of each chronic condition given BMI by category of predictors such as age level. Here, conditional probabilities involving marginal covariates are estimated as follows: Let $\mathbf{x}_i$ be the vector of covariates for patient i, $z_1, z_2 \in \mathbb{R}$ and $\mathbf{Y}$ the vector of outcomes. Then

$$P(Y_{hyp.} = 1|BMI = z_1, age \leq z_2) := \sum_{i|x_{i,age} \leq z_2} \frac{P(Y_{hyp.} = 1|BMI = z_1, \mathbf{x}_i)}{\#\{i|x_{i,age} \leq z_2\}},$$

i.e. we average over all relevant covariate vectors in the population. When not conditioning on marginal covariates, we have

$$P(Y_{hyp.} = 1|BMI = z_1) := \sum_{i=1}^{N} \frac{P(Y_{hyp.} = 1|BMI = z_1, \mathbf{x}_i)}{N},$$

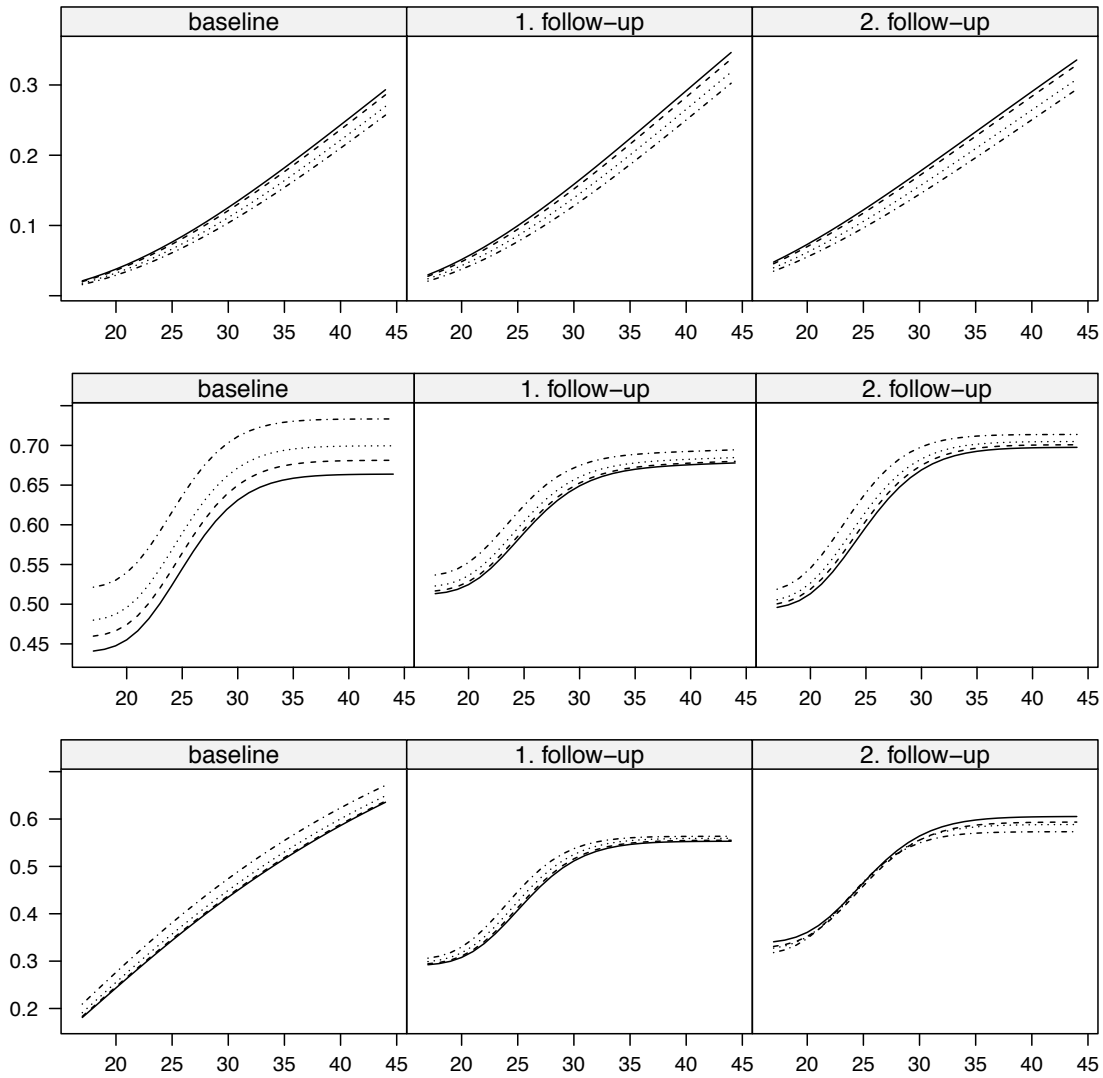
where N is the number of patients. Figure 6.1 depicts the relationship between a subject's BMI and the conditional probability of diabetes, arthritis and hypertension for the different

time periods. The top, middle, and bottom rows represent the patients with hypertension, diabetes and arthritis, respectively. The columns correspond to the different time periods; the first column corresponds to the baseline, the second and third column to the first and second follow-up, respectively. The different lines in each plot correspond to patients of different ages. The solid line is the mean level for patients of age ≤ 72 at the beginning of the study, the dashed for $72 <$ age ≤ 77, the dotted for $75 <$ age ≤ 78, and the dash-dotted for age > 78. In Figure 6.1, we can see that higher probabilities of observing the three diseases (hypertension, diabetes and arthritis) are associated with increasing BMI values.

Prevalence of diabetes given BMI First, the probability of diabetes (upper panel) is almost linearly increasing with BMI for all three periods. This positive association between diabetes and BMI (or obesity) has been reported for all ages (Nguyen et al. 2011) and it is widely accepted that BMI is one of the strongest predictors for diabetes. Therefore, sustained weight loss can bring a reduced risk of diabetes, as studied in Moore et al. (2000). Meanwhile, it is interesting to note that the prevalence of diabetes is slightly lower for the oldest age group in our sample, which might be explained by a decline in BMI which is generally observed after about 60 years of age (Elia 2001).

Prevalence of arthritis given BMI A different trend is observed for the prevalence of arthritis with respect to BMI: a family of S-shaped curves in the mid panel. This confirms a general positive association between BMI and arthritis which has previously been reported in studies for the overall population (Zakkak et al. 2009). However, these studies suggest a stronger increase for the heavily obese (BMI > 40) as compared to the group with $30 <$ BMI < 40 than we find in our sample. This different behavior which we observe might be attributable to a general decline in physical activity in the elderly population, since physical activity is associated both with obesity and with arthritis (Shih et al. 2006). For arthritis (and hypertension), the gap in disease prevalence between elderly with high and low BMI is reduced during the follow-ups compared to the baseline. This indicates that there are more important risk factors other than BMI which broadly affect the probability of observing arthritis among elderly. The influence of the age at the beginning of the study was most significant for arthritis at wave 1: older age groups are more prone to suffering from arthritis. This confirms the observation by e.g. Abyad and Boyer (1992) that arthritis increases with age. However, during the follow-up, the conditional probability gap of the arthritis between the oldest and youngest age group was not significant, which might be due to dropouts. Overall, the effect of BMI on the probability of arthritis exceeds the age effect.

Figure 6.1: Conditional probability of observing diabetes (upper panel), arthritis (mid panel) and hypertension (lower panel), respectively, given a certain value for BMI. The first column is the baseline; the second column is the first follow-up; the third column is the second follow-up. The solid line is the mean level for patients with ($age \leq 72$) at the beginning of the study (dashed: $72 < age \leq 75$, dotted: $75 < age \leq 78$, dash-dotted: $age > 78$).



Prevalence of hypertension given BMI Similar shapes are observable also for hypertension in the follow-up surveys, although our model suggests an almost linear association for baseline. Though systematic studies are scarce, a general increase of blood pressure with

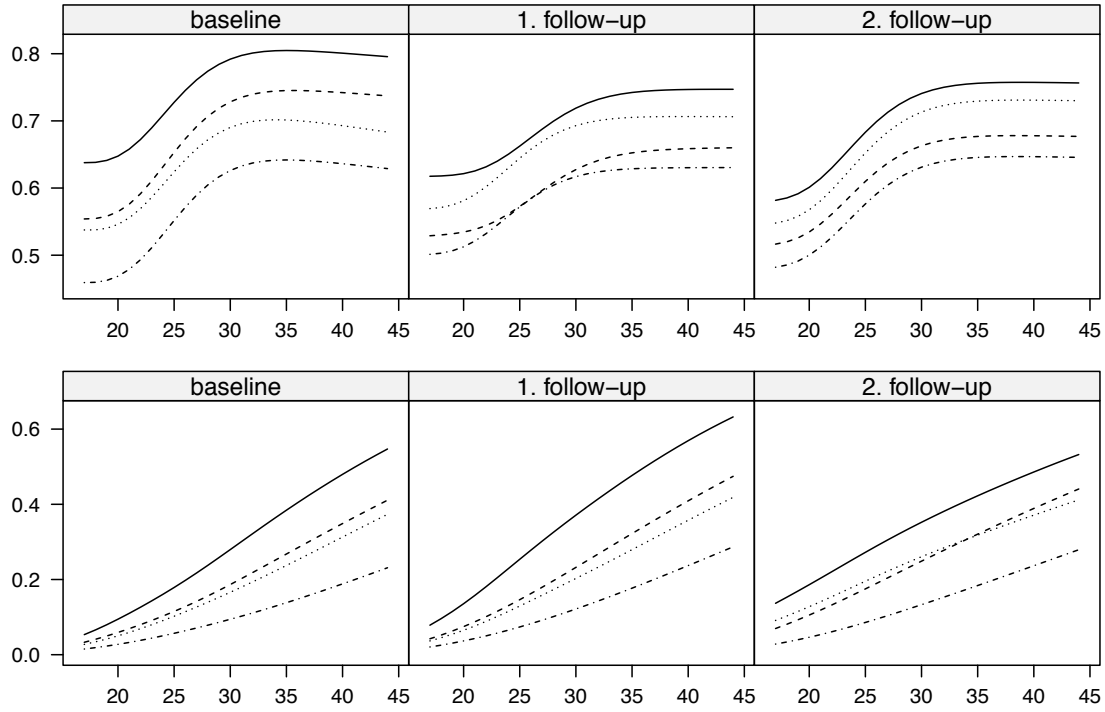
Table 6.3: The limiting behavior of conditional distribution functions $\partial_2 C(u_1, u_2)$ corresponding to well-known bivariate copula families (see Schepsmeier and Stöber (2012) for details on the pair copula families and parametrization). These limiting distributions do also appear in Cooke et al. (2011) and Hua and Joe (2012), where their relation to tail dependence is studied.

Copula Family	$u_2 \to 0$	$u_2 \to 1$
Clayton	1	$(u_1^{-\theta} - 1)^{-(1+1/\theta)}$
Gumbel	1	0
Joe	$(1-u_1)^{\theta-1}$	0
Frank	$\frac{e^\theta}{e^{\theta u_1}} \frac{e^{\theta u_1}-1}{e^\theta-1}$	$\frac{e^{\theta u_1}-1}{e^\theta-1}$
Gauss	1	0

BMI has been previously reported for elderly populations (Masaki et al. 1997). The change to an S-shaped trend in the follow-ups might again be explainable by the general decline in BMI with age. Although the difference between the age groups is non-significant especially for the second follow-up, note that while the older age groups had a higher prevalence of hypertension across all BMI levels at the baseline this trend is reversed in the follow-ups. The shift in the shape of the conditional probability curve of hypertension is expressed in the model by the fact that a Gaussian copula was selected for the baseline while a Frank copula was chosen for the first and second follow-ups by the proposed model. The shape of the curve is governed by the limits of the conditional distribution (Table 6.3). While the Frank copula has a finite limit for arbitrarily small BMI values, the limit for the Clayton and Gaussian copula is 1. Thus, the estimated probabilities continue to increase.

Prevalence of arthritis given BMI and other conditions Figure 6.2 leverages our joint dependence model to show the complex dependence of the probability of observing arthritis with BMI and other chronic conditions. The upper panel of Figure 6.2 shows the probability of arthritis given BMI with the presence/absence of diabetes and the presence/absence of hypertension, thus producing four different lines. This enables us to see the complete picture of arthritis prevalence at the baseline and the follow-up periods. The top panel of Figure 6.2 indicates positive dependence between arthritis and the other two diseases (diabetes and hypertension). When a subject had both diabetes and hypertension, the probability of having arthritis was higher compared to a subject who suffers from only one or no chronic condition. Likewise, the probability of having arthritis was higher with the obese people (BMI $\geq$ 30). In

Figure 6.2: Upper panel: conditional probability of observing arthritis given BMI and other chronic conditions: solid (diabetes, hypertension), dashed (diabetes, no hypertension), dotted (no diabetes, hypertension), dash-dotted (no diabetes, no hypertersion).
Lower panel: probability of observing diabetes given BMI and other chronic conditions: solid (heart disease, stroke), dashed (heart disease, no stroke), dotted (no heart disease, stroke), dash-dotted (no heart disease, no stroke).



addition, Figure 6.2 provides useful information when multiple chronic conditions are present and their causal relationship is unclear. In the follow-up, some changes in terms of the impact of hypertension and diabetes on the risk of arthritis were observed; in contrast to the baseline observation, a person with hypertension faced a higher chance of developing arthritis than one with diabetes provided that other conditions remained the same.

Prevalence of diabetes given BMI and other conditions The lower panel of Figure 6.2 presents the conditional probability of observing diabetes given BMI and two other chronic conditions (heart disease and stroke). The probability of diabetes is not affected strongly by the presence/absence of heart diseases and stroke when the BMI is low, however, as the BMI level increases the probability of observing diabetes is getting larger depending on the presence

of the cardiovascular diseases (CVD). Compared to the case when elderly have either heart disease or stroke, the risk of diabetes jumped by more than 15% for obese patients with both heart disease and stroke, indicating that diabetes is associated with CVD. We want to note that caution is needed when interpreting probability plots since the displayed associations do not imply causations. For instance, though we plotted the probability of diabetes given presence/absence of CVD, diabetes is usually considered as the risk factor of CVD in medical literatures. Our plots only serve as a reference to illustrate the multivariate association among the diseases. To illustrate the merits of our method, we have presented results for several variable groups in this chapter. Similar plots can be produced for all variable groups which are of medical interest.

6.4 Conclusions

The aim of the study presented in this section was to help understanding comorbidity among the elderly and give new clues about its pathways. We have fitted an R-vine PCC for mixed discrete and continuous variables and demonstrated a model selection heuristic as well as parameter inference in the maximum likelihood framework. Since PCCs allow to combine different copula families, different limiting behavior of conditional probabilities for the presence of diseases given the BMI could also be modeled. This improves the predictive performance of the copula model compared to models where all bivariate families are the same as cross-validation shows.

As a final note, we want to point out some limitations of our study, both considering the available data and the statistical model. Currently unavailable information such as the time elapsed since the diagnosis of chronic conditions may help to explain the pathways of comorbidity more accurately in future research. Also, the response variables in our analysis were collected in a self-report survey. Although the study of Kriegsman et al. (1996) implicates that patients' self-reports on chronic diseases are fairly accurate, the use of self-reported diagnoses might have introduced systematic bias. In particular, Kriegsman et al. (1996) find that self-reports on arthritis were often incorrect; using general practitioners information or clinical interviews might be more reliable ways to obtain data. Also, the data set considered in our analysis only included people who survived throughout all three waves. Dropouts due to death would not be missing at random, which is another potential source of systematic bias. Also, the longitudinal pattern in the LSOA II data was not modeled explicitly here. While including informative dropouts and a systematic joint longitudinal modeling approach could help to better understand the pathways and development of chronic conditions, this is

beyond the scope of the current work. Finally, while the inference procedures demonstrated here allow to estimate standard errors for parameter estimates, the model uncertainty cannot be addressed. This could be done in a computationally more intense RJMCMC framework (cf. Gruber et al. (2012)).

Appendix A

Calculation of the observed information in regular vine copula models

To derive an algorithm similar to Algorithm 2.2.3 which recursively determines all terms of the second derivatives of the log-likelihood of an R-vine copula model with respect to parameters θ, γ, we need to distinguish 7 basic cases of dependence on the two parameters which can occur for a term $c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}), F_{V|\mathbf{Z}}(v|\mathbf{z})\right)$.

Table A.1: 7 cases of how a term $c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}), F_{V|\mathbf{Z}}(v|\mathbf{z})\right)$ can depend on θ and γ.

cases	dependence on	θ	γ	Example: $c_{7,3;1,4,5,6}$	
case 1		$F_{U\|\mathbf{Z}}$	$F_{V\|\mathbf{Z}}$	$\theta_{7,4\|1,5,6}$	$\theta_{6,3\|1,4,5}$
case 2		$F_{U\|\mathbf{Z}}, F_{V\|\mathbf{Z}}$	$F_{V\|\mathbf{Z}}$	$\theta_{4,1}$	$\theta_{6,3\|1,4,5}$
case 3		$F_{U\|\mathbf{Z}}, F_{V\|\mathbf{Z}}$	$F_{U\|\mathbf{Z}}$, $F_{V\|\mathbf{Z}}$	$\theta_{4,1}$	$\theta_{5,1}$
case 4	through	$F_{U\|\mathbf{Z}}$	$c_{U,V;\mathbf{Z}}$	$\theta_{7,4\|1,5,6}$	$\theta_{7,3\|1,4,5,6}$
case 5		$F_{U\|\mathbf{Z}}, F_{V\|\mathbf{Z}}$	$c_{U,V;\mathbf{Z}}$	$\theta_{4,1}$	$\theta_{3,7\|1,4,5,6}$
case 6		$F_{U\|\mathbf{Z}}$	$F_{U\|\mathbf{Z}}$	$\theta_{7,4\|1,5,6}$	$\theta_{7,1\|5,6}$
case 7		$c_{U,V;\mathbf{Z}}$	$c_{U,V;\mathbf{Z}}$	$\theta_{7,3\|1,4,5,6}$	$\theta_{7,3\|1,4,5,6}$

Here, case 7 is relevant only for derivatives where $\theta = \gamma$, since we assume that all bivariate copulas occurring in the vine density have one parameter in $\mathbb{R}$. Because of symmetry in the parameters, all other possible combinations are already included in these cases, we only have to exchange θ and γ. As an example let us consider the term corresponding to $c_{7,3|1,4,5,6}$ in our 8-dimensional example. Choices for θ and γ for which the 7 cases arise are listed in the rightmost columns of Table A.1.

In case 1 we determine

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma} \ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta), F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right)\right) \\
&= \frac{\partial_1\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta), F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta), F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right)} \cdot \left(\frac{\partial}{\partial\theta} F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\right) \cdot \left(\frac{\partial}{\partial\gamma} F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right) \\
&- \left(\frac{\partial}{\partial\theta} \ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta), F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right)\right)\right) \\
&\cdot \left(\frac{\partial}{\partial\gamma} \ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta), F_{V|\mathbf{Z}}(v|\mathbf{z},\gamma)\right)\right)\right),
\end{aligned} \tag{A.1}$$

for case 2

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\right)\\
&=\frac{\partial_1\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\right)\cdot\left(\frac{\partial}{\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&-\left(\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\right)\right)\\
&\cdot\left(\frac{\partial}{\partial\gamma}\ln(\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\right)\right)\\
&+\frac{\partial_2\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\cdot\left(\frac{\partial}{\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial^2}{\partial\theta\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right),
\end{aligned}
\tag{A.2}
$$

and case 3 yields

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln(c_{U,V;\mathbf{Z}}(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)))\\
&=\frac{\partial_1\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\left(\frac{\partial}{\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&-\left(\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\right)\right)\\
&\cdot\left(\frac{\partial}{\partial\gamma}\ln(\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\right)\right)\\
&+\frac{\partial_1\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\left(\frac{\partial}{\partial\gamma}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_1\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\left(\frac{\partial}{\partial\gamma}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_1\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\left(\frac{\partial}{\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_2\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\left(\frac{\partial}{\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial^2}{\partial\theta\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)\\
&+\frac{\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right)}\cdot\left(\frac{\partial^2}{\partial\theta\partial\gamma}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta,\gamma)\right).
\end{aligned}
\tag{A.3}
$$

Similarly, we have for case 4 that

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)\right)\\
&=\left(\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)\right)\right)\cdot\frac{-\partial_\gamma c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)}\\
&+\frac{\partial_\gamma\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z})|\gamma\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\right),
\end{aligned}
\tag{A.4}
$$

and

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)\right)\\
&=\left(\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)\right)\right)\\
&\cdot\frac{-\partial_\gamma c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}\\
&+\frac{\partial_\gamma\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta)\right)\\
&+\frac{\partial_\gamma\partial_2 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta),F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)|\gamma\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{V|\mathbf{Z}}(v|\mathbf{z},\theta)\right),
\end{aligned}\tag{A.5}$$

for the fifth case. Finally,

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\right)\\
&=\frac{\partial_1\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}\cdot\left(\frac{\partial}{\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\cdot\left(\frac{\partial}{\partial\gamma}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right)\\
&-\left(\frac{\partial}{\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\right)\right)\\
&\cdot\left(\frac{\partial}{\partial\theta}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)\right)\right)\\
&+\frac{\partial_1 c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma),F_{V|\mathbf{Z}}(v|\mathbf{z})\right)}\cdot\left(\frac{\partial^2}{\partial\gamma\partial\theta}F_{U|\mathbf{Z}}(u|\mathbf{z},\theta,\gamma)\right),\\
&\frac{\partial^2}{\partial\theta\partial\gamma}\ln\left(c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta,\gamma\right)\right)\\
&=\frac{\partial_\theta\partial_\gamma c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta,\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta,\gamma\right)}\\
&-\frac{\partial_\theta c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta,\gamma\right)\cdot\partial_\gamma c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v)|\theta,\gamma\right)}{c_{U,V;\mathbf{Z}}\left(F_{U|\mathbf{Z}}(u|\mathbf{z}),F_{V|\mathbf{Z}}(v|\mathbf{z})|\theta,\gamma\right)^2}.
\end{aligned}\tag{A.6}$$

Algorithm A.0.1 Second derivative with respect to the parameters $\theta^{\tilde{k},\tilde{i}}$ and $\theta^{\hat{k},\hat{i}}$.

The input of the algorithm is a d-dimensional R-vine matrix M with maximum matrix $\tilde{M}$ and parameter matrix $\boldsymbol{\theta}$, and matrices $C^{\tilde{k},\tilde{i}}$, $C^{\hat{k},\hat{i}}$ determined using Algorithm 2.2.2 for parameters $\theta^{\tilde{k},\tilde{i}}$ and $\theta^{\hat{k},\hat{i}}$ of the R-vine parameter matrix. Further, we assume the matrices V^{direct}, $V^{indirect}$ and V^{values}, the matrices $S1^{direct,\tilde{k},\tilde{i}}$, $S1^{indirect,\tilde{k},\tilde{i}}$ and $S1^{values,\tilde{k},\tilde{i}}$ and $S1^{direct,\hat{k},\hat{i}}$, $S1^{indirect,\hat{k},\hat{i}}$ and $S1^{values,\hat{k},\hat{i}}$ to be given. The output will be the value of the second derivative of the copula log-likelihood for the given observation with respect to parameters $\theta^{\tilde{k},\tilde{i}}$ and $\theta^{\hat{k},\hat{i}}$. Without loss of generality, we assume that $\hat{i}\geq\tilde{i}$, and $\hat{k}\geq\tilde{k}$ if $\hat{i}=\tilde{i}$.

1: **if** $c^{\hat{k},\hat{i}}_{\tilde{k},\tilde{i}} == 1$ **then**
2: Set $m=\tilde{m}_{\tilde{k},\tilde{i}}$
3: Set $z_1=v^{direct}_{\tilde{k},\tilde{i}}$, $\tilde{z}_1=s1^{direct,\hat{k},\hat{i}}_{\tilde{k},\tilde{i}}$
4: **if** $m == m_{\tilde{k},\tilde{i}}$ **then**
5: Set $z_2=v^{direct}_{k,d-m+1}$, $\tilde{z}_2=s1^{direct,\hat{k},\hat{i}}_{\tilde{k},d-m+1}$

else

Set $z_2 = v^{indirect}_{\tilde{k},d-m+1}$, $\tilde{z}_2 = s1^{indirect,\hat{k},\hat{i}}_{\tilde{k},d-m+1}$

end if

Set $s2^{direct}_{\tilde{k}-1,\tilde{i}} = 0, s2^{indirect}_{\tilde{k}-1,\tilde{i}} = 0, s2^{values}_{\tilde{k},\tilde{i}} = 0$

if $\tilde{k} == \hat{k}$ & $\tilde{i} == \hat{i}$ **then**

Set $s2^{direct}_{\tilde{k}-1,\tilde{i}} = \partial_{\theta^{\tilde{k},\tilde{i}}}\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_1, z_2|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}})$

Set $s2^{indirect}_{\tilde{k}-1,\tilde{i}} = \partial_{\theta^{\tilde{k},\tilde{i}}}\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_2, z_1|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}})$

Set $s2^{values}_{\tilde{k},\tilde{i}} = \frac{\partial_{\theta^{\tilde{k},\tilde{i}}}\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})} - (s1^{values,\hat{k},\hat{i}}_{\tilde{k},\tilde{i}})^2$

end if

if $c^{\hat{k},\hat{i}}_{\tilde{k}+1,\tilde{i}} == 1$ **then**

Set $s2^{values}_{\tilde{k},\tilde{i}} = s1^{values,\hat{k},\hat{i}}_{\tilde{k},\tilde{i}} \cdot \frac{-\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})} + \frac{\partial_1\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})} \cdot \tilde{z_1}$

Set $s2^{direct}_{\tilde{k}-1,\tilde{i}} = \partial_1\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_1, z_2|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}}) \cdot \tilde{z_1}$

Set $s2^{indirect}_{\tilde{k}-1,\tilde{i}} = \partial_2\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_2, z_1|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}}) \cdot \tilde{z_1}$

end if

if $c^{\hat{k},\hat{i}}_{\tilde{k}+1,d-m+1} == 1$ **then**

Set $s2^{values}_{\tilde{k},\tilde{i}} = s2^{values}_{\tilde{k},\tilde{i}} + \frac{\partial_2\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})} \cdot \tilde{z_2}$

if $c^{\hat{k},\hat{i}}_{k+1,i} == 0$ **then**

Set $s2^{values}_{\tilde{k},\tilde{i}} = s2^{values}_{\tilde{k},\tilde{i}} + s1^{values,\hat{k},\hat{i}}_{\tilde{k},\tilde{i}} \cdot \frac{-\partial_{\theta^{\tilde{k},\tilde{i}}} c(z_1,z_2|\mathcal{B}^{\tilde{k},\tilde{i}},\theta^{\tilde{k},\tilde{i}})}{exp(v^{values}_{\tilde{k},\tilde{i}})}$

end if

Set $s2^{direct}_{\tilde{k}-1,\tilde{i}} = s2^{direct}_{\tilde{k}-1,\tilde{i}} + \partial_2\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_1, z_2|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}}) \cdot \tilde{z_2}$

Set $s2^{indirect}_{\tilde{k}-1,\tilde{i}} = s2^{indirect}_{\tilde{k}-1,\tilde{i}} + \partial_1\partial_{\theta^{\tilde{k},\tilde{i}}} h(z_2, z_1|\mathcal{B}^{\tilde{k},\tilde{i}}, \theta^{\tilde{k},\tilde{i}}) \cdot \tilde{z_2}$

end if

end if

for $i = \tilde{i}, \ldots, 1$ **do**

for $k = \tilde{k} - 1, \ldots, i + 1$ **do**

Set $m = \tilde{m}_{k,i}$

Set $z_1 = v^{direct}_{k,i}$, $\tilde{z}^{\hat{k},\hat{i}}_1 = s1^{direct,\hat{k},\hat{i}}_{k,i}$, $\tilde{z}^{\tilde{k},\tilde{i}}_1 = s1^{direct,\tilde{k},\tilde{i}}_{k,i}$, $\bar{z}_1 = s2^{direct}_{k,i}$

if $m == m_{k,i}$ **then**

Set $z_2 = v^{direct}_{k,d-m+1}$, $\tilde{z}^{\hat{k},\hat{i}}_2 = s1^{direct,\hat{k},\hat{i}}_{k,d-m+1}$, $\tilde{z}^{\tilde{k},\tilde{i}}_2 = s1^{direct,\tilde{k},\tilde{i}}_{k,d-m+1}$, $\bar{z}_2 = s2^{direct}_{k,d-m+1}$

else

Set $z_2 = v^{indirect}_{k,d-m+1}$, $\tilde{z}^{\hat{k},\hat{i}}_2 = s1^{indirect,\hat{k},\hat{i}}_{k,d-m+1}$, $\tilde{z}^{\tilde{k},\tilde{i}}_2 = s1^{indirect,\tilde{k},\tilde{i}}_{k,d-m+1}$, $\bar{z}_2 = s2^{indirect}_{k,d-m+1}$

end if

Set $s2^{values}_{k,i} = -s1^{values,\hat{k},\hat{i}}_{k,i} \cdot s1^{values,\tilde{k},\tilde{i}}_{k,i}, s2^{direct}_{k-1,i} = 0, s2^{indirect}_{k-1,i} = 0$

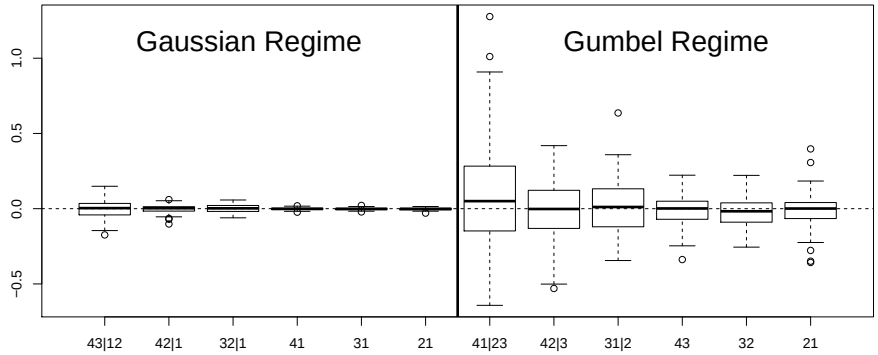
if $c^{\hat{k},\hat{i}}_{k+1,i} == 1 \;\;\&\;\; c^{\tilde{k},\tilde{i}}_{k+1,i} == 1$ **then**
  Set $s2^{values}_{k,i} = s2^{values}_{k,i} + \frac{\partial_1\partial_1 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1 + \frac{\partial_1 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \bar{z}_1$
  Set $s2^{direct}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_1 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \bar{z}_1 + \partial_1\partial_1 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1$
  Set $s2^{indirect}_{k-1,i} = s2^{indirect}_{k-1,i} + \partial_2 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \bar{z}_1 + \partial_2\partial_2 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1$
end if
if $c^{\hat{k},\hat{i}}_{k+1,d-m+1} == 1 \;\;\&\;\; c^{\tilde{k},\tilde{i}}_{k+1,d-m+1} == 1$ **then**
  Set $s2^{values}_{k,i} = s2^{values}_{k,i} + \frac{\partial_2\partial_2 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2 + \frac{\partial_2 c(z_1,z_2|\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \bar{z}_2$
  Set $s2^{direct}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_2 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \bar{z}_2 + \partial_2\partial_2 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2$
  Set $s2^{indirect}_{k-1,i} = s2^{indirect}_{k-1,i} + \partial_1 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \bar{z}_2 + \partial_1\partial_1 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2$
end if
if $c^{\hat{k},\hat{i}}_{k+1,i} == 1 \;\;\&\;\; c^{\tilde{k},\tilde{i}}_{k+1,d-m+1} == 1$ **then**
  Set $s2^{values}_{k,i} = s2^{values}_{k,i} + \frac{\partial_1\partial_2 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2$
  Set $s2^{direct}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_1\partial_2 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2$
  Set $s2^{indirect}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_1\partial_2 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_1 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_2$
end if
if $c^{\hat{k},\hat{i}}_{k+1,d-m+1} == 1 \;\;\&\;\; c^{\tilde{k},\tilde{i}}_{k+1,i} == 1$ **then**
  Set $s2^{values}_{k,i} = s2^{values}_{k,i} + \frac{\partial_2\partial_1 c(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i})}{exp(v^{values}_{k,i})} \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1$
  Set $s2^{direct}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_1\partial_2 h(z_1,z_2|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1$
  Set $s2^{indirect}_{k-1,i} = s2^{direct}_{k-1,i} + \partial_1\partial_2 h(z_2,z_1|\mathcal{B}^{k,i},\theta^{k,i}) \cdot \tilde{z}^{\hat{k},\hat{i}}_2 \cdot \tilde{z}^{\tilde{k},\tilde{i}}_1$
end if
end for
end for
return $\sum_{k,i=1,\ldots,d} s2^{values}_{k,i}$

Appendix B

Additional results from the simulation study in Section 4.1.3

The following tables show the relative bias and relative MSE for the parameters of two selected scenarios in the simulation study (Scenario 2 & Scenario 4). For comparison purposes all copula parameters have been transformed to the Kendall's τ level. Notice the large bias for the posterior mean estimate of the second Markov chain parameter in Scenario 4 (Table B.2) where identification issues where observed. In this case the Gibbs sampler with objective priors fails to capture the underlying Markov structure correctly and the Bayesian procedure needs to be started with strong subjective prior beliefs. In general, we observe that the estimation error in the second and third tree is higher than on the first tree and that the uncertainty in the Gumbel regime, from which less realizations are included in the data set, is higher than in the Gaussian regime.

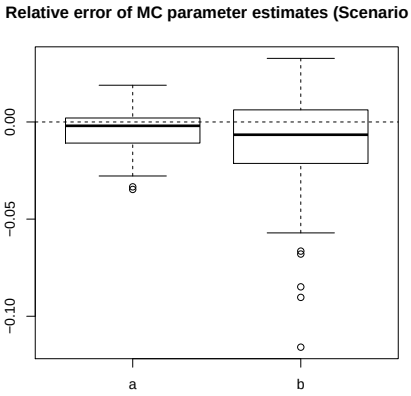
Table B.1: Scenario 2: Relative error of Kendall's τ estimates (top figure) and relative bias / MSE for the Gaussian regime (upper table) and the Gumbel regime (lower table), respectively.

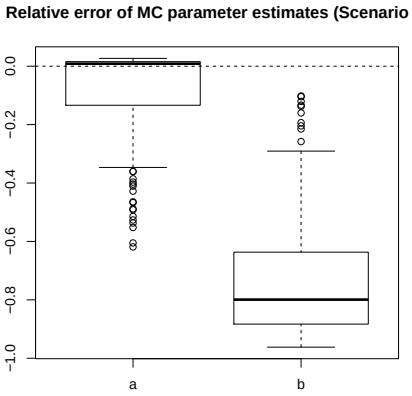


Gaussian regime	$\tau_{43\|12}$	$\tau_{42\|1}$	$\tau_{32\|1}$	τ_{41}	τ_{31}	τ_{21}
relative bias	-2.1 $\cdot 10^{-3}$	-1.7 $\cdot 10^{-3}$	7.2 $\cdot 10^{-4}$	-8.1 $\cdot 10^{-4}$	-1.1 $\cdot 10^{-3}$	-1.9 $\cdot 10^{-3}$
relative MSE	1.4 $\cdot 10^{-3}$	4.3 $\cdot 10^{-4}$	4.3 $\cdot 10^{-4}$	4.6 $\cdot 10^{-5}$	4.1 $\cdot 10^{-5}$	5.5 $\cdot 10^{-5}$

Gumbel regime	$\tau_{41\|23}$	$\tau_{42\|3}$	$\tau_{31\|2}$	τ_{43}	τ_{32}	τ_{21}
relative bias	9.5 $\cdot 10^{-2}$	5.2 $\cdot 10^{-3}$	1.7 $\cdot 10^{-2}$	-1.1 $\cdot 10^{-2}$	-1.6 $\cdot 10^{-2}$	-1.0 $\cdot 10^{-2}$
relative MSE	1.3 $\cdot 10^{-2}$	7.2 $\cdot 10^{-3}$	5.6 $\cdot 10^{-3}$	3.3 $\cdot 10^{-3}$	3.2 $\cdot 10^{-3}$	3.6 $\cdot 10^{-3}$

Table B.2: Relative error of Markov chain parameter estimates in Scenarios 2 (left figure) and 4 (right figure), and relative bias / MSE for Scenario 2 (left table) and Scenario 4 (right table).

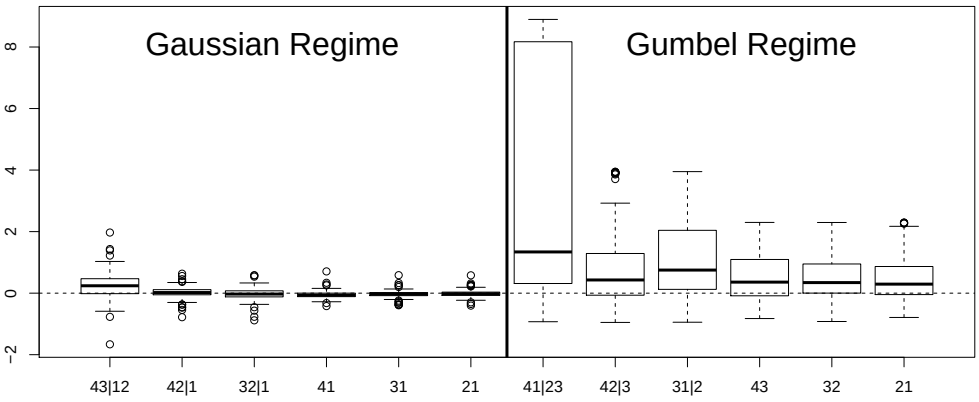




MC parameters	a	b
relative bias	-3.7 $\cdot 10^{-3}$	-1.0 $\cdot 10^{-2}$
relative MSE	1.4 $\cdot 10^{-4}$	6.5 $\cdot 10^{-4}$

MC parameters	a	b
relative bias	-9.7 $\cdot 10^{-2}$	-7.2 $\cdot 10^{-1}$
relative MSE	4.0 $\cdot 10^{-2}$	5.1 $\cdot 10^{-1}$

Table B.3: Scenario 4: Relative error of Kendall's τ estimates (top figure) and relative bias / MSE for the Gaussian regime (upper table) and the Gumbel regime (lower table), respectively.



Gaussian regime	$\tau_{43\|12}$	$\tau_{42\|1}$	$\tau_{32\|1}$	τ_{41}	τ_{31}	τ_{21}
relative bias	2.4 $\cdot 10^{-1}$	1.9 $\cdot 10^{-2}$	-2.6 $\cdot 10^{-2}$	-5.1 $\cdot 10^{-2}$	-3.7 $\cdot 10^{-2}$	-1.5 $\cdot 10^{-2}$
relative MSE	2.5 $\cdot 10^{-2}$	7.5 $\cdot 10^{-3}$	8.3 $\cdot 10^{-3}$	5.7 $\cdot 10^{-3}$	5.4 $\cdot 10^{-3}$	4.4 $\cdot 10^{-3}$

Gumbel regime	$\tau_{41\|23}$	$\tau_{42\|3}$	$\tau_{31\|2}$	τ_{43}	τ_{32}	τ_{21}
relative bias	3.1 $\cdot 10^{0}$	8.7 $\cdot 10^{-1}$	1.2 $\cdot 10^{0}$	5.2 $\cdot 10^{1}$	5.1 $\cdot 10^{-1}$	4.5 $\cdot 10^{-1}$
relative MSE	2.3 $\cdot 10^{0}$	5.3 $\cdot 10^{-1}$	7.5 $\cdot 10^{-1}$	2.7 $\cdot 10^{-1}$	2.3 $\cdot 10^{-1}$	2.1 $\cdot 10^{-1}$

Appendix C

Conditional copula of the Student's t distribution

Let us denote a d-dimensional Student's t distribution with mean vector $\mathbf{0}$, correlation matrix R and degrees of freedom ν is denoted as $t_d(\mathbf{0}, R, \nu)$. Its pdf is $f_{t,d}(\cdot; R, \nu)$ and we write $F_{t,d}(\cdot; R, \nu)$ for the cdf. We consider a d-dimensional random vector $\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_B) = (X_1, X_2, \mathbf{X}_B)$, with $A = \{1, 2\}$ and B=$\{3, \ldots, d\}$, distributed according to a multivariate Student's t distribution with ν degrees of freedom, mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$ and scale matrix

$$R = (R_{i,j})_{i,j=1,\ldots,d} = \begin{pmatrix} R_A & R_{AB} \\ R_{AB}^T & R_B \end{pmatrix}, \text{where } R_{AB} = \begin{pmatrix} R_{1B} \\ R_{2B} \end{pmatrix}, R_A = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}.$$

Let us define

$$V_{A|B} := \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} - \begin{pmatrix} R_{1B} \\ R_{2B} \end{pmatrix} R_B^{-1} \begin{pmatrix} R_{1B}^T & R_{2B}^T \end{pmatrix} =: \begin{pmatrix} V_{1|B} & V_{12|B} \\ V_{21|B} & V_{2|B} \end{pmatrix},$$

$$R_{A|B} := \operatorname{diag}(V_{A|B})^{-1/2} \, V_{A|B} \, \operatorname{diag}(V_{A|B})^{-1/2}, \gamma(\mathbf{x}_B) := \sqrt{\frac{1 + (1/\nu)\mathbf{x}_B^T R_B^{-1} \mathbf{x}_B}{(\nu + d - 2)/\nu}},$$

then we have for the conditional distribution of $\mathbf{X}_A$ given $\mathbf{X}_B = \mathbf{x}_B$:

$$F_{A|B}(\mathbf{x}_A|\mathbf{x}_B) = F_{t,2}\left(\frac{x_1 - \mu_{1|B}(\mathbf{x}_B)}{\sqrt{V_{1|B}} \cdot \gamma(\mathbf{x}_B)}, \frac{x_2 - \mu_{2|B}(\mathbf{x}_B)}{\sqrt{V_{2|B}} \cdot \gamma(\mathbf{x}_B)}; R_{A|B}, \nu + d - 2\right), \tag{C.1}$$

cf. Nikoloulopoulos et al. (2009, Lemma 2.2). Taking $x_2 \to \infty$ ($x_1 \to \infty$) yields

$$\begin{aligned} F_{1|B}(x_1|\mathbf{x}_B) &= F_{t,1}\left(\frac{x_1 - \mu_{1|B}(\mathbf{x}_B)}{\sqrt{V_{1|B}} \cdot \gamma(\mathbf{x}_B)}; \nu + d - 2\right) \\ F_{2|B}(x_2|\mathbf{x}_B) &= F_{t,1}\left(\frac{x_2 - \mu_{2|B}(\mathbf{x}_B)}{\sqrt{V_{2|B}} \cdot \gamma(\mathbf{x}_B)}; \nu + d - 2\right), \end{aligned} \tag{C.2}$$

so that we can now determine the corresponding copula:

$$\begin{aligned} C_{1,2;3:d}(u_1, u_2) &= C_{1,2;3:d}(u_1, u_2|\mathbf{x}_B) = F_{12|3:d}\left(F_{1|3:d}^{-1}(u_1|\mathbf{x}_B), F_{2|3:d}^{-1}(u_2|\mathbf{x}_B)|\mathbf{x}_B\right) \\ &= F_{t,2}\left({F_{t,1}}^{-1}(u_1; \nu + d - 2), {F_{t,1}}^{-1}(u_2; \nu + d - 2); R_{A|B}, \nu + d - 2\right), \end{aligned}$$

where the additive constants and scaling factors in Equations (C.1) and (C.2) cancel. This is a bivariate Student's t copula with $\nu + d - 2$ degrees of freedom and correlation matrix $R_{A|B}$ and does not depend on $\mathbf{x}_B$ anymore.

Appendix D

Application 1: additional material

D.1 Exchange rate data: selected R-vines

This appendix lists the parameter estimates and copula structures corresponding to the MS models considered in Section 5. For models (1) - (3) which are analyzed in detail in the application section, we provide the selected R-vine tree structures, as well as parameter estimates resulting from the EM algorithm and the Bayesian estimation procedure. To make comparisons easier across different copula families, all parameters have been transformed to the corresponding values of Kendall's τ. As summary statistics for the posterior distributions in the Bayesian setup we list the posterior mean estimates as well as the 5% and 95% quantiles.

D.1.1 Model (1)

Here, only the copula parameters are switching while tree structure (Figure D.1) and copula families are common to both regimes.

Figure D.1: First and second tree of the tree structure $\mathcal{V}_1$ of Model (1). This structure represents also the non-crisis regime in Models (2a) - (2c) and (3).

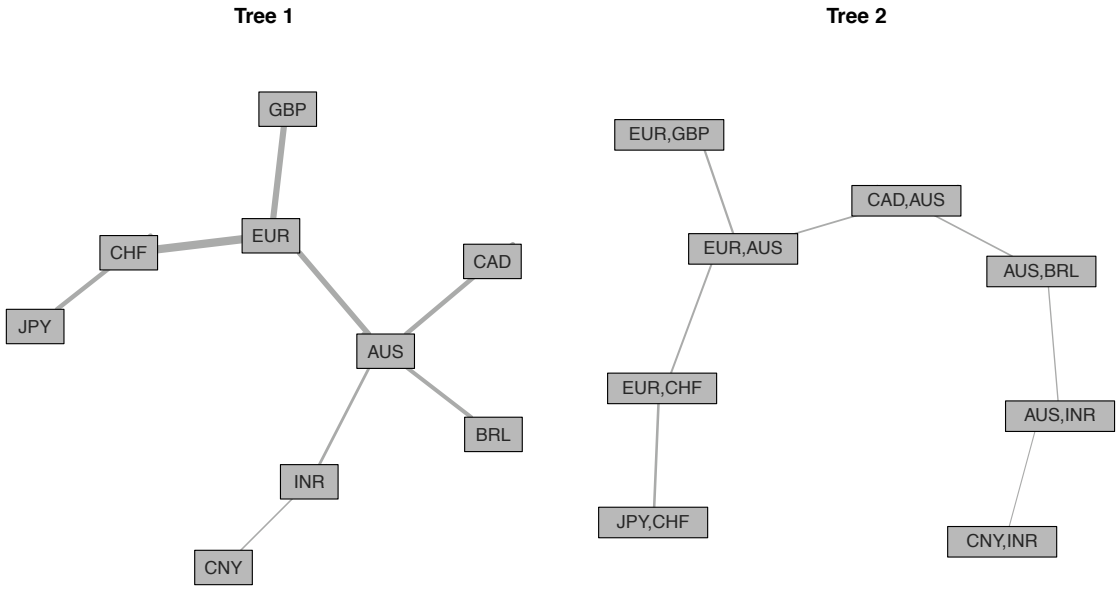


Table D.1: Estimated Kendall's τ for the first and second tree of Model (1).

Tree 1	GBP,EUR	EUR,CHF	CHF,JPY	AUS,EUR	AUS,BRL	INR,AUS	CAD,AUS	CNY,INR
cop. fam.	SG	SG	SG	N	G	N	N	G
				Regime 1				
$\hat{\boldsymbol{\tau}}_1^{EM}$	0.55	0.78	0.46	0.46	0.19	0.14	0.29	0.11
$\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.56	0.79	0.47	0.46	0.18	0.14	0.28	0.13
5% quant.	0.53	0.77	0.43	0.43	0.14	0.10	0.24	0.09
95% quant.	0.60	0.81	0.50	0.49	0.23	0.20	0.32	0.17
				Regime 2				
$\hat{\boldsymbol{\tau}}_2^{EM}$	0.44	0.58	0.24	0.41	0.45	0.26	0.44	0.07
$\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.44	0.58	0.22	0.40	0.43	0.25	0.44	0.05
5% quant.	0.40	0.55	0.17	0.36	0.39	0.20	0.40	0.02
95% quant.	0.47	0.60	0.27	0.43	0.47	0.29	0.50	0.10

Tree 2	GBP,AUS| EUR	CAD,EUR| AUS	CAD,BRL| AUS	BRL,INR| AUS	CNY,AUS| INR	JPY,EUR| CHF	CHF,AUS| EUR
cop. fam.	G	G	SG	0	SG	G270	G270
			Regime 1				
$\hat{\boldsymbol{\tau}}_1^{EM}$	0.15	0.11	0.07	0.02	0.01	-0.06	-0.03
$\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.15	0.11	0.07	0.01	0.02	-0.06	-0.03
5% quant.	0.10	0.07	0.03	-0.04	0.00	-0.10	-0.07
95% quant.	0.20	0.15	0.13	0.06	0.05	-0.02	-0.00
			Regime 2				
$\hat{\boldsymbol{\tau}}_2^{EM}$	0.15	0.11	0.11	0.11	0.10	-0.31	-0.24
$\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.16	0.13	0.11	0.11	0.11	-0.31	-0.24
5% quant.	0.10	0.08	0.04	0.06	0.07	-0.36	-0.28
95% quant.	0.21	0.17	0.17	0.16	0.16	-0.26	-0.19

D.1.2 Models (2)

For Models (2a) - (2c), two R-vine tree structures have been selected. The R-vine $\mathcal{V}_1$, corresponding to normal times, has again the structure displayed in Figure D.1, the second R-vine ($\mathcal{V}_2$), is given in Figure D.2.

Table D.2: Estimated Kendall's τ values corresponding to the first tree of Models (2a) - (2c), respectively. For the t-copula used in Model (2c), the first parameter is Kendall's τ, the second gives the degrees of freedom (with $\nu_{max} = 30$).

"normal", $\mathcal{V}_1$	GBP,EUR	EUR,CHF	CHF,JPY	AUS,EUR	AUS,BRL	INR,AUS	CAD,AUS	CNY,INR
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.53	0.75	0.45	0.46	0.28	0.21	0.35	0.12
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.56	0.78	0.45	0.48	0.24	0.21	0.33	0.15
5% quant.	0.52	0.75	0.41	0.45	0.19	0.17	0.29	0.10
95% quant.	0.60	0.80	0.49	0.51	0.29	0.25	0.36	0.20
(2b) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.54	0.75	0.44	0.47	0.29	0.22	0.35	0.12
(2b) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.52	0.74	0.43	0.44	0.29	0.21	0.34	0.11
5% quant.	0.49	0.72	0.40	0.41	0.26	0.17	0.31	0.07
95% quant.	0.55	0.76	0.46	0.48	0.32	0.24	0.37	0.14
(2c) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.60	0.81	0.47	0.48	0.21	0.19	0.32	0.16
(2c) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.60	0.80	0.47	0.49	0.21	0.21	0.31	0.17
5% quant.	0.57	0.79	0.44	0.46	0.16	0.17	0.27	0.13
95% quant.	0.62	0.82	0.51	0.52	0.26	0.22	0.35	0.22

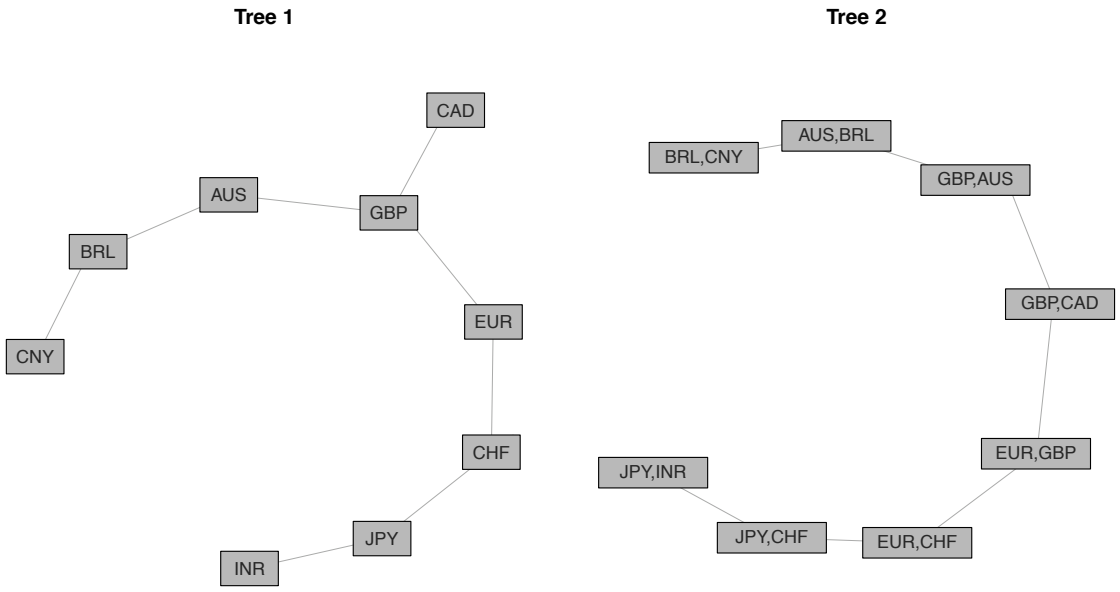
"crisis", $\mathcal{V}_2$	GBP,EUR	EUR,CHF	CHF,JPY	JPY,INR	AUS,GBP	BRL,AUS	BRL,CNY	CAD,GBP
(2a) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.44	0.45	0.11	0.00	0.41	0.49	0.11	0.41
(2a) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.42	0.52	0.22	0.01	0.37	0.47	0.07	0.37
5% quant.	0.36	0.47	0.11	0.00	0.30	0.41	0.01	0.29
95% quant.	0.49	0.56	0.30	0.02	0.44	0.53	0.13	0.43
(2b) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.37	0.37	0.10	0.00	0.32	0.41	0.08	0.36
(2b) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.45	0.37	0.05	0.01	0.35	0.44	0.13	0.38
5% quant.	0.34	0.27	0.00	0.00	0.25	0.36	0.03	0.30
95% quant.	0.55	0.45	0.15	0.04	0.44	0.53	0.24	0.46
(2c) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.43, 10.8	0.58, 8.6	0.27, 7.9	-0.13, 30	0.37, 10.7	0.44, 5.9	0.06, 30	0.33, 30
(2c) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.42, 14.2	0.56, 9.7	0.25, 9.8	-0.15, 21.4	0.35, 15.4	0.45, 9.3	0.05, 20.8	0.34, 21.8
5% quant.	0.38, 7.0	0.53, 5.6	0.21, 5.2	-0.21, 11.6	0.31, 7.0	0.45, 4.8	-0.01, 10.7	0.29, 11.0
95 % quant.	0.46, 25.7	0.60, 16.3	0.29, 18.3	-0.11, 29.2	0.40, 28.1	0.49, 19.0	0.11, 29.2	0.39, 29.3

Table D.3: Estimated Kendall's τ, corresponding to the second tree of (2a) - (2c).

"normal", $\mathcal{V}_1$	JPY,EUR| CHF	AUS,CHF| EUR	AUS,GBP| EUR	CAD,EUR| AUS	CAD,BRL| AUS	INR,BRL| AUS	CNY,AUS| INR
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	-0.14	-0.13	0.14	0.10	0.12	0.07	0.01
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	-0.10	-0.08	0.13	0.10	0.11	0.06	0.00
5 % quantile	-0.16	-0.14	0.09	0.06	0.07	0.01	-0.05
95 % quantile	-0.02	0.00	0.18	0.14	0.15	0.10	0.04
(2b) $\hat{\boldsymbol{\tau}}_1^{EM}$	-0.16	-0.14	0.15	0.10	0.12	0.07	0.01
(2b) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	-0.17	-0.17	0.16	0.10	0.11	0.07	0.02
5 % quantile	-0.21	-0.20	0.13	0.06	0.07	0.03	-0.02
95 % quantile	-0.13	-0.13	0.20	0.13	0.14	0.11	0.06
(2c) $\hat{\boldsymbol{\tau}}_1^{EM}$	-0.01	0.02	0.15	0.10	0.12	0.03	0.00
(2c) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	-0.03	0.00	0.14	0.10	0.11	0.04	0.00
5 % quantile	-0.08	-0.06	0.09	0.05	0.07	-0.01	-0.05
95 % quantile	0.03	0.05	0.18	0.15	0.16	0.09	0.04

"crisis", $\mathcal{V}_2$	CNY,AUS| BRL	GBP,BRL| AUS	AUS,CAD| GBP	CAD,EUR| GBP	GBP,CHF| EUR	EUR,JPY| CHF	CHF,INR| JPY
(2a) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.21	0.04	0.24	0.14	-0.22	-0.42	0.03
(2a) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.17	0.04	0.28	0.15	-0.19	-0.36	0.07
5 % quantile	0.09	-0.03	0.19	0.07	-0.26	-0.45	0.01
95 % quantile	0.25	0.12	0.35	0.21	-0.13	-0.31	0.15
(2b) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.19	0.11	0.26	0.17	-0.16	-0.36	-0.02
(2b) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.22	0.14	0.39	0.28	-0.14	-0.42	-0.02
5 % quantile	0.11	0.02	0.21	0.14	-0.23	-0.52	-0.10
95 % quantile	0.34	0.28	0.56	0.42	-0.04	-0.31	0.07
(2c) $\hat{\boldsymbol{\tau}}_2^{EM}$	0.10	0.03	0.30	0.18	-0.15	-0.35	0.12
(2c) $\hat{\boldsymbol{\tau}}_2^{MCMC}$	0.11	0.04	0.30	0.18	-0.16	-0.36	0.11
5 % quantile	0.06	-0.01	0.25	0.12	-0.20	-0.40	0.05
95 % quantile	0.16	0.10	0.35	0.22	-0.11	-0.31	0.16

Figure D.2: First and second tree of the "crisis" R-vine structure $\mathcal{V}_2$ which we have chosen for Model (2).



D.1.3 Model (3)

The structure for the first regime is $\mathcal{V}_1$ with copulas selected by AIC, the tree structure for the second regime is $\mathcal{V}_3$ (Figure D.3).

Table D.4: Estimated Kendall's τ, corresponding to the first and second tree of the normal regime in Model (3).

"normal", $\mathcal{V}_1$	GBP,EUR	EUR,CHF	CHF,JPY	AUS,EUR	AUS,BRL	INR,AUS	CAD,AUS	CNY,INR
cop. fam.	SG	N	N	N	G	N	N	G
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.51	0.76	0.49	0.45	0.22	0.16	0.30	0.11
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.52	0.76	0.49	0.45	0.22	0.16	0.30	0.11
5 % quantile	0.49	0.75	0.46	0.42	0.17	0.12	0.26	0.07
95 % quantile	0.55	0.78	0.52	0.48	0.26	0.20	0.34	0.16

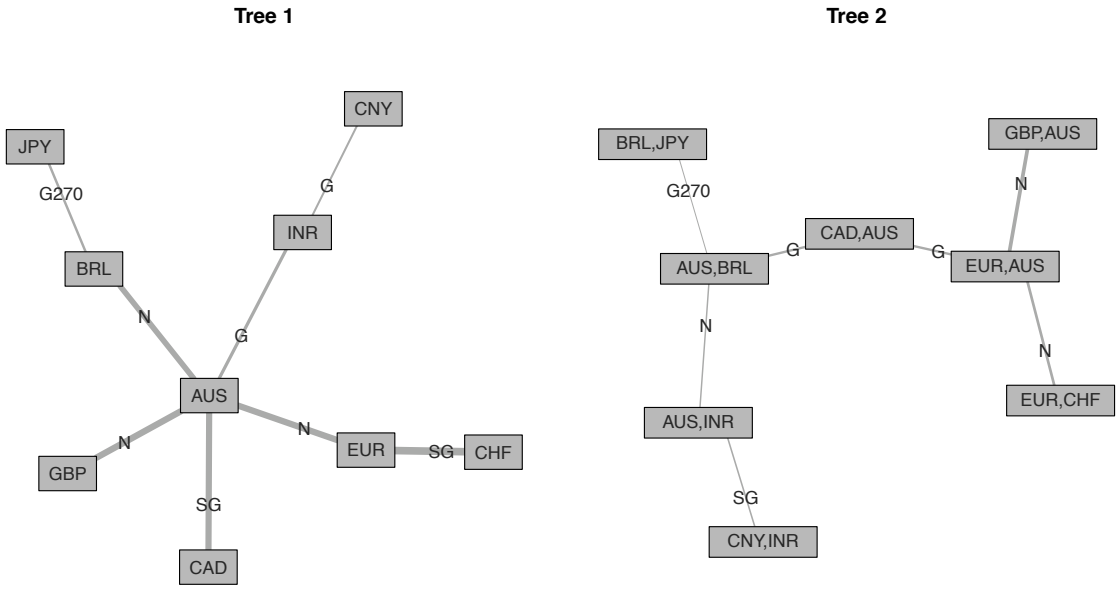
"normal", $\mathcal{V}_1$	JPY,EUR| CHF	AUS,CHF| EUR	AUS,GBP| EUR	CAD,EUR| AUS	CAD,BRL| AUS	INR,BRL| AUS	CNY,AUS| INR
cop. fam.	G270	G 270	G	G	N	N	G
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	-0.09	-0.05	0.14	0.10	0.08	0.05	0.03
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	-0.09	-0.05	0.14	0.11	0.08	0.04	0.04
5% quant.	-0.14	-0.10	0.10	0.07	0.04	-0.01	0.01
95% quant.	-0.04	-0.02	0.18	0.15	0.13	0.09	0.07

Table D.5: Estimated Kendall's τ, corresponding to the first and second tree of the crisis regime in Model (3).

"crisis", $\mathcal{V}_3$	CHF,EUR	EUR,AUS	GBP,AUS	AUS,CAD	AUS,BRL	BRL,JPY	INR,AUS	CNY,INR
cop. fam.	SG	N	N	SG	N	G270	G	G
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.54	0.42	0.42	0.50	0.52	-0.34	0.23	0.06
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.55	0.40	0.41	0.49	0.52	-0.35	0.23	0.06
5% quant.	0.50	0.35	0.35	0.45	0.48	-0.40	0.16	0.00
95% quant.	0.58	0.44	0.46	0.53	0.56	-0.30	0.29	0.12

"crisis", $\mathcal{V}_3$	CNY,AUS| INR	INR,BRL| AUS	AUS,JPY| BRL	CAD,BRL| AUS	CAD,EUR| AUS	GBP,EUR| AUS	AUS,CHF| EUR
cop. fam.	SG	N	G270	G	G	N	N
(2a) $\hat{\boldsymbol{\tau}}_1^{EM}$	0.10	0.11	-0.14	0.10	0.15	0.34	-0.32
(2a) $\hat{\boldsymbol{\tau}}_1^{MCMC}$	0.11	0.11	-0.17	0.13	0.16	0.34	-0.32
5% quant.	0.05	0.05	-0.24	0.05	0.10	0.28	-0.37
95% quant.	0.19	0.17	-0.10	0.19	0.22	0.41	-0.27

Figure D.3: First and second tree of the R-vine structure representing the "crisis" regime of Model (3). We refer to this structure as $\mathcal{V}_3$.



D.1.4 Model (4)

The structures for the two regimes of Model (4) were selected using the model selection heuristic outlined in Section 4.1.4. Other than for Models (1) - (3), we did not truncate after the second tree, but allowed for copulas on all trees to be specified while employing bivariate independence tests as a pre-test for parsimony. While the first and second trees (Figures D.4 and D.5) are similar for the "normal" and the "crisis" regime, the dependence regimes differ

in the corresponding pair copulas (mainly Gaussian copulas for the first tree of the "normal" regime and Gumbel copulas for the "crisis" regime). The higher trees are omitted for brevity and since many of the associated copulas (9 in the "normal" regime and 16 in the "crisis" regime) were chosen to be independence copulas.

Figure D.4: First and second tree of the R-vine structure representing the "normal" regime of Model (4). We refer to this structure as $\mathcal{V}_{4n}$. The edge labels are the copula families and the values of Kendall's τ corresponding to posterior mean estimates.

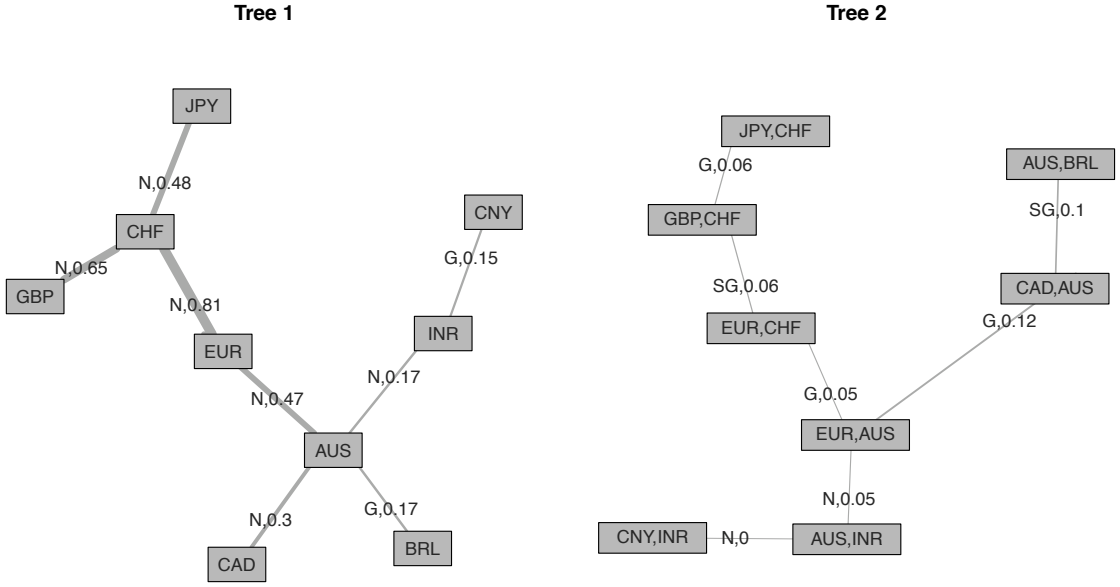
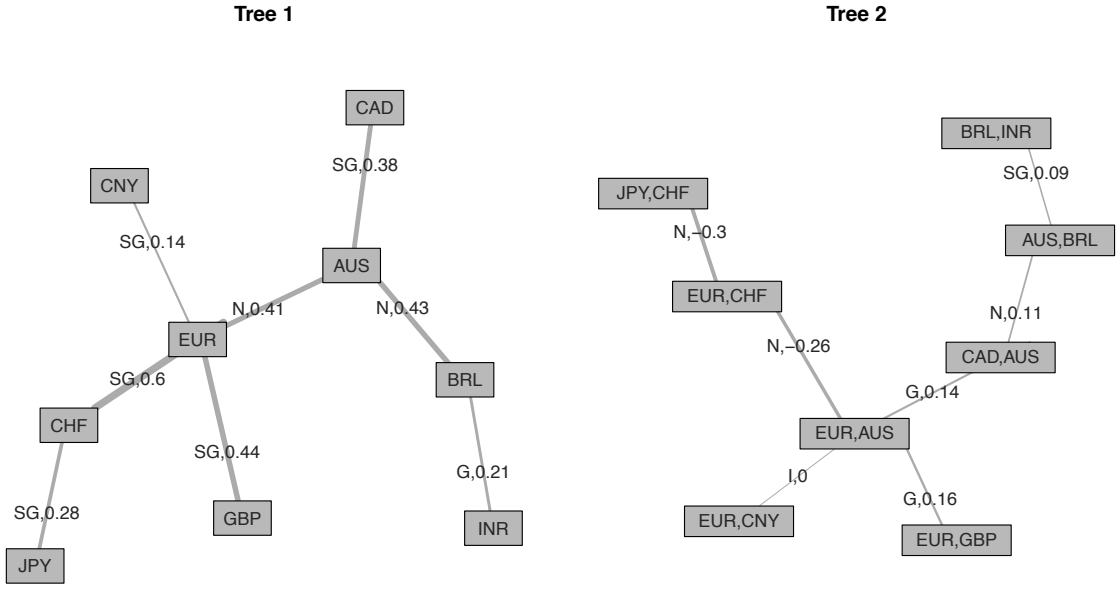


Figure D.5: First and second tree of the R-vine structure representing the "crisis" regime of Model (4). We refer to this structure as $\mathcal{V}_{4c}$. The edge labels are the copula families and the values of Kendall's τ corresponding to posterior mean estimates.



The higher trees and corresponding parameter estimates are available upon request.

D.2 Regime switching marginal distributions

In this section, we provide the remaining plots of smoothed "crisis" probabilities (Figures D.7 - D.6) and the parameter estimates corresponding to the models fitted to the exchange rate data in Section 5.2. The marginal parameter estimates are given in Table D.6, Figures D.9 and D.9 give the selected R-vine tree structures with selected copula families and corresponding parameter estimates being annotated as edge labels.

Figure D.6: Smoothed probabilities of being in the "crisis" regime for the INR/USD and CNY/USD exchange rates. For these time series, we observe less persistent regimes and a higher fluctuation in the period before 2008 than for the other marginal time series.

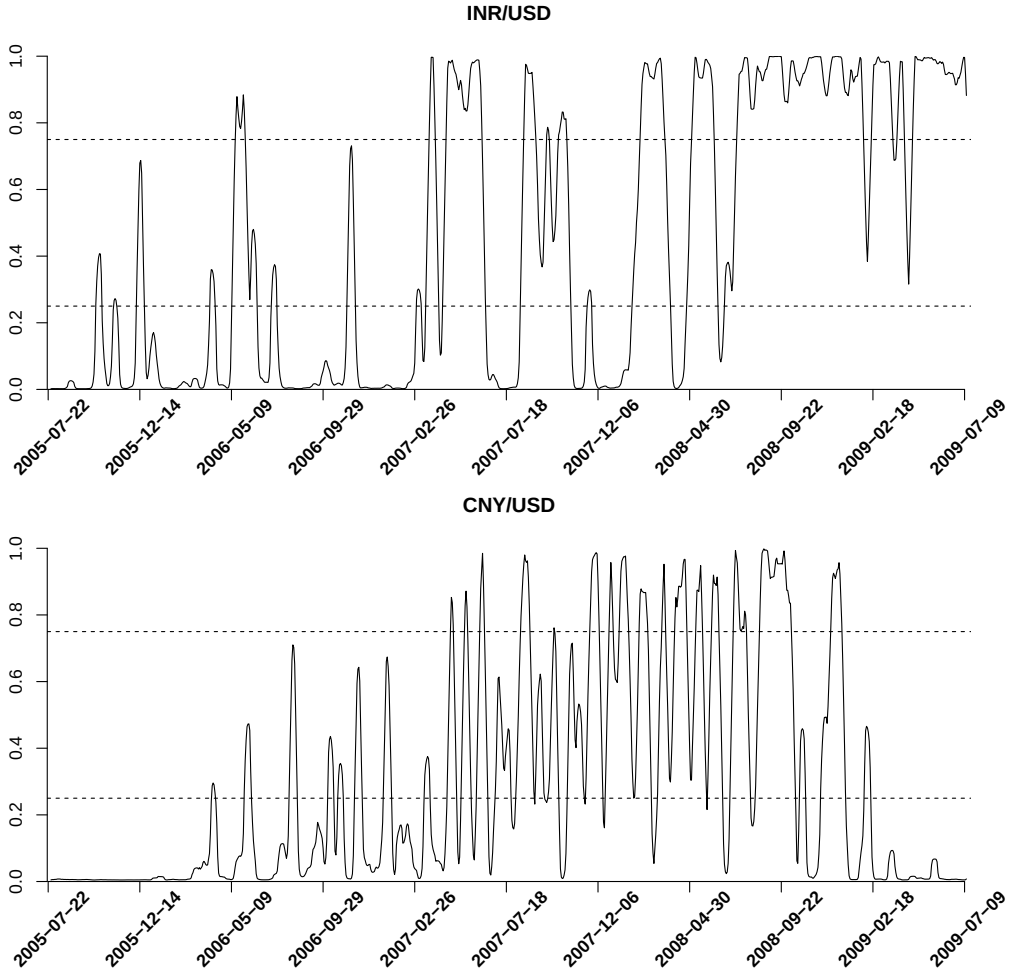


Figure D.7: Smoothed probabilities of being in the "crisis" (high volatility) regime from top to bottom for the GBP/USD, CAD/USD and JPY/USD exchange rates.

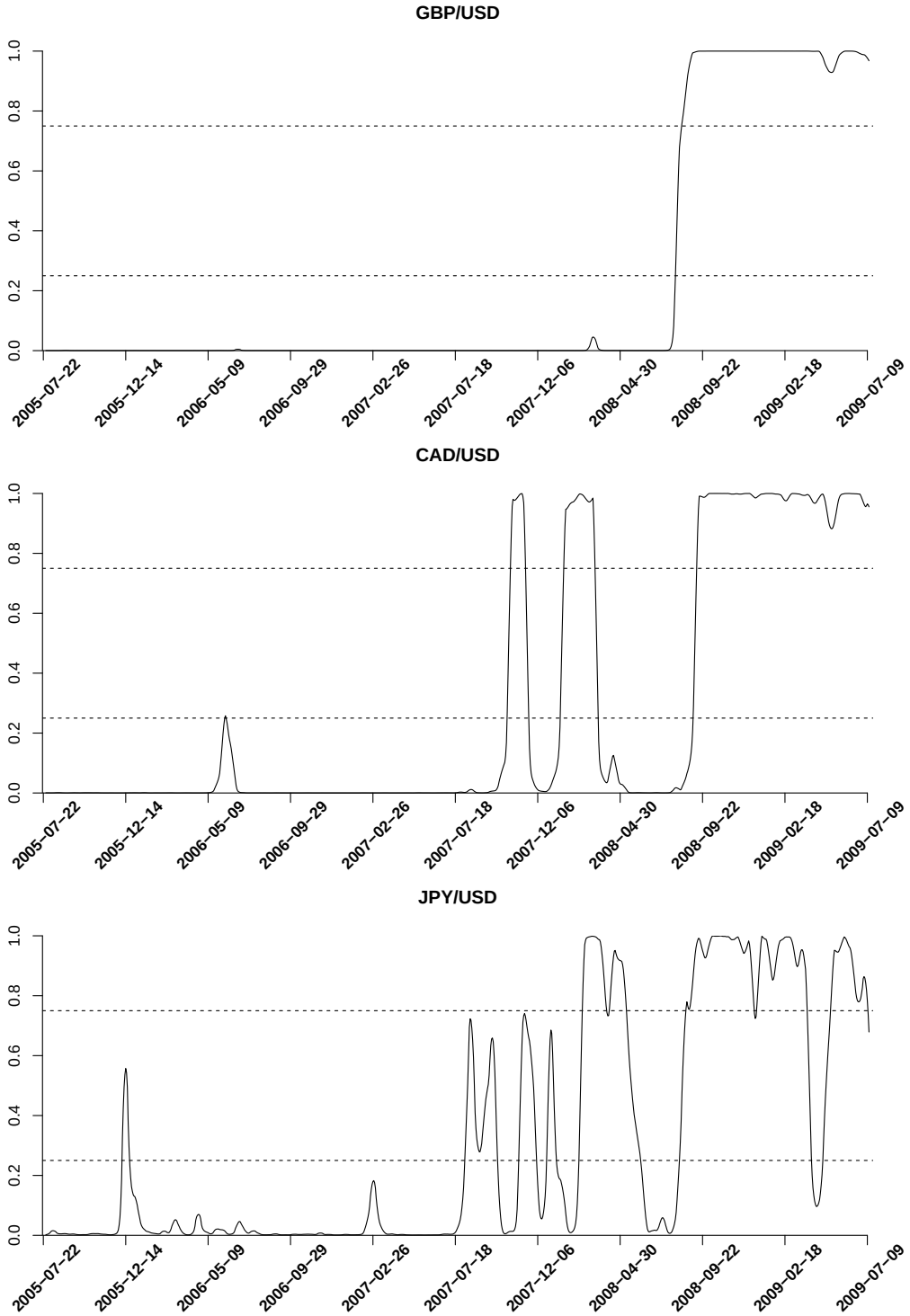


Figure D.8: Smoothed probabilities of being the "crisis" regime for the CHF/USD exchange rate (top panel) the AUD/USD exchange rate (middle panel) and the BRL/USD exchange rate (lower panel). For the BRL/USD exchange rate returns, we observe spikes of high "crisis" probability already in 2005, 2006 and 2007.

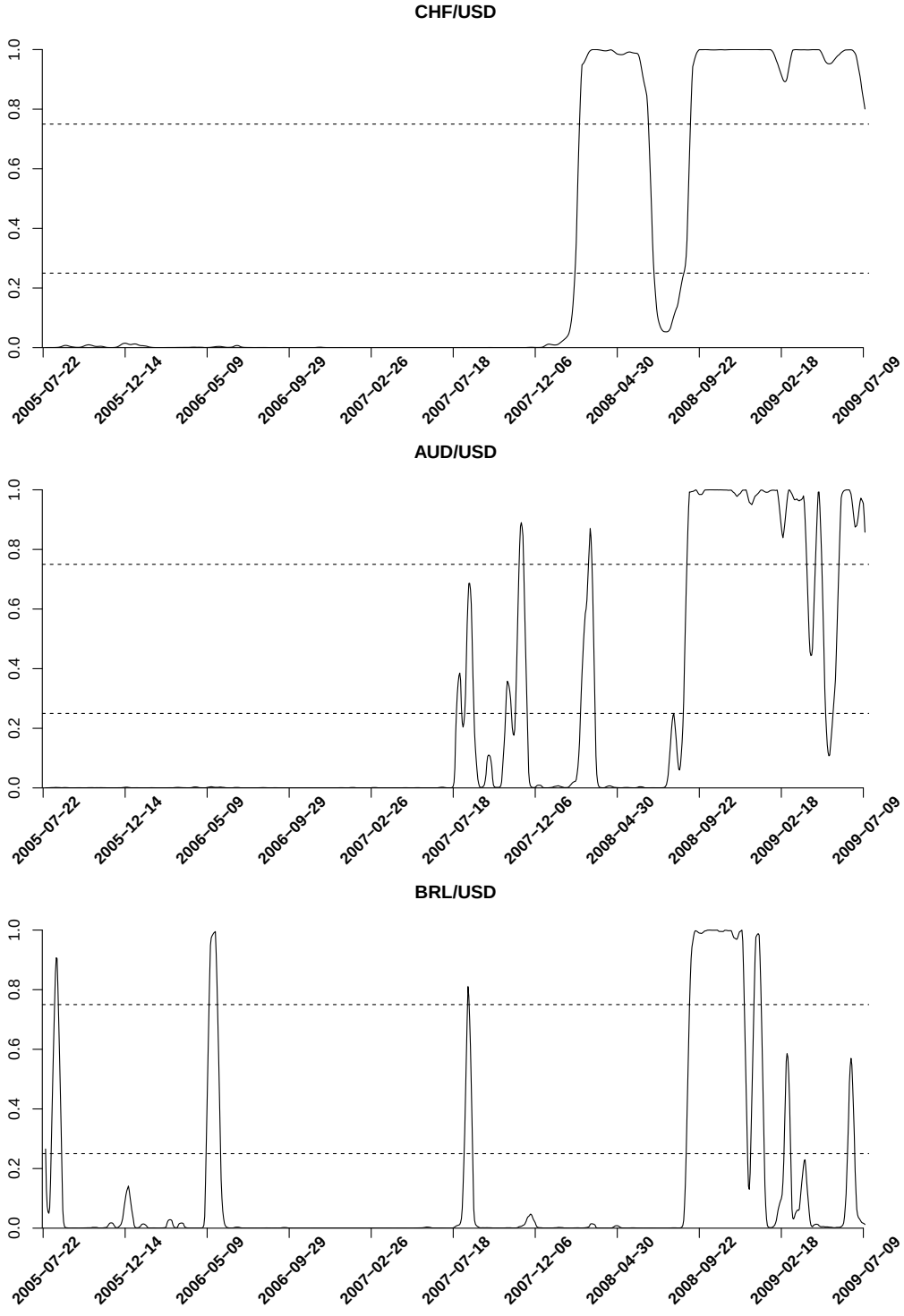
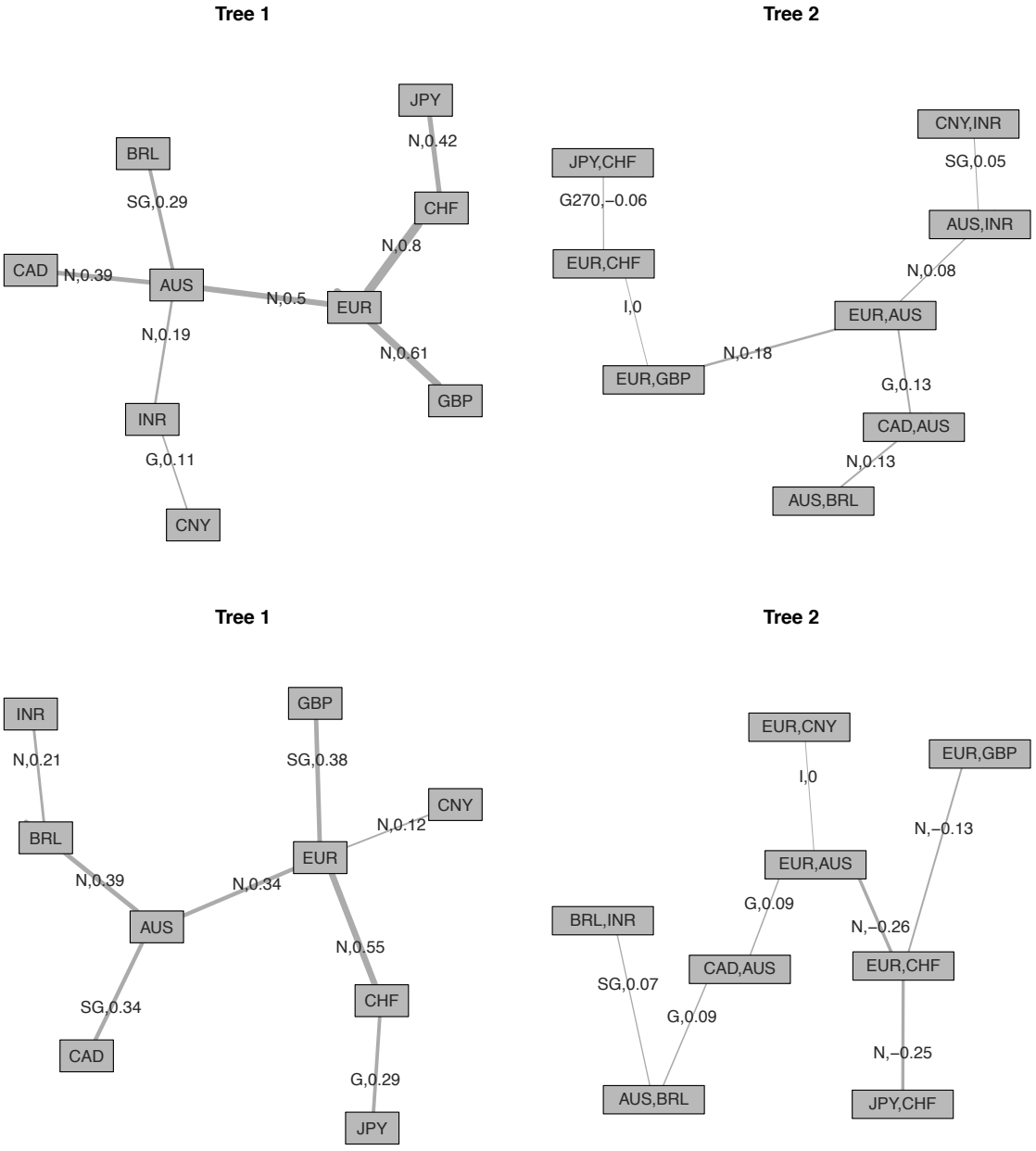
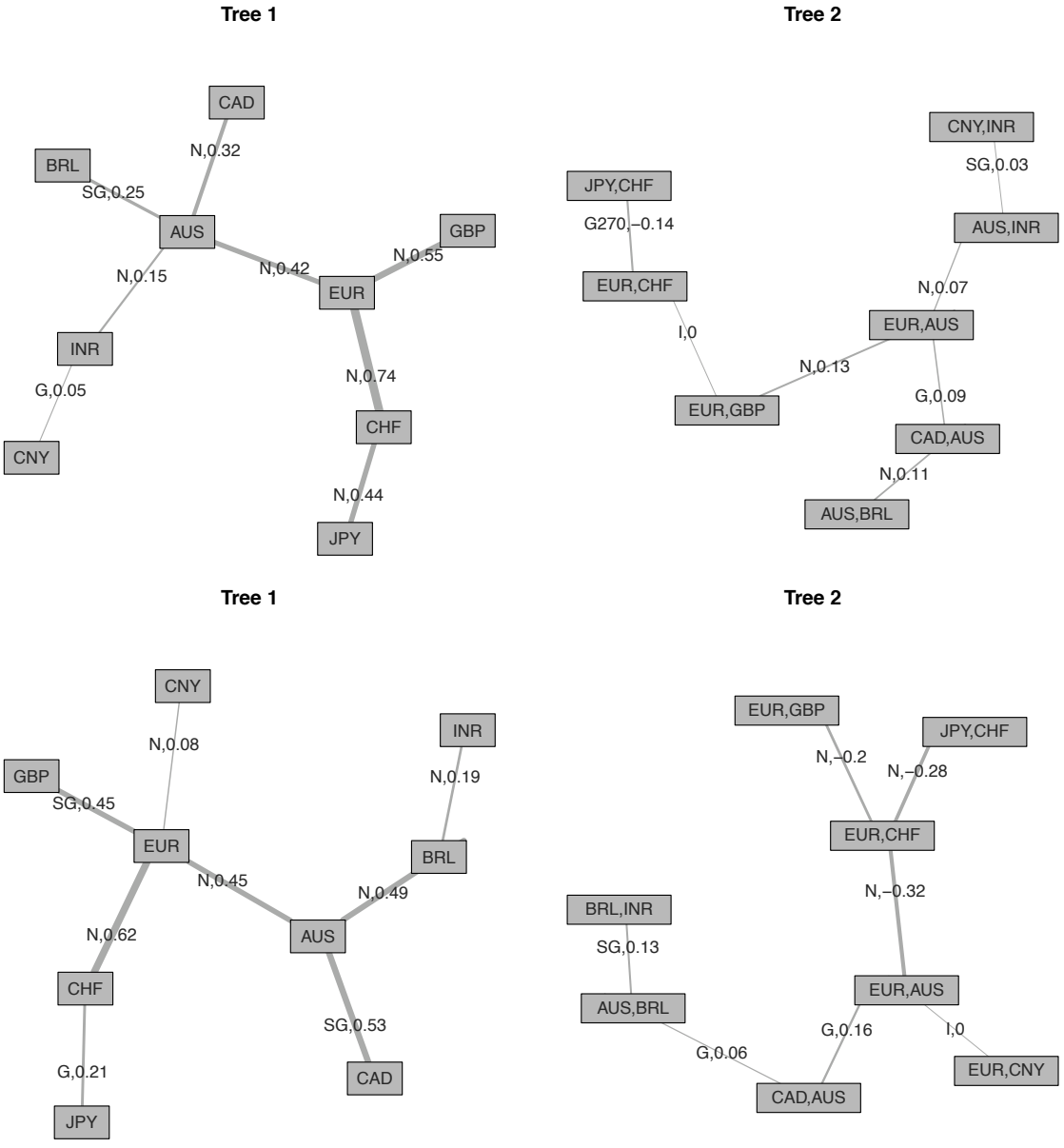


Figure D.9: First and second tree of the R-vine structure representing the "normal" regime (upper panel) and the "crisis" regime (lower panel). The edge labels are the copula families and the values of Kendall's τ corresponding to the estimates of the two-step approach.



The higher trees and corresponding parameter estimates are available upon request.

Figure D.10: First and second tree of the R-vine structure representing the "normal" regime (upper panel) and the "crisis" regime (lower panel). The Kendall's τ values in the edge labels correspond to the estimates for the joint model.



The higher trees and corresponding parameter estimates are available upon request.

Table D.6: Parameters estimated for the "normal" and the "crisis" regime of the marginal time series obtained via separate application of MS models to all marginals (two-step) and the model assuming a joint state variable (joint).

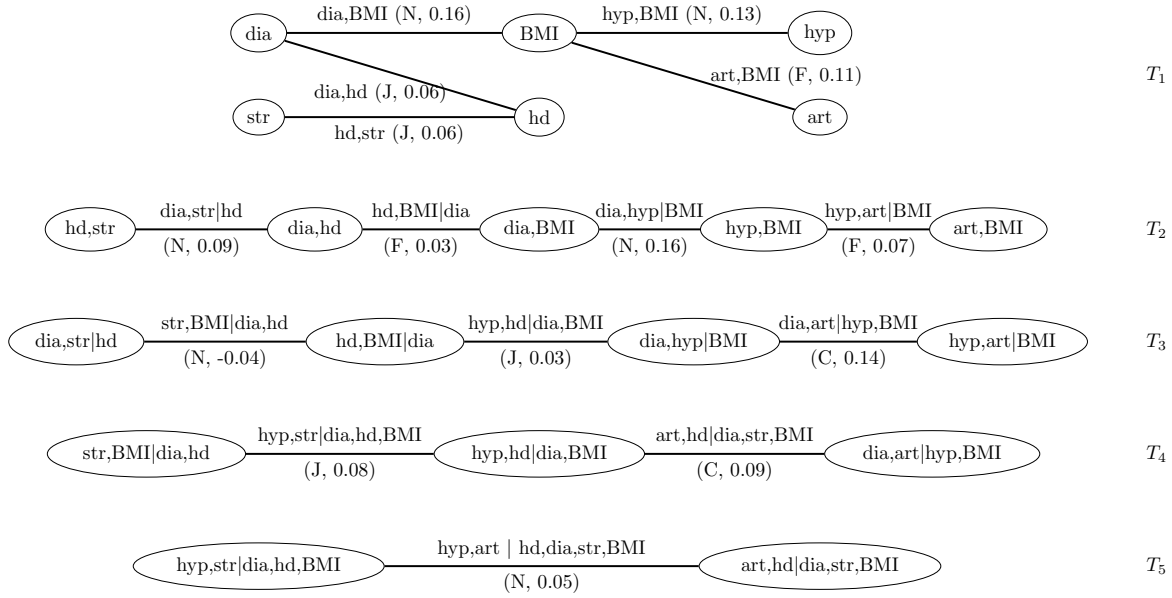
		EUR	GBP	CAD	AUD	BRL	CNY	JPY	CHF	INR
two-step	$100\mu_1$	-0.03	-0.01	-0.02	-0.05	-0.09	-0.01	0.02	-0.02	-0.01
	$100\sigma_1$	0.48	0.48	0.47	0.58	0.76	0.05	0.51	0.53	0.22
	$100\mu_2$	0.03	0.07	0.02	0.14	0.46	-0.04	-0.10	-0.02	0.04
	$100\sigma_2$	1.09	1.26	1.19	2.05	2.87	0.21	1.08	1.04	0.77
joint	$100\mu_1$	-0.03	-0.02	-0.04	-0.05	-0.09	-0.02	0.01	-0.02	-0.01
	$100\sigma_1$	0.47	0.47	0.49	0.57	0.72	0.08	0.53	0.55	0.32
	$100\mu_2$	0.03	0.09	0.08	0.12	0.15	-0.01	-0.09	-0.01	0.07
	$100\sigma_2$	1.05	1.19	1.19	1.93	2.11	0.18	1.12	1.06	0.88

Appendix E

Application 2 (LSOA II model): parameter estimates

In this section, we provide the selected covariates and R-vine tree structures and the parameter estimates for the model we have developed for the LSOA II data in Section 6. The covariates which have been chosen using the AIC criterion for the marginal GLMs are given in Table E.1.

Figure E.1: The R-vine tree structure, pair copulas and corresponding parameter estimates for the first wave of observations of the six response variables. Here, the pair-copulas are parametrized in terms of the theoretical Kendall's tau values which would result in the purely continuous case.



Since we are dealing with data containing both discrete and continuous variables, the theoretical Kendall's rank correlations which would correspond to the estimated parameters in a continuous setup give less information about the strength of dependence. The actual rank correlation values do depend on the marginal distributions here and will be different for different sets of covariates. For these reasons, we give all copula parameter estimates and corresponding standard errors using their standard parameterizations in the tables below (see Schepsmeier and Stöber (2012)). For readers who are more comfortable with the parameterization by Kendall's τ values however, we include these in Figures E.1 and E.2, which show the selected R-vine tree structures and corresponding copula families.

Figure E.2: The R-vine structures which were selected for the follow-up interviews (top: wave 2 and bottom: wave 3).

Table E.1: Included covariates and interaction terms. Covariates which were not selected for all time points are marked in bold.

	Baseline	1. follow-up	2. follow-up
BMI	male, age, edu, income, smoke, male:edu, age:income, age:smoke, income:smoke	male, age, edu, income, smoke, male:edu, age:income, age:smoke, income:smoke	male, age, edu, income, smoke, male:edu, age:income, age:smoke, income:smoke
Hypertension	male, age, edu, income, male:age, male:edu, edu:income	male, age, edu, income, male:age, male:edu, **age:income**, edu:income	male, age, edu, income, male:age, male:edu, edu:income
Diabetes	male, age, edu, income, smoke, male:edu	male, age, edu, income, smoke, male:edu, **edu:income**	male, age, edu, income, smoke, male:edu
Arthritis	male, **age**, edu, income, smoke, **male:age**, male:edu, **male:income**, **edu:income**	male, edu, income, smoke, male:edu	male, edu, income, smoke, male:edu, **male:income**
HD	male, age, edu, **income**, smoke, male:age, age:smoke	male, age, edu, smoke, male:age **male:edu**, **age:edu**, age:smoke, **income:smoke**	male, age, edu, smoke, male:age, **age:edu**, age:smoke
Stroke	male, **age**, **income**	male, **age**, **income**, **smoke**, **male:age**, **income:smoke**	male, **edu**

Table E.2: Estimates of copula parameters for the baseline and the first (f.-up$_1$) and second (f.-up$_2$) follow-up, respectively. The numbers in brackets are the estimated standard errors (std.). The corresponding copula families are given in Figure 3 and Figure 4.

parameter	baseline	(std.)	f.-up$_1$	(std.)	f.-up$_2$	(std.)
art,str|hd,dia,hyp,BMI	0.0749	(0.0551)				
art,hd|dia,hyp,BMI	0.2173	(0.0780)				
dia,str|hd,art,hyp,BMI					1.066	(0.0299)
dia,hd|art,hyp,BMI					0.2357	(0.0449)
hyp,art|str,hd,dia,BMI			0.1477	(0.0486)		
hyp,str|hd,dia,BMI			0.1779	(0.0603)		
hyp,str|hd,art,BMI					1.1365	(0.4073)
BMI,art|str,hd,dia			0.6926	(0.1516)		
hyp,str|hd,dia,BMI	1.1461	(0.0409)				
hyp,hd|art,BMI		(0.0000)			0.3575	(0.0833)
art,dia|hyp,BMI	0.3215	(0.1230)			0.2842	(0.2769)
dia,art|str,hd			0.0377	(0.0462)		
BMI,str|hd,art					-0.5814	(0.3472)
BMI,str|hd,dia	-0.0655	(0.0424)	-0.1952	(0.3870)		
hyp,hd|dia,BMI	1.0569	(0.0293)	0.6819	(0.2239)		
hyp,dia|BMI	0.2424	(0.0468)	1.1434	(0.0321)	1.1914	(0.0427)
BMI,hd|art		(0.0000)			-0.3236	(0.1754)
art,hyp|BMI	0.6752	(0.1819)			0.1747	(0.0479)
BMI,hd|dia	0.2807	(0.1929)	-0.0311	(0.0304)		
dia,str|hd	0.1338	(0.0700)	0.7346	(0.4190)		
art,str|hd			0.4236	(0.1894)	0.3084	(0.1592)
hyp,BMI	0.2046	(0.0253)	1.1202	(0.1494)	1.11	(0.1476)
art,BMI	0.9609	(0.1527)			0.8864	(0.1521)
BMI,dia	0.2415	(0.0363)	0.2458	(0.0337)	0.2175	(0.0322)
dia,hd	1.1162	(0.0307)	1.1211	(0.0303)		
art,hd			0.5027	(0.0930)	0.3956	(0.0781)
str,hd	1.1108	(0.0317)	3.0662	(0.6099)	0.2219	(0.0552)

Table E.3: Estimates of regression parameters for BMI, hypertension (hyp), diabetes (dia) and arthritis (art). Columns correspond to baseline, first (f.-up$_1$) and second (f.-up$_2$) follow-up, respectively. The standard errors (std.) and corresponding p-values (p.val.) are given in brackets.

parameter	baseline	(std., p.val.)	f.-up$_1$	(std., p.val.)	f.-up$_2$	(std., p.val.)
BMI.intercept	4.1008	(0.1480, 0.00)	4.1544	(0.1489, 0.00)	4.2513	(0.1528, 0.00)
BMI.male	-0.0767	(0.0270, 0.00)	-0.0559	(0.0272, 0.04)	-0.0641	(0.0280, 0.02)
BMI.age	-0.009	(0.0019, 0.00)	-0.0101	(0.0019, 0.00)	-0.0115	(0.0019, 0.00)
BMI.edu	-0.0096	(0.0015, 0.00)	-0.0085	(0.0015, 0.00)	-0.0094	(0.0015, 0.00)
BMI.income	-0.0196	(0.0078, 0.01)	-0.0181	(0.0079, 0.02)	-0.0218	(0.0081, 0.01)
BMI.smoke	-0.4173	(0.1123, 0.00)	-0.4008	(0.1128, 0.00)	-0.2765	(0.1161, 0.02)
BMI.male:edu	0.0077	(0.0021, 0.00)	0.0065	(0.0021, 0.00)	0.007	(0.0022, 0.00)
BMI.age:income	0.0002	(0.0001, 0.04)	0.0002	(0.0001, 0.06)	0.0002	(0.0001, 0.02)
BMI.age:smoke	0.0048	(0.0014, 0.00)	0.0045	(0.0014, 0.00)	0.0028	(0.0015, 0.05)
BMI.income:smoke	0.0027	(0.0010, 0.01)	0.0026	(0.0010, 0.01)	0.0026	(0.0011, 0.01)
BMI.lambda	1018.4743	(29.1238, -)	997.1133	(28.5142, -)	927.2608	(26.5108, -)
hyp.intercept	-1.568	(0.9714, 0.11)	1.9823	(1.8848, 0.29)	0.2036	(0.9475, 0.83)
hyp.male	1.2926	(1.5542, 0.41)	3.3813	(1.6180, 0.04)	1.4553	(1.4943, 0.33)
hyp.age	0.0139	(0.0111, 0.21)	-0.0264	(0.0235, 0.26)	-0.0047	(0.0108, 0.67)
hyp.edu	0.0358	(0.0367, 0.33)	0.0121	(0.0361, 0.74)	0.0201	(0.0358, 0.57)
hyp.income	0.0374	(0.0251, 0.14)	-0.1721	(0.1124, 0.13)	0.0402	(0.0240, 0.09)
hyp.male:age	-0.035	(0.0199, 0.08)	-0.0579	(0.0210, 0.01)	-0.0367	(0.0192, 0.06)
hyp.male:edu	0.0838	(0.0298, 0.00)	0.051	(0.0284, 0.07)	0.079	(0.0281, 0.00)
hyp.age:income			0.0025	(0.0014, 0.07)		
dia.edu:income	-0.0047	(0.0021, 0.02)	-0.003	(0.0020, 0.14)	-0.0039	(0.0020, 0.05)
dia.intercept	2.865	(1.4146, 0.04)	2.4728	(1.4093, 0.08)	3.555	(1.2201, 0.00)
dia.male	-0.9745	(0.5611, 0.08)	-1.1107	(0.5693, 0.05)	-0.6109	(0.4830, 0.21)
dia.age	-0.0453	(0.0175, 0.01)	-0.0512	(0.0162, 0.00)	-0.0519	(0.0150, 0.00)
dia.edu	-0.1329	(0.0323, 0.00)	-0.0367	(0.0550, 0.50)	-0.107	(0.0288, 0.00)
dia.income	-0.0218	(0.0132, 0.10)	0.0404	(0.0379, 0.29)	-0.0305	(0.0112, 0.01)
dia.smoke	-0.2645	(0.1590, 0.10)	-0.24	(0.1442, 0.10)	-0.2115	(0.1342, 0.12)
dia.male:edu	0.1059	(0.0458, 0.02)	0.1215	(0.0471, 0.01)	0.085	(0.0393, 0.03)
art.edu:income			-0.0055	(0.0032, 0.09)		
art.intercept	0.8704	(1.0282, 0.40)	1.6874	(0.2360, 0.00)	1.688	(0.2375, 0.00)
art.male	-2.4908	(1.5313, 0.10)	-1.2162	(0.3521, 0.00)	-1.0247	(0.3848, 0.01)
art.age	0.0147	(0.0116, 0.21)				
art.edu	-0.1168	(0.0402, 0.00)	-0.0734	(0.0202, 0.00)	-0.0799	(0.0207, 0.00)
art.income	-0.0503	(0.0266, 0.06)	-0.016	(0.0076, 0.04)	-0.0097	(0.0089, 0.28)
art.smoke	0.1954	(0.0880, 0.03)	0.1681	(0.0871, 0.05)	0.15	(0.0872, 0.09)
art.male:age	0.0281	(0.0194, 0.15)				
art.male:edu	0.0514	(0.0317, 0.11)	0.0619	(0.0273, 0.02)	0.0889	(0.0305, 0.00)
art.male:income	-0.0422	(0.0171, 0.01)			-0.0282	(0.0170, 0.10)

Table E.4: Estimates of marginal parameters for heart disease (hd) and stroke (str) for the baseline and the first (f.-up$_1$) and second (f.-up$_2$) follow-up, respectively. The numbers in brackets are the estimated standard errors (std.) and corresponding p-values (p.val.).

parameter	baseline	(std., p.val.)	f.-up$_1$	(std., p.val.)	f.-up$_2$	(std., p.val.)
hd.edu:income	0.0036	(0.0021, 0.09)				
hd.intercept	-2.9038	(1.3315, 0.03)	5.2583	(3.3673, 0.12)	2.0884	(2.9508, 0.48)
hd.male	3.7663	(1.8113, 0.04)	1.8378	(1.9222, 0.34)	3.212	(1.6912, 0.06)
hd.age	0.0221	(0.0169, 0.19)	-0.0836	(0.0436, 0.06)	-0.0414	(0.0386, 0.28)
hd.edu	-0.0333	(0.0184, 0.07)	-0.5928	(0.2711, 0.03)	-0.4357	(0.2376, 0.07)
hd.income	-0.0146	(0.0097, 0.13)	0.008	(0.0121, 0.51)		
hd.smoke	-4.1742	(1.7500, 0.02)	-5.4277	(1.8272, 0.00)	-3.4623	(1.6127, 0.03)
hd.male:age	-0.0429	(0.0239, 0.07)	-0.0365	(0.0250, 0.15)	-0.0393	(0.0223, 0.08)
hd.male:edu			0.0876	(0.0340, 0.01)		
hd.age:edu			0.0071	(0.0035, 0.05)	0.0054	(0.0031, 0.08)
hd.age:smoke	0.0593	(0.0230, 0.01)	0.0823	(0.0232, 0.00)	0.0471	(0.0212, 0.03)
str.income:smoke			-0.0373	(0.0164, 0.02)		
str.intercept	-4.8408	(1.4335, 0.00)	-5.8913	(2.2146, 0.01)	-2.6835	(0.3562, 0.00)
str.male	0.482	(0.1846, 0.01)	6.1507	(3.7395, 0.10)	0.5452	(0.1913, 0.00)
str.age	0.0314	(0.0180, 0.08)	0.0302	(0.0274, 0.27)		
str.edu					-0.0457	(0.0290, 0.12)
str.income	-0.0315	(0.0139, 0.02)	0.0119	(0.0235, 0.61)		
str.smoke			1.1686	(0.6097, 0.06)		
str.male:age			-0.0773	(0.0498, 0.12)		
str.income:smoke			-0.0652	(0.0331, 0.05)		

Bibliography

Aas, K., C. Czado, A. Frigessi, and H. Bakken. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44:182–198.

Abyad, A, and J. T. Boyer. 1992. Arthritis and aging. *Current opinion in Rheumatology* 4 (2): 153–159.

Acar, E. F., C. Genest, and J. Nešlehová. 2012a. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis* 110:74–90.

Acar, E. F., R. V. Craiu, and F. Yao. 2012b. Statistical testing for conditional copulas. Preprint, arXiv:1204.6644.

Ahmadi Javid, A. 2009. Copulas with truncation-invariance property. *Communications in Statistics - Theory and Methods* 38:3756–3771.

Akaike, H. 1974. A new look at statistical model identification. *IEEE Transaction on Automatic Control* 19 (6): 716–723.

Albert, J. H., and S. Chib. 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics* 11 (1): 1–15.

Almeida, C., and C. Czado. 2011. Efficient Bayesian inference for stochastic time-varying copula models. *Computational Statistics and Data Analysis* 56 (6): 1511–1527.

Ang, A., and G. Bekaert. 2002. International asset allocation with regime switching. *Review of Financial Studies* 11:1137–1187.

Ang, A., and J. Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–494.

Arlot, S., and A. Celisse. 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys* 4:40–79.

Bartram, S., S. Taylor, and Y. Wang. 2007. The Euro and European financial market dependence. *Journal of Banking and Finance* 31:1461–1481.

Bauer, A., C. Czado, and T. Klein. 2012. Pair-copula constructions for non Gaussian DAG models. *The Canadian Journal of Statistics* 40 (2): 86–109.

Bedford, T., and R. 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. Cooke. 2002. Vines - a new graphical model for dependent random variables. *Annals of Statistics* 30:1031–1068.

Boyer, C., and U. Merzbach. 2011. *A history of mathematics.* Wiley.

Brechmann, E., C. Czado, and K. Aas. 2012. Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics* 40 (1): 68–85.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research* 33 (2): 261–304.

Cambanis, S., S. Huang, and G. Simons. 1981. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* 11:386–385.

Carlin, B. P., and T. A. Louis. 2009. *Bayesian methods for data analysis.* Third. Chapman & Hall/CRC, Boca Raton, Florida.

Cerra, V., and S. C. Saxena. 2005. Did output recover from the Asian crisis? *IMF staff papers* 52 (1): 1–23.

Chauvet, M., and J. D. Hamilton. 2006. Dating business cycle turning points. In *Nonlinear analysis of business cycles,* 1–54. Elsevier.

Chen, X., and Y. Fan. 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econometrics* 135:125–154.

Chollete, L., A. Heinen, and A. Valdesogo. 2009. Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics* 7 (4): 437–480.

Clayton, D. G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrica* 65 (1): 141–151.

Cook, R. D., and M. E. Johnson. 1981. A family of distributions for modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society Series B* 43:210–218.

Cooke, R. M., C. Kousky, and H. Joe. 2011. Micro correlations and tail dependence. Chap. 5 in *Dependence Modeling: Vine Copula Handbook,* ed. D. Kurowicka and H. Joe, 89–112. Singapore: World Scientific Publishing.

Cornish, E. A. 1954. The multivariate t-distribution associated with a set of normal sample deviates. *Australian Journal of Physics* 7:531–542.

Czado, C. 2010. Pair-copula constructions of multivariate copulas. In *Copula Theory and Its Applications, Lecture Notes in Statistics,* 93–109. Vol. 198. Heidelberg: Springer-Verlag.

Czado, C., F. Gärtner, and A. Min. 2011. Joint Bayesian inference of D-vines with AR(1) margins. Chap. 13 in *Dependence Modeling: Vine Copula Handbook,* ed. D. Kurowicka and H. Joe, 330–359. Singapore: World Scientific Publishing.

Czado, C., U. Schepsmeier, and A. Min. 2012. Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling* 12 (3): 229–255.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39 (1): 1–38.

Dickey, J. M. 1967. Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematical Statistics* 38 (2): 511–518.

Dißmann, J. 2010. Statistical inference for regular vines and application. Master's thesis, Technische Universität München, Germany.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka. 2013. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis* 59:52–69.

Elia, M. 2001. Obesity in the elderly. *Obesity Research* 9:244S–248S.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond,* 176–223. Cambridge University Press.

Engel, C. 1994. Can the Markov switching model forecast exchange rates? *Journal of International Economics* 36:151–165.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4): 987–1007.

European Banking Authority. 2012. EBA guidelines on stressed Value at Risk (Stressed VaR). EBA/GL/2012/2.

Feller, W. 1971. *An introduction to probability theory and its applications.* Second. Vol. 2. New York: Wiley.

Fleischer, N., R. A. Diez, M Alazraqui, H Spinelli, and F De Maio. 2011. Socioeconomic gradients in chronic disease risk factors in middle-income countries: evidence of effect modification by urbanicity in argentina. *American Journal of Public Health* 101:294–301.

Frühwirth-Schnatter, S. 2001. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 96 (453): 194–209.

Gamerman, D., and H. F. Lopes. 2006. *Markov Chain Monte Carlo - stochastic simulation for Bayesian inference.* Chapman & Hall / CRC.

Garcia, R., and G. Tsafack. 2011. Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking & Finance* 35 (8): 1954–1970.

Genest, C., and A. Favre. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12:347–368.

Genest, C., and R. MacKay. 1986. Copules archimédiennes et familles de lois bidimensionelles dont les marges sont données. *Canadian Journal of Statistics* 14:145–159.

Genest, C., K. Ghoudi, and L.-P. Rivest. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552.

Gneiting, T., and A. E. Raftery. 2007. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association* 102 (477): 359–378.

Green, P. J. 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)* 46 (2): 149–192.

Gruber, L. F., C. Czado, and J. Stöber. 2012. Bayesian model selection for R-vine copulas using reversible jump MCMC. Preprint.

Hamilton, J. D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384.

Hamilton, J. D. 1990. Analysis of time series subject to regime changes. *Journal of Econometrics* 45:39–70.

Hamilton, J. D. 2005. What's real about the business cycle? *Federal Reserve Bank of St. Louis Review* 87 (4): 435–452.

Heinen, A., and A. Valdesogo. 2009. Asymmetric capm dependence for large dimensions: the canonical vine autoregressive model. CORE Discussion Papers from Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Hobæk Haff, I., K. Aas, and A. Frigessi. 2010. On the simplified pair-copula construction - simply useful or too simplistic? *Journal of Multivariate Analysis* 101:1296–1310.

Hobæk Haff, I. 2012. Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis* (Orlando, FL, USA) 110:91–105.

Hobæk Haff, I. 2013. Parameter estimation for pair-copula constructions. *Bernoulli* 19:462–491.

Hua, L., and H. Joe. 2012. Tail comonotonicity and conservative risk measures. *ASTIN Bulletin* 42 (2): 601–629.

Hull, J. C., and A. D. White. 2004. Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation. *The Journal of Derivatives* 12 (2): 8–23.

Hull, J. C., and A. D. White. 2006. Valuing credit derivatives using an implied copula approach. *The Journal of Derivatives* 14 (2): 8–28.

Joe, H. 1996. Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In *Distributions with fixed marginals and related topics,* ed. L. Rüschendorf and B. Schweizer and M. D. Taylor, 120–141. Vol. 28. Hayward, CA: Institute of Mathematical Statistics.

Joe, H. 1997. *Multivariate models and dependence concepts.* Chapman & Hall, London.

Joe, H., and J. J. Xu. 1996. *The estimation method of inference functions for margins for multivariate models.* Technical report. University of British Columbia.

Joe, H., H. Li, and A. K. Nikoloulopoulos. 2010. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis* 101 (1): 252–270.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal. 1998. Markov Chain Monte Carlo in practice: a roundtable discussion. *The American Statistician* 52 (2): 93–100.

Kim, C.-J., and C. R. Nelson. 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics* 80 (2): 188–201.

Kim, C.-J., and C. R. Nelson. 2006. *State-space models with regime switching.* MIT Press, Cambridge.

Kim, G., M. J. Silvapulle, and P. Silvapulle. 2007. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis* 51 (6): 2836 –2850.

Kimeldorf, G., and A. R. Sampson. 1975. Uniform representations of bivariate distributions. *Communications in Statistics - Simulation and Computation* 4:617–627.

König, D. 1936. *Theorie der endlichen und unendlichen graphen: kombinatorische topologie der streckenkomplexe. akademische verlagsgesellschaft.* Reprinted in 2001 as *Theorie Der Endlichen Und Unendlichen Graphen.* AMS Chelsea Publishing. American Mathematical Society. Leipzig: Akademische Verlagsgesellschaft.

Kotz, S., N. Balakrishnan, and N. L. Johnson. 2000. *Continuous multivariate distributions, volume 1: models and applications.* Second. John Wiley & Sons, New York.

Kriegsman, D., B. Penninx, J. van Eijk, A. Boeke, and D. Deeg. 1996. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *Journal of Clinical Epidemiology* 49:1407–1417.

Kurowicka, D., and R. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modelling.* John Wiley & Sons Ltd, Chichester.

Kurowicka, D., and H. Joe. 2011. *Dependence modeling: vine copula handbook.* Wiley Series in Probability and Statistics. Singapore: World Scientific Publishing.

Lewandowski, D., D. Kurowicka, and H. Joe. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* (Orlando, FL, USA) 100 (9): 1989–2001.

Li, D. X. 2000. On default correlation: a copula function approach. *Journal of Fixed Income* 9 (4): 42–54.

Lindskog, F., A. McNeil, and U. Schmock. 2003. Kendall's tau for elliptical distributions. In *Credit risk: measurement, evaluation and management,* ed. G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder, and K.-H. Vollmer, 149–156. Physica.

Longin, F., and B. Solnik. 1995. Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance* 14 (1): 3–26.

Longin, F., and B. Solnik. 2001. Extreme correlations in international equity markets. *Journal of Finance* 56:649–676.

Lugtenberg, M., J. S. Burgers, C. Clancy, G. P. Westert, and E. C. Schneider. 2011. Current guidelines have limited applicability to patients with comorbid conditions: a systematic analysis of evidence-based guidelines. *PLoS ONE* 6, no. 10 (Oct.): e25987.

Lukacs, E. 1955. A characterization of the gamma distribution. *The Annals of Mathematical Statistics* 26 (2): 319–324.

Mai, J.-F., and M. Scherer. 2012. *Simulating copulas: stochastic models, sampling algorithms, and applications.* Series in Quantitative Finance. World Scientific.

Manner, H., and J. Segers. 2011. Tails of correlation mixtures of elliptical copulas. *Insurance: Mathematics and Economics* 48 (1): 153 –160.

Mardia, K. V. 1962. Multivariate pareto distributions. *Ann. Math. Statist.* 33:1008–1015.

Marshall, A. W., and I. Olkin. 2007. *Life distributions: structure of nonparametric, semiparametric and parametric families.* New York: Springer.

Masaki, K. H., J. D. Curb, D. Chiu, H. Petrovitch, and B. L. Rodriguez. 1997. Association of body mass index with blood pressure in elderly japanese american men: the honolulu heart program. *Hypertension* 29 (2): 673–677.

McNeil, A., and J. Nešlehová. 2009. Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *The Annals of Statistics* 37:3059–3097.

McNeil, A., R. Frey, and P. Embrechts. 2005. *Quantitative risk management: concepts, techniques, and tools.* Princeton Series in Finance. Princeton University Press.

Mesfioui, M., and J.-F. Quessy. 2008. Dependence structure of conditional archimedean copulas. *Journal of Multivariate Analysis* 99:372–385.

Mikosch, T. 2006. Copulas: tales and facts. *Extremes* 9 (1): 3–20.

Min, A., and C. Czado. 2010. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics* 8 (4): 511–546.

Moore, L., A. Visioni, P. Wilson, R. D'Agostino, W. Finkle, and R. Ellison. 2000. Can sustained weight loss in overweight individuals reduce the risk of diabetes mellitus? *Epidemiology* 11 (3): 269–273.

Morales-Nápoles, O. 2011. Counting vines. Chap. 9 in *Dependence modeling: vine copula handbook,* ed. D. Kurowicka and H. Joe, 189–218. Singapore: World Scientific Publishing.

Müller, A., and M. Scarsini. 2005. Archimedean copulae and positive dependence. *Journal of Multivariate Analysis* 93:434–445.

Nelsen, R. B. 2006. *An introduction to copulas.* Springer, New York.

Nguyen, N., X. Nguyen, J Lane, and P Wang. 2011. Relationship between obesity and diabetes in a us adult population: findings from the national health and nutrition examination survey, 1999-2006. *Obes Surg* 21:351–355.

Nikoloulopoulos, A., H. Joe, and H. Li. 2009. Extreme value properties of multivariate t copulas. *Extremes* 12 (2): 129–148.

Oh, D.-H., and A. J. Patton. 2012. Modelling dependence in high dimensions with factor copulas. Working paper, Duke University.

Panagiotelis, A., C. Czado, and H. Joe. 2012. Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association* 107 (499): 1063–1072.

Patton, A. J. 2012. Copula methods for forecasting multivariate time series. In *Handbook of economic forecasting, volume 2,* ed. G. Elliott and A. Timmermann. Forthcoming. Oxford: Elsevier.

Patton, A. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* 47 (2): 527–556.

Pelletier, D. 2006. Regime switching for dynamic correlations. *Journal of Econometrics* 131 (1-2): 445–473.

Pozzi, F., T. Matteo, and T. Aste. 2012. Exponential smoothing weighted correlations. *The European Physical Journal B - Condensed Matter and Complex Systems* 85 (6): 1–21.

Prim, R. C. 1957. Shortest connection networks and some generalizations. *Bell System Technology Journal* 36:1389–1401.

R Development Core Team. 2011. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. `http://www.R-project.org/`.

Rockafellar, R. T. 1970. *Convex analysis.* Princeton University Press, Princeton, NJ.

Salmon, F. 2009. Recipe for Disaster: The Formula That Killed Wall Street. *Wired* 17, no. 3 (Mar.): 74–79, 112.

Schäfer, I., E.-C. von Leitner, G. Schön, D. Koller, H. Hansen, T. Kolonko, H. Kaduszkiewicz, K. Wegscheider, G. Glaeske, and H. van den Bussche. 2010. Multimorbidity patterns in the elderly: a new approach of disease clustering identifies complex interrelations between chronic conditions. *PLoS ONE* 5 (12): e15941.

Schepsmeier, U. 2010. Maximum likelihood estimation of C-vine pair-copula constructions on bivariate copulas from different families. Diplomarbeit, Center of Mathematical Sciences, Munich University of Technology.

Schepsmeier, U., and J. Stöber. 2012. Derivatives and fisher information of bivariate copulas. Forthcoming in Statistical Papers.

Schepsmeier, U., J. Stöber, and E. C. Brechmann. 2012. *Vinecopula: statistical inference of vine copulas.* Technische Universität München. `http://cran.r-project.org/web/packages/VineCopula/`.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6 (2): 461–464.

Shih, M, J. Hootman, J Kruger, and C. Helmick. 2006. Physical activity in men and women with arthritis: national health interview survey, 2002. *American Journal of Preventive Medicine* 30:385–393.

Sklar, M. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8:229–231.

Smith, M. S., A. Min, C. Almeida, and C. Czado. 2010. Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association* 105 (492): 1467–1479.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64 (4): 583–639.

Stöber, J., and U. Schepsmeier. 2013. Estimating standard errors in regular vine copula models. Forthcoming in Computational Statistics.

Stöber, J., and C. Czado. 2012. Sampling pair copula constructions with applications to mathematical finance. In *Simulating copulas, stochastic models, sampling algorithms, and applications.*

Stöber, J., and C. Czado. 2013. Regime switches in the dependence structure of multivariate financial data. Forthcoming in Computational Statistics and Data Analysis.

Stöber, J., H. G. Hong, C. Czado, and P. Ghosh. 2012. Comorbidity of chronic diseases in the elderly: longitudinal patterns identified by a copula design for mixed responses. Preprint.

Stöber, J., H. Joe, and C. Czado. 2013. Simplified pair copula constructions — limitations and extensions. *Journal of Multivariate Analysis* 119:101–118.

Takahasi, K. 1965. Note on the multivariate burr's distribution. *Annals of the Institute of Statistical Mathematics* 17:257–260.

Thulasiraman, K., and M. N. S. Swamy. 1992. *Graphs: theory and algorithms.* New York, NY, USA: John Wiley & Sons, Inc.

Timmermann, A. 2000. Moments of Markov switching models. *Journal of Econometrics* 96 (1): 75–111.

U.S. Census Bureau. 2012. *Percent Distribution of the Projected Population by Selected Age Groups and Sex for the United States: 2015 to 2060 (NP2012-T3).* `http://www.census.gov/population/projections/files/summary/NP2012-T3.csv`.

Werner, C. A. 2010. The older population: 2010. *The US Census Bureau, 2010 Census briefs* 9.

White, H. 1996. *Estimation, inference and specification analysis.* Econometric Society Monographs. Cambridge University Press.

Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *The Annals of Statistics* 11 (1): 95–103.

Zakkak, J., D. Wilson, and J. Lanier. 2009. The association between body mass index and arthritis among US adults: CDC's surveillance case definition. *Preventing Chronic Disease* 6 (2): 14:1–14:11.