Technische Universität München
Zentrum Mathematik
Lehrstuhl für mathematische Statistik

Bayesian time series modeling with copula structures

Alexander Kreuzer

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:		Prof. Dr. Matthias Scherer
Prüfer der Dissertation:	1.	Prof. Claudia Czado, Ph.D.
	2.	Prof. Dr. Sylvia Frühwirth–Schnatter
	3.	Prof. Anastasios Panagiotelis, Ph.D.

Die Dissertation wurde am 27.11.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 17.02.2020 angenommen.

Zusammenfassung

Viele multivariate Zeitreihenmodelle basieren auf der multivariaten Normalverteilung, obwohl das nicht immer gerechtfertigt ist. In dieser Arbeit wird eine Reihe von copula-basierten Ansätzen präsentiert, um Zeitreihen zu modellieren. Mit der Hilfe von Copulas können z.B. symmetrische und asymmetrische Tailabhängigkeiten modelliert werden. Dies wäre mit einer multivariaten Normalverteilung nicht möglich. Zum Schätzen der neu-entwickelten Modelle verwenden wir Markov-Ketten-Monte-Carlo (MCMC) Methoden, wie z.B. Hamiltonian Monte Carlo oder Elliptical Slice Sampling.

Zuerst wird ein Ein-Faktor Copula Modell vorgestellt. Dieses Modell ist eine Verallgemeinerung von stochastischen Volatilitätsmodellen mit einem Faktor und konstanten Korrelationen. Zur Parameterschätzung wird ein MCMC Verfahren entwickelt, bei dem die Marginalmodelle und die Copula gemeinsam geschätzt werden. Im Gegensatz zu einem zweistufigem Schätzverfahren wird die Unsicherheit in den Schätzungen der Marginalmodelle nicht mehr ignoriert und wir erwarten präzisere Schätzwerte für die Modellparameter.

Außerdem wird ein effizientes MCMC Verfahren entwickelt mit dem wir die Parameter einer bestimmten Klasse von Zustandsraummodellen mit univariater autoregressiver Zustandsgleichung schätzen können. Diese Klasse beinhaltet univariate stochastische Volatilitätsmodelle und dynamische bivariate Copula Modelle. Unser Ansatz basiert auf Elliptical Slice Sampling, einer adaptiven Metropolis-Hastings Methode und auf einer Ancillarity-Sufficiency Interweaving Strategie. In einer Simulationsstudie wird mit dynamischen bivariaten Copula Modellen untersucht, wie effizient das neue Verfahren ist.

Mithilfe der Vine Copula Theorie, werden die dynamischen bivariaten Copula Modelle zu beliebigen Dimensionen erweitert. Dadurch erhalten wir eine neue Klasse von dynamischen Vine Copula Modellen. In dieser Modellklasse sind allgemeine reguläre Vine Strukturen erlaubt. Wir beschränken uns also nicht nur auf C-vine oder D-vine Modelle. Außerdem wird ein erstes Bayesianisches Schätzverfahren für solche Modelle entwickelt.

Die bis jetzt erwähnten Modelle sind Zustandsraummodelle, bei denen die Zustandsgleichungen durch normalverteilte autoregressive Prozesse beschrieben werden. Zusätzlich werden auch flexiblere copula-basierte Zustandsraummodelle entwickelt, bei denen sowohl die Beobachtungsgleichung als auch die Zustandsgleichung durch Copulas spezifiziert sind. Wir betrachten dabei ein eindimensionales und ein mehrdimensionales Modell. Für die Schätzung dieser Modelle verwenden wir den No-U-Turn Sampler, eine Erweiterung von Hamiltonian Monte Carlo.

Alle entwickelten Modelle werden auf echte Daten angewandt, wie z.B. auf Finanzzeitreihen und Luftverschmutzungsdaten, und mit relevanten Benchmark-Modellen verglichen.

Abstract

Many time series models rely on the assumption of multivariate normality, although this assumption is not always appropriate. We present a variety of Bayesian (vine) copula based approaches to provide more flexible time series models. Relying on copulas allows to model features which cannot be described with a multivariate normal distribution, such as symmetric or asymmetric tail dependence. Estimation of these models is handled with Markov Chain Monte Carlo (MCMC) methods. This includes Hamiltonian Monte Carlo and elliptical slice sampling.

The first contribution is a single factor copula based stochastic volatility model, a generalization of Gaussian stochastic volatility models with one factor and constant correlations. For this model we develop joint Bayesian inference using Hamiltonian Monte Carlo within Gibbs sampling, instead of relying on the popular two-step approach. In contrast to the two-step approach, we expect more accurate estimates since uncertainty in the estimation of the marginal distribution is no longer ignored.

Next, an efficient MCMC approach for a class of nonlinear state space models with univariate autoregressive state equation is developed. This class includes stochastic volatility models and dynamic bivariate copula models. The sampler is based on elliptical slice sampling, adaptive Metropolis-Hastings and on an ancillarity-sufficiency interweaving strategy. Its sampling efficiency is investigated through an extensive simulation study for bivariate dynamic copula models.

Using the vine copula framework, we scale the dynamic bivariate copula model to arbitrary dimensions. This yields a class of dynamic vine copula models. In contrast to previous work on dynamic vine copulas, we develop a first Bayesian estimation procedure and our class allows for general vine structures instead of restricting the approach to only C-vine or D-vine copulas. The Bayesian approach is based on a novel approximation of the posterior distribution.

While the above mentioned models can be considered as state space models, where the state equations are described by Gaussian autoregressive processes, we also develop a very flexible class of copula based state space models. We study an univariate and a multivariate copula based state space model. In the multivariate model we assume a single factor structure in the observation equation. These novel state space models allow for great flexibility by specifying the observation as well as the state equation with copulas. For the estimation of these models we employ STAN's No-U-Turn sampler, an extension of Hamiltonian Monte Carlo.

All the proposed models are illustrated with real data, including financial returns data and atmospheric pollution measurements data, and are compared to relevant benchmark models showing superior performance.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Claudia Czado for making the PhD possible. You already provided great guidance during my Bachelor's and Master's thesis where I discovered my interest in applied statistics. This lead me to the decision to pursue a PhD. During the PhD you supported me to develop my own ideas and taught me how to approach problems as a researcher. I really appreciate all the time and efforts you spend on guiding me through my PhD. I could learn a lot from all our discussions. Furthermore, I am grateful to Luciana Dalla Valle for a great cooperation. I enjoyed our joint work on copula state space models. I also would like to thank Prof. Sylvia Frühwirth–Schnatter and Prof. Anastasios Panagiotelis for acting as referees for my thesis. Further, I appreciate all the good times I had with my (former) colleagues Dani, Dominik, Matthias, Nicole and Thomas, be it at the office or at the beer garden. Most importantly, I would like to express my gratitude to my family Sarah, Brigitte and Paul and to Estela for all their love and encouragement. You are such a great support in my life.

Contents

1 Introduction

1.1 Motivation

In today's world, more and more data is becoming available. Many data sets are collected over time and can therefore be analyzed within the time series framework, such as financial returns data or air pollution measurements. The availability of this data together with increasing computing power offers great potentials for more accurate data-driven decision making. Therefore, it is necessary to correctly understand and analyze data sets. Since the data sets usually consist of not only one but several variables, it is necessary to understand the dependency among these variables in addition to the individual behavior. Misspecification of the dependence structure can have severe effects. For example, in finance, the risk associated with a portfolio is highly influenced by the dependence structure among the assets that make up the portfolio (Embrechts et al. (2002)).

There are multivariate distributions, such as the multivariate normal distribution, which are just not flexible enough to describe complex dependence structures. A multivariate normal distribution can neither accomodate symmetric nor asymmetric tail dependence. Sklar (1959) provides a very useful framework to analyze dependencies. We can separate the individual behavior, described by the marginal distributions, from the joint behavior, characterized by the copula. The copula, a multivariate distribution function with uniform on $[0, 1]$ distributed margins, contains all information about the dependence structure. There exist different classes of copulas, such as elliptical and Archimedean copulas, with different properties and limitations. For example, all bivariate margins of exchangeable Archimedean copulas have the same distribution. Vine copulas provide a very rich and flexible model class. They are constructed from bivariate building blocks, which can be chosen from potentially different bivariate copula families. Due to their flexibility, vine copulas have been applied in many different areas, including finance (Brechmann and Czado (2013), Aas (2016)), medicine (Killiches and Czado (2018), Barthel et al. (2018)) and environmental sciences (Erhardt et al. (2015), Möller et al. (2018)).

The main motivation of this thesis is to develop novel Bayesian time series methodology, utilizing flexible bivariate copula specifications as building blocks. In particular, we investigate a generalization of single factor stochastic volatility models using the factor copula model of Krupskii and Joe (2013). Further, we study bivariate copulas and vine copulas with dynamic parameters, and lastly, we extend Gaussian linear state space models using copula specifications in the observation and in the state equation. To make these models applicable, efficient estimation procedures are required. For this purpose, we either use existing or develop our own Markov Chain Monte Carlo methods. Further, we illustrate the novel methodology with real data and show that it is capable to describe

complex multivariate time series more appropriate than conventional approaches.

1.2 Outline of the thesis

The thesis is based on five research papers.

- Chapter 3: Kreuzer A. and Czado C. (2019a). Bayesian inference for a single factor copula stochastic volatility model using Hamiltonian Monte Carlo. *Submitted to Econometrics and Statistics.* (arXiv:1808.08624v2)
- Chapter 4: Kreuzer A. and Czado C. (2019c). Efficient Bayesian inference for nonlinear state space models with univariate autoregressive state equation. *Under revision at the Journal of Computational and Graphical Statistics.* (arXiv:1902.10412v3)
- Chapter 5: Kreuzer A. and Czado C. (2019b). Bayesian inference for dynamic vine copulas in higher dimensions. *Submitted to Econometrics and Statistics.* (arXiv:1911.00702v1)
- Chapter 6: Kreuzer A., Dalla Valle L. and Czado C. (2019a). A Bayesian Non-linear State Space Copula Model to Predict Air Pollution in Beijing. *Under revision at the Annals of Applied Statistics.* (arXiv:1903.08421v2)
- Chapter 7: Kreuzer A., Dalla Valle L. and Czado C. (2019b). Bayesian Multivariate Nonlinear State Space Copula Models. *Submitted to the Journal of the American Statistical Association.* (arXiv:1911.00448v1)

After the introduction we review basic concepts which are needed throughout the thesis, such as state space models, copulas and Markov Chain Monte Carlo (MCMC) methods. Each of the Chapters 3 to 7 is associated with one of the research papers. All papers include novel approaches to model time series data. The presented approaches belong to the general class of state space models, and can be characterized by an observation and a state equation, as shown in Table 1.1. Chapter 8 provides ideas for future research and concludes the thesis.

Chapter	domain	observation equation	dynamic	state equation
3	$\mathbb{R}^d$	factor copula+ normal margins	volatility	Gauss-AR(1)
4	$[0,1]^2$	bivariate copula	copula parameter	Gauss-AR(1)
5	$[0,1]^d$	vine copula	copula parameter	Gauss-AR(1)
6	$[0,1]$	conditional bivariate copula	latent factor	D-vine (1-truncated)
7	$[0,1]^d$	factor copula	latent factor	D-vine (1-truncated)

Table 1.1: Overview of the different models proposed in the thesis. Each chapter is associated with one model. The models are defined on different domains: $\mathbb{R}^d$, $[0,1]$, $[0,1]^2$ or $[0,1]^d$, with an arbitrary $d \in \{2,3,4,\ldots\}$. Common to all models is that they can be analyzed within the state space framework and are therefore characterized by an observation and a state equation. In the state equation we either use a Gaussian autoregressive process of order 1 (Gauss-AR(1)) or a D-vine truncated after the first tree (D-vine (1-truncated)). The column "dynamic" indicates which parameters are modeled dynamically through the state equation.

In the following we give a more detailed outline for Chapters 3 to 7. In Chapter 3 we propose a single factor copula based stochastic volatility model for multivariate financial time series. The model combines stochastic volatility models for the margins with a single factor copula for the dependence. Factor copula models, as proposed by Krupskii and Joe (2013), provide more flexibility than Gaussian factor models by allowing for different bivariate copulas that link observed variables to the latent factor. Instead of relying on a two-step estimation approach, we provide full Bayesian inference and estimate all model parameters jointly using Hamiltonian Monte Carlo within Gibbs sampling. This approach also includes automatic pair copula family selection. Further, it is demonstrated that the proposed approach results in more accurate value at risk forecasts than a two-step procedure and other relevant benchmark models.

While in Chapter 3 the dependence parameters are assumed to be constant over time, this assumption might not always be appropriate. To allow for time-varying dependence, Almeida and Czado (2012) propose a dynamic bivariate copula model, which is a state space model with univariate autoregressive state equation. Two problems are considered in Chapter 4: First, we deal with the estimation of nonlinear state space models with univariate autoregressive state equation. This is an important model class containing established models, such as stochastic volatility models or the above mentioned dynamic bivariate copula model. We develop a MCMC sampler, which relies on elliptical slice sampling, adaptive Metropolis-Hastings updates and on an ancillarity-sufficiency interweaving strategy. The efficiency of the proposed sampler is illustrated through simulated data. Second, we deal with modeling time-varying asymmetric dependence structures. Therefore, we propose a dynamic mixture copula model, which can be estimated with our proposed approach for nonlinear state space models with univariate autoregressive state equation. The dynamic mixture copula is used to model the volatility return relationship, the relationship between an index and the corresponding volatility index. We demonstrate that our model yields more accurate one-day ahead predictions than conventional approaches, such as dynamic or constant Student t copula models or a bivariate DCC-GARCH model.

In Chapter 5 the vine copula framework is utilized to scale the dynamic bivariate copula model of Almeida and Czado (2012), which has already been studied in Chapter 4, to arbitrary dimensions, yielding the dynamic vine copula model. The bivariate dynamic copula model was already extended to dynamic D-vine and dynamic C-vine copula models by Almeida et al. (2016) and Goel and Mehra (2019), respectively. Our contributions are the development of a first Bayesian estimation approach and the extension to general vine structures. Our estimation approach is based on an approximation of the posterior distribution and is motivated by the algorithm of Dissmann et al. (2013). The model is employed to study the time-varying dependence among 21 exchange rates. For comparison we also estimate a constant vine copula, a dynamic D-vine and a dynamic C-vine copula model. This comparison shows superior performance of the proposed dynamic vine copula model with respect to one-day ahead predictive accuracy.

While in Chapters 3, 4 and 5 we discuss state space models, where the state equations are described by Gaussian autoregressive processes of order 1, we introduce in Chapter 6 novel state space methodology based on copulas. We propose a univariate state space model, where we specify the observation and the state equation with copulas. This allows for flexible modeling and generalizes linear Gaussian state space models. Our application

to air pollution measurements shows that the extension is necessary, since the linear Gaussian state space model is not able to describe the time-dynamics of the air pollution data appropriately. Further, our approach is illustrated with predictions of future air pollution levels under different climate conditions.

Chapter 7 deals with the extension of the univariate copula state space model, introduced in Chapter 6, to a copula state space model with multivariate observations. To capture the dependence among different observed variables, we propose to use a single factor copula in the observation equation. The approach is illustrated with an application to air pollution measurements. It is shown that the structure with a single factor is appropriate for the analyzed data set and that the proposed approach yields more accurate predictions than a Gaussian state space model and than Bayesian additive regression trees.

2 Preliminaries

In this chapter, basic concepts which are used in later chapters of the thesis are introduced: State space models (Section 2.1), copulas (Section 2.2) and Markov Chain Monte Carlo methods (Section 2.3).

2.1 State space models

State space models play a key role in this thesis, since all models we propose can be analyzed within the state space framework. We also discuss a specific class of state space models, namely stochastic volatility models, which are employed in Chapters 3, 4 and 5.

2.1.1 General state space model formulation

State space models provide a very general and powerful framework to analyze dynamical systems that vary over time and have been applied in many different fields, including ecology (Patterson et al. (2008)), neuroscience (Paninski et al. (2010)) and economics (Harvey et al. (1994), Kim et al. (1998), Koop et al. (2010)). Within the general state space framework, a model contains an observation and a state equation. The observation equation describes how the observations relate to latent states through a conditional density of the observation vector at time t, $\boldsymbol{Y_t} \in \mathbb{R}^d$, given the latent states at time t, $\boldsymbol{s_t} \in \mathbb{R}^p$, while the state equation describes the evolution of the latent states over time. A general *state space model* (Durbin and Koopman (2012), Chapter 9) is given by

$$\boldsymbol{Y_t}|\boldsymbol{s_t} \sim f(\boldsymbol{y_t}|\boldsymbol{s_t}) \tag{2.1}$$

$$\boldsymbol{s_t}|\boldsymbol{s_{t-1}} \sim \pi(\boldsymbol{s_t}|\boldsymbol{s_{t-1}}) \tag{2.2}$$

with initial condition $\boldsymbol{s_0} \sim \pi(\boldsymbol{s_0})$, for $t = 1, \dots, T$. Here (2.1) is called *observation equation* and (2.2) is the *state equation*. Similar to Durbin and Koopman (2012) (Chapter 9), we assume throughout the thesis that Equations (2.1) and (2.2) allow the following density factorization

$$\begin{aligned} f(\boldsymbol{y_1}, \dots, \boldsymbol{y_T}|\boldsymbol{s_1}, \dots \boldsymbol{s_T}) &= \prod_{t=1}^{T} f(\boldsymbol{y_t}|\boldsymbol{s_t}) \\ \pi(\boldsymbol{s_0}, \dots \boldsymbol{s_T}) &= \pi(\boldsymbol{s_0}) \prod_{t=1}^{T} \pi(\boldsymbol{s_t}|\boldsymbol{s_{t-1}}). \end{aligned} \tag{2.3}$$

A special class of state space models are linear Gaussian state space models (Durbin and Koopman (2012), Chapter 3), where the observation and state equations are specified by multivariate normal distributions. In the following we denote by $N_k(\boldsymbol{\mu}, \Sigma)$ a k-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. A *linear Gaussian state space model* may be formulated as

$$\begin{aligned} \boldsymbol{Y_t} &= M_t^{obs} \boldsymbol{s_t} + \boldsymbol{\epsilon_t} \\ \boldsymbol{s_t} &= M_t^{lat} \boldsymbol{s_{t-1}} + \boldsymbol{\eta_t}, \boldsymbol{s_0} \sim N_p(\mathbf{0}, \Sigma_0^{lat}) \end{aligned}$$

for $t = 1, \ldots, T$. Here, the error terms satisfy $\boldsymbol{\epsilon_t} \sim N_d(\mathbf{0}, \Sigma_t^{obs})$ independently, $\boldsymbol{\eta_t} \sim N_p(\mathbf{0}, \Sigma_t^{lat})$ independently, and $\boldsymbol{\epsilon_t}$ is independent of $\boldsymbol{\eta_{t'}}$ for $t, t' \in \{1, \ldots, T\}$. Further, $M_t^{obs} \in \mathbb{R}^{d \times p}$, $M_t^{lat} \in \mathbb{R}^{p \times p}$ and Σ_t^{obs} and Σ_t^{lat} are $d \times d$ and $p \times p$ covariance matrices.

2.1.2 Univariate stochastic volatility models

Univariate stochastic volatility (SV) models (Kim et al. (1998)) are popular for modeling financial time series. They are state space models, where the state equation describes the evolution of the log variance over time, i.e. the model allows for time-varying volatility, which is characteristic for financial data. In the following we introduce the SV model with Gaussian errors based on Kreuzer and Czado (2019a) and the SV model with skew Student t errors based on Kreuzer and Czado (2019c).

The stochastic volatility model with Gaussian errors

In the SV model (Kim et al. (1998)), the log variances $(s_1, \ldots, s_T)$ of a conditionally normally distributed vector $(Z_1, \ldots, Z_T)$ are modeled with a latent autoregressive process of order 1 (AR(1) process). This AR(1) process has mean parameter $\mu \in \mathbb{R}$, persistence parameter $\phi \in (-1, 1)$ and standard deviation parameter $\sigma \in (0, \infty)$. More precisely, the *stochastic volatility (SV) model* is given by

$$\begin{aligned} Z_t &= \exp\left(\frac{s_t}{2}\right)\epsilon_t, \\ s_t &= \mu + \phi(s_{t-1} - \mu) + \sigma\eta_t, \end{aligned} \tag{2.4}$$

where $s_0 | \mu, \phi, \sigma \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$ and $\epsilon_t, \eta_t, \sim N(0, 1)$ independently, for $t = 1, \ldots, T$.

Kastner and Frühwirth-Schnatter (2014) develop a MCMC approach for this model, which uses the ancillarity-sufficiency interweaving strategy proposed by Yu and Meng (2011). This strategy leads to an efficient MCMC procedure which is implemented in the `R` package `stochvol` (Kastner (2016)). We discuss the prior densities proposed by Kastner (2016), since we also utilize them later. The following priors for μ, ϕ and σ are chosen

$$\mu \sim N(0, 100^2), \quad \frac{\phi + 1}{2} \sim Beta(5, 1.5), \quad \sigma^2 \sim \chi_1^2. \tag{2.5}$$

The prior for μ is rather uninformative, whereas the prior for ϕ puts more mass on higher values for the persistence parameter. High persistence parameters are characteristic for financial time series. The prior choice for σ^2 differs from the popular inverse Gamma

prior. The χ^2_1 prior is equivalent to a standard normal prior on $\pm\sqrt{\sigma^2}$, which was suggested by Frühwirth-Schnatter and Wagner (2010). Frühwirth-Schnatter and Wagner (2010) argue that this prior is less influential than an inverse Gamma prior for true values of σ^2 close to zero. In contrast to the inverse Gamma prior, the χ^2_1 prior has more mass close to zero.

The chosen prior density of $(\mu, \phi, \sigma, s_0, \ldots, s_T)$ is given by

$$\begin{aligned}\pi(\mu, \phi, \sigma, s_0, \ldots, s_T) &= \pi(s_0, \ldots, s_T|\mu, \phi, \sigma)\pi(\mu, \phi, \sigma) \\ &= \varphi\left(s_0\middle|\mu, \frac{\sigma^2}{1-\phi^2}\right)\prod_{t=1}^{T}\varphi\left(s_t|\mu+\phi(s_{t-1}-\mu), \sigma^2\right)\pi(\mu)\pi(\phi)\pi(\sigma),\end{aligned} \tag{2.6}$$

where $\varphi\left(\cdot|\mu_{normal}, \sigma^2_{normal}\right)$ denotes the univariate normal density with mean μ_{normal} and variance σ^2_{normal} and $\pi(\cdot)$ denotes the corresponding prior density as specified in (2.5).

The stochastic volatility model with skew Student t errors

To introduce the SV model with skew Student t errors, we first discuss the skew Student t distribution. The density of the *univariate skew Student t distribution* (Azzalini and Capitanio (2003), Frühwirth-Schnatter and Pyne (2010)) with parameters $\xi \in \mathbb{R}, \omega \in (0, \infty), \alpha \in \mathbb{R}$ and $df \in (0, \infty)$ is given by

$$st(x|\xi, \omega, \alpha, df) = \frac{2}{\omega}t\left(\frac{x-\xi}{\omega}\middle|df\right)T\left(\alpha\frac{x-\xi}{\omega}\sqrt{\frac{df+1}{\left(\frac{x-\xi}{\omega}\right)^2+df}}\middle|df+1\right),$$

where $t(\cdot|df)$ is the density function of the univariate Student t distribution with df degrees of freedom and $T(\cdot|df)$ denotes the corresponding distribution function. The expectation and variance of a random variable X following a skew Student t distribution with parameters ξ, ω, α as above and $df > 2$ are given by

$$E(X) = \xi + \omega b_{df}\delta, \text{ and } Var(X) = \omega^2\left(\frac{df}{df-2} - b^2_{df}\delta^2\right),$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$ and $b_{df} = \sqrt{\frac{df}{\pi}}\frac{\Gamma(\frac{df-1}{2})}{\Gamma(\frac{df}{2})}$ (Arellano-Valle and Azzalini (2013)). If we set

$$\omega = \sqrt{\frac{1}{\left(\frac{df}{df-2} - b^2_{df}\delta^2\right)}} \text{ and } \xi = -\omega b_{df}\delta = -\sqrt{\frac{1}{\left(\frac{df}{df-2} - b^2_{df}\delta^2\right)}}b_{df}\delta,$$

only the parameters α and df remain unknown and the random variable has zero mean and a variance of one. We refer to the corresponding distribution as the *standardized skew Student t distribution*. Its density is denoted by sst and is obtained as

$$sst(x|\alpha, df) = st\left(x\middle| -\sqrt{\frac{1}{\left(\frac{df}{df-2} - b^2_{df}\delta^2\right)}}b_{df}\delta, \sqrt{\frac{1}{\left(\frac{df}{df-2} - b^2_{df}\delta^2\right)}}, \alpha, df\right). \tag{2.7}$$

The density function $sst(\cdot|\alpha, df)$ is visualized in Figure 2.1 for different values of α and df.

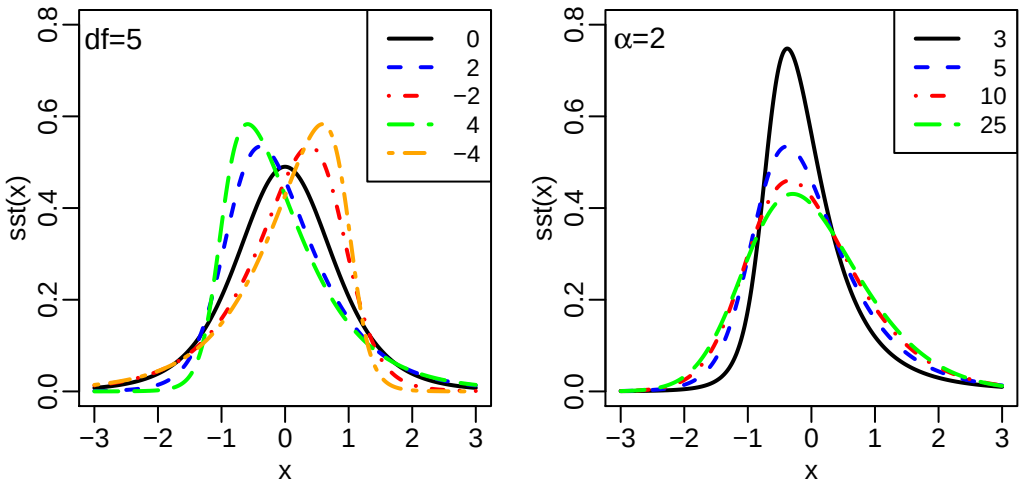

Figure 2.1: Visualization of the standardized skew Student t density function. The left plot shows density curves with $df = 5$ and different values for α as specified in the legend. The right plots shows density curves with $\alpha = 2$ and different values for df as specified in the legend.

To allow for heavy tails and skewness in the SV model, the normal distribution in the observation equation of the SV model in (2.4) is replaced by a standardized skew Student t distribution. This yields the *stochastic volaility (SV) model with skew Student t errors* as considered by Abanto-Valle et al. (2015), which is given by

$$\begin{aligned} Y_t &= \exp\left(\frac{s_t}{2}\right)\epsilon_t \\ s_t &= \mu + \phi(s_{t-1} - \mu) + \sigma\eta_t, \end{aligned} \tag{2.8}$$

where $\epsilon_t|\alpha, df \sim sst(\epsilon_t|\alpha, df)$ independently for $t = 1, \ldots, T$. Further, μ, ϕ, σ, s_0 and η_t are chosen as in (2.4). To complete a Bayesian model specification, we equip the parameters with prior distributions. For μ, ϕ and σ we use the same priors as given in (2.6) for the SV model with Gaussian errors. For the additional parameters, α and df, we choose the following prior distributions

$$\alpha \sim N(0, 100), \qquad df \sim N_{>2}(5, 25), \tag{2.9}$$

where $N_{>2}$ denotes the normal distribution truncated to $(2, \infty)$. We need to ensure that $df > 2$, since the standardized skew Student t distribution would not be well defined otherwise.

2.2 Copulas

Copulas provide a useful tool to describe dependencies. A d-dimensional *copula* $\mathbb{C} : [0, 1]^d \to [0, 1]$ is a multivariate distribution function, where the corresponding marginals

are uniformly distributed on $[0,1]$. Thus, different copulas have the same margins and therefore copulas can only differ in how the different margins interact with each other. In this section we follow Czado (2019), unless stated otherwise.

2.2.1 Sklar's theorem

Sklar (1959) showed that every multivariate distribution can be separated into its marginal distributions and a copula. More precisely, for a d-dimensional random vector $(Y_1, \dots, Y_d)$ with joint distribution function F and marginal distribution functions $F_1, \dots, F_d$, it holds that

$$F(y_1, \dots, y_d) = \mathbb{C}(F_1(y_1), \dots, F_d(y_d)), \tag{2.10}$$

where $\mathbb{C}$ is a copula. The copula $\mathbb{C}$ is unique for absolutely continuous distributions.

This representation allows for a very flexible modeling approach. Each marginal can be modeled separately in a first step. In the second step the marginals are glued together with the copula $\mathbb{C}$. Estimation of such a model can follow a similar procedure. In a first step we obtain estimates for the marginal distribution functions denoted by $\hat{F}_1, \dots, \hat{F}_d$. Then the marginals of the vector $(\hat{F}_1(Y_1), \dots, \hat{F}_d(Y_d))$ are approximately uniformly distributed on $[0,1]$ and the copula $\mathbb{C}$ can be estimated as the joint distribution function of $(\hat{F}_1(Y_1), \dots, \hat{F}_d(Y_d))$. This two-step procedure is also called inference for margins and was proposed by Joe and Xu (1996). For data $\mathtt{Y} = (y_{tj})_{t=1,\dots,T, j=1,\dots,d} \in \mathbb{R}^{T \times d}$ containing observations of $(Y_1, \dots, Y_d)$, we further call $(F_j(y_{tj}))_{t=1,\dots,T, j=1,\dots,d}$ *copula data* and $(\hat{F}_j(y_{tj}))_{t=1,\dots,T, j=1,\dots,d}$ *pseudo copula data*.

2.2.2 Dependence measures

Copula models are able to describe complex dependence structures which go beyond linear correlation. Since Pearson's correlation coefficient can only measure linear relationships and depends on the marginal distribution, it might not be a good choice in the copula world. A frequently used global dependence measure in copula modeling is Kendall's τ (Kendall (1938)). Kendall's τ is a rank based dependence measure, which depends only on the copula and not on the margins of a distribution.

For a continuous bivariate random vector (Y_1, Y_2) we denote by (Y_{11}, Y_{12}) and (Y_{21}, Y_{22}) two independent and identically distributed copies of (Y_1, Y_2). Then *Kendall's* τ between Y_1 and Y_2 is defined as

$$\tau(Y_1, Y_2) \coloneqq P((Y_{11} - Y_{21})(Y_{12} - Y_{22}) > 0) - P((Y_{11} - Y_{21})(Y_{12} - Y_{22}) < 0). \tag{2.11}$$

To estimate Kendall's τ from n observations of the bivariate random vector (Y_1, Y_2), denoted by $(\boldsymbol{y_i})_{i=1,\dots,n}$ with $\boldsymbol{y_i} = (y_{i1}, y_{i2})$, we first define concordant and discordant pairs. For $i, j \in \{1, \dots, n\}$, we call the pair $(\boldsymbol{y_i}, \boldsymbol{y_j})$

- *concordant* if $y_{i1} < y_{j1}$ and $y_{i2} < y_{j2}$ or if $y_{i1} > y_{j1}$ and $y_{i2} > y_{j2}$ holds,
- *discordant* if $y_{i1} < y_{j1}$ and $y_{i2} > y_{j2}$ or if $y_{i1} > y_{j1}$ and $y_{i2} < y_{j2}$ holds,
- *extra* y_1 *pair* if $y_{i1} = y_{j1}$,
- *extra* y_2 *pair* if $y_{i2} = y_{j2}$.

Note that the above mentioned condition for concordance is equivalent to $(y_{i1} - y_{j1})(y_{i2} - y_{j2}) > 0$ and the condition for discordance is equivalent to $(y_{i1} - y_{j1})(y_{i2} - y_{j2}) < 0$. We consider all $\binom{n}{2}$ unordered pairs $(\boldsymbol{y_i}, \boldsymbol{y_j})$ with $i \neq j$ and denote by N_c the number of concordant pairs, by N_d the number of discordant pairs, by N_1 the number of extra y_1 pairs and by N_2 the number of extra y_2 pairs. Then Kendall's τ of (Y_1, Y_2) can be estimated by the *empirical Kendall's* τ

$$\hat{\tau} := \frac{N_c - N_d}{\sqrt{N_c + N_d + N_1}\sqrt{N_c + N_d + N_2}}.$$

In addition to the overall dependency, as measured with a global dependence measure such as Kendall's τ, one might be interested in more specific aspects of the dependence structure. Tail dependence coefficients allow to measure the dependency of extreme events in the tails. The *upper and lower tail dependence coefficients*, λ^U and λ^L, of a bivariate distribution with copula $\mathbb{C}$ are defined as

$$\lambda^U := \lim_{v \to 1^-} \frac{1 - 2v + \mathbb{C}(v, v)}{1 - v} \quad \text{and} \quad \lambda^L := \lim_{v \to 0^+} \frac{\mathbb{C}(v, v)}{v}.$$

2.2.3 Parametric copula families

In this section we introduce several parametric copula families, which are frequently used throughout the thesis. A more detailed overview of different parametric copula families and their properties is provided by Joe (2014) (Chapter 4) and by Czado (2019) (Chapter 3). Restricted to two dimensions, these families form the building blocks for vine copula models, which we introduce in the next section. We denote by $(u_1, \dots, u_d)$ a vector in the d-dimensional hypercube, i.e. $(u_1, \dots, u_d) \in [0, 1]^d$. If the marginals are independent, the corresponding copula is the independence copula, which is our first example.

Independence copula: The d-dimensional *independence copula* $\mathbb{C}$ is given by

$$\mathbb{C}(u_1, \dots, u_d) = \prod_{i=1}^{d} u_i.$$

The inversion of Sklar's theorem (see Section 2.2.1) provides a construction method for copulas. For a joint distribution function F with invertible marginal distribution functions $F_1, \dots, F_d$, the corresponding copula $\mathbb{C}$ is obtained as

$$\mathbb{C}(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)).$$

Applying this method to the multivariate normal and multivariate Student t distribution yields the Gaussian and the Student t copula, respectively. These two copulas are derived from elliptical distributions and belong to the class of elliptical copulas.

Gaussian copula: We denote by Φ the univariate standard normal distribution function and by Φ_R the distribution function of the d-dimensional multivariate normal distribution with zero means, unit marginal variances and correlation matrix R. Then the d-dimensional *Gaussian copula* with parameter R is given by

$$\mathbb{C}(u_1, \dots, u_d) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)).$$

Student t copula: Let T_ν be the distribution function of the univariate Student t distribution with $\nu > 0$ degrees of freedom. Further, we denote by $T_{R,\nu}$ the distribution function of the multivariate central Student t distribution with ν degrees of freedom and scale parameter matrix equal to a correlation matrix R. The d-dimensional *Student t copula* with parameters ν and R is given by

$$\mathbb{C}(u_1, \dots, u_d) = T_{R,\nu}(T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_d)).$$

Another class of copulas is provided by Archimedean copulas. These copulas are constructed from generator functions. These generator functions are not discussed here, since they are not used in the thesis. For more details about Archimedean copulas and their generator functions we refer to Nelsen (2007) (Chapter 4). We now introduce three Archimedean copulas: The bivariate Clayton, Gumbel and Frank copula.

Clayton copula: For $\theta > 0$, the *Clayton copula* is given by

$$\mathbb{C}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}. \tag{2.12}$$

The independence copula is obtained for $\theta \to 0$.

Gumbel copula: The *Gumbel copula* with parameter $\theta \geq 1$ is given by

$$\mathbb{C}(u_1, u_2) = \exp\Big(-[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta]^{\frac{1}{\theta}}\Big). \tag{2.13}$$

Frank copula: For $\theta \in (-\infty, \infty) \setminus \{0\}$,

$$\mathbb{C}(u_1, u_2) = -\tfrac{1}{\theta} \ln\Big(\tfrac{1}{1-\exp(-\theta)}[(1 - \exp(-\theta)) - (1 - \exp(-\theta u_1))(1 - \exp(-\theta u_2))]\Big)$$

is the *Frank copula* with parameter θ.

For the above mentioned bivariate copulas, there is a one-to-one correspondence between the copula parameter and the associated Kendall's τ. These relationships are summarized in Table 2.1.

Family	Kendall's τ
Gaussian	$2/\pi \arcsin(\rho)$
Student t	$2/\pi \arcsin(\rho)$
Frank	$1 + 4(D(\theta) - 1)/\theta$
Clayton	$\theta/(\theta + 2)$
Gumbel	$1 - 1/\theta$

Table 2.1: Copula families and the corresponding Kendall's τ as a function of the copula parameter. Here we consider the bivariate Gaussian and Student t copula. In the bivariate case the matrix R of these copulas can be parametrized with one scalar parameter, the correlation parameter ρ. Further, $D(\theta)$ is given by $D(\theta) = \int_0^\theta \frac{x/\theta}{\exp(x)-1} dx$.

Throughout the thesis, we will parametrize the copulas in Table 2.1 in terms of Kendall's τ, instead of the copula parameter ρ or θ. While the copula parameters ρ or θ might have different interpretations for different copula families, the Kendall's τ

parameters can easily be compared among different families. To express that a copula $\mathbb{C}$ depends on the Kendall's τ parameter τ, we write $\mathbb{C}(u_1, u_2; \tau)$ and the corresponding density is denoted by $c(u_1, u_2; \tau)$.

The Gumbel and the Clayton copula cannot handle negative Kendall's τ values. To allow for more flexibility, copula densities can be rotated. For example, we can handle negative dependencies by rotating the density of a Gumbel copula by 90°. For a bivariate copula $\mathbb{C}$ with copula density c, we consider 90°, 180° and 270° counterclockwise rotations. The corresponding densities are denoted by c_{90}, c_{180} and c_{270} and are obtained as

- $c_{90}(u_1, u_2) := c(1 - u_2, u_1)$
- $c_{180}(u_1, u_2) := c(1 - u_1, 1 - u_2)$
- $c_{270}(u_1, u_2) := c(u_2, 1 - u_1)$ for all $u_1, u_2 \in [0, 1]$.

The copula with density c_{180} obtained from a 180° rotation is the corresponding *survival copula*. Further, we employ rotations to define the *extended Gumbel (eGumbel) copula*, with density

$$c(u_1, u_2; \tau) = \begin{cases} c_{Gumbel}(u_1, u_2; \tau) \text{ if } \tau \geq 0 \\ c_{Gumbel}(1 - u_2, u_1; -\tau) \text{ if } \tau < 0, \end{cases}$$

where $c_{Gumbel}(\cdot, \cdot; \tau)$ is the density of the bivariate Gumbel copula parametrized in terms of Kendall's τ. The *extended Clayton (eClayton) copula* is obtained similarly.

2.2.4 Vine copulas

This section is based on Kreuzer and Czado (2019b) and Czado (2019). Vine copulas (Bedford and Cooke (2001), Aas et al. (2009)) are a popular class in dependence modeling. They allow for great flexibility by constructing a density of arbitrary dimension from two-dimensional densities.

For this construction vine copulas are represented as graphical models. A *regular vine (R-vine) tree sequence* on d elements is a sequence of trees $\mathcal{V} = (T_1, \ldots, T_{d-1})$ satisfying the following conditions

- T_1 is a tree with nodes $N_1 = \{1, \ldots d\}$ and edge set E_1,
- for $j = 2, \ldots, d-1$, it holds that T_j is a tree with nodes $N_j = E_{j-1}$ and edge set E_j,
- for $j = 2, \ldots, d-1$ and $\{a, b\} \in E_j$, it holds that the edges corresponding to a and b in tree T_{j-1} share a common node (*proximity condition*).

The *complete union* of an edge $e \in E_i$ is defined as

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

Further, the *conditioning set* of edge $e = \{a, b\}$ is obtained as

$$\boldsymbol{D_e} := \boldsymbol{A_a} \cap \boldsymbol{A_b}$$

and the *conditioned sets* are defined as

$$a_e := \boldsymbol{A_a} \setminus \boldsymbol{D_e}, \quad b_e := \boldsymbol{A_b} \setminus \boldsymbol{D_e}.$$

Note that for $e = \{a, b\} \in E_1$, we set $a_e = \{a\}, b_e = \{b\}$ and $\boldsymbol{D_e}$ is the empty set.

So for each edge e, we have an associated conditioning set $\boldsymbol{D_e}$ and two associated conditioned sets a_e and b_e. By construction, a_e and b_e have a single element, respectively. For a random vector $(U_1, \dots, U_d)$ with uniform $[0, 1]$ margins, bivariate copulas of conditional distributions can be identified with the conditioning and conditioned sets. For an edge e, we denote by $\mathbb{C}_{a_e,b_e;\boldsymbol{D_e}}$ the copula associated with the conditional distribution $(U_{a_e}, U_{b_e})|\boldsymbol{U_{D_e}} = \boldsymbol{u_{D_e}}$, where $\boldsymbol{u_F} = (u_i)_{i\in\boldsymbol{F}}$ for a set $\boldsymbol{F}$. The corresponding copula density is denoted by $c_{a_e,b_e;\boldsymbol{D_e}}$. Many researchers assume that $c_{a_e,b_e;\boldsymbol{D_e}}$ does not depend on $\boldsymbol{u_{D_e}}$, which is called the simplifying assumption (Haff et al. (2010), Stoeber et al. (2013)). This assumption allows for sequential estimation and selection of vine copula models (Brechmann and Czado (2013), Dissmann et al. (2013)).

Based on these graphical definitions, Bedford and Cooke (2001) build a d-dimensional *(regular) vine copula model* with joint density

$$c(u_1, \dots u_d) = \prod_{i=1}^{d-1} \prod_{e\in E_i} c_{a_e,b_e;\boldsymbol{D_e}}(u_{a_e|\boldsymbol{D_e}}, u_{b_e|\boldsymbol{D_e}}). \tag{2.14}$$

It is a simplified vine copula model since $c_{a_e,b_e;\boldsymbol{D_e}}$ does not depend on the conditioning value $\boldsymbol{u_{D_e}}$. Here, $u_{a_e|\boldsymbol{D_e}}$ and $u_{b_e|\boldsymbol{D_e}}$ are called *pseudo (copula) data*. They are obtained as $u_{a_e|\boldsymbol{D_e}} = \mathbb{C}_{a_e|\boldsymbol{D_e}}(u_{a_e}|\boldsymbol{u_{D_e}})$ and $u_{b_e|\boldsymbol{D_e}} = \mathbb{C}_{b_e|\boldsymbol{D_e}}(u_{b_e}|\boldsymbol{u_{D_e}})$, where $\mathbb{C}_{a_e|\boldsymbol{D_e}}(\cdot|\boldsymbol{u_{D_e}})$ and $\mathbb{C}_{b_e|\boldsymbol{D_e}}(\cdot|\boldsymbol{u_{D_e}})$ are the conditional distribution functions of $U_{a_e}|\boldsymbol{U_{D_e}} = \boldsymbol{u_{D_e}}$ and $U_{b_e}|\boldsymbol{U_{D_e}} = \boldsymbol{u_{D_e}}$, respectively. In the first tree $\boldsymbol{D_e}$ is the empty set and the pseudo data of the first tree is just $u_1, \dots u_d$. For $a_e = \{i\}$ and $b_e = \{j\}$, we also write c_{ij} or $c_{i,j}$ instead of $c_{\{i\},\{j\}}$, and similarly, we may omit the brackets $\{\}$ for the set $\boldsymbol{D_e}$.

Note that the class of simplified vine copulas is broad, including multivariate Gaussian and Student t copulas (Joe (2014), Chapter 3).

To evaluate the conditional distribution functions and to obtain the corresponding pseudo data, the h *functions* (Aas et al. (2009)) for an edge e are defined as

$$\begin{aligned} h_{a_e|b_e;\boldsymbol{D_e}}(u_1|u_2) &:= \frac{d}{du_2}\mathbb{C}_{a_e,b_e;\boldsymbol{D_e}}(u_1, u_2) \\ h_{b_e|a_e;\boldsymbol{D_e}}(u_2|u_1) &:= \frac{d}{du_1}\mathbb{C}_{a_e,b_e;\boldsymbol{D_e}}(u_1, u_2). \end{aligned} \tag{2.15}$$

The h functions are conditional distribution functions. For example, for the copula $\mathbb{C}_{12}$, $h_{1|2}(u_1|u_2) = \mathbb{C}_{1|2}(u_1|u_2)$. If the copula $\mathbb{C}_{a_e,b_e;\boldsymbol{D_e}}$ depends on a set of parameters $\boldsymbol{\delta}$, we write $h_{a_e|b_e;\boldsymbol{D_e}}(u_1|u_2;\boldsymbol{\delta})$ and $h_{b_e|a_e;\boldsymbol{D_e}}(u_2|u_1;\boldsymbol{\delta})$. Based on pseudo data $u_{a_e|\boldsymbol{D_e}}$ and $u_{b_e|\boldsymbol{D_e}}$, we obtain pseudo data for the next tree as

$$\begin{aligned} u_{a_e|b_e\cup\boldsymbol{D_e}} &= h_{a_e|b_e;\boldsymbol{D_e}}(u_{a_e|\boldsymbol{D_e}}|u_{b_e|\boldsymbol{D_e}}), \\ u_{b_e|a_e\cup\boldsymbol{D_e}} &= h_{b_e|a_e;\boldsymbol{D_e}}(u_{b_e|\boldsymbol{D_e}}|u_{a_e|\boldsymbol{D_e}}) \end{aligned} \tag{2.16}$$

(Czado (2019), Chapter 4). So, the pseudo data can be calculated recursively and for the calculation of pseudo data at a certain tree level only bivariate copulas and pseudo data from lower trees are involved.

Figure 2.2 visualizes the first three trees of a six-dimensional vine copula. Assuming that there are only independence copulas in trees higher than Tree 3, the associated density is given by

$$\begin{aligned} c(u_1,\dots,u_6) =& c_{1,2}(u_1,u_2)\cdot c_{2,6}(u_2,u_6)\cdot c_{3,6}(u_3,u_6)\cdot c_{4,6}(u_4,u_6)\cdot c_{5,6}(u_5,u_6) \\ &\cdot c_{4,5;6}(u_{4|6},u_{5|6})\cdot c_{3,5;6}(u_{3|6},u_{5|6})\cdot c_{2,5;6}(u_{2|6},u_{5|6})\cdot c_{1,6;2}(u_{1|2},u_{6|2}) \\ &\cdot c_{3,4;5,6}(u_{3|5,6},u_{4|5,6})\cdot c_{2,4;5,6}(u_{2|5,6},u_{4|5,6})\cdot c_{1,5;2,6}(u_{1|2,6},u_{5|2,6}) \end{aligned} \tag{2.17}$$

The pseudo data can be determined as outlined in (2.16). For example, $u_{4|6} = h_{4|6}(u_4|u_6)$, $u_{5|6} = h_{5|6}(u_5|u_6)$ and $u_{4|56} = h_{4|5;6}(u_{4|6}|u_{5|6}) = h_{4|5;6}(h_{4|6}(u_4|u_6)|h_{5|6}(u_5|u_6))$.

Such vine copulas, where copulas above a certain tree level are set to the independence copula are called *truncated* (Brechmann et al. (2012)). With truncation we can achieve different levels of sparsity.

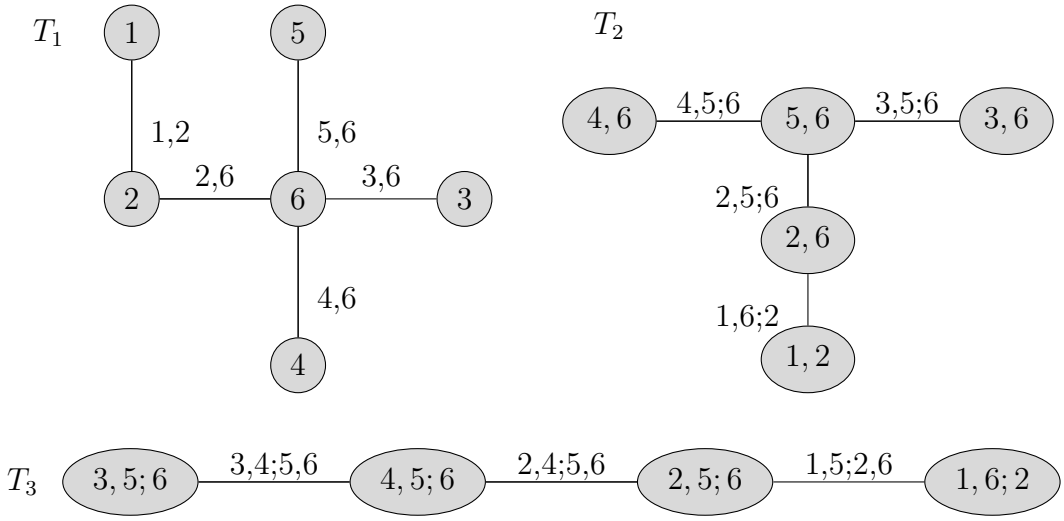

Figure 2.2: Tree structure of a vine copula model.

Drawable and canonical vines are two subclasses of vine copulas that are associated with specific tree structures. The density of a simplified *drawable vine (D-vine)* copula can be written as

$$c(u_1,\dots,u_d) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j;\{i+1,\dots,i+j-1\}}(u_{i|\{i+1,\dots,i+j-1\}}, u_{i+j|\{i+1,\dots,i+j-1\}}).$$

This results in structures where each tree is a line. This structure is often used for time series data (Smith et al. (2010)). The density of a simplified *canonical vine (C-vine)* copula can be expressed as

$$c(u_1,\dots,u_d) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+i;\{1,\dots,j-1\}}(u_{j|\{1,\dots,j-1\}}, u_{j+i|\{1,\dots,j-1\}})$$

resulting in structures where each tree forms a star, i.e. in each tree there is a central node to which all other nodes are connected.

2.3 Markov Chain Monte Carlo

Within this thesis, all models are analyzed within the Bayesian framework. For a thorough introduction to Bayesian data analysis we refer to Gelman et al. (2014a). A Bayesian model is fully specified by a likelihood function and a prior distribution of the corresponding parameters. We denote by $\pi(\boldsymbol{\theta})$ the prior density or prior probability mass function of a parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ and by $\ell(\boldsymbol{\theta}|\mathrm{Y})$ the likelihood function, where $\mathrm{Y} \in \mathbb{R}^{T \times d}$ is a data matrix. The goal is to infer properties of the posterior distribution, which is the distribution of the parameter vector $\boldsymbol{\theta}$ given the data Y. According to Bayes theorem, the *posterior density* or *posterior probability mass function* $f(\boldsymbol{\theta}|\mathrm{Y})$ is obtained as

$$f(\boldsymbol{\theta}|\mathrm{Y}) = \frac{\ell(\boldsymbol{\theta}|\mathrm{Y})\pi(\boldsymbol{\theta})}{f(\mathrm{Y})} \propto \ell(\boldsymbol{\theta}|\mathrm{Y})\pi(\boldsymbol{\theta}), \tag{2.18}$$

where the *marginal likelihood* $f(\mathrm{Y}) = \int_\Theta \ell(\boldsymbol{\theta}|\mathrm{Y})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ ($f(\mathrm{Y}) = \sum_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}|\mathrm{Y})\pi(\boldsymbol{\theta})$ for discrete parameters) can be considered as a normalizing constant. It is not straightforward to infer properties of the posterior density, which is analytically tractable only in few cases. Especially the marginal likelihood $f(\mathrm{Y})$, which involves an integral over the parameter space, might be very expensive to evaluate.

Very popular in this context are Markov Chain Monte Carlo (MCMC) methods. The idea is to construct a Markov Chain whose stationary distribution is the desired posterior distribution. From simulations of the Markov Chain, quantities of interest such as the posterior mean, mode or quantiles can easily be estimated. In the following, we briefly review two basic MCMC methods, Gibbs sampling and the Metropolis-Hastings algorithm. After that we discuss Hamiltonian Monte Carlo and elliptical slice sampling. For more details about Markov chains and MCMC methods we refer to Meyn and Tweedie (2012) and Robert and Casella (2013).

Gibbs sampling

In Gibbs sampling (Geman and Geman (1984)), we partition the parameter vector $\boldsymbol{\theta}$ as follows $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_n})$, where $\boldsymbol{\theta_i}$ can be a scalar or a vector and denote by $\boldsymbol{\theta_{-i}}$ the parameter vector $\boldsymbol{\theta}$ with the component $\boldsymbol{\theta_i}$ removed. The distribution with density or probability mass function $f(\boldsymbol{\theta_i}|\boldsymbol{\theta_{-i}}, \mathrm{Y})$ is called full conditional distribution and we assume that we can sample from it for $i = 1, \ldots, n$. The Gibbs sampler iteratively samples from the full conditional distributions and to obtain R draws from a Markov chain, with stationary distribution equal to the desired posterior distribution, we proceed as follows

- Set initial values $\boldsymbol{\theta^0}$.
- For $r = 1, \ldots, R$:
  Obtain $\boldsymbol{\theta_1^r}$ as a sample from $f(\boldsymbol{\theta_1}|\boldsymbol{\theta_2^{r-1}}, \ldots, \boldsymbol{\theta_n^{r-1}}, \mathrm{Y})$.
  Obtain $\boldsymbol{\theta_2^r}$ as a sample from $f(\boldsymbol{\theta_2}|\boldsymbol{\theta_1^r}, \boldsymbol{\theta_3^{r-1}}, \ldots, \boldsymbol{\theta_n^{r-1}}, \mathrm{Y})$.
  $\vdots$
  Obtain $\boldsymbol{\theta_n^r}$ as a sample from $f(\boldsymbol{\theta_n}|\boldsymbol{\theta_1^r}, \ldots, \boldsymbol{\theta_{n-1}^r}, \mathrm{Y})$.

Metropolis-Hastings

In the Metropolis-Hastings algorithm (Metropolis et al. (1953), Hastings (1970)), we generate proposals from a density or probability mass function that is easy to sample and then accept the proposal as a new value in the Markov chain with some probability. This probability is chosen such that the stationary distribution of the chain is still the desired posterior distribution. We denote the proposal density or proposal probability mass function by $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{\boldsymbol{r-1}})$. In the r-th iteration it may depend on the previous value of the chain $\boldsymbol{\theta}^{\boldsymbol{r-1}}$. To obtain R draws we proceed as follows

- Set initial values $\boldsymbol{\theta}^{\boldsymbol{0}}$.
- For $r = 1, \ldots, R$:
  Generate a proposal $\boldsymbol{\theta}'$ from $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{\boldsymbol{r-1}})$.

  Set $\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}'|\mathrm{Y})q(\boldsymbol{\theta}^{\boldsymbol{r-1}}|\boldsymbol{\theta}')}{f(\boldsymbol{\theta}^{\boldsymbol{r-1}}|\mathrm{Y})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{\boldsymbol{r-1}})}\right)$.

  With probability α we accept the proposal and set $\boldsymbol{\theta}^{\boldsymbol{r}} = \boldsymbol{\theta}'$, otherwise we set $\boldsymbol{\theta}^{\boldsymbol{r}} = \boldsymbol{\theta}^{\boldsymbol{r-1}}$.

2.3.1 Hamiltonian Monte Carlo

This section provides a short introduction to Hamiltonian Monte Carlo (HMC) based on Kreuzer and Czado (2019a), where we follow Neal et al. (2011). For more details about HMC we refer to Neal et al. (2011) and Betancourt (2017).

Hamiltonian dynamics describe the time evolution of a physical system through differential equations. In Hamiltonian Monte Carlo (HMC) the posterior density is connected to the energy function of a physical system. This makes it possible to propose states in the sampling process which are guided by appropriate differential equations. New states are chosen utilizing information of the gradient of the log posterior density, which can lead to more efficient sampling procedures. Therefore HMC has become popular. For example, Hartmann and Ehlers (2017) demonstrate how to estimate parameters of generalized extreme value distributions with HMC, while Pakman and Paninski (2014) use HMC to sample from truncated multivariate Gaussian distributions. Especially with the development of the probabilistic programming language STAN by Carpenter et al. (2017), its popularity is growing. STAN allows easy model specification and utilizes the No-U-Turn sampler of Hoffman and Gelman (2014), an extension of HMC. A brief sketch of the No-U-Turn sampler is provided in the end of this section. We start with the introduction of the Hamiltonian dynamics.

Hamiltonian dynamics

We consider a position vector $\boldsymbol{q} = \boldsymbol{q}(t) = (q_1(t), \ldots, q_d(t))^\top \in \mathbb{R}^d$ with associated momentum vector $\boldsymbol{p} = \boldsymbol{p}(t) = (p_1(t), \ldots, p_d(t))^\top \in \mathbb{R}^d$ at time t. Their change over time is described through the function $H(\boldsymbol{q}, \boldsymbol{p})$ ($H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$), the *Hamiltonian*, which

satisfies the following differential equations:

$$\begin{aligned} \frac{d}{dt}q_i(t) &= \frac{d}{dp_i}H(\boldsymbol{q}(t),\boldsymbol{p}(t)) \\ \frac{d}{dt}p_i(t) &= -\frac{d}{dq_i}H(\boldsymbol{q}(t),\boldsymbol{p}(t)), i = 1,\dots,d. \end{aligned} \tag{2.19}$$

Here $\frac{d}{dp_i}H(\boldsymbol{q}(t),\boldsymbol{p}(t))$ means that we obtain the derivative of H with respect to the component p_i and then evaluate this derivative at $(\boldsymbol{q}(t),\boldsymbol{p}(t))$. Further, H represents the total energy of the system. In HMC, it is usually assumed that H can be expressed as

$$H(\boldsymbol{q},\boldsymbol{p}) = U(\boldsymbol{q}) + K(\boldsymbol{p}) = U(\boldsymbol{q}) + \boldsymbol{p}^\top M^{-1}\boldsymbol{p}/2, \tag{2.20}$$

where $U(\boldsymbol{q})$ is called the potential energy and $K(\boldsymbol{p})$ the kinetic energy. Further, M is a symmetric positive definite mass matrix, which is usually assumed to be diagonal. The Hamiltonian dynamics, specified in (2.19), can therefore be rewritten as

$$\begin{aligned} \frac{d}{dt}q_i(t) &= (M^{-1}\boldsymbol{p}(t))_i \\ \frac{d}{dt}p_i(t) &= -\frac{d}{dq_i}U(\boldsymbol{q}(t)), i = 1,\dots,d. \end{aligned} \tag{2.21}$$

Leapfrog method

Since it is usually not possible to solve the system of differential equations given in (2.21) analytically, we need to find iterative approximations. Therefore we use the Leapfrog method. We assume that M is a diagonal matrix with diagonal entries $m_1,\dots,m_d$ and then the state one-step ahead of time t with step size ϵ, i.e. the state at time $t+\epsilon$, is approximated by

$$\begin{aligned} p_i(t+\epsilon/2) &= p_i(t) - \frac{\epsilon}{2}\frac{d}{dq_i}U(\boldsymbol{q}(t)) \\ q_i(t+\epsilon) &= q_i(t) + \epsilon\frac{p_i(t+\epsilon/2)}{m_i} \\ p_i(t+\epsilon) &= p_i(t+\epsilon/2) - \frac{\epsilon}{2}\frac{d}{dq_i}U(\boldsymbol{q}(t+\epsilon)), \text{ for } i = 1,\dots d. \end{aligned} \tag{2.22}$$

Canonical distribution

To use Hamiltonian dynamics within MCMC sampling, we need to relate the energy function to a probability distribution. Therefore we utilize the *canonical distribution* $P(\boldsymbol{x})$ associated with a general energy function $E(\boldsymbol{x})$ with state $\boldsymbol{x}$ defined through the density

$$p(\boldsymbol{x}) \coloneqq \frac{1}{Z}\exp(-E(\boldsymbol{x})/Temp),$$

where $Temp$ is the system's temperature and Z is the normalizing constant needed to satisfy the density constraint. So the Hamiltonian $H(\boldsymbol{q},\boldsymbol{p})$ specified in (2.20) defines a probability density given by

$$p(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{Z}\exp(-H(\boldsymbol{q},\boldsymbol{p})/Temp) = \frac{1}{Z}\exp(-U(\boldsymbol{q})/Temp)\exp(-K(\boldsymbol{p})/Temp).$$

Thus, $\boldsymbol{q}$ and $\boldsymbol{p}$ are independent. In the following we assume $Temp = 1$.

Sampling with Hamiltonian Monte Carlo

In HMC we specify the corresponding energy function of $\boldsymbol{q}$ and $\boldsymbol{p}$, i.e. the Hamiltonian, and sample from the corresponding canonical distribution of $\boldsymbol{q}$ and $\boldsymbol{p}$. In a Bayesian setup we identify $\boldsymbol{q}$ as our parameters of interest and $\boldsymbol{p}$ are auxiliary variables. Therefore we set

$$U(\boldsymbol{q}) \coloneqq -\ln(\ell(\boldsymbol{q}|\mathrm{Y})\pi(\boldsymbol{q})),$$

where $\pi(\boldsymbol{q})$ is the prior density and $\ell(\boldsymbol{q}|\mathrm{Y})$ the likelihood function for the given data Y. Therefore the canonical distribution of $\boldsymbol{q}$ corresponds to the posterior distribution of $\boldsymbol{q}$, when $Temp = 1$.

Since $K(\boldsymbol{p}) = \boldsymbol{p}^\top M^{-1}\boldsymbol{p}/2$, it holds that the auxiliary parameter vector $\boldsymbol{p}$ is multivariate normal distributed with zero mean vector and covariance matrix M. A MCMC update is then obtained as follows.

- Sample the auxiliary parameter $\boldsymbol{p}$ from the normal distribution with zero mean vector and covariance matrix M.
- Metropolis update: Start with the current state $(\boldsymbol{q}, \boldsymbol{p})$ and simulate L steps of Hamiltonian dynamics with step size ϵ using the Leapfrog method. Obtain $(\boldsymbol{q}', \boldsymbol{p}')$ and accept this proposal with Metropolis acceptance probability

$$\min(1, \exp(-H(\boldsymbol{q}', \boldsymbol{p}') + H(\boldsymbol{q}, \boldsymbol{p}))) = \min\left(1, \frac{\pi(\boldsymbol{q}')l(\boldsymbol{q}'|\mathrm{Y})\exp\left(\boldsymbol{p}^\top M^{-1}\boldsymbol{p}/2\right)}{\pi(\boldsymbol{q})l(\boldsymbol{q}|\mathrm{Y})\exp(\boldsymbol{p}'^\top M^{-1}\boldsymbol{p}'/2)}\right). \tag{2.23}$$

The No-U-Turn sampler

In conventional HMC, the stepsize ϵ and the number of Leapfrog steps L are tuning parameters that need to be specified by the user. The choice of these tuning parameters is usually not straightforward and optimal choices might even differ for different regions of the state space. The No-U-Turn sampler of Hoffman and Gelman (2014) is an extension of HMC that automatically and adaptively selects these tuning parameters.

Hoffman and Gelman (2014) first consider the choice of L for a fixed stepsize ϵ. Assume we have an initial position vector at time 0, $\boldsymbol{q}(0)$. When we start evolving the Hamiltonian differential equation with the Leapfrog scheme (see (2.22)), the obtained position vectors $\boldsymbol{q}(t)$ with a $t > 0$ (for each step within the Leapfrog scheme a new position vector is obtained) will increase the distance from the initial position $\boldsymbol{q}(0)$ until some point in time. Then the position vector is making an U-turn and the distance decreases. Hoffman and Gelman (2014) carefully design an updating scheme, where in each iteration Leapfrog steps are obtained until a U-turn is reached. Instead of using the last obtained state as a proposal, the new state is sampled from a subset of the states visited during the Leapfrog method. This is necessary to guarantee the correct stationary distribution of the resulting Markov Chain. The stepsize parameter ϵ is updated during the burn-in period. In the No-U-Turn sampler, there is not a single acceptance-rejection step as in conventional HMC. The idea is to adapt ϵ such that states visited within one iteration would have an average

acceptance probability of δ in conventional HMC, i.e. this average acceptance probability is calculated using a formula similar to (2.23). A typical choice is $\delta = 0.8$. To achieve this, Hoffman and Gelman (2014) use a stochastic optimization method, which relies on ideas of the dual averaging scheme of Nesterov (2009). For more details and pseudo code we refer the reader to Hoffman and Gelman (2014).

The No-U-Turn sampler is implemented within the user-friendly STAN framework (Carpenter et al. (2017)). Within this framework, the required gradients are obtained through automatic differentiation (Carpenter et al. (2015)).

2.3.2 Elliptical slice sampling

In this section we follow Kreuzer and Czado (2019c). Elliptical slice sampling was proposed by Murray et al. (2010). It was developed for models where dependencies are generated through a latent multivariate normal distribution.

We assume that the posterior density for a parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ given data Y is proportional to

$$f(\boldsymbol{\theta}|\mathrm{Y}) \propto \ell(\boldsymbol{\theta}|\mathrm{Y})\varphi(\boldsymbol{\theta}|\mathbf{0}, \Sigma), \tag{2.24}$$

where $\ell(\boldsymbol{\theta}|\mathrm{Y})$ is the likelihood function and $\varphi(\boldsymbol{\theta}|\mathbf{0}, \Sigma)$ is the multivariate normal density with zero mean vector and covariance matrix Σ. Murray et al. (2010) consider the Metropolis-Hastings sampler of Neal (1998), where a proposal $\boldsymbol{\theta'}$, given the current value $\boldsymbol{\theta}$, is obtained as

$$\begin{aligned} \boldsymbol{v} &\sim N_p(\mathbf{0}, \Sigma), \\ \boldsymbol{\theta'} &= \sqrt{1-\alpha^2}\boldsymbol{\theta} + \alpha\boldsymbol{v}. \end{aligned} \tag{2.25}$$

Here $\alpha \in [-1, 1]$ is a fixed step size parameter. The proposal is accepted with probability

$$\min\left(1, \frac{\ell(\boldsymbol{\theta'}|\mathrm{Y})}{\ell(\boldsymbol{\theta}|\mathrm{Y})}\right). \tag{2.26}$$

Elliptical slice sampling adapts the step size parameter α during sampling. This eliminates the need to select the parameter before sampling and it may be a better approach for situations, where good choices of the step size parameter depend on the region of the state space. Murray et al. (2010) first suggest to propose a new state by

$$\begin{aligned} \boldsymbol{v} &\sim N_p(\mathbf{0}, \Sigma), \\ \boldsymbol{\theta'} &= \cos(\omega)\boldsymbol{\theta} + \sin(\omega)\boldsymbol{v}. \end{aligned} \tag{2.27}$$

Here the angle ω corresponds to the step size. As we move ω towards zero the proposal gets closer to the initial value $\boldsymbol{\theta}$. Murray et al. (2010) argue that (2.27) provides a more flexible choice for the proposals compared to (2.25), if the parameter ω is also updated, which is here the case. In elliptical slice sampling we first draw an angle ω from the uniform distribution on $[0, 2\pi]$ and obtain a proposal as outlined in (2.27). This proposal is accepted according to (2.26). If the proposal is not accepted, a new angle is selected with a slice sampling approach (Neal et al. (2003)) such that the angle approaches zero as more samples are rejected. This ensures that at some point the proposal will be accepted. One iteration of the approach is outlined in Algorithm 1. Here, we denote by $\mathrm{U}(a, b)$

the uniform distribution on the interval (a, b) for $a < b$. Murray et al. (2010) show that Algorithm 1 samples from a Markov chain, where (2.24) is the corresponding stationary distribution.

Algorithm 1 Elliptical slice sampling

$\boldsymbol{v} \sim N_p(\mathbf{0}, \Sigma)$
$u \sim \mathrm{U}(0, 1)$
$\omega \sim \mathrm{U}(0, 2\pi)$
$\omega_{min} = \omega - 2\pi, \quad \omega_{max} = \omega$
$\boldsymbol{\theta'} = \cos(\omega)\boldsymbol{\theta} + \sin(\omega)\boldsymbol{v}$
while $\frac{l(\boldsymbol{\theta'}|\mathrm{Y})}{l(\boldsymbol{\theta}|\mathrm{Y})} \leq u$ **do**
  if $\omega < 0$ **then**
    $\omega_{min} = \omega$
  else
    $\omega_{max} = \omega$
  end if
  $\omega \sim \mathrm{U}(\omega_{min}, \omega_{max})$
  $\boldsymbol{\theta'} = \cos(\omega)\boldsymbol{\theta} + \sin(w)\boldsymbol{v}$
end while

3 A single factor copula stochastic volatility model

This chapter is a reproduction of Kreuzer and Czado (2019a) with minor changes.

3.1 Introduction

Vine copulas (see Section 2.2.4) have proven to be a useful tool to facilitate complex dependence structures (Nikoloulopoulos et al. (2012), Brechmann and Czado (2013), Aas (2016), Fink et al. (2017), Nagler et al. (2019)). A vine copula model consists of $\frac{d(d-1)}{2}$ pair copulas, where d is the number of variables. So the number of parameters grows quadratically with d. Krupskii and Joe (2013) proposed the factor copula model, where the number of parameters grows only linearly in d. This model can be seen as a generalization of the Gaussian factor model. The factor copula model provides much more flexibility, compared to the Gaussian one, as it is made up of different pair copulas that can be chosen arbitrarily. Thus it covers a broad range of dependence structures that can accommodate symmetric as well as asymmetric tail dependence.

One way to construct multivariate time series models is to combine a univariate time series model for the margins with a dependence model, such as the factor copula. Univariate time series models for financial data need to account for typical characteristics like time-varying volatility and volatility clustering. Popular examples of such models include generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev (1986)), the more recently developed generalized autoregressive score (GAS) models (Creal et al. (2013)) and stochastic volatility (SV) models (Kim et al. (1998)). Using the classification of Cox et al. (1981), GARCH and GAS models are observation driven models, whereas the SV model is a parameter driven model. In observation driven models volatility is modeled deterministically through the observed past. Inference for these models is often easier, since evaluation of the likelihood is straightforward. Inference for SV models is more involved, since likelihood evaluation requires high-dimensional integration. But efficient MCMC algorithms have been developed (Kastner and Frühwirth-Schnatter (2014)). In the SV model volatility is modeled as latent variables that follow an autoregressive process of order 1. This representation has compared favorably to GARCH specifications in several data sets (Yu (2002), Chan and Grant (2016)).

We propose a copula based SV model. The marginals follow a SV model and the dependence is modeled through a single factor copula. In contrast to other factor SV models, as considered by Han (2005) or Kastner et al. (2017), we only allow for one

factor and dependence parameters remain constant over time. But we do not assume that conditional on the volatilities the observed data is multivariate normal or Student t distributed. Here we provide more flexibility through the choice of different pair copula families.

The single factor copula model has also been applied to financial data by Schamberger et al. (2017) who use dynamic linear models (West and Harrison (2006)) as marginals and by Krupskii and Joe (2013) who use GARCH models as marginals. As it is common in copula modeling, Schamberger et al. (2017) and Krupskii and Joe (2013) both use a two-step approach for estimation. They first estimate marginal parameters and based on these estimates they infer the dependence parameters. Tan et al. (2019) provide full Bayesian inference for a single factor copula based model, but their marginal models have only few parameters and the proposals for MCMC are built using independence among components. However, for SV margins we need to estimate all T log volatilities, where T denotes the length of the time series. Thus, we have more than T parameters to estimate per margin. These more sophisticated marginal models for financial data are difficult to handle within a full Bayesian approach. Nevertheless, we are able to overcome the two-step approach commonly used in copula modeling and provide full Bayesian inference. For this we develop and implement a Hamiltonian Monte Carlo (HMC) (Duane et al. (1987), Neal et al. (2011)) within Gibbs sampler. In HMC information of the gradient of the log posterior density is used to propose new states, which leads to an efficient sampling procedure.

The main contributions of this chapter are: joint Bayesian inference of a single factor copula model with SV margins using HMC, automated selection of linking copula families and improved value at risk (VaR) forecasting over benchmark models in a financial application. More precisely, we first demonstrate how HMC can be employed for the single factor copula model and compare the HMC approach for the copula part to the MCMC approach of Schamberger et al. (2017) who use adaptive rejection Metropolis within Gibbs sampling (Gilks et al. (1995)). HMC shows superior performance in terms of effective sample size, mean squared error and observed coverage probabilities. Further, the HMC scheme is integrated within a Gibbs approach that allows for full Bayesian inference of the proposed single factor copula based SV model, including copula family selection. Copula families are modeled with discrete indicator variables, which can be sampled directly from their full conditionals within our Gibbs approach. Continuous parameters are updated with HMC. Within the Bayesian procedure, marginal and dependence parameters are estimated jointly. We stress that the joint estimation of marginal and dependence parameters is very demanding and is therefore most commonly avoided. Instead, a two-step approach is used where the marginal parameters are considered fixed when estimating dependence parameters, i.e. uncertainty in the estimation of the marginal parameters is ignored. An advantage of the full Bayesian approach is that this uncertainty is not ignored and full uncertainty quantification is straightforward through credible intervals. We further demonstrate the usefulness of the proposed single factor copula SV model with one-day ahead VaR prediction for financial data involving six stocks. Within our full Bayesian approach a VaR forecast is obtained as an empirical quantile of simulations from the predictive distribution. In addition, we show that joint estimation leads to more accurate VaR forecasts than VaR forecasts obtained from a two-step approach.

This chapter is organized as follows. In Sections 3.2 and 3.3 we discuss the single

factor copula model and the single factor copula SV model, respectively. Both sections follow a similar structure. We first specify a Bayesian model, propose a Bayesian inference approach and evaluate the performance of the approach with simulated data. In Section 3.4 the proposed single factor copula SV model is applied to financial returns data. Section 3.5 concludes.

3.2 Bayesian inference for single factor copulas using the HMC approach

In this section we illustrate how HMC (see Section 2.3.1) can be used to estimate the parameters of single factor copulas. Instead of relying on STAN (see Section 2.3.1), we provide our own implementation of HMC. This allows us to use the HMC updates developed for single factor copulas within other samplers, as we will see in Section 3.3.

3.2.1 Model specification

To illustrate the viability of HMC for factor copula models we start with the single factor copula model as a special case of the p factor copula model according to Krupskii and Joe (2013). We consider d uniform on $[0,1]$ distributed variables $U_1, \dots U_d$ together with a uniform on $[0,1]$ distributed latent factor V. In the *single factor copula model* we assume that, given V, the variables $U_1, \dots, U_d$ are independent. This implies that the joint density of $(U_1, \dots U_d)$ can be written as

$$c(u_1, \dots, u_d) = \int_0^1 \prod_{j=1}^{d} c_{j|V}(u_j|v)dv = \int_0^1 \prod_{j=1}^{d} c_j(u_j, v)dv, \tag{3.1}$$

where c_j is the density of $\mathbb{C}_j$, the copula of (U_j, V). The copulas $\mathbb{C}_1, \dots, \mathbb{C}_d$ are called *linking copulas*, since they link each of the observed copula variables U_j to the latent factor V.

For inference we use single-parameter copula families, for which there is a one-to-one correspondence between the copula parameter and Kendall's τ (see Section 2.2.3). We equip each linking copula density with a corresponding Kendall's τ parameter τ_j, and (3.1) becomes

$$c(u_1, \dots, u_d; \boldsymbol{\tau}) = \int_0^1 \prod_{j=1}^{d} c_j(u_j, v; \tau_j)dv,$$

where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)$. As it is common in Bayesian statistics, we treat the latent variable V as a parameter v. The joint density of $(U_1, \dots, U_d)$ given the parameters $\boldsymbol{\tau}, v$ is obtained as

$$c(u_1, \dots, u_d; \boldsymbol{\tau}, v) = \prod_{j=1}^{d} c_j(u_j, v; \tau_j).$$

Since the latent variable V is random for each observation vector $(u_{t1}, \dots, u_{td})$, we have T latent parameters $\boldsymbol{v} = (v_1, \dots, v_T)$ for T time points. The likelihood of the parameters

$(\boldsymbol{\tau}, \boldsymbol{v})$ given T independent observations $\mathrm{U} = (u_{tj})_{t=1,\dots,T, j=1,\dots,d} \in [0,1]^{T\times d}$ is therefore

$$\ell(\boldsymbol{\tau}, \boldsymbol{v}|\mathrm{U}) = \prod_{t=1}^{T}\prod_{j=1}^{d} c_j(u_{tj}, v_t; \tau_j). \tag{3.2}$$

3.2.2 Bayesian inference

So far, Bayesian inference for the single factor copula model was addressed by Schamberger et al. (2017) and Tan et al. (2019). Both approaches use Gibbs sampling where one can exploit the fact that the factors $v_1, \dots v_T$ are independent given the Kendall's τ parameters $\tau_1, \dots \tau_d$ and vice versa. We now show how HMC can be used for the single factor copula model. Sampling with HMC is slower, since it requires several evaluations of the gradient of the log posterior density. However, with HMC there is no blocking involved and we update the whole parameter vector, with well chosen proposals obtained from the Leapfrog approximation, at once. We expect more accurate samples since this sampler suffers less from the dependence between the latent factors and the Kendall's τ parameters. To support this statement, we compare HMC to adaptive rejection Metropolis sampling within Gibbs sampling (ARMGS) (Gilks et al. (1995)). ARMGS is the sampler that worked best among several samplers that have been investigated by Schamberger et al. (2017) for single factor copula models.

Parametrization

Since HMC operates on unconstrained parameters, we need to provide parameter transformations to remove the constraints present in our problem. Furthermore, we restrict the Kendall's τ values to be in $(0,1)$ to avoid problems that might occur due to multimodal posterior distributions. While for this purpose it might also be enough to only restrict one of the Kendall's τ parameters to be positive (Tan et al. (2019)), restricting all of them is not a severe restriction for applications, whenever the signs of the Kendall's τ values are clear from the context. Since $\tau_{U_1,U_2} = -\tau_{U_1,1-U_2}$[1] we can replace U_2 by $1-U_2$ if we want to model negative dependence between U_1 and U_2. The components of the latent factors $\boldsymbol{v}$ are also in $(0,1)$. To transform parameters on the $(0,1)$ scale to the unconstrained scale the logit function is a common choice. Therefore, we use the following transformations for the Kendall's τ parameters $\boldsymbol{\tau}$ and the latent factors $\boldsymbol{v}$

$$\delta_j = \ln\left(\frac{\tau_j}{1-\tau_j}\right), \quad w_t = \ln\left(\frac{v_t}{1-v_t}\right), \tag{3.3}$$

and obtain unconstrained parameters $\delta_j, w_t \in \mathbb{R}$ for $j = 1, \dots d, t = 1, \dots, T$.

Prior densities

We specify the prior distributions for $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)$ and $\boldsymbol{w} = (w_1, \dots, w_T)$ such that the distributions implied for the corresponding Kendall's τ and for v_t are independently

[1] $\tau_{U_1,1-U_2} = P((U_1-\tilde{U}_1)((1-U_2)-(1-\tilde{U}_2)) > 0) - P((U_1-\tilde{U}_1)((1-U_2)-(1-\tilde{U}_2)) < 0) = P((U_1-\tilde{U}_1)(U_2-\tilde{U}_2) < 0) - P((U_1-\tilde{U}_1)(U_2-\tilde{U}_2) > 0) = -\tau_{U_1,U_2}$, where $(\tilde{U}_1, \tilde{U}_2)$ is an independent copy of (U_1, U_2).

uniform on the interval $(0, 1)$. Applying the density transformation law, this implies that the corresponding prior density can be expressed as

$$\pi(\boldsymbol{\delta}, \boldsymbol{w}) = \prod_{j=1}^{d} \pi_u(\delta_j) \prod_{t=1}^{T} \pi_u(w_t), \tag{3.4}$$

where $\pi_u(x) = (1 + \exp(-x))^{-2} \exp(-x), x \in \mathbb{R}$.

Posterior density

With these choices in (3.2) and (3.4), the posterior density is proportional to

$$f(\boldsymbol{\delta}, \boldsymbol{w}|\mathsf{U}) \propto \ell(\boldsymbol{\tau}, \boldsymbol{v}|\mathsf{U}) \cdot \pi(\boldsymbol{\delta}, \boldsymbol{w}), \tag{3.5}$$

where τ_j and v_t are functions of δ_j and w_t respectively. Therefore the log posterior density is given by

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{w}|U) = \sum_{t=1}^{T} \sum_{j=1}^{d} \ln(c_j(u_{tj}, v_t; \tau_j)) + \sum_{t=1}^{T} \ln(\pi_u(w_t)) + \sum_{j=1}^{d} \ln(\pi_u(\delta_j)) + const,$$

where $const \in \mathbb{R}$ is a constant, independent of the parameters $\boldsymbol{\delta}, \boldsymbol{w}$.

Sampling with HMC

Derivatives of the log posterior density with respect to all parameters are determined to perform Leapfrog approximations (see Appendix A.1). With this at hand, HMC can be implemented as any Metropolis-Hastings sampler (see Section 2.3.1). To run the algorithm we need to set values to the hyper parameters: the Leapfrog stepsize ϵ, the number of Leapfrog steps L and the mass matrix M. Choosing ϵ and L is not easy, since good choices of these parameters can vary depending on different regions of the state space. Neal et al. (2011) suggest to randomly select ϵ and L from a set of values that may be appropriate for different regions. This is the approach that we follow. For our simulation study we have seen that choosing ϵ uniformly between 0 and 0.2 and choosing L uniformly between 0 and 40 leads to reasonable mixing as measured by the effective sample size (Gelman et al. (2014a), page 286). The mass matrix M is set equal to the identity matrix. The MCMC procedure is implemented in `R` using the `R` package `Rcpp` by Eddelbuettel et al. (2011) which allows the integration of `C++`. Effective sample sizes are calculated with the `R` package `coda` of Plummer et al. (2008).

3.2.3 Simulation study

To compare our approach, we conduct the same simulation study as in Schamberger et al. (2017). For each of three scenarios, we simulate 100 data sets from the single factor copula model with $T = 200$ and $d = 5$. The three scenarios are characterized by the values of Kendall's τ of the linking copulas and are denoted by the **low** τ, the **high** τ and the **mixed** τ scenario. The Kendall's τ values are shown in Table 3.1. As linking copulas

	$\mathbb{C}_1$	$\mathbb{C}_2$	$\mathbb{C}_3$	$\mathbb{C}_4$	$\mathbb{C}_5$
low τ	0.10	0.12	0.15	0.18	0.20
high τ	0.50	0.57	0.65	0.73	0.80
mixed τ	0.10	0.28	0.45	0.62	0.80

Table 3.1: Kendall's τ values for the linking copulas $\mathbb{C}_1, \dots \mathbb{C}_5$ in the three scenarios.

only Gumbel copulas are considered. Based on these simulated data sets, the samplers are run for 11000 iterations, whereas the first 1000 iterations are discarded for burn-in.

	τ		V	
	ARMGS	HMC	ARMGS	HMC
Low τ				
MAD	0.1088	0.0564	0.2808	0.2158
MSE	0.0314	0.0059	0.1248	0.0716
ESS/min	6	92	26	246
90% C.I.	0.91	0.94	0.84	0.88
95% C.I.	0.96	0.98	0.91	0.94
High τ				
MAD	0.0292	0.0201	0.0709	0.0502
MSE	0.0014	0.0007	0.0095	0.0046
ESS/min	24	268	44	278
90% C.I.	0.89	0.90	0.89	0.91
95% C.I.	0.95	0.94	0.95	0.95
Mixed τ				
MAD	0.0509	0.0340	0.0828	0.0684
MSE	0.0043	0.0019	0.0132	0.0082
ESS/min	21	132	26	210
90% C.I.	0.87	0.89	0.79	0.85
95% C.I.	0.93	0.93	0.88	0.93

Table 3.2: Comparison of the ARMGS and HMC method in terms of estimated mean absolute deviation (MAD), mean squared error (MSE), effective sample size per minute (ESS/min) and observed coverage probability of the credible intervals (C.I.).

	τ_1	τ_2	τ_3	τ_4	τ_5	v_{10}	v_{50}	v_{100}	v_{150}	v_{190}
Low τ										
MAD	0.0500	0.0493	0.0568	0.0591	0.0666	0.2530	0.2204	0.2376	0.2190	0.2203
MSE	0.0047	0.0043	0.0049	0.0058	0.0096	0.0969	0.0719	0.0819	0.0712	0.0703
ESS/min	121	114	91	76	58	239	255	268	247	262
90% C.I.	0.94	0.94	0.91	0.92	0.98	0.85	0.88	0.83	0.88	0.90
95% C.I.	0.98	0.98	0.97	0.96	0.99	0.93	0.96	0.90	0.91	0.93
High τ										
MAD	0.0253	0.0214	0.0204	0.0158	0.0174	0.0549	0.0549	0.0475	0.0474	0.0474
MSE	0.0011	0.0007	0.0007	0.0004	0.0005	0.0058	0.0049	0.0043	0.0043	0.0044
ESS/min	320	319	312	265	125	277	278	275	279	278
90% C.I.	0.84	0.91	0.88	0.92	0.94	0.86	0.86	0.88	0.95	0.90
95% C.I.	0.89	0.95	0.94	0.97	0.97	0.92	0.96	0.94	0.96	0.95
Mixed τ										
MAD	0.0375	0.0358	0.0278	0.0256	0.0431	0.0690	0.0668	0.0636	0.0712	0.0663
MSE	0.0021	0.0020	0.0012	0.0010	0.0031	0.0098	0.0078	0.0080	0.0084	0.0085
ESS/min	147	250	201	50	11	212	207	224	218	221
90% C.I.	0.89	0.85	0.88	0.93	0.89	0.87	0.87	0.87	0.81	0.87
95% C.I.	0.95	0.89	0.94	0.94	0.94	0.89	0.94	0.94	0.88	0.92

Table 3.3: Detailed simulation results for the HMC method. We show the estimated mean absolute deviation (MAD), mean squared error (MSE), effective sample size per minute (ESS/min) and observed coverage probability of the credible intervals (C.I.) for $\tau_1, \dots, \tau_5$ and five selected latent variables $v_t, t = 10, 50, 100, 150, 190$.

Table 3.2 shows the results of the simulation study and compares them to the re-

sults obtained by Schamberger et al. (2017) using adaptive rejection Metropolis sampling within Gibbs sampling (ARMGS). The corresponding error statistics (e.g. mean absolute deviation (MAD), mean squared error (MSE)) for each parameter is obtained from 100 replications. Then, e.g. the MSE for τ in Table 3.2 is computed as the average of MSE for τ_1 , ..., MSE for τ_5. Since the objective is the comparison of our method to the method of Schamberger et al. (2017), we follow their approach and calculate the error statistics from point estimates (marginal posterior mode estimates obtained from univariate kernel density estimates). Further, we calculate the error statistics for $\tau_1, \dots, \tau_5, v_1, \dots, v_{200}$ which are one-to-one transformations of $\delta_1, \dots, \delta_5, w_1, \dots, w_{200}$. We see that a more accurate credible interval, a lower mean absolute deviation and a lower mean squared error is achieved in most cases by HMC compared to ARMGS. Furthermore, the effective sample size per minute is much higher for HMC. Table 3.3 shows the results of the simulation study in more detail, i.e. we do not average over values of $\tau_1, \dots, \tau_5$ and $v_1, \dots v_{200}$. It is noticable that mixing is worse for higher values of Kendall's τ in every scenario, whereas it is most extreme in the mixed τ scenario. This was also observed for ARMGS (see Schamberger et al. (2017) Table 9 in the appendix).

3.3 The single factor copula stochastic volatility model

Now we combine the single factor copula with margins driven by a SV model with Gaussian errors (see Section 2.1.2) and develop a Bayesian approach to jointly estimate the parameters of the proposed model.

3.3.1 Model specification

We propose a multivariate dynamic model where each marginal follows a SV model and the dependence between the marginals is captured by a single factor copula, the *single factor copula stochastic volatility (factor copula SV) model*. In particular for $t = 1, \dots, T, j = 1, \dots, d$ we assume that

$$\begin{aligned} Z_{tj} &= \exp\left(\frac{s_{tj}}{2}\right)\epsilon_{tj} \\ s_{tj} &= \mu_j + \phi_j(s_{t-1j} - \mu_j) + \sigma_j \eta_{tj}, \end{aligned}$$

where $\mu_j \in \mathbb{R}, \phi_j \in (-1, 1), \sigma_j \in (0, \infty), s_{0j}|\mu_j, \phi_j, \sigma_j \sim N\left(\mu_j, \frac{\sigma_j^2}{1-\phi_j^2}\right)$ and $\eta_{tj} \sim N(0, 1)$ iid holds. The joint distribtion of the errors ϵ_{tj} is now considered. We model the dependence among the marginals by employing a factor copula model on the errors. Similar to Tan et al. (2019), we further allow for Bayesian selection of the d linking copula families of this factor copula instead of assuming that they were known as in Section 3.2 and as in Schamberger et al. (2017). The families are chosen from a set $\mathcal{M}$ of single-parameter copula families, e.g. $\mathcal{M} = \{\text{Gaussian, Gumbel, Clayton}\}$. Schamberger et al. (2017) estimated one model for each specification of the linking copula families. Since there are $|\mathcal{M}|^d$ different specifications, they only considered factor copulas where all linking copulas belong to the same family. With our Bayesian family selection, we can profit from the full flexibility of the factor copula model by allowing for all $|\mathcal{M}|^d$ specifications. In particular,

our modeling approach allows to combine different copula families. Therefore we define d family indicator variables $m_j \in \mathcal{M}, j = 1, \dots, d$. Further, we introduce, similar to Section 3.2, Kendall's τ parameters of the linking copulas $\tau_j \in (0,1), j = 1, \dots, d$. Note that the parameter τ_j has the same interpretation for different copula families, since it is the associated Kendall's τ value. As already noted by Tan et al. (2019), this allows to share this parameter among different copula families. More details about this argument are given in Appendix G.

Since we model the dependence among the errors with a single factor copula, we assume that there exists a latent factor $v_t \sim \mathrm{U}(0,1)$ for each t such that the following holds for the error vector at time t, $\boldsymbol{\epsilon_t} = (\epsilon_{t1}, \dots \epsilon_{td})$,

$$f(\boldsymbol{\epsilon_t}|v_t, m_1, \dots, m_d, \tau_1, \dots, \tau_d) = \prod_{j=1}^{d} \left[c_j^{m_j}(\Phi(\epsilon_{tj}), v_t; \tau_j)\varphi(\epsilon_{tj}) \right], \tag{3.6}$$

where φ and Φ denote the standard normal density and distribution function, respectively. In particular $\epsilon_{tj} \sim N(0,1)$ for any t and j. Here $c_j^{m_j}(\cdot, \cdot; \tau_j)$ is the density of the bivariate copula family m_j with Kendall's τ parameter τ_j. Integrating out the factor v_t in (3.6) yields

$$f(\boldsymbol{\epsilon_t}|m_1, \dots, m_d, \tau_1, \dots, \tau_d) = \left[\int_{(0,1)} \prod_{j=1}^{d} c_j^{m_j}(\Phi(\epsilon_{tj}), v_t; \tau_j) dv_t \right] \prod_{j=1}^{d} \varphi(\epsilon_{tj}). \tag{3.7}$$

Furthermore, we assume that the T components of ($\boldsymbol{\epsilon_1}$, …, $\boldsymbol{\epsilon_T}$) are independent given the family indicators $m_1, \dots, m_d$ and the dependence parameters $\tau_1, \dots, \tau_d, v_1, \dots, v_T$. To shorten notation we use the following abbreviations:

- $\mathrm{Z} = (z_{tj})_{t=1,\dots,T, j=1,\dots d}$ the matrix of observations,
- $\mathcal{E} = (\epsilon_{tj})_{t=1,\dots,T, j=1,\dots d}$ the matrix of errors,
- $\boldsymbol{\mu} = (\mu_j)_{j=1,\dots d}$ the vector of means of the marginal stochastic volatility models,
- $\boldsymbol{\phi} = (\phi_j)_{j=1,\dots d}$ the vector of persistence parameters of the marginal stochastic volatility models,
- $\boldsymbol{\sigma} = (\sigma_j)_{j=1,\dots d}$ the vector of standard deviations of the marginal stochastic volatility models,
- $S = (s_{tj})_{t=0,\dots,T, j=1,\dots d}$ the matrix of log variances,
- $\boldsymbol{s_{\cdot j}} = (s_{tj})_{t=0,\dots,T}$ the vector of log variances of the j-th marginal,
- $\boldsymbol{v} = (v_t)_{t=1,\dots,T}$ the vector of latent factors,
- $\boldsymbol{\tau} = (\tau_j)_{j=1,\dots,d}$ the vector of the Kendall's τ parameters of the linking copulas,
- $\boldsymbol{m} = (m_j)_{j=1,\dots,d}$ the vector of copula family indicators.

Utilizing these abbreviations, we can summarize the parameters of our model as $\{\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{v}, \boldsymbol{\tau}, \boldsymbol{m}\}$.

The model with Gaussian linking copulas

For the special case where all linking copulas are Gaussian (i.e. m_j = Gaussian for $j = 1, \dots, d$), Krupskii and Joe (2013) show that the errors ϵ_{tj} specified in (3.7) allow for the following stochastic representation

$$\epsilon_{tj} = \rho_j w_t + \sqrt{1 - \rho_j^2}\xi_{tj},$$

where $w_t \sim N(0,1)$ and $\xi_{tj} \sim N(0,1)$ independently and $\rho_j = \sin\left(\frac{\pi}{2}\tau_j\right)$ is the correlation parameter of the bivariate Gaussian copula (see Section 2.2.3). Therefore we obtain the following additive structure

$$Z_{tj} = \rho_j \exp\left(\frac{s_{tj}}{2}\right) w_t + \exp\left(\frac{s_{tj}}{2}\right)\sqrt{1 - \rho_j^2}\xi_{tj}. \tag{3.8}$$

This implies a time-varying covariance matrix with elements

$$\operatorname{cov}(Z_{tj}, Z_{tk}) = \rho_j \rho_k \exp\left(\frac{s_{tj}}{2}\right)\exp\left(\frac{s_{tk}}{2}\right) \text{ for } j \neq k.$$

The correlation matrix however remains constant as time evolves and its off-diagonal elements are given by

$$\operatorname{cor}(Z_{tj}, Z_{tk}) = \rho_j \rho_k \text{ for } j \neq k.$$

The additive structure in (3.8) shows connections to other multivariate factor stochastic volatility models (see Chib et al. (2006), Kastner et al. (2017)). This can be seen by considering the following reparametrization

$$s'_{tj} := s_{tj} + \ln\left(1 - \rho_j^2\right), \quad \lambda_j := \frac{\rho_j}{\sqrt{1 - \rho_j^2}},$$

which implies the following representation of (3.8)

$$Z_{tj} = \lambda_j \exp\left(\frac{s'_{tj}}{2}\right) w_t + \exp\left(\frac{s'_{tj}}{2}\right)\xi_{tj}. \tag{3.9}$$

Here s'_{tj} is an AR(1) process with mean $\mu_j + \ln\left(1 - \rho_j^2\right)$, persistence parameter ϕ_j and standard deviation parameter σ_j. For comparison, the model of Kastner et al. (2017) with one factor is given by

$$Z_{tj} = \lambda_j \exp\left(\frac{s'_{td+1}}{2}\right) w_t + \exp\left(\frac{s'_{tj}}{2}\right)\xi_{tj},$$

with one additional latent AR(1) process $s'_{td+1}, t = 1, \dots, T$. This implies time-varying correlations given by

$$\operatorname{cor}(Z_{tj}, Z_{tk}) = \frac{\lambda_j \lambda_k \exp\left(s'_{td+1}\right)}{\sqrt{\lambda_j^2 \exp\left(s'_{td+1}\right) + \exp\left(s'_{tj}\right)}\sqrt{\lambda_k^2 \exp\left(s'_{td+1}\right) + \exp(s'_{tk})}} \text{ for } j \neq k.$$

Dividing Z_{tj} by $\exp\left(\frac{s'_{tj}}{2}\right)$ in (3.9), we recognize the structure of a standard factor model for $Z'_{tj} = \frac{Z_{tj}}{\exp\left(\frac{s'_{tj}}{2}\right)}$ given by

$$Z'_{tj} = \lambda_j w_t + \xi_{tj}, \tag{3.10}$$

with factor loadings $\lambda_1, \ldots, \lambda_d$ and factor w_t. In representation (3.10) the variance of ξ_{tj} is restricted to 1, whereas in the standard factor model (see e.g. Lopes and West (2004)) it is usually modeled through an additional variance parameter. Since the variance of ϵ_{tj} is already determined ($\epsilon_{tj} \sim N(0,1)$), we have this additional restriction compared to factor models with flexible marginal variance. Note that Z_{tj} still has flexible variance and the restriction for ϵ_{tj} is necessary to ensure identifiability.

If all copula families are Gaussian, other multivariate factor stochastic volatility models provide generalizations by allowing for more factors and for a time-varying correlation. We provide generalization with respect to the error distribution. The choice of different pair copula families provides a flexible modeling approach and our model can accommodate features that cannot be modeled with a multivariate normal distribution, as e.g. symmetric or asymmetric tail dependence.

Schamberger et al. (2017) also use factor copulas to model dependence among financial assets. Their approach differs to our approach in the choice of the marginal model. They use dynamic linear models (West and Harrison (2006)). Secondly, they assume the copula families to be known and they perform a two-step estimation approach, whereas we provide full Bayesian inference.

3.3.2 Bayesian inference

In the following we develop a full Bayesian approach for the proposed model. We use a block Gibbs sampler to sample from the posterior distribution. We use d blocks for the marginal parameters $(\mu_j, \phi_j, \sigma_j, \boldsymbol{s}_{\cdot j})$, $j = 1, \ldots, d$, one block for the dependence parameters $(\boldsymbol{\tau}, \boldsymbol{\nu})$ and d blocks for the copula family indicators $\boldsymbol{m}$. Sampling from the full conditionals is done with HMC for the first $d+1$ blocks. Conditioning the dependence parameters on the marginal parameters and on the copula family indicators, we are in the single factor copula framework of Section 3.2. We have seen that HMC provides an efficient way to sample the dependence parameters. Conditioned on the dependence parameters and on the family indicators, the marginal parameters corresponding to different dimensions are independent. Each dimension can be considered as a generalized stochastic volatility model, where the distribution of the errors is determined by the corresponding linking copula. Sampling from the posterior distribution is more involved than in the Gaussian case. In the Gaussian case one can use an approximation based on a mixture of normal distributions and rewrite the observation equation $Z_{tj} = \exp\left(\frac{s_{tj}}{2}\right)\epsilon_{tj}$ to obtain a linear, conditionally Gaussian state space model (Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014)). This is not possible in our case and therefore HMC, which has already shown good performance for the copula part and only requires derivation of the derivatives, is our method of choice. The family indicators $\boldsymbol{m}$ are discrete variables which can be sampled directly from their full conditionals.

Prior densities

For the copula family indicators we use independent discrete uniform priors, i.e

$$\pi(m_j) = \frac{1}{|\mathcal{M}|} \tag{3.11}$$

for $m_j \in \mathcal{M}, j = 1, \ldots, d$ independently. The prior density of the other parameters is chosen as the product of the prior densities used for the single factor copula model and for the marginal stochastic volatility model in Section 2.1.2, i.e.

$$\pi(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}) = \prod_{j=1}^{d} \pi(\mu_j, \phi_j, \sigma_j, \boldsymbol{s_{\cdot j}}), \tag{3.12}$$

where $\pi(\cdot)$ is specified in (2.6). Note that for the components of $\boldsymbol{\tau}$ and $\boldsymbol{\upsilon}$, we utilize uniform priors on $(0, 1)$. Further, we assume that the family indicators are a priori independent of the parameters in (3.12).

Likelihood

The conditional independence of the T components of $(\boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_T})$ implies that the conditional distribution of the errors given the dependence parameters and the copula family indicators is

$$f(\mathcal{E}|\boldsymbol{\upsilon}, \boldsymbol{\tau}, \boldsymbol{m}) = \prod_{t=1}^{T} \prod_{j=1}^{d} \left[c_j^{m_j}(\Phi(\epsilon_{tj}), \upsilon_t; \tau_j)\varphi(\epsilon_{tj}) \right].$$

Using the density transformation rule, the likelihood of parameters $(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m})$ given the observation matrix $\mathrm{Z} = (z_{tj})_{t=1,\ldots,T, j=1,\ldots,d} \in \mathbb{R}^{T \times d}$ is obtained as

$$\ell(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m}|\mathrm{Z}) = \prod_{t=1}^{T} \prod_{j=1}^{d} \left[c_j^{m_j} \left(\Phi \left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)} \right), \upsilon_t; \tau_j \right) \varphi \left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)} \right) \frac{1}{\exp\left(\frac{s_{tj}}{2}\right)} \right].$$

Sampling the marginal parameters

The conditional density we need to sample from is given by

$$\begin{aligned} f(\mu_j, \phi_j, \sigma_j, \boldsymbol{s_{\cdot j}}|\mathrm{Z}, \boldsymbol{\mu_{-j}}, \boldsymbol{\phi_{-j}}, \boldsymbol{\sigma_{-j}}, S_{\cdot -j}, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m}) \\ \propto \ell(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m}|\mathrm{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}) \\ \propto \prod_{t=1}^{T} \left[c_j^{m_j} \left(\Phi \left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)} \right), \upsilon_t; \tau_j \right) \varphi \left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)} \right) \frac{1}{\exp\left(\frac{s_{tj}}{2}\right)} \right] \pi(\mu_j, \phi_j, \sigma_j, \boldsymbol{s_{\cdot j}}). \end{aligned}$$

Here the abbreviation $\boldsymbol{x_{-j}}$ refers to the vector $\boldsymbol{x}$ with the $j-$th component removed and $X_{\cdot -j}$ is the matrix X with the j-th column removed. We sample from this density with HMC as will be outlined below.

Parametrization As in Section 3.2, we need to provide parametrizations such that resulting parameters are unconstrained. In particular we use the following transformations

$$\xi_j = F_Z(\phi_j), \qquad \psi_j = \ln(\sigma_j),$$

where $F_Z(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$ is Fisher's Z transformation. Although the latent log variances are already unconstrained, we make use of the following reparametrization

$$\begin{aligned} \tilde{s}_{0j} &= \frac{(s_{0j} - \mu_j)\sqrt{1-\phi_j^2}}{\sigma_j} \\ \tilde{s}_{tj} &= \frac{s_{tj} - \mu_j - \phi_j(s_{t-1j} - \mu_j)}{\sigma_j}, t = 1, \dots, T. \end{aligned} \tag{3.13}$$

The transformation for $\boldsymbol{s_{\cdot j}}$ was proposed by the Stan Team (2015) for the univariate stochastic volatility model and implies that $\tilde{\boldsymbol{s}}_{\cdot \boldsymbol{j}}|\mu_j, \phi_j, \sigma_j \sim N_{T+1}(\mathbf{0}, I_{T+1})$, where I_{T+1} denotes the $(T+1)$-dimensional identity matrix. According to Yu and Meng (2011), the original parametrization in terms of s_{tj} is a sufficient augmentation scheme, whereas the parametrization in terms of $\tilde{s}_{tj}$ is an ancillary augmentation. The performance of Markov Chain Monte Carlo methods can vary a lot for different parametrizations (Frühwirth-Schnatter and Sögner (2003), Strickland et al. (2008)). Betancourt and Girolami (2015) have seen better performance for the ancillary augmentation when sampling from the posterior distribution of hierarchical models with HMC. Their explanation is that within the ancillary augmentation variables may be less correlated. Here we also rely on the ancillary augmentation, since we have seen much better performance for this parametrization in terms of effective sample size.

Prior densities We consider the joint prior density of the parameters μ_j, ξ_j, ψ_j and $\tilde{\boldsymbol{s}}_{\cdot \boldsymbol{j}}$. The log of this joint prior density is given by

$$\ln(\pi(\mu_j, \xi_j, \psi_j, \tilde{\boldsymbol{s}}_{\cdot j})) \propto \ln(\pi(\mu_j)) + \ln(\pi(\xi_j)) + \ln(\pi(\psi_j)) - \frac{1}{2}\sum_{t=0}^{T} \tilde{s}_{tj}^2 + const1,$$

where $\pi(\cdot)$ are the corresponding prior densities implied by (3.12) (see Appendix A.2 for details) and $const1 \in \mathbb{R}$ is a constant.

Posterior density The log density we need to sample from is given by

$$\begin{aligned} \mathcal{L}(\mu_j, \xi_j, \psi_j, \tilde{\boldsymbol{s}}_{\cdot j}|\mathrm{Z}, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{m}) = \\ \sum_{t=1}^{T}\left[\ln\left(c_j^{m_j}\left(\Phi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right), v_t; \tau_j\right)\right) + \ln\left(\varphi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right)\right) - \frac{s_{tj}}{2}\right] \\ + \ln(\pi(\mu_j, \xi_j, \psi_j, \tilde{\boldsymbol{s}}_{\cdot \boldsymbol{j}})) + const2, \end{aligned}$$

where $\boldsymbol{s_{\cdot j}}$ is a function of $\tilde{\boldsymbol{s}}_{\cdot \boldsymbol{j}}$ (see (3.13)) and $const2 \in \mathbb{R}$ is a constant. The necessary derivatives of this log density are derived (see Appendix A.3) for the Leapfrog approximations and then sampling of the marginal parameters is straightforward.

Sampling the dependence parameters

The conditional density we need to sample from for the dependence parameters is proportional to

$$\begin{aligned} f(\boldsymbol{\tau}, \boldsymbol{\upsilon}|\mathrm{Z}, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{m}) &\propto \ell(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m}|\mathrm{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}) \\ &\propto \ell(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m}|\mathrm{Z}) \\ &\propto \prod_{t=1}^{T}\prod_{j=1}^{d} c_j^{m_j}\left(\Phi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right), \upsilon_t; \tau_j\right). \end{aligned}$$

To sample from this density we use the same HMC approach as in Section 3.2.

Sampling the copula family indicators

The full conditional of m_j is obtained as

$$f(m_j|\mathrm{Z}, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, S, \boldsymbol{\tau}, \boldsymbol{\upsilon}, \boldsymbol{m_{-j}}) = \frac{\prod_{t=1}^{T} c_j^{m_j}\left(\Phi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right), \upsilon_t; \tau_j\right)}{\sum_{m'_j \in \mathcal{M}} \prod_{t=1}^{T} c_j^{m'_j}\left(\Phi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right), \upsilon_t; \tau_j\right)}.$$

We can sample directly from this discrete distribution and no MCMC updates are required here.

3.3.3 Simulation study

We conduct a simulation study to evaluate the performance of the proposed joint HMC sampler. We consider one scenario in five dimensions and one scenario in ten dimensions, as specified in Table 3.4. We choose rather high values for the marginal persistence parameter ϕ and moderate values for the dependence parameter Kendall's τ. These choices roughly correspond to what we expect to see in financial data. For each scenario we simulate 100 data sets from the model introduced in Section 3.3.1. The proposed MCMC sampler with HMC updates is then applied to the simulated data. The sampler is run for 2500 iterations, whereas the first 500 iterations are discarded for burn-in. For family selection we consider the following set of single-parameter copula families {Gaussian, Student t(df=4), Clayton, Gumbel}.

$$\begin{aligned} \boldsymbol{\mu_{sim}} &= (-6, -6, -7, -7, -8) \\ \boldsymbol{\phi_{sim}} &= (0.7, 0.8, 0.85, 0.9, 0.95) \\ \boldsymbol{\sigma_{sim}} &= (0.2, 0.2, 0.3, 0.3, 0.4) \\ \boldsymbol{\tau_{sim}} &= (0.3, 0.4, 0.5, 0.6, 0.7) \\ \boldsymbol{m_{sim}} &= (\text{Gaussian, Student t(df=4), Clayton, Gumbel, Gaussian}) \end{aligned} \tag{3.14}$$

Scenario	d	T	$\boldsymbol{\mu}$	$\boldsymbol{\phi}$	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\boldsymbol{m}$
1	5	1000	$\boldsymbol{\mu_{sim}}$	$\boldsymbol{\phi_{sim}}$	$\boldsymbol{\sigma_{sim}}$	$\boldsymbol{\tau_{sim}}$	$\boldsymbol{m_{sim}}$
2	10	1000	$(\boldsymbol{\mu_{sim}}, \boldsymbol{\mu_{sim}})$	$(\boldsymbol{\phi_{sim}}, \boldsymbol{\phi_{sim}})$	$(\boldsymbol{\sigma_{sim}}, \boldsymbol{\sigma_{sim}})$	$(\boldsymbol{\tau_{sim}}, \boldsymbol{\tau_{sim}})$	$(\boldsymbol{m_{sim}}, \boldsymbol{m_{sim}})$

Table 3.4: Parameter specification for the two different scenarios in the simulation study.

Figures 3.1 and 3.2 show estimated posterior densities and trace plots of one MCMC run for the five-dimensional setup. The trace plots cover the true values and suggest that we achieve proper mixing.

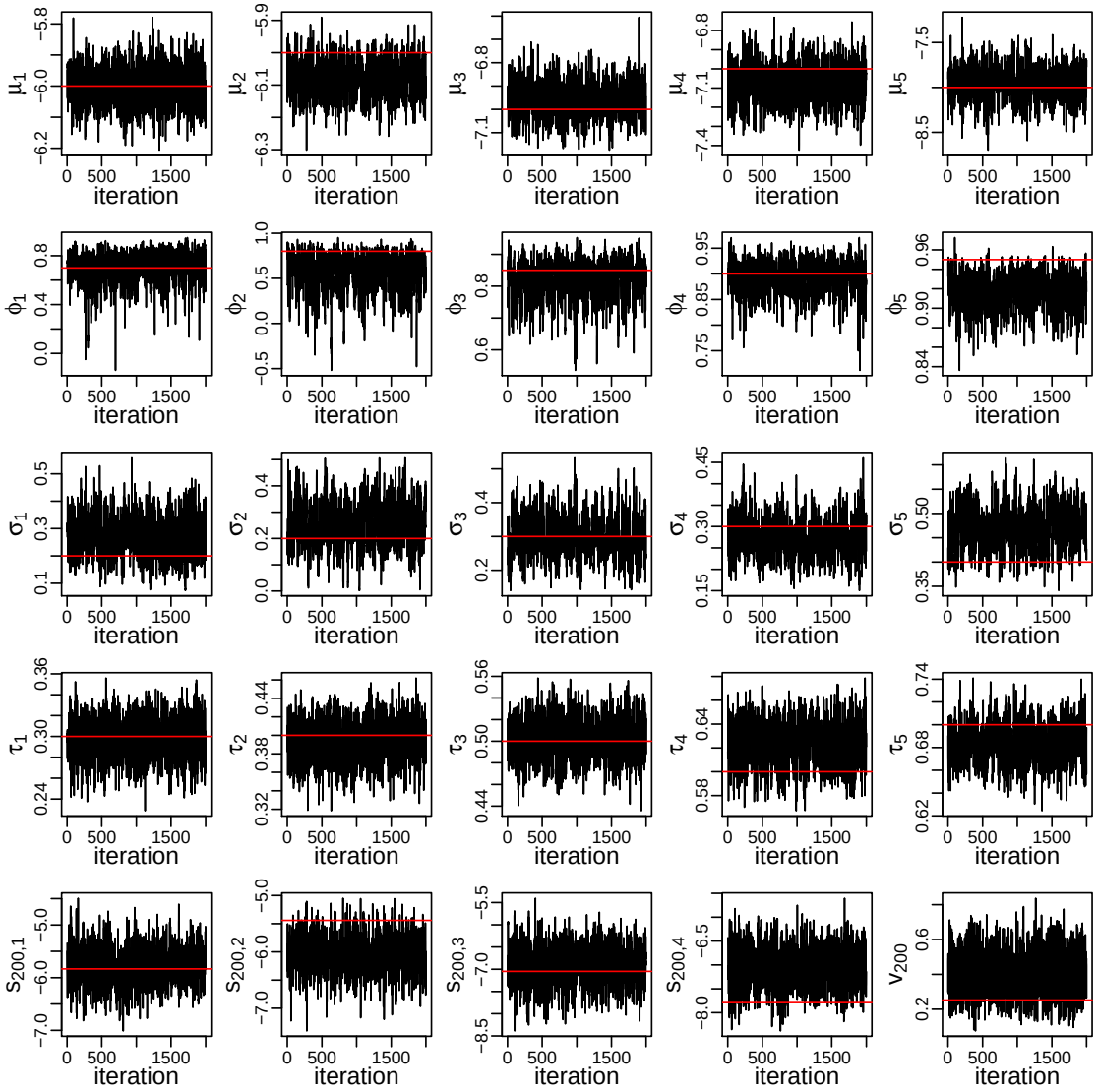

Figure 3.1: Trace plots of selected parameters obtained from a single MCMC run in Scenario 1. The trace plots show 2000 MCMC iterations after a burn-in of 500. The true parameter value is added in red.

The simulation results are summarized in Tables 3.5 and 3.6 for the five-dimensional scenario and in Tables A.1 and A.2 in Appendix A.4 for the ten-dimensional setup. Comparing these two setups, we see that the effective sample sizes are better for the five-dimensional scenario. Besides that, the results of the five and ten-dimensional setups are only slightly different and therefore we discuss only the five-dimensional scenario. Comparing the simulation results for the factor copula parameters to the results of Section 3.2.3, we see that we perform worse in terms of observed coverage probabilities and MSE. But this is not surprising, because here we consider a much more complex model and

also update the copula families and the marginal parameters. Further, we see that the ESS decreases from τ_1 up to τ_5. This is in line with our findings in Section 3.2.3, where we have seen that mixing is worse for higher Kendall's τ values. We can also observe differences with respect to the observed coverage probability of credible intervals. For a low marginal persistence parameter (ϕ_1), coverage probabilities are very high suggesting a broad posterior distribution. For a high persistence parameter (ϕ_5), the observed coverage probabilities are lower. We also see from Figure 3.2 that the estimated posterior density of ϕ_1 is more dispersed compared to the estimated posterior density of ϕ_5. Table 3.6 shows that the correct copula family was selected in at least 66 out of 100 cases. This frequency is best for the first linking copula which has a low Kendall's τ value and worst for the linking copula with the highest Kendall's τ value.

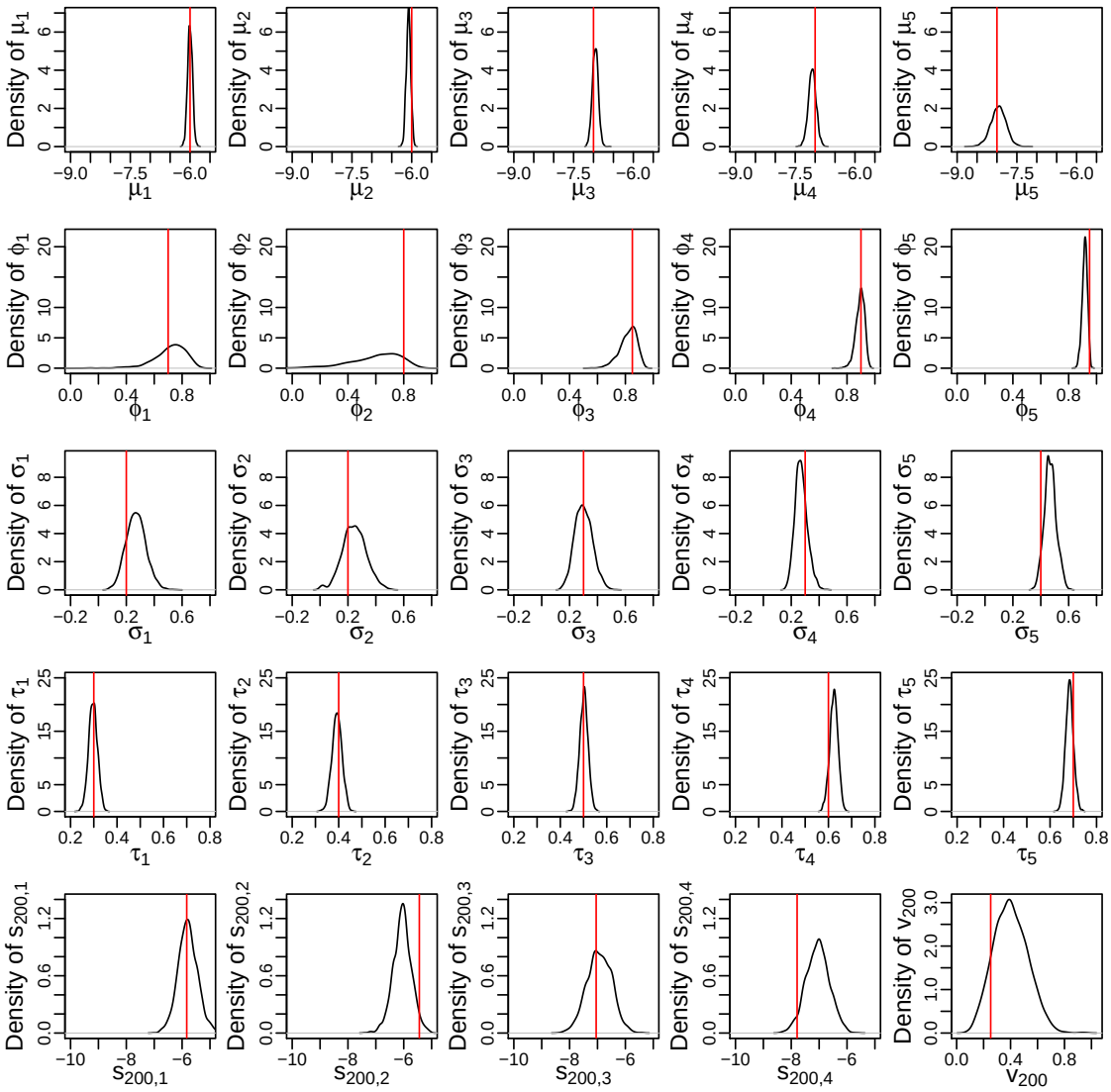

Figure 3.2: Kernel density estimates of the posterior density of selected parameters obtained from a single MCMC run in Scenario 1. The estimates are based on 2000 MCMC iterations after a burn-in of 500. The true parameter value is added in red.

Overall, the results suggest that the method performs well. For all parameters we obtain reasonable MSE and ESS values and our method is able to select the correct copula family in most cases. In particular, our HMC schemes do a good job at jointly updating more than $T = 1000$ parameters of one Gibbs block.

Scenario 1	μ_1	μ_2	μ_3	μ_4	μ_5	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
MSE	0.0027	0.0038	0.0059	0.0107	0.0851	0.0362	0.0408	0.0059	0.0017	0.0003
C.I. 90%	0.91	0.86	0.92	0.92	0.82	0.97	0.86	0.89	0.90	0.83
C.I. 95%	0.95	0.90	0.96	0.94	0.87	0.99	0.94	0.94	0.93	0.89
ESS	1022	666	761	942	505	644	433	399	461	325
	σ_1	σ_2	σ_3	σ_4	σ_5	τ_1	τ_2	τ_3	τ_4	τ_5
MSE	0.0077	0.0055	0.0037	0.0024	0.0026	0.0094	0.0164	0.0255	0.0364	0.0503
C.I. 90%	0.95	0.93	0.91	0.91	0.81	0.78	0.79	0.68	0.79	0.72
C.I. 95%	0.98	0.93	0.93	0.97	0.88	0.84	0.85	0.77	0.81	0.74
ESS	391	368	326	360	255	879	770	528	480	280
	$s_{300,1}$	$s_{300,2}$	$s_{300,3}$	$s_{300,4}$	$s_{300,5}$	v_{100}	v_{200}	v_{500}	v_{800}	v_{900}
MSE	0.0564	0.0892	0.2234	0.1836	0.2132	0.0239	0.0241	0.0283	0.0222	0.0202
C.I. 90%	0.94	0.88	0.93	0.91	0.94	0.86	0.91	0.83	0.86	0.84
C.I. 95%	0.98	0.95	0.95	0.96	0.97	0.91	0.94	0.88	0.88	0.89
ESS	1448	433	1433	1343	1334	997	1104	1036	1111	1085

Table 3.5: MSE estimated using the posterior mode, observed coverage probability of the credible intervals (C.I.) and effective sample sizes (ESS) calculated from 2000 posterior draws for selected parameters (Scenario 1).

m_1	m_2	m_3	m_4	m_5
94%	90%	87%	77%	66%

Table 3.6: Proportion of how often the correct copula family was selected. The selected copula family is the posterior mode estimate of m_j for $j = 1, \ldots, 5$ (Scenario 1).

3.4 Application: Value at risk prediction

We illustrate our approach with one-day ahead value at risk (VaR) prediction for a portfolio consisting of several stocks. These predictions can be obtained from simulations of the predictive distribution. As before, Z is the data matrix containing T observations of the d stocks. We need to sample from the predictive distribution of the log returns at time $T+1$, $\boldsymbol{Z_{T+1}} = (Z_{T+1,1}, \ldots Z_{T+1,d})$, given Z. The parameters $\boldsymbol{s_{T+1}} = (s_{T+1,1}, \ldots, s_{T+1,d}), v_{T+1}$ are associated with the new time point $T + 1$ and we obtain simulations from the joint density

$$f(\boldsymbol{z_{T+1}}, \boldsymbol{s_{T+1}}, S, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\tau}, v_{T+1}, \boldsymbol{v}, \boldsymbol{m}|Z), \tag{3.15}$$

with the following steps:

- Simulate $S, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{m}$ from the corresponding posterior distribution given the data Z with our sampler developed in Section 3.3. We discard the first 500 samples for burn-in and denote the remaining $R = 2000$ samples by $S^r, \boldsymbol{\mu^r}, \boldsymbol{\phi}^r, \boldsymbol{\sigma}^r, \boldsymbol{\tau^r}, \boldsymbol{v^r}, \boldsymbol{m^r}$, $r = 1, \ldots, R$.

We proceed as follows for $r = 1, \ldots, R$:

- Simulate $v^r_{T+1} \sim \mathrm{U}(0, 1)$.

- For $j = 1, \ldots, d$ simulate $s^r_{T+1,j} \sim N(\mu^r_j + \phi^r_j(s^r_{Tj} - \mu^r_j), (\sigma^r_j)^2)$.
- To obtain the sample $\boldsymbol{z^r_{T+1}}$ from

$$f(\boldsymbol{z}_{T+1}|\boldsymbol{s^r_{T+1}}, S^r, \boldsymbol{\mu^r}, \boldsymbol{\phi^r}, \boldsymbol{\sigma^r}, \boldsymbol{\tau^r}, v^r_{T+1}, \boldsymbol{v^r}, \boldsymbol{m^r}, Z) = \prod_{j=1}^{d} \left[c_j^{m^r_j} \left(\frac{z_{T+1,j}}{\exp\left(\frac{s^r_{T+1,j}}{2}\right)}, v^r_{T+1}; \tau^r_j \right) \varphi \left(\frac{z_{T+1,j}}{\exp\left(\frac{s^r_{T+1,j}}{2}\right)} \right) \cdot \frac{1}{\exp\left(\frac{s^r_{T+1j}}{2}\right)} \right],$$

  we simulate u^r_j from $\mathbb{C}_j^{m^r_j}\left(\cdot|v^r_{T+1}; \tau^r_j\right)$ and set $z^r_{T+1,j} = \Phi^{-1}\left(u^r_j|0, \exp\left(s^r_{T+1,j}\right)\right)$ for $j = 1, \ldots, d$. Here $\Phi(\cdot|\mu_{normal}, \sigma^2_{normal})$ is the distribution function of a normally distributed random variable with mean μ_{normal} and variance σ^2_{normal}.

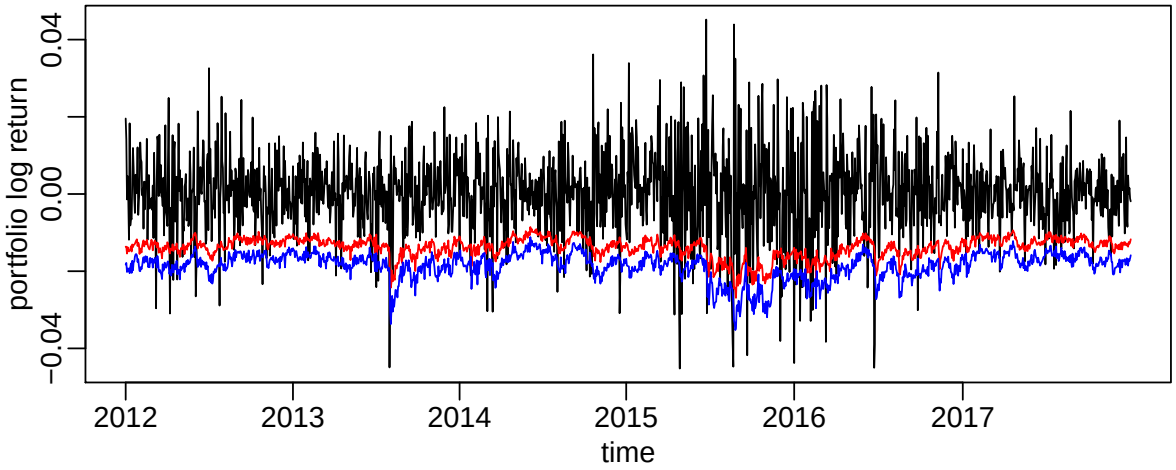

Figure 3.3: Observed daily log return of the portfolio and the estimated one-day ahead 90% VaR (red) and 95% VaR (blue) plotted against time in years.

We consider an equally weighted portfolio consisting of 6 stocks from German companies (BASF, Fresenius Medical Care, Fresenius SE, Linde, Merck, K+S). Since all companies are chosen from the chemical/pharmaceutical/medical industry, we assume that a model with one factor is suitable to capture the dependence structure. Our data, obtained from Yahoo Finance (https://finance.yahoo.com), contains daily log returns of these stocks from 2008 to 2017. We use 1000 days as training period, which corresponds to data of approximately four years. We set $T = 1000$ and obtain simulations of the one-day ahead predictive distribution as described above for the first trading day in 2012. Instead of refitting the model for each day, we fix parameters that do not change over time $(\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{m})$ at their posterior mode estimates and only update dynamic parameters $(S, \boldsymbol{v})$ for the remaining one-day ahead predicitve simulations (see Appendix F). For updating only the dynamic parameters we have seen that it is enough to use the last 100 time points and time needed for computation is reduced a lot. We obtain 2000 simulations of the one-day ahead predictive distribution for each trading day in the period

from January 2012 to December 2017. From the simulations we calculate the portfolio value, and take the corresponding quantile to obtain the VaR prediction. We consider the same VaR level of 90% as in Schamberger et al. (2017) and additionally the 95% VaR. The linking copulas are chosen from the following set of single-parameter copula families: {Gaussian, Student t with 4 degrees of freedom, (survival) Gumbel and (survival) Clayton}. With these choices, we cover a range of different tail dependence structures: no tail dependence (Gauss), symmetric tail dependence (Student t) and asymmetric tail dependence ((survival) Gumbel and (survival) Clayton). As explained above, the copula family indicator was only updated for the first model we fitted and then kept fixed. The linking copula families of BASF, Fresenius Medical Care, Fresenius SE, Linde, Merck and K+S with the highest posterior probabilities are Student t, survival Gumbel, survival Gumbel, Gaussian, survival Gumbel and Student t, respectively. In particular, we obtain a model with asymmetric tail dependence structure. Predicting the VaR for each trading day in six years results in 1521 VaR predictions. The portfolio log returns and corresponding 90% and 95% VaR predictions are visualized in Figure 3.3. We observe that the one-day ahead VaR forecast adapts to changes in the volatility.

To benchmark the proposed model (factor copula SV (**fc SV**)), we repeated the procedure for VaR prediction with two other models: marginal dynamic linear models combined with single factor copulas (**fc dlm**) estimated with a two-step procedure as proposed by Schamberger et al. (2017) and a multivariate factor stochastic volatility model with dynamic factors (**df Gauss SV**) as proposed by Kastner et al. (2017). The df Gauss SV model is here restricted to one factor. To illustrate the necessity of copula family selection, we further consider **fc SV** models with the restriction that all linking copulas are chosen from the same family. We consider the three copula families that were selected as linking copulas for the **fc SV** model and obtain the three restricted models **fc SV (Ga)**, **fc SV (t)** and **fc SV (sGu)** which have only Gaussian, Student t(df=4) and survival Gumbel linking copulas, respectively. Additionally, we compare the proposed approach to a two-step estimation of the factor copula SV model (**fc SV (ts)**). In this two-step approach we obtain simulations from the predictive distribution of the log returns at time $T+1$, $\boldsymbol{Z_{T+1}}$, given Z as follows:

- Estimate a SV model for each margin separately and obtain marginal posterior mode estimates for the latent log variances denoted by $\hat{s}_{tj}$ for $t = 1, \dots, T, j = 1, \dots, d$.
- Use the probability integral transform to obtain pseudo copula data on the $[0, 1]$ scale, $\hat{u}_{tj} = \Phi\left(z_{tj} \cdot \exp\left(-\frac{\hat{s}_{tj}}{2}\right)\right)$.
- For the data $\hat{u}_{tj}, t = 1, \dots, T, j = 1, \dots, d$, we fit the single factor copula model with HMC as explained in Section 3.2, where we allow for Bayesian copula family selection and obtain posterior mode estimates of the corresponding parameters denoted by $\hat{\tau}_1, \dots \hat{\tau}_d, \hat{m}_1, \dots \hat{m}_d$.
- For each margin, we simulate from the predictive distribution of the log variances at time $T+1$, i.e. from $s_{T+1,j}|z_{1j}, \dots, z_{Tj}$, and obtain marginal posterior mode estimates $\hat{s}_{T+1,j}$ for $j = 1, \dots, d$.

For $r = 1, \dots, R$ we proceed as follows:

- We simulate $u_1^r, \dots u_d^r$ from the single factor copula with parameters $\hat{\tau}_1, \dots \hat{\tau}_d, \hat{m}_1, \dots \hat{m}_d$.
- We set $z_{T+1,j}^r = \Phi^{-1}\left(u_j^r | 0, \exp(\hat{s}_{T+1,j})\right)$ for $j = 1, \dots, d$.

Standard measures to compare the predictive accuracy between different models are the continuous ranked probability score (Gneiting and Raftery (2007)) or log predictive scores as used in Kastner (2019). These scores evaluate the overall performance. But we are interested in the VaR, a quantile of the predictive distribution, which is only one specific aspect. Therefore we use the rate of VaR violations and the conditional coverage test (Christoffersen (2012), Chapter 13), which are commonly used to compare VaR forecasts, as in Schamberger et al. (2017) and Nagler et al. (2019). From an optimal VaR measure at level p we would expect that there are $(1-p) \cdot 100\%$ VaR violations and that violations occur independently. This constitutes the null hypotheses of the conditional coverage test. The VaR violation rates for the different models are shown in Table 3.7. For the 90% VaR, the violation rate of the df Gauss SV model is closest to the optimal rate of 10%, whereas for the 95% VaR, the fc SV model performs best. According to the p-values of the conditional coverage test in Table 3.8, only the fc SV (sGu) model can be rejected at the 5% level with respect to 90% VaR prediction. But with respect to 95% VaR prediction, every model except the fc SV model is rejected at the 5% or 10% level. We conclude that the preferred model in this scenario is the fc SV model.

	fc SV	**fc SV (Gauss)**	**fc SV (t)**	**fc SV (sGu)**	**fc SV (ts)**	**fc dlm**	**df Gauss SV**
90% VaR	9.07%	8.74%	9.27%	8.68%	10.52%	8.74%	**10.32%**
95% VaR	**4.93%**	5.39%	5.26%	4.14%	6.51%	4.80%	5.85%

Table 3.7: The rate of 90% and 95% VaR violations for the seven models: fc SV, fc SV(Gauss), fc SV(t), fc SV(sGu), fc SV(ts), fc dlm, df Gauss SV. The violation rate closest to the optimal value of 5% or 10% is marked in bold.

	fc SV	**fc SV (Gauss)**	**fc SV (t)**	**fc SV (sGu)**	**fc SV (ts)**	**fc dlm**	**df Gauss SV**
90% VaR	0.13	0.1	0.43	0.05	**0.78**	0.19	0.69
95% VaR	**0.12**	0.04	0.03	0	0.01	0.03	0.02

Table 3.8: The p-value of the conditional coverage test for the 90% and 95% VaR predictions of seven models: fc SV, fc SV(Gauss), fc SV(t), fc SV(sGu), fc SV(ts), fc dlm, df Gauss SV. The highest p-value per row is marked in bold.

3.5 Conclusion

We propose a single factor copula SV model, a combination of the SV model for the margins and factor copulas for the dependence. Dependence and marginal parameters are estimated jointly within a Bayesian approach, avoiding a two-step estimation procedure which is commonly used for copula models. The proposed model can be seen as one way to extend factor SV models that rely on Gaussian dependence to more complex dependence structures. The necessity of such models was illustrated with one-day ahead value at risk prediction. In the application our stocks were chosen such that one factor is suitable to describe dependencies. However, this might not be appropriate for different

portfolios and the extension of the proposed model to multiple factors will be subject to future research. This extension to multiple factors could exploit the partition of different stocks into sectors as in the structured factor copula model proposed by Krupskii and Joe (2015). Another extension could allow for time-varying dependence parameters or for copula families with two and more parameters.

4 Bayesian inference for nonlinear state space models with univariate autoregressive state equation

This chapter is a reproduction of Kreuzer and Czado (2019c) with minor changes.

4.1 Introduction

There are many situations where statistical models with static (time-constant) parameters are no longer sufficient to appropriately represent certain aspects of the economy. For example, it is well known that volatility of financial assets changes over time (Schwert (1989)). This is why many models that allow for variation in the parameter have been proposed. There are time-varying vector autoregressive models (Primiceri (2005), Nakajima et al. (2011)), stochastic volatility models (Kim et al. (1998)), GAM copula models (Vatter and Chavez-Demoulin (2015), Vatter and Nagler (2018)) and many more. Stochastic volatility models and the bivariate dynamic copula model of Almeida and Czado (2012) assume that the parameter follows a latent autoregressive process of order 1 (AR(1) process). These two models belong to the class of models that we will study.

In a general time-varying parameter framework, we consider a d-dimensional random variable at time t, $\boldsymbol{Y_t} \in \mathbb{R}^d$, which is generated from a d-dimensional density $f(\cdot|s_t)$. We are interested in models, where the density $f(\cdot|s_t)$ has a univariate dynamic parameter $s_t \in \mathbb{R}$ following an AR(1) process. These models can be formulated as state space models with observation equation

$$\boldsymbol{Y_t}|s_t \sim f(\boldsymbol{y_t}|s_t) \tag{4.1}$$

for $t = 1, \dots, T$. The state equation describes an AR(1) process with mean parameter $\mu \in \mathbb{R}$, persistence parameter $\phi \in (-1, 1)$ and standard deviation parameter $\sigma \in (0, \infty)$ and is given by

$$s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t, \tag{4.2}$$

where $\epsilon_t \sim N(0, 1)$ iid for $t = 1, \dots, T$ and $s_0|\mu, \phi, \sigma \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$. In the state equation we assume Gaussian innovations ϵ_t, but in the observation equation we do not put any restrictions on the density f. Thus, we allow for nonlinear and non-Gaussian state space models. Several established models can be analyzed within this framework.

By choosing $f(\cdot|s_t)$ as the univariate normal density with mean 0 and variance $\exp(s_t)$, denoted by $\varphi(\cdot|0, \exp(s_t))$, we obtain the stochastic volatility model (see Section 2.1.2)

given by

$$\begin{aligned} Y_t|s_t &\sim \varphi(y_t|0, \exp(s_t)), \\ s_t &= \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t \end{aligned} \tag{4.3}$$

for $t = 1, \ldots, T$. By modeling the log variance as a latent AR(1) process, this model allows for time-varying volatility. To allow for heavy tails and skewness, other distributions have been considered in the observation equation. One example is the stochastic volatility model with skew Student t errors (see Section 2.1.2), which can also be analyzed within our framework.

Dependence modeling is another research area, where models that allow for time-varying parameters have been introduced. Vine copulas (see Section 2.2.4) are widely used models to capture complex dependence structures. To name a few, Brechmann and Czado (2013) and Nagler et al. (2019) employ vine copulas for forecasting the value at risk of a protfolio, Aas (2016) gives an overview of applications of vine copulas in finance including asset pricing, credit risk management and portfolio optimization and Barthel et al. (2018) model the association pattern between gap times with D-vine copulas to study asthma attacks. A vine copula model is made up of different bivariate copulas with corresponding dependence parameters. Since dependencies may change over time, extensions that allow for variation in the dependence parameter have been proposed. Vatter and Chavez-Demoulin (2015) introduce a bivariate copula model, where the dependence parameter follows a generalized additive model. Another approach is the dynamic bivariate copula model proposed by Almeida and Czado (2012) and Hafner and Manner (2012), which we will analyze within our state space framework. For this model, we consider single-parameter copula families for which there is a one-to-one correspondence between the copula parameter and Kendall's τ. This allows to parametrize copula families in terms of Kendall's τ (see Section 2.2.3). The restriction of Kendall's τ to the interval $(-1, 1)$ is removed by applying the Fisher' Z transformation $F_Z(x) = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right)$. This transformed time-varying Kendall's τ is then modeled by an AR(1) process. More precisely, we consider T bivariate random vectors, $(U_{t1}, U_{t2})_{t=1,\ldots,T} \in [0, 1]^{T\times 2}$, corresponding to T time points. We assume for $t = 1, \ldots, T$ that

$$\begin{aligned} &(U_{t1}, U_{t2})|\tau_t \sim c(u_{t1}, u_{t2}; \tau_t), \\ &s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t, \text{ with } s_t = F_Z(\tau_t), \end{aligned} \tag{4.4}$$

where $c(u_{t1}, u_{t2}; \tau_t)$ is a bivariate copula density with Kendall's τ parameter τ_t.

Nonlinear state space models as specified with (4.1) and (4.2) are typically difficult to estimate, since there is a large number of parameters and likelihood evaluation requires high-dimensional integration. This often makes maximum likelihood approaches infeasible. Gibbs sampling (Geman and Geman (1984)) is a frequently used Bayesian approach to infer parameters of such nonlinear state space models (Carlin et al. (1992)). But the posterior samples, resulting from conventional Gibbs sampling, often suffer from high autocorrelation. Furthermore, the availability of the full conditional distributions or at least an efficient MCMC approach to sample from them is required. This is often tailored to specific situations. We present a Gibbs sampling approach that is designed to handle models with a latent AR(1) process and general likelihood functions as specified by the state space formulation in (4.1) and (4.2). To sample from the associated posterior distribution

we rely on elliptical slice sampling (Murray et al. (2010)) and on an ancillarity-sufficiency interweaving strategy (Yu and Meng (2011)). Elliptical slice sampling is used to sample the latent states. This allows us to exploit the Gaussian dependence structure, that is implied by the AR(1) process. But even if we provide efficient methods to sample from the full conditionals, the sampler may still suffer from the dependence among the parameters in the posterior distribution. Additionally, its performance may vary for different model parametrizations (Frühwirth-Schnatter and Sögner (2003), Strickland et al. (2008)). This problem is tackled with the ancillarity-sufficiency interweaving strategy, where the parameters of the latent AR(1) process are sampled from two different parametrizations. The decision between two parametrizations is avoided by using both. This approach has already shown good results for several models, including univariate and multivariate stochastic volatility models (Kastner and Frühwirth-Schnatter (2014), Kastner et al. (2017)). The efficiency of our proposed sampler is illustrated with a simulation study.

The second part of this chapter has a more applied focus and deals with modeling the volatility return relationship, i.e. the dependence between an index and the corresponding volatility index. More precisely, we investigate the American index S&P500 and its volatility index the VIX as well as the German index DAX and the VDAX. It is important to provide appropriate models for this relationship, since it has influence on hedging and risk management decisions (Allen et al. (2012)). For our analysis we make use of a two-step approach commonly used in copula modeling, motivated by Sklar's Theorem (see Section 2.2.1). We first model the marginal distribution with a univariate skew Student t stochastic volatility model. In the second step we model the dependence for which we propose a dynamic copula model allowing for asymmetric tail dependence. This model is a dynamic mixture of a Gumbel and a Student t copula and can be seen as an alternative to the symmetrized Joe-Clayton copula of Patton (2006). Estimation is carried out through a two-step approach, where we first estimate the marginal stochastic volatility models, fix their parameters at point estimates and then estimate the dynamic mixture copula. At both steps, estimation is straightforward with the proposed sampler. Our model is able to capture several characteristics of the joint distribution of volatility and return. With respect to the marginal distribution, we observe positive skewness for volatility indices compared to slight negative skewness for the return indices. In the dependence structure we identify asymmetry and time-variation. Finally, we compare the proposed model to several restricted models with static (time-constant) or symmetric dependence and to a bivariate DCC-GARCH model (Engle (2002)). Model comparison with respect to predictive accuracy shows the superiority of our approach.

To summarize, the main contribution of this chapter is an approach to efficiently sample from the posterior distribution of general nonlinear state space models as specified by (4.1) and (4.2). In addition, we propose a dynamic mixture copula for time-varying asymmetric tail dependence. We discuss Bayesian inference for this model class and demonstrate how it can be utilized to model the volatility return relationship.

The outline of this chapter is as follows: After the introduction, we discuss the proposed MCMC approach in Section 4.2. In Section 4.3 we investigate the efficiency of the sampler for bivariate dynamic copula models through an extensive simulation study. Section 4.4 deals with modeling the volatility return relationship and Section 4.5 concludes.

4.2 Bayesian inference

We consider the state space model as specified by (4.1) and (4.2). To obtain a fully specified Bayesian model, we equip the parameters μ, ϕ and σ with prior distributions. We employ the same prior distributions as for the latent AR(1) process of the stochastic volatility model (see Section 2.1.2), i.e.

$$\mu \sim N(0, 100^2), \quad \frac{\phi + 1}{2} \sim Beta(5, 1.5), \quad \sigma^2 \sim \chi^2_1. \tag{4.5}$$

With these prior distributions our Bayesian model is complete. For sampling from the posterior distribution of this model, we should take into account that sampling efficiency may highly depend on the model parametrization (Frühwirth-Schnatter and Sögner (2003), Strickland et al. (2008)). Yu and Meng (2011) differentiate between two parametrizations: A sufficient augmentation and an ancillary augmentation scheme. In our case a sufficient augmentation is characterized by an observation equation that is free of the parameters μ, ϕ and σ and only depends on the latent states $\boldsymbol{s_{1:T}} = (s_1, \ldots, s_T)$. In this case $\boldsymbol{s_{1:T}}$ is a sufficient statistics for the parameters μ, ϕ and σ. In an ancillary augmentation the state equation is independent of the parameters μ, ϕ and σ, then $\boldsymbol{s_{1:T}}$ is an ancillary statistics for the parameters μ, ϕ and σ. The standard parametrization of our model is already a sufficient augmentation and we refer to this parametrization as given by (4.1) and (4.2) as (SA).

$$(SA): \quad \begin{aligned} \boldsymbol{Y_t}|s_t &\sim f(\boldsymbol{y_t}|s_t), \\ s_t &= \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t. \end{aligned}$$

An ancillary augmentation is obtained by the following parametrization

$$\tilde{s}_t = \frac{s_t - \mu - \phi(s_{t-1} - \mu)}{\sigma}, \text{ with inverse } s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\tilde{s}_t, \tag{4.6}$$

for $t = 1, \ldots, T$. This reparametrization is obtained by solving Equation (4.2) for ϵ_t and implies that the state space model is given by

$$(AA): \quad \begin{aligned} &\boldsymbol{Y_t}|\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}, s_0, \mu, \phi, \sigma \sim f(\boldsymbol{y_t}|s_t(\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}, s_0, \mu, \phi, \sigma)), \\ &\tilde{s}_t \sim N(0, 1) \text{ independently}, \end{aligned}$$

where $s_t(\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}, s_0, \mu, \phi, \sigma)$ is the function that calculates s_t recursively according to (4.6) and $\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}} = (\tilde{s}_1, \ldots, \tilde{s}_T)$. We refer to this model representation as (AA). Instead of deciding between (SA) and (AA), we combine them in an ancillarity-sufficiency interweaving strategy (Yu and Meng (2011)), given by

- a) Sample $\boldsymbol{s_{0:T}}$ in (SA) from $\boldsymbol{s_{0:T}}|\mathrm{Y}, \mu, \phi, \sigma$.
- b) Sample (μ, ϕ, σ) in (SA) from $\mu, \phi, \sigma|\mathrm{Y}, \boldsymbol{s_{0:T}}$.
- c) Move to (AA) via $\tilde{s}_t = \frac{s_t - \mu - \phi(s_{t-1} - \mu)}{\sigma}$ for $t = 1, \ldots, T$.
- d) Sample (μ, ϕ, σ) in (AA) from $\mu, \phi, \sigma|\mathrm{Y}, s_0, \tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}$.
- e) Move back to (SA) via the recursion $s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\tilde{s}_t$ for $t = 1, \ldots, T$.

Here $\mathrm{Y} = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_T})^\top$ is the data matrix, where $\boldsymbol{y_t}$ is an observation of the random vector $\boldsymbol{Y_t}$, for $t = 1, \ldots, T$, and $\boldsymbol{s_{0:T}} = (s_0, \boldsymbol{s_{1:T}})$. Kastner and Frühwirth-Schnatter (2014) employed interweaving for the stochastic volatility model and showed its superior performance with an extensive simulation study. For the stochastic volatility model, Kastner and Frühwirth-Schnatter (2014) propose to move between a sufficient augmentation and a reparametrization of the latent states $\boldsymbol{s_{1:T}}$ given by $s_t^K = \frac{s_t - \mu}{\sigma}$. Within this reparametrization parameters can be sampled conveniently from its full conditional distribution by recognizing a linear regression model. This is possible for the standard stochastic volatility model, but not in our case, since our sampler is designed to handle more general likelihood functions. Therefore we have chosen the reparametrization of (SA) such that it is optimal in the sense of Yu and Meng (2011), i.e. we move between a sufficient and a ancillary augmentation.

Note that reducing the sampler to the first two steps a) and b) results in a standard Gibbs sampler in (SA). This sampler typically suffers from the dependence among the parameters μ, ϕ, σ and the latent states $\boldsymbol{s_{1:T}}$ in the posterior distribution.

Step a: Sampling of the latent states in the sufficient augmentation

To sample the latent states $\boldsymbol{s_{0:T}}$ from its full conditional in (SA) we make use of elliptical slice sampling (see Section 2.3.2). In (SA), the AR(1) structure implies, that the vector $\boldsymbol{s_{0:T}}|\mu, \phi, \sigma$ has a $(T+1)$-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu^{AR}}$ and covariance matrix Σ^{AR} given by

$$\boldsymbol{\mu^{AR}} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} \in \mathbb{R}^{T+1}, \quad \Sigma^{AR} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \ldots & \phi^T \\ \phi & 1 & \phi & \ldots & \phi^{T-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \phi^T & \phi^{T-1} & \phi^{T-2} & \ldots & 1 \end{pmatrix} \in \mathbb{R}^{(T+1)\times(T+1)}. \tag{4.7}$$

(see e.g. Brockwell et al. (2002), Chapter 2). The posterior density is proportional to

$$\left(\prod_{t=1}^{T} f(\boldsymbol{y_t}|s_t) \right) \varphi(\boldsymbol{s_{0:T}}|\boldsymbol{\mu^{AR}}, \Sigma^{AR}) \pi(\mu)\pi(\phi)\pi(\sigma),$$

where $\varphi(\cdot|\boldsymbol{\mu^{AR}}, \Sigma^{AR})$ denotes the multivariate normal density with mean vector $\boldsymbol{\mu^{AR}}$ and covariance matrix Σ^{AR} as given in (4.7) and $\pi(\cdot)$ denotes the corresponding prior density as specified in (4.5). The initial state can be sampled from its full conditional density given by

$$f(s_0|\mathrm{Y}, \boldsymbol{s_{1:T}}, \mu, \phi, \sigma) = \varphi(s_0|\mu + \phi(s_1 - \mu), \sigma^2).$$

The full conditional density of the latent states $\boldsymbol{s_{1:T}}$ is given by

$$f(\boldsymbol{s_{1:T}}|\mathrm{Y}, s_0, \mu, \phi, \sigma) \propto \left(\prod_{t=1}^{T} f(\boldsymbol{y_t}|s_t) \right) \varphi(\boldsymbol{s_{1:T}}|\boldsymbol{\mu_{1:T|0}}, \Sigma_{1:T|0}),$$

with a corresponding mean vector $\boldsymbol{\mu_{1:T|0}}$ and covariance matrix $\Sigma_{1:T|0}$. The mean vector and the covariance matrix of the conditional distribution are derived in a more general way in Appendix B.1. By reparametrizing the model with $\boldsymbol{s'_{1:T}} = \boldsymbol{s_{1:T}} - \boldsymbol{\mu_{1:T|0}}$, we impose a multivariate normal prior with zero mean. We obtain the situation elliptical slice sampling

was designed for. However, updating the whole T-dimensional vector $\boldsymbol{s_{1:T}}$ with elliptical slice sampling at once will lead to high autocorrelation in the posterior draws. This is illustrated in Section 4.3 and was also observed by Hahn et al. (2019), where elliptical slice sampling was used for linear regression models. Hahn et al. (2019) circumvent this problem by partitioning the vector $\boldsymbol{s_{1:T}}$ into smaller blocks. We follow this approach and partition the set $\{1, \dots, T\}$ into m different blocks $B_1, \dots, B_m \subset \{1, \dots, T\}$. Let $a_i = \min_{s \in B_i} s$ denote the minimal and $b_i = \max_{s \in B_i} s$ denote the maximal index in the i-th block. The blocks are chosen such that $B_i = \{t \in \{1, \dots, T\} : a_i \leq t \leq b_i\}$ and $a_i < a_j$ for $i < j$ and $i, j = 1, \dots, m$. The full conditional density for the i-th block can be expressed as $f(\boldsymbol{s_{B_i}}|\mathrm{Y}, s_0, \boldsymbol{s_{-B_i}}, \mu, \phi, \sigma) \propto \left(\prod_{t \in B_i} f(\boldsymbol{y_t}|s_t)\right) f(\boldsymbol{s_{B_i}}|s_0, \boldsymbol{s_{-B_i}}, \mu, \phi, \sigma)$, where $\boldsymbol{s_A} = (s_i)_{i \in A}$ for a set of indices A and $-B_i = \{1, \dots, T\} \setminus B_i$. The vector $\boldsymbol{s_{B_i}}|s_0, \boldsymbol{s_{-B_i}}, \mu, \phi, \sigma$ is multivariate normal distributed with mean denoted by $\boldsymbol{\mu_{B_i|}}$ and covariance matrix $\Sigma_{B_i|}$ (see Appendix B.1) and therefore the full conditional density can be written as

$$f(\boldsymbol{s_{B_i}}|\mathrm{Y}, s_0, \boldsymbol{s_{-B_i}}, \mu, \phi, \sigma) \propto \left(\prod_{t \in B_i} f(\boldsymbol{y_t}|s_t)\right) \varphi(\boldsymbol{s_{B_i}}|\boldsymbol{\mu_{B_i|}}, \Sigma_{B_i|}).$$

To sample the latent states of the i-th block, $\boldsymbol{s_{B_i}}$, from its full conditional, we proceed as follows

- Set $\boldsymbol{s'_{B_i}} = \boldsymbol{s_{B_i}} - \boldsymbol{\mu_{B_i|}}$
- Draw $\boldsymbol{s'_{B_i}}$ from the density

  $$f(\boldsymbol{s'_{B_i}}|\mathrm{Y}, s_0, \boldsymbol{s_{-B_i}}, \mu, \phi, \sigma) \propto \left(\prod_{t \in B_i} f(\boldsymbol{y_t}|s_t)\right) \varphi(\boldsymbol{s'_{B_i}}|\boldsymbol{0}, \Sigma_{B_i|}),$$

  using elliptical slice sampling, where $\varphi(\boldsymbol{s'_{B_i}}|\boldsymbol{0}, \Sigma_{B_i|})$ is interpreted as the prior density for $\boldsymbol{s'_{B_i}}$.
- Set $\boldsymbol{s_{B_i}} = \boldsymbol{s'_{B_i}} + \boldsymbol{\mu_{B_i|}}$

Step b: Sampling of the static parameters in the sufficient augmentation

In (SA) the observation equation only depends on $\boldsymbol{s_{1:T}}$ and is independent of the parameters μ, ϕ and σ. The parameters μ, ϕ and σ only depend on $\boldsymbol{s_{0:T}}$. This allows to use the same approach as in Kastner and Frühwirth-Schnatter (2014) to sample the parameters μ, ϕ and σ in (SA). We reparametrize the model such that proposals can be found using Bayesian linear regression. We define $\gamma = \mu(1 - \phi)$ and the state equation is given by

$$s_t = \gamma + \phi s_{t-1} + \sigma \epsilon_t,$$

where $\epsilon_t \sim N(0, 1)$. For fixed $\boldsymbol{s_{0:T}}$, this is a linear regression model with regression parameters γ, ϕ and variance σ^2. Proposals for (μ, ϕ, σ) are found and accepted or rejected as described in Kastner and Frühwirth-Schnatter (2014), Section 2.4 (two block sampler).

Step d: Sampling of the static parameters in the ancillary augmentation

To sample μ, ϕ and σ in (AA), we deploy a random walk Metropolis-Hastings scheme with Gaussian proposal, where the proposal variance or covariance matrix is adapted during the burn-in period. For the adaptions we use the Robbins Monro process (Robbins and Monro (1985)) as suggested by Garthwaite et al. (2016). More details are given in Appendix B.1.

Implementation

For the implementation of the sampler we use `Rcpp` (Eddelbuettel et al. (2011)) which allows to embed C++ code into `R`. In addition, we make use of `rvinecopulib` (Nagler and Vatter (2018)) to evaluate copula densities and of `RcppEigen` (Bates et al. (2013)). For sampling (μ, ϕ, σ) in (SA) we use corresponding parts of the implementation of the `R` package `stochvol` (Kastner (2016)). The `R` package `coda` (Plummer et al. (2008)) is used to compute effective sample sizes in the following section.

4.3 Illustration of the proposed sampler for bivariate dynamic copula models

We illustrate the MCMC sampler, we proposed in the previous section, for the bivariate dynamic copula model of Almeida and Czado (2012). Kastner and Frühwirth-Schnatter (2014) have already shown that interweaving improves sampling efficiency a lot for the stochastic volatility model. We investigate if this is also the case for the bivariate dynamic copula model. Further, we study how the sampling efficiency is affected by the chosen block size and by the data generating process (DGP). Therefore, we perform an extensive simulation study. We consider different modifications of the sampler. A sampler is specified by a vector (b,i) which indicates its blocksize (b) and if interweaving is used (i=I) or not (i=NI). We consider 10 different sampler specifications (b,i) $\in \{1, 5, 20, 100, T\} \times \{\text{I}, \text{NI}\}$, where T is the length of the time series. By using blocks of size T, we obtain the sampler which updates the parameters $\boldsymbol{s_{1:T}}$ jointly with elliptical slice sampling. If we turn off interweaving (i=NI), we obtain a standard Gibbs sampler updating parameters in the sufficient augmentation. These samplers are run for different simulated data sets. A data set is simulated from the bivariate dynamic copula model (see (4.4)) with parameters: Family, T, μ, ϕ, σ. The parameters are chosen from the following grid (Family, T, μ, ϕ, σ) $\in \{\text{Gauss}, \text{eClayton}\} \times \{500, 1000, 1500\} \times \{0, 1\} \times \{0, 0.1, 0.5, 0.9, 0.99\} \times \{0.05, 0.1, 0.2\}$. The Gaussian copula is an elliptical copula with symmetric tails, whereas the eClayton copula is an Archimedean copula with asymmetric tail dependence (see Section 2.2.3). Among the different DGPs, most distinct values are considered for ϕ. We expect its choice to be influential, since it controls the dependence among the latent variables. With this grid, we obtain 180 different DGPs. For each of the different DGPs, we generate 100 simulated data sets and for each data set we run the 10 different samplers with the correctly specified copula family for 25000 iterations and discard the first 5000 iterations for burn-in. So, in total we obtain 18000 simulated data sets and each of the 10 samplers is run 18000 times.

Figure 4.1 shows trace plots of different parameters (μ, ϕ, σ, s_{300}) based on one sim-

ulated data set for three different sampler specifications. We consider specification (5,I), and the same specification with interweaving turned off, i.e. (5,NI) and the specification (T,I). We observe that all three samplers produce posterior samples covering the true values. Further, the trace plots suggest that the (5,I) sampler achieves better mixing than the other two considered samplers.

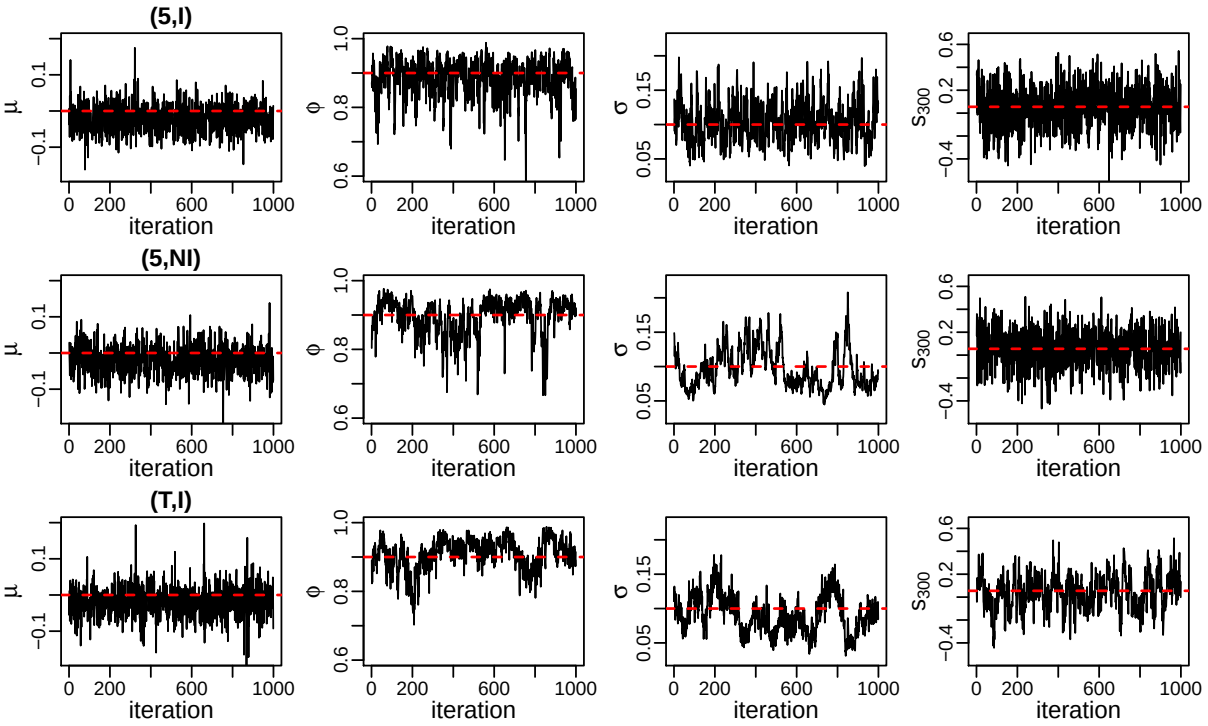

Figure 4.1: Trace plots of 1000 MCMC draws based on a total of 25000 iterations, where the first 5000 draws are discarded for burn-in and the remaining 20000 draws are thinned with factor 20. The trace plots are shown for the parameters μ, ϕ, σ and s_{300} for three different sampler specifications: (5,I) (top row), (5,NI) (middle row), (T,I) (bottom row). The corresponding data was generated from the following DGP: $T = 1000$, Family=eClayton, $\mu = 0$, $\phi = 0.9$, $\sigma = 0.1$. True values are added in red (dashed).

The runtime of the sampler is mainly affected by the choice of T and the sampler specification. From Table 4.1 we see that the runtime is increasing in T and that interweaving adds considerable additional runtime.

	(1,I)	(5,I)	(20,I)	(100,I)	(T,I)	(1,NI)	(5,NI)	(20,NI)	(100,NI)	(T,NI)
$T = 500$	0.6	0.5	0.5	0.7	0.8	0.3	0.3	0.3	0.4	0.5
$T = 1000$	1.1	1.0	1.1	1.3	1.6	0.7	0.5	0.6	0.8	1.1
$T = 1500$	1.7	1.4	1.6	1.9	2.6	1.0	0.7	0.9	1.2	1.8

Table 4.1: Average runtime in minutes for 25000 draws. We consider averages for different sampler specifications and different values of T. The sampler was run on a Linux cluster with CPU Intel Xeon E5-2697 v3.

To measure efficiency of the samplers, we consider the effective sample size per minute which we call effective sampling rate, similar to Hosszejni and Kastner (2019). We average

the effective sampling rate of 100 runs, where the same sampler and the same DGP was used. Then we obtain 180 average effective sampling rates (AESR) per sampler. Since the AESR decreases for higher values of T for every sampler, we compare the 10 different samplers among DGPs with the same value for T. For a fixed T we have 60 AESR values per sampler. For each sampler, the minimum of these 60 values (mAESR) is given in Table 4.2. We consider the minimum since we are interested in samplers that are reliable for all DGPs. We see that for the parameters μ, ϕ and σ sampler specifications with interweaving have higher mAESR values, while for the latent states specifications without interweaving perform better. Although interweaving adds additional runtime, it still increases the mAESR of μ, ϕ and σ considerably. Further, we observe that choosing the blocksize too big results in very low mAESR values for the latent states. In this case many parameters are updated jointly with elliptical slice sampling which results in high autocorrelation among consecutive draws. The performance of samplers with blocksize T is especially poor. For these samplers the length of the time series ($T = 500$, 1000, 1500) also has strong effects. For example the mAESR for μ for specification (T,NI) decreases by 85 % from 78 to 11 when the length of the time series is increased from 500 to 1500. The most inefficient sampler is (T,NI), the sampler without interweaving and with the largest blocksize. In our opinion, the best results are obtained for sampler specification (5,I). It provides the highest mAESR values for μ, ϕ and σ for all choices of T and also provides rather high mAESR values for the latent states.

	(1,I)	(5,I)	(20,I)	(100,I)	(T,I)	(1,NI)	(5,NI)	(20,NI)	(100,NI)	(T,NI)
					$T = 500$					
μ	2829	3178	1813	642	381	437	838	780	277	78
ϕ	471	749	556	191	60	266	326	232	78	24
σ	272	411	263	58	33	53	74	56	23	6
s(a)	938	2982	2484	199	36	1347	4942	3908	273	49
s(m)	346	1092	441	35	5	496	1011	627	41	5
					$T = 1000$					
μ	1167	1365	840	309	180	172	330	371	121	21
ϕ	202	269	212	73	17	81	97	76	27	6
σ	97	175	133	25	9	18	25	21	12	1
s(a)	298	1011	1312	99	12	400	1911	1982	133	16
s(m)	115	417	200	15	1	161	559	296	17	2
					$T = 1500$					
μ	711	833	578	200	111	89	218	231	75	11
ϕ	120	145	107	43	9	38	49	42	14	3
σ	63	109	84	16	3	10	14	13	8	1
s(a)	162	583	879	66	7	247	1168	1359	90	10
s(m)	61	239	112	9	1	95	426	176	11	1

Table 4.2: For different lengths of the time series ($T = 500, 1000, 1500$), the minimum of 60 AESR values (mAESR), corresponding to 60 different DGPs, is shown for different parameters and 10 different sampler specifications. In the $s(a)$ row we calculate the mAESR based on the average of the effective sampling rates of $s_0, \ldots, s_T$, while in the $s(m)$ row the mAESR values are calculated based on the minimum of the effective sampling rates of $s_0, \ldots, s_T$.

In addition to the previous analysis, we investigate how different DGPs affect the samplers. Therefore we consider the best sampler according to Table 4.2, i.e. sampler

specification (5,I). In addition, we have a look at the same sampler specification without interweaving (5,NI) and the same sampler with joint updates of the latent states, i.e. (T,I). We consider the AESR for σ, which is usually the parameter which causes most problems. Table 4.3 shows for each of these three samplers the DGPs with $T = 1000$ which resulted in the lowest and the highest AESR for σ. For the (5,I) specification, satisfactory AESR values, ranging from 175 to 456, are obtained for all DGPs. In the (T,I) specification the latent states are updated jointly. In scenarios with strong dependence among the latent states ($\phi = 0.99$) this sampler performs poorly, whereas the best performance of the sampler was seen for a DGP with low dependence among the latent states (ϕ=0). The (5,NI) specification is a standard Gibbs sampler in the sufficient augmentation. We see that for this specification the AESR values vary a lot, ranging from 25 to 558. The best performance of this sampler specification was seen for a DGP with high persistence ($\phi = 0.99$).

		(5,I)		(T,I)		(5,NI)	
AESR(σ)		min	max	min	max	min	max
		175	456	9	121	25	558
DGP	family	Gauss	eClayton	eClayton	eClayton	Gauss	eClayton
	μ	0	1	0	0	0	0
	ϕ	0.1	0.9	0.99	0	0.1	0.99
	σ	0.2	0.2	0.2	0.05	0.05	0.2

Table 4.3: For three sampler specifications ((5,I), (T,I), (5,NI)) we show the DGPs with $T = 1000$ which resulted in the highest and in the lowest AESR values for σ, respectively. The corresponding AESR is also shown.

DGP				AESR(μ)		AESR(ϕ)		AESR(σ)	
family	μ	ϕ	σ	(5,I)	CG	(5,I)	CG	(5,I)	CG
Gauss	1	0.9	0.1	8187	2130	527	81	393	50
Gauss	1	0.1	0.2	1637	305	364	83	437	69
Gauss	0	0.9	0.2	9935	4150	549	150	316	85
eClayton	1	0.9	0.1	8382	2088	539	74	397	46
eClayton	1	0.1	0.2	1778	279	378	71	416	52
eClayton	0	0.9	0.2	10079	4375	574	175	323	97

Table 4.4: Comparison of sampler specification (5,I) and the coarse grid sampler (CG) of Almeida and Czado (2012) for six different DGPs with $T = 1000$ with respect to the AESR of μ, ϕ and σ.

Lastly, we compare our sampler to the coarse grid sampler employed by Almeida and Czado (2012). We already covered six DGPs that were also analyzed by Almeida and Czado (2012). Instead of running their sampler, we make the comparison with respect to these six cases. There are several points which make the comparison slightly less reliable. First, Almeida and Czado (2012) did not report exact computation times but they note that 100 000 iterations of their sampler take about 15 minutes. We use this number to calculate the AESR values from the effective sample sizes they report in their paper. Second, their calculations were performed on a different computer and third they use different prior distributions for ϕ and σ^2. But we think that this comparison should still give us a rough idea of how the sampling efficiencies compare to each other. From Table

4.4 we see that the (5,I) specification considerably outperforms the coarse grid sampler (CG). For every parameter (μ, ϕ, σ) we obtain way higher AESR values.

4.4 Application: Modeling the volatility return relationship

We investigate the volatility return relationship through the bivariate joint distribution of a stock index and the corresponding volatility index. The joint distribution of return and volatility incorporates all the marginal information as well as information about the dependence, which are both relevant for hedging and risk management (Allen et al. (2012)). Since there has already been evidence for asymmetry in the joint distribution of volatility and return (Allen et al. (2012), Fink et al. (2017)), models which are able to handle such characteristics are necessary.

The two-step copula modeling approach motivated by Sklar's theorem (see Section 2.2.1) provides a very flexible method for the construction of multivariate distributions. We can combine arbitrary marginal distributions with any copula. Here, we propose a bivariate model that combines the skew Student t stochastic volatility model for the margins with a novel dynamic mixture copula. This model allows for asymmetry and heavy tails in the marginal distribution as well as for time-varying asymmetric tail dependence in the dependence structure. Both, the marginal as well as the copula model can be estimated with the proposed sampler.

Marginal model

We utilize the stochastic volatility model with skew Student t errors (see Section 2.1.2). The model is given by

$$\begin{aligned} Y_t &= \exp\left(\frac{s_t}{2}\right)\epsilon_t^{sst} \\ s_t &= \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t, \end{aligned} \tag{4.8}$$

where $\epsilon_t^{sst}|\alpha, df \sim sst(\epsilon_t^{sst}|\alpha, df)$ independently for $t = 1, \ldots, T$ and $sst(\epsilon_t^{sst}|\alpha, df)$ denotes the density of the standardized skew Student t distribution with parameters $\alpha \in \mathbb{R}$ and $df > 2$. Compared to our framework, this model has two additional parameters, α and df. For these additional parameters we choose the following prior distributions (see Section 2.1.2)

$$\alpha \sim N(0, 100), \qquad df \sim N_{>2}(5, 25). \tag{4.9}$$

Conditional on α and df, our sampler can be applied directly to sample $(\mu, \phi, \sigma, \boldsymbol{s_{0:T}})$ from its full conditional. Another approach, which lead to better mixing, is to include the parameters α and df in the interweaving strategy. The sampler is slightly modified in the following way:

- a) Sample $\boldsymbol{s_{0:T}}$ from $\boldsymbol{s_{0:T}}|\mathrm{Y}, \mu, \phi, \sigma, \alpha, df$.
- b) Sample $(\mu, \phi, \sigma, \alpha, df)$ in (SA) from $\mu, \phi, \sigma, \alpha, df|\mathrm{Y}, \boldsymbol{s_{0:T}}$.
- c) Move to (AA) via $\tilde{s}_t = \frac{s_t - \mu - \phi(s_{t-1}-\mu)}{\sigma}$, for $t = 1, \ldots, T$.
- d) Sample $(\mu, \phi, \sigma, \alpha, df)$ in (AA) from $\mu, \phi, \sigma, \alpha, df|\mathrm{Y}, s_0, \boldsymbol{\tilde{s}_{1:T}}$.
- e) Move back to (SA) via the recursion $s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\tilde{s}_t$ for $t = 1, \ldots, T$.

For step a) we proceed as described in Section 4.2. For step b) we draw α and df from its univariate full conditional distributions using Metropolis-Hastings, similar to Step d) in Section 4.2. The parameters (μ, ϕ, σ) are drawn from its full conditional as described in Section 4.2. For step d) we investigated different blocking strategies for the parameters $(\mu, \phi, \sigma, \alpha, df)$. We compared the different strategies with respect to effective sample sizes and decided to use the following three blocks: $(\mu, df), (\phi, \sigma)$ and α. Each block is updated using Metropolis-Hastings as in Step d) in Section 4.2.

Dependence model

Dependence among financial assets is often modeled with a Student t copula. This copula allows for tail dependence symmetric in the upper and lower tail. Evidence against the assumption of symmetric tail dependence has been provided and models to handle this characteristic have become necessary (Patton (2006), Nikoloulopoulos et al. (2012), Jondeau (2016)). Patton (2006) proposes the symmetrized Joe-Clayton copula. This is a modification of the BB7 copula (Joe (2014), Chapter 4) that is symmetric if upper and lower tail dependence coincide, which he describes as a desirable property. In the application of Nikoloulopoulos et al. (2012) the Student t copula provides the best fit in terms of the likelihood. But they argue that if the focus is on the tails, a BB1 or BB7 copula might be more appropriate. The BB1 and BB7 copulas have two parameters which might not be enough, if we want to model three characteristics in a flexible way: upper tail dependence, lower tail dependence and overall dependence as measured with Kendall's τ. We provide another approach to relax the symmetric tail dependence assumption. We propose a mixture of a Student t and an extended Gumbel copula with parameters $\tau \in (-1, 1)$, $\nu > 2$ and $p \in [0, 1]$ given by

$$\mathbb{C}^M(u_1, u_2; \tau, \nu, p) = p\mathbb{C}^t(u_1, u_2; \tau, \nu) + (1-p)\mathbb{C}^G(u_1, u_2; \tau), \tag{4.10}$$

where $\mathbb{C}^t$ is the bivariate Student t copula specified by Kendall's τ and the degree of freedom ν and $\mathbb{C}^G$ is the bivariate extended Gumbel copula specified by Kendall's τ (see Section 2.2.3). Both copulas $\mathbb{C}^t$ and $\mathbb{C}^G$ share the dependence parameter τ and we expect the mixture copula to have a similar strength of dependence. We evaluated the Kendall's τ of the mixture copula, τ^M, for different values of τ, p and ν numerically and observed only negligible difference between τ and τ^M. The lower and upper tail dependence coefficients

λ_M^L and λ_M^U of the mixture copula can, for $\tau > 0$, be obtained as

$$\begin{aligned}
\lambda_M^L(\tau, p, \nu) &= \lim_{u \to 0^+} \frac{\mathbb{C}^M(u,u)}{u} = \lim_{u \to 0^+} \frac{p\mathbb{C}^t(u,u;\tau,\nu) + (1-p)\mathbb{C}^G(u,u;\tau)}{u} \\
&= p2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\sin\left(\pi\frac{\tau}{2}\right))}{1+\sin\left(\pi\frac{\tau}{2}\right)}}\right) + 0, \\
\lambda_M^U(\tau, p, \nu) &= p2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\sin\left(\pi\frac{\tau}{2}\right))}{1+\sin\left(\pi\frac{\tau}{2}\right)}}\right) + (1-p)(2-2^{1-\tau}),
\end{aligned}$$

where we used the well known formulas for the tail dependence coefficients of the Student t and the Gumbel copula (Joe (2014), Chapter 4) and $T_{\nu+1}(\cdot)$ denotes the distribution function of the Student t distribution with $\nu + 1$ degrees of freedom. Whereas the upper and lower tail dependence coefficients measure dependence in the upper right and lower left corner, we are also interested in the dependence in the upper left and the lower right corner when $\tau < 0$. We consider the following tail dependence coefficients in the upper left corner λ_M^{UL} and in the lower right corner λ_M^{LR} if $\tau < 0$

$$\lambda_M^{LR} := \lambda_M^L(-\tau, p, \nu), \quad \lambda_M^{UL} := \lambda_M^U(-\tau, p, \nu)$$

analogous to the definition of quarter tail dependence in Fink et al. (2017).

The tail dependence coefficient of the mixture copula is a linear combination of the tail dependence coefficients of its two components, the Student t and the Gumbel copula. The Student t copula has symmetric tail dependence, whereas the Gumbel copula has upper but no lower tail dependence. So we expect upper tail dependence to be higher than lower tail dependence in the mixture copula. The amount of asymmetry in the tails is controlled by p, whereas the copula is symmetric in the tails for $p = 1$ and the level of asymmetry increases as we decrease p. So this copula allows for great flexibility: The overall dependence can be described by Kendall's τ, the degrees of freedom parameter controls the upper and lower tail dependence coefficient and p controls the difference between upper and lower tail dependence. This is visualized in Figure B.2 in Appendix B.2. Note that the desirable property according to Patton (2006) of symmetry in case of coinciding upper and lower tail dependence is here fulfilled. If we expected higher lower than upper tail dependence, we can replace the Gumbel copula by a survival Gumbel copula which has the density $c^{SG}(u_1, u_2) = c^G(1-u_1, 1-u_2)$. To allow for time-variation, we use the mixture copula $\mathbb{C}^M$ of (4.10) within the dynamic bivariate copula model of Almeida and Czado (2012). A nonlinear state space model for T bivariate random vectors $(U_{t1}, U_{t2})_{t=1,\dots,T} \in [0,1]^{T\times 2}$, corresponding to T time points, is given by

$$\begin{aligned}
&(U_{t1}, U_{t2}) \sim c^M(u_{t1}, u_{t2}; \tau_t, \nu, p) \\
&s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\epsilon_t \quad \text{with} \quad s_t = F_Z(\tau_t)
\end{aligned} \tag{4.11}$$

for $t = 1, \dots, T$. We assign a vague uniform prior on $[0,1]$ for p, a vague normal prior with mean 5 and standard deviation 20 truncated to the interval $(2, \infty)$ for ν and the same priors as in (4.5) for the remaining parameters. Sampling is done in the following way:

- Draw $\ln\left(\frac{p}{1-p}\right)$ and $\ln(\nu-2)$ from its univariate full conditionals with random walk Metropolis-Hastings with Gaussian proposal (proposal standard deviation: 0.3).
- Draw $\mu, \phi, \sigma, \boldsymbol{s_{0:T}}$ conditioned on p and ν as in Section 4.2.

Two-step estimation

We consider the S&P500 (SPX) and its volatility index the VIX as well as the DAX and its volatility index the VDAX. The daily log returns from 2006 to 2013 of these indices are obtained from Yahoo finance (`https://finance.yahoo.com`). With approximately 250 trading days per year, this results in 2063 observations, visualized in Figure B.3 in Appendix B.3. The corresponding data matrix with 2063 rows and 4 columns is denoted by $\mathrm{Y} = (y_{tj})_{t=1,\ldots,2063, j=1,\ldots,4}$.

Combining the marginal and the dependence model, we obtain that for T bivariate random vectors $(Y_{t1}, Y_{t2})_{t=1,\ldots,T} \in \mathbb{R}^{T\times 2}$ the following holds

$$(Y_{t1}, Y_{t2})|s_{t1}^{st}, \alpha_1^{st}, df_1^{st}, s_{t2}^{st}, \alpha_2^{st}, df_2^{st}, s_t^{cop}, \nu^{cop}, p^{cop} \sim$$
$$\mathbb{C}^M\left(ssT\left(\frac{y_{t1}}{\exp(s_{t1}^{st}/2)}\middle|\alpha_1^{st}, df_1^{st}\right), ssT\left(\frac{y_{t2}}{\exp(s_{t2}^{st}/2)}\middle|\alpha_2^{st}, df_2^{st}\right); F_Z^{-1}(s_t^{cop}), \nu^{cop}, p^{cop}\right) \tag{4.12}$$

$$s_{tj}^{st} = \mu_j^{st} + \phi_j^{st}(s_{t-1;j}^{st} - \mu_j^{st}) + \sigma_j^{st}\epsilon_{tj}^{st},$$
$$s_t^{cop} = \mu^{cop} + \phi^{cop}(s_{t-1}^{cop} - \mu^{cop}) + \sigma^{cop}\epsilon_t^{cop},$$

where $\epsilon_{tj}^{st}, \epsilon_t^{cop} \sim N(0,1)$ iid, α_j^{st}, df_j^{st} as in (4.9), ν^{cop}, p^{cop} as in (4.11) and μ_j^{st}, μ^{cop}, ϕ_j^{st}, ϕ^{cop}, σ_j^{st}, σ^{cop}, s_{0j}^{st}, s_0^{cop} as in (4.2) and (4.5) for $j = 1, 2$ and $t = 1, \ldots T$. Here, ssT denotes the distribution function of the standardized skew Student t distribution (see Section 2.1.2). We refer to the probability integral transforms $ssT\left(y_{t1}\exp(-s_{t1}^{st}/2)\middle|\alpha_1^{st}, df_1^{st}\right)$ and $ssT\left(y_{t2}\exp(-s_{t2}^{st}/2)\middle|\alpha_2^{st}, df_2^{st}\right)$ for $t = 1, \ldots, T$ as copula data.

For inference we rely on a two-step approach. We first estimate marginal distributions and based on the resulting estimated copula data we estimate the copula parameters. This approach is also called inference for margins (see Section 2.2.1) and is commonly used in (Bayesian) copula modeling (Min and Czado (2011), Almeida and Czado (2012), Smith (2015), Gruber and Czado (2015), Loaiza-Maya et al. (2018)).

First, we fit a skew Student t stochastic volatility model for each of the indices. For each index we run the sampler (5,I) for 31000 iterations and discard the first 1000 draws as burn-in. As it is typical for financial data, all indices show a high persistence parameter ϕ (Posterior mode estimates for ϕ: SPX: 0.99, VIX: 0.90, DAX: 0.99, VDAX: 0.96). A notable difference is that for stock indices we observe negative skewness, whereas for the volatility indices positive skewness is observed (Posterior mode estimates for α: SPX: -0.51, VIX: 1.33, DAX: -0.48, VDAX: 0.96). Evidence for negative skewness has also been observed for the log returns of other stock indices, as e.g., for the NASDAQ by Abanto-Valle et al. (2015). Estimated posterior modes, posterior quantiles and effective samples sizes for several parameters of the four marginal models are summarized in Table B.1 in Appendix B.3. The estimated daily log variances are shown in Figure 4.2. In the end of 2008, the estimated variances are high for all indices due to the financial crisis.

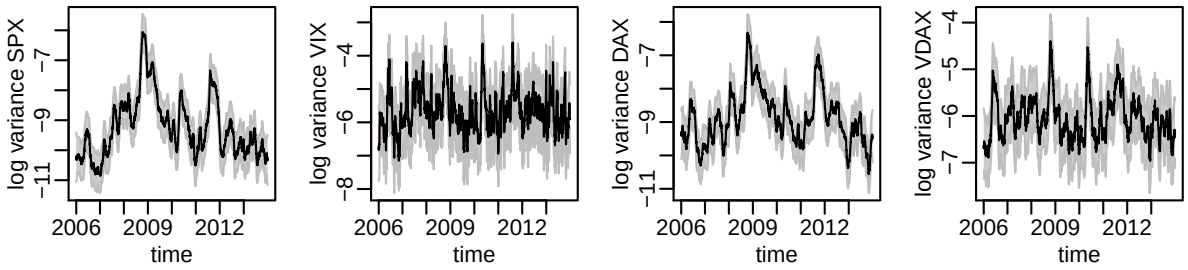

Figure 4.2: For each of the four skew Student t stochastic volatility models for the indices SPX, VIX, DAX, VDAX, posterior mode estimates (obtained from univariate kernel density estimates) of the daily log variances are plotted against time. The 90% credible region, added in grey, is constructed from the estimated 5% and 95% posterior quantiles.

In the next step we obtain data on the [0,1] scale by applying the probability integral transform using the posterior mode estimates (obtained from univariate kernel density estimates) of the marginal parameters. We refer to this data as pseudo copula data and it is obtained as

$$\hat{u}_{tj} = ssT\left(y_{tj}\exp\left(-\frac{\hat{s}^{st}_{tj}}{2}\right)\middle|\hat{\alpha}^{st}_{j}, \hat{df}^{st}_{j}\right),$$

where $\hat{s}^{st}_{tj}, \hat{\alpha}^{st}_{j}, \hat{df}^{st}_{j}$ are the posterior mode estimates of the corresponding marginal skew Student t stochastic volatility model for $t = 1, \ldots, T, j = 1, \ldots, 4$. In the copula data marginal characteristics are removed and what is left is information about the dependence structure. Based on the pseudo copula data, two dynamic mixture copula models are fitted, one corresponding to the pair (SPX,VIX) and one corresponding to the pair (DAX,VDAX). For each pair we obtain 31000 iterations with the sampler specification (5,I) and discard the first 1000 draws as burn-in. The posterior mode estimates for p are 0.29 for the model for (SPX,VIX) and 0.66 for the model for (DAX,VDAX), respectively. (Further, posterior statistics for the model parameters μ, ϕ, σ, p and ν are shown in Table B.2 in Appendix B.3). So both fitted models allow for asymmetric tail dependence, whereas the asymmetry is stronger for the (SPX,VIX) model. For these models tail dependence in the upper left corner λ^{UL}_{M} is stronger than the one in the lower right corner λ^{LR}_{M}. This means that joint extreme comovements, where the stock index decreases and the volatility index increases are more likely to occur than vice versa, which agrees with the statement that the market reacts more extreme in bad market situations (Sun and Wu (2018)).

The time-varying estimates for Kendall's τ and the tail dependence coefficients are shown in Figure 4.3. The figure visualizes the asymmetry in tail dependence. We also observe changes in tail dependence as time evolves: for (SPX,VIX), λ^{UL}_{M} ranges from 0.41 to 0.71 and λ^{LR}_{M} from 0.004 to 0.10 . For the pair (DAX,VDAX), λ^{UL}_{M} ranges from 0.12 to 0.67 and λ^{LR}_{M} from 0.003 to 0.35. This variation over time in tail dependence goes hand in hand with variation in Kendall's τ. For (SPX,VIX), Kendall's τ ranges from -0.76 to -0.48 and for (DAX,VDAX) from -0.80 to -0.30.

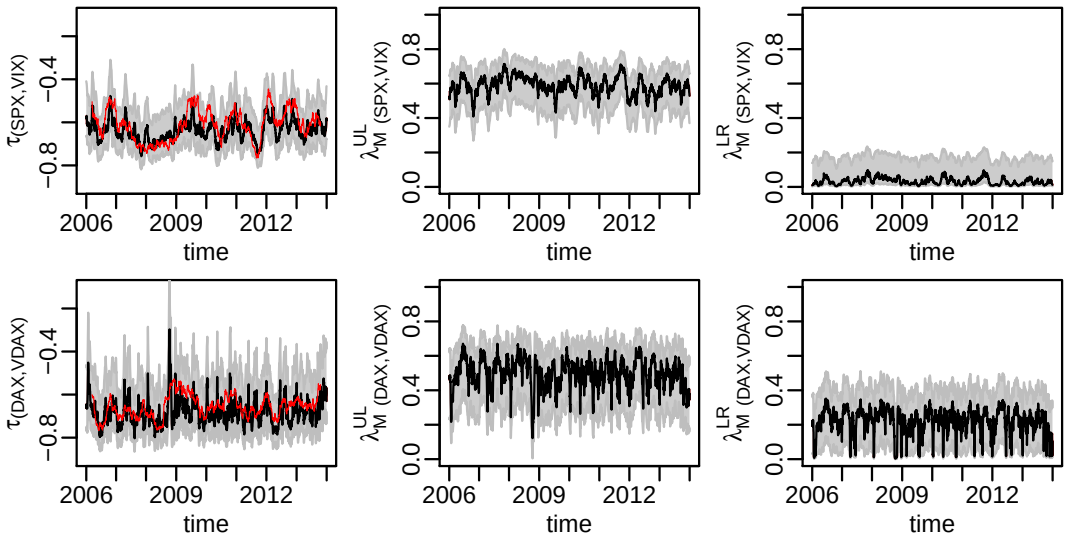

Figure 4.3: In these plots the top row corresponds to the pair (SPX,VIX), the bottom row to the pair (DAX,VDAX). The first column shows posterior mode estimates of Kendall's τ, where the Kendall's τ obtained from rolling window estimates (Kendall's τ at time t is estimated as empirical Kendall's τ based on the 50 observations before and after time t) is added in red. The middle column shows posterior mode estimates of λ_M^{UL} and the right column shows estimates of λ_M^{LR}. Credible regions, constructed from the estimated 5% and 95% posterior quantiles, are added in grey.

Out-of-sample predictions

We aim to further support the findings we obtained through the dynamic copula model, i.e. that the dependence structure is asymmetric and varies over time. Therefore we consider several restrictions with respect to the dependence structure. Giving up time-variation leads a static mixture copula, giving up asymmetry leads to a dynamic Student t copula and giving up time-variation and asymmetry leads to a static Student t copula. In addition, we compare our model to the frequently used dynamic conditional correlation (DCC) GARCH model of Engle (2002). The DCC-GARCH allows for time-varying symmetric dependence. So we take five different models into consideration. These models are summarized in Table 4.5. The model $\mathcal{M}_{dyn}^{mix}$ is the model given in (4.12), which was also used for the previous analysis.

We compare the models with respect to pseudo log predictive scores (Kastner (2019)), which are obtained by evaluating the corresponding density at point estimates, instead of averaging over all posterior draws. In comparison to other multivariate scoring rules, such as the energy score or the variogram score (Scheuerer and Hamill (2015)), pseudo log predictive scores have here the advantage that they can be computed fast, since they require only one evaluation of the density per observation. We consider $T+K$ observations of dimension two, stored in the data matrix $\mathrm{Y}_{1:(T+K);1:2} = (y_{tj})_{t=1,\ldots,T+k,j=1,2}$, where the first T observations are used to train the model and the last K are used for testing.

The DCC-GARCH model is estimated based on T data points in the training period. Based on this model, we obtain rolling one-day ahead forecasts of the covariance matrix

for each day in the test set. The pseudo log predictive scores are obtained by evaluating the corresponding multivariate normal log densities at the observations. More precisely, we denote by Σ^{T+k} the one-day ahead forecast of the covariance matrix for day $T+k$ with $1 \leq k \leq K$ and the pseudo log predictive score for day $T+k$ is obtained as $\ln\left(\varphi\left((y_{T+k;1}, y_{T+k;2})|\mathbf{0}, \Sigma^{T+k}\right)\right)$. To estimate the DCC-GARCH model and to obtain the rolling one-day ahead forecasts of the covariance matrix we used the R package `rmgarch` of Ghalanos (2012).

model	specification	margin	dependence	
		asymmetric	asymmetric	dynamic
$\mathcal{M}_{dyn}^{mix}$	sstSV + dynamic mixture copula	Yes	Yes	Yes
$\mathcal{M}_{static}^{mix}$	sstSV + static mixture copula	Yes	Yes	No
$\mathcal{M}_{dyn}^{t}$	sstSV + dynamic Student t copula	Yes	No	Yes
$\mathcal{M}_{static}^{t}$	sstSV + static Student t copula	Yes	No	No
$\mathcal{M}_{DCC}$	DCC(1,1)-GARCH(1,1)	No	No	Yes

Table 4.5: Different models considered for comparison. Models are specified by: marginal model + copula model. The skew Student t stochastic volatility model as given in (4.8) is denoted by sstSV. The mixture copula is defined in (4.10). If a copula is dynamic, the corresponding copula is considered within the dynamic bivariate copula model framework of Almeida and Czado (2012) given in (4.4).

Similarly, the model $\mathcal{M}_{dyn}^{mix}$ is estimated with the training data. Instead of computing daily updates for all model parameters, we fix the static parameters at their posterior mode estimates to save computation time, as in Section 3.4. The dynamic parameters (i.e. parameters with an index t) are updated daily and the one-day ahead forecasts of the dynamic parameters are obtained by evolving the AR(1) process. To obtain the pseudo log predictive score for time point $T+k$ with $1 \leq k \leq K$, we evaluate the log density implied by (4.12) at the corresponding observation $(y_{T+k;1}, y_{T+k;2})$, using the one-day ahead forecasts for day $T+k$ for the dynamic parameters and the estimates from the training period for the static parameters. Summing up the K pseudo log predictive scores for the days $T+1, \ldots, T+K$ yields the cumulative pseudo log predictive score. Appendix B.3 contains a detailed description of this procedure. For the models $\mathcal{M}_{static}^{mix}$, $\mathcal{M}_{dyn}^{t}$ and $\mathcal{M}_{static}^{t}$ we proceed similarly.

This procedure for calculating the cumulative pseudo log predictive score is applied to both data sets corresponding to the pairs (SPX,VIX) and (DAX,VDAX). Using the last two years (2012 - 2013) of our data set as test data yields $K = 517$. As training period we use $T = 1000$ which corresponds to a training period of approximately four years.

Table 4.6 summarizes the cumulative pseudo log predictive scores. In both cases, for the (SPX,VIX) as well as for the (DAX,VDAX) data, the best model is provided by the dynamic mixture copula model $\mathcal{M}_{dyn}^{mix}$. Furthermore, we see that in both cases the static mixture copula model $\mathcal{M}_{static}^{mix}$ is preferred over the static and dynamic Student t copula models $\mathcal{M}_{static}^{t}$ and $\mathcal{M}_{dyn}^{t}$. For the (DAX,VDAX) data, the second best model is provided by the static mixture copula model $\mathcal{M}_{static}^{mix}$. For this data the rolling window estimates of Kendall's τ in Figure 4.3 vary less than for the (SPX,VIX) data. For the (SPX,VIX) data, the DCC-GARCH $\mathcal{M}_{DCC}$ yields the second best cumulative pseudo log predictive

score.

	$\mathcal{M}_{dyn}^{mix}$	$\mathcal{M}_{static}^{mix}$	$\mathcal{M}_{dyn}^{t}$	$\mathcal{M}_{static}^{t}$	$\mathcal{M}_{DCC}$
(SPX,VIX)	**2740.7**	2733.5	2732.0	2725.9	2737.2
(DAX,VDAX)	**2817.7**	2814.1	2810.1	2809.9	2789.0

Table 4.6: Cumulative pseudo log predictive scores for the models $\mathcal{M}_{dyn}^{mix}$, $\mathcal{M}_{static}^{mix}$, $\mathcal{M}_{dyn}^{t}$, $\mathcal{M}_{static}^{t}$ and $\mathcal{M}_{DCC}$.

4.5 Conclusion

We propose a sampler, applicable to general nonlinear state space models with univariate autoregressive state equation. Sampling efficiency is demonstrated for bivariate dynamic copula models within a simulation study. Furthermore, we use the sampler to estimate the parameters of a dynamic bivariate mixture copula model. This mixture copula model turns out to be a good candidate to model the volatility return relationship, since in our application it produces more accurate forecasts than a bivariate DCC-GARCH model or a Student t copula model.

In this work, there are two objectives that might be extended: The sampler and the bivariate mixture copula model. The sampler could be extended to allow for a broader class of models. For example, we might consider autoregressive processes of higher order in the state equation. In this case we can still rely on elliptical slice sampling and on an interweaving strategy. Another extension could relax the assumption of a Gaussian dependence structure in the state equation by replacing the autoregressive process by a D-vine copula model. In this case elliptical slice sampling can no longer be applied to sample the latent states and an alternative sampling method is required.

The bivariate dynamic mixture copula could serve as a building block for regular vine copula models. Thus, we could extend the bivariate model to arbitrary dimensions. This is interesting if we study not only the bivariate volatility return relationship, but for example the dependence structure among several exchange rates.

5 Dynamic vine copulas

This chapter is a reproduction of Kreuzer and Czado (2019b) with minor changes.

5.1 Introduction

In many applications it is assumed that the dependence does not change over time, but this assumption is often not appropriate. For example, there is evidence that the correlations between the returns of stocks and bonds change over time (Baele et al. (2010)). Popular models for financial data that account for dynamic dependence are multivariate GARCH models with time-varying correlations, such as the DCC-GARCH (Engle (2002)) and multivariate factor stochastic volatility models (Harvey et al. (1994), Pitt and Shephard (1999), Kastner et al. (2017)). But these models rely (conditionally) on multivariate normal distributions. New models have been proposed to overcome the shortcomings of the multivariate normal distribution and to allow for more flexible time-varying dependence structures. One example for such a model is the dynamic copula model of Oh and Patton (2018). Another approach to construct dynamic dependence models in higher dimensions is to extend the flexible class of vine copulas (see Section 2.2.4). Since vine copulas are constructed from bivariate copulas, the vine copula framework allows us to scale dynamic bivariate copula models to arbitrary dimensions. Acar et al. (2019) use nonparametric smoothing techniques to allow for time-variation in bivariate copula models and extend this bivariate approach to higher dimensions using vine copulas. Vatter and Chavez-Demoulin (2015) propose a bivariate copula model, where the copula parameters depend on covariates through generalized additive models. Using the vine copula framework, this bivariate model is extended to higher dimensions by Vatter and Nagler (2018). Similarly the bivariate dynamic copula model as proposed by Almeida and Czado (2012) and Hafner and Manner (2012) is extended to dynamic D-vine copulas by Almeida et al. (2016) and later to dynamic C-vine copulas by Goel and Mehra (2019).

The bivariate dynamic copula model of Almeida and Czado (2012) provides a flexible building block by modeling time-varying dependencies with latent AR(1) processes. However, estimation is no longer straightforward since the likelihood involves high-dimensional integration. Goel and Mehra (2019) follow Almeida et al. (2016), who use a frequentist estimation approach with approximation of the likelihood utilizing efficient importance sampling (Richard and Zhang (2007)). In this approach, parameters are estimated sequentially tree by tree. For estimating parameters of higher trees, the parameters of lower trees are fixed at point estimates. Thus, uncertainty of parameters in lower trees is ignored and therefore uncertainty quantification cannot be provided.

This chapter contains two major contributions: So far dynamic vine copula models,

as generalization of the dynamic bivariate copula model of Almeida and Czado (2012), were restricted to D-vine (Almeida et al. (2016)) and C-vine (Goel and Mehra (2019)) structures. First, we develop an approach to allow for general regular vine tree structures. D-vine structures are especially suited to describe temporal dependence. But when it comes to cross-sectional dependence structures, such as the dependence among several stocks, the D-vine structure might be too restrictive. General regular vine tree structures are more flexible and include C-vine and D-vine structures as special cases.

Second, we present a novel Bayesian estimation approach. To our knowledge, Bayesian estimation of vine copula models, including structure selection, was only tackled by Gruber and Czado (2015) and Gruber and Czado (2018). These approaches only allow for static pair copulas and have not been applied in more than 10 dimensions, while our approach allows for static as well as dynamic pair copulas and can handle higher dimensions.

Our methodology is based on an approximation of the posterior distribution. Approximations to the posterior are also used in variational Bayesian inference (Wainwright et al. (2008)) and have become popular since they make estimation feasible in high-dimensional settings. Variational Bayesian approaches assume that the posterior distribution belongs to some family of distributions, such as the multivariate normal distribution. Our approach does not rely on such an assumption. We propose an approximation, which uses ideas of the frequentist sequential procedure of Dissmann et al. (2013) and which allows to estimate pair copulas of one tree independently of each other using Markov Chain Monte Carlo (MCMC) schemes developed in Chapter 4. The posterior approximation also enables the user to run several MCMC chains in parallel leading to faster computation and making the approach applicable to higher dimensions than the approach of Gruber and Czado (2018). Further, our Bayesian approach includes pair copula family selection based on a set of candidate families. Here we exploit the fact that for several copula families, there is a one-to-one correspondence between the copula parameter and Kendall's τ. This allows to share the Kendall's τ parameter among different copula families, which reduces the parameter space and simplifies estimation. Additionally, our approach also contributes to the selection of sparse models by assessing, whether a pair copula term needs to be modeled dynamically or not. For this the information criteria of Watanabe (2010) is adapted. Another advantage of our Bayesian approach is that it is not necessary to fix pair copula parameters at point estimates although our procedure is sequential. Uncertainty of parameter estimates and information in lower trees is no longer ignored, but propagated as we move up to higher trees in the estimation procedure. All ideas are investigated through simulation and illustrated with real data.

The outline of this chapter is as follows: Section 5.2 discusses the bivariate building blocks that are needed to construct the dynamic vine copula model. We introduce dynamic and static building blocks and show how to select among them. The selection procedure is illustrated with a small simulation study. Section 5.3 introduces the dynamic vine copula model and a novel algorithm for parameter estimation. The performance of the estimation procedure is evaluated with simulated data. In Section 5.4 we model the dependence among 21 exchange rates. Within this application, we compare the predictive accuracy of the proposed dynamic vine copula model to competitor models: a static vine copula, a dynamic C-vine copula and a dynamic D-vine copula. We conclude with providing ideas for future research in Section 5.5.

5.2 Bivariate building blocks for the dynamic vine copula model

The dynamic vine copula model, we introduce in Section 5.3, relies on dynamic and static bivariate copula models. These bivariate models are introduced in Sections 5.2.1 and 5.2.2. Section 5.2.3 discusses selection among different bivariate copula models.

5.2.1 Dynamic bivariate copulas

Model specification

We extend the dynamic bivariate copula model of Almeida and Czado (2012), which was already discussed in Chapter 4, by a model indicator m to allow for Bayesian copula family selection. We consider a set $\mathcal{M}$ of single-parameter copula families for which there is a one-to-one correspondence between the copula parameter and Kendall's τ. The extended model for T bivariate random vectors $(U_{t1}, U_{t2})_{t=1,\dots,T} \in [0,1]^{T\times 2}$ is given by

$$\begin{aligned} &(U_{t1}, U_{t2})|m, s_t \sim c^m(u_{t1}, u_{t2}; \tau_t), \\ &s_t = \mu + \phi(s_{t-1} - \mu) + \sigma\eta_t, \quad \tau_t = F_Z^{-1}(s_t), \end{aligned} \tag{5.1}$$

with $\eta_t, s_0, \dots, s_T, \mu, \phi, \sigma$ similar to (4.2) in Chapter 4 and $c^m(\cdot,\cdot;\tau_t)$ is the bivariate density of copula family $m \in \mathcal{M}$ with Kendall's τ parameter τ_t. The state s_t has the same interpretation for different copula families as the Fisher's Z transform of the corresponding Kendall's τ value. This allows us to share the parameters $\boldsymbol{s_{0:T}} = (s_0, \dots, s_T), \mu, \phi, \sigma$ among different copula families, which keeps the parameter space smaller and simplifies estimation. More details about parameter sharing are given in Appendix G.

A Bayesian model specification is complete by introducing priors for the model parameters. Let $\varphi(\cdot|\mu_{Normal}, \sigma^2_{Normal})$ denote the univariate normal density with mean μ_{Normal} and variance σ^2_{Normal}. This allows us to express the prior for $\boldsymbol{s_{0:T}}$ conditional on μ, ϕ and σ, implied by the AR(1) process in (5.1), as

$$\pi(\boldsymbol{s_{0:T}}|\mu, \phi, \sigma) = \varphi(s_0|\mu, \sigma^2(1-\phi^2)^{-1}) \prod_{t=1}^{T} \varphi(s_t|\mu + \phi(s_{t-1} - \mu), \sigma^2). \tag{5.2}$$

For μ, ϕ, and σ we use the same prior distribution as in Chapter 4 (Equation (4.5)) and for $m \in \mathcal{M}$ we assume a discrete uniform prior, i.e.

$$\pi(m) = \frac{1}{|\mathcal{M}|}. \tag{5.3}$$

We further assume prior independence among the parameters μ, ϕ, σ and m.

Bayesian inference

Model selection procedures often have to deal with model specific parameters. They might have different dimensions. A popular approach in this context is reversible jump MCMC (see Green (1995)), which requires dimension matching. Min and Czado (2011) and

Gruber and Czado (2015) use reversible jump MCMC for selection among vine copula models, while Tan et al. (2019) apply it for model selection among static single factor copula models. The way we constructed our model, reversible jump MCMC is not needed, since we share the parameters $s_0, \dots, s_T, \mu$, ϕ and σ among different models. This allows us to use an efficient Gibbs approach as outlined in the following.

Here the likelihood given the data $\mathsf{U} = (u_{t1}, u_{t2})_{t=1,\dots,T}$ can be expressed as

$$\ell(\mu, \phi, \sigma, \boldsymbol{s_{0:T}}, m|\mathsf{U}) = \prod_{t=1}^{T} \ell_t(s_t, m|u_{t1}, u_{t2}) = \prod_{t=1}^{T} c^m(u_{t1}, u_{t2}; \tau_t). \tag{5.4}$$

The quantity $\ell_t(s_t, m|u_{t1}, u_{t2}) = c^m(u_{t1}, u_{t2}; \tau_t)$ is the contribution to the likelihood at time t and $\tau_t = F_Z^{-1}(s_t)$.

We now employ a Gibbs sampler for parameter estimation. The indicator m is sampled from its full conditional, given by

$$\begin{aligned} P(m|U, \mu, \phi, \sigma, \boldsymbol{s_{0:T}}) &= \frac{f(U|\mu, \phi, \sigma, \boldsymbol{s_{0:T}}, m)\pi(\mu, \phi, \sigma, \boldsymbol{s_{0:T}})\pi(m)}{\sum_{m' \in \mathcal{M}} f(U|\mu, \phi, \sigma, \boldsymbol{s_{0:T}}, m')\pi(\mu, \phi, \sigma, \boldsymbol{s_{0:T}})\pi(m')} \\ &= \frac{\ell(\mu, \phi, \sigma, \boldsymbol{s_{0:T}}, m|U)}{\sum_{m' \in \mathcal{M}} \ell(\mu, \phi, \sigma, \boldsymbol{s_{0:T}}, m'|U)} \\ &= \frac{\prod_{t=1}^{T} c^m(u_{t1}, u_{t2}; \tau_t)}{\sum_{m' \in \mathcal{M}} \prod_{t=1}^{T} c^{m'}(u_{t1}, u_{t2}; \tau_t)}, \end{aligned} \tag{5.5}$$

where $\pi(\mu, \phi, \sigma, \boldsymbol{s_{0:T}}) = \pi(\boldsymbol{s_{0:T}}|\mu, \phi, \sigma)\pi(\mu)\pi(\phi)\pi(\sigma)$ and $\pi(\cdot)$ as specified in (5.2), (4.5) and (5.3). To sample $\mu, \phi, \sigma, \boldsymbol{s_{0:T}}$ conditioned on the indicator m, we use the same approach as described in Chapter 4 for multivariate state space models with a univariate autoregressive state equation.

5.2.2 Static bivariate copulas

Model specification

To allow for static (time-constant) copulas, we consider a static state $s \in \mathbb{R}$ which is mapped to the Kendall's τ parameter similar to (5.1), i.e.

$$\tau = F_Z^{-1}(s) \tag{5.6}$$

We assume that T bivariate random vectors $(U_{t1}, U_{t2})_{t=1,\dots,T}$ are generated as follows

$$(U_{t1}, U_{t2})|m, s \sim c^m(u_{t1}, u_{t2}; \tau), \text{ independently} \tag{5.7}$$

with $m \in \mathcal{M}$. The prior for the state s is chosen such that the corresponding Kendall's τ is uniformly distributed on $(-1, 1)$. The prior for m is chosen as above in (5.3). The priors reflect the fact that we do not have any prior information about the parameters.

Bayesian inference

The parameters of this reduced model are also estimated utilizing a Gibbs sampling approach. Here we sample $m|U, s$ directly from its full conditional, which can be derived similar to (5.5). The parameter s is updated conditional on (U, m) with random walk Metropolis-Hastings with Gaussian proposal and adaptive proposal variance as in Garthwaite et al. (2016).

5.2.3 Model selection among dynamic and static pair copulas using the widely applicable information criteria (WAIC)

One might be interested in how the dynamic copula model compares to the static copula model and to the independence model. We refer to these model classes as dynamic, static and zero dependence, respectively. To also allow for bivariate static and independence copulas is especially important, if the bivariate dynamic copula is used as a building block for vine copula models in higher dimensions. For vine copula models, independence copulas might be useful in higher trees to avoid overfitting.

Bayes factors or the commonly used information critera AIC and BIC are here not tractable choices for selecting the type of dependence (dynamic, static or zero), since their evaluation would require high-dimensional integration. Instead, we rely on the *widely applicable information criteria (WAIC)* (Watanabe (2010), Gelman et al. (2014b)). For the proposed dynamic copula model it is given by

$$\text{WAIC} = -2\left(\sum_{t=1}^{T}\ln(\text{E}(\ell_t(s_t,m|u_{t1},u_{t2}))) - \sum_{t=1}^{T}\text{Var}\left(\ln(\ell_t(s_t,m|u_{t1},u_{t2}))\right)\right) \tag{5.8}$$

with ℓ_t as in (5.4). The expectation and variance are taken with respect to $P(s_t,m)$, the probability measure of the posterior distribution of s_t and m, i.e.

$$\text{E}(\ell_t(s_t,m|u_{t1},u_{t2})) = \int_{\mathbb{R}\times\mathcal{M}} \ell_t(s_t,m|u_{t1},u_{t2})dP(s_t,m) \tag{5.9}$$

and

$$\begin{aligned}&\text{Var}(\ln(\ell_t(s_t,m|u_{t1},u_{t2}))) = \\ &\quad= \int_{\mathbb{R}\times\mathcal{M}}(\ln(\ell_t(s_t,m|u_{t1},u_{t2})))^2 dP(s_t,m) - \left(\text{E}(\ln(\ell_t(s_t,m|u_{t1},u_{t2})))\right)^2.\end{aligned} \tag{5.10}$$

The WAIC can be seen as a Bayesian version of the AIC, where, instead of the number of parameters, $\sum_{t=1}^{T}\text{Var}(\ln(\ell_t(s_t,m|u_{t1},u_{t2}))$ is used as a penalty.

For R observed quantities $(x^r)_{r=1,\dots,R}$, we denote by $\hat{E}((x^r)_{r=1,\dots,R}) = \frac{1}{R}\sum_{r=1}^{R}x^r$ the sample mean and by $\widehat{\text{Var}}((x^r)_{r=1,\dots,R}) = \frac{1}{R-1}\sum_{r=1}^{R}(x^r - \hat{E}((x^r)_{r=1,\dots,R}))^2$ the sample variance. Following Vehtari et al. (2017), the WAIC can be estimated from R samples of the posterior distribution $(s_t^1,m^1),\dots,(s_t^R,m^R)$ with $t=1,\dots,T$, by

$$\widehat{\text{WAIC}} = -2\left(\sum_{t=1}^{T}\ln\left(\hat{E}((\ell_t^r)_{r=1,\dots,R})\right) - \sum_{t=1}^{T}\widehat{\text{Var}}\left((\ln(\ell_t^r))_{r=1,\dots,R}\right)\right), \tag{5.11}$$

where $\ell_t^r = \ell_t(s_t^r,m^r|u_{t1},u_{t2})$. By setting

$$\widehat{\text{WAIC}}_t = -2\left(\ln\left(\hat{E}((\ell_t^r)_{r=1,\dots,R})\right) - \widehat{\text{Var}}\left((\ln(\ell_t^r))_{r=1,\dots,R}\right)\right),$$

we can express $\widehat{\text{WAIC}}$ as $\widehat{\text{WAIC}} = \sum_{t=1}^{T}\widehat{\text{WAIC}}_t$.

To compare between two models with estimated WAIC values $\widehat{\text{WAIC}}^A$ and $\widehat{\text{WAIC}}^B$, Vehtari et al. (2017) suggest to consider the difference in the estimated WAIC given by

$$\widehat{\text{WAIC}}^A - \widehat{\text{WAIC}}^B = \sum_{t=1}^{T}(\widehat{\text{WAIC}}_t^{\,A} - \widehat{\text{WAIC}}_t^{\,B}) \tag{5.12}$$

with corresponding standard error estimate

$$\hat{se}(\widehat{\text{WAIC}}^A - \widehat{\text{WAIC}}^B) = \sqrt{T \cdot \widehat{\text{VAR}}\left((\widehat{\text{WAIC}}_t^A - \widehat{\text{WAIC}}_t^B)_{t=1,\dots,T}\right)}. \tag{5.13}$$

To estimate the standard error, Vehtari et al. (2017) assume independence among the components $\widehat{\text{WAIC}}_t, t = 1, \dots, T$. For the static copula model we use $\ell_t(s, m|u_{t1}, u_{t2}) = c^m(u_{t1}, u_{t2}; \tau)$. Further, WAIC is zero for the independence copula. In our framework, the dynamic model is considered to be more complex than the static model. Similarly, the static and the dynamic model are considered to be more complex than the independence model. Here, we select the more complex model if its estimated WAIC is at least 2 estimated standard errors smaller than the estimated WAIC of the other model.

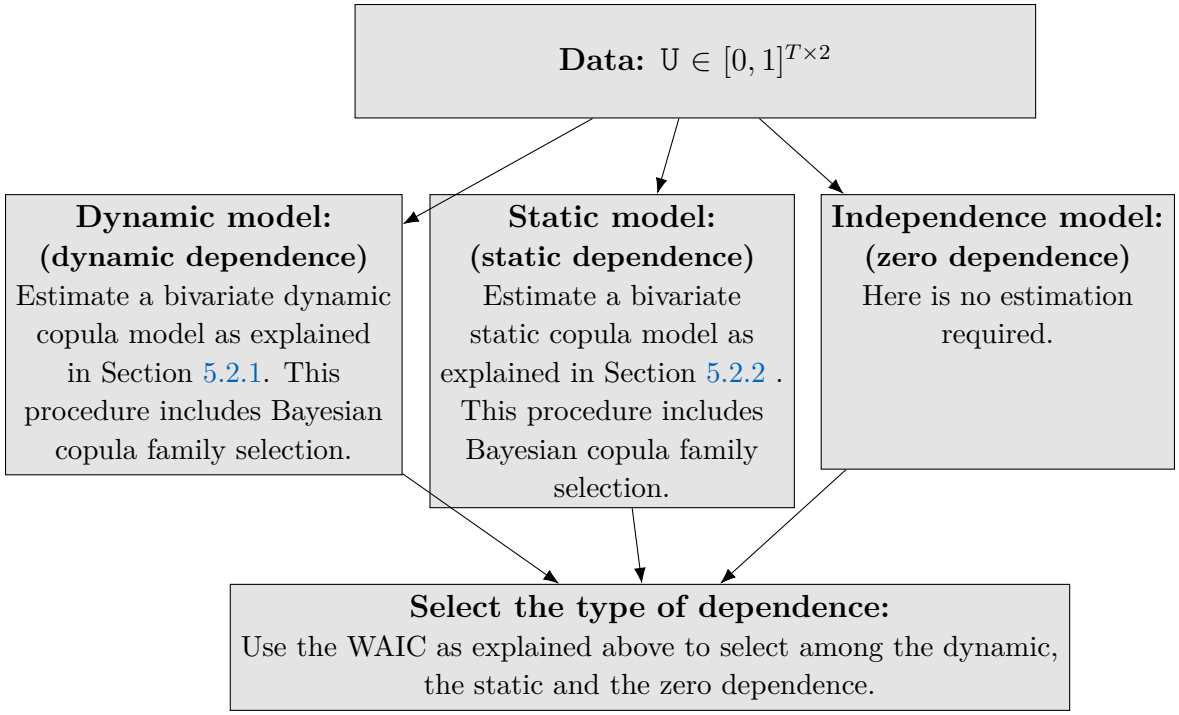

Figure 5.1: Model selection procedure for bivariate copula models.

Alternatively, we could have incorporated the independence, the static and the dynamic copula within one sampler. In this case we would need to move between models with different dimensions by employing e.g. reversible jump MCMC. But proposing moves efficiently from the parameter-free independence copula or from the static copula to dynamic copulas with more than T parameters is difficult and chains might take very long to converge. Another alternative is to select among all models, including family choices, with WAIC. But this would require to estimate a dynamic bivariate copula model for each copula family, which is computationally expensive. So we propose to use the Gibbs sampler to move between models with the same dimension, where the parameters can be shared among the different models. The WAIC is used to select between models with different parameter dimensions where parameters cannot be shared. The whole selection procedure for the bivariate copula models is visualized in Figure 5.1.

5.2.4 Simulation study

We conduct a small simulation study to investigate the ability of WAIC to select the type of dependence. We consider five scenarios specified in Table 5.1. In Scenarios 1 and 2 we simulate from the dynamic model specified in (5.1), in Scenarios 3 and 4 from the static one specified in (5.7) and in Scenario 5 from the bivariate independence copula. For each scenario we simulate $T = 1000$ observations. Based on these observations we fit the dynamic and the static model and select among different models with WAIC as explained in Section 5.2.3. We consider the following set for copula family selection $\mathcal{M} =$ {Independence, Gaussian, Student t(df=4), eClayton, eGumbel}, with copula families as introduced in Section 2.2.3. The degrees of freedom parameter of the Student t copula is here considered fixed. We repeat the simulations for each scenario 100 times. From Table 5.2 we see that in each scenario, the correct type of dependence was selected in at least 84 out of 100 cases. The correct family was selected in at least 98 out of 100 cases according to Table 5.2. We conclude that our Bayesian family selection procedure performs well and that WAIC can be utilized to select the appropriate type of dependence.

	Dynamic		Static		Independence
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
m	Gaussian	eClayton	Student t(df=4)	eGumbel	Independence
μ	0.4	0.4			
ϕ	0.95	0.8			
σ	0.1	0.2			
s			1	0.4	

Table 5.1: Parameter specification for the simulation study for bivariate copula models.

		Dynamic		Static		Independence
	Scenario	1	2	3	4	5
Copula family	Independence	0	0	0	0	**100**
	Gaussian	**100**	2	1	0	0
	Student t(df=4)	0	0	**99**	0	0
	eClayton	0	**98**	0	0	0
	eGumbel	0	0	0	**100**	0
Type of dependence	Dynamic	**100**	**100**	16	0	1
	Static	0	0	**84**	**100**	0
	Zero	0	0	0	0	**100**

Table 5.2: For each scenario of the simulation study, we show how often the different copula families were selected and how often each type of dependence (dynamic, static and zero) was selected. The selected copula family is the marginal posterior mode estimate of m, i.e. the family that occurs most frequently among the posterior samples for m. The true copula family and the true type of dependence, which we used for simulation, is marked in bold.

5.3 Dynamic vine copulas

5.3.1 Model specification

Within the vine copula framework, we only need to specify bivariate copulas and can scale to arbitrary dimensions. Here, we replace each bivariate copula of a vine copula model, as introduced in Section 2.2.4, by the dynamic bivariate copula model specified in (5.1), the static bivariate copula model specified in (5.7) or the independence copula. We denote by E_i^{dyn}, E_i^{static} and E_i^{ind} the set of edges in the i-th tree where the corresponding pair copulas are dynamic, static or independence copulas, respectively, and $E_i = E_i^{dyn} \cup E_i^{static} \cup E_i^{ind}$. For each edge $e \in E_i^{dyn}, i = 1, \dots, d-1$, where d is the number of variables, we have a corresponding family indicator m_e and a corresponding latent AR(1) process given by

$$s_{t,e} = \mu_e + \phi_e(s_{t-1,e} - \mu_e) + \sigma_e \eta_{t,e}, \eta_{t,e} \sim N(0,1) \text{ iid}, \tag{5.14}$$

with μ_e, ϕ_e σ_e and $s_{0,e}$ as in Section 5.2.1. For each edge e in E_i^{static} we have a corresponding family indicator m_e and a state s_e as in Section 5.2.2. The states $s_{t,e}$ and s_e are mapped to the Kendall's τ parameter as in (5.1) and (5.6), i.e.

$$\tau_{t,e} = F_Z^{-1}(s_{t,e}) \text{ and } \tau_e = F_Z^{-1}(s_e), \text{ respectively.} \tag{5.15}$$

This yields the following parameter set for a d-dimensional dynamic vine copula model

$$\begin{aligned}\boldsymbol{\theta^V} =& \{\mu_e, \phi_e, \sigma_e, s_{0,e}, \dots, s_{T,e}, m_e | e \in E_i^{dyn}, i = 1, \dots, d-1\} \cup \\ & \cup \{s_e, m_e | e \in E_i^{static}, i = 1, \dots, d-1\}\end{aligned}$$

Within the *dynamic vine copula model*, we assume that T random vectors of dimension d, $(U_{t1}, \dots, U_{td})_{t=1,\dots,T}$, are generated as follows

$$\begin{aligned}(U_{t1}, \dots U_{td}) | \boldsymbol{\theta^V} \sim & \prod_{i=1}^{d-1} \prod_{e E_i^{dyn}} c^{m_e}_{a_e,b_e;\boldsymbol{D_e}}(u_{t,a_e|\boldsymbol{D_e}}, u_{t,b_e|\boldsymbol{D_e}}; \tau_{t,e}) \cdot \\ & \cdot \prod_{i=1}^{d-1} \prod_{e \in E_i^{static}} c^{m_e}_{a_e,b_e;\boldsymbol{D_e}}(u_{t,a_e|\boldsymbol{D_e}}, u_{t,b_e|\boldsymbol{D_e}}; \tau_e), \text{ independently}\end{aligned} \tag{5.16}$$

for $t = 1, \dots, T$. Further, we assume that parameters for different edges are a priori independent and we use the same priors as specified in Sections 5.2.1 and 5.2.2 for the parameters of dynamic and static pair copulas, respectively. Here, the conditioned and conditioning sets a_e, b_e and $\boldsymbol{D_e}$ (defined in Section 2.2.4) do not depend on the time, i.e. the tree structure does not change over time.

5.3.2 Sequential estimation

Since there exist $\frac{d!}{2} \cdot 2^{\binom{d-2}{2}}$ different regular vine tree structures in d dimensions (Morales-Nápoles (2010)), model selection is complex and it is not possible to take all possible structures into account as the dimension grows. Gruber and Czado (2018) estimate all

trees and parameters jointly using a Bayesian approach, but their procedure is only suitable in lower dimensions and requires substantial computations. Earlier, in a frequentist setup, Dissmann et al. (2013) proposed a sequential selection and estimation approach for static copula parameters. This sequential approach makes model selection feasible in higher dimensions and therefore this idea is often used for vine copula based models. Gruber and Czado (2015) employ a Bayesian approach, where they estimate parameters of static vine copula models tree by tree and fix parameters at point estimates before proceeding to the next tree. Vatter and Nagler (2018) sequentially estimate the parameters of a vine copula model, where the copula parameters are modeled by generalized additive models. We propose a Bayesian procedure, where the copula parameters do not need to be collapsed to point estimates before proceeding to the next tree. The procedure is based on an approximation of the posterior density inspired by the frequentist sequential approach of Dissmann et al. (2013).

To simplify notation we denote by $\boldsymbol{\theta_{T_i}}$ the parameters of a dynamic vine copula corresponding to tree T_i , i.e.

$$\boldsymbol{\theta_{T_i}} = \{\mu_e, \phi_e, \sigma_e, s_{0,e}, \dots, s_{T,e}, m_e | e \in E_i^{dyn}\} \cup \{s_e, m_e | e \in E_i^{static}\}. \tag{5.17}$$

Further, we define by $\mathrm{U} = (u_{tj})_{t=1,\dots,T, j=1,\dots,d}$ the data matrix and the likelihood contribution of the i-th tree is given by

$$\begin{aligned} \ell_i(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_i}} | \mathrm{U}) = \prod_{t=1}^{T} \Bigg(& \prod_{e \in E_i^{dyn}} c^{m_e}_{a_e,b_e;\boldsymbol{D_e}}(u_{t,a_e|\boldsymbol{D_e}}, u_{t,b_e|\boldsymbol{D_e}}; \tau_{t,e}) \cdot \\ & \cdot \prod_{e \in E_i^{static}} c^{m_e}_{a_e,b_e;\boldsymbol{D_e}}(u_{t,a_e|\boldsymbol{D_e}}, u_{t,b_e|\boldsymbol{D_e}}; \tau_e) \Bigg). \end{aligned} \tag{5.18}$$

The complete likelihood is obtained as

$$\ell(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_{d-1}}} | \mathrm{U}) = \prod_{i=1}^{d-1} \ell_i(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_i}} | \mathrm{U}), \tag{5.19}$$

and the posterior density is proportional to

$$f(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_{d-1}}} | \mathrm{U}) \propto \prod_{i=1}^{d-1} \ell_i(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_i}} | \mathrm{U}) \pi(\boldsymbol{\theta_{T_i}}). \tag{5.20}$$

Note that the posterior density is a joint density of continuous and discrete parameters. For discrete parameters $\boldsymbol{\delta^{disc}}$ and continuous parameters $\boldsymbol{\delta^{cont}}$, the joint density is obtained as $f(\boldsymbol{\delta^{cont}}, \boldsymbol{\delta^{disc}}) = f(\boldsymbol{\delta^{cont}} | \boldsymbol{\delta^{disc}}) f(\boldsymbol{\delta^{disc}})$, where $f(\boldsymbol{\delta^{cont}} | \boldsymbol{\delta^{disc}})$ is a joint probability density function and $f(\boldsymbol{\delta^{disc}})$ is a joint probability mass function.

Four-dimensional illustration for a model with static pair copulas and known tree structure

To illustrate our idea in four dimensions, we consider only static pair copulas and the tree structure to be known. The tree structure contains the pair copulas as specified in

(5.21), (5.22) and (5.23). We assume, that we observe data $\mathtt{U} = (u_{tj})_{t=1,\dots,T,j=1,\dots,4}$. The contributions to the likelihood corresponding to trees 1, 2 and 3 are given by

$$\ell_1(\boldsymbol{\theta_{T_1}}|\mathtt{U}) = \prod_{t=1}^{T} c_{13}^{m_{13}}(u_{t1}, u_{t3}; \tau_{13}) c_{23}^{m_{23}}(u_{t2}, u_{t3}; \tau_{23}) c_{34}^{m_{34}}(u_{t3}, u_{t4}; \tau_{34}), \tag{5.21}$$

$$\ell_2(\boldsymbol{\theta_{T_1}}, \boldsymbol{\theta_{T_2}}|\mathtt{U}) = \prod_{t=1}^{T} c_{12;3}^{m_{12;3}}(u_{t1|3}, u_{t2|3}; \tau_{12;3}) c_{24;3}^{m_{24;3}}(u_{t2|3}, u_{t4|3}; \tau_{24;3}), \tag{5.22}$$

$$\ell_3(\boldsymbol{\theta_{T_1}}, \boldsymbol{\theta_{T_2}}, \boldsymbol{\theta_{T_3}}|\mathtt{U}) = \prod_{t=1}^{T} c_{14;23}^{m_{14;23}}(u_{t1|23}, u_{t4|23}; \tau_{14;23}). \tag{5.23}$$

In a frequentist sequential procedure the parameters of the first tree are estimated by considering the part of the likelihood corresponding to the first tree as given in (5.21), ignoring the likelihood contributions of higher trees (5.22) and (5.23) to the parameters of the first tree. This allows to maximize $\prod_{t=1}^{T} c_{13}^{m_{13}}(u_{t1}, u_{t3}; \tau_{13})$, $\prod_{t=1}^{T} c_{23}^{m_{23}}(u_{t2}, u_{t3}; \tau_{23})$ and $\prod_{t=1}^{T} c_{34}^{m_{34}}(u_{t3}, u_{t4}; \tau_{34})$ independently.

In the Bayesian setup the marginal posterior density of the parameters corresponding to the first tree is obtained by integrating out parameters of higher trees, i.e.

$$f(\boldsymbol{\theta_{T_1}}|\mathtt{U}) \propto \ell_1(\boldsymbol{\theta_{T_1}}|\mathtt{U})\pi(\boldsymbol{\theta_{T_1}}) \int_{domain(\boldsymbol{\theta_{T_2}}, \boldsymbol{\theta_{T_3}})} \left(\prod_{i=2}^{3} \ell_i(\boldsymbol{\theta_{T_1}}, \dots, \boldsymbol{\theta_{T_i}}|\mathtt{U})\pi(\boldsymbol{\theta_{T_i}}) \right) d\boldsymbol{\theta_{T_2}} d\boldsymbol{\theta_{T_3}}, \tag{5.24}$$

where $domain(\boldsymbol{\theta_{T_2}}, \boldsymbol{\theta_{T_3}})$ is the domain of the parameters $\boldsymbol{\theta_{T_2}}, \boldsymbol{\theta_{T_3}}$ (For the family indicator m_e the integral is replaced by a sum). While in this illustrative example it might be possible to work with the marginal posterior (5.24), its complexity grows fast if we consider more dimensions or allow for dynamic copulas. For example, if the second and third tree were modeled with dynamic bivariate copulas, we would need to integrate out several thousand parameters for $T = 1000$. To reduce complexity, we approximate $f(\boldsymbol{\theta_{T_1}}|U)$. We make use of the following notation

$$g(\boldsymbol{\delta}) \approx h(\boldsymbol{\delta})$$

to denote that a density g is approximately proportional to a non negative and integrable function h. This means that the density g is approximated by the density $h^{normalized}$ given by $h^{normalized}(\boldsymbol{\delta}) = h(\boldsymbol{\delta}) \left(\int_{domain(\boldsymbol{\delta})} h(\boldsymbol{\delta}) d\boldsymbol{\delta} \right)^{-1}$.

Following the idea of a frequentist sequential estimation, we approximate the marginal posterior $\boldsymbol{\theta_{T_1}}|U$ by considering only the part of the likelihood corresponding to the first tree, i.e.

$$\begin{aligned} f(\boldsymbol{\theta_{T_1}}|\mathtt{U}) \approx \ell_1(\boldsymbol{\theta_{T_1}}|\mathtt{U})\pi(\boldsymbol{\theta_{T_1}}) = &\left(\prod_{t=1}^{T} c_{13}^{m_{13}}(u_{t1}, u_{t3}; \tau_{13}) \right) \pi(s_{13})\pi(m_{13}) \\ &\left(\prod_{t=1}^{T} c_{23}^{m_{23}}(u_{t2}, u_{t3}; \tau_{23}) \right) \pi(s_{23})\pi(m_{23}) \left(\prod_{t=1}^{T} c_{34}^{m_{34}}(u_{t3}, u_{t4}; \tau_{34}) \right) \pi(s_{34})\pi(m_{34}). \end{aligned} \tag{5.25}$$

This approximation simplifies sampling enormously. We do not only get rid of the integral, but in addition, the parameters corresponding to different edges are independent. To obtain samples from the posterior approximation in (5.25), we can sample

parameters of different static bivariate copula models independently by utilizing the algorithm of Section 5.2.2. In particular, the parameters m_{13}, s_{13} are sampled from a static bivariate copula model with corresponding posterior density proportional to $\left(\prod_{t=1}^{T} c_{13}^{m_{13}}(u_{t1}, u_{t3}; \tau_{13})\right) \pi(s_{13})\pi(m_{13})$. Approximations of the posterior distribution that induce independence among parameters are commonly used in variational Bayesian approaches (Wainwright et al. (2008)). For example, in mean field variational inference it is assumed that all parameters are independent in the posterior distribution. Our assumptions are less restrictive, since we do not assume that parameters corresponding to one pair copula are independent. Further, these parameters are updated jointly.

When estimating parameters of higher trees in a sequential frequentist procedure, we condition on estimates from lower trees. Therefore we consider now the following density

$$f(\boldsymbol{\theta}_{\boldsymbol{T_2}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}, \mathbb{U}) \propto \ell_2(\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}|\mathbb{U})\pi(\boldsymbol{\theta}_{\boldsymbol{T_2}}) \int_{domain(\boldsymbol{\theta}_{\boldsymbol{T_3}})} \ell_3(\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}, \boldsymbol{\theta}_{\boldsymbol{T_3}}|\mathbb{U})\pi(\boldsymbol{\theta}_{\boldsymbol{T_3}})d\boldsymbol{\theta}_{\boldsymbol{T_3}}. \tag{5.26}$$

We utilize a similar approximation as in (5.25) and obtain

$$\begin{aligned} f(\boldsymbol{\theta}_{\boldsymbol{T_2}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}, \mathbb{U}) \approx \ell_2(\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}|\mathbb{U})\pi(\boldsymbol{\theta}_{\boldsymbol{T_2}}) = &\left(\prod_{t=1}^{T} c_{12;3}^{m_{12;3}}(u_{t1|3}, u_{t2|3}; \tau_{12;3})\right) \pi(s_{12;3})\pi(m_{12;3}) \\ &\left(\prod_{t=1}^{T} c_{24;3}^{m_{24;3}}(u_{t2|3}, u_{t4|3}; \tau_{24;3})\right) \pi(s_{24;3})\pi(m_{24;3}). \end{aligned} \tag{5.27}$$

The pseudo data $u_{t1|3} = h_{1|3}(u_{t1}|u_{t3}; \tau_{13}, m_{13}) = \frac{d}{du_3}\mathbb{C}_{13}^{m_{13}}(u_{t1}, u_3; \tau_{13})\Big|_{u_3=u_{t3}}$, $u_{t2|3} = h_{2|3}(u_{t2}|u_{t3}; \tau_{23}, m_{23})$ and $u_{t4|3} = h_{4|3}(u_{t4}|u_{t3}, \tau_{34}, m_{34}), t = 1, \ldots, T$ only depend on parameters of the first tree, on which we condition on (see Section 2.2.4 for the definition of the h functions). Further, the posterior density factorizes as in (5.25) and we can sample parameters corresponding to different edges independently. In particular $s_{12;3}, m_{12;3}$ are sampled from a static bivariate copula model with posterior density proportional to $\left(\prod_{t=1}^{T} c_{12;3}^{m_{12;3}}(u_{t1|3}, u_{t2|3}; \tau_{12;3})\right) \pi(s_{12;3})\pi(m_{12;3})$, where $\{u_{t1|3}, u_{t2|3}, t = 1, \ldots, T\}$ is interpreted as observed data. For the third tree we obtain

$$\begin{aligned} f(\boldsymbol{\theta}_{\boldsymbol{T_3}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}, \mathbb{U}) &\propto \ell_3(\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}, \boldsymbol{\theta}_{\boldsymbol{T_3}}|\mathbb{U})\pi(\boldsymbol{\theta}_{\boldsymbol{T_3}}) \\ &= \left(\prod_{t=1}^{T} c_{14;23}^{m_{14;23}}(u_{t1|23}, u_{t4|23}; \tau_{14;23})\right) \pi(s_{14;23})\pi(m_{14;23}). \end{aligned} \tag{5.28}$$

As before, $u_{t1|23}$ and $u_{t4|23}$ only depend on parameters from lower trees, on which we condition on. Interpreting $\{u_{t1|23}, u_{t4|23}, t = 1, \ldots, T\}$ as observed data, (5.28) is the posterior density of a static bivariate copula model as introduced in Section 5.2.2.

To obtain samples from the posterior density $f(\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}, \boldsymbol{\theta}_{\boldsymbol{T_3}}|\mathbb{U})$, we utilize the approximations above to first sample $\boldsymbol{\theta}_{\boldsymbol{T_1}}$ from $f(\boldsymbol{\theta}_{\boldsymbol{T_1}}|\mathbb{U})$, then $\boldsymbol{\theta}_{\boldsymbol{T_2}}$ from $f(\boldsymbol{\theta}_{\boldsymbol{T_2}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}, \mathbb{U})$ and then $\boldsymbol{\theta}_{\boldsymbol{T_3}}$ from $f(\boldsymbol{\theta}_{\boldsymbol{T_3}}|\boldsymbol{\theta}_{\boldsymbol{T_2}}, \boldsymbol{\theta}_{\boldsymbol{T_1}}, \mathbb{U})$, i.e. we employ a collapsed Gibbs sampler (Liu (1994)). To draw the parameters we use the sampling procedure of Section 5.2.2. This sampler simulates a Markov chain, where subsequent draws are autocorrelated. So, by applying this sampler we obtain a sample of $\boldsymbol{\theta}_{\boldsymbol{T_i}}$ in the r-th iteration, denoted by $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r}}$, which depends on the previous value $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r-1}}$. While this is not a problem for conventional Gibbs samplers, as in

Metropolis-Hastings within Gibbs, this can lead to undesired samples for collapsed Gibbs schemes as shown by Van Dyk and Jiao (2015). Following Van Dyk and Jiao (2015), we can circumvent this problem by running the updates for $\boldsymbol{\theta}_{\boldsymbol{T_i}}$ with starting value $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r-1}}$ for k iterations. We set $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r}}$ equal to the update obtained in the k-th step. If we choose k large enough, $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r}}$ will be almost independent of $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r-1}}$. Thus, in total, $R \cdot k$ draws are obtained and R draws are stored for each parameter. We obtain the following procedure

– Set starting values $\boldsymbol{\theta}_{\boldsymbol{T_1}}^{\boldsymbol{0}}$, $\boldsymbol{\theta}_{\boldsymbol{T_2}}^{\boldsymbol{0}}$, $\boldsymbol{\theta}_{\boldsymbol{T_3}}^{\boldsymbol{0}}$

– For $r = 1, \dots, R$ do

– For $i = 1, \dots, 3$ do

– Use the sampler of Section 5.2.2 and the corresponding approximations to sample from $f(\boldsymbol{\theta}_{\boldsymbol{T_1}}|\mathrm{U})$ if $i = 1$, from $f(\boldsymbol{\theta}_{\boldsymbol{T_2}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}^{\boldsymbol{r}}, \mathrm{U})$ if $i = 2$ or from $f(\boldsymbol{\theta}_{\boldsymbol{T_3}}|\boldsymbol{\theta}_{\boldsymbol{T_1}}^{\boldsymbol{r}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}^{\boldsymbol{r}}, \mathrm{U})$ if $i = 3$. The sampler is run for k iterations using $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r-1}}$ as starting value. We set $\boldsymbol{\theta}_{\boldsymbol{T_i}}^{\boldsymbol{r}}$ equal to the sample obtained in the k-th iteration.

– The pseudo data for the next tree is constructed utilizing the h functions defined in Section 2.2.4:

For $i = 1$ the pseudo data is determined as $u_{t1|3}^r = h_{1|3}(u_{t1}|u_{t3}; \tau_{13}^r, m_{13}^r)$, $u_{t2|3}^r = h_{2|3}(u_{t2}|u_{t3}; \tau_{23}^r, m_{23}^r)$ and $u_{t4|3}^r = h_{4|3}(u_{t4}|u_{t3}; \tau_{34}^r, m_{34}^r), t = 1, \dots, T$.

For $i = 2$ the pseudo data is determined as $u_{t1|23}^r = h_{1|2;3}(u_{t1|3}^r, u_{t2|3}^r; \tau_{12;3}^r, m_{12;3}^r)$ and $u_{t4|23}^r = h_{4|2;3}(u_{t4|3}^r, u_{t2|3}^r; \tau_{24;3}^r, m_{24;3}^r)$ $, t = 1, \dots, T$.

In this procedure, no point estimates of the parameters $\boldsymbol{\theta}_{\boldsymbol{T_1}}, \boldsymbol{\theta}_{\boldsymbol{T_2}}, \boldsymbol{\theta}_{\boldsymbol{T_3}}$ are required. We can further extend the procedure in the following ways.

- a) The loops over i and r can be exchanged. We can first obtain R samples from the first tree, then obtain R samples from the second tree and then all R samples from the third tree. This is visualized in Figure 5.2. If the tree structure was not known, we could select the tree structure of the first tree and then obtain R samples from the parameters of the first tree. Based on these samples, we can construct the pseudo data, which can be used to select the tree structure of the second tree and so on. Based on the (pseudo) data of a certain tree level, the corresponding structure can be selected as a maximum spanning tree. This is similar to the algorithm of Dissmann et al. (2013).

- b) The parameters of different edges of a tree are sampled independently by utilizing the sampler of Section 5.2.2 for the static copula model. If we did not know that Kendall's τ was static, we could in addition run the sampler of Section 5.2.1 for the dynamic bivariate copula model. We can decide between the dynamic, the static and the independence model as outlined in Section 5.2.3. Here it is important that these decisions for the type of dependence can be made independently for each edge of the tree.

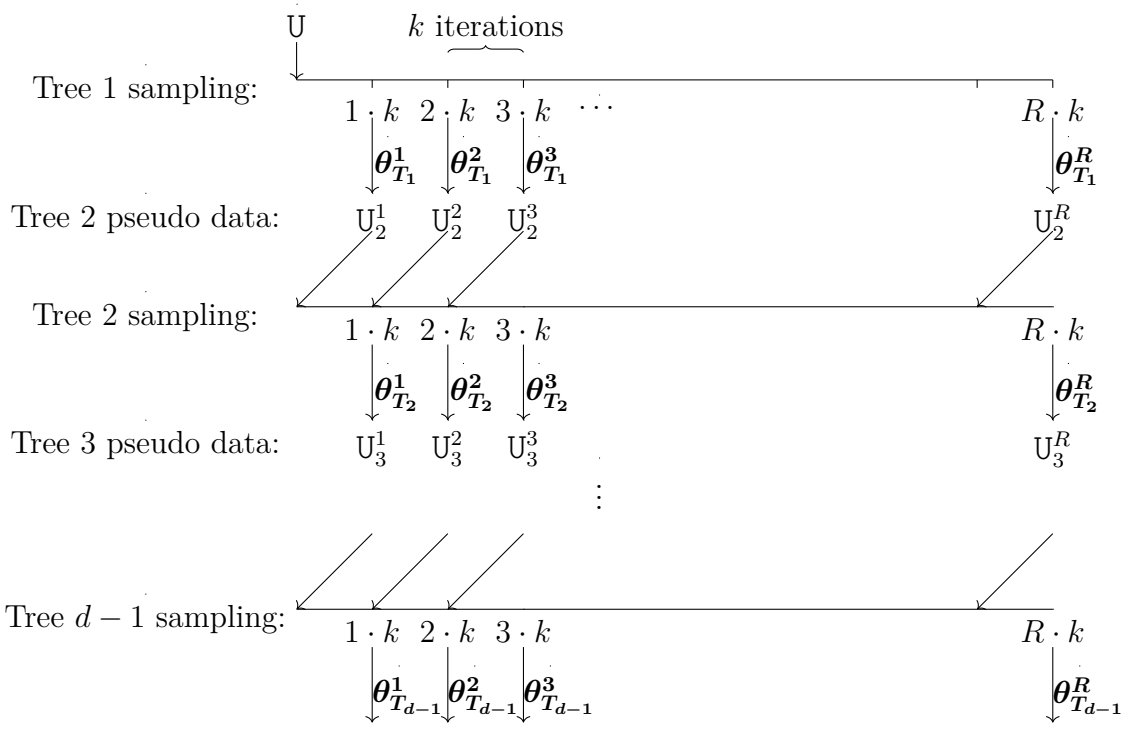

Figure 5.2: Graphical representation of the proposed sampler without selection of the type of dependence and without structure selection. Here $\mathrm{U} \in [0,1]^{T \times d}$ denotes the data matrix used for fitting the model and U_l^r denotes the pseudo data for tree l obtained from parameter draws of the previous tree (tree $l-1$) in iteration r.

The general procedure in d dimensions with vine structure selection

Based on the four-dimensional illustration, we now formulate our procedure for a d-dimensional dynamic vine copula as introduced in Section 5.3.1, incorporating extensions a) and b). The tree structure and the sets $E_i^{dyn}, E_i^{static}, E_i^{ind}$ are selected sequentially as we move up the trees and are fixed at point estimates. In Gruber and Czado (2018) searching among different structures within a full Bayesian procedure resulted in very long computation times for static copula models. Here it would be even worse, since we deal with more complex dynamic pair copulas. Note that both, the sets $E_i^{dyn}, E_i^{static}, E_i^{ind}$ and the tree structure do not change over time.

We propose the following approach with iterations parameter R, burn-in parameter *burnin* and thinning parameter k for structure selection and parameter estimation. Note that, as mentioned above, $R \cdot k$ draws are obtained in total and R iterations are stored for each parameter.

(i) Select the tree structure of the first tree T_1: For all edges e that are allowed in the first tree T_1, i.e. for all pairs (a_e, b_e) with $1 \leq a_e < b_e \leq d$, estimate τ_{a_e,b_e} by the empirical Kendall's τ using $\{u_{t,a_e}, u_{t,b_e}, t = 1, \ldots, T\}$. The structure of tree T_1 is selected as the maximum spanning tree among those edges, where the absolute value of empirical Kendall's τ serves as the corresponding weight.

(ii) (a) For each edge $e \in E_1$ in tree T_1, with corresponding observations $\{u_{t,a_e}, u_{t,b_e}, t =$

$1, \dots, T\}$, run the samplers of Sections 5.2.1 and 5.2.2 for the bivariate dynamic and static copula models. The samplers are run for $R \cdot k$ iterations and we thin the samples with factor k.

(b) For each edge $e \in E_1$, we select among the three bivariate copula models (dynamic, static, independence) as discussed in Section 5.2.3.

(c) For each edge $e \in E_1$, the pseudo data for the next tree is obtained as

$$\begin{aligned} u^r_{t,a_e|b_e} &= h_{a_e|b_e}(u_{t,a_e}|u_{t,b_e}; \tau^r_{t,e}, m^r_e), \\ u^r_{t,b_e|a_e} &= h_{b_e|a_e}(u_{t,b_e}|u_{t,a_e}; \tau^r_{t,e}, m^r_e), \end{aligned} \tag{5.29}$$

for $r = 1, \dots, R, t = 1, \dots, T$, if the dynamic copula was selected for edge e. If the static copula was selected we replace $\tau^r_{t,e}$ by τ^r_e in (5.29). For the independence copula model we use $u^r_{t,a_e|b_e} = u_{t,a_e}$ and $u^r_{t,b_e|a_e} = u_{t,b_e}$.

(iii) Set $l = 2$.

(iv) Select the tree structure of the l-th tree T_l: For all edges that are allowed in T_l according to the proximity condition, estimate Kendall's τ of edge $e = (a_e, b_e; \boldsymbol{D_e})$ denoted by $\tau_{a_e,b_e;\boldsymbol{D_e}}$ by the empirical Kendall's τ. Therefore we use posterior mode estimates of the pseudo data $\{\hat{u}_{t,a_e|\boldsymbol{D_e}}, \hat{u}_{t,b_e|\boldsymbol{D_e}}, t = 1, \dots, T\}$, where $\hat{u}_{t,a_e|\boldsymbol{D_e}}$ is the mode estimate of the univariate kernel density estimate of $\{u^r_{t,a_e|\boldsymbol{D_e}}, r = burnin + 1, \dots, R\}$ and $\hat{u}_{t,b_e|\boldsymbol{D_e}}$ is obtained similarly. Here $\{\hat{u}_{t,a_e|\boldsymbol{D_e}}, \hat{u}_{t,b_e|\boldsymbol{D_e}}, t = 1, \dots, T\}$ are treated as an iid sample for the estimation of $\tau_{a_e,b_e;\boldsymbol{D_e}}$. The structure of tree T_l is selected as the maximum spanning tree among those edges, where the absolute value of empirical Kendall's τ serves as the corresponding weight.

(v) (a) For each edge $e \in E_l$ in tree T_l, with corresponding pseudo data $\{u^r_{t,a_e|\boldsymbol{D_e}}, u^r_{t,b_e|\boldsymbol{D_e}}, t = 1, \dots, T, r = 1, \dots, R\}$, obtain R samples (based on a total of $R \cdot k$ MCMC draws) for the bivariate dynamic and static copula models utilizing the approaches of Sections 5.2.1 and 5.2.2. For the static bivariate copula we proceed as follows for an edge e.

- Set starting values s^0_e, m^0_e.
- For $r = 1, \dots, R$: obtain k samples of s_e, m_e for a static bivariate copula model based on data $\{u^r_{t,a_e|\boldsymbol{D_e}}, u^r_{t,b_e|\boldsymbol{D_e}}, t = 1, \dots, T\}$. We use s^{r-1}_e, m^{r-1}_e as starting value and set s^r_e, m^r_e to the sample obtained in the k-th iteration.

For the dynamic copula model we proceed similarly.

(b) We select for each edge $e \in E_l$ among the three bivariate copula models (dynamic, static, independence) as explained in Section 5.2.3.

(c) For each edge $e \in E_l$, the pseudo data for the next tree is obtained as

$$\begin{aligned} u^r_{t,a_e|b_e \cup \boldsymbol{D_e}} &= h_{a_e|b_e;\boldsymbol{D_e}}(u^r_{t,a_e|\boldsymbol{D_e}}|u^r_{t,b_e|\boldsymbol{D_e}}; \tau^r_{t,e}, m^r_e), \\ u^r_{t,b_e|a_e \cup \boldsymbol{D_e}} &= h_{b_e|a_e;\boldsymbol{D_e}}(u^r_{t,b_e|\boldsymbol{D_e}}|u^r_{t,a_e|\boldsymbol{D_e}}; \tau^r_{t,e}, m^r_e), \end{aligned} \tag{5.30}$$

for $r = 1, \dots, R, t = 1, \dots, T$, if the dynamic copula was selected for edge e. If the static copula was selected we replace $\tau^r_{t,e}$ by τ^r_e in (5.30). For the independence copula model we set $u^r_{t,a_e|b_e \cup \boldsymbol{D_e}} = u^r_{t,a_e|\boldsymbol{D_e}}$ and $u^r_{t,b_e|a_e \cup \boldsymbol{D_e}} = u^r_{t,b_e|\boldsymbol{D_e}}$.

(vi) If $l < d - 1$, set $l = l + 1$ and go to 4.

Runtime and scalability

The MCMC samplers in step (ii) (a) and the ones in step (v) (a) can be run in parallel, respectively. The MCMC samplers are the main drivers for the runtime and therefore parallelization speeds up computation a lot. When enough cores, i.e. at least $d-1$ cores for d-dimensional data, are available, we observed that the computation time for one tree is no more than 40 minutes for time series data of length $T = 1000$ with $R = 1100, burnin = 100, k = 25$, independently of d. For estimating a full vine (i.e. a vine without truncation), we expect that the computation time grows roughly linearly with the dimension d and a full vine in 11 dimension (containing 10 trees) should take no more than $10 \cdot 40$ minutes. But in higher dimensions it is often not necessary to estimate all trees, e.g. we expect the runtime for a 100-dimensional vine, truncated after the 10-th tree, to be not much more than $10 \cdot 40$ minutes. Thus, in combination with truncation, we expect our method to scale very well to higher dimensions.

5.3.3 Simulation study

With this simulation study, we aim to obtain a first impression of the ability of the procedure proposed in Section 5.3.2 to recover trajectories of Kendall's τ and of the ability to select copula families and the type of dependence (dynamic, static, zero). A more extensive simulation study to investigate the potential of the novel approach in more detail is planned for the future.

First, we assume the tree structure to be known and the steps for the vine structure selection in our procedure are left out. This allows to compare the true and estimated Kendall's τ values for each pair copula. Afterwards, we allow for vine structure selection. In this case our procedure might select different tree structures. Thus, the true trajectories of Kendall's τ and the copula family for some pair copulas included in the selected vine structure may not be directly known. We deal with this case by comparing simulations from the true and the estimated model. In addition, we compare average log-likelihoods of true and estimated models.

Known tree structure

We consider the tree structure presented in Section 2.2.4 in Figure 2.2. The corresponding families are chosen from the set: {Independence, Gaussian, Student t(df=2), Student t(df=4), Student t(df=8), eGumbel, eClayton}. For each pair copula we simulate one trajectory of length $T = 1000$ for Kendall's τ from an AR(1) process. The chosen copula families and the parameters of the AR(1) processes are specified in Appendix C.1. We keep the tree structure, the choice of the families and the trajectories for Kendall's τ fixed and simulate 100 times from this model.

For each of the 100 simulated data sets, we run the algorithm proposed in Section 5.3.2 without tree structure selection. Within our sequential procedure, we set $R = 1100$, $k = 25$ and $burnin = 100$. This means that within the procedure $1100 \cdot 25$ draws have been obtained for each parameter, whereas 1100 iterations are stored. Of these 1100 stored iterations the first 100 are discarded for burn-in. From Table 5.3 we see that for each pair copula in the first two trees the correct family was selected in at least 94 out of 100 cases. In Tree 3, the two independence copulas were detected in 69 and 100 out of 100

cases. The static copula in Tree 3 was detected in 88 out of 100 cases. The independence copulas in Trees 4 and 5 were detected every time. Here we count a Student t copula as correctly detected if it was selected as a Student t copula, independently of the degrees of freedom parameter. Table 5.3 also shows how often the correct type of dependence (dynamic, static, zero) was selected. For one pair copula in the first tree, the correct type was only detected in 75 out of 100 cases. The corresponding Kendall's τ is shown in the fifth row, third column in Figure 5.3. We see that this Kendall's τ does not change a lot over time. So it is difficult to distinguish between the dynamic and the static model for this pair copula. Except for this pair copula, the correct type of dependence of pair copulas in the first two trees was detected in at least 93 out of 100 cases. In trees, higher than Tree 2, the correct type was selected in at least 69 out of 100 cases. We think that these are reasonable results for the selection of the family and of the type of dependence for the pair copulas, that make up the dynamic vine copula. In addition, Figures 5.3 and 5.4 illustrate that our procedure can recover the simulated trajectories of Kendall's τ. In these figures, we show marginal (univariate) posterior mode estimates of Kendall's τ parameters, which will be utilized later (Section 5.4) as point estimates.

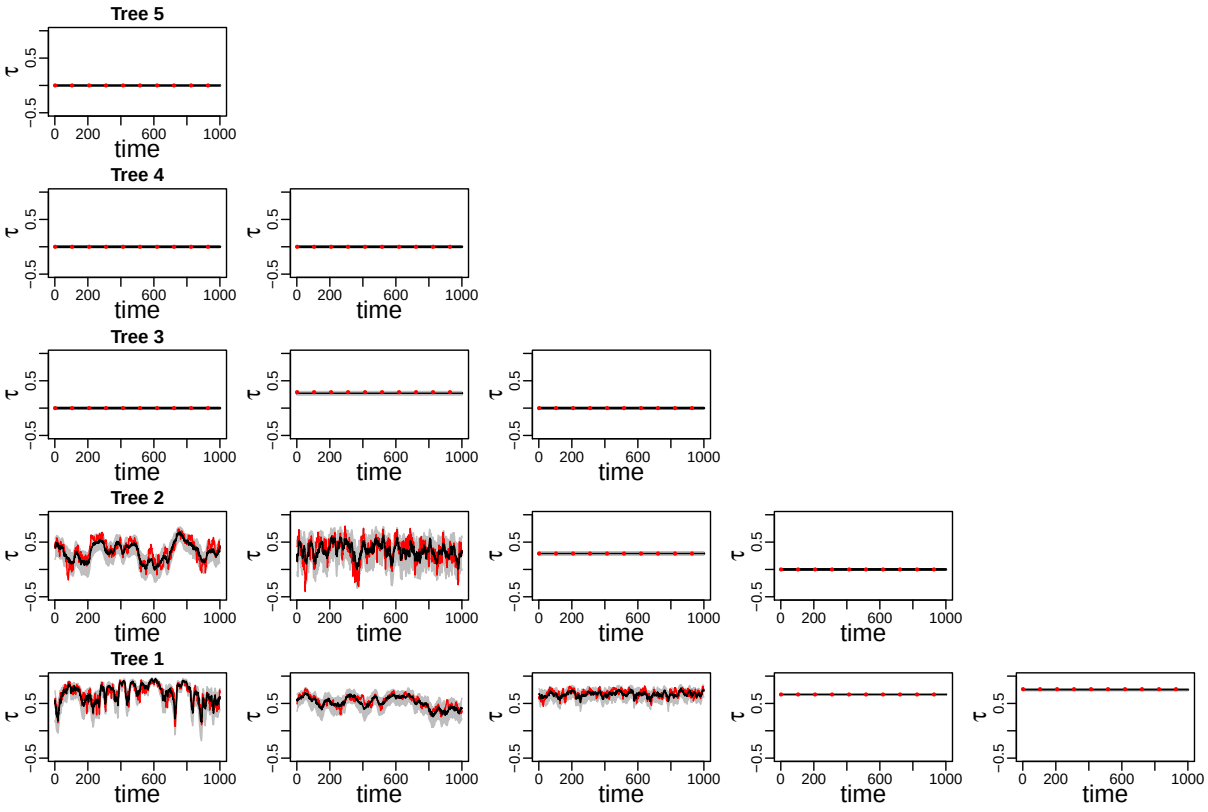

Figure 5.3: This plot corresponds to a fitted model for one simulated data set. Posterior mode estimates of Kendall's τ at time t are plotted against t for each pair copula (black lines). The posterior mode estimates are obtained from marginal (univariate) kernel density estimates of the corresponding Kendall's τ parameter. A 90% credible region constructed from the estimated 5% and 95% posterior quantiles is added in grey. True values of Kendall's τ are added in red.

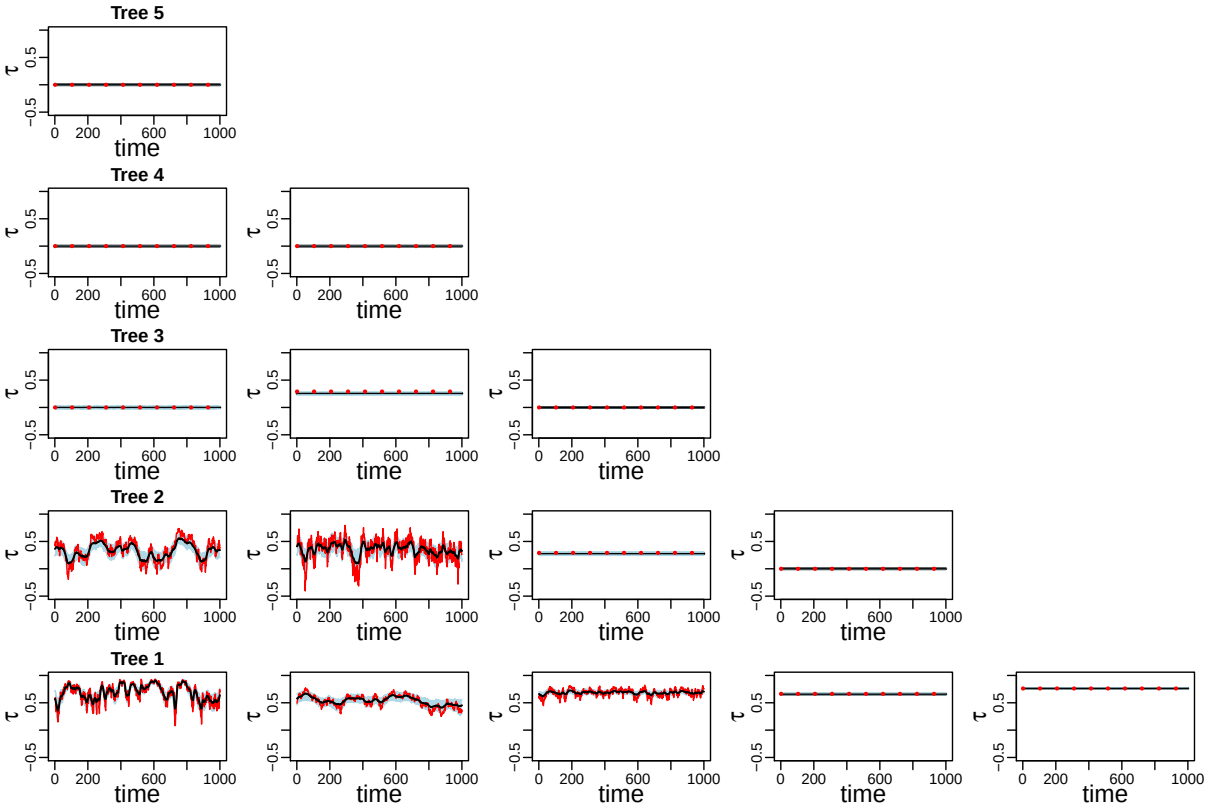

Figure 5.4: In this plot we consider 100 estimated models. The mean of 100 posterior mode estimates of Kendall's τ at time t is plotted against t for each pair copula (black line). The posterior mode estimates are obtained from marginal (univariate) kernel density estimates of the corresponding Kendall's τ parameter. The blue region is constructed from the empirical 5% and 95% quantiles of the 100 posterior mode estimates. True values of Kendall's τ are added in red.

Tree	Copula family					Type of dependence				
5	100					100				
4	100	100				100	100			
3	69	88	100			69	78	100		
2	100	100	94	100		100	96	97	100	
1	94	100	100	99	100	100	93	75	98	97

Table 5.3: This table shows how often the correct copula family and how often the correct type of dependence (dynamic, static, zero) was selected out of the 100 simulations for each pair copula of the dynamic vine copula. There are $6-i$ pair copulas in the i-th tree. The selected copula family for an edge e is the marginal posterior mode estimate of m_e, i.e. the family that occurs most frequently among the posterior samples for m_e.

Unknown tree structure

We use the same 100 simulated data sets as in the case with known tree structure but allow here for structure selection within our procedure. As already mentioned, evaluating

our results is not straightforward in this case. Our estimated model may contain pair copulas for which we do not know the true copula families and Kendall's τ values directly. In this case we simulate 500 times from the true model and from the estimated model. Then we can calculate empirical Kendall's τ values for each of the $\frac{6 \cdot 5}{2} = 15$ pairs $(U_1, U_2), (U_1, U_3), \ldots$. We compare trajectories of the empirical Kendall's τ values in Figures 5.5 and 5.6. These trajectories look similar for the true and the estimated models. We see that also our procedure with structure selection is able to recover the simulated trajectories of Kendall's τ.

For further evaluation of the proposed procedure, we compare log-likelihoods of estimated and true models, as in Gruber and Czado (2015). To save computation time, we evaluate the likelihoods of estimated models based on point estimates (marginal posterior mode estimates) of the parameters, instead of evaluating the likelihoods for all posterior draws. The average log-likelihood of models without structure selection was 94% of the log-likelihood of the true model, whereas the log-likelihood of the models estimated with structure selection was on average 89% of the log-likelihood of the true model. It is not surprising that we perform a bit better if we assume the vine structure to be known. But the difference is not very big and in both cases, with and without vine structure selection, we obtain reasonable results. For further comparison, we also estimated dynamic C-vine and D-vine copulas. Therefore we just restrict our structure selection procedure in Section 5.3.2 to C-vine and D-vine structures, respectively. The dynamic C-vine and D-vine copulas achieved 86% and 88% of the log-likelihood of the true model, respectively. Thus, in this scenario, allowing for general vine structures improves the fit compared to restricting the structure to C-vines or D-vines.

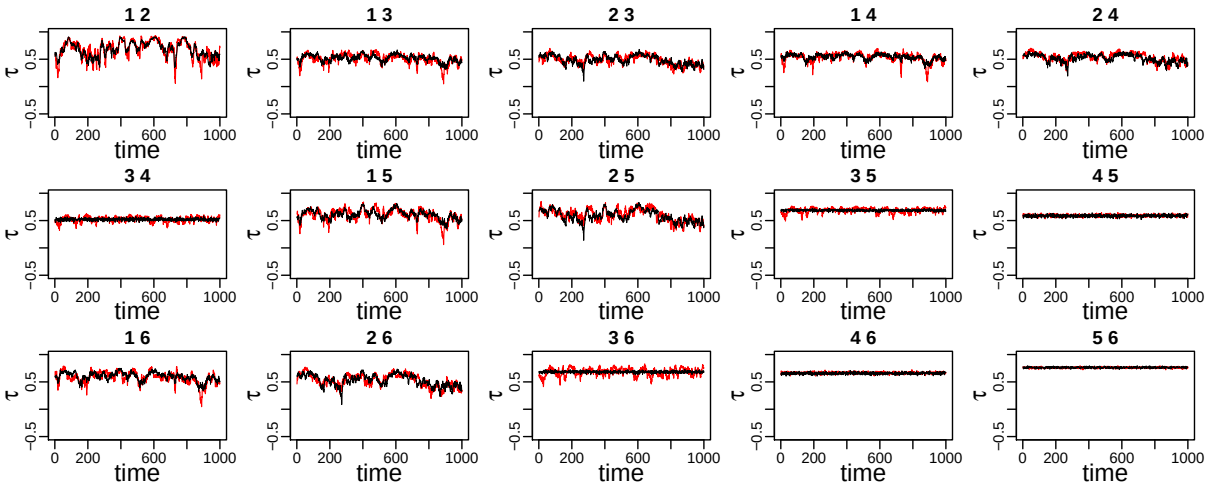

Figure 5.5: This plot corresponds to a fitted model for one simulated data set. The empirical unconditional Kendall's τ estimate at time t, obtained from simulations from the fitted model, is plotted against t for each pair $(U_i, U_j), i, j \in \{1, \ldots, 6\}, i < j$ (black line). True Kendall's τ values determined by simulating from the true model 500 times are added in red.

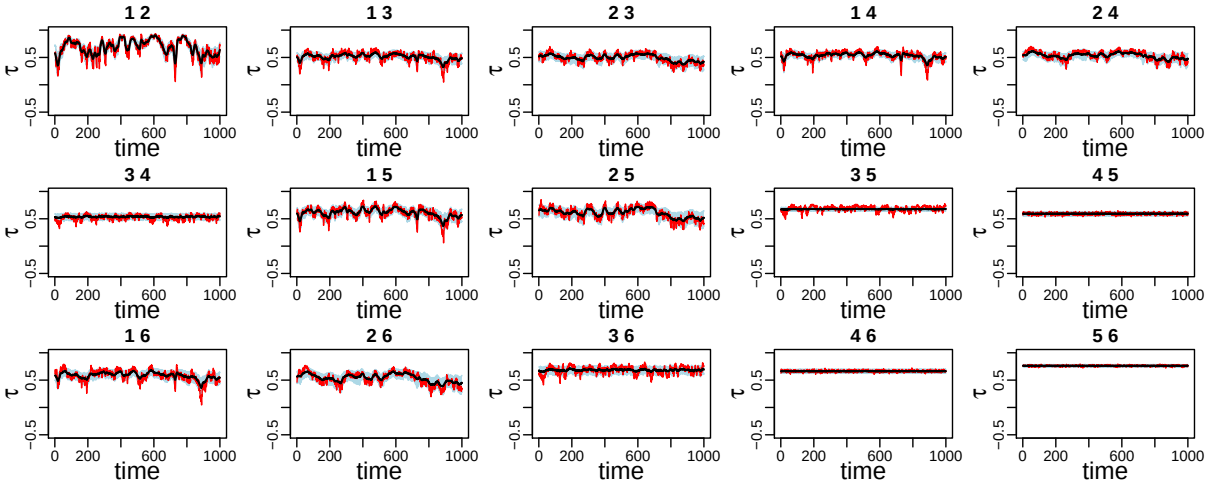

Figure 5.6: In this plot we consider 100 fitted models. The mean of 100 empirical estimates of the unconditional Kendall's τ at time t, obtained from simulations from the fitted models, is plotted against t for each pair $(U_i, U_j), i, j \in \{1, \dots, 6\}, i < j$ (black line). The blue region is constructed from the 5% and 95% empirical quantiles of the 100 empirical estimates of Kendall's τ. True Kendall's τ values determined by simulating from the true model 500 times are added in red.

5.4 Application: Dynamic exchange rates dependence

We employ the proposed dynamic vine copula model to model the dependence among 21 exchange rates with respect to the US Dollar (USD). For this we use data obtained from the FRED database of the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/categories/94) which comprises daily log returns of 21 exchange rates with respect to the USD from 2007 to 2018, resulting in 3130 observations. The 21 currencies and their ticker symbols are summarized in Appendix C.2. We estimate our model based on the first 1500 observations and evaluate its predictive performance based on the remaining 1630 observations. First, the data is demeaned based on the first 1500 observations and we collect the demeaned log returns in the data matrix $\mathrm{Y} = (y_{tj})_{t=1,\dots,3130, j=1,\dots,21} \in \mathbb{R}^{3130\times 21}$.

For the marginals we use skew Student t stochastic volatility models (see Section 2.1.2), i.e. we assume that

$$\begin{aligned} Y_{tj} &= \exp\left(\frac{s_{tj}^{st}}{2}\right)\epsilon_{tj}^{st} \\ s_{tj}^{st} &= \mu_j^{st} + \phi_j^{st}(s_{t-1j}^{st} - \mu_j^{st}) + \sigma_j^{st}\eta_{tj}^{st} \end{aligned} \tag{5.31}$$

with $\eta_{tj}^{st} \sim N(0,1)$ independently, $\mu_j^{st} \in \mathbb{R}, \phi_j^{st} \in (-1,1), \sigma_j^{st} \in (0,\infty)$, $s_{0j}^{st}|\mu_j^{st}, \phi_j^{st}, \sigma_j^{st} \sim N\left(\mu_j^{st}, \frac{(\sigma_j^{st})^2}{1-(\phi_j^{st})^2}\right)$, $\epsilon_{tj}^{st}|\alpha_j^{st}, df_j^{st} \sim sst(\epsilon_{tj}^{st}|\alpha_j^{st}, df_j^{st})$ independently, $\alpha_j^{st} \in \mathbb{R}, df_j^{st} \in (2,\infty)$ for $t = 1, \dots, 3130$. Further, the same prior distributions as in Section 2.1.2 are utilized. The joint distribution among the errors is modeled by the proposed dynamic vine copula model. We follow ideas of the two-step approach, commonly used in copula modeling,

and assume independence among the errors ϵ_{tj}^{st} for estimating the margins. But instead of collapsing parameters of the marginal skew Student t stochastic volatility models to point estimates and obtain the copula data based on these point estimates, we follow ideas from Section 5.3.2. For each of the 21 marginal time series $y_{1j}, \dots, y_{1500j}$, we estimate a skew Student t stochastic volatility model as explained in Section 4.4. The sampler is run for $1100 \cdot 25$ iterations and then we thin the samples with factor 25. The parameter draws of the skewness parameters, of the degrees of freedom parameters and of the latent log variances are denoted by $(\alpha_j^{st})^r, (df_j^{st})^r, (s_{0j}^{st})^r, \dots, (s_{1500j}^{st})^r, r = 1, \dots, 1100$. For each parameter draw, we obtain pseudo copula data as follows

$$u_{tj}^r = ssT\left(y_{tj} \exp\left(-\frac{(s_{tj}^{st})^r}{2}\right) \middle| (\alpha_j^{st})^r, (df_j^{st})^r\right) \tag{5.32}$$

for $t = 1, \dots, 1500$, $j = 1, \dots, 21$, $r = 1, \dots, 1100$. Here ssT denotes the standardized skew Student t distribution function (see Section 2.1.2). Based on these pseudo copula data sets, we fit a dynamic vine copula model. The algorithm of Section 5.3.2 is slightly modified. We start with Step (iii) and set $l = 1$, since we fit our model with a collection of copula data sets instead of only one copula data set. Further, we set $R = 1100, k = 25$ and $burnin = 100$. The copula families are selected from the following set $\mathcal{M} = \{$Independence, Gaussian, Student t(df=2), Student t(df=4), Student t(df=8), eGumbel, eClayton$\}$. The estimated dynamic vine copula model is now analyzed in more detail.

The first tree of the selected vine tree structure is shown in Figure 5.7. We see that some currencies that are connected by an edge are from countries of the same region. For example, the currencies GBP/USD (British Pound to USD) and DKK/USD (Danish Krone to USD) are connected to the EUR/USD (Euro to USD). Since the vine structure is selected as the maximum spanning tree, where the absolute value of Kendall's τ serves as weight, this indicates high dependence among those currencies. Further, we see that the selected vine structure is neither a C-vine nor a D-vine structure. The generalization of C-vine and D-vine structures to R-vine structures seems to be necessary.

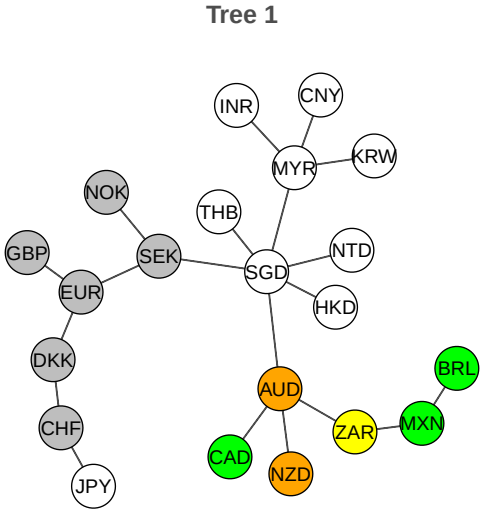

Figure 5.7: The first tree of the vine tree structure selected for the 21-dimensional exchange rates data set. Nodes which belong to the same region have the same color (Europe: grey, Asia: white, America: green, Australasia: orange, Africa: yellow).

In Table 5.4 we show the selected types of dependence per tree level. Above tree nine, all selected copulas are equal to the independence copula, i.e. the type of dependence is estimated to be zero. Further, we see that only few static copulas were selected. The number of dynamic copulas selected decreases as we move up to higher tree levels. In total $\frac{20 \cdot 21}{2} = 210$ pair copulas are estimated, of which 150 were set equal to the independence copula. Our proposed procedure is able to detect sparse structures. Note that the level of sparsity can be increased by adjusting the selection of the type of dependence accordingly. As mentioned in Section 5.2.3, we decide for the more complex type of dependence if the WAIC of the more complex model is at least 2 standard errors smaller. By increasing 2 to for example 4 standard errors, we achieve more sparsity. This might be interesting in higher dimensional settings. Since for most pair copulas in the first trees the dynamic type of dependence is selected, a vine copula model with static dependence might not be appropriate for those selected pair copula terms with time-varying dependence.

Tree	Dynamic	Static	Zero
1	18	2	0
2	16	1	2
3	7	0	11
4	6	1	10
5	2	0	14
6	3	0	12
7	2	0	12
8	1	0	12
9	1	0	11

Table 5.4: We show how often the different types of dependence (dynamic, static, zero) were selected per tree level for the first nine trees.

Figure 5.8 shows how the dynamic Kendall's τ values evolve over time. We see that the dependence between the exchange rates AUD/USD (Australian Dollar to USD) and ZAR/USD (South African Rand to USD) varies more in 2007 and 2008, during the financial crisis, and remains almost constant after that period. Further, we observe that the Kendall's τ between SGD/USD (Singapore Dollar to USD) and THB/USD (Thai Baht to USD) is close to zero in 2007 and then starts to increase after 2007. The dependence between DKK/USD (Danish Krone to USD) and CHF/USD (Swiss Franc to USD) is rather high but decreases in 2010 and reaches its lowest point in 2011. This might be the effect of the introduction of the cap on the Swiss Franc on 6 September 2011 by the Swiss National Bank. The minimum exchange rate was set at 1.2 CHF (Swiss Franc) per EUR (Euro). The second row of Figure 5.8 shows fitted conditional Kendall's τ values. For example, we see how the Kendall'sτ of the exchange rates DKK/USD (Danish Krone to USD) and JPY/USD (Japanese Yen to USD) conditional on CHF/USD (Swiss Franc to USD) evolves over time. The conditional dependence mostly varies between -0.5 and 0. We also provide quantification of uncertainty of the Kendall's τ values through credible intervals. This is an advantage of our Bayesian approach compared to the frequentist approach of Almeida et al. (2016) for dynamic D-vine copulas, where uncertainty quantification was not provided.

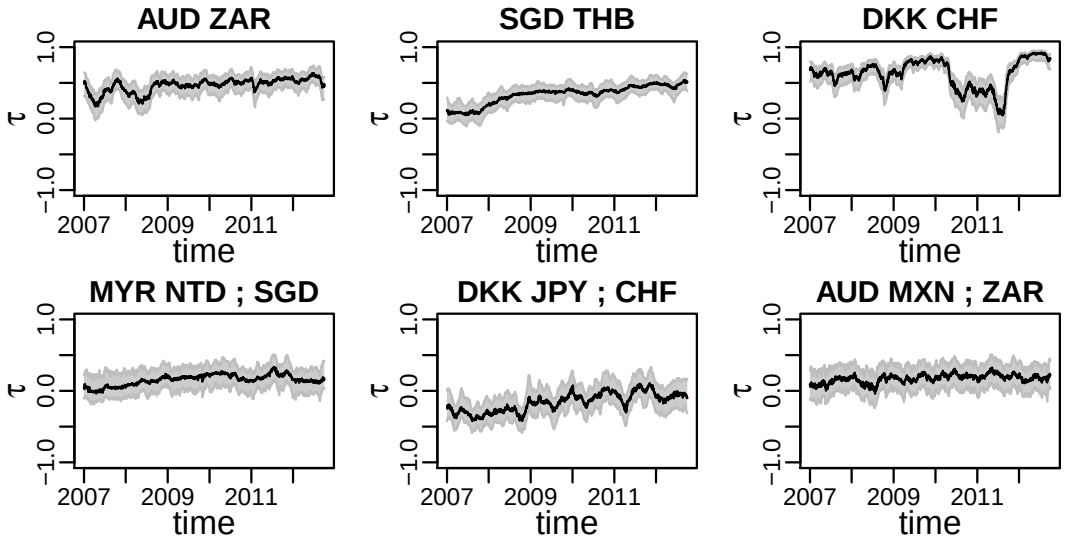

Figure 5.8: Visualization of the dynamic Kendall's τ for some chosen pair copulas of the dynamic vine copula model estimated for the 21-dimensional exchange rates data set. Rows 1 and 2 of this plot correspond to pair copulas in Trees 1 and 2, respectively. The black line shows marginal posterior mode estimates of Kendall's τ plotted against time t. A 90% credible region constructed from the estimated 5% and 95% posterior quantiles is added in grey.

As already mentioned, the proposed dynamic vine copula model can be seen as a generalization of static vine copulas and as a generalization of the dynamic C-vine and D-vine copula models of Goel and Mehra (2019) and Almeida et al. (2016). We would like to further support the hypothesis that our proposed dynamic vine copula model is a needed generalization and is able to describe the dependence structure more appropriate than its competitor models: a dynamic C-vine, a dynamic D-vine and a static vine copula. Therefore we compare these models with respect to their predictive accuracy. To estimate the dynamic C-vine and D-vine copula model, we adjust our structure selection procedure in Section 5.3.2 accordingly. The families are selected from the same set $\mathcal{M}$ that we used for the dynamic vine copula. The static vine copula model is estimated with the algorithm of Dissmann et al. (2013), as implemented in the R-package `rvinecopulib` (Nagler and Vatter (2018)). Here we allow for all parametric copula families that are implemented in the `rvinecopulib` package. For all models we use skew Student t stochastic volatility models for the margins. For the static vine copula model the pseudo copula data is obtained by fixing the parameters of the skew Student t stochastic volatility models at marginal posterior mode estimates. The competitor models are also estimated based on the first 1500 observations of our data. For all models we obtain one-day ahead predictive scores for the other 1630 days in our data set. We proceed as in Sections 3.4 and 4.4, i.e. instead of refitting the models 1629 times, we keep static model parameters fixed at point estimates (marginal posterior mode estimates for the dynamic models, maximum likelihood estimates for the static vine copula) and only update dynamic parameters (see also Appendix F). This reduces computation time a lot. Similar to Section 4.4, we evaluate the corresponding densities at point estimates instead of averaging over all posterior draws and obtain pseudo log predictive scores (plps). The plps at time $t > 1500$ has the following

structure

$$\text{plps}_t = \ln(\hat{c}_t(\hat{u}_{t1}, \dots, \hat{u}_{t21})) + \sum_{j=1}^{21} \ln\left(sst\left(y_{tj} \exp\left(-\frac{\hat{s}_{tj}^{st}}{2}\right) \middle| \hat{\alpha}_j^{st}, \hat{df}_j^{st} \right) \right) - \frac{\hat{s}_{tj}^{st}}{2}, \tag{5.33}$$

with $\hat{u}_{tj} = ssT\left(y_{tj} \exp\left(-\frac{\hat{s}_{tj}^{st}}{2}\right) \middle| \hat{\alpha}_j^{st}, \hat{df}_j^{st} \right)$ for $j = 1, \dots, 21$. Here $\hat{c}_t$ is the estimated copula density of one of the four considered models obtained by fixing the corresponding parameters at point estimates (marginal posterior mode estimates for the dynamic models, maximum likelihood estimates for the static vine copula). Note that for a dynamic parameter, we use a one-day ahead point forecast in (5.33), while for a static parameter we use the estimate obtained from the first 1500 observations. For example $\hat{s}_{tj}^{st}$ is a one-day ahead point forecast and $\hat{\alpha}_j^{st}, \hat{df}_j^{st}$ are marginal posterior mode estimates obtained from the first 1500 observations. The contribution $\sum_{j=1}^{21} \ln\left(sst\left(y_{tj} \exp\left(-\frac{\hat{s}_{tj}^{st}}{2}\right) \middle| \hat{\alpha}_j^{st}, \hat{df}_j^{st} \right) \right) - \frac{\hat{s}_{tj}^{st}}{2}$ of the margins is the same for all considered models. So we compare the models with respect to the copula contributions $\ln(\hat{c}(\hat{u}_{t1}, \dots, \hat{u}_{t21}))$ to which we refer as copula plps. Note that a higher (copula) plps is an indication for better forecasting accuracy.

Table 5.5 shows the cumulative copula plps, i.e. the sum over all 1630 copula plps. We see that the two vine copula models with flexible tree structure outperform the dynamic C-vine and D-vine copulas. Further, the dynamic vine copula, for which the selected structure deviates clearly from a C-vine and a D-vine structure and for which many pair copulas have a dynamic type of dependence, provides the most accurate forecasts. Our conclusion is that the dynamic vine copula model provides a useful generalization of static vine copula models as well as of dynamic C-vine and D-vine copula models.

	Dynamic vine	Dynamic C-vine	Dynamic D-vine	Static vine
copula plps	**11643**	11132	11126	11267

Table 5.5: Cumulative one-day ahead copula plps for the four considered models: Dynamic vine, dynamic C-vine, dynamic D-vine and static vine copula.

5.5 Conclusion and future research

We introduced a class of dynamic vine copula models and provided a novel Bayesian estimation procedure based on an approximation of the posterior distribution allowing for simplification of the sampler. Here we allowed for the selection of the pair copula family, the selection of the type of dependence for each pair copula term and the sequential selection of a static (time-constant) vine structure. The application showed that the dynamic vine copula model is a useful extension of static vine copulas and of dynamic C-vine and D-vine copulas.

Our estimation procedure propagates uncertainty of copula parameter estimation from lower to higher trees. But the type of dependence is selected with WAIC and then fixed before we move to the next tree. Instead of using the WAIC or any other information criteria, shrinkage priors as proposed by Bitto and Frühwirth-Schnatter (2019) that allow

to shrink dynamic parameters to static ones might be an interesting alternative to be studied in future research.

One restriction of our approach is that both the vine tree structure as well as the pair copula families are assumed not to change over time. In the future we are interested in overcoming these restrictions.

Further, it would be interesting to study in more detail how the dynamic vine copula model performs in situations, where appropriate dependence modeling is crucial, for example in financial risk management. The dynamic vine copula model might lead to more accurate value at risk predictions than those obtained from static vine copula models. Another example is pairs trading. Stübinger et al. (2018) showed that profitable trading strategies can be constructed with static vine copula models. These strategies might be improved by allowing for dynamic dependencies.

Lastly, we think that the ability of the vine copula framework to scale bivariate copula models to copula models of arbitrary dimensions has not been fully exploited yet. We are sure that there is a variety of useful extensions of static vine copula models that build on sophisticated bivariate copula models. For example, one could allow for bivariate dynamic copulas with more than one parameter, such as the BB1 family, or for bivariate dynamic mixture copulas as studied in Section 4.4, where both mixture components share the same dynamic on Kendall's τ. Alternatively, one could also study bivariate mixture copula models with one dynamic and one static component.

6 A univariate copula state space model to predict air pollution in Beijing

This chapter is a reproduction of Kreuzer et al. (2019a) with minor changes.

6.1 Introduction

Air pollution has serious effects on human health. It is related to several cardiovascular and respiratory illnesses. It even shortens life expectancy and has severe effects on the economy (Song et al. (2017)). The great smog of London in winter 1952 is an extreme example for the effects of air pollution. Bell and Davis (2001) estimate that the smog caused about 12000 excess deaths in the following year.

It is clear that accurate modeling and prediction of air pollution concentration is of high importance. Statistical techniques have been proposed for this purpose. Several of these approaches focus on the concentration of PM2.5, atmospheric particulate matter with a diameter of less than 2.5 micrometers. Sahu et al. (2006) employ a hierarchical Bayesian space-time model to predict PM2.5 concentrations in the US. Sahu and Mardia (2005) use a Bayesian kriged Kalman filtering approach for short term prediction of PM2.5 levels in New York City. Ippoliti et al. (2012) rely on a linear Gaussian state space model to forecast pollution concentrations in Italy.

We propose a nonlinear non-Gaussian state space model based on copulas with a dynamic latent smoothing effect. The state variables can be interpreted as non-measured effects, not captured by the covariates. Thus, we can identify time points with unusual high levels of air pollution, which cannot be explained by the covariates. We will demonstrate that the proposed methodology describes the time-dynamics of the air pollution data more accurate than a linear Gaussian state space model. In contrast to just smoothing observations, our approach allows to investigate the effect of changed climate conditions on the predicted PM2.5 levels. Illustrations of such climate simulations are also given.

Linear Gaussian state space models

A linear Gaussian state space model (see Section 2.1.1) with univariate state and observation equations can be formulated as follows

$$Z_t = \rho_{obs,t} w_t + \sigma_{obs,t} \eta_{obs,t} \tag{6.1}$$

$$w_t = \rho_{lat,t} w_{t-1} + \sigma_{lat,t} \eta_{lat,t} \tag{6.2}$$

for $t = 1, \dots, T$. Here, Z_t is a random variable corresponding to the observation at time t, w_t is an unobserved univariate state and $\eta_{obs,t}$, $\eta_{lat,t} \sim N(0,1)$ independently for $t = 1, \dots, T$. Further, it holds that $\rho_{obs,t} \in (-1,1)$, $\rho_{lat,t} \in (-1,1)$, $\sigma_{obs,t} \in (0,\infty)$ and $\sigma_{lat,t} \in (0,\infty)$. It is also assumed that $w_0 \sim N(\mu_{lat,0}, \sigma^2_{lat,0})$, where $\mu_{lat,0}$ and $\sigma_{lat,0}$ are generally known.

The linear Gaussian state space model can also be expressed as

$$Z_t|w_t, \rho_{obs,t}, \sigma_{obs,t} \sim N\left(\rho_{obs,t} w_t, \sigma^2_{obs,t}\right) \tag{6.3}$$

$$w_t|w_{t-1}, \rho_{lat,t}, \sigma_{lat,t} \sim N\left(\rho_{lat,t} w_{t-1}, \sigma^2_{lat,t}\right). \tag{6.4}$$

Beijing ambient air pollution data

In this paper, we aim at accurately estimating and predicting the concentration of airborne particulate matter using a flexible state space model. We consider a data set of hourly PM2.5 readings (μg/m^3) and meteorological measurements, such as dew point (DEWP, degrees Celsius), temperature (TEMP, degrees Celsius), pressure (PRES, hPa), wind direction (CBWD, taking values: northwest (NW), northeast (NE), southeast (SE) and calm and variable (CV)), cumulated wind speed (IWS, m/s) and precipitations (PREC), collected in Beijing in 2014, and we split the data into 12 monthly sub-sets [1] (Liang et al. (2015)). This allows us to adjust the model over time periods. In order to consider the effects of meteorological conditions on airborne particulate matter concentrations, we assume a generalized additive model (GAM) (Hastie and Tibshirani (1986)). More precisely, we suppose that, for each month, the relationship between the logarithm of PM2.5 concentrations Y_t and the covariates $\boldsymbol{x}_t$ is described by a GAM, such that

$$Y_t = f(\boldsymbol{x}_t) + \sigma\varepsilon_t \tag{6.5}$$

for $t = 1, \dots, T$ (T is the number of monthly observations), where $\boldsymbol{x}_t$ contains the meteorological covariates and seasonal covariates capturing within-day and -week patterns. Further, $f(\cdot)$ is a smooth function of the covariates, expressing the mean of the GAM and $\varepsilon_t \sim N(0,1)$ iid. For estimation we make use of the two-step approach which is commonly used for copula models: we first estimate the GAM, fix the GAM parameters at point estimates, and then estimate the copula model (see Section 2.2.1). We define the standardized errors Z_t as

$$Z_t = \frac{Y_t - f(\boldsymbol{x}_t)}{\sigma} \tag{6.6}$$

for $t = 1, \dots, T$. This step allows us to account for weather and seasonal patterns. High values of Z_t are then of interest to detect unusual high levels of pollution so far not accounted for. Using the estimates $\hat{f}(\boldsymbol{x}_t)$ and $\hat{\sigma}$ of the GAM, we obtain approximately standard normal data $\hat{z}_t$ as

$$\hat{z}_t = \frac{y_t - \hat{f}(\boldsymbol{x}_t)}{\hat{\sigma}} \tag{6.7}$$

for $t = 1, \dots, T$, where y_t is an observation of Y_t. The empirical autocorrelation function of $(\hat{z}_t)_{t=1,\dots,T}$ is shown for each month in Figure 6.1. We observe dependence among

[1] The data set used here is part of a larger data set collected in Beijing during a 5-year time period, from January 1st, 2010 to December 31st, 2014, for a total of 43,824 observations. The data is available at https://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data

succeeding observations and therefore the independence assumption for the errors ε_t of the standard GAM model in (6.5) does not seem to be appropriate. We employ a state space model, as specified in (6.1) and (6.2), to allow for time effects in the errors of the GAM. Here $\rho_{obs,t}$ and $\rho_{lat,t}$ will be estimated from the data. Further, we assume that they do not depend on time, i.e. we set $\rho_{obs,t} = \rho_{obs}$ and $\rho_{lat,t} = \rho_{lat}$. In our data application we split the data into monthly periods to make this assumption more plausible.

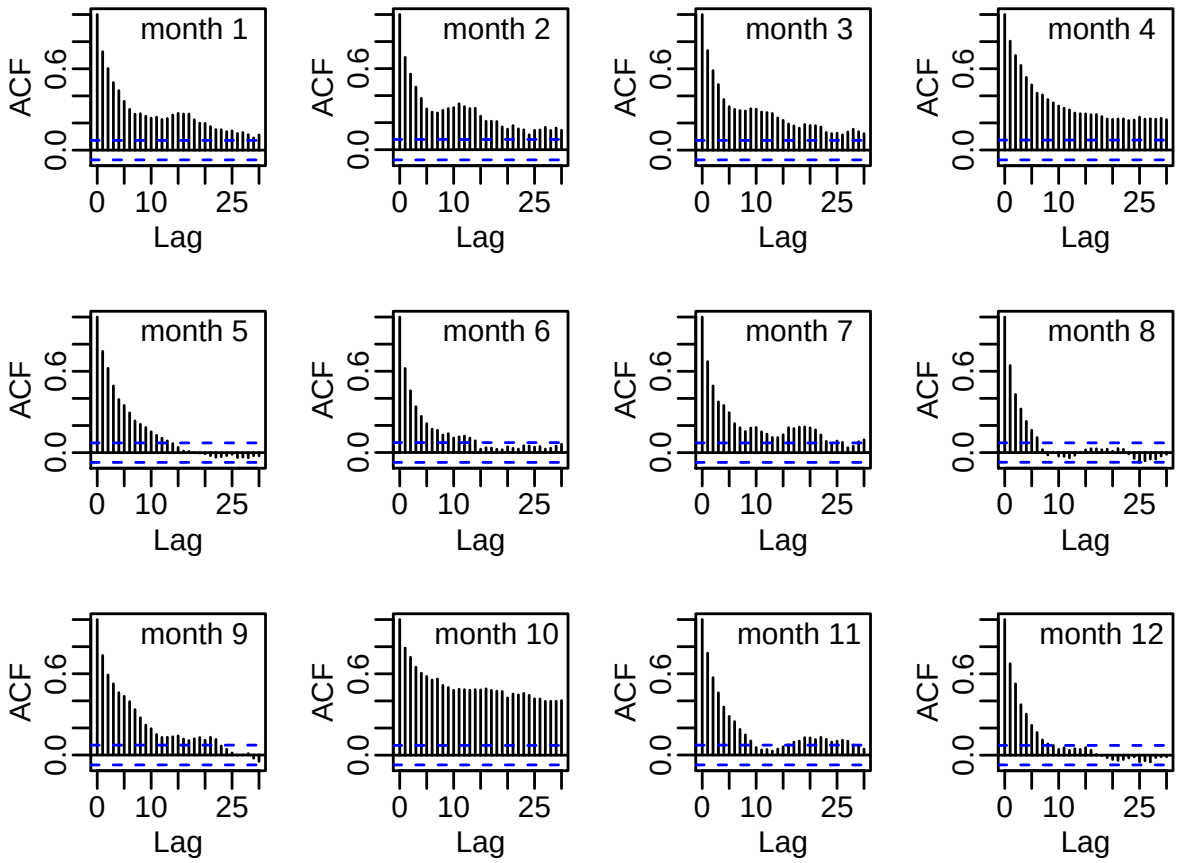

Figure 6.1: Autocorrelation functions (acf) of $(\hat{z}_t)_{t=1,\dots,T}$ for all 12 monthly data sets.

We now consider a state space model for Z_t, which is standardized by a GAM. We set $\sigma_{obs,t} = \sqrt{1-\rho_{obs}^2}$ for $\rho_{obs} \in (-1,1)$ and $\sigma_{lat,t} = \sqrt{1-\rho_{lat}^2}$ for $\rho_{lat} \in (-1,1)$. For the initial conditions we assume $\mu_{lat,0} = 0$ and $\sigma_{lat,0} = 1$. This ensures that Z_t has unit marginal variance. With these assumptions, the state space model in (6.1) and (6.2) becomes

$$\begin{aligned} Z_t &= \rho_{obs} w_t + \sqrt{1-\rho_{obs}^2}\eta_{obs,t} \\ w_t &= \rho_{lat} w_{t-1} + \sqrt{1-\rho_{lat}^2}\eta_{lat,t} \end{aligned} \tag{6.8}$$

with $\eta_{obs,t}, \eta_{lat,t}, w_0 \sim N(0,1)$ iid. Note that representation (6.8) induces the following

bivariate normal distributions

$$\begin{pmatrix} Z_t \\ w_t \end{pmatrix} | \rho_{obs} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{obs} \\ \rho_{obs} & 1 \end{pmatrix} \right)$$
$$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} | \rho_{lat} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{lat} \\ \rho_{lat} & 1 \end{pmatrix} \right).$$

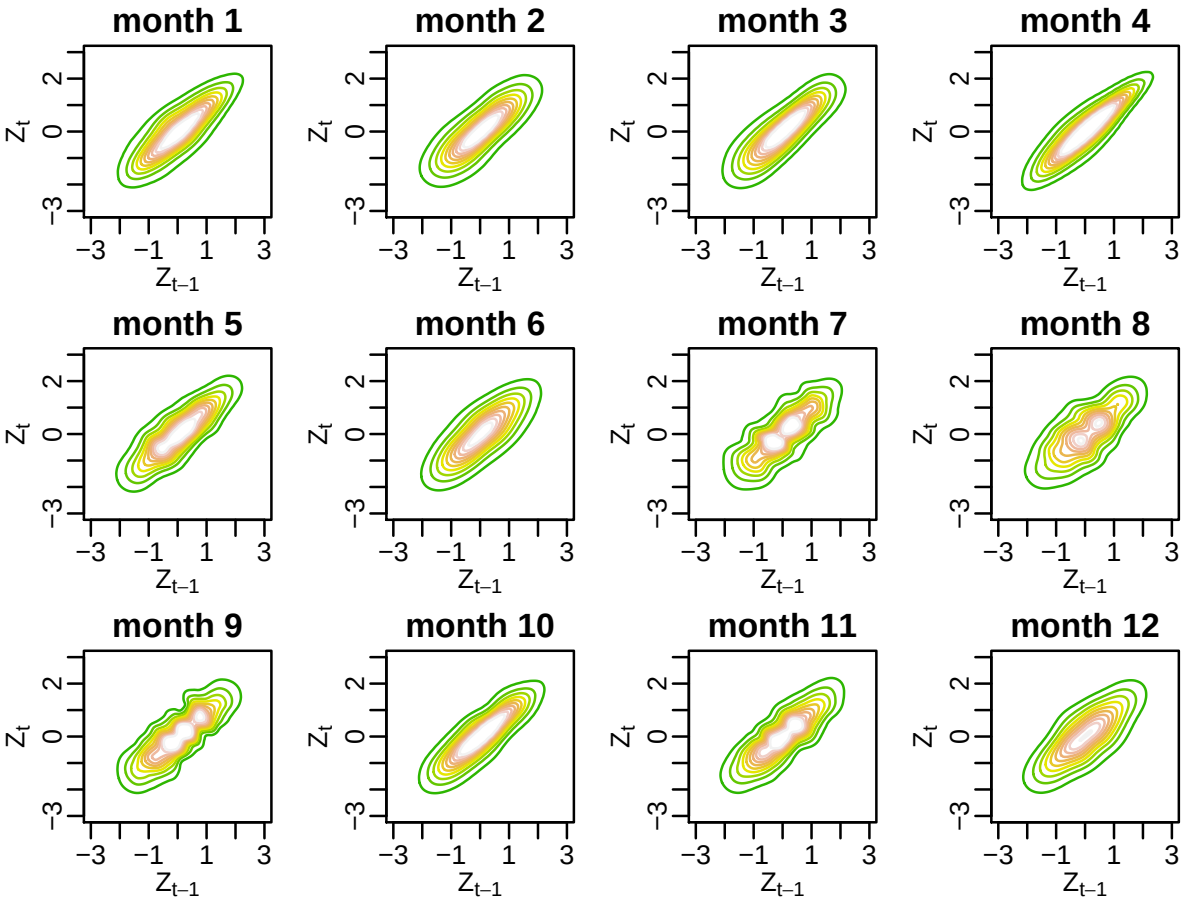

Figure 6.2: Contour plots of bivariate kernel density estimates based on pairs $(\hat{z}_t, \hat{z}_{t-1})_{t=2,\ldots,T}$, ignoring serial dependence, for each of the 12 Beijing air pollution monthly data sets.

In order to assess the suitability of the linear Gaussian state space model to the Beijing air pollution data, we display in Figure 6.2 contour plots of bivariate kernel density estimates based on pairs $(\hat{z}_t, \hat{z}_{t-1})_{t=2,\ldots,T}$ for each month. This visualizes the dependence structure between two successive time points in the series. More details about such contour plots are given in Czado (2019), Chapter 3. Using (6.8), we see that Z_t can be written as a linear function of Z_{t-1} and independent normally distributed disturbances. Since Z_1 is normally distributed, it follows that (Z_t, Z_{t-1}) are jointly normal. In particular, we have

$$Z_t \sim N(0,1) \qquad \text{and} \qquad \text{cov}(Z_t, Z_{t-1}) = \rho_{obs}^2 \rho_{lat} \qquad \text{for all } t \geq 1.$$

However, Figure 6.2 reveals that the contour plots of the Beijing monthly data deviate from the elliptical shape of a Gaussian dependence structure (which, to aid comparisons, is depicted in the top left panel of Figure D.1 in Appendix D). For example, the contour plots for January and October (months 1 and 10) show tail dependence and/or asymmetry in the tails, which cannot be modeled with a Gaussian distribution. This suggests that the linear Gaussian state space model is too restrictive for the Beijing air pollution data and a more flexible approach needs to be adopted.

Our proposal

Extensions of the linear Gaussian state space model, relaxing the assumptions of linearity and normality, have been studied. Chen et al. (2012) propose a state space approach to predict measles infections, where they relax the linearity assumption but still rely on normal errors in the state and in the observation equation. Johns and Shumway (2005) apply a nonlinear non-Gaussian state space model to censored air pollution measurements. Although the errors are assumed to be non-Gaussian, Johns and Shumway (2005) still rely on conditional normality.

We propose a very flexible Bayesian nonlinear and non-Gaussian state space model, where both the observation and the state equations are described by copulas. First, we find an equivalent formulation of the Gaussian state space model in (6.8) in terms of copulas. The representation is given by

$$\begin{aligned}(U_t, v_t)|\tau_{obs} &\sim \mathbb{C}^{Gauss}_{U,V}(\,\cdot\,,\,\cdot\,;\,\tau_{obs})\\ (v_t, v_{t-1})|\tau_{lat} &\sim \mathbb{C}^{Gauss}_{V_2,V_1}(\,\cdot\,,\,\cdot\,;\,\tau_{lat}),\end{aligned} \tag{6.9}$$

where

$$U_t = \Phi\left(Z_t\right), v_t = \Phi\left(w_t\right), \tag{6.10}$$

with Φ denoting the standard normal distribution function. The variables U_t and v_t are marginally uniformly distributed on $[0, 1]$ and Z_t and w_t are standard normal. Here the Gaussian copulas $\mathbb{C}^{Gauss}_{U,V}$ and $\mathbb{C}^{Gauss}_{V_2,V_1}$ are parametrized by Kendall's τ, obtained as $\tau_{obs} = \frac{2}{\pi}\arcsin(\rho_{obs})$ and $\tau_{lat} = \frac{2}{\pi}\arcsin(\rho_{lat})$. Corresponding approximately uniform pseudo copula data, that can be used for estimating the model in (6.9), is obtained as

$$\hat{u}_t = \Phi\left(\hat{z}_t\right). \tag{6.11}$$

By reformulating the state space representation in (6.8) in terms of copulas in (6.9), it is straightforward to see how we can generalize the Gaussian linear state space model by replacing the Gaussian copulas in (6.9) with arbitrary bivariate copulas. Typical restrictions of the Gaussian copula, such as symmetric tails, can be circumvented. For example, a Gumbel copula would allow for asymmetric tails. The proposed Bayesian copula based state space model allows us to specify various types of dependence structures to model the relationships between the observations and the underlying states, and to describe the states evolution over time. We will show that our methodology is able to accurately model and predict the levels of PM2.5 in Beijing.

The remainder of the paper is organized as follows. Section 6.2 introduces a copula based state space model, Section 6.3 illustrates the Bayesian inference for the proposed approach and Section 6.4 is devoted to the application of the copula state space model

to the Beijing pollution data. It also includes some simulations to study the PM2.5 predictions under different climate scenarios. Concluding remarks are given in Section 6.5.

6.2 The copula state space model

The copula state space model extends the linear Gaussian state space approach, allowing for copula specifications in place of normal distributions in the observation equation (6.3) as well as in the state equation (6.4). In particular, we assume that the errors $Z_t = \Phi^{-1}(U_t)$ of the GAM model introduced in (6.5), with $Z_t \sim N(0,1)$ and $U_t \sim \mathrm{U}(0,1)$ defined as in (6.10), depend on the latent state $w_t = \Phi^{-1}(v_t)$, with $w_t \sim N(0,1)$ and $v_t \sim \mathrm{U}(0,1)$, according to a bivariate copula given in the observation equation. The evolution of the latent variable w_t over time is also described by a bivariate copula, which defines the state equation of the model. The copulas defining the observation and state equations of the proposed state space approach do not necessarily belong to the same family, allowing for maximum flexibility in the specification of the model. However, we restrict our model to bivariate copula families with a single parameter. This gives still a flexible class of copula families, including e.g. Gaussian, Gumbel, Clayton or Frank copulas. The Student t copula can also be included if we fix the degrees of freedom parameter. Further, for the considered copula families, we are able to express the copula parameters in the observation and state equations in terms of Kendall's τ. More formally, for $t = 1, \ldots, T$, we assume the following joint distributions for the uniformly transformed variables U_t and v_t

$$\begin{aligned} (U_t, v_t)|\tau_{obs} &\sim \mathbb{C}^{obs}_{U,V}(\,\cdot\,,\,\cdot;\,\tau_{obs}) \\ (v_t, v_{t-1})|\tau_{lat} &\sim \mathbb{C}^{lat}_{V_2,V_1}(\,\cdot\,,\,\cdot;\,\tau_{lat}), \end{aligned}$$

where the bivariate copulas $\mathbb{C}^{obs}_{U,V}$ and $\mathbb{C}^{lat}_{V_2,V_1}$ are parametrized in terms of Kendall's τ (see Section 2.2.3). The *copula state space model* is defined on the uniform scale as follows

$$U_t|v_t, \tau_{obs} \sim \mathbb{C}^{obs}_{U|V}(\,\cdot\,|\,v_t;\,\tau_{obs}) \tag{6.12}$$

$$v_t|v_{t-1}, \tau_{lat} \sim \mathbb{C}^{lat}_{V_2|V_1}(\,\cdot\,|\,v_{t-1};\,\tau_{lat}) \tag{6.13}$$

for $t = 1, \ldots, T$ with initial distribution $v_0 \sim \mathrm{U}(0,1)$, where (6.12) is the observation equation and (6.13) is the state equation. The copula state space model introduced in (6.12) and (6.13) can be visualized as in Figure 6.3.

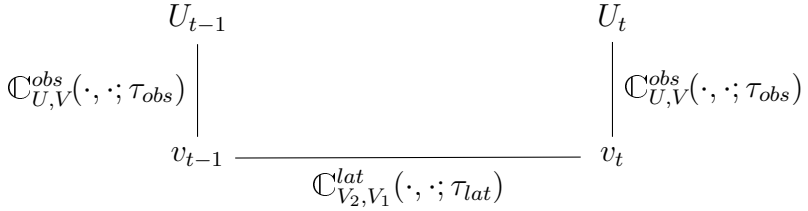

Figure 6.3: Graphical visualization of the copula state space model.

We now derive the joint distributions for the normalized variables Z_t and w_t

$$\begin{aligned} (Z_t, w_t) &\sim F_{Z_t,w_t} && (6.14)\\ (w_t, w_{t-1}) &\sim F_{w_t,w_{t-1}}. && (6.15) \end{aligned}$$

By Sklar's theorem (see Section 2.2.1), the distribution in (6.14) can be expressed as

$$\begin{aligned} F_{Z_t,w_t}(z_t, w_t) &= \mathbb{C}^{obs}_{U,V}(\,\Phi(z_t)\,,\,\Phi(w_t);\,\tau_{obs})\\ &= \mathbb{C}^{obs}_{U,V}(\,u_t\,,\,v_t;\,\tau_{obs}). \end{aligned}$$

Hence,

$$\begin{aligned} F_{Z_t|w_t}(z_t|w_t) &= \frac{d}{dv_t}\mathbb{C}^{obs}_{U,V}(\,\Phi(z_t)\,,\,v_t;\,\tau_{obs})\Big|_{v_t=\Phi(w_t)}\\ &= \mathbb{C}^{obs}_{U|V}(\,u_t\,|\,v_t;\,\tau_{obs})\Big|_{u_t=\Phi(z_t),v_t=\Phi(w_t)}\\ &= \mathbb{C}^{obs}_{U|V}(\,\Phi(z_t)\,|\,\Phi(w_t);\,\tau_{obs}) \end{aligned}$$

(see Czado (2019), Chapter 1, Lemma 1.15 for the first equality). Similarly, we obtain

$$F_{w_t|w_{t-1}}(w_t|w_{t-1}) = \mathbb{C}^{lat}_{V_2|V_1}(\,\Phi(w_t)\,|\,\Phi(w_{t-1});\,\tau_{lat}).$$

Therefore, the model can also be expressed on the normalized scale as follows

$$\begin{aligned} Z_t|w_t, \tau_{obs} &\sim \mathbb{C}^{obs}_{U|V}(\,\Phi(\cdot)\,|\,\Phi(w_t);\,\tau_{obs}) && (6.16)\\ w_t|w_{t-1}, \tau_{lat} &\sim \mathbb{C}^{lat}_{V_2|V_1}(\,\Phi(\cdot)\,|\,\Phi(w_{t-1});\,\tau_{lat}), && (6.17) \end{aligned}$$

where (6.16) is the observation equation and (6.17) is the state equation. Contour plots of the bivariate density of (Z_t, Z_{t-1}) of this model, for different choices of bivariate copulas, are shown in Figure 6.4, illustrating different shapes that the model can deal with. The copula state space model has the advantage of allowing flexibility in the specification of the observation and state equations, and thus is able to accommodate a wide variety of dependence structures in the air pollution data dynamics.

In the standard GAM, the errors are assumed to be independent. Our methodology allows us to account for autoregressive effects in the error through the underlying latent variable $\sigma \cdot w_t$, as defined on the original scale of the GAM residuals, or via the proxy v_t, on the uniform scale. These latent variables can be interpreted as non-measured autoregressive effects. As we will see in Section 6.4.3, the flexibility of our model allows us to detect extreme air pollution levels, which cannot be explained by the covariates. Capturing unusual air pollution levels is very important, since exposure to pollution spikes has a substantial impact on general health, such as causing severe cardiovascular and respiratory illnesses.

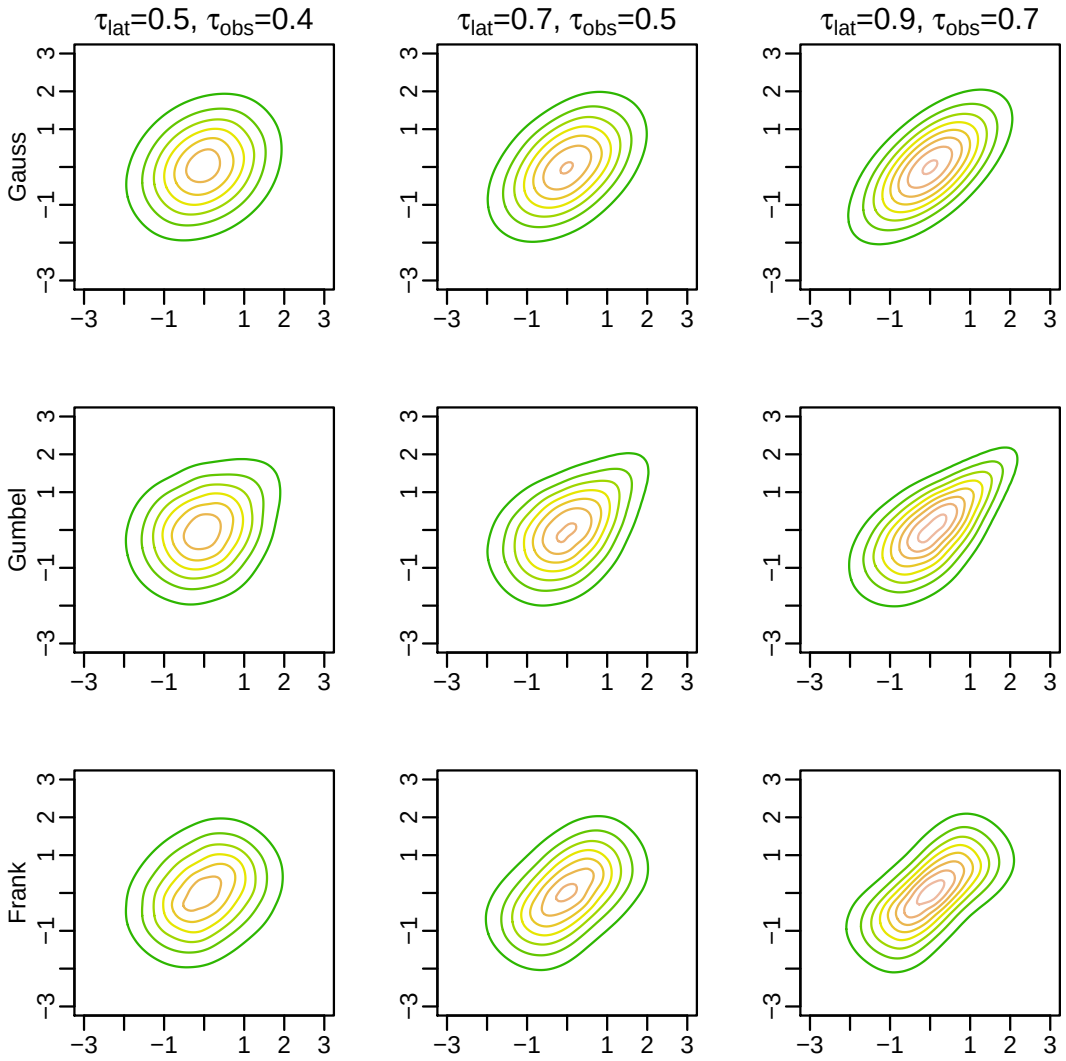

Figure 6.4: Contour plots of the density of (Z_t, Z_{t-1}) for the copula state space model for different choices of bivariate copula families. In the state and observation equation we choose the same copula family. The contour plots are constructed from bivariate kernel density estimates of simulated data.

Identifiability constraints

We notice some identifiability issues related to the model. For example, if we set τ_{obs} close to 1 and $\tau_{lat} = 0$, the observed and latent variables become equivalent and the latent variables $(v_t)_{t=1,\dots,T}$ at different time points become (nearly) independent. The high value of τ_{obs} will typically result in high values of the likelihood. This can make it difficult to recover the true values of τ_{obs} and τ_{lat}. Therefore, we need to set identifiability constraints for the copula state space model by establishing a relationship between τ_{obs} and τ_{lat}. In order to do that, we notice that the dependence between two successive time points U_{t-1} and U_t is determined by both τ_{lat} and τ_{obs}. The form of the correlation between $Z_{t-1} = \Phi^{-1}(U_{t-1})$ and $Z_t = \Phi^{-1}(U_t)$ can be derived exactly when $\mathbb{C}^{obs}_{U,V}$ and $\mathbb{C}^{lat}_{V_2,V_1}$ are both Gaussian copulas. Since in the Gaussian case, the parameter of the observation

equation copula is the correlation coefficient ρ_{obs} ($\tau_{obs} = \frac{2}{\pi}\arcsin(\rho_{obs})$) and the parameter of the state equation copula is the correlation coefficient ρ_{lat} ($\tau_{lat} = \frac{2}{\pi}\arcsin(\rho_{lat})$), the correlation between Z_{t-1} and Z_t is $\text{cor}(Z_{t-1}, Z_t) = \rho_{obs}^2\rho_{lat}$. We now derive the identifiability constraint for $\rho_{obs}, \rho_{lat} \geq 0$ (We do not expect negative dependence in our application). The higher the value of ρ_{lat}, the smoother the latent states are. Higher smoothness of the latent states induces a lower prediction uncertainty for the latent states. To guarantee a certain degree of smoothness, we need to set ρ_{lat} greater than some specific value and therefore impose $\rho_{obs} \leq \rho_{lat}$ in our approach. In particular, we assume the identifiability constraint in the Gaussian case

$$\rho_{obs} = \rho_{lat}^c \quad \text{for some suitable value} \quad c \geq 1.$$

In this case, the correlation between Z_{t-1} and Z_t becomes $\text{cor}(Z_{t-1}, Z_t) = \rho_{lat}^{2c+1}$. Transforming the correlation coefficients into Kendall's τ, in the Gaussian case, we obtain the following relationships

$$\tau_{obs} = \frac{2}{\pi}\arcsin(\rho_{lat}^c) \qquad \text{and} \qquad \tau_{lat} = \frac{2}{\pi}\arcsin(\rho_{lat}),$$

and therefore, τ_{obs} can be considered as a function of τ_{lat} and c. Figure 6.5 visualizes the relationship between the parameters τ_{obs} and τ_{lat} in the Gaussian case for different values of $c = 1, 3, 6, 10$. Considering that the strength of dependence between U_{t-1} and U_t is increasing in τ_{lat} and in τ_{obs}, Figure 6.5 shows that the higher the value of c the higher τ_{lat} needs to be to achieve a fixed strength of dependence between U_{t-1} and U_t. Therefore, for higher values of c, we expect to obtain a smoother behavior of the latent states $(v_t)_{t=1,\dots,T}$.

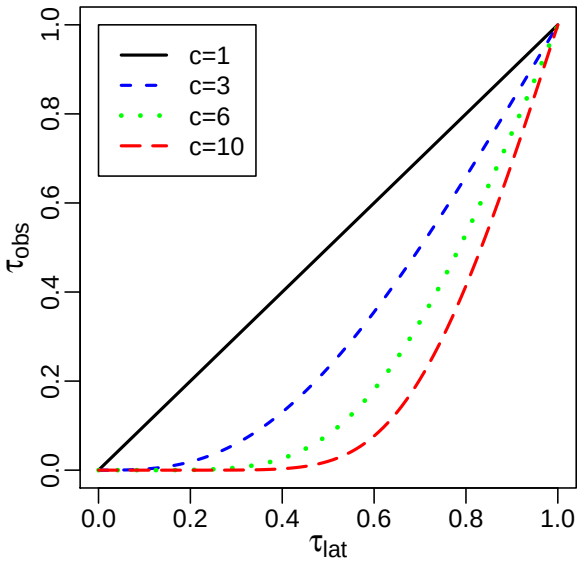

Figure 6.5: Graphical representation of the relationship between the parameter τ_{obs} and τ_{lat}. Here, τ_{obs} is plotted against τ_{lat} in the Gaussian case for different values of $c = 1, 3, 6, 10$.

We propose to use a similar relationship between τ_{lat}, τ_{obs} and c, not only in the Gaussian case, but also for arbitrary bivariate copula families. Therefore, in general,

we impose the following identifiability constraint on the Kendall's τ parameter for all bivariate copula families

$$\sin\left(\frac{\pi}{2}\tau_{obs}\right) = \left(\sin\left(\frac{\pi}{2}\tau_{lat}\right)\right)^c \qquad \text{for some suitable value } c \geq 1. \tag{6.18}$$

6.3 Bayesian analysis of the copula state space model

The copula state space model is a nonlinear and non-Gaussian model, which provides great flexibility by allowing for different bivariate copulas. The downside of this flexibility is that inference for this model is not straightforward, e.g. it is not possible to rely on the Kalman filter for linear Gaussian state space models (Durbin and Koopman (2012), Chapter 4) or to implement a Gibbs sampler, where we can directly sample from the corresponding full conditionals. For inference, we rely on the No-U-Turn sampler of Hoffman and Gelman (2014) implemented within the STAN framework (see Section 2.3.1).

6.3.1 Posterior inference

As prior distribution for τ_{lat} we use a uniform prior on (0,1), which is a vague prior restricted to positive dependence, since we do not expect negative dependence in our application. With this prior choice, the posterior density, for a specified value of c, is obtained as

$$\pi(\tau_{lat}, v_1, \ldots, v_T | \hat{u}_1, \ldots, \hat{u}_T) = \prod_{t=1}^{T} c_{U,V}(\hat{u}_t, v_t; \tau_{obs}) \prod_{t=1}^{T} c_{V_2,V_1}(v_t, v_{t-1}; \tau_{lat}),$$

where τ_{obs} is a function of τ_{lat} as given in (6.18). Note that the pseudo copula data $\hat{u}_1, \ldots, \hat{u}_T$ (see (6.11)) is here treated as the observed data. We run the No-U-Turn sampler to sample from this posterior density. We obtain a posterior sample for τ_{lat}

$$\tau_{lat}^r(c), \quad r = 1, \ldots, R$$

for a chosen c, and similarly for τ_{obs}, using the relationship in (6.18),

$$\tau_{obs}^r(c), \quad r = 1, \ldots, R,$$

where R is the number of MCMC iterations. Additionally, posterior samples for the latent states v_t, for $t = 1, \ldots, T$, are denoted by

$$v_t^r(c), \quad t = 1, \ldots, T, \; r = 1, \ldots, R.$$

6.3.2 Predictive simulation

An advantage of the Bayesian approach is that our model already specifies the predictive distribution, which is the distribution of the response for new data points conditional on observed data points. From this distribution, uncertainty is easy to be quantified through credible intervals. We consider a posterior sample of the model parameters given by the set $\{\tau_{lat}^r(c), \; v_t^r(c), \; r = 1, \ldots, R, \; t = 1, \ldots, T\}$. Simulations for a new value at time $t \in \{1, \ldots, T\}$ on the copula scale can be obtained by

- simulate $u_t^r(c)$ from $\mathbb{C}_{U|V}^{obs}\left(\cdot|v_t^r(c);\tau_{obs}^r(c)\right), r = 1,\ldots,R.$

We refer to the corresponding distribution as the in-sample predictive distribution on the copula scale. The out-of-sample predictive distribution refers to new values at time $t > T$. Simulated values from the one-day ahead predictive distribution of U_{T+1} given $\hat{u}_1,\ldots,\hat{u}_T$ can be obtained as follows

- simulate $v_{T+1}^r(c)$ from $\mathbb{C}_{V_2|V_1}^{lat}\left(\cdot|v_T^r(c);\tau_{lat}^r(c)\right)$,
- simulate $u_{T+1}^r(c)$ from $\mathbb{C}_{U|V}^{obs}\left(\cdot|v_{T+1}^r(c);\tau_{obs}^r(c)\right)$

for $r = 1,\ldots,R$. In general, simulations from the i-days ahead out-of-sample predictive distribution on the copula scale can be obtained recursively through:

- simulate $v_{T+i}^r(c)$ from $\mathbb{C}_{V_2|V_1}^{lat}\left(\cdot|v_{T+i-1}^r(c);\tau_{lat}^r(c)\right)$,
- simulate $u_{T+i}^r(c)$ from $\mathbb{C}_{U|V}^{obs}\left(\cdot|v_{T+i}^r(c);\tau_{obs}^r(c)\right)$

for $r = 1,\ldots,R$. Based on a simulation of the (in-sample or out-of-sample) predictive distribution on the copula scale, $u_t^r(c)$, we further define

$$\varepsilon_t^r(c) = \Phi^{-1}\left(u_t^r(c)\right)$$

as a sample of the predictive distribution of the error of the GAM model specified in (6.5). In particular, we estimate $E(Y_t)$ by $\hat{f}(\boldsymbol{x}_t)$ with estimated error variance $\hat{\sigma}^2$. So

$$y_t^r(c) = \hat{f}(\boldsymbol{x}_t) + \hat{\sigma}\varepsilon_t^r(c)$$

gives a sample of the predictive distribution of the response. Note that, to obtain this predictive sample, we ignore the uncertainty in the estimation of the marginal distribution.

6.4 Data analysis

Recall the hourly data set discussed in Section 6.1 divided into 12 sub data sets, one data set for each month.

6.4.1 Marginal models

For each of the 12 monthly data sets, we fit a GAM using the R package `mgcv` of Wood and Wood (2015), where the response is the logarithm of PM2.5 and the covariates are DEWP, TEMP, PRES, IWS, PREC and CBWD, as described in Section 6.1. We define an additional covariate PREC_ind, which indicates if there is precipitation, i.e. PREC_ind $= \mathbb{1}_{\text{PREC}>0}$. We also use the hour, denoted by H, and the weekday, denoted by D, as covariates. Liang et al. (2015) showed that the wind direction not only has influence on the response itself, but might also influence the relationship between the other covariates

DEWP, TEMP, PRES, IWS, PREC and the response. Therefore we allow for different smooth terms corresponding to different wind directions. More precisely, we create four indicator variables corresponding to the four wind directions $\mathbb{1}_{\text{CBWD=CV}}$, $\mathbb{1}_{\text{CBWD=NE}}$, $\mathbb{1}_{\text{CBWD=NW}}$ and $\mathbb{1}_{\text{CBWD=SE}}$. Then we replicate the part of the model matrix corresponding to a covariate x four times and multiply each of the four parts with one of the indicator variables $\mathbb{1}_{\text{CBWD=CV}}$, $\mathbb{1}_{\text{CBWD=NE}}$, $\mathbb{1}_{\text{CBWD=NW}}$ and $\mathbb{1}_{\text{CBWD=SE}}$. So, we obtain four smooth terms for each of the covariates DEWP, TEMP, PRES and IWS. We do not allow for these interactions with the covariate PREC, since this variable has only few values not equal to zero. For variable selection the approach of Marra and Wood (2011) is used, which allows terms to be penalized to zero.

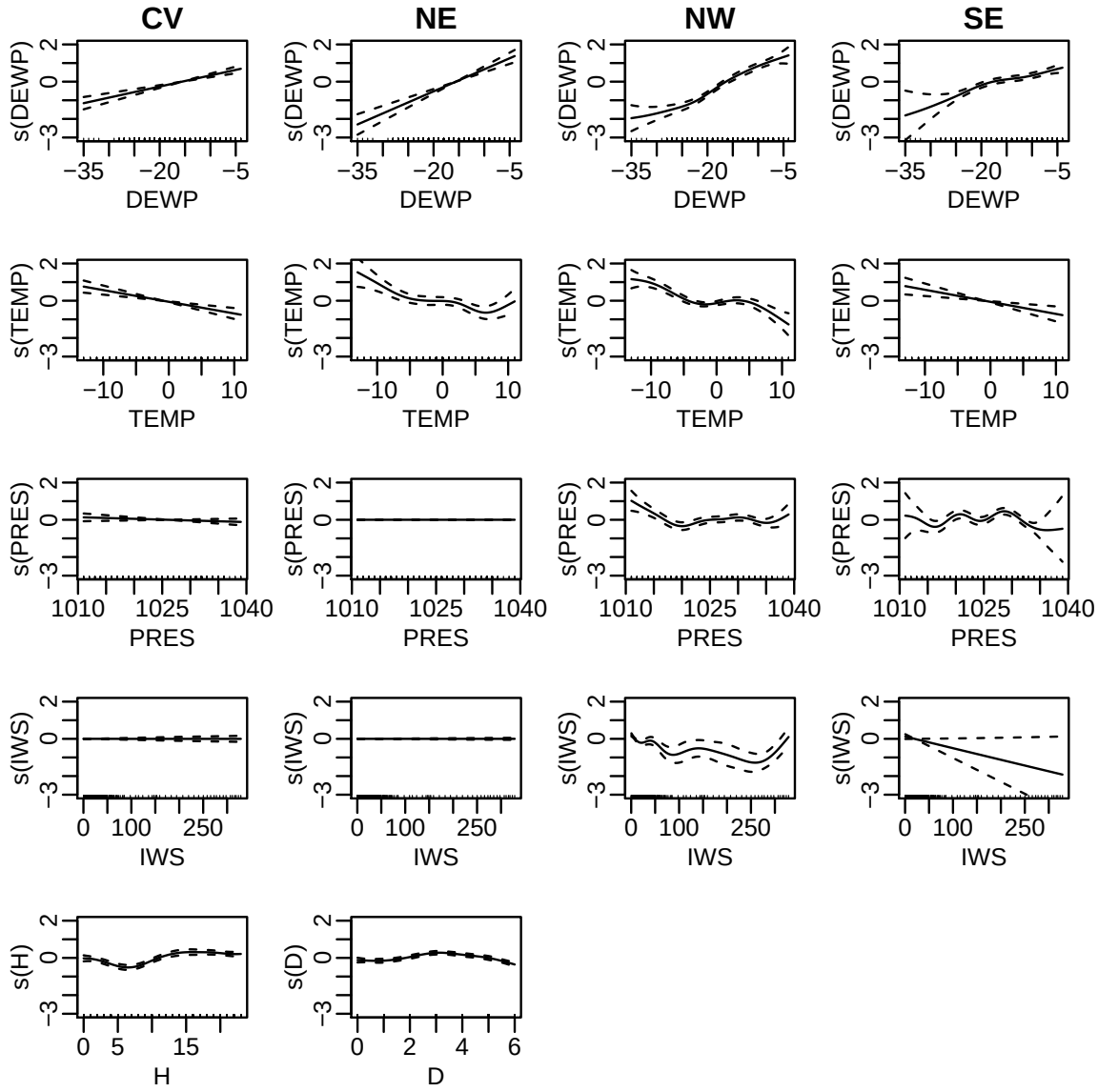

Figure 6.6: Estimated smooth components of the GAM for month 1 for the covariates DEWP, TEMP, PRES, IWS, H and D. For each of the covariates DEWP, TEMP, PRES and IWS we have four different smooth terms corresponding to the four wind directions: CV, NE, NW, and SE. The dashed lines represent a pointwise 95% confidence band.

Plots of the different estimated smooth components are shown in Figure 6.6 for the January (month 1) data set. Plots of the estimated smooth terms in Figure 6.6 indicate the covariate effects on PM2.5. For example, with northwestern winds (NW), PM2.5 is lower for higher temperatures (TEMP). Furthermore, we draw the same conclusion as Liang et al. (2015), that different smooth terms are necessary for different wind directions. For example, with northeastern winds (NE), we do not see any influence of the covariate PRES on PM2.5, whereas with northwestern winds (NW), we observe some nonlinear relationship between PRES and PM2.5.

6.4.2 Model selection of monthly copula family and value of c based on the widely applicable information criterion (in-sample)

We now consider model selection for the copula state space model. This includes the selection of the copula families and the selection of the value of c. We fit models with different copula families and different values of c and select the model which minimizes the widely applicable information criterion (WAIC) (Watanabe (2010), Gelman et al. (2014b)). For our model AIC and BIC would require to integrate out all the latent variables. Therefore we stick to the WAIC which is easy to evaluate for such Bayesian models with latent variables. We denote by $\ell_t^r = c(\hat{u}_t, v_t^r; \tau_{obs}^r(c))$ the likelihood contribution of iteration r at time t. Following Vehtari et al. (2017), the WAIC can then be estimated by

$$\widehat{\text{WAIC}} = -2 \sum_{t=1}^{T} \left[\ln \left(\hat{\text{E}}((\ell_t^r)_{r=1,\dots,R}) \right) - \widehat{\text{Var}} \left((\ln(\ell_t^r))_{r=1,\dots,R} \right) \right],$$

where $\hat{\text{E}}$ denotes the sample mean and $\widehat{\text{Var}}$ the sample variance.

We have one GAM specification for each month and obtain, for each month, approximately uniform pseudo copula data $\hat{u}_t$ by the probability integral transform $\hat{u}_t = \Phi\left(\frac{y_t - \hat{f}(\boldsymbol{x}_t)}{\hat{\sigma}}\right)$ for $t = 1, \dots, T$ as in (6.11). Here $\hat{f}$ and $\hat{\sigma}$ are the estimates of the GAM and T denotes the number of observations in the corresponding monthly data set. To simplify notation, we avoid indexing the models by month. In the following we study several models that can be divided into three model classes.

- **Gaussian state space model** $\mathscr{M}_{Gauss}$: $\mathbb{C}_{U,V}^{obs}$ and $\mathbb{C}_{V_2,V_1}^{lat}$ are both Gaussian copulas.
- **Copula based state space model** $\mathscr{M}_{Cop}$: $\mathbb{C}_{U,V}^{obs}$ and $\mathbb{C}_{V_2,V_1}^{lat}$ are from the same bivariate copula family.
- **GAM model with independent errors** $\mathscr{M}_{Ind}$: $\mathbb{C}_{U,V}^{obs}$ and $\mathbb{C}_{V_2,V_1}^{lat}$ are both independence copulas. This corresponds to a standard GAM model with independent errors.

For each of the 12 monthly data sets on the copula scale, the three model classes are fitted. To estimate model parameters, we run the No-U-Turn sampler with 2 chains, where each chain contains 2000 iterations. The first 500 iterations are discarded for burn-in. Preliminary analysis showed that this burn-in choice is sufficient. We fit the independence

model $\mathscr{M}_{Ind}$, the Gaussian model $\mathscr{M}_{Gauss}$ for every value of $c = 1, 3, 6, 10$ and several latent copula models for the class $\mathscr{M}_{Cop}$. The different copula state space models correspond to all combinations of the values of $c = 1, 3, 6, 10$ and the following bivariate parametric copula families: Student t (df=3), Student t (df=6), Gumbel, Clayton and Frank. This set includes copula families that are appropriate for the observed contour plots in Figure 6.2. So, for one specific monthly data set, a model is specified by the value of c and the copula family.

As an example, we have a closer look at the model for January with Student t copulas with 6 degrees of freedom and $c = 1$. Figure 6.7 shows the trace plots of the dependence parameter τ_{lat} and the latent state at time point 100 (v_{100}) for the first chain. The trace plots suggest that the chains have converged. The chain for τ_{lat} converges to values far away from zero, thus showing dependence. Figure 6.8 illustrates the effect of the different values of c on the posterior mode estimates of the latent states $\hat{v}_t$. As expected, we observe that the size of the oscillations decreases as the value of c increases.

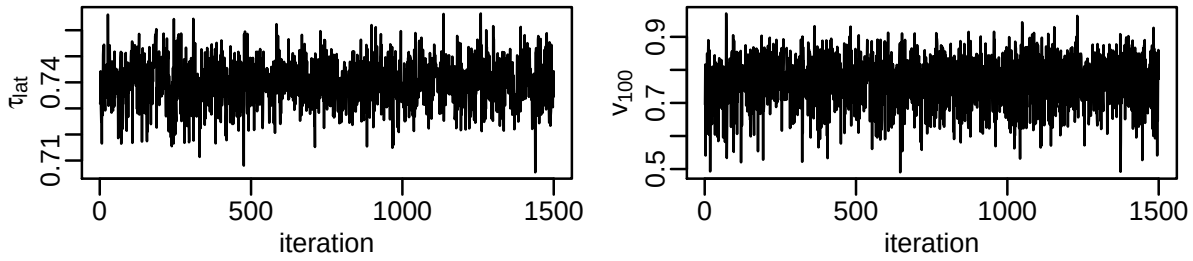

Figure 6.7: Trace plots of 1500 posterior draws after a burn-in of 500 iterations for τ_{lat} (left) and v_{100} (right) for the first chain of the No-U-Turn sampler for the model with Student t copulas with 6 degrees of freedom and $c = 1$ using the data set for January.

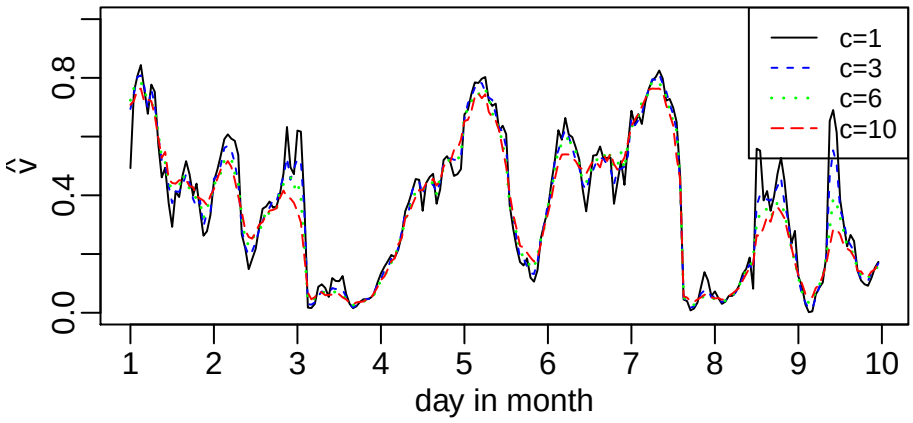

Figure 6.8: Estimated hourly posterior mode of the latent state $\hat{v}_t$ at time t plotted against t for the first 9 days of January for models with Student t copulas with 6 degrees of freedom and different values of c ($c = 1, 3, 6, 10$). The posterior mode estimates are obtained from univariate kernel density estimates and are based on 3000 iterations from two chains.

Table 6.1 shows the best model in $\mathscr{M}_{Cop}$, characterized by the value of c and the copula family, and the best model in $\mathscr{M}_{Gauss}$, characterized by the value of c. In addition, Table 6.1 shows the WAIC of the best model within the model classes $\mathscr{M}_{Cop}$, $\mathscr{M}_{Gauss}$ and $\mathscr{M}_{Ind}$. We see that for $\mathscr{M}_{Gauss}$ and $\mathscr{M}_{Cop}$ the value of c of the best model is always equal to 1, thus allowing for higher oscillations in the posterior of the latent states. The best model according to the WAIC is provided by the copula based model class $\mathscr{M}_{Cop}$ for every month, since this model is always associated to the smallest WAIC.

	family	c		$\widehat{\text{WAIC}}$		
month	$\mathscr{M}_{Cop}$	$\mathscr{M}_{Cop}$	$\mathscr{M}_{Gauss}$	$\mathscr{M}_{Cop}$	$\mathscr{M}_{Gauss}$	$\mathscr{M}_{Ind}$
1	t(6)	1	1	-926	-887	0
2	Frank	1	1	-755	-702	0
3	Frank	1	1	-1000	-898	0
4	t(3)	1	1	-1200	-1103	0
5	t(6)	1	1	-982	-945	0
6	t(3)	1	1	-672	-604	0
7	t(3)	1	1	-808	-722	0
8	t(3)	1	1	-680	-653	0
9	t(6)	1	1	-972	-873	0
10	Gumbel	1	1	-1130	-1102	0
11	t(6)	1	1	-910	-900	0
12	t(6)	1	1	-765	-758	0

Table 6.1: Family of the best model in $\mathscr{M}_{Cop}$, value of c of the best model in $\mathscr{M}_{Cop}$ and the best model in $\mathscr{M}_{Gauss}$ and the estimated WAIC of the best model within each class $\mathscr{M}_{Cop}$, $\mathscr{M}_{Gauss}$ and $\mathscr{M}_{Ind}$. The best model is selected with respect to the WAIC.

6.4.3 Analysis of fitted models

In the previous section we selected the best copula state space models according to the lowest WAIC. This gave the copula family choice and the value of c for $\mathscr{M}_{Cop}$ and the value of c for $\mathscr{M}_{Gauss}$. Figure 6.9 shows the estimated posterior densities for the dependence parameter τ_{lat} for these models. We observe that most of the mass of the posterior density concentrates between 0.6 and 0.8 for all monthly models. This range for τ_{lat} coincides with positive dependence between two succeeding time points. We also see that the Kendall's τ values of the $\mathscr{M}_{Cop}$ model class are slightly higher than those of the $\mathscr{M}_{Gauss}$ model class for all months.

The copula based state space model was fitted to data $\hat{u}_t = \Phi(\hat{z}_t)$, where $\hat{z}_t$ is a standardized residual of the GAM, as defined in (6.7). To further evaluate our model, we simulate from the (in-sample) predictive distribution of the error for each $t \in \{1, \dots, T\}$, as explained in Section 6.3.2, and compare these simulations to the standardized residuals of the corresponding GAM model. Figure 6.10 shows that the copula based state space model is able to recover the dynamics of the standardized residuals. If we ignored the latent effect (i.e. assumed independent errors), the errors ϵ_t would independently follow standard normal distributions. Simulating from the predictive distribution of the error can be considered as taking the latent effect into account. Therefore a concentration of

the predictive distribution that is far away from zero indicates time points where the latent variable has higher effects. These are time points where the level of the response is unusually high or low for the corresponding specification of the covariates. We see from Figure 6.10 that on January 17th/18th, the estimated mode of the predictive density of the error is high. During the corresponding week unusual high pollution was recorded in Beijing [2]. The copula based state space model with a Student t copula has a high peak and is able to capture unusual behavior. Since many studies have shown that high air pollution levels may have severe effects on health and can cause economic losses (Anderson et al. (2012), Kim et al. (2015)), capturing air pollution peaks is very important.

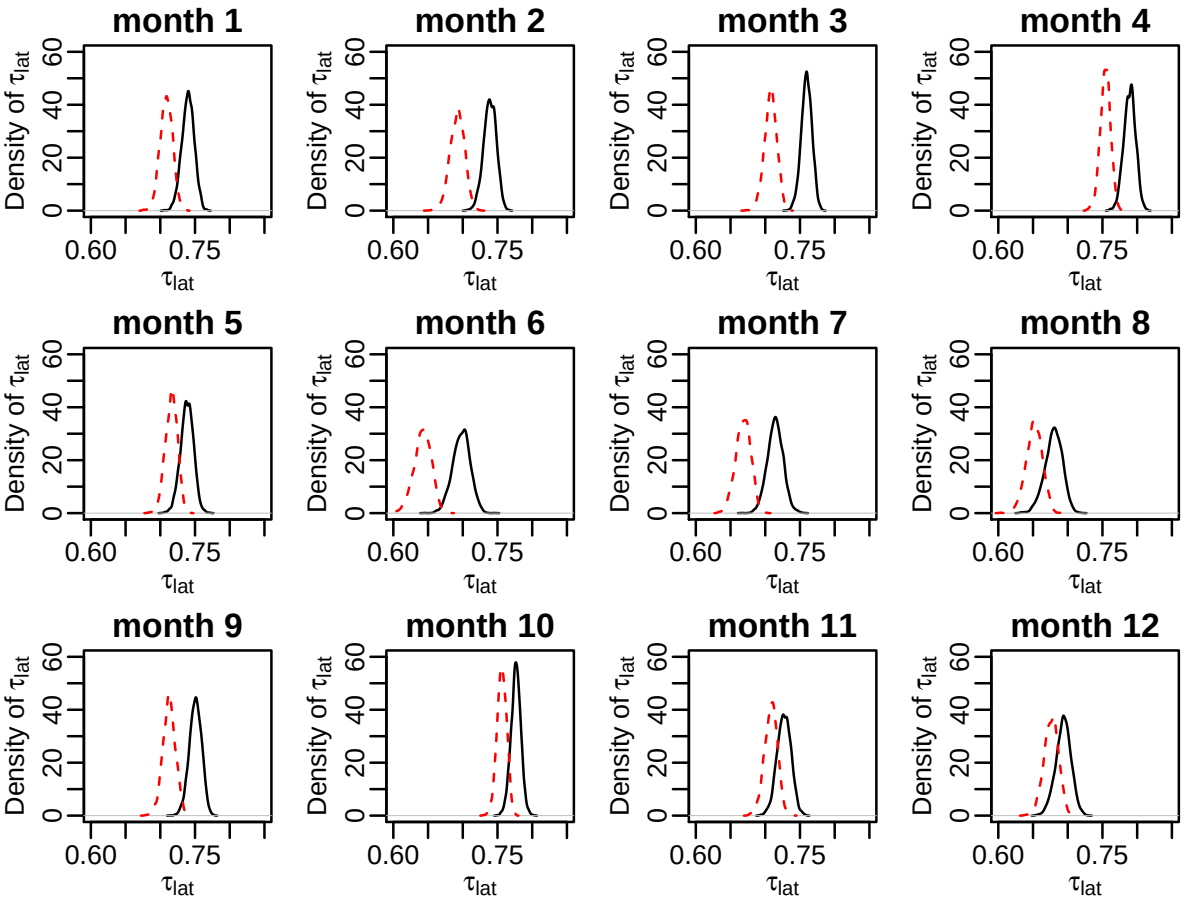

Figure 6.9: Kernel density estimate of the posterior density of the dependence parameter τ_{lat} for the best model in $\mathscr{M}_{Cop}$ (black) and $\mathscr{M}_{Gauss}$ (red, dashed) for all 12 monthly data sets.

[2] See http://www.takepart.com/article/2014/01/18/beijing-china-air-pollution-billboard

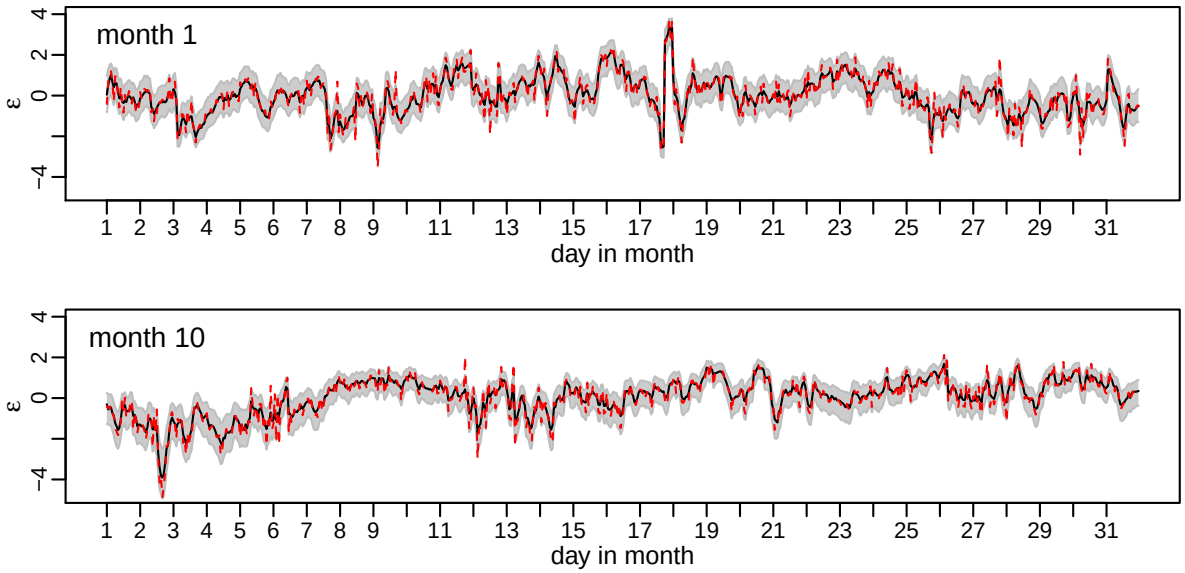

Figure 6.10: Estimated mode of the predictive density of the error ϵ_t plotted against t (black line) for every data point in January (top row) and October (bottom row) using the best models in $\mathscr{M}_{Cop}$ as selected by WAIC. A 90% credible region, constructed from the 5% and 95% empirical quantiles of simulations from the predictive distribution of the error, is added in grey. Further, the standardized residual of the GAM, $\hat{z}_t$, is added in red (dashed).

6.4.4 Out-of-sample predictions

Short term predictions of PM2.5 levels can be used to alert citizens of high pollution periods which are dangerous to health. In this section we construct predictions several hours up to two days ahead. More precisely, we consider the best copula state space model for March and use it to predict the first 48 hours of April. We choose March, since it is a month for which the non-elliptical Frank copula was selected.

We first simulate from the out-of-sample predictive distribution of the error as explained in Section 6.3.2. Figure 6.11 shows predictive densities for different time steps ahead for this model, more precisely the estimated forecast density of ϵ_{T+t} for $t = 1, 12, 24, 48$ hours based on 3000 MCMC iterations from two chains. As we see from Figure 6.11, we obtain non-Gaussian forecast densities. Further, the densities are more disperse for a longer time period ahead, reflecting the fact that uncertainty increases if we predict a longer time period ahead.

To obtain predictions for the PM2.5 levels, the simulations for the error need to be combined with the mean prediction of the GAM, according to our model

$$Y_t = f(\boldsymbol{x_t}) + \sigma\varepsilon_t.$$

To obtain the predicted mean of the GAM, the covariate values are required. Except for the weekday D and the hour H, future covariate levels are not known. As a proxy for an unknown covariate vector with hour H=h, we use the covariate specifications of the last

observed time point with the same hour H=h. We denote this covariate vector by $\boldsymbol{x}_t^l$ and obtain predictive simulations of the response (the logarithm of PM2.5) at time $t > T$ as follows

$$y_t^r = \hat{f}(\boldsymbol{x}_t^l) + \hat{\sigma}\varepsilon_t^r \tag{6.19}$$

for $r = 1, \ldots, R$. These predictive simulations are visualized in Figure 6.12. We see that the observed values are most of the time within the 90% credible interval.

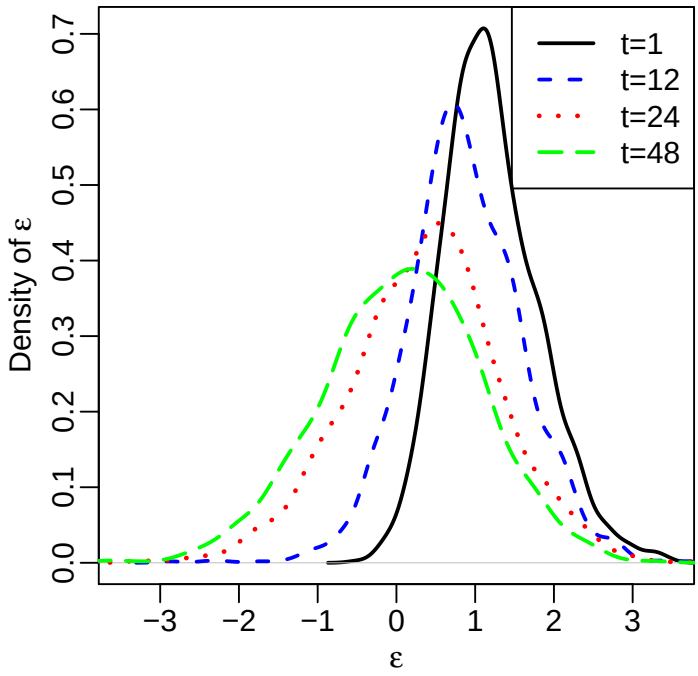

Figure 6.11: Estimated predictive density of ϵ_{T+t} using the best copula state space model for March for different time steps (hours) ahead ($t = 1, 12, 24, 48$). The estimated predictive density is the kernel density estimate of simulations from the corresponding predictive distribution.

In addition, the simulations for the error may be combined with mean predictions obtained from the GAM with different covariate specifications. Since the covariates several hours ahead are random, different scenarios as specified by different covariate levels are possible and should be taken into account. Here, we first consider two cases where the temperature at each time point in $\boldsymbol{x}_t^l$ is increased and decreased by 1 degree. Second, we also investigate more extreme scenarios for $\boldsymbol{x}_t^l$, where we decrease and increase the temperature at each time point by 4 degrees and in addition change the wind direction at each time point to the same value. The value for the wind direction CBWD is set to either CV or SE. This yields four different scenarios. The mode estimates of the resulting predictive densities are visualized in Figure 6.13. It is not surprising that the first case, where we only change the temperature by 1 degree, results in less changes in the mode estimates compared to the more extreme case. There are many more scenarios that can be analyzed in a similar fashion. In particular, relevant scenarios suggested by experts could be analyzed. A conservative warning system could alert citizens if at least one of the scenarios results in dangerous air pollution levels.

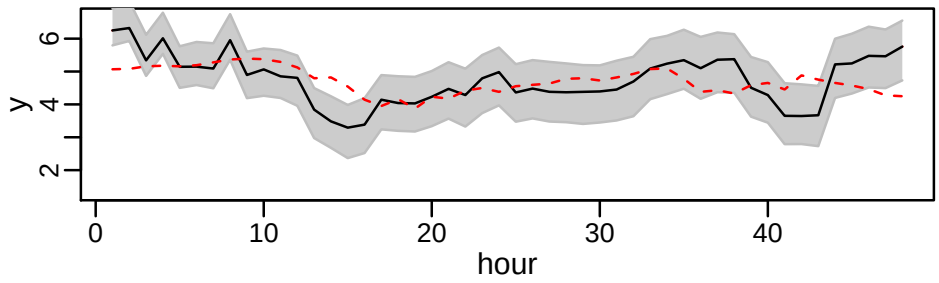

Figure 6.12: Estimated mode of the predictive density of the response t hours ahead plotted against t (black line). A 90% credible region, constructed from the 5% and 95% empirical quantiles of simulations from the predictive distribution of the response, is added in grey. Further, the observed response values are added in red (dashed). The simulations from the predictive distribution of the response 1 up to 48 hours ahead are obtained according to (6.19) based on the best copula state space model for March.

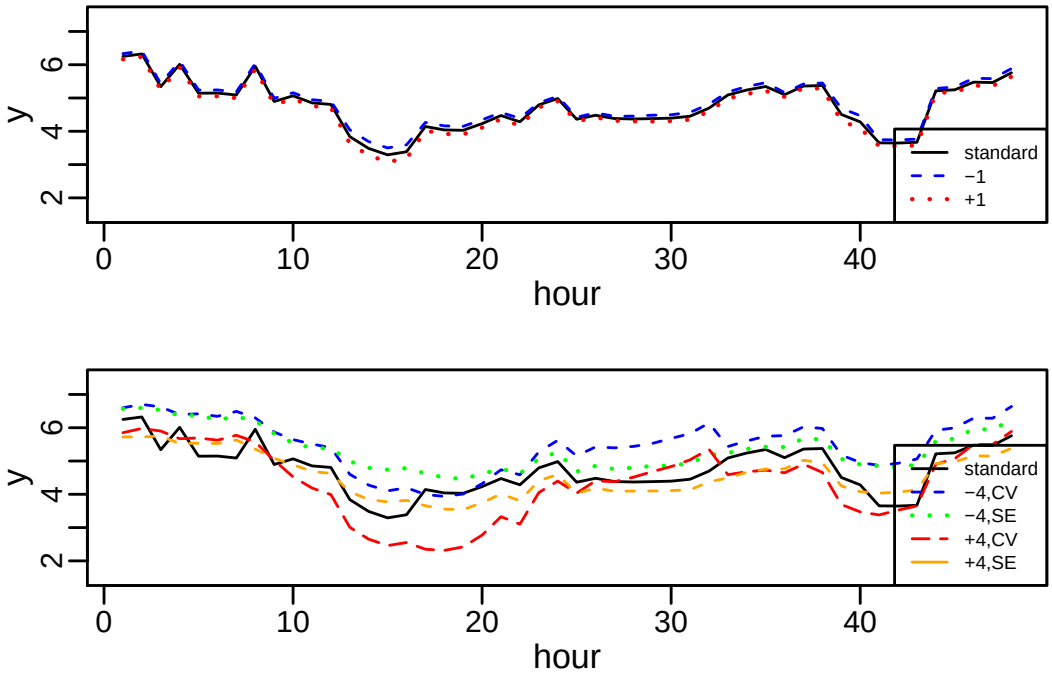

Figure 6.13: We show the estimated mode of the predictive density of the response t hours ahead plotted against t for different specifications of the covariates. For the estimation of the mode we use simulations from the corresponding predictive distribution of the response 1 up to 48 hours ahead, obtained according to (6.19) based on the best copula state space model for March. We consider the predictive distribution obtained from the unchanged covariate vector $\boldsymbol{x}_t^l$ (black line). In the top row, we consider additionally predictive distributions where the temperature of $\boldsymbol{x}_t^l$ is changed by ± 1 degree. In the bottom row, we consider additionally predictive distributions where the temperature of $\boldsymbol{x}_t^l$ is changed by ± 4 degree and the covariate CBWD is set equal to SE or CV.

6.4.5 Simulated scenarios

Instead of only considering predictions several hours or days ahead, our model allows us to simulate typical air pollution levels that might occur in the same month in another year with different covariate levels. We may consider a different covariate vector $\boldsymbol{x}_t^{new}$ and obtain

$$(PM_t^{new})^r = \exp\Big(\hat{f}(\boldsymbol{x}_t^{new}) + \hat{\sigma}\varepsilon_t^r\Big), \tag{6.20}$$

for $r = 1, \ldots, R, t = 1, \ldots, T$, where $\hat{f}$ and $\hat{\sigma}$ are estimates from the marginal GAM models and ε_t^r is a simulation from the in-sample predictive distribution of the error, based on the data for 2014. The values of $(PM_t^{new})^r$ give rise to typical air pollution levels that might occur in the same month in another year with covariate levels $\boldsymbol{x}_t^{new}$.

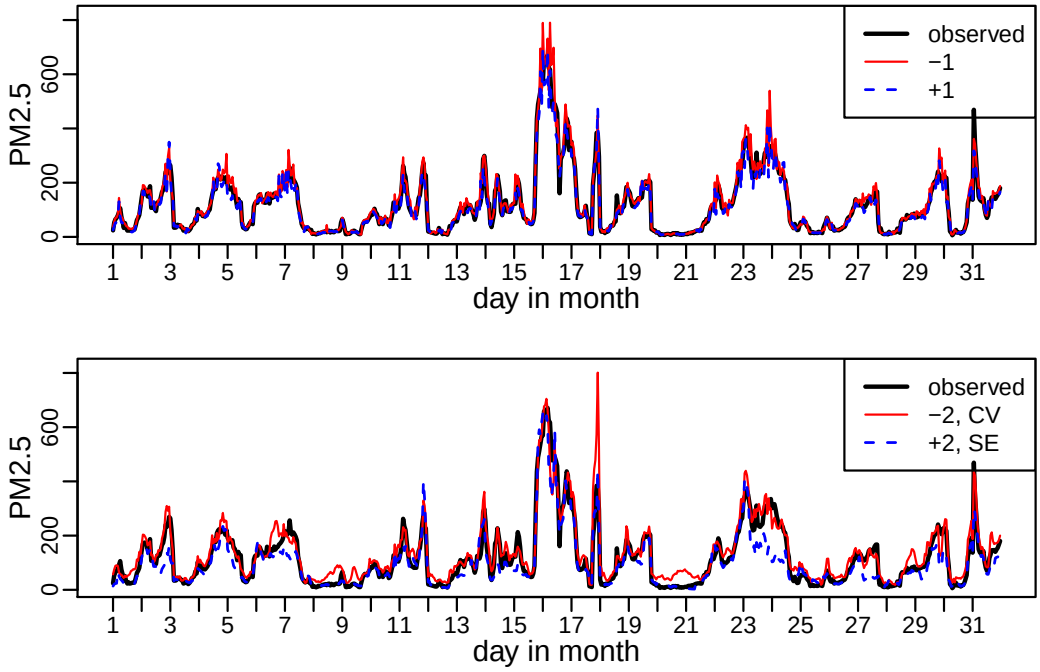

Figure 6.14: We show typical PM2.5 levels for January for different specifications of the covariates. The typical PM2.5 level is estimated as the mode of the kernel density estimate of simulations obtained as explained in (6.20). The top row shows typical air pollution levels where the temperature was changed by ±1 degree. In the bottom row we show one case where the temperature was decreased by 2 degree and the covariate CBWD was set equal to CV and another case where the temperature was increased by 2 degree and the covariate CBWD was set equal to SE. The other covariates are kept at the same levels as they were observed in 2014. The PM2.5 level observed in 2014 is added in black.

Here we analyze different scenarios for January. First, we consider scenarios where we only change the temperature, leaving all the other covariates as they were observed in January 2014. We consider one case where we increase the original temperature variable at each time point by 1 degree and one case where it is decreased by 1 degree. From the data set analyzed by Liang et al. (2015), of which our data set is a subset, we can

see that differences of about 1 degree in the monthly average temperature between two different years are common. Second, we investigate more extreme scenarios where we shift temperatures by ± 2 degree and also change the wind direction. The dominant wind direction in January 2014 was NW (northwestern). The wind direction CBWD at each time point is now changed to the same value. The value is set equal to CV (calm and variable), NE (northeastern) or SE (southeastern). Combining these three choices for the wind direction with two different choices for the temperature leads to 6 different scenarios.

In Figure 6.14 we compare the mode estimates of the density of PM_t^{new} to the observed PM2.5 values in 2014. We see that in January, a decrease in temperature by one degree leads to higher pollution levels. We obtain higher peaks and the average PM2.5 level of this month increases from 118 $\mu\text{g/m}^3$, as observed in January 2014, to 127 $\mu\text{g/m}^3$. Further, we show in Figure 6.14 the two out of the six more extreme cases that lead to the largest increase and decrease in the average PM2.5 level. Increasing the temperature by 2 degree and setting the wind direction equal to SE, leads to the largest decrease in the PM2.5 level. The average PM2.5 level decreases from 118 $\mu\text{g/m}^3$ to 95 $\mu\text{g/m}^3$. By decreasing the temperature by 2 degrees and setting the wind direction equal to CV (calm and variable), the average PM2.5 level increases from 118 $\mu\text{g/m}^3$ to 137 $\mu\text{g/m}^3$. Further, this scenario leads to higher peaks of the air pollution level. Our analysis shows that it is not unlikely to observe higher air pollution levels in future Januaries compared to those of January 2014.

6.5 Summary and outlook

The aim of this chapter was to model air pollution measurements in Beijing. In a first step meteorological and seasonal patterns were removed utilizing GAMs. However, there was still autocorrelation in the GAM residuals. We have seen that this time dependence among the residuals is non-Gaussian and therefore a linear Gaussian state space model is not appropriate. To overcome the limitations of the linear Gaussian state space model, we propose a generalization of this model. We obtain a nonlinear copula based state space model which allows for great flexibility by specifying both, the observation and the state equation with copulas. We showed that our model is able to capture peaks of air pollution, such as the one observed on January 17th/18th. Further, we demonstrated how the proposed model can be utilized to predict future PM2.5 levels under different climate conditions.

There are several aspects of the proposed model that might be extended. The copula parameters of the bivariate copulas could depend on covariates as in Vatter and Chavez-Demoulin (2015). Further, we could allow for time-variation in the copula parameter as in the bivariate dynamic copua model of Almeida and Czado (2012) . Another important future research direction is the extension of the proposed model to higher-dimensional observation and state equations. Here we could rely on vine copulas (Aas et al. (2009)) or on a factor copula structure (Krupskii and Joe (2013)).

7 A multivariate copula state space model

This chapter is a reproduction of Kreuzer et al. (2019b) with minor changes.

7.1 Introduction

Linear Gaussian state space models are widely used. For example, Ippoliti et al. (2012) use a linear Gaussian state space approach to model pollutant measurements in Italy and ozone levels in Mexico. Van den Brakel and Roels (2010) apply a linear Gaussian state space model to Dutch survey data. However, the strong assumptions of linear Gaussian state space models prevent their applicability to data showing departures from linearity and normality. In order to overcome these limitations, extensions have been studied. For example, Chen et al. (2012) propose a nonlinear state space model with Gaussian errors which they use to predict measles infections. Johns and Shumway (2005) develop a nonlinear non-Gaussian state space approach for censored air pollution data, which assumes conditional normality for the errors.

Copula based approaches have proven to be particularly suitable for modeling data showing departures from multivariate normality. They allow to model symmetric and asymmetric tail dependence. Further, copulas focus entirely on the dependence structure and can be combined with different marginal distributions.

Hafner and Manner (2012) and Almeida and Czado (2012) suggest a bivariate state space model, with a bivariate copula in the observation equation and a Gaussian autoregressive process of order one, which describes the time evolution of the copula parameter, in the state equation. In Chapter 6 a univariate nonlinear non-Gaussian state space model was proposed, where both the observation and the state equation are defined in terms of copula specifications. However, for the application it was assumed that the copulas describing the observation and the state equation belong to the same family.

In this chapter, we propose a multivariate nonlinear non-Gaussian state space model, which extends the approach introduced in Chapter 6 to multivariate observations, which we assume to be related to an underlying latent variable. This approach allows us to capture cross-sectional as well as temporal dependence in a very flexible way, since the copulas specifying the model can be different. For each time point, the proposed model can be described as a C-vine truncated after the first tree, with the latent state being the root node. The latent states are treated as parameters, with prior distribution given by a D-vine truncated after the first tree to capture temporal dependence. An advantage of

our approach is that missing values are handled in a natural way, since they are treated as latent variables. But for model estimation, we cannot rely on the standard Kalman filter approach developed for linear Gaussian state space models. Therefore, we suggest a Bayesian approach using Hamiltonian Monte Carlo (HMC) (Neal et al. (2011), Carpenter et al. (2017)), where we introduce an indicator variable for the copula families specifying the state space model equations.

We demonstrate the usefulness of our method in a data set containing different air pollutant measurements. Three different pollutants are considered, and for each pollutant, measurements from a high-cost and from a low-cost sensor are utilized. In addition, covariates such as the temperature are available. To model this data we follow a flexible two-step modeling approach, motivated by Sklar's Theorem (see Section 2.2.1). First we model the marginal distributions with generalized additive models (Hastie and Tibshirani (1986)) and in the second step we model dependencies with the novel copula state space model. We utilize our model to reconstruct high-cost measurements from low-cost measurements as in De Vito et al. (2008) and show that the copula based state space model, in combination with marginal generalized additive models, does a good job at predicting high-cost measurements. We show that it outperforms a Gaussian state space model and Bayesian additive regression trees with respect to the continuous ranked probability score (Gneiting and Raftery (2007)).

The rest of this chapter is organized as follows: Section 7.2 introduces the novel multivariate copula state space model, Section 7.3 discusses Bayesian inference for the novel approach, Section 7.4 is devoted to the air pollutant measurements application and Section 7.5 concludes.

7.2 The model

Copula approaches are very flexible, since they can be combined with different marginal distributions. For the air pollution measurements data with additional covariates, as analyzed in Section 7.4, we propose generalized additive models (GAMs) for the margins in combination with the novel copula state space model to capture dependencies. The GAM explains the effect of the covariates, while the copula based state space model handles temporal and cross-sectional dependence. In this section, we first introduce the marginal models (Section 7.2.1), which yield data on the copula scale. Then we review the linear Gaussian state space model (Section 7.2.2) and show an equivalent formulation in terms of Gaussian copulas (Section 7.2.3). In Section 7.2.4 we finally introduce the multivariate copula state space model as a generalization of the linear Gaussian state space model. The behavior of this model is illustrated with simulated data in Section 7.2.5.

7.2.1 Marginal models

We consider random vectors $\boldsymbol{Y_t} = (Y_{t1}, \dots, Y_{td})$ corresponding to d-dimensional continuous data, observed at the time points $t = 1, \dots, T$, that may depend on a q-dimensional covariate vector $\boldsymbol{x_t} = (x_{t1}, \dots, x_{tq})$.

In order to allow for more flexibility, we consider Box-Cox transformations (Box and Cox (1964)) of the response variables, i.e. we consider the transformed variables

$$BC(Y_{tj}, \lambda_j) = \begin{cases} \frac{Y_{tj}^{\lambda_j} - 1}{\lambda_j}, & \text{for } \lambda_j \neq 0 \\ \ln(Y_{tj}), & \text{for } \lambda_j = 0 \end{cases} . \tag{7.1}$$

for $t = 1, \ldots, T, j = 1, \ldots, d$. The relationship between the Box-Cox transformed variables and the covariates can be expressed in various ways using linear or nonlinear regression models. We assume a GAM (Hastie and Tibshirani (1986)) such that

$$BC(Y_{tj}, \lambda_j) = f_j(\boldsymbol{x_t}) + \sigma_j \varepsilon_{tj},$$

where $f_j(\cdot)$ is a smooth function of the covariates, expressing the mean of the GAM, and $\varepsilon_{tj} \sim N(0, 1)$. The standardized errors of the GAM are defined as

$$Z_{tj} = \frac{BC(Y_{tj}, \lambda_j) - f_j(\boldsymbol{x_t})}{\sigma_j}. \tag{7.2}$$

Note that $Z_{tj} \sim N(0, 1)$ holds.

We aim at modeling the errors $\boldsymbol{Z_t} = (Z_{t1}, \ldots, Z_{td})$ with a multivariate nonlinear non-Gaussian state space model based on copulas.

7.2.2 Linear Gaussian state space models

Suppose that we model the errors $\boldsymbol{Z_t}$, with $t = 1, \ldots, T$, extracted from the GAM as explained in Section 7.2.1, as a linear Gaussian state space model (see Section 2.1.1). Here, we assume that the variables Z_{tj}, $j = 1, \ldots, d$, are connected to a common univariate continuous state w_t. Hence, the model can be formulated as

$$Z_{tj} = \rho_{obs,tj}\, w_t + \sigma_{obs,tj}\, \eta_{obs,tj} \tag{7.3}$$

$$w_t = \rho_{lat,t}\, w_{t-1} + \sigma_{lat,t}\, \eta_{lat,t}, \tag{7.4}$$

where $\eta_{obs,tj}$, $\eta_{lat,t} \sim N(0, 1)$ iid, $\rho_{obs,tj}$, $\rho_{lat,t}$, $\sigma_{obs,tj}$ and $\sigma_{lat,t}$ are model parameters and $w_0 \sim N(\mu_{lat,0}, \sigma_{lat,0})$ with $\mu_{lat,0}$ and $\sigma_{lat,0}$ generally known. The linear Gaussian state space model can also be expressed using conditional distributions as

$$Z_{tj} | w_t, \rho_{obs,tj}, \sigma_{obs,tj} \sim N\left(\rho_{obs,tj}\, w_t, \sigma^2_{obs,tj}\right)$$
$$w_t | w_{t-1}, \rho_{lat,t}, \sigma_{lat,t} \sim N\left(\rho_{lat,t}\, w_{t-1}, \sigma^2_{lat,t}\right).$$

We assume time stationarity, i.e. $\rho_{obs,tj} = \rho_{obs,j}$, for $j = 1, \ldots, d$, and $\rho_{lat,t} = \rho_{lat}$. Since the model is applied to standardized errors with unit variance, we also set $\sigma^2_{obs,tj} = 1 - \rho^2_{obs,j}$ and $\sigma^2_{lat,t} = 1 - \rho^2_{lat}$. In addition, we assume that $\mu_{lat,0} = 0$ and $\sigma_{lat,0} = 1$. These assumptions imply that $Z_{tj} \sim N(0, 1)$ unconditionally. Hence, the model expressed through conditional distributions becomes

$$Z_{tj} | w_t, \rho_{obs,j} \sim N\left(\rho_{obs,j}\, w_t, 1 - \rho^2_{obs,j}\right)$$
$$w_t | w_{t-1}, \rho_{lat} \sim N\left(\rho_{lat}\, w_{t-1}, 1 - \rho^2_{lat}\right).$$

Thus, the state space model induces the following bivariate Gaussian distribution

$$\begin{pmatrix} Z_{tj} \\ w_t \end{pmatrix} | \rho_{obs,j} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{obs,j} \\ \rho_{obs,j} & 1 \end{pmatrix} \right) \tag{7.5}$$

$$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} | \rho_{lat} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{lat} \\ \rho_{lat} & 1 \end{pmatrix} \right). \tag{7.6}$$

Therefore, we obtain the joint distribution

$$(Z_{11}, \dots, Z_{d1}, w_1; Z_{12}, \dots, Z_{d2}, w_2; \dots, Z_{1T}, \dots, Z_{dT}, w_T) | (\rho_{obs,j})_{j=1,\dots,d}, \rho_{lat} \sim N_{dT+T}(\mathbf{0}, \Sigma)$$

with covariance matrix Σ (see Appendix E.1). Thus, the joint distribution of Z_{tj} and Z_{t-1j} is given by

$$\begin{pmatrix} Z_{tj} \\ Z_{t-1j} \end{pmatrix} | \rho_{obs,j}, \rho_{lat} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^2_{obs,j}\rho_{lat} \\ \rho^2_{obs,j}\rho_{lat} & 1 \end{pmatrix} \right).$$

7.2.3 Copula formulation of a Gaussian state space model

The linear Gaussian state space model in (7.5) and (7.6) can be equivalently expressed in the copula space using Gaussian copulas as follows

$$\begin{aligned} (U_{tj}, v_t) | \boldsymbol{\tau_{obs}} &\sim \mathbb{C}^{Gauss}_{U_j,V}(\cdot, \cdot; \tau_{obs,j}) \\ (v_t, v_{t-1}) | \tau_{lat} &\sim \mathbb{C}^{Gauss}_{V_2,V_1}(\cdot, \cdot; \tau_{lat}), \end{aligned} \tag{7.7}$$

where

$$U_{tj} = \Phi(Z_{tj}), \quad v_t = \Phi(w_t), \quad j = 1, \dots, d, \quad t = 1, \dots, T, \tag{7.8}$$

with Φ denoting the standard normal cumulative distribution function and $\boldsymbol{\tau_{obs}} = (\tau_{obs,1}, \dots, \tau_{obs,d})$. The variables U_{tj} and v_t are uniformly distributed as $U_{tj} \sim \mathrm{U}(0,1)$, $v_t \sim \mathrm{U}(0,1)$, while the variables Z_{tj} and w_t are normally distributed as $Z_{tj} \sim N(0,1)$, $w_t \sim N(0,1)$. The Gaussian copulas in (7.7) are parametrized by Kendall's τ, such that $\tau_{obs,j} = \frac{2}{\pi}\arcsin(\rho_{obs,j})$, $\tau_{lat} = \frac{2}{\pi}\arcsin(\rho_{lat})$.

7.2.4 Multivariate nonlinear non-Gaussian copula state space model

The multivariate nonlinear non-Gaussian copula state space model allows the copula families in (7.7) to be different from the Gaussian, thus gaining a much greater flexibility to accommodate a wide range of dependence structures.

More precisely, the proposed model can be expressed, in the copula scale, as follows

$$\begin{aligned} (U_{tj}, v_t) | \boldsymbol{\tau_{obs}}, \boldsymbol{m_{obs}} &\sim \mathbb{C}^{m_{obs,j}}_{U_j,V}(\cdot, \cdot; \tau_{obs,j}) \\ (v_t, v_{t-1}) | \tau_{lat}, m_{lat} &\sim \mathbb{C}^{m_{lat}}_{V_2,V_1}(\cdot, \cdot; \tau_{lat}), \end{aligned} \tag{7.9}$$

where $\boldsymbol{m_{obs}} = (m_{obs,1}, \dots, m_{obs,d})$ and the copula families $m_{obs,j}$, for $j = 1, \dots, d$, and m_{lat} are not necessarily equal and belong to a set $\mathcal{M}$ of single-parameter copula families,

parametrized by $\tau_{obs,j}$ and τ_{lat} (see Section 2.2.3). As in Sections 3.3.1 and 5.2.1, the Kendall' τ parameters are shared among different copula families.

The proposed *multivariate copula state space model* can also be specified in terms of conditional distribution functions as follows

$$\begin{aligned} U_{tj}|v_t, \boldsymbol{\tau_{obs}}, \boldsymbol{m_{obs}} &\sim \mathbb{C}^{m_{obs,j}}_{U_j|V}(\,\cdot\,|\,v_t;\,\tau_{obs,j}) \\ v_t|v_{t-1}, \tau_{lat}, m_{lat} &\sim \mathbb{C}^{m_{lat}}_{V_2|V_1}(\,\cdot\,|\,v_{t-1};\,\tau_{lat}) \end{aligned} \tag{7.10}$$

for $t = 1, \ldots, T, j = 1, \ldots, d$ with initial distribution $v_0 \sim \mathrm{U}(0,1)$.

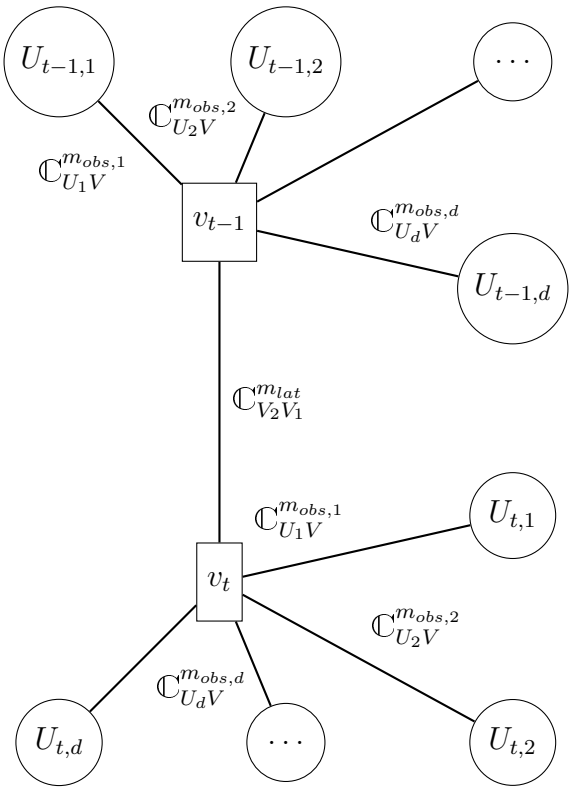

Figure 7.1: Graphical visualization of the multivariate copula state space model as specified in (7.9).

Figure 7.1 shows a graphical representation of the multivariate copula state space model. Each observed variable U_{tj} is linked to the latent state v_t via a copula $\mathbb{C}^{m_{obs,j}}_{U_jV}$ and the dependence between the latent states is modeled by the copula $\mathbb{C}^{m_{lat}}_{V_2V_1}$. In the following we denote by $c^{m_{obs,j}}_{U_j,V}$ and $c^{m_{lat}}_{V_2,V_1}$ the density functions of $\mathbb{C}^{m_{obs,j}}_{U_j,V}$ and $\mathbb{C}^{m_{lat}}_{V_2,V_1}$, respectively.

7.2.5 Illustration of the copula state space model with simulated data

We visualize bivariate dependence structures that are obtained from our model with normalized contour plots (see Czado (2019), Chapter 3). We consider three scenarios which differ in the choice of the family m_{lat} of the latent copula. The parameters are

chosen as follows

$$\begin{aligned} &T = 1000 \\ &d = 6 \\ &\boldsymbol{m_{obs}} = (\text{Gaussian, Gaussian, Clayton, Clayton, Gumbel, Gumbel}) \\ &\boldsymbol{\tau_{obs}} = (0.5, 0.7, 0.5, 0.7, 0.5, 0.7) \\ &\tau_{lat} = 0.7 \\ &m_{lat} = \begin{cases} \text{Gaussian, Scenario 1} \\ \text{Clayton, Scenario 2} \\ \text{Gumbel, Scenario 3} \end{cases} \end{aligned} \tag{7.11}$$

We consider one symmetric bivariate copula (Gaussian) and two asymmetric bivariate copulas (Gumbel, Clayton). We investigate two types of dependence: cross-sectional and temporal. For the cross-sectional dependence, we consider the pairs $(U_{tj}, U_{tj'})$ with $j \neq j'$ and corresponding bivariate copula density

$$c(u_{tj}, u_{tj'}) = \int_0^1 c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t) c_{U_{j'}V}^{m_{obs,j'}}(u_{tj'}, v_t) dv_t. \tag{7.12}$$

The bivariate marginal density of $(U_{tj}, U_{tj'})$ given in (7.12) is neither affected by the time t nor by the copula $\mathbb{C}_{V_2V_1}^{m_{lat}}$. So the cross-sectional dependence is not affected by the copula $\mathbb{C}_{V_2V_1}^{m_{lat}}$ and the corresponding theoretical contour plots are the same for all three scenarios. The empirical normalized contour plots for pairs $(U_{tj}, U_{tj'})$ are shown in Figure 7.2 for Scenario 1. The contour plots are constructed from 5000 independent simulations of the density in (7.12) for a fixed $t \in \{1, \ldots, T\}$.

We see that if both copulas $\mathbb{C}_{U_jV}^{m_{obs,j}}$ and $\mathbb{C}_{U_{j'}V}^{m_{obs,j'}}$ are Gaussian, the contour of $(U_{tj}, U_{tj'})$ looks Gaussian as well (see the panel in the second row and the first column in Figure 7.2). In this case $\mathbb{C}(u_{tj}, u_{tj'})$ is indeed a Gaussian copula. If we mix a Gaussian and an asymmetric copula (see the entries below Row 2 in Columns 1 and 2 in Figure 7.2) or if we combine two asymmetric copulas (see the lower triangular entries in Columns 3, 4 and 5 in Figure 7.2), we can obtain a variety of different asymmetric contour shapes.

For the temporal dependence, we consider the pairs (U_{tj}, U_{t-1j}) with bivariate copula density

$$c(u_{tj}, u_{t-1j}) = \int_{(0,1)^2} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t) c_{V_2V_1}^{m_{lat}}(v_t, v_{t-1}) c_{U_jV}^{m_{obs,j}}(u_{t-1j}, v_{t-1}) dv_t dv_{t-1}. \tag{7.13}$$

This dependence is affected by three copulas. Figure 7.3 shows empirical normalized contour plots of the density in (7.13) obtained from 5000 independent simulations. We can see that if at least one of these copulas is asymmetric, we may obtain an asymmetric dependence structure.

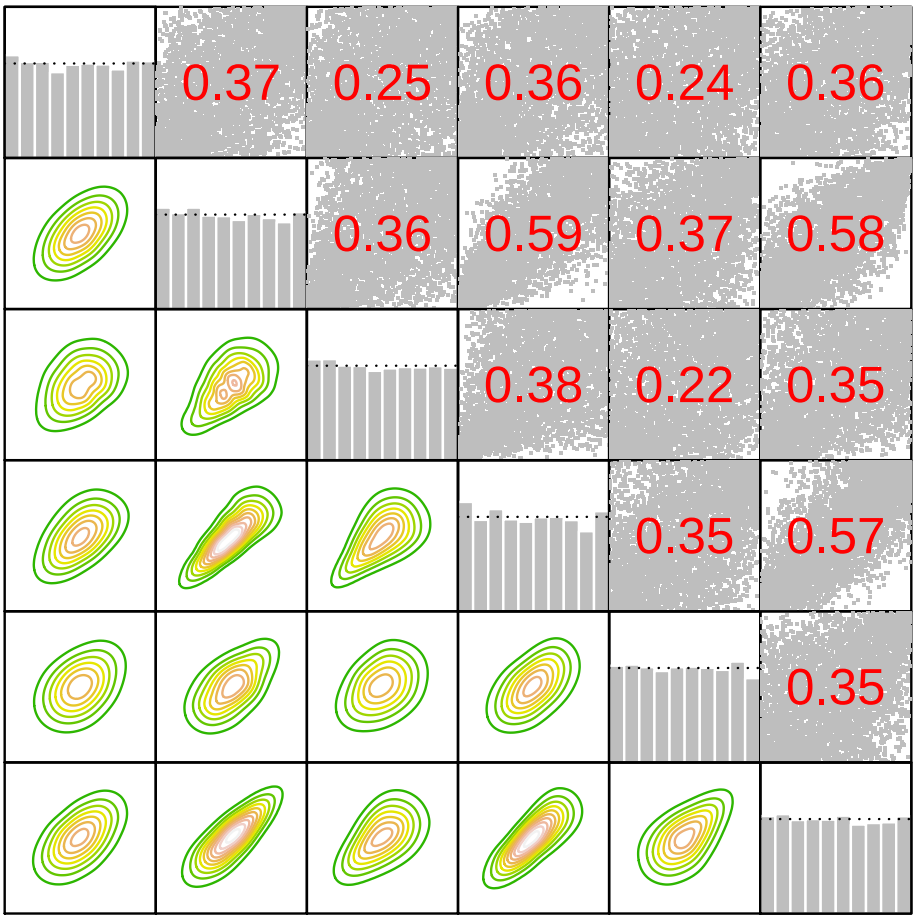

Figure 7.2: This plot is based on independently simulated data $(u_{tj}^r)_{r=1,\ldots,5000,j=1,\ldots,6}$ from Scenario 1 for a fixed $t \in \{1,\ldots,T\}$. The lower triangular part shows contour plots of bivariate kernel density estimates for all pairs of $(z_{t1}^r,\ldots,z_{t6}^r)$, where $z_{tj}^r = \Phi^{-1}(u_{tj}^r)$. The upper triangular part shows corresponding scatter plots and the empirical Kendall's τ for each pair $(u_{tj}^r, u_{tj'}^r)$. The diagonal shows the histograms of the univariate marginals. More precisely, the plot in the i-th row and j-th column shows the contour plot of the bivariate kernel density estimate based on the pair (z_{ti}^r, z_{tj}^r) if $i > j$, the scatter plot of (u_{ti}^r, u_{tj}^r) if $i < j$, or the histogram of u_{ti}^r, if $i = j$, with $r = 1,\ldots,5000$.

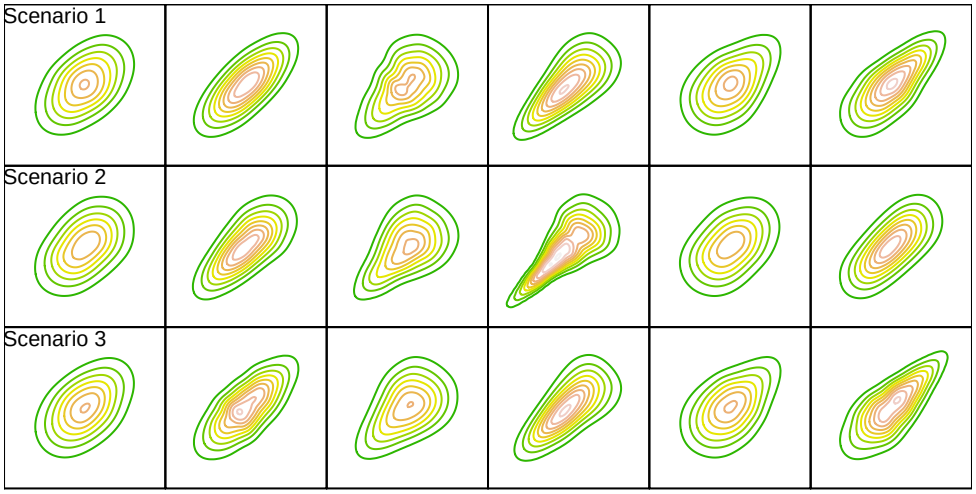

Figure 7.3: This plot is based on independently simulated data $(u^r_{t-1j}, u^r_{tj})_{r=1,\dots,5000,j=1,\dots,6}$ from Scenarios 1–3 for a fixed $t \in \{2,\dots,T\}$. The data is transformed to the normalized scale as $z^r_{t'j} = \Phi^{-1}(u^r_{t'j})$, $t' = t-1, t$. The plot in row m and column j shows the contour plot of the bivariate kernel density estimate based on $(z^r_{tj}, z^r_{t-1j})_{r=1,\dots,5000}$, simulated from the parameter specification of Scenario m.

7.3 Bayesian inference for the multivariate copula state space model

For the type of data we are dealing with, missing values are common. We denote the set of time indices of observed/non-missing values for dimension j by $\mathcal{T}_j^{obs}$ and the set of missing values by $\mathcal{T}_j^{miss} = \{1,\dots,T\} \setminus \mathcal{T}_j^{obs}$, $j = 1,\dots,d$. Further, we call $\mathsf{U}^{obs} = (u_{tj})_{t\in\mathcal{T}_j^{obs},j=1,\dots,d}$ the observed and $U^{miss} = (u_{tj})_{t\in\mathcal{T}_j^{miss},j=1,\dots,d}$ the missing values. The missing values can be treated as latent variables. Integrating out the missing values yields the following likelihood for the observed values U^{obs}

$$\begin{aligned}
\ell(\boldsymbol{v}, \boldsymbol{\tau_{obs}}, \boldsymbol{m_{obs}} | \mathsf{U}^{obs}) &= \int_{(0,1)^{|U_{miss}|}} \prod_{j=1}^{d} \prod_{t=1}^{T} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j}) dU^{miss} = \\
&= \prod_{j=1}^{d} \left(\prod_{t\in\mathcal{T}_j^{obs}} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j}) \prod_{t\in\mathcal{T}_j^{miss}} \int_{(0,1)} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j}) du_{tj} \right) \qquad (7.14) \\
&= \prod_{j=1}^{d} \prod_{t\in\mathcal{T}_j^{obs}} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j}).
\end{aligned}$$

Here, $\boldsymbol{v} = (v_0,\dots,v_T)$, $\boldsymbol{\tau_{obs}} = (\tau_{obs,1},\dots,\tau_{obs,d})$ and $\boldsymbol{m_{obs}} = (m_{obs,1},\dots,m_{obs,d})$. In contrast to a complete case analysis, information from all observed components is utilized in (7.14). The last equality in (7.14) uses the fact that in a copula the margins are uniform.

As mentioned above, we use a D-vine truncated after the first tree to capture temporal

dependence among the latent states, i.e.

$$\pi(\boldsymbol{v}|\tau_{lat}, m_{lat}) = \prod_{t=1}^{T} c_{V_2V_1}^{m_{lat}}(v_t, v_{t-1}; \tau_{lat}) \tag{7.15}$$

with Kendall's τ parameter τ_{lat} and copula family indicator $m_{lat} \in \mathcal{M}$. This is a general Markov model of order 1 and collapses to a Gaussian AR(1) process if the Gaussian copula is used.

We restrict $\tau_{obs,1} \in (0,1)$ to be positive to ensure identifiability. This restriction corresponds to restricting the diagonal entries of the factor loading matrix in conventional Gaussian factor models to be positive (see e.g. Lopes and West (2004)). For the Kendall's τ values of the remaining components we use a vague uniform prior on $(-1,1)$, reflecting the fact that we do not have prior knowledge about these quantities. The following priors are used

$$\tau_{obs,1} \sim Beta(10, 1.5), \;\; \tau_{obs,j} \sim \mathrm{U}(-1,1), \;\; j = 2, \dots, d, \;\; \tau_{lat} \sim \mathrm{U}(-1,1). \tag{7.16}$$

For the copula family indicators we use discrete uniform priors, i.e.

$$\pi(m_{obs,j}) = \pi(m_{lat}) = \frac{1}{|\mathcal{M}|} \tag{7.17}$$

for $j = 1, \dots, d$. Further, we assume that the Kendall's τ values and the copula family indicators are a priori independent such that the joint prior density is proportional to

$$\pi(\boldsymbol{\tau_{obs}}, \boldsymbol{m_{obs}}, \tau_{lat}, m_{lat}, \boldsymbol{v}) \propto \left(\prod_{t=1}^{T} c_{V_2V_1}^{m_{lat}}(v_t, v_{t-1}; \tau_{lat}) \right) \pi(\tau_{obs,1}),$$

where $\pi(\tau_{obs,1})$ is the prior density specified in (7.16). This prior density is a joint density of continuous and discrete parameters. For discrete parameters $\boldsymbol{\delta^{disc}}$ and continuous parameters $\boldsymbol{\delta^{cont}}$ the joint density is defined as

$$f(\boldsymbol{\delta^{cont}}, \boldsymbol{\delta^{disc}}) = f(\boldsymbol{\delta^{cont}}|\boldsymbol{\delta^{disc}}) f(\boldsymbol{\delta^{disc}}),$$

where $f(\boldsymbol{\delta^{cont}}|\boldsymbol{\delta^{disc}})$ is a conditional probability density function and $f(\boldsymbol{\delta^{disc}})$ is a joint probability mass function.

The set of parameters can be summarized as $\mathcal{P} = \{\tau_{lat}, \boldsymbol{\tau_{obs}}, m_{lat}, \boldsymbol{m_{obs}}, \boldsymbol{v}\}$. The posterior density of our model is proportional to

$$f(\mathcal{P}|\mathrm{U}^{obs}) \propto \left(\prod_{j=1}^{d} \prod_{t \in \mathcal{T}_j^{obs}} c_{U_jV}^{m_{obs,j}}(u_{tj}, v_t, \tau_{obs,j}) \right) \left(\prod_{t=1}^{T} c_{V_2V_1}^{m_{lat}}(v_t, v_{t-1}; \tau_{lat}) \right) \pi(\tau_{obs,1}). \tag{7.18}$$

As in Chapter 6, sampling from the posterior in (7.18) is not straightforward, e.g. Kalman filter recursions cannot be applied. Since the No-U-turn sampler of Hoffman and Gelman (2014) has shown good performance for the univariate copula state space model, discussed in Chapter 6, we also use it here.

Updating continuous parameters

Since Hamiltonian Monte Carlo cannot deal with discrete variables, we integrate over the discrete family indicators which corresponds to summing over them, i.e.

$$\begin{aligned} f(\tau_{lat}, \boldsymbol{\tau_{obs}}, \boldsymbol{v}|\mathbb{U}^{obs}) &= \sum_{(m_{lat}, \boldsymbol{m_{obs}}) \in \mathcal{M}^{d+1}} f(\tau_{lat}, \boldsymbol{\tau_{obs}}, m_{lat}, \boldsymbol{m_{obs}}, \boldsymbol{v}|\mathbb{U}^{obs}) \\ &\propto \prod_{j=1}^{d} \left(\sum_{m_{obs,j} \in \mathcal{M}} \prod_{t \in \mathcal{T}_j^{obs}} c_{U_j V}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j}) \right) \cdot \\ &\quad \cdot \left(\sum_{m_{lat} \in \mathcal{M}} \prod_{t=1}^{T} c_{V_2 V_1}^{m_{lat}}(v_t, v_{t-1}; \tau_{lat}) \right) \pi(\tau_{obs,1}) \end{aligned} \tag{7.19}$$

To sample from this density we use STAN's No-U-Turn sampler.

Updating the (discrete) copula family indicators

In $f(\boldsymbol{m_{obs}}, m_{lat}|\tau_{lat}, \boldsymbol{\tau_{obs}}, \boldsymbol{v}, U^{obs})$, all components of $(\boldsymbol{m_{obs}}, m_{lat})$ are independent. We have that

$$\begin{aligned} f(m_{obs,j}|\tau_{lat}, \boldsymbol{\tau}_{obs}, \boldsymbol{v}, \boldsymbol{m_{obs,-j}}, m_{lat}, \mathbb{U}^{obs}) = \\ \frac{f(\tau_{lat}, \boldsymbol{\tau_{obs}}, m_{lat}, \boldsymbol{m_{obs}}, \boldsymbol{v}|\mathbb{U}^{obs})}{\sum_{m'_{obs,j} \in \mathcal{M}} f(\tau_{lat}, \boldsymbol{\tau_{obs}}, m_{lat}, \boldsymbol{m_{obs,-j}}, m'_{obs,j}, \boldsymbol{v}|\mathbb{U}^{obs})}, \end{aligned}$$

where $\boldsymbol{m_{obs,-j}}$ is equal to $\boldsymbol{m_{obs}}$ with the j-th component removed. Therefore, we obtain

$$f(m_{obs,j}|\tau_{lat}, \boldsymbol{\tau_{obs}}, \boldsymbol{v}, \boldsymbol{m_{obs,-j}}, m_{lat}, \mathbb{U}^{obs}) = \frac{\prod_{t \in \mathcal{T}_j^{obs}} c_{U_j V}^{m_{obs,j}}(u_{tj}, v_t; \tau_{obs,j})}{\sum_{m'_{obs,j} \in \mathcal{M}} \prod_{t \in \mathcal{T}_j^{obs}} c_{U_j V}^{m'_{obs,j}}(u_{tj}, v_t; \tau_{obs,j})}. \tag{7.20}$$

Similarly, we obtain

$$f(m_{lat}|\tau_{lat}, \boldsymbol{\tau_{obs}}, \boldsymbol{v}, \boldsymbol{m_{obs}}, \mathbb{U}^{obs}) = \frac{\prod_{t=1}^{T} c_{V_2 V_1}^{m_{lat}}(v_t, v_{t-1}; \tau_{lat})}{\sum_{m'_{lat} \in \mathcal{M}} \prod_{t=1}^{T} c_{V_2 V_1}^{m'_{lat}}(v_t, v_{t-1}; \tau_{lat})}. \tag{7.21}$$

Obtaining updates for the joint posterior density

To obtain R samples from the posterior density given in (7.18), we first obtain R samples of $\tau_{lat}, \boldsymbol{\tau_{obs}}, \boldsymbol{v}$ from the density given in (7.19) using STAN. We denote the samples by τ_{lat}^r, $\boldsymbol{\tau_{obs}^r}$, $\boldsymbol{v^r}$, $r = 1, \ldots, R$. Then we sample $m_{obs,j}$ from $f(m_{obs,j}|\tau_{lat}^r, \boldsymbol{\tau_{obs}^r}, \boldsymbol{v^r}, \mathbb{U}^{obs})$ (see (7.20)) to obtain $m_{obs,j}^r$, for $r = 1, \ldots, R$ and $j = 1, \ldots, d$. Further, m_{lat}^r is obtained by sampling from $f(m_{lat}|\tau_{lat}^r, \boldsymbol{\tau_{obs}^r}, \boldsymbol{v^r}, \mathbb{U}^{obs})$ (see (7.21)) , for $r = 1, \ldots, R$.

Predictive distribution (in-sample period)

The predictive density of a new value u_{tj}^{new} for margin j at time $t \in \{1, \ldots, T\}$ is the conditional density of u_{tj}^{new} given $\mathbb{U}^{obs}$, obtained as

$$f(u_{tj}^{new}|\mathbb{U}^{obs}) = \int_{domain(\mathcal{P})} f(u_{tj}^{new}, \mathcal{P}|\mathbb{U}^{obs}) d\mathcal{P} = \int_{domain(\mathcal{P})} f(u_{tj}^{new}|\mathcal{P}, \mathbb{U}^{obs}) f(\mathcal{P}|\mathbb{U}^{obs}) d\mathcal{P}$$

with $f(u_{tj}^{new}|\mathcal{P}, \mathbb{U}^{obs}) = c_{U_jV}^{m_{obs,j}}(u_{tj}^{new}, v_t; \tau_{obs,j})$ and $domain(\mathcal{P})$ is the domain of the parameter space $\mathcal{P}$. Note that for the discrete indicator variables the integral is a sum.

To obtain samples from the (in-sample period) predictive distribution, we sample from the following density

$$f(u_{tj}^{new}, \mathcal{P}|\mathbb{U}^{obs}) = f(u_{tj}^{new}|\mathcal{P}, \mathbb{U}^{obs}) f(\mathcal{P}|\mathbb{U}^{obs}).$$

We proceed as follows:

- We first simulate R samples of $\mathcal{P}$ from $f(\mathcal{P}|\mathbb{U}^{obs})$ as described above.
- The r-th sample of u_{tj}^{new}, denoted by $(u_{tj}^{new})^r$, is simulated from $\mathbb{C}_{U_j|V}^{m_{obs,j}^r}(\cdot|v_t^r; \tau_{obs,j}^r)$, for $r = 1, \ldots, R$.

For $t \in \mathcal{T}_j^{miss}$, we can obtain simulated values for the missing values.

Predictive distribution (out-of-sample period)

To obtain samples from the (out-of-sample period) predictive distribution of a new value u_{tj}^{new} for margin j at time $t \in \{T+1, T+2, \ldots\}$, we consider the following density

$$f(u_{tj}^{new}, \mathcal{P}|\mathbb{U}^{obs}) = f(u_{tj}^{new}|\mathcal{P}, \mathbb{U}^{obs}) f(\mathcal{P}|\mathbb{U}^{obs})$$

with

$$f(u_{tj}^{new}|\mathcal{P}, \mathbb{U}^{obs}) = \int_{(0,1)^{t-T}} c_{U_jV}^{m_{obs,j}}(u_{tj}^{new}, v_t; \tau_{obs,j}) \prod_{t'=T+1}^{t} c_{V_2V_1}^{m_{lat}}(v_{t'}, v_{t'-1}; \tau_{lat}) dv_{T+1} \ldots, dv_t.$$

We proceed as follows to obtain samples from this density

- We first simulate R samples of $\mathcal{P}$ from $f(\mathcal{P}|\mathbb{U}^{obs})$ as described above.
- For $r = 1, \ldots, R$ and for $t' = T+1, \ldots, t$:
  Sample $v_{t'}$ from $\mathbb{C}_{V_2V_1}^{m_{lat}^r}(\cdot|v_{t'-1}^r; \tau_{lat}^r)$ and denote the sample by $v_{t'}^r$.
- For $r = 1, \ldots, R$: Sample u_{tj}^{new} from $\mathbb{C}_{U_j|V}^{m_{obs,j}^r}(\cdot|v_t^r; \tau_{obs,j}^r)$ and denote the sample by $(u_{tj}^{new})^r$.

Note that the recursive sampling avoids the evaluation of the $t-T$ dimensional integral.

7.4 Data analysis

7.4.1 Data description

We consider a subset of the data set available at `http://archive.ics.uci.edu/ml/datasets/Air+Quality` (De Vito et al. (2008, 2009, 2012)). The data set contains hourly averaged concentration measurements for different atmospheric pollutants obtained at a main road in an Italian city. Here we analyze measurements from June to September 2004, which results in 2928 observations. The measurements for the pollutants were taken from

two different sensors, standard (high-cost) sensors and new low-cost (lc) sensors. We refer to a value measured with the standard (high-cost) sensor as a ground truth (gt) value. Ground truth values are available for CO (mg/m^3), NOx (ppb) and NO2 (μg/m^3) and the aim is to predict these values. For each ground truth value we are given a corresponding value obtained from a low-cost sensor, resulting in six different pollution measurements for one time point. The measurements in July for the pollutant CO are visualized in Figure 7.4. We see that the measurements of the ground truth sensor for CO are missing for several days, i.e. missing observations are present in this data set. The missing values per pollutant range from 4% to 24%, whereas ground truth values have a higher portion of missing values. In addition to the pollution measurements, hourly measurements of the temperature and of relative humidity are also available.

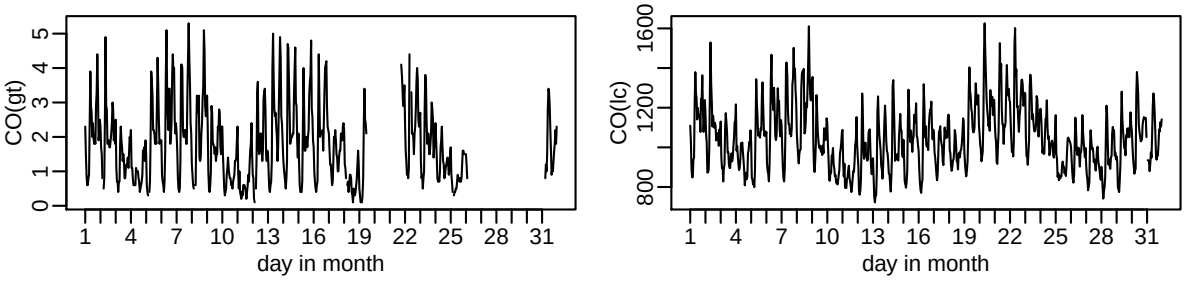

Figure 7.4: Hourly observed values of one pollutant (CO) from the ground truth (gt) and low-cost (lc) sensors in July 2004. When missing values are present, no observations are drawn for the corresponding time points.

In the following, $(y_{tj})_{t=1,\dots,T,j=1,\dots,6}$ denotes the data containing the pollutant measurements, where $T = 2928$ is the length of the time series. As before, $\mathcal{T}_j^{obs}$ is the set of time indices for which observed values are available for the j-th marginal time series. The measurements of relative humidity and temperature are denoted by RH_t and TEMP_t, respectively for $t = 1, \dots, T$.

7.4.2 Marginal models

We fit a generalized additive model (GAM) for each pollutant, where temperature, relative humidity, the hour at time t, $\mathrm{H}_t \in \{0, \dots, 23\}$, and the day at time t, $\mathrm{D}_t \in \{0, \dots, 6\}$, are used as covariates. We denote the covariates by $\boldsymbol{x_t} = (\mathrm{TEMP}_t, \mathrm{RH}_t, \mathrm{H}_t, \mathrm{D}_t)$. As explained in Section 7.2.1, we allow for Box-Cox transformations (Box and Cox (1964)) and assume that

$$BC(Y_{tj}, \lambda_j) = f_j(\boldsymbol{x_t}) + \sigma_j \epsilon_{tj} \tag{7.22}$$

with $\epsilon_{tj} \sim N(0,1)$ for $t = 1, \dots, T, j = 1, \dots, 6$ and $BC(Y_{tj}, \lambda_j)$ as in (7.1).

For estimating the conditional mean function f_j and σ_j, we assume that the errors ϵ_{tj} are independent. Later the dependence among the errors will be modeled with the proposed state space model. For each pollutant, associated with a $j \in \{1, \dots, 6\}$, we estimate a GAM for different values of λ_j and then choose the model which maximizes the likelihood for given data $y_{tj}, t \in \mathcal{T}_j^{obs}$. For each GAM we remove the corresponding missing values and rely on the R package `mgcv` of Wood and Wood (2015) for parameter

estimation. We obtain estimates $\hat{f}_j$, $\hat{\sigma}_j$ and $\hat{\lambda}_j$ for $j = 1, \ldots, 6$. From Table 7.1 we see that the estimates for λ_j deviate from 1, which indicates that the Box-Cox transformations are necessary. Figure 7.5 shows the smooth components of the GAM for four different pollutants. We see, for example, a nonlinear effect of the Hour on the pollution measurement. The pollution is high at around 8 am and at around 6 pm, which may correspond to the hours with the highest traffic due to commuting workers.

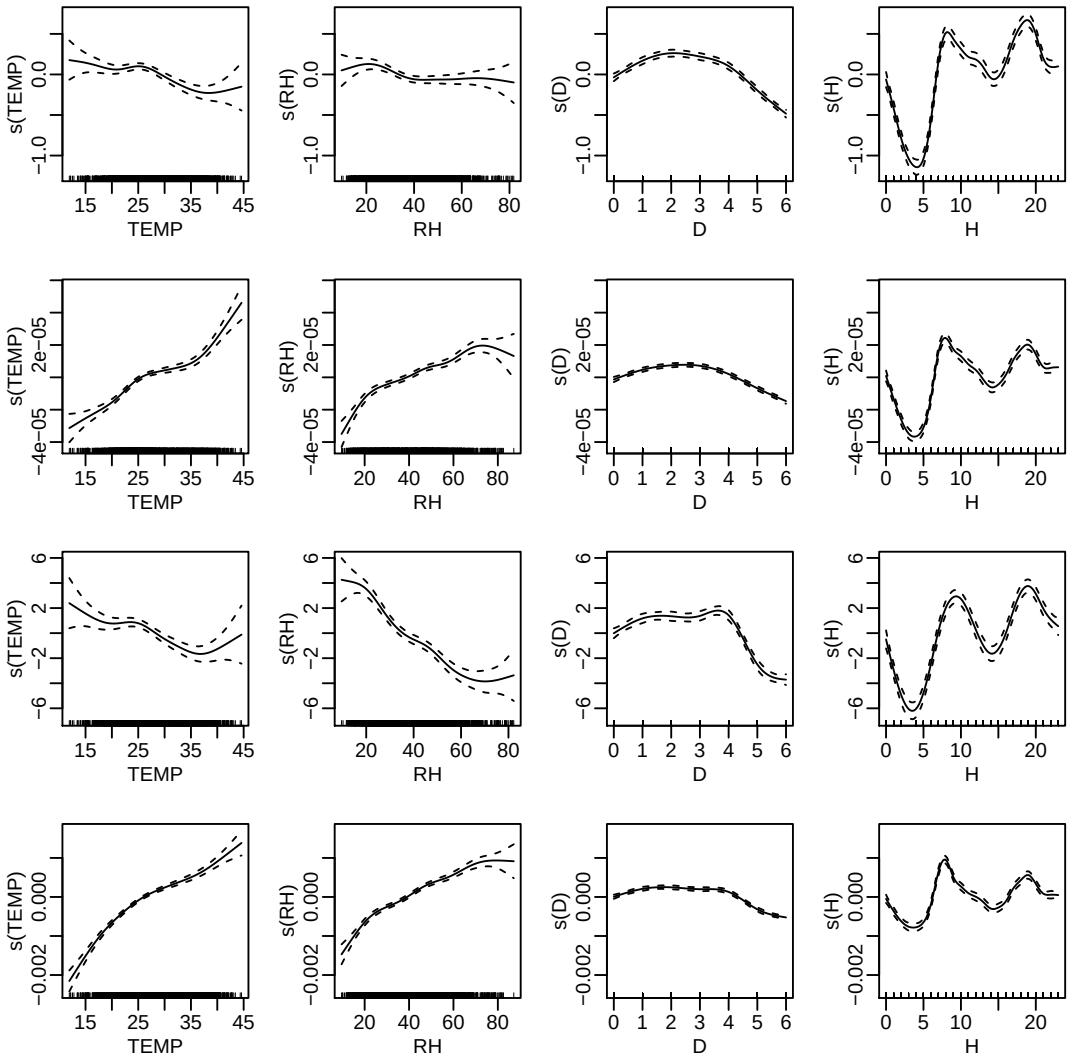

Figure 7.5: Estimated smooth components of the GAMs for four Box-Cox transformed pollutants: CO(gt), CO(lc), NO2(gt), NO2(lc) (top to bottom row). Each GAM has four covariates, TEMP, RH, D and H. The dashed lines represent a pointwise 95% confidence band.

	CO(gt)	CO(lc)	NOx(gt)	NOx(lc)	NO2(gt)	NO2(lc)
$\hat{\lambda}_j$	0.15	-1.25	0.05	0.05	0.55	-0.70

Table 7.1: Estimates of $\lambda_1, \ldots, \lambda_6$ for the six GAMs fitted to the six pollution measurements.

7.4.3 Dependence model

Recall the standardized errors Z_{tj}, defined in (7.2), as

$$Z_{tj} = \frac{BC(Y_{tj}, \lambda_j) - f_j(\boldsymbol{x_t})}{\sigma_j}$$

which are $N(0,1)$ distributed. Pseudo observations of Z_{tj} can be obtained from the estimates $\hat{f}_j$, $\hat{\sigma}_j$ and $\hat{\lambda}_j$ as

$$\hat{z}_{tj} = \frac{BC(y_{tj}, \hat{\lambda}_j) - \hat{f}_j(\boldsymbol{x_t})}{\hat{\sigma}_j} \tag{7.23}$$

for $t = 1, \ldots, T, j = 1, \ldots 6$. To visualize temporal dependence among the variables Z_{tj}, we show contour plots of bivariate kernel density estimates of pairs $(\hat{z}_{tj}, \hat{z}_{t-1j}), t = 2, \ldots, T$ for $j = 1, \ldots, 6$ in Figure 7.6. In addition, we examine cross-sectional dependencies through bivariate contour plots for all pairs of $(\hat{z}_{t1}, \ldots, \hat{z}_{t6}), t = 1, \ldots, T$ in Figure 7.7, whereas we ignore serial dependence. We observe temporal and cross-sectional dependence. Further, the dependence structures seem to be different from a Gaussian one since we observe asymmetries in the contour plots. For example, the contour plot in the bottom left corner of Figure 7.7 indicates stronger dependence in the upper right corner than in the bottom left corner. Therefore, a linear Gaussian state space model might not be appropriate here, but the proposed copula based state space model can be a good candidate for this data.

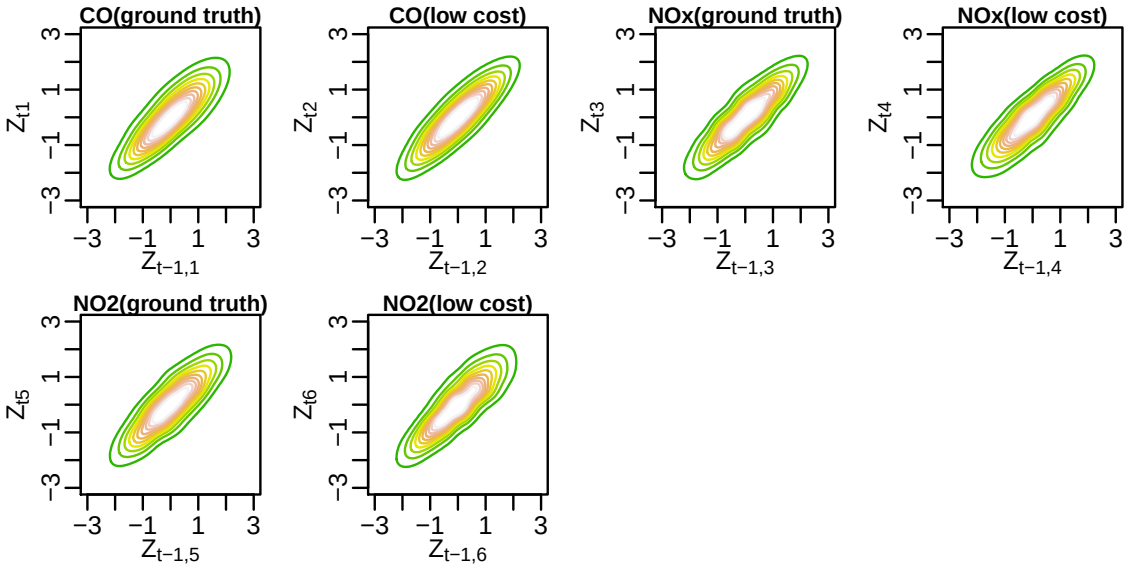

Figure 7.6: Contour plots of bivariate kernel density estimates based on pairs $(\hat{z}_{tj}, \hat{z}_{t-1j})_{t=2,\ldots,T}$ for $j = 1, \ldots, 6$ ignoring serial dependence.

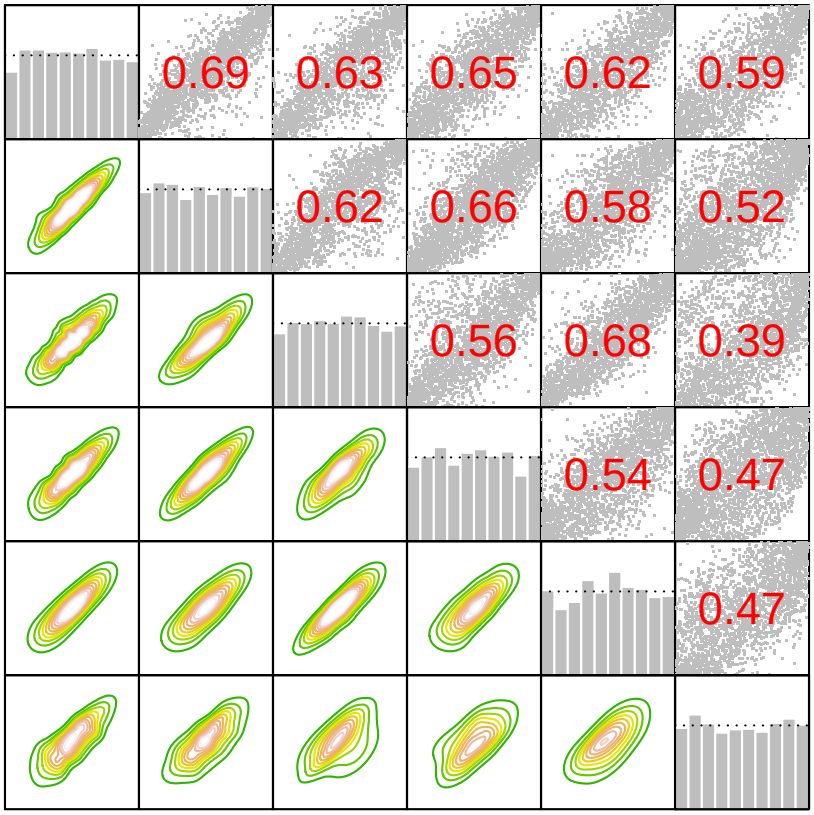

Figure 7.7: The lower triangular part shows contour plots of bivariate kernel density estimates for all pairs of $(\hat{z}_{t1}, \ldots, \hat{z}_{t6})$, ignoring serial dependence. The upper triangular part shows corresponding scatter plots of all pairs of $(\hat{u}_{t1}, \ldots, \hat{u}_{t6})$ with $\hat{u}_{tj} = \Phi(\hat{z}_{tj})$ and the empirical Kendall's τ for each pair. The diagonal shows the histograms of the univariate marginals. More precisely, the plot in the i-th row and j-th column shows the contour plot of the bivariate kernel density estimate based on the pair $(\hat{z}_{ti}, \hat{z}_{tj})$ if $i > j$, the scatter plot of $(\hat{u}_{ti}, \hat{u}_{tj})$ if $i < j$, or the histogram of $\hat{u}_{ti}$, if $i = j$ with $t = 1, \ldots, T$. The variables are ordered as follows: 1: CO(gt), 2: CO(lc), 3: NOx(gt), 4: NOx(lc), 5: NO2(gt), 6: NO2(lc).

Since our multivariate copula state space model operates on marginally uniform$(0, 1)$ distributed data, we obtain uniform on $(0, 1)$ distributed random variables as $U_{tj} = \Phi(Z_{tj})$ with corresponding pseudo copula data

$$\hat{u}_{tj} = \Phi(\hat{z}_{tj}) \tag{7.24}$$

for $t \in \mathcal{T}_j^{obs}, j = 1, \ldots, 6$. The proposed multivariate copula based state space model is fitted to the data $\hat{u}_{tj}, t \in \mathcal{T}_j^{obs}, j = 1, \ldots, 6$, whereas the copula families are selected from $\mathcal{M}$={Gaussian, Student t(df=4), Clayton, Gumbel}. The MCMC approach of Section 7.3 is run for 3000 iterations and the first 1000 draws are discarded for burn-in. Plots of the estimated posterior densities and trace plots are shown in Appendix E.2. These plots indicate proper mixing of the Markov Chain. Table 7.2 shows the selected copula

families corresponding to the estimated posterior modes of $m_{obs,j}$ or m_{obs}. We see that four Gaussian, one Student t and two Gumbel copulas were selected. In particular, our model features an asymmetric dependence structure, since the Gumbel copula is included. Simulations from the in-sample period predictive distribution can be obtained as explained in Section 7.3. Transforming these simulations with the standard normal quantile function, we obtain predictive simulations for the standardized errors, i.e. we obtain draws from the predictive distribution of the standardized error as

$$\epsilon_{tj}^r = \Phi^{-1}((u_{tj}^{new})^r), \tag{7.25}$$

for $r = 1, \ldots, 3000, t = 1, \ldots, T$, where $(u_{tj}^{new})^r$ is a draw from the in-sample period predictive distribution on the copula scale (see Section 7.3). These simulations are compared to the observed standardized residual of the GAM, $\hat{z}_{tj}$, to assess how well our model fits the data. In particular, we want to asses if a single factor structure is appropriate or if it is too restrictive. According to Figure 7.8, the model seems to be appropriate. The single factor structure is able to capture the time-dynamics of the residuals. The ground truth values for CO are missing from day 26 to day 30. We see that within this period, the time-dynamic is learned from other series, where data is available within this period. While Figure 7.8 shows plots for two pollutants in July, plots for different pollutants in different months looked similar.

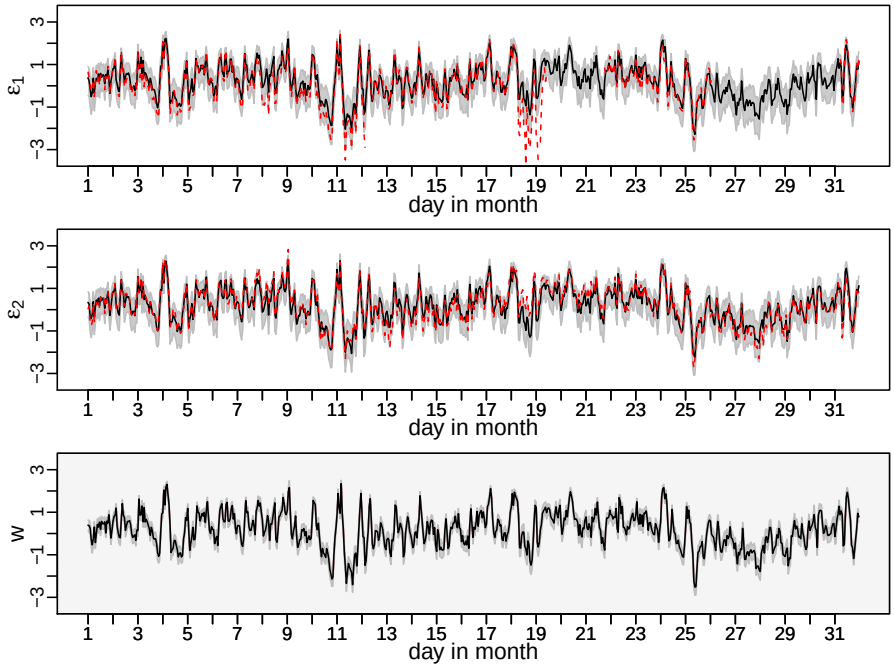

Figure 7.8: This plot is based on data for July. The first two rows show the estimated mode of the predictive density of the standardized error ϵ_{tj} plotted against time t for $j = 1, 2$, corresponding to CO(gt) and CO(lc) (black lines). Draws from the predictive distribution of the standardized error, which are used to estimate the modes, are obtained as in (7.25). The observed standardized residual from the GAM is added in red (dashed). The third row shows the estimated posterior mode of $w_t = \Phi^{-1}(v_t)$ plotted against t. To all plots we add a 90% credible region constructed from the estimated 5% and 95% posterior quantiles. The mode estimates are obtained from univariate kernel density estimates.

	$\hat{m}_{obs,1}$	$\hat{m}_{obs,2}$	$\hat{m}_{obs,3}$	$\hat{m}_{obs,4}$	$\hat{m}_{obs,5}$	$\hat{m}_{obs,6}$	$\hat{m}_{lat}$
Copula family	Gu	Gu	Ga	S	Ga	Ga	Ga

Table 7.2: The marginal posterior mode estimates of the copula family indicators $\boldsymbol{m}_{obs}, m_{lat}$. (Ga: Gaussian, S: Student t(df=4), C: Clayton, Gu: Gumbel).

7.4.4 Predictions

We evaluate the proposed model's ability to predict the ground truth values. Therefore we compare the copula state space model to a Gaussian state space model and to Bayesian additive regression trees (Chipman et al. (2010)), as a representative for a popular machine learning algorithm. Compared to other machine learning techniques, Bayesian additive regression trees have the advantage that a predictive distribution is obtained instead of a single point estimate. Therefore we can compare models with respect to their forecast distribution, for which we utilize the continuous ranked probability score (Gneiting and Raftery (2007)). The continuous ranked probability score (CRPS) for an observed value $y \in \mathbb{R}$ and a univariate forecast distribution function F is defined as

$$CRPS = \int_{\mathbb{R}} (F(z) - \mathbb{1}_{y \leq z})^2 dz. \tag{7.26}$$

For each of the ground truth values we remove the observations in the last month of the data set and treat them as missing values, which yields the training set. Further, we denote by $\mathcal{T}_j^{test}$ the removed time indices for margin j. Based on the training set, we proceed similarly to what we described above, i.e. we first estimate the GAMs, and then estimate the state space model on the copula scale. Here two state space models are estimated: the copula state space model where the family set $\mathcal{M}$ is chosen as in Section 7.4.3 and the Gaussian state space model where we restrict the family set to $\mathcal{M} = \{\text{Gaussian}\}$. For each of the two state space models, we obtain 2000 simulations from the in-sample period predictive distribution u_{tj}^r, $r = 1, \ldots, 2000$, whereas our MCMC approach of Section 7.3 is run for 3000 iterations and the first 1000 draws are discarded for burn-in. Here t is a time point which is among the newly selected missing values for the ground truth value that corresponds to margin j, i.e. $t \in \mathcal{T}_j^{test}$. Based on these simulations, we obtain simulations from the predictive distribution of the Box-Cox transformed response as follows

$$(y_{tj}^{bc})^r = \hat{f}_j(\boldsymbol{x_t}) + \hat{\sigma}_j \Phi^{-1}(u_{tj}^r) \tag{7.27}$$

for $r = 1, \ldots, 2000$.

Since Bayesian additive regression trees rely on the normal distribution, we expect that Box-Cox transformations might also improve the fit for this model. We assume that

$$BC(Y_{tj}, \lambda_j) = g_j(\boldsymbol{x_{tj}^{BART}}) + \sigma_j \epsilon_{tj}, \tag{7.28}$$

where $\boldsymbol{x_{tj}^{BART}}$ are the covariates, $g_j(\cdot)$ is a sum of regression trees and $\epsilon_{tj} \sim N(0,1)$ iid. In addition to the covariates used for the GAM model, all pollutant measurements except the one corresponding to margin j are included in the covariate vector $\boldsymbol{x_{tj}^{BART}}$. For λ_j we use the same value as for the previously fitted GAM. We have seen that this transformation improves the performance of the Bayesian additive regression trees. McCulloch et al.

(2018) implement a MCMC sampler in the `R` package `BART` which we use to obtain draws g_j^r, σ_j^r, $r = 1, \ldots, 10000$ from the corresponding posterior distribution. We discard the first 5000 of these draws and then 5000 simulations from the predictive distribution of the Box-Cox transformed response are obtained as

$$(y_{tj}^{bc})^r \sim N(g_j^r(\boldsymbol{x}_{tj}^{\boldsymbol{BART}}), (\sigma_j^r)^2) \tag{7.29}$$

for $r = 1, \ldots, 5000$.

Based on the simulations, $(y_{tj}^{bc})^r, r = 1, \ldots, 5000$, we calculate the empirical distribution function and use this to approximate the CRPS (this is implemented in the `R` package `scoringRules` of Jordan et al. (2017)). For each of the ground truth indices, associated with an index j, we calculate the CRPS for the time points $\mathcal{T}_j^{test}$ and sum them up to obtain the cumulative CRPS. For each of the three methods, we obtain a cumulative CRPS for each of the three ground truth indices. In addition, we consider reduced bivariate data sets, where each data set consists of the ground truth value of a pollutant, the corresponding low-cost value and the covariates as in Section 7.4.2. This yields three reduced data sets, each associated with one of the three pollutants. For each of the reduced data sets we proceed as above, i.e. we first remove ground truth observations in the last month, fit the three different models and calculate the CRPS values.

We refer to the models fitted to the reduced data as bivariate state space models and reduced Bayesian additive regression trees. The models estimated with the full data are referred to as joint models. We want to investigate how the bivariate state space models compare to the six-dimensional ones. The cumulative CRPS values are shown in Table 7.3. For the pollutant NOx, the state space approach seems not to be the best choice. We have seen (see Appendix E.3) that for this pollutant, the dependence between the ground truth and the low-cost values varies more over time than for the other pollutants. Relaxing the assumption of a time-constant Kendall's τ might improve the predictive accuracy for this pollutant. This model extension is subject to future research. Overall, the copula state space model is the best performing model within this comparison, since it outperforms the Gaussian state space model and the Bayesian additive regression trees in two out of three cases.

	CO	NOx	NO2
joint copula state space model	**74.27**	594.50	**569.03**
bivariate copula state space model	84.92	559.22	845.95
joint Gaussian state space model	76.91	594.30	570.55
bivariate Gaussian state space model	87.90	559.18	844.64
joint Bayesian additive regression trees	183.49	**379.31**	1330.90
reduced Bayesian additive regression trees	89.39	520.93	1095.40

Table 7.3: Cumulative CRPS for the three ground truth values (CO, NOx, NO2) obtained from six different models: joint/bivariate copula state space model, joint/bivariate Gaussian state space model, joint/reduced Bayesian additive regression trees. The best, i.e. the lowest, cumulative CRPS value in marked in bold.

7.5 Concluding remarks

We proposed a multivariate nonlinear non-Gaussian copula based state space model. The model is very flexible: the observation and the state equation are specified with copulas and the model can be combined with different marginal distributions. We illustrated the model with air pollution measurements data and have shown that the novel copula state space model outperforms a Gaussian state space model and Bayesian additive regression trees. As we have seen in Section 7.4.4, the assumption of a time-constant dependence might not always be appropriate. A first extension of the model could allow for dynamic dependence parameters. For this, ideas of the dynamic bivariate copula model of Almeida and Czado (2012) might be used. Another area of future research is the extension to higher-dimensional latent states.

8 Conclusion and outlook

In this thesis we presented a variety of copula based time series models: A single factor copula stochastic volatility model, a dynamic bivariate mixture copula model, dynamic vine copula models and a univariate and a multivariate copula based state space model. For the estimation of these models we developed our own MCMC methods or relied on STAN's No-U-Turn sampler. All models were illustrated with real data and comparison to relevant benchmark models showed satisfactory performance of the proposed models.

In addition to suggestions for future research that were already given at the end of each chapter, there are more promising ideas for future research.

To extend the single factor copula stochastic volatility model, discussed in Chapter 3, to allow for dynamic parameters, we could make use of the methodology developed in Chapter 4. We can allow for time-variation by modeling copula parameters with latent AR(1) processes. Instead of using Hamiltonian Monte Carlo, as in Chapter 3, we can utilize the sampler developed in Chapter 4. The sampler could be used within a Gibbs approach. Conditional on the latent factor, we need to sample from d independent bivariate dynamic copula models, where d is the number of variables. To further improve the efficiency of this sampler, we could develop an additional interweaving strategy that involves the latent factor. We might follow ideas of Kastner et al. (2017), who proposed an interweaving strategy for multivariate factor stochastic volatility models. It involves the latent factor and resulted in huge efficiency gains. Further, it might be interesting to investigate if the methodology of Chapter 5 can be applied to allow for dynamic factor models with more than one factor. We would start with the factor copula model of Krupskii and Joe (2013) with multiple factors and extend it by allowing for dynamic dependence parameters. Conditional on all latent factors, we would already be in the framework of Chapter 5, where the corresponding tree structure is a C-vine and the steps for structure selection would not be necessary in this case.

Similarly, an extension of the multivariate copula state space model proposed in Chapter 7 could benefit from the methodology developed in Chapters 4 and 5. We would employ the factor copula model with multiple factors in the observation equation, describe the time evolution of latent factors with independent D-vines and model the associated copula parameters with latent AR(1) processes. Conditional on the latent factors, we obtain again a dynamic C-vine copula which fits into the framework of Chapter 5.

Since the models of Chapters 3, 4 and 5 are state space models, where the state equations are Gaussian AR(1) processes, the methodology developed in Chapter 6 yields straightforward generalizations of these models: We can replace the Gaussian AR(1) state equations with latent D-vines. So a latent D-vine could describe the time evolution of the copula parameter in the dynamic bivariate copula model discussed in Chapter 4. In this

case, we face new challenges with respect to parameter estimation. As a first approach we could try STAN's No-U-Turn sampler. Alternatively, we might be able to develop a sampling procedure that exploits the latent D-vine structure, similar to how elliptical slice sampling exploits the latent Gaussian AR(1) structure. Instead of sampling from a multivariate normal distribution, as in elliptical slice sampling, we would sample from a D-vine copula to obtain proposals.

Bibliography

Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4):43.

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Abanto-Valle, C., Lachos, V., and Dey, D. K. (2015). Bayesian estimation of a skew-student-t stochastic volatility model. *Methodology and Computing in Applied Probability*, 17(3):721–738.

Acar, E. F., Czado, C., and Lysy, M. (2019). Flexible dynamic vine copula models for multivariate time series data. *Econometrics and Statistics*, 12:181–197.

Allen, D., Singh, A., Powell, R., McAleer, M., and Taylor, J. (2012). The Volatility-Return Relationship: Insights from Linear and Non-Linear Quantile Regressions. School of Accounting. *Finance and Economics & FEMARC Working Paper Series*.

Almeida, C. and Czado, C. (2012). Efficient Bayesian inference for stochastic time-varying copula models. *Computational Statistics & Data Analysis*, 56(6):1511–1527.

Almeida, C., Czado, C., and Manner, H. (2016). Modeling high-dimensional time-varying dependence using dynamic D-vine models. *Applied Stochastic Models in Business and Industry*, 32(5):621–638.

Anderson, J. O., Thundiyil, J. G., and Stolbach, A. (2012). Clearing the air: a review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology*, 8(2):166–175.

Arellano-Valle, R. B. and Azzalini, A. (2013). The centred parameterization and related quantities of the skew-t distribution. *Journal of Multivariate Analysis*, 113:73–90.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Baele, L., Bekaert, G., and Inghelbrecht, K. (2010). The determinants of stock and bond return comovements. *The Review of Financial Studies*, 23(6):2374–2428.

Barthel, N., Geerdens, C., Czado, C., and Janssen, P. (2018). Dependence modeling for recurrent event times subject to right-censoring with D-vine copulas. *Biometrics*, 75:439–451.

Bates, D., Eddelbuettel, D., et al. (2013). Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software*, 52(5):1–24.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):245–268.

Bell, M. L. and Davis, D. L. (2001). Reassessment of the lethal London fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environmental health perspectives*, 109(suppl 3):389–394.

Bennett, J., Grout, R., Pébay, P., Roe, D., and Thompson, D. (2009). Numerically stable, single-pass, parallel statistics algorithms. In *Cluster Computing and Workshops, 2009. CLUSTER'09. IEEE International Conference on*, pages 1–8. IEEE.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Betancourt, M. and Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 79:30.

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.

Brechmann, E. C. and Czado, C. (2013). Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50. *Statistics & Risk Modeling*, 30(4):307–342.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to Time Series and Forecasting*, volume 2. Springer.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Carpenter, B., Hoffman, M. D., Brubaker, M., Lee, D., Li, P., and Betancourt, M. (2015). The Stan math library: Reverse-mode automatic differentiation in C++. *arXiv preprint arXiv:1509.07164*.

Chan, J. C. and Grant, A. L. (2016). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics*, 54:182–189.

Chen, S., Fricks, J., and Ferrari, M. J. (2012). Tracking measles infection through non-linear state space models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(1):117–134.

Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371.

Chipman, H. A., George, E. I., McCulloch, R. E., et al. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

Christoffersen, P. F. (2012). *Elements of Financial Risk Management*. Academic Press.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas*. Lecture Notes in Statistics. Springer.

De Vito, S., Fattoruso, G., Pardo, M., Tortorella, F., and Di Francia, G. (2012). Semi-supervised learning techniques in artificial olfaction: A novel approach to classification problems and drift counteraction. *IEEE Sensors Journal*, 12(11):3215–3224.

De Vito, S., Massera, E., Piga, M., Martinotto, L., and Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757.

De Vito, S., Piga, M., Martinotto, L., and Di Francia, G. (2009). CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic Bayesian regularization. *Sensors and Actuators B: Chemical*, 143(1):182–191.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, 195(2):216–222.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, 1:176–223.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332.

Fink, H., Klimova, Y., Czado, C., and Stöber, J. (2017). Regime switching vine copula models for global equity and volatility indices. *Econometrics*, 5(1):3.

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11(2):317–336.

Frühwirth-Schnatter, S. and Sögner, L. (2003). Bayesian estimation of the Heston stochastic volatility model. In *Operations Research Proceedings 2002*, pages 480–485. Springer.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Garthwaite, P. H., Fan, Y., and Sisson, S. A. (2016). Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process. *Communications in Statistics-Theory and Methods*, 45(17):5098–5111.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014a). *Bayesian Data Analysis*, volume 2. CRC press Boca Raton, FL.

Gelman, A., Hwang, J., and Vehtari, A. (2014b). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741.

Ghalanos, A. (2012). rmgarch: Multivariate GARCH models. *R package version 0.98*.

Gilks, W. R., Best, N., and Tan, K. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, pages 455–472.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Goel, A. and Mehra, A. (2019). Analyzing Contagion Effect in Markets During Financial Crisis Using Stochastic Autoregressive Canonical Vine Model. *Computational Economics*, 53(3):921–950.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Gruber, L. F. and Czado, C. (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis*, 10(4):937–963.

Gruber, L. F. and Czado, C. (2018). Bayesian model selection of regular vine copulas. *Bayesian Analysis*, 13(4):1107–1131.

Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Hafner, C. M. and Manner, H. (2012). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics*, 27(2):269–295.

Hahn, P. R., He, J., and Lopes, H. F. (2019). Efficient sampling for Gaussian linear regression with arbitrary priors. *Journal of Computational and Graphical Statistics*, 28(1):142–154.

Han, Y. (2005). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1):237–271.

Hartmann, M. and Ehlers, R. S. (2017). Bayesian inference for generalized extreme value distributions via Hamiltonian Monte Carlo. *Communications in Statistics-Simulation and Computation*, pages 1–18.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):297–318.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, pages 97–109.

Hoffman, M. D. and Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Hosszejni, D. and Kastner, G. (2019). Approaches Toward the Bayesian Estimation of the Stochastic Volatility Model with Leverage. *arXiv preprint arXiv:1901.11491*.

Ippoliti, L., Valentini, P., and Gamerman, D. (2012). Space–time modelling of coupled spatiotemporal environmental variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(2):175–200.

Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Johns, C. J. and Shumway, R. H. (2005). A non-linear and non-Gaussian state-space model for censored air pollution data. *Environmetrics: The official journal of the International Environmetrics Society*, 16(2):167–180.

Jondeau, E. (2016). Asymmetry in tail dependence in equity portfolios. *Computational Statistics & Data Analysis*, 100:351–368.

Jordan, A., Krüger, F., and Lerch, S. (2017). Evaluating probabilistic forecasts with scoringRules. *arXiv preprint arXiv:1709.04743*.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(5):1–30.

Kastner, G. (2019). Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics*, 210(1):98–115.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.

Kastner, G., Frühwirth-Schnatter, S., and Lopes, H. F. (2017). Efficient Bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, 26(4):905–917.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.

Killiches, M. and Czado, C. (2018). A D-vine copula-based model for repeated measurements extending linear mixed models with homogeneous correlation structure. *Biometrics*, 74(3):997–1005.

Kim, K.-H., Kabir, E., and Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, 74:136–143.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

Koop, G., Korobilis, D., et al. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends® in Econometrics*, 3(4):267–358.

Kreuzer, A. and Czado, C. (2019a). Bayesian inference for a single factor copula stochastic volatility model using Hamiltonian Monte Carlo. *arXiv preprint arXiv:1808.08624v2*.

Kreuzer, A. and Czado, C. (2019b). Bayesian inference for dynamic vine copulas in higher dimensions. *arXiv preprint arXiv:1911.00702v1*.

Kreuzer, A. and Czado, C. (2019c). Efficient Bayesian inference for nonlinear state space models with univariate autoregressive state equation. *arXiv preprint arXiv:1902.10412v3*.

Kreuzer, A., Dalla Valle, L., and Czado, C. (2019a). A Bayesian Non-linear State Space Copula Model to Predict Air Pollution in Beijing. *arXiv preprint arXiv:1903.08421v2*.

Kreuzer, A., Dalla Valle, L., and Czado, C. (2019b). Bayesian Multivariate Nonlinear State Space Copula Models. *arXiv preprint arXiv:1911.00448v1*.

Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

Krupskii, P. and Joe, H. (2015). Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., and Chen, S. X. (2015). Assessing Beijing's PM2. 5 pollution: severity, weather impact, APEC and winter heating. *Proc. R. Soc. A*, 471(2182):20150257.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.

Loaiza-Maya, R., Smith, M. S., and Maneesoonthorn, W. (2018). Time series copulas for heteroskedastic data. *Journal of Applied Econometrics*, 33(3):332–354.

Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, pages 41–67.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7):2372–2387.

McCulloch, R., Sparapani, R., Gramacy, R., Spanbauer, C., and Pratola, M. (2018). BART: Bayesian additive regression trees. *R package version*, 1.

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

Meyn, S. P. and Tweedie, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media.

Min, A. and Czado, C. (2011). Bayesian model selection for D-vine pair-copula constructions. *Canadian Journal of Statistics*, 39(2):239–258.

Möller, A., Spazzini, L., Kraus, D., Nagler, T., and Czado, C. (2018). Vine copula based post-processing of ensemble forecasts for temperature. *arXiv preprint arXiv:1811.02255*.

Morales-Nápoles, O. (2010). Counting vines. In *Dependence Modeling: Vine Copula Handbook*, pages 189–218. World Scientific.

Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical slice sampling. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 9:541–548.

Nagler, T., Bumann, C., and Czado, C. (2019). Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *Journal of Multivariate Analysis*, 172:180–192.

Nagler, T. and Vatter, T. (2018). rvinecopulib: High performance algorithms for vine copula modeling. *R package version 0.2*, 8(0).

Nakajima, J., Kasuya, M., and Watanabe, T. (2011). Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy. *Journal of the Japanese and International Economies*, 25(3):225–245.

Neal, R. M. (1998). Regression and classification using Gaussian process priors. *Bayesian Statistics*, 6:475.

Neal, R. M. et al. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.

Neal, R. M. et al. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media.

Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, 56(11):3659–3673.

Oh, D. H. and Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, 36(2):181–195.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Pakman, A. and Paninski, L. (2014). Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542.

Paninski, L., Ahmadian, Y., Ferreira, D. G., Koyama, S., Rad, K. R., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1-2):107–126.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2):87–94.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Pitt, M. and Shephard, N. (1999). Time varying covariances: a factor stochastic volatility approach. *Bayesian Statistics*, 6:547–570.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2008). coda: Output analysis and diagnostics for MCMC. *R package version 0.13-3, URL http://CRAN. R-project. org/package= coda*.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.

Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.

Robbins, H. and Monro, S. (1985). A stochastic approximation method. In *Herbert Robbins Selected Papers*, pages 102–109. Springer.

Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367.

Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):61.

Sahu, S. K. and Mardia, K. V. (2005). A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):223–244.

Schamberger, B., Gruber, L. F., and Czado, C. (2017). Bayesian Inference for Latent Factor Copulas and Application to Financial Risk Forecasting. *Econometrics*, 5(2):21.

Schepsmeier, U. and Stöber, J. (2014). Derivatives and Fisher information of bivariate copulas. *Statistical Papers*, 55(2):525–542.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., et al. (2018). Package 'VineCopula'. *R package version*, 2(5).

Scheuerer, M. and Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479.

Smith, M. S. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, 31(3):815–833.

Song, C., Wu, L., Xie, Y., He, J., Chen, X., Wang, T., Lin, Y., Jin, T., Wang, A., Liu, Y., et al. (2017). Air pollution in China: status and spatiotemporal variations. *Environmental Pollution*, 227:334–347.

Stoeber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions—limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.

Strickland, C. M., Martin, G. M., and Forbes, C. S. (2008). Parameterisation and efficient MCMC estimation of non-Gaussian state space models. *Computational Statistics & Data Analysis*, 52(6):2911–2930.

Stübinger, J., Mangold, B., and Krauss, C. (2018). Statistical arbitrage with vine copulas. *Quantitative Finance*, 18(11):1831–1849.

Sun, Y. and Wu, X. (2018). Leverage and Volatility Feedback Effects and Conditional Dependence Index: A Nonparametric Study. *Journal of Risk and Financial Management*, 11(2):29.

Tan, B. K., Panagiotelis, A., and Athanasopoulos, G. (2019). Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics*, 28(1):155–173.

Team, S. D. (2015). Stan Modeling Language: User's Guide and Reference Manual. *Version 2.12*.

Van den Brakel, J. and Roels, J. (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *The Annals of Applied Statistics*, pages 1105–1138.

Van Dyk, D. A. and Jiao, X. (2015). Metropolis-Hastings within partially collapsed Gibbs samplers. *Journal of Computational and Graphical Statistics*, 24(2):301–327.

Vatter, T. and Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147–167.

Vatter, T. and Nagler, T. (2018). Generalized additive models for pair-copula constructions. *Journal of Computational and Graphical Statistics*, 27(4):715–727.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Wainwright, M. J., Jordan, M. I., et al. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

West, M. and Harrison, J. (2006). *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media.

Wood, S. and Wood, M. S. (2015). Package 'mgcv'. *R package version*, 1:29.

Yu, J. (2002). Forecasting volatility in the New Zealand stock market. *Applied Financial Economics*, 12(3):193–202.

Yu, Y. and Meng, X.-L. (2011). To center or not to center: That is not the question—an Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570.

Appendix A Supplementary material to Chapter 3

A.1 Derivatives for HMC for the single factor copula model

For all considered copula families there is a one-to-one correspondence between the copula parameter and Kendall's τ. So there is a function $\theta_j = g(\tau_j)$ such that θ_j is the corresponding copula parameter (see Section 2.2.3). We use here that the derivatives of the copula density with respect to θ_j have already been derived by Schepsmeier and Stöber (2014). There is a one-to-one correspondence between θ_j, τ_j and δ_j, given by

$$\tau_j = (1 + \exp(-\delta_j))^{-1}, \qquad \theta_j = g(\tau_j) \tag{A.1}$$

The derivatives of the log posterior density with respect to the parameters $\delta_{j'}$ and $w_{t'}$ are given by

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\delta_{j'}}\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{w}|\mathrm{U}) &= \sum_{j=1}^{d}\sum_{t=1}^{T} \frac{\mathrm{d}}{\mathrm{d}\delta_{j'}} \ln(c_j(u_{tj}, v_t; \tau_j)) + \frac{\mathrm{d}}{\mathrm{d}\delta_{j'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})) \\
&= \sum_{t=1}^{T} \frac{\mathrm{d}}{\mathrm{d}\delta_{j'}} \ln(c_j(u_{tj'}, v_t; \tau_{j'})) + \frac{\mathrm{d}}{\mathrm{d}\delta_{j'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})) \\
&= \sum_{t=1}^{T} \frac{\mathrm{d}}{\mathrm{d}\theta_{j'}} \ln(c_j(u_{tj'}, v_t; \tau_{j'})) \frac{\mathrm{d}\theta_{j'}}{\mathrm{d}\delta_{j'}} + \frac{\mathrm{d}}{\mathrm{d}\delta_{j'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}w_{t'}}\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{w}|\mathrm{U}) &= \sum_{j=1}^{d}\sum_{t=1}^{T} \frac{\mathrm{d}}{\mathrm{d}w_{t'}} \ln(c_j(u_{tj}, v_t; \tau_j)) + \frac{\mathrm{d}}{\mathrm{d}w_{t'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})) \\
&= \sum_{j=1}^{d} \frac{\mathrm{d}}{\mathrm{d}w_{t'}} \ln(c_j(u_{t'j}, v_{t'}; \tau_j)) + \frac{\mathrm{d}}{\mathrm{d}w_{t'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})) \\
&= \sum_{j=1}^{d} \frac{\mathrm{d}}{\mathrm{d}v_{t'}} \ln(c_j(u_{t'j}, v_{t'}; \tau_j)) \frac{\mathrm{d}v_{t'}}{\mathrm{d}w_{t'}} + \frac{\mathrm{d}}{\mathrm{d}w_{t'}} \ln(\pi(\boldsymbol{\delta}, \boldsymbol{w})).
\end{aligned}$$

The components of the derivative of the log posterior density are derived in the following.

Derivatives of the log prior density

The derivative of the log prior density π_u is given by

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln(\pi_u(x)) = \frac{\mathrm{d}}{\mathrm{d}x} - 2\ln(1+\exp(-x)) - x = 2(1+\exp(x))^{-1} - 1.$$

Derivatives of the parameter transformation

We consider derivatives of the parameter transformation, i.e. $\frac{\mathrm{d}\theta_{j'}}{\mathrm{d}\delta_{j'}}$ and $\frac{\mathrm{d}v_{t'}}{\mathrm{d}w_{t'}}$. In this part we suppress the indices j' and t'. We have that

$$v = (1+\exp(-w))^{-1},$$

and the derivative is given by

$$\frac{\mathrm{d}v}{\mathrm{d}w} = (1+\exp(-w))^{-2}\exp(-w).$$

Now we address the derivative $\frac{\mathrm{d}\theta}{\mathrm{d}\delta}$. The parameter δ was chosen to be the logit transform of the corresponding Kendall's τ and so τ can be written as

$$\tau = (1+\exp(-\delta))^{-1},$$

with corresponding derivative

$$\frac{\mathrm{d}\tau}{\mathrm{d}\delta} = (1+\exp(-\delta))^{-2}\exp(-\delta).$$

The copula parameter θ is a function of Kendall's τ and dependent on the copula family considered we obtain the following derivatives.

- Gauss and Student t copula

$$\begin{aligned}
\theta &= \sin\left(\frac{1}{2}\pi\tau\right) \\
\frac{\mathrm{d}\theta}{\mathrm{d}\delta} &= \frac{\mathrm{d}\theta}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}\delta} \\
&= \frac{1}{2}\pi\cos\left(\frac{1}{2}\pi\tau\right)\frac{\mathrm{d}\tau}{\mathrm{d}\delta} \\
&= \frac{1}{2}\pi\cos\left(\frac{1}{2}\pi(1+\exp(-\delta))^{-1}\right)(1+\exp(-\delta))^{-2}\exp(-\delta)
\end{aligned}$$

- Clayton copula

$$\begin{aligned}
\theta &= \frac{2\tau}{1-\tau} \\
&= \frac{2}{\tau^{-1}-1} \\
&= \frac{2}{1+\exp(-\delta)-1} \\
&= 2\exp(\delta) \\
\frac{\mathrm{d}\theta}{\mathrm{d}\delta} &= 2\exp(\delta)
\end{aligned}$$

- Gumbel copula

$$\begin{aligned}
\theta &= (1-\tau)^{-1} \\
\frac{\mathrm{d}\theta}{\mathrm{d}\delta} &= \frac{\mathrm{d}\theta}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}\delta} \\
&= (1-\tau)^{-2}\frac{\mathrm{d}\tau}{\mathrm{d}\delta} \\
&= \{1-[1+\exp(-\delta)]^{-1}\}^{-2}[1+\exp(-\delta)]^{-2}\exp(-\delta) \\
&= [1+\exp(-\delta)-1]^{-2}\exp(-\delta) \\
&= \exp(-\delta)^{-2}\exp(-\delta) \\
&= \exp(\delta)
\end{aligned}$$

Derivatives of log copula densities

For all considered copula families Schepsmeier and Stöber (2014) calculate the derivatives of the copula density with respect to the copula parameter θ_j and with respect to the argument v_t. Based on their results the derivatives of the log copula density are easily derived. The derivatives are also implemented in the `R` package `VineCopula` by Schepsmeier et al. (2018).

A.2 Prior densities for transformed parameters

The prior densities for μ_j, ϕ_j and σ_j^2 in (3.12) imply the following prior densities for μ_j, ξ_j and ψ_j.

- We have that $\mu_j \sim N(0, \sigma_\mu^2)$. So the prior density for μ_j is up to a constant given by

$$\pi(x) \propto \exp\left(-\frac{x^2}{2\sigma_\mu^2}\right).$$

- We have that $\frac{\phi_j+1}{2} \sim Beta(a,b)$. So the density of ϕ_j is given by

$$f_\phi(x) = f_{Beta}\left(\frac{x+1}{2}\right)\frac{1}{2}.$$

  This implies that the prior density of ξ_j is

$$\begin{aligned}
\pi(x) &= f_\phi(F_Z^{-1}(x))\left|\frac{\mathrm{d}}{\mathrm{d}x}F_Z^{-1}(x)\right| \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{F_Z^{-1}(x)+1}{2}\right)^{a-1}\left(1-\frac{F_Z^{-1}(x)+1}{2}\right)^{b-1}\frac{1}{2}\left(1-(F_Z^{-1}(x))^2\right).
\end{aligned}$$

- We have that $\sigma_j^2 \sim \chi_1^2$, i.e.

$$f_\sigma(x) = 2x f_{\chi_1^2}(x^2).$$

So the prior density for ψ_j is given by

$$\begin{aligned}\pi(x) &= f_\sigma(\exp(x))\exp(x)\\ &= 2\exp(x)\exp(x)\frac{1}{\sqrt{2}\Gamma(\frac{1}{2})}\exp(-x)\exp\left(-\frac{\exp(2x)}{2}\right)\\ &= 2\frac{1}{\sqrt{2}\Gamma(\frac{1}{2})}\exp(x)\exp\left(-\frac{\exp(2x)}{2}\right).\end{aligned}$$

A.3 Derivatives for HMC for the stochastic volatility model

We need to calculate derivatives of the function

$$\begin{aligned}\mathcal{L}(\mu_j,\xi_j,\psi_j,\tilde{\boldsymbol{s}}_{\boldsymbol{\cdot j}}|Z,\boldsymbol{\tau},\boldsymbol{v},\boldsymbol{m}) =&\\ &\sum_{t=1}^{T}\left[\ln\left(c_j^{m_j}\left(\Phi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right),v_t;\tau_j\right)\right)+\ln\left(\varphi\left(\frac{z_{tj}}{\exp\left(\frac{s_{tj}}{2}\right)}\right)\right)-\frac{s_{tj}}{2}\right]\\ &+\ln(\pi(\mu_j,\xi_j,\psi_j,\tilde{\boldsymbol{s}}_{\boldsymbol{\cdot j}}))+const,\end{aligned}$$

where $const\in\mathbb{R}$ is a constant.

To subset a vector $\boldsymbol{x}$, we use the notation $\boldsymbol{x_{k:n}}=(x_k,\dots,x_n)$. To further shorten notation, we omit the index j in the following and consider the function

$$\begin{aligned}\mathcal{L}_2(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}) =&\sum_{t=1}^{T}\left[\ln\left(c^m\left(\Phi\left(\frac{z_t}{\exp\left(\frac{s_t}{2}\right)}\right),v_t;\tau\right)\right)+\ln\left(\varphi\left(\frac{z_t}{\exp\left(\frac{s_t}{2}\right)}\right)\right)-\frac{s_t}{2}\right]\\ &+\ln(\pi(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})).\end{aligned}$$

We define

$$\Omega(\boldsymbol{s_{1:T}})=\sum_{t=1}^{T}\left(\ln\left(c^m\left(\Phi\left(\frac{z_t}{\exp\left(\frac{s_t}{2}\right)}\right),v_t;\tau\right)\right)+\ln\left(\varphi\left(\frac{z_t}{\exp\left(\frac{s_t}{2}\right)}\right)\right)-\frac{s_t}{2}\right),$$

and the derivatives can be expressed as

- $\frac{\mathrm{d}}{\mathrm{d}\mu}\mathcal{L}_2(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})=\frac{\mathrm{d}\Omega(\boldsymbol{s_{1:T}})}{\mathrm{d}\boldsymbol{s_{1:T}}}\frac{\mathrm{d}\boldsymbol{s_{1:T}}}{\mathrm{d}\mu}+\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\pi(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}))$
- $\frac{\mathrm{d}}{\mathrm{d}\xi}\mathcal{L}_2(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})=\frac{\mathrm{d}\Omega(\boldsymbol{s_{1:T}})}{\mathrm{d}\boldsymbol{s_{1:T}}}\frac{\mathrm{d}\boldsymbol{s_{1:T}}}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}\xi}+\frac{\mathrm{d}}{\mathrm{d}\xi}\ln(\pi(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}))$
- $\frac{\mathrm{d}}{\mathrm{d}\psi}\mathcal{L}_2(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})=\frac{\mathrm{d}\Omega(\boldsymbol{s_{1:T}})}{\mathrm{d}\boldsymbol{s_{1:T}}}\frac{\mathrm{d}\boldsymbol{s_{1:T}}}{\mathrm{d}\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\psi}+\frac{\mathrm{d}}{\mathrm{d}\psi}\ln(\pi(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}))$
- $\frac{\mathrm{d}}{\mathrm{d}\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}}\mathcal{L}_2(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})=\frac{\mathrm{d}\Omega(\boldsymbol{s_{1:T}})}{\mathrm{d}\boldsymbol{s_{0:T}}}J+\frac{\mathrm{d}}{\mathrm{d}\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}}\ln(\pi(\mu,\xi,\psi,\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})),$

where $J\in\mathbb{R}^{(T+1)\times(T+1)}$ denotes the corresponding Jacobian matrix, i.e. $J_{tj}=\frac{\mathrm{d}s_t}{\mathrm{d}\tilde{s}_j}$. The derivatives are calculated in the following.

- $\frac{\mathrm{d}}{\mathrm{d}s_i}\Omega(\boldsymbol{s_{1:T}}) = \frac{\mathrm{d}}{\mathrm{d}x}\ln(c^m(x, v_i; \tau))\Big|_{x=\Phi\left(\frac{z_i}{\exp\left(\frac{s_i}{2}\right)}\right)} \varphi\left(\frac{z_i}{\exp\left(\frac{s_i}{2}\right)}\right)\frac{z_i}{\exp\left(\frac{s_i}{2}\right)}(-\frac{1}{2}) + \frac{z_i^2}{2\exp(s_i)} - \frac{1}{2}$ for $i = 1, \ldots, T$

- We have that $s_0 = \frac{\tilde{s}_0\sigma}{\sqrt{1-\phi^2}} + \mu, \quad s_t = \tilde{s}_t\sigma + \mu + \phi(s_{t-1} - \mu), t = 1, \ldots, T$ and obtain

$$\begin{array}{ll}
\frac{\mathrm{d}s_0}{\mathrm{d}\mu} = 1 & \frac{\mathrm{d}s_t}{\mathrm{d}\mu} = 1 - \phi + \phi\frac{\mathrm{d}}{\mathrm{d}\mu}s_{t-1}, t = 1, \ldots, T \\
\frac{\mathrm{d}s_0}{\mathrm{d}\phi} = \tilde{s}_0\sigma(1-\phi^2)^{-\frac{3}{2}}\phi & \frac{\mathrm{d}s_t}{\mathrm{d}\phi} = s_{t-1} - \mu + \phi\frac{\mathrm{d}}{\mathrm{d}\phi}s_{t-1}, t = 1, \ldots, T \\
\frac{\mathrm{d}s_0}{\mathrm{d}\sigma} = \frac{\tilde{s}_0}{\sqrt{1-\phi^2}} & \frac{\mathrm{d}s_t}{\mathrm{d}\sigma} = \tilde{s}_t + \phi\frac{\mathrm{d}}{\mathrm{d}\sigma}s_{t-1}, t = 1, \ldots, T \\
\frac{\mathrm{d}s_t}{\mathrm{d}\tilde{s}_0} = \phi^t\frac{\sigma}{\sqrt{1-\phi^2}}, t = 0, \ldots, T & \frac{\mathrm{d}s_t}{\mathrm{d}\tilde{s}_j} = \phi^{t-j}\sigma\mathbb{1}_{t\geq j}, t = 0, \ldots, T, j = 1, \ldots T
\end{array}$$

- $\frac{\mathrm{d}\phi}{\mathrm{d}\xi} = 1 - F^{-1}(\xi)^2, \quad \frac{\mathrm{d}\sigma}{\mathrm{d}\psi} = \exp(\psi)$

- $\frac{\mathrm{d}}{\mathrm{d}\mu}\ln(\pi(\mu, \xi, \psi, \tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})) = -\frac{\mu}{\sigma_\mu^2}$

- $\frac{\mathrm{d}}{\mathrm{d}\xi}\ln(\pi(\mu, \xi, \psi, \tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})) = (a-1)\frac{(1-F_Z^{-1}(\xi)^2)}{(F_Z^{-1}(\xi)+1)} - (b-1)(1+F_Z^{-1}(\xi)) - 2F_Z^{-1}(\xi)$ where $a = 5$ and $b = 1.5$ are the parameters of the beta distribution.

- $\frac{\mathrm{d}}{\mathrm{d}\psi}\ln(\pi(\mu, \xi, \psi, \tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})) = 1 - \exp(2\psi)$

- $\frac{\mathrm{d}}{\mathrm{d}\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}}\ln(\pi(\mu, \xi, \psi, \tilde{\boldsymbol{s}}_{\boldsymbol{0:T}})) = -\tilde{\boldsymbol{s}}_{\boldsymbol{0:T}}$

A.4 Results of the simulation study for $d = 10$

m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
94%	89%	92%	93%	78%	91%	88%	91%	91%	81%

Table A.1: Proportion of how often the correct copula family was selected. The selected copula family is the posterior mode estimate of m_j for $j = 1, \ldots, 10$ (Scenario 2).

Scenario 2	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
MSE	0.0033	0.0031	0.0050	0.0097	0.0560	0.0031	0.0036	0.0068	0.0121	0.0629
C.I. 90%	0.88	0.90	0.93	0.90	0.92	0.88	0.83	0.87	0.90	0.86
C.I. 95%	0.95	0.97	0.97	0.94	0.98	0.96	0.93	0.92	0.94	0.89
ESS	779	499	616	695	465	781	548	555	659	457
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
MSE	0.0462	0.0239	0.0038	0.0009	0.0003	0.0321	0.0340	0.0035	0.0010	0.0003
C.I. 90%	0.98	0.95	0.83	0.90	0.81	0.98	0.96	0.90	0.92	0.82
C.I. 95%	0.98	0.98	0.89	0.96	0.92	1.00	0.99	0.92	0.97	0.92
ESS	480	385	362	402	319	478	412	369	393	325
	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9	σ_{10}
MSE	0.0082	0.0047	0.0028	0.0018	0.0026	0.0068	0.0053	0.0028	0.0018	0.0027
C.I. 90%	0.95	0.97	0.85	0.89	0.81	0.96	0.98	0.91	0.93	0.84
C.I. 95%	0.97	0.99	0.95	0.96	0.89	0.98	0.99	0.96	0.96	0.90
ESS	283	297	298	300	242	288	317	295	295	236
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}
MSE	0.0112	0.0195	0.0305	0.0440	0.0590	0.0112	0.0195	0.0309	0.0435	0.0590
C.I. 90%	0.78	0.85	0.77	0.74	0.78	0.79	0.79	0.81	0.77	0.76
C.I. 95%	0.83	0.85	0.84	0.80	0.82	0.82	0.87	0.84	0.81	0.80
ESS	520	474	279	261	181	498	459	275	253	164
	$s_{300,1}$	$s_{300,2}$	$s_{300,3}$	$s_{300,4}$	$s_{300,5}$	$s_{300,6}$	$s_{300,7}$	$s_{300,8}$	$s_{300,9}$	$s_{300,10}$
MSE	0.0815	0.0915	0.1839	0.1536	0.2546	0.0877	0.0941	0.1639	0.1711	0.2437
C.I. 90%	0.86	0.84	0.92	0.91	0.87	0.85	0.83	0.92	0.91	0.90
C.I. 95%	0.91	0.93	0.94	0.95	0.95	0.94	0.92	0.96	0.96	0.94
ESS	1073	1074	1043	1007	620	1087	1100	1010	999	628
	v_{100}	v_{200}	v_{500}	v_{800}	v_{900}					
MSE	0.0269	0.0263	0.0190	0.0157	0.0146					
C.I. 90%	0.82	0.91	0.82	0.90	0.85					
C.I. 95%	0.88	0.93	0.85	0.95	0.90					
ESS	393	406	395	405	406					

Table A.2: MSE estimated using the posterior mode, observed coverage probability of the credible intervals (C.I.) and effective sample sizes (ESS) calculated from 2000 posterior draws for selected parameters (Scenario 2).

Appendix B Supplementary material to Chapter 4

In this chapter we make use of the following notation to subset vectors and matrices: For sets of indices A and B we set $\boldsymbol{x}_A = (x_i)_{i\in A}$ for a vector $\boldsymbol{x}$ and $X_{A;B} = (x_{ij})_{i\in A, j\in B}$ for a matrix X. We use a capital letter to refer to a matrix and small letters to refer to its components. The set $\{n, \dots, k\}$ of integers will be abbreviated by $n : k$.

B.1 Details on the sampling procedure

Sampling of the latent states in the sufficient augmentation

Here we derive $\boldsymbol{\mu}_{\boldsymbol{B_i}|}$ and $\Sigma_{B_i|}$. By the conditional independence assumptions of the AR(1) process and the way we defined the blocks $B_1, \dots, B_m$, we obtain

$$\begin{aligned} f(\boldsymbol{s}_{\boldsymbol{B_i}}|s_0, \boldsymbol{s}_{-\boldsymbol{B_i}}, \mu, \phi, \sigma) &= f(\boldsymbol{s}_{\boldsymbol{B_i}}|s_{a_i-1}, s_{b_i+1}, \mu, \phi, \sigma), \text{ for } i = 1, \dots, m-1, \text{ and} \\ f(\boldsymbol{s}_{\boldsymbol{B_m}}|s_0, \boldsymbol{s}_{-\boldsymbol{B_m}}, \mu, \phi, \sigma) &= f(\boldsymbol{s}_{\boldsymbol{B_m}}|s_{a_m-1}, \mu, \phi, \sigma). \end{aligned}$$

Conditional on μ, ϕ and σ, the vector $(s_{a_i-1}, \boldsymbol{s}_{\boldsymbol{B_i}}, s_{b_i+1})$ is multivariate normal distributed with mean vector $\boldsymbol{\mu}^{\boldsymbol{AR}}_{\boldsymbol{(a_i-1, B_i, b_i+1)}} \in \mathbb{R}^{c_i+2}$ and covariance matrix $\Sigma^{AR}_{(a_i-1,B_i,b_i+1);(a_i-1,B_i,b_i+1)} \in \mathbb{R}^{(c_i+2)\times(c_i+2)}$, where c_i is the cardinality of B_i. Thus the vector $\boldsymbol{s}_{\boldsymbol{B_i}}|s_{a_i-1}, s_{b_i+1}, \mu, \phi, \sigma$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{B_i}|}$ and covariance matrix $\Sigma_{B_i|}$ given by

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{B_i}|} &= \boldsymbol{\mu}^{\boldsymbol{AR}}_{\boldsymbol{B_i}} + \Sigma^{AR}_{B_i;(a_i-1,b_i+1)} \frac{1-\phi^2}{(1-\phi^{2(c_i+1)})\sigma^2} \begin{pmatrix} 1 & -\phi^{c_i+1} \\ -\phi^{c_i+1} & 1 \end{pmatrix} \begin{pmatrix} s_{a_i-1} - \mu \\ s_{b_i+1} - \mu \end{pmatrix}, \\ \Sigma_{B_i|} &= \Sigma^{AR}_{B_i;B_i} - \Sigma^{AR}_{B_i;(a_i-1,b_i+1)} \frac{1-\phi^2}{(1-\phi^{2(c_i+1)})\sigma^2} \begin{pmatrix} 1 & -\phi^{c_i+1} \\ -\phi^{c_i+1} & 1 \end{pmatrix} \Sigma^{AR}_{(a_i-1,b_i+1);B_i}. \end{aligned} \tag{B.1}$$

The vector $\boldsymbol{s}_{\boldsymbol{B_m}}|s_{a_m-1}, \mu, \phi, \sigma$ corresponding to the last block is multivariate normal distributed with mean vector $\boldsymbol{\mu}_{\boldsymbol{B_m}|}$ and covariance matrix $\Sigma_{B_m|}$ obtained as

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{B_m}|} &= \boldsymbol{\mu}^{\boldsymbol{AR}}_{\boldsymbol{B_m}} + \Sigma^{AR}_{B_m;a_m-1} \frac{1-\phi^2}{\sigma^2}(s_{a_m-1} - \mu), \\ \Sigma_{B_m|} &= \Sigma^{AR}_{B_m;B_m} - \Sigma^{AR}_{B_m;a_m-1} \frac{1-\phi^2}{\sigma^2} \Sigma^{AR}_{a_m-1;B_m}. \end{aligned}$$

We need to sample from $N_{c_i}(\mathbf{0}, \Sigma_{B_i|})$ several times during elliptical slice sampling. Instead of working with the $c_i \times c_i$ covariance matrix $\Sigma_{B_i|}$ we can more efficiently sample from the c_i-dimensional normal distribution by using the conditional independence assumptions of the AR(1) process. It holds that

$$\begin{aligned} f(\boldsymbol{s_{B_i}}|s_{a_i-1}, s_{b_i+1}, \mu, \phi, \sigma) &= \prod_{t=0}^{c_i-1} f(s_{a_i+t}|\boldsymbol{s_{a_i-1:a_i+t-1}}, s_{b_i+1}, \mu, \phi, \sigma) \\ &= \prod_{t=0}^{c_i-1} f(s_{a_i+t}|s_{a_i+t-1}, s_{b_i+1}, \mu, \phi, \sigma), \end{aligned}$$

where $f(s_{a_i+t}|s_{a_i+t-1}, s_{b_i+1}, \mu, \phi, \sigma)$ is the univariate normal density with mean

$$\mu_{a_i+t|a_i+t-1,b_i+1} = \mu + \frac{1}{1-\phi^{2(c_i+1-t))}} \left((\phi - \phi^{2c_i+1-2t})(s_{a_i+t-1} - \mu) + (\phi^{c_i-t} - \phi^{c_i+2-t}(s_{b_i+1} - \mu)) \right),$$

and variance

$$\sigma^2_{a_i+t|a_i+t-1,b_i+1} = \frac{\sigma^2}{1-\phi^2} \left(1 - \frac{1}{1-\phi^{2(c_i+1-t)}} \left(\phi^2 - 2\phi^{2(c_i-t+1)} + \phi^{2(c_i-t)} \right) \right).$$

So we can sample $\boldsymbol{s_{B_i}} = (s_{a_i+t})_{t=0,\dots,c_i-1}$ conditioned on $s_{a_i-1}, s_{b_i+1}, \mu, \phi, \sigma$ recursively by

$$s_{a_i+t} \sim N(\mu_{a_i+t|a_i+t-1,b_i+1}, \sigma^2_{a_i+t|a_i+t-1,b_i+1}),$$

for $t = 0, \dots, c_i - 1$ and then $\boldsymbol{s_{B_i}} - \boldsymbol{\mu_{B_i|}}$ is a sample from $N_{c_i}(\mathbf{0}, \Sigma_{B_i|})$.

Sampling of the constant parameters in the ancillary augmentation

To sample μ, ϕ and σ in (AA), we deploy an adaptive random walk Metropolis-Hastings scheme as suggested by Garthwaite et al. (2016). For sampling, it is convenient to move to unconstrained parameter spaces which is achieved by the following transformations

$$\psi = \ln(\sigma), \quad \xi = F_Z(\phi).$$

Here $F_Z(x) = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right)$ is Fisher's Z transformation. The from (4.5) implied log prior densities for ξ and ψ are given by

$$\begin{aligned} \ln(\pi(\xi)) &= (5-1)\ln\left(F_Z^{-1}(\xi) + 1\right) + (1.5-1)\ln\left(1 - F_Z^{-1}(\xi)\right) + \ln\left(1 - (F_Z^{-1}(\xi))^2\right) + c_1 \\ \ln(\pi(\psi)) &= \psi - \frac{1}{2}\exp(2\psi) + c_2, \end{aligned},$$

where $c_1 \in \mathbb{R}$ and $c_2 \in \mathbb{R}$ are constants. The log posterior density in (AA) is obtained as

$$\begin{aligned} lp_{(AA)}(\mu,\xi,\psi,s_0,\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}|\mathrm{Y}) = &\sum_{t=1}^{T} \ln\left(f(\boldsymbol{y_t}|s_t(\tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}, \mu, F_Z^{-1}(\xi), \exp(\psi)))\right) - \frac{1}{2}\sum_{t=1}^{T} \tilde{s}_t^2 \\ &+ \ln\left(\varphi\left(s_0|\mu, \frac{\exp(\psi)^2}{1 - F_Z^{-1}(\xi)^2}\right)\right) + \ln(\pi(\mu)) + \ln(\pi(\xi)) + \ln(\pi(\psi)) + c_3, \end{aligned}$$

where $c_3 \in \mathbb{R}$ is a constant. We sample (μ, ϕ, σ) in two blocks, one block for μ and one block for (ϕ, σ).

Update for μ

To sample the mean parameter μ from its full conditional, we propose a new state μ_{prop} in the r-th iteration of the MCMC procedure by

$$\mu_{prop} \sim N(\mu_{cur}, \sigma^{r-1}_{MH,\mu}),$$

where μ_{cur} is the current value for μ. The proposal μ_{prop} is accepted with probability

$$R = \exp\Big(lp_{(AA)}(\mu_{prop}, \xi, \psi, s_0, \tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}|\mathrm{Y}) - lp_{(AA)}(\mu_{cur}, \xi, \psi, s_0, \tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}|\mathrm{Y})\Big)$$

and the scaling parameter $\sigma^r_{MH,\mu}$ is updated according to Garthwaite et al. (2016) by

$$\ln\Big(\sigma^r_{MH,\mu}\Big) = \ln\Big(\sigma^{r-1}_{MH,\mu}\Big) + 4.058\frac{(R - 0.44)}{r-1}.$$

The scaling parameter is increased, if the acceptance probability is larger than 0.44 and decreased if the acceptance probability is smaller than 0.44. We target an average acceptance probability of 0.44, as recommended by Roberts et al. (2001) for univariate random walk Metropolis-Hastings. The constant 4.058 controls the step size and is chosen as suggested by Garthwaite et al. (2016).

Joint update for ϕ and σ

In the r-th iteration, a two dimensional proposal $(\xi_{prop}, \psi_{prop})$ for (ξ, ψ) is obtained by

$$(\xi_{prop}, \psi_{prop})^\top \sim N_2((\xi_{cur}, \psi_{cur})^\top, \Sigma^{r-1}_{MH,\xi,\psi}),$$

where (ξ_{cur}, ψ_{cur}) are the current values. The proposal is accepted with probability

$$R = \exp\Big(lp_{(AA)}(\mu, \xi_{prop}, \psi_{prop}, s_0, \tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}|\mathrm{Y}) - lp_{(AA)}(\mu, \xi_{cur}, \psi_{cur}, s_0, \tilde{\boldsymbol{s}}_{\boldsymbol{1:T}}|\mathrm{Y})\Big).$$

For adapting the covariance matrix, we follow a suggestion of Garthwaite et al. (2016). Let I_n denote the n-dimensional identity matrix. We set $\Sigma^r_{MH,\xi,s} = I_2$ if $r < 100$ and

$$\Sigma^r_{MH,\xi,s} = (\sigma^r_{MH,\xi,s})^2 \left(\hat{\Sigma}^r + \frac{(\sigma^r_{MH,\xi,s})^2}{r} I_2\right) \qquad \text{if } r \geq 100.$$

Here $\hat{\Sigma}^r$ is the empirical covariance matrix of $(\xi^i, \psi^i)_{i=1,\dots r}$, the first r samples for (ξ, ψ), and

$$\ln\Big(\sigma^r_{MH,\xi,s}\Big) = \ln\Big(\sigma^{r-1}_{MH,\xi,s}\Big) + 6.534\frac{(R - 0.234)}{r-1}.$$

The matrix $\hat{\Sigma}^r + \frac{(\sigma^r_{MH,\xi,\psi})^2}{r} I_2$ is a positive definite estimate of the covariance matrix. This covariance estimate is scaled by $(\sigma^r_{MH,\xi,\psi})^2$ to obtain the covariance matrix for the

proposal in the next iteration. The scaling $(\sigma^r_{MH,\xi,\psi})^2$ is tuned to achieve an average acceptance probability of 0.234 as suggested by Roberts et al. (1997) for multivariate random walk Metropolis-Hastings. To reduce computational cost, the empirical covariance matrix $\hat{\Sigma}^r$ can be updated in every step by the following recursion (see e.g. Bennett et al. (2009))

$$\hat{\Sigma}^r = \frac{r-2}{r-1}\hat{\Sigma}^{r-1} + \frac{1}{r}((\xi^r, \psi^r)^\top - \hat{\boldsymbol{\mu}}^{r-1})((\xi^r, \psi^r)^\top - \hat{\boldsymbol{\mu}}^{r-1})^\top,$$

where $\hat{\boldsymbol{\mu}}^{r-1}$ is the sample mean of $(\xi^i, \psi^i)^\top_{i=1,\dots,r-1}$. We also update the sample mean recursively by

$$\hat{\boldsymbol{\mu}}^r = \frac{1}{r}((r-1)\hat{\boldsymbol{\mu}}^{r-1} + (\xi^r, \psi^r)^\top).$$

We have seen that the adaptions for the μ and the (ϕ, σ) updates tend to be very small after burn-in and therefore we only adapt during the burn-in period. This also ensures a correct sampling procedure without the need to verify the validity of an adaptive MCMC scheme.

B.2 Additional material for the bivariate dynamic mixture copula (Section 4.4)

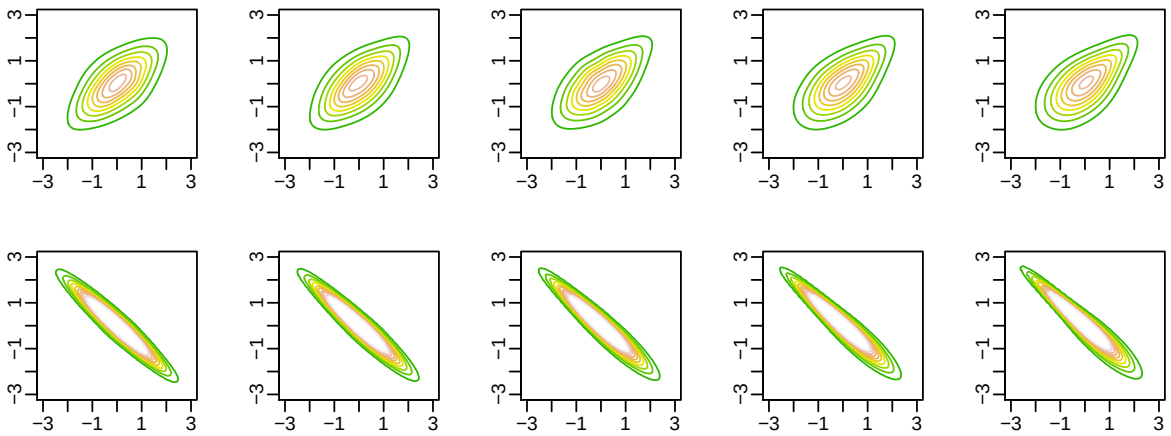

Figure B.1: Normalized contour plots (see Czado (2019), Chapter 3) of the bivariate density of the mixture copula model in (4.10) with $\tau = 0.4$ (top row), $\tau = -0.8$ (bottom row), $\nu = 5$ and $p = 1, 0.75, 0.5, 0.25, 0$ (from left to right).

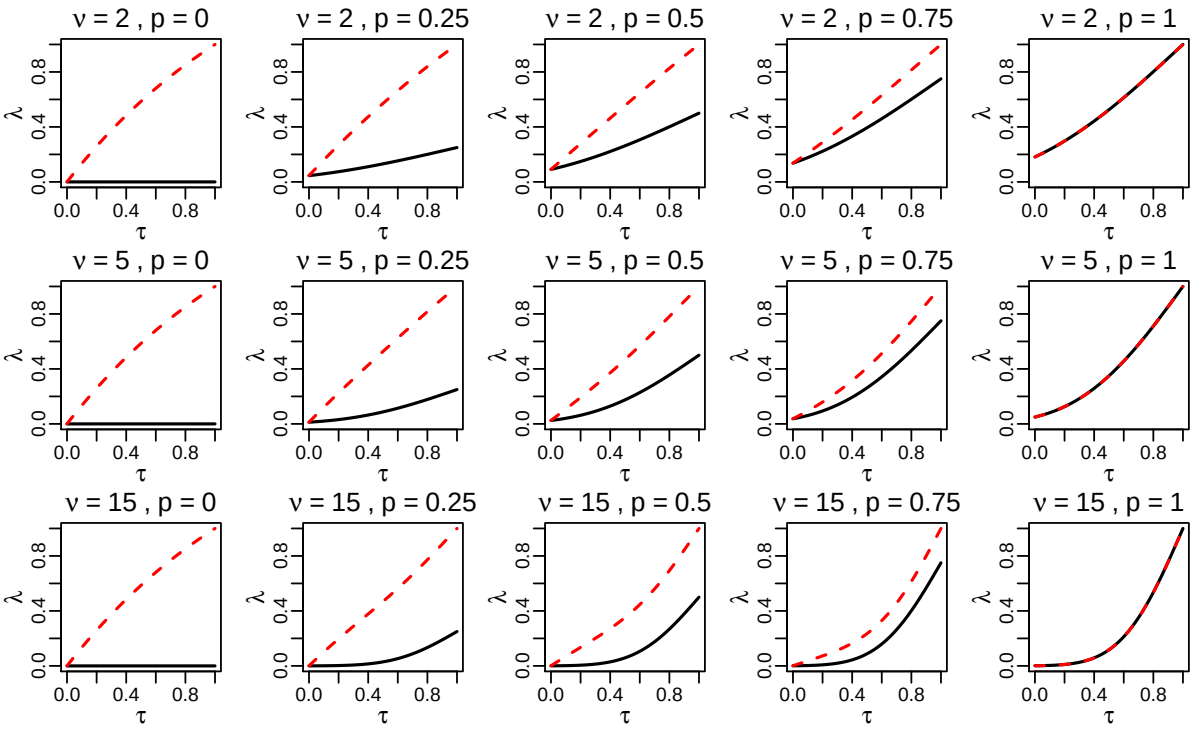

Figure B.2: Upper (red, dashed) and lower (black) tail dependence coefficient of the mixture copula defined in (4.10) plotted against Kendall's τ for different values of ν and p.

B.3 Further results for the application in Section 4.4

Daily log returns

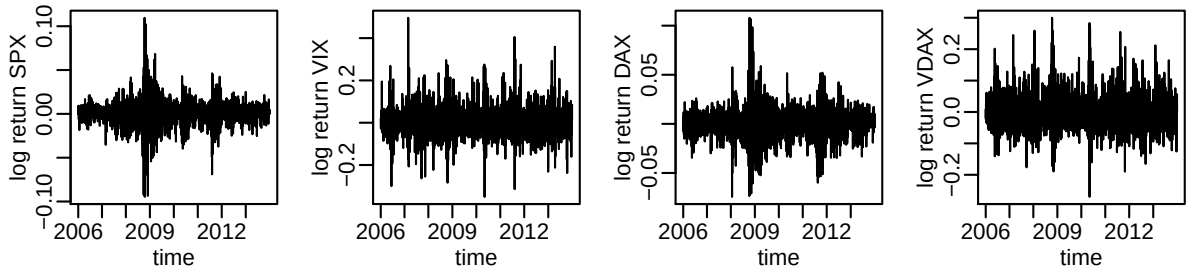

Figure B.3: Daily log returns of the four indices SPX, VIX, DAX, VDAX from 2006 to 2013 plotted against time.

Posterior statistics

	mode	5% quantile	95% quantile	effective sample size
		SPX		
μ	-9.32	-9.94	-8.70	13896.29
ϕ	0.99	0.98	1.00	565.10
σ	0.15	0.13	0.19	208.31
α	-0.51	-0.80	-0.22	4293.41
df	6.84	5.46	10.40	1821.93
		VIX		
μ	-5.65	-5.80	-5.50	3586.38
ϕ	0.90	0.84	0.93	362.48
σ	0.36	0.28	0.48	311.76
α	1.33	0.97	1.73	1376.24
df	9.30	6.50	15.14	1251.32
		DAX		
μ	-8.89	-9.30	-8.50	19573.46
ϕ	0.99	0.97	0.99	598.27
σ	0.15	0.12	0.19	249.07
α	-0.48	-0.80	-0.06	5662.26
df	9.74	7.31	15.11	2483.44
		VDAX		
μ	-6.06	-6.21	-5.89	8086.52
ϕ	0.96	0.92	0.97	416.02
σ	0.18	0.13	0.24	293.01
α	0.96	0.66	1.27	3305.73
df	8.35	6.44	12.88	1386.65

Table B.1: Estimated posterior modes, posterior quantiles and effective sample sizes for parameters of the univariate skew Student t stochastic volatility models for the four indices SPX, VIX, DAX, VDAX. Posterior mode estimates are obtained from univariate kernel density estimates.

	mode	5% quantile	95% quantile	effective sample size
		(SPX,VIX)		
μ	-0.74	-0.77	-0.71	1862.77
ϕ	0.94	0.85	0.97	306.33
σ	0.05	0.03	0.08	215.92
p	0.29	0.13	0.44	1436.78
ν	9.03	5.29	41.14	1039.38
		(DAX,VDAX)		
μ	-0.81	-0.84	-0.78	1785.30
ϕ	0.86	0.73	0.92	285.29
σ	0.10	0.06	0.13	207.29
p	0.66	0.50	0.81	1406.11
ν	8.30	5.91	34.32	756.50

Table B.2: Estimated posterior modes, posterior quantiles and effective sample sizes for parameters of the dynamic mixture copula models for the pairs (SPX,VIX) and (DAX,VDAX). Posterior mode estimates are obtained from univariate kernel density estimates.

Plots for the marginal models

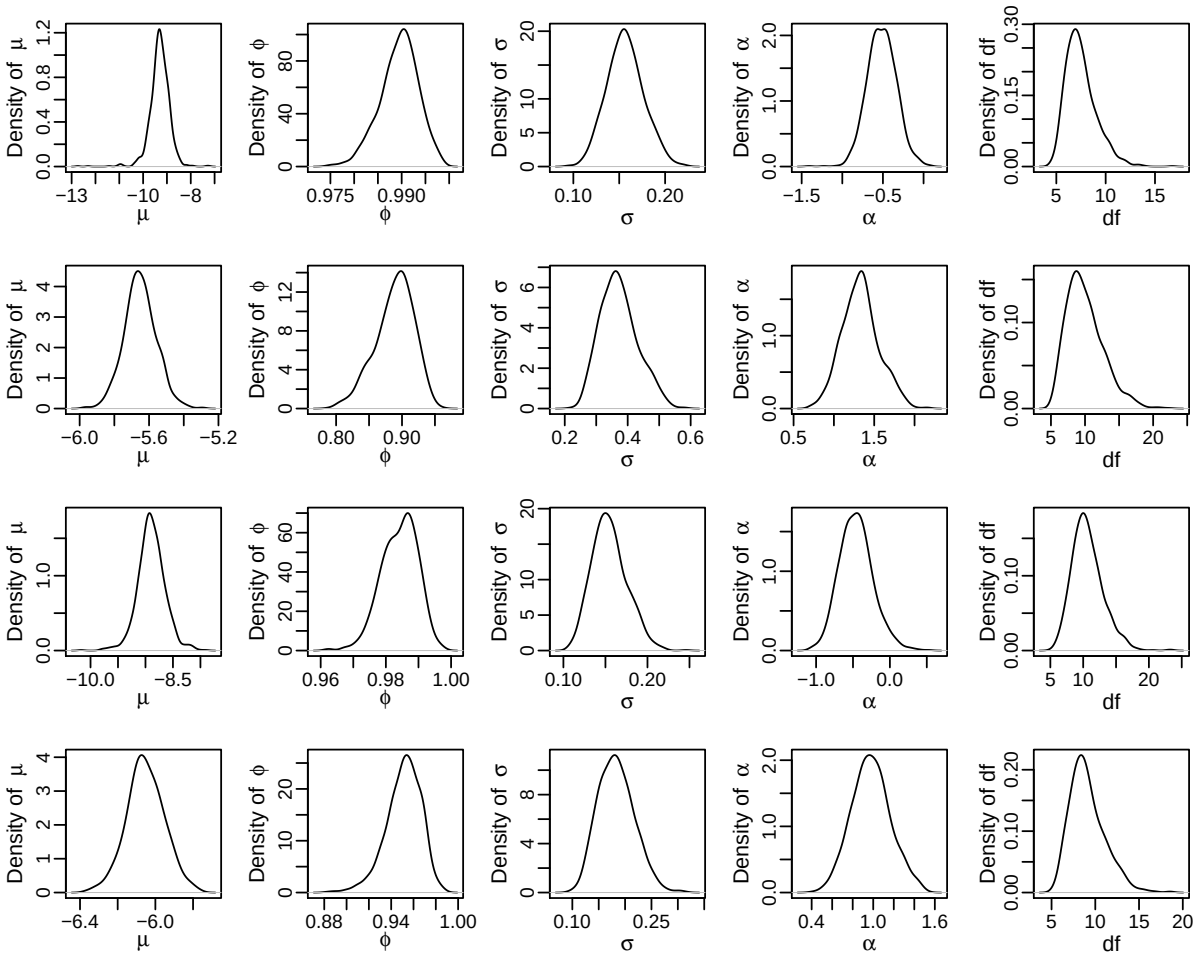

Figure B.4: Kernel density estimates of the posterior densities based on 30000 MCMC iterations after a burn-in of 1000 for the parameters of the univariate skew Student t stochastic volatility models for SPX, VIX, DAX and VDAX (from top to bottom row).

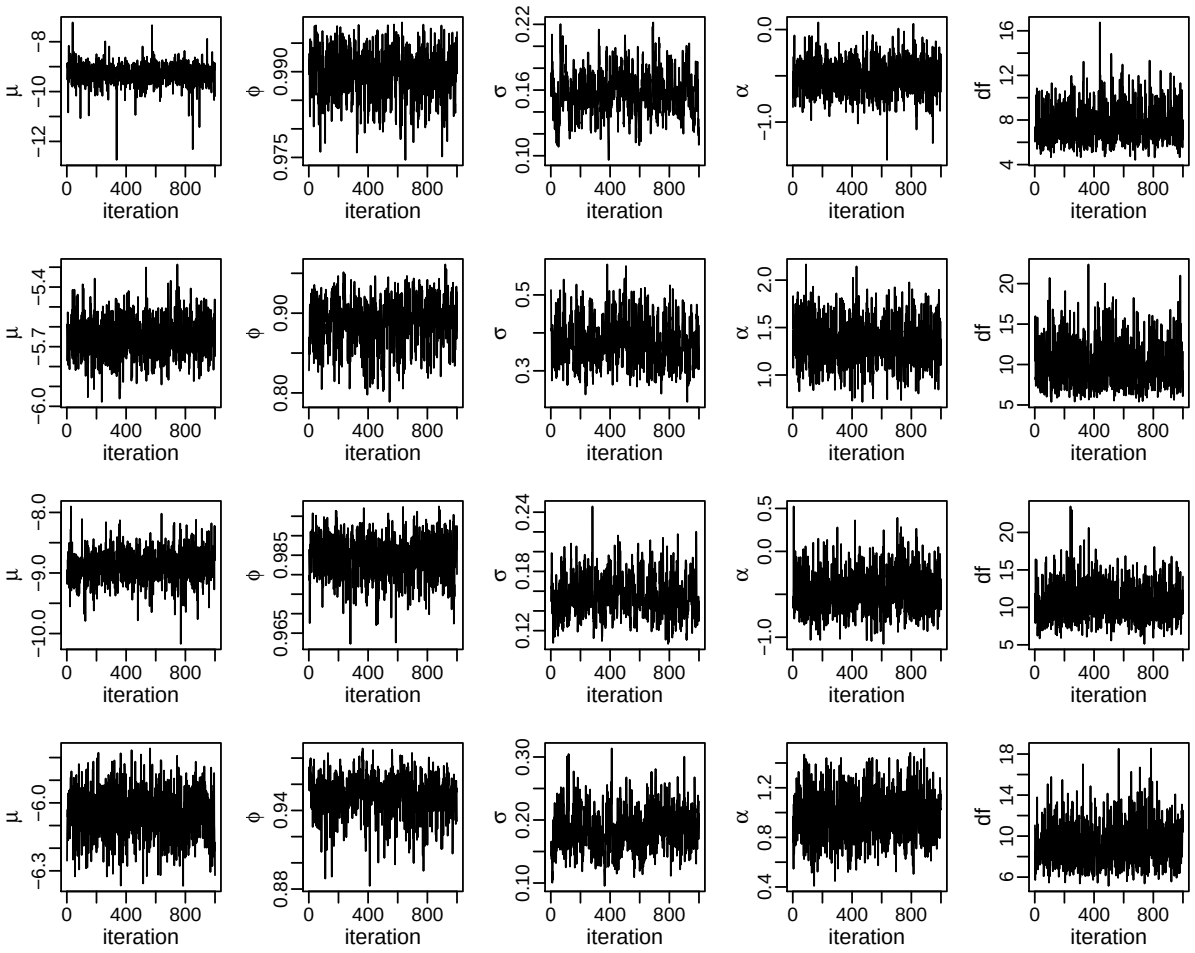

Figure B.5: Trace plots of 1000 MCMC draws based on a total of 31000 iterations, where the first 1000 draws are discarded for burn-in and the remaining 30000 draws are thinned with factor 30. The trace plots are shown for parameters of the univariate skew Student t stochastic volatility models for SPX, VIX, DAX and VDAX (from top to bottom row).

Plots for the dependence models

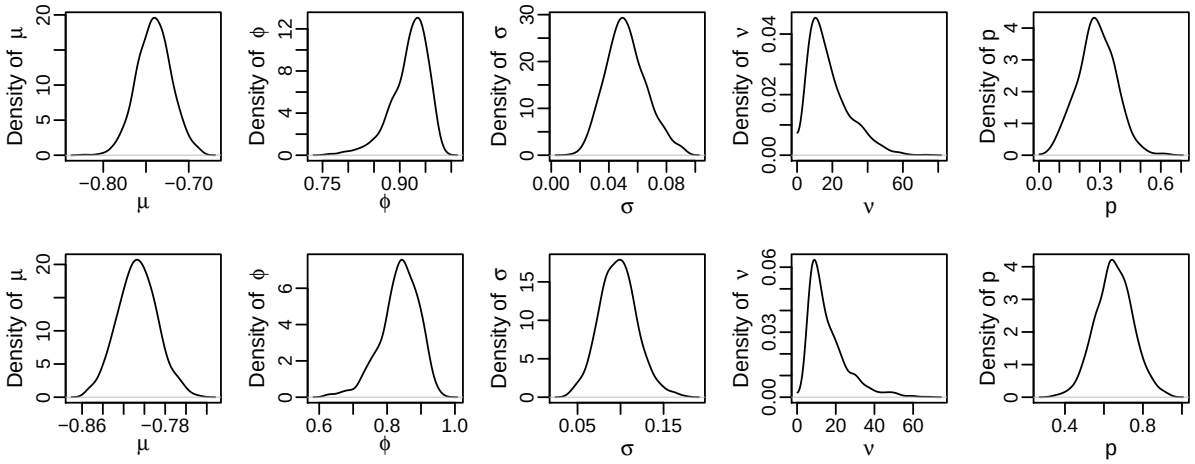

Figure B.6: Kernel density estimates of the posterior densities based on 30000 MCMC iterations after a burn-in of 1000 for parameters of the dynamic mixture copula model for (SPX,VIX) in the top row and for (DAX,VDAX) in the bottom row.

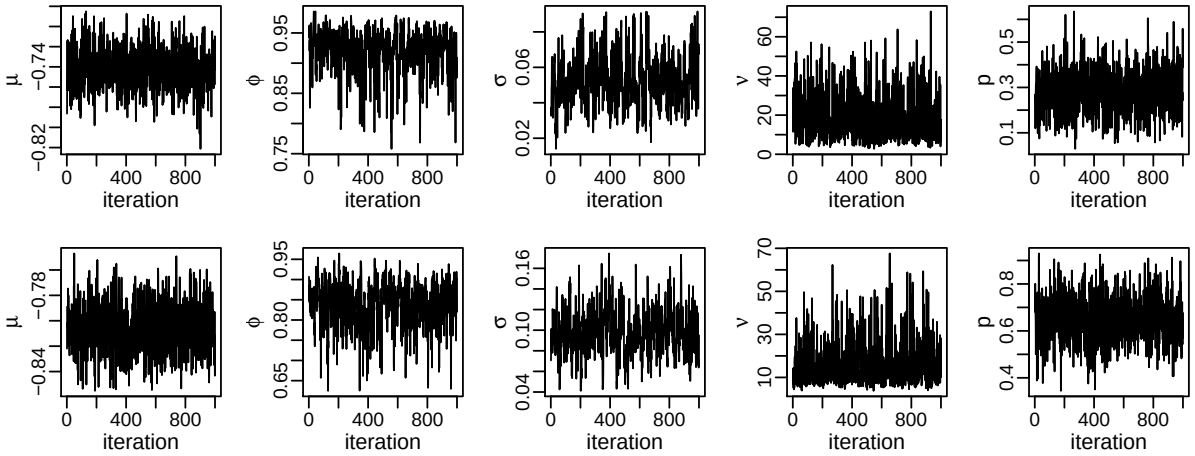

Figure B.7: Trace plots of 1000 MCMC draws based on a total of 31000 iterations, where the first 1000 draws are discarded for burn-in and the remaining 30000 draws are thinned with factor 30. The trace plots are shown for parameters of the dynamic mixture copula model for (SPX,VIX) in the top row and for (DAX,VDAX) in the bottom row.

Calculating the pseudo log predictive score

We describe in detail how we proceed for model $\mathcal{M}_{dyn}^{mix}$. We consider $T+K$ observations of dimension two, stored in the data matrix $\mathrm{Y}_{1:(T+K);1:2} = (y_{tj})_{t=1,\ldots,T+k,j=1,2}$, where the first T observations are used to train the model and the last K are used for evaluation.

Step 1: (Model fitting based on the training period)

- We fit two marginal skew Student t stochastic volatility models to $\boldsymbol{y_{1:T;1}}$ and $\boldsymbol{y_{1:T;2}}$. This yields R_{train} draws of the parameters denoted by $\boldsymbol{s_{1:T;j}^{st,r}}$, $\mu_j^{st,r}$, $\phi_j^{st,r}$, $\sigma_j^{st,r}$, $\alpha_j^{st,r}$ and $df_j^{st,r}$, $r = 1, \ldots, R_{train}$ and corresponding posterior mode estimates $\hat{\boldsymbol{s}}_{\boldsymbol{1:T;j}}^{\boldsymbol{st}}$, $\hat{\mu}_j^{st}$, $\hat{\phi}_j^{st}$, $\hat{\sigma}_j^{st}$, $\hat{\alpha}_j^{st}$ and $\hat{df}_j^{st}$ for $j = 1, 2$.
- We estimate the copula data

$$\hat{u}_{tj} = ssT\left(y_{tj}\exp\left(-\frac{\hat{s}_{tj}^{st}}{2}\right)\middle|\hat{\alpha}_j^{st}, \hat{df}_j^{st}\right)$$

  for $t = 1, \ldots, T, j = 1, 2$.
- We fit the dynamic bivariate mixture copula model introduced in (4.11) based on the pseudo copula data $\hat{U}_{1:T;1:2}$ and obtain posterior draws $\boldsymbol{s_{1:T}^{cop,r}}$, $\mu^{cop,r}$, $\phi^{cop,r}$, $\sigma^{cop,r}$, $\nu^{cop,r}$, $p^{cop,r}$ for $r = 1, \ldots, R_{train}$ and corresponding posterior mode estimates $\hat{\boldsymbol{s}}_{\boldsymbol{1:T}}^{\boldsymbol{cop}}$, $\hat{\mu}^{cop}$, $\hat{\phi}^{cop}$, $\hat{\sigma}^{cop}$, $\hat{\nu}^{cop}$, $\hat{p}^{cop}$.

Step 2: (The one-day ahead predictive density)

Estimating the one-day ahead predictive density at time $T+k, 1 \leq k \leq K$ would usually require to fit daily models with observations up to time $T+k-1$ for $k = 1, \ldots, K$. In order to save computational resources, we use another approach, where we only update the dynamic parameters, i.e. the log variances and Kendall's τ. For the constant parameters we use the estimates from the training period $1, \ldots, T$. In this case we found that it is enough to only consider a time horizon of 100 time points, i.e. to estimate a dynamic parameter at time $T+k$ we consider data in the period $T+k-100, \ldots, T+k-1$. We proceed as follows to obtain the one-day ahead predictive density at time point $T+k$ with $1 \leq k \leq K$.

- We consider a skew Student t stochastic volatility model as in (4.8), where we keep the parameters μ, ϕ, σ, α and df fixed and only update the latent log variances. Therefore we draw the latent log variances $\boldsymbol{s_{(T+k-100):(T+k-1);j}^{st}}$ conditional on $\boldsymbol{y_{(T+k-100):(T+k-1);j}}$, $\hat{\mu}_j^{st}$, $\hat{\phi}_j^{st}$, $\hat{\sigma}_j^{st}$, $\hat{df}_j^{st}$ and $\hat{\alpha}_j^{st}$ for $j = 1, 2$. We denote the draws by $\boldsymbol{s_{(T+k-100):(T+k-1);j}^{st,r}}$, $r = 1, \ldots, R_{test}$ for $j = 1, 2$. Corresponding posterior mode estimates are denoted by $\hat{\boldsymbol{s}}_{\boldsymbol{(T+k-100):(T+k-1);j}}^{\boldsymbol{st}}, j = 1, 2$.
- We estimate the copula data via the probability integral transform, i.e. for $j = 1, 2$ and $t = T+k-100, \ldots T+k-1$, we calculate

$$\hat{u}_{tj} = ssT\left(y_{tj}\exp\left(-\frac{\hat{s}_{tj}^{st}}{2}\right)\middle|\hat{\alpha}_j^{st}, \hat{df}_j^{st}\right).$$

- We fit the dynamic mixture copula model to the data $\hat{U}_{(T+k-100):(T+k-1);(1:2)}$ where we keep the constant parameters fixed. We only update $\boldsymbol{s}^{\boldsymbol{cop}}_{\boldsymbol{(T+k-100):(T+k-1)}}$ conditional on $\hat{U}_{(T+k-100):(T+k-1);(1:2)}$, $\hat{\mu}^{cop}$, $\hat{\phi}^{cop}$, $\hat{\sigma}^{cop}$, $\hat{\nu}^{cop}$, $\hat{p}^{cop}$. The corresponding draws are denoted by $\boldsymbol{s}^{\boldsymbol{cop,r}}_{\boldsymbol{(T+k-100):(T+k-1)}}, r = 1, \ldots, R_{test}$ and the posterior mode estimates by $\hat{\boldsymbol{s}}^{\boldsymbol{cop}}_{\boldsymbol{(T+k-100):(T+k-1)}}$.
- For $j = 1, 2$, we obtain an estimate for the log variance at time point $T + k$ as $\hat{s}^{st}_{T+k;j} = \hat{\mu}^{st}_j + \hat{\phi}^{st}_j(\hat{s}^{st}_{T+k-1;j} - \hat{\mu}^{st}_j)$.
- We obtain an estimate for Fisher's Z transform of Kendall's τ at time point $T + k$ as $\hat{s}^{cop}_{T+k} = \hat{\mu}^{cop} + \hat{\phi}^{cop}(\hat{s}^{cop}_{T+k-1} - \hat{\mu}^{cop})$.
- The predictive density evaluated at (z_1, z_2) is given by

$$f^p_{T+k}(z_1, z_2) = c^p_{T+k}(z_1, z_2) g^p_{T+k}(z_1, z_2),$$

with

$$c^p_{T+k}(z_1, z_2) = c^M\left(ssT\left(x_1 \middle| \hat{\alpha}^{st}_1, \hat{df}^{st}_1\right), ssT\left(x_2 \middle| \hat{\alpha}^{st}_2, \hat{df}^{st}_2\right); F_Z^{-1}(\hat{s}^{cop}_{T+k}), \hat{\nu}^{cop}, \hat{p}^{cop}\right),$$

where c^M is the density of the mixture copula defined in (4.10) and

$$g^p_{T+k}(z_1, z_2) = sst\left(x_1 \middle| \hat{\alpha}^{st}_1, \hat{df}^{st}_1\right) sst\left(x_2 \middle| \hat{\alpha}^{st}_2, \hat{df}^{st}_2\right) \exp\left(-\frac{\hat{s}^{st}_{T+k;1}}{2}\right) \exp\left(-\frac{\hat{s}^{st}_{T+k;2}}{2}\right),$$

with $x_j = z_j \exp\left(-\hat{s}^{st}_{T+k;j}/2\right)$ for $j = 1, 2$.

Step 3: (The cumulative pseudo log predictive score)

The cumulative pseudo log predictive score is obtained as

$$LP = \sum_{k=1}^{K} \ln\left(f^p_{T+k}(y_{T+k;1}, y_{T+k;2})\right).$$

During the training period we run $R_{train} = 31000$ iterations with a burn-in of 1000, while for updating only the dynamic parameters 11000 iterations with a burn-in of 1000 is enough, i.e. we use $R_{test} = 11000$.

Appendix C Supplementary material to Chapter 5

C.1 Parameter specification for the simulation study in Section 5.3.3

Parameters of a dynamic vine copula model are here specified through matrices. The last row shows parameters corresponding to pair copulas in the first tree, the second last row parameters corresponding to pair copulas in the second tree and so on. If we set the dispersion parameter and the standard deviation parameter to zero, we obtain a static copula model.

$$\mu = \begin{pmatrix} 0.0 & & & & \\ 0.0 & 0.0 & & & \\ 0.0 & 0.3 & 0.0 & & \\ 0.3 & 0.4 & 0.3 & 0.0 & \\ 0.9 & 0.6 & 0.8 & 0.8 & 1.0 \end{pmatrix} \quad \phi = \begin{pmatrix} 0.00 & & & & \\ 0.00 & 0.00 & & & \\ 0.00 & 0.00 & 0.00 & & \\ 0.98 & 0.90 & 0.00 & 0.00 & \\ 0.95 & 0.98 & 0.90 & 0.00 & 0.00 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.00 & & & & \\ 0.00 & 0.00 & & & \\ 0.00 & 0.00 & 0.00 & & \\ 0.05 & 0.10 & 0.00 & 0.00 & \\ 0.10 & 0.03 & 0.05 & 0.00 & 0.00 \end{pmatrix}$$

$$\text{family} = \begin{pmatrix} \text{Independence} & & & & \\ \text{Independence} & \text{Independence} & & & \\ \text{Independence} & \text{eClayton} & \text{Independence} & & \\ \text{Gaussian} & \text{Student t(df=4)} & \text{eGumbel} & \text{Independence} & \\ \text{Gaussian} & \text{Student t(df=4)} & \text{eClayton} & \text{eGumbel} & \text{Gaussian} \end{pmatrix}$$

Note that, within the dynamic bivariate copula model, the stationary distribution of the AR(1) process is given by

$$s|\mu, \phi, \sigma \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right).$$

for a state s. Using the density transformation rule, this implies the following density for Kendall's τ (the inverse Fisher's Z transform of s)

$$f(\tau|\mu,\phi,\sigma) = \varphi\left(F_Z(\tau)|\mu, \frac{\sigma^2}{1-\phi^2}\right)\frac{1}{1-\tau^2}, \tau \in (-1,1). \tag{C.1}$$

To obtain an understanding of what different choices of μ, ϕ and σ imply for τ, we show the density given in (C.1) in Figure C.1.

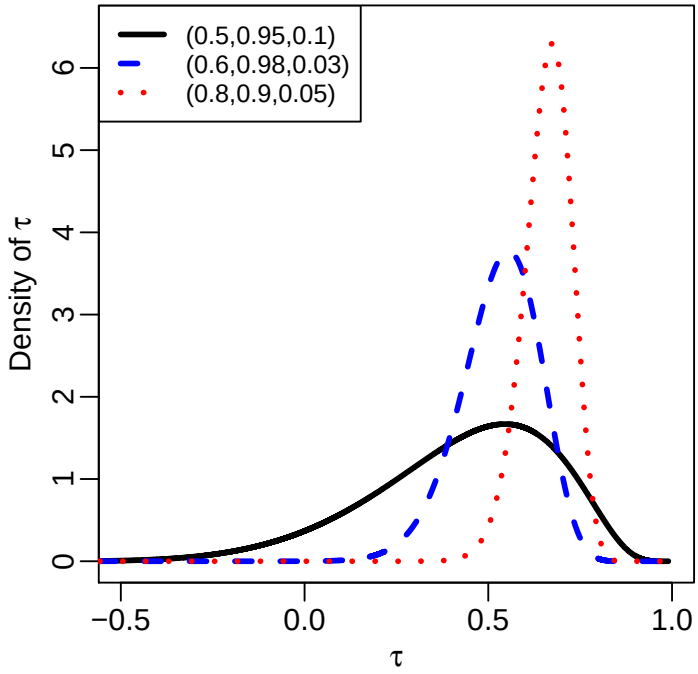

Figure C.1: We show the stationary density of τ given in (C.1) for different values of μ, ϕ and σ. Each line is associated with a vector (μ, ϕ, σ) given in the legend.

C.2 Exchange rates (to the US Dollar) data set

Ticker	Currency
BRL	Brazilian Real
CAD	Canadian Dollar
CNY	Chinese Yuan
DKK	Danish Krone
HKD	Hong Kong Dollar
INR	Indian Rupees
JPY	Japanese Yen
KRW	South Korean Won
MYR	Malaysian Ringgit
MXN	Mexican New Pesos
NOK	Norwegian Krone
SEK	Swedish Krona
ZAR	South African Rand
SGD	Singapore Dollar
CHF	Swiss Franc
NTD	New Taiwan Dollar
THB	Thai Baht
AUD	Australian Dollar
EUR	Euro
NZD	New Zealand Dollar
GBP	British Pound

Table C.1: The 21 currencies with corresponding ticker symbols used in the application in Section 5.4.

Appendix D Supplementary material to Chapter 6

Contour plots of bivariate copula densities

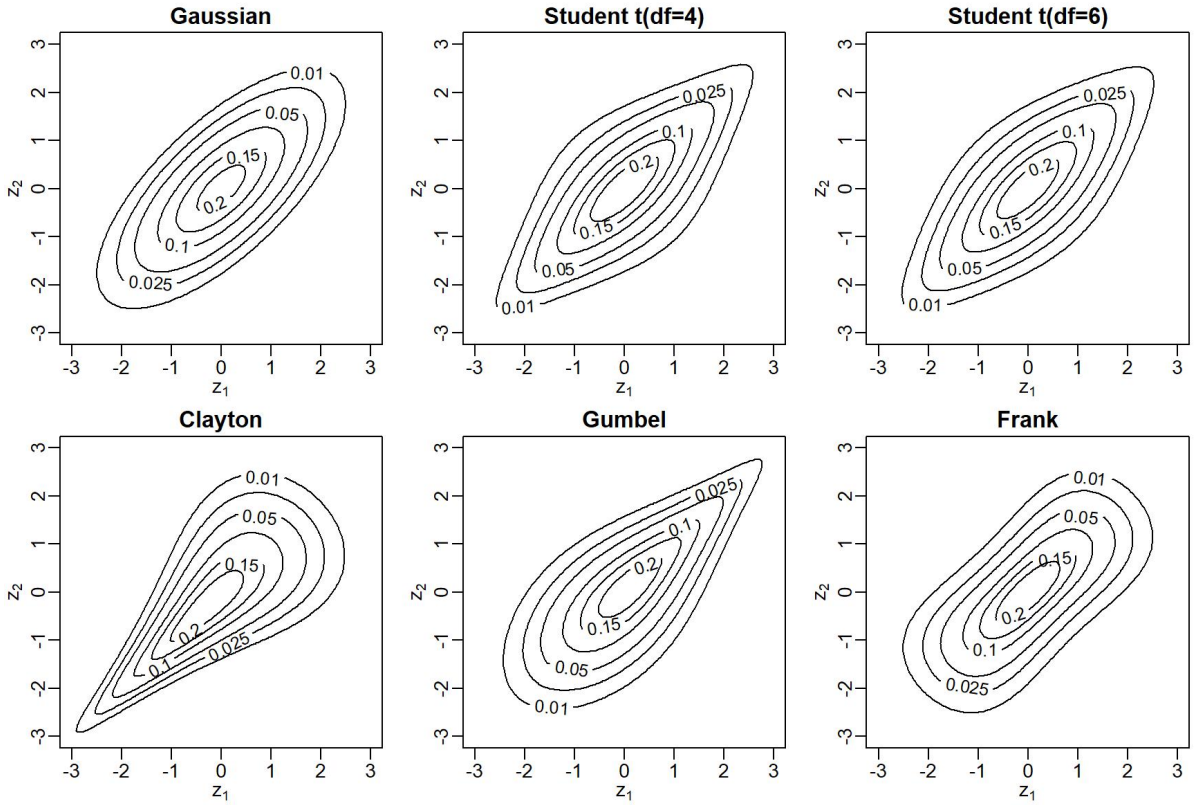

Figure D.1: Normalized contour plots (see Czado (2019), Chapter 3) of bivariate copula densities with Kendall's $\tau = 0.5$.

Appendix E Supplementary material to Chapter 7

E.1 Additional material for Section 7.2.2

The covariance matrix Σ of the joint distribution

$$(Z_{11}, \dots, Z_{d1}, w_1; Z_{12}, \dots, Z_{d2}, w_2; \dots, Z_{1T}, \dots, Z_{dT}, w_T) | (\rho_{obs,j})_{j=1,\dots,d}, \rho_{lat} \sim N_{dT+T}(\mathbf{0}, \Sigma)$$

takes the form

$$\Sigma = \begin{pmatrix} A & \rho_{lat}(A+B) & \rho_{lat}^2(A+B) & \dots & \rho_{lat}^{T-1}(A+B) \\ \rho_{lat}(A+B) & A & \rho_{lat}(A+B) & \dots & \rho_{lat}^{T-2}(A+B) \\ \rho_{lat}^2(A+B) & \rho_{lat}(A+B) & A & \dots & \rho_{lat}^{j-2}(A+B) \\ \rho_{lat}^3(A+B) & \rho_{lat}^2(A+B) & \rho_{lat}(A+B) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{lat}^{T-1}(A+B) & \rho_{lat}^{T-2}(A+B) & \rho_{lat}^{T-3}(A+B) & \dots & A \end{pmatrix}$$

where the matrices A and B take the following forms

$$A = \begin{pmatrix} 1 & \rho_{obs,1}\rho_{obs,2} & \rho_{obs,1}\rho_{obs,3} & \dots & \rho_{obs,1}\rho_{obs,d} & \rho_{obs,1} \\ \rho_{obs,1}\rho_{obs,2} & 1 & \rho_{obs,2}\rho_{obs,3} & \dots & \rho_{obs,2}\rho_{obs,d} & \rho_{obs,2} \\ \rho_{obs,1}\rho_{obs,3} & \rho_{obs,2}\rho_{obs,3} & 1 & \dots & \rho_{obs,3}\rho_{obs,d} & \rho_{obs,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{obs,1}\rho_{obs,d} & \rho_{obs,2}\rho_{obs,d} & \rho_{obs,3}\rho_{obs,d} & \dots & 1 & \rho_{obs,d} \\ \rho_{obs,1} & \rho_{obs,2} & \rho_{obs,3} & \dots & \rho_{obs,d} & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} \rho_{obs,1}^2 - 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \rho_{obs,2}^2 - 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \rho_{obs,3}^2 - 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \rho_{obs,d}^2 - 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

E.2 Further results for the application in Section 7.4

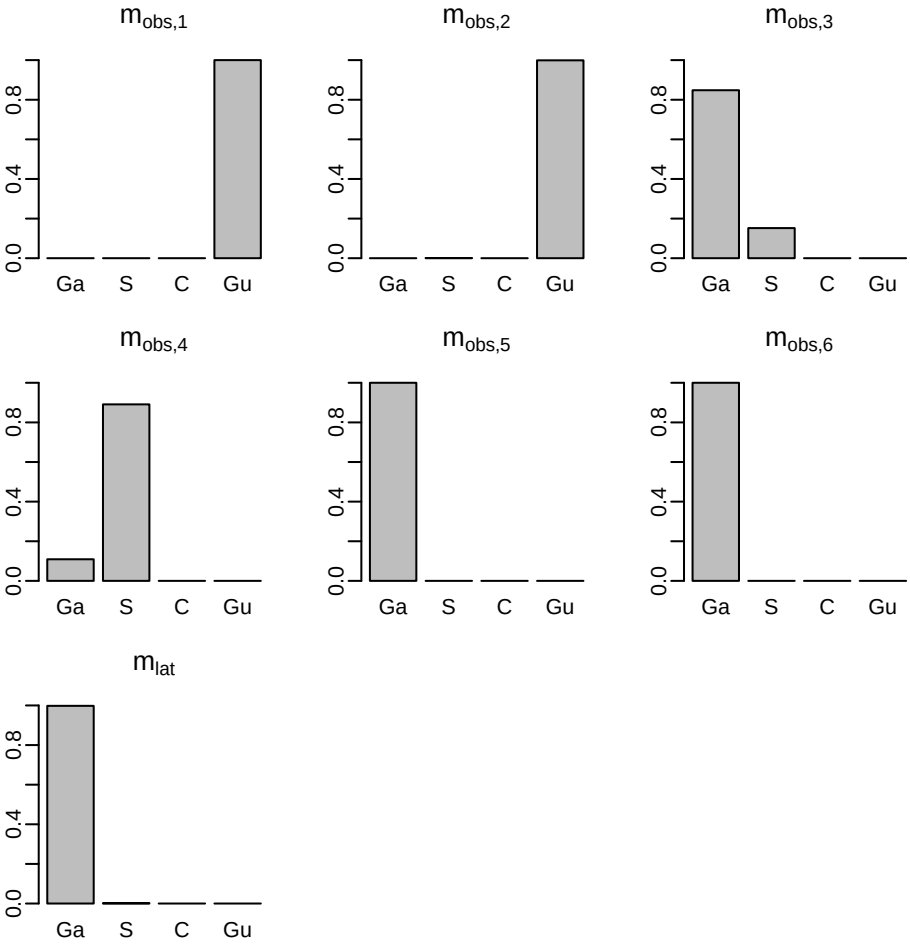

Figure E.1: Estimated posterior distribution of the copula family indicators $m_{obs1}, \dots, m_{obs,6}, m_{lat}$ obtained from 2000 iterations after a burn-in of 1000 (Ga: Gaussian, S: Student t(df=4), C: Clayton, Gu: Gumbel).

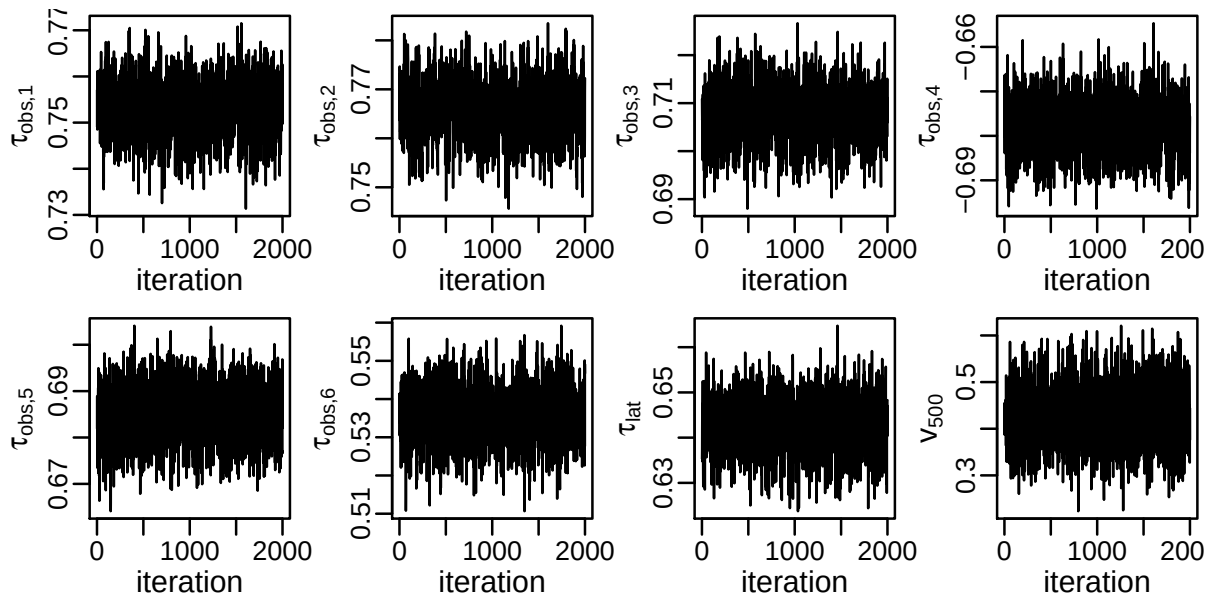

Figure E.2: Trace plots of 2000 draws after a burn-in of 1000 for selected parameters of the copula state space model. The variables are ordered as follows: 1: CO(gt), 2: CO(lc), 3: NOx(gt), 4: NOx(lc), 5: NO2(gt), 6: NO2(lc).

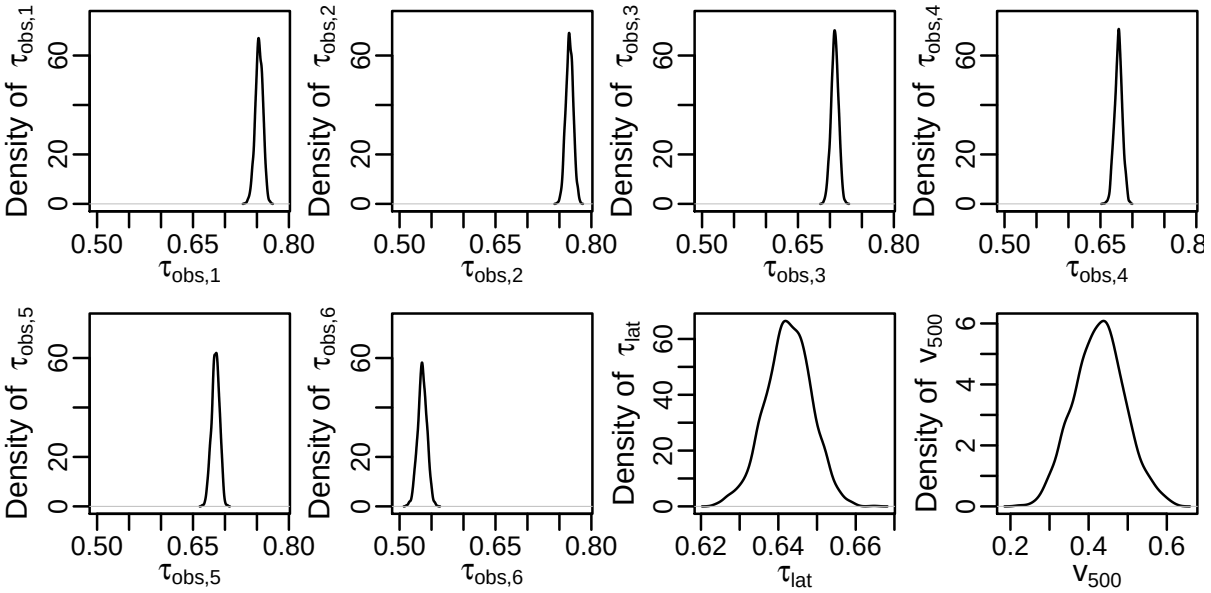

Figure E.3: Estimated posterior density for selected parameters of the copula state space model. The posterior density is estimated as the kernel density estimate based on 2000 draws after a burn-in of 1000. For better comparability we multiplied the draws of $\tau_{obs,4}$ by -1. The variables are ordered as follows: 1: CO(gt), 2: CO(lc), 3: NOx(gt), 4: NOx(lc), 5: NO2(gt), 6: NO2(lc).

E.3 Contour plots over different time periods (Section 7.4.4)

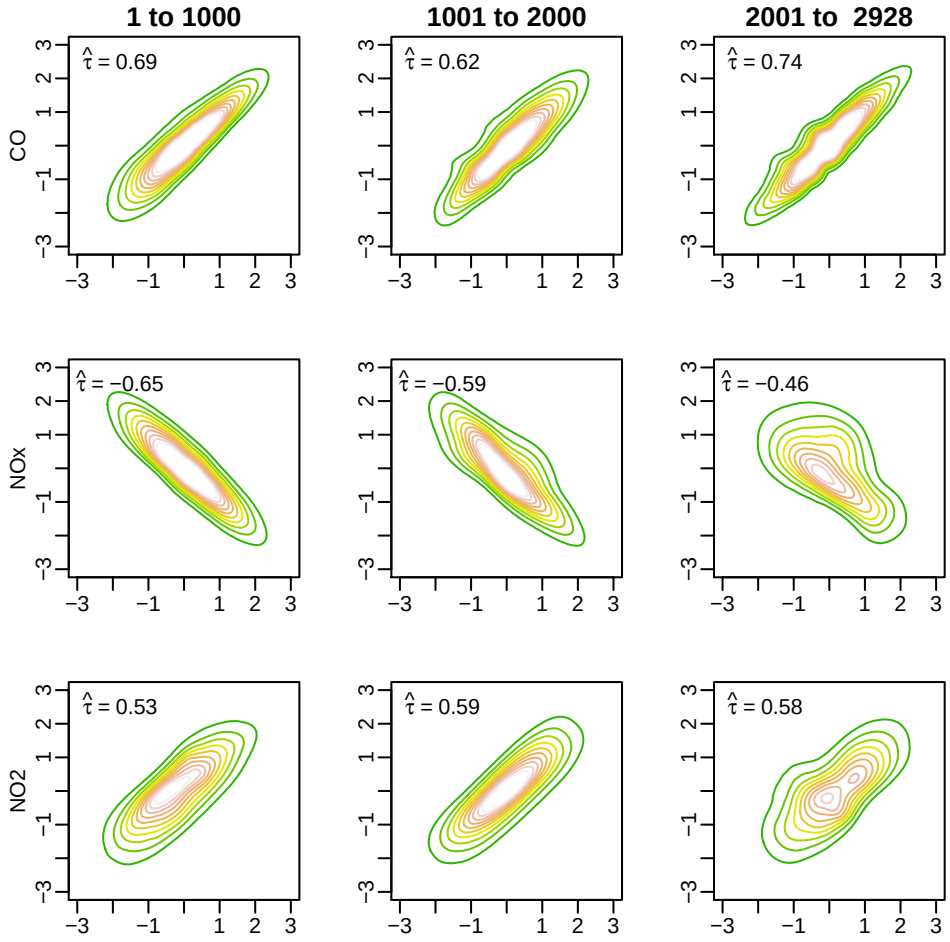

Figure E.4: Contour plots of bivariate kernel density estimates based on pairs $(\hat{z}_{tj}, \hat{z}_{tj'})_{t \in P_i}$, $i = 1, 2, 3$ where j corresponds to a ground truth and j' to the corresponding low-cost value within a time period P_i ($P_1 : 1, \ldots, 1000$, $P_2 : 1001, \ldots, 2000$, $P_3 : 2001, \ldots, 2928$). For example, the top row shows contour plots for the (CO(gt), CO(lc)) pair for the three different time periods. In the top left corner we added the corresponding empirical Kendall's τ, based on the pair $(\hat{z}_{tj}, \hat{z}_{tj'})_{t \in P_i}$.

Appendix F Updating time-varying parameters

For several models we consider, we require one-day ahead predictions/predictive simulations (see Sections 3.4, 4.4 and 5.4). Refitting the model k times, to obtain these one-day ahead predictions/predictive simulations for k different time points can be very expensive. Generally, models considered here for random variables $\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_T} \in \mathbb{R}^d$, observed at T time points can be formulated as

$$\begin{aligned} \boldsymbol{Y_t}|\boldsymbol{s_t}, \boldsymbol{\delta} &\sim f(\boldsymbol{y_t}|\boldsymbol{s_t}, \boldsymbol{\delta}) \\ \boldsymbol{s_t}|\boldsymbol{s_{t-1}}, \boldsymbol{\theta} &\sim f(\boldsymbol{s_t}|\boldsymbol{s_{t-1}}, \boldsymbol{\theta}), \end{aligned} \tag{F.1}$$

for $t = 1, \ldots, T$, where $\boldsymbol{s_t}$ are dynamic (time-varying) parameters and $\boldsymbol{\delta}, \boldsymbol{\theta}$ are static (time-constant) parameters. We suggest to fit the model in (F.1) once, obtain point estimates for the static parameters, denoted by $\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\theta}}$, and then consider the model

$$\begin{aligned} \boldsymbol{Y_t}|\boldsymbol{s_t} &\sim f(\boldsymbol{y_t}|\boldsymbol{s_t}, \hat{\boldsymbol{\delta}}) \\ \boldsymbol{s_t}|\boldsymbol{s_{t-1}} &\sim f(\boldsymbol{s_t}|\boldsymbol{s_{t-1}}, \hat{\boldsymbol{\theta}}), \end{aligned} \tag{F.2}$$

where the static parameters are fixed at point estimates. The model in (F.2) has only dynamic parameters. To obtain k one-day ahead predictions/predictive simulations, the model in (F.2) is then estimated k times. In other words: We fix the static parameters at point estimates and only update dynamic parameters. Further, we have seen that the model in (F.2) requires less data points for estimation than the model in (F.1). For the model in (F.1), we typically use a sample size of 1000 or larger to estimate parameters. But for the model in (F.2), i.e. when the static parameters are fixed at point estimates, we have seen that it might be enough to consider only the last 100 data points before the time point we want to predict.

Appendix G Additional details for parameter sharing

This chapter is based on Kreuzer and Czado (2019b). Our approaches in Sections 3.3.1, 5.2.1 and 7.2.4 share parameters among different copula families. This is motivated by the fact that the Kendall's τ parameter is similar for different copula families. To support this statement, we conduct the following experiment: We simulate 100 bivariate data sets, each containing 1000 observations, from the bivariate Student t copula with 4 degrees of freedom and copula parameter ρ_{true}. The corresponding Kendall's τ is obtained as $\tau_{true} = \frac{2}{\pi}\arcsin(\rho_{true})$. For each data set, we estimate the copula parameter of the Gaussian, Student t, Clayton and Gumbel copula by maximizing the likelihood and transform the estimates to the corresponding Kendall's τ values. We obtain 100 estimated Kendall's τ values for each copula family and take the average of those 100 values, which we denote by $\hat{\tau}$. This results in four different $\hat{\tau}$ values corresponding to four different copula families. This procedure is repeated for different values of τ_{true} and the average Kendall's τ estimate, $\hat{\tau}$, is shown in Figure G.1 for each value of τ_{true}. We see that the estimated Kendall's τ values for the Gaussian, Student t and Gumbel copula are close to each other. Although the Kendall's τ estimates for the Clayton copula are a bit further apart, we think that they are still reasonable close to justify parameter sharing.

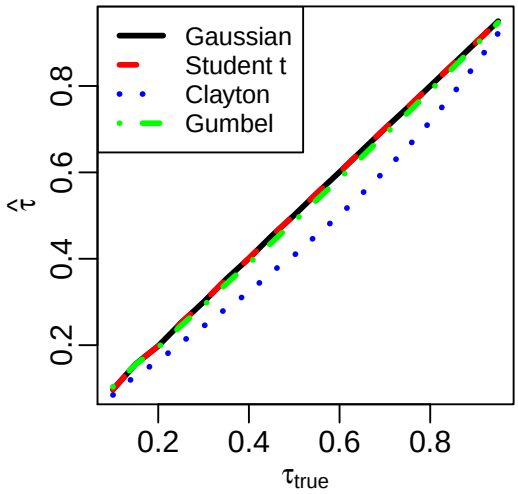

Figure G.1: This plot shows average Kendall's τ estimates, $\hat{\tau}$, for different copula families (Gaussian, Student t(df=4), Clayton, Gumbel) plotted against the Kendall's τ that was used for simulation, τ_{true}.