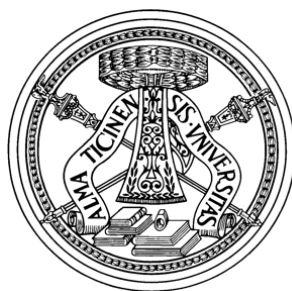


UNIVERSITÀ DEGLI STUDI DI PAVIA
Dipartimento di Matematica “Felice Casorati”

Dottorato di Ricerca in Matematica e Statistica

XXIX CICLO



Bayesian Analysis of AR Copula Models with Tree Structural Representation

Advisor:

Prof. Federico Bassetti

Co-Advisor:

Prof. Claudia Tarantola

Ph.D. thesis of:

Enrica Nicolino

Contents

Introduction	i
1 Preliminaries	1
1.1 Copula Functions	1
1.1.1 Basic Properties	2
1.1.2 Elliptical Copulas	3
1.1.3 Archimedean Copulas	4
1.2 Dependence Measures	6
1.2.1 Kendall's Tau	7
1.2.2 Tail dependence	8
1.3 Rotated Copulas	9
1.4 Copulas and Graphical Models	10
1.4.1 Tree Copula	12
1.4.2 Mixture of Tree Copulas	14
1.4.3 Factor Copula	15
1.5 Copula and Time Series Models	16
1.5.1 Frequentist Inference	17
1.5.2 Bayesian Inference	18
1.5.3 Bayesian AR-Copula Model	19
1.5.4 Bayesian AR-Markov Switching Copula Model	20
1.5.5 Model Selection Analysis	22
2 Tree Copula Model and Tree Copula Mixture Model	24
2.1 Bayesian inference for Tree Copula Model	24
2.2 MCMC for Tree Copula Model	26
2.3 Bayesian Inference for Finite Tree Copula Mixture Model	29
2.4 MCMC for Mixture of Tree Copulas Model	30
2.5 Simulation Study	33

3	DP-Tree Copula Model	47
3.1	Dirichlet Process	47
3.1.1	Stick Breaking Construction	48
3.1.2	Infinite Mixture Models	48
3.2	Bayesian Inference for DP-Tree Copula Model	50
3.3	MCMC for DP-Tree Copula Model	52
3.4	Simulation Study	55
4	Factor Copula Model and Markov Switching Factor Copula Model	63
4.1	Bayesian Inference for Factor Copula Model	63
4.2	MCMC for Factor Copula Model	65
4.3	Bayesian Inference for Markov Switching Factor Copula Model	67
4.4	MCMC for Markov Switching Factor Copula Model	69
4.5	Simulation Study	72
5	Energy Data Analysis	82
5.1	Model Selection	83
5.2	Energy Market Portfolio Analysis	87
5.3	Dependence Structure Analysis	94
5.3.1	Tree Copula and Mixture of Tree Copulas Models	94
5.3.2	DP-Tree Copula Mixture Model	100
5.3.3	Factor Copula Model	102
5.4	Markov Switching Factor Copula Model	104
6	Conclusion and Extensions	115
6.1	Conclusion	115
6.2	Future Work	116

Introduction

Modeling the multivariate dependencies between a set of random variables is an interesting problem in statistical science that has received considerable attention in the last years. The accurate analysis of the joint distribution among random variables is very important in many applications, such as insurance, risk management, energy market, finance and many other areas. One increasingly popular approach for modeling multivariate dependencies is based on copula functions.

In this thesis, we deal with a Bayesian analysis of AR-copula models, in which the joint distribution of the innovations of a panel of AR time series is described via a suitable multivariate copula. In particular, we focus on three alternative copula structures: the tree copulas, the mixture of tree copulas and the factor copulas.

A copula is a multivariate distribution with uniform margins on the simplex and permits to obtain the joint multivariate distribution embedding the variable's dependence structure. Indeed, according to Sklar's Theorem, Sklar (1959), each multivariate distribution can be decomposed into its marginal distributions and a copula \mathcal{C} that represents the dependence structure. Given a random vector $\mathbf{X} = (X_1, \dots, X_N)$ with absolutely continuous distribution F and density f , this relationship can be written in terms of density as

$$f(x_1, \dots, x_N) = c(F_1(x_1), \dots, F_N(x_N))f_1(x_1) \cdots f_N(x_N)$$

where c is a copula density and F_j and f_j are the marginal distribution function and the density, respectively, of X_j .

In the literature, there are many different bivariate copula families available that are widely applied because of their flexibility and the fact that they are easy to compute. Unfortunately, for the multidimensional setting, the choice of a multivariate copula is rather scarce, due to theoretical and computational limitations.

Recently, graphical models have been applied in order to simplify the construction of multivariate copulas. Indeed, the use of graphical structures allow to represent a multivariate copula via a set of suitable bivariate ones; see e.g. Joe (1996), Bedford and Cooke (2002), Kirshner (2007), Aas et al. (2009), Silva and Gramacy (2009), Elidan (2013), Stöber and Czado (2014), Krupskii and Joe (2015), Dalla Valle et al. (2016) and Oh and Patton (2016).

One possible approach for the construction of multivariate copulas in terms of bivariate ones is the Pair Copula Construction (PCC), introduced by Joe (1996) and largely applied in the literature. The PCC permits to decompose a multivariate distribution into bivariate unconditional and conditional copulas, called pair copulas or linking copulas. This factorization is not unique and therefore, in order to organize all the possible factorizations, Bedford and Cooke (2001) and Bedford and Cooke (2002) introduced graphical structures, called regular vines (R-vine), as pictorial representation of PCCs. R-vines are built using a nested set of trees called vine structure. Usually, in concrete inferential problem, in order to estimate the copula parameters and the vine structure, a two step procedure is applied. Following Aas and Berg (2009) and Aas et al. (2009), the underlying graphical structure is chosen a priori, by means of a preliminary dependence analysis, and then, given that structure, the parameters of the pair copulas are estimated. Indeed, since the number of possible pair copula structures increase very rapidly with the number of variables, the problem of a complete inference of both the graphical structure and the copula parameters is an extremely difficult task. Examples of frequentist two step procedures are the Inference Function for Margins (IFM), Joe (1997), and the Canonical Marginal Likelihood (CLM), Genest et al. (1995). The combinatorial complexity of the vine structures turns out to be problematic also in the Bayesian framework. This leads to the use of suitable Bayesian two step approaches which avoid a direct complete Bayesian inference. Recently, Gruber and Czado (2015a) and Gruber and Czado (2015b) propose complete Bayesian model selection procedures for the analysis of regular vines. Unfortunately, due to the nested structure of these kinds of copulas, their algorithms are very demanding.

In this thesis, in order to reduce the complexity of a fully Bayesian learning procedure, we propose alternative Bayesian models based on tree copulas, introduced by Kirshner (2007), and on factor copulas, see e.g. Krupskii and Joe (2013). These copulas are special cases of truncated regular vines, Kurowicka (2011), and hence present a simpler underlying graphical structure.

Our interest is to apply copula functions to model and to study the dependence of time series. In this context, the joint density of the innovations is typically described via a suitable copula, replacing in this way the common assumption of joint normality, see e.g. Hofmann and Czado (2010), Min and Czado (2010), Czado et al. (2011). In this thesis, we focus on a specific time series model, the $AR(p)$. We assume to have a set of N $AR(p)$ series in which the k -th component is given by

$$X_{k,t} = \sum_{i=1}^p \alpha_{k,i} X_{k,t-i} + \varepsilon_{k,t}$$

where $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})$ is the vector of the innovations, assumed to be independent and identically distributed. In this work, the alternative Bayesian copula models mentioned above are applied to make inference on the law of the innovations of the $AR(p)$ models.

In our first model, we assume that the multivariate density of the innovations is represented via a tree copula distribution which combines a Markov tree network with suitable bivariate linking copulas. Through the use of the global Markov property associated to the tree structure, one reads conditional independencies from the graph and obtains a useful decomposition of the multivariate density in terms of bivariate linking copulas. More precisely, if the multivariate density f of the innovations ε has Markov tree dependence structure described by a tree copula with set of edges \mathcal{E} , then it factorizes as

$$f(\varepsilon_1, \dots, \varepsilon_N) = \prod_{(l,m) \in \mathcal{E}} c_{l,m}(F_l(\varepsilon_l), F_m(\varepsilon_m)) \prod_{k=1}^N f_k(\varepsilon_k),$$

where (l, m) denotes the edge between ε_l and ε_m and $c_{l,m}$ a bivariate copula density associated to (l, m) . In order to relax the limitations imposed from the assumption of the Markov tree structure, we also study Bayesian models where the multivariate density is represented via a mixture (finite and infinite) of tree copula distributions. In particular, for the infinite mixture model, we rely on the Dirichlet process mixture model. In all these models, we proceed with a fully Bayesian estimation that allows us to make inference also on the underlying graphical structure associated to the multivariate density.

As an alternative to the tree copula models, we analyze the one-factor copula model. In this case, the dependence among the random variables is explained via a latent variable V and again the multidimensional density is decomposed in the product of suitable linking copulas. In the one-factor copula model, one assumes that the variables $\varepsilon_1, \dots, \varepsilon_N$ are conditionally independent given a latent variable V and, therefore, the joint density can be re-written as

$$f(\varepsilon_1, \dots, \varepsilon_N) = \int_0^1 \prod_{k=1}^N \left(c_{\theta_k, 0}(F_k(\varepsilon_k), v) f_k(\varepsilon_k) \right) dv$$

where $c_{k,0}$ denotes the bivariate copula density between ε_k and V . The underlying graph associated to a one-factor copula is fixed, but the presence of an additional latent variable allows for more flexibility.

Finally, we study a Markov switching factor copula model. As usual, in a Markov switching copula model, one assumes the presence of two or more distinct regimes that are characterized by different levels of dependence among the examined variables. The switching mechanism is controlled by an hidden latent state variable S_t that determines the specific copula structure of the innovations at time t . $(S_t)_{t \geq 1}$ is assumed to be a first order Markov chain in discrete time which takes values in a finite set $\{1, \dots, R\}$. In particular, we suppose that the joint dependence structure in each regime is given by a suitable one-factor copula. By permitting switching between the different regimes, this model is able to capture more complex dependence patterns.

For all the models, since we obtain posterior densities not in closed form, we develop suitable MCMC algorithms. We rely on Metropolis-within-Gibbs methods in order to approximate posterior quantities of interest. The most delicate step for the tree copula-based models is related to the tree structures and copula parameters. To sample these parameters, we use a local move, called tree-angular move introduced in Silva and Gramacy (2009). In this way, one proposes a new tree structure and, then, given the new tree, the copula parameters are sampled using a suitable random walk proposal. Some MCMC code, written in MatLab, can be found at the link <http://matematica.unipv.it/nicolino/>.

As a case study, we present an application of our methodologies to the analysis of Italian and German energy markets. Our aim is to understand the different roles of the drivers of the energy price, and to identify the dependence structure characterizing the market. Indeed, the power price is strongly related to the price and quantity of raw materials used to produce it, such as coal, oil, gas and also the carbon emission price (CO₂). In particular, we consider daily time series of one-year forward contracts for the commodities Power Italy, Power Germany, Brent (oil), TTF (gas), PSV (gas), CO₂ and Api2 (coal). We apply the different models to each domestic market in order to obtain information on the dependence structure and we also proceed with portfolio analysis and evaluation.

The general outline of this thesis is as follows. **Chapter 1** presents the mathematical concepts and the notations that will be used throughout this work. We provide the definition and main results on copula functions and we introduce important dependence measures and their relations with copulas. Moreover, we describe the particular copula structures used in the remaining Chapters. Finally, we show how copula functions can be applied in time series models. In particular, we focus on Bayesian AR-copula models where the joint density of the innovations is represented through a suitable multivariate copula.

In **Chapter 2**, based on material from Bassetti et al. (2016), we study two alternative Bayesian AR-copula methodologies where a tree copula and a finite mixture of tree copulas are, respectively, considered. First, we focus on a tree copula model, in which the joint density of the innovations is represented through a tree copula distribution. In our Bayesian approach, we assume that both the underlying graphical structure associated to the tree copula and the linking copulas are unknown. Hence, we consider a complete Bayesian inference for estimate the copula parameters and the tree structure. In the second model, named tree copula mixture model, in order to overcome the limitations imposed by a tree copula distribution, we consider a model in which the joint density of the innovations is represented by a finite mixture of tree copulas. Also in this case, both the tree structures and the copula parameters are assumed unknown and are inferred with a complete Bayesian approach. For both models, we present the related MCMC algorithms developed to approximate

posterior quantities of interest. Finally, we study the performance of our models using different simulation studies.

As an alternative to the mixture tree copula model where one needs to choose a priori the number of mixture components, in **Chapter 3** we develop a Bayesian non parametric approach by using the Dirichlet Process tree copula model. In this case, the number of components is unknown a priori and can be inferred from the data. At the beginning of the Chapter, we give a brief introduction to the Dirichlet process and then we present our Bayesian model and the corresponding MCMC algorithm based on the slice sampling approach, proposed by Walker (2007) and Kalli et al. (2011). Also in this case, we perform a complete Bayesian inference on the copula parameters and on the tree structures. Finally, we report the results obtained in a simulation study.

In **Chapter 4**, we analyze two models. In the first, named factor copula model, the innovations of the AR time series are described via a one-factor copula. With this structure we have an additional latent variable V and the innovations are assumed to be conditionally independent given V . In the second case, we consider a Markov switching copula model to represent the multivariate distribution of the innovations. In particular, we assume that, in each regime, the dependence structure is given through a suitable one-factor copula. For each model, we present the corresponding MCMC algorithm. For the Markov switching copula model, in order to simulate the hidden Markov chain, we apply the forward filtering-backward sampling method, Carter and Kohn (1994) and Frühwirth-Schnatter (1994). Finally, we consider alternative simulation studies in order to investigate the performance of our models.

Finally, in **Chapter 5**, we apply the different copula-based models to the analysis of the Italian and the German market and present the corresponding results. In particular, we focus on the portfolio evaluations and on dependence structure analysis. For all the models, we obtain a graphical tree representation of the dependence structure, via maximum a posteriori probability tree in case of tree copula model or via the minimum spanning tree construction for the other cases. Concerning dependence analysis, with the tree copula, mixture of tree copulas and factor copula models, we discover some interesting findings. For the Italian market, the commodity that shows in most of the cases direct connection with the Italian power price is the TTF. This commodity, in the factor model, turns out also to be identified as “root” node. This is coherent with the fact that in the Italian energy market, the power is mainly produced using gas. Moreover, we find always a strong correlation between TTF and PSV, since they are both gas price, and a path of length three connecting them to Api2, the coal price. For the German market, an edge between Api2 and TTF is always presented and we find that Api2 is the commodity that plays a central role in the factor copula models. Regarding portfolio analysis, our fully Bayesian procedure presents an overall good performance for both the markets. For sake of comparison, we also consider a Bayesian IFM procedure. We find that our fully Bayesian approach shows best results respect to the Bayesian

IFM one in term of mean distances between the historical and predictive portfolios. Among all the models, we obtain that the DP-tree model shows the lowest values of mean distances for both market.

With regard to the Markov switching factor copula model, we consider the data from January 2013 to December 2014 and find 3 interesting regimes: state 1 associated to the observations of year 2013, state 3 to year 2014 and a third regime (state 2) that identifies the period of change from one year to another. We discover that each regime is characterized by significantly different dependence structures. In particular, state 2 shows higher volatility and higher correlations among the variables.

Acknowledgment. I'm grateful to ENEL Group (Global Trading, Modeling & Pricing Division) for providing the data and for helpful comments and discussions. I also would like to thank the referees for carefully reading my dissertation and for giving constructive comments and observations.

Chapter 1

Preliminaries

In this Chapter we introduce some results that we will use in the rest of the thesis. In particular we provide the main definitions and results on copula functions and the different copula-based models studied in the following part.

In Section 1.1 are presented fundamentals of copula theory, including the seminal theorem by Sklar and some important copula families used in the literature. In Section 1.2 are described two relevant dependence measures (upper tail and Kendall's tau) and their relations to copulas. In Section 1.3 we show how to construct particular copulas, named rotated copulas, using suitable decreasing functions, while in Section 1.4 we focus on the copula structures analyzed in this thesis, i.e. tree copula, mixture of tree copulas and factor copula. Finally, in Section 1.5 we give an introduction to the use of copula in time series models, with particular attention to the AR-copula models.

1.1 Copula Functions

Copula functions are useful instruments for studying the dependence between random variables. They can be used to joint (or couple) one-dimensional marginal distributions, eventually belonging to different families, to obtain a joint multivariate distribution. Due to their flexibility, copulas showed their effectiveness in the dependence analysis in different areas. Examples of applications of copulas include general multivariate time series [e.g. Hofmann and Czado (2010), Min and Czado (2010), Czado et al. (2011)], energy market [e.g. Liu (2011), Wen et al. (2012), Wu et al. (2012), Jaschke (2014), Marimoutou and Soury (2015)], financial data [e.g. Gruber and Czado (2015a), Gruber and Czado (2015b), Laih (2014)] and many others.

In the following we briefly present the main results concerning the copula function and some commonly used copula families.

In the rest of the thesis, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the underlying probability space where all the random variables are defined.

1.1.1 Basic Properties

This section provides the definition and a few of the fundamental properties of copulas.

Definition 1.1. A N -dimensional *Copula* \mathcal{C} is a multivariate distribution function with standard uniform margins on $[0, 1]$.

Definition 1.2. If \mathcal{C} is an N -dimensional absolutely continuous copula function, then its *density* c is defined as

$$c(u_1, \dots, u_N) = \frac{\partial \mathcal{C}(u_1, \dots, u_N)}{\partial u_1 \cdots \partial u_N}. \quad (1.1)$$

In the above definition, the derivative is well-defined almost surely.

The following theorem elucidates the role that copula plays in linking a multivariate distribution with its univariate margins.

Theorem 1.3 (Sklar, 1959). *Let X_1, X_2, \dots, X_N be random variables with distribution function F_1, F_2, \dots, F_N respectively, and joint distribution function F . Then there exists a N -copula \mathcal{C} such that for all $\mathbf{x} \in \mathbb{R}^N$*

$$F(x_1, x_2, \dots, x_N) = \mathcal{C}(F_1(x_1), F_2(x_2), \dots, F_N(x_N)). \quad (1.2)$$

If F is a continuous distribution, then \mathcal{C} is uniquely defined as

$$\mathcal{C}(u_1, \dots, u_N) = F(F_1^{-1}(u_1), \dots, F_N^{-1}(u_N)), \quad (1.3)$$

where $F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}$; otherwise, \mathcal{C} is uniquely determined on $\text{Ran}(F_1) \times \text{Ran}(F_2) \times \dots \times \text{Ran}(F_N)$, where $\text{Ran}(F_i)$ is the range of the marginal distribution.

Conversely, if \mathcal{C} is an N -copula and F_1, F_2, \dots, F_N are distribution functions, then the function F defined by (1.2) is an N -dimensional distribution function with margins F_1, F_2, \dots, F_N .

This theorem is due to Sklar (1959). For a proof see Theorem 2 in Schweizer and Sklar (1974) or Theorem 2.10.9 in Nelsen (2006).

If F is absolutely continuous, its probability density function (pdf) can be easily derived from the corresponding copula function \mathcal{C} . In this case (1.2) can be rewritten in terms of densities and it turns out that the joint pdf f is

$$f(x_1, \dots, x_N) = c(F_1(x_1), \dots, F_N(x_N)) f_1(x_1) \cdots f_N(x_N) \quad (1.4)$$

where f_1, \dots, f_N are the marginal densities of F .

Another important property of the copula functions is that they are invariant under increasing transformations. For simplicity we state the next result only for the case $N = 2$.

Theorem 1.4. *Let X_1, X_2 be random variables with joint distribution function F , marginal distribution functions F_1, F_2 and corresponding copula \mathcal{C} .*

- 1) If h_1 and h_2 are strictly increasing functions on $\text{Ran}(F_1)$ and $\text{Ran}(F_2)$, respectively, then the copula function of the vector $(h_1(X_1), h_2(X_2))$ is \mathcal{C} .
- 2) If h_1 and h_2 are strictly decreasing functions on $\text{Ran}(F_1)$ and $\text{Ran}(F_2)$, respectively, then the copulas $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ of the pairs $(h_1(X_1), X_2), (h_1(X_1), h_2(X_2)), (X_1, h_2(X_2))$, respectively, are independent of the particular choices of h_1 and h_2 and are given by

$$\begin{aligned}\mathcal{C}_1(u_1, u_2) &= u_2 - \mathcal{C}(1 - u_1, u_2) \\ \mathcal{C}_2(u_1, u_2) &= u_1 + u_2 - 1 + \mathcal{C}(1 - u_1, 1 - u_2) \\ \mathcal{C}_3(u_1, u_2) &= u_1 - \mathcal{C}(u_1, 1 - u_2).\end{aligned}\tag{1.5}$$

For a proof of the previous theorem see Theorem 3 in Schweizer and Wolf (1981).

Sometimes, the copula defined in Equation (1.5) are called rotated copulas. More precisely, \mathcal{C}_1 is usually referred as the 90° rotated copula, \mathcal{C}_2 as the 180° rotated copula and \mathcal{C}_3 as the 270° rotated copula.

1.1.2 Elliptical Copulas

Elliptical copulas are copulas derived from elliptical distributions by the application of the Sklar's theorem. The class of elliptical distributions provides an important source of multivariate distributions such as the multivariate Normal and Student's t distributions. In general, a N -dimensional random vector $\mathbf{X} = (X_1, \dots, X_N)$ is said to be elliptically distributed (or simply elliptical) if and only if there exist a vector $\boldsymbol{\mu} \in \mathbb{R}^N$, a positive semidefinite symmetric matrix $\Sigma \in \mathbb{R}^{N \times N}$, and a function $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}$ such that the characteristic function $t \mapsto \varphi_{\mathbf{X}-\boldsymbol{\mu}}(t)$ of $\mathbf{X} - \boldsymbol{\mu}$ corresponds to $t \mapsto \phi(t' \Sigma t)$, $t \in \mathbb{R}^N$. We recall that the characteristic function $\varphi_{\mathbf{X}}$ of a vector \mathbf{X} is defined as $\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[\exp\{i\mathbf{t}' \mathbf{X}\}]$.

Definition 1.5. If F is an elliptical distribution function, then a copula \mathcal{C} defined via Equation (1.2) is called *elliptical copula*.

The two most relevant elliptical copulas are the Gaussian and Student's t copula that we briefly introduce.

Gaussian Copula. Let Φ_N be the distribution function of a N -dimensional Normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$ with mean $\mathbf{0}$ and correlation matrix Σ such that it belongs to $[-1, 1]^{N \times N}$. Then, the multivariate Gaussian copula \mathcal{C} is given by

$$\begin{aligned}\mathcal{C}_\Sigma(u_1, \dots, u_N) &= \Phi_N(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_N)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_N)} \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \exp\left\{-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}\right\} d\mathbf{x}\end{aligned}$$

where $|\Sigma|$ is the determinant of Σ , $\mathbf{x} = (x_1, \dots, x_N)^T$ and Φ^{-1} is the inverse of the univariate standard normal distribution.

If $N = 2$ and $\rho \in (-1, 1)$ is the correlation coefficient, the expression reduces to

$$C_\rho(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right\} dx_1 dx_2.$$

The scatter plot of 200 i.i.d. observations simulated from a bivariate Gaussian copula is reported in Figure 1.1.

Student's t Copula. Let $t_{\nu, \Sigma}$ be the multivariate Student's t distribution with a symmetric positive matrix Σ belonging to $[-1, 1]^{N \times N}$ and ν degrees of freedom. Then, the multivariate Student's t copula is

$$\begin{aligned} C_{\nu, \Sigma}(u_1, \dots, u_N) &= t_{\nu, \Sigma}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_N)) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \dots \int_{-\infty}^{t_\nu^{-1}(u_N)} \frac{\Gamma(\frac{\nu+N}{2}) |\Sigma|^{-1/2}}{\Gamma(\frac{\nu}{2}) (\nu\pi)^{N/2}} \left(1 + \frac{1}{\nu} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right)^{-\frac{\nu+N}{2}} d\mathbf{x} \end{aligned}$$

where t_ν^{-1} is the inverse of the univariate distribution of Student's t with ν degrees of freedom.

If $N = 2$, $\rho \in (-1, 1)$ is the correlation coefficient and $\nu > 2$, the expression reduces to

$$C_{\nu, \rho}(u_1, u_2) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dx_1 dx_2.$$

1.1.3 Archimedean Copulas

The Archimedean copulas are another important class of copulas, firstly named by Ling (1965). They offer more flexibility than the elliptical ones, for example allowing asymmetric tails, and, at least in the bidimensional case, they are easy to build and use.

Archimedean copulas are constructed using a function $\phi : [0, 1] \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ continuous, decreasing, convex and such that $\phi(1) = 0$ called *generator*. It is said strict generator whenever $\phi(0) = +\infty$.

The *pseudo-inverse* $\phi^{[-1]}$ of ϕ is defined as

$$\phi^{[-1]}(u) = \begin{cases} \phi^{-1}(u) & 0 \leq u \leq \phi(0) \\ 0 & \phi(0) \leq u \leq +\infty. \end{cases}$$

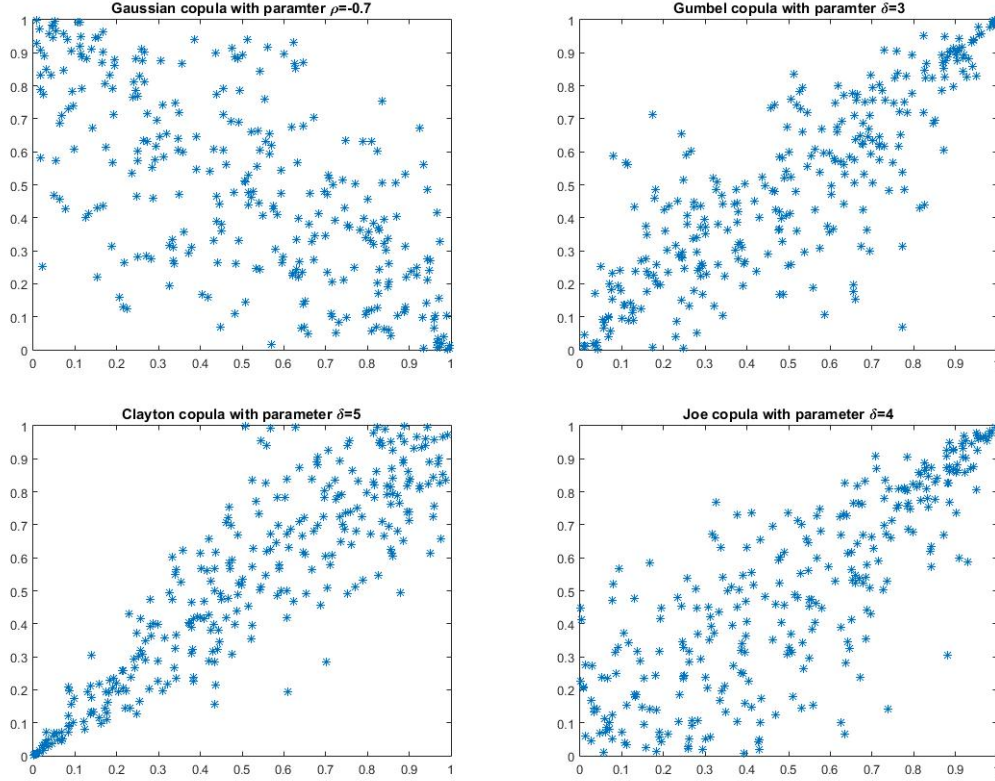
The pseudo-inverse coincides with the usual inverse if ϕ is a strict generator.

Definition 1.6. Given a generator ϕ and its pseudo-inverse, an N -dimensional *Archimedean copula* \mathcal{C} is defined as

$$C_\phi(u_1, \dots, u_N) = \phi^{[-1]}(\phi(u_1) + \dots + \phi(u_N)).$$

The Archimedean copulas are easily related to measures of association such as the Kendall's tau or the tail dependence measures, as shown in Section 1.2.

In the following, we present some commonly used families of one parametric Archimedean bivariate copulas.

Figure 1.1: Scatterplots of $N = 200$ i.i.d. observations from different bivariate copula densities.

Gumbel copula. The generator function for the Gumbel copula is

$$\phi_\delta(x) = (-\ln(x))^\delta$$

with pseudo-inverse function

$$\phi_\delta^{[-1]}(y) = e^{-y^{1/\delta}}$$

with $x, y \in (0, 1)$ and $\delta \in [1, +\infty)$. The 2-dimensional Gumbel copula distribution is given by

$$\mathcal{C}_\delta(u_1, u_2) = \exp \left\{ - [(-\ln u_1)^\delta + (-\ln u_2)^\delta]^{1/\delta} \right\}$$

with density

$$c_\delta(u_1, u_2) = \exp \left(x_1 + x_2 - (x_1^\delta + x_2^\delta)^{1/\delta} \right) (x_1^\delta + x_2^\delta)^{-2+\frac{2}{\delta}} (x_1 x_2)^{\delta-1} \left(1 + (\delta-1)(x_1^\delta + x_2^\delta)^{-\frac{1}{\delta}} \right)$$

where $x_i = (-\log u_i)$ for $i = 1, 2$ and $u_1, u_2 \in [0, 1]$. The scatter plot of 200 i.i.d. observations simulated from a bivariate Gumbel copula is reported in Figure 1.1.

Clayton copula. The generator function for the Clayton copula is

$$\phi_\delta(x) = \frac{1}{\delta} (x^{-\delta} - 1)$$

with pseudo-inverse function

$$\phi_{\delta}^{[-1]}(y) = (1 + \delta y)^{-1/\delta},$$

$x, y \in (0, 1)$ and $\delta \in (0, +\infty)$. The 2-dimensional Clayton copula distribution is equal to

$$\mathcal{C}_{\delta}(u_1, u_2) = \left[\max(u_1^{-\delta} + u_2^{-\delta}, 0) \right]^{-1/\delta}$$

with density

$$c_{\delta}(u_1, u_2) = (1 + \delta)(u_1 u_2)^{-1-\delta} (u_1^{-\delta} + u_2^{-\delta} - 1)^{-\frac{1}{\delta}-2}$$

$u_1, u_2 \in [0, 1]$. The scatter plot of 200 i.i.d. observations simulated from a bivariate Clayton copula is reported in Figure 1.1.

Joe copula. The generator function for the Joe copula is

$$\phi_{\delta}(x) = -\ln(1 - (1 - x)^{\delta})$$

with pseudo-inverse function

$$\phi_{\delta}^{[-1]}(y) = 1 - (1 - \exp(-y))^{1/\delta},$$

$x, y \in (0, 1)$ and $\delta \in [1, +\infty)$. The 2-dimensional Joe copula distribution is

$$\mathcal{C}_{\delta}(u_1, u_2) = 1 - \left[(1 - u)^{\delta} + (1 - v)^{\delta} - (1 - u)^{\delta}(1 - v)^{\delta} \right]^{1/\delta}$$

with density

$$c_{\delta}(u_1, u_2) = \left[x_1^{\delta} + x_2^{\delta} - x_1^{\delta} x_2^{\delta} \right]^{\frac{1}{\delta}-2} x_1^{\delta-1} x_2^{\delta-1} \left[\delta - 1 + x_1^{\delta} + x_2^{\delta} - x_1^{\delta} x_2^{\delta} \right]$$

where $x_i = (1 - u_i)$ for $i = 1, 2$ and $u_1, u_2 \in [0, 1]$. The scatter plot of 200 i.i.d. observations simulated from a bivariate Joe copula is reported in Figure 1.1.

1.2 Dependence Measures

Copulas provide a natural way to study and measure dependence between random variables. Indeed, as shown in Theorem 1.4, a copula is invariant under strictly increasing transformations while the margins may be changed. This suggests that a copula captures the property of a joint distribution which are invariant under strictly increasing transformations.

In this section we introduce important dependence measures and the main relations with the copula functions. In particular we focus on the Kendall's tau and the tail dependence measures that are commonly used in applications and that will be also considered in the rest of the thesis.

1.2.1 Kendall's Tau

In this Sections, we present an important measures of association (concordance) known as Kendall's tau and its relation with copula functions.

Definition 1.7. The dependence measure *Kendall's tau* of the random vector (X_1, X_2) with joint distribution function F is defined as

$$\tau_{X_1, X_2} = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x_1, x_2) dF(x_1, x_2) - 1.$$

One can shown that the Kendall's tau is simply the probability of concordance minus the probability of discordance, that is

$$\tau = \tau_{X_1, X_2} = \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0),$$

where (X'_1, X'_2) is an independent and identically distributed (i.i.d.) copy of (X_1, X_2) ; see e.g. Kruskal (1958).

The empirical version of the Kendall's tau is defined in terms of concordance as follows. Let $\{(x_{1,i}, x_{2,i}), i = 1, \dots, n\}$ a random sample of n observations from a random vector (X_1, X_2) . Then the *empirical Kendall's tau* for the sample is given by

$$\tau_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}((x_{1,i} - x_{1,j})(x_{2,i} - x_{2,j})).$$

Equivalently, τ_n is the fraction of the number of concordant pairs minus the number of discordant pairs of the sample over the total number of possible pairs.

The following theorem establishes a connection between a copula and the Kendall's tau, and provides a useful formula to calculate its values for some copula families.

Theorem 1.8. *Let (X_1, X_2) be a vector with absolutely continuous copula \mathcal{C} . Then the Kendall's tau for (X_1, X_2) is given by*

$$\tau_{X_1, X_2} = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1.$$

The proof follows immediately by a change of variable from Definition 1.7, see Schweizer and Wolf (1981) or Theorem 5.1.3 in Nelsen (2006).

Corollary 1.9. *Let (X_1, X_2) be a vector with absolutely continuous Archimedean copula \mathcal{C} generated by ϕ . Then the Kendall's tau is given by*

$$\tau_{X_1, X_2} = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

For a proof see e.g. Theorem 5.1.4 in Nelsen (2006).

In the following, as an example, we report the Kendall's tau for some specific copula families.

Example 1.1. If \mathcal{C} is a bivariate Gaussian \mathcal{C}_ρ or a Student's t copula $\mathcal{C}_{\nu,\rho}$, then the Kendall's tau is

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

If \mathcal{C} is a Gumbel copula \mathcal{C}_δ , its Kendall's tau is

$$\tau_\delta = 1 - \frac{1}{\delta};$$

if \mathcal{C}_δ is a Clayton copula then

$$\tau_\delta = \frac{\delta}{\delta + 2};$$

and in case of a Joe copula the Kendall's tau is

$$\tau_\delta = 1 - 4 \sum_{k=1}^{\infty} \frac{1}{(k(\delta k + 2)(\delta(k-1) + 2))}.$$

1.2.2 Tail dependence

Tail dependence coefficients are important in modeling dependence of extreme events. Roughly speaking they represent the probability that a random variable exceeds a certain threshold given that another random variable has already exceeded that threshold. More formally, the upper and lower tail dependence coefficients are defined as follows.

Definition 1.10. Let (X_1, X_2) a random vector with marginal distribution functions F_1 and F_2 , respectively. The *upper tail dependence coefficient* of (X_1, X_2) is defined as:

$$\lambda^U = \lim_{t \rightarrow 1^-} \mathbb{P}[X_2 > F_2^{-1}(t) | X_1 > F_1^{-1}(t)].$$

Analogously, the *lower tail dependence coefficient* of (X_1, X_2) is defined as:

$$\lambda^L = \lim_{t \rightarrow 0^+} \mathbb{P}[X_2 \leq F_2^{-1}(t) | X_1 \leq F_1^{-1}(t)].$$

These parameters, as the Kendall's tau, depend only on the copula of X_1 and X_2 as shown in following theorem.

Theorem 1.11. *Let (X_1, X_2) be a vector with absolutely continuous copula \mathcal{C} . If the upper and lower tail coefficient exist, then they are given by*

$$\lambda^U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - \mathcal{C}(t, t)}{1 - t}$$

$$\lambda^L = \lim_{t \rightarrow 0^+} \frac{\mathcal{C}(t, t)}{t}.$$

For a proof of the theorem see Nelsen (2006), Theorem 5.4.2.

Corollary 1.12. *Let (X_1, X_2) be a vector with absolutely continuous Archimedean copula \mathcal{C} generated by ϕ . Then*

$$\begin{aligned}\lambda^U &= 2 - \lim_{t \rightarrow 1^-} \frac{1 - \phi^{[-1]}(2\phi(t))}{1 - t} = 2 - \lim_{x \rightarrow 0^+} \frac{1 - \phi^{[-1]}(2x)}{1 - \phi^{[-1]}(x)} \\ \lambda^L &= \lim_{t \rightarrow 0^+} \frac{\phi^{[-1]}(2\phi(t))}{t} = \lim_{t \rightarrow +\infty} \frac{\phi^{[-1]}(2x)}{\phi^{[-1]}(x)}.\end{aligned}$$

For a proof see Theorem 5.4.3. in Nelsen (2006).

As an example, we report the upper and lower tail measures for some specific copula families.

Example 1.2. If \mathcal{C} is a bivariate Gaussian \mathcal{C}_ρ then the tail coefficients are

$$\lambda^U = \lambda^L = 0,$$

while, for a Student's t copula $\mathcal{C}_{\nu, \rho}$,

$$\lambda^U = \lambda^L = 2t_{\nu+1} \left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right).$$

If \mathcal{C} is a Gumbel copula or a Joe copula \mathcal{C}_δ , then the tail coefficients are

$$\lambda^U = 2 - 2^{1/\delta} \quad \lambda^L = 0;$$

and, in case of Clayton copula, they are equal to

$$\lambda^U = 0 \quad \lambda^L = 2^{-1/\delta}.$$

1.3 Rotated Copulas

In order to allow for more general forms of dependence it is possible to modify existing bivariate copula families using rotations. For example, if a copula \mathcal{C} admits only positive dependence, as the Archimedean copulas presented in 1.1.3, the 90° rotated version exhibits only negative dependence. The list of the possible rotations of a bivariate copula density c is given below:

$$\begin{aligned}c^{90}(u_1, u_2) &= c(1 - u_1, u_2) \quad 90^\circ \text{ rotated copula,} \\ c^{180}(u_1, u_2) &= c(1 - u_1, 1 - u_2) \quad 180^\circ \text{ rotated copula,} \\ c^{270}(u_1, u_2) &= c(u_1, 1 - u_2) \quad 270^\circ \text{ rotated copula.}\end{aligned}\tag{1.6}$$

Through the rotations we are able to study both positive and negative dependence. Note that the functions in (1.6) are the densities of the copula distributions defined in Theorem 1.4.

For example, using rotations, we obtain the Double Clayton and Double Gumbel copulas. These copulas combine different rotations of the usual Gumbel and Clayton copula to allow the modeling of negative dependence. In order to define the different rotations it is useful to re-parameterize

these copulas in terms of Kendall's tau. Using the relation between the copula parameter δ and the Kendall' tau τ , shown in Example 1.1, we can work with Gumbel and Clayton copula with parameter τ .

Starting from $c^G(u, v; \tau)$, the Gumbel copula density with Kendall' tau equal to τ , the Double Gumbel copula of first kind of parameter τ is defined by

$$c^{DG_1}(u, v; \tau) = \begin{cases} c^G(u, v; \tau) & \text{for } \tau > 0 \\ c^G(1 - u, v; -\tau) & \text{for } \tau < 0 \end{cases}$$

while the Double Gumbel copula of second kind is

$$c^{DG_2}(u, v; \tau) = \begin{cases} c^G(1 - u, 1 - v; \tau) & \text{for } \tau > 0, \\ c^G(u, 1 - v; -\tau) & \text{for } \tau < 0. \end{cases}$$

An example of Double Gumbel is reported in Figure 1.2.

Analogously, using the Clayton copula $c^C(u, v; \tau)$ re-parameterized with the Kendall' tau, we define the Double Clayton of first kind DC_1 and of second kind DC_2 as

$$c^{DC_1}(u, v; \tau) = \begin{cases} c^C(u, v; \tau) & \text{for } \tau > 0 \\ c^C(1 - u, v; -\tau) & \text{for } \tau < 0 \end{cases}$$

$$c^{DC_2}(u, v; \tau) = \begin{cases} c^C(1 - u, 1 - v; \tau) & \text{for } \tau > 0, \\ c^C(u, 1 - v; -\tau) & \text{for } \tau < 0. \end{cases}$$

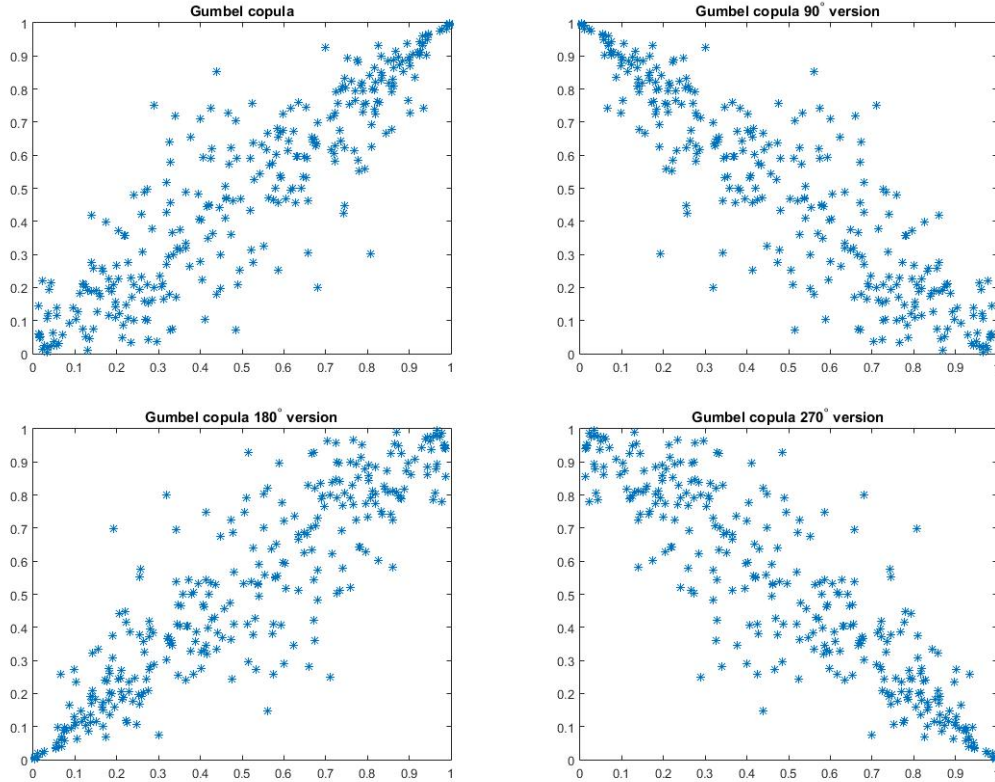
1.4 Copulas and Graphical Models

As already noted, there are many parametric copula families available for the two dimensional case, see e.g. Joe (1997), Nelsen (2006). On the contrary, in the multivariate setting, the use of families different from Normal and Student's t is rather scarce, due to computational and theoretical limitations.

Recently, graphical models have been used to represent a multivariate copula via a set of bivariate ones, simplifying in this way the construction of multivariate copulas. One approach, widely studied in literature, is the Pair Copula Constructions (PCCs), firstly introduced by Joe (1996). Through this construction, the multivariate distribution is represented via a cascade of bivariate conditional copulas, also known as pair copulas or linking copulas. The PCC is order dependent and the factorization is not unique. For this reason, Bedford and Cooke (2001) and Bedford and Cooke (2002) introduced graphical structure called regular vine (R-vine) as pictorial representation of PCCs in order to organize all the possible factorizations. R-vines represent the decomposition of the joint distribution into bivariate components using a nested set of trees. The most commonly used regular vines are the C-vines and D-vines, see Kurowicka and Cooke (2006).

In order to use PCC in a concrete inferential problem, one needs to estimate both all the pair copulas and the particular vine structure. Due to the fact that the number of possible vine structures is increasing very rapidly with the number of variables, the problem of fitting an optimal pair copula

Figure 1.2: Scatterplots of $N = 200$ i.i.d. observations from Double Gumbel copula with parameter $|\tau| = 0.75$.



for a given set of data is an extremely difficult task. This problem arises both in the frequentist and Bayesian framework. For this reason, following e.g. Aas and Berg (2009) and Aas et al. (2009), usually the underlying structure of a vine copula is chosen a priori and, given that structure, the parameters of the linking copulas are estimated in a second step. Examples of a frequentist two step procedure can be found in Liu (2011), Lu et al. (2011), Wu et al. (2012), Jaschke (2014), Arreola (2014), Lai (2014).

Also in the Bayesian framework, the combinatorial complexity of the vine structures turns out to be problematic and leads to the use of suitable Bayesian two step approach where the underlying copula structure is selected a priori, see e.g. Hofmann and Czado (2010), Min and Czado (2010) and Czado et al. (2011). Recently, complete Bayesian model selection procedures for the analysis of regular vines have been proposed in Gruber and Czado (2015a) and Gruber and Czado (2015b). These authors developed a Bayesian approach for vine with structural learning. Unfortunately, due to the nested structure of these kinds of copulas, the algorithms are computationally demanding.

In order to reduce the complexity of the learning procedure and to develop an efficient fully Bayesian approach to estimate all the parameters of the model, we focus on two different copula

structures which are strongly related to vines: the tree copulas and the factor copulas, which are both particular types of truncated regular vines, Kurowicka (2011).

1.4.1 Tree Copula

A tree copula structure combines copula functions and Markov tree network, a particular graphical model, in order to construct multivariate distributions. They were introduced by Kirshner (2007) and extensions to mixture of tree copulas were studied by Kirshner and Smyth (2007) and Silva and Gramacy (2009).

It is well-known that joint probability distributions can be described in term of probabilistic graphical models, where the structure of the graph is used to provide a pictorial representation of the conditional independence relationships between the examined variables.

A graph is a collection of nodes and edges between the nodes; each node in the graph represents a random variable and the edges represent the dependencies between these variables. In order to read conditional independencies from the graph one assumes the global Markov property. According to this property nonadjacent variables on a path are conditionally independent given any set of variables separating them. For more details on graphical models see e.g. Lauritzen (1996).

In this thesis, we consider a Markov tree network (hereafter Markov tree), a particular type of graphical model that has an undirected tree as underlying graph. An undirected tree is defined by a set of nodes $\mathcal{V} = \{1, \dots, N\}$ and a set of edges \mathcal{E} (unordered pair of nodes) with no cycles. Given the set of nodes, the structure of the tree is uniquely defined by the set of edges \mathcal{E} . In the following we identify the structure of a tree only via its edge set \mathcal{E} , and henceforth we denote with \mathcal{E} the tree structure. Over N nodes there are N^{N-2} possible tree structures, and we indicate the space of these tree structures by \mathcal{E}_N .

Let \mathbf{X} be a random vector with multivariate (positive) pdf f on $\mathcal{X} \subset \mathbb{R}^N$ that can be represented by a Markov tree with edge set \mathcal{E} . It follows that its density factorizes as

$$f(x_1, \dots, x_N) = \left[\prod_{(l,m) \in \mathcal{E}} \frac{f_{l,m}(x_l, x_m)}{f_l(x_l)f_m(x_m)} \right] \prod_{k=1}^N f_k(x_k), \quad (1.7)$$

where f_k is the marginal density of X_k and $f_{l,m}$ is the joint density of (X_l, X_m) .

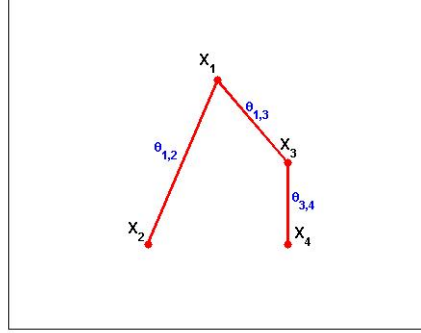
In the tree copula construction of Kirshner (2007) each density $f_{l,m}$ in (1.7) is represented by the corresponding bivariate copula density. More precisely, for every edge $(l, m) \in \mathcal{E}$, using (1.4) and Sklar's theorem, there is a bivariate density copula $c_{l,m}$ such that

$$\frac{f_{l,m}(x_l, x_m)}{f_l(x_l)f_m(x_m)} = c_{l,m}(F_l(x_l), F_m(x_m)), \quad (1.8)$$

where F_l and F_m are the marginal cdfs of X_l and X_m . Hence, using (1.8), equation (1.7) can be rewritten as

$$f(x_1, \dots, x_N) = \prod_{(l,m) \in \mathcal{E}} c_{l,m}(F_l(x_l), F_m(x_m)) \prod_{k=1}^N f_k(x_k),$$

Figure 1.3: Example of tree copula structure on 4 variables.



and the corresponding copula density is $c(u_1, \dots, u_N) = \prod_{(l,m) \in \mathcal{E}} c_{l,m}(u_l, u_m)$.

Conversely, if a bivariate copula $c_{l,m}(u_l, u_m | \theta_{l,m})$, parameterized through a parameter $\theta_{l,m}$, is associated to every edge (l, m) of \mathcal{E} , then the function $c_\theta(u_1, \dots, u_N) = \prod_{(l,m) \in \mathcal{E}} c_{l,m}(u_l, u_m | \theta_{l,m})$ is an admissible copula density, and

$$f_\theta(x_1, \dots, x_N) = \prod_{(l,m) \in \mathcal{E}} c_{l,m}(F_l(x_l), F_m(x_m) | \theta_{l,m}) \prod_{k=1}^N f_k(x_k) \quad (1.9)$$

is an admissible density for \mathbf{X} with margins f_k 's; see Kirshner (2007) for more details.

In the rest of the thesis, we simplify the notation as follows. If m is the parent node of l in the directed version of \mathcal{E} with root node 1, we write $c_{\theta_{l,m}}(u_l, u_m)$ in place of $c_{l,m}(u_l, u_m | \theta_{l,m})$; otherwise if l is the parent node of m we set $c_{l,m}(u_l, u_m | \theta_{l,m}) = c_{\theta_{m,l}}(u_m, u_l)$. Hence, from now on in place of (1.9) we write $\prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}}(F_l(x_l), F_m(x_m)) \prod_{k=1}^N f_k(x_k)$ with the above convention.

An example of tree copula distribution on 4 variables is reported in the following.

Example 1.3. The joint density f of the variables X_1, X_2, X_3, X_4 can be always decompose as

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) f_{2|1}(x_2|x_1) f_{3|1,2}(x_3|x_1, x_2) f_{4|1,2,3}(x_4|x_1, x_2, x_3).$$

If we also assume that f has a Markov tree dependence structure described by the tree copula \mathcal{E} of Figure 1.3, then it further factorizes as

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= f_1(x_1) f_{2|1}(x_2|x_1) f_{3|1}(x_3|x_1) f_{4|3}(x_4|x_3) \\ &= f_1(x_1) \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} \frac{f_{1,3}(x_1, x_3)}{f_1(x_1)} \frac{f_{3,4}(x_3, x_4)}{f_3(x_3)} \\ &= c_{\theta_{1,2}}(F_1(x_1), F_2(x_2)) c_{\theta_{1,3}}(F_1(x_1), F_3(x_3)) c_{\theta_{3,4}}(F_3(x_3), F_4(x_4)) \prod_{k=1}^4 f_k(x_k) \end{aligned} \quad (1.10)$$

where in (1.10) we use the conditional independencies encoded by \mathcal{E} .

1.4.2 Mixture of Tree Copulas

The assumption of Markov tree dependence could be quite restrictive. A possible solution is to consider a copula obtained as a mixture of tree copulas. In fact any convex combinatorial of copula functions is a copula. In this way, we preserve the relative low complexity of the Markov tree structures, taking also into account richer dependencies between the variables.

A mixture of Markov tree copulas is given by

$$c(u_1, \dots, u_N) = \sum_{d=1}^D w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(u_l, u_m),$$

where $D \leq +\infty$ is the number of mixture components, $(w_d)_{d=1, \dots, D}$ are positive weights with $\sum_{d=1}^D w_d = 1$, \mathcal{E}_d ($d = 1, \dots, D$) is the tree structure of the d -th component of the mixture, and $\{\theta_{l,m}^{(d)}\}$ are the copula parameters corresponding to the tree structure \mathcal{E}_d . Consequently, if $\mathbf{X} = (X_1, \dots, X_N)$ is a random vector with density f that is a mixture of tree copulas, then it factorizes as

$$f(x_1, \dots, x_N) = \sum_{d=1}^D w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(F_l(x_l), F_m(x_m)) \prod_{k=1}^N f_k(x_k).$$

Note that the graphical model associated to a mixture of different trees structures cannot be straightforwardly identified. For example, even if two variables X_i and X_j are separated by another variable X_k in every single tree of the mixture, this will not necessary imply that X_i and X_j are separated by X_k in the mixture structure; see Meilă and Jordan (2000) for more details.

In order to obtain a representative graphical structure for a mixture distribution, one can built a complete weighted graph in which the weight of each edge, say (l, m) , is related to some measure of dependence between the variables X_l and X_m . For example, one can consider weights functions of an appropriate pairwise dependence measure such as the Kendall's tau or the correlation coefficient. If one is interested in a representative tree structure associated to the mixture, one can apply the Minimum Spanning Tree (MST) approach to extract a tree from the weighted graph, see e.g. Wang and Xie (2016). Let us recall that the Minimum Spanning Tree is the spanning tree that minimize the sum of edge weights.

It is worth noticing that, in general, from the MST structure we can not read any conditional independence of the variables. In particular there is no connection between this tree and a possible Markov Tree structure of the joint distribution of the variables.

In this thesis we consider two weighted graphs based on different quantities. In the first case, we associate to each edge (l, m) the absolute value of the Kendall's tau between the variables X_l and X_m , obtaining the weighted graph Γ_τ . In the second case, we consider the quantity

$$\Upsilon_{(l,m)}(\mathfrak{E}, \mathbf{w}) = \sum_{d=1}^D w_d \mathbb{1}\{(l, m) \in \mathcal{E}_d\}, \quad (1.11)$$

that represents the global weight associated to edge (l, m) in the mixture with trees $\mathfrak{E} = (\mathcal{E}_1, \dots, \mathcal{E}_D)$ and weights $\mathbf{w} = (w_1, \dots, w_D)$, constructing the weighted graph Γ_w . In order to find the tree structures that maximize the considered quantities, we need to redefine the weights accordingly. More precisely, if one uses the Kendall's tau τ , the weights will be defined as $1 - |\tau|$, while in the other case as $1 - \Upsilon_{(l,m)}$. Starting from these two new weighted graphs, we construct the corresponding MST, that in the following will be denoted by τMST and $wMST$, respectively.

In practice, in order to obtain the previous graphs, one needs to estimate the previous quantities (i.e. the Kendall's taus and the $\Upsilon_{(l,m)}$ s).

1.4.3 Factor Copula

In a factor copula model the dependence among the random variables is explained via one or several latent factors and it is based on bivariate copulas. Factor copulas extend multivariate normal conditional independence models as shown in Joe (2014). The case with one latent factor was originally developed in Joe (2011) and later generalized to arbitrary numbers of factors in Krupskii and Joe (2013). The proposed structured copulas are special cases of truncated vines with latent variables, where the root nodes are represented by the latent factors. Krupskii and Joe (2013) shows that, for some financial return data, the factor copula model presents a better fit than truncated vines in terms of the Akaike or Bayesian information criteria.

Let $\mathbf{X} = (X_1, \dots, X_N)$ be a random vector with multivariate (positive) cdf F on $\mathcal{X} \subset \mathbb{R}^N$ with density f . In a p -factor copula model, we assume that X_1, \dots, X_N are conditional independent given p latent variable V_1, \dots, V_p and that these variables are independent and identically distributed uniformly on $(0, 1)$. It turns out that the joint distribution function F is equal to

$$F(x_1, \dots, x_N) = \int_0^1 \prod_{k=1}^N F_{k|V_1, \dots, V_p}(x_k | v_1, \dots, v_p) dv_p \cdots dv_1 \quad (1.12)$$

where $F_{k|V_1, \dots, V_p}$ denotes the conditional distribution of X_k given V_1, \dots, V_p .

In this thesis we focus on the case with $p = 1$. In the one-factor copula model, we can easily rewrite equation (1.12) in terms of suitable bivariate copula density. In this case, the joint distribution is given by

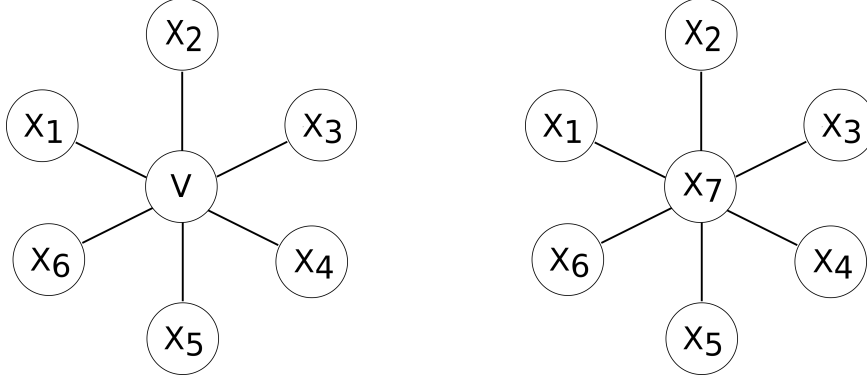
$$F(x_1, \dots, x_N) = \int_0^1 \prod_{k=1}^N F_{k|V}(x_k | v) dv$$

and, since $F_{k|V}(x_k | v) = \frac{\partial}{\partial v} F_{k,V}(x_k, v)$ almost everywhere, the joint pdf is

$$f(x_1, \dots, x_N) = \int_0^1 \prod_{k=1}^N f_{k,V}(x_k, v) dv$$

where $f_{k,V}$ is the bivariate pdf of (X_k, V) .

Figure 1.4: Graphical structure of a one-factor copula with $N=6$ (on the left) and the corresponding truncated C-vine at the first level (on the right).



As in the tree copula structure, to each bivariate density $f_{k,V}$ we can associate a corresponding bivariate copula density $c_{\theta_{k,0}}$ parameterized through a parameters $\theta_{k,0}$. Through the use of the bivariate copulas, we obtain that the joint density is given by

$$f(x_1, \dots, x_N) = \int_0^1 \prod_{k=1}^N (c_{\theta_{k,0}}(F_k(x_k), v) f_k(x_k)) dv.$$

For each copula parameter we use the notation $\theta_{k,0}$ to emphasize the connection with the tree copula. Indeed, the one-factor copula model coincides with a tree copula, i.e. a truncated C-vine at the first level, where the root node is represented by the latent variable. The copula with parameter $\theta_{k,0}$ is associated to an edge of this structure with neighbor node X_k and V . In Figure 1.4 is reported the graphical structure of a one-factor copula with $N = 6$ and the structure of the corresponding truncated C-vine at the first level.

In the factor model, the pictorial representation of the dependence structure among the variable (X_1, \dots, X_N, V) is obtained straightforwardly. On the contrary there is no direct representation of the dependence structure of (X_1, \dots, X_N) . Once again, through the use of the MST approach, we can build the representative tree for the variables of interest.

1.5 Copula and Time Series Models

In many time series applications copula functions are applied to model and study the dependence structure among the variables of interest. In this context, the copula are used to represent the joint density of the innovations, replacing the common assumption of joint normality.

Very broadly, a set of N time series can be represented as

$$X_{t,k} = G_t(\mathcal{O}_{t-1}, \varepsilon_{k,1:t}; \boldsymbol{\alpha}_k) \quad k = 1, \dots, N$$

$$\varepsilon_{t,k} \sim F_{\nu_k}(\cdot) \quad i.i.d.$$

where G_t is a suitable function that depends on the innovations $\varepsilon_{k,1:t} = (\varepsilon_{k,1}, \dots, \varepsilon_{k,t})$, on some parameters α_k , and on \mathcal{O}_{t-1} that denotes the information up to time t , for $t = 1 \dots, T$. The innovations are assumed to be i.i.d. with distribution F_{ν_k} , eventually affected by parameters ν_k and with density f_{ν_k} .

A classical example of time series is the *GARCH*(m, n) model defined as

$$\begin{aligned} X_{k,t} &= \sigma_{k,t} \varepsilon_{k,t} \\ \sigma_{k,t}^2 &= w + \sum_{i=1}^m a_{k,i} X_{k,t-i} + \sum_{j=1}^n b_{k,j} \sigma_{k,t-j}^2 \end{aligned}$$

where $\varepsilon_{k,t} \sim F_{\nu_k}(\cdot)$ i.i.d. In this case, $\alpha_k = (w, \mathbf{a}_k, \mathbf{b}_k)$ with $\mathbf{a}_k = (a_{k,1}, \dots, a_{k,m})$, $\mathbf{b}_k = (b_{k,1}, \dots, b_{k,n})$ and $\mathcal{O}_{t-1} = \{(x_{1,s}, \dots, x_{N,s}) \mid s = 1, \dots, t-1\}$. Another typical example is the *ARMA*(p, q) model described as

$$X_{k,t} = \sum_{i=1}^m a_{k,i} X_{k,t-i} + \sum_{j=1}^n b_{k,j} \varepsilon_{k,t-j} + \varepsilon_{k,t}$$

where the innovations are assumed i.i.d. with marginal distribution F_{ν_k} .

In the copula-based models, the joint density of the innovations is represented through a copula function $c(\cdot; \Phi)$ depending on some parameter Φ . With this assumption, for any fixed t , the joint density of $\varepsilon_t := (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})$ is given by

$$f_{\varepsilon_t}(\varepsilon_{1,t}, \dots, \varepsilon_{N,t} | \Phi, \boldsymbol{\nu}) = c\left(F_{\nu_1}(\varepsilon_{1,t}), \dots, F_{\nu_N}(\varepsilon_{N,t}); \Phi\right) \prod_{k=1}^N f_{\nu_k}(\varepsilon_{k,t}), \quad (1.13)$$

where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N)$ is the collection of the marginal parameters.

1.5.1 Frequentist Inference

As for the inference on the parameters, in literature, most applications of copulas and time series rely on frequentist approach where a two steps estimation procedure is usually applied. The idea is to get approximate i.i.d. residuals and then pseudo copula data to be used in estimating the copula parameter Φ . Under the approach known as Inference Function for Margins (IFM hereafter), in a first step, the parameters $(\hat{\alpha}_k, \hat{\nu}_k)$ of each univariate marginal model are estimated, for example via maximum likelihood or least squares. Using the time series parameters $\hat{\alpha}_k$, the residuals $(\hat{\varepsilon}_{1,t}, \dots, \hat{\varepsilon}_{N,t})$ are obtained and then, applying the estimated marginal cumulative distribution functions (cdfs), the pseudo copula data $\hat{u}_{1,t} = F_{\hat{\nu}_1}(\hat{\varepsilon}_{1,t}), \dots, \hat{u}_{N,t} = F_{\hat{\nu}_N}(\hat{\varepsilon}_{N,t})$ are determined. In a second step, one uses the pseudo copula data to estimate the copula parameters by maximizing the pseudo likelihood function

$$\ell(\Phi | \mathcal{O}_T, \hat{\boldsymbol{\nu}}, \hat{\mathbf{A}}) = \prod_{t=1}^T c(\hat{u}_{1,t}, \dots, \hat{u}_{N,t}; \Phi)$$

where $\hat{\mathbf{A}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_N)$ and $\hat{\boldsymbol{\nu}} = (\hat{\nu}_1, \dots, \hat{\nu}_N)$, see Joe (1997).

Another possible procedure is the Canonical Marginal Likelihood (CML) in which, firstly the parameters $(\hat{\boldsymbol{\alpha}}_1, \dots, \hat{\boldsymbol{\alpha}}_N)$ are estimated and used to obtain the residuals. Secondly, the pseudo copula data $(\hat{u}_{1,t}, \dots, \hat{u}_{N,t})$ are determined through the use of the empirical distribution function \hat{F}_k defined as

$$\hat{F}_k(\cdot) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}(\hat{\varepsilon}_{k,t} \leq \cdot)$$

where $\mathbb{1}(\cdot)$ is the indicator function. Then, through $(\hat{u}_{1,t}, \dots, \hat{u}_{N,t})$, the copula parameters are obtained by maximizing the pseudo copula likelihood, see e.g. Genest et al. (1995).

Examples of these frequentist approaches applied to bivariate time series can be found in Lu et al. (2011), Wen et al. (2012), Wu et al. (2012), Jaschke (2014), Laih (2014); for higher dimensional cases see, e.g., Liu (2011) and Arreola (2014).

1.5.2 Bayesian Inference

As an alternative to the frequentist procedure, a Bayesian approach can be considered. In this case, it is required to specify a prior distribution on the parameters of the model and proceed with Bayesian learning. If we denote with $\pi_0(\cdot)$ the prior density on the parameters $(\Phi, \boldsymbol{\nu}, \mathbf{A})$, through the Bayes' theorem, we obtain the posterior density

$$\pi(\Phi, \boldsymbol{\nu}, \mathbf{A} | \mathcal{O}_T) \propto L(\mathcal{O}_T | \Phi, \boldsymbol{\nu}, \mathbf{A}) \pi_0(\Phi, \boldsymbol{\nu}, \mathbf{A}) \quad (1.14)$$

where $L(\cdot)$ is the likelihood of the observations and $\mathbf{A} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N)$.

In some cases, to assume a prior for all the parameters and proceed with a fully Bayesian approach can be computationally demanding and a two step estimation method, similar to the IFM/CML, can be considered. For example, Min and Czado (2010) applies a similar procedure to GARCH models where the law of the innovations is represented through a pair copula. Firstly, GARCH margins are fitted, and transformed non-parametrically to pseudo copula data; then the parameters of the pair copula, with bivariate t -copulas as building blocks, are estimated. A more complete Bayesian approach is considered by Czado et al. (2011), where fully Bayesian inference in a pair copula with t -copula building blocks and AR(1) margins is studied. A fully Bayesian inference for models with GARCH(1,1) margins with t -innovations coupled with a pair copula has been developed by Hofmann and Czado (2010). It is worth noticing that, in the previous works, the structure of the pair copula is chosen a priori, following Aas and Berg (2009). Recently, Gruber and Czado (2015a) and Gruber and Czado (2015b) propose Bayesian model selection procedures for the analysis of the structure of regular vines. In the first work, marginal DLM time series models are considered and a Bayesian two-step process is applied for estimating the parameters of interest. Firstly, the marginals parameters are computed and then they are used in a sequential Bayesian strategy where the pair copulas are selected. Gruber and Czado (2015b) present a fully Bayesian approach applied to DLM time series models that allows to select the structure of the vine copula.

1.5.3 Bayesian AR-Copula Model

In this thesis, we focus on one of the simplest time series models, the autoregressive one. Let us recall that a time series $(X_t)_{t \geq 1}$ follows an autoregressive model of order p , $AR(p)$, if

$$X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \varepsilon_t \quad \text{for } t = 1, 2, \dots$$

where the innovations $\varepsilon_1, \varepsilon_2, \dots$ are independent and identically distributed.

We shall consider N time series, $(X_{k,t})$, $t = 1, \dots, T$ for $k = 1, \dots, N$, where every series follows an $AR(p)$ model with a fixed p . The k -th component of the panel is given by

$$\begin{aligned} X_{k,t} &= \sum_{i=1}^p \alpha_{k,i} X_{k,t-i} + \varepsilon_{k,t} \\ \varepsilon_{k,t} &\sim F_{\nu_k}, \end{aligned}$$

where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \dots, \alpha_{kp})$ and ν_k are unknown parameters. In the previous expression F_{ν_k} denotes the univariate marginal cdf of the innovations with density function f_{ν_k} . The innovation vectors $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})$ for $t = 1, \dots, T$ are assumed to be independent and identically distributed.

We suppose that the joint density of the innovations is represented through a suitable copula function $c(\cdot; \Phi)$ as in Equation (1.13).

For $t > p$ and for each k , using the transformation

$$\epsilon_{k,t} = x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}, \quad (1.15)$$

we can express the joint conditional density of $\mathbf{X}_t = (X_{1,t}, \dots, X_{N,t})$ given the previous observations $\mathcal{O}_{t-1} = \{(x_{1,s}, \dots, x_{N,s}), s = 1, \dots, t-1\}$, as

$$\begin{aligned} f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t} | \Phi, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1}) &= c \left(F_{\nu_1}(x_{1,t} - \sum_{i=1}^p \alpha_{1,i} x_{1,t-i}), \dots, F_{\nu_N}(x_{N,t} - \sum_{i=1}^p \alpha_{N,i} x_{N,t-i}); \Phi \right) \times \\ &\quad \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right). \end{aligned}$$

Hence, the likelihood of the observations is

$$\begin{aligned} L(\mathcal{O}_T | \Phi, \boldsymbol{\nu}, \mathbf{A}) &= \prod_{t=1}^T c \left(F_{\nu_1}(x_{1,t} - \sum_{i=1}^p \alpha_{1,i} x_{1,t-i}), \dots, F_{\nu_N}(x_{N,t} - \sum_{i=1}^p \alpha_{N,i} x_{N,t-i}); \Phi \right) \times \\ &\quad \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \end{aligned}$$

where the first p observations, denoted by (x_0, \dots, x_{-p+1}) , are set equal to zero for simplicity.

In this thesis, we propose different Bayesian copula-based models to represent the joint density of the innovations. In particular we study the following models: a tree copula model where the dependence structure among the innovations is given by a tree copula (Section 2.1), a finite and an

infinite mixture of tree copula models (Sections 2.3 and 3.2 respectively) and, finally, a one-factor copula model (Section 4.1). One of the advantages of all these models is that they allow us to write the joint density of the innovations in terms of suitable bivariate copulas. In the rest of the thesis, we consider a given set of bivariate copula densities depending on some parameters $\theta_{l,m}$ where $l, m = 1, \dots, N - 1$ for the tree copula models and $l = 1, \dots, N$ and $m = 0$ for the factor copula model. Moreover, we assume that the innovations have univariate Normal marginal distributions with $\nu_k = (\mu_k, p_k)$ mean and precision parameters.

As for the linking copulas, we restrict our attention to Gumbel, Clayton and Joe copulas, presented in Section 1.1.3, that allow us to study upper/lower tail dependence, and that can be easily re-parameterized through these measures, as showed in Example 1.2. For the Gumbel and Joe copulas we use the upper tail parameter, setting $\theta_{l,m} = \lambda_{l,m}^U$, while for the Clayton copula the lower tail parameter, $\theta_{l,m} = \lambda_{l,m}^L$.

One possibility is to study a model in which a single family is fixed for all the linking copulas. As an alternative, one can assume to have different copulas families associated to the edges of the underlying structure. In this case, it can be useful to express the alternative copulas in terms of a common parameter. For this reason, we consider only Gumbel and Clayton copulas rewritten in term of Kendall's tau. We don't use the Joe copulas, since for this type of copula, the relation between the copula parameter δ and the Kendall's tau does not have a closed form expression, see Example 1.1. Finally, since Clayton and Gumbel copulas have only positive correlation, we also consider their rotations in order to capture negative dependence, i.e. we use Double Gumbel and Double Clayton copulas introduced in Section 1.3.

Summarizing, in the following chapters, we assume $\theta_{l,m} = \lambda_{l,m}^{L/U} \in (0, 1)$ if a given family of copulas (Clayton, Gumbel or Joe) is fixed for all the bivariate copulas, while we set $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m}) \in (-1, 1) \times \mathcal{H}$ where $\mathcal{H} = \{DG_1, DG_2, DC_1, DC_2\}$ if each bivariate copula can be a Double Gumbel or a Double Clayton (Combined model).

1.5.4 Bayesian AR-Markov Switching Copula Model

Markov switching models, also called regime switching or hidden Markov models, are time series models which allow for two or more distinct regimes. The regime changes reflect changes in the behavior of time series characterized by different level of dependence, that may be caused by various events, such as economic recession or abrupt variations in government policy. The switching mechanism is controlled by an unobservable state variable that follows a Markov Chain process. Regime-switching models were introduced in econometric by Hamilton (1989) and have been widely applied in economics and finance.

One approach to account for the different levels of dependence is to switch between different copula structures. Examples of regime-switching copula models can be found in Rodriguez (2007),

Okimoto (2008), Cholette et al. (2009), Garcia and Tsafack (2011), Stöber and Czado (2014).

In a regime switching copula model, the dependence among N time series $(X_{k,t})$, $t = 1, \dots, T$, is driven by an hidden latent state variable S_t , that determines the specific copula structure of the innovations.

As usual in the Markov Switching approach, $(S_t)_{t \geq 1}$ is assumed to be a first order Markov chain (MC) in discrete time which takes values in a finite set $\mathcal{R} = \{1, \dots, R\}$. The law of $(S_t)_{t \geq 1}$ is characterized by its transition matrix \mathbf{Q} with elements $\mathbf{q}_{i,j} := \mathbb{P}(S_t = i | S_{t-1} = j)$ for $i, j = 1, \dots, R$ and initial distribution (ρ_1, \dots, ρ_R) .

In an AR-Markov Switching copula model, one deals with a set of N $AR(p)$ time series and assumes that the conditional joint density of the innovations $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})$, for any fixed t , given the regime $S_t = s_t$, is equal to

$$f_{\boldsymbol{\varepsilon}_t}(\varepsilon_{1,t}, \dots, \varepsilon_{N,t} | s_t, \Phi, \boldsymbol{\nu}) = c \left(F_{\nu_1^{(s_t)}}(\varepsilon_{1,t}), \dots, F_{\nu_N^{(s_t)}}(\varepsilon_{N,t}); \Phi^{(s_t)} \right) \prod_{k=1}^N f_{\nu_k^{(s_t)}}(\varepsilon_{k,t}) \quad (1.16)$$

where $\Phi = (\Phi^{(1)}, \dots, \Phi^{(R)})$ collects all the copula parameters and, analogously, $\boldsymbol{\nu} = (\boldsymbol{\nu}^{(1)}, \dots, \boldsymbol{\nu}^{(R)})$ the marginal parameters, with $\boldsymbol{\nu}^{(r)} = (\nu_1^{(r)}, \dots, \nu_N^{(r)})$.

In general, one can assume that also the time series coefficients vary at the changes of regimes. In particular, if we denote with \mathbf{A} the collection of the AR parameters in each regime, i.e. $\mathbf{A} = \{(\boldsymbol{\alpha}_1^r, \dots, \boldsymbol{\alpha}_N^r), r = 1, \dots, R\}$, using the transformation (1.15), we obtained that the joint conditional density of \mathbf{X}_t given the latent chain $S_t = s_t$, the parameters and the previous observations \mathcal{O}_{t-1} is

$$\begin{aligned} f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t} | s_t, \Phi, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1}) &= \\ &= c \left(F_{\nu_1^{(s_t)}} \left(x_{1,t} - \sum_{i=1}^p \alpha_{1,i}^{(s_t)} x_{1,t-i} \right), \dots, F_{\nu_N^{(s_t)}} \left(x_{N,t} - \sum_{i=1}^p \alpha_{N,i}^{(s_t)} x_{N,t-i} \right); \Phi_{s_t} \right) \times \\ &\quad \prod_{k=1}^N f_{\nu_k^{(s_t)}} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i}^{(s_t)} x_{k,t-i} \right). \end{aligned}$$

In this thesis, we focus on regime switching copula model where the dependence structure in each regime is represented via a one factor copula and we suppose that the AR parameters do not depend on the hidden states, see Section 4.3. Also in this case, we consider models in which a single copula family is assumed for all the bivariate copulas (i.e Gumbel, Clayton or Joe copulas re-parameterized through tail measure) or models that allow to have different copula families associated to the linking copulas (i.e. Double Gumbel and Double Clayton re-parameterized through Kendall's tau).

It is important to observe that the Bayesian models we will describe in Sections 2.1, 2.3, 3.2, 4.1 and 4.3 can be easily adapted to other type of copulas and/or marginal distributions.

1.5.5 Model Selection Analysis

In this thesis, we propose different models to represent the joint density of the innovations and hence we face the problem of model selection. Many authors have examined this problem, from both frequentist and Bayesian perspectives, and many tools have been suggested in the literature, see e.g. Burnham and Anderson (2002), Claeskens and Hjort (2008).

Here, we rely on the Bayesian Deviance Information Criterion (*DIC*) introduced by Spiegelhalter et al. (2002), and on the DIC_3 for mixture models, see Celeux et al. (2006) and Richardson (2002).

Given a set of data \mathbf{y} with likelihood function $f(\mathbf{y}|\boldsymbol{\varphi})$ depending on a set of parameters $\boldsymbol{\varphi}$, the *DIC* is defined as

$$DIC = -4\mathbb{E}_{\boldsymbol{\varphi}}[\log f(\mathbf{y}|\boldsymbol{\varphi})|\mathbf{y}] + 2\log f(\mathbf{y}|\tilde{\boldsymbol{\varphi}})$$

where $\tilde{\boldsymbol{\varphi}}$ is a posterior estimate of $\boldsymbol{\varphi}$ (a common choice is the posterior mean).

In mixture model the set of parameters $\boldsymbol{\varphi}$ is not always identifiable, and consequently we cannot obtain $\tilde{\boldsymbol{\varphi}}$ in a straightforward way. For this reason, we replace the term depending on $\tilde{\boldsymbol{\varphi}}$ in equation (1.17) with a function invariant under permutation. A natural choice is to consider an estimator of the density $f(\mathbf{y}|\boldsymbol{\varphi})$, that is invariant under permutation of the parameters. One possibility is to use the posterior expectation $\mathbb{E}_{\boldsymbol{\varphi}}[f(\mathbf{y}|\boldsymbol{\varphi})|\mathbf{y}]$, obtaining in this way

$$DIC_3 = -4\mathbb{E}_{\boldsymbol{\varphi}}[\log f(\mathbf{y}|\boldsymbol{\varphi})|\mathbf{y}] + 2\log \left[\mathbb{E}_{\boldsymbol{\varphi}}[f(\mathbf{y}|\boldsymbol{\varphi})|\mathbf{y}] \right].$$

Lower is the value of the DIC/DIC_3 associated to a specific model, better is the fit of that model.

The *DIC* measure presents some limitations, in particular related to its derivation that lacks theoretical justification (see e.g. Plummer, Robert and Titterton, in discussion of Spiegelhalter et al., 2002). Moreover, since commonly $\tilde{\boldsymbol{\varphi}}$ is equal to the posterior mean, the *DIC* must be used carefully. Indeed, if the examined distribution is not symmetric or not unimodal, the *DIC* may not be appropriate, due to the fact that it assumes the posterior mean as good measure of central location of the distribution. However, the *DIC* is simple to calculate using Markov chain Monte Carlo (MCMC) and therefore, despite the criticisms, widely used in literature.

An alternative Bayesian criterion for choosing between models is the Bayes factor. Given a collection of models $\{\mathcal{M}_1, \dots, \mathcal{M}_L\}$, we can compare the different models using the pairwise Bayes factor. Let $\boldsymbol{\varphi}_r$ denote the parameter vector that characterized \mathcal{M}_r , $\pi_r(\boldsymbol{\varphi}_r|\mathcal{M}_r)$ the prior density and $f(\mathbf{y}|\mathcal{M}_r, \boldsymbol{\varphi}_r)$ the likelihood under the model \mathcal{M}_r ; then for any two models \mathcal{M}_r and \mathcal{M}_s , the Bayes Factor is given by

$$B_{rs} = \frac{m(\mathbf{y}|\mathcal{M}_r)}{m(\mathbf{y}|\mathcal{M}_s)}$$

where

$$m(\mathbf{y}|\mathcal{M}_r) = \int f(\mathbf{y}|\mathcal{M}_r, \boldsymbol{\varphi}_r)\pi_r(d\boldsymbol{\varphi}_r|\mathcal{M}_r)$$

is the marginal likelihood. In order to compute the Bayes factor, one needs to evaluate the marginal likelihood, which is obtained by integrating $f(\mathbf{y}|\mathcal{M}_r, \boldsymbol{\varphi}_r)$ with respect to the prior distribution of the parameters, and not the posterior distribution. For this reason, the MCMC output cannot be used directly to estimate $m(\mathbf{y}|\mathcal{M}_r)$. Alternative methods have been proposed to approximate the marginal likelihood using the MCMC output, see e.g. Chib (1995), Chib and Jeliazkov (2001), Chib et. al (2002), Basu and Chib (2003), Weinberg (2012). All these approaches require non trivial computational work in general and the presence of the copula structure in all our model makes the approximation of the marginal distribution an even more challenging task. For this reason, in the rest of the thesis, we rely on the DIC/DIC₃ criteria for model selection analysis and we leave the use of the Bayes factor for future developments.

Chapter 2

Tree Copula Model and Tree Copula Mixture Model

In this Chapter, we consider the AR-copula model described in Section 1.5.3 and we assume that the joint density of the innovations is given via a tree copula or a finite mixture of tree copulas distributions. We propose a fully Bayesian approach to estimate all the quantities of interest including the underlying tree dependence structure.

The tree copula model is presented in Section 2.1 and the MCMC algorithm for posterior inference is detailed in Section 2.2. The finite mixture of tree copulas model and the corresponding MCMC algorithm for posterior inference are described, respectively, in Sections 2.3 and 2.4. Finally some results with simulated data are shown in Section 2.5.

2.1 Bayesian inference for Tree Copula Model

Let us assume that for any fixed time t , $t = 1, \dots, T$ the dependence structure among the innovations is given by a tree copula distribution with unknown underlying structure \mathcal{E} . With this assumption, Equation (1.13) becomes

$$f_{\boldsymbol{\epsilon}_t}(\epsilon_{1,t}, \dots, \epsilon_{N,t} | \Phi, \boldsymbol{\nu}) = \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \quad (2.1)$$

and via (1.15), we obtain that the joint density of \mathbf{X}_t given \mathcal{O}_{t-1} is

$$f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t} | \Phi, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1}) = \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l} \left(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i} \right), F_{\nu_m} \left(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i} \right) \right) \times \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right).$$

In this model, the parameter Φ indicated both the underlying tree structure \mathcal{E} and the collection of the copula parameters for every possible edge $(l, m) \in \mathcal{E}$, denoted from now on by $\boldsymbol{\theta}$.

We recall that, in our model, every copula $c_{\theta_{l,m}}$ belongs to one of the copula families described in Section 1.5.3 and we suppose univariate Normal marginal pdfs for the innovations with mean μ_k and precision parameter p_k .

In order to proceed with a Bayesian inference, we assign the following independent prior distributions

$$\begin{aligned}\nu_k &= (\mu_k, p_k) \sim \mathcal{NG}(a_k, b_k, m_k, r_k) \\ \boldsymbol{\alpha}_k &\sim \mathcal{N}_p(\mathbf{M}_k, \boldsymbol{\Sigma}_k) \\ \theta_{l,m} &\sim p_{l,m} \\ \mathcal{E} &\sim \Psi(\cdot),\end{aligned}\tag{2.2}$$

for $k = 1, \dots, N$ and $(l, m) \in \mathcal{E}$. In (2.2), $\mathcal{NG}(a, b, m, r)$ is a Normal-Gamma distribution, with density proportional to $p^{a-1} \exp\{-\frac{p}{b}\} \sqrt{pr} \exp\{-\frac{pr}{2}(\mu - m)^2\}$ and $\mathcal{N}_p(\mathbf{M}, \boldsymbol{\Sigma})$ is a p -dimensional Normal distribution with mean \mathbf{M} and covariance matrix $\boldsymbol{\Sigma}$. On the one hand the use of these priors represents the usual choice, since the marginal distributions are assumed to be Normal; on the other hand, even if, due to the presence of the copula structure, this choice does not allow for a closed form of the full conditionals, it still gives more simplifications in the computational steps. If $\theta_{l,m} = \lambda_{l,m}^{U/L}$, then $p_{l,m}$ is a Beta distribution with parameters $(\gamma_{l,m}, \delta_{l,m})$, otherwise $p_{l,m}$ is the product of a translated Beta¹ on $(-1, 1)$ with parameters $(\gamma_{l,m}, \delta_{l,m})$ and an uniform distribution on $\mathcal{H} = \{DG_1, DG_2, DC_1, DC_2\}$ that we will indicate with $Unif_{\mathcal{H}}(\cdot)$. Finally, Ψ is a suitable (discrete) prior over the space $\mathcal{E}_{\mathcal{N}}$.

Regarding Ψ , Meilä and Jaakkola (2006) proposed to use a decomposable prior of the form

$$\Psi(\mathcal{E}|\boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\beta})} \prod_{(l,m) \in \mathcal{E}} \beta_{l,m}$$

where $\boldsymbol{\beta}$ is a symmetric $N \times N$ matrix with non-negative entries and zero on the diagonal. The normalizing constant $Z(\boldsymbol{\beta})$ can be computed explicitly, see Jakkola et al. (1999); moreover there are exact and randomized algorithms for sampling from such prior distribution, see Kirshner and Smyth (2007). In our case, in absence of specific prior information on the dependence structure, we assign a uniform prior on $\mathcal{E}_{\mathcal{N}}$; this is equivalent to setting $\beta_{l,m} = 1$ for every (l, m) .

The joint posterior density (1.14) in this case is equal to

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{E}|\mathcal{O}_T) &\propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \\ &\quad \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \pi_0(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{E}),\end{aligned}$$

¹ X has translated Beta distribution on $(-1, 1)$ with parameters (γ, δ) , if $X = 2Y - 1$ where Y has Beta distribution with parameters (γ, δ) . We denote this distribution with $Beta_{(-1,1)}(\gamma, \delta)$.

where $\pi_0(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{E})$ is the prior density described above.

The previous posterior density cannot be obtained in closed form, hence we propose a suitable MCMC algorithm to approximate posterior quantities of interest.

2.2 MCMC for Tree Copula Model

The MCMC algorithm developed for the analysis of the tree copula model is a Metropolis within Gibbs methods, based on the works of Silva and Gramacy (2009), Gruber and Czado (2015a), and Gruber and Czado (2015b). The algorithm iteratively samples from the full conditionals:

- $\boldsymbol{\nu}$ given $[\mathcal{E}, \boldsymbol{\theta}, \mathbf{A}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- \mathbf{A} given $[\mathcal{E}, \boldsymbol{\theta}, \boldsymbol{\nu}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- $\mathcal{E}, \boldsymbol{\theta}$ given $[\boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T]$ (Metropolis-Hasting step).

In the following we present in details the full conditional densities and the proposal densities required for the algorithm. As we will see, in many Metropolis-Hasting steps we will consider random walk proposals. In these cases, the variances are tuned to achieve acceptance rates between 20% and 80% as proposed in Besag et al. (1995).

Full conditional of $\boldsymbol{\nu}$. We start with the marginal parameters. The full conditional of $\boldsymbol{\nu}$ given $(\mathcal{E}, \boldsymbol{\theta}, \mathbf{A}, \mathcal{O}_T)$ is

$$\pi(\boldsymbol{\nu} | \mathcal{E}, \boldsymbol{\theta}, \mathbf{A}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k)$$

where $\pi_{\nu,k}(\cdot)$ denotes the prior Normal-Gamma $\mathcal{NG}(a_k, b_k, m_k, r_k)$ and $\epsilon_{k,t}$ is given by (1.15). Samples from this full conditional are obtain using a Metropolis-Hasting step, with proposal density

$$q_{\nu}(\boldsymbol{\nu}^* | \boldsymbol{\nu}) = \prod_{k=1}^N q_{\nu,k}(\nu_k^* | \nu_k).$$

Recalling that $\nu_k = (\mu_k, p_k)$, we take

$$q_{\nu,k}(\nu_k^* | \nu_k) = q_{p,k}(p_k^* | p_k) q_{\mu,k}(\mu_k^* | \mu_k),$$

where the proposal density on p_k^* is a gamma density with mean p_k and fixed variance σ_p^2 , while the proposal on μ_k^* is a Normal density with mean μ_k and fixed variance σ_μ^2 .

The acceptance probability of the MH step for $\boldsymbol{\nu}$ is given by

$$\min \left\{ 1, \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} \frac{c_{\theta_{l,m}}(F_{\nu_l^*}(\epsilon_{l,t}), F_{\nu_m^*}(\epsilon_{m,t}))}{c_{\theta_{l,m}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}))} \prod_{k=1}^N \frac{f_{\nu_k^*}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k^*) q_{p,k}(p_k^* | p_k)}{f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k) q_{p,k}(p_k^* | p_k)} \right\}$$

where $q_{\mu,k}(\cdot)$ simplified since it is a symmetric proposal density.

Full conditional of \mathbf{A} . The full conditional of \mathbf{A} given $(\mathcal{E}, \boldsymbol{\theta}, \boldsymbol{\nu}, \mathcal{O}_T)$ is

$$\pi(\mathbf{A}|\mathcal{E}, \boldsymbol{\theta}, \boldsymbol{\nu}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k)$$

where $\pi_{\mathbf{A},k}(\cdot)$ is the p -dimensional Normal prior $\mathcal{N}_p(\mathbf{M}_k, \boldsymbol{\Sigma}_k)$. We proceed again with a Metropolis-Hasting step.

The proposal density $q_{\mathbf{A}}(\mathbf{A}^*|\mathbf{A})$ is a multivariate Normal distribution with mean \mathbf{A} and diagonal covariance matrix. The acceptance probability is

$$\min \left\{ 1, \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} \frac{c_{\theta_{l,m}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i}^* x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i}^* x_{m,t-i}) \right)}{c_{\theta_{l,m}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right)} \times \prod_{k=1}^N \frac{f_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i}^* x_{k,t-i}) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k^*)}{f_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k)} \right\}.$$

Also in this case, the proposal density $q_{\mathbf{A}}(\cdot)$ simplifies since it is a symmetric proposal.

Full conditional of $(\mathcal{E}, \boldsymbol{\theta})$. This is the most delicate step in the algorithm. Samples from the full conditional of $(\mathcal{E}, \boldsymbol{\theta})$ given $(\boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T)$ are obtained again by using a Metropolis-Hasting step. The full conditional is

$$\pi(\mathcal{E}, \boldsymbol{\theta}|\mathbf{A}, \boldsymbol{\nu}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \pi_{\mathcal{E},\boldsymbol{\theta}}(\mathcal{E}, \boldsymbol{\theta})$$

where $\pi_{\mathcal{E},\boldsymbol{\theta}}$ is the prior over $(\mathcal{E}, \boldsymbol{\theta})$ specified in Section 2.1.

First, we sample the new tree \mathcal{E}^* and then, conditionally on the new tree \mathcal{E}^* , we (independently) sample the parameters $\theta_{l,m}$. Formally, the proposal is

$$q_{\mathcal{E},\boldsymbol{\theta}}(\mathcal{E}^*, \boldsymbol{\theta}^*|\mathcal{E}, \boldsymbol{\theta}) = q_{\mathcal{E}}(\mathcal{E}^*|\mathcal{E}) \prod_{(l,m)} q_{\theta_{l,m}}(\theta_{l,m}^*|\mathcal{E}, \theta_{l,m}, \mathcal{E}^*).$$

In order to propose a new tree \mathcal{E}^* , one can use different proposals. For instance one can simply sample from the prior or alternatively, as we have done in the simulations, one can use the local move by Silva and Gramacy (2009). Further details on this local move are given in the next paragraph. As for the copula parameters, we start describing the proposal density for the case $\boldsymbol{\theta} = (\tau, \zeta)$. We distinguish between the edges $(l, m) \notin \mathcal{E}^*$, the edges $(l, m) \in \mathcal{E} \cap \mathcal{E}^*$, and the edges $(l, m) \in (\mathcal{E}^* \setminus \mathcal{E})$. The proposal for the parameters associated to edges $(l, m) \notin \mathcal{E}^*$ is chosen equal to the prior. Note that this parameters do not appear in the copula construction and, since they are not used in the acceptance probability, one need not to sample them. For every $(l, m) \in \mathcal{E} \cap \mathcal{E}^*$ we sample $\tau_{l,m}^*$ (after an appropriate change of variable) with a random walk proposal on \mathbb{R} , while the corresponding

parameter $\zeta_{l,m}$ is left unchanged. Finally, for the parameters corresponding to $(l, m) \in \mathcal{E}^* \setminus \mathcal{E}$, we sample $\tau_{l,m}^*$ (after an appropriate change of variable) from a Normal distribution centered on the (transformed) empirical estimate $\tilde{\tau}_{l,m}$ (based on the residuals $(\epsilon_{l,t}, \epsilon_{m,t})_{t=1,\dots,T}$) of the Kendall's tau and we sample $\zeta_{l,m}^*$ from the discrete distribution on \mathcal{H}

$$q_{\zeta}(\zeta_{l,m}^* = j | \tau_{l,m}^*) = \frac{\prod_{t=1}^T c(\tau_{l,m}^*; j) \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right)}{\sum_{h \in \mathcal{H}} \prod_{t=1}^T c(\tau_{l,m}^*; h) \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right)}$$

for $j \in \mathcal{H}$. The acceptance probability of the MH step is

$$\min \left\{ 1, \prod_{t=1}^T \frac{\prod_{(l,m) \in \mathcal{E}^*} c_{\theta_{l,m}^*} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}^*) q_{\mathcal{E}}(\mathcal{E} | \mathcal{E}^*) \prod_{(l,m) \in \mathcal{E}^*} p_{l,m}(\theta_{l,m}^*)}{\prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}) q_{\mathcal{E}}(\mathcal{E}^* | \mathcal{E}) \prod_{(l,m) \in \mathcal{E}} p_{l,m}(\theta_{l,m})} \times \right. \\ \left. \frac{\prod_{(l,m) \in \mathcal{E}^* \cap \mathcal{E}} q_{\tau}(\tau_{l,m} | \tau_{l,m}^*) \prod_{(l,m) \in \mathcal{E} \setminus \mathcal{E}^*} q_{\zeta}(\zeta_{l,m} | \tau_{l,m}) q_{\tau}(\tau_{l,m} | \tilde{\tau}_{l,m})}{\prod_{(l,m) \in \mathcal{E}^* \cap \mathcal{E}} q_{\tau}(\tau_{l,m}^* | \tau_{l,m}) \prod_{(l,m) \in \mathcal{E}^* \setminus \mathcal{E}} q_{\zeta}(\zeta_{l,m}^* | \tau_{l,m}^*) q_{\tau}(\tau_{l,m}^* | \tilde{\tau}_{l,m})} \right\},$$

where $q_{\tau}(\cdot | \tau_0)$ is the (transformed) Normal proposal centered on τ_0 .

In the alternative case, where $\theta_{l,m} = \lambda_{l,m}^{U/L}$, for the parameters associated to the edges $(l, m) \in \mathcal{E}^* \cap \mathcal{E}$ we sample $\theta_{l,m}$ (after a suitable change of variable) with a random walk proposal on \mathbb{R} . For all the other parameters corresponding to $(l, m) \in (\mathcal{E}^* \cap \mathcal{E})^C$ we take the prior density as proposal. In this case, the MH acceptance probability is equal to

$$\min \left\{ 1, \prod_{t=1}^T \frac{\prod_{(l,m) \in \mathcal{E}^*} c_{\theta_{l,m}^*} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}^*) q_{\mathcal{E}}(\mathcal{E} | \mathcal{E}^*) \prod_{(l,m) \in \mathcal{E}^* \cap \mathcal{E}} p_{l,m}(\theta_{l,m}^*) q_{\theta}(\theta_{l,m} | \theta_{l,m}^*)}{\prod_{(l,m) \in \mathcal{E}} c_{\theta_{l,m}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}) q_{\mathcal{E}}(\mathcal{E}^* | \mathcal{E}) \prod_{(l,m) \in \mathcal{E}^* \cap \mathcal{E}} p_{l,m}(\theta_{l,m}) q_{\theta}(\theta_{l,m}^* | \theta_{l,m})} \right\}$$

where $q_{\theta}(\cdot | \theta_0)$ is the (transformed) random walk proposal centered on θ_0 .

Random walk proposal for sampling trees: the tree-angular proposal. Sampling a tree uniformly or from the prior distribution can frequently produce a tree with very low acceptance probability. For this reason, we use the local proposal introduced in Silva and Gramacy (2009). The move leaves the tree unchanged with the exception of a path say $i - j - k$ which is changed to $j - i - k$. It is possible to traverse the whole space of spanning trees with sequences of tree-angular moves, see Proposition 1, as proved in Silva and Gramacy (2009). The tree-angular proposal, say $q_{\mathcal{E}}^{TA}(\cdot | \mathcal{E})$, given a tree \mathcal{E} , propose a new tree \mathcal{E}^* as follows:

1. choose an edge (u, v) in \mathcal{E} (with some probability);
2. choose a neighbor node w of (u, v) in the tree \mathcal{E} , which is neither u nor v , uniformly at random;
3. return a new tree \mathcal{E}^* that results from removing from \mathcal{E} the edge (u, v) and adding the edge (z, w) , where $z = u$ if v and w are adjacent in \mathcal{E} and otherwise $z = v$.

In order to improve the mixing of the chain for the copula parameters, we choose

$$q_{\mathcal{E}}(\mathcal{E}^*|\mathcal{E}) = \eta\delta_{\mathcal{E}}(\mathcal{E}^*) + (1 - \eta)q_{\mathcal{E}}^{TA}(\mathcal{E}^*|\mathcal{E})$$

with some fixed probability η .

2.3 Bayesian Inference for Finite Tree Copula Mixture Model

We now consider the case in which the joint distribution of the innovations is represented via a finite mixture of D tree copulas. In this model, the pdf of the innovations is equal to

$$f_{\boldsymbol{\varepsilon}_t}(\varepsilon_{1,t}, \dots, \varepsilon_{N,t}|\Phi, \boldsymbol{\nu}, \mathbf{w}) = \sum_{d=1}^D w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(F_{\nu_l}(\varepsilon_{l,t}), F_{\nu_m}(\varepsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\varepsilon_{k,t}).$$

We denote with $\mathbf{w} = (w_1, \dots, w_D)$ the vector of weights, and with \mathcal{E}_d and $\boldsymbol{\theta}_d = \{\theta_{l,m}^{(d)}\}$, respectively, the tree copula structure and the vector of copula parameters for the d -th component. Finally, we set $\boldsymbol{\mathcal{E}} = \{\mathcal{E}_1, \dots, \mathcal{E}_D\}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D\}$. In this model the parameter Φ correspond to $(\boldsymbol{\mathcal{E}}, \boldsymbol{\Theta})$, the collection of all the tree structures and of all the copula parameters.

The conditional density of the observations at time t given \mathcal{O}_{t-1} is obtained using transformation (1.15) and is given by

$$\begin{aligned} f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t}|\Phi, \boldsymbol{\nu}, \mathbf{w}, \mathbf{A}, \mathcal{O}_{t-1}) &= \\ &= \sum_{d=1}^D w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \\ &\quad \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right). \end{aligned} \tag{2.3}$$

In this model, we assign the following independent prior densities

$$\begin{aligned} \mathbf{w} &\sim \text{Dir}(\psi_1, \dots, \psi_D) \\ \nu_k &= (\mu_k, p_k) \sim \mathcal{NG}(a_k, b_k, m_k, r_k) \\ \boldsymbol{\alpha}_k &\sim \mathcal{N}_p(\mathbf{M}_k, \boldsymbol{\Sigma}_k) \\ \theta_{l,m}^{(d)} &\sim p_{l,m} \\ \mathcal{E}_d &\sim \Psi(\cdot) \end{aligned} \tag{2.4}$$

for $k = 1, \dots, N$ and $d = 1, \dots, D$. $\text{Dir}(\psi_1, \dots, \psi_D)$ is a Dirichlet distribution with density proportional to $\prod_{d=1}^{D-1} w_d^{\psi_d - 1} (1 - \sum_{j=1}^{D-1} w_j)$, with $w_D = 1 - \sum_{d=1}^{D-1} w_d$, $\Psi(\cdot)$ is a uniform prior on \mathcal{E}_N and, $p_{l,m}$ is the same as in Section 2.1. Also in this case, the choice of the prior on the marginal and AR parameters leads to a simplification on the computational steps, while the assumption of a Dirichlet priors on the weights allows to obtain the corresponding full conditional in closed form.

The posterior distribution of the finite mixture can be easily obtained from (2.3). In order to sample from it we adopt a data augmentation approach and introduce an allocation variable for each observation, $I_t \in \{1, \dots, D\}$ for $t = 1, \dots, T$. The complete data likelihood becomes

$$f(\mathcal{O}_T, \mathbf{I} | \Phi, \boldsymbol{\nu}, \mathbf{w}, \mathbf{A}) = \prod_{t=1}^T w_{I_t} \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \quad (2.5)$$

where $\mathbf{I} = (I_1, \dots, I_T)$. The likelihood of the finite mixture model can be easily obtained as marginal distribution of (2.5). The posterior density of the allocation variable \mathbf{I} and the parameters given the observations \mathcal{O}_T is given by

$$\pi(\boldsymbol{\mathfrak{E}}, \Theta, \boldsymbol{\nu}, \mathbf{A}, \mathbf{w}, \mathbf{I} | \mathcal{O}_T) \propto \prod_{t=1}^T w_{I_t} \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \pi_0(\boldsymbol{\mathfrak{E}}, \Theta, \boldsymbol{\nu}, \mathbf{A}, \mathbf{w})$$

where $\pi_0(\cdot)$ is the prior density described in (2.4).

Also in this case, the posterior density cannot be obtained in closed form; hence, for posterior inference, we adopt the Metropolis within Gibbs algorithm described in the next Section.

2.4 MCMC for Mixture of Tree Copulas Model

In case of mixture of tree copulas model, the Metropolis within Gibbs algorithm samples iteratively from the full conditionals:

- \mathbf{w} given $[(\boldsymbol{\mathfrak{E}}, \Theta), \mathbf{A}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I}]$ (closed form);
- $\boldsymbol{\nu}$ given $[(\boldsymbol{\mathfrak{E}}, \Theta), \mathbf{A}, \mathbf{w}, \mathcal{O}_T, \mathbf{I}]$ (Metropolis-Hasting step);
- \mathbf{A} given $[(\boldsymbol{\mathfrak{E}}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I}]$ (Metropolis-Hasting step);
- $(\boldsymbol{\mathfrak{E}}, \Theta)$ given $[\mathbf{w}, \mathbf{A}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I}]$ (Metropolis-Hasting step);
- \mathbf{I} given $[(\boldsymbol{\mathfrak{E}}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T]$ (closed form).

The variance of the random walk proposal densities are tuned to achieve acceptance rates between 20% and 80%. In the following, we present in details the steps of the algorithm.

Full conditional of \mathbf{w} . The full conditional of \mathbf{w} given $((\mathfrak{E}, \Theta), \mathbf{A}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I})$ is

$$\begin{aligned}\pi(\mathbf{w} | ((\mathfrak{E}, \Theta), \mathbf{A}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I})) &\propto \prod_{t=1}^T w_{I_t} \pi_{\mathbf{w}}(\mathbf{w}) \\ &= \prod_{d=1}^D w_d^{N_d(I)} \pi_{\mathbf{w}}(\mathbf{w})\end{aligned}$$

where $\pi_{\mathbf{w}}$ is the Dirichlet prior over \mathbf{w} and $N_d(I) = \#\{I_t = d\}$ is the number of the element I_t equal to d .

This full conditional is in closed form, since it turns out to be a $Dir(\psi_1 + N_1(I), \dots, \psi_D + N_D(I))$ and we can sample directly from it.

Full conditional of $\boldsymbol{\nu}$. The full conditional of $\boldsymbol{\nu}$ given $((\mathfrak{E}, \Theta), \mathbf{A}, \mathbf{w}, \mathcal{O}_T, \mathbf{I})$ is proportional to

$$\pi(\boldsymbol{\nu} | ((\mathfrak{E}, \Theta), \mathbf{A}, \mathcal{O}_T, \mathbf{I})) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k)$$

where $\pi_{\nu,k}$ is the prior over ν_k . We sample $\boldsymbol{\nu}$ using a MH step, where we sample every ν_k^* from the proposal density

$$q_{\nu,k}(\nu_k^* | \nu_k) = q_{p,k}(p_k^* | p_k) q_{\mu,k}(\mu_k^* | \mu_k).$$

The proposal $q_{p,k}$ and $q_{\mu,k}$ are chosen, respectively, equal to a gamma density with mean p_k and variance σ_p^2 , and a Normal density with mean μ_k and variance σ_μ^2 . The acceptance probability is

$$\min \left\{ 1, \frac{\prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l^*}(\epsilon_{l,t}), F_{\nu_m^*}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k^*}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k^*) q_{p,k}(p_k | p_k^*)}{\prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k) q_{p,k}(p_k^* | p_k)} \right\}.$$

Full conditional of \mathbf{A} . We sample \mathbf{A} using a MH step. The full conditional of \mathbf{A} given $((\mathfrak{E}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I})$ is

$$\begin{aligned}\pi(\mathbf{A} | ((\mathfrak{E}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathcal{O}_T, \mathbf{I})) &\propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \\ &\quad \prod_{t=1}^T f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k)\end{aligned}$$

where $\pi_{\mathbf{A},k}(\cdot)$ is the p-dimensional Normal prior $\mathcal{N}_p(\mathbf{M}_k, \boldsymbol{\Sigma}_k)$. We sample every $\boldsymbol{\alpha}_k^*$ from the proposal density equal to a p-dimensional Normal density with mean $\boldsymbol{\alpha}_k$ and fixed diagonal covariance matrix.

The acceptance probability is

$$\min \left\{ 1, \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_t} \frac{c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i}^* x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i}^* x_{m,t-i}) \right)}{c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right)} \times \right. \\ \left. \prod_{k=1}^N \frac{f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i}^* x_{k,t-i} \right) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k^*)}{f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \pi_{\mathbf{A},k}(\boldsymbol{\alpha}_k)} \right\}.$$

Full conditional of (\mathfrak{E}, Θ) . In this step we analyze each element of the mixture one at a time.

For every $(\mathcal{E}_d, \boldsymbol{\theta}_d)$, $d = 1, \dots, D$ the full conditional is proportional to

$$\pi(\mathcal{E}_d, \boldsymbol{\theta}_d | \mathbf{w}, \mathbf{A}, \boldsymbol{\nu}, \mathcal{I}\mathcal{O}_T) \propto \prod_{t: I_t=d} \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \pi_{\mathcal{E},\boldsymbol{\theta}}(\mathcal{E}_d, \boldsymbol{\theta}_d)$$

where $\pi_{\mathcal{E},\boldsymbol{\theta}}$ is the prior specified in Section 2.3.

We proceed with a MH step. Let us start with the case in which $\boldsymbol{\theta}_d = (\boldsymbol{\tau}_d, \boldsymbol{\zeta}_d)$. For every component $(\mathcal{E}_d, \boldsymbol{\theta}_d)$, we choose the proposal

$$q_{\mathcal{E},\boldsymbol{\theta}}(\mathcal{E}_d^*, \boldsymbol{\theta}_d^* | \mathcal{E}, \boldsymbol{\theta}) = q_{\mathcal{E}}(\mathcal{E}_d^* | \mathcal{E}_d) \prod_{(l,m)} q_{\theta_{l,m}} \left(\theta_{l,m}^{(d)*} | \mathcal{E}_d, \theta_{l,m}^{(d)}, \mathcal{E}_d^* \right).$$

As in the tree copula model, the proposal over the tree structure can be choose equal to the prior or one can use a local move. For the parameters $\boldsymbol{\theta}_d$ we consider the same proposal density describe in the previous algorithm.

The acceptance probability of each MH step in case of Double Gumbel and Double Clayton linking copulas, i.e. $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\zeta})$, is given by

$$\min \left\{ 1, \prod_{t: I_t=d} \frac{\prod_{(l,m) \in \mathcal{E}_d^*} c_{\theta_{l,m}^{(d)*}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}_d^*) q_{\mathcal{E}}(\mathcal{E}_d | \mathcal{E}_d^*) \prod_{(l,m) \in \mathcal{E}_d^*} p_{l,m} \left(\theta_{l,m}^{(d)*} \right)}{\prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}_d) q_{\mathcal{E}}(\mathcal{E}_d^* | \mathcal{E}_d) \prod_{(l,m) \in \mathcal{E}_d} p_{l,m} \left(\theta_{l,m}^{(d)} \right)} \times \right. \\ \left. \frac{\prod_{(l,m) \in \mathcal{E}_d^* \cap \mathcal{E}_d} q_{\boldsymbol{\tau}} \left(\boldsymbol{\tau}_{l,m}^{(d)} | \boldsymbol{\tau}_{l,m}^{(d)*} \right) \prod_{(l,m) \in \mathcal{E}_d \setminus \mathcal{E}_d^*} q_{\boldsymbol{\zeta}} \left(\boldsymbol{\zeta}_{l,m}^{(d)} | \boldsymbol{\tau}_{l,m}^{(d)} \right) q_{\boldsymbol{\tau}} \left(\boldsymbol{\tau}_{l,m}^{(d)} | \tilde{\boldsymbol{\tau}}_{l,m}^{(d)} \right)}{\prod_{(l,m) \in \mathcal{E}_d^* \cap \mathcal{E}_d} q_{\boldsymbol{\tau}} \left(\boldsymbol{\tau}_{l,m}^{(d)*} | \boldsymbol{\tau}_{l,m}^{(d)} \right) \prod_{(l,m) \in \mathcal{E}_d \setminus \mathcal{E}_d^*} q_{\boldsymbol{\zeta}} \left(\boldsymbol{\zeta}_{l,m}^{(d)*} | \boldsymbol{\tau}_{l,m}^{(d)*} \right) q_{\boldsymbol{\tau}} \left(\boldsymbol{\tau}_{l,m}^{(d)*} | \tilde{\boldsymbol{\tau}}_{l,m}^{(d)} \right)} \right\},$$

where $q_{\boldsymbol{\tau}}(\cdot | \boldsymbol{\tau}_0)$ is the (transformed) Normal proposal centered on $\boldsymbol{\tau}_0$.

In the alternative case, where each parameters $\theta_{l,m}$ corresponds to the upper or lower tail parameter, the acceptance probability of the MH step is equal to

$$\min \left\{ 1, \prod_{t=I_t=d} \frac{\prod_{(l,m) \in \mathcal{E}_d^*} c_{\theta_{l,m}^{(d)*}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}_d^*) q_{\mathcal{E}}(\mathcal{E}_d | \mathcal{E}_d^*)}{\prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right) \Psi(\mathcal{E}_d) q_{\mathcal{E}}(\mathcal{E}_d^* | \mathcal{E}_d)} \right. \\ \left. \times \frac{\prod_{(l,m) \in \mathcal{E}_d^* \cap \mathcal{E}_d} p_{l,m} \left(\theta_{l,m}^{(d)*} \right) q_{\boldsymbol{\theta}} \left(\theta_{l,m}^{(d)} | \theta_{l,m}^{(d)*} \right)}{\prod_{(l,m) \in \mathcal{E}_d^* \cap \mathcal{E}_d} p_{l,m} \left(\theta_{l,m}^{(d)} \right) q_{\boldsymbol{\theta}} \left(\theta_{l,m}^{(d)*} | \theta_{l,m}^{(d)} \right)} \right\}$$

where $q_{\boldsymbol{\theta}}(\cdot | \boldsymbol{\theta}_0)$ is the (transformed) random walk proposal centered on $\boldsymbol{\theta}_0$.

Full conditional of I . The step on the allocation variable is in closed form. The full conditional of I given $((\mathcal{E}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T)$ is

$$\pi(I | ((\mathcal{E}, \Theta), \mathbf{w}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T)) \propto \prod_{t=1}^T w_{I_t} \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})).$$

We sample each I_t from a discrete distribution on $\{1, \dots, D\}$ with probability

$$\mathbb{P}\{I_t = d\} = \frac{w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}))}{\sum_{h=1}^D w_h \prod_{(l,m) \in \mathcal{E}_h} c_{\theta_{l,m}^{(h)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}))}$$

for $d = 1, \dots, D$ and $t = 1, \dots, T$.

2.5 Simulation Study

We use simulated data to investigate the performance of the methodologies proposed in Sections 2.1 and 2.3. We first present the case in which all copulas $c_{\theta_{l,m}}$ belong to the same family (Gumbel, Clayton or Joe) and we use the tail parameterization ($\theta_{l,m} = \lambda_{l,m}^{U/L}$).

We examined different alternative scenarios. In the following we describe and present the results regarding only the most interesting ones. In scenarios 1-3, we consider a simulated dataset of $T = 300$ observations obtained from a multivariate AR model with fixed parameters and a specific tree copula distribution. We apply the MCMC algorithm described in Section 2.2 with 70000 iterations and a burn-in of 45000 iterations. In scenario 4, we consider a simulated dataset of 300 observations obtained from a multivariate AR model with fixed parameters, and a mixture of tree copula distributions. We employ the MCMC algorithm described in Section 2.4 with 100000 iterations and a burn-in of 50000. For a tree copula model with five variables and a mixture tree model with five variables and three components, the algorithms required 3 and 6 minutes, respectively, per 10000 iterations on a Intel(R) Core(TM) i7 personal computer. The programs were written in MATLAB. For the choice of the prior settings, we perform sensitivity analysis with different values for the hyper-parameters. The results show that the choice of the prior hyper-parameters does not affect significantly the posterior estimates. For the sake of comparison, for all the scenarios we present the results obtained with two alternative prior settings.

We now describe in more details the examined scenarios. In scenarios 1, we simulate AR-observations using i.i.d. innovations with tree copula distribution based on Gumbel copulas. Then we apply the corresponding Bayesian model of Section 2.1 with the same linking copulas (reparameterized through the upper tail parameters) and prior specification as in (2.2) for the inference exercise. In scenarios 2, we generate the data starting from a tree copula distribution with bivariate Clayton copulas and we apply a tree copula model with linking Clayton copulas. Analogous, in

Table 2.1: Simulation study: Scenario 1. Tree copula model with bivariate Gumbel copulas re-parameterized through upper tail dependence parameters. For each prior settings, we report the posterior mean of the parameters $\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}$ and the MAP tree structure. For the tree, the posterior probability is reported in brackets.

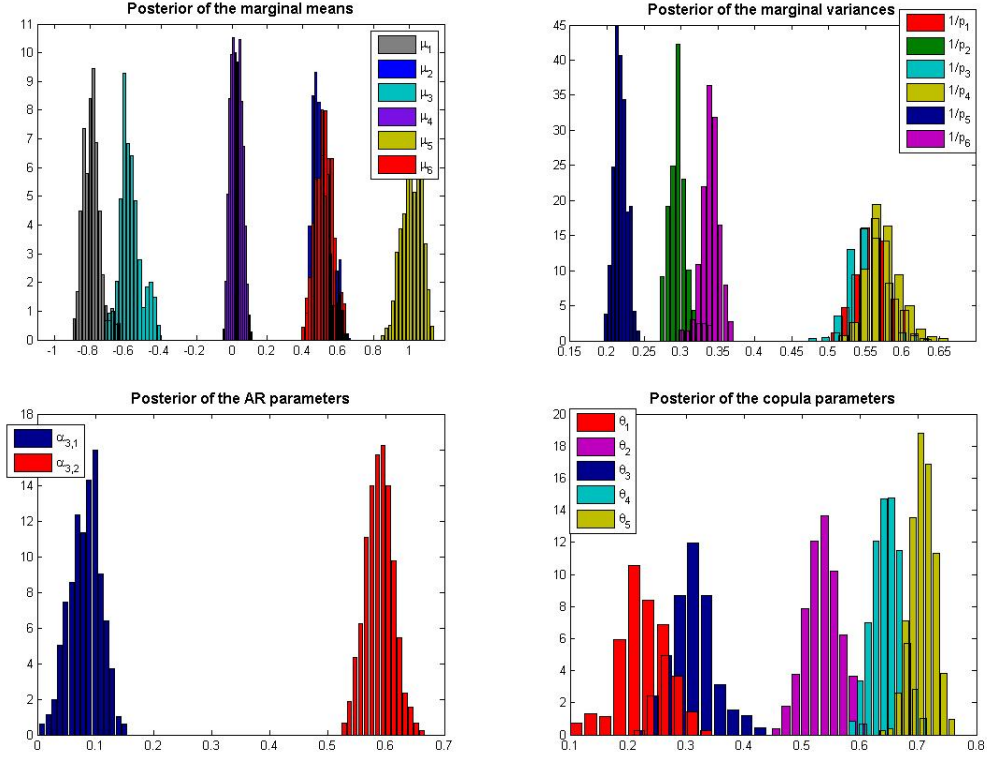
Scenario 1		Prior Settings	
$N = 6$ $p = 2$	$\theta_{i,j}$ (μ_k, p_k) α_k	$Beta(1, 1)$ $\mathcal{NG}(1, 10, 1, 0.1)$ $\mathcal{N}_p(\mathbf{0.4}, 100\mathbf{I}_p)$	$Beta(1, 1)$ $\mathcal{NG}(1, 10, 1, 0.1)$ $\mathcal{N}_p([0.4, 0.3], 70\mathbf{I}_p)$
	True Values	Posterior Means	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(-0.8, 0.5) (0.5, 0.3) (-0.5, 0.5) (0, 0.6) (1, 0.2) (0.5, 0.3)	(-0.68, 0.49) (0.48, 0.31) (-0.42, 0.50) (0.04, 0.59) (0.91, 0.20) (0.47, 0.30)	(-0.82, 0.56) (0.54, 0.29) (-0.53, 0.54) (0.02, 0.57) (0.99, 0.21) (0.51, 0.33)
α_k	(0.3, 0.5) (-0.3, 0.4) (0.1, 0.6) (-0.2, 0.4) (0.1, 0.5) (0.3, 0.5)	(0.35, 0.47) (-0.28, 0.36) (0.13, 0.60) (-0.14, 0.42) (0.11, 0.53) (0.31, 0.49)	(0.30, 0.49) (-0.34, 0.35) (0.08, 0.58) (-0.06, 0.44) (0.13, 0.46) (0.28, 0.50)
$\boldsymbol{\theta}$	0.2 0.5 0.3 0.6 0.7	0.13 0.53 0.33 0.56 0.71	0.20 0.53 0.30 0.64 0.70
\mathcal{E}	[1, 1, 3, 3]*	[1, 1, 3, 3](0.8704)	[1, 1, 3, 3](0.9911)

*Prüfer code corresponding to the tree structure $\{(2, 1), (3, 1), (4, 1), (5, 3), (6, 3)\}$

scenario 3, we simulate from a tree copula distribution with Joe copulas and we consider the corresponding tree copula model for inference. The results are summarized in Tables 2.1, 2.2 and 2.3, respectively. In these tables, we present the posterior estimates obtained with two alternative prior settings. In each table we show the values of the parameters $(\mathcal{E}, \boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A})$ used to generate the data, the prior specifications, the posterior means for $(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A})$ and the maximum a posteriori probability (MAP) tree structure. The trees are denoted by their Prüfer code, see Prüfer (1918).

For each scenario we report the histograms of the simulated values of the posterior distributions of the parameters $\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}$ in Figures 2.1, 2.3 and 2.4 respectively. For the AR parameters we present the histograms only for one AR series; for the others AR coefficients the results are analogous. For all scenarios the posterior estimates are consistent with the parameters used to generate the data. Further experiments of this kind have been performed, obtaining similar results. Concerning the convergence of the algorithm, in Figure 2.2 we report the trace plot and ergodic mean plot for some parameters of scenario 1, after a discarding of 45000 iterations. From these diagnostic plots we can

Figure 2.1: Histograms of the simulated values of the posterior distributions for the parameters of scenario 1 (second prior setting of Table 2.1), where we consider a tree copula model with bivariate Gumbel copulas re-parameterized through upper tail parameters. For the parameters of the AR series, we report only the histograms for the coefficients of the third series, i.e. α_3 .

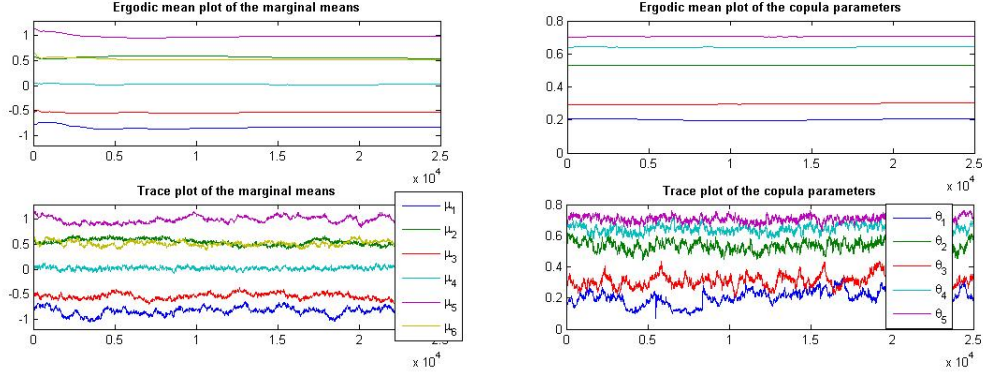


note a good convergence of our tree copula MCMC algorithm. Also the other scenarios show similar results and, hence, the related diagnostic plots are not reported.

In scenario 4, we consider simulated AR-observations starting from a mixture of tree copula distributions with three components and with bivariate Gumbel copulas. We apply our Bayesian mixture tree copula model presented in Section 2.3 with $D = 3$, the same linking copulas and prior specification as in (2.4). In Table 2.4, we present the results obtained with two alternative prior settings. In this table, we show the values of the parameters used to generate the data, the prior settings and, the posterior means for $(\mathbf{w}, \boldsymbol{\nu}, \mathbf{A})$. Also in this case the posterior estimates are consistent with the parameters used to generate the data.

Due to the label switching problem, more attention should be paid to evaluate the posterior tree probabilities for scenario 4. In fact, in mixture models, inference on (functions of) parameters, which are not invariant to parameter permutation, is very delicate if not essentially meaningless. Different solutions to the label switching problem have been proposed in the literature, see e.g. Jasra et al. (2005). One possibility is to identify and work only with statistical quantities that are not affected

Figure 2.2: Diagnostic plots for the marginal parameters of scenario 1 (first prior setting of Table 2.4), where we consider a tree copula model with bivariate Gumbel copulas re-parameterized through upper tail parameters.



by it. Naturally, if the function of interest is invariant to parameter permutations, label switching does not create any problem. Hence, in our case, instead of estimating the posterior probability of the tree appearing in each component, one can simply evaluate the (posterior) probability that a given tree (i.e. a given conditional dependence structure) belongs to the mixture structure. To this end, for a tree \mathcal{E}_0 we can introduce the following permutation invariant function

$$\Psi_{\mathcal{E}_0}(\mathfrak{E}, \mathbf{w}) = \sum_{d=1}^D w_d \mathbb{1}\{\mathcal{E}_0 = \mathcal{E}_d\}. \quad (2.6)$$

Equation (2.6) defines the weight of the tree \mathcal{E}_0 in the mixture and its posterior mean is equal to the posterior probability that the tree \mathcal{E}_0 is contained in the mixture.

In last part of Table 2.4 we list the highest posterior probability tree structures. The trees are sorted in decreasing ordered according to the posterior probability score given in (2.6). More precisely we list the trees up to the one, used to generate the data, with the smallest posterior probability. For scenario 4, in Figure 2.5, we present the histograms of the simulated values of the posterior distributions of the parameters $(\boldsymbol{\nu}, \mathbf{A})$ and in Figure 2.6 the trace plot and ergodic mean plot for some parameters of this scenario.

We also consider model selection analysis. For model comparison, we rely on the DIC and DIC₃ criteria introduced in Section 1.5.5.

For comparative purposes, we estimate the parameters of each scenario using a combination of IFM and Bayesian procedures (shortly B-IFM). In this two steps estimation procedure, the univariate AR models are fitted separately to each marginal (forward-backward approach), and the residuals straightforwardly obtained. In the second step the copula and marginal parameters are estimated using the Bayesian models considered in this Chapter. More precisely, to the simulated data of scenario 1 we apply a B-IFM procedure using a tree copula distribution with linking Gumbel copulas

Table 2.2: Simulation study: Scenario 2. Tree copula model with bivariate Clayton copulas reparameterized through lower tail dependence parameters. For each prior settings, we report the posterior mean of the parameters θ, ν, \mathbf{A} and the MAP tree structure. For the tree, the posterior probability is reported in brackets.

Scenario 2		Prior Settings	
$N = 5$	$\theta_{i,j}$	$Beta(1, 1)$	$Beta(1, 5)$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 10, 1, 0.1)$	$\mathcal{NG}(1, 10, 0, 0.1)$
	α_k	$\mathcal{N}_p(\mathbf{0.4}, 10\mathbf{I}_p)$	$\mathcal{N}_p([0.7, 0.1], 10\mathbf{I}_p)$
	True Values	Posterior Means	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(-0.8, 0.5)	(-0.74, 0.49)	(-0.91, 0.49)
	(0.5, 0.3)	(0.48, 0.29)	(0.46, 0.30)
	(-0.5, 0.5)	(-0.53, 0.50)	(-0.56, 0.48)
	(0, 0.6)	(-0.04, 0.60)	(-0.02, 0.59)
	(0.2, 0.2)	(0.17, 0.20)	(0.21, 0.20)
α_k	(0.3, 0.5)	(0.29, 0.52)	(0.30, 0.46)
	(-0.3, 0.4)	(-0.30, 0.40)	(-0.25, 0.41)
	(0.1, 0.6)	(0.11, 0.58)	(0.09, 0.57)
	(-0.2, 0.4)	(-0.17, 0.45)	(-0.20, 0.39)
	(0.1, 0.7)	(0.12, 0.68)	(0.09, 0.69)
θ	0.7	0.67	0.69
	0.5	0.50	0.50
	0.1	0.15	0.07
	0.6	0.63	0.62
\mathcal{E}	[1, 4, 5]*	[1, 4, 5](0.9926)	[1, 4, 5](0.8110)

*Prüfer code corresponding to the tree structure $\{(2, 1), (3, 5), (4, 1), (5, 4)\}$

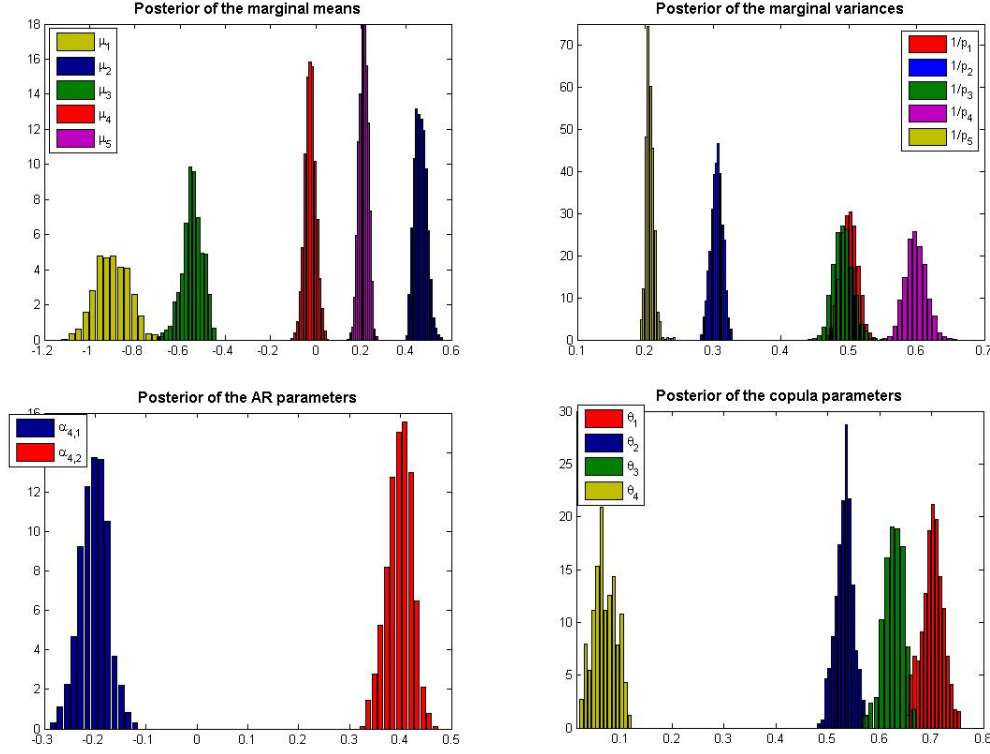
while to the data of scenario 2/scenario 3 a B-IFM procedure with tree copula distribution based on bivariate Clayton/Joe copulas, respectively. Finally, for the data of scenario 4 we consider a B-IFM procedure using a mixture of three tree copula distributions with bivariate Gumbel copulas. For the sake of parsimony, we do not report here all the results obtained via B-IFM, but only the values of the score functions.

Table 2.5 reports the value of the DIC\|DIC₃ of the models estimated by our fully Bayesian procedure using Gumbel, Clayton and Joe tree copula (columns 2-4). Column 5 of the same table reports the DIC\|DIC₃ of the model estimated by the B-IFM procedure using the tree copula corresponding to the generating processes (Gumbel, Clayton, Joe and Gumbel Mixture for rows 2 to 5). As expected, the model from which we generated the data shows the best fits in comparison with the others. Moreover, the models estimated by the fully Bayesian approach have a lower DIC\|DIC₃ with respect to the corresponding models estimated by B-IFM procedure.

In the last two scenarios, we focus on models in which Double Gumbel and Double Clayton copulas with $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m})$ are used as linking copula (Combined models).

In scenario 5 we simulated a dataset of $T = 300$ observations from a multivariate AR model with

Figure 2.3: Histograms of the simulated values of the posterior distributions for the parameters of scenario 2 (second prior setting of Table 2.2), where we consider a tree copula model with bivariate Clayton copulas re-parameterized through lower tail parameters. For the parameters of the AR series, we report only the histograms for the coefficients of the fourth series, i.e. α_4 .



fixed parameters, and a specific tree structure with Double Gumbel or Double Clayton bivariate copulas. We apply the tree copula model with Double Gumbel/Double Clayton as linking copulas and with parameters $\theta = (\tau, \zeta)$. The results obtained with two alternative prior setting are summarized in Table 2.6. In this table, we report the values of the parameters used to generate the data, the two prior settings, the maximum a posteriori probability tree structure, the posterior mode of ζ , and the posterior means of (τ, ν, \mathbf{A}) . The histograms of the simulated values of the posterior distributions for the parameters (τ, ν, \mathbf{A}) are shown in Figure 2.7.

In scenario 6, we consider a dataset of 300 AR-observations generated from a mixture of three tree copula distributions with bivariate Double Gumbel and Double Clayton copulas. We employ the mixture of tree copulas model with same linking copulas and apply the algorithm presented in Section 2.4 with 100000 iterations and a burn-in of 50000. In Table 2.7 are reported the values of the parameters used to generate the simulated data and the corresponding posterior estimates obtained with two alternative prior settings. More precisely, for the parameters $(\mathbf{w}, \nu, \mathbf{A})$ we present the posterior means, while for the underlying graphical structure, we list the trees with the highest

Table 2.3: Simulation study: Scenario 3. Tree copula model with bivariate Joe copulas reparameterized through upper tail dependence parameters. For each prior settings, we report the posterior mean of the parameters $\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}$ and the MAP tree structure. For the tree, the posterior probability is reported in brackets.

Scenario 3		Prior Settings	
$N = 5$	$\theta_{i,j}$	$Beta(2, 2)$	$Beta(1, 2)$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 1, 1, 1)$	$\mathcal{NG}(1, 1, 0.5, 0.1)$
	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.7, 0.1], 100\mathbf{I}_p)$	$\mathcal{N}_p([0.7, 0.1], 10\mathbf{I}_p)$
	True Values	Posterior Means	
$\boldsymbol{\nu}_k = (\mu_k, \frac{1}{p_k})$	$(-0.7, 0.5)$	$(-0.51, 0.52)$	$(-0.62, 0.50)$
	$(0.5, 0.6)$	$(0.50, 0.62)$	$(0.58, 0.62)$
	$(-0.5, 0.5)$	$(-0.39, 0.54)$	$(-0.56, 0.53)$
	$(0, 0.3)$	$(0.02, 0.30)$	$(-0.003, 0.31)$
	$(1.5, 0.2)$	$(1.41, 0.24)$	$(1.43, 0.24)$
$\boldsymbol{\alpha}_k$	$(0.3, 0.5)$	$(0.33, 0.45)$	$(0.30, 0.50)$
	$(-0.3, 0.4)$	$(-0.41, 0.26)$	$(-0.32, 0.38)$
	$(0.1, 0.6)$	$(0.16, 0.56)$	$(0.05, 0.58)$
	$(-0.2, 0.4)$	$(-0.20, 0.32)$	$(-0.11, 0.46)$
	$(0.1, 0.5)$	$(0.16, 0.46)$	$(0.11, 0.49)$
$\boldsymbol{\theta}$	0.5	0.50	0.51
	0.7	0.67	0.73
	0.3	0.36	0.30
	0.6	0.63	0.61
\mathcal{E}	$[1, 1, 3]^*$	$[1, 1, 3](0.993)$	$[1, 1, 3](0.964)$

*Prüfer code corresponding to the tree structure $\{(2, 1), (3, 1), (4, 1), (5, 3)\}$

posterior probability obtained via (2.6). The histograms of the simulated values of the posterior distributions of parameters $(\boldsymbol{\nu}, \mathbf{A})$ are shown in Figure 2.8.

Also for the Combined models we can evaluate the DIC and DIC₃ criteria. In Table 2.8 are reported the values of the DIC and DIC₃ for scenario 5 and 6 computed with our fully Bayesian approach. We also present the DIC/DIC₃ estimated with the B-IFM procedure. We apply a B-IFM procedure using a tree copula with Double Gumbel and Double Clayton copula to the data of scenario 5 and a B-IFM approach with mixture of tree copulas based on Double Gumbel/Double Clayton copulas and with $D = 3$ to the dataset of scenario 6. We note again that the models estimated by the fully Bayesian approach have a lower DIC/DIC₃ with respect to the corresponding models estimated by B-IFM procedure.

Figure 2.4: Histograms of the simulated values of the posterior distributions for the parameters of scenario 3 (second prior setting of Table 2.3), where we consider a tree copula model with bivariate Joe copulas re-parameterized through upper tail parameters. For the parameters of the AR series, we report only the histograms for the coefficients of the fourth series, i.e. α_4 .

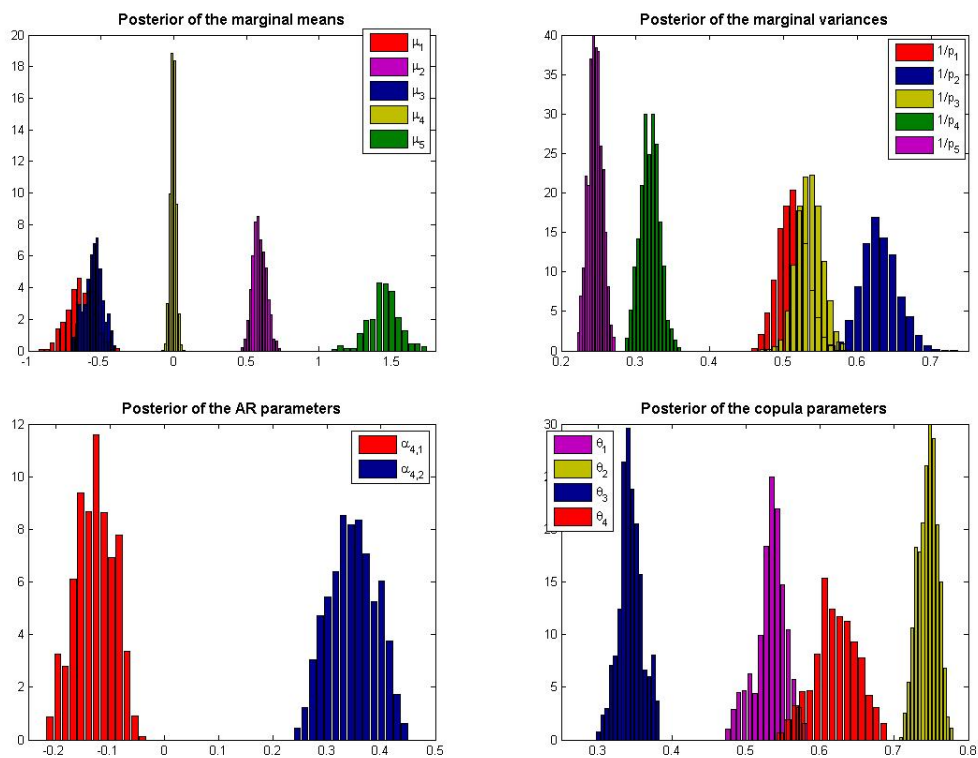


Table 2.4: Simulation study: Scenario 4. Mixture of tree copulas model with bivariate Gumbel copulas re-parameterized through upper tail dependence parameters. For each prior settings, we present the posterior means of $(\boldsymbol{\nu}, \mathbf{A}, \mathbf{w})$. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (2.6) is reported in brackets.

Scenario 4		Prior Settings	
$N = 5$	$\theta_{i,j}$	$Beta(1, 1)$	$Beta(1, 1)$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 1, 1, 0.1)$	$\mathcal{NG}(1, 10, 0, 0.1)$
$D = 3$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p(0.4, 0.1\mathbf{I}_p)$	$\mathcal{N}_p(0.4, 10\mathbf{I}_p)$
	\mathbf{w}	$Dir(10, 10, 10)$	$Dir(10, 10, 10)$
	True Values	Posterior Means	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(1, 0.5)	(1.07, 0.51)	(1.04, 0.53)
	(-1, 0.3)	(-0.91, 0.31)	(-0.89, 0.32)
	(0.5, 0.4)	(-0.45, 0.41)	(0.43, 0.42)
	(1, 0.5)	(1.08, 0.51)	(0.95, 0.50)
	(-0.5, 0.9)	(-0.48, 0.92)	(-0.57, 0.91)
$\boldsymbol{\alpha}_k$	(0.2, 0.6)	(0.20, 0.58)	(0.21, 0.57)
	(-0.3, 0.5)	(-0.26, 0.52)	(-0.28, 0.57)
	(0.1, 0.5)	(0.12, 0.51)	(0.14, 0.49)
	(-0.1, 0.6)	(-0.11, 0.59)	(-0.11, 0.60)
	(0.2, 0.5)	(0.21, 0.49)	(0.17, 0.51)
\mathbf{w}	0.33	0.349	0.329
	0.33	0.333	0.362
	0.33	0.317	0.307
\mathcal{E}	[1, 1, 5]	[4, 5, 5](0.168)	[5, 5, 5](0.185)
	[2, 3, 5]	[2, 3, 5](0.096)	[1, 1, 5](0.091)
	[4, 5, 5]	[5, 5, 5](0.078)	[2, 1, 5](0.060)
		[4, 3, 5](0.060)	[4, 5, 5](0.057)
		[2, 2, 5](0.051)	[2, 3, 5](0.054)
		[1, 2, 5](0.047)	
		[1, 1, 5](0.046)	

Table 2.5: Simulated Data: DIC and DIC₃ Table. The DIC value is reported in brackets.

True Model	Fully Bayesian			B-IFM
	Gumbel	Clayton	Joe	
Scenatio 1 (Gumbel)	899 (913)	1357 (1347)	1614 (1614)	1030 (1039)
Scenario 2 (Clayton)	1509 (1495)	1200 (1197)	2846 (2783)	1327 (1316)
Scenario 3 (Joe)	1962 (1961)	2333 (2264)	1281 (1293)	1828 (1837)
Scenario 4(Mixture Gumbel)	Fully Bayesian			B-IFM
	Mixture Gumbel	Mixture Clayton	Mixture Joe	Mixture B-IFM
	1815	2038	1877	1882

Figure 2.5: Histograms of the simulated values of the posterior distributions for the parameters of scenario 4 (first prior setting of Table 2.4), where we consider a mixture of tree copulas model with bivariate Gumbel copulas re-parameterized through upper tail parameters. For the parameters of the AR series, we report the histograms for the coefficients of the first and fourth series, i.e. α_1 and α_4 .

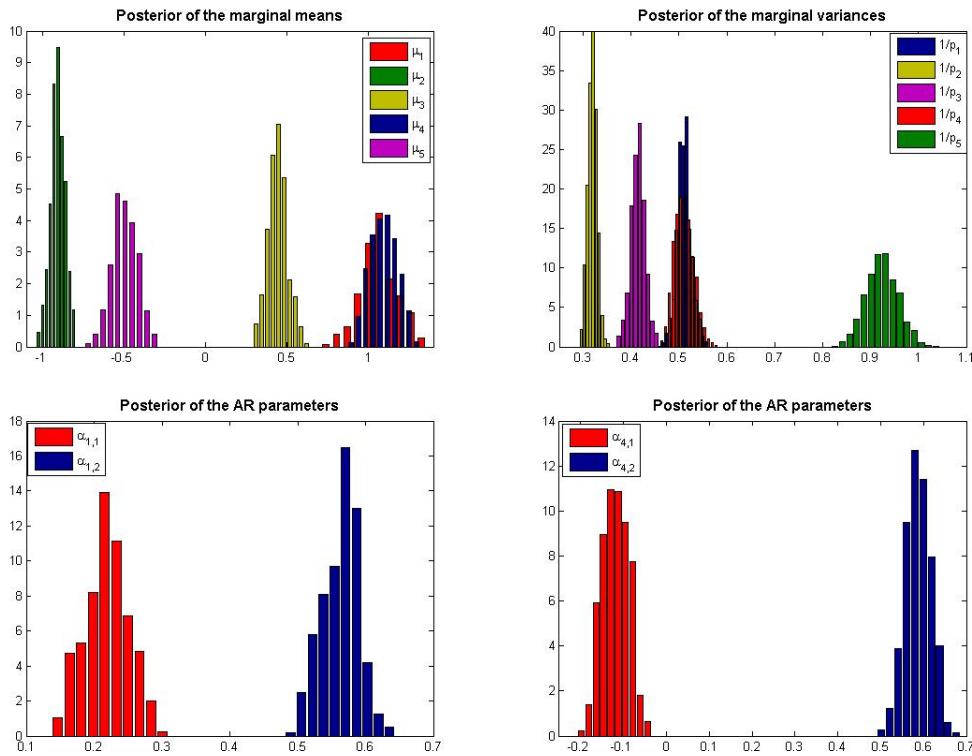


Figure 2.6: Diagnostic plots for the marginal parameters of scenario 4 (first prior setting of Table 2.4), where we consider a mixture of tree copulas model with bivariate Gumbel copulas re-parameterized through upper tail parameters.

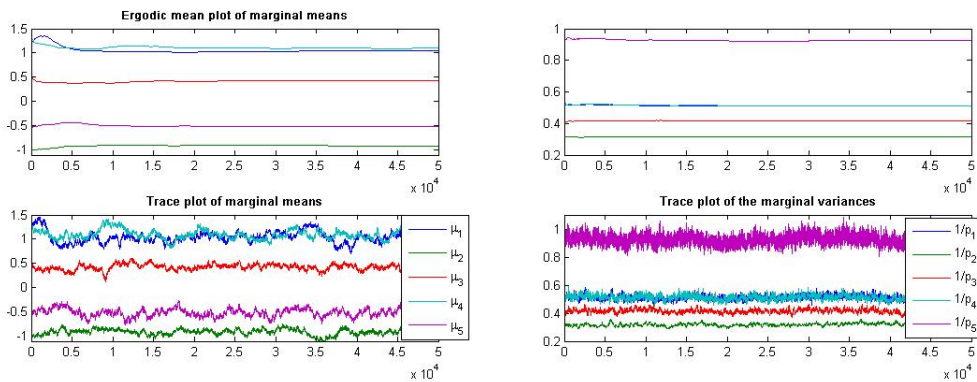


Table 2.6: Simulation study: Scenario 5. Tree copula model with bivariate Double Gumbel or Double Clayton copulas re-parameterized through the Kendall's tau. For each prior settings, we report the posterior mean of the parameters $\boldsymbol{\tau}, \boldsymbol{\nu}, \mathbf{A}$, the posterior mode for $\boldsymbol{\zeta}$ and, the MAP tree structure. For the tree structure, the posterior probability is reported in brackets.

Scenario 5		Prior Settings	
$N = 5$ $p = 2$	$(\pi_{l,m}, \zeta_{l,m})$ (μ_k, p_k) $\boldsymbol{\alpha}_k$	$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H})$ $\mathcal{NG}(1, 0.5, 0, 0.1)$ $\mathcal{N}_p([0.7, 0.1], 100\mathbf{I}_p)$	$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H})$ $\mathcal{NG}(1, 1, 0.5, 0.1)$ $\mathcal{N}_p([0.3, 0.4], 10\mathbf{I}_p)$
	True Values	Posterior Means/Modes	
$\nu_k = (\mu_k, \frac{1}{p_k})$	$(-0.8, 0.5)$ $(0.5, 0.3)$ $(-0.5, 0.4)$ $(0, 0.6)$ $(1, 0.2)$	$(-0.73, 0.49)$ $(0.57, 0.32)$ $(-0.47, 0.43)$ $(-0.04, 0.63)$ $(0.98, 0.22)$	$(-0.74, 0.54)$ $(0.49, 0.33)$ $(-0.58, 0.39)$ $(-0.01, 0.59)$ $(1.04, 0.22)$
$\boldsymbol{\alpha}_k$	$(0.3, 0.5)$ $(-0.3, 0.4)$ $(0.1, 0.6)$ $(-0.2, 0.4)$ $(0.1, 0.5)$	$(0.25, 0.55)$ $(-0.33, 0.39)$ $(0.04, 0.69)$ $(-0.18, 0.41)$ $(0.08, 0.52)$	$(0.32, 0.48)$ $(-0.26, 0.39)$ $(0.14, 0.49)$ $(-0.17, 0.35)$ $(0.12, 0.46)$
$(\boldsymbol{\tau}, \boldsymbol{\zeta})$	$(0.7, DC_1)$ $(-0.2, DG_1)$ $(0.3, DG_2)$ $(-0.6, DG_1)$	$(0.54, DG_2)$ $(-0.23, DG_1)$ $(0.29, DG_2)$ $(-0.59, DG_1)$	$(0.69, DC_1)$ $(-0.25, DG_1)$ $(0.37, DG_2)$ $(-0.61, DG_2)$
\mathcal{E}	$[1, 2, 5]^*$	$[1, 2, 5](0.895)$	$[1, 2, 5](0.892)$

*Prüfer code corresponding to the tree structure $\{(2, 1), (3, 1), (4, 5), (5, 2)\}$

Figure 2.7: Histograms of the simulated values of the posterior distributions for the parameters of scenario 5 (second prior setting of Table 2.6), where we consider a tree copula model with bivariate Double Gumbel and Double Clayton copulas re-parameterized through Kendall's tau parameters. For the parameters of the AR series, we report only the histograms for the coefficients of the first series, i.e. α_1 .

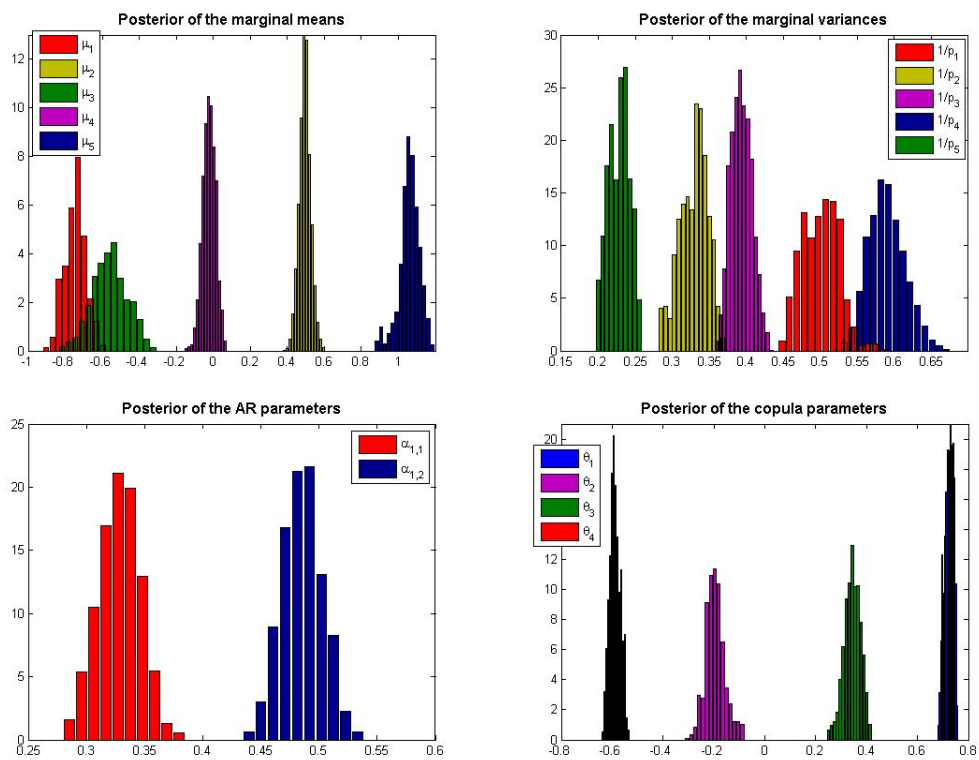


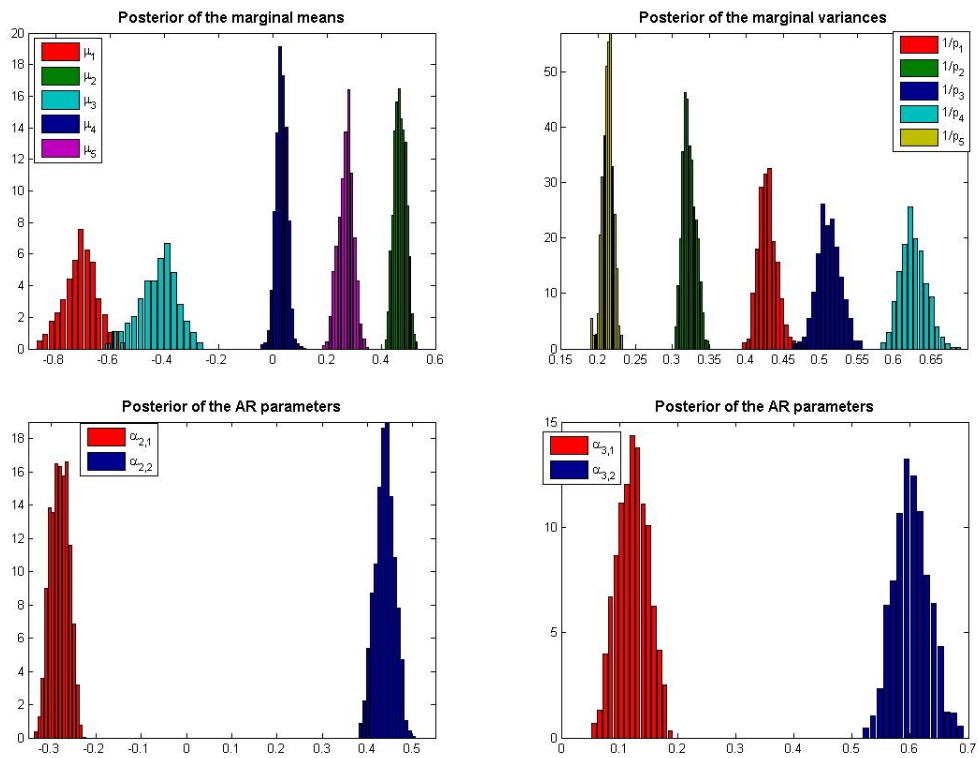
Table 2.7: Simulation study: Scenario 6. Mixture of tree copulas model with Double Gumbel and Double Clayton linking copulas re-parameterized through the Kendall' tau. For each prior settings, we present the posterior means of $(\boldsymbol{\nu}, \mathbf{A}, \mathbf{w})$. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (2.6) is reported in brackets.

Scenario 6		Prior Settings	
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H})$	$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H})$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 0.5, 0.5, 0.1)$	$\mathcal{NG}(1, 1, 0, 0.1)$
$D = 3$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.3, 0.1], 10\mathbf{I}_p)$	$\mathcal{N}_p([0.3, -0.1], 10\mathbf{I}_p)$
	\mathbf{w}	$Dir(5, 5, 5)$	$Dir(10, 10, 10)$
	True Values	Posterior Means	
$\nu_k = (\mu_k, \frac{1}{p_k})$	$(-0.8, 0.4)$	$(-0.82, 0.44)$	$(-0.71, 0.42)$
	$(0.5, 0.3)$	$(0.49, 0.31)$	$(0.46, 0.31)$
	$(-0.5, 0.5)$	$(-0.47, 0.54)$	$(-0.42, 0.51)$
	$(0, 0.6)$	$(-0.01, 0.62)$	$(0.03, 0.62)$
	$(0.3, 0.2)$	$(0.27, 0.21)$	$(0.26, 0.21)$
$\boldsymbol{\alpha}_k$	$(0.3, 0.5)$	$(0.31, 0.46)$	$(0.32, 0.49)$
	$(-0.3, 0.4)$	$(-0.27, 0.32)$	$(-0.28, 0.43)$
	$(0.1, 0.6)$	$(0.06, 0.62)$	$(0.12, 0.60)$
	$(-0.2, 0.4)$	$(-0.16, 0.37)$	$(-0.21, 0.39)$
	$(0.1, 0.5)$	$(0.13, 0.50)$	$(0.17, 0.48)$
\mathbf{w}	0.2	0.11	0.15
	0.3	0.16	0.40
	0.5	0.74	0.45
\mathcal{E}	$[2, 1, 3]$	$[4, 1, 3](\mathbf{0.720})$	$[2, 1, 3](\mathbf{0.2708})$
	$[1, 3, 5]$	$[1, 3, 5](\mathbf{0.116})$	$[1, 3, 5](\mathbf{0.154})$
	$[4, 1, 3]$	$[2, 1, 3](\mathbf{0.04})$	$[4, 1, 3](\mathbf{0.120})$

Table 2.8: Simulated Data: DIC and DIC₃ Table for Combined models. The DIC value is reported in brackets.

	Fully Bayesian	B-IFM
True Model		
Scenario 5 (Combined Tree Copula)	528 (532)	579 (582)
Scenario 6 (Mixture of Combined Tree Copula)	1067	1322

Figure 2.8: Histograms of the simulated values of the posterior distributions for the parameters of scenario 6 (second prior setting of Table 2.7), where we consider a mixture of tree copulas model with bivariate Double Gumbel and Double Clayton copulas re-parameterized through Kendall' tau parameters. For the parameters of the AR series, we report only the histograms for the coefficients of the second and third series, i.e. α_2 and α_3 .



Chapter 3

DP-Tree Copula Model

As an extension of the models presented in Chapter 2, here we consider a Bayesian nonparametric approach based on a Dirichlet process mixture model, named from now on DP-tree copula model. We focus on a mixture of tree copulas where we assume that the number of mixture components is unknown a priori. In this framework we consider a Dirichlet process as prior for the copula parameters and the tree structures.

The Dirichlet process is currently one of the most popular Bayesian nonparametric models. It was first formalized by Ferguson (1973) for general Bayesian statistical modeling, as a prior over distributions with wide support yet tractable posteriors. A common application of the Dirichlet process is in clustering data using mixture models, see e.g. Lo (1984), Neal (1992), Escobar and West (1995). In order to sample from the posterior distribution of a Dirichlet process mixture models, many Markov chain methods have been proposed, see e.g. MacEachern and Müller (1998), Neal (2000), Walker (2007), Kalli et al. (2011), Ishwaran and James (2001). In this work we rely on the slice sampling approach of Walker (2007) and Kalli et al. (2011).

The outline of the Chapter is the following. In Section 3.1 we briefly introduce the Dirichlet process, focusing on its stick breaking construction and on related mixture models. In Section 3.2 is described our DP-tree copula model while in Section 3.3 is illustrated the MCMC algorithm. Finally, a simulation study is reported in Section 3.4.

3.1 Dirichlet Process

Let G_0 be a probability measure on a polish space $(\mathfrak{B}, \mathcal{B})$ and let ψ be a positive real number. A random probability measure G over $(\mathfrak{B}, \mathcal{B})$ has a *Dirichlet process distribution* with concentration parameter ψ and base measure G_0 (in symbol $DP(\psi, G_0)$) if, for any finite measurable partition (B_1, \dots, B_K) of \mathfrak{B} , the random vector $(G(B_1), \dots, G(B_K))$ is distributed as a finite dimensional Dirichlet distribution with parameters $(\psi G_0(B_1), \dots, \psi G_0(B_K))$, i.e.

$$(G(B_1), \dots, G(B_K)) \sim Dir(\psi G_0(B_1), \dots, \psi G_0(B_K)).$$

We write $G \sim DP(\psi, G_0)$ if G has a Dirichlet process distribution, or briefly if G is a Dirichlet process (DP).

The existence of such an object was proved by Ferguson (1973).

3.1.1 Stick Breaking Construction

One property of the Dirichlet process distribution is that measures drawn from it are discrete with probability one. In the stick breaking construction, see Sethuraman (1994), the property of discreteness is made explicit. Through this construction, the Dirichlet process can be represented as

$$G(\cdot) = \sum_{d \geq 1} w_d \delta_{\zeta_d}(\cdot). \quad (3.1)$$

The weights w_d are generated randomly using the stick-breaking construction where $w_1 = v_1$ and

$$w_d = v_d \prod_{l < d} (1 - v_l)$$

with v_d i.i.d. random variables with $Beta(1, \psi)$ distribution. The atoms ζ_d are i.i.d. random variables with distribution G_0 and δ_{ζ_d} is a probability measure concentrated on ζ_d . This stick-breaking construction is so-called because proportions v_d are sequentially broken from the remaining length $\prod_{l < d} (1 - v_l)$ of a unit-length stick.

Sethuraman (1994) showed that G defined in Equation (3.1) is a random probability measure distributed according to $DP(\psi, G_0)$.

An alternative representation of the DP, that we don't report here, is based on a Pòlya urn scheme. The interested reader is referred to Blackwell and MacQueen (1973).

3.1.2 Infinite Mixture Models

One of the uses of the Dirichlet process is as a nonparametric prior on the parameters of a mixture model. In general, in a mixture model the pdf of an observation \mathbf{X} is defined as

$$f(\mathbf{x}|G) = \int_{\mathfrak{B}} K(\mathbf{x}|\zeta)G(d\zeta) \quad (3.2)$$

where $K(\cdot|\zeta)$ is a suitable kernel distribution with parameters ζ and G is a suitable random measure on the parameters space $(\mathfrak{B}, \mathcal{B})$.

If G is a Dirichlet process, we obtain a mixture model with a countably infinite number of components usually named *Dirichlet process mixture model*. In this case, using the stick breaking representation given above, we obtain that the pdf (3.2) can be rewritten as

$$f(\mathbf{x}|\zeta, \mathbf{V}) = \sum_{d \geq 1} w_d K(\mathbf{x}|\zeta_d) \quad (3.3)$$

where $\mathbf{V} = (v_1, v_2, \dots)$, and $\zeta = (\zeta_1, \zeta_2, \dots)$.

The Dirichlet process mixture model can be derived as the limit of a finite mixture model where the number of the components is taken to infinity. More precisely, let us consider a finite mixture model with D components where the measure G can be written as

$$G = G_D = \sum_{d=1}^D w_d \delta_{\zeta_d}$$

with $\sum_{d=1}^D w_d = 1$. If we consider the following model

$$\begin{aligned} \zeta_d | G_0 &\stackrel{\text{iid}}{\sim} G_0 \\ \mathbf{w} = (w_1, \dots, w_D) | \psi &\sim \text{Dir}\left(\frac{\psi}{D}, \dots, \frac{\psi}{D}\right) \end{aligned} \quad (3.4)$$

then, for each measurable function g , integrable with respect to G_0 , we have, as $D \rightarrow +\infty$:

$$\int_{\mathfrak{B}} g(\zeta) G_D(d\zeta) \xrightarrow{\mathcal{D}} \int_{\mathfrak{B}} g(\zeta) G(d\zeta)$$

where $G \sim DP(\psi, G_0)$; see Ishwaran and Zarepour (2002).

In case we model a set of observations $(\mathbf{X}_1, \dots, \mathbf{X}_T)$ via a Dirichlet process mixture model, we have the following hierarchical representation:

$$\begin{aligned} \mathbf{X}_i | \zeta_i &\stackrel{\text{iid}}{\sim} K(\cdot | \zeta_i) \\ \zeta_i | G &\stackrel{\text{iid}}{\sim} G \\ G | \psi, G_0 &\sim DP(\psi, G_0). \end{aligned}$$

Due to the property of discreteness of G , the variables ζ_i can take the same value simultaneously and the above model can be seen as a mixture model in which the observations \mathbf{X}_i with the same values of ζ_i belong to the same cluster. The number of clusters D in the first T observations is random. Antoniak (1974) showed that the (prior) distribution of D in the first T observation is equal to

$$\mathbb{P}(D = d | \psi, T) = \frac{T! \Gamma(\psi)}{\Gamma(\psi + T)} |s_{Td}| \psi^d$$

for $d \geq 1$, where s_{Td} is the signed Stirling number and, $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$ is the Gamma function. The expected number of components is directly affected by the value of the concentration parameter ψ . Indeed the expected number of clusters is

$$\mathbb{E}(D) = \sum_{t=1}^T \frac{\psi}{t-1+\psi}$$

As the number of observations $T \rightarrow +\infty$, this gives

$$\mathbb{E}(D) \simeq \psi \log T. \quad (3.5)$$

For large value of ψ , the probability of the presence of a new component in the mixture increases. The use of DP mixture model allows for the inclusion of the number of components uncertainty and the posterior distribution of D given a sample of data provide information on the unknown number of components.

3.2 Bayesian Inference for DP-Tree Copula Model

Coming back to the AR-copula model, in this Chapter we assume that the joint density of the innovations is represented through a Dirichlet process mixture distribution on the parameters Φ of the copula structure. More precisely, we assume that the pdf of the innovations at time t has the form

$$f_{\epsilon_t}(\epsilon_{1,t}, \dots, \epsilon_{N,t} | G, \nu) = \int K(\epsilon_{1,t}, \dots, \epsilon_{N,t} | \Phi, \nu) G(d\Phi)$$

where $G \sim DP(\psi, G_0)$ and ν has some distribution Q_0 . We also suppose that the kernel density K is represented by a tree copula density, Equation (2.1). Following the stick breaking representation of the Dirichlet process, we can write the joint density of the innovations as

$$f_{\epsilon_t}(\epsilon_{1,t}, \dots, \epsilon_{N,t} | \Phi, \nu, \mathbf{V}) = \sum_{d \geq 1} w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t})$$

where Φ represents the collection of all the copula parameters and tree structures, i.e. $\Phi = ((\theta_1, \mathcal{E}_1), (\theta_2, \mathcal{E}_2), \dots)$. Let us recall that \mathbf{V} denotes the elements (v_1, v_2, \dots) of the stick breaking construction, where $w_1 = v_1$ and $w_d = v_d \prod_{l < d} (1 - v_l)$. As in the previous Chapter, we assume that each bivariate copula $c_{\theta_{l,m}^{(d)}}$ belongs to one of the copula families listed in Section 1.5.3, i.e. Gumbel, Clayton, Joe or Double Gumbel/Double Clayton copulas.

Using transformation (1.15) that involves the AR models, we obtain the conditional joint density of the observations at time t

$$f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t} | \Phi, \nu, \mathbf{V}, \mathbf{A}, \mathcal{O}_{t-1}) = \sum_{d \geq 1} w_d \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l} \left(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i} \right), F_{\nu_m} \left(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i} \right) \right) \times \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right). \quad (3.6)$$

In our model, each ζ_d of Equation (3.3) is equal to $(\theta_d, \mathcal{E}_d)$, therefore G_0 decomposes in the product of a prior on θ_d and a prior on \mathcal{E}_d . We also assume that the concentration parameter ψ of the DP is unknown, as well as \mathbf{A} , the collection of the AR parameters.

Summarizing, assuming independent prior distributions for the unknown parameters, we have the following prior setting

$$\begin{aligned} \nu_k &= (\mu_k, p_k) \sim \mathcal{NG}(a_k, b_k, m_k, r_k) \\ \alpha_k &\sim \mathcal{N}_p(\mathbf{M}_k, \Sigma_k) \\ v_d &\sim \text{Beta}(1, \psi) \\ (\theta_d, \mathcal{E}_d) &\sim G_0(\cdot) \\ \psi &\sim \text{Gamma}(a_\psi, b_\psi) \end{aligned} \quad (3.7)$$

for $k = 1, \dots, N$ and $d \geq 1$. In (3.7), $\text{Gamma}(a, b)$ denotes a Gamma density with (a, b) shape and scale parameters, respectively. Moreover, we assume that G_0 is equal to the product between a prior on $\boldsymbol{\theta}_d$ with density p_0 and a uniform prior on the tree structure \mathcal{E}_d . Furthermore, since $\boldsymbol{\theta}_d = \{\theta_{l,m}^{(d)}\}$, we set $p_0(\cdot) = \prod_{(l,m)} p_{l,m}(\cdot)$ where $p_{l,m}$ is the prior on each copula parameter introduced in Section 2.1. Also in this case, the choice of the prior on the marginal and AR parameters leads to a simplification on the computational steps.

In order to sample from the posterior distribution one needs to use a simulation method. Silva and Gramacy (2009) rely on the corresponding truncated version of the DP- model as explained in Section 3.1.2, following the work of Ishwaran and James (2001). In this thesis, we consider another approach known as slice sampling algorithm, proposed by Walker (2007) and Kalli et al. (2011).

The slice sampling algorithm introduces a latent variable $u_t, t = 1, \dots, T$ such that the joint density of $(x_{1,t}, \dots, x_{N,t}, u_t)$ given all the parameters $(\boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \mathbf{V}, \psi)$ and the previous observations becomes

$$f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t}, u_t | \boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \psi, \mathbf{V}, \mathcal{O}_{t-1}) = \sum_{d \geq 1} \mathbb{1}(w_d > u_t) K_t(x_{1,t}, \dots, x_{N,t} | (\boldsymbol{\theta}_d, \mathcal{E}_d), \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1})$$

where

$$\begin{aligned} & K_t(x_{1,t}, \dots, x_{N,t} | (\boldsymbol{\theta}_d, \mathcal{E}_d), \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1}) \\ &= \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}} \left(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i}) \right) \times \\ & \quad \prod_{k=1}^N f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right). \end{aligned}$$

If we define the set

$$B_t = \{d : w_d > u_t\},$$

then the likelihood can be rewritten as

$$f(\mathcal{O}_T, \mathbf{U} | \boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \psi, \mathbf{V}) = \prod_{t=1}^T \sum_{d \in B_t} K_t(x_{1,t}, \dots, x_{N,t} | (\boldsymbol{\theta}_d, \mathcal{E}_d), \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1})$$

where $\mathbf{U} = (u_1, \dots, u_T)$. Note that, with the introduction of this latent variable, since B_t is a finite set for all t , we obtain a finite mixture model.

Now, we can introduce an allocation variable I_t that identify the component of the mixture model from which \mathbf{X}_t is drawn. The complete data likelihood becomes

$$f(\mathcal{O}_T, \mathbf{U}, \mathbf{I} | \boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \psi, \mathbf{V}) = \prod_{t=1}^T \mathbb{1}(w_{I_t} > u_t) K_t(x_{1,t}, \dots, x_{N,t} | (\boldsymbol{\theta}_{I_t}, \mathcal{E}_{I_t}), \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1})$$

where $\mathbf{I} = (I_1, \dots, I_T)$. From the complete likelihood and the prior distribution setting in (3.7), we obtain that the augmented posterior density is given by

$$\pi(\boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I} | \mathcal{O}_T) \propto \prod_{t=1}^T \mathbb{1}(w_{I_t} > u_t) K_t(x_{1,t}, \dots, x_{N,t} | (\boldsymbol{\theta}_{I_t}, \mathcal{E}_{I_t}), \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_{t-1}) \pi_0(\boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{A}, \psi, \mathbf{V})$$

where π_0 is the prior distribution defined above. Samples from this posterior probability are obtained following the work of Walker (2007) and Kalli et al. (2011). The details of our MCMC algorithm are presented in the following Section.

3.3 MCMC for DP-Tree Copula Model

We define the set $\mathcal{D} = \{d : \exists t \text{ such that } I_t = d\}$ that represents the non empty mixture components. The cardinality of \mathcal{D} gives the number of the mixture components, denote in Section 3.1.2 with D , and $\mathcal{D}^* = \max \mathcal{D}$ is the number of stick breaking components used in the mixture.

The Metropolis within Gibbs algorithm developed for the DP-tree model samples iteratively from the full conditionals:

- ν given $[\Phi, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- \mathbf{A} given $[\Phi, \nu, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- Φ given $[\nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- $\mathbf{U}, \mathbf{V}, \psi$ given $[\Phi, \nu, \mathbf{A}, \mathbf{I}, \mathcal{O}_T]$ sampled as a block:
 - ψ given $[\Phi, \nu, \mathbf{A}, \mathbf{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);
 - \mathbf{V} given $[\Phi, \nu, \mathbf{A}, \psi, \mathbf{I}, \mathcal{O}_T]$ (closed form);
 - \mathbf{U} given $[\Phi, \nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{I}, \mathcal{O}_T]$ (closed form);
- \mathbf{I} given $[\Phi, \nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T]$ (closed form).

The variance of the random walk proposal densities are tuned to achieve acceptance rates between 20% and 80%. In the following, we present in details the steps of the algorithm.

Full conditional of ν . The full conditional of ν given $(\Phi, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T)$ is

$$\pi(\nu | \Phi, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) \prod_{k=1}^N f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k)$$

where $\pi_{\nu,k}(\cdot)$ denotes the prior Normal-Gamma $\mathcal{NG}(a_k, b_k, m_k, r_k)$ introduced in (3.7). We proceed with a MH step similar to the one presented for the finite mixture model of Section 2.4. For each ν_k ($k = 1, \dots, N$) we consider a proposal density of the form $q_{\nu,k}(\nu_k^* | \nu_k) = q_{p,k}(p_k^* | p_k) q_{\mu,k}(\mu_k^* | \mu_k)$ where the proposal density on p_k^* is a gamma density with mean p_k and fixed variance σ_p^2 , while for μ_k^* we assume a Normal proposal density with mean μ_k and fixed variance σ_μ^2 .

Full conditional of \mathbf{A} . The full conditional of \mathbf{A} given $(\Phi, \nu, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T)$ is

$$\pi(\mathbf{A}|\Phi, \nu, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}(F_{\nu_l}(x_{l,t} - \sum_{i=1}^p \alpha_{l,i} x_{l,t-i}), F_{\nu_m}(x_{m,t} - \sum_{i=1}^p \alpha_{m,i} x_{m,t-i})) \times \prod_{k=1}^N f_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}) \pi_{\mathbf{A},k}(\alpha_k)$$

where $\pi_{\mathbf{A},k}(\cdot)$ denotes the normal prior $\mathcal{N}_p(\mathbf{M}_k, \Sigma_k)$ on α_k . We apply a MH step as in the case of the finite mixture model where a random walk proposal is assumed. We set the proposal density $q_{\mathbf{A}}(\mathbf{A}^*|\mathbf{A})$ equal to a multivariate Normal distribution with mean \mathbf{A} and fixed diagonal covariance matrix.

Full conditional of Φ . We recall that in this case $\Phi = ((\theta_1, \mathcal{E}_1), (\theta_2, \mathcal{E}_2), \dots)$ and its full conditional given $(\nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T)$ is

$$\pi(\Phi|\nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathbf{I}, \mathcal{O}_T) = \prod_{d:I_t=d} \prod_{(l,m) \in \mathcal{E}_d} c_{\theta_{l,m}^{(d)}}(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})) g_0(\theta_d, \mathcal{E}_d)$$

where g_0 is the prior density on $(\theta_d, \mathcal{E}_d)$.

We proceed by distinguishing between the element $d \in \mathcal{D}$ and the elements $d \notin \mathcal{D}$. For every $d \in \mathcal{D}$ we used a MH step with a proposal density of the form

$$q_{\mathcal{E},\theta}(\mathcal{E}_d^*, \theta_d^*|\mathcal{E}_d, \theta_d) = q_{\mathcal{E}}(\mathcal{E}_d^*|\mathcal{E}_d) \prod_{(l,m)} q_{\theta_{l,m}^{(d)}}(\theta_{l,m}^{(d)*}|\mathcal{E}_d, \theta_{l,m}^{(d)}, \mathcal{E}_d^*).$$

as in the finite mixture model. For each $\theta_{l,m}^{(d)} \in \theta_d$ we use the same proposal density presented in Section 2.2 and for the tree structure we apply the tree-angular proposal described in Section 2.2.

For the element $d \notin \mathcal{D}$ we don't need a MH step and we sample the parameters directly from the prior. In theory, we need to sample an infinite number of $\theta_{l,m}^{(d)}$, but actually, in order to proceed with the algorithm, only a finite number parameters is necessary, i.e. only the elements involved in the full conditional of I .

Full conditional of $(\mathbf{U}, \mathbf{V}, \psi)$. At this step we sample $(\mathbf{U}, \mathbf{V}, \psi)$ as a block. This means sampling ψ given the rest excluded (\mathbf{U}, \mathbf{V}) and then sampling (\mathbf{U}, \mathbf{V}) given all the rest.

Firstly, we present how to sample (\mathbf{U}, \mathbf{V}) . Also for this variables we consider a blocking, and we sample \mathbf{V} given the rest excluded \mathbf{U} and \mathbf{U} given all the other parameters.

The full conditional of \mathbf{V} given $(\Phi, \nu, \mathbf{A}, \psi, \mathbf{I}, \mathcal{O}_T)$ is

$$\pi(\mathbf{V}|\Phi, \nu, \mathbf{A}, \psi, \mathbf{I}, \mathcal{O}_T) \propto \prod_{d:I_t=d} \left(v_d \prod_{l < d} (1 - v_d) \right) \pi_{\mathbf{V}}(v_d)$$

where $\pi_{\mathbf{V}}$ is the prior density $Beta(1, \psi)$. We don't need a MH step, but we have to distinguish in three different cases.

If $d \in \mathcal{D}$ the full conditional is in closed form and proportional to a $Beta(\gamma_d, \delta_d)$ with parameters

$$\begin{aligned}\gamma_d &= 1 + \sum_{t=1}^T \mathbb{1}(I_t = d) \\ \delta_d &= \psi + \sum_{t=1}^T \mathbb{1}(I_t > d).\end{aligned}$$

If $d \notin \mathcal{D}$ but $d \leq \mathcal{D}^*$, we sample v_d from a $Beta(1, \delta_d)$, otherwise for all $d > \mathcal{D}^*$ we sample directly from the prior density $Beta(1, \psi)$. Also in this case, we don't need to sample an infinite number of v_d , since only the elements affected the full conditional of I are necessary.

Now we can consider the full conditional of \mathbf{U} given $(\Phi, \nu, \mathbf{A}, \mathbf{V}, \mathbf{I}, \psi, \mathcal{O}_T)$, that is given by

$$\pi(\mathbf{U} | \Phi, \nu, \mathbf{A}, \mathbf{V}, \mathbf{I}, \psi, \mathcal{O}_T) \propto \prod_{t=1}^T \mathbb{1}(w_{I_t} > u_t).$$

and therefore we can sample each u_t from a uniform distribution on $(0, w_{I_t})$, for $t = 1, \dots, T$.

Finally, we can compute the full conditional of ψ given $(\Phi, \nu, \mathbf{A}, \mathbf{I}, \mathcal{O}_T)$. The information on ψ are provided only by the sample size and the different cluster, that is the number of different I_t . The full conditional of ψ involves only $(\mathbf{I}, \mathcal{O}_T)$ and is given by

$$\pi(\psi | \mathbf{I}, \mathcal{O}_T) \propto \frac{\psi^D \Gamma(\psi)}{\Gamma(\psi + T)} \pi_\psi(\psi)$$

where D is the number of cluster, π_ψ the Gamma prior on ψ and, $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$ is the Gamma function, see Escobar and West (1995). In order to sample from this full conditional, we consider a MH where as proposal density we choose the prior $Beta(a_\psi, b_\psi)$. The acceptance probability is

$$\min \left\{ 1, \frac{\psi^{*D} \Gamma(\psi^*) \Gamma(\psi + T)}{\psi^D \Gamma(\psi) \Gamma(\psi^* + T)} \right\}.$$

Full conditional of \mathbf{I} . The full conditional of \mathbf{I} given $(\Phi, \nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathcal{O}_T)$ is

$$\pi(\mathbf{I} | \Phi, \nu, \mathbf{A}, \psi, \mathbf{V}, \mathbf{U}, \mathcal{O}_T) \propto \prod_{t=1}^T \mathbb{1}(w_{I_t} > u_t) K_t(x_{1,t}, \dots, x_{N,t} | (\theta_{I_t}, \mathcal{E}_{I_t}), \nu, \mathbf{A}, \mathcal{O}_{t-1}).$$

For each t , we sample I_t from a discrete distribution with probability

$$\mathbb{P}(I_t = k | \mathcal{O}_T, \mathbf{U}, \zeta, \eta, \mathbf{V}, \psi) \propto \prod_{t=1}^T \mathbb{1}(w_k > u_t) K_t(x_{1,t}, \dots, x_{N,t} | (\theta_k, \mathcal{E}_k), \nu, \mathbf{A}, \mathcal{O}_{t-1}).$$

As shows in Walker (2007) and Kalli et al. (2011), we have to sample, almost surely, from a finite set. More precisely, if we define N_t the smallest integer for which

$$\sum_{k=1}^{N_t} w_k > 1 - u_t, \quad \text{for } t = 1, \dots, T$$

and $N^* = \max(N_t)$, we can observe that for every $k > N^*$ and for all t , $w_k < u_t$, and therefore $\mathbb{1}(w_k > u_t) = 0$.

Table 3.1: Simulation study: Scenario 1. DP-tree copula model with bivariate Double Gumbel or Double Clayton linking copulas re-parameterized through Kendall's tau. For each prior setting we show the posterior mean of the parameters $(\boldsymbol{\nu}, \mathbf{A})$ and the posterior mode \hat{D} of the number of cluster. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (2.6) is reported in brackets.

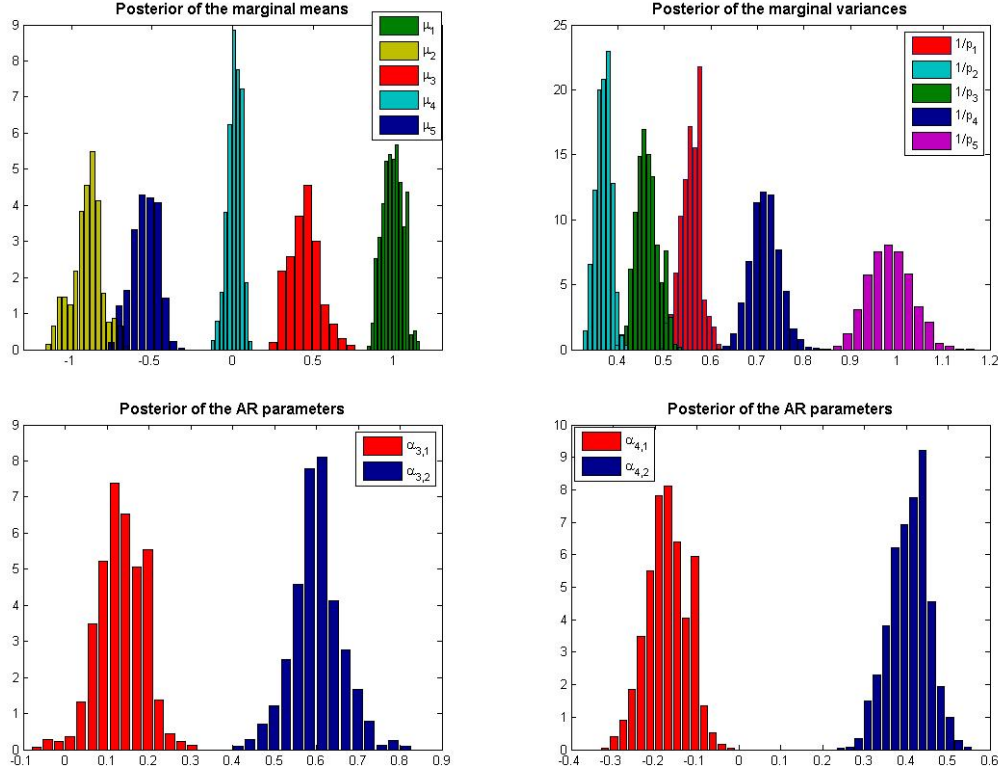
Scenario 1		Prior Settings	
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$Beta_{(-1,1)}(1, 2) \times Unif(\mathcal{H})$	$Beta_{(-1,2)}(1, 1) \times Unif(\mathcal{H})$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 0.5, 1, 0.01)$	$\mathcal{NG}(0.5, 1, 1, 0.01)$
$D = 2$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.1, 0.3], 10\mathbf{I}_p)$	$\mathcal{N}_p([0.2, 0.3], 10\mathbf{I}_p)$
	ψ	$Gamma(0.2, 1)$	$Gamma(0.08, 25)$
	True Values	Posterior Means/Modes	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(1, 0.5)	(0.99, 0.56)	(0.99, 0.53)
	(-1, 0.3)	(-0.89, 0.36)	(-1.08, 0.30)
	(0.5, 0.4)	(0.46, 0.46)	(0.52, 0.42)
	(0, 0.7)	(0.01, 0.71)	(-0.04, 0.71)
	(-0.5, 0.9)	(-0.53, 0.98)	(-0.56, 0.84)
$\boldsymbol{\alpha}_k$	(0.3, -0.4)	(0.27, -0.35)	(0.28, -0.40)
	(-0.3, 0.4)	(-0.23, 0.44)	(-0.34, 0.35)
	(0.1, 0.6)	(0.12, 0.59)	(0.06, 0.62)
	(-0.2, 0.4)	(-0.16, 0.41)	(-0.18, 0.40)
	(0.1, 0.5)	(0.10, 0.46)	(0.05, 0.52)
\hat{D}	2	2	2
\mathcal{E}	[1, 2, 5]	[1, 2, 5](0.451)	[1, 4, 5](0.4745)
	[1, 4, 5]	[1, 4, 5](0.190)	[1, 2, 5](0.1711)

3.4 Simulation Study

We use simulated data to investigate the performance of our DP-tree copula model. We present 3 alternative scenarios. For each scenario we use a dataset of $T = 200$ observations generated from a multivariate AR model with fixed parameters and a specific mixture of tree copula distributions with fixed number of components D . Also in this case, we perform sensitivity analysis. As is often the case, the concentration parameter ψ turns out to be more sensible with respect to the other parameters. Therefore, in the following prior settings, we consider two alternative choice for the hyper-parameters of the Gamma priors. The first one is an informative setting with small variance, while the second setting is less informative with a bigger value for the variance.

More precisely, in scenario 1 we consider a dataset simulated from a mixture of two tree copula distributions with bivariate Double Gumbel/Double Clayton copulas, weights $\boldsymbol{w} = [0.5, 0.5]$ and copula parameters $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m})$. In this case we consider our DP-tree copula model of Section 3.2 based on Double linking copulas and with prior specification as in (3.7). In scenario 2, the AR-observations are generated starting from a mixture of three tree copulas where each bivariate copulas belongs to the Gumbel family and $\boldsymbol{w} = [1/3, 1/3, 1/3]$. In this case, we apply the DP-tree copula

Figure 3.1: Histograms of the simulated values of the posterior distributions for the marginal and AR parameters of scenario 1 (first prior setting of Table 3.1), where we consider a DP-tree model with linking Double Gumbel/Double Clayton copulas re-parameterized through the Kendall's tau. We report the results for the AR coefficients of the third and fourth series, i.e. α_3 and α_4 .



model with same linking copulas and parameters $\theta_{l,m} = \lambda_{l,m}^U$. For these two scenarios, we consider the MCMC algorithm presented in Section 3.3 with 70000 iterations and a burn-in of 50000. Finally, in scenario 3, we used simulated data from a mixture of four tree copulas with linking Double copulas and weights $\boldsymbol{w} = [0.25, 0.25, 0.25, 0.25]$. Then, we apply our Bayesian DP-tree model with bivariate Double copulas without assuming a prior for ψ , i.e. fixing a specific value for the concentration parameter.

For scenario 1 and 2, the results are shown in Tables 3.1 and 3.2, respectively. On each tables we show the values of the parameters used to generate the data and the corresponding posterior estimates obtained with two alternative prior settings. In particular, we present the posterior means of the parameters $(\boldsymbol{\nu}, \mathbf{A})$, the posterior mode \hat{D} of the number of clusters and the list of the higher posterior probability tree structures obtained via Equation (2.6). For both the scenarios, the posterior estimates are consistent with the parameters used to generate the data. Furthermore, in Figures 3.1 and 3.4 we display the histograms of the simulated values of the posterior distributions for $\boldsymbol{\nu}$ and for the coefficients of two AR series of the DP-tree model with Double linking copulas

Figure 3.2: Diagnostic plots for the marginal parameters of scenario 1 (second prior setting of Table 3.1), where we consider a DP-tree model with linking Double Gumbel/Double Clayton copulas re-parameterized through the Kendall's tau.

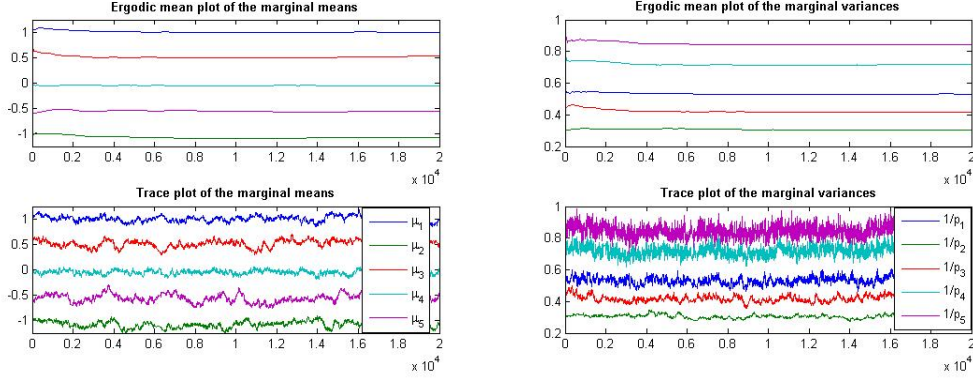
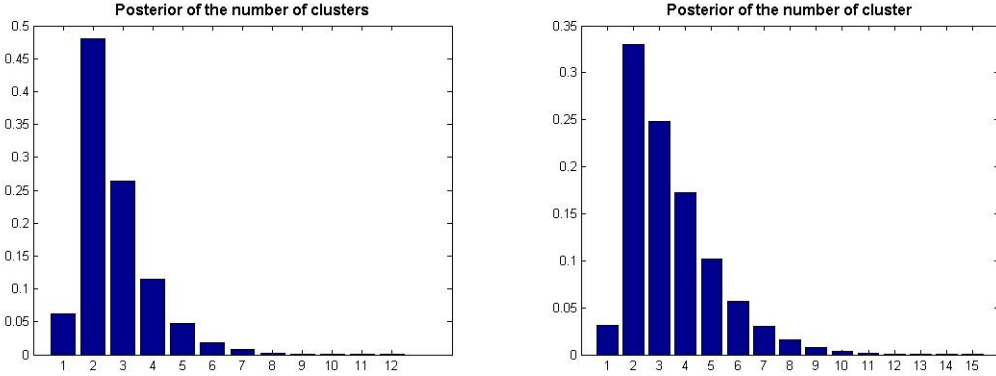


Figure 3.3: Histograms of the simulated values of the posterior distributions of number of components for scenario 1, where we study a DP-tree copula model with Double Gumbel/Double Clayton linking copulas and $D = 2$. On the left, the result with the first prior setting of Table 3.1 and on the right with the second prior setting.



and of the DP-tree model with Gumbel linking copulas, respectively. Concerning the convergence of the algorithm, in Figure 3.2, as an example, we report the diagnostic plots for some parameters of scenario 1. Finally, the histograms of the simulated values of the posterior distributions of the number of cluster for each prior setting of scenario 1 and of scenario 2 are presented in Figures 3.3 and 3.5. All these histograms are centered on the corresponding true value of D , i.e. $D = 2$ for scenario 1 and $D = 3$ for scenario 2. Note that, for both scenarios, the expected number of cluster a priori $\mathbb{E}(D)$, computed via (3.5), is approximately 1.06 for the first prior settings, while for the second prior setting is approximately 10.6. Moreover, we observe that the distributions associated to the first prior setting of each scenario are more concentrated than the distributions related to the second prior settings. Indeed, for both scenarios, the value of the variance of the Gamma prior

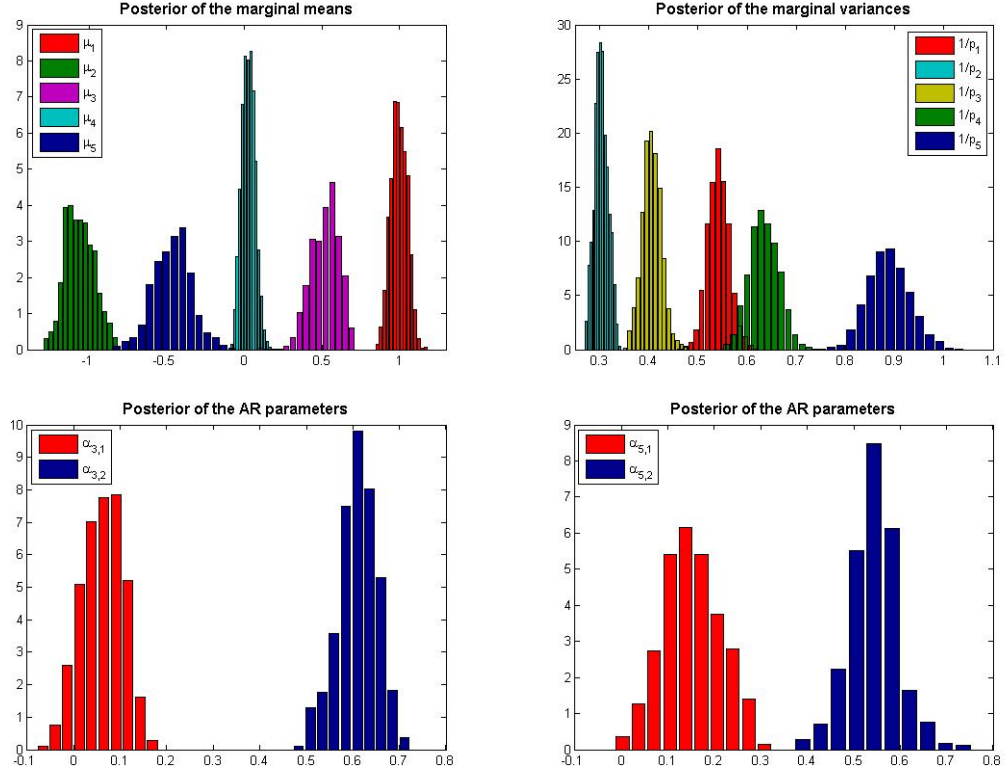
Table 3.2: Simulation study: Scenario 2. DP-tree copula model with bivariate Gumbel linking copulas re-parameterized through upper tail parameters. For each prior setting we show the posterior mean of the parameters $(\boldsymbol{\nu}, \mathbf{A})$ and the posterior mode \hat{D} of the number of cluster. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (2.6) is reported in brackets.

Scenario 2		Prior Settings	
$N = 5$	$\theta_{l,m}$	$Beta(1, 1)$	$Beta(1, 2)$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 1, 0, 0.1)$	$\mathcal{NG}(1, 2, 0, 0.1)$
$D = 3$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.1, 0.3], 10\mathbf{I}_p)$	$\mathcal{N}_p[0.2, 0.3], 10\mathbf{I}_p)$
	ψ	$Gamma(0.2, 1)$	$Gamma(0.08, 25)$
	True Values	Posterior Means/Modes	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(1, 0.5)	(1.00, 0.54)	(0.91, 0.50)
	(-1, 0.3)	(-1.08, 0.30)	(-1.12, 0.31)
	(0.5, 0.4)	(0.52, 0.40)	(0.36, 0.38)
	(0, 0.6)	(0.02, 0.63)	(0.05, 0.62)
	(-0.5, 0.9)	(-0.45, 0.88)	(-0.53, 0.87)
$\boldsymbol{\alpha}_k$	(0.3, -0.4)	(0.25, -0.40)	(0.33, -0.33)
	(-0.3, 0.4)	(-0.35, 0.39)	(-0.35, 0.35)
	(0.1, 0.6)	(0.06, 0.61)	(0.07, 0.70)
	(-0.2, 0.4)	(-0.17, 0.34)	(-0.14, 0.44)
	(0.1, 0.5)	(0.15, 0.54)	(0.03, 0.52)
\hat{D}	3	3	3
\mathcal{E}	[1, 1, 1]	[1, 1, 3](0.326)	[1, 1, 1](0.262)
	[1, 1, 3]	[1, 5, 1](0.123)	[1, 1, 3](0.173)
	[1, 5, 5]	[1, 1, 1](0.100)	[1, 5, 5](0.157)
		[1, 5, 5](0.065)	

associated to ψ in the first prior setting is equal to 0.2, while in the second prior settings is higher, i.e. equal to 50.

In scenario 3, we illustrate the results obtained with a DP-tree copula model based on linking Double Gumbel and Double Clayton copulas where the value of the concentration parameter ψ is fixed. The values of the parameters used to generate the data and the posterior estimates computed with two different prior settings are reported in Table 3.3. In particular, the prior settings present different value for the concentration parameter, respectively $\psi = 0.1$ and $\psi = 0.5$. In this table, for each prior setting, we show the posterior means for the parameters $(\boldsymbol{\nu}, \mathbf{A})$, the posterior mode of the number of clusters and the trees with the higher posterior probability estimated via (2.6). The histograms of the simulated values of the posterior distributions for the marginal parameters and for the coefficients of two AR series are presented in Figure 3.6. The posterior distributions of the number of clusters obtained with each prior setting of scenario 3 are shown in Figure 3.7. Also in this case, both the distributions are centered around the true value of D . As expected, the distribution obtained with the first prior setting is more concentrated than the distribution obtained

Figure 3.4: Histograms of the simulated values of the posterior distributions for the marginal and AR parameters of scenario 2 (first prior setting of Table 3.2), where we consider a DP-tree model with linking Gumbel copulas re-parameterized through upper tail parameters. We report the results for the AR coefficients of the third and fifth series, i.e. α_3 and α_5 .



with the second prior setting that, indeed, shows a higher values of ψ .

Finally, in Table 3.4 are listed the values of the DIC_3 criterion computed for each scenario with our fully Bayesian approach and the B-IFM procedure. As for the B-IFM, to the data of scenario 1/ scenario 2 we apply a B-IFM procedure with, respectively, a mixture of two tree copulas based on Double copulas and a mixture of three tree copulas with linking Gumbel copulas. To the data of scenario 3, we apply a B-IFM procedure with mixture of four tree copulas based on Double copulas and with fixed value of ψ . Also in this case, we observe that the DIC_3 values obtained with the fully Bayesian approach are lower than the DIC_3 of the corresponding B-IFM procedure.

Figure 3.5: Histograms of the simulated values of the posterior distributions of number of components for scenario 2, where we study a DP-tree copula model with Gumbel linking copulas and $D = 3$. On the left, the result with the first prior setting of Table 3.2 and on the right with the second prior setting.

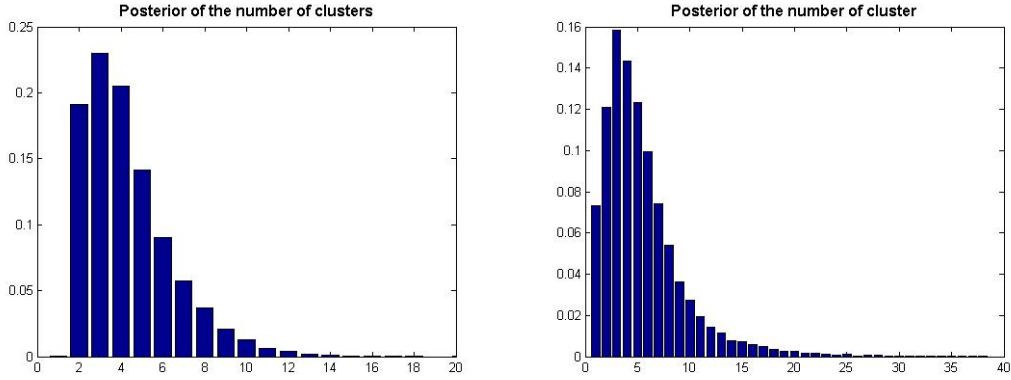


Figure 3.6: Histograms of the simulated values of the posterior distributions for the marginal and AR parameters of scenario 3 (first prior setting of Table 3.3, i.e. $\psi = 0.1$), where we consider a DP-tree copula model with Double Gumbel and Double Clayton copulas re-parameterized through Kendall's tau and fixed value of ψ . We report the results for the AR coefficients of the first and fourth series, i.e. α_1 and α_4 .

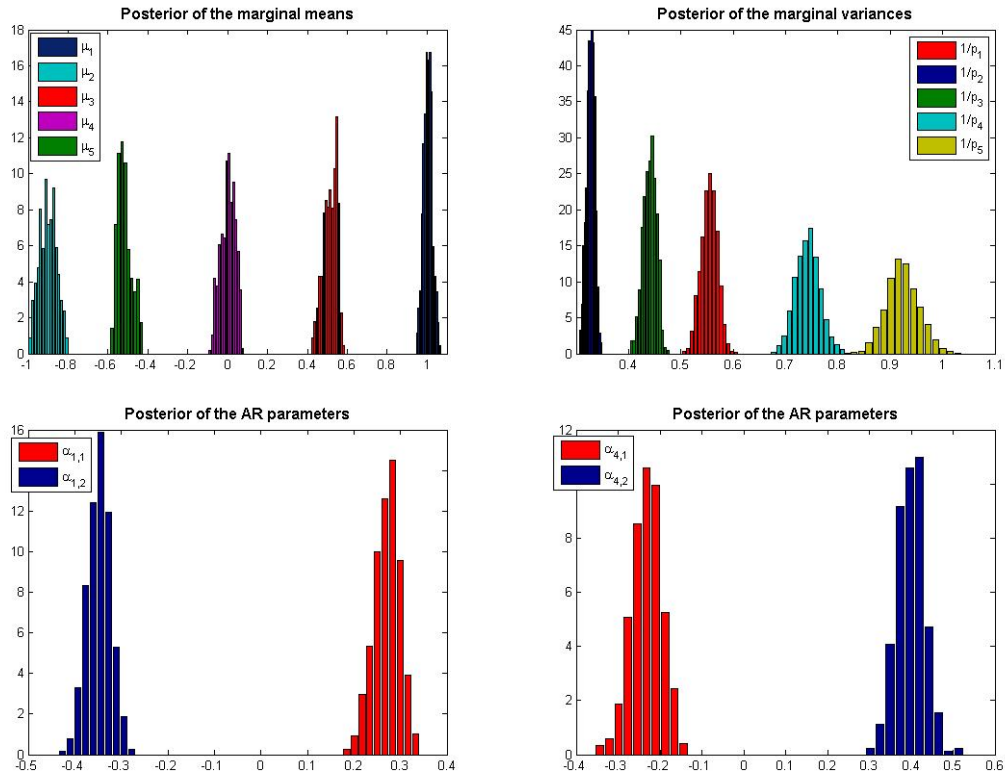


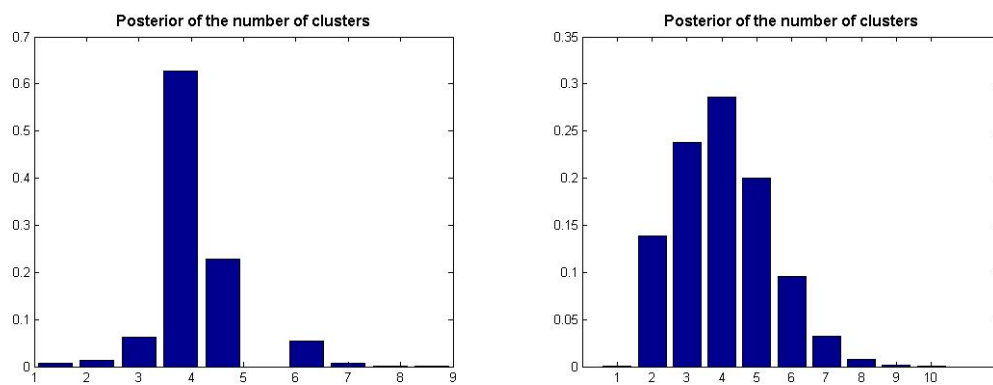
Table 3.3: Simulation study: Scenario 3. DP-tree copula model with with Double Gumbel and Double Clayton copulas re-parameterized through Kendall's tau parameter, and with fixed value of ψ . For each prior setting we show the posterior mean of the parameters $(\boldsymbol{\nu}, \mathbf{A})$ and the posterior mode \hat{D} of the number of cluster. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (2.6) is reported in brackets.

Scenario 3		Prior Settings	
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$Beta_{(-1,1)}(2, 2) \times Unif(\mathcal{H})$	$Beta_{(-1,1)}(1, 2) \times Unif(\mathcal{H})$
$p = 2$	(μ_k, p_k)	$\mathcal{NG}(1, 0.5, 1, 0.01)$	$\mathcal{NG}(1, 0.5, 1, 0.01)$
$D = 4$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.1, 0.3], 10\mathbf{I}_p)$	$\mathcal{N}_p([0.1, 0.3], 10\mathbf{I}_p)$
	ψ	$\psi = 0.1$	$\psi = 0.5$
	True Values	Posterior Means/Modes	
$\nu_k = (\mu_k, \frac{1}{p_k})$	(1, 0.5)	(1.00, 0.55)	(0.91, 0.57)
	(-1, 0.3)	(-0.90, 0.32)	(-0.92, 0.34)
	(0.5, 0.4)	(0.52, 0.44)	(0.41, 0.47)
	(0, 0.7)	(0.01, 0.74)	(0.002, 0.68)
	(-0.5, 0.9)	(-0.53, 0.93)	(-0.38, 0.90)
$\boldsymbol{\alpha}_k$	(0.3, -0.4)	(0.27, -0.34)	(0.33, -0.39)
	(-0.3, 0.4)	(-0.25, 0.44)	(-0.24, 0.40)
	(0.1, 0.6)	(0.15, 0.53)	(0.12, 0.62)
	(-0.2, 0.4)	(-0.23, 0.39)	(-0.16, 0.48)
	(0.1, 0.5)	(0.10, 0.47)	(0.23, 0.46)
\hat{D}	4	4	4
\mathcal{E}	[1, 2, 5]	[2, 3, 5](0.292)	[2, 3, 5](0.193)
	[1, 4, 5]	[3, 5, 4](0.080)	[1, 4, 5](0.108)
	[2, 1, 5]	[5, 1, 4](0.078)	[1, 2, 5](0.080)
	[2, 3, 5]	[1, 2, 3](0.055)	[1, 1, 3](0.057)
		[1, 5, 1](0.041)	[5, 1, 5](0.049)
		[1, 4, 5](0.036)	[2, 2, 5](0.045)
		[3, 2, 5](0.0298)	[2, 1, 5](0.034)
		[2, 1, 5](0.293)	[3, 5, 5](0.030)
		[1, 2, 5](0.278)	[1, 4, 5](0.274)

Table 3.4: Simulated Data: DIC₃ Table for DP-tree copula models.

	Fully Bayesian	B-IFM
True Model		
Scenario 1 (Infinite Mixture of Combined Tree Copula)	1208	1446
Scenario 2 (Infinite Mixture of Gumbel Tree Copula)	1079	1185
Scenario 3 (Infinite Mixture of Combined Tree Copula, fixed ψ)	1341	1465

Figure 3.7: Histograms of the simulated values of the posterior distributions of number of components for scenario 3, where we study a DP-tree copula model with Double Gumbel/Double Clayton linking copulas, with fixed value of ψ and $D = 4$. On the left, the result with the first prior setting of Table 3.3 ($\psi = 0.1$) and on the right with the second prior setting ($\psi = 0.5$).



Chapter 4

Factor Copula Model and Markov Switching Factor Copula Model

As an alternative to the tree copula model and the (finite or infinite) mixture of tree copulas model presented in Chapters 2 and 3, here we introduce the one-factor copula model. We consider the AR-model of Section 1.5.3 and assume that the pdf of the innovations is represented via suitable bivariate copula that link observed data to a latent variable V . Also in this case, a graphical tree structure can be associated to the joint density of the innovations and V , indeed, as already noted in Section 1.4.3, the one-factor copula is equivalent to a truncated C-vine at first level with a latent node. In contrast to the previous models, in the factor copula the underlying tree structure is fixed and we can not proceed with structural learning but the presence of the latent node allows to study a more general dependence structure respect to the one associated to a single tree copula distribution.

In this Chapter, we also present a Markov switching factor copula model in which a hidden latent variable S_t determines the copula structure of the innovations at time t . $(S_t)_{t \geq 1}$ is assumed to be a first order Markov chain in discrete time that takes values in $\{1, \dots, R\}$, see Section 1.5.4. In particular, we suppose that in each state $r = 1, \dots, R$ the dependence structure of the innovations is represented through a specific one factor copula. The presence of different regimes allows to have a more flexible model.

The outline of this Chapter is the following. The Bayesian one factor copula model is presented in Section 4.1 and the corresponding MCMC algorithm for posterior estimates is described in Section 4.2. The Markov switching factor copula model and the corresponding MCMC algorithm are detailed in Sections 4.3 and 4.4. Finally, an application of these methodologies applied to simulated data is shown in Section 4.5.

4.1 Bayesian Inference for Factor Copula Model

In this model we assume that for any fixed t the joint density of the innovations is represented via a one-factor copula with latent variable V as described in Section 1.4. With this assumption,

the joint density of ε_t is given by

$$f_{\varepsilon_t}(\varepsilon_{1,t}, \dots, \varepsilon_{N,t} | \Phi, \nu) = \int_0^1 \prod_{k=1}^N c_{\theta_{k,0}}(F_{\nu_k}(\varepsilon_{k,t}), v) f_{\nu_k}(\varepsilon_{k,t}) dv.$$

Using again the transformation (1.15) that involves the AR-models, we obtain that the joint density of \mathbf{X}_t given the past is equal to

$$f_{\mathbf{X}_t}(x_{1,t}, \dots, x_{N,t} | \Phi, \nu, \mathbf{A}, \mathcal{O}_{t-1}) = \int_0^1 \prod_{k=1}^N c_{\theta_{k,0}} \left(F_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right), v \right) f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) dv.$$

For each observation $\mathbf{X}_t = (X_{1,t}, \dots, X_{N,t})$ there is a distinct, independent realization v_t of latent variable V . Hence, for Bayesian inference, $\mathbf{V}_{1:T} = (V_1, \dots, V_T)$ represents additional independent latent variables distributed uniformly on $[0, 1]$. Thus, the complete augmented likelihood is

$$f(\mathcal{O}_T, v_{1:T} | \Phi, \nu, \mathbf{A}) = \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}} \left(F_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right), v_t \right) f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right).$$

In this model, since we have T more latent variables $\mathbf{V}_{1:T}$, we assume for simplicity that the means of the Normal marginal distributions are known, i.e. $\nu_k = p_k$ for $k = 1, \dots, N$. The parameter Φ denotes only the copula parameters $\theta = (\theta_{1,0}, \dots, \theta_{N,0})$, since we cannot make inference on the underlying graphical structure.

In order to proceed with Bayesian inference, we assign the following independent prior distributions

$$\begin{aligned} \nu_k &\sim \text{Gamma}(a_k, b_k) \\ \alpha_k &\sim \mathcal{N}_p(\mathbf{M}_k, \Sigma_k) \\ \theta_{k,0} &\sim p_{k,0} \end{aligned} \tag{4.1}$$

for $k = 1, \dots, N$ and $t = 1, \dots, T$. Also for this model, we assume that each bivariate copula belongs to one of the families presented in Section 1.5.3, i.e. Gumbel, Clayton or Joe copulas re-parameterized in terms of tail dependence parameters or Double Gumbel/Double Clayton copulas re-parameterized through Kendall's tau. Also in this case, for the marginal and AR parameters we consider the usual prior densities that permit to simplify the computational steps.

If $\theta_{k,0}$ is equal to upper or lower tail parameter, $p_{k,0}$ is a Beta density with parameters $(\gamma_{k,0}, \delta_{k,0})$. For the model that involves the rotated copulas, i.e. $\theta_{k,0} = (\tau_{k,0}, \zeta_{k,0})$, more attention should be paid. Indeed, if we consider a one-factor copula $c(u_1, u_2)$ with linking bivariate Gumbel copulas $c(\cdot, \cdot | \tau_1)$ and $c^{180}(\cdot, \cdot | \tau_2)$ re-parameterized via the Kendall's tau, $\tau_1, \tau_2 > 0$, respectively, then

$$\begin{aligned} c(u_1, u_2) &= \int_0^1 c(u_1, v | \tau_1) c^{180}(u_2, v | \tau_2) dv \\ &= \int_0^1 c(u_1, v | \tau_1) c(1 - u_2, 1 - v | \tau_2) dv \end{aligned} \tag{4.2}$$

where in (4.2) we use the relation between a copula density and its 180° rotation, see Equation (1.6). By the change of variable $v' = 1 - v$, we have

$$\begin{aligned} c(u_1, u_2) &= \int_0^1 c(u_1, v|\tau_1)c^{180}(u_2, v|\tau_2)dv \\ &= \int_0^1 c(u_1, 1 - v'|\tau_1)c(1 - u_2, v'|\tau_2)dv' \\ &= \int_0^1 c^{270}(u_1, v'|\tau_1)c^{90}(u_2, v'|\tau_2)dv'. \end{aligned}$$

Therefore the joint density is not uniquely defined by its linking copulas. The problem of identifiability holds also for the Double Clayton copula, and for any number of variable $N \geq 2$. In order to obtain a well-defined model, we have to choose only one rotation between 0° and 270° and one between 90° and 180°. Hence, the prior $p_{k,0}$ on $\theta_{k,0} = (\tau_{k,0}, \zeta_{k,0})$ is assumed to be the product of a shifted Beta on $(-1, 1)$ with parameters $(\gamma_{k,0}, \delta_{k,0})$ and a uniform density on \mathcal{H}_1 , where $\mathcal{H}_1 = \{DG_1, DC_1\}$, i.e. the rotations 0° and 90° for each copula family.

With this assumptions, the posterior density is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathbf{V}_{1:T}|\mathcal{O}_T) &= \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}} \left(F_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right), V_t \right) \times \\ &\quad f_{\nu_k} \left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i} \right) \pi_0(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathbf{V}) \end{aligned}$$

where $\pi_0(\cdot)$ is the prior setting presented in (4.1).

Also in this case the previous posterior density is not in closed form and hence, in order to obtain posterior estimates, we adopt the Metropolis within Gibbs algorithm described in the following Section.

4.2 MCMC for Factor Copula Model

The MCMC algorithm developed for the analysis of the factor copula model iteratively samples from the full conditionals:

- $\boldsymbol{\nu}$ given $[\boldsymbol{\theta}, \mathbf{A}, \mathbf{V}_{1:T}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- \mathbf{A} given $[\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{V}_{1:T}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- $\mathbf{V}_{1:T}$ given $[\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- $\boldsymbol{\theta}$ given $[\boldsymbol{\nu}, \mathbf{A}, \mathbf{V}_{1:T}, \mathcal{O}_T]$ (Metropolis-Hasting step).

The variance of the random walk proposal are tuned to achieve acceptance rates between 20% and 80%. The details of each step of the algorithm are given in the following.

Full conditional of ν . The full conditional of the marginal parameters given $(\theta, \mathbf{A}, \mathbf{V}_{1:T}, \mathcal{O}_T)$ is given by

$$\pi(\nu | \theta, \mathbf{A}, \mathbf{V}_{1:T}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}}(F_{\nu_k}(\epsilon_{k,t}), V_t) f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k)$$

where $\pi_{\nu,k}$ is a Gamma prior since in this model the marginal means are assumed known. As in the previous models we have to use a Metropolis Hasting step where we consider the proposal density

$$q(\nu^* | \nu) = \prod_{k=1}^N q_{p,k}(p_k^* | p_k)$$

where each $q_{p,k}(p_k^* | p_k)$ is a Gamma distribution with mean p_k and fixed variance σ_p^2 . This choice allow us to proceed with N different MH steps where each acceptance probability is

$$\min \left\{ 1, \prod_{t=1}^T \frac{c_{\theta_{k,0}}(F_{\nu_k^*}(\epsilon_{k,t}), V_t) f_{\nu_k^*}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k^*) q_{p,k}(p_k^* | p_k)}{c_{\theta_{k,0}}(F_{\nu_k}(\epsilon_{k,t}), V_t) f_{\nu_k}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k) q_{p,k}(p_k^* | p_k)} \right\}$$

for $k = 1, \dots, N$.

Full conditional of \mathbf{A} . The likelihood of the $AR(p)$ parameters given $(\theta, \nu, \mathbf{V}_{1:T}, \mathcal{O}_T)$ is

$$\pi(\mathbf{A} | \theta, \nu, \mathbf{V}_{1:T}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}}\left(F_{\nu_k}\left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}\right), V_t\right) f_{\nu_k}\left(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}\right) \pi_{\alpha,k}(\alpha_k)$$

where $\pi_{\alpha,k}$ indicates the multivariate Normal prior on α_k . As in the step on ν , we can consider a proposal density of the form

$$q(\mathbf{A}^* | \mathbf{A}) = \prod_{k=1}^N q_{\alpha,k}(\alpha_k^* | \alpha_k)$$

and proceed with a MH step for each α_k . In particular, we choose each proposal $q_{\alpha,k}$ equal to a p dimensional Normal density with mean α_k and fixed variance σ_{α}^2 . Each acceptance probability is given by

$$\min \left\{ 1, \prod_{t=1}^T \frac{c_{\theta_{k,0}}(F_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i}^* x_{k,t-i}), V_t) f_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i}^* x_{k,t-i}) \pi_{\alpha,k}(\alpha_k^*) q_{\alpha,k}(\alpha_k | \alpha_k^*)}{c_{\theta_{k,0}}(F_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}), V_t) f_{\nu_k}(x_{k,t} - \sum_{i=1}^p \alpha_{k,i} x_{k,t-i}) \pi_{\alpha,k}(\alpha_k) q_{\alpha,k}(\alpha_k^* | \alpha_k)} \right\}$$

for $k = 1, \dots, N$.

Full conditional of $\mathbf{V}_{1:T}$. For the latent variable the full conditional, given $(\theta, \nu, \mathbf{A}, \mathcal{O}_T)$, is

$$\pi(\mathbf{V}_{1:T} | \theta, \nu, \mathbf{A}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}}(F_{\nu_k}(\epsilon_{k,t}), V_t).$$

We proceed with a MH step for each V_t and we choose an uniform proposal density. We obtain for each t the acceptance probability equal to

$$\min \left\{ 1, \prod_{k=1}^N \frac{c_{\theta_{k,0}}(F_{\nu_k}(\epsilon_{k,t}), V_t^*)}{c_{\theta_{k,0}}(F_{\nu_k}(\epsilon_{k,t}), V_t)} \right\}.$$

Full conditional of θ . The last full conditional is on the copula parameters and it is equal to

$$\pi(\boldsymbol{\theta}|\boldsymbol{\nu}, \mathbf{A}, \mathbf{V}_{1:T}, \mathcal{O}_T) \propto \prod_{t=1}^T \prod_{k=1}^N c_{\theta_{k,0}} \left(F_{\nu_k}(\epsilon_{k,t}), V_t \right) p_{k,0}(\theta_{k,0}).$$

We recall that the prior $p_{k,0}(\cdot)$ is a Beta prior if the parameter $\theta_{k,0}$ corresponds to the upper or lower tail measures, otherwise if $\theta_{k,0} = (\tau_{k,0}, \zeta_{k,0})$ it is the product of a shifted Beta on $(-1, 1)$ and a uniform distribution on $\mathcal{H}_1 = \{DG_1, DC_1\}$. We consider a proposal density of the form

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = \prod_{k=1}^N q_{\theta,k}(\theta_{k,0}^*|\theta_{k,0}).$$

In case of $\theta_{k,0}^* = (\tau_{k,0}^*, \zeta_{k,0}^*)$ we sample $\zeta_{k,0}^*$ from the discrete distribution on \mathcal{H}_1 with probabilities

$$q_{\zeta,k}(\zeta_{k,0}^* = j|\tau_{k,0}^*) = \frac{\prod_{t=1}^T c_{(\tau_{k,0}^*, j)} \left(F_{\nu_k}(\epsilon_{k,t}), V_t \right)}{\sum_{h \in \mathcal{H}_1} \prod_{t=1}^T c_{(\tau_{k,0}^*, h)} \left(F_{\nu_k}(\epsilon_{k,t}), V_t \right)},$$

while for $\tau_{k,0}^*$, through the transformation $h : [-1, 1] \rightarrow \mathbb{R}$ defined as

$$h(x) = \log \left(\frac{1-x}{1+x} \right),$$

we use a random walk proposal defined on \mathbb{R} . In particular, in order to avoid numerical problems, we take $\tau_{k,0}^* \in [-0.99, 0.99]$ and therefore we consider a truncated random walk proposal defined on $[-5.29, 5.29]$ since $h(-0.99) = -5.29$ and $h(0.99) = 5.29$.

The acceptance probability of the MH step on $(\tau_{k,0}, \zeta_{k,0})$, for $k = 1, \dots, N$ is

$$\min \left\{ 1, \prod_{t=1}^T \frac{c_{\theta_{k,0}^*} \left(F_{\nu_k}(\epsilon_{k,t}), V_t \right) p_{k,0}(\theta_{k,0}^*) q_{\tau,k}(\tau_{k,0}|\tau_{k,0}^*) q_{\zeta,k}(\zeta_{k,0}^*|\tau_{k,0}^*)}{c_{\theta_{k,0}} \left(F_{\nu_k}(\epsilon_{k,t}), V_t \right) p_{k,0}(\theta_{k,0}) q_{\tau,k}(\tau_{k,0}^*|\tau_{k,0}) q_{\zeta,k}(\zeta_{k,0}|\tau_{k,0})} \right\} \quad (4.3)$$

In the case of $\theta_{k,0} = \lambda_{k,0}^{U/L}$, the situation is more simple with only one parameter that we sample again with a truncated random walk on $[-4.59, 4.59]$ where the adopted transformation is $h(x) = \log \left(\frac{1-y}{y} \right)$.

The acceptance rate that we obtained is similar to (4.3).

4.3 Bayesian Inference for Markov Switching Factor Copula Model

In this Section, we consider an AR- Markov switching model - see Section 1.5.4 - where, in each regime, the dependence structure among the examined variables is represented via a one factor copula. For the presence of the latent node V of the factor copula and the hidden Markov chain, we have $2T$ additional latent variables, and hence, for simplicity, in this model we suppose that the AR-parameters are known and work directly with the innovations. From the inference perspective, this is equivalent to consider a B-IFM procedure as described in Section 2.5.

In the Markov switching factor copula model, Equation (1.16), that represents the conditional pdf of the innovations at time t , becomes

$$f_{\boldsymbol{\varepsilon}_t}(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N | s_t, \boldsymbol{\Phi}, \boldsymbol{\nu}) = \int_0^1 \prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_{k,t}), v) f_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_{k,t}) dv$$

where the hidden first order Markov chain $(S_t)_{t=1}^T$ has transition matrix \mathbf{Q} and state space $\mathcal{R} = \{1, \dots, R\}$. The initial distribution of S_t is denoted with (ρ_1, \dots, ρ_R) . In this model, $\boldsymbol{\Phi}$ denotes the copula parameters of the factor copula of each regime, i.e. $\boldsymbol{\Phi} = \{\boldsymbol{\theta}^{(r)}, r = 1, \dots, R\}$ with $\boldsymbol{\theta}^{(r)} = (\theta_{1,0}^{(r)}, \dots, \theta_{N,0}^{(r)})$ and, analogously, $\boldsymbol{\nu}$ collects all the parameters of the marginal distributions, i.e. $\boldsymbol{\nu} = \{\boldsymbol{\nu}^{(r)}, r = 1, \dots, R\}$ with $\boldsymbol{\nu}^{(r)} = (\nu_1^{(r)}, \dots, \nu_N^{(r)})$. As in the factor copula model of Section 4.1, we assume marginal Normal distributions with unknown precision $p_k^{(r)}$, and, therefore, in the following, $\nu_k^{(r)} = p_k^{(r)}$ for $k = 1, \dots, N$ and $r = 1, \dots, R$.

The joint density of the innovations $\boldsymbol{\varepsilon}_{1:T} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)$ and of the Markov chain $\mathbf{S}_{1:T} = (S_1, \dots, S_T)$ is equal to

$$\begin{aligned} f(\boldsymbol{\varepsilon}_{1:T}, \mathbf{s}_{1:T} | \boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{Q}) &= \\ &= f(\boldsymbol{\varepsilon}_1 | s_1, \boldsymbol{\Phi}, \boldsymbol{\nu}) \rho_{s_1} \prod_{t=2}^T f(\boldsymbol{\varepsilon}_t | s_t, \boldsymbol{\Phi}, \boldsymbol{\nu}) \mathbf{q}_{s_t, s_{t-1}} \\ &= \prod_{t=1}^T \left[\int_0^1 \prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_k), v) f_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_k) dv \right] \rho_{s_1} \prod_{t=2}^T \mathbf{q}_{s_t, s_{t-1}}. \end{aligned}$$

As in the factor copula model, for each observation one can consider an additional random variable V_t related to the latent node. Hence, we consider $\mathbf{V}_{1:T} = (V_1, \dots, V_T)$ additional independent random variables distributed uniformly on $[0, 1]$. In this way, we obtain the augmented likelihood of the Markov switching factor copula model

$$f(\boldsymbol{\varepsilon}_{1:T}, \mathbf{s}_{1:T}, \mathbf{v}_{1:T} | \boldsymbol{\Phi}, \boldsymbol{\nu}, \mathbf{Q}) = \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_k), v_t) f_{\nu_k^{(s_t)}}(\boldsymbol{\varepsilon}_k) \right] \rho_{s_1} \prod_{t=2}^T \mathbf{q}_{s_t, s_{t-1}}.$$

Once again, we assume that the linking copulas belong to one of the copula families presented in Section 1.5.3, i.e. Gumbel/Clayton or Joe copulas re-parameterized through the tail dependence measure or Double Gumbel/Double Clayton copulas with Kendall's tau parameter.

In order to proceed with Bayesian inference, we assume the following independent prior distributions

$$\begin{aligned} \nu_k^{(r)} &\sim \text{Gamma}(a_k^{(r)}, b_k^{(r)}) \\ \mathbf{Q}_{r\cdot} &\sim \text{Dir}(\psi_{1,r}, \dots, \psi_{R,r}) \\ \theta_{k,0}^{(r)} &\sim p_{k,0}^{(r)} \end{aligned} \tag{4.4}$$

for $k = 1, \dots, N$, $r = 1, \dots, R$ and where $\mathbf{Q}_{r\cdot}$ denotes the r -th row of the transition matrix \mathbf{Q} . The prior $p_{k,0}^{(r)}$ is equal to a Beta density with parameters $(\gamma_k^{(r)}, \delta_k^{(r)})$ if the copula parameters are equal to

upper or lower tail dependence. Otherwise, in case of Double copulas and $\theta_{k,0}^{(r)} = (\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$, $p_{k,0}^{(r)}$ is equal to the product of a shifted Beta on $(-1, 1)$ with parameters $(\gamma_k^{(r)}, \delta_k^{(r)})$ and a uniform density on $\mathcal{H}_1 = \{DG_1, DC_1\}$. The choice of the Dirichlet prior on each row of the transition matrix allows a closed form of the related full conditional, while the prior on the marginal parameters permits to simplify the computational steps.

Assuming the prior setting of (4.4), we obtain the following posterior density

$$\pi(\Phi, \nu, \mathbf{Q}, \mathbf{v}_{1:T}, \mathbf{s}_{1:T} | \epsilon_{1:T}) \propto \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\epsilon_k), v_t) f_{\nu_k^{(s_t)}}(\epsilon_k) \right] \rho_{s_1} \prod_{t=2}^T \mathbf{q}_{s_t, s_{t-1}} \pi_0(\Phi, \nu, \mathbf{Q})$$

where π_0 denotes the prior described above.

The previous posterior density is not in closed form, hence in order to obtain posterior samples we adopt a suitable Metropolis within Gibbs algorithm described in the following Section.

4.4 MCMC for Markov Switching Factor Copula Model

The MCMC algorithm developed for the analysis of the Markov switching factor model iteratively samples from the full conditional:

- ν given $[\Phi, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T}]$ (Metropolis-Hasting step);
- Φ given $[\nu, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T}]$ (Metropolis-Hasting step);
- $\mathbf{S}_{1:T}$ given $[\Phi, \nu, \mathbf{Q}, \mathbf{V}_{1:T}, \epsilon_{1:T}]$ (forward filtering-backward sampling);
- \mathbf{Q} given $[\Phi, \nu, \mathbf{S}_{1:T}, \mathbf{V}_{1:T}, \epsilon_{1:T}]$ (closed form);
- $\mathbf{V}_{1:T}$ given $[\Phi, \nu, \mathbf{S}_{1:T}, \mathbf{Q}, \epsilon_{1:T}]$ (Metropolis-Hasting step).

The variance of the random walk proposal are tuned to achieve acceptance rates between 20% and 80%. The details of each step of the algorithm are given in the following.

Full conditional of ν . The full conditional of ν given $(\Phi, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T})$ is proportional to

$$\begin{aligned} \pi(\nu | \Phi, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T}) &\propto \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\epsilon_k), V_t) f_{\nu_k^{(s_t)}}(\epsilon_k) \right] \pi_{\nu,k}(\nu_k^{(s_t)}) \\ &= \prod_{r=1}^R \prod_{t: S_t=r} \left[\prod_{k=1}^N c_{\theta_{k,0}^{(r)}}(F_{\nu_k^{(r)}}(\epsilon_k), V_t) f_{\nu_k^{(r)}}(\epsilon_k) \right] \pi_{\nu,k}(\nu_k^{(r)}) \end{aligned} \quad (4.5)$$

where $\pi_{\nu,k}$ is the Gamma prior introduced in (4.4). We proceed with a Metropolis-Hasting step for each marginal parameter $\nu_k^{(r)}$ for $k = 1, \dots, N$ and $r = 1, \dots, R$. We recall that, in this case, $\nu_k^{(r)} = p_k^{(r)}$ since the marginal means are supposed to be known. As proposal density $q_{p,k} \left(p_k^{(r)*} | p_k^{(r)} \right)$,

we take a Gamma distribution with mean $p_k^{(r)}$ and fixed variance. The acceptance probability of each MH step is equal to

$$\min \left\{ 1, \prod_{t=1}^T \frac{c_{\theta_{k,0}^{(r)}}(F_{\nu_k^{(r)*}}(\epsilon_{k,t}), V_t) f_{\nu_k^{(r)*}}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k^{(r)*}) q_{p,k}(p_k^{(r)} | p_k^{(r)*})}{c_{\theta_{k,0}^{(r)}}(F_{\nu_k^{(r)}}(\epsilon_{k,t}), V_t) f_{\nu_k^{(r)}}(\epsilon_{k,t}) \pi_{\nu,k}(\nu_k^{(r)}) q_{p,k}(p_k^{(r)*} | p_k^{(r)})} \right\}$$

for $k = 1, \dots, N$ and $r = 1, \dots, R$.

Full conditional of Φ . The full conditional of Φ given $(\nu, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T})$ is

$$\begin{aligned} \pi(\Phi | \nu, \mathbf{Q}, \mathbf{V}_{1:T}, \mathbf{S}_{1:T}, \epsilon_{1:T}) &\propto \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\epsilon_k), V_t) \right] p_{k,0}^{(s_t)}(\theta_{k,0}^{(s_t)}) \\ &= \prod_{r=1}^R \prod_{t: S_t=r} \left[\prod_{k=1}^N c_{\theta_{k,0}^{(r)}}(F_{\nu_k^{(r)}}(\epsilon_k), V_t) \right] p_{k,0}^{(r)}(\theta_{k,0}^{(r)}) \end{aligned}$$

where the prior $p_{k,0}^{(r)}$ is a Beta density if as linking copulas we have Gumbel/Clayton or Joe copulas with parameter upper or lower tail, otherwise, if $\theta_{k,0}^{(r)} = (\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$, $p_{k,0}^{(r)}$ is the product between a shifted Beta on $(-1, 1)$ and a uniform distribution on \mathcal{H}_1 .

We proceed with a MH step for each copula parameter $\theta_{k,0}^{(r)}$ with proposal $q_{\theta,k}(\theta_{k,0}^{(r)*} | \theta_{k,0}^{(r)})$. More precisely, in case of $\theta_{k,0}^{(r)*} = (\tau_{k,0}^{(r)*}, \zeta_{k,0}^{(r)*})$ we sample $\zeta_{k,0}^{(r)*}$ from the discrete distribution on \mathcal{H}_1 with probabilities

$$q_{\zeta,k}(\zeta_{k,0}^{(r)*} = j | \tau_{k,0}^{(r)*}) = \frac{\prod_{t: S_t=r} c_{(\tau_{k,0}^{(r)*}, j)}(F_{\nu_k^{(r)}}(\epsilon_{k,t}), V_t)}{\sum_{h \in \mathcal{H}_1} \prod_{t: S_t=r} c_{(\tau_{k,0}^{(r)*}, h)}(F_{\nu_k^{(r)}}(\epsilon_{k,t}), V_t)},$$

and for $\tau_{k,0}^{(r)*}$, we use a random walk proposal defined on \mathbb{R} using a suitable transformation. The acceptance probability of the MH step for $(\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$ is

$$\min \left\{ 1, \prod_{t: S_t=r} \frac{c_{\theta_{k,0}^{(r)*}}(F_{\nu_k^{(r)}}(\epsilon_{k,t}), V_t) p_{k,0}(\theta_{k,0}^{(r)*}) q_{\tau,k}(\tau_{k,0}^{(r)} | \tau_{k,0}^{(r)*}) q_{\zeta,k}(\zeta_{k,0}^{(r)*} | \tau_{k,0}^{(r)*})}{c_{\theta_{k,0}^{(r)}}(F_{\nu_k^{(r)}}(\epsilon_{k,t}), V_t) p_{k,0}(\theta_{k,0}^{(r)}) q_{\tau,k}(\tau_{k,0}^{(r)*} | \tau_{k,0}^{(r)}) q_{\zeta,k}(\zeta_{k,0}^{(r)} | \tau_{k,0}^{(r)})} \right\}. \quad (4.6)$$

Analogously, if $\theta_{k,0}^{(r)}$ is equal to the upper or lower tail parameter, we use a random walk proposal on \mathbb{R} and the acceptance probability of the MH step is similar to (4.6).

Full conditional of $\mathbf{S}_{1:T}$. The full conditional of the $\mathbf{S}_{1:T}$ given $(\Phi, \nu, \mathbf{Q}, \mathbf{V}_{1:T}, \epsilon_{1:T})$ is equal to

$$\pi(\mathbf{S}_{1:T} | \Phi, \nu, \mathbf{Q}, \mathbf{V}, \epsilon_{1:T}) \propto \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta_{k,0}^{(s_t)}}(F_{\nu_k^{(s_t)}}(\epsilon_k), V_t) f_{\nu_k^{(s_t)}}(\epsilon_k) \right] \rho_{S_1} \prod_{t=2}^T q_{S_t, S_{t-1}}. \quad (4.7)$$

In order to sample from the full conditional (4.7), we rely on the forward filtering-backward sampling algorithm, see e.g Carter and Kohn (1994) and Frühwirth-Schnatter (1994). Indeed, one can rewrite the conditional probability of $\mathbf{S}_{1:T}$, given the innovations $\epsilon_{1:T}$ and the other parameters

$\Lambda := (\Phi, \nu, \mathbf{Q}, \mathbf{V}_{1:T})$, as

$$\begin{aligned} \mathbb{P}(\mathbf{S}_{1:T} = \mathbf{s}_{1:T} | \epsilon_{1:T}, \Lambda) &= \mathbb{P}(S_T = s_T | \epsilon_{1:T}, \Lambda) \mathbb{P}(S_{T-1} = s_{T-1} | S_T, \epsilon_{1:T-1}, \Lambda) \times \\ &\quad \mathbb{P}(S_{T-2} = s_{T-2} | S_{T-1}, \epsilon_{1:T-2}, \Lambda) \cdots \mathbb{P}(S_1 = s_1 | S_2, \epsilon_1, \Lambda) \\ &= \mathbb{P}(S_T = s_T | \epsilon_{1:T}, \Lambda) \prod_{t=1}^{T-1} \mathbb{P}(S_t = s_t | S_{t+1}, \epsilon_{1:t}, \Lambda), \end{aligned} \quad (4.8)$$

where we use that S_t , given S_{t+1} , does not depend on S_{t+2}, \dots, S_T and on $\epsilon_{t+2}, \dots, \epsilon_T$. Now Equation (4.8) suggests that we can simulate the whole chain $\mathbf{S}_{1:T}$ sequentially backward in time, i.e. we can sample S_T conditionally on $(\epsilon_{1:T}, \Lambda)$ and then, for $t = T - 1, \dots, 1$ generate S_t given $(\epsilon_{1:t}, \Lambda)$ and S_{t+1} .

The forward filtering-backward sampling algorithm is based on two step. In the first, one computes the filtered probabilities $\mathbb{P}(S_t = s | \epsilon_{1:t}, \Phi, \nu, \mathbf{Q}, \mathbf{V}_{1:T})$. In the second step, using the filtered probabilities, the state $(S_t)_{t=1:T}$ are drawn sequentially backward in time from $\mathbb{P}(S_t = s | S_{t+1} = r, \epsilon_{1:t}, \Phi, \nu, \mathbf{Q}, \mathbf{V}_{1:T})$.

For simplicity, we define $h(\epsilon_t | \Lambda, s) = \prod_{k=1}^N c_{\theta_k^{(s)}}(F_{\nu_k^{(s)}}(\epsilon_{k,t}), V_t) f_{\nu_k^{(s)}}(\epsilon_{k,t})$.

The two steps in details are:

- Forward filtering: for $t = 1$, the filtered probabilities are defined as

$$\mathbb{P}(S_1 = s | \epsilon_1, \Lambda) = h(\epsilon_1 | \Lambda, s) \rho_s \quad \text{for } s = 1, \dots, R.$$

For $t = 2, \dots, T$ the filtered probabilities are

$$\begin{aligned} \mathbb{P}(S_t = s | \epsilon_{1:t}, \Lambda) &\propto f_t(\epsilon_{1:t}, s | \Lambda) \\ &\propto \sum_{r=1}^R f_{t,t-1}(\epsilon_{1:t}, S_t = s, S_{t-1} = r | \Lambda) \\ &\propto h(\epsilon_t | \Lambda, s) \sum_{r=1}^R \mathbb{P}(S_{t-1} = r | \epsilon_{1:t-1}, \Lambda) \mathbf{q}_{s,r} \end{aligned}$$

where we denote with $f_t(\epsilon_{1:T}, s | \Lambda)$ the joint density of $\epsilon_{1:T}$ and $S_t = s$ given Λ . Analogously, $f_{t,t-1}(\epsilon_{1:t}, s, r | \Lambda)$ is the joint density of $\epsilon_{1:t}, S_t = s, S_{t-1} = r$ given Λ .

- Backward sampling: we sample S_T from $\mathbb{P}(S_T = s | \epsilon_{1:T}, \Lambda)$, the filtering probability at time T , obtained from the previous step. Then, for $t = T - 1, \dots, 1$ we sample each S_t from $\mathbb{P}(S_t = s | S_{t+1} = r, \epsilon_{1:t}, \Lambda)$, where

$$\begin{aligned} \mathbb{P}(S_t = s | S_{t+1} = r, \epsilon_{1:t}, \Lambda) &\propto \mathbb{P}(S_t = s, S_{t+1} = r, \epsilon_{1:t} | \Lambda) \\ &\propto \mathbf{q}_{r,s} \mathbb{P}(S_t = s | \epsilon_{1:t}, \Lambda). \end{aligned}$$

The quantities $\mathbb{P}(S_t = s | \epsilon_{1:t}, \Lambda)$ are known from the filtering step.

Full conditional of \mathbf{Q} . The full conditional of the transition matrix \mathbf{Q} given $(\Phi, \nu, \mathbf{S}_{1:T}, \mathbf{V}_{1:T}, \epsilon_{1:T})$ is

$$\begin{aligned} \pi(\mathbf{Q}|\Phi, \nu, \mathbf{S}_{1:T}, \mathbf{V}_{1:T}, \epsilon_{1:T}) &\propto \prod_{t=2}^T \mathbf{q}_{S_t, S_{t-1}} \prod_{r=1}^R \pi_{\mathbf{Q}}(\mathbf{Q}_{r\cdot}) = \\ &= \prod_{r=1}^R \prod_{s=1}^R \mathbf{q}_{r,s}^{N_S(r,s)} \pi_{\mathbf{Q}}(\mathbf{Q}_{r\cdot}) \end{aligned}$$

where $N_S(r, s) = \#\{t : S_t = r, S_{t-1} = s\}$, the number of transitions from state s to r , and $\pi_{\mathbf{Q}}(\mathbf{Q}_{r\cdot})$ is the Dirichlet prior on each row of \mathbf{Q} .

Since the Dirichlet and the multinomial distribution are conjugate distributions, also the conditional posterior distribution of each row r is a Dirichlet distributions with parameters $(\Psi_{1,r} + N_S(r, 1), \dots, \psi_{R,r} + N_S(r, R))$.

Full conditional of $\mathbf{V}_{1:T}$. The full conditional of $\mathbf{V}_{1:T}$ given $(\Phi, \nu, \mathbf{Q}, \mathbf{S}_{1:T}, \epsilon_{1:T})$ is proportional to

$$\pi(\mathbf{V}_{1:T}|\Phi, \nu, \mathbf{Q}, \mathbf{S}_{1:T}, \epsilon_{1:T}) \propto \prod_{t=1}^T \left[\prod_{k=1}^N c_{\theta^{(S_t)}}(F_{\nu_k^{(S_t)}}(\epsilon_k), V_t) \right].$$

We proceed with a MH step for each V_t with an uniform proposal density. For each t , we have the following acceptance probability

$$\min \left\{ 1, \prod_{k=1}^N \frac{c_{\theta^{(S_t)}}(F_{\nu_k^{(S_t)}}(\epsilon_k), V_t^*)}{c_{\theta^{(S_t)}}(F_{\nu_k^{(S_t)}}(\epsilon_k), V_t)} \right\}.$$

4.5 Simulation Study

We study the performance of the models described in Sections 4.1 and 4.3 with simulated data. In the following, first we show the results for the factor copula model and, then, for the Markov switching factor copula model. For both models, we perform sensitivity analysis with different values for the hyper-parameters and the results show that the choice of the prior hyper-parameters does not affect significantly the posterior estimates.

As for the factor copula model, we present two alternative scenarios in which a simulated dataset of $T = 300$ observations is obtained from a multivariate AR model with fixed parameters and a specific one-factor copula distribution. More precisely, in scenario 1 we simulate AR-observations using i.i.d. innovations with one factor copula based on Double Gumbel and Double Clayton linking copulas. Then, we apply the corresponding Bayesian model of Section 4.1 with prior specification as in (4.1). In scenario 2, we generate the data starting from a one factor copula with Gumbel linking copulas and we apply the factor copula model with bivariate Gumbel copulas. For both scenarios, we employ the MCMC algorithm described in Section 4.2 with 70000 iterations and a burn-in of 45000.

Table 4.1: Simulation study: Scenario 1. Factor copula model with Double Gumbel and Double Clayton linking copulas re-parameterized through the Kendall's tau. For each prior setting we show the posterior means of the parameters $\boldsymbol{\nu}$, \mathbf{A} , $\boldsymbol{\tau}$ and the posterior mode for $\boldsymbol{\zeta}$.

Scenario 1		Prior Settings	
$N = 6$	$(\tau_{k,0}, \zeta_{k,0})$	$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H}_1)$	$Beta_{(-1,1)}(2, 2) \times Unif(\mathcal{H}_1)$
$p = 2$	p_k	$Gamma(1, 0.5)$	$Gamma(1, 1)$
	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.4, 0.4], 10\mathbf{I}_p)$	$\mathcal{N}_p([0.4, 0.4], 100\mathbf{I}_p)$
	True Values	Posterior Means/Modes	
$\nu_k = \frac{1}{p_k}$	0.4	0.41	0.41
	0.5	0.49	0.53
	0.6	0.58	0.61
	0.9	0.87	0.93
	0.8	0.79	0.78
	0.7	0.72	0.70
$\boldsymbol{\alpha}_k$	(0.3, 0.5)	(0.31, 0.51)	(0.30, 0.46)
	(-0.3, 0.4)	(-0.36, 0.32)	(-0.13, 0.41)
	(0.1, 0.6)	(0.16, 0.56)	(0.07, 0.49)
	(-0.2, 0.4)	(-0.14, 0.39)	(-0.15, 0.48)
	(0.1, 0.5)	(0.22, 0.39)	(0.11, 0.48)
	(0.3, 0.5)	(0.26, 0.46)	(0.37, 0.47)
$(\tau_{k,0}, \zeta_{k,0})$	(0.2, DG_1)	(0.24, DG_1)	(0.17, DG_1)
	(-0.2, DC_1)	(-0.12, DC_1)	(-0.19, DG_1)
	(-0.5, DG_1)	(-0.58, DG_1)	(-0.58, DG_1)
	(0.3, DC_1)	(0.21, DC_1)	(0.24, DC_1)
	(-0.5, DG_1)	(-0.44, DG_1)	(-0.47, DG_1)
	(0.1, DC_1)	(0.11, DC_1)	(0.10, DC_1)

The results obtained with two alternative prior settings are reported in Tables 4.1 and 4.2, respectively. On each table are presented the two different prior settings, the values of the parameters used to generate the data and, the posterior means of $(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A})$. In case of Double copulas, we also report the posterior mode for $\boldsymbol{\zeta}$. For each scenario, we show the histograms of the simulated values of the posterior distributions of the parameters $(\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{A})$, see Figures 4.1 and 4.4. For the AR parameters, we report the histograms only for two series; for the other AR coefficients the results are similar. Furthermore, for scenario 1 and 2, the trace plots of the sample of two latent variables are displayed in Figures 4.3 and 4.5, respectively. Finally, in Figure 4.2 we report the trace plot and ergodic mean plot for some parameters of scenario 1. These diagnostic plots show a good convergence of the model. For the other scenario we obtain similar results.

Also for this model, we compute the DIC₃ criterion. In particular, in Table 4.3, we report the value of DIC₃ evaluated, for each scenario, with our fully Bayesian approach and the B-IFM procedure. To the data of scenario 1, we apply a B-IFM procedure using a one factor copula with linking Double copulas, while to the data of scenario 2 a B-IFM procedure with one factor copula distribution based on bivariate Gumbel copulas. Also in this case, the fully Bayesian models present

Figure 4.1: Histograms of the simulated values of the posterior distributions for the parameters of scenario 1 (first prior setting of Table 4.1), where we consider a factor copula model with Double Gumbel/Double Clayton bivariate copulas. For the parameters of the AR series, we report only the histograms only for the coefficients of the first and fourth series, i.e. α_1 and α_4 .

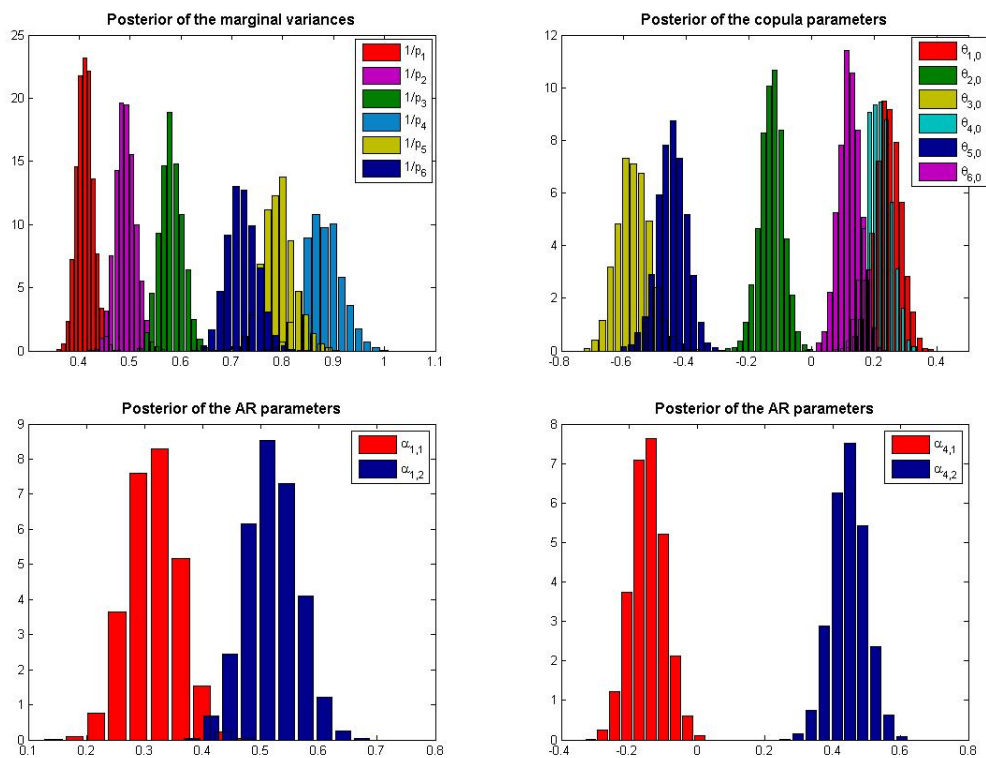


Figure 4.2: Diagnostic plot for the marginal and copula parameters of scenario 1 (first prior setting of Table 4.1), where we consider a factor copula model with Double Gumbel/Double Clayton bivariate copulas. For the parameters of the AR series, we report only the histograms only for the coefficients of the first and fourth series, i.e. α_1 and α_4 .

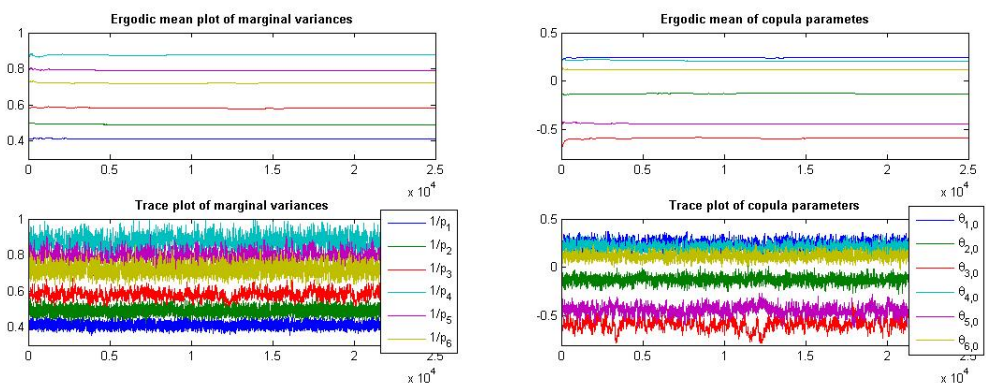
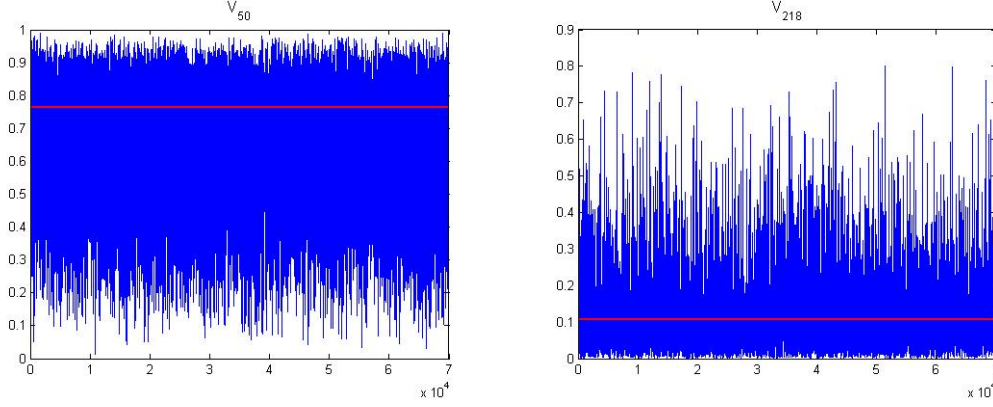


Figure 4.3: Posterior sample of two latent variables, V_{50} and V_{218} , for scenario 1 where we study a one factor copula models with linking Double Gumbel and Double Clayton copulas. The red line denotes the true value of the latent variable.



a better fit respect to the B-IFM ones.

Table 4.2: Simulation study: Scenario 2. Factor copula model with Gumbel linking copulas reparameterized through the upper tail parameter. For each prior setting we show the posterior means of the parameters ν , \mathbf{A} , θ .

Scenario 2		Prior Settings	
$N = 5$	$\theta_{k,0}$	$Beta(1, 1)$	$Beta(1, 2)$
$p = 3$	p_k	$Gamma(1, 1)$	$Gamma(1, 1)$
	α_k	$\mathcal{N}_p([0.3, 0.4, -0.1], 10\mathbf{I}_p)$	$\mathcal{N}_p([0.3, 0.4, -0.1], 100\mathbf{I}_p)$
	True Values	Posterior Means	
$\nu_k = \frac{1}{p_k}$	0.4	0.40	0.41
	0.5	0.52	0.53
	0.6	0.64	0.66
	0.7	0.77	0.73
	0.8	0.80	0.79
α_k	(0.3, -0.4, 0.2)	(0.28, -0.42, 0.24)	(0.34, -0.47, 0.29)
	(-0.3, 0.4, 0.2)	(-0.34, 0.38, 0.26)	(-0.21, 0.44, 0.19)
	(0.1, 0.6, -0.2)	(0.10, 0.58, -0.22)	(0.11, 0.60, -0.21)
	(-0.2, 0.4, 0.3)	(-0.15, 0.43, 0.25)	(-0.22, 0.37, 0.31)
	(0.1, 0.5, -0.1)	(0.13, 0.46, -0.16)	(0.07, 0.47, -0.06)
$\theta_{k,0}$	0.7	0.73	0.70
	0.2	0.20	0.15
	0.4	0.37	0.36
	0.3	0.29	0.28
	0.5	0.48	0.49

Figure 4.4: Histograms of the simulated values of the posterior distributions for the parameters of scenario 2 (first prior setting of Table 4.2), where we consider a factor copula model with Gumbel bivariate copulas. For the parameters of the AR series, we report only the histograms only for the coefficients of the second and fifth series, i.e. α_2 and α_5 .

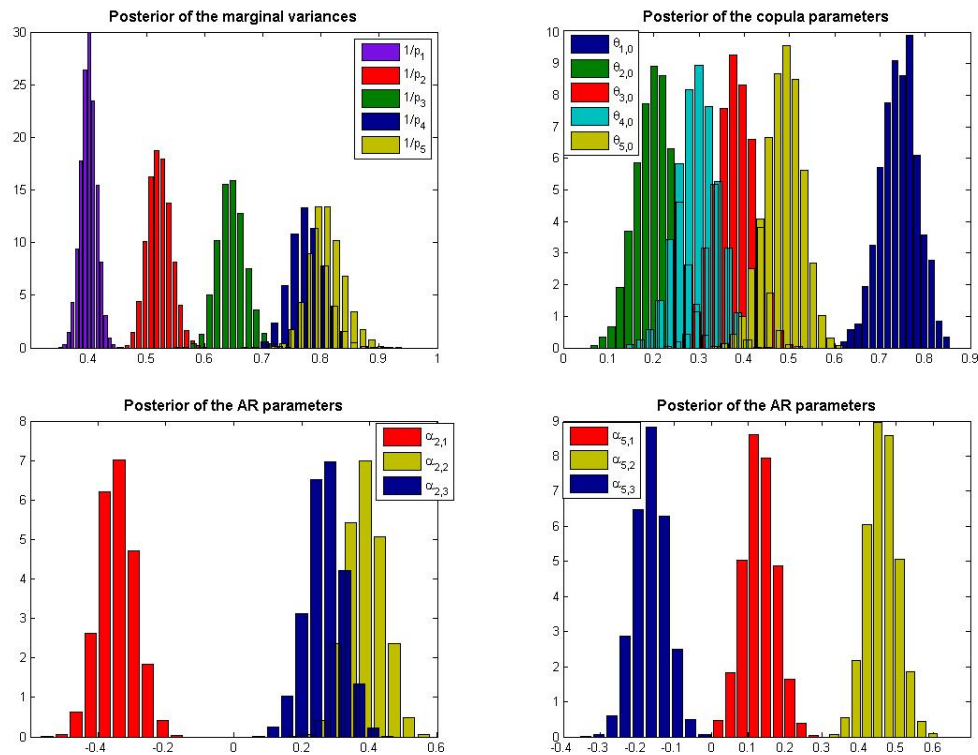


Figure 4.5: Posterior sample of two latent variables, V_{50} and V_{89} , for scenario 2 where we study a factor copula models with linking Gumbel copulas. The red line denotes the true value of the latent variables.

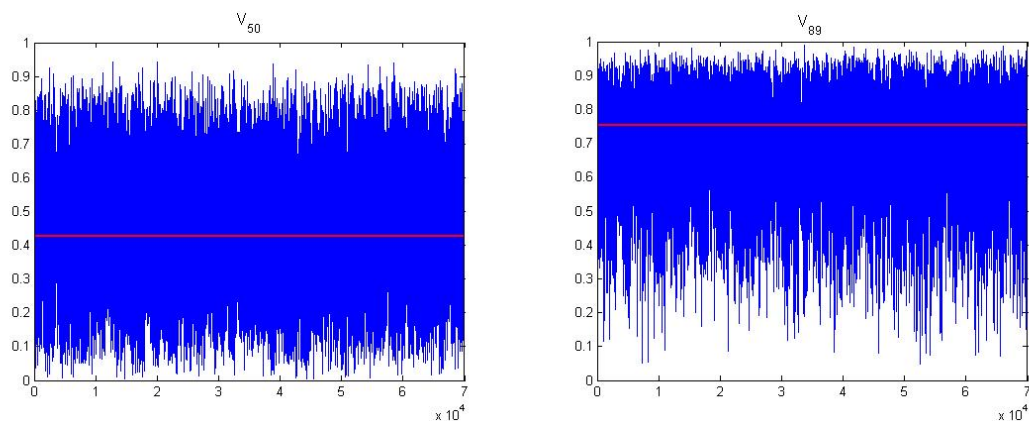


Table 4.3: Simulated Data: DIC₃ Table for factor copula models

	Fully Bayesian	B-IFM
Scenario 1 (Combined Factor Copula)	2540	2620
Scenario 2 (Gumbel Factor Copula)	2350	2470

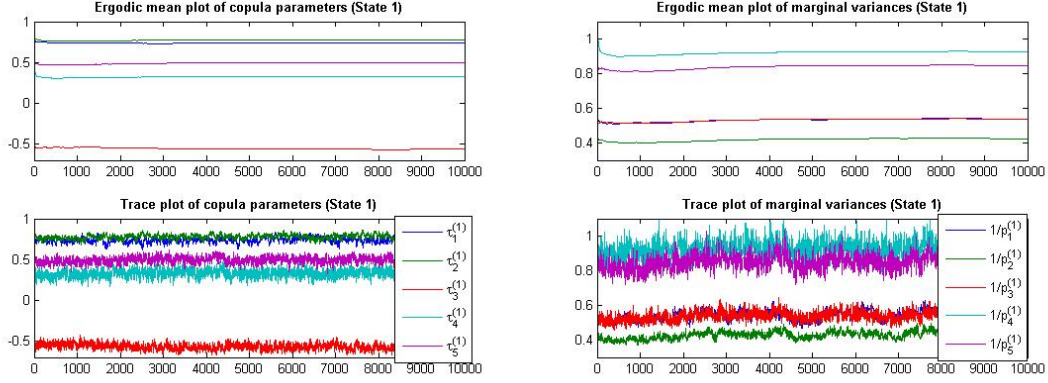
Table 4.4: Simulation study: Scenario 3. Markov Switching factor copula model with Double Gumbel and Double Clayton linking copulas re-parameterized through the Kendall's tau. For each prior setting we show the posterior means of the parameters ν , τ , \mathbf{Q} and the posterior mode for ζ .

Scenario 3	Prior Settings					
$N = 5$	$(\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$		$Beta_{(-1,1)}(2, 2) \times Unif(\mathcal{H}_1)$		$Beta_{(-1,1)}(1, 1) \times Unif(\mathcal{H}_1)$	
$R = 2$	$p_k^{(r)}$		$Gamma(2, 2)$		$Gamma(1, 2)$	
	$\mathbf{Q}_{r \cdot}$		$Dir(5, 5)$		$Dir(10, 10)$	
	True Values		Posterior Means/Modes			
\mathbf{Q}	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$		$\begin{bmatrix} 0.74 & 0.26 \\ 0.19 & 0.81 \end{bmatrix}$		$\begin{bmatrix} 0.79 & 0.21 \\ 0.28 & 0.72 \end{bmatrix}$	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$	$r = 1$	$r = 2$
$1/p_k^{(r)}$	0.5	0.7	0.52	0.80	0.52	0.70
	0.4	0.8	0.42	0.90	0.42	0.88
	0.5	0.9	0.53	0.99	0.52	1.03
	0.9	0.3	0.92	0.36	0.96	0.34
	0.8	0.2	0.84	0.26	0.83	0.24
$(\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$	(0.7, DG_1)	(-0.3, DC_1)	(0.73, DG_1)	(-0.38, DC_1)	(0.74, DG_1)	(-0.21, DC_1)
	(0.8, DC_1)	(0.5, DG_1)	(0.78, DC_1)	(0.58, DG_1)	(0.77, DC_1)	(0.56, DG_1)
	(-0.5, DG_1)	(0.6, DG_1)	(-0.56, DG_1)	(0.66, DG_1)	(-0.50, DG_1)	(0.63, DG_1)
	(0.4, DG_1)	(0.7, DG_1)	(0.33, DG_1)	(0.67, DG_1)	(0.43, DG_1)	(0.67, DG_1)
	(0.5, DG_1)	(0.6, DG_1)	(0.49, DG_1)	(0.66, DG_1)	(0.50, DG_1)	(0.68, DG_1)

Regarding the Markov switching model of Section 4.3, we present two alternative scenarios in which a simulated dataset of $T = 300$ observations is obtained from a Markov switching model with fixed parameters, fixed number of regimes and a specific one-factor copula distribution in each regime. More precisely, in scenario 3 we consider innovations simulated from a Markov switching factor model with 2 regimes, a specific transition matrix \mathbf{Q} and, in each regime, a one factor copula based on Double Gumbel and Double Clayton linking copulas. Then, we apply the corresponding Bayesian model of Section 4.3 with prior specification as in (4.4). In scenario 4 we generate the data starting from a Markov switching factor model with 2 regimes and one factor copulas with Gumbel linking copulas in each regime. We apply the Markov switching factor copula model with bivariate Gumbel copulas. For both the scenarios, we employ the MCMC algorithm described in Section 4.4 with 30000 iterations and a burn-in of 20000.

The results for scenario 3 and scenario 4 obtained with two alternative prior settings are reported in Tables 4.4 and 4.5, respectively. On each table are presented the two different prior settings, the

Figure 4.6: Diagnostic plots for the marginal and copula parameters of State 1 of scenario 3 (first prior setting of Table 4.4) where we study a Markov Switching factor copula models with linking Double copulas.



values of the parameters used to generate the data and the posterior means of (Φ, ν, Q) . In case of Double copulas, we also report the posterior mode for ζ . Regarding the convergence of the algorithm, as an example, in Figure 4.6 we display diagnostic plots for some parameters of scenario 3.

As for the hidden Markov chain, one can compute, for each time t , the posterior probability to be in state r , i.e.

$$\hat{\pi}_t(r) := \mathbb{P}(S_t = r | \epsilon_{1:T}),$$

for $r = 1, \dots, R$. In Figure 4.7 we report the posterior estimates of being in state r , $r = 1, 2$, obtained for scenario 3. In the same figure we also present the posterior mode \hat{S}_t of the hidden Markov chain and the corresponding true value. The analogous results for scenario 4 are presented in Figure 4.8. We note that the posterior mode of the hidden Markov chain associates correctly most of the observations to the true regime.

Figure 4.7: In the first two figures, we present the posterior probabilities of being in each state r ($r = 1, 2$) overtime for scenario3. In the third, we show the comparison between the posterior mode \hat{S}_t of the hidden Markov chain and the corresponding true value for the same scenario. The results refer to the first prior setting of Table 4.4

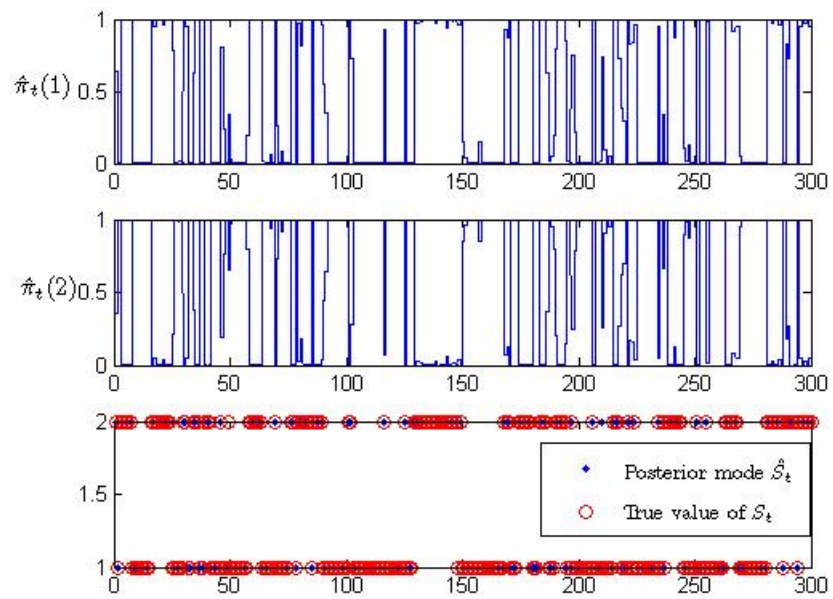
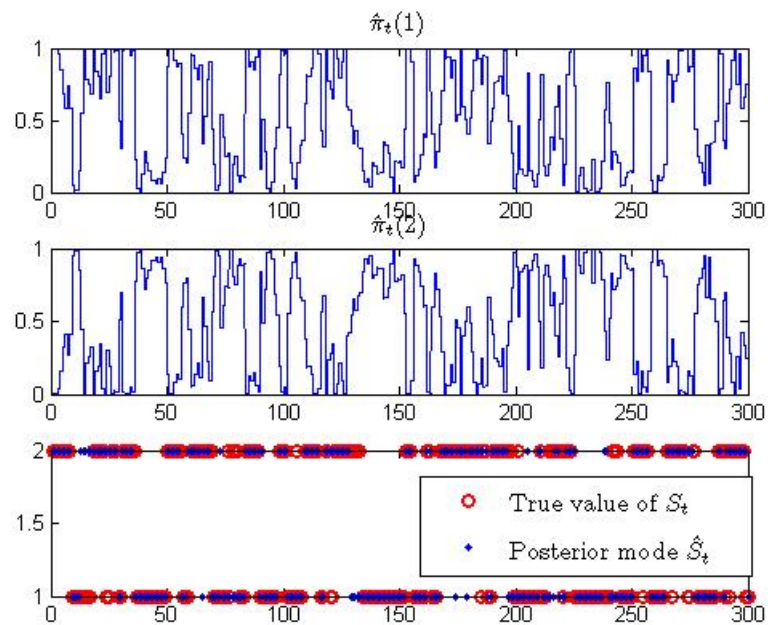


Table 4.5: Simulation study: Scenario 4. Markov Switching factor copula model with Gumbel linking copulas re-parameterized through the upper tail parameter. For each prior setting we show the posterior means of the parameters ν, θ, \mathbf{Q} .

Scenario 4	Prior Settings					
$N = 6$	$\theta_{k,0}^{(r)}$		$Beta(1, 1)$		$Beta(2, 2)$	
$R = 2$	$p_k^{(r)}$		$Gamma(1, 1)$		$Gamma(1, 2)$	
	\mathbf{Q}_r		$Dir(5, 5)$		$Dir(3, 3)$	
	True Values		Posterior Means			
\mathbf{Q}	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$		$\begin{bmatrix} 0.74 & 0.26 \\ 0.27 & 0.73 \end{bmatrix}$		$\begin{bmatrix} 0.76 & 0.24 \\ 0.24 & 0.76 \end{bmatrix}$	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$	$r = 1$	$r = 2$
$1/p_k^{(r)}$	0.8	0.5	0.78	0.53	0.78	0.54
	0.7	0.6	0.73	0.67	0.73	0.69
	0.9	0.7	0.91	0.76	0.89	0.81
	0.8	0.4	0.79	0.46	0.81	0.42
	0.7	0.8	0.65	0.88	0.69	0.83
	0.3	0.2	0.33	0.23	0.32	0.24
$\theta_{k,0}^{(r)}$	0.6	0.5	0.61	0.54	0.64	0.53
	0.7	0.6	0.75	0.57	0.67	0.59
	0.8	0.7	0.81	0.76	0.82	0.73
	0.9	0.5	0.89	0.54	0.96	0.56
	0.2	0.6	0.17	0.60	0.24	0.60
	0.3	0.5	0.35	0.55	0.28	0.52

Figure 4.8: In the first two figures, we present the posterior probabilities of being in each state r ($r = 1, 2$) overtime for scenario 4. In the third, we show the comparison between the posterior mode \hat{S}_t of the hidden Markov chain and the corresponding true value for the same scenario. The results refer to the first prior setting of Table 4.5



Chapter 5

Energy Data Analysis

In this Chapter we apply the different models presented in the previous part of the thesis to the analysis of energy market data in order to study the dependence structures among the examined commodities and to proceed with portfolio analysis and evaluation. Since for the Markov switching factor copula model we have only preliminary results, we discuss this model in a distinct Section and, first, we focus on tree copula, mixture (finite and infinite) of tree copulas and, one factor copula models.

For the tree copula, mixture of tree copulas and factor copula models, we consider daily time series of one-year forward contracts for Power Italy, Power Germany, Brent, TTF, PSV, CO_2 and Api2. The power prices are obtained from EEX¹, the remaining commodities from Reuters². All commodities prices are expressed in Euros. The data cover the period from January 2014 to December 2014. The limited amount of data is due to the intrinsic nature of the examined one-year contracts. Since the dependence structure can differ significantly from one year to another, we only consider data referring to a specific year. For the Markov switching model, that is a more flexible model, we consider daily time series of one-year forward contracts from January 2013 to December 2014 of the same commodities listed previously.

Power Italy and Power Germany are the energy prices for Italian and German markets; TTF and PSV are prices for natural gas deriving from transactions in virtual trading points in Netherlands and Italy; Api2 index is the standard reference price benchmark for coal imported into northwest Europe; Brent is one of the major classifications of oil and can serve as a major benchmark price for purchases of oil worldwide, and CO_2 represents the price to pay for the emission of carbon dioxide into the atmosphere. Under the Kyoto protocol, OECD countries must reduce their emissions of greenhouse gas. European plants with large CO_2 emissions obtain from their governments allowances to emit metric tons of CO_2 equivalent and these permissions can be traded in spot, future and option markets. In the European Union, the higher production of CO_2 emissions is concentrated on the

¹<https://www.eex.com>

²<http://www.reuters.com>

Table 5.1: Descriptive statistics for the examined commodities.

	Power Italy	Power Germany	Brent	Api2	TTF	PSV	CO_2
Mean	53.71	35.07	72.84	58.82	25.77	27.47	6.095
Std	1.69	0.82	6.59	1.63	0.92	0.86	0.63
Min	49.80	33.77	50.67	54.68	23.05	24.81	4.48
Max	58.45	36.90	79.66	63.85	27.81	29.10	7.42
Kurt	3.67	2.31	6.94	2.43	2.89	2.70	2.32
Skew	0.86	0.64	-1.97	-0.55	-0.20	-0.38	-0.13

power generation sector. As a consequence of the introduction of CO_2 emission constraints on power generators in the European Union, climate policy is starting to have notable effects on energy market, see e.g. Reinaud (2007). For this reason, CO_2 should be included in the list of the variables that influence the energy price.

Our aim is to model the dependence structure among each power price (Italian or German) and the other commodities (gas, coal, oil and CO_2) by means of suitable copulas. It is worth noticing that PSV refers only to the Italian market, we don't consider it in the analysis of the German market. Table 5.1 lists descriptive measures for the considered commodities while in Figure 5.1 are shown the examined time series.

For each series we compute the monthly logarithmic return rates $X_{t,k} = \ln \left\{ S_{t+20,k} / S_{t,k} \right\}$ where $S_{t,k}$ is the price at each day t of commodity i . The monthly logarithmic return rates, $X_{t,i}$, are modeled with an AR(3) and we suppose dependent innovations with Normal marginal distributions. The lag of the AR models has been selected via preliminary analysis on the data, and using AIC/BIC criteria.

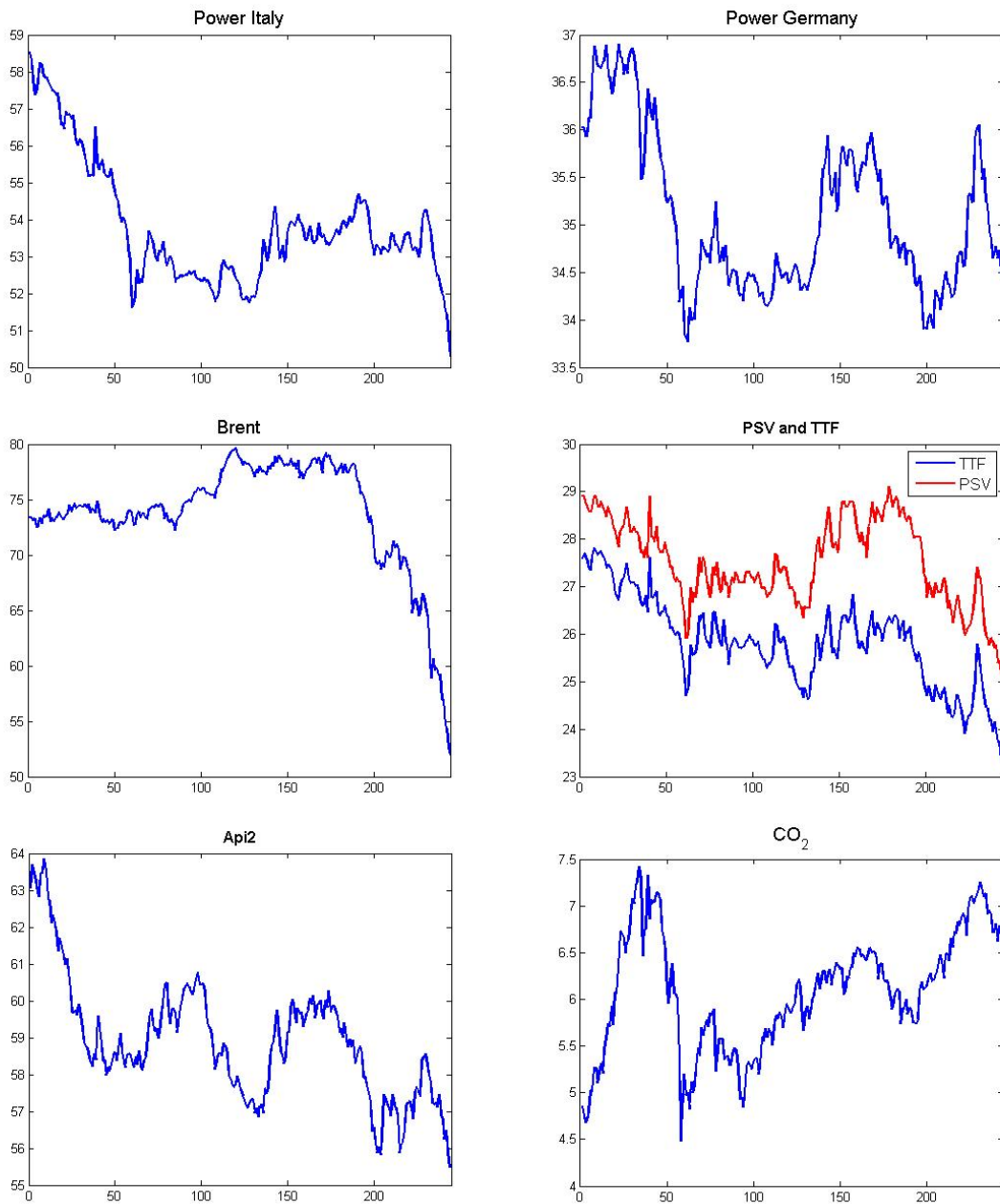
The outline of this Chapter is the following. In Section 5.1 we present the alternative models based on tree copula, mixture of tree copulas and factor copula distributions considered for energy data analysis and the respective DIC/DIC₃ values. In Section 5.2 we proceed with portfolio evaluations for each domestic market and in Section 5.3 we consider dependence structure analysis among each power price and the remaining commodities with the models listed in Section 5.1. Finally, some preliminary results obtained with Markov switching factor copula model are showed in Section 5.4.

5.1 Model Selection

In order to analyze the dependence among the examined commodities, we consider different models where the likelihood of the innovations is represented via one of the copula structure described in Chapters 2, 3 and 4, i.e. tree copula, mixture of tree copulas and factor copula.

In the case of tree copula and finite mixture of tree copulas (Chapter 2), for the innovations of the AR(3) series, we consider the following models :

Figure 5.1: Time series plots of the examined commodities.



- i) Gumbel tree model; innovations distributed according to a tree copula where all $c_{\theta_{l,m}}$ are Gumbel copulas and $\theta_{l,m} = \lambda_{l,m}^U$,
- ii) Joe tree model; innovations distributed according to a tree copula where all $c_{\theta_{l,m}}$ are Joe copulas and $\theta_{l,m} = \lambda_{l,m}^U$,
- iii) Clayton tree model; innovations distributed according to a tree copula where all $c_{\theta_{l,m}}$ are Clayton copulas and $\theta_{l,m} = \lambda_{l,m}^L$,
- iv) Combined tree model; innovations distributed according to a tree copula where on each edge the copula can be either a Double Clayton or a Double Gumbel and $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m})$,
- v) Gumbel mixture model; innovations distributed according to a mixture of tree copulas where all $c_{\theta_{l,m}^{(d)}}$ are Gumbel copulas and $\theta_{l,m}^{(d)} = \lambda_{l,m}^{U(d)}$,
- vi) Joe mixture model; innovations distributed according to a mixture of tree copulas where all $c_{\theta_{l,m}^{(d)}}$ are Joe copulas and $\theta_{l,m}^{(d)} = \lambda_{l,m}^{U(d)}$,
- vii) Clayton mixture model; innovations distributed according to a mixture of tree copulas where all $c_{\theta_{l,m}^{(d)}}$ are Clayton copulas and $\theta_{l,m}^{(d)} = \lambda_{l,m}^{L(d)}$,
- viii) Combined mixture model; innovations distributed according to a mixture of tree copulas where on each edge the copula can be either a Double Clayton or a Double Gumbel and $\theta_{l,m}^{(d)} = (\tau_{l,m}^{(d)}, \zeta_{l,m}^{(d)})$.

For these models, we consider the following independent prior distributions. On $\theta_{l,m}$ and on $\mathcal{E}_{\mathcal{N}}$ we assign uniform priors; on ν_k we assume a Normal-Gamma prior with parameters $(1, 1, 0, 0.1)$. On α_k we set a 3-dimensional Normal prior centered on the forward-backward estimates of the AR coefficients and covariance matrix $10\mathbf{I}_p$. For the weights of the mixture models we use a Dirichlet prior with hyperparameters equal to 10. Sensitivity analysis shows that the choice of the prior settings does not affect significantly the posterior estimates.

In case of infinite tree copula mixture model (Chapter 3), we apply to the innovations of the AR(3) series the following structures:

- ix) Combined DP-tree model; innovations distributed according to an infinite mixture of tree copulas where on each edge $c_{\theta_{l,m}^{(d)}}$ can be either a Double Clayton or a Double Gumbel copula and $\theta_{l,m}^{(d)} = (\tau_{l,m}^{(d)}, \zeta_{l,m}^{(d)})$;
- x) Gumbel DP-tree model; innovations distributed according to an infinite mixture of tree copulas where all $c_{\theta_{l,m}^{(d)}}$ are Gumbel copulas and $\theta_{l,m}^{(d)} = \lambda_{l,m}^{U(d)}$.

Table 5.2: Energy Data: DIC and DIC₃ Table. The DIC value is reported in brackets.

	Italy	Germany
Gumbel Tree Model	706 (726)	714 (731)
Clayton Tree Model	779 (792)	782 (794)
Joe Tree Model	905 (920)	772 (786)
Combined Tree Model	653 (675)	668 (694)
Gumbel Mixture Model 2 Comp.	669	704
Clayton Mixture Model 2 Comp.	714	729
Joe Mixture Model 2 Comp.	892	754
Combined Mixture Model 2 Comp.	641	680
Gumbel Mixture Model 5 Comp.	660	695
Clayton Mixture Model 5 Comp.	696	726
Joe Mixture Model 5 Comp.	862	749
Combined Mixture Model 5 Comp.	695	775
Combined DP-tree Model	597	645
Gumbel DP-tree Model	667	709
Combined Factor Model	856	769
Gumbel Factor Model	865	774

For these models, we use the following independent prior distributions. On $\theta_{l,m}$ and on $\mathcal{E}_{\mathcal{N}}$ we assign uniform priors; on ν_k we assume a Normal-Gamma prior with parameters $(1, 1, 0, 0.1)$. On α_k , the parameters of the AR models, we set a 3-dimensional Normal prior centered on the forward-backward estimates of the AR coefficients and covariance matrix $10\mathbf{I}_p$. On the concentration parameter ψ we assume a Gamma prior with mean 0.4 and variance 0.8 for the Italian market, while for the German market a Gamma prior with mean 0.3 and variance 0.6. We consider other values for the Gamma prior on ψ , but these choices showed the best results in terms of DIC₃ criterion.

Finally, for the factor copula structure (Chapter 4), we apply to the innovations the following models

- xi) Combined factor model; innovations distributed according to a one-factor copula where on each edge $c_{\theta_{l,0}}$ can be either a Double Clayton or a Double Gumbel copula and $\theta_{l,0} = (\tau_{l,0}, \zeta_{l,0})$;
- xii) Gumbel factor model; innovations distributed according to a one-factor copula where all $c_{\theta_{l,0}}$ are Gumbel copulas and $\theta_{l,0} = \lambda_{l,0}^U$.

In this case, we consider the next independent prior distributions: on $\theta_{l,0}$ we use an uniform prior and on α_k we assume a 3-dimensional Normal distribution centered on the forward-backward estimates of the AR coefficients and with covariance matrix $10\mathbf{I}_p$. Finally, on the marginals parameters we assign a Gamma prior with parameters $(1, 1)$; indeed in the factor model we assume that the marginal means are known and set equal to the empirical mean.

The alternative models are compared via DIC₃ and DIC (when feasible), see Table 5.2. Using these criteria, the Combined DP-tree model presents the lowest values of DIC₃, both for the Italian

Table 5.3: VaR and ES for Italian and German portfolios at time $T+1$, T corresponding to 15th December 2014.

	Italy		Germany	
	VaR	ES	VaR	ES
Joe tree model	24.63	24.44	-2.09	-2.28
Gumbel tree model	24.74	24.58	-2.09	-2.20
Clayton tree model	24.71	24.55	-1.97	-2.15
Combined tree model	24.78	24.59	-2.05	-2.22
Joe mixture model 2comp.	24.71	24.61	-1.99	-2.18
Gumbel mixture model 2comp.	24.85	24.62	-2.05	-2.27
Clayton mixture model 2comp.	24.95	24.77	-1.98	-2.18
Combined mixture model 2comp.	24.55	24.28	-2.13	-2.34
Combined DP-tree model	24.69	24.51	-2.10	-2.34
Gumbel DP-tree model	24.64	24.27	-2.13	-2.33
Combined factor model	24.69	24.49	-2.17	-2.36
Gumbel factor model	24.73	24.54	-2.14	-2.35

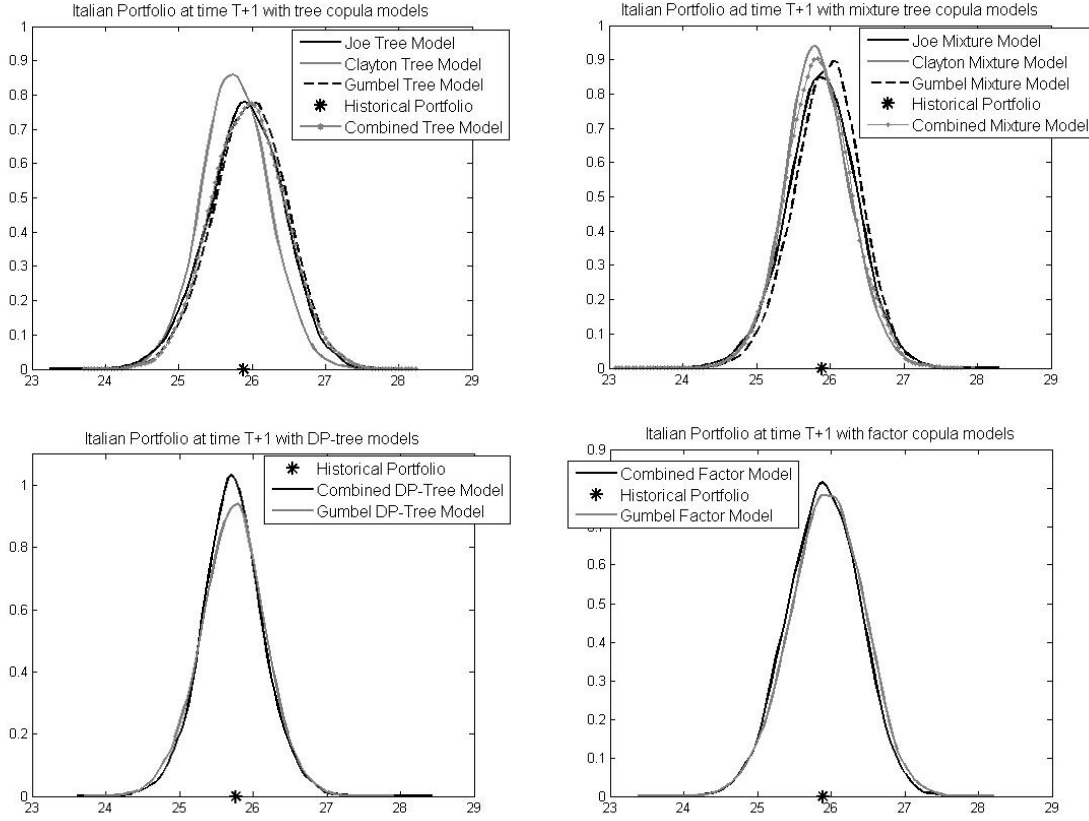
and the German market. Among the tree and mixture of tree models, we note that the Combined tree model and the Combined mixture model with two components show better results. Moreover, according to $DIC \setminus DIC_3$, the mixture models with more than two components do not improve substantially the fitting of the data. As an exemplification, in Table 5.2, we report the results regarding the mixture with five components for models v), vi) vii) and viii). This fact is coherent with the results obtained with the DP-tree models. Indeed, with the Combined DP-tree model, for both the Italian and the German market we find that the posterior mode of the number of cluster is 2, see Section 5.3.2. We also observe that the DIC_3 obtained for the factor models are higher than most of the other values. As we will show in Section 5.3.3, these models estimate strong correlation between the latent node and a commodity (TTF for Italian market and Api2 for German market). Therefore, the presence of the latent node do not give other substantial information with respect to a tree model and the presence of a fixed underlying tree structure, that it is not the optimal one, can explain the higher values of the DIC_3 .

The DIC_3/DIC associated to the B-IFM models are similar to the ones obtained with the fully Bayesian procedure. Nevertheless, for the portfolio analysis, see Section 5.2, the B-IFM results are substantially inferior to the fully Bayesian ones.

5.2 Energy Market Portfolio Analysis

In energy market framework, it is important to perform portfolio analysis and evaluation. For each domestic market, we consider a portfolio made of one power asset and the remaining commodities. We work in the perspective of an energy company that sells energy and buys the other commodities in order to produce it. Hence, in the portfolio composition, power has a positive weight,

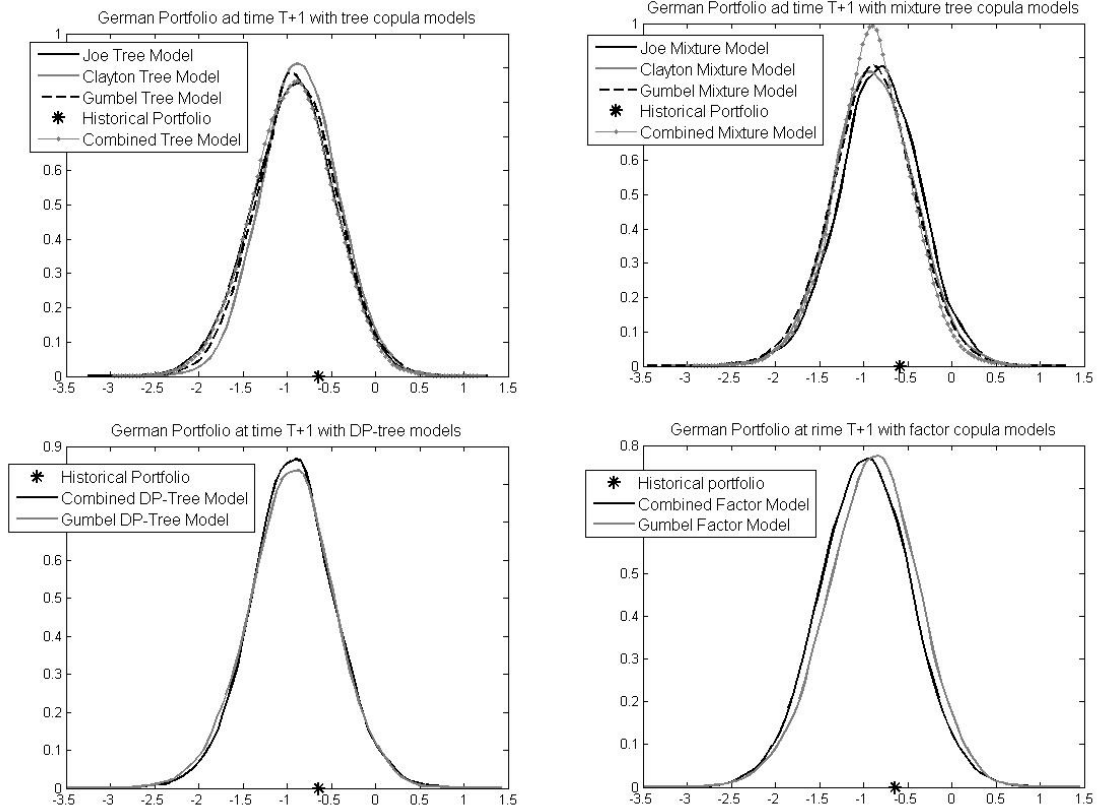
Figure 5.2: Italian predictive portfolio density at time $T+1$, T corresponding to 15th December 2014 with tree models (first row on the right), mixture of trees models with 2 components (first row on the left), DP-tree models (second row on the right) and factor models (second row on the left).



q_1 , while the remaining components should have negative weights, $-q_i$ ($i = 2, \dots, N$). Consequently, the value of the portfolio at time t is given by $V_t = q_1 S_{t,1} - \sum_{i=2}^N q_i S_{t,i}$.

For portfolio evaluation, we compute two well-known quantile risk measures: the Value-at-Risk (VaR) and the Expected Shortfall (ES), see e.g. Klumngman et al. (2008), Szegö (2005). The VaR of a portfolio, at a given future time $T+k$ (at given probability level β), is $P\{V_{T+k} \leq VaR(T+k, \beta)\} = \beta$. It is one of the most used risk measure in applications, since it is easy to estimate and to explain even to non-experts. The 99% VaR for a horizon of two weeks is acceptable for the Basel Committee on Banking and Supervision of Banks for International Settlement (Basel Committee (1995) and following amendments). Nevertheless, many authors have criticized its adequacy as a measure of risk, see e.g. Acerbi and Tasche (2002). One drawback is that it does not provide any information about the potential size of loss that exceeds the VaR level and this may lead to extremely risky investment strategies. Furthermore, it is not a coherent risk measure since it lacks the property of sub-additivity. In order to overcome the previous problems, Artzner et al. (1999) proposed to use the Expected Shortfall, defined as $E[V_{T+k} | V_{T+k} < VaR(T+k, \beta)]$. It measures the expected value

Figure 5.3: German predictive portfolio density at time $T+1$, T corresponding to 15th December 2014 with tree models (first row on the right), mixture of trees models with two components (first row on the left), DP-tree models (second row on the right) and factor models (second row on the left).



of a portfolio given that the VaR has been exceeded. It has at least two important advantages over the VaR; it does quantifies losses exceeding the VaR level and it is a sub-additive coherent measure of risk.

Using a Bayesian approach, we construct the portfolio predictive distribution at time $T+k$ on the basis of the information up to time T and we then compute the related Bayesian predictive VaR and ES, $P\{V_{T+k} \leq VaR_{T+k}(\beta)|\mathcal{O}_T\} = \beta$ and $E[V_{T+k}|V_{T+k} < VaR(T+k, \beta), \mathcal{O}_T]$, respectively. The previous quantities can be easily approximated using the MCMC output, see e.g. Osiewalski and Pajor (2010).

For the analysis of the portfolio, we consider the models described in Section 5.1. For the Italian market we construct an industrial portfolio made of Power Italy, TTF, PSV, Api2, CO_2 , Brent with weights $[1, 0, 0.14, 0.28, 0.69, 0]$. For the German market, we use a trading portfolio made of Power Germany, TTF, Api2, CO_2 , Brent with weights $[1, 0.61, 0.27, 0.76, 0]$. The composition of each portfolio has been chosen according to information provided by experts of Enel group (Global

Figure 5.4: One day ahead Italian portfolio estimates with 95% credible intervals (Combined mixture model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.

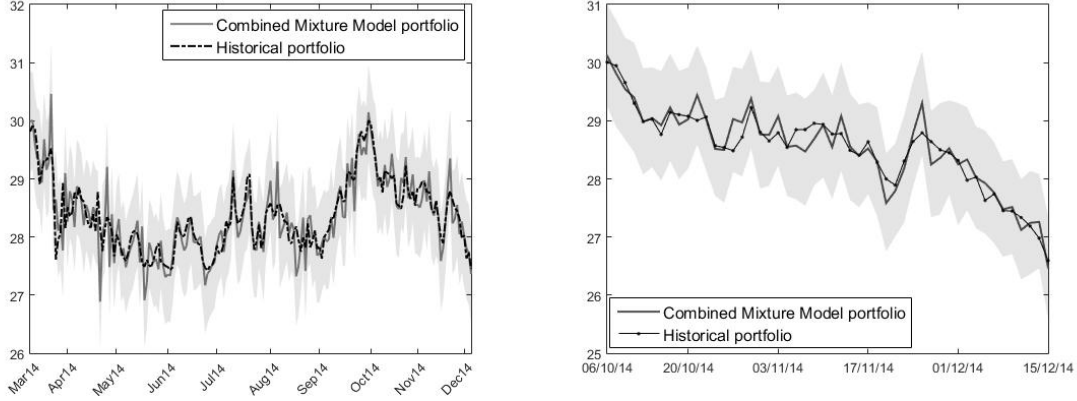
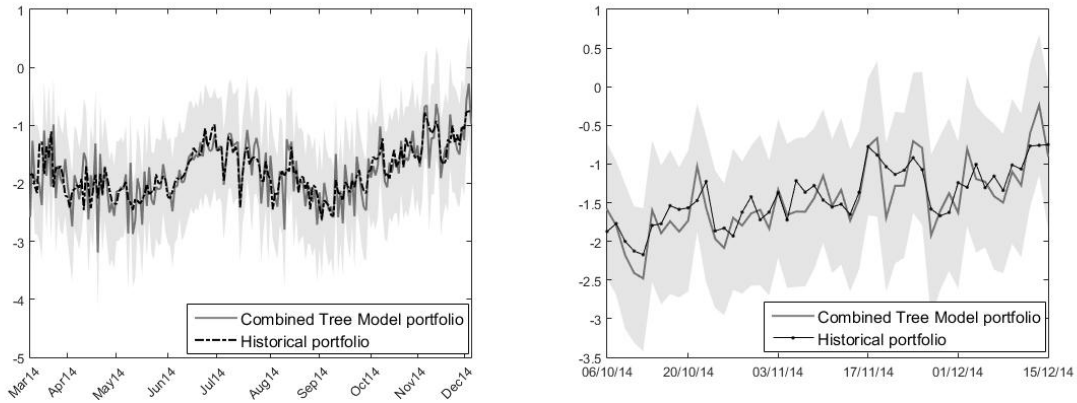


Figure 5.5: One day ahead German portfolio estimates with 95% credible intervals (Combined tree model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.



Trading, Modeling & Pricing Division).

The predictive portfolio distribution at time $T + 1$ for Italian/German market obtained with all the alternative models are shown in Figure 5.2/Figure 5.3. We assume $T = 224$, corresponding to 15th December 2014. In addition, the VaR and ES for each portfolio distribution are reported in Table 5.3. The predictive distributions are centered around the historical portfolio value.

We also analyze the forecasting performance of our models using both in-sample and an out-of-sample analysis for one day ahead portfolio estimation. For the in-sample analysis, we consider the last 195 observations, while for the out-of-sample portfolio we consider the last 50 observations and we estimate the portfolio by the output of the MCMC algorithm (50000 iterations) for each $t = 174, \dots, 224$. As for the tree and mixture of tree models, we present the results obtained with the Combined mixture model for the Italian portfolio (Figure 5.4) and with the Combined tree model

Figure 5.6: One day ahead Italian portfolio estimates with 95% credible intervals (Combined DP-tree model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.

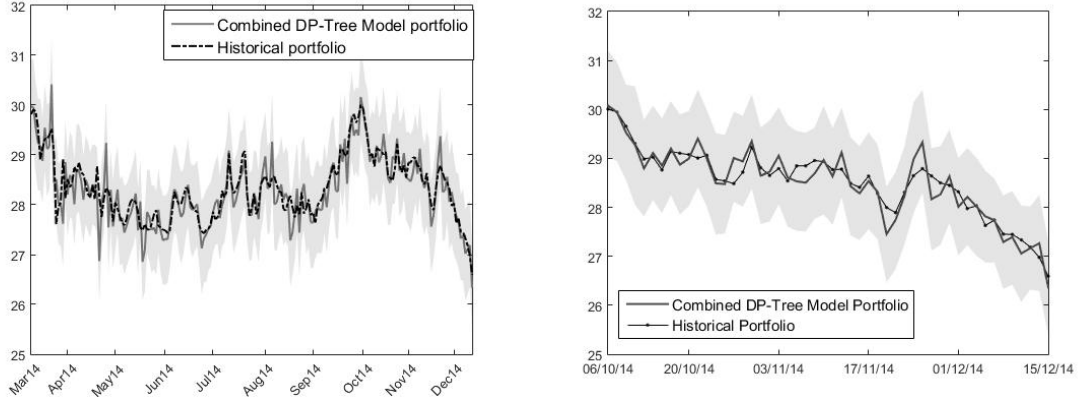
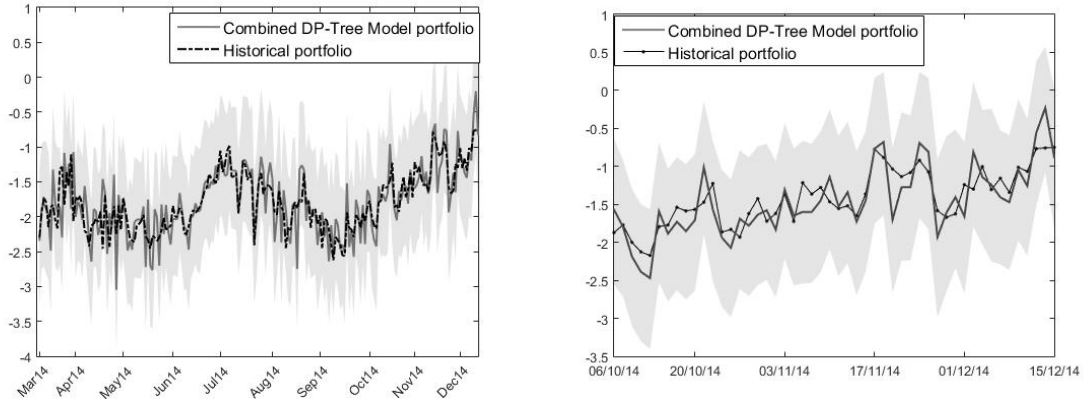


Figure 5.7: One day ahead German portfolio estimates with 95% credible intervals (Combined DP-tree model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.



for the German portfolio (Figure 5.5). In Figures 5.6/5.7 are shown the Italian and the German forecasting portfolios computed with the Combined DP-tree model and the analogous results with the Combined factor model are reported in Figures 5.8/5.9. Note that, there is a good agreement between the behaviors of predictive portfolios and the historical ones.

Finally, Table 5.4 reports the mean distances between the in-sample forecasting portfolio and the historical portfolio, that is

$$\frac{1}{\tau_1 - \tau_0} \sum_{t=\tau_0}^{\tau_1} \left| \mathbb{E}[V_{t+1} | \mathcal{O}_t] - V_{t+1} \right|,$$

where $\tau_0 = 174$ and $\tau_1 = 224$. For each model, we report the mean distance obtained using our fully

Figure 5.8: One day ahead Italian portfolio estimates with 95% credible intervals (Combined factor model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.

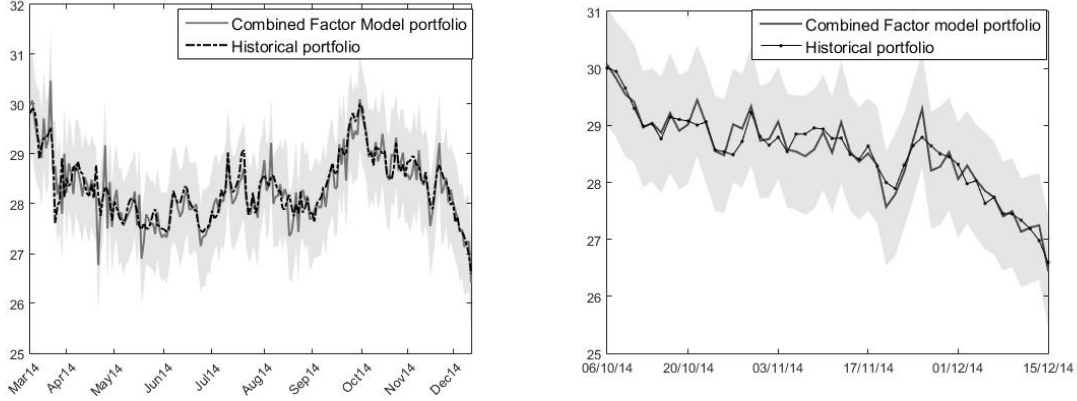
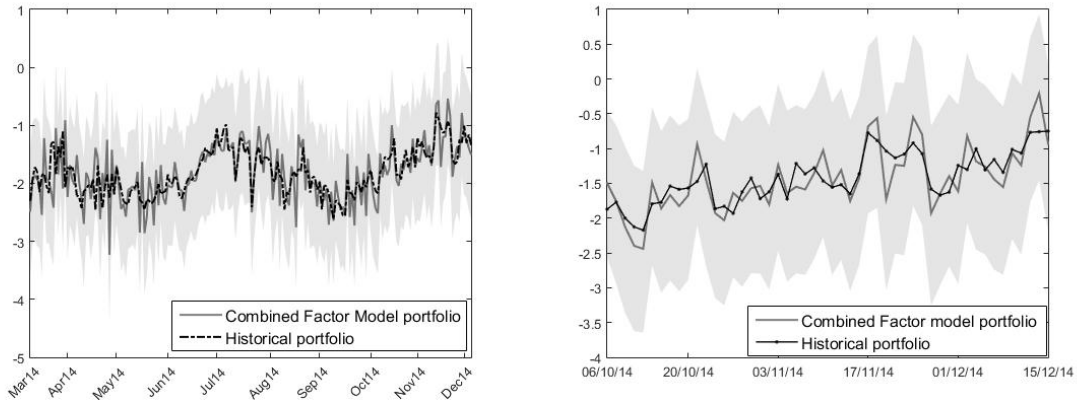


Figure 5.9: One day ahead German portfolio estimates with 95% credible intervals (Combined factor model); in-sample analysis on the left, out-of-sample (on the last 50 observations) on the right.



Bayesian approach and the B-IFM. For comparative purposes, we also considered and estimated a model in which the innovations are assumed multivariate normally distributed. In this case, for the marginal parameters we assume a Normal-Wishart prior distribution and a p -dimensional Normal prior on the AR parameters. For both markets, the Combined mixture model and the Combined DP-tree model show the minimum mean distances among all the examined models (including the multivariate Normal one).

We can observe that the distances obtained with the alternative models are not so different, while the values obtained with the B-IFM are systematically higher than the corresponding distances computed with our fully Bayesian procedure.

As a further analysis on the performance of our models, for each domestic market, we consider the equally weighted portfolio, i.e. $q_1 = q_2 = \dots = q_N$ and $\sum_{i=1}^N q_i = 1$. Also for this kind of portfolio,

Table 5.4: Mean distance between the in-sample predictive portfolio and historical portfolio on the last 50 observations for the Italian market and the German market with fully Bayesian and B-IFM procedures.

	Italy		Germany	
	Fully Bayesian Mean Distance	B-IFM Mean Distance	Fully Bayesian Mean Distance	B-IFM Mean Distance
Joe tree model	0.186	0.191	0.206	0.215
Gumbel tree model	0.168	0.193	0.198	0.221
Clayton tree model	0.195	0.193	0.206	0.213
Combined tree model	0.175	0.188	0.201	0.218
Joe mixture model 2 Comp.	0.179	0.193	0.206	0.217
Gumbel mixture model 2 Comp.	0.170	0.188	0.204	0.215
Clayton mixture model 2 Comp.	0.170	0.189	0.199	0.210
Combine mixture model 2 Comp.	0.165	0.188	0.195	0.226
Combined DP-tree model	0.162	0.196	0.196	0.227
Gumbel DP-tree model	0.167	0.186	0.201	0.216
Combined factor model	0.172	0.190	0.220	0.215
Gumbel factor model	0.175	0.192	0.207	0.214
Multivariate normal model	0.192	0.192	0.199	0.211

Table 5.5: Equally weighted portfolio: mean distance between the in-sample predictive portfolio and historical portfolio on the last 50 observations for the Italian market and the German market with a fully Bayesian procedure.

	Italy	Germany
	Fully Bayesian Mean Distance	Fully Bayesian Mean Distance
Joe tree model	0.177	0.180
Gumbel tree model	0.161	0.182
Clayton tree model	0.157	0.180
Combined tree model	0.160	0.176
Joe mixture model 2 Comp.	0.158	0.184
Gumbel mixture model 2 Comp.	0.159	0.183
Clayton mixture model 2 Comp.	0.157	0.183
Combine mixture model 2 Comp.	0.157	0.178
Combined DP-tree model	0.154	0.169
Gumbel DP-tree model	0.156	0.175
Combined factor model	0.161	0.177
Gumbel factor model	0.163	0.177
Multivariate normal model	0.164	0.186

we compute the mean distances between the in-sample forecasting portfolio and the historical one on the last 50 observations using our fully Bayesian procedures. The results are presented in Table 5.5. As in the previous case, the distances related to the German market are higher than the ones obtained for the Italian market. Furthermore, we observe that, also for the equally weighted portfolio, the Combined DP-tree model shows the best results for both markets respect to all the other examined models.

Table 5.6: Posterior mean of the Kendall's tau between the innovations, for the Italian market (first table) and the German market (second table), obtained respectively with the Combined tree model (first row) and the Combined mixture model with two components (second row).

Combined Tree Model						
	PowIT	TTF	PSV	Api2	CO ₂	Brent
PowIT	1	0.27	0.25	0.11	-0.02	0.04
TTF		1	0.72	0.33	-0.06	0.07
PSV			1	0.30	-0.06	0.07
Api2				1	-0.12	0.09
CO ₂					1	-0.01
Brent						1

Combined Mixture Model						
	PowIT	TTF	PSV	Api2	CO ₂	Brent
PowIT	1	0.20	0.18	0.13	0.09	0.04
TTF		1	0.68	0.31	-0.05	0.07
PSV			1	0.29	-0.02	0.06
Api2				1	-0.11	0.09
CO ₂					1	0.001
Brent						1

Combined Tree Model						
	PowDE	TTF	Api2	CO ₂	Brent	
PowDE	1	-0.04	-0.02	0.13	0.01	
TTF		1	0.34	-0.07	0.07	
Api2			1	-0.09	0.10	
CO ₂				1	-0.006	
Brent					1	

Combined Mixture Model!						
	PowDE	TTF	Api2	CO ₂	Brent	
PowDE	1	0.06	-0.02	0.07	0.008	
TTF		1	0.32	0.005	0.07	
Api2			1	-0.04	0.10	
CO ₂				1	-0.01	
Brent					1	

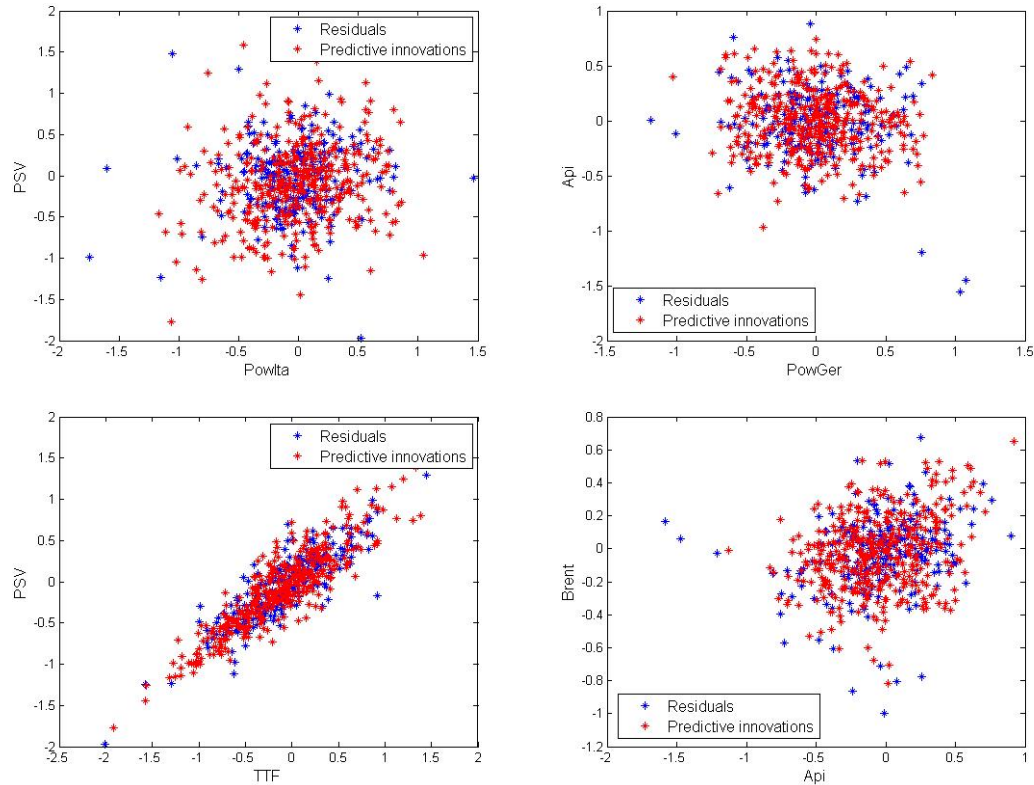
5.3 Dependence Structure Analysis

Since we are interested in dependence analysis among the commodities of each domestic market, we apply the models listed in Section 5.1 in order to get information about the dependence structures. In the following we present the results obtained with the alternative copula-based structures. For every model, as an estimate of the Kendall's tau between the innovations of the AR models, we compute its posterior mean. Moreover, for the tree copula model, we show the MAP tree structure that gives a graphical representation of the dependence among the examined variables. For the others models, we present the weighted graph Γ_τ based on the absolute values of the estimated Kendall's tau and, using the MST construction, we get the associated representative tree τMST . In case of mixture and DP-tree models, we also show the weighted graph Γ_w and the related minimum spanning tree $wMST$ built using the posterior mean of the quantity $\Upsilon_{(l,m)}$, see Equation (1.11).

5.3.1 Tree Copula and Mixture of Tree Copulas Models

In this Section we focus on models i)-viii) presented in Section 5.1. For each model, we compute the posterior mean of the Kendall's tau between the innovations of the AR series. In Table 5.6 we report the results for the Combined tree model and the Combined mixture model that, among the tree and mixture of tree models, present the lowest DIC₃ value, respectively, for the German and the Italian market. We can observe that some values of the Kendall's tau are quite small. This is not surprising since we are dealing with the innovations of log-returns and not directly with prices of the commodities. In order to check that these small values are not due to mis-specification of the linking copulas, we estimated these quantities by the forward-backward frequentist approach obtaining

Figure 5.10: Scatter plots for some commodities. In red the scatter plot of 400 predictive innovations from the Combined mixture model and in blue the scatter plot of the forward-backward residuals.



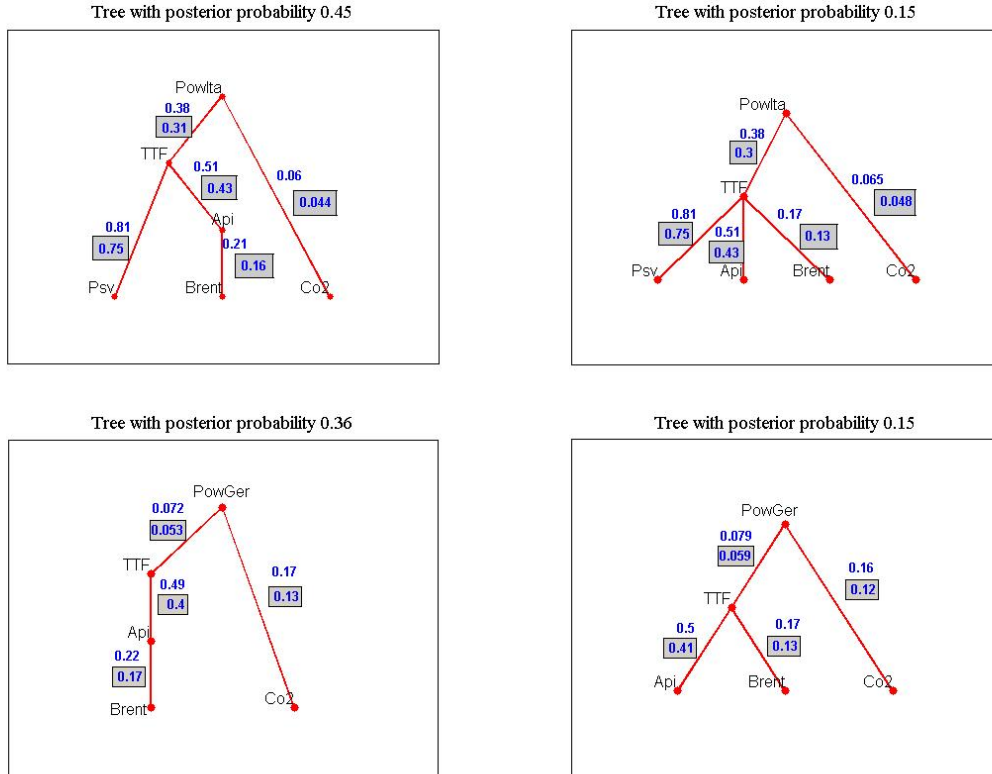
comparable results. Also a graphical comparison of the behavior of the predictive innovations with respect to the forward-backward residuals suggests similar conclusions. As an example, in Figure 5.10 we report the scatter plots of the predictive innovations and the forward-backward residual for some commodities.

For Joe, Gumbel, Clayton and Combined tree models we can directly visualize the dependence structure using the trees with highest posterior probabilities. For the mixture models, we consider the estimated MST.

For the tree copula models, the posterior probability is essentially concentrated on two structures. This can explain why mixture models with more than two components do not improve relevantly the fitting in comparison with the two components models.

The two structures with highest posterior probability for each domestic market, obtained using Gumbel and Clayton tree models, are depicted in Figures 5.11 and 5.12. The highest posterior probability tree structures obtained with the Joe copula model are the same as the ones obtained with the Gumbel tree model, with similar values of the parameters. Hence, the results are omitted. The posterior mean of the dependence parameters and of the Kendall's tau are reported on each

Figure 5.11: Tree copula structure with Gumbel copulas for the Italian market (first row) and the German market (second row). On each edge is reported in blue the posterior mean of the upper tail parameter between the neighbor nodes, and in a grey box the Kendall's tau posterior mean.



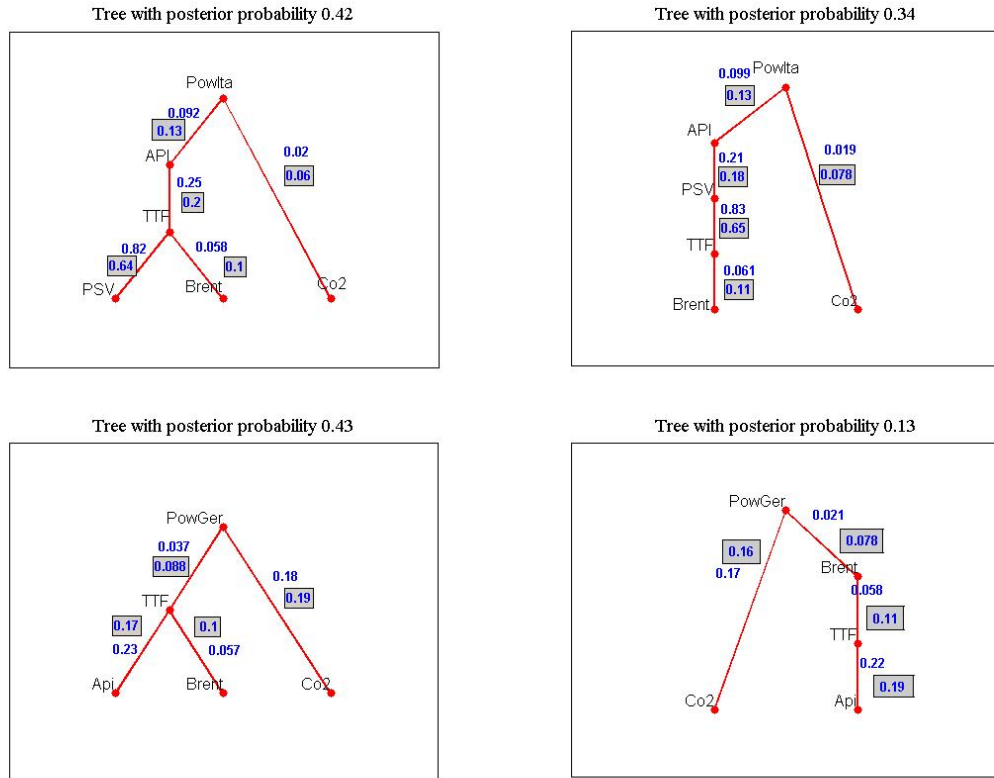
edge. It is worth noticing that the previous measures are computed conditionally on the considered tree structure.

We now analyze more in details the selected tree structures for the Italian market. The node corresponding to coal price is always connected to a gas node. TTF and PSV, the two gas nodes, are strongly connected in all structures and Brent is always an endpoint node. Finally, CO_2 is only connected with Power Italy. A possible explanation of this, can be found in the different nature of CO_2 with respect to the other variables that represent raw materials prices.

Also for the German market, we can identify some common edges in all the estimated trees. TTF and Api2 are always connected, also for this market CO_2 is only connected to Power Germany.

As already noted, the Gumbel and Joe copulas are suitable for the description of possible upper tail dependence, while Clayton copula for lower tail dependence. This can explain the differences in the estimated tree structures between the models. For instance, in the Italian market, the power price is directly connected with TTF in the Gumbel case, while in the Clayton tree model it is linked with Api2.

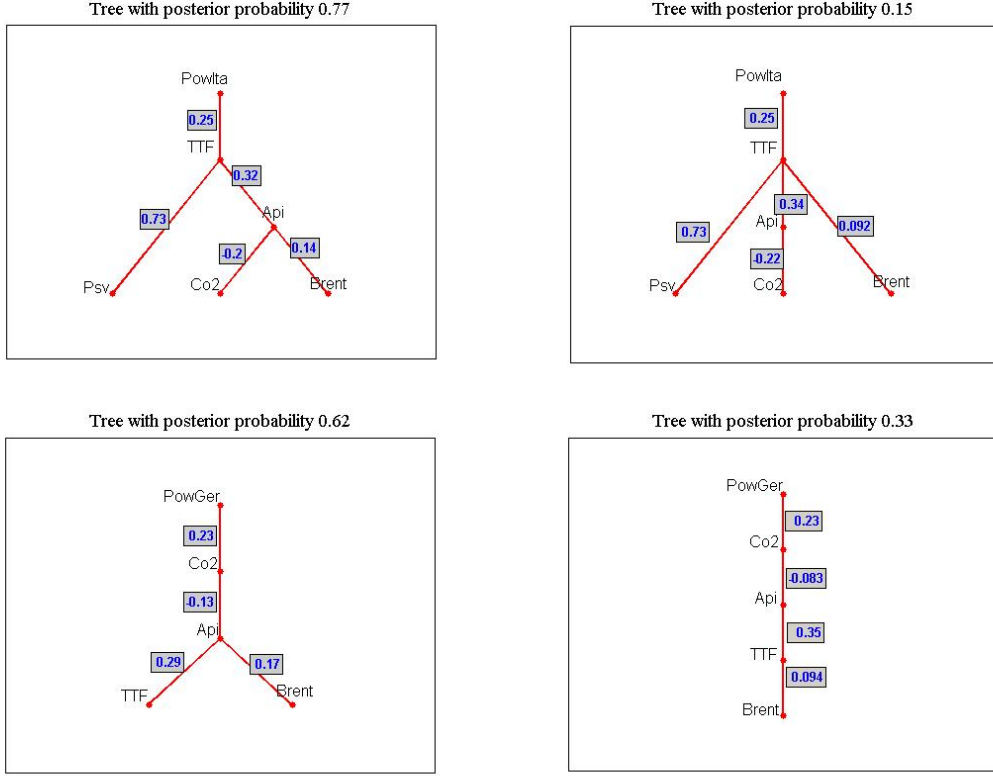
Figure 5.12: Tree copula structure with Clayton copulas for the Italian market (first row) and the German market (second row). On each edge is reported in blue the posterior mean of the lower tail parameter between the neighbor nodes, and in a gray box the Kendall's tau posterior mean.



In the German market, Gumbel and Clayton trees present more similarities. TTF plays a central role between power and the other raw materials prices, except for the second tree estimated via Clayton model which, nevertheless, has a quite low posterior probability. For both markets, the selected trees are in line with some well-known characteristics of the current energy market. Indeed, in both markets the power is mainly produced using gas (TTF or PSV) and coal (Api2), while the use of Brent has drastically decreased in the last years. Coherently, we find the Brent as endpoint node in all the estimated trees, with the exception of the second structure in the Clayton model.

Finally, in Figure 5.13 are represented the maximum a posteriori trees selected via Combined tree models. Comparing Figures 5.11-5.12 and 5.13 we note some structural similarities. As in the previous cases, Api2 is always connected to TTF (a gas node), TTF and PSV is an edge and Brent is still an endpoint node. There also some structural differences, that can be due to the fact that, in contrast with the previous model, the combined model can capture negative dependencies. More in details, one can observe that the edges with the lowest Kendall's tau coefficient in the previous models are absent in the corresponding tree structures in Figure 5.13 and edges with negative dependence

Figure 5.13: Tree copula structure with Combined tree model for the Italian market (first row) and the German market (second row). On each edge is reported the posterior mean of the Kendall's tau between the neighbor nodes.



are now added.

For the mixture models, we use the τMST and the $wMST$ in order to construct a representative tree. We recall that the τMST is the minimum spanning tree based on the absolute value of the Kendall's tau matrix, while the $wMST$ is based on the quantity

$$\Upsilon_{(l,m)}(\mathfrak{E}, \mathbf{w}) = \sum_{d=1}^D w_d \mathbb{1}\{(l,m) \in \mathcal{E}_d\}.$$

The $\tau MSTs$ obtained for the mixture of Gumbel, mixture of Clayton and mixture of Joe coincide with the MAP tree selected by the Gumbel model, for both markets. For the Combined mixture model, the τMST for the Italian market corresponds to the MAP structure of the Combined model while for the German market it corresponds to the MAP tree of the Gumbel model. Note that Gumbel tree model presents a lower value of $DIC \setminus DIC_3$ in comparison with the Clayton and Joe tree models. In Figure 5.14 are reported the weighted graphs Γ_τ and the corresponding MSTs computed with the Combined mixture model for the Italian and German markets.

Regarding the minimum spanning tree based on $\Upsilon_{(l,m)}$, in case of mixture of Gumbel, mixture of

Figure 5.14: Weighted graphs Γ_τ computed with the Combined mixture model for Italian (on the left) and German (on the right) markets. On each edge we show the absolute value of the posterior mean of the Kendall's tau between the neighbor nodes. For the Italian market, the edges with weight <0.05 are not depicted, while for the German case the edge with weight <0.01 are not drawn. The associated τMST structures are reported in green on each graph.

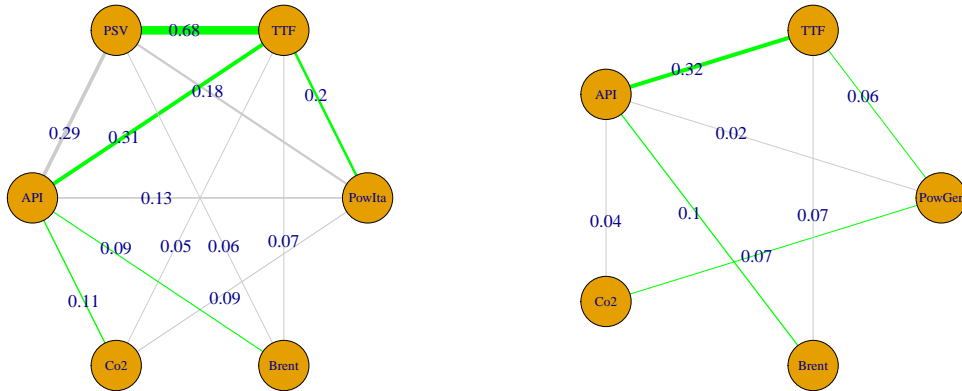


Figure 5.15: Weighted graphs Γ_w computed with the Combined mixture model for Italian (on the left) and German (on the right) markets. On each edge we show the absolute value of the posterior mean of the quantity $\Upsilon_{(l,m)}$ between the neighbor nodes. For the Italian market, the edges with weight <0.05 are not depicted, while for the German case the edge with weight <0.01 are not drawn. The associated $wMST$ structures are reported in green on each graph.

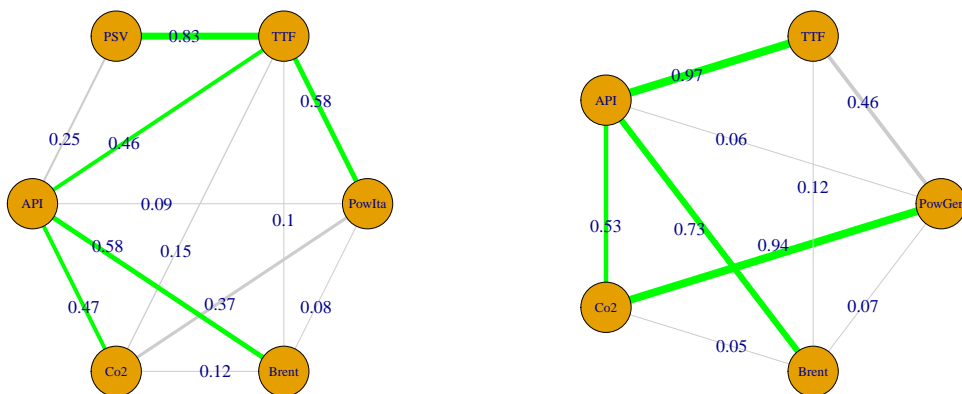
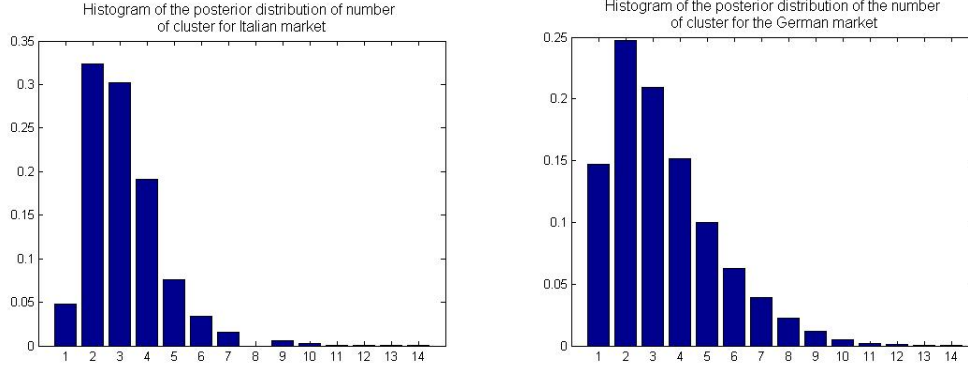


Figure 5.16: Histograms of the posterior distributions of number of clusters with the Combined DP-tree model for Italian (on the left) and German (on the right) markets.



Clayton and mixture of Joe the $wMSTs$ coincide with the MAP trees of the Gumbel tree model for both market. For the Combined mixture model, both for Italian and German markets, the $wMST$ are equal to the MAP structures of the Combined tree model. In Figure 5.15, for example, we report the weighted graphs Γ_w and the $wMSTs$ obtained with the Combined mixture model for each domestic market.

5.3.2 DP-Tree Copula Mixture Model

In case of DP-tree models (models ix-x listed in Section 5.1), an additional information that we obtain is the posterior estimate of the number of clusters. As already noted in Section 5.1, with the Combined DP-tree model the posterior mode of the number of clusters is 2 both for the Italian market and the German market. The posterior histograms of the number of clusters are reported in Figure 5.16. We have similar results with the Gumbel DP-tree model for which the posterior mode of the number of cluster is 3 both for the Italian and the German market, respectively.

Concerning dependence analysis, also for the DP-tree models we compute the posterior mean of the Kendall's tau measure between the innovations of the AR series. The results for the each markets, obtained with the Combined DP-tree model, that presents the lowest DIC_3 value, are reported in Table 5.7. We observe that the values of the Kendall's tau computed with the DP-tree model are similar to the values reported in Table 5.6.

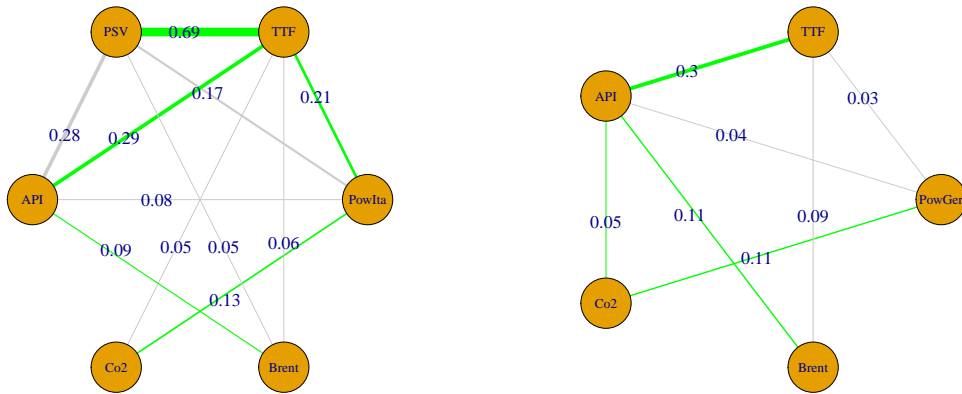
In order to have a representative dependence graph among the innovations, we consider the MST construction based on the absolute value of the estimated Kendall's tau matrix. For both markets, the $\tau MSTs$ obtained with the Combined DP-tree model and the Gumbel DP-tree model coincide with the MAP trees selected by the Gumbel model. In Figure 5.17 we present the weighted graphs Γ_τ and the $\tau MSTs$ obtained with the Combined DP-tree model for the Italian and German markets.

Table 5.7: Posterior mean of the Kendall’s tau between the innovations, for the Italian market (first table) and the German market (second table), obtained with the Combined DP-tree model.

Combined DP-Tree Model						
	PowIT	TTF	PSV	Api2	CO ₂	Brent
PowIT	1	0.21	0.17	0.08	0.13	0.03
TTF		1	0.69	0.29	0.05	0.06
PSV			1	0.28	0.04	0.05
Api2				1	-0.04	0.09
CO ₂					1	0.02
Brent						1

Combined DP-Tree Model					
	PowDE	TTF	Api2	CO ₂	Brent
PowDE	1	0.03	0.04	0.11	0.01
TTF		1	0.30	0.01	0.09
Api2			1	-0.05	0.11
CO ₂				1	-0.01
Brent					1

Figure 5.17: Weighted graphs Γ_τ computed with the Combined DP-tree model for Italian (on the left) and German (on the right) markets. On each edge we show the absolute value of the posterior mean of the Kendall’s tau between the neighbor nodes. For the Italian market, the edges with weight <0.05 are not depicted, while for the German case the edge with weight <0.01 are not drawn. The associated τMST structures are reported in green on each graph.



As an alternative to the τMST , we also construct the $wMST$ based on the posterior mean of the quantity $\Upsilon_{(l,m)}$, see Equation (1.11). In Figure 5.18 are depicted the graphs Γ_w and the related MSTs for the Italian and the German market obtained with the Combined DP-tree model. In the $wMST$ of the Italian market we observe that the commodity Api2 plays a central role, while for the German market the $wMST$ coincide with the MAP tree of the Combined tree model. For the Gumbel DP-tree copula model, the $wMST$ for the Italian market corresponds to the MAP structure of the Gumbel tree model, while in the German case, the MST structure is similar to the MAP tree of Gumbel tree model with the edge Brent-Power Germany instead of TTF-Power Germany.

Figure 5.18: Weighted graphs Γ_w computed with the Combined DP-tree model for Italian (on the left) and German (on the right) markets. On each edge we show the absolute value of the posterior mean of the quantity $\Upsilon_{(l,m)}$ between the neighbor nodes. For the Italian market, the edges with weight <0.05 are not depicted, while for the German case the edge with weight <0.01 are not drawn. The associated $wMST$ structures are reported in green on each graph.

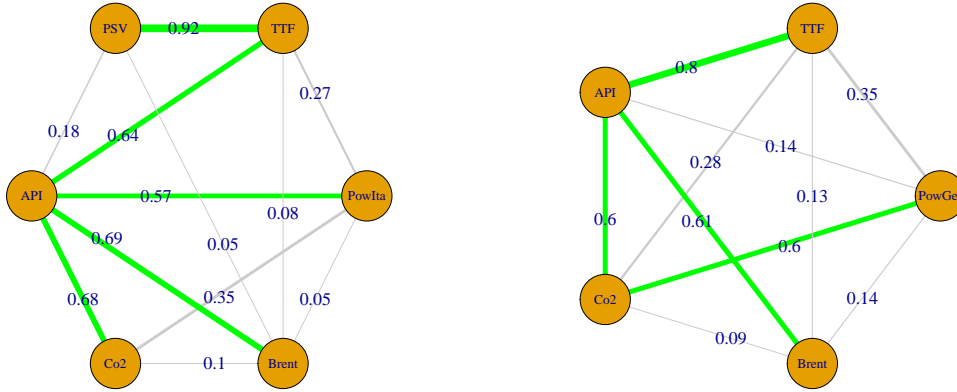


Table 5.8: Posterior mean of the Kendall’s tau between the innovations, for the Italian market (first table) and the German market (second table), obtained with the Combined factor model.

Combined Factor Model						
	PowIT	TTF	PSV	Api2	CO ₂	Brent
PowIT	1	0.22	0.19	0.13	-0.01	0.01
TTF		1	0.70	0.37	-0.03	0.09
PSV			1	0.35	-0.01	0.05
Api2				1	-0.02	0.05
CO ₂					1	-0.01
Brent						1

Combined Factor Model					
	PowDE	TTF	Api2	CO ₂	Brent
PowDE	1	-0.05	-0.07	0.02	-0.01
TTF		1	0.35	-0.04	0.12
Api2			1	-0.06	0.20
CO ₂				1	-0.004
Brent					1

5.3.3 Factor Copula Model

We now present the results obtained with the one-factor copula model (models xi) and xii)). For both models we compute the posterior mean of the Kendall’s tau measures. In Table 5.8 are reported the posterior estimates for the Italian and the German market obtained with the Combined factor model. Comparing Table 5.8 with Tables 5.6/5.7, we can observe that the values are quite similar.

Concerning dependence structure analysis, through the use of a one-factor copula, we have a fixed underlying tree structure with a latent variable V as root node. Hence, we cannot proceed with structure learning but we compute the posterior estimates of each bivariate copula parameter. The results obtained with the Combined factor model and Gumbel factor model for each domestic

Figure 5.19: Factor copula structure with Double copulas for the Italian market and the German market. The Kendall's tau posterior mean is reported on each edge in a gray box.

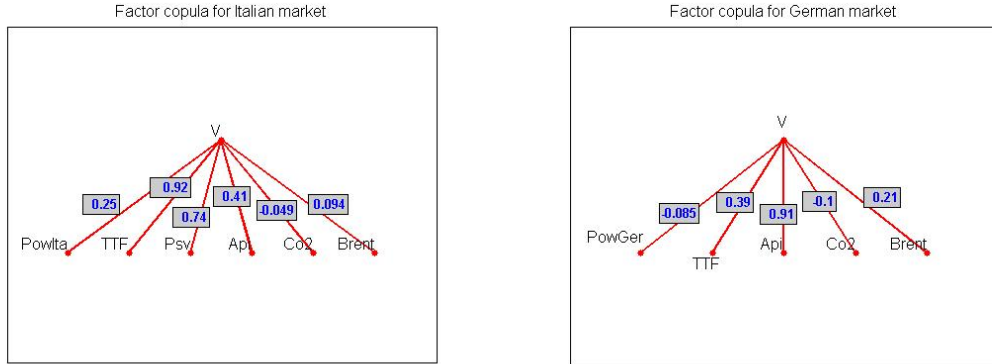
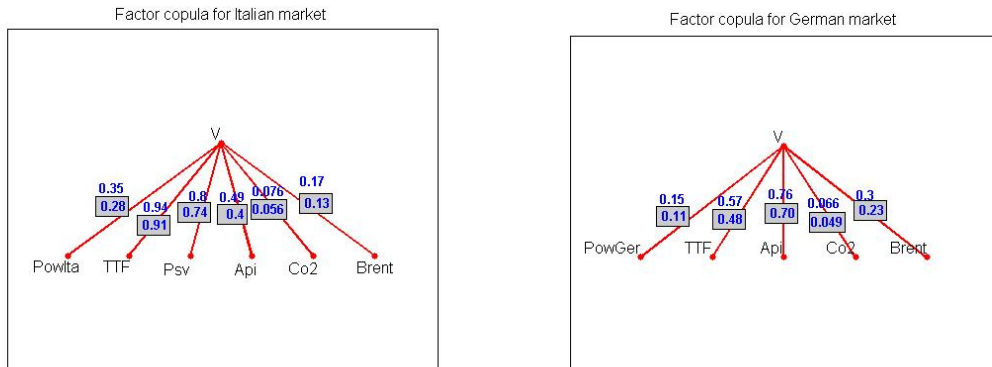
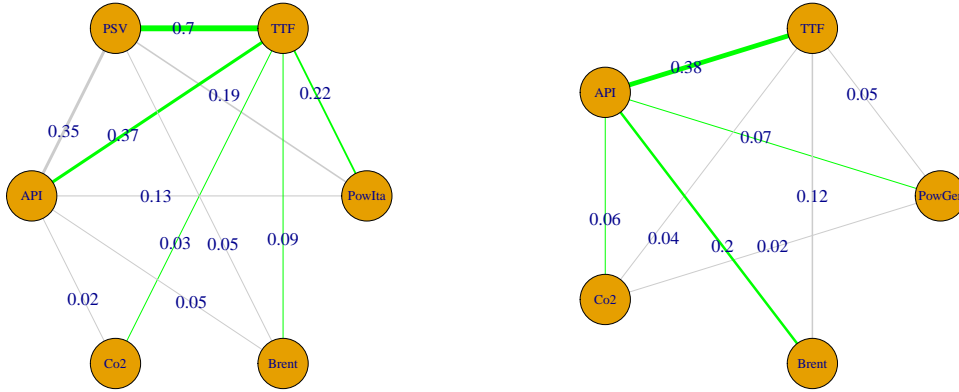


Figure 5.20: Factor copula structures with Gumbel copulas for the Italian market and the German market. On each edge is reported the posterior mean of the upper tail parameter between the neighbor nodes in blue and the Kendall's tau posterior mean in a gray box.



market are reported in Figure 5.19 and 5.20, respectively. For the Italian market, we can observe that the copula parameter associated to the edge V-TTF, in both models, has value ≥ 0.9 . A high value for $\theta_{l,0}$ on an edge means that the latent node V is strongly correlated with the commodity on that edge, in this case the TTF. Therefore, using the factor copula model, we find that TTF is a commodity that plays a central role in the Italian market. This observation is coherent with the MAP structures of the Combined tree model, Figure 5.13, where the TTF is the node that presents higher correlation with its neighbor and higher number of connections with other commodities. For the German market we can note a similar effect in the Combined factor model between the latent node V and Api2. Also this result is coherent with the role of Api2 in the MAP trees for the German market of Figure 5.13.

Figure 5.21: Weighted graphs Γ_τ computed with the Combined factor model for Italian (on the left) and German (on the right) markets. On each edge we show the absolute value of the posterior mean of the Kendall's tau between the neighbor nodes. For the Italian and the German market, the edges with weight <0.05 are not depicted. The associated τMST structures are reported in green on each graph.



In order to obtain a graphical representation of the dependence structure without the latent node, we rely on MST construction based on the absolute value of the Kendall's tau. For both models, Combined and Gumbel factor models, we obtain the same representative tree structures for each domestic market. In Figure 5.21 are reported the weighted graphs Γ_τ and, in green, the related τMST s obtained with the Combined factor model. We observe that also for the τMST s, there is a commodity, TTF for Italian market and Api2 for German market, that turns out to be a “root” node.

5.4 Markov Switching Factor Copula Model

In this Section, we present some preliminary results obtained for the Italian and the German energy market using the Markov switching factor model introduced in Section 4.3.

In this case, for each commodity, we consider daily time series of one year forward contracts from January 2013 to December 2014. Once again, for each series we compute the monthly logarithmic return rates $X_{t,k}$ and model each $X_{t,k}$ using an AR(3). The innovations are assumed to have Normal marginal distributions.

In order to apply our Markov switching factor model, we proceed with a B-IFM procedure. First, the AR models are fitted marginally to each logarithmic return rates $X_{k,t}$ (forward-backward approach) and then the Markov switching model is applied to the estimated residuals.

Table 5.9: Posterior mean of the marginal variances of the innovations in each state, for the Italian market (first table) and the German market (second table), obtained with the Combined Markov switching factor model.

Italian Market				German Market			
	State 1	State 2	State 3		State 1	State 2	State 3
PowIT	0.27	0.92	0.30	PowDE	0.35	0.86	0.29
TTF	0.31	0.86	0.65	TTF	0.28	0.87	0.62
PSV	0.27	0.82	0.61	Api2	0.32	0.65	0.36
Api2	0.34	0.70	0.37	CO ₂	0.37	1.00	0.25
CO ₂	0.38	1.01	0.29	Brent	0.29	0.64	0.33
Brent	0.29	0.63	0.32				

In particular, we analyze the following Markov switching factor copula structures:

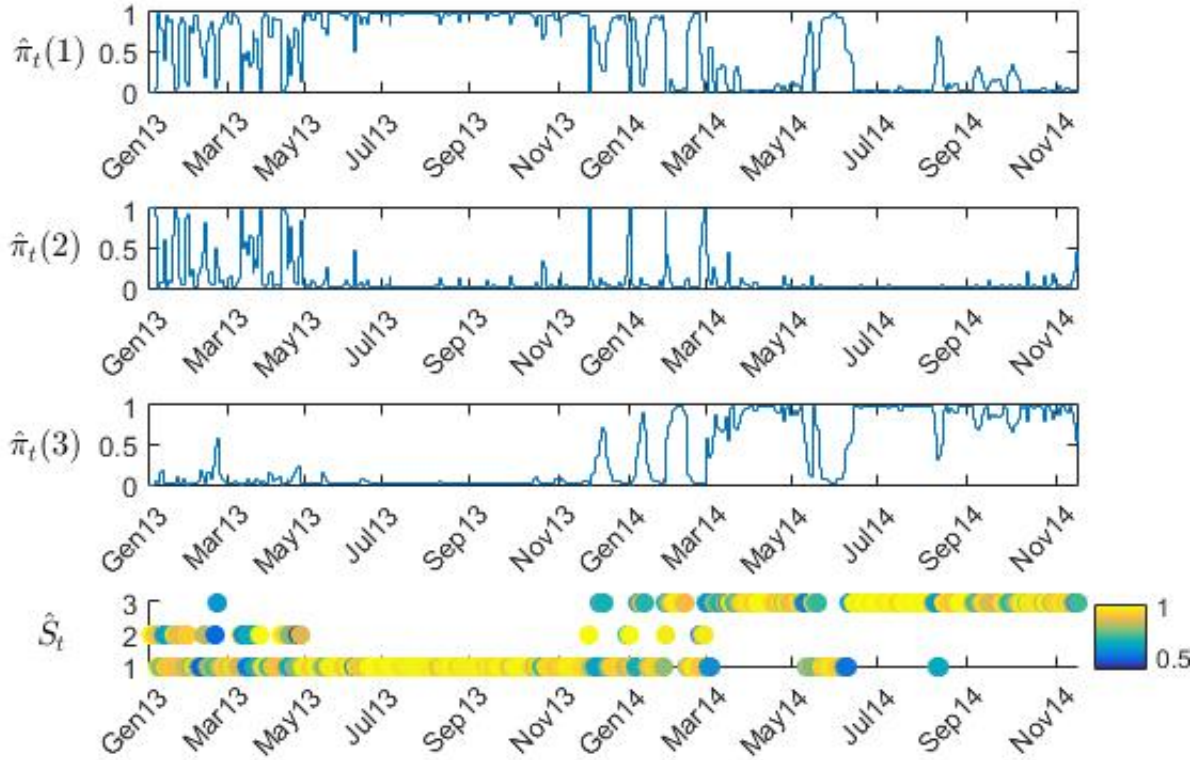
- Combined Markov switching model: innovations distributed according to a Markov switching factor copula where on each edge the copula $c_{\theta_{k,0}^{(r)}}$ can be either a Double Gumbel or a double Clayton and $\theta_{k,0}^{(r)} = (\tau_{k,0}^{(r)}, \zeta_{k,0}^{(r)})$;
- Gumbel Markov switching model: innovations distributed according to a Markov switching factor copula where all the copulas $c_{\theta_{k,0}^{(r)}}$ are Gumbel copulas re-parameterized through the upper tail measure.

For these models, we use the following independent prior distributions. On $\theta_{k,0}^{(r)}$ we assign a uniform prior; on the marginal parameters (i.e. the precision) we use a Gamma prior with parameters (1,1) while on each row of the transition matrix \mathbf{Q} we set a Dirichlet distribution with hyperparameters equal to 10. For each model, we assume the presence of three regimes.

Figure 5.22 shows the posterior estimates $\hat{\pi}_t(r)$ of being in state r , $r = 1, 2, 3$ for the Italian market computed with the Combined Markov switching model. In the same Figure, we also report the posterior mode \hat{S}_t of the hidden Markov chain (last picture). The analogous results for the German market are presented in Figure 5.23. We observe that the first regime identifies, approximately, the observations of the first year and the third regime the observations of the second year. Finally, to the second state are associated observations at the beginning or at the end of a year. The second regime identifies the period of change from one year to another in which usually the data are characterized by higher volatility. Therefore, our model recognizes three different dependence structures, one for each year (state 1 and state 3) and one for the period with higher volatility (state 2). In Table 5.9 we list the posterior mean of the marginal variances of each regime, for Italian and German markets, computed with the Combined Markov switching model. As expected, the posterior variances associated to state 2 show the highest values.

The posterior estimates $\hat{\pi}_t(r)$, $r = 1, 2, 3$ and the posterior mode \hat{S}_t of the hidden chain, obtained with the Gumbel Markov switching model, are reported in Figures 5.24 and 5.25 for the Italian and the German market, respectively. We note that the observations are associated to the different regimes in the same way as in the Combined Markov switching model, i.e. state 1 and state 3

Figure 5.22: In the first three pictures, the posterior probability of being in state r ($r = 1, 2, 3$) overtime for the Italian market obtained with the Combined Markov switching factor model. In the last image, we present the posterior mode \hat{S}_t of the hidden Markov chain. Each point is colored accordingly to the posterior probability of being in the state identified by \hat{S}_t .

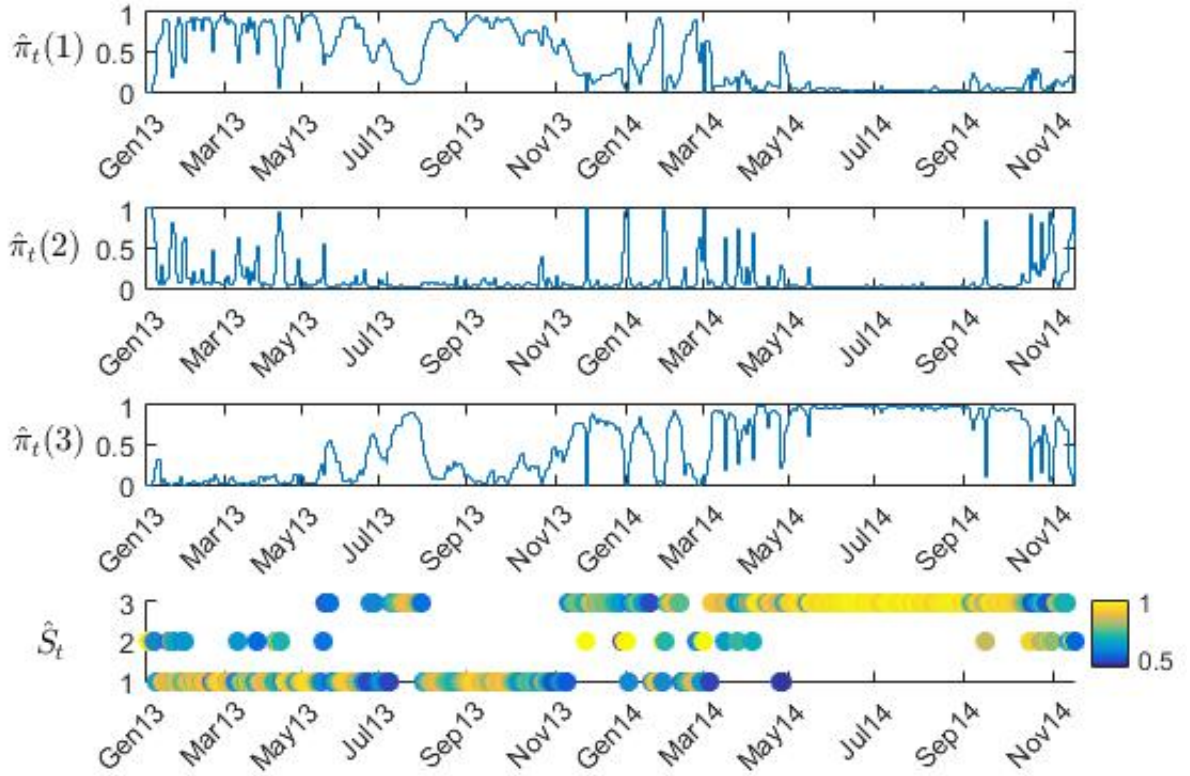


identify the observations of year 2013 and 2014, respectively, and state 2 the period of change from one year to another. Also for the posterior means of the marginal variances computed with the Gumbel Markov switching model, we have that the highest values are related to the second regime.

In Figures 5.26 and 5.27 we show the posterior density for each probabilities $q_{i,j}$, $i, j = 1, 2, 3$ of the transition matrix Q obtained for Italian and German market with the Combined Markov switching model and the Gumbel Markov switching model, respectively. For both models and both markets, state 1 and state 3 present higher persistent probabilities and lower probabilities to move to other regimes. Instead, as expected, state 2 shows higher probabilities to move from itself to the other regimes.

Concerning dependence analysis, in case of the Markov switching model, one can study the dependence structure among the examined variables in each regime. In particular, conditional to $S_t = r$, $r = 1, 2, 3$, we compute the posterior mean of the Kendall's tau between the innovations

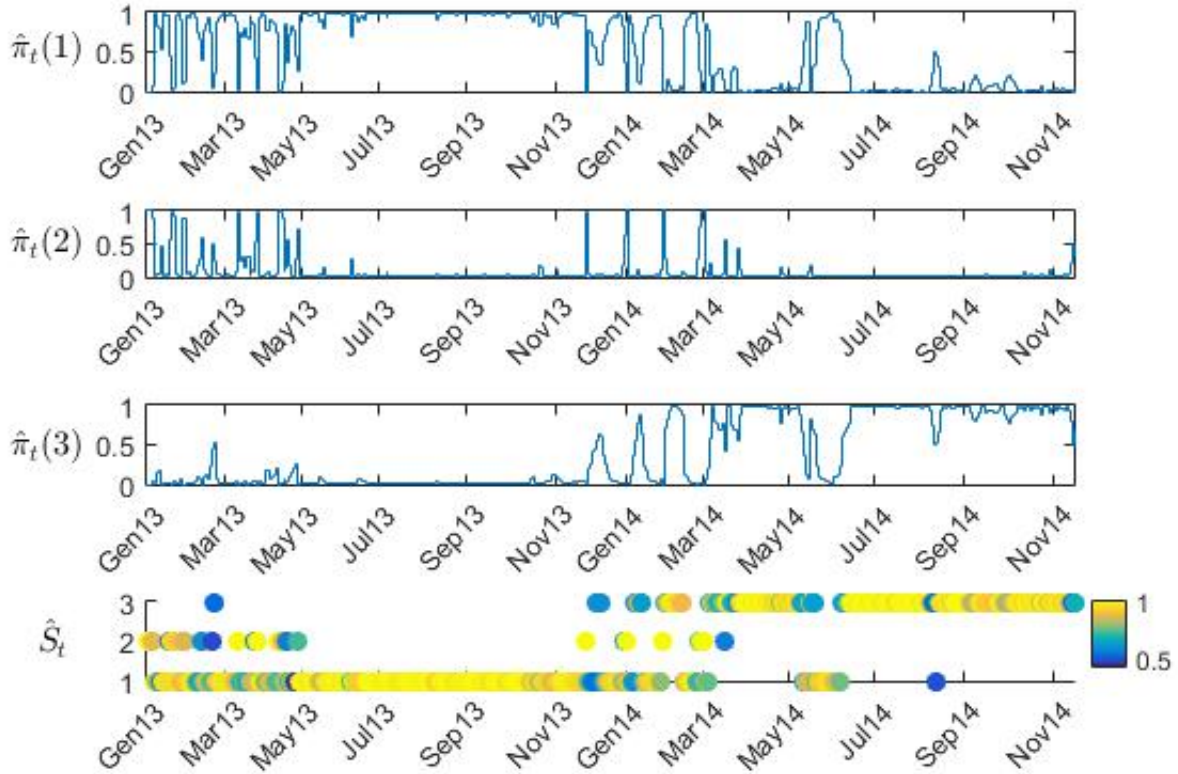
Figure 5.23: In the first three pictures, the posterior probability of being in state r ($r = 1, 2, 3$) overtime for the German market obtained with the Combined Markov switching factor model. In the last image, we present the posterior mode \hat{S}_t of the hidden Markov chain. Each point is colored accordingly to the posterior probability of being in the state identified by \hat{S}_t .



in order to get information about the dependence in each regime. Then, for each state r , we built the weighted graph Γ_r based on the posterior estimates of the Kendall's tau computed in regime r and the corresponding minimum spanning tree. The graphs Γ_r and the associated τMST s obtained with the Combined Markov switching model for the Italian and the German market are reported in Figures 5.28 and 5.29, respectively. In each regime we obtain significantly different dependence structures. In particular, for both markets, we note that, conditionally to the state that identify the observations of year 2014 (state 1), the MSTs are equal to the minimum spanning trees obtained with the Combined factor model, see Figures 5.21. Moreover, we observe that the posterior means of the Kendall's tau associated to the regime 2, that identifies the period of transition from one year to another, are higher than the corresponding posterior estimates in the other states.

We also proceed with portfolio analysis and evaluation with the Markov switching model. For each domestic market, we consider the portfolio composition presented in Section 5.2. Also for this

Figure 5.24: In the first three pictures, the posterior probability of being in state r ($r = 1, 2, 3$) overtime for the Italian market obtained with the Gumbel Markov switching factor model. In the last image, we present the posterior mode \hat{S}_t of the hidden Markov chain. Each point is colored accordingly to the posterior probability of being in the state identified by \hat{S}_t .

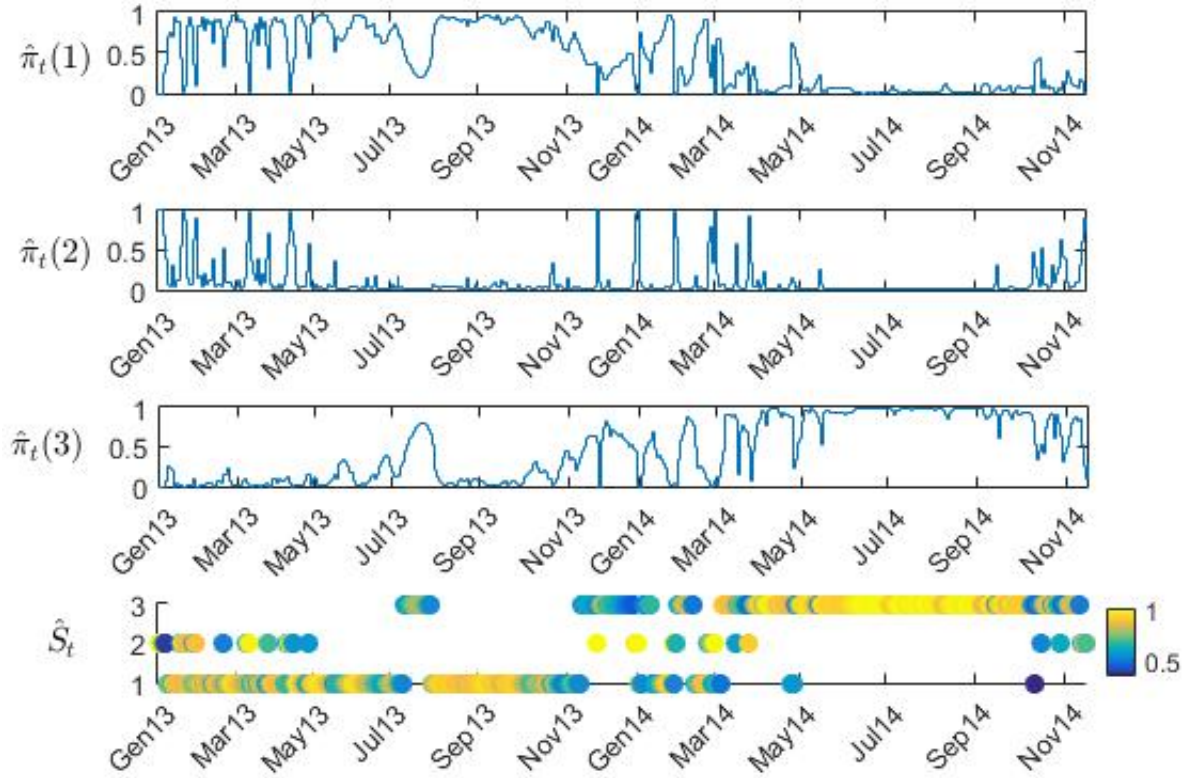


methodology, we construct the portfolio predictive distribution at time $T+1$, where T corresponds to 15th December 2014. The predictive portfolio distributions at time $T + 1$ for each domestic market are reported in Figure 5.30.

In order to analyze the forecasting performance of our Markov switching models, we compute the in-sample portfolio on the last 195 observations, as in Section 5.2. The one day ahead domestic portfolios obtained with the Combined Markov switching model are reported in Figure 5.31. The analogous one day ahead portfolios computed with the Gumbel Markov switching model are shown in Figure 5.32. We can note again that there is a good agreement between the behaviors of predictive portfolios and the historical ones.

Finally, in Table 5.10 we report the mean distances between the in-sample forecasting portfolio and the historical one on the last 50 observations computed with the Markov switching models for each market. For comparative purposes, we compute the mean distances with the DP-mixture

Figure 5.25: In the first three pictures, the posterior probability of being in state r ($r = 1, 2, 3$) overtime for the German market obtained with the Gumbel Markov switching factor model. In the last image, we present the posterior mode \hat{S}_t of the hidden Markov chain. Each point is colored accordingly to the posterior probability of being in the state identified by \hat{S}_t .



models and the factor copula models applied to the same data used for the Markov switching model (from January 2013 to December 2014) with a B-IFM procedure. The results are reported in Table 5.10. We observe that both Markov switching models present the lower values of mean distances respect to the DP-tree and factor models. In particular, the best results are obtained with the Combined Markov switching model. Moreover, if we compare the values in Table 5.10 with the corresponding B-IFM estimates of Table 5.4, computed using only the observations of year 2014, we note that, for the Italian market, the Combined Markov switching models shows the best results also respect to all the B-IFM mean distances obtained with the other models. Regarding German market, we have that the distances of the Markov switching models are in line with the B-IFM values of Table 5.4.

Figure 5.26: Posterior density of each transition probability for Italian and German markets obtained with the Combined Markov switching model.

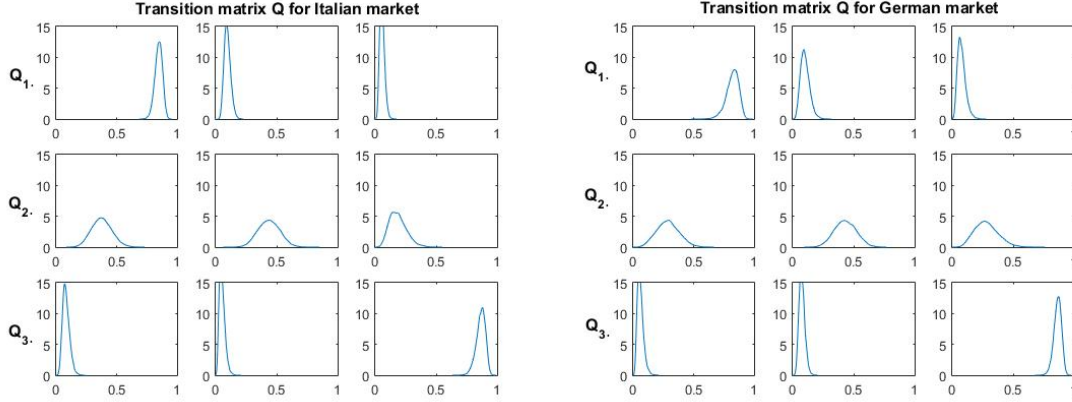


Figure 5.27: Posterior density of each transition probability for Italian and German markets obtained with the Gumbel Markov switching model.

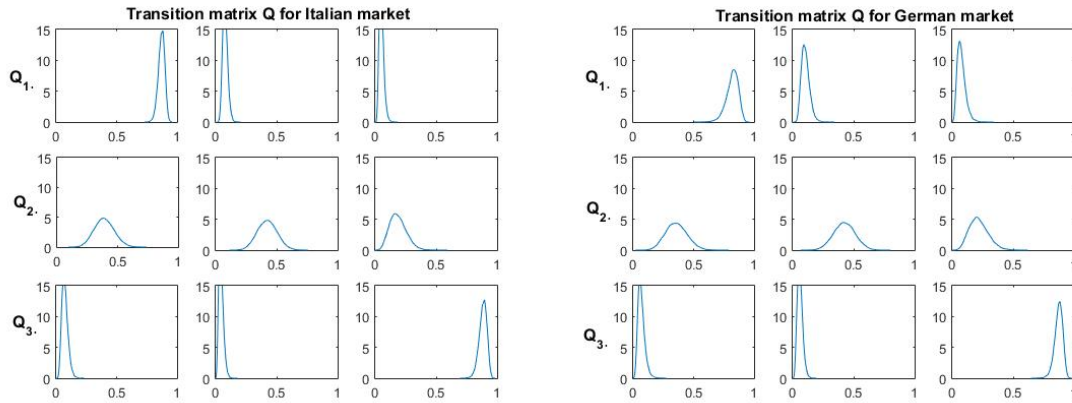


Table 5.10: Mean distance between the in-sample predictive portfolio and historical portfolio on the last 50 observations for the Italian market and the German market with different models (using the data from January 2013 to December 2014).

	B-IFM Mean Distance	
	Italy	Germany
Combined Markov Switching model	0.177	0.213
Gumbel Markov Switching model	0.186	0.214
Combined DP-tree model	0.196	0.219
Gumbel DP-tree model	0.191	0.220
Combined factor model	0.192	0.215
Gumbel factor model	0.194	0.215

Figure 5.28: Weighted graphs Γ_τ for each regime computed with the Combined Markov Switching factor model for Italian market. On each edge we show the absolute value of the posterior mean of the Kendall's tau between the neighbor nodes. For the Italian market, the edges with weight < 0.08 are not depicted. The associated τMST structures are reported in green on each graph.

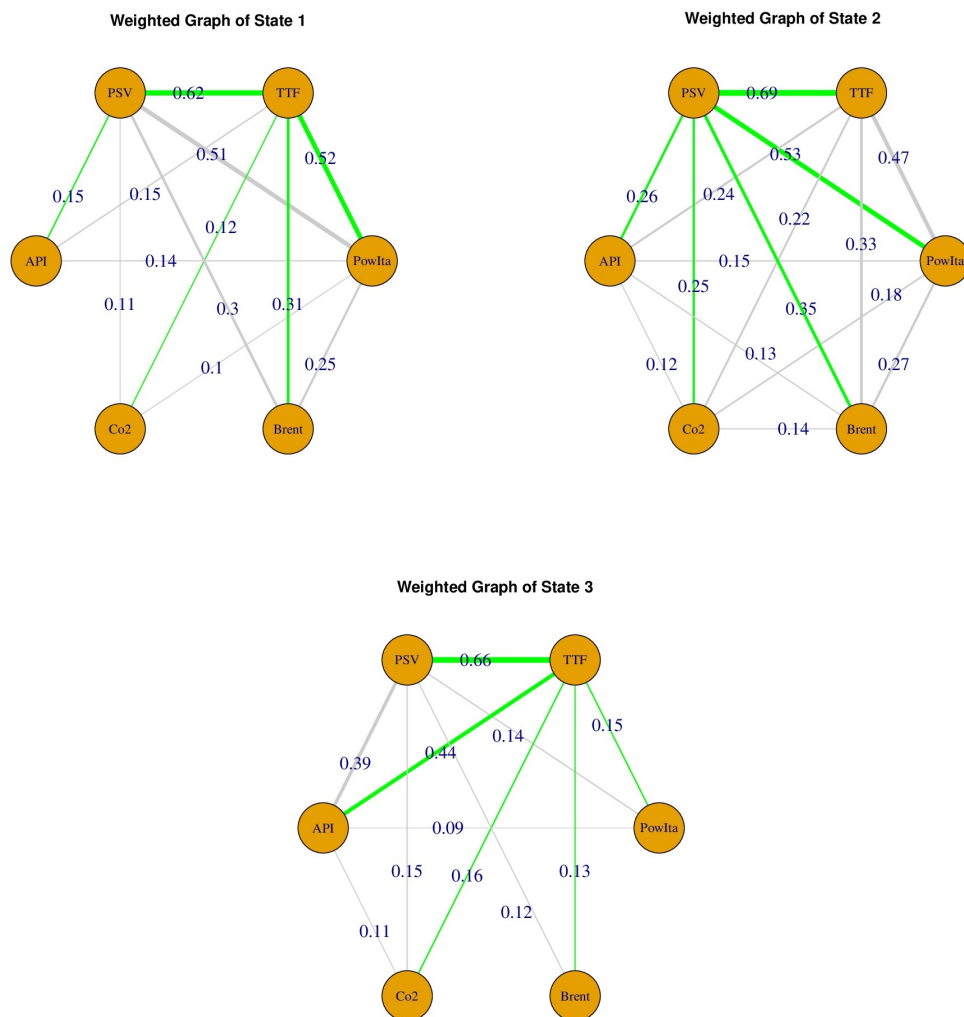


Figure 5.29: Weighted graphs Γ_τ for each regime computed with the Combined Markov Switching factor model for German market. On each edge we show the absolute value of the posterior mean of the Kendall's tau between the neighbor nodes. For the Italian market, the edges with weight <0.03 are not depicted. The associated τMST structures are reported in green on each graph.

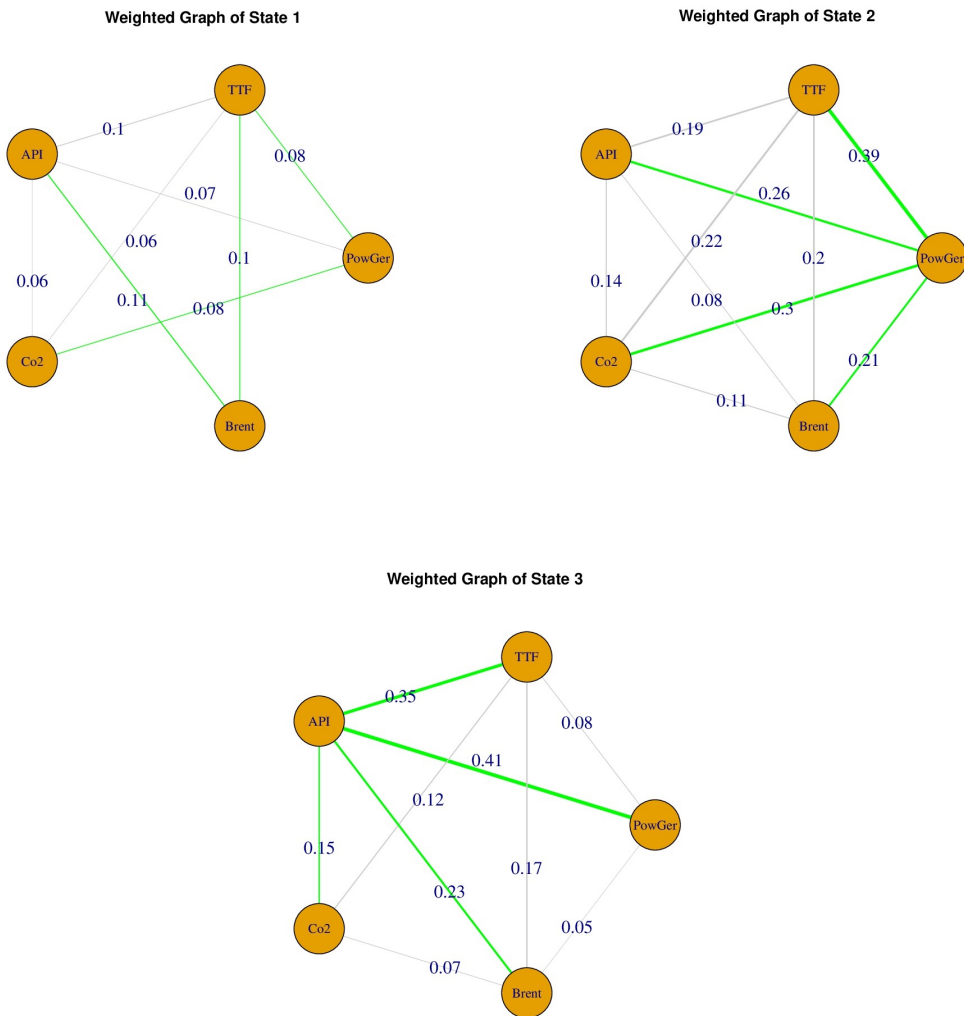


Figure 5.30: Italian and German predictive portfolio densities at time $T+1$, T corresponding to 15th December 2014 with Markov switching models.

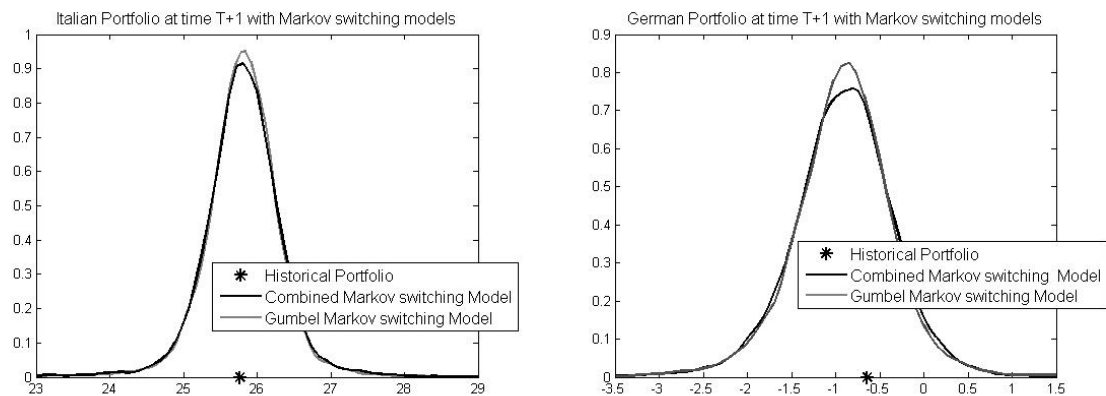


Figure 5.31: One day ahead Italian (on the left) and German (on the right) portfolios values with 95% credible intervals obtained with the Combined Markov switching model (in-sample analysis).

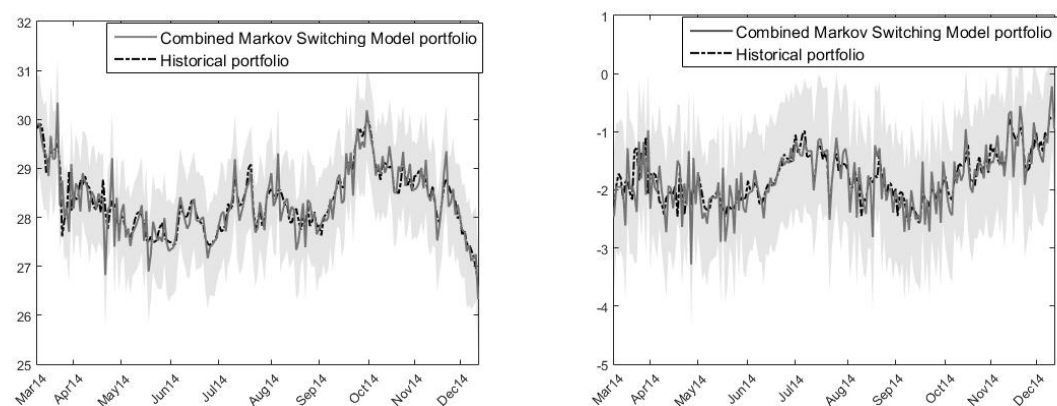
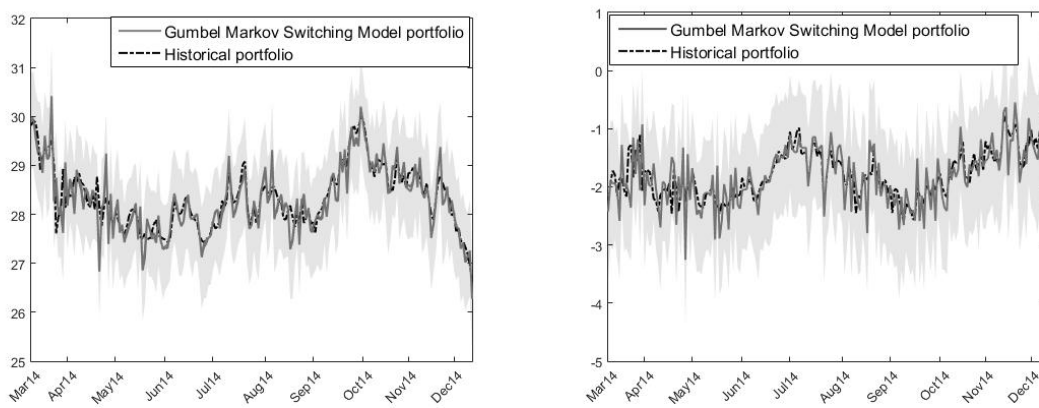


Figure 5.32: One day ahead Italian (on the left) and German (on the right) portfolios values with 95% credible intervals obtained with the Gumbel Markov switching model (in-sample analysis).



Chapter 6

Conclusion and Extensions

6.1 Conclusion

In this dissertation we investigated alternative copula-based models in order to study the multivariate dependence among AR(p) time series. The use of copula structures allows us to analyze different type of dependence with respect to than the one induced by the multivariate Normal distribution. In particular, we presented Bayesian models based on tree copula, finite mixture of tree copulas and factor copula distributions. In these models, the joint density is built using bivariate copulas and a suitable graphical structure. For each approach, we proposed a fully Bayesian procedure to estimate all quantities of interest, including, in case of tree copula-based models, the underlying dependence graphical structure. Additional information on the dependence among the data are deduced from posterior estimates of tail parameters and Kendall's tau. We proposed suitable MCMC schemes for posterior estimation and predictive analysis.

Moreover, as an extension of the factor copula structure, we developed the Markov switching factor copula model in which the joint dependence structure in each regime is given by a suitable one-factor copula. Also for this methodology, we proposed a Bayesian procedure in order to estimate all quantities and a suitable MCMC scheme for posterior estimates.

Our methodologies have been used to study the behavior of Italian and German energy markets. In particular, we analyzed the dependence structure among the drivers of each market, and we evaluated and forecasted each domestic portfolio using suitable risk measures.

In order to obtain a graphical representation of the dependence structure, we proceeded with structural learning with the tree copula models, while for the other methodologies we relied on the MST construction based on the posterior mean of the Kendall's tau. Comparing the tree structures obtained with the tree copula, mixture of tree copulas and factor copula models, we discovered some interesting findings. For the Italian market, we found that the commodity mostly connected with the Italian power price is the TTF. This commodity, both in the Combined factor copula model and Gumbel factor copula model, turned out also to be identified as "root" node. This is coherent with

the fact that, in the Italian energy market, the power is mainly produced by gas. Furthermore, we found always a strong correlation between the two gas nodes, TTF and PSV, and a path of length three connecting them to the coal price (Api2). For the German market, an edge between PSV and Api2 is always included and, moreover, Api2 is the commodity that plays a central role in the factor copula models, in particular in the Combined ones. As an alternative to the τMST , for the mixture tree copula models and DP-tree models, we also built the $wMST$ where the weights are computed starting from the quantities $\Upsilon_{l,m}$.

Regarding portfolio analysis, our fully Bayesian procedure presents a better performance with respect to B-IFM in terms of mean distances between the historical and predictive portfolios. The Combined DP-tree model shows the best performance both for dependence analysis, in terms of DIC_3 criterion, and portfolio analysis, in terms of mean distances.

Finally, we presented some preliminary results obtained with the Markov switching factor copula model. In this case, we considered the data from January 2013 to December 2014 and assumed to have 3 regimes. We found that to regimes 1 and 3 are associated to the observations of year 2013 and of year 2014, respectively, while state 2 identifies the period of change from one year to another. We also computed the posterior mean of the Kendall's tau in each regime and built the weighted graphs Γ_τ and the corresponding τMST s. For both market, we discovered significantly different dependence structures for each regime. In particular, state 2 shows the higher values of Kendall's tau and also the higher values of posterior mean of the marginal variances.

6.2 Future Work

With regard to future perspectives, we are planning to develop different extensions for the alternative methodologies proposed in this thesis.

Concerning model selection analysis, we relied on the DIC/DIC_3 criteria that, however, present some limitations. As first step, we aim to consider another criterion for model comparison. In particular, one could evaluate the Bayes factor using one of the alternative method described in literature for the approximation of the marginal likelihood. Furthermore, we plan to design a more systematic comparison of the various models by a large scale simulation study. More precisely, one could generate different dataset from the same simulated scheme and compare the boxplot of posterior estimates, obtained with the different models, with the true values. We leave this comparison for future work since the choice of the simulated scheme is delicate and require further investigation in order to obtain a fair comparison.

As further extension, we propose to adapt the different approaches assuming other type of copulas and marginal distributions and develop suitable Bayesian models. In particular, as alternative margins, one can consider skewness distributions, such as the skew Normal or the skew t , that allow to analyze dataset that do not present symmetric distributions.

Another idea is to proceed with a Bayesian inference also on the order p of the AR series or to consider the Vector Autoregressive (VAR) model that generalize the univariate AR model. We can also extend the methodologies including alternative type of time series such as the ARMA or GARCH models, widely used in applications.

Other future works can concern the development of a fully Bayesian approach for the Markov switching copula model that involves also the AR parameters, assuming a suitable prior distribution on the time series parameters. Furthermore, it may be interesting to consider different time series coefficients in each regime of the Markov switching model.

Concerning the DP-tree copula model, one can study an alternative non parametric prior in place of the Dirichlet process, such as, for example, the Pitman Yor process. Finally, as an extension for the factor copula model, one can generalized the one-factor copula using a p -factor copula. In this case, the dependence among the variable is represented by p latent variables and each bivariate copula links the observed variables to all the latent factors.

Bibliography

- Aas, K. & Berg, D., Models for construction of multivariate dependence - a comparison study, *Eur. J. Financ.*, **15**, 639-659, 2009.
- Aas, K., Czado, C., Frigessi, A., & Bakken, H., Pair-copula constructions of multiple dependence. *Insur. Math. Econ.*, **44**, 182-198, 2009.
- Acerbi, C. & Tasche, D., Expected shortfall: a natural coherent alternative to value-at-risk, *Economic notes*, **31**, 379-388, 2002.
- Antoniak, C. E., Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The annals of statistics*, 1152-1174, 1974.
- Arreola Hernandez, J., Are oil and gas stocks from the Australian market riskier than coal and uranium stocks? Dependence risk analysis and portfolio optimization, *Energ. Econ.*, **45**, 528-536, 2014.
- Artzner P., Delbaen F., Eber J.M. & Heath D., Coherent measures of risk, *Math. Financ.*, **9**, 203-228, 1999.
- Basel Committee on Banking Supervision, *An Internal Model-Based Approach to Market Risk Capital Requirements*, Bank for International Settlements, Basel, Switzerland, 1995.
- Bassetti F., De Giuli M.E., Nicolino E., Tarantola C., Multivariate Dependence Analysis via Tree Copula Models: an Application to Energy Market, *Submitted*.
- Basu, S., & Chib, S., Marginal likelihood and Bayes factors for Dirichlet process mixture models, *Journal of the American Statistical Association*, **98**, 224-235, 2003.
- Besag, J., Green, P., Higdon, D. and Mengersen, K., Bayesian computation and stochastic systems, *Statist. Sci.*, **10**, 3-66, with comments and a reply by the authors, 1995.
- Blackwell, D., & MacQueen, J. B., Ferguson distributions via Pólya urn schemes, *The annals of statistics*, 353-355, 1973.

- Bedford, T. & Cooke, R.M., Probability density decomposition for conditionally dependent random variables modeled by vines, *Ann. Math. Artif. Intel.*, **32**, 245-268, 2001.
- Bedford, T. & Cooke, R.M., Vines - a new graphical model for dependent random variables, *Ann. Stat.*, **30**, 1031-1068, 2002.
- Burnham, K. P. & Anderson, D. R., *Model selection and multimodel inference. A practical information-theoretic approach*, Second edition, Springer-Verlag, New York, 2002.
- Carter, C. K., & Kohn, R., On Gibbs sampling for state space models, *Biometrika*, **81**, 541-553, 1994.
- Celeux, G., Forbes, F., Robert, C. P. & Titterton, D. M., Deviance information criteria for missing data models, *Bayesian analysis*, **1**, 651-673, 2006.
- Chib, S., Marginal likelihood from the Gibbs output *Journal of the American Statistical Association*, **90**, 1313-1321, 1995.
- Chib, S., & Jeliazkov, I., Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association*, **96**, 270-28, 2001.
- Chib, S., Nardari, F., Shephard, N., Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics*, **108**, 281-316, 2002.
- Chollete, L., Heinen, A., Valdesogo, A., Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics*, **7**, 437-480, 2009.
- Claeskens, G., & Hjort, N. L., *Model selection and model averaging*, Vol. 330, Cambridge: Cambridge University Press, 2008.
- Czado, C., Gärtner, F. & Min, A., Analysis of Australian Electricity Loads using Joint Bayesian Inference of D-Vines with Autoregressive Margins, in *Dependence Modeling: Vine Copula Handbook*, 265-280, World Scientific, London, 2011.
- Dalla Valle, L., De Giuli, M. E., Tarantola, C., & Manelli, C., Default probability estimation via pair copula constructions, *Eur. J. Oper. Res.*, **249**, 298-311, 2016.
- Elidan, G., Copulas and Machine Learning, *Copulae in Mathematical and Quantitative Finance*, Springer Berlin, 39-60, 2013.
- Escobar, M. D., & West, M., Bayesian density estimation and inference using mixtures, *Journal of the american statistical association*, **90**, 577-588, 1995.
- Ferguson, T. S., A Bayesian analysis of some nonparametric problems, *The annals of statistics*, 209-230, 1973.

- Frühwirth-Schnatter, S., Data augmentation and dynamic linear models, *Journal of time series analysis*, **15**, 183-202, 1994.
- Garcia, R., & Tsafack, G., Dependence Structure and Extreme Comovements in International Equity and Bond Markets, *Journal of Banking & Finance*, **35**, 1954-1970, 2011.
- Genest, C., Ghoudi, K., & Rivest, L. P., A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**, 543-552, 1995.
- Gruber, L.F., & Czado, C., Sequential Bayesian model selection of regular vine copulas, *Bayesian Analysis*, 2015.
- Gruber, L.F., & Czado, C., Bayesian model selection of regular vine copulas, *submitted*, 2015.
- Hamilton, J. D., A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the Econometric Society*, 357-384, 1989.
- Hofmann, M. & Czado, C., Assessing the VaR of a portfolio using D-vine copula based multivariate GARCH models, Working Paper, Technische Universitaät München Zentrum Mathematik, 2010.
- Ishwaran, H. & James, L. F., Gibbs sampling methods for stick-breaking priors, *Journal of the American Statistical Association*, **96**, 161-173, 2001.
- Ishwaran, H. & Zarepour, M., Exact and approximate sum representations for the Dirichlet process, *Canadian Journal of Statistics*, **30**, 269-283, 2002.
- Jaakkola, T., Meilä, M., Jebara, T., Maximum entropy discrimination, *Adv. Neur. In. 12*, MIT Press, 1999.
- Jaschke, S., Estimation of risk measures in energy portfolios using modern copula techniques, *Comput. Stat. Data An.*, **76**, 359-376, 2014.
- Jasra, A., Holmes, C.C., Stephens, D.A., Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling, *Stat. Sci.*, **20**, 50-67, 2005.
- Joe, H., Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters, *IMS Lecture Notes-Monograph Series*, **28**, 120-141, 1996.
- Joe, H., Multivariate model and dependence concepts, *Monographs on Statistics an Applied Probability*, **73**, Chapman, Hall, London, 1997.
- Joe, H., Asymptotic efficiency of the two-stage estimation method for copula-based models, *J. Multivariate Anal.*, **94**, 401-419, 2005.
- Joe, H, Tail dependence in vine copulae. *Dependence Modeling: Vine Handbook*, 165-189, 2011.

- Joe, H., *Dependence Modeling with Copulas*, CRC Press, 2014.
- Kalli, M., Griffin, J. E., & Walker, S. G., Slice sampling mixture models. *Statistics and computing*, **21(1)**, 93-105, 2011.
- Kirshner, S., Learning with tree-averaged densities and distribution, *Adv. Neur. In.*, **20**, 2007.
- Kirshner, S. & Smyth, P., Infinite mixture of trees, *Proceedings of the 24th international conference on Machine learning*, 417-423, 2007.
- Klugman, S.A., Panjer, H.H., Willmot, G.E., *Loss Models: From Data to Decisions*, 3rd ed., John Wiley & Sons, NJ, 2008.
- Krupskii, P. & Joe, H., Factor Copula Models for Multivariate Data, *J. of Multivariate Anal.*, **120**,85-101, 2013.
- Krupskii, P. & Joe, H., Structured factor copula models: theory, inference and computation, *J. Multivariate Anal.*, **138**, 53-73, 2015.
- Kruskal, W. H., Ordinal measures of association, *Journal of the American Statistical Association*, **53**, 814-861, 1958.
- Kurowicka, D. Optimal truncation of vines, in *Dependence Modeling: Vine Copula Handbook*, World Scientific Publishing Co., 2011.
- Kurowicka, D. & Cooke, R., *Uncertainty analysis with high dimensional dependence modelling*, Wiley, Chichester, 2006.
- Laih, Y. W., Measuring rank correlation coefficients between financial time series: A GARCH-copula based sequence alignment algorithm, *Eur. J. Oper. Res.*, **232**, 375-382, 2014.
- Lauritzen, S., *Graphical Models*, Clarendon Press, Oxford, 1996.
- Ling, C.-H., Representation of associative functions, *Publ. Math. Debrecen*, **12**, 189-212, 1965.
- Liu, J., Extreme value theory and copula theory: a risk management application with energy futures, *Ph.D. Thesis, University of Victoria*, 2011.
- Lo, A. Y., On a class of Bayesian nonparametric estimates: I. Density estimates. *The annals of statistics*, **12**, 351-357, 1984.
- Lu, X.F., Lai, K.K., Liang, L., Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model, *Ann. Oper. Res.*, **219**, 333-357, 2011.
- Marimoutou, V. & Soury, M., Energy Markets and CO2 Emissions: Analysis by Stochastic Copula Autoregressive Model, *Energy*, **88**, 417-429, 2015.

- MacEachern, S. N., & Müller, P., Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics*, 7(2), 223-238, 1998.
- Meilä, M. & Jordan, M., Learning with mixtures of trees, *J. Mach. Learn. Res.*, 1-48, 2000.
- Meilä, M. & Jaakkola, T., Tractable Bayesian learning of tree belief networks, *Stat. Comput.*, **16**, 77-92, 2006.
- Min, A. & Czado, C., Bayesian Inference for Multivariate Copulas using Pair-copula Constructions, *J. Financ. Economet.*, **8**, 511-546, 2010.
- Neal, R. M., Bayesian mixture modeling, in *Maximum Entropy and Bayesian Methods*, 197-211, Springer Netherlands, 1992.
- Neal, R. M., Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, **9**, 249-265, 2000.
- Nelsen, R. B., *An introduction to copulas, Second Edition*, Springer-Verlag, New York, 2006.
- Oh, D. H., & Patton, A. J., Modelling dependence in high dimensions with factor copulas, *J. Bus. Econ. Stat.*, to appear, 2016.
- Okimoto, T., New evidence of asymmetric dependence structures in international equity markets, *Journal of Financial and Quantitative Analysis*, **43**, 787-815, 2008.
- Osiewalski, J. & Pajor, A., Bayesian Value-at-Risk for a Portfolio: Multi- and Univariate Approaches using MSF-SBEKK Models, *Central European Journal of Economic Modelling and Econometrics*, **2**, 253-277, 2010.
- Prüfer H., Neuer Beweis eines Satzes über Permutationen, *Archiv für Mhatematik und Physik*, **27**, 742-744, 1918.
- Reinaud, J., CO2 allowance & electricity price interaction, *Impact on industry's electricity purchasing strategies in Europe. IEA information paper*, 2007.
- Richardson, S., Discussion of Spiegelhalter et al. *J. R. Stat. Soc.*, Series B, **64**, 651-655, 2002.
- Rodriguez, J., Measuring financial contagion: A copula approach, *Journal of Empirical Finance*, **14**, 401-423, 2007.
- Schweizer, B., & Sklar, A., Operations on distribution functions not derivable from operations on random variables, *Studia Mathematica*, **52**,43-52, 1974.
- Schweizer B. & Wolff E.F., On nonparametric measures of dependence for random variables, *Ann Statist*, **9**, 879-885, 1981.

- Sethuraman, J., A constructive definition of Dirichlet priors, *Statistica sinica*, 639-650, 1994.
- Silva, R. & Gramacy R. B., MCMC Methods for Bayesian Mixtures of Copula, *J. Mach. Learn. Res.*, **5**, 512-519, 2009.
- Sklar, A., Fonctions de répartition á ndimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, **8**, 229-231, 1959.
- Sklar, A., Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*, 1-14, 1996.
- Spiegelhalter, D. J. et al., Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B*, **64**, 583-639, 2002.
- Stöber, J., & Czado, C., Regime switches in the dependence structure of multidimensional financial data, *Computational Statistics & Data Analysis*, **76**, 672-686, 2014.
- Szegö, G., Measures of risk, *Eur. J. Oper. Res.*, **163**, 5-19, 2005.
- Walker, S. G., Sampling the Dirichlet mixture model with slices, *Communications in Statistic - Simulation and Computation*, **36(1)**, 45-54, 2007.
- Wang, G.J., Xie, C., Tail dependence structure of the foreign exchange market: A network view, *Expert Syst. Appl.*, **46**, 164-179, 2016.
- Weinberg, M. D., Computing the Bayes factor from a Markov chain Monte Carlo simulation of the posterior distribution, *Bayesian Analysis*, **7**, 737-770, 2012.
- Wen, X., Wei, Y., Huang, D., Measuring contagion between energy market and stock market during financial crisis: A copula approach , *Energ. Econ.*, **34**, 1435-1446, 2012.
- Wu, C-C., Chung, H., Chang, Y-H., The economic value of co-movement between oil price and exchange rate using copula-based GARCH models, *Energ. Econ.*, **34**, 270-282, 2012.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Federico Bassetti for the continuous support during my Ph.D studies. He always made himself available to clarify my doubts and helped me through my research. His guidance has been really fundamental for me.

My sincere thanks also goes to my co-advisor Prof. Claudia Tarantola who always encouraged and supported me and gave me the opportunity to join her project and to work with the Enel Group. A special thanks also to Prof. Elena De Giuli for her constant support and to all the Enel guys, Roberto, Mirko, Federico, Carlo and Flavio, that gave me a lot of information and suggestions for the development of this work.

A lot of people made these last three years a wonderful period, old and new friends. Thanks to all my colleagues for making the Dimat a pleasant and funny place. Special thanks to all of you that are now not just colleagues, but great friends. Thanks also to all my friends that always supported and encouraged me, usually with wine, and are now ready to celebrate another special moment with me.

Last but not least, I thank all my amazing family that always believed in me and gave me all the support to overcome the bad moments. Among all, thanks to you, Sandro.