Multivariate Dependence Analysis via Tree Copula Models: an Application to One-year Forward Energy Contracts

Federico Bassetti[a], Maria Elena De Giuli[b], Enrica Nicolino[b,c], Claudia Tarantola[b,*]

[a] *Department of Mathematics, University of Pavia, Via Ferrata 5, 27100 Pavia, Italy*
[b] *Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy*
[c] *RGI S.p.A., Corso Svizzera 185, 10149 Torino, Italy.*

Abstract

We propose a novel multivariate approach for dependence analysis in the energy market. The methodology is based on tree copulas and GARCH type processes. We use it to study the dependence structure among the main factors affecting energy price, and to perform portfolio risk evaluation. The temporal dynamic of the examined variables is described via a set of GARCH type models where the joint distribution of the standardised residuals is represented via suitable tree copula structures. Working in a Bayesian framework, we perform both qualitative and quantitative learning. Posterior summaries of the quantities of interest are obtained via MCMC methods.

Keywords: Multivariate Analysis, Bayesian Analysis, Copula Model, Energy Market, Markov Chain Monte Carlo.

1. Introduction

In recent years, the behaviour of the energy market has assumed a crucial role in the global economy, impacting and influencing both economic and social activities. Energy price directly affects industrial costs, becoming a fundamental element in the decision-making process of companies and entrepreneurs. The energy price is related to the cost and quantity of raw materials used to produce it. Moreover, since 2005, it is also related to the price for carbon emission (CO_2). Indeed, when applying the Kyoto Protocol, the European Union Emissions Trading Scheme (EU-ETS) of 2005 set up caps for the CO_2

*Address for correspondence: Claudia Tarantola, Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy. e-mail: claudia.tarantola@unipv.it, phone: +39 0382 986213

Email addresses: federico.bassetti@unipv.it (Federico Bassetti), elena.degiuli@unipv.it (Maria Elena De Giuli), enrica.nicolino@rgigroup.com (Enrica Nicolino), claudia.tarantola@unipv.it (Claudia Tarantola)

emissions of plants. Installations can increase emissions above their caps by acquiring emission allowances. Furthermore, installations with emissions below caps are allowed to sell unused allowances. Permits can be traded in spot, future and option markets and the power sector is a key player in the EU-ETS, see e.g. Reinaud (2007). Finally, the elements determining the energy price have become increasingly interconnected in the last years.

In this paper, we use Bayesian AR-GARCH copula models to study the behaviour and the connections among the main factors affecting energy price (coal, gas, oil and CO_2 prices). Our aim is to identify the dependence structure characterising the market, with particular attention to tail behavior. We focus on two representative European markets, the Italian and the German. For both markets we consider daily observations of one year forward contracts subscribed in 2014. Differently from the German case, only in 2014 Italian power prices were traded on a regulated market. In order to investigate the effect of this event we also analyse the Italian market in the period 2013-2016. We work in terms of monthly logarithmic return rates and we model their temporal dynamic via AR-GARCH processes. We account for the dependence between the variables by fitting alternative copula models to the distribution of the standardised residuals.

We perform both qualitative and quantitative Bayesian analysis, and we show how suitable pictorial representations of the dependence structure of the processes can be obtained. Finally, we illustrate how market risk of energy portfolios can be easily estimated via Bayesian predictive measures.

The estimated dependence structures are in line with some specific characteristics of the current energy market. In particular, we observe that the price of Brent (one of the major classifications of oil) has a marginal influence on the power price and the commodities that mostly impact the energy price are natural gas and coal. Furthermore, for the Italian case we find that the pairwise dependence between variables increases for almost all the examined quantities from 2013 to 2016.

Among possible alternative models for dependence analysis, we focus on copula functions, which are nowadays very popular in finance, insurance, econometrics and recently in the analysis of commodity markets; see e.g. Czado et al. (2011); Wen et al. (2012); Wu et al. (2012); Jaschke (2014); Marimoutou and Soury (2015); Dalla Valle et al. (2016) and Oh and Patton (2017). Although there are different types of bivariate copulas available,

the choice of multivariate copulas is rather limited, due to computational and theoretical limitations. To overcome this issue, Joe (1996) introduced the Pair Copula Construction (PCC), as an instrument for building flexible multivariate copulas starting from a set of bivariate ones, referred to as linking copulas. The core of this approach is its graphical representation, called R-vines, that consists of a nested set of trees, each edge of which is associated with a linking copula, see Bredford and Cooke (2001, 2002). Unfortunately, R-vines present a combinatorial complexity that may create difficulties in structural learning and parameter estimation both in frequentist and in Bayesian settings. In the frequentist approach, a two steps estimation procedure, known as Inference Function for Margins (IFM) proposed by Joe (1997), is usually applied. Also in the Bayesian framework it is common the use of a suitable two steps procedure 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, Gruber and Czado (2015a,b) developed a Bayesian approach for vine with structural learning. Nevertheless, due to the nested structure of the R-vine, the algorithms used to simulate from the posterior distributions are computationally demanding.

In order to reduce the complexity of the learning procedure and develop an efficient Bayesian approach to jointly estimate the copula structure and its parameters, in this paper we rely on tree copula models introduced by Kirshner (2007). Tree copulas are truncated R-vines, see Kurowicka (2011), whose underlying graphical structure, simpler than the R-vine structure, allows the inference procedure to be simplified. Furthermore, they provide a pictorial representation of the dependence structure that is easy to explain to non-experts. Nevertheless, considering only tree structures may be too restrictive to represent a realistic dependence among variables. Hence, following Silva and Gramcy (2009) and Elidan (2013), we also examine finite and infinite mixture of tree copulas. In the latter case we assume a non-parametric Dirichlet Process prior.

The use of Bayesian techniques in contrast to frequentist methods is motivated by the fact that the latter are not asymptotically efficient when applied to copula models, see Joe (2005). Moreover, in the Bayesian setting parameters uncertainty can be considered in the prediction. Another advantage of our approach based on Markov Chain Monte Carlo (MCMC) methods is that it allows mixture models to be estimated easily. Finally,

portfolio predictive cumulative distributions, risk measures and credible intervals for all the estimated parameters can be straightforwardly approximated by using the output of the MCMC.

The outline of the paper is as follows. In Section 2 our tree copula models are presented. In Section 3 the Bayesian estimation methodology is outlined. Section 4 describes the application of the proposed methodology to the analysis of real data. Readers primarily interested in the application may wish to browse lightly through Sections 2-3 and focus on Section 4. Concluding remarks are given in Section 5. The details of the MCMC algorithms and further results on simulated data are provided in the Supplementary Material.

2. AR-GARCH copula models specification

In order to describe the dynamic of the prices of the commodities we rely on AR-GARCH copula models. Let $S_{t,k}$ be the price at day t of commodity k, and $X_{t,k} = \log\left\{S_{t+20,k}/S_{t,k}\right\}$ be the corresponding monthly logarithmic return rate. Varying t over the set of working days, we obtain for each commodity $k = 1, \ldots, N$ a time series $(X_{k,t})$ that we model via an $AR(p) - GARCH(q, r)$ structure. More precisely,

$$\begin{aligned} X_{k,t} &= \sum_{i=1}^{p} a_{k,i} X_{k,t-i} + \varepsilon_{k,t}, \\ \varepsilon_{k,t} &= \sigma_{k,t} Z_{k,t}, \\ \sigma_{k,t}^2 &= \sigma_k^2 + \sum_{i=1}^{q} b_{k,i} \sigma_{k,t-i}^2 + \sum_{j=1}^{r} c_{k,j} \varepsilon_{k,t-j}^2. \end{aligned} \tag{1}$$

Setting q and r equal to 0, one obtains an $AR(p)$ model with $\sigma_{k,t}^2 = \sigma_k^2$ for every $t \geq 1$.

The vectors $\boldsymbol{Z_t} = (Z_{1,t}, \ldots, Z_{N,t})$ for $t = 1, \ldots, T$ are usually assumed to be independent and identically distributed. A common assumption is that $Z_{k,t} = \varepsilon_{k,t}/\sigma_{k,t}$ are standardised residuals normally distributed with zero mean and unit variance, and are jontly normally distributed with unknown correlation matrix. As an alternative, in this work we propose copula based models for the vector $\boldsymbol{Z_t}$ of the standardised residuals, see Section 2.4.

In Sections 2.1, 2.2 and 2.3 we briefly introduce copula functions, the related notation and terminology needed to define our models.

2.1. Copula Functions

A popular and efficient tool in multivariate dependence analysis is the copula function. The advantage of copulas is the ability to obtain the joint multivariate distribution embedding the dependence structure of the variables. A copula is a multivariate distribution with uniform margins on the unit interval. It is used to couple one-dimensional marginal distributions in order to obtain the corresponding joint multivariate distribution. Sklar's theorem (Sklar, 1959) states that any N-dimensional cumulative distribution function (cdf) F, with univariate cumulative marginal distributions $F_1, \ldots, F_N$, can be written as $F(z_1, \ldots, z_N) = C\Big(F_1(z_1), \ldots, F_N(z_N)\Big)$, where C is a suitable copula function. Consequently, if F is absolutely continuous, the corresponding joint probability density function (pdf) is given by

$$f(z_1, \ldots, z_N) = c\Big(F_1(z_1), \ldots, F_N(z_N)\Big) f_1(z_1) \cdots f_N(z_N),$$

where c is the copula density function.

2.2. Tree Copula

As mentioned in Section 1 graphical models can be used to simplify the construction of multivariate copulas. In a graphical model, the structure of the graph provides a pictorial representation of the conditional independence relationships between the variables; for a detailed presentation and the relevant terminology see Lauritzen (1996).

In this paper, we consider a Markov tree model, a particular type of graphical model having as underling graph an undirected tree with set of nodes $\mathcal{V} = \{1, \ldots, N\}$ and set of edges $\mathcal{E}$ (unordered pair of nodes). A random variable is associated with each node of the tree and the global Markov property is used to read conditional independencies among them. According to this property, disconnected sets of variables are conditionally independent given a separating set. Since a tree is uniquely defined by its edge set, in the following we use $\mathcal{E}$ to denote the tree structure. We indicate with $\mathscr{E}_N$ the set, of cardinality N^{N-2}, of all tree structures with N nodes.

If $\boldsymbol{Z}$ is a random vector with multivariate (positive) pdf f on $\mathcal{Z} \subset \mathbb{R}^N$ represented by a Markov tree $\mathcal{E}$, then its joint density can be factorised as

$$f(z_1, \ldots, z_N) = \left[\prod_{(l,m) \in \mathcal{E}} \frac{f_{l,m}(z_l, z_m)}{f_l(z_l) f_m(z_m)} \right] \prod_{i=1}^{N} f_i(z_i), \tag{2}$$

where f_l is the marginal density of Z_l and $f_{l,m}$ is the joint density of (Z_l, Z_m).

Following the tree copula construction of Kirshner (2007), we represent each density $f_{l,m}$ in (2) via the corresponding bivariate linking copula density $c_{l,m}$ associated with the edge (l, m). Hence, equation (2) can be re-written as

$$f(z_1, \ldots, z_N) = \prod_{(l,m)\in\mathcal{E}} c_{l,m}\Big(F_l(z_l), F_m(z_m)\Big) \prod_{i=1}^{N} f_i(z_i),$$

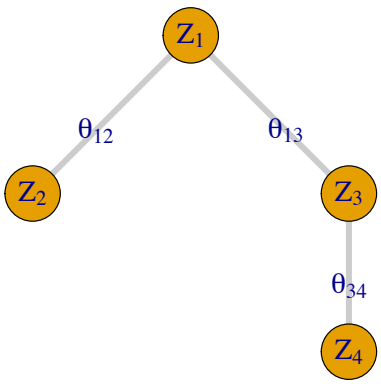
where F_l and F_m are the marginal cdfs of Z_l and Z_m.

Conversely, given a tree structure $\mathcal{E}$ and a family of a bivariate copula $c_{l,m}(u_l, u_m|\theta_{l,m})$ (parameterised through a parameter $\theta_{l,m}$), $c_\theta(u_1, \ldots, u_N) = \prod_{(l,m)\in\mathcal{E}} c_{l,m}(u_l, u_m|\theta_{l,m})$ is an admissible copula density. To simplify the notation, if m is the parent node of l in the directed version of $\mathcal{E}$ with root node 1, we set $c_{\theta_{l,m}}(u_l, u_m) = c_{l,m}(u_l, u_m|\theta_{l,m})$; otherwise if l is the parent node of m, $c_{\theta_{m,l}}(u_m, u_l) = c_{l,m}(u_l, u_m|\theta_{l,m})$. Consequently, $f_\theta(z_1, \ldots, z_N) = \prod_{(l,m)\in\mathcal{E}} c_{\theta_{l,m}}\Big(F_l(z_l), F_m(z_m)\Big) \prod_{i=1}^{N} f_i(z_i)$, is a density with margins f_i, $i = 1, \ldots, N$.

For example, for the tree copula reported in Figure 1, the joint density factorises as

$$f(z_1, z_2, z_3, z_4) = c_{\theta_{1,2}}\left(F_1(z_1), F_2(z_2)\right) c_{\theta_{1,3}}\left(F_1(z_1), F_3(z_3)\right) c_{\theta_{3,4}}\left(F_3(z_3), F_4(z_4)\right) \prod_{k=1}^{4} f_k(z_k).$$

Figure 1: Graphical representation of a tree copula on 4 variables.



2.3. Mixture of Tree Copula

In order to overcome the restrictions imposed by the tree structure, we consider copula functions obtained as the mixture of tree copulas. This strategy allows us to preserve

the relative low complexity of the Markov tree structures, taking into account richer dependencies between the variables. A mixture of Markov tree copulas is given by

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

where $D \leq +\infty$ is the number of mixture components, $(w_d)_{d=1,\ldots,D}$ are positive weights with $\sum_{d=1}^{D} w_d = 1$, $\mathcal{E}_d$ $(d = 1, \ldots, D)$ 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 $\mathcal{E}_d$.

A possible drawback of the mixture of tree copula is that its graphical model cannot be straightforwardly identified. See Meilă and Jordan (2000) for more details.

2.4. Distribution of the Standardised Residuals

In our AR-GARCH models we consider two alternative copula-based distributions for the vectors of standardised residuals. According to the first one, we assume that for any fixed t the dependence structure among the standardised residuals is given by a tree copula distribution with unknown underlying structure $\mathcal{E}$. Therefore, the pdf of the vector $\boldsymbol{Z}_t$ is given by

$$f_{\boldsymbol{Z}_t}(z_{1,t}, \ldots, z_{N,t}|\boldsymbol{\theta}, \mathcal{E}) = \prod_{(l,m)\in\mathcal{E}} c_{\theta_{l,m}}\Big(F_{\nu_l}(z_{l,t}), F_{\nu_m}(z_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}(z_{k,t}), \tag{3}$$

where $\boldsymbol{\theta}$ is the collection of all the copula parameters $\theta_{l,m}$ with $(l, m) \in \mathcal{E}$, F_{ν_k} is the marginal cumulative distribution function with parameter ν_k and f_{ν_k} the corresponding density. In the second model, we represent the joint density of the standardised residuals via a mixture of tree copulas, i.e.

$$f_{\boldsymbol{Z}_t}(z_{1,t}, \ldots, z_{N,t}|\boldsymbol{w}, \mathfrak{E}, \Theta) = \sum_{d=1}^{D} w_d \prod_{(l,m)\in\mathcal{E}_d} c_{\theta_{l,m}^{(d)}}\Big(F_{\nu_l}(z_{l,t}), F_{\nu_m}(z_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}(z_{k,t}),$$

where $\boldsymbol{w} = (w_1, \ldots, w_D)$, $\mathfrak{E} = \{\mathcal{E}_1, \ldots, \mathcal{E}_D\}$, $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_D\}$, $\boldsymbol{\theta}_d$ denoting the collection of the copula parameters $\theta_{l,m}^{(d)}$ for the d-th component.

The assumption of normality for the marginal distribution function F_{ν_k} of the standardised residuals may be not adequate due to possible heavy tails and asymmetry in the data. Hence, in addition to the case of Normal residuals, we also consider the case in which the standardised residuals follow a Skew Student-t distribution of parameters $\nu_k = (\lambda_k, \eta_k)$.

Following Hansen (1994), the Skew-t density with parameters λ (skewness) and η (degree of freedom) is defined as

$$f_{\lambda,\eta}(z) = \begin{cases} b\,c\left(1+\frac{1}{\eta-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & z < -a/b \\ b\,c\left(1+\frac{1}{\eta-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & z \geq -a/b \end{cases}$$

with $\eta > 2$ and $-1 < \lambda < 1$. The constants a, b and c are given by $a = 4\lambda c\left(\frac{\eta-2}{\eta-1}\right)$, $b^2 = 1 + 3\lambda^2 - a^2$, and $c = \Gamma\left(\frac{\eta+1}{2}\right)/\sqrt{\pi(\eta-2)\Gamma(\eta/2)}$.

Hansen (1994) shows that this is a proper density function with mean 0 and unit variance. Furthermore, if $\lambda = 0$ the Skew-t is reduced to the standard Student-t distribution. If $\lambda > 0$ ($\lambda < 0$), then this function is positively (negatively) skewed.

2.5. Linking Copulas

We assume that every bivariate copula density $c_{\theta_{l,m}}$ belongs to a specific family depending on a parameter $\theta_{l,m}$. We focus on two well-known family of copulas for tail dependence, Gumbel and Clayton, and their rotations (Double Gumbel and Double Clayton). These copulas have been widely used in applied analyses to study tail dependence between variables of interest. More sophisticated families of copulas for tail dependence can be effectively approximated for practical purposes by Gumbel and Clayton ones, see Demarta and McNeil (2005).

Double Gumbel and Double Clayton copulas are based on the rotations of standard Gumbel and Clayton copula family. The Gumbel copula is given by $C^G(u,v) = \exp\left\{-\left[(-\log u)^\phi + (-\log v)^\phi\right]^{1/\phi}\right\}$, where ϕ is a parameter assuming value in $[1,\infty)$. The Clayton copula is defined as $C^C(u,v) = \left[\max\left(u^{-\phi}+v^{-\phi}-1,0\right)\right]^{-1/\phi}$, where $\phi \in (0,\infty)$. In order to define Double Clayton and Double Gumbel copulas, we first reparameterise the copulas defined above in term of Kendall's tau measure; $\tau = (\phi-1)/\phi$ for the Gumbel copula, while $\tau = \phi/(2+\phi)$ in the case of Clayton copula.

Subsequently, rotations are used to obtain the Double Clayton and Double Gumbel copulas. If $c^G(u,v;\tau)$ is the Gumbel copula density reparameterised by the Kendall's tau, 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

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

In an analogous way we define the Double Clayton of first and second kind, DC_1 and DC_2, in term of rotations and Kendall's tau. Therefore, any linking copula $c_{\theta_{l,m}}$ is completely described by the parameter $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m})$ taking values in $(-1,1) \times \mathcal{H}$, where $\mathcal{H} = \{DG_1, DG_2, DC_1, DC_2\}$.

3. Methodology

In this section, we illustrate how to perform Bayesian inference for AR-GARCH copula models. We denote with Φ the collection of all the parameters describing the copula structures (for instance in the case of the simple tree copula distribution one has $\Phi = \{\boldsymbol{\theta}, \mathcal{E}\}$), with $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ the collection of the parameters of the marginal distributions and with $\boldsymbol{A}$ the collection of the parameters of the marginal time series models. In a fully Bayesian approach, if $\pi(\cdot)$ is the prior density on the parameters $(\Phi, \boldsymbol{\nu}, \boldsymbol{A})$ and $L(\mathcal{O}_T|\Phi, \boldsymbol{\nu}, \boldsymbol{A})$ is the likelihood of the observations $\mathcal{O}_T = \{(x_{1,t}, \ldots, x_{N,t}),\, t = 1, \ldots, T\}$, we obtain the posterior density using Bayes' theorem as

$$\pi(\Phi, \boldsymbol{\nu}, \boldsymbol{A}|\mathcal{O}_T) \propto L(\mathcal{O}_T|\Phi, \boldsymbol{\nu}, \boldsymbol{A})\pi(\Phi, \boldsymbol{\nu}, \boldsymbol{A}).$$

In principle, from the posterior distribution one can extract all the information needed for inference. In particular, one can compute posterior quantities, such as the mean or mode, as estimators for the parameters and evaluate predictive distributions for forecasting. Often the posterior distribution does not have a closed-form analytical expression. One possibility is to use MCMC methods to produce samples from a Markov chain having as stationary distribution the posterior of the model parameters, see e.g. Robert and Casella (2004). In the fully Bayesian approach, even MCMC algorithms can be inefficient and computationally demanding due to the structure of the likelihood.

In order to reduce the computational complexity of the fully Bayesian approach, in this paper we apply a combination of IFM and Bayesian procedures (shortly B-IFM), see e.g. Min and Czado (2010) and Dalla Valle et al. (2016). Following Min and Czado (2010) and Czado and Min (2011), in a first step we use a frequentist approach to estimate

the parameters $\boldsymbol{A}$ of the univariate marginal AR-GARCH models. This is done via the forward-backward approach implemented in the package *rugarch* of R. Subsequently, estimates of copula parameters Φ and of marginals parameters $\boldsymbol{\nu}$ (given the estimated parameters $\boldsymbol{A}$) are obtained in a Bayesian way. The posterior distributions for the copula parameters Φ and for the marginals parameters $\boldsymbol{\nu}$ (given the estimated parameters $\boldsymbol{A}$) can be obtained by

$$\pi(\Phi, \boldsymbol{\nu} | \boldsymbol{A}, \mathcal{O}_T) \propto L(\mathcal{O}_T | \Phi, \boldsymbol{\nu}, \boldsymbol{A}) \pi(\Phi, \boldsymbol{\nu}).$$

Although such a two steps procedure may lead to an underestimate of the uncertainty in the other parameters, in practice one usually does not see any significative difference between parameters estimated with a fully Bayesian approach and those estimated with this two steps approach. In Section 4 we compare the results obtained via our B-IFM procedures to the ones deriving from a fully Bayesian approach. Further comparisons are presented in the Supplementary Material.

In the following subsections, we describe the choice of the prior in the B-IFM setting for AR-GARCH copula models with Skew-t marginals. The case of AR copula models and/or Normal marginals, as well as the fully Bayesian setting, can be obtained via straightforward modifications. More details are provided in Section 4 and in the Supplementary Material. It is important to observe that the Bayesian models that will be described in Sections 3.1–3.3, can be easily adapted to employ other types of copulas and/or marginal distributions.

3.1. Bayesian Tree Copula Model

In this model, we assume that the standardised residuals have a tree copula distribution (3), where each $c_{\theta_{l,m}}$ is a Double Copula with parameter $\theta_{l,m}$, as described in Section 2.5. We suppose that the standardised residuals follow the univariate Hansen Skew-t pdfs with parameters $\nu_k = (\lambda_k, \eta_k)$ described in Section 2.4.

We assign the following independent prior distributions

$$\begin{aligned} \lambda_k &\sim \text{Beta}_{(-1,1)}(\alpha_k, \beta_k) \\ \eta_k &\sim \text{Exp}(\Lambda_k) \\ \theta_{l,m} = (\tau_{l,m}, \zeta_{l,m}) &\sim \text{Beta}_{(-1,1)}(\delta_{l,m}, \gamma_{l,m}) \times \text{Unif}_{\mathcal{H}}(\cdot) \\ \mathcal{E} &\sim \mathcal{U}(\cdot), \end{aligned} \tag{4}$$

for $k = 1, \dots, N$ and $(l, m) \in \mathcal{E}$. We denote with $\pi(\boldsymbol{\nu}, \boldsymbol{\theta}, \mathcal{E})$ the corresponding joint prior density. In (4), $\text{Beta}_{(-1,1)}(\alpha, \beta)$ indicates a translated Beta distribution on $(-1, 1)$ of parameters (α, β) [1] and $\text{Unif}_{\mathcal{H}}(\cdot)$ is a uniform distribution on $\mathcal{H} = \{DG_1, DG_2, DC_1, DC_2\}$. For the prior on the degrees of freedom, following Geweke (1993), we consider an exponential distribution $\text{Exp}(\Lambda)$ with parameter Λ. In the absence of specific prior information on the dependence structure, we use uniform prior on $\mathscr{E}_N$. Note that this prior is a special case of the default decomposable prior proposed by Meilă and Jaakkola (2006). The joint posterior density is

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

where $\boldsymbol{\theta}$ is the collection of all the copula parameters $\theta_{l,m}$.

Since the previous posterior density cannot be obtained in closed form, we rely on a Metropolis within Gibbs algorithm, based on the works of Silva and Gramcy (2009), Gruber and Czado (2015a,b). For more details, see the Supplementary Material.

3.2. Bayesian Tree Copula Mixture Model

We now consider the case in which the joint distribution of the standardised residuals is represented via a finite mixture of D tree copulas with double linking copulas and Hansen Skew-t margins. The pdf of the standardised residuals is equal to

$$f_{\boldsymbol{Z}_t}(z_{1,t}, \dots, z_{N,t}) = \sum_{d=1}^{D} w_d \prod_{(l,m)\in\mathcal{E}_d} c_{\theta_{l,m}^{(d)}}\left(F_{\nu_l}(z_{l,t}), F_{\nu_m}(z_{m,t})\right) \prod_{k=1}^{N} f_{\nu_k}(z_{k,t}).$$

In the previous equation, $\boldsymbol{w} = (w_1, \dots, w_D)$ is the vector of weights and $\mathcal{E}_d$ and $\boldsymbol{\theta}_d = \{\theta_{l,m}^{(d)}\}$ are the tree copula structure and the vector of copula parameters for the d-th component, respectively. Finally, we set $\mathfrak{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_D\}$ and $\Theta = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D\}$.

[1] X has translated Beta distribution on $(-1, 1)$ with parameters (α, β), if $X = 2Y - 1$ where Y has Beta distribution with parameters (α, β).

We assign the following independent prior distributions to the parameters

$$\begin{aligned} \boldsymbol{w} &\sim Dir(\psi_1, \ldots, \psi_D) \\ \lambda_k &\sim \text{Beta}_{(-1,1)}(\alpha_k, \beta_k) \\ \eta_k &\sim \text{Exp}(\Lambda_k) \\ \theta_{l,m}^{(d)} = (\tau_{l,m}^{(d)}, \zeta_{l,m}^{(d)}) &\sim \text{Beta}_{(-1,1)}(\delta_{l,m}, \gamma_{l,m}) \times \text{Unif}_{\mathcal{H}}(\cdot) \\ \mathcal{E}_d &\sim \mathcal{U}(\cdot) \end{aligned} \tag{5}$$

for $k = 1, \ldots, N$, $d = 1, \ldots, D$ and we denote with π the corresponding prior density. In the above $Dir(\psi_1, \ldots, \psi_D)$ 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 $w_D = 1 - \sum_{d=1}^{D-1} w_d$.

In order to sample from the posterior distribution of the finite mixture of tree copulas we adopt the data augmentation procedure by Van Dyk and Meng (2001). The details of the resulting Metropolis within Gibbs algorithm are described in the Supplementary Material.

3.3. DP-Tree Copula Model

As an extension of the tree copula mixture model described in the previous subsection, we propose a Bayesian nonparametric approach based on a Dirichlet process (DP) mixture model. In this way we do not need to fix a priori the number of the mixture components.

The Dirichlet process $DP(G_0, \psi)$ is a measure on measures. It has two parameters, a scaling parameter $\psi > 0$ and a base probability measure G_0. It was first formalised by Ferguson (1973) for general Bayesian statistical modelling, as a prior over distributions with wide support yet tractable posteriors. It is currently one of the most popular Bayesian nonparametric model. The so-called stick breaking construction, Sethuraman (1994), shows that if G is a draw from a $DP(G_0, \psi)$, then $G(\cdot) = \sum_{d \geq 1} w_d \delta_{\Psi_d}(\cdot)$ where the atoms Ψ_d are i.i.d. random variables with distribution G_0, δ_{Ψ_d} is a probability measure concentrated on Ψ_d and the weights w_d are generated using the stick-breaking construction. More precisely, $w_1 = v_1$ 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 DP mixture model can be derived as the limit of a finite mixture model where the number of the components tends to infinity, see e.g. Ishwaran and Zarepour (2002), and allows for the inclusion of the the uncertainty about the number of components.

In the DP-Tree Copula Model we assume that the pdf of the standardised residuals at time t has the form

$$f_{\boldsymbol{Z}_t}(z_{1,t},\dots,z_{N,t}) = \int \prod_{(l,m)\in\mathcal{E}} c_{\theta_{l,m}}\Big(F_{\nu_l}(z_{l,t}), F_{\nu_m}(z_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}(z_{k,t}) G(d\boldsymbol{\theta} d\mathcal{E}),$$

where $G \sim DP(\psi, G_0)$. By the stick breaking representation we obtain

$$f_{\boldsymbol{Z}_t}(z_{1,t},\dots,z_{N,t}) = \sum_{d\geq 1} w_d \prod_{(l,m)\in\mathcal{E}_d} c_{\theta_{l,m}^{(d)}}\Big(F_{\nu_l}(z_{l,t}), F_{\nu_m}(z_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}(z_{k,t}).$$

The base measure G_0 decomposes into the product of a prior on $\boldsymbol{\theta}_d$ and a prior on $\mathcal{E}_d$. We also assume that the concentration parameter ψ of the DP and the parameters $\boldsymbol{\nu}$ are unknown.

Summarising, in the DP-Tree Copula model the prior π is described by

$$\begin{aligned} \lambda_k &\sim \text{Beta}_{(-1,1)}(\alpha_k, \beta_k) \\ \eta_k &\sim \text{Exp}(\Lambda_k) \\ v_d &\sim \text{Beta}(1, \psi) \\ (\boldsymbol{\theta}_d, \mathcal{E}_d) &\sim G_0(\cdot) \\ \psi &\sim \text{Gamma}(a_\psi, b_\psi) \end{aligned} \tag{6}$$

for $k = 1,\dots,N$ and $d \geq 1$. In (6), $Gamma(a,b)$ denotes a Gamma density with shape parameter a and scale parameter b. 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 $\mathcal{E}_d$. Since $\boldsymbol{\theta}_d = \{\theta_{l,m}^{(d)}\}$, we set $p_0 = \prod_{(l,m)} \text{Beta}_{(-1,1)}(\delta_{l,m}, \gamma_{l,m})$.

In order to sample from the posterior distribution we adopt a variant of the slice sampling MCMC algorithm, proposed by Walker (2007) and Kalli et al. (2011). See the Supplementary Material.

3.4. Model comparison

A commonly used criterion for model comparison is the Deviance Information Criterion (DIC) introduced by Spiegelhalter et al. (2002). Since DIC cannot be applied to mixture models, we rely on DIC_3, a variant of the original criterion, see Richardson (2002) and Celeux et al. (2006). Given a density function $f(\boldsymbol{x}_{1:T}|\boldsymbol{\varphi})$ depending on a set of parameters $\boldsymbol{\varphi}$, the DIC is defined as

$$DIC = -4\mathbb{E}_{\boldsymbol{\varphi}}\Big[\log f(\boldsymbol{x}_{1:T}|\boldsymbol{\varphi})|\boldsymbol{x}_{1:T}\Big] + 2\log f(\boldsymbol{x}_{1:T}|\tilde{\boldsymbol{\varphi}}), \tag{7}$$

where $\tilde{\boldsymbol{\varphi}}$ is a posterior estimate of $\boldsymbol{\varphi}$ (a common choice is the posterior mean) and $\boldsymbol{x}_{1:T} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is the set of observations up to time T.

In the mixture model the set of parameters $\boldsymbol{\varphi}$ is not always identifiable, and consequently we cannot obtain $\tilde{\boldsymbol{\varphi}}$ straightforwardly. To overcome this problem, one can replace the term depending on $\tilde{\boldsymbol{\varphi}}$ in equation (7) with a function invariant under permutation. A natural choice is to consider the posterior predictive $\mathbb{E}_{\boldsymbol{\varphi}}\Big[f(\boldsymbol{x}_{1:T}|\boldsymbol{\varphi})|\boldsymbol{x}_{1:T}\Big]$, obtaining in this way the DIC_3 as

$$DIC_3 = -4\mathbb{E}_{\boldsymbol{\varphi}}\Big[\log f(\boldsymbol{x}_{1:T}|\boldsymbol{\varphi})|\boldsymbol{x}_{1:T}\Big] + 2\log\Big[\mathbb{E}_{\boldsymbol{\varphi}}\Big[f(\boldsymbol{x}_{1:T}|\boldsymbol{\varphi})|\boldsymbol{x}_{1:T}\Big]\Big].$$

Although these measures present some limitations (see the discussion of the paper of Spiegelhalter et al. (2002)), they are simple to calculate using MCMC and therefore, despite the criticisms, widely used in the Literature.

As an alternative measure for model comparison, following the econometric Bayesian literature, we consider the cumulative log-score, see e.g. Gneiting and Raftery (2007). It is defined as

$$\text{log-score} = \sum_{t=\tau_0}^{\tau_1} \log\Big\{\mathbb{E}_{\boldsymbol{\varphi}}\Big[f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{\varphi})|\boldsymbol{x}_{1:t}\Big]\Big\}, \tag{8}$$

where $\boldsymbol{x}_{1:t} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ is the set of observations up to time t, $\boldsymbol{\varphi}$ is the set of parameters, $f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{\varphi})$ is the conditional likelihood and $\mathbb{E}_{\boldsymbol{\varphi}}\Big[f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{1:t}, \boldsymbol{\varphi})|\boldsymbol{x}_{1:t}\Big]$ is the posterior predictive distribution. Higher values of the log-score correspond to better fits.

3.5. Dependence Analysis

For the tree copula model of Section 3.1 the dependence structure between the standardised residuals is directly encoded by the underling graphical structure $\mathcal{E}$. A simple estimate of the unknown tree structure $\mathcal{E}$ is the maximum a posteriori probability (MAP) tree structure defined as $\mathcal{E}_{MAP} = \text{argmax}_{\mathcal{E}}\pi(\mathcal{E}|\mathcal{O}_T)$. In addition to the MAP tree one can evaluate the values of the Kendall's tau between pairs of standardised residuals. As an estimate of this measure we compute its predictive posterior mean

$$\hat{\tau}_{i,j} = \mathbb{E}[\tau(z_{i,T+1}, z_{j,T+1}|\boldsymbol{\Phi}, \boldsymbol{\nu})|\mathcal{O}_T] = \int \tau(z_{i,T+1}, z_{j,T+1}|\boldsymbol{\Phi}, \boldsymbol{\nu}, \boldsymbol{A})\pi(d\boldsymbol{\Phi}d\boldsymbol{\nu}|\mathcal{O}_T, \boldsymbol{A}),$$

where $\tau(z_{i,T+1}, z_{j,T+1}|\boldsymbol{\Phi}, \boldsymbol{\nu}, \boldsymbol{A})$ is the Kendall's tau between the variables $z_{i,T+1}$ and $z_{j,T+1}$ under the AR-GARCH copula model given the unknown parameters $\boldsymbol{\Phi}, \boldsymbol{\nu}$ and the marginally estimated parameters $\boldsymbol{A}$. For the models described in Sections 3.1-3.3, numerical

approximations of the previous quantities can be easily obtained from the output of the corresponding MCMC algorithm.

For the mixture of tree copula models, in order to obtain a representative graphical structure, we apply the Minimum Spanning Tree (MST) approach. This procedure allows us to obtain a representative tree from the weighted graph based on Kendall's tau, see e.g. Wang and Xie (2016). More precisely, we consider a complete weighted graph in which the weight of each edge (i, j) is the absolute value of the (estimated) Kendall's tau between the variables i and j, i.e. $|\hat{\tau}_{i,j}|$. The Minimum Spanning Tree is the spanning tree that maximises the sum of edge weights or, equivalently, it is the spanning tree $\mathcal{E}^*$ such that $\sum_{(i,j)\in\mathcal{E}^*}(1-|\hat{\tau}_{i,j}|) = \min_{\mathcal{E}} \sum_{(i,j)\in\mathcal{E}}(1-|\hat{\tau}_{i,j}|)$, where the minimum is taken over the set of all possible spanning trees.

It is worth noticing that there is no direct correspondence between missing edges in the MST and conditional independences between the variables. In particular there is no connection between this tree and a possible Markov Tree structure associated to the joint distribution of the variables.

3.6. Risk Measures

Different risk measures are usually used to analyse and quantify the tail risk exposure, see e.g. Klumgman et al. (2008) and Szegö (2005). In order to evaluate the market risk of an energy portfolio we focus on two well-known quantile risk measures: Value-at-Risk (VaR) and Expected Shortfall (ES).

Following Artzner et al. (1999), the VaR at given probability level α is defined as

$$VaR_\alpha(V) = -\inf\{v : F_V(v) \geq \alpha\},$$

where F_V is the cdf of the net worth V of a portfolio. Typically α is set equal to 0.01 or 0.05, corresponding to the so-called 99% and 95% VaR. The VaR is one of the most commonly used risk measure by practitioners, it is easy to estimate and to explain even to non-experts. The 99% VaR for a horizon of two weeks is acceptable measure of risk according to the Basel Committee on Banking and Supervision of Banks for International Settlement (Basel Committee (1995) and following amendments). Nevertheless, many authors have criticised its adequacy as a measure of risk for different reasons, see e.g.

Acerbi and Tasche (2002). The main problems are the following. First of all, it considers only a single quantile of the portfolio distribution, so that it does not provide any information about the potential size of loss that exceeds its value. Secondly, it does not satisfy the sub-additivity property and, consequently, it may underestimate the portfolio risk. A measure that overcomes the previous problems is the Expected Shortfall (ES), see eg Artzner et al. (1999). For a significance level α, the ES is (minus) the conditional expectation of V, given that V is below $-VaR_\alpha(V)$, i.e.

$$ES_\alpha = -\mathbb{E}[V|V < -VaR_\alpha(V)].$$

ES is a coherent risk measure and, in contrast to VaR, is sensitive to the severity of losses beyond VaR. For a comprehensive and critical comparison between VaR and ES, see e.g. Embrechts et al. (2014) and Emmer et al. (2015).

Using a Bayesian approach, given the data observed until time T, the k-step-ahead VaR at level α, i.e. the $VaR_\alpha(V_{T+k})$, can be estimated using the α-quantile of the k-step-ahead (posterior) predictive distribution of the net worth. This posterior predictive distribution at time $T+k$ is given by

$$F_{V_{T+k}}(v|\mathcal{O}_T) := P\{V_{T+k} \leq v|\mathcal{O}_T\} = \int P\{V_{T+k} \leq v|\mathcal{O}_T, \boldsymbol{\Phi}, \boldsymbol{\nu}\}\pi(d\boldsymbol{\Phi}d\boldsymbol{\nu}|\mathcal{O}_T).$$

In the previous equation, $P\{V_{T+k} \leq v|\mathcal{O}_T, \boldsymbol{\Phi}, \boldsymbol{\nu}\}$ is the predictive distribution of V_{T+k} given the observations $\mathcal{O}_T$ and the unknown parameters $\boldsymbol{\Phi}, \boldsymbol{\nu}$, and $\pi(d\boldsymbol{\Phi}d\boldsymbol{\nu}|\mathcal{O}_T)$ is the posterior distribution of the parameters given the observations $\mathcal{O}_T$. Hence, $VaR_\alpha(V_{T+k})$, is estimated by

$$VaR_{\alpha,T+k} = -\inf\{v : F_{V_{T+k}}(v|\mathcal{O}_T) \geq \alpha\}$$

and the k step ahead ES at level α is estimated by

$$ES_{\alpha,T+k} = -\mathbb{E}[V_{T+k}|V_{T+k} < -VaR_{\alpha,T+k}, \mathcal{O}_T].$$

The previous quantities can be easily approximated using the MCMC output, see e.g. Osiewalski and Pajor (2010).

4. Empirical analysis

4.1. Data description

We apply the proposed models to the analysis of the Italian and German energy markets. We consider the following variables: Power (Italy/German), Brent, TTF (natural gas price deriving from transactions in virtual trading points in the Netherlands), CO_2 (price to pay for the emission of carbon dioxide into the atmosphere) and API (index representing a reference price benchmark for coal imported into North-West Europe). For the Italian market we also take into account PSV (natural gas price deriving from transactions in virtual trading points in Italy). The data are daily observations of one-year forward contracts. Power prices are obtained from the European Energy Exchange (EEX)[2], the leading energy exchange in Central Europe, while the remaining data are obtained from Reuters[3]; all values are expressed in Euros. Due to the intrinsic nature of these contracts, the dependence structure among the considered variables can differ significantly from one year to another. Since our model does not take into account changes of regime, in our analysis we focus on a single year at the time. In the following we present a detailed analysis of the behaviour of Italian and German markets from January 2014 to December 2014. For completeness of the analysis in Section 4.3 we present a multi-year analysis of the Italian market. As described in Section 2, in our analysis we work with the monthly logarithmic return rates $X_{t,i} = \log\left\{S_{t+20,i}/S_{t,i}\right\}$ where $S_{t,i}$ is the price at each day t of commodity i and t assumes values in the set of all working days. Descriptive measures for the considered commodities and for the corresponding logarithmic returns are reported in Table 1.

4.2. Prior settings and models comparison

In order to describe the marginal time series, we considered a variety of AR-GARCH models of different orders with Normal or Skew-t marginal distributions. Following a frequentist approach we selected the model and the order that best fits the data through the AIC criterion. Table 2 lists the AIC values for a set of representative cases.

[2]https://www.eex.com

[3]http://www.reuters.com

Table 1: Descriptive statistics for the examined commodities (daily prices) and the corresponding (monthly) logarithmic returns (from January 2014 to December 2014).

	Commodities						
	Power Italy	Power Germany	Brent	API	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
	Logarithmic return rates						
	Power Italy	Power Germany	Brent	API	TTF	PSV	CO_2
Mean	-0.34	-0.13	-0.37	-0.25	-0.58	0.19	-0.28
Std	0.022	0.023	0.028	0.030	0.027	0.131	0.062
Min	-3.18	-3.07	-2.84	-2.60	-2.59	-3.74	-3.83
Max	2.02	2.15	2.42	2.61	1.47	2.76	0.78
Kurt	3.46	2.99	3.27	3.44	2.08	4.90	6.20
Skew	-0.32	-0.17	0.22	0.20	-0.080	-0.59	-1.80

We note that none of the models has the lowest AIC value for all marginals. On the other hand, we observe that the GARCH model shows the worst fit and it will not be further investigated. We select AR(3)-GARCH(1,1) and AR(3) models that show a good fit for a high number of marginal series.

In the following, we apply the Bayesian models described in Sections 3.1-3.3: the tree copula model (TCM), the finite mixture of tree copulas model (Mix-TCM), and the DP-tree copula model (DP-TCM). In addition to the Skew-t margins case, for the sake of comparison, we also consider standard Normal marginal distributions. If an AR model is examined, we include in the parameters list also the precision $p_k = 1/\sigma_k^2$, see (1).

Table 2: AIC values for representative alternative models for the marginal series.

	PowIT	PowDE	TTF	PSV	API	CO_2	Brent
AR(2)-GARCH(2,2) Norm	140	131	312	263	67	75	-147
AR(2)-GARCH(2,2) Skew-t	112	132	299	250	69	73	-147
AR(3)-GARCH(1,1) Norm	141	**131**	307	255	63	75	-147
AR(3)-GARCH(1,1) Skew-t	**112**	132	293	243	66	**67**	**-148**
AR(3) Norm	144	134	303	252	**60**	144	-73
AR(3) Skew-t	118	134	**290**	**240**	64	96	-107
AR(2) Norm	142	132	304	257	60	140	-75
AR(2) Skew-t	115	132	291	245	63	94	-110
GARCH(1,1) Norm	469	498	517	510	532	390	218
GARCH(1,1) Skew-t	449	491	519	504	483	369	204

This means that we assume $\nu_k = (\lambda_k, \eta_k, p_k)$ for Skew-t margins and $\nu_k = p_k$ for Normal margins. Finally, for both cases we assume a standard Gamma prior for p_k.

For the prior distributions we consider the following hyperparameters values. We choose $(\delta_{l,m}, \gamma_{l,m}) = (1, 1)$, i.e. a uniform prior for each $\theta_{l,m}$. In the finite mixture models we use a symmetric Dirichlet prior with hyperparameters equal to 10 for the weights $\mathbf{w}$. In DP-tree copula model the hyperparameters of the Gamma prior on ψ are $(a_\psi, b_\psi) = (16, 0.25)$. In case of marginal Skew-t distribution we consider a translated Beta prior with parameters $(1, 1)$ on λ_k (skewness parameter) and an Exponential prior on η_k (degree of freedom) with mean 10. On the parameters precision p_k we assign a Gamma prior with parameters $(1, 1)$. Sensitivity analysis shows that the choice of the prior settings does not affect significantly the posterior estimates. The DIC_3 is slightly sensitive to the choice of the hyperparameters for the prior on ψ. Hence we tuned them to get the best result in term of DIC_3.

In Table 3 we compare the alternative models in terms of DIC_3. The DP tree model with Skew-t margins shows the best results for both the Italian and the German markets. Moreover, the models with Skew-t margins are always better than the corresponding ones with Normal margins. Log-scores, computed out of sample on the last 50 observations

Table 3: DIC_3 values for the alternative models

	Italy		**Germany**	
	Norm	Skew-t	Norm	Skew-t
AR(3) TCM	654	412	655	365
AR(3) Mix-TCM 2 Comp.	624	367	645	367
AR(3) Mix-TCM 5 Comp.	638	373	656	376
AR(3) DP-TCM	564	341	625	354
AR(3)-GARCH(1,1) TCM	235	217	291	273
AR(3)-GARCH(1,1)Mix-TCM 2 Comp.	235	216	287	276
AR(3)-GARCH(1,1)Mix-TCM 5 Comp.	237	237	291	287
AR(3)-GARCH(1,1) DP-TCM	223	207	281	270

and in sample on the last 200 observations, confirm these findings. In Figure 2, we report the comparison among the out of sample cumulative log-score obtained with the DP-tree models applied to the AR(3)-GARCH(1,1) and AR(3) estimated standardised residuals for each domestic market.

With regard to the marginal distribution of the standardised residuals, we computed the posterior mean and 95% credible interval for the parameters (λ, η) of the Skew-t. Table 4 shows the results for the DP-TCM. One can observe that the estimates of the skewness parameters λ_k are very close to zero for all standardised residuals. The 0.95 credible intervals suggest that there is posterior support for zero for all commodities, although the ones associated with CO_2 and Brent have zero very close to one of the end points. On the basis of the posterior means of λ_k, one can suppose that the standardised residuals of the CO_2 series are slightly positively skewed, while Brent's standardised residuals are slightly negatively skewed. As far as the estimated degrees of freedom is concerned, the tails of the API standardised residuals are the most close to Normal (with the estimates of η between 44 and 49 in the AR copula models and between 17 and 18 in the AR-GARCH copula models). For all the other commodities the estimated degrees of freedom range between 4 and 14, supporting the hypothesis of heavy tails.

As discussed in Section 3.3, using the DP-tree copula models one can also estimate the posterior distribution of the number of clusters. With the AR(3)-GARCH(1,1) DP-

Figure 2: Out of sample cumulative log-score (on the last 50 observations) for the DP-tree models: Italian market (first row) and German market (second row).

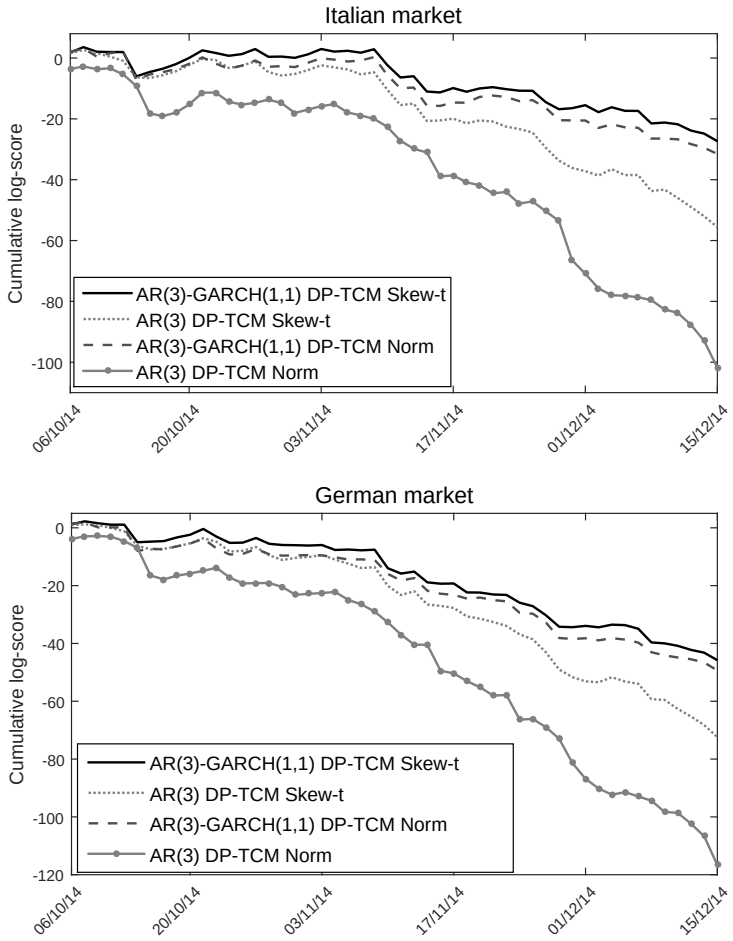


Table 4: Posterior mean and 95% credible interval (CI) for the parameters (η, λ) obtained with the DP-TCM with Skew-t margins.

		AR(3)-GARCH(1,1) marginal				AR(3) marginal			
Market		η	95% CI	λ	95%CI	η	95% CI	λ	95%CI
Italy	PowIT	4.69	[3.31,7.29]	0.03	[-0.12,0.21]	6.4	[3.33,14.0]	0.03	[-0.11,0.17]
	PSV	7.81	[4.49,15.02]	0.05	[-0.08,0.20]	7.9	[4.24,13.3]	-0.04	[-0.27,0.13]
	TTF	4.82	[3.39,7.65]	0.04	[-0.07,0.18]	7.8	[4.63,12.1]	0.03	[-0.16,0.19]
	API	17.7	[9.1,27.9]	-0.06	[-0.22,0.10]	44.0	[25.7,64.0]	-0.04	[-0.21,0.11]
	CO_2	4.71	[3.43,7.42]	0.12	[-0.03,0.27]	4.11	[2.80,6.41]	0.136	[-0.01,0.30]
	Brent	11.5	[6.33,19.32]	-0.16	[-0.33,0.01]	3.54	[2.25,5.91]	-0.10	[-0.24,0.04]
German	PowDE	13.58	[9.70,27.2]	0.10	[-0.09,0.26]	12.28	[4.30,20.8]	0.06	[-0.12,0.24]
	TTF	4.54	[3.20,7.31]	-0.03	[-0.1,0.13]	7.8	[3.35,15.3]	-0.03	[-0.19,0.11]
	API	18.6	[6.7,34.7]	-0.05	[-0.21,0.13]	49.1	[26.6,73.1]	-0.04	[-0.22,0.15]
	CO_2	4.38	[3.53,6.23]	0.12	[-0.01,0.30]	4.00	[2.84,6.11]	0.14	[0.01,0.30]
	Brent	12.5	[6.33,18.9]	-0.16	[-0.28,0.05]	3.90	[2.83,6.04]	-0.09	[-0.22,0.04]

tree the posterior mode of the number of clusters is 6 for the Italian market and 8 for the German market. We have similar results with the AR(3) DP-tree copula model for which the posterior mode of the number of clusters is 5 for the Italian and 8 for the German one. The posterior histograms of the number of clusters are reported in Figure S5 and S6 in the Supplementary Material.

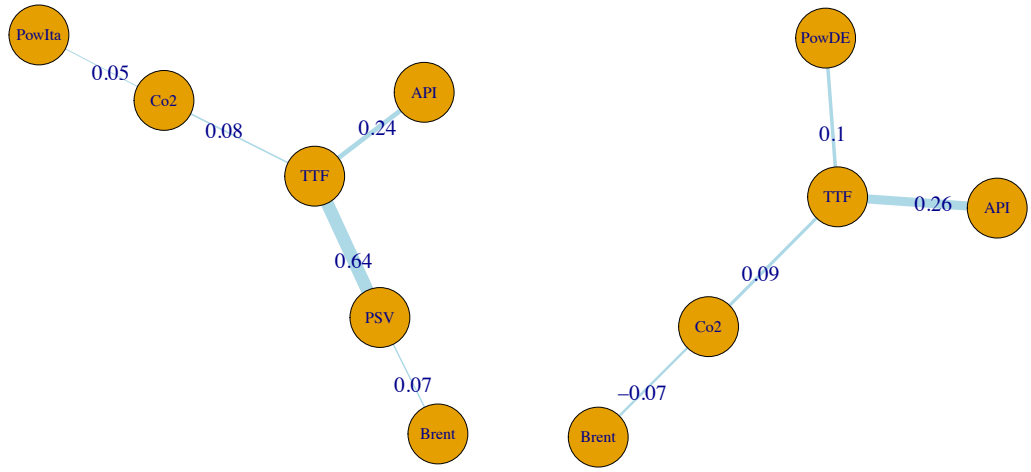
4.3. Dependence Structure Analysis

In this subsection we study the dependence between the commodities in each domestic market. We start by considering the tree copula model with Skew-t margins described in Section 3.1 applied to the estimated standardised residuals of the AR(3)-GARCH(1,1) and of the AR(3) models. The posterior distribution on tree structures is quite flat and the probabilities of the MAP trees ranges between 0.08 and 0.22 (in the alternative models/markets). As an example, in Figure 3 we report the MAP structures obtained for the AR(3)-GARCH(1,1) model for each domestic market, with posterior probabilities equal to 0.22 and 0.13, respectively. In Figure 5 we display the corresponding results

obtained with the AR(3) estimated standardised residuals. In this case, the posterior probabilities are 0.09 for the Italian market and 0.08 for the German market.

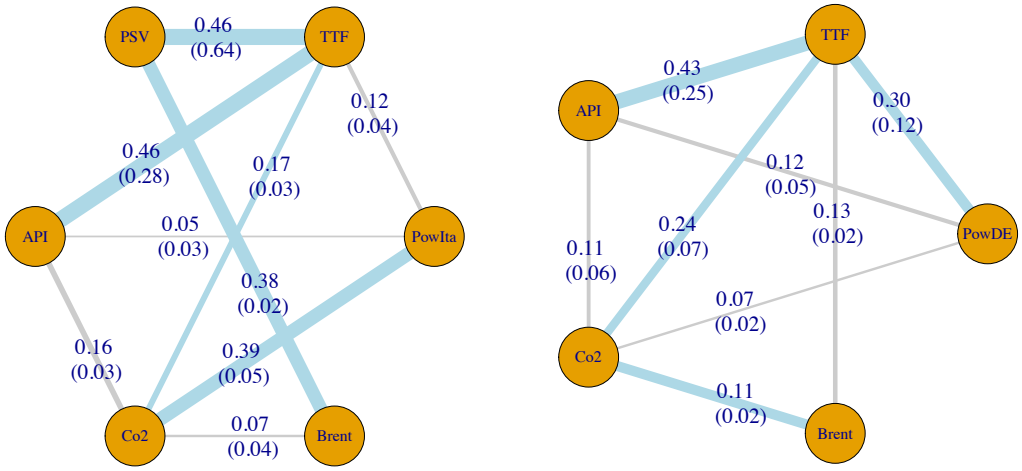
Ranking the trees with respect to their posterior probabilities, we find that the highest ranked trees share some interesting features. In order to obtain a cumulative posterior probability at least equal to 0.5, in the AR(3)-GARCH(1,1) model we need to consider 6 structures for the Italian market and 8 structures for the German one. In all these structures Brent is an end node. Moreover, in the Italian case, Power Italy is always connected to TTF or CO_2 and, in addition, the path PSV-TTF-API is always included. For the German market, in each of these structures we find the edge TTF-API and Power Germany is connected to TTF or CO_2. In case of AR(3) estimated standardised residuals, we need to consider the first 8 ranked structures for the Italian market and the first 10 for the German market in order to obtain a cumulative posterior probability at least equal to 0.5. Also in this case, for the Italian market we find the path PSV-TTF-API and Power Italy is connected to CO_2. In the German market there are always the edges TTF-API and Power Germany-CO_2 and Brent is an end node in both markets.

Figure 3: AR(3)-GARCH(1,1) TCM with Skew-t margins: MAP tree structure for the Italian market (left) and German market (right). The posterior mean of the Kendall's tau (given the MAP tree structure) is reported on each edge.



The ranked trees are in line with some well-known characteristics of the current energy markets. Indeed, in both markets, Brent is always an end node and the value of the Kendall's tau of the corresponding edge is quite low confirming that the use of this

Figure 4: AR(3)-GARCH(1,1) TCM with Skew-t margins: weighted graph for the Italian market (left) and German market (right). The weights on each edge correspond to the posterior probability that the same edge belongs to the ranked trees with at least 0.5 cumulative probability. The edges of the MST are reported in light blue. The (global) posterior mean of the Kendall's tau is reported within brackets on each edge.



commodity for energy production has drastically decreased in the last years. Moreover, as expected, the two gas nodes, TTF and PSV, are always connected in the Italian market and present high values of the Kendall's tau. The commodity that impacts mostly on the energy price is TTF, although also CO_2 plays an important role in the market. This is consistent with the fact that TTF is one of the main raw material used for energy production and the cost of CO_2 permissions cannot be neglected in the analysis of energy price behaviour, see e.g. Marimoutou and Soury (2015) and references therein.

Starting from the ranked tree structures, we constructed a graph with weights associated to each edge corresponding to the posterior probability that the same edge belongs to the ranked trees with at least 0.5 cumulative probability. Finally, we built the corresponding MST by maximising the posterior probabilities. The results are reported in Figures 4 and 6. The edges defining the MST are in light blue. The grey edges are those not included in the MST. Edges with weight < 0.02 are not depicted. We note that the MST structures are equal to the corresponding MAP trees, suggesting that the above discussed dependence paths are meaningful.

Following Section 3.5, we estimated pairwise Kendall's tau between the standardised residuals by computing the corresponding posterior mean for all the models, i.e. TCM,

Figure 5: AR(3) TCM with Skew-t margins: MAP tree structure for the Italian market (left) and German market (right). The posterior mean of the Kendall's tau (given the MAP tree structure) is reported on each edge.

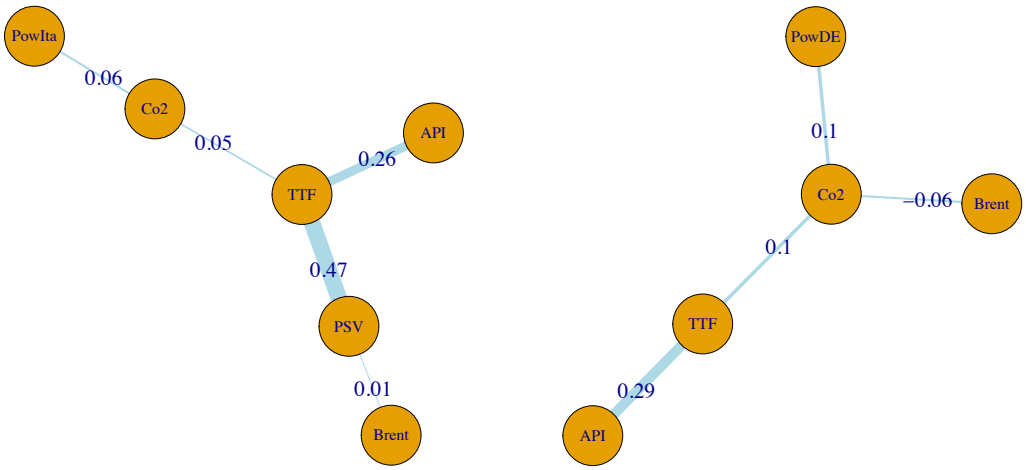


Figure 6: AR(3) TCM with Skew-t margins: weighted graph for the Italian market (left) and German market (right). The weights on each edge correspond to the posterior probability that the same edge belongs to the ranked trees with at least 0.5 cumulative probability. The edges of the MST are reported in light blue. The (global) posterior mean of the Kendall's tau is reported in brackets on each edge.

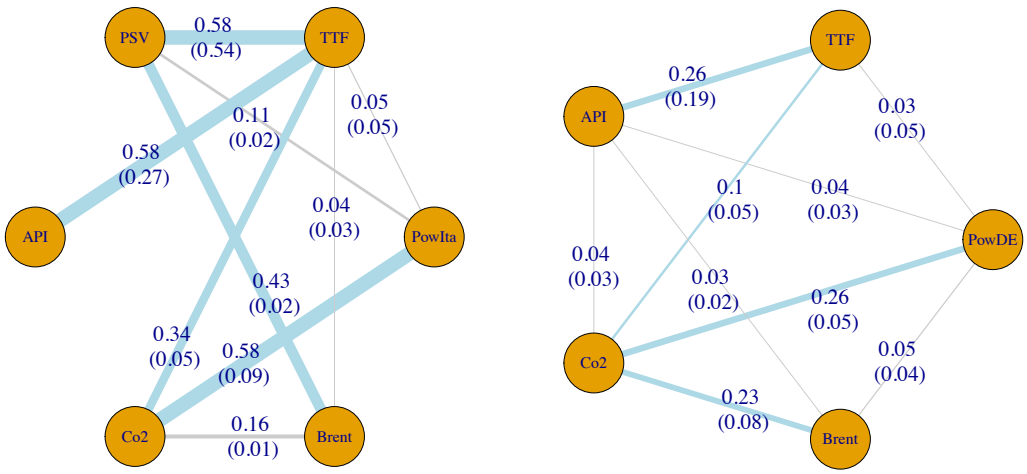


Figure 7: Kendall's tau weighted graph for the AR(3)-GARCH(1,1) DP-TCM with Skew-t margins. Italian market (left) and German market (right). The absolute value of the posterior mean of the Kendall's tau between the adjacent nodes is reported on each edge. The edges corresponding to MST are in ligth blue.

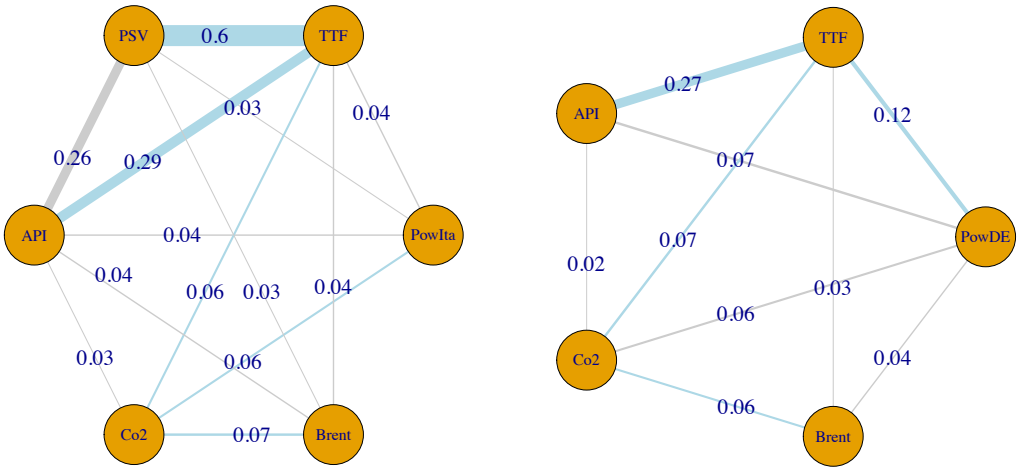
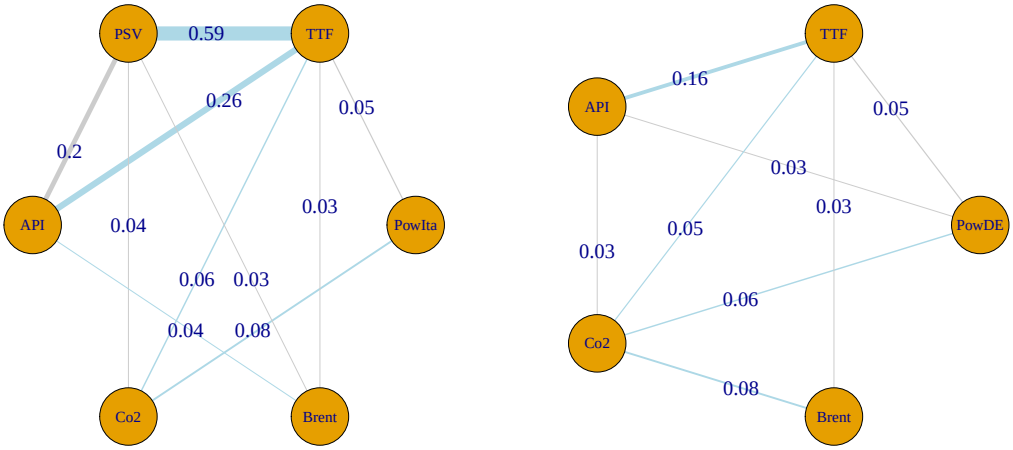


Figure 8: Kendall's tau weighted graph for the AR(3) DP-TCM with Skew-t margins Italian market (left) and German market (right). The absolute value of the posterior mean of the Kendall's tau between the adjacent nodes is reported on each edge. The edges corresponding to MST are in ligth blue.



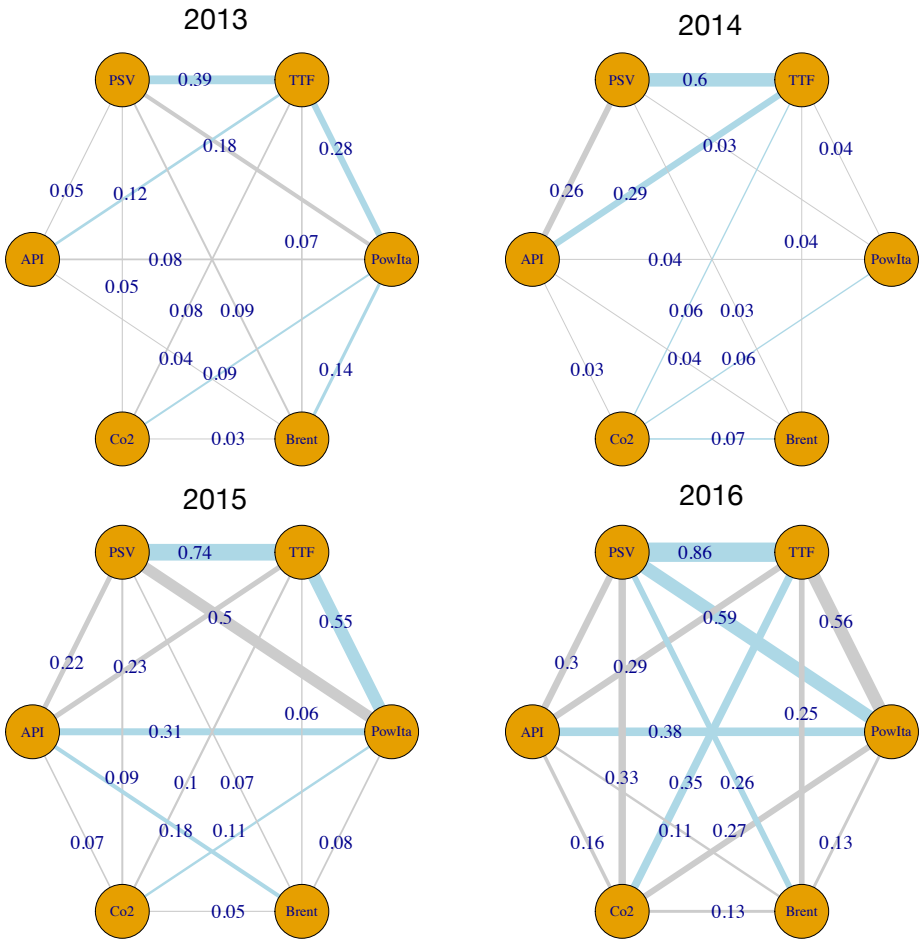
Mix-TCM and DP-TCM. The estimated values are then used to construct the weighted graphs and the MST structures. In Figures 7 and 8 we show the results for DP-TCM with Skew-t margins applied to the estimated standardised residuals of the AR(3)-GARCH(1,1) and AR(3). We can observe that some values of the Kendall's tau are very small. This is not surprising since we are dealing with the residuals 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 or to our estimation procedure, we also estimated these quantities by frequentist approach obtaining comparable results. Also a graphical comparison of the behaviour of the predictive residuals (obtained with our Bayesian model) with respect to the residuals of a frequentist IFM model suggests similar conclusions.

Comparing the MST structures for the DP-TCM models (Figures 7 and 8) and the MAP trees for the TCM models (Figures 3 and 5) we can note some similarities. For the German market the MST and the MAP tree structures are the same. On the other hand, for the Italian market they differ by the edge connecting Brent to the remaining part of each tree. In all structures Brent is an end point node with different neighbour sets. More precisely, in the MAPs it is adjacent to PSV, in the AR(3)-GARCH(1,1) MST it is connected to CO_2 and in the AR(3) MST is linked to API. All these edges present very low values of the estimated Kendall's tau.

We conclude this section with some results regarding a multi-year analysis of the Italian market. Italian energy contracts were traded for the first time by EEX in 2014. Here, we examine data from the previous year 2013 up to December 2016. Data for 2013 have been provided by traders of a leading Italian energy company. In Figure 9 we report, for each year, the weighted graph based on the absolute values of the Kendall's tau obtained with the AR(3)-GARCH(1,1) DP-TCM. It is interesting to note that the estimated values of the pairwise Kendall's tau increases for almost all variables from 2013 to 2016. More precisely, Kendall's tau between PSV and TTF increases from 0.37 to 0.86, reflecting the reduction of price differential between PSV and TTF from 2013 to 2016, see the annual reports of Italian Regulatory Authority for Electricity and Gas (http://www.autorita.energia.it). An analogous behaviour characterises the dependence between API and Power Italy as well TTF/PSV and Power Italy. In contrast, Kendall's tau between Power Italy and Brent does not present an increasing trend, ranging between

0.02 and 0.13. This can be possibly explained by the minor rule of Brent in the energy production.

Figure 9: Kendall's tau weighted graph for the AR(3)-GARCH(1,1) DP-TCM with Skew-t margins Italian market: 2013, 2014, 2015, 2016. The absolute value of the posterior mean of the Kendall's tau between the adjacent nodes is reported on each edge. The edges corresponding to MST are reported in light blue.



4.4. Energy Market Portfolio Analysis

In order to evaluate the market risk of an energy portfolio, we compute the Bayesian predictive VaR and ES. For each domestic market, we consider a portfolio made of one power asset and the remaining examined commodities. In our analysis, we work in the perspective of an energy company that sells energy and buys the other commodities in order to produce it. Consequently, in the portfolio composition, power has a positive weight, q_1, while the remaining components should have negative weights, $-q_i$,

$i = 2, \ldots, N$. Hence, the portfolio value (corresponding to its net worth) at time t is given by $V_t = q_1 S_{t,1} - \sum_{i=2}^{N} q_i S_{t,i}$. The composition of each portfolio has been provided us by experts of one of the major Italian energy company. More precisely, for the Italian market we consider a portfolio made of Power Italy, TTF, PSV, API, CO_2, Brent with weights $[1, 0, 0.14, 0.28, 0.69, 0]$, while for the German one, we use a portfolio made of Power Germany, TTF, API, CO_2, Brent with weights $[1, 0.61, 0.27, 0.76, 0]$. The weights are a good approximation to the real ones used by the energy company. In the following we refer to this portfolio as the "realistic" one.

For completeness, we tested our methodology using also alternative portfolio weights. Some results for an equally weighted portfolio are discussed in the last part of this subsection.

As described in Section 3.5, one can simulate the portfolio predictive distribution at time $T + 1$ on the basis of the information up to time T and, then, compute the related Bayesian predictive VaR and ES. As an example, the predictive portfolio value distribution at time $T + 1$ (T = 15th December 2014) for Italian/German market for AR(3)-GARCH(1,1) and AR(3) DP-TCM with Skew-t and Normal margins are shown in Figure 10. One can clearly note that the portfolio value corresponding to the AR(3) DP-TCM has larger variance with respect to the analogous portfolio estimated with AR(3)-GARCH(1,1) DP-TCM. In all the cases the portfolio distribution corresponding to models with Skew-t margins are more peaked although they show heavier tails.

We analysed the forecasting performance of our models using both in sample and an out of sample analysis for one day ahead portfolio value estimation. For the out of sample portfolio we considered the last 50 observations and we estimated the portfolio by the output of the MCMC algorithm (10000 iterations) for each $t = 174, \ldots, 224$. The estimated portfolios values are obtained by taking the predictive mean of the portfolio value, i.e. $\hat{V}_{t+1} = \mathbb{E}[V_{t+1}|\mathcal{O}_t]$. This value can be easily approximated by using the output of the MCMC. The results for the Italian and for the German portfolio obtained with the AR(3)-GARCH(1,1) DP-TCM models with Skew-t margins are reported in Figures 11 and 12, respectively. The estimated portfolio values (continuous line) are compared with the true portfolio values (dotted line). The 99% credible intervals of the estimated portfolio values are in grey and minus the 99% ES is plotted with a blue line. Clearly,

Figure 10: One day ahead predictive distributions of the Italian and German portfolio values (15th December 2014).

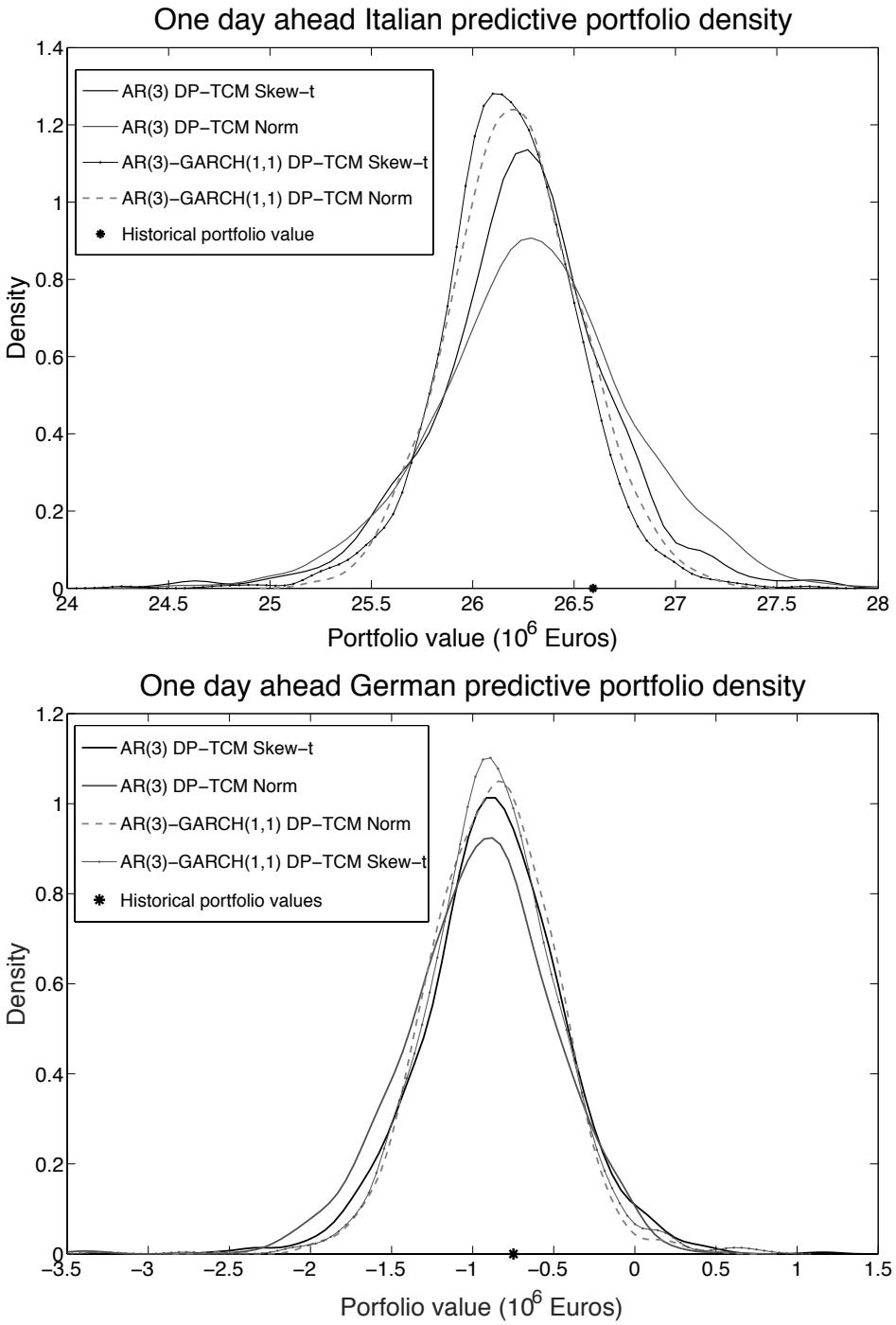
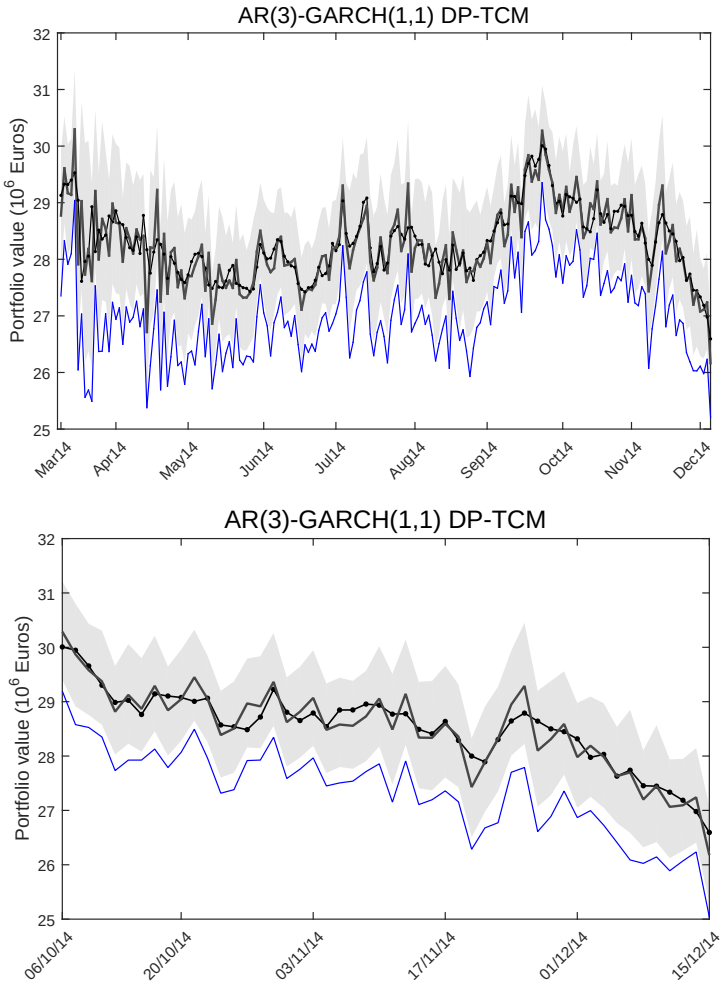


Figure 11: One day ahead estimated Italian portfolio values (continuous black line) compared with the true portfolio values (dotted black line), with 99% credible intervals (grey areas) and minus the 99% ES (blue line) under the AR(3)-GARCH(1,1) DP-TCM with Skew-t margins. In sample (on the last 200 observations) first row, out of sample (on the last 50 observations) second row.



minus the 99% VaR coincides with the lower bound of the grey areas. Overall, there is good agreement between the behaviours of predictive portfolios and the historical ones. In the out of sample case, the true values are always above minus the estimated 99% VaR.

Finally, Table 5 reports the Mean Absolute Error (MAE) between the out of sample forecasting portfolio and the historical portfolio, that is

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

where $\tau_0 = 174$ and $\tau_1 = 224$, corresponding to October 6, 2014 and December 15, 2014.

For the "realistic" portfolio the best performance in term of MAE is obtained by the

Figure 12: One day ahead estimated German portfolio values (continuous black line) compared with the true portfolio values (dotted black line) with 99% credible intervals (grey areas) and minus the 99% ES (blue line) under the AR(3)-GARCH(1,1) DP-TCM with Skew-t margins. In sample (on the last 200 observations) first row, out of sample (on the last 50 observations) second row.

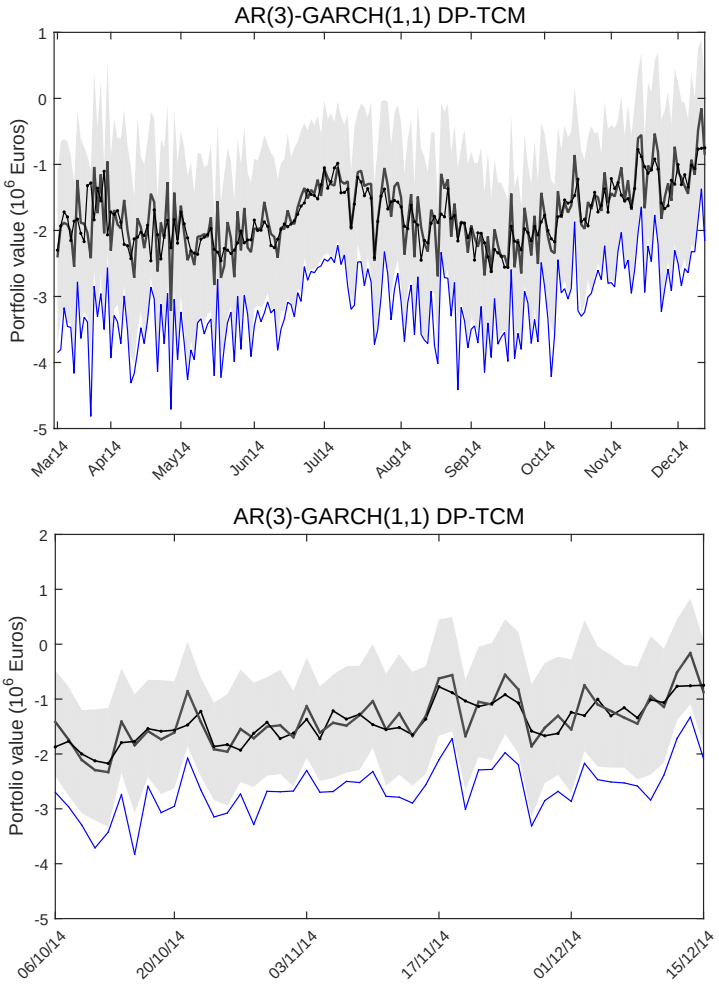


Table 5: MAEs between the out of sample forecasting portfolio (on the last 50 observations) and the historical portfolio.

	"Realistic" Portfolio				**Equally Weighted Portfolio**			
	Italy		**Germany**		**Italy**		**Germany**	
	Norm	Skew-t	Norm	Skew-t	Norm	Skew-t	Norm	Skew-t
AR(3) TCM	0.182	0.190	0.210	0.202	0.154	0.160	0.168	0.179
AR(3) Mixt-TCM	0.179	0.189	0.209	0.200	0.153	0.159	0.167	0.179
AR(3) DP-TCM	0.178	0.188	0.210	0.201	0.153	0.160	0.168	0.178
AR(3)-GARCH(1,1) CTM	0.194	0.193	0.207	0.201	0.161	0.161	0.178	0.181
AR(3)-GARCH(1,1) Mix-TCM	0.193	0.192	0.206	0.200	0.160	0.161	0.177	0.180
AR(3)-GARCH(1,1) DP-TCM	0.193	0.190	0.205	0.198	0.159	0.160	0.176	0.180

AR(3)-DP-TCM with Normal margins for the Italian portfolio and by AR(3)-GARCH(1,1)-DP-TCM with Skew-t margins for the German portfolio. It should be noted that the MAEs of the AR(3) models differ significantly from the corresponding MAEs for the AR(3)-GARCH(1,1) models in the Italian case, while in the German portfolio exercise all the MAEs are very close. In sample MAEs on 200 observations (see Table S13 in the Supplementary Material) show similar trends, in particular all the methods have similar MAEs when applied to the forecasting of the German portfolio. For this reason we computed the MAEs for other type of portfolio weights. As an example, in Table 5, we show the (out of sample) estimates for an equally weighted portfolio (see Table S13 in Supplementary Materials for the in sample results). For this portfolio, the trends are similar for both the Italian and the German market and the AR models show the best results in term of MAE (both in the in sample and in the out of sample experiments).

We conclude this subsection with a comparison of the previous results with the ones obtained with a fully Bayesian approach with AR tree copula models. In the fully Bayesian case, in addition to the priors (4), (5) and (6) we also consider a Normal prior on the AR parameters:

$$\boldsymbol{\alpha}_k \sim \mathcal{N}_p(\boldsymbol{M_k}, \boldsymbol{\Sigma_k})$$

where $\mathcal{N}_p(\boldsymbol{M}, \boldsymbol{\Sigma})$ is a p-dimensional Normal distribution with mean $\boldsymbol{M}$ and covariance matrix $\boldsymbol{\Sigma}$. Table 6 reports the MAEs between the out of sample forecasting portfolio

values and the historical ones. It is worth noticing that the values are very similar to the ones obtained with the B-IFM procedures (see Table 5), although for the "realistic" portfolio they are slightly better. Overall, the improvement in the MAEs does not justify the additional computational work needed for a fully Bayesian estimation of the proposed models.

Table 6: MAEs between the out of sample (on the last 50 observations) forecasting portfolio for the fully Bayesian AR(3) tree copula models and the historical portfolio.

	"Realistic" Portfolio				**Equally Weighted Portfolio**			
	Italy		**Germany**		**Italy**		**Germany**	
	Norm	Skew-t	Norm	Skew-t	Norm	Skew-t	Norm	Skew-t
AR(3) TCM	0.179	0.181	0.208	0.200	0.155	0.159	0.171	0.178
AR(3) Mixt-TCM	0.178	0.178	0.208	0.198	0.155	0.159	0.170	0.177
AR(3) DP-TCM	0.174	0.176	0.205	0.195	0.153	0.158	0.169	0.176

5. Summary and concluding remarks

The aim of this paper was to propose a novel Bayesian methodology for multivariate dependence analysis in the energy market. Our final goal was to study the connections between the main factors affecting energy price, and to provide efficient tools for portfolio risk evaluations.

We presented a Bayesian analysis of AR-GARCH copula models, in which the joint distribution of the standardised residuals of a panel of AR-GARCH time series is described via suitable tree copula models. Tree copulas are a particular type of R-vines, whose simple underlying graphical structure allows for an efficient inferential engine. In addition, via the graphical representation of the model, dependencies among the variables can be easily explained to non-experts. Nevertheless, in some cases the independence constraints implied by the tree structure can be too stringent. For this reason, we also considered mixtures of tree copulas. Using this strategy we preserved the relative low complexity of the tree copula structures, taking into account richer dependencies between the variables. We examined both the case in which the joint distribution of the standardised residuals is represented via a finite mixture of tree copulas and by an infinite mixture.

We applied our methodology to the analysis of two representative European markets, the Italian and German one. Our data are daily observations of one-year forward contracts. Due to the nature of the data, the dependence structure among variables can differ significantly from one year to another. Hence, we focused on a single year at the time. We presented a detailed analysis of year 2014 for both markets, and we provided a multi-year analysis of the Italian market from 2013 to 2016. We select this specific time period to to investigate the effect of the entrance of Italy in the EEX market in 2014. The estimated dependence structure are in line with specific characteristics of the current energy market. Bayesian predictive estimates of standard risk measures, i.e. VaR and ES, together with portfolio predictive distribution are easily obtained from the MCMC output.

Even if in our models we assumed that the marginal distributions of the standardised residuals are Normal or Skew-t and the linking copulas are Double Clayton and/or Double Gumbel, the methodology and the corresponding computational algorithms can be easily adapted to employ other types of copulas and/or marginal distributions.

Acknowledgements

The authors would like to thank the Referees and the Associate Editor for their very careful reading of our manuscript and the manifold pieces of advice for improvement. The authors are grateful to ENEL Group for providing the data, and for helpful comments and discussions. We also would like to thank Erika Blanc and Giorgia Oggioni for helpful comments and suggestions. The work was financially supported by the project "Multivariate Statistical Analysis for Extreme Value Risk Management in Energy Markets" (Contratto Aperto 8400057081) by Enel S.p.A. The work of the first author was partially supported by MIUR, Italy, PRIN 2015SNS29B-002. The work of the second author was partially supported by MIUR, Italy, PRIN MISURA 2010RHAHPL.

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**, 182-198.

Acerbi, C., & Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value-at-risk, *Economic Notes*, **31**, 379-388.

Arreola Hernandez, J. (2014). Are oil and gas stocks from the Australian market riskier than coal and uranium stocks? Dependence risk analysis and portfolio optimization, *Energy Economics*, **45**, 528-536.

Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1999). Coherent measures of risk, *Mathematical Finance*, **9**, 203-228.

Basel Committee on Banking Supervision (1995). *An Internal Model-Based Approach to Market Risk Capital Requirements*, Bank for International Settlements, Basel, Switzerland.

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

Bedford, T., & Cooke, R.M. (2002). Vines - a new graphical model for dependent random variables, *The Annals of Statistics*, **30**, 1031-1068.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models, *Bayesian Analysis*, **1**, 651-673.

Czado, C., Gärtner, F., & Min, A. (2011). 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.

Czado, C., & Min, A. (2011). Bayesian model selction fo D-vine pair copula constructions, *The Canadian Journal of Statistics*, **39**, 239-258.

Dalla Valle, L., De Giuli, M.E., Tarantola, C., & Manelli, C. (2016). Default probability estimation via pair copula constructions, *European Journal of Operational Research*, **249**, 298-311.

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

Elidan, G. (2013). Copulas and Machine Learning, *Copulae in Mathematical and Quantitative Finance*, Springer Berlin, 39-60.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., & Beleraj, A. (2014). An academic response to Basel 3.5, *Risks*, **2**, 25-48.

Emmer, S., Kratz, M., & Tasche, D. (2015). What Is the Best Risk Measure in Practice? A Comparison of Standard Measures, *Journal of Risk*, **18**, 31-60.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, 209-230.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics*, **8**, S19-S40.

Gneiting, T., & Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association,* **102**, 359-378.

Gruber, L.F., & Czado, C. (2015a). Sequential Bayesian model selection of regular vine copulas, *Bayesian Analysis*.

Gruber, L.F., & Czado, C. (2015b). Bayesian model selection of regular vine copulas, (Submitted for publication).

Hansen, B.E. (1994). Autoregressive conditional density estimation, *International Economic Review*, 705-730.

Hofmann, M., & Czado, C. (2010). Assessing the VaR of a portfolio using D-vine copula based multivariate GARCH models, Working Paper, Technische Universitaät München Zentrum Mathematik.

Ishwaran, H., & Zarepour, M. (2002). Exact and approximate sum representations for the Dirichlet process, *Canadian Journal of Statistics*, **30**, 269-283.

Jaschke, S. (2014). Estimation of risk measures in energy portfolios using modern copula techniques, *Computational Statistics & Data Analysis* , **76**, 359-376.

Jasra, A., Holmes, C.C., & Stephens, D.A. (2005). Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling, *Statistical Science*, **20**, 50-67.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters, *IMS Lecture Notes-Monograph Series*, **28**, 120-141.

Joe, H. (1997). Multivariate model and dependence concepts, *Monographs on Statistics an Applied Probability*, **73**, Chapman, Hall, London.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models, *Journal of Multivariate Analysis*, **94**, 401-419.

Kalli, M., Griffin, J.E., & Walker, S.G. (2011). Slice sampling mixture models. *Statistics and Computing*, **21(1)**, 93-105.

Kirshner, S. (2007). Learning with tree-averaged densities and distribution, *Advances in Neural Information Processing Systems*, **20**.

Klugman, S.A., Panjer, H.H., & Willmot, G.E. (2008). *Loss Models: From Data to Decisions*, John Wiley & Sons, NJ.

Kurowicka, D. (2011). Optimal truncation of vines, in *Dependence Modeling: Vine Copula Handbook*, World Scientific Publishing Co.

Laih, Y.W. (2014). Measuring rank correlation coefficients between financial time series: A GARCH-copula based sequence alignment algorithm, *European Journal of Operational Research*, **232**, 375-382.

Lauritzen, S. (1996). *Graphical Models*, Clarendon Press, Oxford.

Marimoutou, V., & Soury, M. (2015). Energy Markets and CO_2 Emissions: Analysis by Stochastic Copula Autoregressive Model, (2015), *Energy*, **88**, 417-429.

Meilă, M., & Jordan, M. (2000). Learning with mixtures of trees, *Journal of Machine Learning Research*, 1-48.

Meilă, M., & Jaakkola, T. (2006). Tractable Bayesian learning of tree belief networks, *Statistics and Computing*, **16**, 77-92.

Min, A., & Czado, C. (2010). Bayesian Inference for Multivariate Copulas using Pair-copula Constructions, *Journal of Financial Econometrics*, **8**, 511-546.

Oh, D.H., & Patton, A.J. (2017). Modelling dependence in high dimensions with factor copulas, *Journal of Business & Economic Statistic*, **35**, 139-145.

Osiewalski, J., & Pajor, A. (2010). 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.

Reinaud, J. (2007). CO_2 allowance & electricity price interaction, *Impact on industry's electricity purchasing strategies in Europe. IEA information paper.*

Richardson, S. (2002). Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society*, Series B, **64**, 651-655.

Robert, C., & Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, 639-650.

Silva, R., & Gramcy R.B. (2009). MCMC Methods for Bayesian Mixtures of Copula, *Journal of Machine Learning Research*, **5**, 512-519.

Skiena, S. (1990). *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*. Reading, MA, Addison-Wesley, 1990.

Sklar, A. (1959). Fonctions de répartition á *n*dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, **8**, 229-231.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society*, Series B, **64**, 583-639.

Szegö, G. (2005). Measures of risk, *European Journal of Operational Research*, **163**, 5-19.

Van Dyk, D.A., & Meng, X.L. (2001). The art of data augmentation, *Journal of Computational and Graphical Statistics*, **10**, 1-50.

Walker, S.G. (2007). Sampling the Dirichlet mixture model with slices, *Communications in Statistic - Simulation and Computation*, **36(1)**, 45-54.

Wang, G.J., & Xie, C. (2016). Tail dependence structure of the foreign exchange market: A network view, *Expert Systems with Applications*, **46**, 164-179.

Wen, X., Wei, Y., & Huang, D. (2012). Measuring contagion between energy market and stock market during financial crisis: A copula approach, *Energy Economics*, **34**, 1435-1446.

Wu, C-C., Chung, H., & Chang, Y-H. (2012). The economic value of co-movement between oil price and exchange rate using copula-based GARCH models, *Energy Economics*, **34**, 270-282.

Supplementary Material for:

"Multivariate Dependence Analysis via Tree Copula Models: an Application to One-year Forward Energy Contracts"

by F. Bassetti, M.E. De Giuli, E. Nicolino, C. Tarantola.

S1. Computational details

In this section we present the MCMC algorithms for the particular case of the fully Bayesian AR copula models with Skew-t margins. Inference on B-IFM AR-GARCH copula models is obtained by discarding the metropolis step on the parameters $\mathbf{A}$, which are estimated marginally and hence are fixed in the sampler. Finally, if the Normal margins are used the only change is the full conditional of $\boldsymbol{\nu}$.

In what follows we compactly write

$$\epsilon_{m,t} := x_{m,t} - \sum_{i=1}^{p} a_{m,i} x_{m,t-i}. \tag{S1}$$

In order to obtain the algorithms for the B-IFM AR-GARCH copula models one simply need to assume

$$\epsilon_{m,t} := (x_{m,t} - \sum_{i=1}^{p} a_{m,i} x_{m,t-i})\sigma_{m,t}^{-1} \tag{S2}$$

and remember that $\sigma_{m,t}$ are estimated marginally and hence are fixed in the MCMC algorithm.

S1.1. MCMC for the Tree Copula Model

The algorithm iteratively samples from the full conditionals:

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

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

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

The variance of the random walk proposal densities are tuned to achieve acceptance rates between 20% and 80%.

• *Full conditional of* $\boldsymbol{\nu}$. Recall that in the case of the fully Bayesian AR copula models with Skew-t margins, $\nu_k = (\lambda_k, \eta_k, p_k)$. The full conditional of $\boldsymbol{\nu}$ given $(\mathcal{E}, \boldsymbol{\theta}, \boldsymbol{A}, \mathcal{O}_T)$ is

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

where $\pi(\lambda_k)$ denotes the Beta prior on the skewness parameters and $\pi(\eta_k)$ the Exponential prior on the degrees of freedom and $\pi(p_k)$ the Gamma prior on the precision $p_k = 1/\sigma_k$. In order to sample the marginal parameters, we consider three separate Metropolis-Hasting steps, one for the skewness parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, one for the degrees of freedom $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ and one for $\boldsymbol{p} = (p_1, \ldots, p_N)$.

As for $\boldsymbol{\lambda}$, we assume a proposal density of the form

$$q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}) = \prod_{k=1}^{N} q(\lambda_k^*|\lambda_k)$$

where we sample each λ_k^* (after an appropriate change of variable) with a random walk proposal on $\mathbb{R}$. The acceptance probability of the corresponding MH step is given by

$$\min\left\{1, \prod_{t=1}^{T} \prod_{(l,m)\in\mathcal{E}} \frac{c_{\theta_{l,m}}\Big(F_{(\lambda_l^*,\eta_l,p_l)}(\epsilon_{l,t}), F_{(\lambda_m^*,\eta_m,p_m)}(\epsilon_{m,t})\Big)}{c_{\theta_{l,m}}\Big(F_{(\lambda_l,\eta_l,p_l)}(\epsilon_{l,t}), F_{(\lambda_m,\eta_m,p_m)}(\epsilon_{m,t})\Big)} \prod_{k=1}^{N} \frac{f_{(\lambda_k^*,\eta_k,p_k)}\Big(\epsilon_{k,t}\Big)}{f_{(\lambda_k,\eta_k,p_k)}\Big(\epsilon_{k,t}\Big)}\right\}.$$

For the parameters $\boldsymbol{\eta}$ and $\boldsymbol{p}$ we proceed similarly.

• *Full conditional of* $\Big(\mathcal{E}, \boldsymbol{\theta}\Big)$. Samples from the full conditional of $\Big(\mathcal{E}, \boldsymbol{\theta}\Big)$ given $\Big(\boldsymbol{\nu}, \boldsymbol{A}\Big)$ are obtained again by using a Metropolis-Hasting step. The full conditional is

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

where $\pi(\mathcal{E}, \boldsymbol{\theta})$ is the prior over $(\mathcal{E}, \boldsymbol{\theta})$ specified in Section 3.1. First, we sample the new tree $\mathcal{E}^*$ and then, conditionally on the new tree $\mathcal{E}^*$, we (independently) sample the parameters $\theta_{l,m} = (\tau_{l,m}, \zeta_{l,m})$. Formally, the proposal is

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

In order to propose a new tree $\mathcal{E}^*$, one can use different proposals. For instance one can simply sample from the prior or, as we have done in the present work, one can use the local move by Silva and Gramcy (2009). Further details on this local move are given

in the next paragraph. As for the copula parameters, 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 these parameters do not appear in the copula construction and, since they are not used in the acceptance probability, one does not need 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 cantered 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 $H = \{DG_1, DG_2, DC_1, DC_2\}$

$$q\Big(\zeta^*_{l,m} = j | \tau^*_{l,m}\Big) = \frac{\prod_{t=1}^T c_{\left(\tau^*_{l,m}, j\right)}\Big(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\Big)}{\sum_{h \in H} \prod_{t=1}^T c_{\left(\tau^*_{l,m}, h\right)}\Big(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\Big)}.$$

The acceptance probability of the MH step is

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

where $q(\tau|\tau_0)$ is the (transformed) Normal proposal centred on τ_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 Gramcy (2009). This move leaves the tree unchanged except for a triangular path say $i-j-k$ which is changed to $j-i-k$. As proved in Silva and Gramcy (2009), it is possible to traverse the whole space of spanning trees with sequences of tree-angular moves. More precisely, the tree-angular proposal, say $q^{TA}(\cdot|\mathcal{E})$, given a tree $\mathcal{E}$, propose a new tree $\mathcal{E}^*$ as follows:

1. choose an edge (u,v) in $\mathcal{E}$ (with uniform probability);
2. choose a neighbour node w of (u,v) in the tree $\mathcal{E}$, which is neither u nor v, uniformly at random;

3. obtain 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

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

with some fixed probability η.

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

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

where $\pi(\boldsymbol{\alpha}_k)$ is the p-dimensional Normal prior $\mathcal{N}_p\Big(\boldsymbol{M}_k, \boldsymbol{\Sigma_k}\Big)$. We proceed again with a Metropolis-Hasting step.

The proposal density $q(\boldsymbol{A}^*|\boldsymbol{A})$ is a multivariate Normal distribution with mean $\boldsymbol{A}$ and covariance matrix $\sigma^2_{\boldsymbol{A}} I_{p\times N}$. The acceptance probability is

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

S1.2. **MCMC for the Finite Mixture of Tree Copulas**

We introduce an allocation variable for each observation, $I_t \in \{1, \ldots, D\}$ for $t = 1, \ldots, T$. The complete data likelihood is given by

$$L(\mathcal{O}_T, \boldsymbol{I}|\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{w}, \boldsymbol{A}) == \prod_{t=1}^{T} w_{I_t} \prod_{(l,m)\in\mathcal{E}_{I_t}} c_{\theta^{(I_t)}_{l,m}}\Big(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}\left(\epsilon_{k,t}\right) \quad \text{(S3)}$$

where $\boldsymbol{I} = (I_1, \ldots, I_T)$. The likelihood of the finite mixture model can be easy obtained as marginal distribution of (S3). The posterior density of the allocation variables $\boldsymbol{I}$ and the parameters is

$$\pi(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{w}, \boldsymbol{I}, \boldsymbol{A}|\mathcal{O}_T) \propto L(\mathcal{O}_T, \boldsymbol{I}|\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{w}, \boldsymbol{A})\pi(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{w}, \boldsymbol{A})$$

where $\pi(\cdot)$ is the prior setting described in (5).

The algorithm samples iteratively from the full conditionals:

i) $\boldsymbol{w}$ given $[\mathfrak{E}, \Theta, \boldsymbol{A}, \boldsymbol{\nu}, \mathcal{O}_T, \boldsymbol{I}]$ (closed form);

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

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

iv) $\boldsymbol{I}$ given $[\mathfrak{E}, \Theta, \boldsymbol{w}, \boldsymbol{\nu}, \boldsymbol{A}, \mathcal{O}_T]$ (closed form);

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

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

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

where $\pi(\boldsymbol{w})$ is the Dirichlet prior over $\boldsymbol{w}$ 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, i.e. $Dir\left(\psi_1 + N_1(I), \ldots, \psi_D + N_D(I)\right)$, and we can sample directly from it.

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

$$\begin{aligned}&\pi(\boldsymbol{\nu}|\mathfrak{E}, \Theta, \boldsymbol{A}, \mathcal{O}_T, \boldsymbol{I}) \propto \\ &\qquad \prod_{t=1}^{T} \prod_{(l,m)\in\mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}\Big(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\Big) \prod_{k=1}^{N} f_{\nu_k}\left(\epsilon_{k,t}\right) \pi(\lambda_k)\pi(\eta_k)\pi(p_k).\end{aligned}$$

We sample the marginal parameters using MH steps, as done in Section S1.1 for the analogous point.

• *Full conditional of* $(\mathfrak{E}, \Theta)$. For every $(\mathcal{E}_d, \boldsymbol{\theta}_d)$, $d = 1, \ldots, D$ the full conditional is

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

where $\pi(\mathcal{E}_d, \boldsymbol{\theta}_d)$ is the prior specified in Section 3.2.
We proceed with a MH step. For every component $(\mathcal{E}_d, \boldsymbol{\theta}_d)$, we choose the proposal

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

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 is

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

where $q(\tau|\tau_0)$ is the (transformed) Normal proposal centred on τ_0.

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

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

We sample each I_t from a discrete distribution on $\{1, \ldots, D\}$ with probability of $\{I_t = d\}$ equal to

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

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

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

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

We consider a MH step similar to the previous algorithm.

S1.3. **MCMC for DP-tree copula model**

The slice sampling algorithm introduces two latent variables u_t and I_t for each observation $t = 1, \dots, T$. The complete augmented data likelihood becomes

$$\prod_{t=1}^{T} \mathbb{1}(w_{I_t} > u_t) \prod_{(l,m)\in\mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}}\left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\right) \prod_{k=1}^{N} f_{\nu_k}(\epsilon_{k,t})$$

where $\boldsymbol{U} = (u_1, \dots, u_T)$ and $\boldsymbol{I} = (I_1, \dots, I_T)$.

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 and $\mathcal{D}^* = \max \mathcal{D}$ is the number of stick breaking components used in the mixture. In the following we denote with $\boldsymbol{V} = (v_1, v_2, \dots)$ the element of the stick breaking construction and with $(\mathfrak{E}, \Theta)$ the collection of all the tree structures and the corresponding copula parameters, respectively.

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

(i) $\boldsymbol{\nu}$ given $[\mathfrak{E}, \Theta, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);

(ii) $(\mathfrak{E}, \Theta)$ given $[\boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);

(iii) $\boldsymbol{U}, \boldsymbol{V}, \psi$ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \boldsymbol{I}, \mathcal{O}_T]$ sampled as a block:

- ψ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \boldsymbol{I}, \mathcal{O}_T]$ (Metropolis-Hasting step);
- $\boldsymbol{V}$ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{I}, \mathcal{O}_T]$ (closed form);
- $\boldsymbol{U}$ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{I}, \mathcal{O}_T]$ (closed form);

(iv) $\boldsymbol{I}$ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T]$ (closed form);

(v) $\boldsymbol{A}$ given $[\mathfrak{E}, \Theta, \boldsymbol{\nu}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T]$ (Metropolis-Hasting step).

- *Full conditional of* $\boldsymbol{\nu}$. As in the previous algorithm.

- *Full conditional of* $\mathfrak{E}, \Theta$. The full conditional of $(\mathfrak{E}, \Theta)$ given $(\boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T)$ is

$$\pi(\mathfrak{E}, \Theta | \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{I}, \mathcal{O}_T) = \prod_{d:I_t=d} \prod_{(l,m)\in\mathcal{E}_d} c_{\theta_{l,m}^{(d)}}\Big(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t})\Big) G_0(\boldsymbol{\theta}_d, \mathcal{E}_d)$$

where G_0 is the prior density on $(\boldsymbol{\theta}_d, \mathcal{E}_d)$.

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

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

as in the finite mixture model. For each $\theta_{l,m}^{(d)} \in \boldsymbol{\theta}_d$ we use the same proposal density for the tree copula model and for the tree structure we apply the tree-angular proposal previously described.

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

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

The full conditional of $\boldsymbol{V}$ given $(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{I}, \mathcal{O}_T)$ is

$$\pi(\boldsymbol{V} | \mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{I}, \mathcal{O}_T) \propto \prod_{d: I_t = d} \Big(v_d \prod_{l<d} (1 - v_d) \Big) \pi(v_d)$$

where $\pi(v_d)$ is the prior density $\text{Beta}(1, \psi)$.

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

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

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

The full conditional of $\boldsymbol{U}$ given $(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \boldsymbol{V}, \boldsymbol{I}, \psi, \mathcal{O}_T)$ is

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

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

Finally, we can compute the full conditional of ψ given $(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \boldsymbol{I}, \mathcal{O}_T)$. 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 $(\boldsymbol{I}, \mathcal{O}_T)$ and is given by

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

where D is the number of cluster, $\pi(\psi)$ the Gamma prior on ψ and, $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$ is the Gamma function. In order to sample from this full conditional, we consider a MH step with proposal given by s $Gamma(a_\psi, b_\psi)$. 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* $\boldsymbol{I}$. The full conditional of $\boldsymbol{I}$ given $(\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \mathcal{O}_T)$ is

$$\pi(\boldsymbol{I}|\mathfrak{E}, \Theta, \boldsymbol{\nu}, \boldsymbol{A}, \psi, \boldsymbol{V}, \boldsymbol{U}, \mathcal{O}_T) \propto \prod_{t=1}^{T} \mathbb{1}(w_{I_t} > u_t) \prod_{(l,m) \in \mathcal{E}_{I_t}} c_{\theta_{l,m}^{(I_t)}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}) \right).$$

For each t, we sample $I_t = k$ from a discrete distribution with probability proportional to

$$\mathbb{1}(w_k > u_t) \prod_{(l,m) \in \mathcal{E}_k} c_{\theta_{l,m}^{(k)}} \left(F_{\nu_l}(\epsilon_{l,t}), F_{\nu_m}(\epsilon_{m,t}). \right)$$

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

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

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

• *Full conditional of* $\boldsymbol{A}$. We proceed as in the previous algorithm.

S2. Simulation Studies

In this section we present the results obtained using simulated data to investigate the performance of the methodologies proposed in Section 3.

We considered different alternative scenarios. For all scenarios the posterior estimates are consistent with parameters used to generate the data. Further experiments of this

kind have been performed, obtaining similar results. For each scenario we present two alternative prior settings, the values of the parameters used to generate the data, the posterior means for the parameters of interest and the MAP tree structure (for the TCM). For brevity, in the rest of the section, each tree is identified by its Prüfer code, see e.g. Skiena (1990). The Prüfer code is a classical way of coding a tree with a (unique) sequence of $n-2$ numbers, each being one of the numbers 1 through n.

Before describing in more details the alternative simulation scenarios, it is worth mentioning that, due to the label switching problem, attention should be paid to evaluate the posterior tree probabilities for the mix-TCM and for the DP-TCM. 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 by it. Indeed, if the function of interest is invariant to parameter permutations, label switching does not create any problem. In our case, instead of estimating the posterior probability of the tree appearing in each component, we 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 introduce the following permutation invariant function

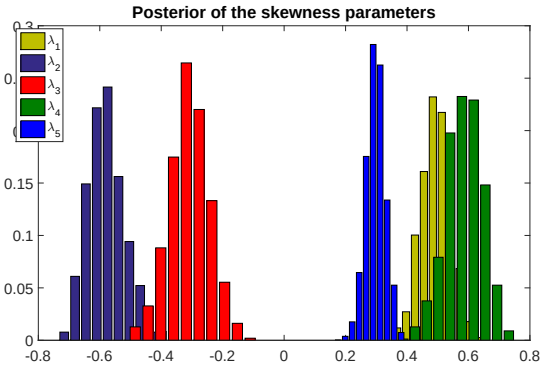
$$\Psi_{\mathcal{E}_0}(\mathfrak{E}, \boldsymbol{w}) = \sum_{d=1}^{D} w_d \mathbb{1}\{\mathcal{E}_0 = \mathcal{E}_d\}. \tag{S1}$$

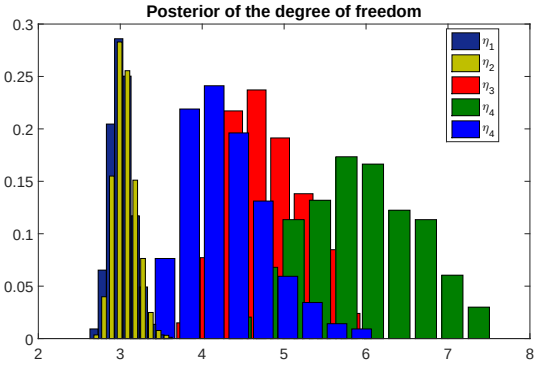
Equation (S1) defines the weight of the tree $\mathcal{E}_0$ in the mixture and its posterior mean is equal to the posterior probability that the tree $\mathcal{E}_0$ is contained in the mixture.

In Scenario 1 we simulated a dataset of $T = 300$ observations from a specific tree copula distributions with Double Gumbel or Double Clayton bivariate linking copulas and Skew-t margins with fixed parameters. We apply the Bayesian Tree Copula Models (TCM) to make inference on all the parameters involved. The results obtained with two alternative prior setting are summarised in Table S7. In this table we report the values of the parameters used to generate the data, the two prior settings, the MAP tree structure, the posterior mode of $\boldsymbol{\zeta}$ and the posterior means of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$.

In Figures S1 we display the histograms of the simulated values of the posterior distributions for $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ (first prior setting of Table S7).

Figure S1: Histograms of the simulated values of the posterior distributions for the parameters of Scenario 1 (first prior setting of Table S7), where we consider a tree copula model with bivariate Double Gumbel and Double Clayton copulas with Skew-t margins.





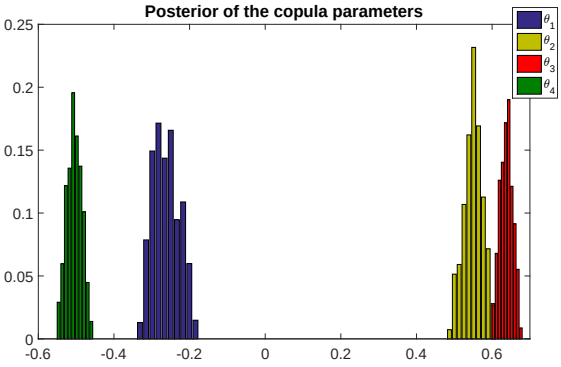


Table S7: Simulation study: Scenario 1. Tree copula model with bivariate Double Gumbel or Double Clayton copulas and Skew-t margins. For each prior settings, we report the posterior mean of the parameters $\boldsymbol{\nu}$, the posterior mode for $\boldsymbol{\zeta}$ and, the MAP tree structure. For the tree structure, the posterior probability is reported within brackets.

Scenario 1		First Prior Setting	Second Prior Setting
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$\text{Beta}_{(-1,1)}(1,1) \times Unif(\mathcal{H})$	$\text{Beta}_{(-1,1)}(1,3) \times Unif(\mathcal{H})$
	λ_k	$\text{Beta}_{(-1,1)}(1,1)$	$\text{Beta}_{(-1,1)}(1,2)$
	η_k	Exp(10)	Exp(5)
	True Values	Posterior Means/Modes	
$\nu_k = (\lambda_k, \eta_k)$	$(0.5\,, 3)$	(0.49, 3.04)	(0.47, 3.16)
	$(-0.6, 3)$	$(-0.58\,, 3.07)$	(−0.62, 2.86)
	$(-0.3, 5)$	$(-0.30\,, 4.55)$	(−0.35, 4.80)
	$(0.6, 6)$	(0.58, 6.46)	(0.64, 5.01)
	$(0.3, 4)$	$(0.29, 4.24)$	(0.22, 4.01)
$(\boldsymbol{\tau}, \boldsymbol{\zeta})$	$(-0.3, DG_1)$	$(-0.27\,, DG_1)$	$(-0.27, DG_1)$
	$(0.5, DC_1)$	$(0.55\,, DG_1)$	$(0.51, DC_1)$
	$(0.7, DG_1)$	$(0.67\,, DG_1)$	$(0.69, DG_1)$
	$(-0.6, DC_1)$	$(-0.56\,, DC_1)$	$(-0.57, DC_2)$
$\mathcal{E}$	$[1, 5, 5]^*$	$[1, 5, 5](0.928)$	$[1, 5, 5](0.895)$

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

In Scenario 2, to test the performance of the DP-TCM, we consider a dataset of $T = 300$ observations generated from a specific mixture of tree copula distributions with fixed number of components $D = 3$, Skew-t margins and $\boldsymbol{w} = [1/3, 1/3, 1/3]$.

The results of Scenario 2 are shown in Table S8, where we report the values of the parameters used to generate the data, the two alternative prior settings, the posterior means of the parameters $\boldsymbol{\nu}$, the posterior mode $\hat{D}$ of the number of clusters and the list of the higher posterior probability tree structures obtained via Equation (S1).

In Figures S2 we display the histograms of the simulated values of the posterior distributions for $\boldsymbol{\nu}$ (first prior setting of Table S8). Finally, the histograms of the simulated

Table S8: Simulation study: Scenario 2. DP-tree copula model with bivariate Double Gumbel or Double Clayton linking copulas and Skew-t margins. For each prior setting we show the posterior mean of the parameters $\boldsymbol{\nu}$ 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 (S1) is reported in brackets.

Scenario 2		First Prior Setting	Second Prior Setting
$N = 4$	$(\tau_{l,m}, \zeta_{l,m})$	$\text{Beta}_{(-1,1)}(1,1) \times Unif(\mathcal{H})$	$\text{Beta}_{(-1,1)}(2,2) \times Unif(\mathcal{H})$
$D = 3$	λ_k	$\text{Beta}_{(-1,1)}(1,1)$	$\text{Beta}_{(-1,1)}(1,3)$
	η_k	$\text{Exp}(10)$	$\text{Exp}(5)$
	ψ	$\text{Gamma}(0.2,1)$	$\text{Gamma}(0.25,2)$
	True Values	Posterior Means/Modes	
$\nu_k = (\lambda_k, \eta_k)$	$(0.2, 6)$	$(0.19, 6.0)$	$(0.17, 6.2)$
	$(-0.2, 4)$	$(-0.20, 3.9)$	$(-0.23,\ 3.8)$
	$(-0.4, 5)$	$(-0.43, 4.7)$	$(-0.33,\ 4.6)$
	$(0.8, 7)$	$(0.80, 6.9)$	$(0.78,\ 6.8)$
D	3	3	3
$\mathcal{E}$	$[1,1]$	$\mathbf{[4,3](0.298)}$	$\mathbf{[1,1](0.354)}$
	$[3,2]$	$\mathbf{[3,2](0.236)}$	$\mathbf{[4,3](0.334)}$
	$[4,3]$	$\mathbf{[1,1](0.129)}$	$\mathbf{[3,2](0.110)}$

values of the posterior distributions of the number of clusters are presented in Figures S3 for both prior setting. The histograms are centred on the true value of D, i.e. $D = 3$.

In Scenario 3 we simulated a dataset of $T = 300$ observations from a multivariate AR model with fixed parameters, a specific tree structure with Double Gumbel or Double Clayton bivariate copulas and Normal margins with fixed precision. We apply the fully Bayesian AR-TCM with Double Gumbel/Double Clayton as linking copulas and Normal margins. The results obtained with two alternative prior setting are summarised in Table S9. We report the values of the parameters used to generate the data, the two prior settings, the MAP tree structure, the posterior mode of $\boldsymbol{\zeta}$ and the posterior means of

Figure S2: Histograms of the simulated values of the posterior distributions for the parameters of Scenario 2 (first prior setting of Table S8), where we consider a DP-tree copula model with bivariate Double Gumbel and Double Clayton copulas with Skew-t margins.

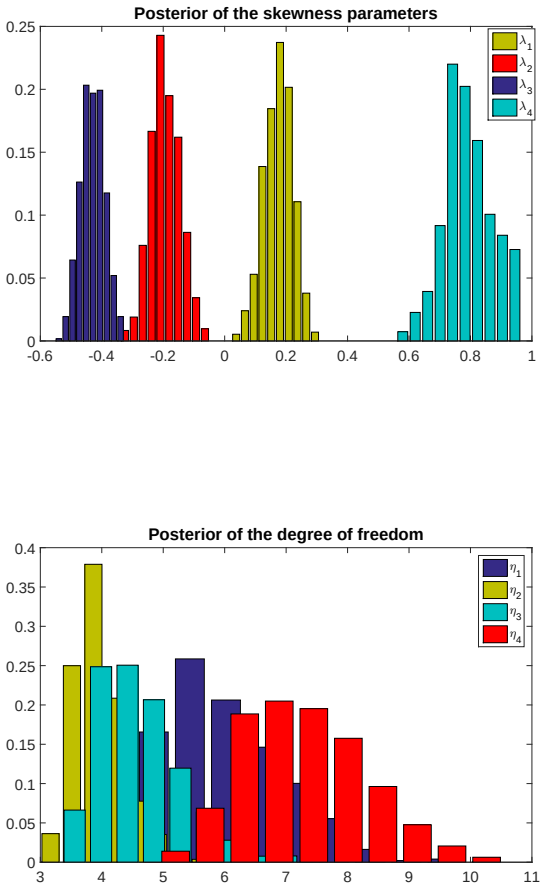
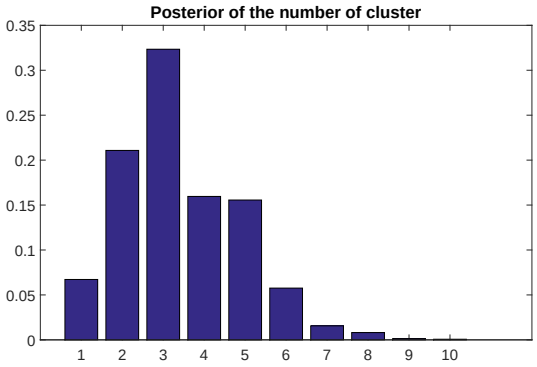
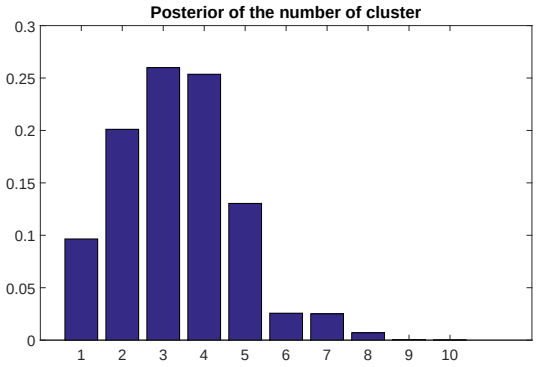


Figure S3: Histograms of the simulated values of the posterior distributions of number of components for Scenario 2, where we study a DP-tree copula model with Double Gumbel/Double Clayton linking copulas. On the left, the result with the first prior setting of Table S8, on the right with the second prior setting. The data are generated from a mixture with $D = 3$ components.





$(\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{A})$.

Table S9: Simulation study: Scenario 3. AR-Tree copula model with bivariate Double Gumbel or Double Clayton copulas and Normal margins. For each prior settings, we report the posterior mean of the parameters $\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{A}$, the posterior mode for $\boldsymbol{\zeta}$ and, the MAP tree structure. For the tree structure, the posterior probability is reported within brackets.

Scenario 3		First Prior Setting	Second Prior Setting
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$\text{Beta}_{(-1,1)}(1,1) \times Unif(\mathcal{H})$	$\text{Beta}_{(-1,1)}(1,1) \times Unif(\mathcal{H})$
$p = 2$	p_k	$\text{Gamma}(0.1, 0.1)$	$\text{Gamma}(0.5, 0.1)$
	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.7, 0.1], 100\boldsymbol{I}_p)$	$\mathcal{N}_p([0.3, 0.4], 10\boldsymbol{I}_p)$
	True Values	Posterior Means/Modes	
$\frac{1}{p_k}$	0.5	0.49	0.54
	0.3	0.32	0.33
	0.4	0.43	0.39
	0.6	0.63	0.59
	0.2	0.22	0.22
$\boldsymbol{\alpha}_k$	$(0.3, 0.5)$	$(0.25, 0.55)$	$(0.32,\ 0.48)$
	$(-0.3, 0.4)$	$(-0.33, 0.39)$	$(-0.26,\ 0.39)$
	$(0.1, 0.6)$	$(0.04, 0.69)$	$(0.14,\ 0.49)$
	$(-0.2, 0.4)$	$(-0.18, 0.41)$	$(-0.17,\ 0.35)$
	$(0.1, 0.5)$	$(0.08, 0.52)$	$(0.12,\ 0.46)$
$(\boldsymbol{\tau}, \boldsymbol{\zeta})$	$(0.7, DC_1)$	$(0.54, DG_2)$	$(0.69, DC_1)$
	$(-0.2, DG_1)$	$(-0.23, DG_1)$	$(-0.25, DG_1)$
	$(0.3, DG_2)$	$(0.29, DG_2)$	$(0.37, DG_2)$
	$(-0.6, DG_1)$	$(-0.59, DG_1)$	$(-0.61, DG_2)$
$\mathcal{E}$	$[1, 2, 5]^*$	$[1, 2, 5](0.895)$	$[1, 2, 5](0.892)$

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

We conclude this Section testing the AR-DP-TCM. We simulate datasets 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 and Normal

margins. As is often the case with Dirichlet Process mixture models, 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) and also the case in which the parameter ψ is fixed.

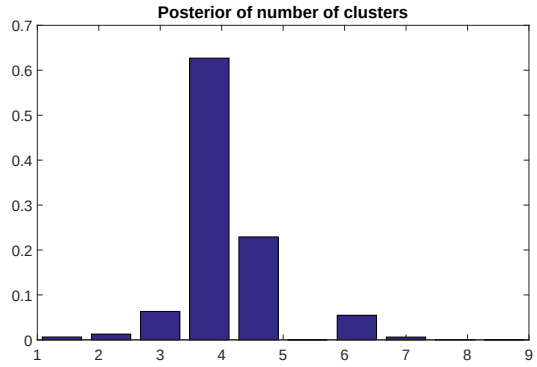
In Scenario 4 we consider a dataset simulated from a mixture of two tree copula distributions with bivariate Double Gumbel/Double Clayton copulas, weights $\boldsymbol{w} = [0.5, 0.5]$ and Normal margins. In Scenario 5, we used simulated data from a mixture of four tree copulas with linking Double copulas and weights $\boldsymbol{w} = [0.25, 0.25, 0.25, 0.25]$ and Normal margins. In both cases, we apply the fully Bayesian DP-TCM. In Scenario 4 we assume a gamma prior for the parameter ψ, while in Scenario 6 we do not make inference on ψ, i.e. we fix a specific value for the concentration parameter.

The results of Scenario 4 are shown in Table S10. 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 $(\boldsymbol{\nu}, \boldsymbol{A})$, the posterior mode $\hat{D}$ of the number of clusters and the list of the higher posterior probability tree structures obtained via Equation (S1).

The values of the parameters used to generate the data in Scenario 5 and the corresponding posterior estimates computed with two different prior settings are reported in Table S11. The prior settings present different value for the concentration parameter, respectively $\psi = 0.1$ and $\psi = 0.5$. The posterior distributions of the number of clusters obtained with each prior setting of Scenario 3 are shown in Figure S4. Also in this case, both the distributions are centred around the true value of D. As expected, the distribution obtained with the first prior setting is more concentrated than the distribution obtained with the second prior setting that, indeed, shows a higher values of ψ.

Finally, in Table S12 we compare the values of the DIC_3 for scenarios 3,4,5 computed with the fully Bayesian approach and with the B-IFM procedure. The models estimated by the fully Bayesian approach have a slighter lower DIC_3 with respect to the corresponding models estimated by B-IFM procedure.

Figure S4: Histograms of the simulated values of the posterior distributions of number of components for Scenario 5. AR-DP-TCM with Double Gumbel and Double Clayton copulas and with fixed value of ψ. On the left, the result with the first prior setting of Table S11 ($\psi = 0.1$) and on the right with the second prior setting ($\psi = 0.5$).



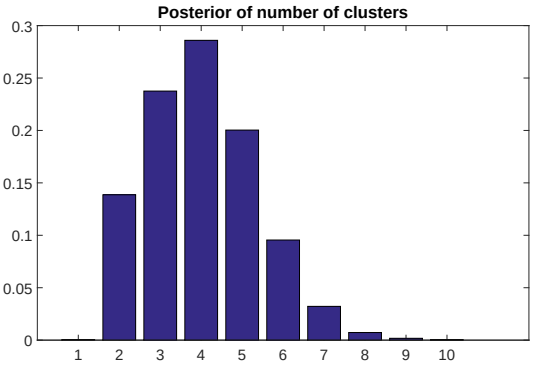


Table S10: Simulation study: Scenario 4. AR-DP-TCM with bivariate Double Gumbel or Double Clayton linking copulas and Normal margins. For each prior setting we show the posterior mean of the parameters $(\boldsymbol{\nu}, \boldsymbol{A})$ and the posterior mode $\hat{D}$ of the number of clusters. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (S1) is reported in brackets.

Scenario 4		First Prior Setting	Second Prior Setting
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$Beta_{(-1,1)}(1,2) \times Unif(\mathcal{H})$	$Beta_{(-1,2)}(1,1) \times Unif(\mathcal{H})$
$p = 2$	p_k	Gamma$(1, 0.01)$	Gamma$(1, 0.01)$
$D = 2$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.1, 0.3], 10\boldsymbol{I}_p)$	$\mathcal{N}_p([0.2, 0.3], 10\boldsymbol{I}_p)$
	ψ	Gamma$(0.2, 1)$	Gamma$(0.08, 25)$
	True Values	Posterior Means/Modes	
$\frac{1}{p_k}$	0.5	0.56	0.53
	0.3	0.36	0.30
	0.4	0.46	0.42
	0.7	0.71	0.71
	0.9	0.98	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
$\mathcal{E}$	$[1, 2, 5]$	$\mathbf{[1, 2, 5](0.451)}$	$\mathbf{[1, 4, 5](0.4745)}$
	$[1, 4, 5]$	$\mathbf{[1, 4, 5](0.190)}$	$\mathbf{[1, 2, 5](0.1711)}$

S3. Additional real data results

This section presents additional details of the application of our AR-GARCH tree copula models to the real data application discussed in Section 4.

Table S11: Simulation study: Scenario 5. AR-DP-TCM with Double Gumbel and Double Clayton copulas and with fixed value of ψ. For each prior setting we show the posterior mean of the parameters $(\boldsymbol{\nu}, \boldsymbol{A})$ and the posterior mode $\hat{D}$ of the number of clusters. In last part of the table, we list the highest posterior probability tree structures. For each tree structure, the posterior mean of (S1) is reported within brackets.

Scenario 5		First Prior Setting	Second Prior Setting
$N = 5$	$(\tau_{l,m}, \zeta_{l,m})$	$Beta_{(-1,1)}(2,2) \times Unif(\mathcal{H})$	$Beta_{(-1,1)}(1,2) \times Unif(\mathcal{H})$
$p = 2$	p_k	Gamma$(1, 0.01)$	Gamma$(1, 0.01)$
$D = 4$	$\boldsymbol{\alpha}_k$	$\mathcal{N}_p([0.1, 0.3], 10\boldsymbol{I}_p)$	$\mathcal{N}_p([0.1, 0.3], 10\boldsymbol{I}_p)$
	ψ	$\psi = 0.1$	$\psi = 0.5$
	True Values	Posterior Means/Modes	
$\frac{1}{p_k}$	0.5	0.55	, 0.57
	0.3	0.32	0.34
	0.4	0.44	0.47
	0.7	0.74	0.68
	0.9	0.93	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
$\mathcal{E}$	$[1, 2, 5]$	$\mathbf{[2, 3, 5](0.292)}$	$\mathbf{[2, 3, 5](0.193)}$
	$[1, 4, 5]$	$[3, 5, 4](0.080)$	$[1, 4, 5](0.108)$
	$[2, 1, 5]$	$[5, 1, 4](0.078)$	$\mathbf{[1, 2, 5](0.080)}$
	$[2, 3, 5]$	$[1, 2, 3](0.055)$	$[1, 1, 3](0.057)$
		$[1, 5, 1](0.041)$	$[5, 1, 5](0.049)$
		$\mathbf{[1, 4, 5](0.036)}$	$[2, 2, 5](0.045)$
		$[3, 2, 5](0.0298)$	$\mathbf{[2, 1, 5](0.034)}$
		$\mathbf{[2, 1, 5](0.293)}$	$[3, 5, 5](0.030)$
		$\mathbf{[1, 2, 5](0.278)}$	$\mathbf{[1, 4, 5](0.274)}$

Figure S5: Histograms of the posterior distributions of number of clusters with the DP-tree model for Italian (on the left) and German (on the right) markets applied to the residuals of the AR(3)-GARCH(1,1) models.

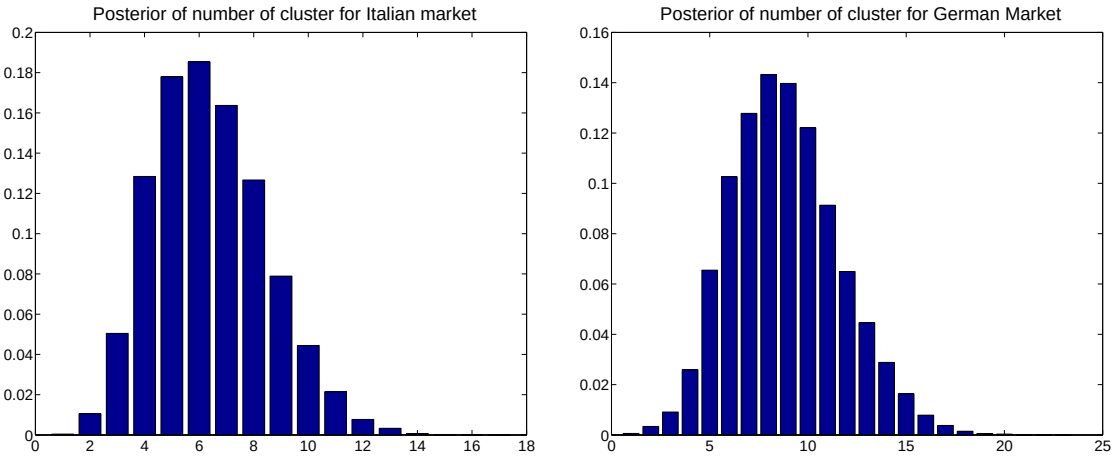


Figure S6: Histograms of the posterior distributions of number of clusters with the DP-tree model for Italian (on the left) and German (on the right) markets applied to the residuals of the AR(3) models.

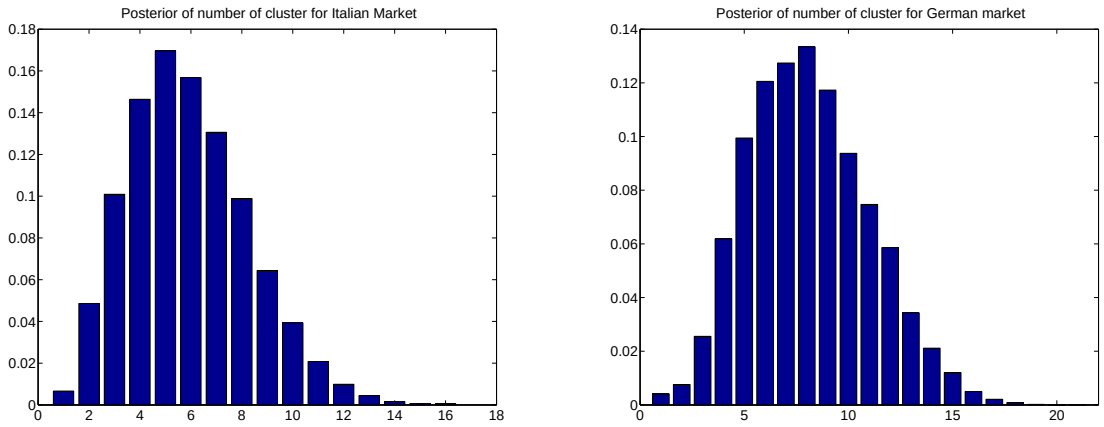


Table S12: Simulated Data: fully Bayesian vs BIF, DIC_3 Table.

True Model	**Fully Bayesian**	**B-IFM**
Scenario 3(TCM)	**528**	579
Scenario 4(DP-TCM)	**1208**	1446
Scenario 5(DP-TCM, fixed ψ)	**1341**	1465

Table S13: MAEs between predictive portfolio and historical one for the "realistic" portfolio described in Section 4.4 and an equally weighted portfolio. In sample results on the last 200 observations.

	"Realistic" Portfolio				**Equally Weighted Portfolio**			
	Italy		**Germany**		**Italy**		**Germany**	
Margins	Norm	Skew-t	Norm	Skew-t	Norm	ST	Norm	Skew-t
AR(3) CTM	0.2360	0.2450	0.2421	0.2439	0.1474	0.1494	0.1498	0.1512
AR(3) DP-TCM	0.2345	0.2440	0.2431	0.2411	0.1468	0.1477	0.1493	0.1511
AR(3)-GARCH(1,1) CTM	0.2446	0.2492	0.2471	0.2435	0.1478	0.1470	0.1501	0.1517
AR(3)-GARCH(1,1) DP-TCM	0.2443	0.2490	0.2450	0.2415	0.1459	0.1473	0.1500	0.1509