ARTICLE IN PRESS
JID: EOR [m5G;March 14, 2018;17:13]
European Journal of Operational Research 000 (2018) 1–15
Contents lists available at ScienceDirect
European Journal of Operational Research
journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

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

ARTICLE INFO

Article history:
Received 17 May 2016
Accepted 16 February 2018
Available online xxx

Keywords:
Multivariate analysis
Bayesian analysis
Copula model
Energy market
Markov Chain Monte Carlo

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.

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 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. Aas, Czado, Frigessi, and Bakken

* Corresponding author.
E-mail addresses: federico.bassetti@unipv.it (F. Bassetti), elena.degiuli@unipv.it (M.E. De Giuli), enrica.nicolino@rgigroup.com (E. Nicolino), claudia.tarantola@unipv.it (C. Tarantola).

https://doi.org/10.1016/j.ejor.2018.02.037
0377-2217/

(2009), Czado, Gärtner, and Min (2011), Arreola Hernandez (2014), Dalla Valle, De Giuli, Tarantola, and Manelli (2016), Jaschke (2014), Laih (2014), Marimoutou and Soury (2015), Wen, Wei, and Huang (2012), Wu, Chung, and Chang (2012) 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 Bedford 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, Gärtner, and Min (2011) . Recently, Gruber and Czado (2014, 2015) 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 $k = 1,\dots,N$ a time series $(X_{k,\,t})$ that we model via an $AR(p) - GARCH(q,r)$ structure. More precisely,

$$X_{k,t} = \sum_{i=1}^{p} a_{k,i} X_{k,t-i} + \varepsilon_{k,t},$$
$$\varepsilon_{k,t} = \sigma_{k,t} Z_{k,t},$$
$$\sigma_{k,t}^2 = \sigma_k^2 + \sum_{i=1}^{q} b_{k,i}\sigma_{k,t-i}^2 + \sum_{j=1}^{r} c_{k,j}\varepsilon_{k,t-j}^2. \quad (1)$$

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 $\boldsymbol{Z_t} = (Z_{1,t},\dots,Z_{N,t})$ for $t = 1,\dots,T$ 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 jointly 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 2.1–2.3 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 $F_1,\dots,F_N$, can be written as $F(z_1,\dots,z_N) = C\big(F_1(z_1),\dots,F_N(z_N)\big)$, where C is a suitable copula function. Consequently, if F is absolutely continuous, the corresponding joint probability density function (pdf) is given by

$$f(z_1,\dots,z_N) = c\big(F_1(z_1),\dots,F_N(z_N)\big) f_1(z_1)\dots f_N(z_N),$$

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 $\mathcal{V} = \{1,\dots,N\}$ 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

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