

# Dependence Modeling in Energy Markets using Sibuya-type Copulas

Nikolai Kolev<sup>1</sup> & Jayme Pinto<sup>1</sup>

<sup>1</sup> Department of Statistics, University of São Paulo, Brazil

Correspondence: Nikolai Kolev, Department of Statistics, University of São Paulo. Rua do Matão, 1010 - ZIP CODE 05508-090 - São Paulo (SP), Brazil. E-mail: kolev.ime@gmail.com

Received: January 13, 2017 Accepted: March 10, 2017 Online Published: March 27, 2017

doi:10.5539/ijsp.v6n3p43 URL: <https://doi.org/10.5539/ijsp.v6n3p43>

## Abstract

The dependence structure between 756 prices for futures on crude oil and natural gas traded on NYMEX is analyzed using a combination of novel time-series and copula tools. We model the log-returns on each commodity individually by Generalized Autoregressive Score models and account for dependence between them by fitting various copulas to corresponding error terms. Our basic assumption is that the dependence structure may vary over time, but the ratio between the joint distribution of error terms and the product of marginal distributions (e.g., Sibuya's dependence function) remains the same, being time-invariant. By performing conventional goodness-of-fit tests, we select the best copula, being member of the currently introduced class of Sibuya-type copulas.

**Keywords:** copulas, dependence function, energy markets, invariance, time-series

## 1. Energy Markets and their Modeling

The evidence of joint movements in energy markets prices can be possibly explained by the simultaneous impact of increasing economic and political activities in different geographic regions. Traditionally, in a first step of statistical analysis, several single equation time-series models are applied to account for autocorrelation and volatility clustering in marginals. The model variability is huge and depend on many circumstances.

Usually, the return series are modeled by state-space time series (in most of the cases GARCH-type) with suitable marginal distributions for error-terms or by factor models. Following the standard theory, the error terms are assumed to be independent and identically distributed. The empirical analysis shows that this theoretical belief is frequently violated in practice, especially during pre- and post-crisis periods. For example, Alexander (2005) studied dependence between prices for futures on crude oil and natural gas. She concluded that when prices of these commodities are fitted by time-series models, the distribution for the error terms are found to be asymmetric, i.e., the dependence can not be modeled correctly by the bivariate Normal distribution.

In the univariate case non-normality is associated to the skewness and leptokurtosis phenomena, or the fat-tail problem. In the multivariate case, the fat-tail problem can be assigned not only to the marginal distributions but also to the probability of large co-movements of the individual returns. In other words, the tail dependence phenomena is a modern statistical standard to describe the amount of extreme dependence, see Jaschke (2014).

Recently, in a second step, professionals select an appropriate copula family and fit it to filtered log-return series in order to gauge their dynamic interdependence. Since the copula is a function of marginal distributions (applied to error terms), an adequate modeling of individual time series is crucial for estimating the dependence structure between underlying log-returns. For example, the copula approach for modeling energy markets was addressed in Gregoire et al. (2008) who analyzed prices of crude oil and natural gas based on one-month-ahead futures contracts traded on the New York Mercantile Exchange (NYMEX) from July 1, 2003 to July 19, 2006. The authors modeled the log-returns individually as time series, selecting appropriate GARCH-type model for marginal series, and selected the "best" copula to depict the existing dependence between error terms.

As a standard, the researches apply the following copulas in their works: Gaussian copula (without tail dependence), Frank and Student's  $t$ -copulas (exhibiting both upper and lower tail dependence), Gumbel and Pareto copulas (having upper tail dependence), Clayton (displaying lower tail dependence), BB1 copula family and few other modifications, see Joe (2015).

An implicit assumption in many studies is that the dependence structure between the prices of crude oil and natural gas is constant (static) in time, i.e., the corresponding copula  $C(u, v)$  joining marginal log-returns should be time-invariant, meaning that

$$C_t(u, v) = C(u, v) \quad \text{for } (u, v) \in [0, 1] \quad \text{and all } t = 1, 2, \dots, T.$$

It is well known that the absolutely continuous copula satisfying the last relation is the Clayton's one, see Oakes (2005). It exhibits lower tail dependence (but not upper one). Thus, the Clayton copula can not be used to describe the extreme dependence between markets, for example. Nevertheless (and curiously) there can be found "contributions" whose authors indicate the Clayton copula as the "best choice" for modeling extreme market movements.

Another common misunderstanding can be found in many papers and business reports indicating as a "best" fit Gaussian and/or  $t$ -copulas. The Gaussian and  $t$ -copulas are popular in applications due to their convenience for computation (being members of the elliptical copula family). The Gaussian copula has no tail dependence, while the  $t$ -copula allows different degrees of symmetric upper or lower tail dependence which is determined by the estimated degrees of freedom ( $df$ ) parameter. A smaller  $df$  implies higher tail dependence. As the  $df$  goes to infinity, the Student's  $t$ -copula converges to the Gaussian copula. The problem is that the static Gaussian and  $t$ -copulas are not time invariant and thus, they are not appropriate to describe the existing dynamic dependence between individual log-returns.

If one googles "Dependence modeling in energy markets with copulas", around 120 related works after 2010 will be listed. Many authors conclude that the overall relationship has increased over time, implying growing tail dependency. Owing to these features, the usual static dependence measures (lower and upper tail indexes) should be considered inadequate after 2008 crisis, indicating a preference to local their versions. We will list and comment several recent contributions:

- Reboredo (2011) used copula based GARCH model to study the dependence structure between the crude oil benchmark prices in 12 international crude oil markets. It was found that in times of market stress, crude oil prices in each market tend to exhibit comovements with the same intensity. Moreover, crude oil and natural gas log-returns exhibit non-linear dependence;
- Alouia et al. (2013) analyzed daily closing prices for the Brent crude oil index and MSCI (Morgan Stanley Capital International) stock market indexes of the six Central East European transition economies (Bulgaria, Czech Republic, Hungary, Poland, Romania and Slovenia) over the period from December 1, 2005 to August 20, 2012, totalizing 1753 observations. The authors investigated the dependence structure between oil and stock market returns through time applying the time-varying copula approach, see, e.g., Cherubini et al. (2012). The conclusion is that bias toward evidence of tail dependence is generated when time-varying parameters in the dependence distribution is not allowed;
- Jaschke et al. (2012) studied the dependence of extreme events in energy markets. In a deep discussion the authors established that the formal consideration of tail dependence indexes may falsely lead to evidence of asymmetric relation between the returns and suggested to use local version based on tail-copula approach, see Schmidt and Stadtmüller (2006). The empirical investigation focused on the dependence between crude oil and natural gas prices based on one-month-ahead futures contracts traded at NYMEX between July 2, 2007 and July 2, 2010;
- Beckmann et al. (2016) analyzed oil prices and exchange rates against the dollar and both experienced long swings over recent decades. The authors focused on the evolution of the relationship between oil prices and dollar exchange rates of 12 oil-exporting and oil-importing countries based on a copula approach. Daily data for two 5 year periods between September 2003 and 2013 were used, taking the crash of Lehman Brothers during 2008 as the dividing point. The growing tail dependence shows that extreme events are more likely to occur simultaneously for both series;
- Pan (2014) investigated the tail dependence structure between energy market and stock markets returns in the BRIC (Brazil, Russia, India and China) countries over the period from 12 January 2000 to 28 December 2012. The data set included the Bovespa index (Brazil), the RTSI index (Russia), the BSE 30 Sensitivity index (India) and the Shanghai Composite index (China) returns. Pan (2014) showed that the tail dependence increased rather substantially in the financial crisis of 2008. Moreover, the lower tail dependence for all the paired returns is larger than the upper one. Finally, the tail dependence is found to be the strongest for Russia and the weakest for China;
- The vine copula model is another powerful tool to analyze the dependence structure in a multivariate setting, consult Kurowicka and Joe (2011). It allows one to define the structure of relationship between the variables by using expert knowledge, concordance of data, or both. Moreover, it can describe the association between the variables through graphical model, or through what is called pair-copulas. We would mention the applied paper by Kiatmanarochy and Sriboonchittaz (2014) who used the GARCH model and the D-vine copula model to analyze the relationship between the crude oil prices of three different continents: Light crude futures of the NYMEX for North America, Brent crude futures of the Intercontinental Exchange (ICE) for Europe, and Oman crude futures of the Dubai Mercantile Exchange (DME) for Asia. The daily closing prices during the period from 26 December 2008 to 28 June 2013 were used to conduct the analysis.

The chance to depict the existing dynamic dependence between commodities in different geographic regions would increase if one starts with a general model as a base. Our choice is the so-called Generalized Autoregressive Score (GAS) models (also known as Dynamic Conditional Score models), recently introduced by Creal et al. (2013) which incorporate many known time-series models as particular cases. We outline main facts about GAS models in Section 2.1.

As we have seen, the time invariance is not a common copula property and depends on the additional appropriate utility criteria applied by the investigator. For example, all copulas corresponding to distributions possessing bivariate lack of memory property are "time-invariant" in the non-aging world (when the distribution of a random vector coincides with the distribution of its residual lifetime vector), see Kulkarni (2006).

As an alternative and complement to the existing methods, we suggest to use Sibuya-type copulas which incorporate many known copula families. Sibuya-class of copulas generate and represent multivariate distributions preserving the Sibuya's dependence function (being the ratio of joint distribution to the product of marginal distributions, e.g., Sibuya (1960)) with respect to the residual lifetime vector. In fact, the dependence structure may vary over time, but Sibuya's dependence function remains in equilibrium, which is a reasonable utility criteria. The class of Sibuya-type copulas is introduced by Pinto and Kolev (2015) and a short description is presented in Section 2.2.

In Section 3 we apply GAS methodology and Sibuya-type copulas to analyze energy markets log-returns for the data used by Gregoire et al. (2008). It happens that the best fit is achieved with the copula of the extended Gumbels's law (see Example 3) connecting the corresponding error terms. We finish the article with a short discussion.

## 2. Methodology

We will introduce briefly two basic tools to be used in our proposal: Generalized Autoregressive Score (GAS) models and Sibuya-class of copulas.

### 2.1 GAS Models

In many settings of empirical interest, time variation in a selection of parameters of a model is important for capturing the dynamic behavior. Creal et al. (2013) introduce a new, general framework for building observation driven approach, where parameters are defined as a function of lagged dependent values, exogenous variables, and past observations. Consult for recent developments <http://gasmodel.com>

Let  $R_t$  denote the dependent variable of interest (log returns, say),  $f_t$  the time-varying parameter vector,  $x_t$  a vector of exogenous variables (covariates), all at time  $t = 1, \dots, n$ . Define  $\mathbf{R}^t = \{R_1, \dots, R_t\}$ ,  $\mathbf{F}^t = \{f_1, \dots, f_t\}$  and  $\mathbf{X}^t = \{x_1, \dots, x_t\}$ . The available information at time  $t$  consists of past observations  $\mathbf{R}^{t-1}$ , the time-varying parameters in  $\mathbf{F}^{t-1}$  and  $\mathbf{X}^t$ . It is assumed that  $R_t$  is generated by the observation density  $p(\mathbf{R}^{t-1}, \mathbf{F}^{t-1}, \mathbf{X}^t; \theta)$ , where  $\theta$  is a vector of static parameters.

The updating approach for the time-varying parameter  $f_t$  is given by the Generalized Autoregressive Score model with orders  $m$  and  $q$ , to be abbreviated  $GAS(m, q)$ , as follows

$$f_t = \omega(\theta) + \sum_{i=0}^{m-1} A_i(\theta) s_{t-i} + \sum_{j=1}^q B_j(\theta) f_{t-j}, \tag{1}$$

where  $\omega(\theta)$  is a vector of unknown parameters. Coefficient matrices  $A_i(\theta)$  and  $B_j(\theta)$  in (1) have appropriate dimensions, while  $s_t$  is a scaled (link) function, which may be treated as "innovation" depending on set  $\{\mathbf{R}^{t-1}, \mathbf{F}^{t-1}; \theta\}$ .

The score vector  $s_t$  in (1) is specified by

$$s_t = S(t, \mathbf{R}^{t-1}, \mathbf{F}^{t-1}; \theta) \nabla_t = S_{t-1} \nabla_t, \tag{2}$$

where  $S_{t-1}$  is the time dependent scaling matrix (approximated by the inverse of Fisher information matrix usually) and

$$\nabla_t = \frac{\partial p(\mathbf{R}^{t-1}, \mathbf{F}^{t-1}, \mathbf{X}^t; \theta)}{\partial f_{t-1}}. \tag{3}$$

The score terms  $s_t$  in (1) can be interpreted as Gauss-Newton updating step for every new observation  $R_t$  that becomes available through time. In addition, the parameter  $f_t$  is amended in the direction of steepest increase of the log-density at time  $t$ .

The main advantage of score  $s_t$  defined by (2) and (3) is the possibility of its particular driving mechanism selection. There are several intuitive choices for the scaling matrix  $S_{t-1}$ . For example,

- When  $S_{t-1}$  is the identity matrix, i.e.,  $S_{t-1} = I$ , the recursion (1) captures models such as the autoregressive conditional multinomial model;

- Another natural choice is to set  $S_{t-1}$  equal to the (pseudo)-inverse information matrix based on  $p(\mathbf{R}^{t-1}, \mathbf{F}^{t-1}, \mathbf{X}^t; \theta)$ , e.g.,

$$S_{t-1} = \mathbf{I}_{t-1}^{-1} = E_{t-1}[\nabla_t \nabla_t']^{-1}.$$

We mention two basic properties of  $GAS(m, q)$  of bivariate continuous distributions only:

- The expectation of the score is zero, i.e.,  $E_{t-1}[\nabla_t] = 0$ . As a result,  $s_t$  is a martingale difference sequence;
- If  $f_t$  is stationary, its unconditional expectation is  $E[f_t] = \omega(I - B_1(1))^{-1}$ .

Specific choices for scaling the score vector  $s_t$  transform the GAS model into well-known observation driven models, see complete list in Creal et al. (2013). Two examples are given below.

**Example 1 (GARCH model).** Consider the basic model

$$R_t = \sigma_{t-1} \varepsilon_t, \quad \text{with disturbance } \varepsilon_t \sim N(0, 1),$$

where  $\sigma_{t-1}$  is a time-varying standard deviation. One can check that the GAS(1,1) model with  $S_{t-1} = \mathbf{I}_{t-1}^{-1}$  for  $f_t = \sigma_t^2$  reduces to

$$f_t = \omega + A_0(R_t^2 - f_{t-1}) + B_1 f_{t-1},$$

being the standard GARCH(1,1) model.

**Example 2 (Linear Gaussian state-space model).** Let  $R_t$  be generated by

$$R_t | f_t \sim p(R_t | f_t, \theta), \quad \text{with } f_t = \Lambda(\alpha_t), \quad t = 1, \dots, n,$$

where  $f_t$  is the time-varying parameter vector,  $\Lambda(\cdot)$  is the link function, and the scalar  $\alpha_t$  has a linear dynamic specification

$$\alpha_t = \delta + \gamma \alpha_{t-1} + \eta_{t-1}, \quad \text{where } \alpha_1 \sim N(a_1, P_1), \quad \eta_{t-1} \sim N(0, \sigma_\eta^2).$$

The static parameter vector  $\theta$  incorporates additional fixed and unknown coefficients. In the above state-space model  $\delta$  is a constant and  $\gamma$  is the autoregressive coefficient.

Alternative frameworks for observation driven models within the exponential family of distributions have been suggested in the literature, including the generalized linear autoregressive models, the generalized autoregressive moving average models and the vector multiplicative error models. In contrast to these proposals, GAS models are able to exploit the complete density structure specified by  $p(\mathbf{R}^{t-1}, \mathbf{F}^{t-1}, \mathbf{X}^t; \theta)$ , rather than only means and higher moments.

Thus, GAS models can postulate different dynamics for volatilities from fat-tailed distributions, depending on the underlying observation density  $p(\mathbf{R}^{t-1}, \mathbf{F}^{t-1}, \mathbf{X}^t; \theta)$ . Extensions of GAS model to asymmetric, long memory, and other more complex dynamics can be considered as well. Therefore, GAS models allow the formulation of a wide range of new models and gives rise to a number of useful observation driven models that have not been investigated before.

### 2.2 Sibuya-type Copulas

Let  $S_{X_1, X_2}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$  be the joint survival function of non-negative continuous random vector  $(X_1, X_2)$ . Define the hazard rate components by  $r_i(x_1, x_2) = -\frac{\partial}{\partial x_i} [\ln S_{X_1, X_2}(x_1, x_2)]$ ,  $i = 1, 2$ .

The Sibuya-type copulas represent a class  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  of bivariate continuous distributions specified by relation

$$r_1(x_1, x_2) + r_2(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2, \quad (x_1, x_2) \in [0, \infty)^2,$$

where  $a_0, a_1, a_2 \geq 0$ , see Pinto and Kolev (2016).

The class  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  can be characterized as follows.

**Theorem 1 (Pinto and Kolev, 2016).** *If the first partial derivatives of  $S_{X_1, X_2}(x_1, x_2)$  exist then  $r_1(x_1, x_2) + r_2(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2$  is fulfilled if and only if the corresponding joint survival function can be represented by*

$$S_{X_1, X_2}(x_1, x_2) = \begin{cases} S_{X_1}(x_1 - x_2) \exp \left\{ -a_0 x_2 - a_1 x_1 x_2 - \frac{a_2 - a_1}{2} x_2^2 \right\}, & \text{if } x_1 \geq x_2, \\ S_{X_2}(x_2 - x_1) \exp \left\{ -a_0 x_1 - a_2 x_1 x_2 - \frac{a_1 - a_2}{2} x_1^2 \right\}, & \text{if } x_2 \geq x_1, \end{cases} \quad (4)$$

where  $S_{X_i}(x_i)$  are the marginal survival functions,  $i = 1, 2$ .

The joint survival function  $S_{X_1, X_2}(x_1, x_2)$  given by (4) defines a new aging notion. It is valid only for certain marginal distributions satisfying  $\frac{\partial^2}{\partial x_1 \partial x_2} S_{X_1, X_2}(x_1, x_2) \geq 0$ . This inequality implies restrictions for the marginal densities  $f_{X_i}(x_i)$ , which determine the parameter space  $\mathcal{A}$  of the  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  class as follows

$$\mathcal{A} = \{a_0 \in [\max(f_{X_1}(0), f_{X_2}(0)), f_{X_1}(0) + f_{X_2}(0)]; a_1, a_2 \geq 0, a_1 + a_2 > 0\}.$$

If  $a_0 = f_{X_1}(0) + f_{X_2}(0)$  the bivariate distribution is absolutely continuous and if  $a_0 < f_{X_1}(0) + f_{X_2}(0)$  the corresponding distribution has a singular component along the line  $x_1 = x_2 = x \geq 0$ .

Particular cases of the class  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  are:

- All distributions possessing bivariate lack of memory property (in particular, Marshall-Olkin’s bivariate exponential distribution, being positive quadrant dependent), e.g., Kulkarni (2006): when  $a_1 = a_2 = 0$ ;
- Gumbel’s (1960) bivariate exponential type I distribution

$$S_{X_1, X_2}(x_1, x_2) = \exp\{-x_1 - x_2 - \theta x_1 x_2\}, \quad \theta \in [0, 1], \tag{5}$$

exhibiting negative quadrant dependence: when  $a_0 = 2$  and  $a_1 = a_2 = \theta$ . The survival copula corresponding to (5) is given by

$$\bar{C}_G(u, v) = uv \exp\{-\theta \ln u \ln v\}, \quad u, v \in [0, 1]; \tag{6}$$

- Some members of the Generalized Marshall-Olkin distributions introduced by Li and Pellerei (2011) via stochastic representation

$$(X_1, X_2) = [\min(T_1, T_3), \min(T_2, T_3)],$$

where  $T_i$ ’s are independent random variables. Put  $a_0 = \lambda_3 > 0$ ,  $a_1 = 2\lambda_1$ ,  $a_2 = 2\lambda_2$  in (4) to get  $S_{X_1, X_2}(x_1, x_2) = \exp\{-\lambda_1 x_1^2 - \lambda_2 x_2^2 - \lambda_3 \max(x_1, x_2)\}$ .

It happens that the class  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  can be also characterized by "non-aging Sibuya dependence function"

$$\Omega_{X_1, X_2}(x_1, x_2) = \frac{S_{X_1, X_2}(x_1, x_2)}{S_{X_1}(x_1)S_{X_2}(x_2)}, \quad x_1, x_2 \geq 0,$$

being time-invariant with respect to the residual life time vector

$$(X_{1t}, X_{2t}) = [(X_1 - x_1, X_2 - x_2) \mid X_1 > t, X_2 > t],$$

i.e.,  $\Omega_{X_1, X_2}(x_1, x_2) = \Omega_{X_{1t}, X_{2t}}(x_1, x_2)$  for all  $t > 0$ .

The class  $\mathcal{L}(\mathbf{x}; \mathbf{a})$  is huge. It contains bivariate distributions that can be symmetric or asymmetric, being absolutely continuous or having a singular component, displaying positive or negative quadrant dependence. Hence, the Sibuya-copula class corresponding to (4) is broad. We will give only an example of a new bivariate distribution extending the Gumbel’s distribution (5).

**Example 3 (Extended Gumbel’s law).** Let  $S_{X_i}(x) = \exp\{-\lambda_i x\}$  for  $x \geq 0$ ,  $\lambda_i > 0$ , i.e.,  $f_{X_i}(x) = \lambda_i \exp\{-\lambda_i x\}$  and  $f_{X_i}(0) = \lambda_i$ ,  $i = 1, 2$ .

The parameter space  $\mathcal{A}$  in this case is

$$\mathcal{A} = \{\max(\lambda_1, \lambda_2) \leq a_0 \leq \lambda_1 + \lambda_2, a_1 + a_2 > 0 \text{ and } 0 \leq a_i \leq \lambda_i(a_0 - \lambda_i), i = 1, 2\}.$$

The bivariate survival function is absolutely continuous if  $a_0 = \lambda_1 + \lambda_2$  and one can deduce that  $a_i = \theta_i \lambda_1 \lambda_2$  for  $\theta_i \in (0, 1]$ ,  $i = 1, 2$ . Substitute these values in (4) to get

$$S_{X_1, X_2}(x_1, x_2) = \begin{cases} \exp\left\{-\left[\lambda_1 x_1 + \lambda_2 x_2 + \lambda_1 \lambda_2 x_2 \left(\theta_1 x_1 + \frac{\theta_2 - \theta_1}{2} x_2\right)\right]\right\}, & \text{if } x_1 \geq x_2, \\ \exp\left\{-\left[\lambda_1 x_1 + \lambda_2 x_2 + \lambda_1 \lambda_2 x_1 \left(\theta_2 x_2 + \frac{\theta_1 - \theta_2}{2} x_1\right)\right]\right\}, & \text{if } x_2 \geq x_1, \end{cases} \tag{7}$$

to be named *Extended Continuous Gumbel’s distribution* with parameters  $\lambda_1, \lambda_2, \theta_1$  and  $\theta_2$ . We will abbreviate it EG( $\lambda_1, \lambda_2, \theta_1, \theta_2$ ). When  $\theta_1 = \theta_2 = \theta$  and  $\lambda_1 = \lambda_2 = 1$  in (7) we obtain the Gumbel’s law (5).

The survival copula generating  $EG(\lambda_1 = 1, \lambda_2 = 1, \theta_1, \theta_2)$  is given by

$$\bar{C}_{EG}(u, v) = \begin{cases} uv \exp\{-\theta_1 \ln u \ln v\} \exp\{-0.5(\theta_2 - \theta_1)(\ln v)^2\}, & \text{if } 0 < u < v \leq 1, \\ uv \exp\{-\theta_2 \ln u \ln v\} \exp\{-0.5(\theta_1 - \theta_2)(\ln u)^2\}, & \text{if } 1 \geq u \geq v > 0. \end{cases} \quad (8)$$

Substituting  $\theta_1 = \theta_2 = \theta$  one will obtain Gumbel's survival copula (6).

We will use copulas (6) and (8) in data analysis presented in Section 3.

### 3. Data Analysis

We perform our analysis on the log-returns of crude oil ( $O_t$ ) and natural gas ( $G_t$ ) prices based on one-month-ahead futures contracts traded on NYMEX from July 1, 2003 to July 19, 2006 ( $n = 756$  observations in total), examined by Gregoire et al. (2008) as well. To detect the presence of autocorrelation and heteroscedasticity in both series, one should perform standard Box-Pierce and Ljung-Box tests on  $O_t, G_t$  and their squared values. To account for dependence between  $O_t$  and  $G_t$ , we estimated the empirical correlation coefficient ( $\rho_n = 0.508$ ) and empirical Kendall's tau ( $\tau_n = 0.349$ ), indicating a possible dependence between marginal series.

Gregoire et al. (2008) have found that the variation of  $G_t$  is best fitted by GARCH(1,6) model, with error term  $E_G$  being skewed  $t$ -distribution with four parameters. With respect to  $O_t$ , let us consider the random variable  $E_O$ , with realizations given by  $(O_t - \hat{\mu})/\hat{\sigma}$ , where  $\hat{\mu}$  and  $\hat{\sigma}$  are, respectively, the empirical mean and standard deviation of  $O_t$ . Gregoire et al. (2008) concluded that  $E_O$  can be accurately modeled by a Student distribution with  $df$  parameter equal to 13.745.

In a second step, assuming a possible inter-relation, the authors fit various families of copulas to  $E_O$  and  $E_G$ . In order to select the best copula  $C(E_O, E_G)$ , a range of goodness-of-fit tests are performed involving Cramer-von Mises test statistic

$$M_n = n \sum_{t=1}^n [C_{\Theta_n}(u_t, v_t) - C_n(u_t, v_t)]^2, \quad (9)$$

where  $C_{\Theta_n}(u_t, v_t)$  is the fitted copula with parameter vector  $\Theta$  and  $C_n(u_t, v_t)$  is the empirical copula,  $n = 756$ . The best choice in Gregoire et al. (2008) is the  $t$ -copula with parameters  $\rho = 0.522$  and  $df = 22$  (with a computed P-value of 3%). With a P-value of 1.4%, the Normal copula with parameter  $\rho = 0.522$  would be viable alternative.

The analysis in Gregoire et al. (2008) clearly shows that the  $O_t$  and  $G_t$  are not independent. Furthermore, Pareto and Frank copulas are inappropriate to model the dependence structure (see third column of Table 1). As to the specific form of the dependence between  $E_O$  and  $E_G$ , no definite conclusion can be drawn. In fact, in view of the extremely low P-values (not exceeding 0.03) none of the six copula models considered by authors provides an adequate description of the dependence.

We begin our analysis applying the GAS methodology for time varying parameters of individual log-returns,  $O_t$  and  $G_t$ . After a careful evaluation of more than 200 GAS models under test (using the available software in <http://gasmodel.com>) we confirm the findings in Gregoire et al. (2008) with respect to the distribution of  $E_G$ . In contrast, we conclude that GARCH(1,1) model provides the best fit to  $O_t$ , with  $O_t = 2.131 \times 10^{-2} + E_{O_t} \sqrt{h_t}$  and  $h_t = 2.131 \times 10^{-2} + 3.223 \times 10^{-2} O_{t-1}$ . The error terms  $E_{O_1}, E_{O_2}, \dots$  form a random sample from  $E_O$ , being  $t$ -distributed with parameter  $df = 11.095$ .

In a second step, to select the "best" copula joining the corresponding error terms in individual O and G series, we use the parametric bootstrap procedure described by Genest et al. (2009). Accounting for this dependence would lead to more adequate prediction of both crude oil and natural gas prices. In addition to copula functions considered by Gregoire et al. (2008) we included the EG copula (8). For each model considered, we generated 2500 bootstrap values of the Cramer-von Mises test statistic (9) and calculated the proportion of these values that are larger than  $M_n$ . The test results together with parameter estimates are summarized in Table 1.

Our analysis shows (see the second column in Table 1) that the null hypothesis cannot be rejected at 5% level for any of the three candidates: Student's  $t$ , Gumbel, and EG copulas. Using the highest P-value as a criterion to select the model with the best fit to the data at hand, we conclude that the EG copula (8) describes best the dependence between error terms of crude oil and natural gas log-returns.

Table 1. Test results for the goodness-of-fit of different copula models

Copula and estimated parameters	P-value	2008 P-value
Independence	0.000	0.0000
Pareto ( $\hat{\theta} = 0.545$ )	0.000	0.0000
Frank ( $\hat{\theta} = 2.001$ )	0.017	0.0095
Gaussian ( $\hat{\rho} = 0.552$ )		<b>0.0140</b>
$t$ -copula ( $\hat{\theta} = 0.552, df = 9$ )		<b>0.0300</b>
Gaussian ( $\hat{\rho} = 0.613$ )	0.028	
$t$ -copula ( $\hat{\theta} = 0.613, df = 8$ )	<b>0.051</b>	
Gumbel's (6) copula ( $\hat{\theta} = 0.772$ )	<b>0.212</b>	
EG (8) copula ( $\hat{\theta}_1 = 0.821 \hat{\theta}_2 = 0.711$ )	<b>0.431</b>	

Of course, Monte Carlo simulations should be performed to compare the observed prices (after July 2, 2006) with those predicted by our models. The main steps to simulate a specific model (EG copula (8), for example) are as follows:

- Generate a pair  $(u_t, v_t)$  from the EG copula (8);
- Set  $E_{O_t} = F^{-1}(u_t)$  and  $E_{G_t} = G^{-1}(v_t)$ , where  $F$  and  $G$  are  $t$ -distributed with parameters 13.745 and 11.095, correspondingly;
- Compute  $O_t = 2.131 \times 10^{-2} + E_{O_t} \sqrt{h_t}$  and  $G_t$  by expression given in Gregoire et al. (2008);
- Repeat above steps to get an empirical predictive realization of O and G and compare with future observations.

#### 4. Conclusions

The dependence among spatially or vertically linked markets has received much attention in the literature. Time-varying attributes in price co-movements can result from many reasons such as government interventions, financial contagion, disease outbreaks and altering consumer tastes. However, there is little available information about the data generating process. It is desirable to start in such scenarios with a general model and then to select a member that agrees with the patterns of the data. In this study we suggest to explore jointly the GAS models and Sibuya-type copulas as an alternative to the existing approaches for modeling price transmission that has appeared in the literature. The analysis in Section 3 shows that the combination of both recent methodologies leads to better results.

We assume that the dependence structure between series may vary over time, but the ratio between the joint distribution of error terms and the product of marginal distributions remains in equilibrium, being time-invariant. This is a restriction, but this approach leads to a huge class of distributions. Of course, the time-varying (dynamic) copula approach could be adapted in a second step, assigning a function of time for copula parameters. In this case, one should develop efficient estimation procedures and goodness of fit tests, specially in higher dimensions. Consult Patton (2009) and Cherubini et al. (2012) for further discussion.

Another important message in this paper is devoted to the practitioners who are ready to use static copulas in order to model temporal dependence between individual series. This is possible, but one should check, at last empirically (via Kendall's tau, say), if the data exhibit roughly constant dependence in time.

#### Acknowledgements

The authors are grateful for the useful referee's comments leading to the improvement of the paper. The first author is partially supported by FPAESP Grant No. 2013/07375-0.

#### References

- Alexander, C. (2005). Correlation in crude oil and natural gas markets. In V. Kaminski (Ed.), *Managing Energy Price Risk: The New Challenges and Solutions*, 573-604. Risk Books, London.
- Aloui, R., Hammoudeh, S., & Nguyen, D. K. (2013). A time-varying copula approach to oil and stock market dependence: the case of transition economies. *Energy Economics*, 39, 208-221. <https://doi.org/10.1016/j.eneco.2013.04.012>
- Beckmann, J., Berger, T., & Czudaj, R. (2016). Oil price and FX-Rates Dependency. *Quantitative Finance* 3, 477-488. <https://doi.org/10.1080/14697688.2015.1045930>
- Cherubini, U., Gobbi, F., Mulinacci, S., & Romagnoli, S. (2012). *Dynamic Copula Models in Finance*. Chichester, UK: John Wiley & Sons.

- Creal, D., Koopman, S., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28, 777-795. <https://doi.org/10.1002/jae.1279>
- Genest, C., Remillard, B., & Beaudoin, D. (2009). Omnibus goodness-of-fit tests for copulas: A review and power study. *Insurance: Mathematics and Economics*, 44, 199-213. <https://doi.org/10.1016/j.insmatheco.2007.10.005>
- Gregoire, V., Genest, C., & Gendron, M. (2008). Using copulas to model price dependence in energy markets. *Energy Risk*, 8, 62-68.
- Gumbel, E. (1960). Bivariate exponential distributions. *Journal of American Statistical Association*, 55, 698-707. <https://doi.org/10.1080/01621459.1960.10483368>
- Jäschke, S., Siburg, K., & Stoimenov, P. (2012). Modeling dependence of extreme events in energy markets using tail copulas. *Journal of Energy Markets*, 5, 60-80. <https://doi.org/10.21314/jem.2012.083>
- Jäschke, S. (2014). Estimation of risk measures in energy portfolios using modern copula techniques. *Computational Statistics and Data Analysis*, 76, 359-376. <https://doi.org/10.1016/j.csda.2014.01.019>
- Joe, H. (2015). *Dependence Modeling with Copulas*. Boca Raton: CRC Press.
- Kiatmanarochoy, T., & Sriboonchitta, S. (2014). Dependence structure between world crude oil prices: evidence from NYMEX, ICE, and DME markets. *Thai Journal of Mathematics*, 19, 181-198.
- Kulkarni, H. (2006). Characterizations and modelling of multivariate lack of memory property. *Metrika*, 64, 167-180. <https://doi.org/10.1007/s00184-006-0042-2>
- Kurowicka, D., & Joe, H. (Eds.). (2011). *Dependence Modeling : Vine Copula Handbook*. Singapore: World Scientific.
- Li, X., & Pellerey, F. (2011). Multivariate generalized Marshall-Olkin distributions and related bivariate properties. *Journal of Multivariate Analysis*, 102, 1399-1409. <https://doi.org/10.1016/j.jmva.2011.05.006>
- Oakes, D. (2005). On the preservation of copula structure under truncation. *The Canadian Journal of Statistics*, 33, 465-468. <https://doi.org/10.1002/cjs.5540330310>
- Pan, Z. (2014). Modeling tail dependence between energy market and stock markets in the BRIC countries. *Applied Economics Letters*, 21, 789-794. <https://doi.org/10.1080/13504851.2014.892188>
- Patton, A. (2009). Copula-based methods for financial time series. In T. G. Andersen, R. A. Davis, J.-P. Kreiss & T. Mikosch (Eds.), *Handbook of Financial Time Series*, 767-785. Springer, New York.
- Pinto, J., & Kolev, N. (2015). Copula representations for invariant dependence functions. In K. Glau, M. Scherer & R. Zagst (Eds.), *Innovations in Quantitative Risk Management*, (pp. 411-421). Springer Series in Mathematics & Statistics 99. [https://doi.org/10.1007/978-3-319-09114-3\\_24](https://doi.org/10.1007/978-3-319-09114-3_24)
- Pinto, J., & Kolev, N. (2016). A class of continuous bivariate distributions with linear sum of hazard gradient components. *Journal of Statistical Distributions and Applications*, 3, 1-17. <https://doi.org/10.1186/s40488-016-0048-x>
- Reboredo, J. C. (2011). How do crude oil prices co-move? A copula approach. *Energy Economics*, 33, 948-955. <https://doi.org/10.1016/j.eneco.2011.04.006>
- Schmidt, R., & Stadtmüller, U. (2006). Nonparametric estimation of tail dependence. *Scandinavian Journal of Statistics. Theory and Applications*, 33, 307-335. <https://doi.org/10.1111/j.1467-9469.2005.00483.x>
- Sibuya, M. (1960). Bivariate extreme statistics I. *Annals of the Institute of Statistical Mathematics*, 11, 195-210. <https://doi.org/10.1007/BF01682329>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).