SFB 823

Discussion Paper

Modelling dependence of extreme events in energy markets using tail copulas

Stefan Jäschke, Karl Friedrich Siburg, Pavel A. Stoimenov

Nr. 8/2011

SFB 823

Modelling dependence of extreme events in energy markets using tail copulas

Stefan Jäschke
RWE Supply & Trading GmbH
Risk Valuation
Altenessener Str. 27
45141 Essen
Germany
Email: stefan.jaeschke@rwe.com
Tel.: +49 201 1247235

Karl Friedrich Siburg
Fakultät für Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund
Germany
Email: karl.f.siburg@math.tu-dortmund.de
Tel.: +49 231 7557211

Pavel A. Stoimenov
Fakultät Statistik
Technische Universität Dortmund
Vogelpothsweg 78
44227 Dortmund
Germany
Email: pavel.stoimenov@tu-dortmund.de
Tel.: +49 231 7553127

January 28, 2011

Abstract

This paper studies, for the first time, the dependence of extreme events in energy markets. Based on a large data set comprising quotes of crude oil and natural gas futures, we estimate and model large co-movements of commodity returns. To detect the presence of tail dependence we apply a new method based on the concept of tail copulas which accounts for different scenarios of joint extreme outcomes. Moreover, we show that the common practice to fit copulas to the data cannot capture the dynamics in the tail of the joint distribution and, therefore, is unsuitable for risk management purposes.

Keywords: Extreme events; Crude oil; Natural gas; Tail dependence; Tail copulas

1 Introduction

As energy and commodity markets have become more volatile and increasingly interconnected in recent years, the need to capture the joint dynamics of commodity prices has become indispensable for managing energy risk. In contrast to the vast literature on modelling dependence between stock returns, there is a relative paucity of related research for the energy market. Alexander (2004) shows that the log-returns of futures on crude oil and natural gas exhibit asymmetric behaviour and strong, nonlinear dependence and are, therefore, a far cry from the assumption that their joint distribution is bivariate normal. The same problem is studied in Grégoire et al. (2008). They model the log-returns individually as time series, and account for the dependence between them by fitting various families of copulas to the error terms. In order to select the best copula, the authors perform a range of goodness-of-fit tests, which again show that the dependence between crude oil and natural gas log-returns cannot be characterized by the Gaussian copula. Accioly and Aiube (2008) studies the dependence of oil and gasoline prices. After adjusting auto regressive GARCH models to filter the linear and the nonlinear time dependence in the series of returns they fit various copulas to the residuals of these models. Dividing the sample in two periods, it is shown that the dependence is well represented by the t-copula and the Plackett copula, respectively.

The departure from normality in a multivariate setting poses an additional problem from the perspective of risk management, whose primary concern is the occurrence of extreme events. In the univariate case non-normality is associated to the skewness and leptokurtosis phenomena, or briefly the fat-tail problem. In the multivariate case, the fat-tail problem can be referred not only to the marginal distributions but also to the probability of large co-movements of the individual returns, i.e. tail dependence. Consequently, a thorough understanding of energy portfolio risk requires an adequate assessment of the probability that large price movements in energy markets occur together.

The latest statistical standard to describe the amount of extreme dependence is represented by the concept of tail dependence. To the best of our knowledge, however, the problem of modelling and estimating tail dependence between returns of energy commodities has not yet been addressed in the literature. The present paper opens this line of research by analyzing a large data set comprising quotes of the Light Sweet Crude Oil Futures and the Henry Hub Natural Gas Futures traded on the NYMEX (New York Mercantile Exchange). From a methodological point of view, the main tool is the theory of copulas, which allows the separate specification of the marginal distributions and the dependence structure.

The purpose of the paper is twofold: First, to provide an applicable model which estimates adequately the likelihood of joint extreme events in energy markets. Second, to illustrate the pitfalls of fitting copulas to data for risk management purposes. Undoubtedly, copulas represent the current standard for modelling stochastic dependence. However, we argue that a general goodness-of-fit test for copulas does not necessarily provide a good model for tail dependence. The reason therefor is that the procedure is based on minimizing the distance between observed and model values over the whole support of the distribution and therefore cannot capture the joint dynamics in the tail of the underlying distribution. Thus, to some extent the paper is also meant as an educative warning against adopting general copula inference techniques for modelling dependence of extreme events.

In order to overcome this difficulty, we apply the concept of tail copulas. A tail copula is a function of the underlying copula, which describes the dependence structure in the tail of multivariate distributions, but works more generally than the simple tail dependence coefficient, which is just the value of the tail copula at the point $(1, 1)$. Therefore, tail copulas enable the modelling of tail dependence of arbitrary form and, thus, account for all possible scenarios of joint extreme outcomes.

The paper is organized as follows. In the next section we briefly review some fundamental properties of copulas and introduce the closely related concept of tail copulas and tail dependence. Section 3 describes the data. To account for serial dependence in the data we estimate numerous univariate GARCH and APARCH models with various distributions for the error term. Section 4 presents our methodology and the empirical results. In a first step we fit different families of copulas to the residuals of the individual time series. Then, we estimate non-parametrically (Schmidt and Stadtmüller, 2006) the lower and upper tail copulas and show that the joint distribution of the log-returns of the crude oil and natural gas futures are both lower and upper tail dependent – a fact, which can not be detected by fitting a copula to the whole data set. The final Section 5 summarizes and provides an outlook for further research.

2 Preliminaries

The theory of copulas dates back to Sklar (1959), but its application to statistical modelling is far more recent. A copula is a function that embodies all the information about the dependence structure between the components of a random vector. From a probabilistic point of view, it is a multivariate distribution function with uniformly distributed margins on the interval $[0, 1]$. For notational convenience, all further definitions and results are provided for the bivariate case only. In the following, we consider a random vector (X, Y) with continuous marginal distribution functions $F(x) := \mathbb{P}[X \leq x]$ and $G(y) := \mathbb{P}[Y \leq y]$, $x, y \in \mathbb{R}$, respectively. By Sklar's theorem, there exists a unique copula C, called the copula of X and Y, such that

$$\mathbb{P}[X \leq x, Y \leq y] = C(F(x), G(y)), \tag{1}$$

for all $x, y \in \mathbb{R}$. Conversely, if C is a copula and F, G are distribution functions, then the function defined via (1) is a bivariate distribution function with margins F, G. It follows that copulas can be interpreted as dependence functions since they separate the marginal distributions from the dependence structure. In fact, the copula of X and Y is the joint distribution function of the probability integral transformations $U := F(X)$ and $V := G(Y)$, which are uniform on $[0, 1]$. It follows that

$$C(u, v) = \mathbb{P}[U \leq u, V \leq v] = \mathbb{P}[X \leq F^{-1}(u), Y \leq G^{-1}(v)], \tag{2}$$

for all $u, v \in [0, 1]$, where F^{-1} and G^{-1} denote the generalized inverses of F and G, respectively, i.e. $F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\}$, for all $u \in [0, 1]$ (analogously for G).

Let $\overline{F}(x) := 1 - F(x) = \mathbb{P}[X > x]$ and $\overline{G}(y) := 1 - G(y) = \mathbb{P}[Y > y]$ denote the corresponding survival functions of X and Y. Define a function $\widehat{C}$ by

$$\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v), \tag{3}$$

for all $u, v \in [0, 1]$. Then, for all $x, y \in \mathbb{R}$, we have

$$\mathbb{P}[X > x, Y > y] = \widehat{C}(\overline{F}(x), \overline{G}(y)). \tag{4}$$

The function $\widehat{C}$ is a copula itself and is called the survival copula of X and Y. In view of (4), the survival copula links the joint survival function to its univariate margins in a manner completely analogous to the one in which the copula connects the joint distribution function to its margins. Hence, for all $u, v \in [0, 1]$, we have

$$\begin{aligned} \widehat{C}(u, v) &= \mathbb{P}[U > 1 - u, V > 1 - v] \\ &= \mathbb{P}[X > F^{-1}(1 - u), Y > G^{-1}(1 - v)]. \end{aligned} \tag{5}$$

For further details regarding the theory of copulas we refer the reader to Nelsen (2006).

As the focus of this paper is to characterize and measure extreme dependence, the rest of this section is devoted to the concept of tail dependence, which concentrates on the upper and lower quadrant tails of the joint distribution. The standard way to determine whether X and Y are tail dependent is to look at the so-called lower and upper tail dependence coefficients, denoted by λ_L and λ_U, respectively. λ_L is the limit (if it exists) of the conditional probability that X is less than or equal to the u-th quantile of F, given that Y is less than or equal to the u-th quantile of G as u approaches 0, i.e.

$$\lambda_L := \lim_{u \to 0^+} \mathbb{P}[X \leq F^{-1}(u) \mid Y \leq G^{-1}(u)]. \tag{6}$$

Similarly, λ_U is the limit (if it exists) of the conditional probability that X is greater than the u-th quantile of F, given that Y is greater than the u-th quantile of G as u approaches 1, i.e.

$$\lambda_U := \lim_{u \to 1^-} \mathbb{P}[X > F^{-1}(u) \mid Y > G^{-1}(u)]. \tag{7}$$

The random vector (X, Y) is said to have lower (upper) tail dependence if $\lambda_L(\lambda_U) \in (0, 1]$, and no lower (upper) tail dependence if $\lambda_L(\lambda_U) = 0$.

The following identities show that the tail dependence coefficients are nonparametric and depend only on the copula C of X and Y. In particular, we have (see Nelsen, 2006)

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u} \tag{8}$$

and

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u} = \lim_{u \to 0^+} \frac{\widehat{C}(u, u)}{u}, \tag{9}$$

where $\widehat{C}$ is the survival copula of X and Y defined in (3). Since copulas are invariant under strictly increasing transformations of the random variables, it follows that λ_L and λ_U exhibit the same invariance property.

From a practitioner's point of view, tail dependence can be interpreted as the limiting likelihood of an asset/portfolio return falling below its Value at Risk at a certain level, given that another asset/portfolio return has fallen below its Value at Risk at the same level. However, like any scalar measure of dependence, λ_L and λ_U suffer from a certain loss of information concerning the joint behaviour in the tails of the distribution. In the context of tail dependence, the immediate analogue to copulas, which describe the entire dependence structure, is given by tail copulas; see Schmidt and Stadtmüller (2006) for further details on tail copulas. The lower tail copula Λ_L associated with X and Y is a function of their copula C and is defined by

$$\Lambda_L(x,y) := \lim_{t\to 0^+} \frac{C(tx,ty)}{t}, \tag{10}$$

if the above limit exists for all $x, y \in [0,\infty)$. The upper tail copula Λ_U associated with X and Y is a function of their survival copula $\widehat{C}$ and is defined by

$$\Lambda_U(x,y) := \lim_{t\to 0^+} \frac{\widehat{C}(tx,ty)}{t} \tag{11}$$

if the above limit exists for all $x, y \in [0,\infty)$.

The probabilistic interpretation of Λ_L and Λ_U is provided by the following relationships:

$$\Lambda_L(x,y) = y \lim_{t\to 0^+} \mathbb{P}[X \le F^{-1}(tx) \mid Y \le G^{-1}(ty)], \tag{12}$$

$$\Lambda_U(x,y) = y \lim_{t\to 0^+} \mathbb{P}[X > F^{-1}(1-tx) \mid Y > G^{-1}(1-ty)]. \tag{13}$$

It is easy to show that the tail dependence coefficients are a special case of the respective tail copulas. More precisely, we have

$$\lambda_L = \Lambda_L(1,1) \quad \text{and} \quad \lambda_U = \Lambda_U(1,1). \tag{14}$$

As pointed out in Schmidt and Stadtmüller (2006), another reason to embed the tail dependence coefficients in the framework of tail copulas is to facilitate their estimation, which is a non-trivial task, especially for non-standard distributions.

3 Data

Our empirical investigation focuses on the dependence between crude oil and natural gas prices. The data set covers quotes of the Light Sweet Crude Oil Futures and the Henry Hub Natural Gas Futures traded on the NYMEX (New York Mercantile Exchange) between July 2, 2007 and July 2, 2010. Both futures contracts are generic one-month-ahead futures. The quotes are collected from Bloomberg's Financial Information Services. The log-returns for both series are plotted in Figure 1. In order to estimate and model the dependence between the two commodities, it is necessary to consider dependence within the individual time series. To detect the presence of heteroscedasticity we perform standard Box-Pierce and Ljung-Box tests on the squared log-returns for three different lags (lag 1, lag 5 and lag 10). For crude oil, both tests are significant at the 1% level for all

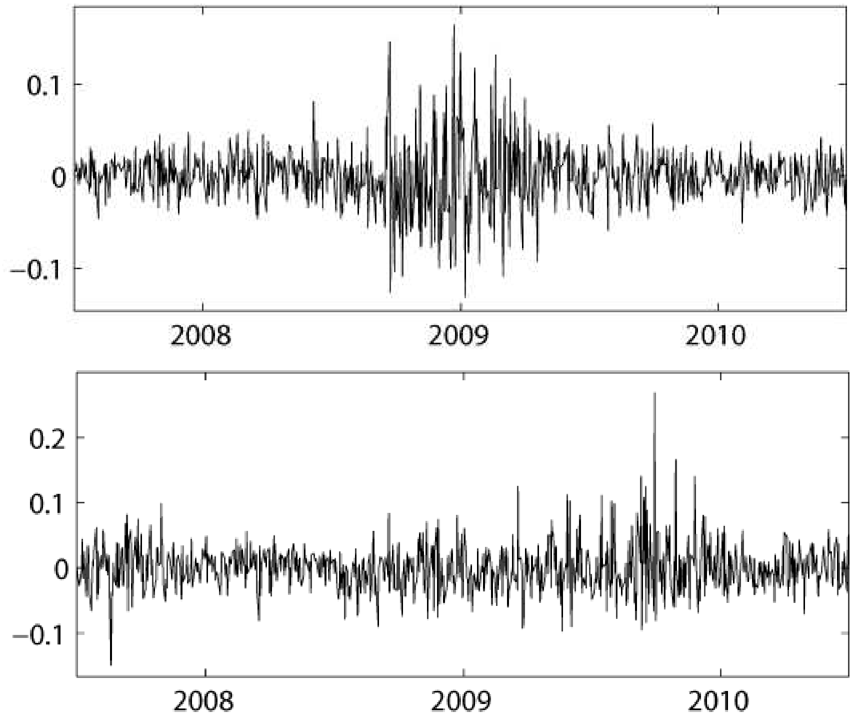

Figure 1: Log-returns of one-month-ahead futures on crude oil (top) and natural gas (bottom). Source: Bloomberg

lags. Applied to the natural gas data, the same is true for lag 5 and lag 10, while the null hypothesis that none of the autocorrelation coefficients up to a certain lag are different from zero cannot be rejected for lag 1 (p-value $= 0.18$).

These findings show that the assumption of an i.i.d. sample is unrealistic and therefore one has to account for heteroscedasticity in the marginal series. For this purpose, we employ GARCH(p, q) and APARCH(p, q)-models, $p, q \in \{1, \dots, 6\}$, with various distributions for the error term, including the normal, the skew normal, the general error, the skewed generalized error and the skew Student's t-distribution. After careful evaluation of the 360 models under test we conclude that the APARCH$(1, 1)$-model with skew normal distribution describes adequately the data generating process for the log-returns

$O_t, t = 1, \ldots, T$ of the crude oil futures:

$$\begin{aligned} O_t &= 2.639 \times 10^{-5} + \sigma_t X_t, \\ \sigma_t^2 &= 1.193 \times 10^{-5} \\ &\quad + 9.542 \times 10^{-2}(|O_{t-1}| - 2.385 \times 10^{-1} O_{t-1})^2 \\ &\quad + 8.880 \times 10^{-1} \sigma_{t-1}^2. \end{aligned} \tag{15}$$

Thus, the standardized residuals $X_1, X_2, \ldots, X_T$ can be viewed as a random sample from a skew normal distribution with skewness parameter $\alpha = 0.986$; see Azzalini and Dalla Valle (1996) for details on the skew normal distribution.

With respect to the log-returns $N_t, t = 1, \ldots, T$ of the natural gas futures, the GARCH(1,1)-model with skew Student's t-distribution

$$\begin{aligned} N_t &= -4.841 \times 10^{-4} + \sigma_t Y_t, \\ \sigma_t^2 &= 2.648 \times 10^{-5} + 5.106 \times 10^{-2} N_{t-1}^2 \\ &\quad + 9.281 \times 10^{-1} \sigma_{t-1}^2 \end{aligned} \tag{16}$$

provides the best fit to the data. Analogously, we conclude that the standardized residuals $Y_1, Y_2, \ldots, Y_T$ are randomly drawn from a skew Student's t-distribution with skewness parameter $\alpha = 1.113$ and $\nu = 8.094$ degrees of freedom. For the definition and properties of the skew Student's t-distribution we refer to Azzalini and Capitano (2003).

Applying now the Box-Pierce and Ljung-Box tests to the squared standardized residuals of both series for lag 1, lag 5 and lag 10, provides completely different test results than initially obtained. None of the tests detects the presence of auto correlation at the 5% level. In fact, apart from the test results for the crude oil futures at lag 1 (p-value = 0.06), the p-values exceed 0.4.

4 Methodology and results

4.1 Copula estimation

Having specified adequate models for the log-returns of the crude oil and natural gas futures, O_t and N_t, we now address the problem of modelling their dependence. For this purpose, we employ the theory of copulas, briefly introduced in Section 2. The main advantage of copulas is that they allow the separate specification of the marginal distributions and the dependence structure.

As shown in the preceding section, $(O_t, N_t), t = 1, \ldots, T$ certainly does not constitute an i.i.d. sample. Therefore, in order to estimate their copula, we consider the sample (X_t, Y_t) of the standardized residuals of the individual time-series models (15) and (16). The next step is to transform each pair of observation into its rank based representation (u_t, v_t), calculated by

$$u_t = \frac{\text{rank}(X_t)}{T+1} \quad \text{and} \quad v_t = \frac{\text{rank}(Y_t)}{T+1}. \tag{17}$$

Figure 2 shows a scatter plot of the 758 pairs (u_t, v_t). It reveals a certain tendency of u_t and v_t, and thus of X_t and Y_t to vary together, regardless of their marginal distributions.

To confirm this judgement, we first estimate the two most common rank correlation coefficients, Spearman's ρ and Kendall's τ, which depend solely on the underlying copula via

$$\rho(C) = -3 + 12 \int_0^1 \int_0^1 C(u,v)\, du\, dv \tag{18}$$

$$\tau(C) = -1 + 4 \int_0^1 \int_0^1 C(u,v) c(u,v)\, du\, dv \tag{19}$$

where $c(u,v) = \partial^2 C(u,v)/\partial u \partial v$ is the density of C (assuming it exists). For the empirical measures ϱ_T and τ_T (Genest and Favre (2007)) we find that $\varrho_T = 0.314$ and $\tau_T = 0.214$, which are both significantly different from zero (p-value < 0.001). For details about the test statistics and their asymptotic distributions we refer to Genest and Favre (2007). These preliminary results are in accordance with the ones reported in Grégoire et al. (2008), who perform the same tests on their data set.

Note that, as any scalar measure of dependence, ρ and τ cannot describe the whole dependence structure of the joint distribution. Therefore, having concluded that the log-returns of the crude oil and natural gas futures are positively dependent, we now address the problem of estimating their copula. For this purpose we first compute the empirical copula $\mathbb{C}_T$ (Deheuvels, 1979) and the empirical survival copula $\widehat{\mathbb{C}}_T$, defined by

$$\mathbb{C}_T(u,v) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(u_t \leq u, v_t \leq v) \tag{20}$$

and

$$\widehat{\mathbb{C}}_T(u,v) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(u_t > u, v_t > v), \tag{21}$$

where $\mathbb{1}$ denotes the indicator function.

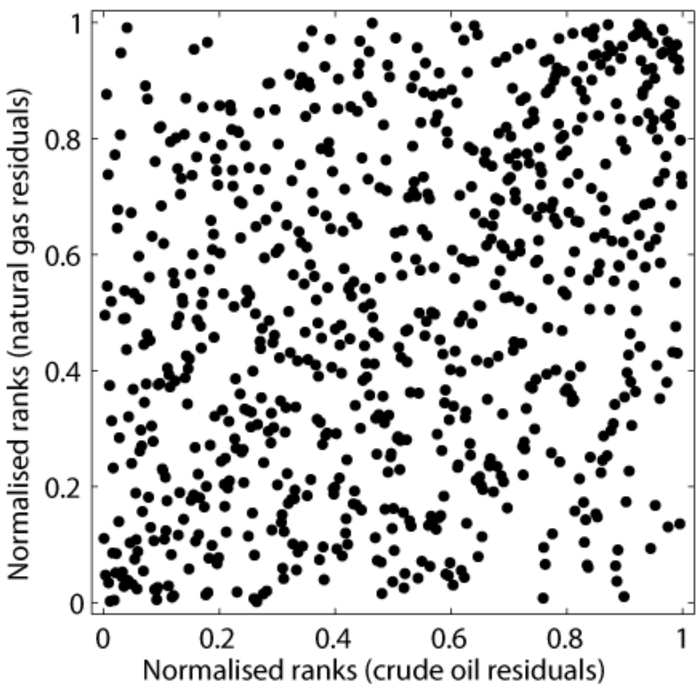

Figure 2: Scatter plot of the ranks of the standardized residuals u_t and v_t.

In order to find an appropriate copula model, we consider different parametric families of copulas commonly used in economics and finance. Beside the independence copula $C(u, v) = uv$, we fit two elliptical copulas, Gaussian and t, and three Archimedean copulas, Pareto, Frank and Gumbel. Each of these families is completely characterized by a single parameter, θ, with exception of the t-copula, which, in addition, requires the specification of its degree of freedom ν. For each copula family, we estimate the unknown parameter θ by inversion of Kendall's τ (Genest and Rémillard, 2008). More precisely, the estimate $\hat{\theta}_T$ is computed by substituting $\hat{\tau}_T$ for τ in formula (19), and then solving the equation to obtain $\hat{\theta}_T$, which gives us the best copula fit $C_{\hat{\theta}_T}$ within the respective copula family. The inversion approach utilizes the fact that, for the considered copula families, θ is a monotone function of τ, which, for example, for the Gaussian and the t-copula is given by $\theta = \sin(\tau\pi/2)$.

Finally, to select the optimal copula model, we apply a goodness-

Copula	p**-value**
Independence	0.000
Pareto ($\theta_T = 0.545$)	0.000
Frank ($\theta_T = 2.001$)	0.027
Gaussian ($\theta_T = 0.330$)	0.289
t ($\theta_T = 0.330, \nu = 9$)	0.311
Gumbel ($\theta_T = 1.272$)	0.712

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

of-fit test, based on the Cramér-von Mises statistic

$$S_T = \sum_{t=1}^{T} \left\{ \mathbb{C}_T(u_t, v_t) - C_{\theta_T}(u_t, v_t) \right\}^2 . \tag{22}$$

A review and comparison of goodness-of-fit procedures is given by Genest et al. (2009). This statistic measures how close the fitted copula C_{θ_T} is from the empirical copula $\mathbb{C}_T$. Since the distribution of S_T depends on the unknown value of θ_T under the null hypothesis that the true copula C is from the respective copula family, we compute the p-values of the test using the parametric bootstrap procedure described by Genest and Rémillard (2008).

The test results together with the estimates θ_T are summarized in Table 1. They clearly show that the log-returns of crude oil and natural gas are not independent. Furthermore, the Pareto copula and, to a great extent, the Frank copula seem inappropriate to model the dependence structure. As to the specific form of the dependence, no definite conclusion can be drawn. The null hypothesis cannot be rejected for any of the other three candidates. In particular, even the Gaussian copula could be an applicable alternative, which of course does not imply that the joint distribution is bivariate normal. Using the highest p-value as a criterion to select the model with the best fit to the data, we conclude that the Gumbel copula, with $\theta_T = 1.272$, describes best the dependence between crude oil and natural gas.

Finally, we point out, that our results differ substantially from those of Grégoire et al. (2008), who find, for example, that for the Gumbel copula the null hypothesis can be rejected on the basis of the same test procedure. In fact, in view of the extremely low p-values, not exceeding 0.03, none of the six copula models considered in Grégoire et al. (2008) provides an adequate description of the dependence between crude oil and natural gas log-returns.

4.2 Tail copula estimation

We now address the question whether the joint distribution of the log-returns of the crude oil and natural gas futures has a tendency to generate extreme values simultaneously and is, in this sense, a dangerous distribution for risk managers. The scatter plot in Figure 2 reveals a pronounced concentration of data points in both tails of the distribution. As concluded above, however, the overall dependence structure is well represented by the Gumbel copula, which exhibits no lower tail dependence since, in view of Equation (8), its lower tail dependence coefficient λ_L is 0. Thus, as to the modelling of extreme dependence in the lower tail of the distribution, the Gumbel copula is no better choice than the independence copula.

We point out that this seeming contradiction suggests neither that the variables are tail independent, nor that the selected copula is unsuitable to model the overall dependence. It rather shows that a general goodness-of-fit test for copulas does not necessarily provide an appropriate model for tail dependence, simply because the procedure is based on minimizing the distance between observed and model values over the whole support of the distribution. In fact, one of the main aims of this paper is to illustrate the lack of effectiveness of fitting copulas to the data in capturing the dependence in the tail of the underlying distribution.

In order to assess the risk of joint extreme events, we apply the theory of tail copulas, briefly introduced in Section 2. For the estimation of the tail copulas, we use the lower and upper empirical tail copulas, denoted by $\mathbb{A}_{L,T}$ and $\mathbb{A}_{U,T}$, respectively. These non-parametric estimators, introduced and studied by Schmidt and Stadtmüller (2006), are defined by

$$\begin{aligned}\mathbb{A}_{L,T}(x,y) &:= \frac{T}{k}\mathbb{C}_T\left(\frac{kx}{T},\frac{ky}{T}\right)\\ &\approx \frac{1}{k}\sum_{t=1}^{T}\mathbb{1}\left(u_t \le \frac{kx}{T+1}, v_t \le \frac{ky}{T+1}\right)\end{aligned} \tag{23}$$

and

$$\begin{aligned}\mathbb{A}_{U,T}(x,y) &:= \frac{T}{k}\widehat{\mathbb{C}}_T\left(\frac{kx}{T},\frac{ky}{T}\right)\\ &\approx \frac{1}{k}\sum_{t=1}^{T}\mathbb{1}\left(u_t > \frac{T-kx}{T+1}, v_t > \frac{T-ky}{T+1}\right)\end{aligned} \tag{24}$$

with some parameter $k \in \{1, \ldots, T\}$ to be chosen by the statistician. Under the assumptions that $k = k(T) \to \infty$ and $k/T \to 0$ for $T \to \infty$, Schmidt and Stadtmüller (2006) establishes weak convergence and strong consistency for $\mathbb{\Lambda}_{L,T}$ and $\mathbb{\Lambda}_{U,T}$. The paper also describes a method of choosing the optimal threshold k via a simple plateau-finding algorithm. Implementing this procedure, we calculate the empirical tail copulas for the log-returns of the crude oil and natural gas futures. These are visualized in Figures 3 and 4.

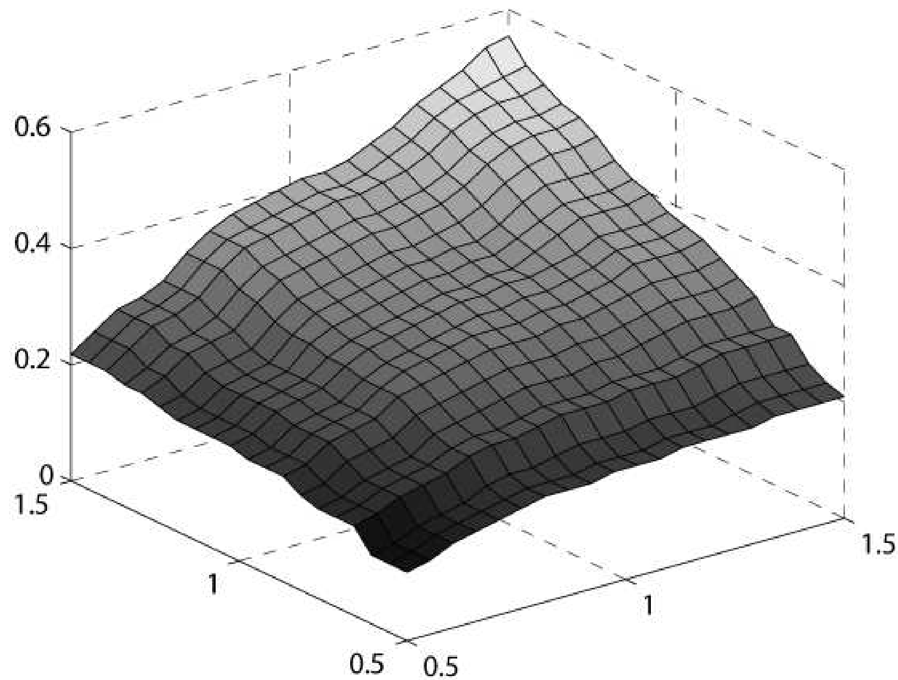

Figure 3: Empirical lower tail copula of the log-returns of the crude oil and natural gas futures.

Schmidt and Stadtmüller (2006) develops a consistent non-parametric estimator for the lower and upper tail dependence coefficients, given by

$$\mathbb{\lambda}_{L,T} := \mathbb{\Lambda}_{L,T}(1,1) \quad \text{and} \quad \mathbb{\lambda}_{U,T} := \mathbb{\Lambda}_{U,T}(1,1), \tag{25}$$

respectively. The estimators $\mathbb{\lambda}_{L,T}$ and $\mathbb{\lambda}_{U,T}$, based on the empirical counterparts of the identities given in (14), emphasize that tail copulas are an intuitive generalization of the tail dependence coefficients via a function describing the complete dependence structure in the tail of a distribution. Therefore, tail copulas constitute a powerful tool

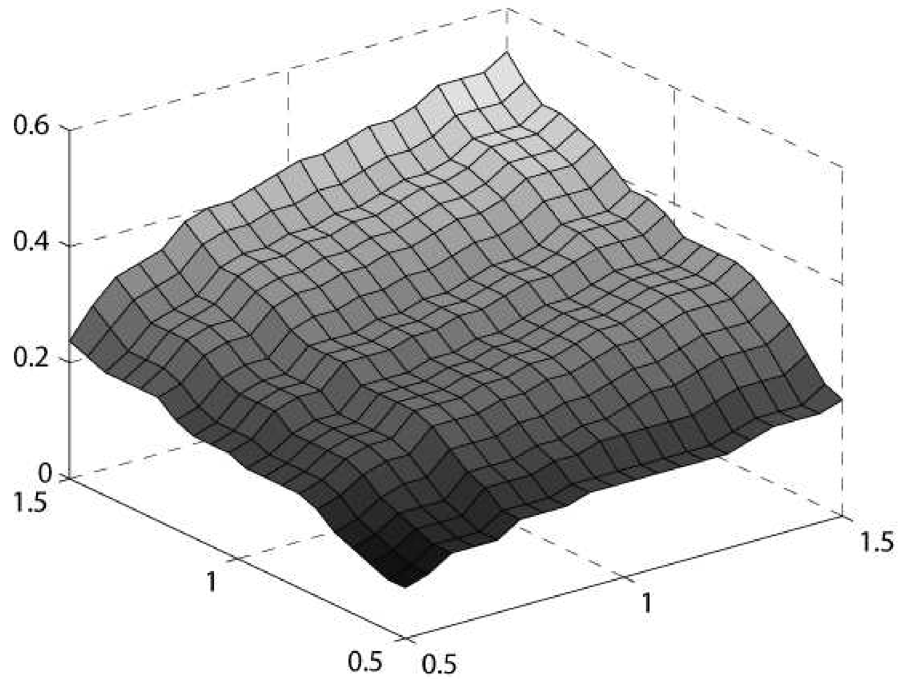

Figure 4: Empirical upper tail copula of the log-returns of the crude oil and natural gas futures.

for modelling tail dependence of arbitrary form. Moreover, as pointed out in Schmidt and Stadtmüller (2006), tail copulas also provide a convenient method for the, otherwise, non-trivial task of estimating λ_L and λ_U.

For our data set, we find $\lambda_{L,T} = 0.32$ and $\lambda_{U,T} = 0.28$, which implies that the log-returns of the crude oil and natural gas futures exhibit tail dependence. Our estimates clearly demonstrate that the Gumbel copula, with $\theta_T = 1.272$, which according to the goodness-of-fit test conducted in Section 4.1 provides the best fit to the data, does not describe adequately the risk of extreme events in the lower tail of the distribution. This raises the question of the optimal copula choice from a risk management perspective.

One possibility to deal with this problem is to select the copula, whose tail dependence coefficients are closest to their empirical counterparts. Table 2 lists the tail-dependence coefficients of the copulas considered for the goodness-of-fit test. Note that the t-copula here has the same value of θ_T, but less degrees of freedom ν than the fitted one in Table 1 since the lower ν, the higher the tail dependence coefficients. Our choice $\nu = 2.7$ is dictated by the fact that for smaller degrees of

Copula	$\boldsymbol{\lambda_L}$	$\boldsymbol{\lambda_U}$
Independence	0.00	0.00
Pareto ($\hat{\theta}_T = 0.545$)	0.28	0.00
Frank ($\hat{\theta}_T = 2.001$)	0.00	0.00
Gaussian ($\hat{\theta}_T = 0.330$)	0.00	0.00
t ($\hat{\theta}_T = 0.330, \nu = 2.7$)	0.25	0.25
Gumbel ($\hat{\theta}_T = 1.272$)	0.00	0.28
Empirical	0.32	0.28

Table 2: Tail dependence coefficients of the copulas considered for the goodness-of-fit test presented in Table 1.

freedom the null hypothesis of the goodness-of-fit test conducted in Section 4.1 must be rejected at the 5%-level.

It turns out that λ_L of the Pareto copula, with $\hat{\theta}_T = 0.545$, is closest to our estimate $\hat{\lambda}_{L,T}$, although according to the goodness-of-fit test results, presented in Table 1, the Pareto copula clearly fails to capture the overall dependence structure of the crude oil and natural gas returns. With respect to the upper tail of the distribution, the situation changes completely. Here the Gumbel copula, $\hat{\theta}_T = 1.272$, which exhibits the highest p-value of the goodness-of-fit test, remains the best model, since its $\lambda_U = 0.28$ even coincides with the value of $\hat{\lambda}_{U,T}$.

At this point, we emphasize again that λ_L and λ_U, as scalar measures of tail dependence, cannot characterize the entire dependence structure in the tails of the distribution. For example, two tail copulas with the same value at $(1, 1)$, corresponding to λ_L or λ_U, respectively, could differ substantially at other points in their domain. Thus, they could incorporate different levels of risk, although judging from the tail dependence coefficients alone, they would seem equally dangerous. In order to account for all possible scenarios of joint extreme outcomes one should utilize more information from the tail copula than simply its value at the point $(1, 1)$.

From Table 2, it is evident that, beside the Pareto and Gumbel copulas, the only alternative for modelling tail dependence between the log-returns of crude oil and natural gas is the t-copula since all other copula models exhibit no tail dependence. Figures 5 and 6 visualize the values of the respective tail copulas and their empirical

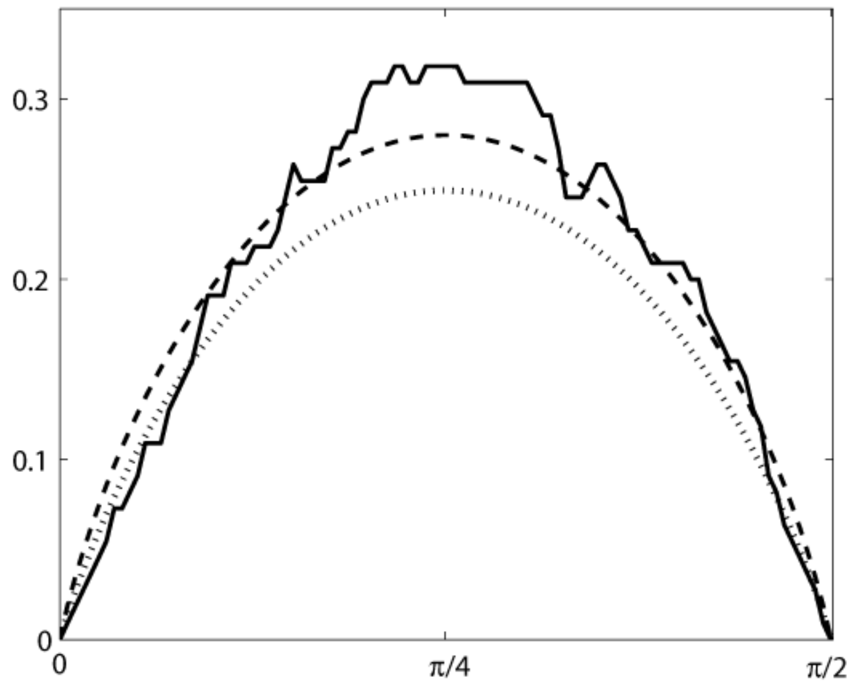

Figure 5: Parametrization of the circle around the origin through the point $(1, 1)$ in the first quadrant with values of the empirical lower tail copula (solid line) as well as the lower tail copulas of the Pareto copula, $\theta_T = 0.545$ (dashed line), and the t-copula, $\theta_T = 0.330, \nu = 2.7$ (dotted line).

counterparts along the circle passing through the point $(1, 1)$. Thus, the (empirical) tail dependence coefficients are the values of the functions at $\pi/4$. Instead of comparing the plotted functions at a single point in their domain, we propose a least squares approach and calculate the Cramér-von Mises statistic for the empirical tail copula and the tail copula of each of the fitted copulas. This approach has the advantage of taking into account simultaneously different scenarios of extreme co-movements of the two variables. We find that the Pareto and Gumbel copulas remain the best models for lower and upper tail dependence, respectively.

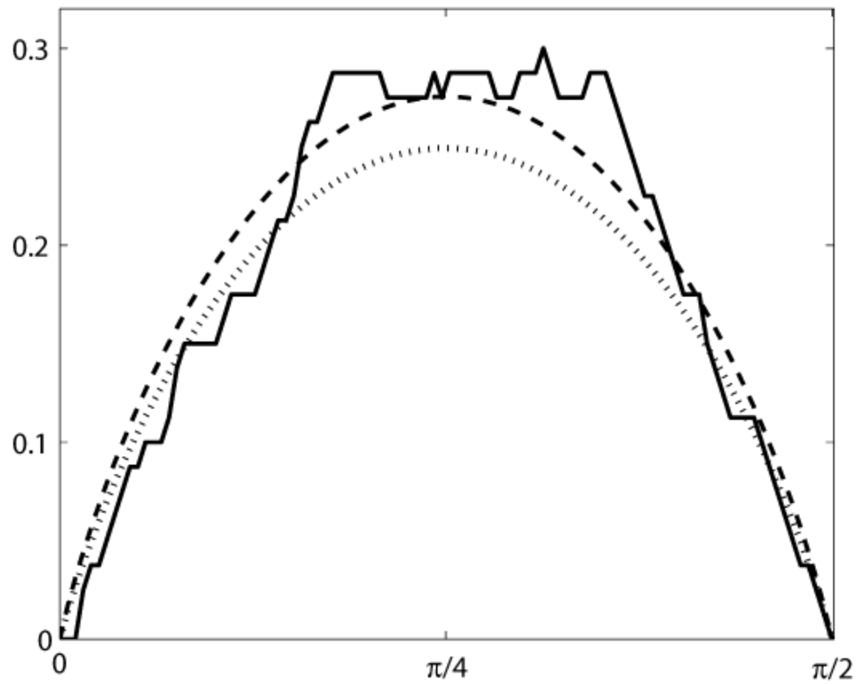

Figure 6: Parametrization of the circle around the origin through the point $(1,1)$ in the first quadrant with values of the empirical upper tail copula (solid line) as well as the upper tail copulas of the Gumbel copula, $\vartheta_T = 1.272$ (dashed line), and the t-copula, $\vartheta_T = 0.330, \nu = 2.7$ (dotted line).

5 Concluding remarks

This paper studies, for the first time, the dependence of extreme events in energy markets. In particular, we estimate and model both the overall and the tail dependence of crude oil and natural gas returns. The main conclusions which can be drawn from our empirical investigation can be summarized as follows.

According to the conducted goodness-of-fit test, the Gumbel copula describes best the overall dependence structure, but is unable to generate the joint occurrence of large drops in the analyzed commodity prices. However, applying the nonparametric technique for estimation of the tail copula introduced in Schmidt and Stadtmüller (2006), we detect the presence of such extreme events. This shows that a good model for the overall dependence between commodity returns can be a very bad model for their tail dependence. In fact, we see that using the copula which provides the best fit to the data can be very misleading for risk management which requires an adequate assessment of the probability that large price movements occur together.

With respect to the applicability of our results, it must be said that although the Pareto and Gumbel copulas describe well the lower and upper tail dependence in the data, respectively, they perform very badly in the other tail of the joint distribution. Among the considered copula families, the only copula which delivers satisfactory results as a general and, at the same time, as a tail dependence model is the t-copula.

Acknowledgements

The third author gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft (SFB 823).

References

Accioly, R. M. S. and F. A. L. Aiube (2008). Analysis of crude oil and gasoline prices through copulas. *Cadernos do IME – Série Estatística 24*, 15–28.

Alexander, C. (2004). Correlations in crude oil and natural gas markets. In V. Kaminsky (Ed.), *Managing energy price risk* (third ed.)., pp. 573–606. London: Risk Books.

Azzalini, A. and A. Capitano (2003). Distributions generated by pertubation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc., B 65*, 367–389.

Azzalini, A. and A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika 83*, 715–726.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Acad. R. Belg., Bull. Cl. Sci. 65*, 274–292.

Genest, C. and A. C. Favre (2007, July). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*(4), 347–368.

Genest, C. and B. Rémillard (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré - Probabilités et Statistique 44*(6), 1096–1127.

Genest, C., B. Rémillard, and D. Beaudoin (2009, April). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom. 44*(2), 199–213.

Grégoire, V., C. Genest, and M. Gendron (2008). Using copulas to model price dependencies in energy markets. *Energy Risk 5*(5), 58–64.

Nelsen, R. B. (2006). *An introduction to copulas* (second ed.). New York: Springer.

Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scand. J. Statist. 33*(2), 307–335.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.