CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
Rio de Janeiro - RJ - Brasil
ISSN 1413-9022 / v. 24, p. 15 - 28, 2008

ANALYSIS OF CRUDE OIL AND GASOLINE PRICES THROUGH COPULAS

Ricardo de Melo e Silva Accioly
Universidade do Estado do Rio de Janeiro – Petróleo Brasileiro S.A.
ricardo.accioly@gmail.com

Fernando Antonio Lucena Aiube
Pontifícia Universidade Católica do Rio de Janeiro - Petróleo Brasileiro S.A.
aiube@puc-rio.br

Abstract

In this paper we investigate the dependence of crude oil and gasoline prices. The understanding of the behavior of this dependence is useful for modeling the portfolio of investments in an integrated oil company. An accurate simulation of the behavior of these prices reveals precisely the risk and return of the portfolio. Morover the movements of these prices is crucial for goverment planing since they affect the overall economy of developed and developing countries. The classical approach which uses elliptical distributions to model the risk factors can be misleading since they are actually not elliptical. We used copula to establicsh such dependence since this methodolgy precludes the use of elliptical distributions. We found a change in the behavior of prices in the recent period compared to those in beginning of the decade and this fact is also reported in the literature. This change is observed through different copula models that were adjusted. These results were confirmed with a bootstrap analysis.

Key-words*: Crude oil prices, Gasoline prices, Copulas.*

1. Introduction

Crude oil is by far the most important commodity in the world. Its relevance is related to its importance as the main source of energy in developed and in developing economies. Its price can affect the economy of many countries for long periods. Recently crude oil and gasoline prices have shown a peculiar behavior. They had been an upward movement since 2003 until the end of the first semester of this year. In the second semester of 2008 they had a sudden drop mainly due to the fear of a global recession. The comprehension of the dynamics of crude oil and gasoline prices and their relationship is relevant for: (i) governmental economic and energetic planning; (ii) oil industry as a supplier; (iii) consumer industry; and (iv) consumer in general.

Traditionally when modeling the dynamics of prices one assumes they have a Gaussian or elliptical distribution. This is a common hypothesis in finance literature both from stochastic processes perspective as well as from econometrics point of view (see Mendes e Souza (2004)). The key point in this framework is that the statistical dependence or the co-movement of the variables can be modeled through a linear correlation as the Pearson coefficient or a covariance matrix a in a multivariate case.

Unfortunately, most of the random variables are not elliptically distributed. For example, prices (or returns) which represent the fundamental risk factor in economics and finance are non elliptical, strictly speaking. As a consequence the structure of dependence of these variables could be compromised if this simplified approach is adopted.

The methodology which is suited in capturing the whole dependence of variables is provided by copulas. Roughly speaking the copula is a function that joins the marginal distributions of random variables in a multivariate distribution describing the joint behavior. Or statistically speaking, the copula allows the description of a multivariate distribution in terms of a specific dependence structure of the marginal distributions of these variables. Copulas were introduced in 1959 in a context of probability, (see Frees and Valdez (1998)). Its use in insurance industry started in 1995, (see Embrechts (2008)) and spread to different fields of finance such as risk, decision analysis, portfolio management, and pricing derivatives. The use of copulas in these fields is well detailed in McNeil *et al*. (2005) and Cherubini *et al.* (2004), among others. Despite the growing interest of copula in finance research, it has been used in different

areas of science such as hydrology, see Genest and Frave (2007), oil field development, see Accioly and Chiyoshi (2004), for example.

The main goal of this paper is to investigate the dependence structure of crude oil prices and gasoline prices. From the perspective of an oil firm this dependence is crucial for the portfolio management of real projects. The understanding of the actual risks involved in such portfolio and the return predicted are dependent on the correct modeling of prices (or risk factors) embedded in this analysis. This way we will investigate the behavior of these variables based on past information. We will adjust auto regressive GARCH (AR-GARCH) models to filter the linear and the non linear time dependence in the series of returns. The residual of these models will be studied through copulas. Hence we are able to capture the true interdependence of the variables.

The paper is organized as follows: section 2 presents an overview of copula functions, section 3 presents the data and the econometric modeling of AR-GARCH models, section 4 analyzes the dependence of the variables and section 5 presents the conclusion.

2. Copula basics

Our starting point is as simple as most real problems. We know, for example, two variables or two risk factors. Then we want to know how these two variables are related or how to describe the sensitivity of one to the other. To achieve this goal we are going to construct the joint distribution using the concept of copula. At this point we refer to the textbooks of Nelsen (1999) and Joe (1997), among others.

Consider X_1 and X_2 two variables that represent two risk factors. The joint distribution function of these two risk factors is given by.

$$F_X(x) = P(X_1 \leq x_1, X_2 \leq x_2) \qquad x \in R^2$$

Consider that we know the information about the distribution function of these two variables F_{X1} and F_{X2}, in other words, we know the marginal distributions of these two risk factors X_1 and X_2, such as $F_{X_i}(x) = P(X_i \leq x)$ $x \in R$. Assume that these marginal distributions are continuous and strictly increasing. The copula is a function C

defined in $[0,1]^2$ with uniform marginals in (0,1) which provides the dependence link between F and the marginals F_{X1} and F_{X2} so that

$$F_X(x) = C(F_1(x_1), F_2(x_2)) \quad (1)$$

Where $F_i = F_{Xi}$. We can also write equation (1) in the converse form

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \quad (2)$$

Sklar (1959) showed what is essential in this formulation: if the marginals are continuous and strictly increasing, the function C is unique. Sklar's Theorem is absolutely general: any joint distribution can be written in copula form.

3. Time series modeling

There are different types of crude oil negotiated in the world. Among them the WTI (West Texas Intermediate) is the most liquid traded crude oil. We are going to use the historical WTI and gasoline first future contracts to proceed with this study. We took the daily close prices for the first future contract for gasoline and crude oil (WTI) from 05/01/1990 until 09/26/2008. Table 1 presents the main statistics of these variables. The sample contains 4614 values and they are reported in US$/bbl.

Table 1: Main Statistics of the Series

Statistic	**Gasoline**	**Crude oil**
Mean	40,68	33.93
Median	28.95	24.13
Maximum	149.98	145.29
Minimum	13.68	10.72
Std. Dev.	26.23	23.99
Skewness	1.77	2.02
Kurtosis	5.71	7.08

There is empirical evidence of the existence of a structural break at end of 2003, see Murat and Tokat (2008). Our study has some similarities with Grégoire *et al.* (2008), but they looked only to the dependence between oil prices and natural gas prices. They used a sample that encompasses the period of July, 2003 until July, 2006. In this article we divided the entire sample into two samples following the evidences of a structural break in 2003. The first period begins in 05/01/1990 and finishes in 12/31/2003. The second period covers the rest of the sample from 01/05/2004 to 09/26/2008.

We proceeded adjusting an AR-GARCH model to the log returns of these series for both periods. This way we filtered the linear and the quadratic dependences (the dependence on variance) in each series. The residuals of each series were tested. The Box-Pierce test showed that they are uncorrelated. We used the ARCH-LM and we could not reject the null hypothesis that there is no ARCH effect. For period 1 we modeled the log returns for each one of the series as:

- Gasoline:

$$r_t = \underset{(-1.96)}{-0.00782} + \underset{(3.46)}{0.002874 d_w} \underset{(-2.97)}{- 0.05278 r_{t-6}} + h_t^{1/2}\varepsilon_t$$

$$h_t = \underset{(5.80)}{8.97\times 10^{-6}} + \underset{(18.14)}{0.07691\varepsilon_{t-1}^2} + \underset{(166.86)}{0.9134 h_{t-1}}$$

- Crude oil:

$$r_t = \underset{(-3.01)}{-0.05347 r_{t-2}} \underset{(-2.57)}{- 0.04473 r_{t-6}} + h_t^{1/2}\varepsilon_t$$

$$h_t = \underset{(4.70)}{3.99\times 10^{-6}} + \underset{(15.20)}{0.075321\varepsilon_{t-1}^2} + \underset{(161.15)}{0.9223 h_{t-1}}$$

where d_w accounts for winter seasonality and $\varepsilon_t \sim N(0,1)$ and the figures in brackets are t-Statistics. For period 2 we proceeded in the same way and got the following results:

– Gasoline:

$$r_t = 0.1354d + h_t^{1/2}\varepsilon_t$$
$$(8.84)$$
$$h_t = 6.99\times10^{-5} + 0.0577\varepsilon_{t-1}^2 + 0.8359h_{t-1}$$
$$(2.10) \qquad (3.00) \qquad (13.57)$$

– Crude oil:

$$r_t = 0.02131d + h_t^{1/2}\varepsilon_t$$
$$(3.29)$$
$$h_t = 2.55\times10^{-5} + 0.07264\varepsilon_{t-1}^2 + 0.8740h_{t-1}$$
$$(2.33) \qquad (5.81) \qquad (27.75)$$

where d accounts for outliers points. Once we had extracted the linear and quadratic dependences, the residuals will retain the true dependence of the variables. So the copula will be adjusted using the residuals of the models above, for period 1 and period 2.

4. Copula modeling

In this section the details of copula modeling are presented. We used the software R version 2.7.2 for Windows with the QRMlib by Alexander McNeil (see McNeil *et al*. (2005)). We also use the Resample Library of Splus.

Our first step to select a copula model begins with the analysis of the dependence between the residuals from econometric models of gasoline and WTI. As pointed by Frees and Valdez (1998), Pearson correlation coefficient is a good choice when the dependence between the variables is linear. So it is always a good practice to use also a nonparametric dependence measure like Kendall's tau. In Table 2 one can observe that there is a considerable positive dependence between gasoline and WTI residuals. It seems that the dependence has increased from the first period to the second. In both periods, as could be expected, Pearson's coefficient indicates a greater dependence.

Table 2: Correlation coefficients

Period	**Pearson coeficient**	**Kendall's τ**
First	0.708	0.549
Second	0.743	0.588

Since we have two periods of data, it will be interesting to check the uncertainty of these dependence measures in both periods. For this purpose we will use the bootstrap method (see Davidson and Hinckley (1997)), a resampling procedure that is very useful to construct confidence intervals when there is no a direct procedure. Because of the computational cost of the Kendall's τ measure, only one thousand bootstrap samples were used to obtain the percentile confidence intervals (95%). Table 3 presents the results. From these confidence intervals one can observe that there is an intersection between both periods. A permutation test for the difference of Pearson correlation coefficient on both samples will indicate if there is a statistically significant difference (see Pesarin (2001)). The results from two thousand permutations are presented in Table 4. The results showed no evidence of a statistically significant difference between both periods.

Table 3: Confidence intervals

Period	**Pearson coeficient**	**Kendall's τ**
	Percentile	**Percentile**
First	(0.681 0.738)	(0.532 0.567)

Second	(0.703 0.782)	(0.562 0.616)

Table 4: Statistics for the difference of Pearson coefficient

Statistic	Observed	Mean	Std. Dev.	Alternative	p-value
Pearson	-0.03492	0.0007153	0.03176	Two sided	0.2679

The next step is to transform each pair of observation $(X_1, Y_1), \ldots, (X_n, Y_n)$ into its rank based representation. The pairs (R_{xi}, R_{yi}), represent pseudo observations from the underlying copula that characterizes the dependence structure, were calculated by

$$R_{xi} = \frac{rank(X_i)}{n+1} \qquad and \qquad R_{yi} = \frac{rank(Y_i)}{n+1}$$

These pairs of residuals are shown in Figure 1 for both periods (period 1 - P1 and period 2 - P2) and they confirm a positive dependence that was previously verified through the computation of Pearson's and Kendall's measures. The pattern is similar and the perceptible difference is due to the number of points: 3419 in period 1 and 1187 in period 2.

Figure 1: Normalized residuals for period 1 and period 2

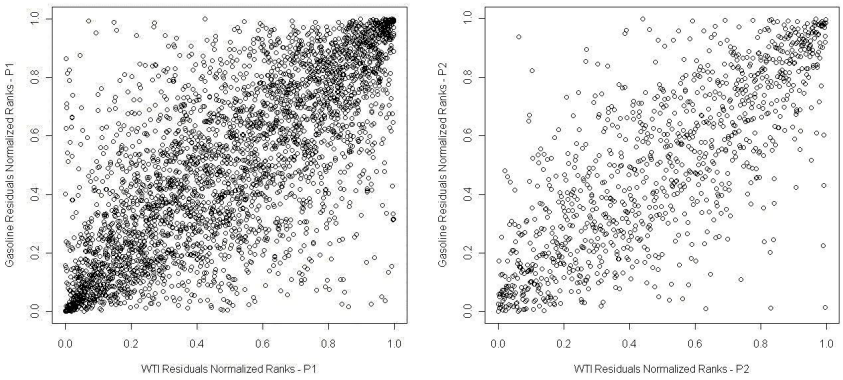


Table 5 contains the families of copulas that will be used to model the dependence between the residuals in this study. Families N.12 and N.14 received their names from Nelsen (1999). The domain of the dependence parameter shows that half of them are only suited for positive dependences. The parameter estimation will be carried out through the maximization of log-likelihood function using the canonical maximum likelihood (CML) process. As shown in Accioly (2005) this method avoids problems in model selection due to improper specification of marginal distributions. Genest *et al*. (1995) proposed such estimation procedure that is appropriate when one does not want to specify any parametric model to describe the marginal distributions. In these cases one could use a nonparametric estimate for the marginal, so the inference about the dependence parameter α should be margin-free.

Table 5: Copulas Families used

Family	$C(u,v)$	$\alpha \in$
Clayton	$(u^{-\alpha}+v^{-\alpha}-1)^{-1/\alpha}$	[-1, ∞) \ {0}
Gumbel	$\exp(-[(-\ln u)^{\alpha}+(-\ln v)^{\alpha}]^{1/\alpha}$	[1, ∞)
Frank	$-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}\right)$	(-∞, ∞) \ {0}
N.12	$\left(1+[(u^{-1}-1)^{\alpha}+(v^{-1}-1)^{\alpha}]^{1/\alpha}\right)^{-1}$	[1, ∞)
N.14	$\left(1+[(u^{-1/\alpha}-1)^{\alpha}+(v^{-1/\alpha}-1)^{\alpha}]^{1/\alpha}\right)^{-\alpha}$	[1, ∞)
Gaussian	$N_{\alpha}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)$	[-1,1]
t	$T_{\alpha,\gamma}\left(T_{\gamma}^{-1}(u),T_{\gamma}^{-1}(v)\right)$	[-1,1]
Plackett	$[1+(\alpha-1)(u+v)-$ $\left[\{1+(\alpha-1)(u+v)\}^2-4uv\alpha(1-\alpha)\right]^{1/2}]/\{2(\alpha-1)\}$	(0,∞) \ {1}

To construct the likelihood function our main concern was the parametric representation of the copulas, specifically the copula density. Given a random sample $\{X_{1k},X_{2k}\}$:k=1,…n} from a distribution $F_{\alpha}(x_1,x_2)=C_{\alpha}(F_1(x_1),F_2(x_2))$, the usual procedure is the selection of parameter α that maximizes the pseudo log-likelihood function:

$$L(\alpha) = \sum_{k=1}^{n} ln\left[c_\alpha(F_1(x_{1k}), F_2(x_{2k}))\right]$$

Once we estimated the dependence parameter α through the procedure above for each copula model, we need to select the best model. This is addressed using the AIC criteria given by

$$AIC = -2L(\alpha) + 2P$$

where P is the number of estimated parameters and works as a penalty. This method is one of the most used in model selection, see for example Frees and Valdez (1998). Burnham and Anderson (2002) recommended the computation of AIC differences, $\Delta_i = AIC_i - AIC_{min}$, over all candidates. For a specific model, the larger Δ_i is, less likely it is to be the best model. They also suggested that a better interpretation can be obtained with Akaike weights, given by

$$w_i = \frac{exp(-0.5\Delta_i)}{\sum_{r=1}^{R} exp(-0.5\Delta_r)}$$

Given R candidates, w_i is the weigh of evidence that model i is the best model. To ensure that the model is properly selected we will use the bootstrap procedure described in Burnham and Anderson (2002). Using this methodology we can verify how different samples can affect the model selection.

Table 6 presents the MLE (Maximum Likelihood Estimation) results for the first period. It also shows the computation of Δ_i and w_i. From the results one can observe that t copula is the best possible model. The difference between t copula and the others is so huge that the Akaike weights of the others can be considered as zero. Due to numerical problems the Gaussian copula was not estimated.

Table 6: MLE results for the first period

Family	$\hat{\alpha}$	$L(\alpha)$	Δ_i	w_i
t	0.75	1456	0	1,0
Plackett	15.34	1426	58	0

N.14	1.66	1394	121	0
N.12	1.43	1347	215	0
Frank	6.81	1339	231	0
Gumbel	2.07	1297	315	0
Clayton	1.61	1110	689	0

As mentioned before we confirmed the conclusion above using bootstrap. After each bootstrap sample we evaluated the AIC to verify the best model. Because of computational cost we limited this analysis to one thousand samples. Table 7 presents the results. Although the t copula model is sometimes worse than Plackett copula, there is no doubt that the former is the best possible model in this group.

Table 7: Bootstrap results for the first period

Family	% Selected through min AIC
t	95,1%
Plackett	4,9%

For the second period the results are shown in Table 8. One can observe that the Plackett copula is the best possible model. The difference between the Plackett copula and the others is not as big as the results of the first period, but we still can consider the Akaike weigths from all other models approach zero.

Again we used the same analysis as before. Table 9 shows the results from bootstrap analysis. One can observe that the results are not as strong as in the first period. The Plackett copula had been overcome by t copula and Frank copula 30% of the times. Comparing the dependence between period 1 and period 2, we can say that the increase in dependence combined with other (undetected) factor has led to the change in the order of model selection.

Table 8: MLE results for the second period

Family	$\widehat{\alpha}$	$L(\alpha)$	Δ_i	w_i
Plackett	17.24	502	0	0,99978
t	0.78	495	17	0,00017
Frank	7.66	492	20	0
N.14	1.79	474	56	0
Gumbel	2.20	459	87	0
Gaussian	0.76	458	89	0

N12	1.49	448	109	0
Clayton	1.62	359	287	0

Table 9: Bootstrap results for the second period

Family	% of Selected through min AIC
Plackett	70,7 %
t	21,9 %
Frank	7,4 %

Figure 2 presents simulated pairs and the original pairs for period 1. It is clear that the dependence behavior is well represented by t copula. We also present the bidensity plot of t copula with dependence parameter 0.754 and 4 degrees of freedom, corresponding to the dependence between gasoline and WTI residuals in the first period.

Figure 3 presents the simulated and the original residuals for period 2. It is clear that the behavior of the dependence between gasoline and WTI has a good representation by Plackett copula. The bidensity of Plackett copula with parameter 17.24 is also presented.

Analyzing the results for gasoline in both periods, one can observe the strong dependence on the tails. The extreme events observed in the variable return is responsible for the fat tails of these distributions, this is a stylized fact of financial time series. Our result is in accordance with the fact that t distribution and fat tail distributions are suited for modeling financial series.

Figure 3: Original and simulated residuals and bivariate t copula

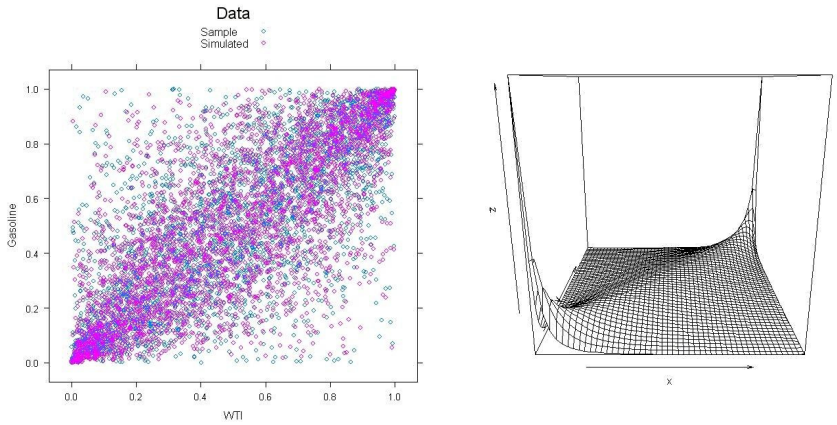
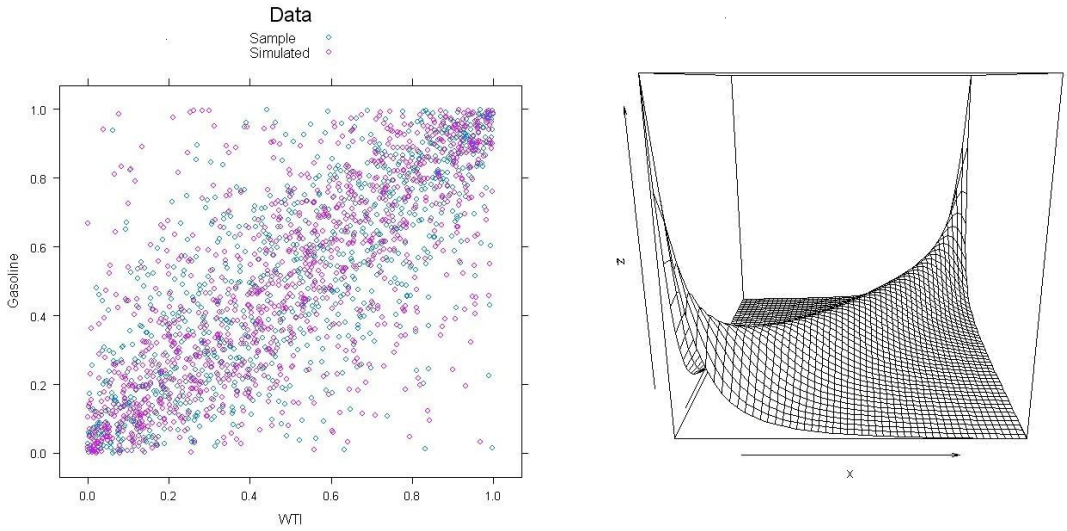


Figure 4: Original and simulated residuals and the bidensity Plackett copula



5. Conclusions

This paper analyzed the dependence of oil and gasoline prices. The main goal was to establish the real dependence instead of the classical approach of linear dependence. Price dependencies have important implications on oil industry mainly in the decision making process for investments. In each series we adjusted an AR-GARCH model, filtering the linear and quadratic dependences. Then we analyzed the dependence through the residuals of these models in each period. We found for the first period that the dependence is well represented by the bivariate t copula. The bootstrap analysis showed that in 95% of the cases the t copula is the best choice. In the second period the behavior of prices changed and the Plackett copula best described the behavior. The bootstrap analysis confirmed that it is the best choice but not as good as the t copula in the first period. It can be observed in the second period that there is an increase in the dependence and this fact is in accordance with the Pearson coefficient. This increased dependence in the second period can be one of the reasons that the competing models (t copula and Plackett copula) have changed the selection order. One natural direction to extend this analysis is the inclusion of other refined products such as

diesel and the natural gas. Moreover, further research can be done on pricing derivatives of oil industry.

References

ACCIOLY, R. M. S. Modeling dependencies with copulas: contributions to uncertainty analysis of exploration and production projects, **Doctoral Thesis**, COPPE, UFRJ, 2005.

ACCIOLY, R. M. S., CHIYOSHI, F. Y. Modeling dependence with copulas: a useful tool for field development decision process, **Journal of Petroleum Science and Engineering**, 44, 83-91, 2004.

BURNHAM, K. P., ANDERSON, D. R. **Model selection and multimodel inference: A practical information-theoretic approach**, 2nd ed., Springer-Verlag, 2002.

CHERUBINI, U., LUCIANO, E., VECCHIATO, W. **Copulas methods in finance**, Wiley, New York, 2004.

DAVIDSON, A. C., HINCKLEY, D. V. **Bootstrap methods and their application**, Cambridge Press, 1997.

EMBRECHTS, P. Copulas: A personal view. **Working paper**, Department of Mathmatics, ETH Zurich, Switzerland, 2008.

FRESS, E. W., VALDEZ, E. A.. Understanding relationships using copulas, **North American Actuarial Journal** 2(1), 1:25, 1998.

GENEST, C., FAVRE, A. C. Everything you always wanted to know about copula modeling but were afraid to ask, **Journal of Hydrologic Engineering** 12, 347-367, 2007.

GENEST, C., GHOUDI, K., RIVEST, L. P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, **Biometrika**, vol 82, issue 3, 543-552, 1995.

GRÉGOIRE, V., GENEST, C., GENDRON, M. **Using copulas to model price dependence in energy markets**, Energy risk 58-64, 2008.

JOE, H. **Multivariate models and dependence concepts**. Monographs on Statistics and Applied Probability 73, Chapman \& Hall/CRC, 1997.

McNEIL, A., FREY, R., EMBRECHTS, P. **Quantitative Risk Management: Concepts, Techniques and Tools**, Princeton University Press, 2005.

MENDES, B. V. M., SOUZA, R. M. **Measuring financial risks with copulas**, International Review of Financial Analisys 13, 27-45, 2004.

MURAT, A., TOKAT, E. **Forecasting oil price movements with crack spread futures**. Energy Economics, Forthcoming, 2008.

NELSEN, R. B. **An introduction to copulas**, Springer, New York, 1999.

PESARIN, F. P. **Multivariate permutation tests with applications in biostatistics**, Wiley, Chichester, 2001.

SKLAR, A. Fonctions de repárticion à n dimensions et leurs marges, vol 8. **Publications de l'Institut de Statistique de l'Université de Paris**, Paris, 229-231, 1959.