GLOBAL NATURAL GAS MARKET: REALITY OR EXPECTATION?

Marcela Ferreira Ramos*++ Amanda Pimenta Carlos*
Rafael Martins de Souza*

* Fundação Getúlio Vargas (Getulio Vargas Foundation – Graduate School of Economics), Rio de Janeiro, Brazil.
++ Corresponding Author: Marcela Ferreira Ramos, Phone: +55(21) 9295 6407, Email: marcela@fgvmail.br

May 6, 2009

Abstract

The natural gas market all over the world has been suffering tremendous changes for the past decades, not only because of market liberalization, but also due to industry restructuring and the strong expansion of liquefied natural gas market (LNG) which brought the missing mechanism to achieve the integration of the regional markets, or, perhaps, the formation of a unique global market for natural gas. Previous related literature utilizes the Cointegration approach and Kalman Filter to test the Law of One Price hypothesis within the global natural gas market. In this article, we use Error Correction Model (ECM), Time Varying Parameters Error Correction Model (TVP-ECM) and the Copula Analysis to examine the dependence between regional markets. Once verified cointegration relationships between the natural gas price series, we estimate an ECM for each pair of prices and, using the Kalman Filter Approach, we ran a TVP-ECM as a way to model the short run dynamics of the pair of prices that presented some structural break. After that, we modeled the short run dependence with the Copula Analysis. This is an innovation of this article: the joint use of Cointegration Technique with Copulas. Basically, the Copulas are cumulative distributions functions (CDF) quite suitable to deal with dependence in multivariate data. The application of Copulas was motivated by the fact that neither the VECM nor the TVP-ECM was enough to account for all the dependence of the series in the short run. We use monthly data of natural gas prices in United States, Europe (United Kingdom, Belgium) and Japan, from 2001 to 2009. There is no statistical evidence of cointegration relationship between the Japanese market and the others, while the American and European markets present price convergence in the long run. The Copulas obtained for the pairs Zeebrugge-NBP and NBP-Henry Hub indicate that these markets move together more intensively during extreme periods. Finally, for the pair Zeebrugge-Henry Hub the Copula fitted shows that there is no dependence at the extremes of the bivariate distribution.

JEL Codes: L9, Q4, C22

Keywords: Natural Gas, Cointegration, Kalman Filter, Error Correction Model, Copula Analysis.
1 Introduction

The global natural gas market has been suffering a series of changes over the past decades, both institutional and economical. We can divide the world market into three large regional markets: North America, Europe and Asia/Pacific. Liberalization in the American market occurred during the 1970’s, but deregulation only occurred in the second half of the 1980’s, thereby allowing free access to the natural gas transportation network. In Continental Europe the opening of the market only occurred in 2003, almost 20 years after the brake up of the market monopoly. In Asia/Pacific the natural gas market is still regulated, strongly related to oil prices through long term contracts.

The observed changes in the American and European markets allowed the development of spot market on the two continents, and also the surge of short term negotiation and non-dedicated contracts, which provoked the restructurization of the regional markets. Henry Hub in United States, National Balancing Point (NBP) in United Kingdom and Zeebrugge Hub, in Belgium, have become reference points in their respective markets, allowing total transparency.

International natural gas trade has increased a lot, mainly because of the development of the liquefied natural gas market (LNG). The LNG provides the link that was missing among the regional markets, once it makes possible arbitrage between them. This results in the markets becoming more and more integrated.

During the past few years, because of the main changes in the natural gas industry, markets have been constantly the focus of study in literature. L’Hegaret, Siliverstovs and Hirschhausen (2005) utilize cointegration analysis to evaluate integration between markets in Europe, North America and Japan and conclude that the American market is not totally integrated and the European and Japanese markets are integrated, but still indexed to oil prices. Within the three markets, the authors demonstrated that, in the long run, the American market will still be disconnected from the others.

Neumann and Hirschhausen (2006) analyze the convergence between North American and European markets using the Kalman Filter technique. Using daily data for the Henry Hub in the United States, and the NBP in the United Kingdom, the authors achieve empirical evidence that natural gas spot prices in the Atlantic Basin are converging towards the Law of One Price.

Other articles investigate convergence inside each specific market. For example, the American market has been analyzed in King and Cuc (1996), Serletis (1997) e Cuddington and Wang (2006). The European market was the subject of analysis in Asche, Osmundsen and Tveteras (2002) and Neumann, Siliverstovs and Hirschhausen (2005).

In this article, we start making a quick analysis of the characteristics among each regional market (North America, Europe and Japan) and the LNG market. Next, we use Cointegration technique to investigate price convergence between the three markets and, once the long run convergence is verified, we estimate Error Correction Models (ECM) for each pair of prices. For the one which present a sign of some structural break, we apply the Kalman Filter to the Error Correction Model to analyze price behavior on the short run. As we have noticed, since these days, related studies on natural gas market integration has only used Cointegration Technique or the Kalman Filter Approach. This is the first paper to use a Time Varying Parameter Error Correction Model (TVP-ECM) to investigate this relation.

Another innovation of this paper is the joint use of Cointegration Technique with Copulas. Basically, as we will introduce in the upcoming sections, the Copulas are cumulative distributions functions (CDF) quite suit-
able to deal with dependence in multivariate data. Embrechts, Lindskog and McNeil (2001), Bouyé, Durrleman, Nikeghbali, Riboulet and Roncalli (2000), Embrechts, McNeil, Straumann (1999a), Embrechts, McNeil, Straumann (1999b), Mendes and Souza (2004) are good examples of applications in financial literature. Its use in the financial literature is growing since late 1990’s, mainly because of the perception that the framework widely used to deal with dependence relies on the Gaussian assumption. The decision to use Copulas was motivated by the fact that neither the VECM nor the TVP-ECM were enough to account for all the dependence of the series in the short run, as we are going to expose. Besides that, more than quantifying correctly, the Copula Technique is suitable to model the “dependence structure” of the series.

The remainder of the article is structured as follows: in the next section we characterize each market. In section 3 we present the related literature. In section 4 we specify our model. Section 5 presents the estimation procedure and data details; in section 6 we have the results and in section 7 the conclusion.

2 Regional Market Characteristics

2.1 The North American Market

The North American natural gas market deregulation occurred in the mid 1980’s in Canada and United States. It had great impact on the natural gas market once it provided the development of a spot market, hubs, a future market, a secondary market for transport capacity, etc. All these changes generated a huge increase not only in natural gas production, but also in its consumption within the region.

Due to competition that took place in the market, natural gas prices are now determined simply by supply and demand. In this new scenario, the main factors that can cause volatility in North American prices are: inventory fluctuations; abrupt demand variation and physical infrastructure damages due to climate (hurricanes, extreme weather, etc.); market expectations; lack of flexibility in market balance (supply inline with demand).

On the other hand, we can verify the fast development of a short run market and, also, a spot market for natural gas in the Atlantic Basin because of LNG. It gives more flexibility to natural gas supply and makes the interconnection easier between North American and other markets, primarily the European market, thereby increasing the possibility of arbitrage between them.

2.2 The European Market

The United Kingdom is a particular market compared to other European countries because there exists a competitive market and a spot market of huge importance. In Continental Europe, this is still not true, prices are still linked to oil prices through long term contracts. However, over the past few years this situation has been progressively changing and, nowadays, in the European continent, one can already find contracts indexing the natural gas price to the spot price in the United Kingdom market (NBP). This link between the English and the continental markets is due to the openness of the “Interconnector”, a pipeline which connects the United Kingdom to Zeebrugge in Belgium, at the end of the 1990’s. Zeebrugge is directly connected to the pipelines in Netherlands, Germany and France.
2.3 The Japanese Market

The Asian market is almost totally supplied by LNG; Japan alone imports approximately 50% of the whole world’s production of LNG (L’Hegaret et al. 2004). Even so, the country does not have a vast pipeline network, but, on the other hand, it has the largest number of regasification terminals in the world. However, natural gas has a only minor participation in the Japanese energy matrix (13% of total energy consumption), while oil represents 50% of it. Because of the great concern about supply warranty, this market is characterized by long term contracts in which prices are indexed to oil prices, thereby establishing a weak dependence on the spot market.

2.4 The LNG Market

The three markets that we have just analyzed are the main LNG consumers in the world nowadays, where 70% of trade occurs on the Asia/Paciﬁc region and the other 30% on the Atlantic Basin. The volume of LNG traded has increased considerably during the last decade, going from 108MMtpa in 2000 to an estimated 247MMtpa in 2010 (Wood Mackenzie, 2008).

Spot market transactions of LNG suﬀered a huge boom in the early 2000’s. This growth brought the missing link to integration between the regions, mainly in the Atlantic Basin, where arbitrage practices can already be seen (Neumann et Hirschhausen, 2006). For the next few years, a huge expansion in import capacity in the United States and also in Europe is expected which will provide an even stronger interaction between the two regions.

Nowadays, LNG trade represents 20% of the whole natural gas trade in the world and it is expected that, over the next few years, due to the development of new liquefaction and regasification plants, this will achieve 30%. The LNG supply of new projects will allow even greater gains in ﬂexible supply.

3 Related Literature

There are some related articles that analyze the integration of natural gas regional markets and the price convergence between them. Now, we are going to review some of the main studies which utilized similar techniques to make these analyses.

Neumann and Hirschhausen (2006) use the Kalman Filter approach to investigate price dynamics between Europe and United States from 1999 until 2006. Utilizing daily spot prices for both natural gas and oil in both respective markets they found clear evidence that prices are converging in the direction of the Law of One Price. However, when they deduce the eﬀect of oil spot prices on natural gas prices and apply again the Kalman Filter to the new series, it seems that prices do not interact that much.

L’Hegaret, Siliverstovs and Hirschhausen (2004) tests the degree of natural gas integration throughout Europe, North America and Japan, between the period of mid 1990’s until 2002. The Principal Component Analysis and Cointegration procedures show a “high level of integration within the European/Japanese and North American markets and that the European/Japanese and the North American markets are connected to a much lesser extent”.

Neumann, Siliverstovs and Hirschhausen (2005) test whether the prices of natural gas in the major European trading spots – United Kingdom (NBP), Belgium (Zeebrugge) and Germany (Bunde) – are converging. They use
daily prices from March 2000 until February 2005 and applied the time-varying Kalman Filter analysis to each pair of prices. They find evidence of full convergence between UK and Belgium. On the other hand, it clearly states that there is not one single Continental European European natural gas market as prices at Zeebrugge and Bunde are not converging, even though these two locations are also connected by a pipeline.

Asche, Osmundsen and Tveteras (2002) examine whether or not there is integration within the German market by using time series analysis of Norwegian, Dutch and Russian gas export prices to Germany between 1990 and 1998. Results from cointegration tests show that “the different border prices for gas to Germany move proportionally over time, indicating an integrated gas market”. However, the Russian gas had been sold at lower prices than Dutch and Norwegian gas.

King and Cuc (1996) used Kalman Filter to measure the degree of price convergence in North American natural gas spot markets from mid 1980’s to mid 1990’s. The results indicate that price convergence in natural gas spot markets had increased significantly since the price deregulation of the mid 1980’s, but a split between prices in the east and west coasts of the United States still existed.

4 The Model

In this section we are going to describe our model. The use of copulas deserves special attention, since its application to natural gas literature is being introduced in this paper. We also will describe our estimation methodology.

4.1 Copulas

The copula concept was first used in the statistics literature in 1959 with the seminal work of Sklar (1959). Since the end of 1990 decade there exist an increasing interest in copula models applied to multivariate finance data. One of the main reasons for this recent interest is the larger and larger understanding that the traditional framework, used to deal with dependence, strongly relies on the Gaussian hypothesis, which has shown to be inadequate to deal with the stylized facts of financial data. There is a plenty of work on copulas applied to financial data, good examples are: Embrechts, Lindskog and McNeil (2001), Bouyé, Durrleman, Nikeghbali, Riboulet and Roncalli (2000), Embrechts, McNeil, Straumann (1999a), Embrechts, McNeil, Straumann (1999b), Mendes and Souza (2004), among others. In this section we present a very short introduction to Copulas. Very good introductory references are Nelsen (1999) and Joe (1997).

Definition 1 A copula is a cumulative distribution function of a random vector in \( \mathbb{R}^p \) with \( U \sim (0,1) \) marginal distributions. Equivalently, a copula is any function \( C : [0,1]^p \rightarrow [0,1] \) with three properties:

1. \( C(x_1, ..., x_p) \) is increasing in each one of its components \( x_i, i = 1, ..., p; \)
2. \( C(1, ..., 1, x_i, ..., 1, ..., 1) = x_i \) \( \forall \ i \in \{1, ..., p\}, x_i \in [0,1]; \)
3. For all \( (a_1, ..., a_p), (b_1, ..., b_p) \in [0,1]^p \) with \( a_i \leq b_i \), we have:

\[
\sum_{i_1=1}^{2} \cdot \sum_{i_p=1}^{2} (-1)^{i_1+\cdots+i_p} C(x_{1i_1}, ..., x_{pi_p}) \geq 0, \tag{1}
\]
in which \( x_{j1} = a_j \) and \( x_{j2} = b_j \) for all \( j \in \{1, ..., p\} \).

As the definition tells by itself, a copula is just a cumulative distribution function (CDF) in which each univariate marginal have the uniform distribution in \((0, 1)\). Sklar(1959) Theorem is one of the most important results about copulas, as follows:

**Theorem 2** Let \( F \) be a \( p \)-dimensional cumulative distribution function with marginals \( F_1, \ldots, F_p \). Then there exist a copula \( C \) such that

\[
F(x_1, \ldots, x_p) = C(F_1(x_1), \ldots, F_p(x_p))
\]

(2)

If \( F_1, \ldots, F_p \) are all continuous, then \( C \) is unique; i.e., \( C \) is determined in \( \text{Im} F_1 \times \ldots \times \text{Im} F_p \). Reversely, for a copula \( C \) in continuous marginals \( F_1, \ldots, F_p \) a distribution function \( F \) is a \( p \)-dimensional distribution function with marginals \( F_1, \ldots, F_p \).

Thus, this important theorem states that vector \((X_1, \ldots, X_p)\) of random variables with CDF \( F \) has a copula \( C \) and that if \((X_1, \ldots, X_p)\) is a vector of continuous random variables, this copula is unique. The following Proposition states about another important property of the copula function:

**Proposition 3** If \((X_1, \ldots, X_p)\) has copula \( C \) and \( T_1, \ldots, T_p \) increasing are functions, so \((T_1(X_1), \ldots, T_p(X_p))\) has also copula \( C \).

Then, when working with a pair of continuous random variables and applying over them an increasing function, the new transformed variables will have the same copula. This result lead us to think about the copula function as the “dependence function” of the variables in the random vector.

As \( C \) is a continuous CDF, we can, obviously, get its probability density function. Taking its total derivate, we have:

\[
c(u_1, \ldots, u_p) = \frac{\partial^p C(u_1, \ldots, u_p)}{\partial u_1 \partial u_2 \cdots \partial u_p}.
\]

\( c(u_1, \ldots, u_p) \) is very usefully when we need to write random vector \((X_1, \ldots, X_p)\) pdf as follows:

\[
f(x_1, \ldots, x_p; (\theta_c, \theta_1, \ldots, \theta_p)) = c(F_1(x_1), \ldots, F_n(x_p); \theta_c) \prod_{i=1}^p f_i(x_i; \theta_p).
\]

(3)

If we consider the loglikelihood for the vector \( \theta \) of parameters constructed based on the pdf above, we get:

\[
\log L(\theta_c, \theta_1, \ldots, \theta_p; x_1, \ldots, x_p) = \log (f(x_1, \ldots, x_p; \theta_c, \theta_1, \ldots, \theta_p)) \prod_{i=1}^p f_i(x_i; \theta_p)) = \log (c(F_1(x_1), \ldots, F_n(x_p); \theta_c) + \sum_{i=1}^p \log (f_i(x_i; \theta_p)).
\]

(4)

Therefore, when working with copulas, the multivariate probability model can be seen as the one that clearly describes the marginal behavior of each of the univariate random variables in the random vector and the joint behavior of them, separately. By the statistical point of view, the following result may help to state these ideas:
Proposition 4 Let $X$ be a random variable with cumulative distribution function $F$ e let $F^{-1}$ be its inverse, i.e., $F^{-1}(\alpha) = \inf\{x|F(x) \geq \alpha\}, \alpha \in (0,1)$. If $F$ is a continuous function i.e., then the random variable $F(X)$ has uniform distribution on the $(0, 1)$, i.e., $F(X) \sim U(0, 1)$.

This result is very important because it opens two streams of estimation of equation 3: full estimation and two step estimation. The full estimation consists on the joint estimation of all the parameters of equation 3 in one step. The two-step estimation consists of, firstly, estimate the parameters of the marginal models and, secondly, using the “uniformized” observations, estimate the copula function. Here, by “uniformized” observation we mean the original observation transformed by the application of estimated CDF, i.e., $u_{j,t} = F(x_{j,t}), j = 1, \ldots, p$ and $t = 1, \ldots, n$. We will discuss about estimation in the upcoming section Model Estimation. Following, we present the copulas functions that we are going to work in this paper.

4.1.1 Some Copulas Functions

As long as the concept was presented, we introduce some important copulas families that are going to be used in this work. Each of these families have particular characteristics. Although we are not able to go further and detail all their properties, the reader is stimulated to search for additional information on the cited references.

Elliptical Copulas These copulas are characterized to have their dependence given by elliptical forms. The first of them is the Gaussian or Normal Copula. This copula is the copula associated to the multivariate normal distribution. As in this paper we will deal only with bivariate data, we write its bivariate version:

$$C_{\rho} = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v))$$

(5)

where $u$ and $v \in [0,1]$ and $\Phi$ denotes the Normal cumulative distribution function.

Another very important elliptical copula is the t-Student Copula, which form is given by:

$$C_{\nu, R}() = t_{\nu, R}^{n}(t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_n))$$

(6)

where $R_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ for $i, j \in \{1, \ldots, n\}$ and where $t_{\nu, R}^{n}$ denotes the distribution function of multivariate $t$-student distribution with $n$ $t_{\nu}$ marginals.

A very important difference between the Gaussian and the t-Student Copula is about their upper tail dependence,

$$\lambda_U = 2 \lim_{x \to \infty} P(X_2 > x|X_1 = x),$$

(7)

and their lower tail dependence

$$\lambda_L = 2 \lim_{x \to -\infty} P(X_2 < x|X_1 = x).$$

(8)

$\lambda_U$ and $\lambda_L$ are important measures. They measure, as their own names suggest, dependence at the extremes of the distribution’s supports. These measures are constant given one copula function, i.e., given a copula function, these values are given independently of the marginals distributions. One of the most important properties of elliptical copulas is that their structure of dependence is symmetric. This implies that $\lambda_U = \lambda_L$. 
An important difference emerges when we use Gaussian and t-Student Copulas: the Normal Copula has $\lambda_L = \lambda_U = 0$, which implies independence at the extremes of the bivariate distribution. On the other hand, the t-Student Copula usually has $\lambda_L > 0$ and $\lambda_L$ increases with $\rho$ and decreases with $\nu$.

**Archimedean Copulas** Archimedean copulas are an important family of copulas, which have a simple properties such as associativity and have a variety of dependence structures. They have closed-form solutions and are not derived from the multivariate functions using Sklar’s Theorem. One particularity of a $n$-dimensional Archimedean copula is

$$C^A(u_1, \ldots, u_n) = \Psi^{-1}\left(\sum_{i=1}^{n} \Psi(u_i)\right)$$

$\Psi$ is known as generator function. A generator satisfies the following properties:

$$\Psi(1) = 0; \quad \lim_{x \to 0} = \infty; \quad \Psi'(x) < 0; \quad \Psi''(0) > 0.$$  \hspace{1cm} (10)

Some Archimedean copulas are used in this paper. The first of them is the Product Copula, that is the copula of independence between the variables. Its density function, that is unity on its support, and its generator function are given as follows:

$$C(u, v) = uv, \quad \Psi(x) = -\ln(x);$$

$$C(u_1, \ldots, u_n) = \Psi^{-1}\left(\sum_{i=1}^{n} \Psi(u_i)\right)$$

Among the Archimedean copulas, we are going to use three models considering different structures of dependence. The first of them is the Gumbel Copula, given by:

$$C^G_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right), \quad \Psi = (-\ln(x))^\alpha;$$

$$C^G_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right), \quad \Psi = (-\ln(x))^\alpha;$$

The Gumbel Copula is such that for $\theta > 0$, $\lambda_U^G = 2 - 2^{1/\theta}$, i. e., has upper tail dependence. Its $\lambda_U^G = 0$. The Clayton Copula is the second density we are going to study:

$$C^C_\theta(u, v) = (u^\theta + v^\theta - 1)^{1/\theta}, \quad \Psi = x^\theta - 1; \quad \theta \leq 0;$$

It goes on the other way Gumbel Copula does: it has $\lambda_U^C = 0$ and $\lambda_L^C > 0$. The third Archimedean copula is the Frank Copula.

$$C^F_\theta(u, v) = (u^\theta + v^\theta - 1)^{1/\theta}, \quad \Psi = \ln\left(\frac{e^{\alpha x} - 1}{e^{\alpha x} - 1}\right)$$

$$C^F_\theta(u, v) = (u^\theta + v^\theta - 1)^{1/\theta}, \quad \Psi = \ln\left(\frac{e^{\alpha x} - 1}{e^{\alpha x} - 1}\right)$$

The Frank copula also satisfies $\lambda_U^C = \lambda_L^F = 0$.

Figure 1 presents some plots of the copulas, for a better understanding of the models used in this paper. Comparing the graphs on the first row, it is easily noticed that the edges of the t-Student Copula are sharper than the ones in the Gaussian Copula. This helps to understand the difference in the tail dependence in these two families. In the second row, we present the Archimedean copulas. Gumbel Copula has an edge in the upper corner, Clayton Copula, in the lower and the Frank Copula has no sharp edges.
4.2 Model Specification

As stated before, the cointegration technique has been widely applied to study the link of the different global natural gas markets. In these works, the underlying VECM model for the cointegrated series is built and estimated using the Gaussian framework. But, natural gas prices are, essentially, financial data. One of the most important stylized facts about financial data is their non-Gaussian distributions. In the univariate case, this non-normal behavior is mainly linked to excess of kurtosis. In the multivariate case, the lack of normality is associated to different structures of dependence of the ones generated by multivariate normal distribution. Besides these important stylized facts, most of the work on cointegration relies on the traditional Gaussian framework, ignoring this important aspect of gas prices.

Therefore, in order to build a VECM to investigate the existence of a global natural gas market, we are going to adapt this technique to identify and model the possible non-normality that natural gas price data may have. To let our ideas clear, consider the conditional joint probability density function (pdf) of the series of log returns of natural gas prices, namely, \( f(\Delta \ln p_{1,t}, \Delta \ln p_{2,t}; \theta_{I_{t-1}}) \). As presented in the previous section, the copula approach is going to have an important role in dealing with the dependence structure of the error terms. Using the proposed factorization, we can write our probability model as follows:

\[
f(\Delta \ln p_{1,t}, \Delta \ln p_{2,t}; \theta_{I_{t-1}}) = c(F(\Delta \ln p_{1,t}|I_{t-1}), F(\Delta \ln p_{2,t}|I_{t-1}); \theta_1) f_1(\Delta \ln p_{1,t}; \theta_{I_{t-1}}) f_2(\Delta \ln p_{2,t}; \theta_{I_{t-1}}) .
\]

(15)

where \( c(F(\Delta \ln p_{1,t}|I_{t-1})) \) is the copula function, \( \theta \) is the vector containing all the parameters of the conditional
model, \( \theta_i \) is the vector of the univariate marginal \( i \). As we want to estimate the parameters of the equation 15, we may write the conditional log likelihood function, as follows:

\[
l(\theta; \Delta \ln p_{1,t}, \Delta \ln p_{2,t}) = l_c(\theta_c; F(\Delta \ln p_{1,t} | I_{t-1}), F(\Delta \ln p_{2,t} | I_{t-1}))) + l_1 (\theta_1; \Delta \ln p_{1,t}) + l_2 (\theta_2; \Delta \ln p_{2,t}) \quad (16)
\]

Equation 16 helps to let clear our aim. Using the factorization above, it is possible to split the multivariate model in two univariate model and one "dependence model". Moreover, this suggests that we can make use of this factorization in the estimation process. Then, our complete model can be written as stated in the equation 17:

\[
\begin{align*}
\Delta \ln p_{1,t} &= \beta_{01} + \beta_{11} \epsilon_{t-1} + \sum_{i=1}^{p} \gamma_{1,i} \Delta p_{1,t-i} + \sum_{i=1}^{q} \lambda_{1,i,t} \Delta p_{2,t-i} + \epsilon_{1,t}, \epsilon_{1,t} \sim NID(0, \sigma^2_{\epsilon}) \\
\Delta \ln p_{2,t} &= \beta_{02} + \beta_{12} \epsilon_{t-1} + \sum_{i=1}^{p} \gamma_{2,i} \Delta p_{1,t-i} + \sum_{i=1}^{q} \lambda_{2,i,t} \Delta p_{2,t-i} + \epsilon_{2,t}, \epsilon_{2,t} \sim NID(0, \sigma^2_{\epsilon})
\end{align*}
\]  

(17)

where the bivariate joint distribution of \( \epsilon_{1,t} \) and \( \epsilon_{2,t} \) will be completed by the copula model and \( \epsilon_{t-1} \) is the deviation of the long run relationship between \( \ln p_{1,t} \) and \( \ln p_{2,t} \) at time \( t - 1 \).

We had evidence that some VECM coefficients of model 17 vary over time. To deal with this feature, we decided to modify our model to accommodate this result. The solution was to use a Time Varying Parameters Error Correction Model (TVP-ECM) which is a combination of the Kalman Filter Approach (Harvey, 1989) and the VECM stated above. This model was first proposed by Hall (1993). Ramajo (2001) and Li, Wong, Song and Witt (2006) also presented some applications of this model. With non-fixed coefficients, this model is able to capture the long run relationship between the variables and allows a flexible way to deal with some possible non stability of short run adjustments. The measurement equation is stated as follows:

\[
\begin{align*}
\Delta \ln p_{1,t} &= \beta_{1,0,t} + \beta_{1,1,t} \epsilon_{t-1} + \sum_{j=1}^{p} \gamma_{1,j,t} \Delta p_{1,t-j} + \sum_{j=1}^{q} \lambda_{1,j,t} \Delta p_{2,t-j} + \epsilon_{1,t}, \epsilon_{1,t} \sim NID(0, \sigma^2_{\epsilon}) \\
\Delta \ln p_{2,t} &= \beta_{2,0,t} + \beta_{2,1,t} \epsilon_{t-1} + \sum_{j=1}^{p} \gamma_{2,j,t} \Delta p_{1,t-j} + \sum_{j=1}^{q} \lambda_{2,j,t} \Delta p_{2,t-j} + \epsilon_{2,t}, \epsilon_{2,t} \sim NID(0, \sigma^2_{\epsilon})
\end{align*}
\]  

(18)

where the bivariate joint distribution of \( \epsilon_{1,t} \) and \( \epsilon_{2,t} \) will be completed by the copula model and \( \epsilon_{t-1} \) is the deviation of the long run relationship between \( \ln p_{1,t} \) and \( \ln p_{2,t} \) at time \( t - 1 \), \( \beta_{i,j,t}, \ i = 1,2, \ j = 0,1, \gamma_{i,j,t}, \ i = 1,2, \ j = 1,....,p, \ t = 1,...,n, \lambda_{i,j,t}, \ i = 1,2, \ j = 1,....,q, \ t = 1,...,n \). To complete the TVP-ECM model we need to specify the state equation, as follows:

\[
\begin{bmatrix}
\beta_{i,t} \\
\gamma_{i,t} \\
\lambda_{i,t}
\end{bmatrix} =
\begin{bmatrix}
\beta_{i,t-1} \\
\gamma_{i,t-1} \\
\lambda_{i,t-1}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{i,t-1}
\end{bmatrix}
\]

(19)

where \( \beta_{i,t}, \ i = 1,2 \) is a vector containing \( \beta_{i,j,t}, \ j = 0,1, \gamma_{i,t}, \) is a vector containing \( \gamma_{i,j,t}, \ j = 1,....,p, \lambda_{i,t}, \) is a vector containing \( \lambda_{i,j,t}, \ j = 1,....,q \) and \( \bar{\varepsilon}_{i,t-1} \) the vector of error terms \((2 + p + q \times 1)\), which we assume to be \( N(0, \Sigma_{\varepsilon}) \), \( \Sigma_{\varepsilon} \) diagonal.

Given the complexity of the proposed model, we decide to implement a three step estimation procedure: in the first step we estimate and test the long run relationship using equation 20, as follows:

\[
\ln p_{1,t} = \alpha_0 + \alpha_1 \ln p_{2,t} + \epsilon_t, \epsilon_t \sim i.i.d. \ D(0, \sigma^2_{\epsilon}).
\]

(20)
Using the estimated residuals of equation 20, we apply Phillips and Ouliaris (1990) to test cointegration. If cointegration is confirmed, we replace the error terms in equation 17 or 18 to estimate the VECM equations.

Next, we used the residuals of the VECM or TVP-VECM model to estimate the Copula Model. In this step, it is necessary to transform the original residuals into bivariate observations in the \((0, 1)\) interval. This may be done using a model for the residuals or not. In this paper we decide to use the empirical distribution function to obtain the “uniformized” observations. This procedure of estimation, known as semi-parametric, has the important advantage of eliminate possible complex and sometimes unreliable assumptions on the marginal distributions.

5 Estimation

5.1 Data

We used logged monthly natural gas prices of the main trading points in United States, United Kingdom and Belgium, Henry Hub, NBP and Zeebrugge respectively. The data is available at Bloomberg. For Japan we use the weighted average of monthly LNG selling prices from the main suppliers of the country (Australia, Indonesia, Malaysia, Qatar, Abu Dhabi, Brunei, Oman and United States), also available at Bloomberg. The period under consideration is from April 2001 to February 2009, totalizing 95 observations.

In Figure 2 is shown the graph of the four monthly series. One can observe that all of them present a co-movement, increasing until the Sub-Prime Crisis. After november 2008, this series still moved together, but decreasing after that. Now, we are going to examine whether this behavior is or not a consequence of some integration between the regional markets.
Table 1 presents the descriptive statistics of the four natural gas price series. As one can see, NBP has the lowest mean and LNGJP the biggest one. The biggest value reached by a series was of US$ 18.53 per MMBtu, reached by Zeebrugge.

<table>
<thead>
<tr>
<th></th>
<th>HH</th>
<th>NBP</th>
<th>LNGJP</th>
<th>ZBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.238792</td>
<td>5.604731</td>
<td>6.690526</td>
<td>5.730849</td>
</tr>
<tr>
<td>Median</td>
<td>6.026800</td>
<td>4.675329</td>
<td>5.690000</td>
<td>4.787310</td>
</tr>
<tr>
<td>Maximum</td>
<td>14.84230</td>
<td>17.92724</td>
<td>15.12000</td>
<td>18.53347</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.760900</td>
<td>1.007813</td>
<td>3.980000</td>
<td>1.594868</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.574337</td>
<td>3.455316</td>
<td>2.767300</td>
<td>3.362999</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.862907</td>
<td>1.132787</td>
<td>1.467827</td>
<td>1.338481</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.045268</td>
<td>3.730036</td>
<td>4.295450</td>
<td>4.523493</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>16.11444</td>
<td>22.42705</td>
<td>40.75599</td>
<td>37.55333</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000317</td>
<td>0.000013</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>Sum</td>
<td>592.6852</td>
<td>532.4495</td>
<td>635.6000</td>
<td>544.4307</td>
</tr>
<tr>
<td>Sum Sqr. Dev</td>
<td>622.9577</td>
<td>1122.285</td>
<td>719.8471</td>
<td>1063.118</td>
</tr>
</tbody>
</table>

### 5.2 Model Estimation

First of all, we ran unit root tests to verify the stationarity of each series. We used the Dickey and Fuller (1979) test and found non-stationarity in all series of this work, as shown in the next section.

After verifying stationarity, we test for the cointegration relation. There are some alternative procedures to test cointegration between variables: Johansen (1988), Stock and Watson (1988), Engle and Granger (1987) and Phillips & Ouliaris (1990). We decided to use a residual based test, specifically the Phillips and Ouliaris (1990) test, which consists in running an ordinary least squares involving pair of prices to get the cointegration vector and the residuals. The null hypothesis is that there is no cointegration. The choosed test has a superior power performance at least in large samples and slower rates of divergence under cointegration than the others tests (Phillips & Ouliaris, 1990).

But cointegration only states for a long run relationship between the variables. Once we find cointegration, it is necessary to search for a better specification with an Error Correction Model (ECM) in order to account for the long and the short run relationships between the variables. We estimate a VECM for each pair of prices. As not all the pairs presented significant VECM, we, then, test for the presence of structural break in each of them. When there is some indication of structural break, a Time Varying Parameter Error Correction Model (TVP-ECM) is indicated. We used the CUSUM test to account for structural break and only the pair NBP-HH showed some sign of it. Consequently, we ran a TVP-ECM for this pair of prices.
Finally, the VECM’s and the TVP-ECM estimations showed that these were not sufficient models to explain all the short run relationship between the variables. However, the Phillips and Ouliaris test showed that in the long run the variables cointegrate in pairs. Therefore, we decided to make Copula Analysis for each pair of prices, attempting to understand how this dependence works in the short run. Although the estimated Copula Functions explain how the joint movements of each pair of prices happens in the short run, it is still necessary more investigative work to figure out the instruments that are making these joint movements possible.

6 Results

Initially we apply the Augmented Dickey-Fuller test to verify series stationarity, which demonstrate that all the four natural gas prices are non-stationary at 5% of significance. Table 2 shows the unit root test for the log of price series.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Components</th>
<th>( \tau )-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHH</td>
<td>Intercept and Trend</td>
<td>-2.2585</td>
</tr>
<tr>
<td>LNBP</td>
<td>Intercept and Trend</td>
<td>-3.365</td>
</tr>
<tr>
<td>LZBG</td>
<td>Intercept and Trend</td>
<td>-3.1801</td>
</tr>
<tr>
<td>LGNLJP</td>
<td>Intercept and Trend</td>
<td>-2.2824</td>
</tr>
</tbody>
</table>

* H0: Series have a unit root at 5% of significance.

5% Critical Value: -3.45

Although the ADF test stated the non-stationarity of the series, a linear relation between them could be stationary, indicating that there exists a cointegration relationship, meaning that, in the long run, they have a common behavior. In our context, the presence of cointegration shows that the natural gas prices are functioning together, what would be a proxy to validate the Law of One Price.

In the next step of our work we test for cointegration between all the combinations of natural gas prices\(^1\) (Henry Hub, NBP, Zeebrugge and “LNGJP”), using Phillipis & Ouliaris (1990) test. Tests including more than two variables are found inconclusive whether there is or not cointegration. Then, it is not possible to evaluate the Law of One Price in a global extent.

When making pair analysis, we find statistical evidence of cointegration between the logs of Zeebrugge and NBP, NBP and Henry Hub and Zeebrugge and Henry Hub, meaning that there exists a long run relationship among them. The LNGJP does not cointegrate with any of the other series. Main results are shown in Table 3.

\(^1\) We applied the lag length criteria in the VAR estimation preceding the cointegration test. The Schwarz Information Criteria indicated one lag for the three pairs of prices analysed.
Figure 3 shows the graph of each pair of prices where we find cointegration. It shows clearly that the pair NBP-Zeebrugge has a strong relationship. The other two pairs also have a common behaviour, but apparently in a weak way.

In Table 4, we report the three cointegrating equations, all of them statistically significant. These results show that the long run relation is much stronger between the natural gas markets in United Kingdom and Continental Europe. Normalizing the variation in the Belgium market price to 1%, the price in United Kingdom varies 0.75%. When analysing the European markets vis-a-vis the American one, this long run relation is not so strong. The Henry Hub varies less than 0.40% while the European prices varies 1%.
After verifying the Cointegration relation between the series, we now test the existence of an Error Correction Model (ECM) for each pair of prices. First, we worked with the traditional ECM from Engle and Granger (1987), which considers fixed parameters. In Tables 5, 6 and 7 we report the results of the ECM Estimation for Zeebrugge-NBP, NBP-Henry Hub and Zeebrugge-Henry Hub respectively. In each column the dependent variable is the first one that appears in the header. We run the ECM in both directions, attempting to identify which market has influence in the other.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Normalized Coefficients</th>
<th>(standard errors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LZBG</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.314</td>
<td>(0.051)</td>
</tr>
<tr>
<td>LNBP</td>
<td>0.745</td>
<td>(0.048)</td>
</tr>
</tbody>
</table>

LZBG - LNBP

<table>
<thead>
<tr>
<th></th>
<th>LZBG - LNBP</th>
<th>LNBP - LZBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.090 (0.029)</td>
<td>0.007 (0.033)</td>
</tr>
<tr>
<td>Cointegration Residual</td>
<td>-0.222 (0.212)</td>
<td>0.532 (0.231)**</td>
</tr>
<tr>
<td>LZBG (-1)</td>
<td>-0.340 (0.174)*</td>
<td>-0.186 (0.191)</td>
</tr>
<tr>
<td>LNBP (-1)</td>
<td>-0.051 (0.149)</td>
<td>-0.180 (0.163)</td>
</tr>
</tbody>
</table>

1Standard Errors between parentheses.

***1%, **5% and *10% of significance.

According to the results in Table 5, one can see that the VECM is not valid for the pair Zeebrugge-NBP, as the coefficients estimates are not significant. In the Equation where the log of Zeebrugge price is the dependent variable, only the auto-regressive coefficient is significant. And, in the other direction, only the long run component (cointegration residual) is significant.
Table 6 - ECM Estimation for LNBP and LHH

<table>
<thead>
<tr>
<th></th>
<th>LNBP - LHH</th>
<th>LHH- LNBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-0.008 (0.1348)</td>
<td>0.246 (0.095)**</td>
</tr>
<tr>
<td>Cointegration Residual</td>
<td>-0.225 (0.077)***</td>
<td>0.049 (0.054)</td>
</tr>
<tr>
<td>LNBP (-1)</td>
<td>-0.229 (0.104)**</td>
<td>0.030 (0.073)</td>
</tr>
<tr>
<td>LHH (-1)</td>
<td>0.007 (0.075)</td>
<td>-0.140 (0.053)***</td>
</tr>
</tbody>
</table>

1Standard Errors between pharenteses.

***1%, **5% and *10% significance.

Table 7 - ECM Estimation for LZBG and LHH

<table>
<thead>
<tr>
<th></th>
<th>LZBG - LHH</th>
<th>LHH- LZBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.006 (0.119)</td>
<td>0.232 (0.095)***</td>
</tr>
<tr>
<td>Cointegration Residual</td>
<td>-0.237 (0.076)***</td>
<td>0.055 (0.060)</td>
</tr>
<tr>
<td>LZBG (-1)</td>
<td>-0.318 (0.101)***</td>
<td>-0.133 (0.052)</td>
</tr>
<tr>
<td>LHH (-1)</td>
<td>0.002 (0.067)</td>
<td>0.069 (0.080)**</td>
</tr>
</tbody>
</table>

1Standard Errors between pharenteses.

***1%, **5% and *10% significance.

Tables 6 and 7 show that the VECM is valid when we analyse the American market against the European ones. Either if we were working with NBP or Zeebrugge data, when the European market is the dependent variable, the long run component parameter is significant and the auto-regressive parameter is also significant and explains the short run adjustments. On the other hand, when the American market is the dependent variable, only the constant and the auto-regressive coefficient are significant. These results suggest that, in some extent, the American market is the one which has influence on the others.

However, the VECM results show that it is not enough to account for all the dependence between these series in the short run, as the coefficients of the explanatory variable related to the external market are not significant. Therefore, we decided to run a Structural Break Test trying to identify if, for any pair of prices, it was observed. When there is some suspect of structural break, a Time Varying Parameter Technique is indicated to provide a better way to model the Error Correction Model.
The CUSUM test provides a plot of the pair of 5 percent critical lines. Movements outside the critical lines is suggestive of parameter instability. The cumulative sum is generally within the 5% significance lines, but, as one can see in figure 4, in 2005, 2006 and 2008 we have some evidence of instability and so, a structural break, for the pair NBP-HH. Then, for this pair, we decided to run a TVP-ECM, which results are shown in Figure 5. The first plot of this figure is the time-varying coefficients for the long run adjustment component of the Error Correction Model. The second one gives the dynamics of the autoregressive parameters and the last one of the explanatory variable parameters. One can see that the estimated coefficients for the long run adjustment and the autoregressive components are not significant, once the confidence interval contains the zero. For the explanatory variable, the estimated coefficients are significant only during the year 2000.
Once the VECM's and the TVP-ECM are not appropriate techniques to model the short run relationship between the variables, we decide to estimate Copula Functions for each pair of prices, attempting to figure out how the short run dependence works. Tables 8, 9 and 10 show the results for the Copula Estimation. The best fitting is given by a combination of the log likelihood and the "Goodness of Fit" (GOF) statistic.

<table>
<thead>
<tr>
<th>Table 8-CopulaAnalysis ZBG-NBP</th>
<th>Table 9-CopulaAnalysis NBP-HH</th>
<th>Table 10-CopulaAnalysis ZBG-HH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Copula</strong></td>
<td><strong>Param.</strong></td>
<td><strong>LogLik.</strong></td>
</tr>
<tr>
<td>Normal</td>
<td>0.765</td>
<td>38.170</td>
</tr>
<tr>
<td>$t_1$</td>
<td>0.738</td>
<td>52.686</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.812</td>
<td>54.615</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.830</td>
<td>53.837</td>
</tr>
<tr>
<td>$t_4$</td>
<td>0.836</td>
<td>52.862</td>
</tr>
<tr>
<td>$t_5$</td>
<td>0.838</td>
<td>51.947</td>
</tr>
<tr>
<td>$t_6$</td>
<td>0.839</td>
<td>51.126</td>
</tr>
<tr>
<td>$t_7$</td>
<td>0.837</td>
<td>50.395</td>
</tr>
<tr>
<td>$t_8$</td>
<td>0.836</td>
<td>49.743</td>
</tr>
<tr>
<td>$t_9$</td>
<td>0.835</td>
<td>49.159</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2.552</td>
<td>48.671</td>
</tr>
<tr>
<td>Clayton</td>
<td>1.938</td>
<td>33.997</td>
</tr>
<tr>
<td>Frank</td>
<td>8.106</td>
<td>43.399</td>
</tr>
</tbody>
</table>

As one can see in Table 8 above, for the pair Zeebrugge-NBP the Copula Function that best explain the behaviour of these prices in the short run is the $t$ Student Copula with 2 degrees of freedom. This means that the prices in these markets move together mainly in extreme periods. As it was exposed in previous sections, in a $t$ Student Copula the lower dependence at the extreme distribution’s support increases with the degree of freedom.

Another important result related to the pair Zeebrugge-NBP is the fact that we reject the Normal Copula. As we have mentioned in Section 4, the Normal Copula has the characteristic of no dependence at the extremes of the bivariate distribution ($\lambda_L = \lambda_U = 0$). The $t$ Student Copula is more realistic once the natural gas prices in these markets move together more strongly during extreme periods.

Table 9 shows the results for the pair NBP-Henry Hub and the $t$ Student Copula with 4 degrees of freedom achieved the best fit, confirming the fact that prices in these markets are more related at extreme occasions. As we explained in Section 2, these markets are huge spot natural gas markets and strongly represents the european and the american markets, respectively.

Finally, Table 10 presents the results for the pair Zeebrugge-Henry Hub and the Copula that best describes the dependence of these series is the Frank Copula. As we have seen in Section 4, the Frank Copula, as the Normal Copula, has no dependence at the extremes of the bivariate distribution ($\lambda_L^F = \lambda_U^F = 0$). This result is very intuitive once the Henry Hub and the Zeebrugge markets are very distant from each other and the Zeebrugge market has not the magnitude observed in the NBP market.
7 Conclusion

The natural gas market has been suffering a lot of changes all over the world during the last decades and all these transformations catch the attention of some literates on the subject in an attempt to explain a possible convergence between the main regional natural gas markets (the North American, the European and the Asian/Pacific markets). The main studies on this matter use Cointegration Technique and Kalman Filter Approach. In this article we use Error Correction Model (VECM), Time Varying Parameters Error Correction Model (TVP-ECM) and Copula Analysis to examine if there exists or not a short run dynamic between these regions and what is the "dependence" between these markets.

Cointegration tests including more than two variables were not conclusive, what makes not possible to evaluate a global integration of the markets. This happened probably because the LNG market, which is the instrument that gives the missing link to connect the regions, is still very incipient.

We then analyze convergence among pair of prices. We use logged monthly natural gas prices for Henry Hub, in the United States, NBP, in the United Kingdom, and Zeebrugge, in Belgium. For the Japanese market we use an average of the main LNG selling prices. All data goes from April 2001 to February 2009. These tests show that the Japanese market does not integrate with any other, probably because the Japanese market is still too much indexed to oil prices through long term contracts, whether in the United States and the United Kingdom, Henry Hub and NBP respectively, both represent spot markets, and in Continental Europe one can already find contracts indexing natural gas prices to the spot market in UK. Continuing the cointegration tests, the pairs Henry Hub-NBP, NBP-Zeebrugge and HH-Zeebrugge presented a long run relationship.

Trying to identify a short run dependence between these prices, we run the VECM for each pair of prices. For the pairs NBP-Henry Hub and Zeebrugge-Henry Hub this model indicate a possible influence of the American market in the European ones. However, the results are not enough to account for all the short run dependence. For the pair NBP-Zeebrugge, this model is not significant at all, what is a sign that these markets move together but not with relevant time delays, that is, they seem to move simultaneously.

Facing these results, we decided to test for structural breaks and only the pair NBP-Henry Hub presented some trace of it. Therefore we ran a TVP-ECM for this pair of prices. Nevertheless, the smoothed states estimates obtained were not significant.

In order to figure out the short run relationship between the natural gas prices, we applied the Copula Approach. For the pairs Zeebrugge-NBP and NBP-Henry Hub we found that the short run relation could be represented by a $t$ – Student Function, indicating that these markets move together strongly during extreme periods. On the other hand, for the pair Zeebrugge-Henry Hub the copula obtained was the Frank Copula, which has the characteristic of no dependence at the extremes of the bivariate distribution.

To improve our results, it would be interesting to implement some additional research, for example:

1. Working with daily frequency data. In this case, when estimating the TVP-ECM, it will be necessary to model the error of the measurement equation with a GARCH component, in order to incorporate the higher volatility coming from the new frequency in the specification.

2. Incorporating a proxy for transport costs among the regions on the state equations of the TVP-ECM.

Nonetheless, the application of VECM, TVP-ECM and Copula Technique is a valid exercise, since, as we
know, there are no articles in the topic of natural gas market integration using these methodologies and it will certainly help for a better comprehension of the behavior of the natural gas prices among the markets.

References


