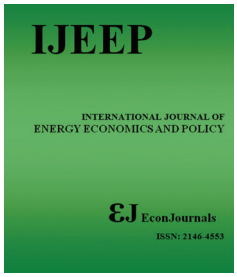


International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2019, 9(5), 322-330.

The Relevance of Crude Oil Prices on Natural Gas Pricing Expectations: A Dynamic Model Based Empirical Study

Andre Assis de Salles*, Ana Beatriz Mendes Campanati

Department of Industrial Engineering, Federal University of Rio de Janeiro, Brazil. *Email: as@ufrj.br

Received: 25 February 2019 **Accepted:** 15 June 2019 **DOI:** https://doi.org/10.32479/ijeep.7755

ABSTRACT

The natural gas price is an important and often decisive variable for economic policy makers. Many studies have been developed in order to establish a stochastic process that can represent the movements or the returns of natural gas prices or variations of such prices time series to forecast price expectations. This work aims to study the relationship between natural gas and crude oil prices in the international market, proposing to investigate its nature and long term equilibrium, through the development of adequate econometric models for determining future expectations of major natural gas price benchmarks, or of their returns. In order to accomplish this, time series for both benchmark crude oil and natural gas prices are subjected to statistical tests with the purpose of verifying the underlying hypotheses behind the appropriate autoregressive dynamic models. The conditional heteroskedasticity and non-normality of the return series, which are prevalent characteristics in energy markets, are considered when elaborating these models. To reach the purpose of this work weekly natural gas and crude oil prices benchmarks traded in the international market were collected.

Keywords: Natural Gas Prices, Crude Oil Prices, Cointegration, Causality, Autoregressive Distributed Lag Model
JEL Classifications: C22, C51, G15, Q40

1. INTRODUCTION

All natural gas is a critical resource for national economies, whose importance is expressed not only by its substantial use in the industry as a source of heat and energy, as well as an input for production, but also through its relevant role in residential and commercial heating systems. A relevant implication of the versatility arising from this wide range of applications is that demand variations for a certain use of natural gas may have a significant impact on prices for other applications. Accordingly, from the beginning of the last decade, gas price volatility has been remarkable. After hitting record levels with the 2005 and 2008 peaks, gas prices have continuously plummeted following the outbreak of the global financial crisis. This volatility is partly associated with difficulties with gas transportation in localities where pipeline infrastructure is not consolidated. Given these limitations, there is no global market for natural gas and local prices may be largely dependent on regional production and availability. Therefore, as observed by the Union of Concerned Scientists (2014), the setting of prices is subjected, to a certain degree, to local supply and demand fundamentals. However, an important factor also to be considered in pricing is the long term relationships in the energy market. As energy sources may be substituted for end uses, it is only reasonable to suppose that energy prices of difference sources are in some way associated. Indeed, many studies have been developed over recent years confirming the cointegrating relationship between different energy commodity prices. Among these studies many are about the relationship between natural gas and crude oil prices, which have in several instances been characterized as stable and possessing a long-term equilibrium. This is consistent with economic theory, as noted by Hartley and Medlock (2014), which suggests that different types of fuels are in competition, though some studies have presented evidence against this relationship.

Important economic implications are derived from ascertaining the existence or not of a stable relationship between crude oil

and natural gas prices. One might mention the importance of permanent shifts in this relationship to policy makers, as such changes may aid or frustrate attempts to promote the use of one fuel type over another when one of these is subject to more significant environmental externalities, as noted by Hartley and Medlock (2014). Similarly, companies and investors are interested in such a link insofar as understanding it may allow for lucrative investments, arbitrage opportunities, speculative strategies and hedging.

This work aims to study the relationship between natural gas and crude oil prices in the international market, proposing to investigate its nature and long term equilibrium, through the development of adequate econometric models for determining future expectations of major natural gas price benchmarks, or of their returns. In order to accomplish this, time series for both benchmark crude oil and natural gas prices are subjected to statistical tests with the purpose of verifying the underlying hypotheses behind the appropriate autoregressive dynamic models. The conditional heteroskedasticity and non-normality of the return series, which are prevalent characteristics in energy markets, are considered when elaborating these models.

The remainder of this paper is structured as follows: Section 2 presents a literature review, with comments on studies that analyze the relationship between natural gas and crude oil prices in the international market. Section 3 introduces the methodological approach employed to fulfill the objectives of this work, whilst Section 4 describes the sample selected to do so. Section 5 shows and provides comments on the results obtained from the econometric procedures undertaken for the purposes of this research. Finally, Section 6 is concerned with the final remarks of this study, and is followed by a list of references.

2. LITERATURE REVIEW

There is an extensive empirical literature on the relationship between natural gas and crude oil prices. Many studies have been developed in recent decades with the purpose of verifying the existence of a long term relationship between these two series, as well as analyze the adequacy of the use of time series models to describe this relationship. Some comments on these studies are presented in the following paragraphs.

Hartley and Medlock (2014) investigate the existence of a stable long-term relationship between natural gas and crude oil prices, identify shocks that cause shifts from this relation, and estimate the duration of the adjustment process. Three time series are analyzed namely, WTI crude oil prices, henry hub (HH) natural gas prices, and residual fuel oil prices, on a monthly frequency from February 1990 to October 2006. An error correction model is employed to study the relationship. Hartley and Medlock (2014) conclude these prices remain linked in their long term evolution through an indirect relationship, which is manifested through the competition between natural gas and residual fuel oil. This contradicts the direct relationship to which most of the previous literature points. The results indicate that crude oil prices are exogenous to the system that includes natural gas and residual fuel oil prices, specifically, this system tends to respond to movements in the international crude oil market, but the reverse does not hold true. Long-term relationships are attained after an adjustment period. Therefore, a rise in the global crude oil price results in a rise in the residual fuel oil price, and, ultimately, in a rise in natural gas price.

Another paper worth mentioning is that of Brown and Yücel (2007). These authors use an error correction model to show that when factors such as weather, storage, and others are taken into account, crude oil price movements have a significant role in the formation of natural gas prices. In order to analyze the relationship between weekly natural gas and crude oil prices, Brown and Yücel (2008) use a sample that covers the period of January 1994 to July 2006, which is further restricted to the period of June 1997 to July 2006 when the influence of climate, seasonality, shocks and storage are all considered. It cannot be inferred from the estimated regression models that movements in oil and gas prices are properly explained, which may have contributed to the view that these variables are independent of each other. A regression model with an error correction mechanism reveals that weekly crude oil and natural gas prices still have an important relationship, conditioned to weather, seasonality and storage. Considering all these additional factors, natural gas price movements are well explained by those of crude oil price.

The work of Panagiotidis and Rutledge (2007) evaluates the relationship between natural gas and Brent crude oil prices in the UK, in order to determine whether these prices have "decoupled", as the orthodox gas market liberalization theory would suggest. To that end, unit root and cointegration testes were employed. It has been argued that in liberalized markets, such as the United Kingdom's, the link between crude oil and natural gas prices disappears. However, between 1990 and the end of 2000, the relationship between these two markets appears to be strong. The work shows evidence that, in the UK, natural gas prices and crude oil prices are cointegrated. Recursive methods show that the cointegration hypothesis is not affected by the Bacton-Zeebrugge gas interconnector. This indicates that, although the UK market is considered to be liberalized, crude oil and natural gas prices possess common stochastic properties.

Leykam and Frauendorfer (2008) used spot price data from the four largest natural gas price benchmarks traded in Europe, namely, National Balancing Point (NBP), Zeebrugge, TTF and Bunde, in order to analyse the interrelationships between these price series. The sample covers the period from March 2005 to May 2008, with a total of 824 observation for each series. The Engle-Granger and Johansen cointegration tests were employed to test cointegration between these four markets. Regression models with error correction mechanisms were also estimated in order to analyse spread between markets, whilst autoregressive conditional heteroskedasticity models and causality tests were applied to study volatilities. Results show that the European natural gas markets are linked by a long term relationship, in which prices do not deviate more that transportation and transaction costs in the long-run. However, the magnitude of this integration is dependent on the market pair being considered. The market pair NBP and Zeebrugge deserves special attention, as it appears to be very integrated.

A more recent study by Brigida (2014) conducts an analysis of the cointegrating relationship between natural gas and crude oil prices by incorporating shifts in the cointegrating vector when estimating the equation. The cointegrating equation is switched between *m* states, according to a first-order Markov process. The cointegrating equation's suitability for the two-state regime-switching model is evidence that there is a switching relationship between natural gas and crude oil prices. Statistical inferences indicate integration between these energy markets, and show that forecast models relating natural gas and crude oil prices should be conditioned on state probability.

In another recent study by Nick and Thoenes (2014), structural vector autoregressive model is developed to be applied to the German natural gas market, analysing key factors that determine natural gas prices. The data collected include NCG natural gas prices, WTI oil prices and North-Western-European coal price, ranging from January 2008 to June 2012. Results suggest that natural gas price is affected by temperature, storage and supply shortfall in the short-run, whilst in the long-run the main factors are crude oil and coal prices, which reflects the economic environment and the substitution relationship between different energy commodities.

Finally, it is also worth mentioning the work of Frey et al. (2009), which investigates the econometric literature regarding crude oil forecasting. A taxonomy for econometric models for oil price forecasts is established, a critical analysis of the different methodologies is conducted, and an interpretation of the heterogeneous findings in the empirical literature is also provided. Econometric models in the existing literature are thus divided into three categories; time series models, which exploit the statistical properties of historical data, financial models based on the relationship between spot and futures prices, and structural models, which describe how economic drivers affect the future values of crude oil prices. It was noted that, for the reviewed studies, the random walk and the autoregressive model never outperform more general models. Some authors suggest combining the performance of different models as a good strategy. By doing so, it is possible to obtain significant improvements in forecasting accuracy. It is not possible, however, to identify which class of models outperforms the rest in terms of accuracy.

3. METHODOLOGICAL APPROACH

Research involving time series presents a set of obstacles. For the most part, empirical research that includes this type of data assumes the underlying time series is stationary, meaning, in a broad sense, that its mean, variance and autocovariance do not change systematically over time. Since time series models are frequently used in forecasting, it is important to verify whether this assumption holds true, or whether the statistical inferences obtained may be considered as valid if the underlying time series is not in fact stationary. One of the most common problems when using non-stationary time series is spurious regression, which is defined as a regression where there is a statistically significant coefficient of determination R^2 between two variables that should not, a priori, be related. Yule (1926) verified that this phenomenon remains in non-stationary series even if the sample is very large. According to Granger and Newbold (1976), $R^2 > D$ is a good rule of thumb for suspecting the existence of spurious regression, where *D* designates the Durbin-Watson statistic.

Thus, verifying the hypothesis of stationarity is extremely relevant for the elaboration of models that describe a time series or a stochastic process. The conditions for stationarity of a time series are fulfilled when its mean, variance and autocovariance, in various lags, do not change over time. Therefore, the series has a tendency towards mean reversal and fluctuations around the mean, measured by the variance, have constant amplitude, as noted by Gujarati (2004).

The most direct way of establishing the existence of stationarity would be to conduct the t test; however, under the null hypothesis, the t value of the estimated coefficient of *Yt−1* does not follow the *t* distribution, even in large samples. An alternative to this issue was developed by Dickey and Fuller (1979), who showed that, under the null hypothesis of δ=0, the estimated *t* value of the *Yt−1* coefficient follows the τ (tau) statistic. Critical values of τ were thus computed through Monte Carlo simulations and in the literature the τ test is commonly referred to as Dickey-Fuller (DF) test. A popular version of this test is the Augmented DF test (ADF), applied when the error term *ut* is correlated. As shown in equation (1), the new test is conducted with the added lagged values of the variable Δ*Yt*.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad (1)$$

Where *εt* is a pure white noise error term, β_1 and β_2 are constants. The number of lagged values to be included for the variable ΔY_t is empirically determined so as to eliminate serial correlation in the error term. The term includes a deterministic trend and the β_1 term is applicable when there is suspicion of a random walk with drift, as opposed to a pure random walk.

Generally speaking, if a time series need to be differentiated *n* times so as to become stationary, it is denoted as integrated of order *n*, which is represented as $Y_t \sim I(n)$. As noted by Wooldridge (2009), if the series in question is stationary, differentiation is not necessary, it is said that such a series is integrated of order zero, or $Y_t \sim I(0)$. Let X_t e Y_t both represent *I* (1) time series; regressing Y_t on X_t yields the following result:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (2)$$

$$u_t = Y_t - \beta_1 - \beta_2 X_t \quad (3)$$

An interesting situation arises when u_t, isolated on the left-hand side of equation (3), is *I*(0), which means that the linear combination of two non-stationary time series is stationary. In this case, the linear combination cancels the stochastic trends of the two series. As a result, the regression of Y_t on X_t is not spurious and the two variables are cointegrated. The regression showed on equation (2) is named cointegrating regression and the β_2 term is denoted as cointegrating parameter, as noted by Gujarati (2004).

Although the possibility of spurious regressions calls for caution when employing the levels of $I(1)$ variables, which makes differentiation the safest course of action, exploring the cointegrating relationship between two variables expands the scope of questions that can be answered. Through the economic perspective, two cointegrated variables have a long-term or equilibrium relationship between them that allows the use of the traditional regression methodology. Thus, if u_t from equation (3) is stationary, the series Y_t and X_t are cointegrated. In order to test whether such is the case, it is enough to simply apply the DF test or the ADF test on the cointegrating equation's residuals. This characterizes the Engle-Granger test, or Augmented Engle-Granger test, respectively, as observed by Gujarati (2004). An important distinction to be noted is that, since the residuals estimated on equation (3) are based on the estimated cointegrating parameter β_2 the DF or ADF test critical values are no longer appropriate. Engle and Granger (1987) therefore computed new critical values for the cointegration test. If Y_t and X_t are not cointegrated, the regression from equation (8) is spurious and meaningless: there is no long term relationship between Y and X. It is still possible to regress the differentiated variables ΔY_t and ΔX_t, however, the interpretation in this case is related to how well the differentiated X variable explains the differentiated Y variable, and this regression says nothing of the relationship between the two variables on level. Conversely, if the variables are cointegrated, more general dynamic models may be employed. For more details, see Wooldridge (2009).

An autoregressive model includes one or more lagged values of the dependent variable among the explanatory variables. Equation (4) shows an example of an autoregressive model:

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t \tag{4}$$

Such models are also known as dynamic, as they represent the dependent variable's path over time in relation to its past values. When the regression model includes not only the current value, but also the lagged values of the explanatory variables, it is called a distributed lag model. Generally, this model, with k lagged periods, may be described by the following formula:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_k X_{t-k} + u_t \tag{5}$$

The β_0 coefficient shown on equation (5) is known as short-run multiplier, because it expresses the change in Y's mean due to a contemporaneous unit variation in X. After k periods, the distributed lag long-run multiplier is given by equation (6):

$$\sum_{i=0}^{k} \beta_i = \beta_0 + \beta_1 + \beta_2 + \ldots + \beta_k = \beta \tag{6}$$

The dependence of the endogenous variable Y on the exogenous variable X is rarely instantaneous but occurs with a delay denoted as lag. Therefore, autoregressive and distributed-lag models are very useful in econometric analysis. There are three main reasons behind the phenomenon of lags. The first set may be characterized as psychological reasons and stems from the force of habit, or the inertia of economic agents. Technological reasons are also partly responsible for lags. These are related to the time taken to implement changes in response to price variations and due to imperfect knowledge of such changes. Limited knowledge is particularly prevalent in dynamic sectors, such as the high technology sector, which may cause hesitation to realizing changes in response to potentially transitory variations. A third reason for lags is related to institutional obstacles.

With the concepts elucidated above, it is possible to define a dynamic time series model denominated autoregressive distributed lag (ARDL) model. As noted by Pickup (2014), such a model is a combination of the autoregressive and distributed lag models described, thus containing lags of both dependent and independent variables on the right-hand side of the equation. Equation (7) represents an ARDL (p, m), with p lags of the dependent variable and m lags of the independent variable:

$$y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \beta_1 x_t + \sum_{i=1}^{m} \beta_{i+1} x_{t-i} + \varepsilon_t \tag{7}$$

Where ε_t is a pure white noise error term.

The previously defined models assume homoscedasticity of the error terms. However, autocorrelation may be present in the variance of the error term on instant t in relation to its past values. This phenomenon was first observed by researchers analyzing financial time series such as stock prices, inflation rates and exchange rates. The noted autocorrelation is denominated autoregressive conditional heteroskedasticity and expressed by the ARCH model presented in the seminal paper by Engle (1982). This model was subsequently widespread in econometric literature for the case when the variance of the stochastic term is related to the square of the lagged value of the stochastic term. If the error variance is related to the squared errors of several past periods, the process is known as generalized autoregressive conditional heteroskedasticity (GARCH). Financial and energy market time series show high volatility and are, for the most part, random walks. Therefore, it would be only natural to model the first differences, which is usually stationary, however, the first differences of these series frequently have high volatility. This suggests that the variance of financial time series changes over time. Consider the dynamic model represented by equation (8) below:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t \tag{8}$$

In order to model the dynamics of the variance of ε_t, it is important to establish the distinction between conditional and unconditional variances. It is assumed is constant and without serial correlation, as shown in equation (9) below:

$$E\left(\varepsilon_t \varepsilon_{t-s}\right) = \begin{cases} \sigma^2, & s = 0 \\ 0, & s \neq 0 \end{cases} \tag{9}$$

Though the assumption of constant unconditional variance still holds, the conditional variance of past error values is allowed to change over time. Therefore, the error term is now modelled as a process with conditional variance $E(\varepsilon_t \varepsilon_{t-s})$, that is, as a function of the past values of the variance. A common way of doing so is by modelling the squared errors as an autoregressive process of order m, which defines the ARCH(m) process:

$$\varepsilon_t^2 = \zeta + \phi_1\varepsilon_{t-1}^2 + \phi_2\varepsilon_{t-2}^2 + \ldots + \phi_m\varepsilon_{t-m}^2 + \omega_t \quad (10)$$

Where ω_t is a white noise process of zero mean and constant unconditional variance. Time series models that include such a process in error modelling are named ARCH models. According to Engle (1982), equation (10) may be estimated using the OLS and the maximum likelihood methods for consistent and similar results. The ARCH test is useful to verify the existence of correlation in the error variance. It consists of estimating equation (10) and subsequently employing the *F* statistic to analyze the following null hypothesis:

$$H_0 : \phi_1 = \phi_2 = \ldots = \phi_m = 0 \quad (11)$$

Should the null hypothesis described by equation (11) hold true, then $\varepsilon_t^2 = \zeta$, that is, the variance is constant and there is no ARCH effect. One of the versions of the ARCH model that has proved most popular in the literature is the GARCH model, first proposed by Engle and Bollerslev (1986). This new approach defines the conditional variance generating process as an ARMA process. The GARCH (*p*, *q*) model is represented by equation (12) ahead.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \quad (12)$$

Equation (12) proposes that the conditional variance of the error term on instant *t* is dependent not only on the squared term of past error, but also on the conditional variance of past periods, Bollerslev (1986) for more details.

As elucidated by Gujarati (2004), the existence of a relationship between two variables, verified through regression analysis, does not prove causality or the direction of this influence. A particular situation arises for the case of time series regressions: if event *A* precedes event *B*, it is possible that *A* may cause *B*, but it is not conceivable that *B* may cause *A*, as, clearly, the future cannot cause the past. Granger (1969) made use of this principle when elaborating a causality test that has become widespread in econometric literature. Let X_t and Y_t represent two time series; the question to be considered is whether X_t "causes" Y_t or whether Y_t "causes" X_t. It is assumed that the relevant information for the prediction of these variables is entirely incorporated in their time series. Conducting the test involves estimating the following regressions models:

$$X_t = \sum_{(i=1)}^{n} \alpha_i Y_{(t-i)} + \sum_{j=1}^{n} \beta_j X_{t-j} + u_{1t} \quad (13)$$

$$Y_t = \sum_{i=1}^{n} \lambda_i Y_{t-i} + \sum_{j=1}^{n} \delta_j X_{t-j} + u_{2t} \quad (14)$$

Where the residual terms u_{1t} and u_{2t} are uncorrelated. Equation (13) proposes that the contemporaneous value of *X* is related to its own past values and those of *Y*, whilst equation (14) predicts a similar behavior for *Y*. There are four possible outcomes, each of which imply different conclusions from the test. The first case is that of unidirectional causality from *Y* to *X*, and is indicated if the estimated coefficients of the lagged values of *Y* in equation (13) are, as a set, statistically different from zero $\left(\sum_{i=1}^{n} \alpha_i \neq 0\right)$, while the set of coefficients of *X* in equation (14) are not $\left(\sum_{j=1}^{n} \delta_j = 0\right)$. A second possibility consists of unidirectional causality from *X* to *Y*, valid for the conditions of $\sum_{i=1}^{n} \alpha_i = 0$ and $\sum_{j=1}^{n} \delta_j \neq 0$. The third outcome, that of bilateral causality, is suggested when the sets of estimated coefficients of *X* and *Y* are both statistically different from zero on both regressions. Finally, a fourth case of independence arises if the *X* and *Y* are not statistically significant in either of the regressions. Thus, if including lagged value of *X* leads to a significant improvement in the prediction of *Y*, it may be say that *X* Granger causes *Y*. However, some assumptions have to be verified in order to ensure the validity of the conclusions derived from the Granger causality test. It is worth mentioning, among those, that the two variables *X* and *Y* must be stationary. As previously noted, the disturbance terms must be uncorrelated. The number of lags to be introduced in the test is an empirical question, to be resolved through information criteria, such as the Akaike or Schwarz criteria. It should be noted that, as observed by Gujarati (2004), the direction of the causality may be highly dependent on the number of lags chosen.

The data used in this study is presented in the following section.

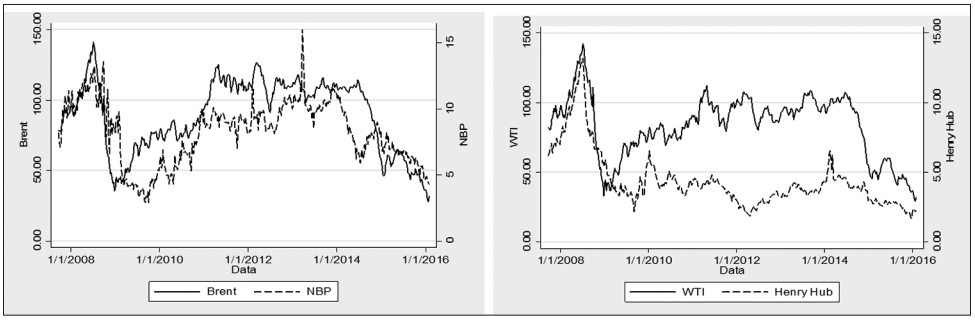
4. THE DATA – SAMPLE USED

The two important benchmarks of natural gas prices in the international market are the HH and the NBP in the US and United Kingdom markets, respectively. The NBP gas market is the oldest in Europe, in operation since the late 1990s. The NBP price is widely used as an indicator of the wholesale market for natural gas in Europe while the well known natural gas HH is used in North America market and HH in Louisiana represents the standard delivery point for natural gas future contracts traded on the New York Mercantile Exchange. As regards crude oil, the two main benchmarks in the international markets are Brent and WTI crude oil types. Brent crude oil refers to crude oil extracted in the North Sea and traded in the London market, while WTI crude oil type originates from US producing region, traded in the New York market.

The sample used in this work comprises the NBP and HH weekly spot prices, collected in the Bloomberg web-site, and Brent and WTI crude oil types weekly spot prices collected from the EIA, the North American energy agency. The sample covers the period from September 2007 up to January 2016, four time series of 437 observations. The Brent and WTI crude oil types were collected in US dollars per barrel while prices for the HH and NBP natural gas prices were converted into US dollars per Million BTU.

The plots shown in Figure 1 below show the WTI and HH weekly prices, the US market references, and the Brent and NBP weekly prices, European market benchmarks. The crude oil prices show a similar behavior of natural gas. The stationarity of these time series appear to be implausible. These plots suggest a large variation for the average during the period studied. The variance also varies over time. It is possible to identify cyclic movements in these time series but not a seasonality pattern. It can also be observed that the time series studied show a financial time series classical behavior, that is, a random walk. These two plots shown in Figure 1

Figure 1: Oil Spot Weekly Prices (US$/Barrel) and Natural Gas Spot Weekly Prices (US$/MMBtu)



demonstrate similar movements, which suggests correlation. As Gujarati (2004) observes the first difference of price time series is usually stationary.prices for the HH and NBP natural gas prices were converted into US dollars per Million BTU.

The Table 1 presents the time series statistical summary, with summary measures, normality and stationarity hypothesis tests results. The average prices differ for the natural gas and crude oil benchmarks, respectively. The average and standard deviations prices allows to infer that the natural gas HH presents a greater volatility than the NBP whereas the Brent and WTI crude oil types prices show a similar behavior. All the skewness coefficients differ from the normal distribution coefficient, apart from the HH that was positive and farthest from zero. The HH prices kurtosis coefficient differs from other price time series used in this work, apart from the HH prices that shows a leptokurtosis. The other price time series are lower than the normal distribution, indicating platykurtosis. The Jarque-Bera test demonstrates that the normality for all time series analyzed could not be accepted. This is a common feature in the financial assets and commodities price time series. As previously noted, the prices time series selected do not show stationarity and their variation or first difference must be stationary. Thus, from the price time series collected the variations of these prices or the returns were calculated.

Therefore the price return time series logarithmic returns were calculated for the weekly prices presented above using the following formula:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (15)$$

where R_t represents the return in period t, and P_t represents the price in period t. The price time series are integrated of order 1, that is $I(1)$, the return time series obtained from the price time series transformation will be stationary, that is $I(0)$. The plots shown in Figure 2 presents the price returns time series. Table 2 below shows a statistical summary of price return time series of the crude oil and natural gas benchmarks, as well as the normality and the stationarity hypothesis tests results. The average values of these time series have similar values and as demonstrated by standard deviation the NBP natural gas return time series show the greatest variability. It is possible to reject the normality hypothesis at a significance level of less than 1%. According to the ADF test, it is possible to reject the unit root hypothesis for the returns time series at a significance level of <1%. Thus, as the plots analysis and as occurs with financial series, the crude oil and natural gas

Table 1: Statistical summary of weekly price time series

	NBP	HH	Brent	WTI
Mean	8.25	4.45	89.43	83.41
Median	8.60	3.89	97.13	88.50
Maximum	16.00	13.20	141.07	142.52
Minimum	2.90	1.68	27.76	29.19
Std deviation	2.35	2.06	25.58	22.26
Skewness	−0.02	1.95	−0.49	−0.04
Kurtosis	2.47	7.13	2.11	2.75
Jarque-Bera	7.539	588.170	31.834	14.380
(P-value)	(0.023)	(0.000)	(0.000)	(0.001)
ADF P value	0.389	0.199	0.382	0.339
Observations	437	437	437	437

NBP: National Balancing Point, HH: Henry hub

Table 2: Statistical summary of weekly price returns time series

	NBP	HH	Brent	WTI
Mean	−0.0012	−0.0023	0.0020	−0.0021
Median	0.0000	−0.0022	−0.0007	−0.0013
Maximum	0.3478	0.3010	0.2002	0.2512
Minimum	−0.3455	−0.3093	−0.1646	−0.1910
Std deviation	0.0857	0.0634	0.0424	0.0450
Skewness	0.2092	0.0747	−0.0968	−0.0619
Kurtosis	6.0357	6.6423	5.7759	6.9453
Jarque-Bera	170.982	588.170	140.992	283.699
(P-value)	(0.000)	(0.000)	(0.000)	(0.000)
ADF P value	0.000	0.000	0.000	0.000
Observations	437	437	437	437

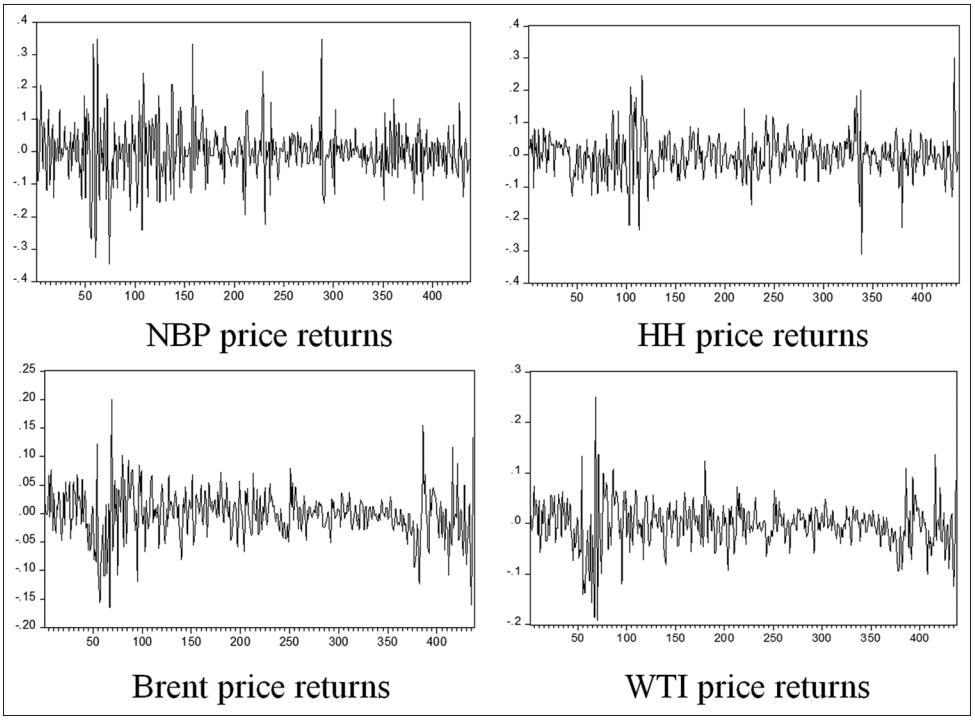
NBP: National Balancing Point, HH: Henry hub

prices are integrated of order 1, or $I(1)$, once their first differences constitute stationary time series. Figure 2 below shows the price return time series, and from the plots it is possible to identify a common behavior in these price return time series, with highly positive or negative observations appearing in clusters, that is, the phenomenon of concentration of higher volatility followed by others periods of relative lower volatility. This shows that the volatility of the current period is related to that of past periods, which presupposes autoregressive heteroskedasticity. Using the ARCH test for price returns time series this proposition can be confirmed.

5. ANALYSIS OF THE RESULTS OBTAINED

The relationship between natural gas and crude oil prices has been investigated for the US and UK markets through the Engle-Granger cointegration test for the US market where the

Figure 2: Crude oil weekly price returns (US$/Barrel) and natural gas weekly prices returns (US$/MMBtu)



natural gas and crude oil benchmarks are the HH and WTI prices respectively. Thus, the model used for the DF unit root test has as its dependent variable the first difference of stochastic terms that was obtained from the HH price regression on WTI prices. A linear regression with intercept of the stochastic terms when using the Engle-Granger cointegration test was implemented. Given the AIC criterion, this model was the most appropriate. The same procedure was applied to the NBP natural gas and the Brent crude oil prices benchmarks in the UK market, where the most appropriate model for this cointegration test was the same mentioned before but without the intercept. The model describe above can be written as follows:

$$\Delta u_t = \beta_1 + \delta u_{t-1} + \varepsilon_t \quad (16)$$

The results obtained with the cointegration hypothesis using the Engle-Granger test are respectively −4.3992 for the HH and WTI weekly prices, and −3.6557 for the NBP and Brent weekly prices. With p values close to 0.0074 and 0.0018, respectively. The results obtained with the cointegration hypothesis using the Engle-Granger test are respectively −4.3992 for the HH and WTI weekly prices, and −3.6557 for the NBP and Brent weekly prices. With p values close to 0.0074 and 0.0018, respectively. The critical values for these tests are −3.922, −3.350 and −3.054 for significance levels of 1%, 5% and 10%, respectively. With these results, the null hypothesis of non-cointegration between the natural gas and crude oil prices in the US and UK markets can not be accepted. Therefore, there is a long-term relationship between the HH and WTI prices as well as between the NPB and Brent crude oil prices, which rule out the possibility of a spurious regression between the natural gas and crude oil prices practiced in the US and UK markets.

The causality is the other hypothesis that must be tested since the existence or not of causal relation can be determinant in the models constructed to explain prices or returns of natural gas traded in the international market. It should emphasized that the Granger's causality test, presented in the methodology of this work, assumes the stationary hypothesis of the variables involved. This hypothesis can be accepted. The causality tests results conducted between the HH and WTI price returns and between the NBP and Brent type crude oil price returns were carried out with lags from 1 to 12 and the results are listed on Table 3. It can be observed from the p value for all lags considered the WTI price returns cause HH price returns, which does not occur in the opposite direction. These results are consistent with the widespread hypothesis that crude oil price cause the natural gas price. On the other hand, the results suggest a unilateral or bilateral causal relationship between the Brent and NBP returns, which differ from the hypothesis in which oil prices are an important factor in the natural gas pricing.

To explain the natural gas price returns a ARDL model up to 12 lags for the variables was used. Among several ARDL models estimated using the AIC criterion an ARDL (3,3) was selected. From this preliminary model, a new model was proposed excluding the variables with non-significant parameters. Thus, a final ARDL model to explain HH price returns was obtained. Using the same criteria the appropriate ARDL model to explain the NBP natural gas price returns was an ARDL (2,8). A final NBP model without no statistical significance parameters was obtained.

The heteroskedasticity suggested by the plots of price returns time series shown in Figure 2 presented in Section 4 was verified. This way the ARCH test for the stochastic terms of the respective models selected mentioned before was implemented using the following model:

$$\varepsilon_t^2 = \zeta + \phi_1 \varepsilon_{t-1}^2 \quad (17)$$

The ARCH test results show the existence of the ARCH effect, once the F statistic for the models referring to the returns of the

Table 3: Granger causality test results for price return time series

Lags	HH →WTI		WTI →HH		NBP →Brent		Brent →NBP	
	F statistics	P-value	F statistics	P-value	F statistics	P-value	F statistics	P-value
1	1.9461	0.1637	4.5946	0.0326	13.1024	0.0003	0.0941	0.7592
2	1.2045	0.3009	3.0446	0.0486	6.8624	0.0012	2.3708	0.0946
3	0.6755	0.5674	2.5863	0.0527	5.3943	0.0012	1.5485	0.2013
4	0.8822	0.4744	2.2729	0.0607	4.1854	0.0025	1.2799	0.2771
5	0.5932	0.7052	1.8291	0.1059	3.3652	0.0054	1.0560	0.3843
6	0.5245	0.7898	1.9232	0.0758	3.3183	0.0033	2.0818	0.0543
7	0.4701	0.8562	1.9863	0.0557	3.5623	0.0010	1.9493	0.0607
8	0.5591	0.8113	2.5955	0.0089	5.4900	0.0000	2.1218	0.0328
9	0.6236	0.7771	2.3104	0.0153	5.1628	0.0000	2.0720	0.0309
10	0.7164	0.7092	2.0425	0.0281	5.0606	0.0000	1.8944	0.0443
11	1.4361	0.1539	1.7859	0.0544	5.0482	0.0000	1.6453	0.0839
12	1.6879	0.0670	1.6574	0.0740	4.7386	0.0000	2.0151	0.0219

HH was 21.4334 and the returns of the NBP prices were 30.3193, respectively. It should be mentioned the p-values are close to zero. Thus, ARCH effect hypothesis was accepted. Furthermore in order to improve the models it is important to include the ARCH processes to deal with the autoregressive heteroskedasticity. For this purpose, several alternatives of GARCH processes (p, q) were included in the natural gas price return models presented above, with p and q varying from zero to five. The Akaike criterion pointed out the best GARCH models are: GARCH (1,2) and GARCH (1,3), for the HH and NBP, respectively.

Another consideration observed was the violation of the normality assumption of the series of price returns. Thus, the Student t distribution was used in the construction of the models to explain the returns of HH and NBP prices.

Therefore the final models estimated to obtain expectations of the price returns and, consequently, of the prices of natural gas in the international market can be described as follows, for the returns of the HH and NBP returns, respectively, *RHH* and *RNBP*.

$$(RHH_t \mid I_{t-1}) \sim Student(\mu_{HHt}; \sigma^2_{HHt}; \nu_{HH})$$

$$_{HHt} = \alpha_1 RHH_{t-1} + \alpha_2 RHH_{t-2} + \alpha_3 RWTI_t + e_{HHt} \quad (18)$$

$$\sigma^2_{HHt} = \delta_1 + \delta_2 e^2_{HHt-1} + \gamma_1 \sigma^2_{HHt-1} + \gamma_2 \sigma^2_{HHt-2}$$

$$(RNBP_t \mid I_{t-1}) \sim Student(\mu_{NBPt}; \sigma^2_{NBPt}; \nu_{NBP})$$

$$_{NBPt} = \alpha_1 RNBP_{t-1} + \alpha_2 RNBP_{t-2} + \alpha_3 RNBP_{t-3} + e_{NBPt} \quad (19)$$

$$\sigma^2_{NBPt} = \delta_1 + \delta_2 e^2_{NBPt-1} + \gamma_1 \sigma^2_{NBPt-1} + \gamma_2 \sigma^2_{NBPt-2} + \gamma_3 \sigma^2_{NBPt-3}$$

Tables 4 and 5 above show respectively the RHH and RNBP models estimates results. These tables presents, besides the statistics of the coefficients: residual sum of squares, Durbin Watson statistic (DW), Akaike criterion (AIC) and mean square error (MSE).

Table 4: Model HH-estimation results

Parameter	Coefficient	Standard error	Statistic t	P-value
α_1	0.1623	0.0462	3,5156	0.0004
α_2	−0.0923	0.0495	−1.8658	0.0621
α_3	0.1945	0.0556	3.4976	0.0005
δ_1	0.0002	0.0000	3.9469	0.0001
δ_2	0.0826	0.0634	3.4453	0.0006
γ_1	1.4729	0.0747	22.4583	0.0000
γ_2	−0.5969	6.6423	−12.3005	0.0000
ν	7.7660	2.5153	3.0875	0.0020
RSS	1.6509		AIC	−2.9372
DW	2.1201		MSE	0.0038

RSS: Residual sum of squares, DW: Durbin Watson

Table 5: Model NBP-estimation results

Parameter	Coefficient	Standard error	Statistic t	p-value
α_1	−0.1779	0.0489	−3.6399	0.0003
α_2	−0.0813	0.0487	−1.6680	0.0953
α_3	0.2624	0.0779	3.3681	0.0008
δ_1	0.0004	0.0002	2,2553	0.0241
δ_2	0.2581	0.0666	3,8771	0.0001
γ_1	1.0441	0.1378	7.5795	0.0000
γ_2	−0.8589	0,1703	−5,0442	0.0000
γ_3	0.5183	0.1085	4.7763	0.0000
ν	5.6189	1.7164	3.2738	0.0011
RSS	2.9815		AIC	−2.4454
DW	1.9845		MSE	0.0068

RSS: Residual sum of squares, DW: Durbin Watson

6. FINAL REMARKS

The objective of this work was to investigate the relationship between natural gas and crude oil price, as well as to verify the crude oil prices relevance models constructed to determine price expectations of natural gas traded in the international market. From this paper overview, it can be observed that the objectives have been achieved. Through the classical statistical inference procedures it was possible to verify the existence of long-term equilibrium relationship which was confirmed by the Engle-Granger cointegration test, with 5% of significance, for both the HH and the WTI, and for the NBP and the Brent. It was also possible to verify the Granger causality which allows us to infer that the crude oil price returns somehow cause the variations of natural gas return prices. This causal relationship demonstrated to be unilateral for the WTI and HH price changes, whereas for the NBP and Brent price returns this appears in two ways. The ARCH test for residuals of the preliminary ARDL model, developed between the natural gas and crude oil price returns pointed out the non-rejection of the autoregressive heteroskedasticity hypothesis,

suggesting that the inclusion of ARCH process in the stochastic models should be mandatory in order to deal with price volatility. The fact that the price returns time series differ from the normal distribution that was treated here using the Student t distribution also deserves attention. It should be noted that imposing limits for the numbers of lags for the GARCH model and the ARDL model constrained the set of models, so that there are other models to describe the dynamic relationship between natural gas and crude oil price returns. However, the models proposed here present a satisfactory indicator of price returns direction.

For future studies it must be highlighted the importance of including alternatives models to test the hypothesis here tested, through more comprehensive lags limits or variations of the presented models. The GARCH model, in particular, has several variations, such as EGARCH or TGARCH that would enable other, and maybe better adjustments. Besides that, the stochastic volatility models can be implemented. The results of the forecasts can be improved by including other relevant explanatory variables once other studies mention that factors such as climate, storage conditions and seasonality play an important role in the relationship between natural gas and crude oil prices. In addition, the use of other methodological approaches would allow to obtain the more robust results.

REFERENCES

Bollerslev, T. (1986), Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31, 307-327.

Brigida, M. (2014), The switching relationship between natural gas and crude oil prices. Energy Economics, 43, 48-55.

Brown, S.P.A., Yücel, M.K. (2008), What drives natural gas prices? The Energy Journal, 28(1), 45-60.

Dickey, D.A., Fuller, W.A. (1979), Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74, 427-431.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987-1007.

Engle, R.F., Bollerslev, T. (1986), Modelling the persistence of conditional variances. Econometric Reviews, 5, 1-50.

Engle, R.F., Granger, C.W. (1987), Co-integration and error correction: Representation, estimation and testing. Econometrica, 55, 251-276.

Frey, G., Manera, M., Markandya, A., Scarpa, E. (2009), Econometric models for oil price forecasting: A critical survey. CESifo Forum, 10(1), 29-44.

Granger, C. (1969), Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), 424-438.

Granger, C., Newbold, P. (1976), R^2 and the transformation of regression variables. Journal of Econometrics, 4, 205-210.

Gujarati, D. (2004), Basic Econometrics. 4th ed. New York: The McGraw-Hill Companies.

Hartley, P.R., Medlock, K.B 3rd. (2014), The relationship between crude oil and natural gas prices: The role of the exchange rate. The Energy Journal, 35(2), 25-44.

Leykam, K., Frauendorfer, K. (2008), Co-integration and Volatility in the European Natural Gas Spot Markets, Master Thesis. St. Gallen, Switzerland: University of St. Gallen.

Nick, S., Thoenes, S. (2014), What drives natural gas prices? A structural VAR approach. Energy Economics, 45, 517-527.

Panagiotidis, T., Rutledge, E. (2007), Oil and gas markets in the UK: Evidence from a cointegrating approach. Energy Economics, 29(2), 329-347.

Pickup, M. (2014), Introduction to Time Series Analysis. Washington DC: Sage Publications.

Union of Concerned Scientists. (2014), Uses of Natural Gas. Available from: http://www.ucsusa.org/clean-energy/coal-and-other-fossil-fuels/uses-of-natural-gas.

Wooldridge, J.M. (2009), Introductory Econometrics: A Modern Approach. 4th ed. Mason: Nelson Education.

Yule, G.U. (1926), Why do we sometimes get nonsense-correlations between Time-Series? A study in sampling and the nature of time-series. Journal of the Royal Statistical Society, 89(1), 1-64.