Advances in Social Sciences Research Journal – Vol.1, No.3
Publication Date: May 15, 2014
DOI: 10.14738/assrj.13.204
De-Salles, A. A., & De-Oliveira, E. M. (2014). Relevance of Future Contracts on Spot Price Formation in Crude Oil Markets. *Advances in Social Sciences Research Journal, 1*(3), 145-156

The Relevance of Future Contracts on Spot Price Formation in Crude Oil Markets

Andre Assis de Salles
Polytechnic School, Federal University of Rio de Janeiro, Brazil;
as@ufrj.br

Erick Meira de Oliveira
Polytechnic School, Federal University of Rio de Janeiro, Brazil;
erickmeira89@gmail.com

ABSTRACT

This paper aims to examine the role different future contracts play on oil spot price formation. Firstly the cointegration and causality hypothesis are tested using appropriate methodologies. Several distributed lag models are estimated in order to forecast spot price behavior, taking into account current information on future price. The results provide evidence of strong predictive power for certain short-term future contracts, using as reference the corresponding expiry date at the time of the transaction. All data have been obtained from daily quotations of the Brent and WTI crude oil prices, in US$ per barrel, in the spot market and their four nearest future contracts. The time period of the analysis spanned from June 2009 to March 2013.

Keywords: Cointegration, Causality, Distributed Lag Models, Crude Oil Market.

INTRODUCTION

Future markets are used nowadays for many different purposes. First and foremost they provide a mechanism of financial risk management, allowing risk to be transferred from the market agents, who wish to avoid it, to speculators willing to take it up. There are also the arbitrators who operate in different markets taking advantage of price differences, when markets present inefficiency, to make clean profits with minimal risk. Moreover current information about future price in efficient markets may be used to forecast future spot price behavior.

The existing literature investigating the dynamics of the association between spot and future markets is extensive. One of the appealing works in this area was proposed by [2], who used a linear regression model to test the wheat and soya markets market efficiency, concluding at that time, that current information on future prices provides no significant estimates of spot price behavior. These results were subsequently contested by [19], who stated that they were based on the F test and, therefore, not valid when prices are non-stationary. On the other hand the cointegration theory first proposed by [10] in a seminal work, and later developed by [8] and [11], provided a new methodology to test market efficiency hypothesis. Many authors committed themselves to investigate the market efficiency hypotheses in future markets, achieving results that did not reject the market efficiency hypothesis. One of the criticisms addressed to the approach proposed by [8] was the possibility of weak inferences for the parameters involved in the regression, which are the key points of the market efficiency hypothesis test. In light of these problems, [18] presented different statistic procedures to test cointegration using maximum likelihood estimation. Such procedures were based on a Vector

Autoregressive (VAR) model, which allows for possible interaction among the parameters involved in the spot and future price estimation.

Despite the huge scientific output in this area, many authors limit themselves to investigate the existence of cointegration between the spot price quotations and those related to the first future contract, sometimes addressed to as front month contract, of financial assets or commodities. Some works also test the causal relationship between future and spot prices, achieving results that confirm the expected result in the short run. However, it must be noted that many markets have developed fast in recent years and offer nowadays a wide range of future contracts. Therefore, very often current information about the future is less consistent especially when distant expiry dates are taken into account. One of the reasons for this situation is the high volatility of international oil markets, whose first future contracts appeared in 1983, and were highly motivated by severe fluctuations in prices, exchange rates and interest rates. Currently, future oil contracts are listed up to nine years in advance. For this and the next five years, it is possible to find monthly contracts for crude oil.

In this context, and in view of the important role that the international oil market plays in the economy as a whole, this paper aims to investigate the association between future oil contracts and oil spot prices, as well as the influence that each contract has on the spot price formation. Firstly, the existence of cointegration and causal relationship between the main crude oil spot prices and their first four future contracts were tested. From these tests several distributed lag models have been proposed and estimated in order to explain oil spot price returns behavior through current information about future price returns on different expiry dates.

DATA – SAMPLE USED

Data Used

In order to achieve the results discussed previously, we make use of daily price time series for the Brent and WTI crude oil in the spot market and in their four future contracts, classified according to the difference between the expiry date and the moment of negotiation. The time period of the analysis spans from June 2009 to March 2013. Spot market and WTI futures data were collected from the U.S. Energy Information Administration (EIA). With regards to the Brent future prices time series, contract data for the Brent crude was taken from the Quandl website.

Future Crude Oil Contracts Negotiation

The nomenclature for oil future contracts is unique, no matter what type of crude oil they are related to, and it is important to understand the results obtained in this work. First of all it is worth highlighting that, in the crude oil market every future contract expires on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading ceases on the third business day prior to the business day preceding the 25th calendar day. After a contract expires, Contract 1 for the remainder of that calendar month is the second following month. In these terms, it would be impossible to build time series referring to a specific contract month, for example the oil future contract expiring in July 2013. Price quotations for this contract date back from some time, which corresponds to the first time it was put on sale, and will cease on June 20th, 2013 (Thursday), the third business day prior to June 25th, 2013 (Tuesday). From that point onwards, there would not be any information regarding the July contract. However, data for the contract expiring in August 2013 would be available until July 22nd, 2013, for example. Therefore, we would have to work with multiple time series for each contract validity period.

To circumvent this problem the terminology used in future markets allows the time period between a contract and the spot price quotations to be kept constant. In other words, instead of using months to reference future contracts, the nomenclature uses numbers - contract 1, contract 2, and so on. That way: Contract 1 refers to contracts traded up to 30 days prior to the expiry date; Contract 2 comprises contracts negotiated between 60 and 30 days prior to the expiry date; and Contract 3 for contracts traded between 90 and 60 days prior to the expiry date. The same procedure applies for subsequent contracts.

It is also important to point out that the expiry date does not exactly match the delivery date. In practice most crude oil transactions take place under long-term contracts that last for years. The contract defines the general conditions of the trade such as the volumes to be traded every month, delivery location, product specification, credit terms and conditions, payment terms and pricing mechanism. Once the contract is set up, the delivery dates follow fairly standard terms that any two counterparties can feel comfortable with.

Stylized Facts

The initial step of this work is to investigate the long run relationship between selected future oil contracts and the oil spot market. In order to do so, we make use of cointegration tests available in the econometric literature. To achieve consistent results, the basic requirement of these tests is that the time series involved must be integrated of order 1, or I(1), as conveyed in [12]. In other words, these series need to be differentiated once to become stationary. Under these terms, after having obtained all data mention in Section 2.1, the price returns time series were generated from the original price time series by making use of the following formula:

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

where represents the price return in the period t and represents the price in t.

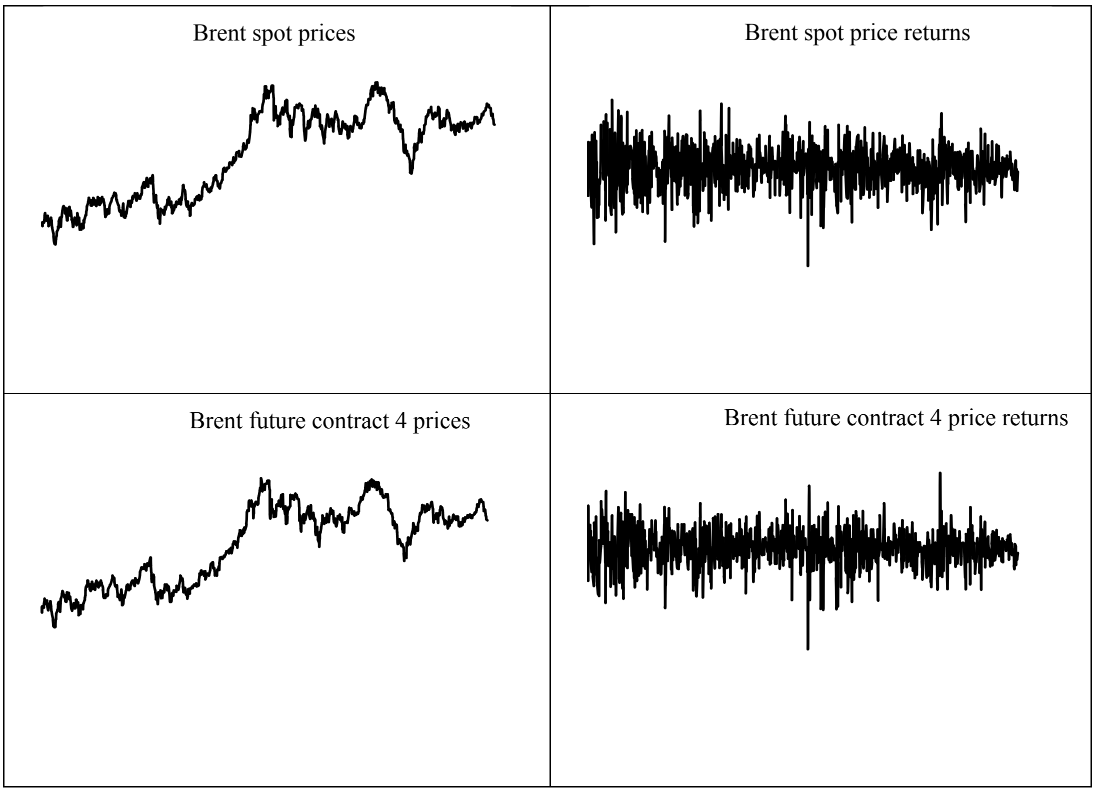

Figure 1: Brent - Prices (US$/barrel) and Price Returns - Spot Market and Future Contract 4

The new generated price returns time series correspond to the first differences of the original data. Therefore, if the price returns are stationary it can be assumed that the price quotations used in this work are indeed integrated of order 1. Figure 1 presents the plots for the daily price and price returns time series related to the Brent crude oil. The first two plots illustrate the price and price returns data corresponding to the spot market, and the following two represent the price and price returns of the 4th future contract, according to the nomenclature mentioned in section 2.2. Illustrations for contracts 1, 2 and 3 were intentionally not included in Figure 1 since they are quite similar to contract 4. The only difference that is worth noting is that, for each time observation, there is an increase in the price values when contracts with long expiry dates are analyzed. In other words, contract 1 is slightly cheaper than contract 2, which is less expensive than contract 3, and so on. All prices displayed in the left side of Figure 1 are quoted in US$/barrel. Figure 2 illustrates the same data explained above for the WTI crude oil.

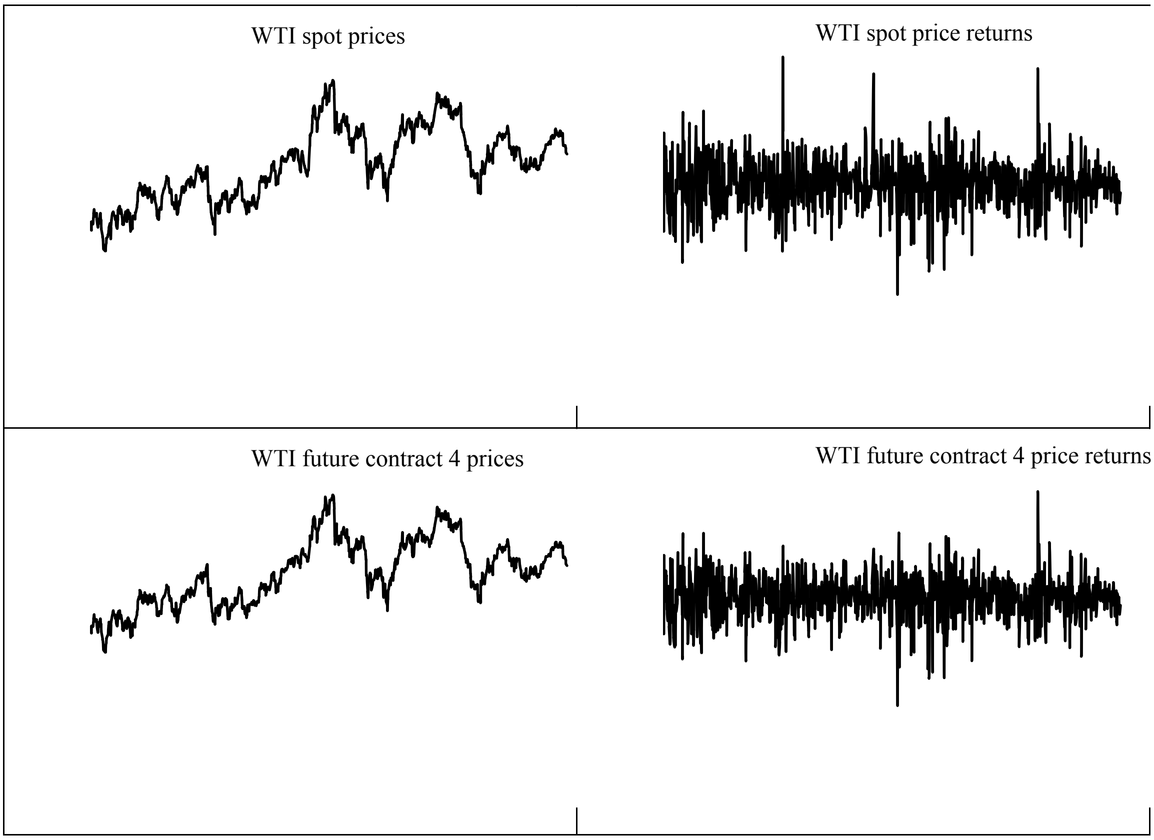

Figure 2: WTI - Prices (US$/barrel) and Price Returns - Spot Market and Future Contract 4

All price time series in Figures 1 and 2 indicate a positive trend in its values over the years. By contrast, there are no notable trends readily apparent in the price returns both Brent and WTI crude oil. Both figures suggest that the oil price returns time series have a constant mean with fluctuations around it. These fluctuations seem to share the same amplitude scaling for the majority of observations, aside from specific time periods. Another remarkable feature is the difference in price returns volatility when comparing the Brent spot market with the Brent future market, as Contract 4 shows, which is considerably higher in the Brent future market. This suggests that oil price returns volatility increases when longer expiry dates are taken into account, due to higher levels of uncertainties. In essence, there is a strong indication that the original price time series are integrated of order 1, since all of them have trends in their values over the years and their corresponding price returns time series seem to be stationary. To test this hypothesis, Tables 1 and 2 give an overview of the statistical summary for each price returns time series involved in this work: spot market and contracts 1, 2, 3 and 4.

Table 1: Statistical Summary of Price Returns Time Series - Brent

Statistics	Spot	Contract 1	Contract 2	Contract 3	Contract 4
Mean	0.0005	0.0005	0.0005	0.0005	0.0005
Median	0.0006	0.0007	0.0008	0.0010	0.0010
Maximum	0.0587	0.0681	0.0680	0.0683	0.0681
Minimum	-0.0825	-0.0896	-0.0898	-0.0898	-0.0899
Standard Deviation	0.0176	0.0174	0.0170	0.0168	0.0166
Skewness	-0.2574	-0.2065	-0.2925	-0.3037	-0.3190
Kurtosis	3.9874	4.4777	4.5684	4.6383	4.7161
Observatios	939	939	939	939	939
Jarque-Bera	183,1943	110,3064	110,0191	126,1632	143,7350
p-value	0,0000	0,0000	0,0000	0,0000	0,0000
Statistic ADF test	-30,5965	-30,6555	-30,8036	-30,8651	-30,9232
T (α = 1%)	-3,9678	-3,9678	-3,9678	-3,9678	-3,9678

Table 2: Statistical Summary of Price Returns Time Series - WTI

Statistics	Spot	Contract 1	Contract 2	Contract 3	Contract 4
Mean	0,0003	0,0003	0,0003	0,0003	0,0003
Median	0,0008	0,0008	0,0005	0,0008	0,0009
Maximum	0,0990	0,0895	0,0890	0,0881	0,0872
Minimum	-0,0854	-0,0904	-0,0895	-0,0889	-0,0884
Standard Deviation	0,0192	0,0190	0,0185	0,0181	0,0177
Skewness	0,0139	-0,1168	-0,1907	-0,2508	-0,2851
Kurtosis	5,1545	4,6556	4,6257	4,7163	4,8214
Observatios	947	947	947	947	947
Jarque-Bera	183,1943	110,3064	110,0191	126,1632	143,7350
p-value	0,0000	0,0000	0,0000	0,0000	0,0000
Statistic ADF test	-30,5965	-30,6555	-30,8036	-30,8651	-30,9232
T (α = 1%)	-3,9678	-3,9678	-3,9678	-3,9678	-3,9678

Apart from the basic descriptive statistics such as mean, median, maximum and minimum values, among others, Tables 1 and 2 also provide the values obtained from the Jarque-Bera (JB) test (see [14]) and Augmented Dickey-Fuller (ADF) tests (see [5]), to verify the presence of normality and stationarity, respectively, in each time series used here. All price returns time series have means and medians close to zero, whose values are also very similar in each crude oil type price returns. This fact alone could indicate the presence of symmetry in the data. However, the values obtained for the skewness and kurtosis suggest that the normality assumption of the time series should not be accepted. That is subsequently confirmed by the Jarque-Bera test results, which demonstrate that the normality assumption of the data may not be accepted, as the p-value obtained in each JB test can be approximated by zero in all time series returns. Therefore, one must take this fact into consideration when constructing models, for instance, choosing to model the error terms of the data by another distribution rather than the normal. The t of Student distribution was widely used given the attractiveness of variation provided by number of degree. The results of the Augmented Dickey-Fuller test stationarity of the data, the results for the τ Statistic obtained from the Augmented Dickey-Fuller τ Statistic test indicates that the existence of unit root cannot be accepted. In other words, the price returns time series are indeed stationary. As for the original time series, it is worth mentioning that the ADF test was used, leading to results that indicated the non-stationarity of the data in all price time series studied. Hence, our data meets the requirements for the cointegration test procedures.

METHODOLOGY

The cointegration test is used to verify if two or more series share a long-term equilibrium relationship. Two or more time series are said to be cointegrated if they share a common stochastic drift. There are many ways to test for the cointegration hypothesis among different time series. In this work the methods proposed by [15] and [18] were used. As mentioned earlier, the standard procedure of these methods is based on a Vector Autoregressive (VAR) model, which can demonstrate possible interactions among the parameters involved in the relation of spot and future prices estimation. A VAR model of *p* can be written as follows:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \ldots\ldots + A_p y_{t-p} + Bx_t + \varepsilon_t \qquad (2)$$

where y_t is a *k* x *1* vector of non-stationary I(1) variables, x_t is a *d* x *1* vector of deterministic variables and ε_t is a vector of innovations. This model can also be represented as

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + Bx_t + \varepsilon_t \,, \qquad (3)$$

where $\Pi = \sum_{i=1}^{p} A_i - I$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j$.

Granger's representation theorem asserts that if the coefficient matrix Π has reduced rank *r* < *k*, then there exist *k* x *r* matrices α and β each with rank *r* such that $\Pi = \alpha\beta'$ and $\beta' y_t$ is I(0). In such terms, *r* represents the number of cointegrating relations, or cointegration rank, and each column of β is a cointegrating vector. Also, the elements of α correspond to the adjustment parameters when estimating a Vector Error Correction (VEC) model. Johansen's test consists of estimating the Π coefficient matrix from an unrestricted VAR model and testing if it is possible not to accept the restrictions implied by the reduced rank of Π .

According [16] the maximum likelihood hypothesis tests used to verify the number of characteristic roots different from zero of the □ coefficient matrix had their asymptotic distributions discriminated, converging to two different statistic test: the trace statistic and the maximum eigenvalue statistic. In brief terms, there are two distinct tests. The first one tests the null hypothesis of *r* cointegrating relations against the alternative of *k* cointegrating relations, where *k* stands for the number of endogenous variables, for $r = 0, 1, \ldots, k - 1$. The alternative of *k* cointegrating relations corresponds to the case where none of the series has a unit root and a stationary VAR may be specified in terms of time series levels. The trace statistic for the null hypothesis of *r* cointegrating relations is computed as

$$LR_{trace}(r|k) = -T \sum_{i=r+1}^{k} \log(1-\lambda_i) \qquad (4)$$

where λ_i is the *i*-th largest eigenvalue of the Π coefficient matrix while *T* stands for the number of observations included in the analysis. On its turn, the second test, which that reports the maximum eigenvalue statistic, tests the null hypothesis of *r* cointegrating relations against the alternative of *r* + 1 cointegrating relations. In such terms, the corresponding statistic test can be computed as follows:

$$\begin{aligned} LR_{trace}(r|r+1) &= -T \log(1-\lambda_{r+1}) \qquad (5) \\ &= LR_{trace}(r|k) - LR_{trace}(r+1|k), \end{aligned}$$

for $r = 0, 1, \ldots\ldots, k - 1$.

It should also be emphasized that the test results are sensitive to trends in the input time series. [17] considered five deterministic trend cases to estimate the cointegration relations, according to different time series behaviors: (i) the level data y_t has no deterministic trends and the cointegrating equations do not have intercepts; (ii) the original level data y_t has no deterministic trends and the cointegrating equations have intercepts; (iii) the level data y_t has linear trends but the cointegrating equations have only intercepts; (iv) Both the level data y_t and cointegrating equations have linear trends; and (v) the level data y_t has quadratic trends and the cointegrating equations have linear trends. However, defining a specific behavior for all time series involved can be quite challenging, especially when there is only a small sample of observations. Therefore, in this work, the cointegration relations are tested for all five time series mentioned above. The critical values of the trace and maximum eigenvalue statistics vary with the number of observations and are listed in [20]. Finally, the selection regarding the optimal number of past lagged values included in the vector autoregressive is made taking into account the results of the Akaike (AIC) and Schwarz (BIC) information criteria presented in [1] and [22], respectively.

Upon completion of the cointegration tests between the spot market and each of the first four future contracts taken into consideration, a study of the causal relationship between these time series was conducted. To that end the Granger causality test, presented in [9], which basically consists of inferring the predictive power of past lagged values as regressors was used. In its simplest form, the causal VAR model proposed by [9] for two time series is computed as follows:

$$\begin{aligned} y_t &= \sum_{j=1}^{m} a_j x_{t-j} + \sum_{j=1}^{m} b_j y_{t-j} + \varepsilon_t \\ x_t &= \sum_{j=1}^{m} c_j x_{t-j} + \sum_{j=1}^{m} d_j y_{t-j} + \eta_t \end{aligned} \tag{6}$$

where ε_t and η_t are random noise time series assumed to be non-correlated. The causality definition implies that if y_t causes x_t then there exists one or more than one in b_j statistically significant and the reciprocal is also true for statistical significance of c_j. In order to test if the null hypothesis that the parameters included in the VAR model are null, which indicates the absence of causality, critical values of the F statistic are used.

Finally, the results obtained for the cointegration and causality tests allows for the construction and estimation of consistent models to forecast or explain the spot market behavior through current information about the future. The elaboration of such models is based on Distributed Lag Models (DLM), generalizations of the autoregressive models which, apart from considering lagged values of the response variable, also take into account past values of other independent variables. In one of its simplest forms, as mentioned in [13], a DLM can be written as follows:

$$y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=0}^{q} \beta_j x_{t-j} + \varepsilon_t \tag{7}$$

where y_t corresponds to the return of the dependent variable, while y_{t-j} stand for its j lagged returns and x_t are the lagged returns of the independent variable. In this formulation, the model is designated as DLM (p, q). It is also possible to include other exogenous variables in the regression. For instance, when an additional independent variable is taken into account, the above mentioned model will be referred to as DLM (p, q, k). In this work, the crude oil spot price returns correspond to the response variable of interest and its behavior can be explained through its first lagged term and the first lagged terms of the future contracts price returns. This representation refers to the mean part of the regression but it is also important to put foward consistent processes to estimate the variance, or the volatility of the price returns. To that end, several Autoregressive Conditional Heteroskedasticity (ARCH) models are proposed. In this work, besides the multivariate version of the original ARCH model, proposed by [6], the following models were tested: the GARCH model, a straightfoward generalization of the ARCH process which also takes into account past lags of the conditional variance, first proposed by [3]; Exponential GARCH, proposed by [21], which considers asymmetric shocks in the price returns; and IGARCH, proposed by [7], a particular case of the GARCH model that is quite similar to the exponentially weighted moving average (EWMA) model. For a better understanding on the existing variations of the original ARCH model, an extensive list is available in [4]). Finally, since the normality assumption is not accepted for the price returns time series here involved, the Student's *t* distribution was used. This distribution has proved to be adequate for the vast majority of financial assets price returns and also has the attractiveness of allowing for estimation with different degrees of freedom, according to each time series involved.

Table 3: Number of cointegrating relations per model – Brent

Brent	Data Trend	None	None	Linear	Linear	Quadratic
Involved Series	Level Data	no intercept	no intercept	no intercept	no intercept	No intercept
	Equation	no trend	no trend	no trend	no trend	no trend
spot – future 1	LR_{trace}	1	1	1	1	2
	LR_{max}	1	1	1	1	2
spot – future 2	LR_{trace}	1	1	1	1	2
	LR_{max}	1	1	1	1	2
spot – future 3	LR_{trace}	0	1	1	1	2
	LR_{max}	1	1	1	1	2
spot – future 4	LR_{trace}	0	1	1	1	2
	LR_{max}	0	1	1	1	2

RESULTS OBTAINED AND DISCUSSION

The results obtained by the Johansen cointegration test (see [18]) between the spot price and each future contract price time series are shown in Tables 3 and 4. Table 3 refers to the Brent crude oil prices whereas Table 4 displays the results for the WTI crude oil prices. As mentioned earlier, the test was applied on the original price time series, not using any log transformations, as this could invalidate the cointegration restrictions. The analysis was carried out for each pair of variables, the first one being the spot market prices and the second corresponding to each of the first four future crude oil contracts taken into account.

Data regarding the optimal number of lagged values of the variables included in each regression and their corresponding values for the information criteria was purposely omitted from the tables to avoid exposing unnecessary information. It is just worth mentioning that the ideal number of lagged terms included in each bivariate analysis ranged between 2 and 4, being unusual to find lower AIC and BIC criteria for higher numbers of lagged terms.

The results for the Brent crude oil indicate the presence of cointegrating relations between the spot prices and all four future contract prices. In practical terms, the information taken from the first two [17] deterministic trend cases should not be taken into consideration, since the plot analysis already shows that the price time series possess some sort of upward trend in their values over the years.

Table 4: Number of cointegrating relations per model – WTI

WTI	Data Trend	None	None	Linear	Linear	Quadratic
Involved Series	**Level Data**	no intercept	no intercept	no intercept	no intercept	no intercept
	Equation	no trend	no trend	no trend	no trend	no trend
spot – future 1	LR_{trace}	1	1	2	1	2
	LR_{max}	1	1	2	1	2
spot – future 2	LR_{trace}	0	1	2	1	2
	LR_{max}	1	0	2	1	0
spot – future 3	LR_{trace}	0	0	0	0	0
	LR_{max}	0	0	0	0	0
spot – future 4	LR_{trace}	0	0	0	0	0
	LR_{max}	0	0	0	0	0

The situation described for the WTI crude oil somehow differs from the previous case: the cointegration test results are only consistent when the first future contract is used in the regression. The relationship between the WTI crude oil spot market and the second WTI future contract is somewhat unclear, with some statistics test indicating the existence of cointegrating relations for specific trends in the data, but not for all of them. For contracts 3 and 4, the long-run relationship practically ceases to exist. Therefore, it can be inferred that the Brent crude oil market presents a higher informational efficiency degree than the WTI crude oil market over recent years. It is also worth noting that, for the WTI market, as longer expiry dates are taken into account, the long run relationship between the spot market and the future market becomes less strong, as this allows for higher levels of uncertainties in real price estimation and the cointegration results bear this out indicating the absence of cointegration relations for long future contracts.

Table 5: Granger causality test results - Brent

Lags	2		5		10		20	
Null Hypothesis (Ho)	F Stat	p-value	F Stat	p-value	F Stat	p-value	F Stat	p-value
Future 1 does not cause Spot	82.007	0.000	34.710	0.000	17.461	0.000	8.684	0.000
Spot does not cause Future 1	0.977	0.377	1.315	0.255	0.992	0.448	1.774	0.019
Future 2 does not cause Spot	69.284	0.000	32.055	0.000	16.184	0.000	8.096	0.000
Spot does not cause Future 2	1.250	0.287	1.120	0.348	0.709	0.717	1.608	0.044
Future 3 does not cause Spot	59.994	0.000	28.623	0.000	14.412	0.000	7.324	0.000
Spot does not cause Future 3	1.938	0.145	1.381	0.229	0.847	0.583	1.636	0.039
Future 4 does not cause Spot	54.785	0.000	26.460	0.000	13.292	0.000	6.850	0.000
Spot does not cause Future 4	2.284	0.102	1.545	0.173	0.954	0.483	1.671	0.033

Upon completion of the cointegration tests, the causality relationships between the spot market and each contract future were also examined, so that it can be inferred the predictive power that each future contract has when estimating the spot market price. The mentioned Granger Causality test results are listed in Tables 5 and 6, respectively, for the Brent and WTI oil prices.

As shown in Table 5 the four contracts studied represent useful information to explain the Brent crude oil spot market while the reciprocal is not always true. There is indeed a causal relationship from the spot market to some future contracts, but this relationship is somewhat weak and only appears when higher lagged terms are taken into consideration in the models. With regards to the WTI crude oil, the causality results ratify what was already to be expected after applying the cointegration tests: only future contract 1, the front month contract, has a considerable causal relationship on spot market price behavior. All other contracts do not constitute useful predictive information on the spot market price formation. As shown in Table 6 the WTI spot price Granger causes future contract null hypothesis could not be rejected for any of the four future contracts studied in this work.

Table 6: Granger causality test results - WTI

Lags	2		5		10		20	
Null Hypothesis (Ho)	F Stat	p-value	F Stat	p-value	F Stat	p-value	F Stat	p-value
Future 1 does not cause Spot	3.722	0.025	1.959	0.083	2.044	0.027	1.298	0.178
Spot does not cause Future 1	0.206	0.814	0.794	0.554	1.408	0.173	1.076	0.373
Future 2 does not cause Spot	0.167	0.846	0.134	0.985	0.733	0.694	8.872	0.622
Spot does not cause Future 2	0.464	0.629	0.342	0.888	0.689	0.735	0.851	0.650
Future 3 does not cause Spot	0.181	0.834	0.153	0.979	0.857	0.574	0.998	0.464
Spot does not cause Future 3	0.677	0.509	0.360	0.876	0.709	0.716	0.949	0.525
Future 4 does not cause Spot	0.255	0.775	0.166	0.975	0.917	0.517	0.989	0.476
Spot does not cause Future 4	0.743	0.476	0.376	0.865	0.721	0.706	0.917	0.565

Observing the results of the cointegration and causality tests, procedures to estimate consistent forecast models were taken. The expectation is that future contracts time series which showed to be cointegrated with the spot prices and presented the highest predictive powers over the spot market, given by causality test, should be the relevant explanatory variables in the spot price returns forecasting models. As both cointegration and causality tests results for the WTI crude oil indicated poor association and explanatory power of future contracts over the spot price formation, this work only proposed and tested models for the Brent crude oil. Hence, in order to explain the Brent spot price returns, a total of 70 models were tested, in which its mean was represented by a Distributed Lag Model using the first lagged term of the spot price returns time series and the first lagged terms of the future contracts price returns. As for the variance estimation, several GARCH models were proposed, as mentioned in Section 3. Among the estimated models, the results were statistically significant in 32 of them. When examining the results of the statistically significant models, it could be noticed that the first lagged price returns of the second contract did not have any influence on spot price returns estimation. Also, it could be stated that the IGARCH volatility model provided the best estimation results for the variance, followed by the EGARCH model. As for the mean estimation, the first past values of contracts 3 and 4 price returns were the ones which presented the most significant results. Finally, it is worth highlighting that the optimal number of degrees of freedom for the *t* statistic chosen to represent the data involved was approximately 8, which indicates fatter tails for the selected dataset distribution. The selection criteria to choose the model which best fits the data were: standard error of the regression; Akaike criterion; and Schwarz criterion. The combination which produced the best results was the following:

$$\begin{aligned} &(R_t | I_{t-1}) \sim Student(\mu_t ; \sigma_t^2 ; \nu) \\ &\mu_t = \alpha R_{t-1} + \beta RF3_{t-1} + \delta RF4_{t-1} \\ &\sigma_t^2 = \lambda e_{t-1}^2 + \gamma \sigma_{t-1}^2 \end{aligned} \quad (8)$$

where: R_t is the Brent crude oil spot price return, at period *t*; I_{t-1} is the available information at period *t* - 1; μ_t is the mean Brent price return at period *t*; $RF3_{t-1}$ is the o Brent price return in future contract 3, at period *t* - 1; $RF4_{t-1}$ $\mathrm{RF4}_{t\text{-}1}$is the o Brent price return in future contract 4, at period *t*-1; σ_t is the standard deviation of the Brent spot price return at period t; and $e_t = (R_t - \mu_t)$. $e_t = (R_t - \mu_t)$ And Table 7 presents the results for the estimation parameters of the selected model.

Table 7: Estimation results of the selected model

Parameter	Value	Std Deviation	t Statistic	p-value
Mean Equation				
α	-0.385	0.042	-9.259	0.000
β	5.512	0.913	6.038	0.000
δ	-5.049	0.924	-5.467	0.000
Variance Equation				
λ	0.035	0.008	4.227	0.000
γ	0.965	0.008	116.881	0.000
Degrees of Freedom	8.406	1.756	4.786	0.000
Standard Error Regression	0.017	AIC criterion		-5.463
Sum Squared Residuals	0.255	BIC criterion		-5.437
Durbin-Watson Statistics	2.112	Log Likelihood		2566.935

CONCLUSIONS AND FINAL REMARKS

This work aimed to investigate the influence that different oil future contracts have on oil spot price formation. The analysis was conducted taking into account daily observations for the first four contracts, classified according to the difference between the expiry date and the moment of negotiation, and for the spot prices of two of the main crude oils traded nowadays - Brent and WTI. At first, the study sought to examine the long-run relationship between the spot market and the future contracts for each. The results were quite satisfactory for the Brent crude, where the tests indicated that cointegration hypothesis could not be rejected in any of the four cases studied. On the other hand, the results for the WTI crude oil were only consistent when the first future contract was taken into account. The relationship between the WTI spot market and the second WTI future contract was somewhat unclear and contracts 3 and 4 did not seem to share any long term equilibrium relationship with the spot market at all.

The causal relationships between spot and future contracts were also examined and the results suggested that all Brent future contracts studied seemed to have strong predictive power over spot market price behavior. Conversely, apart from the first contract, WTI future contracts do not seem to have any considerable influences on WTI spot market prices.

Finally, this study proposes to estimate an effective model to explain the Brent oil spot price returns behavior through previous information regarding the first lagged terms of the spot price returns time series and the first lagged terms of the future contracts price returns. According to the selection models criteria used in this work, among 70 estimated models, the results were statistically significant in 32 of them. Surprisingly, the selected model only took into account information concerning spot prices and future contracts 3 and 4.

It can be stated that all objectives were achieved, as it was possible to establish coherent and consistent criterion to contrast the association between the spot market and each of the four contracts studied for each crude oil type, and it was possible to estimate significant forecast models, from the statistical point of view. It should be emphasized, though, that the results were taken for a specific time period and there may be significant differences when other samples are taken into account. Besides that, it must be highlighted that other statistical inference methodologies may be proposed to investigate the relationship between the involved markets here studied, and the selection of the most appropriate econometric tests and forecast models are linked to a huge variety of criteria.

References

Akaike, H., *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 1974. AC-19 (6): p. 716-723.

Bigman, D., Goldfarb D. and Schechtman E., *Futures market efficiency and the time content of the information sets*. The Journal of Futures Markets, 1983. 3: p. 321-334.

Bollerslev, T., *Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics*, 1986. 31: p. 307-327.

Bollerslev, T., Glossary to ARCH (GARCH): in Bollerslev, T., Russel, J., Watson, M. (Org.). Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle, 2009. Oxford University Press, Oxford.

Dickey, D. A. and Fuller, W. A., *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root*. Econometrica, 1981. 49: p. 1057-1072.

Engle, R., *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation*. Econometrica, 1982. 50: p. 987-1008.

Engle, R. and Bollerslev, T., *Modeling the persistence of conditional variances*. Econometric Reviews, 1986. 5(1): p. 1-50.

Engle, R. and Granger, C., *Cointegration and error correction: representation, estimation and testing*. Econometrica, 1987. 55: p. 251-276.

Granger, C., *Investigating Causal Relations by Econometric Models and Cross-spectral Methods*. Econometrica, 1969. 37(3): p. 424-438.

Granger, C., *Some Properties of Time Series Data and their use in Econometric Model Specification*. Journal of Econometrics, 1981. 16: p. 121-130.

Granger, C. and Hallman, J. J., *Long Memory Series with Attractors*, Oxford Bulletin of Economic and Statistics, 1991. 53: p. 11-46.

Gujarati, D. N., *Basic Econometrics*. 4th Edition. 2003, New York: McGraw-Hill Book Company.

Gujarati, D. N., *Econometrics by Example*. 2012, New York: Palgrave Macmilian.

Jarque, C. M. and Bera, A. K., *A Test for Normality of Observations and Regression Residuals*. Revue Internationale de Statistique, 1987. 55(2): p. 163-172.

Johansen, S., *Statistical analysis of cointegration vectors*. Journal of Economic Dynamics and Control, 1988. 12(2): p. 231-254.

Johansen, S., *Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models*. Econometrica, 1991. 59(6): p. 1551-1580.

Johansen, S., *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: University of Oxford, 1995.

Johansen, S. and Juselius, K., *Maximum likelihood estimation and inference on cointegration — with applications to the demand for money*. Oxford Bulletin of Economics and Statistics, 1990. 52: p. 169-210.

Maberly, E. D., *Testing futures market efficiency—a restatement*. The Journal of Futures Markets, 1985. 5: p. 425-432.

MacKinnon, J., Haug A. and Michelis, L., *Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration*, Journal of Applied Econometrics, 1999. 14(5): p. 563-577.

Nelson, D., *Conditional Heteroskedasticity in Asset Returns: A New Approach*. Econometrica, 1991. 59: p. 347-370.

Schwarz, G., *Estimating the Dimension of a Model*. The Annals of Statistics, 1978. 6 (2): p. 461-464.