

“The Relationship between Crude Oil Spot and Futures Prices: Cointegration, Linear and Nonlinear Causality”

Stelios D. Bekiros^{*}, Cees G.H Diks

*Center for Nonlinear Dynamics in Economics and Finance (CeNDEF)
Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11,
1018 WB Amsterdam, The Netherlands*

Abstract

The present study investigates the linear and nonlinear causal linkages between daily spot and futures prices for maturities of one, two, three and four months of West Texas Intermediate (WTI) crude oil. The data cover two periods October 1991-October 1999 and November 1999-October 2007, with the latter being significantly more turbulent. Apart from the conventional linear Granger test we apply a new nonparametric test for nonlinear causality by Diks and Panchenko after controlling for cointegration. In addition to the traditional pairwise analysis, we test for causality while correcting for the effects of the other variables. To check if any of the observed causality is strictly nonlinear in nature, we also examine the nonlinear causal relationships of VECM filtered residuals. Finally, we investigate the hypothesis of nonlinear non-causality after controlling for conditional heteroskedasticity in the data using a GARCH-BEKK model. Whilst the linear causal relationships disappear after VECM cointegration filtering, nonlinear causal linkages in some cases persist even after GARCH filtering in both periods. This indicates that spot and futures returns may exhibit asymmetries and statistically significant higher-order moments. Moreover, the results imply that if nonlinear effects are accounted for, neither market leads or lags the other consistently, videlicet the pattern of leads and lags changes over time.

JEL classification: C14; C51; Q40

Keywords: Nonparametric nonlinear causality; Oil Futures Market; Cointegration; GARCH-BEKK filtering

^{*} E-mail address: S.Bekiros@uva.nl (corresponding author); Tel.: + 31 20 525 5375; fax.: + 31 20 525 5283.

1. Introduction

The role of futures markets in providing an efficient price discovery mechanism has been an area of extensive empirical research. Several studies have dealt with the lead-lag relationships between spot and futures prices of commodities with the objective of investigating the issue of market efficiency. Garbade and Silber (1983) first presented a model to examine the price discovery role of futures prices and the effect of arbitrage on price changes in spot and futures markets of commodities. The Garbade-Silber model was applied to the feeder cattle market by Oellermann *et al.* (1989) and to the live hog commodity market by Schroeder and Goodwin (1991), while a similar study by Silvapulle and Moosa (1999) examined the oil market. Bopp and Sitzer (1987) tested the hypothesis that futures prices are good predictors of spot prices in the heating oil market, while Serletis and Banack (1990) and Chen and Lin (2004) tested for market efficiency using cointegration analysis. Crowder and Hamed (1993) and Sadorsky (2000) also used cointegration to test the simple efficiency hypothesis and the arbitrage condition for crude oil futures. Finally, Schwarz and Szakmary (1994) examined the price discovery process in the markets of crude and heating oil.

In theory, since both futures and spot prices “reflect” the same aggregate value of the underlying asset and considering that instantaneous arbitrage is possible, futures should neither lead nor lag the spot price. However, the empirical evidence is diverse, although the majority of studies indicate that futures influence spot prices but not vice versa. The usual rationalization of this result is that the futures prices respond to new information more quickly than spot prices, due to lower transaction costs and flexibility of short selling. With reference to the oil market, if new information indicates that oil

prices are likely to rise, perhaps because of an OPEC decision to restrict production, or an imminent harsh winter, a speculator has the choice of either buying crude oil futures or spot. Whilst spot purchases require more initial outlay and may take longer to implement, futures transactions can be implemented immediately by speculators without an interest in the physical commodity per se and with little up-front cash. Moreover, hedgers who are interested for the physical commodity and have storage constraints will buy futures contracts. Therefore, both hedgers and speculators will react to the new information by preferring futures rather than spot transactions. Spot prices will react with a lag because spot transactions cannot be executed so quickly (Silvapulle and Moosa, 1999). Furthermore, the price discovery mechanism, as illustrated by Garbade and Silber (1983), supports the hypothesis that futures prices lead spot prices. Their study of seven commodity markets indicated that, although futures markets lead spot markets, the latter do not just echo the former. Futures trading can also facilitate the allocation of production and consumption over time, particularly by providing a market scheme in inventory holdings (Houthakker, 1992). In this case, if futures prices for late deliveries are above those for early ones, delay of consumption becomes attractive and changes in futures prices result in subsequent changes in spot prices. According to Newberry (1992) futures markets provide opportunities for market manipulation by the better informed or larger at the expense of other market participants. For example, it is profitable for the OPEC to intervene in the futures market to influence the production decisions of its competitors in the spot market. Finally, support for the hypothesis that causality runs from futures to spot prices can also be found in the model of determination of futures prices proposed by Moosa and Al-Loughani (1995). In their model the futures price is determined by

arbitrageurs whose demand depends on the difference between the arbitrage and actual futures price and by speculators whose demand for futures contracts depends on the difference between the expected spot and the actual futures price. The reference point in both cases is the futures price and not the spot price (Silvapulle and Moosa, 1999).

There is also empirical evidence that spot prices lead futures prices. Specifically, in the study of Moosa (1996) a spot price change triggers action from all kinds of market participants and this subsequently changes the futures price. Initially, arbitrageurs will react to the violation of the cost-of-carry condition¹ and then speculators will revise their expectation of the spot price and respond to the disparity between expected spot and futures price. Similarly, speculators who act upon the expected futures price will revise their expectation responding to the disparity between current and expected futures prices. Finally, in few studies causality is reported to be bi-directional. Kawaller *et al.* (1988) introduced the principle that both spot and futures prices are affected by their past history, as well as by current market information. They argue that potential lead - lag patterns dynamically change as new information arrives. At any time point each may lead the other, as market participants filter information relevant to their positions, which may be spot or futures. So far, the hypothesis that futures prices lead spot prices is stronger in terms of empirical evidence and more compelling. Thus, further empirical testing is required to infer on this issue with respect to the crude oil market.

The recent empirical evidence on causality is invariably based on the Granger test (Granger, 1969). The conventional approach of testing for Granger causality is to assume

¹ The relationship between futures and spot prices can be summarized as $F = Se^{(c-y)T}$ in terms of what is known as the cost-of-carry. In that, y is the convenience yield (market's expectations of the future availability of the commodity), T is the period to maturity, and c the cost-of-carry which equals the storage cost plus the cost of financing a commodity minus the income earned on the commodity (Hull, 2000).

a parametric linear, time series model for the conditional mean. Although it requires the linearity assumption this approach is appealing, since the test reduces to determining whether the lags of one variable enter into the equation for another variable. Moreover, tests based on residuals will be sensitive only to causality in the conditional mean while covariables may influence the conditional distribution of the response in nonlinear ways. Baek and Brock (1992) noted that parametric linear Granger causality tests have low power against certain nonlinear alternatives. Recent work has revealed that nonlinear structure indeed exists in spot and futures returns. These nonlinearities are normally attributed to nonlinear transaction cost functions, the role of noise traders, and to market microstructure effects (Abhyankar, 1996; Chen and Lin, 2004; Silvapulle and Moosa, 1999). In view of this, nonparametric techniques are appealing because they place direct emphasis on prediction without imposing a linear functional form. Various nonparametric causality tests have been proposed in the literature. The test by Hiemstra and Jones (1994), which is a modified version of the Baek and Brock (1992) test, is regarded as a test for a nonlinear dynamic relationship. The Hiemstra and Jones test relaxes Baek and Brock's assumption that the time series to which the test is applied are mutually and individually independent and identically distributed. Instead, it allows each series to display weak (or short-term) temporal dependence. When applied to the residuals of vector autoregressions, the Hiemstra and Jones test can be used to determine whether nonlinear dynamic relations exist between variables by testing whether the past values influence present and future values. However, Diks and Panchenko (2005, 2006) demonstrate that the Hiemstra and Jones test can severely over-reject if the null hypothesis of non-causality is true i.e., the Hiemstra and Jones test has serious size

distortion problems. As an alternative Diks and Panchenko (2006) developed a new test statistic that overcomes these limitations.

Empirically it is important to take into account the possible effects of cointegration on both linear and nonlinear Granger causality tests. Nonstationary variables are said to be cointegrated if a stationary linear combination exists (Engle and Granger, 1987). This linear combination is called the cointegrating equation and may be interpreted as a long-run equilibrium relationship among the variables. Controlling for cointegration is necessary because it affects the specification of the model used for causality testing. If the series are cointegrated, then causality testing should be based on a Vector Error Correction model (VECM) rather than an unrestricted VAR model (Johansen, 1988). When cointegration is not modelled, evidence may vary significantly towards detecting linear and nonlinear causality between the predictor variables. Specifically, the absence of cointegration could mean the violation of the necessary condition for the simple efficiency hypothesis, which implies that the futures price is not an unbiased predictor of the spot price at maturity. This implies an absence of a long-run relationship between spot and futures prices, as it was reported in works of Chowdhury (1991), Krehbiel and Adkins (1993), Crodwer and Hamed (1993). Alternatively, based on the cost-of-carry relationship, a failure to find cointegration may be attributed to the nonstationarity of the other components of this relationship such as the interest rate or the convenience yield (Moosa and Al-Loughani, 1995, and Moosa, 1996).

The aim of the present study is to test for the existence of linear and nonlinear causal lead-lag relationships between spot and futures prices of West Texas Intermediate (WTI) crude oil, which is used as an indicator of world oil prices and is the underlying

commodity of New York Mercantile Exchange's (NYMEX) oil futures contracts. We apply a three-step empirical framework for examining dynamic relationships between spot and futures prices. First, we explore nonlinear and linear dynamic linkages applying the nonparametric Diks-Panchenko causality test, and after controlling for cointegration, a parametric linear Granger causality test. In the second step, after filtering the return series using the properly specified VAR or VECM model, the series of residuals are examined by the nonparametric Diks-Panchenko causality test. In addition to applying the usual bivariate VAR or VECM model to each pair of time series, we also consider residuals of a full five-variate model to account for the possible effect of the other variables. This step ensures that any remaining causality is strictly nonlinear in nature, as the VAR or VECM model has already purged the residuals of linear dependence. Finally, in the last step, we investigate the null hypothesis of nonlinear non-causality after controlling for conditional heteroskedasticity in the data using a GARCH-BEKK model, again both in a bivariate and in a five-variate representation. Our approach incorporates the entire variance-covariance structure of the spot and future prices interrelationship. The empirical methodology employed with the multivariate GARCH-BEKK model can not only help to understand the short-run movements, but also explicitly capture the volatility persistence mechanism. Improved knowledge of the direction and nature of causality and interdependence between the spot and futures markets, and consequently the degree of their integration, will expand the information set available to policymakers, international portfolio managers and multinational corporations for decision-making.

The remainder of the paper is organized as follows. Section 2 briefly reviews the linear Granger causality framework and provides a description of the Diks-Panchenko

nonparametric test for nonlinear Granger causality. Section 3 describes the data used and Section 4 presents the results. Section 5 concludes with a summary and suggestions for future research.

2. The Nonparametric Diks – Panchenko Causality Test

Granger (1969) causality has turned out to be a useful notion for characterizing dependence relations between time series in economics and econometrics. Assume that $\{X_t, Y_t; t \geq 1\}$ are two scalar-valued strictly stationary time series. Intuitively $\{X_t\}$ is a strictly Granger cause of $\{Y_t\}$ if past and current values of X contain additional information on future values of Y that is not contained only in the past and current Y_t values. Let $F_{X,t}$ and $F_{Y,t}$ denote the information sets consisting of past observations of X_t and Y_t up to and including time t , and let ‘ \sim ’ denote equivalence in distribution. Then $\{X_t\}$ is a Granger cause of $\{Y_t\}$ if, for $k \geq 1$:

$$(Y_{t+1}, \dots, Y_{t+k}) | (F_{X,t}, F_{Y,t}) \not\sim (Y_{t+1}, \dots, Y_{t+k}) | F_{X,t} \quad (1)$$

In practice $k = 1$ is used most often, in which case testing for Granger non-causality amounts to comparing the one-step-ahead conditional distribution of $\{Y_t\}$ with and without past and current observed values of $\{X_t\}$. A conventional approach of testing for Granger causality among stationary time series is to assume a parametric, linear, time series model for the conditional mean $E(Y_{t+1} | (F_{X,t}, F_{Y,t}))$. Then, causality can be tested by comparing the residuals of a fitted autoregressive model of Y_t with those obtained by regressing Y_t on past values of both $\{X_t\}$ and $\{Y_t\}$ (Granger, 1969). Now, assume delay

vectors $\mathbf{X}_t^{\ell_x} = (X_{t-\ell_x+1}, \dots, X_t)$ and $\mathbf{Y}_t^{\ell_y} = (Y_{t-\ell_y+1}, \dots, Y_t)$, ($\ell_x, \ell_y \geq 1$). In practice the null hypothesis that past observations of $\mathbf{X}_t^{\ell_x}$ contain no additional information (beyond that in $\mathbf{Y}_t^{\ell_y}$) about Y_{t+1} is tested, i.e.:

$$H_0 : Y_{t+1} | (\mathbf{X}_t^{\ell_x}; \mathbf{Y}_t^{\ell_y}) \sim Y_{t+1} | \mathbf{Y}_t^{\ell_y} \quad (2)$$

For a strictly stationary bivariate time series Eq. (2) comes down to a statement about the invariant distribution of the $(\ell_x + \ell_y + 1)$ -dimensional vector $\mathbf{W}_t = (\mathbf{X}_t^{\ell_x}, \mathbf{Y}_t^{\ell_y}, Z_t)$ where $Z_t = Y_{t+1}$. To keep the notation compact, and to bring about the fact that the null hypothesis is a statement about the invariant distribution of $(\mathbf{X}_t^{\ell_x}, \mathbf{Y}_t^{\ell_y}, Z_t)$ we drop the time index and also $\ell_x = \ell_y = 1$ is assumed. Hence, under the null, the conditional distribution of Z given $(X, Y) = (x, y)$ is the same as that of Z given $Y = y$. Further, Eq. (2) can be restated in terms of ratios of joint distributions. Specifically, the joint probability density function $f_{X,Y,Z}(x, y, z)$ and its marginals must satisfy the following relationship:

$$\frac{f_{X,Y,Z}(x, y, z)}{f_Y(y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \cdot \frac{f_{Y,Z}(y, z)}{f_Y(y)} \quad (3)$$

This explicitly states that X and Z are independent conditionally on $Y = y$ for each fixed value of y . Diks and Panchenko (2006) show that this reformulated H_0 implies:

$$q \equiv E[f_{X,Y,Z}(X, Y, Z) f_Y(Y) - f_{X,Y}(X, Y) f_{Y,Z}(Y, Z)] = 0 \quad (4)$$

Let $\hat{f}_W(W_i)$ denote a local density estimator of a d_W - variate random vector \mathbf{W} at W_i defined by $\hat{f}_W(W_i) = (2\varepsilon_n)^{-d_W} (n-1)^{-1} \sum_{j, j \neq i} I_{ij}^W$ where $I_{ij}^W = I(\|W_i - W_j\| < \varepsilon_n)$ with $I(\cdot)$ the indicator function and ε_n the bandwidth, depending on the sample size n . Given this estimator, the test statistic is a scaled sample version of q in Eq. (4):

$$T_n(\varepsilon_n) = \frac{n-1}{n(n-2)} \cdot \sum_i (\hat{f}_{X,Z,Y}(X_i, Z_i, Y_i) \hat{f}_Y(Y_i) - \hat{f}_{X,Y}(X_i, Y_i) \hat{f}_{Y,Z}(Y_i, Z_i)) \quad (5)$$

For $\ell_X = \ell_Y = 1$, if $\varepsilon_n = Cn^{-\beta}$ $\left(C > 0, \frac{1}{4} < \beta < \frac{1}{3} \right)$ then Diks and Panchenko (2006) prove

under strong mixing that the test statistic in Eq. (5) satisfies:

$$\sqrt{n} \frac{(T_n(\varepsilon_n) - q)}{S_n} \xrightarrow{D} N(0,1) \quad (6)$$

where \xrightarrow{D} denotes convergence in distribution and S_n is an estimator of the asymptotic variance of $T_n(\cdot)$ (Diks and Panchenko, 2006).

3. Data and preliminary analysis

The data consist of time series of daily spot and futures prices for maturities of one, two, three and four months of West Texas Intermediate (WTI), also known as Texas Light Sweet, which is a type of crude oil used as a benchmark in oil pricing and the underlying commodity of New York Mercantile Exchange's (NYMEX) oil futures contracts. The NYMEX futures price for crude oil represents, on a per-barrel basis, the market-determined value of a futures contract to either buy or sell 1,000 barrels of WTI at a specified time. The NYMEX market provides important price information to buyers and sellers of crude oil around the world, although relatively few NYMEX crude oil contracts are actually executed for physical delivery.

The data cover two equally sampled periods, namely PI which spans October 21, 1991 to October 29, 1999 (2061 observations) and PII November 1, 1999 to October 30, 2007 (2061 observations). The segmentation of the sample corresponds roughly to the reduction in OPEC spare capacity (defined as the difference between sustainable capacity

and current OPEC crude oil production) and to the increase in the United States' gasoline consumption and imports, both of which occurred after 1999. The effect of these events on price dynamics is evident and it can be summarized in the accelerated rise of the average level of oil prices and in the increased volatility. Additionally, in PII markets witnessed more occasional spikes in crude prices. Figure 1 displays the spot and future price and returns time series. The following notation is used: "WTI Spot" is the spot price and "WTI F1", "WTI F2", "WTI F3" and "WTI F4" are the futures prices for maturities of one, two, three and four months respectively. Descriptive statistics for WTI spot and futures log-daily returns are reported in Table 1. Specifically, the returns are defined as $r_t = \ln(P_t) - \ln(P_{t-1})$, where P_t is the closing price on day t . The differences between the two periods are quite evident in Table 1 where a significant increase in variance can be observed as well as a higher dispersion of the returns distribution in Period II reflected in the lower kurtosis. Additionally, Period II witnessed many occasional negative spikes as it can be also inferred from the skewness. The results from testing nonstationarity are presented in Table 2.

[Insert Table 1 here]

[Insert Figure 1 here]

[Insert Table 2 here]

Specifically, Table 2 reports the Augmented Dickey-Fuller (ADF) test for the logarithmic levels and log-daily returns. The lag lengths which are consistently zero in all cases were selected using the Schwartz Information Criterion (SIC). All the variables appear to be nonstationary in log-levels and stationary in log-returns based on the reported p -values. Table 1 also reports the correlation matrix at lag 0 (contemporaneous correlation) for

both periods. Significant sample cross-correlations are noted for spot and futures returns indicating a high interrelationship between the two markets. However, since linear correlations cannot be expected to fully capture the long-term dynamic linkages in a reliable way, these results should be interpreted with caution. Consequently, what is needed is a long-term causality analysis.

4. Empirical results

The empirical methodology comprises three steps. In the first pre-filtering step, we explore the linear and nonlinear dynamic linkages applying a Granger causality test based on a VECM specification on the log-price levels and the nonparametric Diks-Panchenko test on the log-differenced time series of the spot and futures prices. Then, we implement both pairwise and five-variate VECM filtering on the log-price series, and the residuals are examined by the Diks-Panchenko test. Finally, we investigate the hypothesis of nonlinear Granger non-causality after controlling for conditional heteroskedasticity using a GARCH-BEKK filter, again in a bivariate and a five-variate representation. Additionally, in the last two steps we consistently apply a linear Granger causality test on the “whitened” residuals via a VAR specification (i.e., no cointegration detected on the residuals) in order to investigate whether any remaining causality is strictly nonlinear in nature or not.

The results are reported in the corresponding panels of Tables 3 and 4. In order to overcome the difficulty of presenting large tables with numbers we use the following simplifying notation: “ ** ” indicating that the corresponding p -value of a particular

causality test is smaller than 1% and “ * ” that the corresponding p -value of a test is in the range 1-5%; Directional causalities will be denoted by the functional representation \rightarrow .

4.1 Causality testing on raw data

The linear Granger causality test is usually constructed in the context of a reduced-form vector autoregression (VAR). Let \mathbf{Y}_t the vector of endogenous variables and ℓ number of lags. Then the VAR(ℓ) model is given as follows:

$$\mathbf{Y}_t = \sum_{s=1}^{\ell} \mathbf{A}_s \mathbf{Y}_{t-s} + \boldsymbol{\varepsilon}_t \quad (7)$$

where $\mathbf{Y}_t = [Y_{1t}, \dots, Y_{\ell t}]$ the $\ell \times 1$ vector of endogenous variables, \mathbf{A}_s the $\ell \times \ell$ parameter matrices and $\boldsymbol{\varepsilon}_t$ the residual vector, for which $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \begin{cases} \boldsymbol{\Sigma}_{\varepsilon} & t = s \\ \mathbf{0} & t \neq s \end{cases}$.

Specifically, in case of two stationary time series $\{X_t\}$ and $\{Y_t\}$ the bivariate VAR model is given by:

$$\begin{aligned} X_t &= A(\ell)X_t + B(\ell)Y_t + \varepsilon_{X,t} \\ Y_t &= C(\ell)X_t + D(\ell)Y_t + \varepsilon_{Y,t} \end{aligned} \quad t = 1, 2, \dots, N \quad (8)$$

where $A(\ell), B(\ell), C(\ell)$ and $D(\ell)$ are all polynomials in the lag operator with all roots outside the unit circle. The error terms are separate i.i.d. processes with zero mean and constant variance. The test whether Y strictly Granger causes X is simply a test of the joint restriction that all the coefficients of the lag polynomial $B(\ell)$ are zero, whilst similarly, a test of whether X strictly Granger causes Y is a test regarding $C(\ell)$. In each case, the null hypothesis of no Granger causality is rejected if the exclusion restriction is rejected. If both joint tests for significance show that $B(\ell)$ and $C(\ell)$ are different from

zero, the series are bi-causally related. However, in order to explore effects of possible cointegration, a VAR in error correction form (Vector Error Correction Model-VECM) is estimated using the methodology developed by Engle and Granger (1987) and expanded by Johansen (1988) and Johansen and Juselius (1990). The bi-variate VECM model has the following form:

$$\begin{aligned}\Delta X_t &= -p_1 \left([1 \quad -\lambda] \cdot \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} \right) + A(\ell)\Delta X_t + B(\ell)\Delta Y_t + \varepsilon_{\Delta X,t} \\ \Delta Y_t &= -p_2 \left([1 \quad -\lambda] \cdot \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} \right) + C(\ell)\Delta X_t + D(\ell)\Delta Y_t + \varepsilon_{\Delta Y,t}\end{aligned} \quad t = 1, 2, \dots, N \quad (9)$$

where $\Delta X_t, \Delta Y_t$ the first differences of X and Y , $[1 \quad -\lambda]$ the cointegration vector and λ the cointegration coefficient. Thus, in case of cointegrated time series, linear Granger causality should be investigated via the VECM specification.

For the pairwise implementation the linear causality testing was carried out using the Granger's test based on a VECM model of the log-prices because all series were found to be cointegrated. The lag lengths of the VECM specification were set using the Wald exclusion criterion and for each pair in PI are (in parenthesis): WTI Spot - WTI F1 (3), WTI Spot - WTI F2 (7), WTI Spot - WTI F3 (3) and WTI Spot - WTI F4 (3). Similarly, in period PII: WTI Spot - WTI F1 (3), WTI Spot - WTI F2 (6), WTI Spot - WTI F3 (6) and WTI Spot - WTI F4 (4). In addition, in PI for all pairs the Johansen test identified two (2) cointegrating vectors using the trace statistic and in PII one (1) cointegrating vector. In case of the five-variate implementation cointegration was also detected and in particular in PI the Johansen test identified five (5) cointegrating vectors while in PII three (3). The number of lags for the 5x5 system in PI was eleven (11) and in PII nine (9).

For the Diks-Panchenko test, in what follows we discuss results for lags $\ell_X = \ell_Y = 1$. Moreover, the test was applied directly on log-returns. To implement the test, the constant C for the bandwidth ε_n was set at 7.5, which is close to the value 8.0 for ARCH processes suggested by Diks and Panchenko (2006). With the theoretical optimal rate $\beta = 2/7$ given by Diks and Panchenko (2006), this implies a bandwidth value of approximately one times the standard deviation of the time series for both PI and PII. Selecting bandwidth values smaller (larger) than one times the standard deviation resulted, in general, in larger (smaller) p -values.

The results presented in Tables 3 and 4 allow for the following remarks: In the pairwise implementation of the linear Granger tests (VECM), strong bi-directional Granger causality between spot and futures prices was detected in both periods with small differences regarding the degree of statistical significance. An exception could be that WTI Spot and WTI F4 present only unidirectional linear relationship WTI Spot \rightarrow WTI F4. On the contrary, the linear causality for the five-variate implementation appears to be uni-directional, mainly in the more volatile and trending period PII and from spot to futures prices regardless of maturity, providing evidence that spot tend to lead futures prices. This indicates that spot prices can be useful in the prediction of futures prices under a 5x5 VECM formulation, i.e., accounting for the contributions of all maturities in the causality detection. Further, there is a causal relationship in PI of WTI Spot \rightarrow WTI F1, WTI F3 \rightarrow WTI Spot and WTI F4 \rightarrow WTI Spot. The nonlinear causality test revealed a bi-directional nonlinear relationship in PI, whereas in PII only uni-directional causality was detected from Spot to WTI F1, WTI F2 and WTI F3 returns, excluding WTI F4.

[Insert Table 3 here]

[Insert Table 4 here]

4.2 Causality testing on VECM-filtered residuals

The results from the previous step suggest that there are significant and persistent linear and nonlinear causal linkages between the spot and futures prices. However, even though we found nonlinear causality, the Diks-Panchenko test should be reapplied to the filtered VECM-residuals to ensure that any causality found is strictly nonlinear in nature. The number of lags and the number of cointegrating vectors identified for the VECM specification were reported in the previous section. Moreover, a linear Granger test is applied to the filtered residuals to conclude on a remaining linear structure even after filtering. The causality on the filtered residuals was investigated with a VAR specification (the null of no cointegration was not rejected) and the lags were determined using the Schwartz Information Criterion (SIC).

The pairwise implementation of the Granger tests after VECM filtering shows that the linear causal relationships detected on the raw returns have now disappeared. In fact none of the previously mentioned causalities appear or any other new ones have emerged after linear filtering. Similarly, no causal relationship could be detected after five-variate filtering. The application of the nonlinear test on the VECM residuals, both in the bivariate and five-variate implementation, points towards the preservation of the bi-directional causality reported in PI on the raw log-returns. In PII the nonlinear causal relationships $WTI\ Spot \rightarrow WTI\ F2$, $WTI\ Spot \rightarrow WTI\ F3$ have vanished, while $WTI\ Spot \rightarrow WTI\ F1$ remains, albeit statistically less significant. Interestingly, in the same

period, a uni-directional causality from futures to spot returns has now emerged for all maturities.

The nature and source of the detected nonlinearities are different from that of the linear Granger causality and may also imply a temporary, or long-term, causal relationship between the spot and futures markets. For instance, excess volatility in PII might have induced nonlinear causality. The nature of the volatility transmission mechanism can be investigated after controlling for conditional heteroskedasticity using a GARCH-BEKK model, in a bi-variate and five-variate representation.

4.3 Causality testing on GARCH-BEKK filtered VECM-residuals

The use of the Diks-Panchenko test on filtered data with a multivariate GARCH model enables one to determine whether the posited model is sufficient to describe the relationship among the series. If the statistical evidence of nonlinear Granger causality lies in the conditional variances and covariances then it would be strongly reduced when the appropriate multivariate GARCH model is fitted to the raw or linearly filtered data. However, failure to accept the no-causality null hypothesis may also constitute evidence that the selected multivariate GARCH model was incorrectly specified. This line of analysis is similar to the use of the univariate BDS test on raw data and on GARCH models (Brock et al., 1996; Brooks, 1996; Hsieh, 1989). Many GARCH models can be used for this purpose. In the present study the GARCH-BEKK model of Engle and Kroner (1995) is used. The BEKK (p, q) model is defined as:

$$\mathbf{H}_t = \mathbf{C}'\mathbf{C} + \sum_{j=1}^q \mathbf{A}'_{jk} \boldsymbol{\varepsilon}_{t-j} \boldsymbol{\varepsilon}'_{t-j} \mathbf{A}_{jk} + \sum_{j=1}^p \mathbf{G}'_{jk} \mathbf{H}_{t-j} \mathbf{G}_{jk} \quad , \quad \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \mathbf{v}_t \quad (10)$$

where $\mathbf{C}, \mathbf{A}_{jk}$ and \mathbf{G}_{jk} are $(N \times N)$ matrices and \mathbf{C} is upper triangular. \mathbf{H}_t is the conditional covariance matrix of $\{\varepsilon_t\}$ with $\varepsilon_t | \Phi_{t-1} \sim (\mathbf{0}, \mathbf{H}_t)$ and Φ_{t-1} the information set at time $t - 1$. The residuals are obtained by the whitening matrix transformation $\mathbf{H}^{1/2} \varepsilon_t$. Gouriéroux (1997) gives sufficient conditions for \mathbf{A}_t and \mathbf{G}_t in order to guarantee that \mathbf{H}_t is positive definite.

Tables 3 and 4 show results before and after GARCH-BEKK (1,1) filtering. The order parameters were determined for the time series in terms of the minimal SIC. The linear Granger causality interdependencies remain absent, exactly as after VECM filtering in both periods and for both representations i.e., bivariate and five-variate. After the nonlinear causality testing in some cases the statistical significance is weaker after filtering, particularly in the five-variate GARCH-BEKK implementation. These differences in statistical significance indicate that the nonlinear causality is partially due to simple volatility effects. However, this is not indicative of a general conclusion. Instead, significant nonlinear interdependencies remain after the bi-variate and five-variate GARCH-BEKK filtering, revealing that volatility effects and spillovers are probably not the only ones inducing nonlinear causality. This of course does not apply to all the pairs of spot and futures returns but some main results can be drawn for specific relationships. These are also depicted graphically in Figure 2 where strong causality (“**”) is denoted by a “double arrow”.

In particular, the pairwise nonlinear causality reveals the bi-directional linkages WTI Spot ↔ WTI F1, WTI Spot ↔ WTI F3 and WTI Spot ↔ WTI F4 in PI, and WTI Spot → WTI F1 in PII. In fact, these relationships remain roughly unchanged from the previous VECM filtering stage. Yet, there are two significant changes; the bi-directional

causality $WTI\ Spot \leftrightarrow WTI\ F2$ in PI is reduced to a weakened $WTI\ Spot \rightarrow WTI\ F2$ linkage, and most importantly in PII the uni-directional causality from futures to spot for all maturities, has now vanished. Thus, there is strong evidence that the latter nonlinear causal relationship can be attributed to second moment effects.

Now, incorporating the “contribution” of all variables in a five-variate GARCH-BEKK framework, the whitened residuals present different causal relationships than before. Specifically, in PI the bi-directional linkages $WTI\ Spot \leftrightarrow WTI\ F1$, $WTI\ Spot \leftrightarrow WTI\ F3$ and $WTI\ Spot \leftrightarrow WTI\ F4$ are reduced to uni-directional and the $WTI\ Spot \rightarrow WTI\ F2$ has disappeared. It seems that the nonlinear causality from futures to spot returns which persisted even after the five-variate VECM filtering was induced by conditional heteroskedasticity and thus a five-variate and not a bi-variate GARCH-BEKK filtering of the VECM-residuals is better at “capturing” the volatility transmission mechanism. Instead, in PII the uni-directional linkages $WTI\ F1 \rightarrow WTI\ Spot$ and $WTI\ F4 \rightarrow WTI\ Spot$ were not entirely removed as in the bi-variate GARCH-BEKK filtering of the VECM-residuals. Eventually, in all results, third or higher-order causality may be a significant factor of the remaining interdependence.

5. Conclusions

In the present paper we investigated the existence of linear and nonlinear causal relationships between the daily spot and futures prices for maturities of one, two, three and four months of West Texas Intermediate (WTI), which is the underlying commodity of New York Mercantile Exchange's (NYMEX) oil futures contracts. The data covered two separate periods, namely PI: 10/21/1991-10/29/1999 and PII: 11/1/1999-10/30/2007,

with the latter being significantly more turbulent. The study contributed to the literature on the lead–lag relationships between the spot and futures markets in several ways. In particular, it was shown that the pairwise VECM modeling suggested a strong bi-directional Granger causality between spot and futures prices in both periods, whereas the five-variate implementation resulted in a uni-directional causal linkage from spot to futures prices only in PII. This empirical evidence appears to be in contrast to the results of Silvapulle and Moosa (1999) on the futures to spot prices uni-directional relationship. Additionally, whilst the linear causal relationships have disappeared after the cointegration filtering, nonlinear causal linkages in some cases were revealed and more importantly persisted even after multivariate GARCH filtering during both periods. Interestingly, it was shown that the five-variate implementation of the GARCH-BEKK filtering, as opposed to the bi-variate, captured the volatility transmission mechanism more effectively and removed the nonlinear causality due to second moment spillover effects. Moreover, the results imply that if nonlinear effects are accounted for, neither market leads or lags the other consistently, or in other words the pattern of leads and lags changes over time. Given that causality can vary from one direction to the other at any point in time, a finding of bi-directional causality over the sample period may be taken to imply a changing pattern of leads and lags over time, providing support to the Kawaller *et al.* (1988) hypothesis. Hence it can be safely concluded that, although in theory the futures market play a bigger role in the price discovery process, the spot market also plays an important role in this respect. These conclusions, apart from offering a much better understanding of the dynamic linear and nonlinear relationships underlying the crude oil spot and futures markets, may have important implications for market efficiency.

For instance, they may be useful in future research to quantify the process of market integration or may influence the greater predictability of these markets.

An interesting subject for future research is the nature and source of the nonlinear causal linkages. As presented, volatility effects may partly account for nonlinear causality. The GARCH-BEKK model partially captured the nonlinearity in daily spot and future returns, but only in some cases. An explanation could be that spot and futures returns may exhibit statistically significant higher-order moments. A similar result was reported by Scheinkman and LeBaron, (1989) for stock returns. Alternatively, parameterized asymmetric multivariate GARCH models could be employed in order to accommodate the asymmetric impact of unconditional shocks on the conditional variances.

Acknowledgements

The authors acknowledge financial support from the Netherlands Organization for Scientific Research (NWO) under the project: *Information Flows in Financial Markets*. The usual disclaimer applies.

References

- Abhyankar, A., 1996. Does the stock index futures market tend to lead the cash? New evidence from the FT-SE 100 stock index futures market. Working paper no 96-01, Department of Accounting and Finance, University of Stirling.
- Baek, E., and Brock, W., 1992. A general test for non-linear Granger causality: bivariate model. Working paper, Iowa State University and University of Wisconsin, Madison, WI.
- Bopp, A. E., and Sitzer, S., 1987. Are petroleum futures prices good predictors of cash value? *The Journal of Futures Markets* 7, 705–719.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1996. A test for independence based on the correlation dimension. *Econometric Reviews* 15(3), 197-235.
- Brooks, C., 1996. Testing for nonlinearities in daily pound exchange rates. *Applied Financial Economics* 6, 307-317.
- Chen, A.-S., and Lin, J.W., 2004. Cointegration and detectable linear and nonlinear causality: analysis using the London Metal Exchange lead contract. *Applied Economics* 36, 1157-1167.
- Choudhury, A R., 1991. Futures market efficiency: Evidence from cointegration tests. *The Journal of Futures Markets* 11, 577–589.
- Crowder, W. J. and Hamed, A., 1993. A cointegration test for oil futures market efficiency. *Journal of Futures Markets* 13, 933–41.
- Diks, C. and Panchenko, V., 2005. A note on the Hiemstra-Jones test for Granger noncausality. *Studies in Nonlinear Dynamics and Econometrics* 9, art. 4

- _____, 2006. A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics & Control* 30, 1647-1669.
- Engle, R.F. and Granger, C.W.J., 1987. Cointegration, and error correction: representation, estimation and testing. *Econometrica*, 55, 251–76.
- Engle, R.F. and Kroner, F.K. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11, 122-150.
- Garbade, K.D., and Silber, W. L., 1983. Price movement and price discovery in futures and cash markets. *Review of Economics and Statistics* 65, 289–297.
- Granger, C.W.J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424-438.
- Gourieroux, C. 1997. ARCH models and Financial applications. Springer Verlag
- Hiemstra, C. and Jones, J.D., 1994. Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance* 49, 1639-1664.
- Houthakker, H.S., 1992. Futures trading. In: Newman, P., Milgate, M., and Eatwell, J. (Eds.) *The new Palgrave dictionary of money and finance* 2, 211–213, London: Macmillan.
- Hsieh, D., 1989. Modeling heteroscedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistics* 7, 307–317.
- Hull J., 2000. *Options, Futures and Other Derivatives*. Prentice Hall, New York
- Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic and Dynamics and Control*, 12, 231–54.

- Johansen, S. and Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration with application to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52, 169–209.
- Kawaller, I.G., Koch, P.D., and Koch, T.W., 1988. The relationship between the S&P 500 index and the S&P 500 index futures prices. *Federal Reserve Bank of Atlanta Economic Review* 73 (3), 2–10.
- Krehbiel, T., and Adkins, L.C., 1993. Cointegration tests of the unbiased expectations hypothesis in metals markets. *The Journal of Futures Markets* 13, 753–763.
- Moosa, I.A., 1996. An econometric model of price determination in the crude oil futures markets. In: McAleer, M., Miller, P., and Leong, K. (Eds.) *Proceedings of the Econometric Society Australasian meeting 3*, 373–402, Perth: University of Western Australia.
- Moosa, I.A., and Al-Loughani, N.E., 1995. The effectiveness of arbitrage and speculation in the crude oil futures market. *The Journal of Futures Markets* 15, 167–186.
- Newberry, D.M., 1992. Futures markets: Hedging and speculation. In: Newman, P., Milgate, M., and Eatwell, J. (Eds.) *The new Palgrave dictionary of money and finance 2*, 207–210, London: Macmillan.
- Oellermann, C. M., Brorsen, B. W., and Farris, P. L., 1989. Price discovery for feeder cattle. *The Journal of Futures Markets* 9, 113–121.
- Sadorsky, P., 2000. The empirical relationship between energy futures prices and exchange rates. *Energy Economics* 22 (2), 253-266
- Scheinkman, J., and LeBaron, 1989. Nonlinear Dynamics and Stock Returns. *The Journal of Business* 62 (3), 311-337.

- Schroeder, T. C., and Goodwin, B. K., 1991. Price discovery and cointegration for live hogs. *The Journal of Futures Markets* 11, 685–696.
- Schwarz, T.V., and Szakmary, A.C., 1994. Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis. *The Journal of Futures Markets* 14, 147–167.
- Serletis, A., and Banack, D., 1990. Market efficiency and cointegration: An application to petroleum market. *Review of Futures Markets* 9,372–385.
- Silvapulle, P., and Moosa, I.A., 1999. The Relationship between Spot and Futures Prices: Evidence from the Crude Oil Market. *The Journal of Futures Markets* 19, 175-193.

Figure 1: WTI price and return time series in PI:10/21/1991–10/29/1999 and PII:11/1/1999 – 10/30/2007

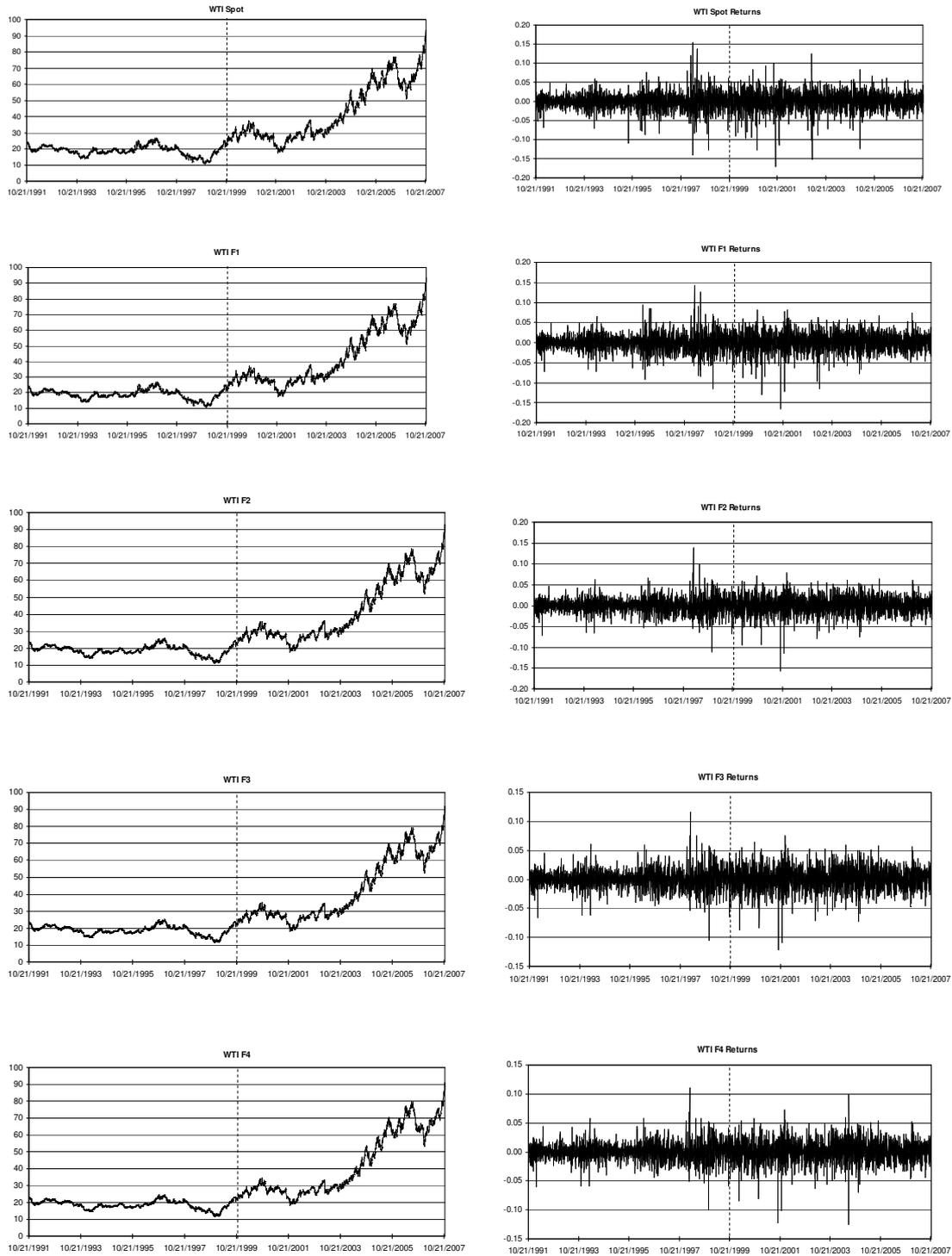
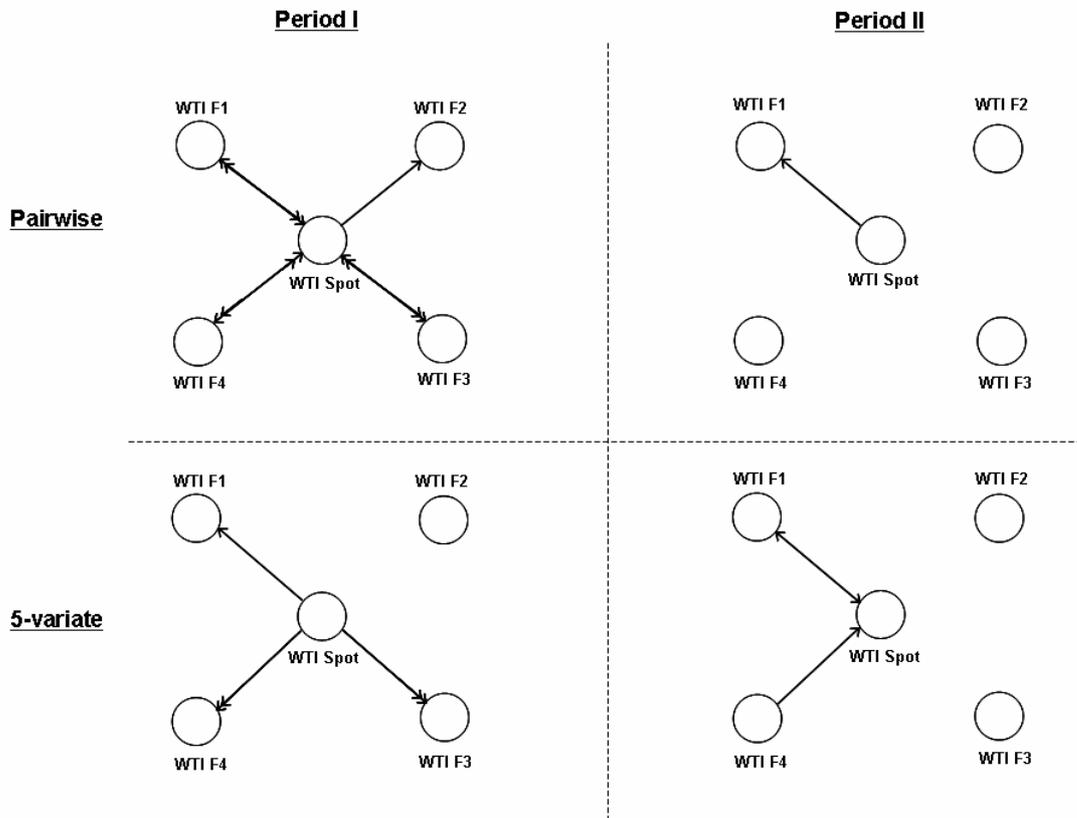


Figure 2: Diagrammatical representation of directional causalities on GARCH-BEKK filtered VECM residuals (Diks-Panchenko test)



Notation: \longrightarrow \longleftarrow denote unidirectional and bi-directional causality corresponding to $5\% \leq p\text{-value} < 1\%$
 \longrightarrow \longleftarrow denote unidirectional and bi-directional causality corresponding to $p\text{-value} \leq 1\%$

Table 1: Descriptive Statistics**Period I (10/21/1991–10/29/1999)**

	WTI Spot	WTI F1	WTI F2	WTI F3	WTI F4
Mean	-0.00004	-0.00005	-0.00005	-0.00004	-0.00004
Standard Deviation	0.02060	0.01975	0.01702	0.01542	0.01432
Sample Variance	0.00042	0.00039	0.00029	0.00024	0.00021
Kurtosis	6.14	4.47	4.72	4.29	4.51
Skewness	-0.01979	0.17027	0.16022	0.09616	0.10314
<i>Correlation Matrix</i>					
	<i>WTI Spot</i>	<i>WTI F1</i>	<i>WTI F2</i>	<i>WTI F3</i>	<i>WTI F4</i>
<i>WTI Spot</i>	1				
<i>WTI F1</i>	0.848	1			
<i>WTI F2</i>	0.835	0.955	1		
<i>WTI F3</i>	0.824	0.936	0.993	1	
<i>WTI F4</i>	0.813	0.917	0.983	0.996	1

Period II (11/1/1999 – 10/30/2007)

	WTI Spot	WTI F1	WTI F2	WTI F3	WTI F4
Mean	0.00069	0.00069	0.00069	0.00068	0.00069
Standard Deviation	0.02388	0.02276	0.02083	0.01945	0.01879
Sample Variance	0.00057	0.00052	0.00043	0.00038	0.00035
Kurtosis	4.04	3.08	2.65	1.86	3.05
Skewness	-0.58017	-0.56054	-0.44623	-0.35836	-0.44760
<i>Correlation Matrix</i>					
	<i>WTI Spot</i>	<i>WTI F1</i>	<i>WTI F2</i>	<i>WTI F3</i>	<i>WTI F4</i>
<i>WTI Spot</i>	1				
<i>WTI F1</i>	0.871	1			
<i>WTI F2</i>	0.859	0.970	1		
<i>WTI F3</i>	0.849	0.957	0.994	1	
<i>WTI F4</i>	0.828	0.932	0.973	0.983	1

Table 2: Unit root tests

Variables	ADF-statistic (PI)	ADF-statistic (PII)
WTI Spot (0)	0.039	0.943
$\Gamma_{\text{WTI Spot}}(0)$	0.000**	0.000**
WTI F1(0)	0.044	0.967
$\Gamma_{\text{WTIF1}}(0)$	0.000**	0.000**
WTI F2 (0)	0.070	0.974
$\Gamma_{\text{WTIF2}}(0)$	0.000**	0.000**
WTI F3 0)	0.085	0.978
$\Gamma_{\text{WTIF3}}(0)$	0.000**	0.000**
WTI F4 0)	0.089	0.979
$\Gamma_{\text{WTIF4}}(0)$	0.000**	0.000**

All variables are in logarithms and reported numbers for the augmented Dickey–Fuller test are p -values. The number of lags in parenthesis is selected using the SIC. (**) denotes p -value corresponding to 99% confidence level.
PI: 10/21/1991–10/29/1999; PII: 11/1/1999 – 10/30/2007

Table 3: Causality results (Pairwise)

Variables		Panel A: Linear Granger Causality						Panel B: Non-Linear Causality									
		Raw data			VECM residuals			GARCH-BEKK residuals			Raw data			VECM residuals			GARCH-BEKK residuals
X	Y	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X
WTI Spot	WTI F1	**	**	*								**	**	*	**	*	**
WTI Spot	WTI F2	**	**	**								**	**	**	*		
WTI Spot	WTI F3	**	**	*	**							**	*	**	**	**	**
WTI Spot	WTI F4	**	**									**	*	**	*	**	**

(*)/(**) Denotes p -value statistical significance at 5% and 1% level. $X \rightarrow Y$: r_x does not Granger Cause r_y . PI: 10/21/1991–10/29/1999; PII: 11/1/1999 – 10/30/2007.

Panel A: Linear Granger Causality

All data (log-levels) were found to be cointegrated and the lag lengths of VECM specification are set using the Wald exclusion criterion. The number of lags (in parenthesis) identified in period PI are: WTI Spot - WTI F1 (3), WTI Spot - WTI F2 (7), WTI Spot - WTI F3 (3), WTI Spot - WTI F4 (3) and in period PII: WTI Spot - WTI F1 (3), WTI Spot - WTI F2 (6), WTI Spot - WTI F3 (6), WTI Spot - WTI F4 (4). In period PI for all pairs the Johansen test identified two (2) cointegrating vectors using the trace statistic and in PII one (1) cointegrating vector again for all pairs. The causality on the VECM residuals was investigated with a VAR specification (the null of no cointegration was not rejected) and the lags were determined using the Schwartz Information Criterion (SIC). The number of lags identified for all variables in all periods is one (1). The second moment filtering was performed with a GARCH-BEKK (1,1) model.

Panel B: Non-Linear Causality

The number of lags used for the nonlinear causality test are $\ell_x = \ell_y = 1$. The data used are log-returns. Since the log-levels were found to be cointegrated (Panel A) the nonlinear causality was investigated on the VECM residuals. The number of lags (Wald exclusion criterion) and the number of cointegrating vectors (Johansen trace statistic test) identified for the VECM specification are reported in Panel A. The second moment filtering was performed with a GARCH-BEKK (1,1) model.

Table 4: Causality results (five-variate)

Variables		Panel A: Linear Granger Causality						Panel B: Non-Linear Causality									
		Raw data			VECM residuals			GARCH-BEKK residuals			Raw data			VECM residuals			GARCH-BEKK residuals
X	Y	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X	X→Y	Y→X
		PI	PII	PI	PII	PI	PII	PI	PII	PI	PII	PI	PII	PI	PII	PI	PII
WTI Spot	WTI F1	*															*
WTI Spot	WTI F2																
WTI Spot	WTI F3		**														
WTI Spot	WTI F4		**														*

(*), (**) Denotes *p*-value statistical significance at 5% and 1% level. X→Y: r_X does not Granger Cause r_Y . PI: 10/21/1991–10/29/1999; PII: 11/1/1999 – 10/30/2007.

Panel A: Linear Granger Causality

The 5x5 system of the data (log-levels) was found to be cointegrated and the lag length of the VECM specification was set using the Wald exclusion criterion. The number of lags identified in PI is eleven (11) and in PII is nine (9). In period PI the Johansen test identified five (5) cointegrating vectors using the trace statistic and in PII three (3). The causality on the VECM residuals was investigated with a VAR specification (the null of no cointegration was not rejected) and the lags were determined using the Schwartz Information Criterion (SIC). The number of lags identified for all variables in all periods is one (1). The second moment filtering was performed with a GARCH-BEKK (1,1) model.

Panel B: Non-Linear Causality

The number of lags used for the nonlinear causality test are $\ell_X = \ell_Y = 1$. The data used are log-returns. Since the 5-variate system of log-levels was found to be cointegrated (Panel A) the nonlinear causality was investigated on the VECM residuals. The number of lags (Wald exclusion criterion) and the number of cointegrating vectors (Johansen trace statistic test) identified for the VECM specification are reported in Panel A. The second moment filtering was performed with a GARCH-BEKK (1,1) model.