The Nexus between Natural Gas Spot and Futures Prices at NYMEX: Do Weather Shocks and Non-Linear Causality in Low Frequencies Matter?

Theologos Dergiades [a, c *], Reinhard Madlener [b] and Georgia Christofidou [c]

[a] Department of International and European Studies, University of Macedonia, 156 Egnatia street, 546 36, Thessaloniki, Greece.

[b] Institute for Future Energy Consumer Needs and Behavior (FCN), School of Business and Economics/E.ON Energy Research Center, RWTH Aachen University, Mathieustrasse 10, 52074 Aachen, Germany.

[c] School of Science and Technology, International Hellenic University, 14th km Thessaloniki/Moudania, 57001 Thermi, Greece.

Abstract

The existence of non-linear dynamics in the prices of commodities is an endemic feature and one of the most fundamental stylized facts in the finance literature. This study, conditioning on weather shocks, investigates the nature of the existing predictive power between natural gas spot and futures price at the NYMEX market. By implementing a causality test in the frequency domain, we find that the short maturity futures market offers a significant predictive power towards the spot market. The identified predictive power over the frequency band proves to be asymmetric with respect to the first- and the second-conditional moments of the series. In particular, our results show that for high frequencies (short-run) predictability is linked to the first-conditional moment while for low frequencies (long-run) predictability is attributed to the second-conditional moment.

Keywords : Natural gas; Spot and futures markets; Frequency domain causality

JEL Classification : C14, C58, G14, G15

[*] Corresponding author, e-mail: dergiades@uom.edu.gr

1. Introduction

The lead-lag relationship between spot and futures markets is a conundrum that continues to attract noteworthy attention from the academic scene. Futures markets can be seen as precious vehicles for hedging, while at the same time it can be argued that they constitute valuable sources for disclosing strong signals for upcoming spot market trends (see, e.g., Daskalakis *et al.*, 2009). Futures contracts are used to offset commodity price risk as well as to limit the exposure to price fluctuations. In particular, companies and investors trade futures contracts to hedge against the undertaken risks, thus protecting their portfolios in cases of abrupt price changes (see, e.g., Lien and Yang, 2008; Chng, 2009). Moreover, there is credible empirical evidence, not without questioning, that futures markets may act as leading indicators for the forthcoming movements in the spot market (see, e.g., Ng and Pirrong, 1996; Darrat *et al.*, 2002; Zhang and Liu, 2018). Within such a framework, futures prices affect market participants' actions, as the former are depicted through the adopted investment strategies of the latter. In particular, as far as the energy market is concerned, where long-term planning is in many cases absolutely essential, knowledge of energy commodities' future price movements is considered imperative. Therefore, there is no doubt that revealing the exact nature of the causality for the natural gas (NG, hereafter) spot and futures prices is of great importance.

Over the last three decades, the forecasting ability of futures markets has attracted much attention from the academic scene (see, e.g., Bierbrauer *et al.*, 2007; Asche *et al.* 2013; Mishra and Smith, 2016). Expectedly, the empirical findings in the relevant literature are quite diverse. More specifically, several studies investigate the connection between spot and futures markets for different commodities. Garbade and Silber (1983) were the first to investigate the linkages between spot and futures prices for a set of commodities. Evidence for a unidirectional forecasting ability running from the futures market towards the spot market, but not vice versa, is frequently found in the literature (see, e.g., Movassagh and Modjtahedi, 2005; Zhang and Liu, 2018). The main argument which supports the predictive content of futures prices with respect to the spot prices is based on the capacity of the former market to incorporate at a faster pace new relevant information (Silvapulle and Moosa, 1999). However, according to a number of other studies, spot prices appear to lead futures prices (for instance, Quan, 1992). Bidirectional causality is also identified in a few cases (see, e.g., Silvapulle and Moosa, 1999; Lee and Zeng, 2011).

To the best of our knowledge, recent work that has been published and investigates the nexus between NG spot and futures markets taking into consideration also the weather conditions is scarce. This sporadic interest in the NG market could be due to the fact that the respective futures market is relatively young compared to futures markets for other energy commodities, such as crude oil. The New York Mercantile Exchange (NYMEX), for instance, only began trading NG futures contracts in 1990. Even the most recent studies, to the extent that we are aware, approach the relationship between spot and futures NG prices within the linear causality paradigm, failing additionally, in most cases, to take into account the indirect effect of other variables such as the weather conditions. Characteristically, Mu (2007) shows that temperature alterations play a crucial role in affecting both the first and the second conditional moment of the NG futures returns, stating at the same time that "Little effort has been devoted to quantifying the weather effect on short-term natural gas price dynamics". Other studies that recognize the essential impact of the weather conditions for the NG market, among others, are those by Elkhafif (1996) and Considine (2000).

Therefore, our analytical framework can be considered as much broader with respect to the rest of the literature, provided that the notion of causality is investigated from a non-linear perspective, considering at the same time the effect of weather conditions. In this sense, our approach constitutes a differentiation with respect to the rest of the literature in the field. More specifically, major merits of our study are: (a) the effort to gain a deeper knowledge of the nature of causality and (b) the implementation, before and after filtering, of a recently advanced causality test in the frequency domain, which remains robust in the presence of volatility clusters,[1] separates between long-run and short-run causality and, finally, enables to identify non-linear relationships. Additionally, for reasons of completeness we also implement the standard Granger causality approach.

Initially, the NG spot and futures returns series are tested for causality (with and without conditioning on weather shocks) by using both the standard Granger causality approach and the Breitung and Candelon (2006) approach (B&C, hereafter). Additionally, given the presence of cointegration, the series are filtered through an appropriate VEC model, and both tests are re-applied (with and without conditioning on weather shocks). Such a treatment enables inference of whether the identified causal relation is attributed to the first conditional moment or to higher-order moments (e.g.

[1] Common attribute for financial time series.

variance, skewness, or kurtosis). In the case where causality persists, after the first-moment filtering, this is evidence that the causality is non-linear and due to higher-order moments. For that reason, we finally implement a second-moment filtering to the series based on a bivariate GARCH-BEKK (1,1) model and both tests are re-conducted.

The remainder of this paper is organized as follows: Section 2 briefly reviews the recent literature which examines the causal dynamics between spot and futures prices for three different energy commodities (electricity, oil, and natural gas). Section 3 illustrates the adopted methodological framework. Section 4 presents the data sources and conducts the necessary preliminary econometric analysis. Section 5 is devoted to a discussion of our empirical findings, while Section 6 concludes and proposes some broad investment strategy recommendations.

2. Literature Review

During recent decades, there is an increasing volume of empirical studies devoted to the nexus between spot and futures markets for several energy commodities. Provided that the focus of the empirical research dealing solely with the NG markets is rather limited, our review will embrace two other major energy commodities: electricity and crude oil. Starting with the former, the existing empirical work that investigates the electricity futures market and its connection with the corresponding spot market can be characterized again as inadequate in terms of volume. This is mainly so because the electricity futures market is relatively new, and competitive derivatives' trading is less widespread than for other energy commodities such as crude oil. Electricity futures were introduced at the NYMEX in 1996. Electricity, as a flow commodity, can be distinguished from other energy commodities due to the fact that it cannot be stored in large quantities. As a result, arbitrage opportunities related to trading electricity over time are limited or even eliminated.

Without testing for causality, Shawky *et al.* (2003) is one of the few studies that concentrates on the electricity market at the NYMEX and investigates elements of predictability between spot and futures prices. In more detail, Shawky *et al.* (2003) derive estimates for the forward risk premium in electricity futures contracts and for the optimal hedge ratio between spot and futures markets. Their findings suggest for the examined commodity a relatively high forward risk premium (4% per month) with respect to other commodities, a fact that is attributed to the idiosyncratic characteristics of electricity. Additionally, through a simple GARCH specification, Shawky *et al.* (2003) estimated a

relatively high optimal hedge ratio (1.63) with respect to other commodities and at the same time they also observe that the resulting volatility in the former market (spot) is more pronounced than that of the latter market (futures). Lastly, the conducted impulse response analysis reveals that both markets' returns move towards the same direction after a one standard deviation shock in the spot market, while in case where a similar shock takes place in the futures market the direction of the effect remains the same for both markets but with a lower magnitude.

In contrast to the case of electricity, the respective literature on crude oil is quite abundant. Many studies investigate the existing linkages for the crude oil spot and futures markets by implementing a diverse set of econometric approaches for different time periods. The majority of these studies have focused on the West Texas Intermediate (WTI) crude oil prices. Specifically, Bopp and Sitzer (1987), implementing simple regression analysis, found that futures contracts with one-month maturity convey significant information about the spot market, while this information vanishes for contracts with longer maturity. In a similar fashion, Bopp and Lady (1991), again by means of a regression analysis, find that both markets embody explanatory power with respect to the actual prices. Quan (1992) is the first who examines the lead-lag interdependence for the spot and futures prices series by conducting cointegration and standard causality analysis. He finds that only spot prices cause futures prices, while the opposite hypothesis is not a validated hypothesis empirically.

The linear causality paradigm was the major workhorse in the early causality literature. Non-linear methods, however, like the one proposed by Hiemstra and Jones (1994), or more recently by Breitung and Candelon (2006), appear to produce more powerful results, since these are capable of detecting hidden non-linear relations that otherwise cannot be captured. For instance, Silvapulle and Moosa (1999), under the linear causality paradigm and using the Hsiao (1981) version of causality, show the existence of a unidirectional causality that runs from the futures market towards the spot market, and not vice versa. When adopting a non-linear causality framework instead, an additional hidden channel of causality was revealed, running in the opposite direction. In other words, the non-linear testing results indicate a bidirectional causal relationship, a fact which was not identifiable within the linear causality paradigm.

Lee and Zeng (2011) is relatively a recent study coping with the nexus of the crude oil spot and futures prices. Compared to previous studies, this one stands out for applying the quantile cointegrating approach. Lee and Zeng (2011) investigate the

existence of possible predictability by implementing linear, non-linear, and quantile methods. For the period 1986 to 2009, both linear and non-linear tests provide strong evidence toward a bidirectional causal relationship. The quantile causal analysis, though, produced quite different results. While spot prices appear to Granger-cause futures prices (irrespective of their maturity) on every quantile, this is not the case when the examined hypothesis runs in the opposite direction. In particular, for futures contracts with maturity of two, three and four months, respectively, there was no causal relationship with the spot prices for all the examined quantiles except the first one. Only for futures contracts with one month-maturity the null hypothesis of no Granger causality is rejected on every quantile.

The NG futures market at the NYMEX commenced in 1990 (April 3) and has been growing ever since. Especially during periods of high volatility, market participants trade futures to hedge their position and lock in a profit. Therefore, efficient price-forecasting mechanisms would be a valuable tool for all market participants. The examination of the forecast ability that the spot market can offer towards the futures market (and vice versa) has been investigated to some extent in the literature but not thoroughly enough. Having a broader look at the literature for the NG market, the majority of studies conclude that the futures market offers some predictive power over the spot market, and that generally the futures prices can be considered as an efficient predictor. In one of the most recent studies, Gebre-Mariam (2011), focusing on the Northwest US NG markets, for the period extending from 1999 to 2004, investigates the causal relationship between NG spot and futures prices. Gebre-Mariam (2011) finds that NG spot prices Granger-cause futures prices with maturity of less than a year, whilst the opposite is true for futures contracts with more than one year to maturity.

In general, the only safe conclusion that can be drawn is that the literature provides mixed results with respect to the direction of causality between spot and futures prices. These differences in the causality inferences can be attributed to several factors, including the methodological framework adopted, the features of the examined energy commodity, and the extent of the investigated period. Regarding the NG spot and futures markets, it has been revealed that a fertile ground exists for further research, provided that the studies already conducted have not taken into account the non-linear characteristics of the market structure. Furthermore, no previous attempt has been made to uncover the true nature of the causal relationship conditioning also on weather shocks.

3. Methodology

To investigate the lead-lag relationship between the NG spot and futures prices we employ (apart the standard Granger causality test) a frequency domain causality test proposed by Breitung and Candelon (2006). The B&C methodological framework is built up on a similar context to that of Geweke (1982) and Hosoya (1991), introduced roughly three decades ago. Based on bivariate or even on higher-order dimensional Vector Autoregressive (VAR) systems, B&C propose a test that can control for equilibrium relationships and at the same time to distinguish between the short-run and the long-run predictive power.

The innovativeness of the frequency domain non-causality test lies on two basic attributes that are usually neglected from the standard causality tests routinely implemented in the literature (e.g. Toda and Yamamoto, 1995; Hsiao, 1981). Firstly, the frequency domain approach can identify non-linear causal relationships and, secondly, the testing of causality is carried out in a dynamic manner since it is implemented over several alternative frequencies, providing this way a more structured understanding about the true nature of the investigated causal relationship (Granger, 1969).

To illustrate briefly the frequency domain approach, let S_t and F_t to be two covariance-stationary time series of length *N*. Under the typical assumptions, the two-dimensional vector $z_t = (S_t, F_t)'$ has a standard form VAR representation of finite order *p*.[2] The Wold representation theorem implies that vector z_t can be projected on current and past values of the error term:

$$z_t = \Theta(L)^{-1}\varepsilon_t = \Phi(L)\varepsilon_t \quad \text{with } \Theta(L)^{-1} = \Phi(L), \tag{1}$$

where $\Theta(L)$ is a 2×2 matrix of polynomials in the lag operator with order *p*, and ε_t is an error term assuming the usual properties, that is $E(\varepsilon_t) = 0$, $E(\varepsilon_t\varepsilon_t') = \Sigma$ and Σ positive definite. A structural representation of the system can be obtained through the implementation of the Cholesky identification process. Therefore, by introducing the lower triangular C matrix, such that $u_t = C\varepsilon_t$ and $E(u_t, u_t') = I$, the system can be re-written in terms of the structural innovations u_t.

---

[2] The representation has the functional form: $\Theta(L)z_t = \varepsilon_t$

$$z_t = \Theta(L)C^{-1}u_t = \Psi(L)u_t \quad \text{with } \Theta(L)^{-1}C^{-1} = \Psi(L) \tag{2}$$

Based on the above structural representation, the spectral density of S_t at frequency ω can be expressed by *Eq.* (3) as follows:

$$f_S(\omega) = (1/2\pi)\left\{\left|\Psi_{11}(e^{-i\omega})\right|^2 + \left|\Psi_{12}(e^{-i\omega})\right|^2\right\} \tag{3}$$

The non-causality hypothesis within the framework of Geweke (1982) is tested from the following Fourier transformation of the moving average coefficients:

$$M_{F\to S}(\omega) = \log\left[\frac{2\pi f_S(\omega)}{\left|\Psi_{11}(e^{-i\omega})\right|^2}\right] = \log\left[1 + \frac{\left|\Psi_{12}(e^{-i\omega})\right|^2}{\left|\Psi_{11}(e^{-i\omega})\right|^2}\right]. \tag{4}$$

In case where F_t does not cause S_t at frequency ω, $M_{F\to S}(\omega)$ has to be equal to zero and consequently $\left|\Psi_{12}(e^{-i\omega})\right|^2 = 0$. This framework can easily be extended to higher-dimensional systems (e.g. trivariate or quadrivariate) by conditioning upon the newly incorporated variables. This process is analytically described in Geweke (1984), with the interpretation of the conditional causality to remain analogous.

Provided that $\left|\Psi_{12}(e^{-i\omega})\right|^2$ is used to test the no-causality hypothesis within the Geweke (1982; 1984) framework and that it is a complicated non-linear function of the estimated VAR parameters (Breitung and Candelon, 2006), B&C offer a remedy for this complexity through a set of linear restrictions imposed on the estimated VAR coefficients. In particular, B&C, by concentrating on the $\Psi_{12}(L)$ element of the $\Psi(L)$ matrix, restate the null hypothesis. The $\Psi_{12}(L)$ element is given in *Eq.* (5) below:

$$\Psi_{12}(L) = -\frac{1}{c_{22}}\frac{\Theta_{12}(L)}{|\Theta(L)|} \tag{5}$$

where $1/c_{22}$ is the positive[3] lower diagonal element of the C^{-1} matrix and $|\Theta(L)|$ is the determinant of $\Theta(L)$. As a result, the null hypothesis of no causality at frequency ω from F_t towards S_t is not rejected whenever *Eq.* (6) holds:

$$\left|\Theta_{12}(e^{-i\omega})\right| = \left|\sum_{k=1}^{p}\theta_{12,k}\cos(k\omega) - \sum_{k=1}^{p}\theta_{12,k}\sin(k\omega)i\right| = 0, \tag{6}$$

where $\theta_{12,k}$ is the upper right element of the Θ_k matrix. Subsequently, the set of restrictions that should be imposed are:[4]

$$\sum_{k=1}^{p}\theta_{12,k}\cos(k\omega) = 0 \quad \text{and} \quad \sum_{k=1}^{p}\theta_{12,k}\sin(k\omega) = 0. \tag{7}$$

The empirical procedure of the B&C approach lies on the validity of the above linear restrictions. For brevity reasons, if we denote $\alpha_j = \theta_{11,j}$ and $\beta_j = \theta_{12,j}$, then the VAR equation that corresponds to the S_t variable may be rewritten as:

$$S_t = \alpha_1 S_{t-1} + \ldots + \alpha_p S_{t-p} + \beta_1 F_{t-1} + \ldots + \beta_p F_{t-p} + \varepsilon_{1t}. \tag{8}$$

Thus, the hypothesis of no causality, $M_{F\to S}(\omega) = 0$, is equivalent to the following set of linear restrictions:

$$R(\omega)\beta = 0, \quad \text{where } \beta = \left(\beta_1, \ldots, \beta_p\right)' \text{ and } R(\omega) = \begin{pmatrix} \cos(\omega) & \ldots & \cos(p\omega) \\ \sin(\omega) & \ldots & \sin(p\omega) \end{pmatrix}. \tag{9}$$

B&C assess the validity of the linear restrictions illustrated in *Eq.* (9) for frequencies ω that receive values within the interval of $(0, \pi)$, and compare the obtained statistic with the 0.05 critical value of the χ^2 distribution with two degrees of freedom.

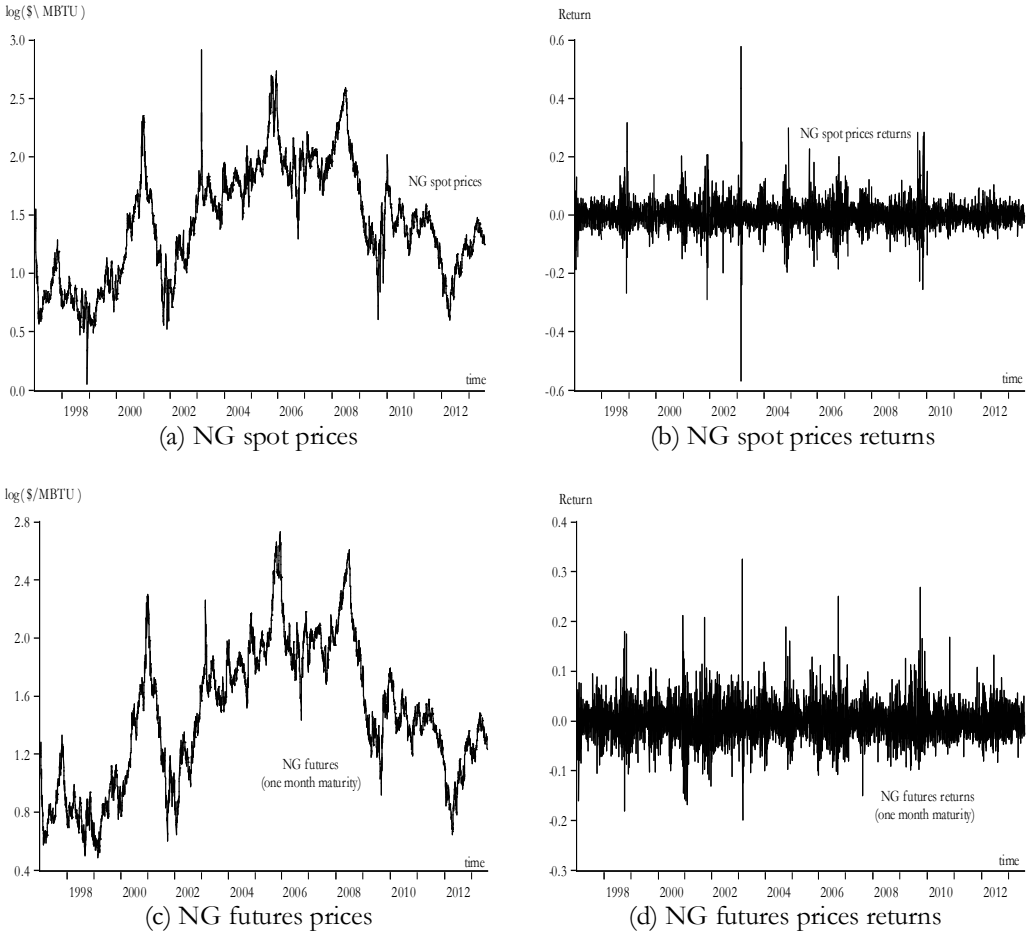
---

[3] This is due to the assumption that the variance-covariance matrix Σ is positive definite.

[4] Given that $\sin(k\omega) = 0$ in the cases where $\omega = 0$ and $\omega = \pi$, the second restriction in *Eq.* (7) can simply be disregarded.

4. Data Sources and Preliminary Analysis

The NG spot prices as well as the NG NYMEX futures prices with one-month maturity are used in this study (both in US Dollars per million BTU, or in short $/MBTU),[5] and were obtained from the US Energy Information Administration (EIA) database.[6] The delivery period for the futures contract used is the calendar month following the trade date. For reasons of brevity, the variables are referred to as *S* (spot price), *F* (futures price for the one-month maturity contract). Finally, the period examined spans from January 7, 1997, to July 30, 2013, and includes 4142 observations in total. The NG spot and futures prices along with the respective return series are depicted in Figure 1.



(a) NG spot prices (b) NG spot prices returns

(c) NG futures prices (d) NG futures prices returns

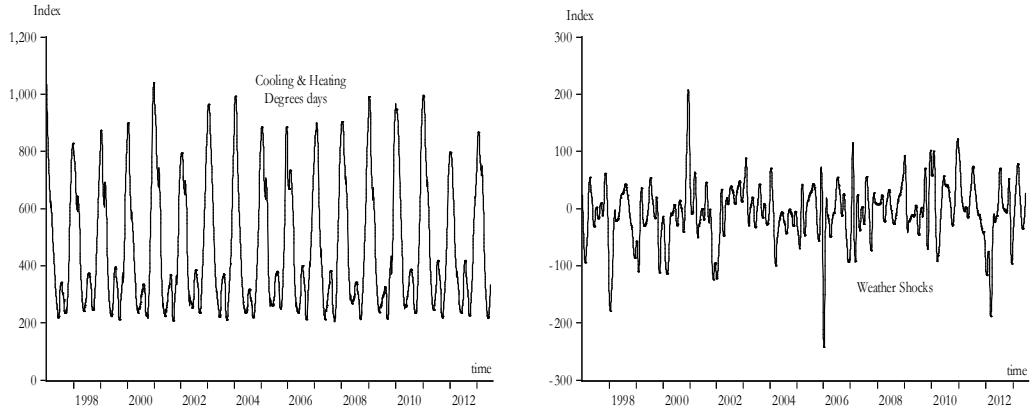
Fig. 1. NG spot and futures prices

[5] The delivery point assumed for the spot trades and futures contracts is the Henry Hub in Louisiana.
[6] Available at: www.eia.gov; the sourcekey codes for the spot and futures series are RNGWHHD and RNGC1, respectively.

The raw data on the cooling and heating degrees days (DD) were obtained from the National Oceanic and Atmospheric Administration (NOAA) of the US Department of Commerce.[7, 8] Using the cooling and heating degrees index data above, we construct the weather shocks variable (*WS*) by adopting Mu's (2007) approach. More specifically, Mu (2007) constructs the weather shocks series as:

$$WS_t = (1/k) * \left(\sum_{i=1}^{k} (DD_{t+i} - NDD_{t+i}) \right), \tag{10}$$

where DD_{t+i} is the realized cooling and heating degree days on day $t+i$,[9] NDD_{t+i} is the normally expected *DD* value on day $t+i$, defined as the thirty-year average of the respective day and, finally, k stands for the weather forecast horizon which typically features values between one and seven (see Mu, 2007). The raw data on the cooling and heating degree days, together with the resulting weather shocks series, are illustrated in Figure 2.



(a) Cooling & heating degree days index (b) Weather shocks

Fig. 2. Degrees days index and weather shocks

To detect the order of integration of the variables, the Augmented Dickey-Fuller unit root test (ADF) (Dickey and Fuller, 1979) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Kwiatkowski *et al.*, 1992) stationarity test are implemented both with and without a time trend. The tests are initially applied to the log levels of the series as well as

[7] Please see: http://www7.ncdc.noaa.gov/CDO/CDODivisionalSelect.jsp#
[8] DD is available at monthly frequency; we convert the series to daily frequency using the quadratic-match average method.
[9] As is the case in Mu (2007), the actual degree days are used instead of the forecasted degree days.

to their first differences. When the ADF test is applied to the levels for the two price series we fail to reject the null hypothesis, while this is not the case for the weather shock series. On the other hand, when the test is re-applied to the first differences, the null is rejected consistently. Clearly, according to the ADF test, the two NG price series (spot and futures) are integrated of order one, $I(1)$, while the weather shock variable is integrated of order zero, $I(0)$. The KPSS test results reach the same conclusion.

Table 1. ADF and KPSS test statistics

Variable	Levels		First differences	
	No trend	Trend	No trend	Trend
	t-stat.(*k*)	*t*-stat.(*k*)	*t*-stat.(*k*)	*t*-stat.(*k*)
Panel *A* – ADF unit root test				
S	-2.561 (2)	-2.655 (2)	-53.619 (1)***	-53.613 (1)***
F	-2.243 (1)	-2.269 (1)	-69.309 (0)***	-69.303 (0)***
WS	-9.346 (29)***	-9.438 (29)***	-11.722 (30)***	-11.723 (30)***
Panel *B* – KPSS stationarity test				
S	2.261***	1.340***	0.036	0.028
F	2.344***	1.387***	0.061	0.036
WS	0.243	0.044	0.008	0.006

Notes: *k* represents the selected lag length (based on the Schwarz criterion with k_{min}=0 and k_{max}=30). *, ** and *** denote rejection of the null hypothesis at the 10%, 5%, and 1% significance levels, respectively.

Additionally, we examine the presence of cointegration. In the case where the NG spot and futures prices conditional on the weather shocks are cointegrated, the delinearization[10] should be conducted within a VECM framework, rather than in a simple VAR framework. Provided that our trivariate system is a mix of $I(1)$ and $I(0)$ variables (the weather shocks are treated as exogenous $I(0)$ variable) the standard Johansen (1995) cointegration approach is inappropriate to test for cointegration given that the Trace statistic depends on nuisance parameters (Rahbek and Mosconi, 1999). As a result, we implement the Rahbek and Mosconi (1999) remedy by incorporating within the cointegrating vector the accumulated sum of the weather shocks together with a time trend.[11] The cointegration testing results, which are presented in Table 2 below, reveal the existence of an equilibrium relationship. The 0.05 critical values from Pesaran *et al.* (2000) (based on a sample of 500 observations and 10,000 replications) suggest two cointegrating vectors, while the 0.01 simulated critical values (based on a tailor-made sample of 4,000 observations and 10,000 replications) imply one cointegrating vector. Overall, it becomes apparent that the delinearization process necessitates the use of a VECM framework.

[10] The delinearization takes place to ensure that the identified causality is solely non-linear in nature.
[11] The critical values of the new limiting distribution of the Trace statistic are provided by Harbo *et al.* (1998).

Table 2. Cointegration test with $I(0)$ exogenous variable (Rahbek & Mosconi fix)

Null	Alternative	Trace statistic	Pesaran *et al.* (2000) 0.05 c.v.	Simulated 0.01 c.v.
$r = 0$	$r = 1$	216.61***	30.77	36.03
$r \leq 1$	$r = 2$	18.58**	15.44	19.59

Notes: **, *** denotes rejection of the null hypothesis at the 0.05 and 00.1 significance levels, respectively, while r is the number of cointegrating vectors. The abbreviation c.v. stands for critical values. The first set of c.v. obtained from Pesaran *et al.* (2000), p.340, Table 6(d), case IV, while the second set of c.v. comes from a simulation, done by the authors, based on a sample of 4,000 observations and 10,000 replications.

Finally, provided that our main interest lies in the identification of the non-linear causality between NG spot and futures prices, it is crucial to verify the existence of a non-linear structure. Consequently, we will test the assumption of independence (i.i.d. assumption) through the BDS test, as suggested by Brock *et al.* (1996). The results are analytically displayed in Table 3. Clearly, the existence of a non-linear structure is verified in both series, irrespective of whether the delinearization takes place within a simple bivariate VAR model or within a VECM model (with one exogenous $I(0)$ variable). In particular, the i.i.d. assumption for the spot and futures prices is rejected at the 0.01 significance level in every case. Thus, we infer that it is essential to re-conduct the causality testing after delinearizing the variables of interest within a VECM specification.

Table 3. BDS test for the VAR and VECM residuals

Variable	VAR residuals		VECM residuals	
	BDS statistic	*p*-value	BDS statistic	*p*-value
S	0.034	0.000	0.035	0.000
F	0.008	0.000	0.009	0.000

Notes: The delinearization takes place within a standard bivariate VAR and within a trivariate VECM with one exogenous variable. The selected lag-order of the VAR and VECM models is 3 for both cases. For presentation brevity, all the illustrated BDS statistics correspond to an embedding dimension equal to 2 (the test inference is identical when the embedding dimension is set equal to 3 and up to 10).

5. Empirical Results

5.1 Causality for the unfiltered series

Enhanced knowledge of the causal nature for a set of series is of major importance for all market participants, since it will allow them to develop more efficient investment strategies. For instance, if we assume that there is short-run causality in the first conditional mean that runs from the NG futures prices to the NG spot prices, then the high frequency components of the futures prices can be utilized by investors to realize excess returns in the spot market. To gain a better knowledge about the true nature of causality between the NG spot and futures prices conditioning on the weather shocks, a

three-step testing procedure is adopted. The first step of our analysis entails implementation of the standard Granger causality test and the B&C frequency domain causality test directly on the returns of the spot and futures prices. Rejection of the null hypothesis in that initial step, however, leaves half the story untold with respect to the nature of causality, since we may infer whether the causal relationship is short- or long-run (based on the B&C results), while uncertainty remains on whether the causality is attributed to the first- or higher-order moments. The second step involves again the implementation of both causality tests after a first-moment filtering that is conducted through a VECM specification (weather shocks are incorporated as an exogenous $I(0)$ variable). The obtained VECM-filtered residuals, which are now free from any linear predictive power, are used to re-conduct the B&C test. In the case where the null hypothesis is still rejected at some frequencies, after the first-moment filtering, this is evidence that the identified causality is due to second- or higher-order moments. For this reason, the third step entails for once more implementation of the two tests after a second-moment filtering based on a bivariate GARCH-BEKK (1,1) specification.

We begin the causality testing by focusing on the return series of the NG spot and futures prices. The frequency domain causality testing results received for examining the hypothesis that spot prices do not cause futures prices ($S \not\rightarrow F$), with and without conditioning on weather shocks, are analytically displayed in Figure 3. Visually inspecting Figure 3, a straightforward inference can be made about the investigated hypothesis. The frequency interval, in which the causality testing is implemented, begins at 0 and terminates at π (3.14). By conducting the causality test on the return series within a bivariate framework (without conditioning), for every frequency we fail to reject, at the conventional levels of significance, the null hypothesis of no causality running from NG spot to futures prices.

A source of criticism for simple bivariate systems comes from the resulting bias in the causality inference due to omitted variables. Specifically, the omitted variables issue may be a serious threat when central channels of causality are disregarded and the resulting causal analysis, consequently, leads to misleading propositions. Consequently, we repeat causality testing for the same hypothesis, but this time we condition on weather shocks. As it can be asserted from the bold line in Figure 3, even though a conditioning on weather shocks takes place, the previously identified causality inference remains unaltered. In short, there is strong statistical evidence supporting the hypothesis that the spot NG prices embody no predictive power towards the futures prices.

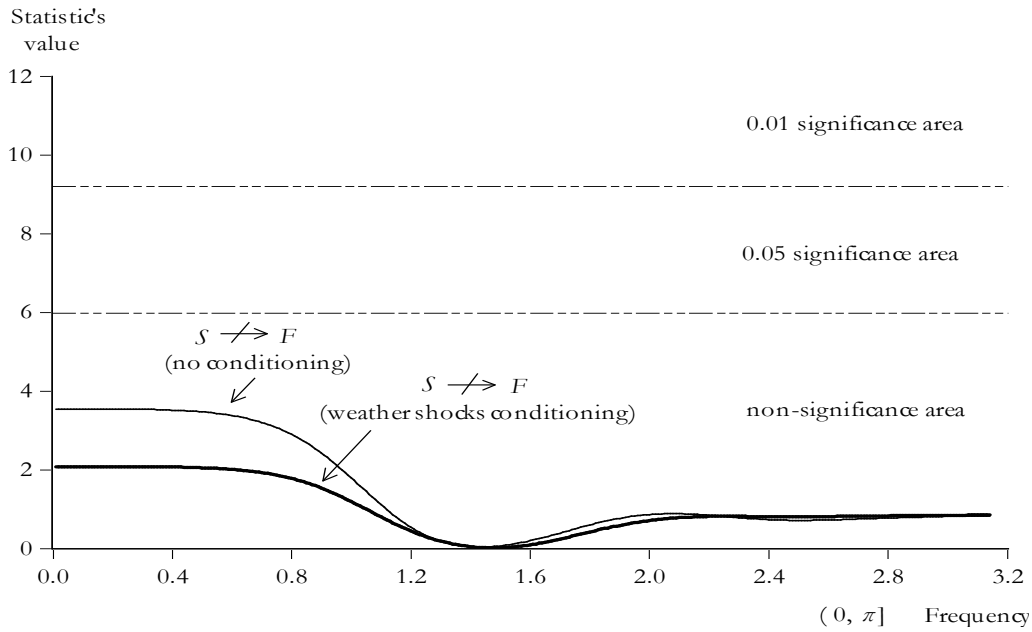


Fig. 3. Spot prices do not cause futures prices ($S \not\rightarrow F$)

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 5 for both cases (with and without weather shocks conditioning).

Turning now our attention to the hypothesis of the opposite direction ($F \not\rightarrow S$), that is futures prices do not cause spot prices, the testing outcome reveals a different significance pattern with respect to the previous case. The causality testing results of the opposite hypothesis ($F \not\rightarrow S$) are illustrated in Figure 4. More specifically, for the return NG series and within a bivariate framework, the derived B&C test statistics show that the null hypothesis of causality is clearly rejected, even at the 0.01 significance level, for every frequency. Furthermore, we dig out no additional information when the testing procedure is re-applied by conditioning on weather shocks, since the significance pattern of the derived statistics is almost identical to the bivariate case. In other words, the null hypothesis is still rejected at the 0.01 significance level. To this point it is worth mentioning that when conditioning is taking place, then the derived statistics are systematically lower than the respective statistics in the bivariate case. Such evidence can be attributed to the key role that the weather conditions play as an indirect channel of causality for the shaping of the NG spot prices. On the whole, the predictive content that the futures prices have with respect to the spot prices is significantly supported.

Additionally, the same inference for both hypotheses, $S \not\rightarrow F$ and $F \not\rightarrow S$, is also confirmed by the standard Granger causality test as this is depicted in Table 4. With and without conditioning, the former hypothesis is not rejected while the latter is rejected, supporting this way the findings of the B&C test.

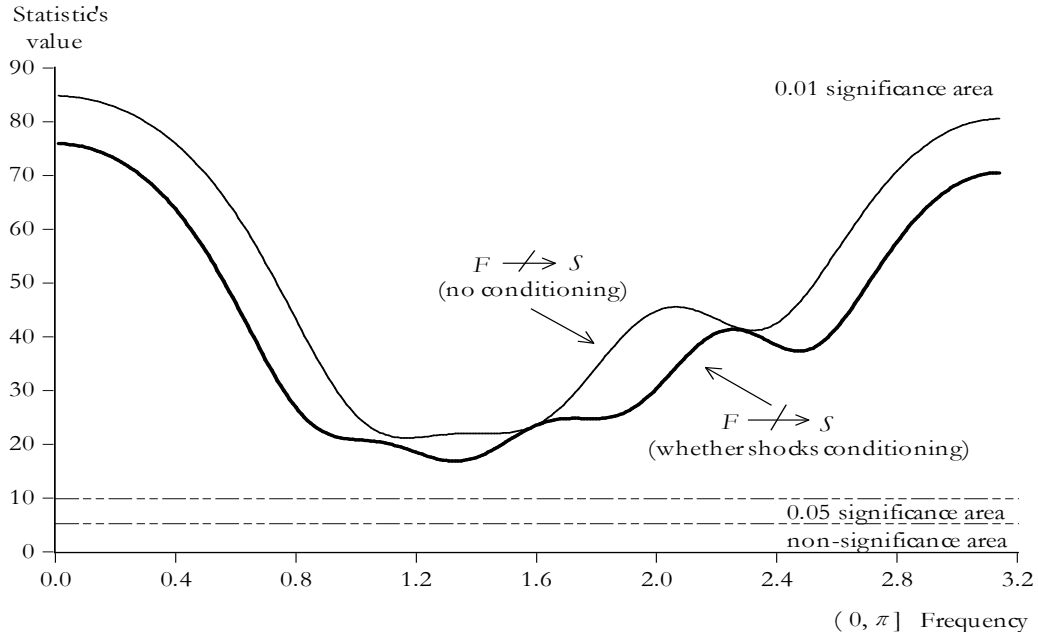


Fig. 4. Futures prices do not cause spot prices ($F \not\rightarrow S$)

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 5 for both cases (with and without weather shocks conditioning).

Table 4. Standard Granger non-causality for the unfiltered series.

Specification	$S \not\rightarrow F$		$F \not\rightarrow S$	
	Statistic	*p*-value	Statistic	*p*-value
VECM model (without conditioning)	1.963	0.580	772.359***	0.000
VECM model (weather shocks conditioning)	2.541	0.467	716.691***	0.000

Notes: The reported statistics are chi-squared statistics which are derived from the respective Wald test. The symbol *** denotes the rejection of the null hypothesis of non-causality at the 0.01 significance level. Finally, the arrow signifies the direction of causality.

Overall, the first-step results indicate, for the entire frequency domain,[12] convincing evidence of unidirectional causality running from the NG futures prices towards the spot prices, and not vice versa. Unfortunately, the performed causal analysis in this first step offers a precarious picture about the true nature of causality. The only clue that we have got so far is the verification of a unidirectional short-run and long-run causality, but we cannot yet reply with certainty to a set of critical questions such as: Is there any non-linear causality? In the case where a non-linear causality exists, does it hold for short-run or long-run components of the series? Therefore, to ensure that the identified causality is solely non-linear in nature we need to perform, the same testing procedure after filtering-out the series through a VECM specification and a GARCH-BEKK (1,1) model.

[12] Significance over the entire examined frequency domain implies short-run and long-run causality.

5.2 Causality for the VECM and GARCH-BEKK filtered series

Having completed the first step, we now continue by re-conducting both causality tests, after filtering-out the linear components of the series within a VECM framework. This second step will allow us to untangle the source of the causality identified in the first step. Under the hypothesis that the NG spot prices do not cause futures prices ($S \not\rightarrow F$), the frequency domain testing results for the first-moment filtered series, before and after the conditioning on weather shocks, are presented in Figure 5. As can be seen, the implemented causality test fails, in both cases, to reject the null hypothesis of no causality. Actually, the revealed significance pattern of the derived statistics, after a first-moment filtering, is indistinguishable from the pattern observed in the first step (see Figure 3). The persistence in the non-significance implies that there is absence of predictability running from the NG spot prices towards the futures prices at all moments of the series (first- and higher-order moments). As a confirmation of this, the GARCH-BEKK (1,1) filtered series, illustrate no predictive content at the entire frequency band for the same examined hypothesis (see Figure 6). The empirical insights that could be derived from our analysis up to this point do not support the existence of a valuable information set in the NG spot market that can improve our ability to predict upcoming movements in the NG futures market.

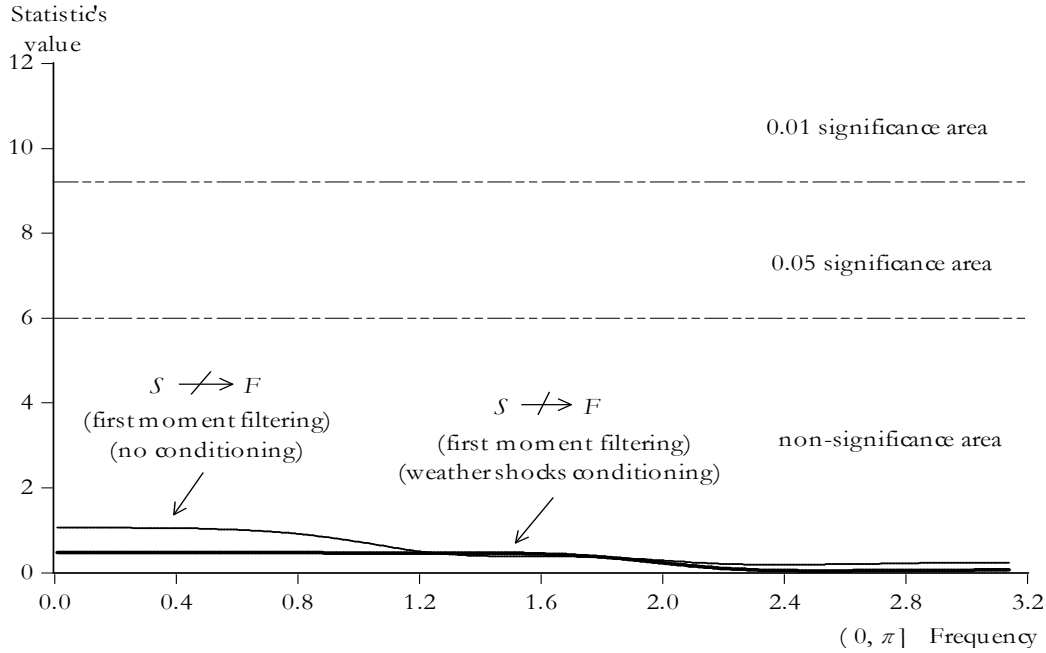


Fig. 5. Spot prices do not cause futures prices ($S \not\rightarrow F$), VECM filtered series

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 4 for both cases (with and without weather shocks conditioning).

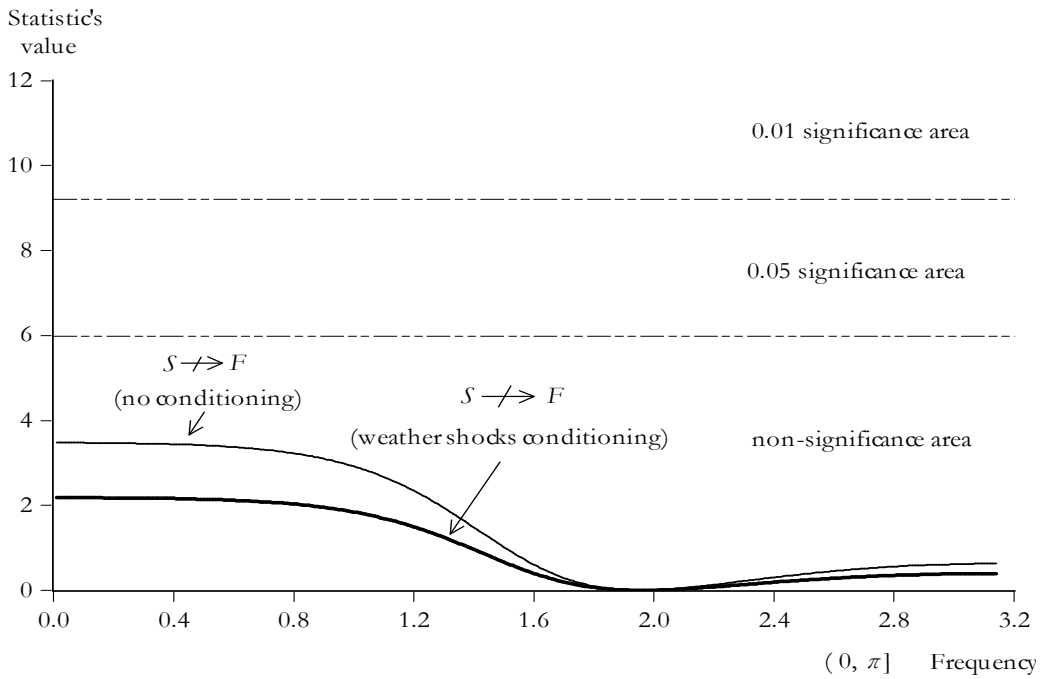


Fig. 6. Spot prices do not cause futures prices ($S \not\rightarrow F$), GARCH-BEKK filtered series

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 4 for both cases (with and without weather shocks conditioning).

Revealing, relative to the first step, is the resulting causality inference when the hypothesis that the NG futures market does not cause the spot market ($F \not\rightarrow S$) is tested. The said hypothesis is once more examined using the VECM and the GARCH-BEKK filtered series before and after conditioning on weather shocks. The testing results for the VECM filtered series with and without conditioning are illustrated in Figure 7. The interesting fact is that the null hypothesis ($F \not\rightarrow S$) is rejected at the 0.01 significance level, however this is not the case for the entire frequency interval. For the bivariate system (no conditioning) the rejection of the null takes place for frequencies that belong to the interval (0, 0.50π], while the respective rejection interval for the trivariate system (conditioning on weather shocks) is (0, 0.52π]. This finding actually implies that long-run cycle components of the de-meaned futures prices are those that offer non-linear predictive power with regard to the de-meaned spot prices. According to Baek and Brock (1992), after a first-moment filtering, any remaining predictive power can be considered as non-linear. Furthermore, significance at frequencies receiving values less than or equal to 0.50π (0, 1.56), implies predictability for wave lengths of more than 4.0 days. Similarly, significance at frequencies below 0.52π (0, 1.62) implies predictability for wave lengths of more than 3.9 days.

On the contrary, at the 0.01 significance level, we fail to reject the null hypothesis for the remaining short-run frequencies, that is (0.50π, π] for the bivariate case and (0.52π, π] for the trivariate case. The fact that the significance persists in low frequencies, after carrying out the first-moment filtering, while it vanishes in high frequencies consists precious information regarding the true nature of the investigated causality. Specifically, we may infer two important conclusions: (a) the short-run cyclical components of the NG futures prices offer predictive power with respect to the conditional mean of the NG spot prices and (b) the long-run cyclical components of the NG futures prices can contribute mainly in predicting the second- or higher-order conditional moments of NG spot prices (non-linear causality). Provided that causality persists, after the first-moment filtering, we implement a second-moment filtering. The B&C testing results for the GARCH-BEKK (1,1) filtered series are illustrated in Figure 8. The results show that the NG futures prices do not offer predictive with respect to the spot prices for the entire frequency range. Therefore, we infer that the identified predictive power is attributed solely to the second conditional moment of the series. Finally, Table 5 below offers, for selected frequencies, a summary of the causality analysis which has been executed in Section 5.

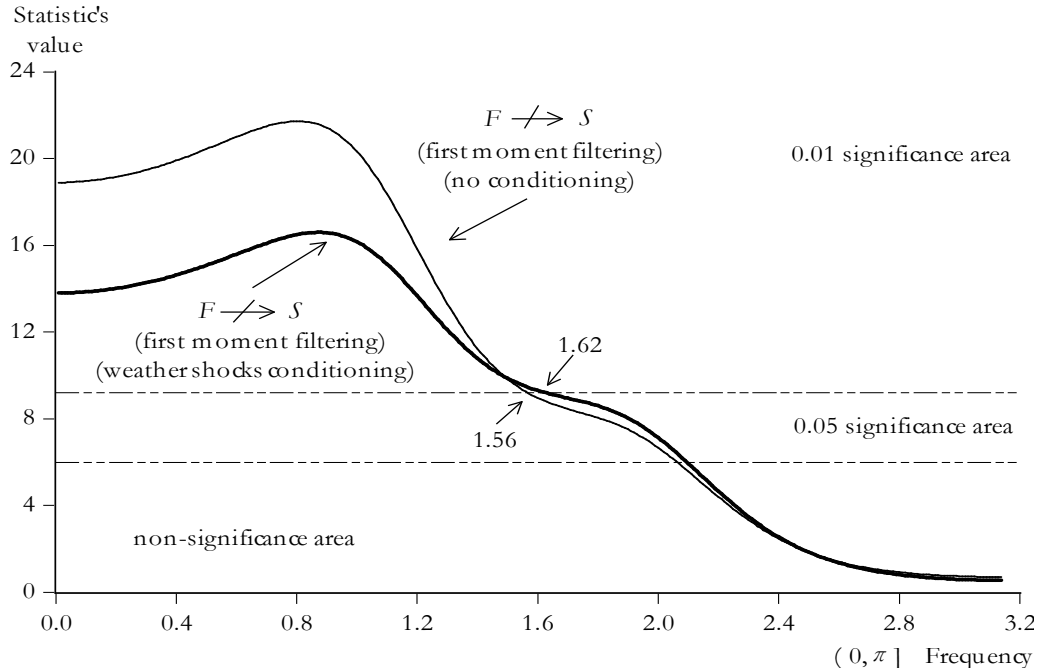


Fig. 7. Futures prices cause spot prices ($F \not\rightarrow S$), VECM filtered series

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 4 for both cases (with and without weather shocks conditioning).

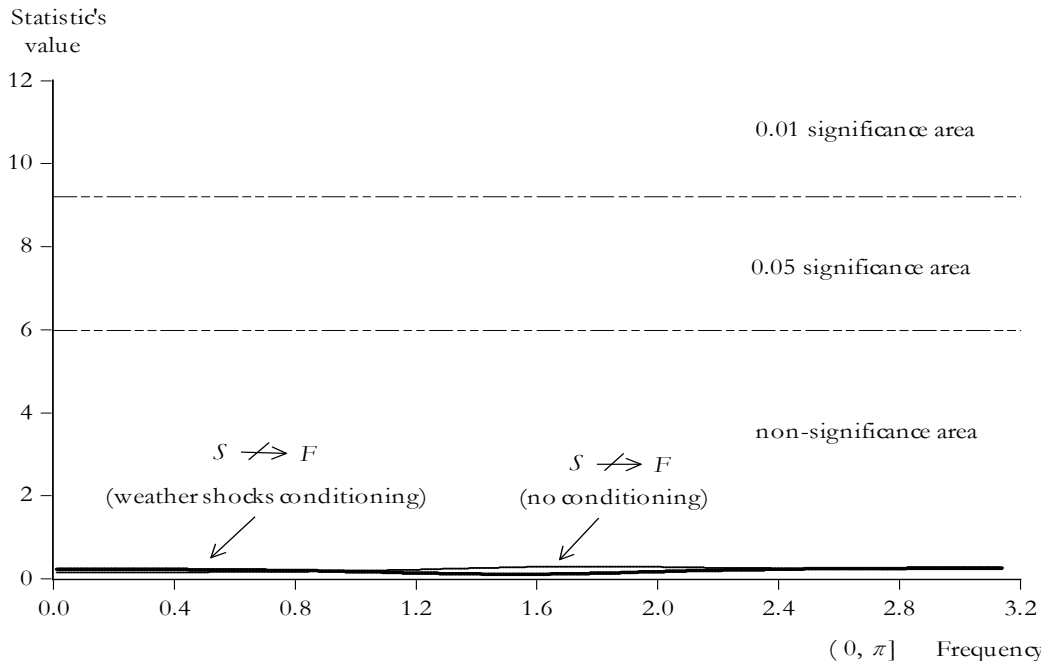


Fig. 8. Futures prices cause spot prices ($F \not\rightarrow S$), GARCH-BEKK filtered series

Note: The VAR lag-length used for the implementation of the frequency domain causality test is 4 for both cases (with and without weather shocks conditioning).

Table 5. Summary of the spectrum causality testing for selected frequencies.

Causality direction	Selected spectrum values					Causality inference	Causality interval
	0^+	$\pi/4$	$\pi/2$	$3\pi/4$	π		
Panel *A* (without conditioning)							
Return series							
$S \not\rightarrow F$	3.55	2.94	1.49	0.75	0.82	$\not\in$	-
$F \not\rightarrow S$	84.80***	44.39***	22.93***	41.53***	80.53***	$\in$	$(0, \pi]$
VECM residuals							
$S \not\rightarrow F$	1.06	0.92	0.38	0.19	0.23	$\not\in$	-
$F \not\rightarrow S$	18.90***	21.27***	9.18**	2.79	0.71	$\underset{\sim}{\in}$	$(0, 0.50\pi]$
GARCH-BEKK residuals							
$S \not\rightarrow F$	3.47	3.24	0.72	0.27	0.62	$\not\in$	-
$F \not\rightarrow S$	0.15	0.17	0.28	0.25	0.23	$\not\in$	-
Panel *B* (weather shocks conditioning)							
Return series							
$S \not\rightarrow F$	2.08	1.80	0.06	0.81	0.85	$\not\in$	-
$F \not\rightarrow S$	75.85***	27.62***	22.76***	39.57***	70.51***	$\in$	$(0, \pi]$
VECM residuals							
$S \not\rightarrow F$	0.48	0.47	0.44	0.05	0.06	$\not\in$	-
$F \not\rightarrow S$	13.81***	16.46***	9.45***	2.90	0.56	$\underset{\sim}{\in}$	$(0, 0.52\pi]$
GARCH-BEKK residuals							
$S \not\rightarrow F$	2.18	2.04	0.47	0.17	0.39	$\not\in$	-
$F \not\rightarrow S$	0.23	0.20	0.11	0.22	0.26	$\not\in$	-

Notes: (a) the *** symbol indicates statistical significance at the 0.01 significance level; (b) the arrow denotes the direction of causality; (c) the symbols $\in$, $\underset{\sim}{\in}$ and $\not\in$ stand for existence of causality over the entire frequency domain, the existence of causality over a segment in the frequency domain and no causality over the entire frequency domain, respectively; (d) the symbol 0^+ indicates a spectrum value that is positive and close to zero and finally; (e) the indicated causality intervals in the last column refer to a significant causality at the 0.01 significance level.

Overall, it can be argued that the frequency domain causality tests applied into the filtered series provide a much clearer picture regarding the existing nature of causality between spot and future prices in the NG market. Given that the significance vanishes for wave lengths of less than or equal to 4 days (after filtering the series with a VECM specification), it becomes clear that in the short-run the predictive power of F with respect to S (established in the first step) is attributed entirely to the conditional mean. Additionally, the persistence of the significance for wave lengths of more than 4 days for the VECM filtered series along with the entire absence of significance for any wave length for the GARCH-BEKK (1,1) filtered series, jointly these findings support the hypothesis that the long-run predictive power of F with respect to S is heavily attributed to the second conditional moment.

Finally, the two hypotheses, $S \not\rightarrow F$ and $F \not\rightarrow S$, for the VECM (see Table 6) and the GARCH-BEKK (1,1) (see Table 7) filtered series, with and without conditioning on weather shocks, are also examined by using the standard Granger non-causality test. In all cases the standard Granger non-causality test does not reject the examined each time hypothesis, failing this way to reveal the true nature of causality between spot and future prices in the NG market.

Table 6. Standard Granger non-causality for the VECM-filtered series.

Specification	$S \rightarrow F$		$F \rightarrow S$	
	Statistic	*p*-value	Statistic	*p*-value
VAR model (without conditioning)	0.079	0.961	0.797	0.671
VAR model (weather shocks conditioning)	0.082	0.959	0.357	0.836

Notes: The reported statistics are chi-squared statistics which are derived from the respective Wald test. Finally, the arrow signifies the direction of causality.

Table 7. Standard Granger non-causality for the GARCH-BEKK (1,1)-filtered series.

Specification	$S \rightarrow F$		$F \rightarrow S$	
	Statistic	*p*-value	Statistic	*p*-value
VAR model (without conditioning)	3.446	0.327	0.306	0.958
VAR model (weather shocks conditioning)	1.814	0.611	0.454	0.928

Notes: The reported statistics are chi-squared statistics which are derived from the respective Wald test. Finally, the arrow signifies the direction of causality.

6. Conclusions

This study examines the nature of causality between NG spot and futures prices (with one-month maturity) at the NYMEX market. The time period covered is fairly extensive, spanning from January 7, 1997, up to July 30, 2013 (4142 observations). Given that the existence of non-linearities in financial markets due, to several factors (transaction costs, noise traders, market frictions, and market microstructure effects), is a well-verified stylized fact, the adoption of a non-linear causality paradigm results as an unquestionable and inescapable necessity. Apart from the non-linear analytical framework, an interesting feature of this study is the fact that we disentangle the identified causality between short-run and long-run causality as well as between causality that is attributed to the first conditional moment and to higher-order moments. In particular, the B&C test and the standard Granger causality test are initially conducted into the return series and, at a later phase, after filtering out the first conditional moment, both tests are re-conducted once again. Given the existence of a long-run equilibrium, the first-moment filtering takes place within a VECM specification. Additionally, to capture possible hidden channels of causality, our analysis is not restricted to the bivariate level, but also extends to a trivariate level, conditioning on weather shocks. As a last step, we implement again the two causal tests after a second-moment filtering based on a bivariate GARCH-BEKK (1,1) specification.

The methodological framework used in this study contributes towards a deeper understanding of the role that the NG futures prices play in predicting NG spot prices. Our empirical findings provide strong statistical evidence in favor of a unidirectional causality that runs from the NG futures prices towards the spot prices. In particular, due to its dynamic features the B&C test reveals that the short-run cyclical components of F, with wave lengths of less than 4 days, encompass predictive power with respect to the conditional mean of S, while the long-run cyclical components of F, with wave lengths of more than 4 days, encompass predictive power with respect to the conditional variance of S. In other words, our findings show the following important characteristics; (a) there is unidirectional causality running from the NG futures prices towards the NG spot prices, and not vice versa; (b) for high frequencies the causality is vastly connected to the first conditional moment and, finally; (c) for low frequencies the causality is attributed to the second conditional moment. In other words, the identified predictive power over the frequency band proves to be asymmetric with respect to the first- and the second-conditional moments of the series.

Overall, it can be argued that the findings of this work shed some interesting new light on the true nature of causality between the NG spot and futures prices at NYMEX. Thus, this additional knowledge of the nature of causality may contribute towards a better understanding of the existing interdependencies between the NG spot and futures markets. Our findings, by enriching the available information set for the NG spot and futures markets, can be used by investors and other market participants in order to develop more efficient investment strategies. For instance, the high frequency components of futures contracts with one-month maturity can be utilized to realize excess returns in the spot market, while the low frequency components of futures contracts with one-month maturity can be used to improve investors' ability to appraise the existing risk in the NG spot market.

References

Asche, F., B. Misund and M. Sikveland (2013). The relationship between spot and contract gas prices in Europe. *Energy Economics* 38, 212-217.

Baek, E.G. and A.W. Brock (1992). A general test for non-linear Granger causality: bivariate model. Technical report, Korean Development Institute and University of Wisconsin, Madison.

Bierbrauer, M., C. Menn, S.T. Rachev and S. Truck (2007). Spot and derivative pricing in the EEX power market. *Journal of Banking and Finance* 31, 3462-3485.

Bopp, A.E. and S. Sitzer (1987). Are petroleum futures prices good predictors of cash value? *Journal of Futures Markets* 7, 705-719.

Bopp, A.E. and G. Lady (1991). A comparison of petroleum futures versus spot prices as predictors of prices in the future. *Energy Economics* 13, 274-282.

Breitung, J. and B. Candelon (2006). Testing for short and long-run causality: a frequency domain approach. *Journal of Econometrics* 132, 363-378.

Brock, W.A., J.A. Scheinkman, W.D. Dechert and B. LeBaron (1996). A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197-235.

Chng, M.T. (2009). Economic linkages across commodity futures: Hedging and trading implications. *Journal of Banking and Finance* 33, 958-970.

Considine, T. (2000). The impacts of weather variations on energy demand and carbon emissions. *Resource and Energy Economics* 22, 295-314.

Darrat, A.F., S. Rahman and M. Zhong (2002). On the role of futures trading in spot market fluctuations: Perpetrator of volatility or victim of regret? *Journal of Financial Research* 25, 431-444.

Daskalakis, G., D. Psychoyios and R. Markellos (2009). Modeling CO_2 emission allowance prices and derivatives: Evidence from the European trading scheme. *Journal of Banking and Finance* 33, 1230-1241.

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

Elkhafif, M (1996). An iterative approach for weather-correcting energy consumption data. *Energy Economics* 18, 211-220.

Garbade, K. D. and W.L. Silber (1983). Price movement and price discovery in futures and cash markets. *Review of Economics and Statistics* 65, 289-297.

Gebre-Mariam, Y.K. (2011). Testing for unit roots, causality, cointegration, and efficiency: The case of the northwest US natural gas market. *Energy* 36, 3489-3500.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77, 304-324.

Geweke, J. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* 79, 907-915.

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

Harbo, I., S. Johansen, B. Nielsen and A. Rahbek (1998). Asymptotic inference on cointegration rank in partial systems. *Journal of Business and Economic Statistics* 16, 388-399.

Hosoya, Y. (1991). The decomposition and measurement of the interdependence between second-order stationary process. *Probability Theory and Related Fields* 88, 429-444.

Hsiao, C. (1981). Autoregressive modelling and money-income causality detection. *Journal of Monetary Economics* 7, 85-106.

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159-178.

Lee, C.C. and J.H. Zeng (2011). Revisiting the relationship between spot and futures oil prices: Evidence from quantile cointegration regression. *Energy Economics* 33, 924-935.

Lien, D. and L. Yang (2008). Asymmetric effect of basis on dynamic futures hedging: Empirical evidence from commodity markets. *Journal of Banking and Finance* 32, 187-198.

Movassagh, N. and B. Modjtahedi (2005). Bias and backwardation in natural gas futures prices. *Journal of Futures Markets* 25, 281-308.

Mishra, V. and Smyth R. (2016). Are natural gas spot and futures prices predictable? *Economic Modelling* 54, 178-186.

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics* 29, 46-63.

Ng, V.K. and S.C. Pirrong (1996). Price dynamics in refined petroleum spot and futures markets. *Journal of Empirical Finance* 2, 359-388.

Pesaran, H., Y. Shin and R. Smith (2000). Structural analysis of vector error correction models with exogenous I(1) variables. *Journal of Econometrics* 97, 293-343.

Quan, J. (1992). Two step testing procedure for price discovery role of futures prices. *Journal of Futures Markets* 12, 139-149.

Rahbek, A. and R. Mosconi (1999). Cointegration rank inference with stationary regressors in VAR models. *Econometrics Journal* 2, 76-91.

Shawky, H.A., A. Marathe and C.L. Barrett (2003). A first look at the empirical relation between spot and futures electricity prices in the United States. *Journal of Futures Markets* 23, 931-955.

Silvapulle P. and I.A. Moosa (1999). The relationship between spot and futures prices: Evidence from the crude oil market. *Journal of Futures Markets* 19, 175-193.

Toda, H.Y. and T. Yamamoto (1995). Statistical inference in vector autoregressions with possible integrated processes. *Journal of Econometrics* 66, 225-250.

Zhang, Y. and Liu, L. (2018). The lead-lag relationships between spot and futures prices of natural gas. Physica A: Statistical Mechanics and its Applications 490, 203-211.