The Informational Efficiency of European Natural Gas Hubs: Price Formation and Intertemporal Arbitrage

*Sebastian Nick**

ABSTRACT

In this study, the informational efficiency of the European natural gas market is analyzed by empirically investigating price formation and arbitrage efficiency between spot and futures markets. Econometric approaches accounting for non-linearities induced by the low liquidity-framework and by technical constraints of the considered gas hubs are specified. The empirical results reveal that price discovery generally takes place on the futures market. Thus, the futures market seems to be more informationally efficient than the spot market. The theory of storage seems to hold at all hubs in the long run. There is empirical evidence of significant market frictions hampering intertemporal arbitrage. UK's NBP and Austria's CEGH seem to be the hubs at which arbitrage opportunities are exhausted most efficiently, although there is convergence in the degree of intertemporal arbitrage efficiency over time at the hubs investigated.

Keywords: Natural gas market, Informational efficiency, Nonlinear causality, Threshold error correction, Kalman filter

http://dx.doi.org/10.5547/01956574.37.2.snic

1. INTRODUCTION

The price signals of commodity spot and futures markets are of economic significance for market participants and various stakeholders, as they tend to ensure an efficient allocation of resources. However, the extent to which commodity spot and futures prices fulfil their function crucially depends on the informational efficiency of the respective market. Economic theory suggests that sufficient market liquidity facilitates the processing of information into valid price signals. Thus, the efficiency of markets that are still immature and suffer a lack of liquidity may be questioned. This holds true for the natural gas wholesale markets within continental Europe. Spot markets for immediate delivery of natural gas as well as futures markets have emerged rather recently as a consequence of the natural gas directives of the European Parliament (EU, 2003; EU, 2009), aiming towards an integrated and competitive European gas market. Liquidity on these markets, though rising, is still low compared to the mature gas markets in the UK or the U.S. The limited liquidity of both spot and futures markets at continental European gas hubs has entered the scientific debate, as European gas pricing is currently undergoing a transition phase from traditional oil indexed pricing of long-term contracts (LTC) to an increase in the significance of hub-based pricing.[1]

1. For an elaborated discussion of the economics fostering the transition from oil-indexation to hub-based pricing, see Stern and Rogers (2011). A real-life illustration are the current renegotiations of LTCs between various continental European gas importers with their suppliers (ICIS, 2013).

* Institute of Energy Economics, University of Cologne, Vogelsanger Straße 321, 50827 Cologne, Germany. E-mail: nick.sebastian.sn@gmail.com, phone: +49 15233913648.

The shifting towards hub-based pricing of natural gas in continental Europe is based on the assumption that the respective hubs are capable of providing valid price signals. In this context, this work seeks to shed light on the informational efficiency of European gas hubs by empirically investigating two areas that allow for valuable insights with regard to market efficiency: The price discovery process at spot and futures markets for the same underlying asset and the efficiency of intertemporal arbitrage between these two markets. It draws upon econometric approaches for six major European gas hubs, where the mature and liquid British hub serves as a benchmark for the other hubs.[2]

This paper extends empirical research on natural gas markets in various ways: Foremost, it is innovative as it analyzes the informational efficiency of the European gas hubs through the investigation of the price formation process and the efficiency of intertemporal arbitrage. Second, it explicitly addresses the specific characteristics of the European gas market, namely low liquidity and technical constraints, by nonlinear econometric approaches. Third, it allows innovative insights into the evolution of informational efficiency at European gas hubs over time. The empirical results of this study yield comprehensive insights into the informational efficiency on European natural gas markets. First, they show that the futures market is more informationally efficient than the spot market as price discovery generally takes place on the futures market. Second, the analysis of intertemporal arbitrage reveals that there is a stable long-run equilibrium between spot and futures markets, but short-run equilibrium deviations can be quite persistent, pointing towards significant frictions in intertemporal arbitrage trading. Third, the increase in liquidity seems to have improved informational efficiency only at two of the hubs considered.

The remainder of the paper is organized as follows: Section 2 provides the underlying economic theory and discusses relevant previous research. Section 3 presents the data used in this study and preliminary statistical tests, while Section 4 provides information with regard to market liquidity and the flexibility potential of gas storages at the European gas hubs. In Section 5, price discovery at European gas hubs is investigated using linear and nonlinear causality testing. Section 6 explores the long-run relationship of spot and futures markets at the considered hubs and analyzes the efficiency of intertemporal arbitrage. A state-space approach to capture the evolution of intertemporal arbitrage efficiency over time is specified in Section 7. Section 8 concludes.

II. THEORETICAL CONSIDERATIONS AND PREVIOUS RESEARCH

Efficient markets are expected to process relevant information instantaneously (Fama, 1970). Within an intertemporal context, this implies that spot and futures markets should react simultaneously to news that affects both markets. Consequently, there should be no structural lead-lag relationship between the two markets (Zhang and Jinghong, 2012). This is in line with the weak-form efficiency hypothesis stating that excess returns on spot and futures markets should be unpredictable as otherwise risk-free profits may be generated (Arouri et al., 2013). However, if one of the markets is more efficient in processing information, this market may become the leading market. In that case, price discovery takes place at the leading market and the price signal is subsequently transmitted to the following market.

2. Although the British gas hub may be considered as an appropriate benchmark for pricing European gas imports in terms of liquidity, the limited cross-border transportation capacity between mainland Europe and the UK as well as the implied currency risks for European gas traders carrying out transactions at this hub suggest the need for a continental European gas price benchmark.

There are various hypotheses with regard to the differences in informational efficiency of spot and futures markets and the resulting systematic relationship. Silvapulle and Moosa (1999) and Bohl et al. (2012) suggest that futures prices may react quicker to the arrival of information, since informationally efficient speculators are only active in this market. As a result, information processing and price discovery occur in the futures market and the spot prices adjust accordingly until an arbitrage-free equilibrium is achieved. In contrast, Moosa and Al-Loughani (1995) argue that the spot market should lead the futures market because arbitrageurs react to spot price movements by engaging in futures market positions. Empirical research on price discovery on natural gas spot and futures markets is scarce. Dergiades et al. (2012) explore linear and nonlinear causality relationships between spot and futures prices at the U.S. gas hub. Focusing on the northwest U.S. natural gas market, Gebre-Mariam (2011) tests for causality among spot and futures market prices and market efficiency by drawing upon cointegration techniques.

Concerning the European gas market, Stern (2014) argues that the oil price indexation of natural gas imports to Europe does not reflect market fundamentals anymore, triggering a switch to hub pricing in Europe. The regional integration of European gas hubs and the efficiency of regional price arbitrage have been empirically explored (e.g., Neumann et al., 2006; Growitsch et al., 2012). In the context of regional market integration, Keyaerts and D'haeseleer (2012) discuss and quantify potential gains in market efficiency induced by cross-border procurement of natural gas balancing services. Asche et al. (2013) empirically analyze the relationship between European natural gas spot prices, the prices of long-term natural gas import contracts and the price of crude oil. In line with Growitsch et al. (2012), Asche et al. (2013) find a high level of regional market integration using cointegration technique while their findings point towards significant influence of crude oil prices on both spot and contract prices for natural gas.

The price formation process at the European spot and futures markets, in contrast, has thus far only received limited attention. Schulz and Swieringa (2013) investigate the price discovery process of the European natural gas market using high-frequency data. Based on a regression approach applied to different European physical and financial natural gas contracts, they conclude that the futures contract of the British hub NBP displays greater price discovery than the other spot and futures markets considered in their study. Moreover, the NBP futures contract price exhibits the largest contribution to the long-run equilibrium between the different markets analyzed. Schulz and Swieringa (2013) attribute this finding to the superior maturity of the British natural gas hub.

The theory of storage suggests that spot and futures markets for storable commodities are linked through transactions of market participants optimizing their portfolios intertemporally, resulting in a stable long-run relationship between these markets (Working, 1949). The cost-of-carry condition is characterized by the equivalence of the price of a futures contract in period t with the delivery in period $t+k$, $F_{t+k|t}$, and the spot price compounded with the respective interest rate $r_{t+k|t}$, $S_t(1+r_{t+k|t})$ plus the storage costs $w_{t+k|t}$ adjusted for the convenience yield $c_{t+k|t}$ (i.e., the economic benefit of physical ownership). This condition can be stated as

$$F_{t+k|t} = S_t(1+r_{t+k|t}) + w_{t+k|t} - c_{t+k|t} \tag{1}$$

Deviations from the intertemporal equilibrium may trigger arbitrage activity by market participants. In this context, arbitrage can be considered as the economic activity of generating risk free profits by taking advantage of the substitutability between commodity spot and futures markets (Schwartz and Szakmary, 1994). As outlined by Huang et al. (2009), a long arbitrage position, i.e., buying the commodity on the spot market and selling a futures contract, is profitable if the basis

$b_t = F_t - S_t$ exceeds the difference of warehouse costs and convenience yield, adjusted for the interest rate r:

$$b_t - S_t r_{t+k|t} > w_{t+k|t} - c_{t+k|t} \tag{2}$$

In contrast, a short arbitrage position, i.e., selling the commodity on the spot market and buying a futures contract, generates profits if

$$b_t - S_t r_{t+k|t} < -(w_{t+k|t} - c_{t+k|t}) \tag{3}$$

The theory of storage has been empirically analyzed for different commodity markets by Fama and French (1987), and more recently by Considine and Larson (2001) and Huang et al. (2009). With regard to the European natural gas market, Stronzik et al. (2009) find significant deviations from the theory of storage equilibrium for three European hubs for the period 2005 to 2008 using indirect testing procedures. However, the efficiency of intertemporal arbitrage activity at European gas hubs has not yet been addressed in the existing literature. The subsequent sections seek to bridge this research gap in the area of gas markets.

III. SAMPLE DESCRIPTION AND PRELIMINARY DATA ANALYSIS

The sample comprises daily spot, one month-ahead (m + 1), two month-ahead (m + 2) and three month-ahead (m + 3) futures prices for the German hubs 'NetConnect Germany' (NCG) and 'Gaspool' (GP)[3], the Dutch gas hub 'Title Transfer Facility' (TTF)[4], UK's 'National Balancing Point' (NBP)[5], French's hub 'Point d'Echange de Gaz Nord' (PEGN)[6] and the Austrian 'Central European Gas Hub' (CEGH)[7] for the period October 2007 to August 2012.[8] All prices represent the settlement prices of the respective trading day. Monthly futures contracts are preferred to quarterly or seasonal products to account for the tendency towards the trading of monthly contracts with short maturity (NMA, 2012). Descriptive statistics of the return series, computed as the differences in the logarithms of two consecutive daily settlement prices, are provided in Table 1.

For the subsequent econometric analysis, the stationarity properties of all price series are investigated using the Augmented Dickey Fuller (ADF) test and the nonparametric Phillips-Perron test to avoid misleading statistical inference. In general, the null hypothesis of a unit root in the log-level cannot be rejected, which is the case for the first differences (i.e., the daily returns).[9] Thus, the cost-of-carry hypothesis between the spot and futures markets at the considered hubs can be investigated using cointegration analysis as proposed by Johansen (1988).[10] The null hypothesis of

3. Spot and futures prices were obtained from the European Energy Exchange.

4. Spot prices were obtained from Endex, futures prices from the Intercontinental Exchange.

5. Spot prices were obtained from Endex, futures prices from the Intercontinental Exchange.

6. Spot and futures prices were obtained from Powernext.

7. Spot and futures prices were obtained from the Central European Gas Hub AG.

8. For the French hub, data is not available before 2009, while for CEGH, no data is available before December 2010. In particular for CEGH, the empirical results should therefore be interpreted cautiously due to the rather small sample size.

9. Only for the spot price series of CEGH, the result of the unit root tests depends on the test specification. However, since economic theory suggests that commodity price series should contain a unit root, the series is assumed to be integrated of order one.

10. This holds only true under the assumption that the determinants of the cost-of-carry relationship (i.e., the interest rate, storage costs and the convenience yield) exhibit stationary character. Both economic intuition and the short maturity of the future contracts considered suggest that this assumption holds true in the context of this research.

Table 1: Descriptive Statistics of Gas Price Returns

	Observations	*Mean*	*Variance*	*Skewness*	*Kurtosis*
NCG Spot	1228	1.88E-04	0.0023	−0.5081	12.3466
NCG m + 1	1228	1.45E-04	0.0008	1.8054	21.6685
NCG m + 2	1228	1.53E-04	0.0007	2.1349	25.2165
NCG m + 3	1228	1.11E-04	0.0006	2.3307	23.7995
GP Spot	1228	2.08E-04	0.0021	−0.2827	11.3990
GP m + 1	1228	1.79E-04	0.0008	1.4077	18.3094
GP m + 2	1228	1.91E-04	0.0008	1.7719	21.4135
GP m + 3	1228	1.47E-04	0.0007	1.8552	19.8192
TTF Spot	1228	2.81E-04	0.0018	−0.1175	8.9574
TTF m + 1	1228	1.51E-04	0.0008	1.3689	14.1179
TTF m + 2	1228	1.55E-04	0.0007	1.596	19.2947
TTF m + 3	1228	1.29E-04	0.0006	1.9247	20.0573
NBP Spot	1268	2.23E-04	0.0062	−0.2147	18.9689
NBP m + 1	1268	2.36E-04	0.0011	2.5508	27.0689
NBP m + 2	1268	1.93E-04	0.0009	1.8292	19.7212
NBP m + 3	1268	2.13E-04	0.0007	1.5505	18.2572
PEGN Spot	900	3.70E-05	0.0016	−0.2831	10.6324
PEGN m + 1	900	1.72E-05	0.0008	1.0362	16.3075
PEGN m + 2	900	3.55E-05	0.0008	1.6671	23.1840
PEGN m + 3	900	9.70E-05	0.0007	0.0682	27.2667
CEGH Spot	427	1.65E-04	0.0009	−0.1259	26.5774
CEGH m + 1	427	2.81E-04	0.0003	0.7636	6.6321
CEGH m + 2	427	3.81E-04	0.0003	0.9432	6.7976
CEGH m + 3	427	4.49E-04	0.0003	0.7442	6.5284

no cointegration between spot and month-ahead prices can be rejected for all hubs.[11] The relevant test statistics are presented in the Appendix.

IV. THE ROLE OF LIQUIDITY AND STORAGE CAPACITY

Differences in the informational efficiency between the European gas hubs may be caused by different sources. Most notably, sufficient market liquidity is regarded as an important element for an efficient price formation process (see, e.g., Chordia et al., 2008). Besides, the availability of flexible storage capacity and a functioning third party access to these facilities may be a prerequisite for efficient intertemporal arbitrage activity. A direct empirical investigation of both potential efficiency determinants seems promising but suffers from the lack of suitable and comprehensive data sets. Nevertheless, the subsequent paragraphs present stylized facts on these potential determinants for the hubs analyzed in order to provide an indication.

The spot and futures markets of the gas hubs considered in this study differ significantly with respect to their liquidity. While the NBP hub can be considered as mature and liquid, the younger continental European hubs suffer from low liquidity despite steadily increasing trading volumes during the last years. The churn rate, defined as the ratio between the number of traded contracts and the number of contracts that result in physical delivery of the underlying asset, can

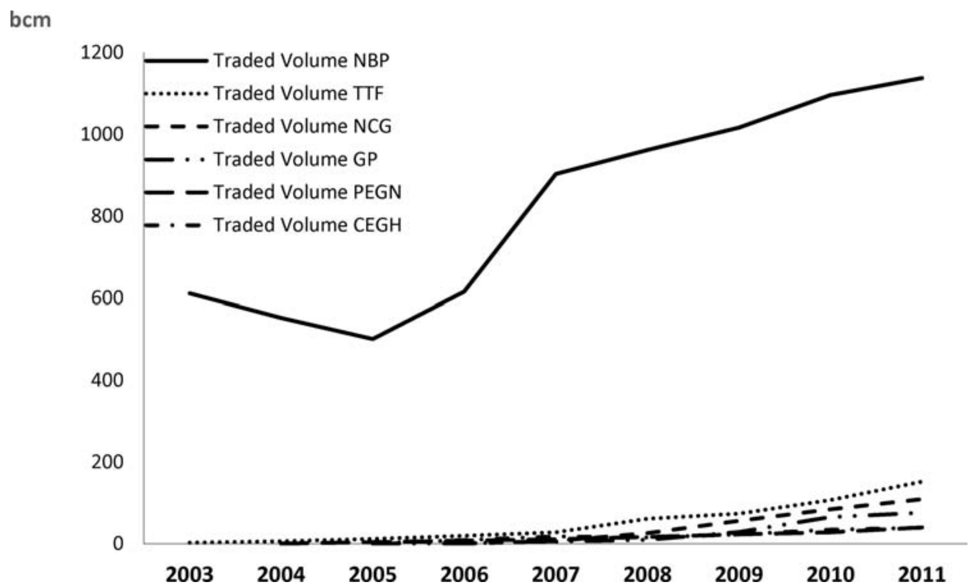
11. In the following, this study focuses on the month-ahead contracts. This is in line with the fact that the trading of futures contracts at the European gas hubs is centered on these contracts. Test statistics for futures contracts with longer maturity are presented in the Appendix. However, the choice of maturity does not alter the main empirical findings significantly.

Table 2: Liquidity at European Gas Hubs in Billion Cubic Meters (as of 2011)

	Physical Volume	*Traded Volume*	*Churn Rate*
NCG	35.5	108.5	3.1
GP	29.6	75.8	2.6
TTF	35.6	151.7	4.3
NBP	79.6	1137.2	14.3
PEG	12.8	39.8	3.1
CEGH	11.6	39.2	3.4

Source: IEA (2012a), Gasunie (2011), NCG (2011). The figures presented refer to total hub trades (sum of trades in the "Over The Counter" (OTC) market and those via exchanges).

Figure 1: Trading Volumes at European Gas Hubs



Source: IEA (2012a)

be used to assess the degree of financialization of commodity markets. Table 2 illustrates the differences among the hubs with regard to their churn rates. The historical development of traded volumes is presented in Figure 1. There is no agreement as to which churn rate is required for a market to be considered as sufficiently liquid. However, a churn rate in the range from eight to fifteen is frequently regarded as critical (IEA, 2012a). As can be seen in Table 2, only the churn rate of NBP is situated within this range. Based on the superior liquidity of the British hub, information processing and thus price formation is expected to be more efficient at NBP compared to the continental European hubs.

With regard to storage capacity, a first indicator is the ratio of aggregated working gas volume to annual gas consumption. In addition, the flexibility potential of the existing storage capacities is crucial for an efficient adjustment of storage flows in order to exploit arbitrage opportunities. Appropriate measures for the degree of gas storage flexibility are the shares of aggre-

Table 3: Storage Capacity and Flexibility Potential (as of 2011)

	WGV (bcm)	*C (bcm/a)*	*WGV/C*	*WC/WGV*	*IC/WGV*
Germany	20.3	77.6	0.2627	0.0215	0.0111
Netherlands	5.3	47.9	0.1098	0.0410	0.0112
UK	4.5	82.6	0.0529	0.0195	0.0055
France	12.9	41.2	0.3083	0.0216	0.0118
Austria	7.2	9.5	0.7554	0.0119	0.0094

Source: IEA (2012b), GIE (2011).

gated injection capacity (IC) and aggregated withdrawal capacity (WC) on aggregated working gas volume (WGV). Table 3 presents data on WGV, measured in billion cubic meters (bcm), consumption (C, in bcm per year) and the three flexibility indicators for Germany, the Netherlands, the UK, France and Austria.

The data emphasize the ample storage capacity of the German, French and Austrian gas markets. In contrast, storage capacity in the UK is rather scarce since the WGV only amounts to approximately 5 % of annual gas consumption. The Netherlands range between Germany and the UK in terms of this indicator. With regard to operational flexibility, Dutch gas storages seem most capable of adjusting operations to changing market conditions in the short-run, while UK storage facilities are fairly inflexible. From a technical point of view, the storage indicators thus suggest that the storage market in the UK is less supportive of efficient intertemporal arbitrage activity.

V. PRICE FORMATION AT EUROPEAN GAS HUBS: LINEAR AND NONLINEAR CAUSALITY TESTING

a. Econometric Methodology and Economic Interpretation

The Fama (1970) hypothesis of simultaneous information processing of markets for the same asset implies that there should be no systematic relationship between price changes on spot and futures markets. Otherwise price returns of one market may be helpful in predicting price returns of another market, allowing for risk-free profits. As a consequence, tests for a systematic relationship, i.e., causality tests, can be used to empirically test the hypothesis of simultaneous information processing. In other words, causality tests are helpful in testing whether a market *A* is quicker in processing information and hence more informationally efficient than another market *B*. In this case, market *B* follows the price changes of market *A*, which acts as the leading market. The finding of causality from a market *A* to a market *B* thus represents evidence of market *A* providing price discovery for market *B* in this example. In contrast, the hypothesis of equal informational efficiency does rule out systematic unidirectional causality between the markets considered. The most established econometric measure of causality is the concept of Granger causality (Granger, 1969). Granger causality exists if one variable is helpful in predicting future changes of another variable, i.e., the availability of current data on a certain variable reduces the forecast error of another variable. In statistical terms, a process x_t is said to cause a process y_t in the sense of Granger if

$$\sum\nolimits_z(h\,|\,\Omega_t) < \sum\nolimits_z(h\,|\,\Omega_t \backslash (x_s\,|\,s \leq t)) \text{ for at least one } h = 1,2,\ldots,N \tag{4}$$

where $\sum_z(h|\Omega_t)$ is the optimal mean squared error of an h-step forecast based on the information set Ω_t reflecting all past and current information (Lütkepohl, 2005).

The Granger causality test outlined above is only capable of investigating linear relationships. However, there is empirical evidence suggesting nonlinearities in the relationship of commodity spot and futures markets, which is usually attributed to the nonlinearity of transaction costs and market microstructure effects such as minimum lot sizes (Bekiros and Diks, 2008; Chen and Lin, 2004; Silvapulle and Moosa, 1999) as well as to asymmetric information and heterogeneous expectations of market participants (Arouri et al., 2013). There are good reasons to believe that these drivers of nonlinear interaction are relevant for the continental European gas hubs, since low liquidity and technical constraints at these hubs may foster market frictions. As a consequence, the empirical investigation of price discovery at the European gas market should take nonlinearities into account.

For this purpose, the nonlinear causality test of Diks and Panchenko (2006) can be applied based on the Hiemstra Jones Test (Hiemstra and Jones, 1994). Their methodology tests whether the whole current conditional distribution of a certain variable *A* has predictive power for the future conditional distribution of variable *B*. Thus, not only causality in the first but also in higher moments of distribution can be investigated. From an economic perspective, this allows the empirical analysis of nonlinear interaction between the two markets caused by transaction costs such as information costs and bid-ask spreads or by technical constraints such as restricted network or storage capacity.

In statistical terms, the null hypothesis of absent nonlinear Granger causality between two series is tested using their conditional distributions. Assuming stationarity, the null hypothesis of *Y* with respect to *X* implies that the conditional distribution of a variable *Z* given its past realization $Y = y$ equals the conditional distribution of *Z* given $Y = y$ and $X = x$. Thus, the joint probability functions and their marginals can be used to state the null hypothesis as

$$\frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)} * \frac{f_{Y,Z}(y,z)}{f_Y(y)} \tag{5}$$

Diks and Panchenko (2006) show that the null hypothesis can be reformulated as

$$q \equiv E[f_{X,Y,Z}(X,Y,Z)f_Y(Y) - f_{X,Y}(X,Y)f_{Y,Z}(Y,Z)] = 0 \tag{6}$$

As outlined by Diks and Panchenko (2005), the test statistic is corrected for possible size bias resulting from time-varying conditional distributions. Diks and Panchenko (2006) show that the adjusted test statistic is

$$T_n(\epsilon_n) = \frac{n-1}{n(n-2)} * \sum_i(\hat{f}_{X,Y,Z}(X_i,Y_i,Z_i)\hat{f}_Y(Y_i) - \hat{f}_{X,Y}(X_i,Y_i)\hat{f}_{Y,Z}(Y_i,Z_i)) \tag{7}$$

where $\hat{f}_W(W_i)$ is the estimator of the local density of a d_W-variate random vector W_i with

$$\hat{f}_W(W_i) = (2\epsilon_n)^{-d_W}(n-1)^{-1}\sum_{j,j\neq i} I_{ij}^W \tag{8}$$

where ϵ_n is the bandwidth depending on the sample size *n* and $I_{ij}^W = I(\|W_i - W_j\| < \epsilon_n)$ is an indicator function. Diks and Panchenko (2006) demonstrate that the distribution of the test statistic equals

$$\sqrt{n}\frac{(T_n(\epsilon_n)-q)}{S_n}\xrightarrow{d}N(0,1) \tag{9}$$

for a lag length of 1 and $\epsilon_n = C_n^{-\beta}$ with $C>0$ and $\frac{1}{4}<\beta<\frac{1}{3}$. S_n is the estimator of the asymptotic variance of $T_n(\cdot)$ (Bekiros and Diks, 2008). Furthermore, Diks and Panchenko (2006) show that the optimal bandwidth (i.e., minimizing the mean squared error of T_n) is

$$\epsilon_n^* = C^*\, n^{(-2/7)} \tag{10}$$

To sum up, causality tests are used in this study to investigate the Fama (1970) simultaneous information processing hypothesis and thus price discovery at European natural gas hubs. Besides linear causality patters, nonlinear causality is explicitly addressed to account for the low-liquidity framework of the continental European gas hubs and for the technical specifics of the natural gas market.

b. Empirical Results

The linear Granger causality test is carried out for the price returns within the vector error correction (VECM) framework.[12] In addition, the VECM-filtered residuals are tested for any remaining linear causality pattern. Table 4 contains the results of the linear Granger causality tests for the spot- and month-ahead return series. For the unfiltered return series, the null hypothesis of absent Granger causality can be rejected for the direction from futures to spot markets at all hub. This means that the change in the month-ahead futures price has explanatory character for the next day's spot price change. In contrast, only at GP there is (weak) empirical evidence of Granger causality from the spot to the futures market. Consequently, information is not processed simultaneously by spot and futures market participants. In fact, information is first processed within the futures market and subsequently transmitted to the spot market. Thus, the futures market seems to be the dominant market in terms of price discovery. The finding of the futures market providing price discovery for the spot market is especially noteworthy in the context of natural gas markets, where the information sets of spot and futures markets partially differ from one another. Most notably, short-run influences such as weather conditions or infrastructure outages are expected to affect spot market returns significantly, whereas their impact on the futures market should be limited. However, despite these specific characteristics of the purely physical spot market, the futures market still has significant explanatory power for the subsequent outcome of the spot market.

The informational superiority of the futures market may result from the broader scope of market participants at this market. The opportunity to trade futures contracts multiple times before maturity (and thus close out the trading position without taking physical delivery) makes the futures market attractive for hedgers and speculators without interest in physical delivery of the underlying asset. These additional market participants may cause a greater efficiency of information processing of the futures market compared to the one of the spot market, as suggested by Silvapulle and Moosa (1999) and Bohl et al. (2012). Overall, the empirical evidence of the month-ahead natural gas futures

12. In doing so, the cointegration relationship between spot and futures prices is explicitly accounted for to avoid misleading inference. Ignoring an existing cointegration relationship may lead to invalid results of linear and nonlinear Granger causality tests, as outlined by Chen and Lin (2004).

Table 4: Pairwise Linear Causality Tests for Gas Price Returns

Direction	*Chi-sq-Statistic*
NCG Spot on NCG m + 1	0.7909
NCG m + 1 on NCG Spot	275.68***
GP Spot on GP m + 1	8.0437**
GP m + 1 on GP Spot	211.18***
TTF Spot on TTF m + 1	4.2355
TTF m + 1 on TTF Spot	2121.7***
NBP Spot on NBP m + 1	3.6395
NBP m + 1 on NBP Spot	229.82***
PEGN Spot on PEGN m + 1	5.2072
PEGN m + 1 on PEGN Spot	146.50***
CEGH Spot on CEGH m + 1	0.6474
CEGH m + 1 on CEGH Spot	102.62***

Notes: *** (**) Denotes significance at the 99 (95) %-level. Granger causality has been investigated within the VECM-framework, explicitly accounting for the cointegration relationship.

market leading the corresponding spot market is in line with the findings of Root and Lien (2003) and Dergiades et al. (2012) for the U.S. natural gas market. For the VECM-filtered series, the null hypothesis of absent Granger causality cannot be rejected in any direction for all hubs (test statistics are provided in the Appendix). This suggests that all linear causality is captured by the VECM-model.

The nonlinear causality testing procedure is applied to the VECM-filtered residuals to ensure that any detected causality can be attributed to nonlinear interaction of the spot and futures markets. Following Diks and Panchenko (2006), the constant term C^* of the bandwidth ϵ_n is set to 8.[13]

As can be seen in Table 5, the null hypothesis of absent nonlinear Granger causality among spot and month-ahead return series can be rejected in both directions for all hubs except for CEGH.[14] However, this finding should be interpreted cautiously: As pointed out by Silvapulle and Moosa (1999), conditional heteroskedasticity of both series may distort the size of the nonlinear causality test. Following this argument, a multivariate GARCH model, the diagonal BEKK GARCH model of Engle and Kroner (1995), is applied to capture the dynamics in the second moment of distribution in both markets, filtering out conditional volatility effects.[15] Subsequently, the nonlinear causality test of Diks and Panchenko (2006) is used for the BEKK GARCH-filtered VECM residuals.

For all hubs except for Gaspool, the nonlinear causality from spot to futures markets disappears after BEKK-GARCH filtering while for some hubs, nonlinear causality from futures to the spot market remains. This suggests that the predictive power of spot return distributions for subsequent distributions of futures market returns is mainly due to conditional volatility effects and thus not a result of informational superiority. Overall, the performed causality analysis suggests that price formation at European gas hubs generally takes place on the more informationally efficient futures markets with the less informationally efficient spot markets adjusting accordingly.

13. Similar values of C^* have been used for comparable empirical approaches (e.g., Bekiros and Diks (2008) set C^* equal to 7.5).

14. The results of the nonlinear causality tests for the other pairs of return series are presented in the Appendix.

15. BEKK refers to the first letters of the names of Baba, Engle, Kroner and Kraft, who jointly developed the model.

Table 5: Pairwise Nonlinear Causality Tests for Gas Price Returns

	Direction	*t-Statistic*
VECM-filtered Data	NCG Spot on NCG m + 1	4.219***
	NCG m + 1 on NCG Spot	5.520***
	GP Spot on GP m + 1	5.740***
	GP m + 1 on GP Spot	5.593***
	TTF Spot on TTF m + 1	3.965***
	TTF m + 1 on TTF Spot	7.703***
	NBP Spot on NBP m + 1	3.305***
	NBP m + 1 on NBP Spot	3.222***
	PEGN Spot on PEGN m + 1	4.960***
	PEGN m + 1 on PEGN Spot	4.882***
	CEGH Spot on CEGH m + 1	0.163
	CEGH m + 1 on CEGH Spot	0.897
BEKK GARCH-filtered Data	NCG Spot on NCG m + 1	−1.944
	NCG m + 1 on NCG Spot	−0.477
	GP Spot on GP m + 1	1.740**
	GP m + 1 on GP Spot	2.426***
	TTF Spot on TTF m + 1	−0.711
	TTF m + 1 on TTF Spot	5.698***
	NBP Spot on NBP m + 1	1.016
	NBP m + 1 on NBP Spot	0.939
	PEGN Spot on PEGN m + 1	1.404
	PEGN m + 1 on PEGN Spot	2.646***
	CEGH Spot on CEGH m + 1	0.228
	CEGH m + 1 on CEGH Spot	0.245

Notes: *** (**) Denotes significance at the 99 (95) %-level.

VI. LONG- AND SHORT-RUN DYNAMICS BETWEEN SPOT AND FUTURES MARKETS: THE EFFICIENCY OF INTERTEMPORAL ARBITRAGE

a. Econometric Methodology and Economic Interpretation

The theory of storage calls for a stable long-run equilibrium between the spot and the futures market for the same underlying asset. The finding of cointegration relationships for the spot and futures market price series at all hubs in Section 3 thus confirms that the theory of storage holds in the long run. From an economic perspective, this means that there is significant arbitrage activity at these hubs in order to prevent that deviations from the long-run equilibrium are infinitely persistent. The intertemporal long-run relationship between spot and futures market can be written as

$$S_t = c + \beta_t F_t + \epsilon_t \tag{11}$$

Here, S_t and F_t are the spot and the futures prices, respectively. The coefficient β represents the degree of price convergence in the long run and ϵ_t captures the deviations from the long run

16. With regard to the cost-of-carry relationship, the intercept in Equation (11) contains the time-invariant spread between futures and spot prices that can be assigned to the convenience yield, storage costs and the interest rate. Assuming time-invariant carrying parameters, ϵ_t represents the deviation from the cost-of-carry relationship, triggering arbitrage trading

relationship.[16] As outlined by Arouri et al. (2013), long-run informational efficiency implies cointegration between both price series and full price convergence of spot and futures markets, i.e., $\beta = 1$.

However, even if the market is informationally efficient in the long run, short-run inefficiencies may exist (Arouri et al., 2013). Such short-run inefficiencies are characterized by transitory deviations from the cost-of-carry conditions that are not immediately exploited by arbitrage activity. In order to assess the short-run informational efficiency of spot and futures markets, their short-term behavior can be modelled by the following VECM:

$$\begin{aligned} \Delta f_t &= \alpha_t^f \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^f \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^f \Delta s_{t-k} + \eta_t^f \\ \Delta s_t &= \alpha_t^s \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^s \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^s \Delta s_{t-k} + \eta_t^s \end{aligned} \tag{12}$$

where α is the adjustment coefficient representing the error correction of the series in case of any deviation from the long-run equilibrium (Lütkepohl, 2005) and k denotes the lag length. To assess the efficiency of arbitrage, the α coefficients are of central interest because they measure the speed of error correction, i.e., the process of arbitrage activity eliminating the deviations from the intertemporal equilibrium. The greater the value of the adjustment coefficient in absolute terms, the more informationally efficient are the market participants in exhausting arbitrage opportunities. The α coefficients hence represent a measure of intertemporal arbitrage efficiency. The γ and δ coefficients account for autoregressive behaviour of the series and thus give an indication about whether lagged changes in the variables are significant for current changes of the variables and are not of interest for the assessment of intertemporal arbitrage efficiency .

The specified VECM assumes linearity in the error correction process. This implies that arbitrage activity starts instantaneously in case of any, arbitrarily small, deviation from the long-run equilibrium, thus neglecting any kind of market frictions. However, the exhaustion of arbitrage opportunities at European gas hubs may be constrained by significant transaction costs resulting from low liquidity and by physical constraints, e.g., limited injection and withdrawal capacity of storage facilities. Thus, intertemporal arbitrage may only be triggered if the deviation from the cost-of-carry equilibrium exceeds a certain threshold, such that the arbitrage traders are compensated for the incurred transaction or information costs (Li, 2010), resulting in a so-called "band of no arbitrage" around the long-run equilibrium.

In order to investigate whether intertemporal arbitrage is constrained by a "band of no arbitrage" due to market frictions, the TVECM approach proposed by Granger and Lee (1989) can be applied.[17] The economic intuition of a TVECM is that arbitrage activity may depend on the magnitude of the deviation from the equilibrium. Thus, the model allows arbitrage efficiency to vary between different regimes. The bivariate TVECM of order k applied to the bivariate system of spot and futures market returns used in this study has the representation

$$\begin{aligned} \Delta f_t &= (I-1)\,\alpha_h^f \epsilon_{t-1} + I\alpha_l^f \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^f \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^f \Delta s_{t-k} + \eta_t^f \\ \Delta s_t &= (I-1)\,\alpha_h^s \epsilon_{t-1} + I\alpha_l^s \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^s \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^s \Delta s_{t-k} + \eta_t^s \end{aligned} \tag{13}$$

between spot and futures markets. One should keep in mind that in case of time-varying carrying parameters (e.g. fluctuations of storage costs), ϵ_t may not completely reflect deviations from the equilibrium condition.

17. TVECMs have proved to be a useful approach for capturing arbitrage dynamics among spot and futures markets for financial assets (e.g., Anderson, 1997) and various commodities (Li, 2010; Huang et al., 2009; Root and Lien, 2003) by explicitly accounting for market frictions.

where I denotes the regime indicator stating whether the lagged deviation from the long-run equilibrium is below or above a certain threshold (in absolute terms). The adjustment coefficient α_h (α_l) represents the error correction dynamic for the case in which the absolute value of the deviation is higher (lower) than the threshold (Enders and Siklos, 2001). As a consequence, the coefficients α_h and α_l represent measures of regime-specific intertemporal arbitrage efficiency. Thus, comparing the significance and magnitude of the regime-specific adjustment coefficients allows for an assessment of the arbitrage efficiency in the different regimes and therefore for the relevance of market frictions. For instance, if α_l is insignificant in the model while α_h is significant, this points towards a “band of no arbitrage”, i.e., arbitrage is not carried out due to market frictions unless the deviation from the intertemporal equilibrium crosses a certain threshold. In contrast, if α_l and α_h are both significant, this suggests that there is no empirical evidence of a “band of no arbitrage” as arbitrage dynamics are similar in both regimes.

b. Empirical Results

In order to investigate the efficiency of intertemporal arbitrage, linear VECMs as specified in Equation (12) are estimated for all hubs. Table 6 presents the estimated cointegration vector and the adjustment coefficients. The estimated β coefficients are statistically significant and close to unity. For all hubs, the hypothesis of full price convergence of the spot and the futures market, i.e., $\beta = 1$, cannot be rejected using likelihood ratio tests.[18] Thus, the hubs analyzed can be considered as informationally efficient in the long run.

Next, the potential short-run inefficiencies are investigated. The adjustment coefficient is statistically significant in all spot price return equations. Hence, deviations from the long-run relationship are corrected within the spot market at all hubs. In contrast, except for Gaspool where the futures price adjusts slightly, the futures price return series do not react to deviations from the equilibrium. This finding is in line with Huang et al. (2009), who obtain similar results for crude oil spot and futures markets in the period 1991 to 2001.[19] The small absolute values of the adjustment coefficients imply a sticky error correction process and thus suggest a rather low efficiency of intertemporal arbitrage.[20] Although this means that none of the considered hubs can be regarded as fully informationally efficient in the short run, arbitrage seems to be most efficiently exploited at NBP and CEGH. However, the empirical results of CEGH should be interpreted with caution due to the small sample size of the respective price series. In turn, the finding of high arbitrage efficiency at NBP is noteworthy as physical storage flexibility is much smaller in the UK than in continental Europe (see Table 3) and may be a result of the superior liquidity of the British hub. However, the difference in the speed of adjustment and hence in the degree of arbitrage efficiency between the hubs is fairly moderate. The TVECM of Equation (13) is estimated using different thresholds for all hubs except for CEGH.[21] The thresholds are assumed to be symmetric and their

18. Test statistics are provided in the Appendix.

19. Similar results are obtained from the VECM estimation for the interaction of spot prices and futures prices with longer maturity. The respective test statistics are presented in the Appendix.

20. For instance, the absolute value of the adjustment coefficient of the NCG spot return series implies a half-life period of error correction of about five days.

21. The small sample size of CEGH does not allow for valid statistical inference when splitting the data in various regimes as done in the TVECM.

Table 6: Normalized Cointegration Vectors and Error Correction Coefficients (Spot - m + 1)

	Parameter	*Standard Error*	*t-Statistic*
c_{NCG}	−0.0276	0.0761	−0.3621
β_{NCG}	0.9836	0.0203	38.849***
$\alpha_{NCG,\ spot}$	−0.1329	0.0176	−7.5461***
$\alpha_{NCG,\ m+1}$	0.0107	0.0106	1.0021
c_{GP}	−0.0426	0.0724	−0.5891
β_{GP}	0.9800	0.0241	40.597***
$\alpha_{GP,\ spot}$	−0.1153	0.0173	−6.6566***
$\alpha_{GP,\ m+1}$	0.0294	0.0110	2.6856***
c_{TTF}	−0.0368	0.0816	−0.4509
β_{TTF}	0.9809	0.0272	36.005***
$\alpha_{TTF,spot}$	−0.1111	0.0130	−8.5230***
$\alpha_{TTF,m+1}$	0.0036	0.0105	0.3453
c_{NBP}	−0.0665	0.1520	−0.4375
β_{NBP}	0.9758	0.0394	24.858***
$\alpha_{NBP,spot}$	−0.1538	0.0196	−7.8323***
$\alpha_{NBP,m+1}$	0.0044	0.0088	0.5035
c_{PEGN}	0.0453	0.0819	0.5539
β_{PEGN}	1.0131	0.0281	36.119***
$\alpha_{PEGN,spot}$	−0.1280	0.0182	−7.0180***
$\alpha_{PEGN,,m+1}$	−0.0174	0.0136	−1.2841
c_{CEGH}	−0.1955	0.3797	−0.5150
β_{CEGH}	0.9836	0.1182	−7.9149***
$\alpha_{CEGH,spot}$	−0.1707	0.0322	−5.3038***
$\alpha_{CEGH,,m+1}$	−0.0068	0.0178	−0.3815

Notes: *** Denotes significance at the 99 %-level. A lag length of 1 for the VECM is selected based on the Schwarz Information Criterion for NCG, Gaspool, TTF, PEGN and CEGH, while the same criterion suggests to include 2 lags for NBP. The autoregressive coefficients are not reported to conserve space.

size is defined in terms of the standard deviation of ϵ_t, the error term of the cointegration regression.[22] Table 7 contains the estimates for the regime-specific adjustment coefficients of the TVECM.

In the TTF spot price return equation, the adjustment coefficient is statistically significant in both regimes. Thus, for the threshold values tested, there is no empirical evidence of a "band of no arbitrage". In contrast, arbitrage at NCG, GP, NBP and PEGN does not start until the deviation from the long-run equilibrium exceeds a certain threshold (i.e., α_l is insignificant for at least one specification). Surprisingly, although NBP is the most liquid hub in the sample, it exhibits a rather broad "band of no arbitrage", indicating significant frictions hampering instantaneous arbitrage. To sum up, intertemporal arbitrage starts most instantaneously at TTF but is executed most efficiently

22. The standard deviations of different ϵ_t-series range between 0.08 and 0.11. The thresholds selected for the TVECM estimation are $0.5\sigma_\epsilon$ and σ_ϵ. In general, smaller and greater thresholds can be used to investigate the regime-dependent arbitrage dynamics. However, these threshold choices result in small regimes with large standard errors of the estimated coefficients, hindering valid statistical inference. The same problem occurs when estimating the thresholds endogenously following the procedure of Balke and Fomby (1997).

Table 7: Estimates of Threshold Vector Error Correction Models

		NCG		*GP*		*TTF*		*NBP*		*PEGN*	
Threshold	Regime	α_{spot}	α_{m1}	α_{spot}	α_{m1}	α_{spot}	α_{m1}	α_{spot}	α_{m1}	α_{spot}	α_{m1}
$0.5\sigma_\epsilon$	high	−0.1359***	0.0238	−0.1175***	0.0292**	−0.1101***	0.0012	−0.1603***	0.0042	−0.1278***	−0.0539
$0.5\sigma_\epsilon$	low	−0.0834	0.0094	−0.0870	0.0242	−0.1312***	0.0421	−0.0465	0.0078	−0.1317	−0.0156
σ_ϵ	high	−0.1358***	0.0148	−0.1224***	0.0288**	−0.1092***	0.0080	−0.1835***	0.0033	−0.1325***	−0.0520
σ_ϵ	low	−0.1203***	−0.0144	−0.0829**	0.0297	−0.1212***	−0.0243	−0.0196	0.0094	−0.0960	−0.0125

Notes: *** (*) Denotes significance at the 99 (95) %-level. The estimation is based on OLS using robust standard errors as proposed by Newey and West (1987). A lag length of 1 for the VECM is selected based on the Schwarz Information Criterion for NCG, Gaspool, TTF and PEGN, while the same criterion suggests including 2 lags for NBP. The autoregressive coefficients are not reported to conserve space.

at NBP once the deviation from the intertemporal equilibrium crosses a certain threshold. The first finding is in line with the high flexibility of Dutch gas storage (see Table 3), while the latter may be attributed to the superior liquidity of NBP (see Table 2).

VII. THE EVOLUTION OF ARBITRAGE EFFICIENCY: A KALMAN FILTER APPROACH

a. Econometric Methodology and Economic Interpretation

Various political and regulatory measures have been introduced to foster the liquidity of the continental European gas hubs.[23] As a consequence, one may expect informational efficiency at these hubs to have increased over time. To test this hypothesis, a dynamic state-space approach can be applied to capture the evolution of intertemporal arbitrage efficiency over time. Such time-varying coefficient models allow for an assessment of the evolution of the economic relationship investigated over time. Time-varying coefficient models have been used for the European gas market in different applications (see Neumann et al., 2006 and Growitsch et al., 2012). However, this paper is the first to apply the state-space methodology within an intertemporal context for the European gas markets. In doing so, the intertemporal arbitrage dynamic is investigated by estimating Equation (14):

$$\begin{aligned} \Delta f_t &= \alpha_t^f \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^f \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^f \Delta s_{t-k} + \eta_t^f \\ \Delta s_t &= \alpha_t^s \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^s \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^s \Delta s_{t-k} + \eta_t^s \end{aligned} \tag{14}$$

with

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{15}$$

where α_t represents the time-varying adjustment coefficient following a random walk and ϵ_{t-1} is the lagged error term of the linear cointegration regression. Thus, α_t represents a time-varying measure of intertemporal arbitrage efficiency at the respective hub. Based on the hypothesis of increasing informational efficiency over time at the continental European hubs due to the rise in liquidity, the absolute values of the respective α_t coefficients are expected to increase over time. In contrast, a decrease in the absolute value of α_t would imply a decrease in intertemporal arbitrage efficiency.

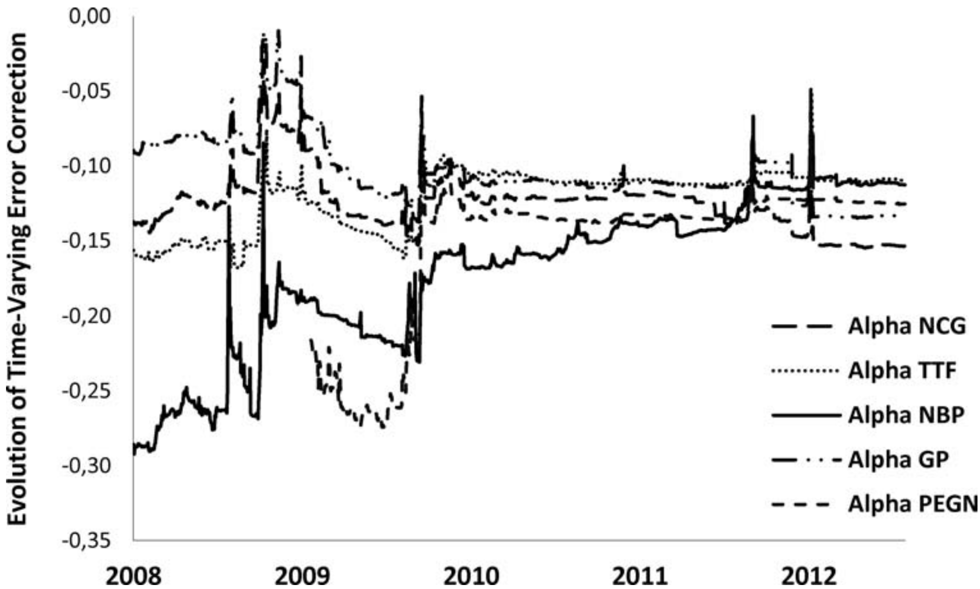
b. Empirical Results

The state-space model of Equation (14) is estimated using the recursive procedure suggested by Kalman (1960).[24] Figure 2 presents the estimated time paths for the adjustment coefficients

23. Most notably, the Third Gas Market Directive of the European Union from 2009 comprises various efforts to improve access to gas infrastructure and thus facilitates the development of liquid natural gas hubs (EU, 2009).

24. As initial value of α, zero is selected assuming informational inefficiency at the beginning of the sample period. The variance of the respective spot return series, σ^2_{rspot}, is selected as initial variance of η_t and ζ_t is set to $\sigma^2_{rspot}/1000$. In line with the linear VECM specified above, one lag is included for NCG, GP, TTF, and PEGN while two lags are used in the specification for NBP. The small sample size of CEGH does not allow for a valid estimation of the state-space model for this hub.

Figure 2: Time-Varying Adjustment Coefficients of Spot Price Return Series



in the spot return equation.[25] Some of the spikes in the plotted series can be attributed to the economic downturn in autumn 2008, and gas market-specific shocks such as the extraordinary supply interruptions resulting from the Russian-Ukrainian crisis in January 2009 and the cold spell in February 2012.[26] There is a distinctive pattern in the evolution of the relative informational efficiency of the hubs considered over time, as can be inferred from the time-varying coefficient estimates: As of the beginning of 2008, the two German hubs NCG and GP are the least informationally efficient hubs. However, the absolute value of the adjustment coefficients grows towards the end of the sample period, indicating an increase in informational efficiency. In contrast, the absolute value of the adjustment coefficient of NBP decreases over time, indicating a decline in the efficiency of intertemporal arbitrage. Similarly, intertemporal arbitrage efficiency has decreased at PEGN despite the growth in liquidity. For the Dutch TTF hub, informational efficiency is stable at a rather low level in the second half of the sample period. Overall, there is convergence in the degree of informational efficiency of the hubs considered and only the informational efficiency of the two German hubs seems to have benefited from the increase in liquidity. Thus, as of 2012, the differences in intertemporal arbitrage efficiency of the hubs considered appear significantly reduced.

VIII. CONCLUSION

The objective of the paper was to analyze the informational efficiency of different European gas hubs by empirically investigating price discovery and arbitrage activity between spot and futures

25. The evolution of the adjustment coefficient in the futures return equation is neglected due to statistical insignificance.

26. In the latter two periods, it seems reasonable to infer that the strong increase in spot price represents an immediate reaction to the physical supply and demand imbalance, independent from the futures market price. For a more detailed discussion of the economic impact of these events on German gas prices, see Nick and Thoenes (2013).

markets. For this purpose, linear and nonlinear econometric approaches were specified to explicitly account for the low-liquidity environment and the physical characteristics of the gas market.

Causality testing reveals that price formation generally takes place on the futures market. This finding is in line with the hypothesis that futures market participants react more efficiently to information than traders at spot markets (Silvapulle and Moosa, 1999; Bohl et al., 2011). It seems intuitive to attribute this finding to the broader scope of market participants on the futures market: Although the futures contracts considered result in physical delivery, the opportunity to trade the contract multiple times before maturity and thus to close out the trading position without taking physical delivery enables their use for hedging and speculation. Thus, in contrast to the purely physical spot market, the futures market is easily accessible for traders without interest in physical delivery. Apparently, this structural difference between both markets yields the futures market to be significantly informational superior compared to the spot market. In the light of hub-based pricing of internationally traded gas, an indexation on futures market prices rather than on spot market prices promises to provide more valid price signals.

The theory of storage seems to hold for all gas hubs considered in the long run, indicating the existence of arbitrage between the respective spot and futures markets. However, the error correction process is rather sticky and subject to significant frictions. From a dynamic perspective, the state-space estimations reveal a convergence in informational efficiency across the hubs during the sample period. With regard to the liquidity of the hubs investigated, the empirical results provide mixed evidence: On the one hand, intertemporal arbitrage opportunities are rather efficiently exploited at the liquid NBP and the rise in liquidity seems to have fostered informational efficiency at NCG and GP. On the other hand, the detected frictions in the price formation process and arbitrage activities are similar for all hubs, regardless of their liquidity. Therefore, it seems reasonable to attribute these frictions at least partly to physical characteristics of the market, e.g., limited storage flexibility or inefficient allocation of storage capacity, rather than exclusively to market liquidity.

A promising field for further research could be the direct empirical analysis of potential determinants of informational inefficiency of the hubs analyzed such as liquidity, storage utilization or network capacity. This approach, however, is currently aggravated by the lack of comprehensive data sets in adequate frequency and is therefore left for future research.

REFERENCES

Anderson, H. M. (1997). "Transaction Costs and Non-Linear Adjustment towards Equilibrium in the US Treasury Bill Market." *Oxford Bulletin of Economics and Statistics* 59(4): 465–484. http://dx.doi.org/10.1111/1468-0084.00078.

Arouri, M. E. H., S. Hammoudeh, A. Lahiani, A. and D.K. Nguyen (2013). "On the Short- and Long-Run Efficiency of Energy and Precious Metal Markets."*Energy Economics* 40: 832–844. http://dx.doi.org/10.1016/j.eneco.2013.10.004.

Asche, F., B. Misund and M. Sikveland (2013). "The Relationship between Spot and Contract Gas Prices in Europe." *Energy Economics* 38: 212–217. http://dx.doi.org/10.1016/j.eneco.2013.02.010.

Balke, N.S. and T.B. Fomby (1997): "Threshold Cointegration." *International Economic Review* 38(3): 627–645. http://dx.doi.org/10.2307/2527284.

Bekiros, S. (2011). "Causality Testing with Stepwise Multivariate Filtering." European University Institute, EUI ECO Working Paper No. 2011-22.

Bekiros, S. D. and C.G. Diks (2008). "The Relationship between Crude Oil Spot and Futures Prices: Cointegration, Linear and Nonlinear Causality." *Energy Economics* 30(5): 2673–2685. http://dx.doi.org/10.1016/j.eneco.2008.03.006.

Bohl, M. T., C.A. Salm and M. Schuppli (2012). "Price Discovery and Investor Structure in Stock Index Futures." *The Journal of Futures Markets* 31(2): 282–306.

Chen, A. and J. Lin (2004). " Cointegration and Detectable Linear and Nonlinear Causality: Analysis Using the London Metal Exchange Lead Contract." *Applied Economics* 36(11): 1157–1167. http://dx.doi.org/10.1080/000368404 2000247352.

Chordia, T., R. Roll and A. Subrahmanyam, (2008). “Liquidity and Market Efficiency.” *Journal of Financial Economics* 87(2): 249–268. http://dx.doi.org/10.1016/j.jfineco.2007.03.005.

Considine, T. J. and D.F. Larson (2001). “Risk Premium on Inventory Assets: The Case of Crude Oil and Natural Gas.” *The Journal of Futures Markets* 21(2): 109–126. http://dx.doi.org/10.1002/1096-9934(200102)21:2<109::AID-FUT1>3.0.CO;2-A.

Dergiades, T., G. Christofidou and R. Madlener (2012). “The Nexus between Natural Gas Spot and Futures Prices at NYMEX: What about Non-Linear Causality” Institute for Future Energy Costumer Needs and Behavior, Working Paper No. 17/2012.

Diks, C. and V. Panchenko (2005). “A Note on the Hiemstra-Jones Test for Granger Non-Causality.” *Studies in Nonlinear Dynamics & Econometrics* 9(2): 1–9. http://dx.doi.org/10.2202/1558-3708.1234.

Diks, C. and V. Panchenko (2006). “A New Statistic and Practical Guidelines for Nonparametric Granger Causality Testing.” *Journal of Economic Dynamics and Control* 30(9–10): 1647–1669. http://dx.doi.org/10.1016/j.jedc.2005.08.008.

Enders, W. and P.L. Siklos (2001). “Cointegration and Threshold Adjustment.” *Journal of Business and Economic Statistics* 19(2): 166–176. http://dx.doi.org/10.1198/073500101316970395.

Engle, R. and C.W.J. Granger (1987). “Co-Integration and Error Correction: Representation, Estimation and Testing.” *Econometrica* 55(2): 251–276. http://dx.doi.org/10.2307/1913236.

Engle, R. and K. Kroner (1995). “Multivariate Simultaneous Generalized GARCH.” *Econometric Theory* 11(1): 122–150. http://dx.doi.org/10.1017/S0266466600009063.

EU (2003). Directive 2003/55/EC of the European Parliament and of the Council.

EU (2009). Directive 2009/73/EC of the European Parliament and of the Council.

Fama, E.F. (1970). “Efficient capital markets: A review of theory and empirical work.” *The Journal of Finance* 25(2): 383–417. http://dx.doi.org/10.2307/2325486.

Fama, E. F. and K.R. French (1987). “Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage.” *The Journal of Business* 60(1): 55–73. http://dx.doi.org/10.1086/296385.

Gebre-Mariam, Y. (2011). “Testing for Unit Roots, Causality, Cointegration and Efficiency: The Case of the Northwest US Natural Gas Market.” *Energy* 36(5): 3489–3500. http://dx.doi.org/10.1016/j.energy.2011.03.055.

GIE (2011). Gas Storage Europe Database. Gas Infrastructure Europe, Technical Report.

Granger, C.W.J. (1969). “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica* 37(3): 424–438. http://dx.doi.org/10.2307/1912791.

Granger, C.W.J. and T. Lee (1989). “Investigation of Production, Sales and Inventory Relationships Using Multicointegration and Nonsymmetric Error-Correction Models.” *Journal of Applied Econometrics* 4(Special Issue): 146–159.

Growitsch, C., M. Stronzik, and R. Nepal (2012). “Price Convergence and Information Efficiency in German Natural Gas Markets.” Institute of Energy Economics at the University of Cologne, EWI Working Paper No. 2012-5.

Heather, P. (2012). “Continental European Gas Hubs: Are They Fit for Purpose?” The Oxford Institute for Energy Studies, OIES Working Paper NG No. 63.

Hiemstra, C. and J. Jones (1994). “Testing for Linear and Nonlinear Granger Causality in the Stock Price Volume Relation.” *Journal of Finance* 49(5): 1639–1664.

Huang, B., C. Yang and M. Hwang (2009). “The Dynamics of a Nonlinear Relationship between Crude Oil Spot and Futures Prices: A Multivariate Threshold Regression Approach.” *Energy Economics* 31(1): 91–98. http://dx.doi.org/10.1016/j.eneco.2008.08.002.

ICIS (2013). Heren European Gas Markets. ICIS, Newsletter 15 March 2013.

IEA (2012a). Medium Term Gas Market Report. International Energy Agency.

IEA (2012b). Natural gas information 2012. International Energy Agency.

Johansen, S. (1988). “Statistical Analysis of Cointegration Vectors.” *Journal of Economics Dynamics and Control* 12(2–3): 231–254. http://dx.doi.org/10.1016/0165-1889(88)90041-3.

Kalman, R. (1960). “A New Approach to Linear Filtering and Prediction Problems.” *Journal of Basic Engineering* 82(1): 35–45. http://dx.doi.org/10.1115/1.3662552.

Keyaerts, N. and W. D’Haeseleer (2012). “Increasing Efficiency through Market-Based Crossborder Procurement of Gas-Balancing Services in Europe.” *Energy* 47(1): 564–576. http://dx.doi.org/10.1016/j.energy.2012.09.049.

Li, M. Y. L. (2010). “Dynamic Hedge Ratio for Stock and Index Futures: Application of Threshold VECM.” *Applied Economics* 42(11): 1403–1417. http://dx.doi.org/10.1080/00036840701721380.

Lütkepohl, H. (2005). “New Introduction to Multiple Time Series Analysis.” Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-27752-1.

Moosa, I. and N. Al-Loughani (1995). “The Effectiveness of Arbitrage and Speculation in the Crude Oil Futures Market.” *The Journal of Futures Markets* 15(2): 167–186. http://dx.doi.org/10.1002/fut.3990150205.

Neumann, A., B. Siliverstovs and C. von Hirschhausen (2006). "Convergence of European Spot Market Prices for Natural Gas? A Real-Time Analysis of Market Integration using the Kalman Filter." *Applied Economics Letters* 13(11): 727–732. http://dx.doi.org/10.1080/13504850500404258.

Newey, W. and K. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3): 703–708. http://dx.doi.org/10.2307/1913610.

Nick, S. and S. Thoenes (2013). "What Drives Natural Gas Prices? – A Structural VAR Approach." Institute of Energy Economics at the University of Cologne, EWI Working Paper No. 2013-2.

NMA (2012). 2012 Liquidity Report: Wholesale Markets for Natural Gas and Electricity. Netherlands Competition Authority - Office of Energy Regulation, Technical Report.

Root, T. H. and D. Lien (2003). "Can Modeling the Natural Gas Futures Market as a Threshold Cointegrated System Improve Hedging and Forecasting Performance?" *International Review of Financial Analysis* 12(2): 117–133. http://dx.doi.org/10.1016/S1057-5219(03)00003-6.

Samuelson, P. A. (1965). "Proof that Properly Anticipated Prices Fluctuate Randomly." *Management Revie*w 6(2): 41–49.

Schultz, E., and J. Swieringa (2013). "Price Discovery in European Natural Gas Markets." *Energy Polic*y 61: 628–634. http://dx.doi.org/10.1016/j.enpol.2013.06.080.

Schwartz, T. and A. Szakmary (1994). "Price Discovery in Petroleum Markets: Arbitrage, Cointegration and the Time Interval of Analysis." *The Journal of Futures Market*s 14(2): 147–167. http://dx.doi.org/10.1002/fut.3990140204.

Silvapulle, P. and I. A. Moosa (1999). "The Relationship between Spot and Futures Prices: Evidence from the Crude Oil Market." *The Journal of Futures Markets* 19(2): 175–193. http://dx.doi.org/10.1002/(SICI)1096-9934(199904)19:2＜175::AID-FUT3＞3.0.CO;2-H.

Stern, J. P. and H. Rogers (2010). "The Transition to Hub-Based Gas Pricing in Continental Europe." The Oxford Institute for Energy Studies, OIES Working Paper NG No. 49.

Stern, J. (2014). "International Gas Pricing in Europe and Asia: A Crisis of Fundamentals." *Energy Policy* 64: 43–48. http://dx.doi.org/10.1016/j.enpol.2013.05.127.

Stronzik, M., M. Rammerstorfer and A. Neumann (2009). "Does the European Natural Gas Market Pass the Competitive Benchmark of the Theory of Storage? Indirect Test for Three Major Trading Points." *Energy Policy* 37(12): 5432–5439. http://dx.doi.org/10.1016/j.enpol.2009.08.003.

Working, H. (1949). "The Theory of the Price of Storage." *American Economic Review* 39(6): 1242–1262.

Zhang, J. and S. Jinghong, S. (2012). "Causality in the VIX Futures Market." *The Journal of Futures Markets* 32(1): 24–46. http://dx.doi.org/10.1002/fut.20506.

APPENDIX: TEST STATISTICS

Table A.8: Results of the Unit Root Tests

	t-Statistic ADF	*p-Value ADF*	*t-Statistic PP*	*p-Value PP*
NCG Spot	−1.5307	0.5178	−1.9745	0.2938
NCG m + 1	−1.3782	0.5943	−1.3513	0.6073
NCG m + 2	−1.8279	0.3671	−1.3083	0.6276
NCG m + 3	−1.1575	0.6945	−1.2410	0.6585
GP Spot	−1.5137	0.5266	−2.1798	0.2140
GP m + 1	−1.3929	0.5871	−1.3508	0.6075
GP m + 2	−1.5173	0.5247	−1.3740	0.5964
GP m + 3	−1.7997	0.3810	−1.2838	0.6391
TTF Spot	−1.5473	0.5093	−2.1754	0.2156
TTF m + 1	−1.3514	0.6072	−1.3401	0.6126
TTF m + 2	−1.4541	0.5567	−1.2593	0.6502
TTF m + 3	−1.1283	0.7065	−1.2117	0.6714
NBP Spot	−1.6091	0.4776	−3.2456	0.0177
NBP m + 1	−1.4889	0.5391	−1.6794	0.4415
NBP m + 2	−1.5543	0.5057	−1.6491	0.4570
NBP m + 3	−1.4122	0.5776	−1.4726	0.5474
PEGN Spot	−1.5370	0.5145	−1.7867	0.3874
PEGN m + 1	−1.1551	0.6954	−1.2264	0.6648
PEGN m + 2	−1.2614	0.6491	−1.2321	0.6623
PEGN m + 3	−0.9294	0.7790	−1.1224	0.7088
CEGH Spot	−3.8982	0.0022	−4.2445	0.0006
CEGH m + 1	−2.2627	0.1848	−2.5310	0.1088
CEGH m + 2	−2.1710	0.2174	−2.3038	0.1713
CEGH m + 3	2.1330	0.2319	−2.1981	0.2078
ΔNCG Spot	−13.2306	0.0000	−40.8718	0.0000
ΔNCG m + 1	−12.7497	0.0000	−32.7785	0.0000
ΔNCG m + 2	−6.3319	0.0000	−33.9596	0.0000
ΔNCG m + 3	−5.0573	0.0000	−33.8859	0.0000
ΔGP Spot	−14.4425	0.0000	−37.4064	0.0000
ΔGP m + 1	−20.2233	0.0000	−32.8511	0.0000
ΔGP m + 2	−10.1769	0.0000	−34.3888	0.0000
ΔGP m + 3	−5.3505	0.0000	−33.7829	0.0000
ΔTTF Spot	−13.1479	0.0000	−34.7274	0.0000
ΔTTF m + 1	−10.8880	0.0000	−34.3284	0.0000
ΔTTF m + 2	−9.9450	0.0000	−33.2840	0.0000
ΔTTF m + 3	−5.2044	0.0000	−32.7979	0.0000
ΔNBP Spot	−10.2739	0.0000	−62.3198	0.0001
ΔNBP m + 1	−20.7571	0.0000	−35.1039	0.0000
ΔNBP m + 2	−22.2504	0.0000	−34.8534	0.0000
ΔNBP m + 3	−21.9632	0.0000	−34.0489	0.0000
ΔPEGN Spot	−11.5735	0.0000	−26.2360	0.0000
ΔPEGN m + 1	−10.6398	0.0000	−29.6818	0.0000
ΔPEGN m + 2	−9.3694	0.0000	−28.8184	0.0000
ΔPEGN m + 3	−9.3722	0.0000	−29.8813	0.0000
ΔCEGH Spot	−8.5586	0.0000	−23.5471	0.0000
ΔCEGH m + 1	−22.9783	0.0000	−22.9737	0.0000
ΔCEGH m + 2	−25.0619	0.0000	−25.3103	0.0000
ΔCEGH m + 3	−17.5181	0.0000	28.6268	0.0000

Notes: The unit root tests are specified with a constant but without a linear trend, as a time trend seemed inappropriate from the first investigation of the price series. The optimization of the lag length included for the ADF test equation was conducted with respect to the Akaike Information Criterion. The selection of the bandwidth for the Phillips-Perron test was based on the Newey-West procedure using a Bartlett kernel.

Table A.9: Results of the Johansen Cointegration Test (Spot and m + 1-Contract)

	Eigenvalue	*Trace Statistic*	*Critical Value (95 %)*	*p-Value*
NCG r = 0	0.0611	78.694	20.262	0.0000
NCG r≤1	0.0013	1.5589	9.1645	0.8627
GP r = 0	0.0652	84.952	20.262	0.0000
GP r≤1	0.0019	2.2744	9.1645	0.7227
TTF r = 0	0.0548	70.511	20.262	0.0000
TTF r≤1	0.0012	1.5125	9.1645	0.8712
NBP r = 0	0.0450	60.508	20.262	0.0000
NBP r≤1	0.0019	2.3415	9.1645	0.7092
PEGN r = 0	0.0609	58.1038	20.262	0.0000
PEGN r≤1	0.0018	1.6588	9.1645	0.8442
CEGH r = 0	0.0965	48.333	20.262	0.0000
CEGH r≤1	0.0127	5.3881	9.1645	0.2436

Table A.10: Results of the Johansen Cointegration Test (Spot- and m + 2)

Hypothesis	*Eigenvalue*	*Trace Statistic*	*Critical Value (95 %)*	*p-Value*
NCG r = 0	0.0256	33.602	20.262	0.0004
NCG r≤1	0.0013	1.5655	91.645	0.8615
GP r = 0	0.0280	36.819	20.262	0.0001
GP r≤1	0.0017	2.0268	9.1645	0.7724
TTF r = 0	0.0227	29.652	20.262	0.0019
TTF r≤1	0.0012	1.5023	91.645	0.8730
NBP r = 0	0.0226	31.429	20.262	0.0010
NBP r≤1	0.0020	2.5516	9.1645	0.6673
PEGN r = 0	0.0276	27.0748	20.262	0.0049
PEGN r≤1	0.0022	1.9537	9.1645	0.7869
CEGH r = 0	0.0666	34.901	20.262	0.0002
CEGH r≤1	0.0135	5.7501	9.1645	0.2109

Table A.11: Results of the Johansen Cointegration Test (Spot- and m + 3)

Hypothesis	*Eigenvalue*	*Trace Statistic*	*Critical Value (95 %)*	*p-Value*
NCG r = 0	0.0167	22.0623	20.262	0.0280
NCG r≤1	0.0013	1.5672	9.1645	0.8612
GP r = 0	0.0166	22.327	20.262	0.0256
GP r≤1	0.0015	1.8390	9.1645	0.8096
TTF r = 0	0.0149	19.8639	20.262	0.0566
TTF r≤1	0.0012	1.5087	9.1645	0.8718
NBP r = 0	0.0226	31.4289	20.262	0.0010
NBP r≤1	0.0012	1.5087	9.1645	0.8718
PEGN r = 0	0.0164	16.5558	20.262	0.1500
PEGN r≤1	0.0018	1.6534	9.1645	0.8452
CEGH r = 0	0.0515	27.547	20.262	0.0041
CEGH r≤1	0.0122	5.1813	9.1645	0.2641

Table A.12: Pairwise Linear Causality Tests for NCG Returns

	Direction	*Chi-sq-Statistic*
Raw Data	NCG Spot on NCG m + 2	0.7768
	NCG m + 2 on NCG Spot	155.8917***
	NCG Spot on NCG m + 3	3.4629
	NCG m + 3 on NCG Spot	97.0596***
	NCG m + 1 on NCG m + 2	7.6506**
	NCG m + 2 on NCG m + 1	16.6513***
	NCG m + 1 on NCG m + 3	5.9699
	NCG m + 3 on NCG m + 1	21.4936***
	NCG m + 2 on NCG m + 3	8.3798**
	NCG m + 3 on NCG m + 2	36.4586***
VECM-filtered Data	NCG Spot on NCG m + 1	0.0001
	NCG m + 1 on NCG Spot	0.0115
	NCG Spot on NCG m + 2	0.0010
	NCG m + 2 on NCG Spot	0.0273
	NCG Spot on NCG m + 3	0.0111
	NCG m + 3 on NCG Spot	0.0234
	NCG m + 1 on NCG m + 2	0.0086
	NCG m + 2 on NCG m + 1	0.0000
	NCG m + 1 on NCG m + 3	0.0308
	NCG m + 3 on NCG m + 1	0.0002
	NCG m + 2 on NCG m + 3	0.0148
	NCG m + 3 on NCG m + 2	0.0040

Notes: *** Denotes significance at the 99 %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.13: Pairwise Linear Causality Tests for GP Returns

	Direction	*Chi-sq-Statistic*
Raw Data	GP Spot on GP m + 2	0.3404
	GP m + 2 on GP Spot	140.0471***
	GP Spot on GP m + 3	0.4956
	GP m + 3 on GP Spot	92.3368***
	GP m + 1 on GP m + 2	10.1855***
	GP m + 2 on GP m + 1	14.4801***
	GP m + 1 on GP m + 3	3.9084
	GP m + 3 on GP m + 1	28.8782***
	GP m + 2 on GP m + 3	7.5078**
	GP m + 3 on GP m + 2	51.2555***
VECM-filtered Data	GP Spot on GP m + 1	0.0122
	GP m + 1 on GP Spot	0.0008
	GP Spot on GP m + 2	0.0159
	GP m + 2 on GP Spot	3.66E-06
	GP Spot on GP m + 3	0.0066
	GP m + 3 on GP Spot	0.0004
	GP m + 1 on GP m + 2	0.0026
	GP m + 2 on GP m + 1	0.0018
	GP m + 1 on GP m + 3	0.0398
	GP m + 3 on GP m + 1	0.0072
	GP m + 2 on GP m + 3	0.0306
	GP m + 3 on GP m + 2	0.0177

Notes: *** (**) Denotes significance at the 99 (95) %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.14: Pairwise Linear Causality Tests for TTF Returns

	Direction	*Chi-sq-Statistic*
Raw Data	TTF Spot on TTF m + 2	9.5571***
	TTF m + 2 on TTF Spot	2306.470***
	TTF Spot on TTF m + 3	10.2639***
	TTF m + 3 on TTF Spot	1922.157***
	TTF m + 1 on TTF m + 2	2.8634
	TTF m + 2 on TTF m + 1	17.2112***
	TTF m + 1 on TTF m + 3	2.0024
	TTF m + 3 on TTF m + 1	24.6698***
	TTF m + 2 on TTF m + 3	6.3951**
	TTF m + 3 on TTF m + 2	45.1625***
VECM-filtered Data	TTF Spot on TTF m + 1	0.0294
	TTF m + 1 on TTF Spot	0.0381
	TTF Spot on TTF m + 2	0.0859
	TTF m + 2 on TTF Spot	0.0067
	TTF Spot on TTF m + 3	0.1358
	TTF m + 3 on TTF Spot	0.0116
	TTF m + 1 on TTF m + 2	0.0025
	TTF m + 2 on TTF m + 1	0.0002
	TTF m + 1 on TTF m + 3	0.0020
	TTF m + 3 on TTF m + 1	0.0063
	TTF m + 2 on TTF m + 3	0.0233
	TTF m + 3 on TTF m + 2	0.0118

Notes: *** (**) Denotes significance at the 99 (95) %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.15: Pairwise Linear Causality Tests for NBP Returns

	Direction	*Chi-sq-Statistic*
Raw Data	NBP Spot on NBP m + 2	4.3714
	NBP m + 2 on NBP Spot	157.2911***
	NBP Spot on NBP m + 3	5.1231
	NBP m + 3 on NBP Spot	99.2291***
	NBP m + 1 on NBP m + 2	216.2309***
	NBP m + 2 on NBP m + 1	27.9609***
	NBP m + 1 on NBP m + 3	27.6634***
	NBP m + 3 on NBP m + 1	23.0742***
	NBP m + 2 on NBP m + 3	6.4175
	NBP m + 3 on NBP m + 2	21.6083***
VECM-filtered Data	NBP Spot on NBP m + 1	0.0073
	NBP m + 1 on NBP Spot	0.0009
	NBP Spot on NBP m + 2	0.0016
	NBP m + 2 on NBP Spot	0.0218
	NBP Spot on NBP m + 3	0.0357
	NBP m + 3 on NBP Spot	0.0288
	NBP m + 1 on NBP m + 2	0.0031
	NBP m + 2 on NBP m + 1	0.0133
	NBP m + 1 on NBP m + 3	0.0115
	NBP m + 3 on NBP m + 1	0.0000
	NBP m + 2 on NBP m + 3	0.0143
	NBP m + 3 on NBP m + 2	0.0063

Notes: *** Denotes significance at the 99 %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.16: Pairwise Linear Causality Tests for PEGN Returns

	Direction	*Chi-sq-Statistic*
Raw Data	PEGN Spot on PEGN m + 2	2.0969
	PEGN m + 2 on PEGN Spot	81.7434***
	PEGN Spot on PEGN m + 3	0.4702
	PEGN m + 3 on PEGN Spot	63.4343***
	PEGN m + 1 on PEGN m + 2	6.8971**
	PEGN m + 2 on PEGN m + 1	20.3777***
	PEGN m + 1 on PEGN m + 3	5.3423
	PEGN m + 3 on PEGN m + 1	21.7213***
	PEGN m + 2 on PEGN m + 3	10.4598***
	PEGN m + 3 on PEGN m + 2	19.8635***
VECM-filtered Data	PEGN Spot on PEGN m + 1	0.1062
	PEGN m + 1 on PEGN Spot	0.1011
	PEGN Spot on PEGN m + 2	0.4615
	PEGN m + 2 on PEGN Spot	0.0428
	PEGN Spot on PEGN m + 3	0.6684
	PEGN m + 3 on PEGN Spot	0.0880
	PEGN m + 1 on PEGN m + 2	0.4023
	PEGN m + 2 on PEGN m + 1	0.0108
	PEGN m + 1 on PEGN m + 3	0.1856
	PEGN m + 3 on PEGN m + 1	0.0307
	PEGN m + 2 on PEGN m + 3	0.7385
	PEGN m + 3 on PEGN m + 2	0.9814

Notes: *** Denotes significance at the 99 %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.17: Pairwise Linear Causality Tests for CEGH Returns

	Direction	*Chi-sq-Statistic*
Raw Data	CEGH Spot on CEGH m + 2	1.7325
	CEGH m + 2 on CEGH Spot	131.7766 ***
	CEGH Spot on CEGH m + 3	6.7756**
	CEGH m + 3 on CEGH Spot	109.2315***
	CEGH m + 1 on CEGH m + 2	5.0231
	CEGH m + 2 on CEGH m + 1	9.3958***
	CEGH m + 1 on CEGH m + 3	9.6474***
	CEGH m + 3 on CEGH m + 1	7.4904**
	CEGH m + 2 on CEGH m + 3	4.7641
	CEGH m + 3 on CEGH m + 2	12.6124***
VECM-filtered Data	CEGH Spot on CEGH m + 1	0.0036
	CEGH m + 1 on CEGH Spot	0.0013
	CEGH Spot on CEGH m + 2	0.0101
	CEGH m + 2 on CEGH Spot	0.0037
	CEGH Spot on CEGH m + 3	0.0170
	CEGH m + 3 on CEGH Spot	0.3131
	CEGH m + 1 on CEGH m + 2	0.6097
	CEGH m + 2 on CEGH m + 1	0.1904
	CEGH m + 1 on CEGH m + 3	0.3149
	CEGH m + 3 on CEGH m + 1	0.0297
	CEGH m + 2 on CEGH m + 3	0.0715
	CEGH m + 3 on CEGH m + 2	0.0405

Notes: *** Denotes significance at the 99 %-level. For the raw return series, Granger causality was investigated within the VECM framework, explicitly taking into account the cointegration relationship. For the VECM-filtered residuals, causality testing is based on a VAR-model of the residuals, where the number of lags is optimized with respect to the Schwarz information criterion, suggesting the inclusion of one lag.

Table A.18: Results of the Likelihood Ratio Test on the Cointegration Vector

	Chi-sq-Statistic	*p-Value*
NCG	0.4036	0.5252
GP	0.6631	0.4155
TTF	0.4726	0.4918
NBP	0.3605	0.5482
PEGN	0.2110	0.6460
CEGH	0.2459	0.6200

Notes: The test was applied to the cointegration vector of the spot and the m + 1 futures prices. The null hypothesis of the LR test is: $\beta = [1; -1]$.

Table A.19: Normalized Cointegration Vectors and Error Correction Coefficients (Spot- m + 2)

	Parameter	*Standard Error*	*t-Statistic*
c_{NCG}	−0.0658	−0.0658	−0.3478
β_{NCG}	0.9605	0.0621	15.4605***
$\alpha_{NCG,spot}$	−0.0630	0.0114	−5.52501***
$\alpha_{NCG,m+2}$	−0.0052	0.0066	−0.7735
c_{GP}	−0.0813	0.1749	−0.4647
β_{GP}	0.9567	0.0577	16.5833***
$\alpha_{GP,\quad spot}$	−0.0594	0.0114	−5.2151***
$\alpha_{GP,m+2}$	0.0032	0.0072	0.4490
c_{TTF}	−0.0760	0.1925	−0.3949
β_{TTF}	0.9571	0.0635	15.0659***
$\alpha_{TTF,spot}$	−0.0486	0.0087	−5.6054***
$\alpha_{TTF,m+2}$	−0.0060	0.0063	−0.9532
c_{NBP}	−0.2412	0.3205	−0.7526
β_{NBP}	0.9214	0.0819	11.2517***
$\alpha_{NBP,spot}$	−0.0807	0.0137	−5.8978***
$\alpha_{NBP,m+2}$	−0.0059	0.0056	−10.503
c_{PEGN}	0.1766	0.2117	0.8342
β_{PEGN}	−1.0480	0.0719	14.5723***
$\alpha_{PEGN,spot}$	−0.0557	0.0116	−4.8144***
$\alpha_{PEGN,m+2}$	−0.0091	0.0083	−1.1011
c_{CEGH}	−0.5329	0.5117	−1.0417
β_{CEGH}	0.8254	0.1583	−5.2138***
$\alpha_{CEGH,spot}$	−0.1342	0.0269	−4.9964***
$\alpha_{CEGH,,m+1}$	−0.0079	0.0153	−0.5182

Notes: *** Denotes significance at the 99 %-level. A lag length of 1 for the both VECMs is selected based on the Schwarz Information Criterion for NCG, GP, TTF, PEGN and CEGH, while the same criterion suggests to include 2 lags for NBP.

Table A.20: Normalized Cointegration Vectors and Error Correction Coefficients (Spot- m + 3)

	Parameter	*Standard Error*	*t-Statistic*
c_{NCG}	−0.1865	0.3334	−0.5595
β_{NCG}	0.9134	0.1086	8.4090***
$\alpha_{NCG,spot}$	−0.0377	0.0086	−4.4097***
$\alpha_{NCG,m+3}$	−0.0045	0.0046	−0.9814
c_{GP}	−0.1791	0.3087	−0.5802
β_{GP}	0.9171	0.1010	9.0846***
$\alpha_{GP,spot}$	−0.0366	0.0086	−4.2516***
$\alpha_{GP,m+3}$	−0.0009	0.0050	−0.1850
c_{TTF}	−0.1852	0.3384	−0.5472
β_{TTF}	0.9142	0.1108	8.2498***
$\alpha_{TTF,spot}$	−0.0280	0.0066	−4.2222***
$\alpha_{TTF,m+3}$	−0.0040	0.0044	−0.9098
c_{NBP}	−0.5174	0.4971	−1.0408
β_{NBP}	0.8448	0.1260	6.7045***
$\alpha_{NBP,spot}$	−0.0531	0.0110	−4.8353***
$\alpha_{NBP,m+3}$	−0.0047	0.0041	−11.493
c_{PEGN}	0.1566	0.3968	0.3947
β_{PEGN}	1.0327	0.1338	7.7162***
$\alpha_{PEGN,spot}$	−0.0322	0.0085	−3.7777***
$\alpha_{PEGN,m+3}$	−0.0061	0.0059	−1.0340
c_{CEGH}	−1.3609	0.6810	−1.9985**
β_{CEGH}	0.5672	0.2098	−2.7041***
$\alpha_{CEGH,spot}$	−0.1031	0.0230	−4.4913***
$\alpha_{CEGH,,m+1}$	−0.0201	0.0134	−1.4953

Notes: *** Denotes significance at the 99 %-level. A lag length of 1 for both VECMs is selected based on the Schwarz Information Criterion for NCG, GP, TTF and PEGN, while the same criterion suggests to include 2 lags for NBP.