Price Formation and Intertemporal Arbitrage within a Low-Liquidity Framework: Empirical Evidence from European Natural Gas Markets

AUTHOR

Sebastian Nick

EWI Working Paper, No 13/14

August 2013

Institute of Energy Economics at the University of Cologne (EWI)
www.ewi.uni-koeln.de

Institute of Energy Economics
at the University of Cologne (EWI)

Alte Wagenfabrik
Vogelsanger Straße 321
50827 Köln
Germany

Tel.: +49 (0)221 277 29-100
Fax: +49 (0)221 277 29-400
www.ewi.uni-koeln.de

CORRESPONDING AUTHOR

Sebastian Nick
Institute of Energy Economics at the University of Cologne (EWI)
Tel: +49 (0)221 277 29-303
Fax: +49 (0)221 277 29-400
sebastian.nick@uni-koeln.de

ISSN: 1862-3808

Price Formation and Intertemporal Arbitrage within a Low-Liquidity Framework: Empirical Evidence from European Natural Gas Markets

Sebastian Nick[a,*]

[a] *Institute of Energy Economics, University of Cologne, Vogelsanger Straße 321, 50827 Cologne, Germany.*

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 are specified that explicitly account for nonlinearities and the low liquidity-framework of the considered gas hubs. The empirical results reveal that price discovery takes place on the futures market, while the spot price subsequently follows the futures market price. Furthermore, there is empirical evidence of significant market frictions hampering intertemporal arbitrage. UK's NBP seems to be the hub 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, liquidity, nonlinear causality, threshold error correction, Kalman filter

JEL classification: Q40, Q41, G14, C58

[☆] The author would like to thank Felix Höffler, Christian Growitsch, Heike Wetzel, Hans Manner, Anne Neumann and Stefan Thoenes for helpful comments on earlier versions of the paper. Julia Bellenbaum provided excellent research assistance.

[*] Corresponding author

Email address: Sebastian.Nick@ewi.uni-koeln.de, +49 22127729303 (Sebastian Nick)

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 fulfill 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 US. 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]

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 the German, Dutch and British gas hubs, where the mature and liquid British hub serves as a benchmark for the other hubs.[2]

[1] For an elaborated discussion of the economics fostering the transition from oil-indexation to hub-based pricing, see Stern and Rogers (2010). A real-life illustration are the current renegotiations of LTCs between various continental European gas importers with their suppliers (ICIS, 2013).

[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.

In the first part of the paper, the price formation process at European natural gas hubs is investigated. Fama (1970) states that all relevant information should be instantaneously reflected in prices within efficient markets, ruling out the predictability of price changes between equally efficient markets. With regard to price formation on spot and futures markets for the same underlying asset, Fama's idea implies that there should be no systematic lead-lag relationship between these markets. Thus, this study uses Granger causality testing to investigate whether there is a structural lead-lag relationship between price changes at spot and futures markets (i.e., one market dominates the price discovery process) or whether both markets react simultaneously to new information. In addition, this work allows for higher-moment interaction between price changes of the spot and futures markets by drawing upon nonlinear causality testing. From an economic perspective, analyzing the price discovery process for natural gas markets is particularly interesting because spot and futures markets for the same underlying commodity may have heterogeneous information sets: The spot market for natural gas may be driven to a greater extent by short-term influences (e.g., weather conditions or infrastructure outages) than the corresponding futures market. Consequently, the application of causality tests on spot and futures markets for natural gas yields insights into the price discovery process of spot and futures markets with the same underlying asset but partially different information sets.

The second part of this work seeks to assess how efficiently intertemporal arbitrage opportunities between spot and futures markets at the European gas hubs are exhausted. The theory of storage states that efficient arbitrage between spot and futures markets, carried out by storage operators, should establish a stable equilibrium between the two markets (Working, 1949). Cointegration techniques are applied to test for a long-run equilibrium between spot and futures markets. The efficiency of intertemporal arbitrage can be measured by the time required to correct a deviation from the long-run equilibrium: Well-informed market participants are expected to react instantaneously to arbitrage opportunities, resulting in a quick correction of deviations from the intertemporal equilibrium, while market participants with low informational efficiency should lead to a stickier adjustment process. To assess the speed of adjustment towards the arbitrage free equilibrium, vector error correction models (VECM) are estimated for the European gas hubs.

In the context of European natural gas hubs, the lack of liquidity as well as the physical characteristics of the market (e.g., restricted withdrawal and injection capacities of gas storages) may affect arbitrage patterns between the spot and the futures market. Within this paper, such frictions are explicitly addressed by allowing for nonlinearities in the arbitrage process estimating threshold VECM (TVECM). The TVECM enables the identification of regimes that exhibit different arbitrage dynamics depending on the magnitude of the deviation from the long-run equilibrium. Thus, the significance of the frictions hampering intertermporal arbitrage can be assessed by comparing the error correction processes across the regimes. Beside allowing for nonlinearities in the arbitrage processes, the efficiency of intertemporal arbitrage is investigted within a dynamic econometric framework. This methodology is particularly promising for the European gas hubs, as the Dutch and the German hub experienced a remarkable growth in liquidity. To capture the resulting effects on the efficiency of intertemporal arbitrage at the continental European hubs over time, the Kalman filter technique is applied to estimate a VECM with time-varying coefficients. This state-space approach permits the analysis of whether the increase in liquidity has fostered the arbitrage efficiency of the two continental European gas hubs throughout the sample period.

This paper extends research on natural gas markets in various ways: Foremost, it is the first to analyze 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. Moreover, the findings of this paper are not limited to natural gas markets and may be transferred to other relatively immature markets suffering from low liquidity.

The causality tests reveal a significant lead-lag relationship, with the month-ahead futures contract leading the spot market. Thus, information is not processed simultaneously on spot and futures markets, indicating that price formation takes place on the futures market despite the partly different information set of the two markets. The empirical results suggest that the theory of storage holds for all hubs in the long-run. The linear VECM approach indicates that intertemporal arbitrage opportunities appear to be most efficiently exhausted at the British hub, a finding that is consistent with the relative maturity of this hub. The results of the TVECM confirm that instantaneous arbitrage between spot and futures markets is hampered by market frictions, at least at some of the hubs considered. From a dynamic perspective, the state-space VECM approach shows convergence in the efficiency of intertemporal arbitrage among the hubs within the sample period.

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 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 informational efficiency over time is specified in Section 7. Section 8 concludes.

2. 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 affect 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.

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. Accordingly, information processing and price discovery occur in the futures market and the spot prices adjust accordingly until an arbitrage-free equilibrium is achieved. Bohl et al. (2012) link the ability of futures markets to provide reliable price discovery to the degree of institutionalization, stating that price discovery takes place at the futures market only in the case that the market is dominated by institutional investors. 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. Apart from a unidirectional lead-lag pattern, the pattern of price discovery may change depending on the kind of information entering the market (Kawaller et al., 1988).

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 US gas hub. Focusing on the northwest US 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, empirical research has centered on the assessment of market integration and the efficiency of regional arbitrage (e.g., Neumann et al., 2006; Growitsch et al., 2012), whereas the price formation process at the European spot and futures markets has thus far been neglected.

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 corresponding cost-of-carry hypothesis states that deviations from the spot-futures equilibrium are only transitory, as efficient arbitrage helps to restore the long-run relationship. 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 compounded spot price $S_t(1+r_{t+k|t})$ 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 (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.

3. 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 hub 'NetConnect Germany' (NCG)[3], the Dutch gas hub 'Title Transfer Facility' (TTF)[4] and UK's 'National Balancing Point' (NBP)[5] for the period

[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.

October 2007 to August 2012.[6] All prices represent the settlement prices of the respective trading day. The selection of the two continental European hubs is motivated by the steady rise in trading activity during the last years, suggesting that (at least) one of theses hubs will emerge as the leading continental European trading area (Heather, 2012). The NBP hub, as the most mature and liquid hub in Europe, serves as benchmark to assess the informational efficiency of NCG and TTF. 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). Table 1 provides descriptive statistics on the gas price returns (calculated as the differences in the logarithms of two consecutive daily settlement prices) of the three hubs considered in this study.

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
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	1.5960	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

All price return series exhibit means close to zero.[7] The Samuelson Hypothesis, stating that the variance of price returns decreases with the maturity (Samuelson, 1965), is confirmed by the data: Spot market returns have the greatest variance, while fluctuations gradually decline from the m+1 to the m+3 contracts. All return series exhibit excess kurtosis, reflecting a fat-tailed distribution that is frequently observed in commodity market return series.

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. For all price series, the null hypothesis of a unit root

[6] The beginning of the sample has been restricted by the availability of NCG prices which were not available before October 2007.

[7] For all return series, the mean is not significantly different from zero when regressing against a constant. Thus, there seems to be no expected return on a daily level.

in the log-level cannot be rejected, which is the case for the first differences (i.e., the daily returns). The results of the unit root tests are presented in the Appendix.

Since the spot and futures price series are integrated of order one, the cost-of-carry hypothesis between the spot and futures markets at the considered hubs can be investigated via cointegration analysis.[8] The concept of cointegration was developed by Engle and Granger (1987). It states that for two time series, both integrated of order n with n greater or equal to one, there exists a linear combination of these series that is integrated of order $n-1$. Following Lütkepohl (2005), the cointegration relationship can be investigated based on a k-dimensional vector autoregressive model (VAR) of order p:

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \tag{4}$$

where cointegration of rank r implies that the matrix

$$\Pi = -(I_k - A_1 - ... - A_p) = \alpha\beta', \tag{5}$$

is of rank r, while α and β (representing the loading matrix and cointegration matrix, respectively) are of dimension (k x r) and of rank r. To determine the rank of Π, the procedure proposed by Johansen (1988) is applied. The test results are presented in Table 2. The null hypothesis of no cointegration between spot- and month-ahead prices can be rejected for all hubs.[9]

4. The Role of Liquidity and Storage Capacity

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 hubs (NCG and TTF) suffer from low liquidity despite steadily increasing trading volumes during the last years.

[8] 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.

[9] 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 empirical findings significantly.

Table 2: Results of the Johansen Cointegration Test (Spot and m+1-Contract)

Hypothesis	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
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

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 be used to assess the degree of financialization of commodity markets. Table 3 illustrates the differences among the three 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 3, 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.

Table 3: 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
TTF	35.6	151.7	4.3
NBP	79.6	1137.2	14.3

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

Beside the frictions resulting from illiquid spot and futures markets, the efficiency of intertemporal arbitrage activity may also be restricted by technical constraints. In particular, scarcity in storage capacity may prevent efficient arbitrage trading (at least in the short run), since the construction of additional storage facilities requires significant amount of time. A first indicator for the availability of sufficient storage capacity 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

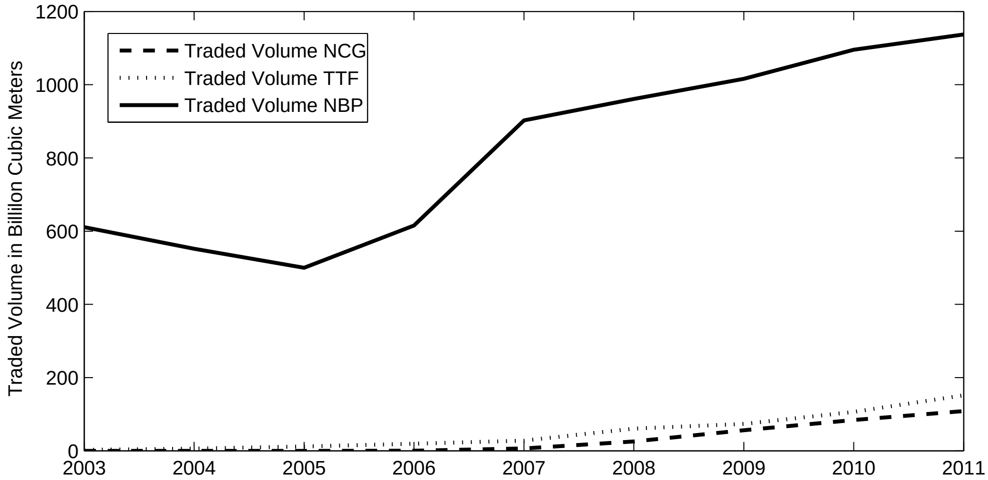

Figure 1: Trading Volumes at European Gas Hubs

Source: IEA (2012a)

measures for the degree of gas storage flexibility are the shares of aggregated injection capacity (IC) and aggregated withdrawal capacity (WC) on aggregated working gas volume (WGV). Table 4 presents data on WGV (measured in billion cubic meters (bcm)), consumption (C, in bcm per year) and the three flexibility indicators for Germany,[10] the Netherlands and the UK.

Table 4: Storage Capacity and Flexibility Potential (as of 2011)

	WGV (bcm)	C (bcm/a)	C/WGV	WC / WGV	IC / WGV
Germany	20.1	77.6	3.8	0.0215	0.0111
Netherlands	5.1	47.9	9.4	0.0410	0.0112
UK	4.5	82.6	18.2	0.0195	0.0055

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

The data emphasize the ample storage capacity of the German gas market. Thus, physical scarcity should not prevent efficient intertemporal arbitrage trading at NCG. In contrast, storage capacity in the UK is rather scarce (in a physical sense), since the WGV only amounts to approximately 5% of annual gas consumption. The Netherlands range between Germany and UK in terms

[10]Since there are two market areas in Germany (NCG and Gaspool) not all German consumption and storage capacity can be allocated to NCG. However, since consumption and storages are split rather equally between NCG and Gaspool, the consumption-to-storage ratio can be considered as a valid approximation for both hubs.

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, all indicators thus suggest that the storage market in the UK is less supportive of efficient intertemporal arbitrage activity. On the other hand, the comparably high level of LNG imports to the UK may constitute another flexibility option.[11] However, the flexiblity in receiving the delivery of LNG is limited due to significant transportation distances and the usage of LTCs. Moreover, once having entered the British market, the LNG has to be stored before being injected into the network, thus suffering from the relatively inflexible and limited storage capacities.

Beside the physical capacity constraints, contractual congestion may hamper intertemporal arbitrage activity. It should be noted, however, that the markets considered do not differ from each other in terms of the way in which they have implemented the third party access to storage facilities required by the third European natural gas directive (EU, 2009): The national regulators generally opted for a negotiated, rather than a regulated, third party access model.[12] Thus, regulatory differences are not expected to explain potential differences in the efficiency of intertemporal arbitrage between the hubs investigated.

5. Price Formation at European Gas Hubs: Linear and Nonlinear Causality Testing

This section investigates the price discovery process on the spot and futures markets of the European gas hubs. Econometric tests are applied to investigate whether Fama's hypothesis of simultaneous information processing (Fama, 1970) holds for the spot and futures markets under consideration. The idea of simultaneous information processing implies that there should be no systematic (i.e., no causal) relationship between price changes on spot and futures markets. Thus, causality testing can be applied to analyze information transmission and price formation. Granger (1969) defined a concept of causality based on a time lag between cause and effects. Accordingly, a process x_t is said to cause a process y_t in the sense of Granger if

[11] In 2011, UK imported 24.8 bcm of LNG, whereas there have been no significant direct LNG imports to the German and Dutch market (IEA, 2012b).

[12] Only in the UK have some LNG storage facilities been subject to regulated access tariffs during the sample period and some UK storages been exempted from the third party access obligation.

$$\Sigma_z(h|\Omega_t) < \Sigma_z(h|\Omega_t \backslash (x_s|s \leq t)) \text{ for at least one } h = 1,2,..., \quad (6)$$

where $\Sigma_z(h|\Omega_t)$ 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). Hence, x_t causes y_t in the sense of Granger if past and current values of x have explanatory power for future values of y. An obvious shortfall of this concept of causality is the fact that it is based on lagged correlation patterns between the series of interest rather than on "real" causal linkages. However, in the context of the study, this feature is not a drawback, as the central interest is exactly this lead-lag pattern among spot and futures markets for natural gas.

The testing procedure proposed by Granger requires covariance stationarity of the investigated series. Thus, we rely on the daily returns of the price series, as the unit root tests in Section 3 suggest that the price series themselves are integrated of order one. Moreover, the test is carried out within the vector error correction (VECM) framework. By doing so, the cointegration relationship between spot and futures prices revealed in Section 2 is explicitly accounted for to avoid misleading inference.[13] In addition, the VECM-filtered residuals (i.e., the residuals obtained from the VECM estimation) are tested for any remaining linear causality pattern. Table 5 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 three hubs. This means that the change in the month-ahead futures price has explanatory character for the next day's spot price change.[14] 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 month-ahead 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

[13]Ignoring an existing cointegration relationship may lead to invalid results of linear and nonlinear Granger causality tests, as outlined by Chen and Lin (2004).

[14]The finding of Granger causality from futures market price returns to spot market price returns at all hubs remains unchanged when controlling for conditional heteroskedasticity of the price return series within a GARCH(1,1)-framework. The results of the Granger causality tests for the spot market and longer-maturity futures markets are presented in the Appendix. For all hubs, there is empirical evidence of futures markets leading 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 due to 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 that 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 market leading the corresponding spot market is in line with the finding of Root and Lien (2003) and Dergiades et al. (2012) for the US 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.

Table 5: Pairwise Linear Causality Tests for Gas Price Returns

Direction	Chi-sq-Statistic
NCG Spot on NCG m+1	0.1193
NCG m+1 on NCG Spot	5.7441**
TTF Spot on TTF m+1	2.8416
TTF m+1 on TTF Spot	306.06***
NBP Spot on NBP m+1	2.1940
NBP m+1 on NBP Spot	8.6832**

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

The econometric methodology as applied thus far 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). Additionally, asymmetric information and heterogeneous expectations of market participants may also induce nonlinearities in the relationship between commodity spot and futures prices (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 the low liquidity at these hubs may foster market frictions such as transaction and information costs. Following this reasoning, the nonlinear causality test proposed by Diks and Panchenko (2006) is applied to investigate nonlinear dynamics among the considered spot and futures markets. The testing procedure of Diks and Panchenko (2006) is based on the Hiemstra Jones Test (Hiemstra and Jones, 1994). 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

$$\frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)} \cdot \frac{f_{Y,Z}(y,z)}{f_Y(y)}. \tag{7}$$

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

$$q \equiv E\left[f_{X,Y,Z}(X,Y,Z)f_Y(Y) - f_{X,Y}(X,Y)f_{Y,Z}(Y,Z)\right] = 0. \tag{8}$$

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

$$T_n(\epsilon_n) = \frac{n-1}{n(n-2)} \cdot \sum_i (\hat{f}_{X,Z,Y}(X_i,Z_i,Y_i)\hat{f}_Y(Y_i) - \hat{f}_{X,Z}(X_i,Y_i)\hat{f}_{Y,Z}(Y_i,Z_i)), \tag{9}$$

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

$$\hat{f}_W(W_i) = (2\epsilon_n)^{-d_W}(n-1)^{-1} \sum_{j,j\neq i} I_{ij}{}^W, \tag{10}$$

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

$$\sqrt{n}\frac{(T_n(\epsilon_n) - q)}{S_n} \xrightarrow{d} N(0, 1), \tag{11}$$

for a lag length of 1 and $\epsilon_n = Cn^{-\beta}$ 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

$$\epsilon_n^* = C^* n^{(-2/7)}. \tag{12}$$

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. In doing so, cointegration is explicitly controlled for in order to avoid cointegration mechanisms to be accidentally interpreted as a nonlinear causal relationship as suggested by Chen and Lin (2004). Following Diks and Panchenko (2006), the constant term C^* of the bandwidth ϵ_n is set to 8.[15] Inserting C^* into equation (12) results in a bandwidth ϵ_n of approximately 1.[16]

As can be seen in Table 6, the null hypothesis of absent nonlinear Granger causality among spot- and month-ahead return series can be rejected in both directions for all three hubs.[17] 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 test. Following this argument, a multivariate GARCH model is applied to capture the dynamics in the second moment of distribution in both markets, filtering out conditional volatility effects. The BEKK GARCH model

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

[16] As a robustness check, the test has been conducted with smaller and larger bandwidths within the range of 0.9 and 1.1. However, the results are not very sensitive to the choice of bandwidth.

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

of Engle and Kroner (1995) is applied to explicitly control for volatility spillover effects between the investigated markets.[18] The BEKK GARCH model can be written as

$$\Sigma_{t|t-1} = C_0^{*\prime} C_0^* + \sum_{n=1}^{N} \sum_{j=1}^{q} \Gamma_{jn}^{*}{}' u_{t-j} u_{ut-j}' \Gamma_{jn}^* \sum_{n=1}^{N} \sum_{j=1}^{m} G^* \Sigma_{t-j|t-j-1} G_{jn}^*, \tag{13}$$

where the matrices C^*, Γ^*, and G^* are N x N with C* are lower triangular matrices. To reduce computational complexity, Γ^* and G^* are restricted to be diagonal, as proposed by Bekiros (2011) and Silvennoinen and Teräsvirtä (2008). Subsequently, the nonlinear causality test of Diks and Panchenko (2006) is applied to the BEKK GARCH-filtered VECM residuals.

For all hubs, the nonlinear causality from spot to futures markets disappears after BEKK-GARCH filtering. This suggests that any predictive power of spot return distributions for subsequent distributions of futures market returns is due to simple volatility effects. Thus, the role of spot markets concerning price discovery is limited (if existent at all) for all hubs. Moreover, the results do not yield evidence of nonlinear causality among spot and futures markets for NCG and NBP. For these hubs, the causal relationship is apparently limited to the first and second moment of distribution. Only for TTF, there is unidirectional nonlinear causality from the month-ahead to the spot market.

Table 6: 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***
	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***
BEKK GARCH-filtered Data	NCG Spot on NCG m+1	-1.944
	NCG m+1 on NCG Spot	-0.477
	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

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

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

6. Long- and Short-Run Dynamics between Spot and Futures Markets: The Efficiency of Intertemporal Arbitrage

The finding of cointegration relationships for the spot and futures market price series at all hubs in Section 3 suggests that the theory of storage holds in the long run. This long-run relationship can be written as

$$S_t = c + \beta_t F_t + \epsilon_t. \tag{14}$$

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 relationship. With regard to the cost-of-carry relationship, c 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 between spot and futures markets.[19] The impact of such a deviation on the short-term behavior of the series can be modeled by the following VECM:

$$\begin{aligned} \Delta f_t = &\quad \alpha_t^f \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^f \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^f \Delta s_{t-k} + \eta_t^f, \\ \Delta s_t = &\quad \alpha_t^s \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma_k^s \Delta f_{t-k} + \sum_{k=1}^{k=n} \delta_k^s \Delta s_{t-k} + \eta_t^s, \end{aligned} \tag{15}$$

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. The γ and δ coefficients account for autoregressive behavior of the series. To asses the efficiency of arbitrage, the α coefficients are of central interest because they measure the speed of error correction: The greater the value of the adjustment coefficient (in absolute terms), the more informationally efficient

[19] 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.

are the market participants in exhausting arbitrage opportunities. In contrast, small absolute values of α indicate a sticky adjustment process and hence a lower level of informational efficiency. In the context of the European natural gas hubs, the comparably low liquidity at NCG and TTF suggests significant transaction costs and may consequently lead to a slower (i.e., less efficient) adjustment process compared to at NBP. To investigate the efficiency of intertemporal arbitrage, linear VECM models are estimated for all hubs. Table 7 presents the estimated cointegration vector and the adjustment coefficients.

Table 7: 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_{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

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 and TTF, while the same criterion suggests to include 2 lags for NBP.The autoregressive coefficients are not reported to conserve space.

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, 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. The insignificant adjustment coefficient in all futures return equations suggests that these series are weakly exogenous with respect to the corresponding spot price series (Urbain, 1992).[20]

[20] 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.

The small absolute values of the adjustment coefficients imply a sticky error correction process and thus suggest a rather low efficiency of intertemporal arbitrage.[21] Although this means that none of the considered hubs can be regarded as fully informationally efficient, arbitrage seems to be most efficiently exploited at NBP. This finding is noteworthy, as physical storage flexibility is much smaller in the UK than in Germany and in the Netherlands (see Table 4) 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 (compared to NCG and TTF) is fairly moderate.

The specified VECM assumes linearity in the adjustment process. This implies that error correction 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 the low liquidity at the respective spot and futures markets 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). To account for market frictions resulting from the low liquidity at European hubs and from gas market specific characteristics, the TVECM approach proposed by Granger and Lee (1989) is applied in the following. TVECMs have proved to be a useful approach for capturing arbitrage dynamics among spot and futures markets for various commodities by explicitly accounting for market frictions (Li, 2010; Huang et al., 2009; Root and Lien, 2003). The bivariate TVECM of order n applied to the bivariate system of spot and futures market returns has the representation

$$\begin{aligned} \Delta f_t = \quad & (I-1)\alpha_h^f \epsilon_{t-1} + I\alpha_l^f \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma^f \Delta f_{t-1} + \sum_{k=1}^{k=n} \delta^f \Delta s_{t-1} + \eta_t^f, \\ \Delta s_t = \quad & (I-1)\alpha_h^s \epsilon_{t-1} + I\alpha_l^s \epsilon_{t-1} + \sum_{k=1}^{k=n} \gamma^f \Delta f_{t-1} + \sum_{k=1}^{k=n} \delta^f \Delta s_{t-1} + \eta_t^s, \end{aligned} \tag{16}$$

[21] 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.

where I denotes the regime indicator stating whether the lagged deviation from the long-run equilibrium is below or above the threshold (in absolute terms). The 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. The γ and δ coefficients account for the autoregressive behavior of the series (Enders and Siklos, 2001).

The model of Equation (16) is estimated using different thresholds. The thresholds are assumed to be symmetric and their size is defined in terms of the standard deviation of ϵ_t, the error term of the cointegration regression.[22] This approach reveals the magnitude of the deviation from the long-run equilibrium that is necessary to trigger arbitrage activity by investigating the statistical significance of α_l, the adjustment coefficient in the "lower regime" (i.e., the regime with small deviations from the long-run equilibrium) for different thresholds. Thus, the results of the estimations yield insights into the significance of market frictions hampering instantaneous arbitrage (e.g., transaction costs or physical constraints). Moreover, the absolute values of the regime-specific adjustment coefficients α_l and α_h reveal the efficiency of intertemporal arbitrage within the respective regime. Table 8 contains the estimates for the regime-specific adjustment coefficients of the TVECM.

Table 8: Estimates of Threshold Vector Error Correction Models

		NCG		TTF		NBP	
Threshold	Regime	α_{spot}	α_{m1}	α_{spot}	α_{m1}	α_{spot}	α_{m1}
$0.5\sigma_\epsilon$	high	-0.1359***	0.0238	-0.1101***	0.0012	-0.1603***	0.0042
$0.5\sigma_\epsilon$	low	-0.0834	0.0094	-0.1312***	0.0421	-0.0465	0.0078
σ_ϵ	high	-0.1358***	0.0148	-0.1092***	0.0080	-0.1835***	0.0033
σ_ϵ	low	-0.1203***	-0.0144	-0.1212***	-0.0243	-0.0196	0.0094

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 and TTF, while the same criterion suggests to include 2 lags for NBP. The autoregressive coefficients are not reported to conserve space.

[22] The standard deviations of ϵ_t are 0.08 for NCG and TTF, and 0.11 for NBP. 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.

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 frictions constraining instantaneous intertemporal arbitrage trading. Not even the efficiency of arbitrage seems to be regime-specific, as both adjustment coefficients have similar magnitude. In contrast, arbitrage at NCG and NBP 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. This finding suggests that characteristics of the gas market other than liquidity (e.g., restricted storage flexibility) prevent the immediate exhaustion of arbitrage opportunities. However, once the deviation from the long-run equilibrium crosses the threshold, arbitrage opportunities are exploited most efficiently at NBP, as can be inferred from the absolute values of the respective adjustment coefficients. In line with the results of the linear VECM, the futures return series does not adjust to restore the long-run equilibrium, as none of the adjustment coefficients in the futures price return equation is statistically significant.

To sum up, intertemporal arbitrage starts most instantaneously at TTF but is executed most efficiently at NBP. The first finding is in line with the high flexibility of Dutch gas storage (see Table 4), while the latter can be attributed to the superior liquidity of NBP (see Table 3).

7. The Evolution of Arbitrage Efficiency: A Kalman Filter Approach

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 is applied to capture the evolution of intertemporal arbitrage efficiency over time. Time-varying coefficient models have been used for the European gas market in different applications. Neumann et al. (2006) draw upon a state-space approach to investigate regional price convergence. Growitsch et al. (2012) estimate a time-varying VECM to assess the evolution of regional price arbitrage

[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).

efficiency over time. However, this paper is the first to apply the state-space methodology within an intertemporal context for the European gas markets. A two-step procedure is applied to estimate a VECM with time-varying coefficients: In a first step, the long-run relationship between spot and futures prices over time is estimated based on the following state-space model:

$$S_t = c + \beta_t F_t + \epsilon_t \tag{17}$$

with

$$\beta_t = \beta_{t-1} + \psi_t, \tag{18}$$

where β represents the state-dependent coefficient, while the error terms ϵ_t and ψ_t are white noise processes with zero mean and variances σ^2_ϵ and σ^2_ψ, respectively. The β coefficients in Equation (16) are estimated using the Kalman filter (Kalman, 1960).[24] In a second step, the intertemporal arbitrage dynamic is investigated by estimating Equation (18), since the time-varying behavior of the adjustment coefficient is of central interest for this study.

$$\begin{aligned} \Delta f_t = & \quad c^f + \alpha^f_t \epsilon_{t-1} + \eta^f_t, \\ \Delta s_t = & \quad c^s + \alpha^s_t \epsilon_{t-1} + \eta^s_t, \end{aligned} \tag{19}$$

with

$$\alpha_t = \alpha_{t-1} + \zeta_t, \tag{20}$$

where α_t represents the time-varying adjustment coefficient. The recursive procedure suggested by Kalman (1960) is applied to estimate the state-space model.[25] 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 coefficients are expected to increase over time. Figure 2 presents the estimated time paths for the adjustment coefficients in the spot return

[24] One is chosen as the initial value for β assuming that futures and spot prices are driven by the same fundamentals. For the initial variances of ϵ_t, the variance of the respective log-spot price, σ^2_{spot}, is selected, while ψ_t is set to $\sigma^2_{spot}/1000$, to achieve an appropriate signal to noise ratio.

[25] 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 $\sigma^2_{rspot}/1000$.

equation.[26] The spikes in the series can be attributed to the economic downturn in autumn 2008, and gas market-specific shocks such as the extraordinary supply interruptions resulting form the Russian-Ukrainian crisis in January 2009 and the cold spell in February 2012.[27] For all three hubs, there is a distinctive pattern in the evolution of the adjustment coefficients over time: For NCG, the efficiency of intertemporal arbitrage is low at the beginning of the sample period. However, it steadily rises (the absolute value of α increases), reaching the efficiency level of TTF at the end of the sample period. For TTF, the α coefficient is rather stable over time. This implies that informational efficiency has remained quite unchanged at this hub. In contrast, the absolute value of the adjustment coefficient of NBP decreases over time, indicating a decline in the efficiency of intertemporal arbitrage. Overall, there is convergence in the degree of informational efficiency of the hubs considered. Thus, as of 2012, the differences in arbitrage efficiency of the hubs considered appear significantly reduced. This finding may be attributed to the increased liquidity of the continental European hubs, with NCG having benefited most in terms of informational efficiency.

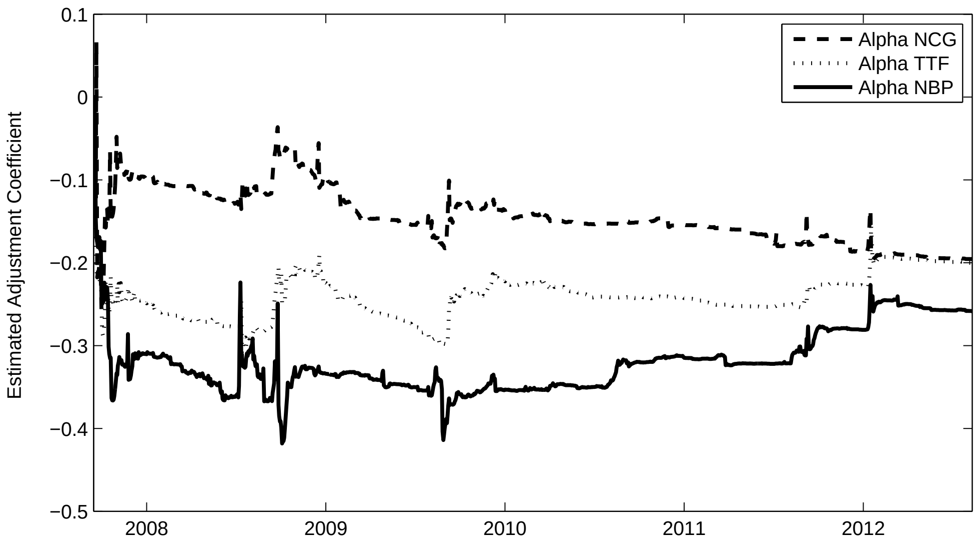

Figure 2: Time-Varying Adjustment Coefficients of Spot Price Return Series

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

[27]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).

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

Causality testing reveals that price formation takes place on the futures market at all hubs. 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., 2012). 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 have significant informational superiority compared to the spot market.

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. Relating the empirical findings to the liquidity of the respective hubs requires careful interpretation: On the one hand, the detected frictions in the price formation process and arbitrage activities are similar for all hubs, regardless of their liquidity. Moreover, instantaneous arbitrage appears to be significantly constrained at NBP, the most liquid hub of the sample. 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. On the other hand, the superior liquidity of NBP (compared to NCG and TTF) apparently does matter when it comes to the efficiency of arbitrage. Despite the restricted capacity and flexibility of gas storages in the UK, intertemporal arbitrage opportunities are most efficiently exploited at NBP (once arbitrage activity has been triggered). Another finding that points towards a positive effect of liquidity on informational efficiency is the increase in intertemporal arbitrage efficiency at NCG throughout the sample period, coinciding with a steady rise in liquidity at this hub.

The insights gained by this work have several implications on the potential hub price indexation of gas LTCs: The finding that price formation takes place at the futures markets suggests that an indexation of LTCs on near-mature futures prices is more likely to capture valid price signals than spot price indexation. Moreover, with regard to the whole sample period, NBP is the hub with the highest informational efficiency. Thus, despite the implied currency risks and the limited physical network integration of the UK and mainland Europe, including NBP into a price benchmark may yield efficiency gains.

A promising field for further research could be the extension of the analysis to intraday data, investigating the interaction of spot and futures markets at an even higher time resolution. Another extension of this study may be the inclusion of futures contracts with longer maturity. Both exercises, however, suffer from the lack of data availability and are therefore left for future research ventures.

References

Arouri, M. E. H., Hammoudeh, S., Lahiani, A., and Nguyen, D. K. (2013). On the short- and long-run efficiency of energy and precious metal markets. HAL Working Papers Series hal-00798036.

Bekiros, S. (2011). Nonlinear causality testing with stepwise multivariate filtering. EUI ECO Working Papers 2011-22, European University Institute.

Bekiros, S. D. and Diks, C. G. (2008). The relationship between crude oil spot and futures prices: Cointegration, linear and nonlinear causality. *Energy Economics*, 30(5):2673–2685.

Bohl, M. T., Salm, C. A., and Schuppli, M. (2012). Price discovery and investor structure in stock index futures. *The Journal of Futures Markets*, 31(2):282–306.

Chen, A. and Lin, J. (2004). Cointegration and detectable linear and nonlinear causality: Analysis using the London Metal Exchange lead contract. *Applied Economics*, 36(11):1157–1167.

Considine, T. J. and Larson, D. F. (2001). Risk premium on inventory assets: The case of crude oil and natural gas. *The Journal of Futures Markets*, 21(2):109–126.

Dergiades, T., Christofidou, G., and Madlener, R. (2012). The nexus between natural gas spot and futures prices at NYMEX: What about non-linear causality? FCN Working Paper Series 17/2012, Institute for Future Energy Costumer Needs and Behavior.

Diks, C. and Panchenko, V. (2005). A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics & Econometrics*, 9(2):1–9.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30:1647–1669.

Enders, W. and Siklos, P. L. (2001). Cointegration and threshold adjustment. *Journal of Business and Economic Statistics*, 19(2):166–176.

Engle, R. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):251–276.

Engle, R. and Kroner, K. (1995). Multivariate simultaneous generalized GARCH. *Econometric Theory*, 11:122150.

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

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *The Journal of Business*, 60(1):55–73.

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.

GIE (2011). Gas Storage Europe database. Technical report, Gas Infrastructure Europe.

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

Granger, C. and Lee, T. (1989). Investigation of production, sales and inventory relationships using multicointegration and nonsymmetric error-correction models. *Journal of Applied Econometrics*, 4(Special Issue on Topics in Applied Econometrics):146–159.

Growitsch, C., Stronzik, M., and Nepal, R. (2012). Price convergence and information efficiency in German natural gas markets. EWI Working Papers 2012-5, Institute of Energy Economics at the University of Cologne.

Heather, P. (2012). Continental european gas hubs: Are they fit for purpose? OIES Working Papers NG 63, The Oxford Institute for Energy Studies.

Hiemstra, C. and Jones, J. (1994). Testing for linear and nonlinear granger causality in the stock pricevolume relation. *Journal of Finance*, 49(5):1639–1664.

Huang, B., Yang, C., and Hwang, M. (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.

ICIS (2013). Heren European gas markets. Newsletter 15 March 2013, ICIS.

IEA (2012a). Medium term gas market report. Technical report, International Energy Agency.

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

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economics Dynamics and Control*, 12(2-3):231–254.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kawaller, I., Koch, P., and Koch, T. (1988). The relationship between the S&P 500 index and the S&P 500 index futures prices. *Federal Reserve Bank of Atlanta Economic Review*, 73(3):2–10.

Li, M.-Y. L. (2010). Dynamic hedge ratio for stock and index futures: Application of threshold VECM. *Applied Economics*, 42(11):1403–1417.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Moosa, I. and Al-Loughani, N. (1995). The effectiveness of arbitrage and speculation in the crude oil futures market. *The Journal of Futures Markets*, 15:167–186.

Neumann, A., Siliverstovs, B., and von Hirschhausen, C. (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.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Nick, S. and Thoenes, S. (2013). What drives natural gas prices? - A structural VAR approach. EWI Working Papers 2013-2, Institute of Energy Economics at the University of Cologne.

NMA (2012). 2012 Liquidity Report: Wholesale markets for natural gas and electricity. Technical report, Netherlands Competition Authority - Office of Energy Regulation.

Root, T. H. and Lien, D. (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.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Management Review*, 6(2):41–49.

Schwartz, T. and Szakmary, A. (1994). Price discovery in petroleum markets: Arbitrage, cointegration and the time interval of analysis. *The Journal of Futures Markets*, 14(2):147167.

Silvapulle, P. and Moosa, I. A. (1999). The relationship between spot and futures prices: Evidence from the crude oil market. *The Journal of Futures Markets*, 19(2):175–193.

Silvennoinen, A. and Teräsvirtä, T. (2008). Multivariate GARCH models. CREATES Working Paper Series 2008-6, Center for Research in Econometric Analysis of Time Series University of Aarhus.

Stern, J. P. and Rogers, H. (2010). The transition to hub-based gas pricing in continental Europe. OIES Working Papers NG 49, The Oxford Institute for Energy Studies.

Stronzik, M., Rammerstorfer, M., and Neumann, A. (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.

Urbain, J. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics*, 54(2):187–207.

Working, H. (1949). The theory of the price of storage. *American Economic Review*, 39(6):1242–1262.

Zhang, J. and Jinghong, S. (2012). Causality in the VIX futures market. *The Journal of Futures Markets*, 32(1):24–46.

Appendix A. Test Statistics

Table A.9: 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
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
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
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

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.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	33.6016	20.262	0.0004
NCG r$\leq$ 1	0.0013	1.5655	9.1645	0.8615
TTF r=0	0.0227	29.6522	20.262	0.0019
TTF r$\leq$ 1	0.0012	1.5023	9.1645	0.8730
NBP r=0	0.0226	31.4289	20.262	0.0010
NBP r$\leq$ 1	0.0020	2.5516	9.1645	0.6673

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$\leq$ 1	0.0013	1.5672	9.1645	0.8612
TTF r=0	0.0149	19.8639	20.262	0.0566
TTF r$\leq$ 1	0.0012	1.5087	9.1645	0.8718
NBP r=0	0.0226	31.4289	20.262	0.0010
NBP r$\leq$ 1	0.0012	1.5087	9.1645	0.8718

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

	Direction	Chi-sq-Statistic
Raw Data	NCG Spot on NCG m+2	0.0593
	NCG m+2 on NCG Spot	12.974***
	NCG Spot on NCG m+3	2.6556
	NCG m+3 on NCG Spot	10.8730***
	NCG m+1 on NCG m+2	3.6889
	NCG m+2 on NCG m+1	0.0989
	NCG m+1 on NCG m+3	3.7935
	NCG m+3 on NCG m+1	1.1040
	NCG m+2 on NCG m+3	3.2389
	NCG m+3 on NCG m+2	2.2918
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 (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.13: Pairwise Linear Causality Tests for TTF Returns

	Direction	Chi-sq-Statistic
Raw Data	TTF Spot on TTF m+2	5.1896
	TTF m+2 on TTF Spot	347.91***
	TTF Spot on TTF m+3	6.3281**
	TTF m+3 on TTF Spot	349.45***
	TTF m+1 on TTF m+2	0.0001
	TTF m+2 on TTF m+1	0.8102
	TTF m+1 on TTF m+3	0.2332
	TTF m+3 on TTF m+1	0.9347
	TTF m+2 on TTF m+3	0.4150
	TTF m+3 on TTF m+2	4.1041**
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.14: Pairwise Linear Causality Tests for NBP Returns

	Direction	Chi-sq-Statistic
Raw Data	NBP Spot on NBP m+2	2.7163
	NBP m+2 on NBP Spot	33.872***
	NBP Spot on NBP m+3	3.2826
	NBP m+3 on NBP Spot	38.780***
	NBP m+1 on NBP m+2	162.84***
	NBP m+2 on NBP m+1	0.1249
	NBP m+1 on NBP m+3	23.098***
	NBP m+3 on NBP m+1	0.8021
	NBP m+2 on NBP m+3	0.4293
	NBP m+3 on NBP m+2	0.9533
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 (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: Results of the Likelihood Ratio Test on the Cointegration Vector

	Chi-sq-Statistic	p-Value
NCG	0.4036	0.5252
TTF	0.4726	0.4918
NBP	0.3605	0.5482

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: β =[-1;-1].

Table A.16: 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_{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	-1.0503

Notes: *** (**) Denotes significance at the 99 (95) %-level. A lag length of 1 for the both VECMs is selected based on the Schwarz Information Criterion for NCG and TTF, while the same criterion suggests to include 2 lags for NBP.

Table A.17: 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_{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	-1.1493

Notes: *** (**) Denotes significance at the 99 (95) %-level. A lag length of 1 for both VECMs is selected based on the Schwarz Information Criterion for NCG and TTF, while the same criterion suggests to include 2 lags for NBP.