ePub^WU Institutional Repository

Margarethe Rammerstorfer and Thomas Kremser

Predictive Performance and Bias - Evidence from Natural Gas Markets

Predictive Performance and Bias: Evidence from Natural Gas Markets

Thomas Kremser[1] & Margarethe Rammerstorfer[2]

[1] Finance, Banking and Insurance, Vienna University of Economics and Business, Vienna, The Republic of Austria

[2] Institute for Finance, Banking and Insurance, Vienna University of Economics and Business, Vienna, The Republic of Austria

Correspondence: Thomas Kremser, Finance, Banking and Insurance, Vienna University of Economics and Business, Vienna, The Republic of Austria. E-mail: thomas.kremser@wu.ac.at; margarethe.rammerstorfer@wu.ac.at

Abstract

This paper sheds light on the differences and similarities in natural gas trading at the National Balancing Point in the UK and the Henry Hub located in the US. For this, we analyze traders' expectations and implement a mechanical forecasting model that allows traders to predict future spot prices. Based on this, we compute the deviations between expected and realized spot prices and analyze possible reasons and dependencies with other market variables. Overall, the mechanical predictor performs well, but a small forecast error remains which can not be characterized by the explanatory variables included.

Keywords: natural gas, commodity pricing, unbiasedness, Kalman filtering, risk premium

1. Introduction

The recent developments of financial markets, but also the financial crisis, have raised the question of the efficiency of certain market places and the predictive performance of futures forecasting subsequent spot prices. Especially, for the developing European energy and natural gas market it is an important topic to analyze possibly existing restrictions and market frictions in order to derive possible policy recommendations to stipulate further development. To provide the reader with a broader understanding of differences between developed and a developing natural gas market, we incorporate in our analysis the most liquid European hub for natural gas (National Balancing Point) as well as the most liquid trading place for natural gas located in the US (Henry Hub).

Natural gas futures contracts provide market participants with a risk management tool, in which the buyer of the contract reduces the risk of price fluctuations of the future spot price by locking in a secured price level at the termination of the futures contract. In an efficient commodity market, futures prices deem as appropriate forecast for the future spot price. Hence, the informational content of futures prices is important, as these prices influence the ability of market participants to predict future spot prices (Note 1). A well-known theory addressing this is given by the unbiasedness hypothesis (UH) which states that rational risk neutral agents should come up with perfect forecasts on futures spot prices (see Taylor, 1995). When testing the UH we analyze whether or not futures prices consistently over- or underpredict the future price level of the spot product. In case the UH is rejected, the reasons could be manifold, as for example a biased information set, market frictions given by regulations or an existing risk premium in the market (see for example Kaminsky & Kumar, 1990; Krehbiel & Adkins, 1993; Beck, 1994). Moreover, if the UH does not hold, the bias causes significant costs for hedging and diversification, as shown by Chang (1985) or Roon et al. (2000).

The empirical literature on the unbiasedness of the European natural gas markets is rather scarce in comparison to the US. Empirical studies dealing with the unbiasedness hypothesis and the existence of a risk premium implied in US natural gas markets are given by Movassagh & Modjtahedi (2005), Modjtahedi & Movassagh (2005), or Wei & Zhu (2006). Movassagh & Modjtahedi (2005) find evidence that natural gas futures contracts traded at Henry Hub (HH) are biased predictors of future spot prices at termination for contracts ranging from 3 to 12 months. In a second study (see Modjtahedi & Movassagh, 2005), the authors highlight the fact that spot and futures prices are non-stationary and, thus, conventional econometric techniques cannot be applied to test the UH. Furthermore, Modjtahedi & Movassagh (2005) find that with increasing contract length also the pricing bias

in the futures contract rises which could be an indicator for a time-varying risk premium. For natural gas forward contracts at HH, Wei & Zhu (2006) find that forward contracts are biased predictors of future spot prices and conclude that there exists a time-varying risk premium. For Europe Hobaeck-Haff et al. (2008) find that prices of forward contracts ranging from one to five months and spot prices are cointegrated and that forward prices are biased predictors of the future spot price. Furthermore, they find a significant positive risk premium inherent in UK natural gas forward markets at the National Balancing Point (NBP). The study of Cartea & Williams (2008) investigates UK forward and spot gas prices at NBP and focuses on the interdependency between risk premium and inventory levels. Cartea & Williams (2008) highlight that the short-term market price of risk is mainly determined by deviations from expected seasonal storage levels.

In this line of reasoning but with respect to other markets Switzer and El-Khoury (2007) investigate the efficiency of futures contracts on light sweet crude oil traded at the New York Mercantile Exchange (NYMEX) and find that futures prices are cointegrated with the subsequent spot price. They reveal that the efficiency in this market is time-dependent and improved significantly over time. However, a study conducted by Shambora & Rossiter (2007) rejected the hypothesis of efficient crude oil futures contracts traded at the NYMEX. Moreover, analyzing weekly WTI and Brent crude oil futures and spot prices, Maslyuk & Smyth (2008) find that these markets can be considered as weakform efficient in the sense of Fama (1970). Alvarez-Ramirez et al. (2008) confirm this results and investigate the improvement of efficiency over time and find consistent with Maslyuk & Smyth (2008) that the crude oil markets converged to a state of weak-form efficiency.

Based on these studies, our analysis emphasizes the predictive performance of the futures with different maturities for the future spot price. We find that for both markets the corresponding futures and spot prices share a common stochastic trend, however, the observable futures prices are only biased predictors such that the UH has to be rejected for NBP and HH. In a second step, we implement a mechanical predictor that allows market participants to predict the future spot price and the according risk premium at every day within a futures' maturity and, thus, with increasing information set. Our model provides rather good estimates for the uncertain future spot prices although a certain forecast error remains. In a third step, we analyze possible driving factors for the forecast error between realized and expected spot prices. We find that the forecast error can partly be explained by the included influencing variables.

Overall, the contribution of this article is threefold. First, we contribute to the existing literature by empirically testing the short- and long-run unbiasedness relationship of futures and spot prices traded at NBP and HH. For this, we investigate on the basis of the corresponding natural gas price series the appropriate econometric evaluation process to test whether market participants are endowed with rational expectations and are risk neutral on the short- and long-term. Secondly, we allow market participants to form expectations on the future spot prices with respect to the dynamic behavior of the data set and an inherent risk premium in the markets for natural gas. This requires the implementation of a mechanical prediction technique given by the state-space model (SSM) using the Kalman filter technique. Lastly, we determine the differences between realized and expected future spot prices, also known as forecast errors, and study its relationship to other market variables. To our knowledge we are the first to investigate the stochastical properties of the natural gas futures markets at HH and NBP with an integrated approach.

The remainder of this article is organized as follows. The next section outlines the test procedure for the unbiasedness hypothesis. Section 3 describes the data and the natural gas trading hubs, and highlights the evolution of spot and futures prices. In Section (4) we describe the results of the empirical tests on the unbiasedness hypothesis for the markets considered in this study. In the second part of the paper starting with Section (5), the expected future spot price and the expected risk premium are determined with respect to the Kalman filter forecasting procedure. Furthermore, we outline the empirical methodology to analyze the factors that affect the forecast error and their impact deduced from existing empirical evidence. Finally, Section (7) draws together the main findings and discusses the considerable and significant impact of the empirical results.

2. The Unbiasedness Hypothesis

Under the assumption of unbiased expectations, the information set at time t-1 available to market participants (denoted by Ω_{t-1}) should fully reflect all information about historical and current futures and spot prices, as well as the stochastic price process of the underlying commodity. As a result, the current futures price should be an unbiased predictor of the corresponding future spot price at maturity, such that on average the difference between the price of the spot product at maturity and the price of the corresponding futures contract at time t-1 should be equal to zero. In this context, Hsieh & Kulatilaka (1982), Lai & Lai (1991), and Moosa & Al-Loughani (1994) conclude that the UH is associated with the fact that there exists no strategy from which market participants can benefit continuously by speculating in the futures market on the future spot price. This implies

further all variables included in the information set have to be uncorrelated with the forecast error.

Thus, in line with Geman (2005) the futures-spot-price-relation can be stated by:

$$F_{t-1} = E_Q[S_t|\Omega_{t-1}], \tag{1}$$

with F_{t-1} as the futures price at issuance, S_t as the spot price at the futures' termination, whereas E_Q expresses the expectations operator under the risk-neutral probability measure Q (Note 2). Following this assumption, we can state that:

$$S_t = F_{t-1} + \varepsilon_t, \tag{2}$$

with ε_t representing the rational expectations error term with zero mean and finite variance ($E[\varepsilon_t] = 0$).

Nevertheless, if the assumption of risk-neutrality is violated and market participants are risk averse we have to include a measure of risk for the arising uncertainty. This uncertainty is mainly explained by irrational expectations of market participants or a persistent nonzero risk premium (Note 3). Hence, Equation (1) can be modified to:

$$F_{t-1} = E_Q[S_t|\Omega_{t-1}] - \pi_{t-1}, \tag{3}$$

with π_{t-1} as the risk premium which can be further specified by the following equation (see Moosa & Al-Loughani, 1994):

$$\pi_{t-1} = \mu + \upsilon_{t-1}, \tag{4}$$

with μ as the mean of the risk premium process and υ_{t-1} as the error term with zero mean and finite variance (white noise).

Thus, in line with Cornell (1977), Frenkel & Mussa (1980), Hsieh & Kulatilaka (1982), Crowder & Hamed (1993), Moosa & Al-Loughani (1994), Kellard et al. (1999), McKenzie & Holt (2002), and Haugom & Ullrich (2012), we substitute Equation (4) into (3) and reshuffle terms which allows us to test the UH with the following equation which is in general specified in logarithms (Note 4):

$$s_t = \delta + \gamma f_{t-1} + u_t. \tag{5}$$

Hence, the log futures price (f_{t-1}) is an unbiased predictor of the log of s_t if the joint hypothesis that $\delta = 0$ and $\gamma = 1$ holds. Additionally, the futures contract can also be considered as an efficient predictor if the error term does not contain price sensitive information (Note 5).

In line with Elam & Dixon (1988), several studies identify that financial price series are non-stationary, such that a standard F-Test is not appropriate to examine whether the UH is accepted or rejected. As a consequence, a stationarity test has to be conducted to determine the order of integration of the different price series. In case the price series are stationary, the UH could be verified by running a standard OLS-regression on Equation (5) (Note 6). Under absence of stationarity, the cointegration relations between spot and futures prices has to be analyzed (Note 7). In this case the Johansen Full Information Maximum Likelihood approach (see Johansen, 1996) can be applied to test for cointegration (longrun relationship between variables) which is based on the Vector Error Correction model framework (VECM) of a Vector Autoregressive model (VAR) and takes the following generalized form:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t, \tag{6}$$

where y_t represents the *m*-vector of endogenous (non-stationary) variables of *k* lags, Δ corresponds to the first difference term, and the matrix Π highlights the long-run dynamics of the process y_t. Furthermore, Γ_i captures the short-run dynamics, and ε_t corresponds to the i.i.d. Gaussian error term (Note 8).

If there is no strong evidence against a long-run equilibrium, we are able to test also a short-run regression relating the change in spot price to the cointegration relation and the lagged changes in spot and futures prices:

$$\Delta s_t = \alpha(u_{t-1}) + \sum_{k=1}^{m} \lambda_k \Delta f_{t-k} + \sum_{j=1}^{m} \rho_j \Delta s_{t-j} + \varepsilon_t, \tag{7}$$

where Δs_t represents the first difference operator of the spot price series, u_{t-1} the cointegrating vector, and Δf_{t-k} as well as Δs_{t-j} the lags of changes in the futures and spot prices.

Nevertheless, this type of analysis cannot be applied to estimate the short-run relationship between two series if the data are sampled more frequently than the length of a certain futures contract. As this is the case for the sample used in this article, we have to shorten the series significantly, such that we end up with 12 observations per year for a futures with 1 month maturity, 6 observations for a futures with 2 months maturity, and 4 observations for the 3 months maturity futures contracts. For this the notationallycorrected cointegration regression becomes:

$$s_t = \delta + \gamma f_{t-\tau} + u_t. \tag{8}$$

Herein, we match the futures price at initiation $(t - \tau)$ with the spot price at the futures final termination (t). To clarify the approach taken here, we have to understand that the lag suggested by Equation (5) is different for each futures considered and depending on the futures contract length. Based on this, we also have to rewrite Equation (7) in order to end with a quasi-VEC model as suggested by Kellard et al. (1999):

$$\Delta s_t = \alpha(u_{t-\tau}) + \sum_{k=1}^{m} \lambda_k \Delta f_{t-k} + \sum_{j=1}^{m} \rho_j \Delta s_{t-j} + \varepsilon_t, \tag{9}$$

which can be further solved for S_t:

$$S_t = (1 + \alpha)S_{t-\tau} + (\lambda_1 - \alpha\gamma)f_{t-\tau-1} + \alpha\delta + \lambda_1 f_{t-\tau} + \sum_{k=2}^{m} \lambda_k \Delta f_{t-\tau-k} + \sum_{j=1}^{m} \rho_j \Delta s_{t-j} + \varepsilon_t. \tag{10}$$

Kellard et al. (1999) show that if there exists a long-run equilibrium, it is permissible to run an additional regression on the lagged changes in spot and futures prices as given by Equation (9) in which α, λ_k and ρ_j are employed to test whether the markets are unbiased in the short-run. For this the following restrictions have to hold: $\alpha = -1$ as well as $\alpha\gamma = \lambda_1 \neq 0$ and $\lambda_k = \rho_j = 0 \ \forall k \geq 2$ and $j \geq 1$ on Equation (10) (Note 9). The adjustment parameter α indicates the speed of adjustment to the long-run equilibrium and implies that any disequilibrium reverts back to the long-run equilibrium. In case $\alpha = -1$ we observe an immediate reversion back to the long-run relationship between the price series. λ_1 represents the coefficient on the first difference of the futures price and if $\lambda_1 \neq 0$ any change in the futures price is reflected in Δs_t (Note 10). Furthermore, the coefficients λ_k and ρ_j define the significant lags ($k \geq 2$ and $j \geq 1$) on the difference operators (Δ) of the corresponding spot and futures price series included in the model. The latter constraint indicates that all past information should be incorporated in the first difference of the futures price and thus all significant lags are equal to zero.

3. Hubs and Data

The comparison of NBP and HH is motivated by the historical different evolution of these trading places for natural gas. In general, in Europe, the wholesale market for natural gas has a much shorter history than in the US, as its market opening traces back to an EU Directive 98/30/EC established in 1998. Before that, regulation in the EU member states was solely driven by national authorities. Consequently, the European wholesale market for natural gas is still developing. Nevertheless, only a small fraction of the gas consumed in Europe is traded via these hubs, which also explains the lower liquidity and churn rate at the European trading places when compared to the US. As a consequence, we base our study on the National Balancing Point which is the most liquid hub in Europe and a centralized point for natural gas spot and futures trading at the Intercontinental Exchange (ICE). According to a study by Prospex Research Ltd, approximately 80% of Europe's gas trading in 2010 took place in the UK. In North America, regulation started in 1985 with FERC Order No. 436, which enabled distribution companies to purchase gas directly from producers and separately pay the pipeline companies to transport the gas. This order also provided the regulatory framework for the formation of several trading points in the US. In a second step of market liberalization, FERC Order No. 636 was passed in 1992. Here, the lack of liquidity at the trading points lies in the center of the market adjustments. Due to this order, network operators are obliged to unbundle their sales and transport services to support the development of liquid trading places. One of these is the so-called Henry Hub in Louisiana, which was established in 1988. It is operated by Sabine Pipe, a subsidiary of Chevron Texaco, and connects nine states and four intrastate pipelines. HH is the largest centralized point for natural gas spot and futures trading at the NYMEX.

Overall, our data set comprises of spot and futures prices with maturities from one to three month, traded at ICE or NYMEX, respectively. The futures time series are generic and constructed by different consecutive futures contracts (Note 11).The data consists of end of day mid-quotes provided by Bloomberg. At NBP natural gas is traded in GBP/therm whereas in the United States the trading unit is given by USD/mmBtu. Due to data availability constraints and in order to facilitate the comparison of results between NBP and HH, the sample period spans from March 1997 to November 2012.

3.1 National Balancing Point—Preliminary Analysis

Figure 1 shows the evolution of spot and futures prices traded at NBP. During the first and second quarter of the gas year 2006, two extraordinary peaks can be observed (Note 12). These peaks (on November 22nd, 2005 and March 13th, 2006) coincide with shortages in the Norwegian gas fields, accompanied by low temperatures across Europe at that time (see for example Neumann et al., 2008). While both, spot and futures prices show the first peak, the second peak is only observable in the spot price series. Moreover, the impact of the financial crisis of 2008 is also evident, as this event led to a substantial decrease in natural gas prices. Apart from that, price differentials between spot and futures prices increase significantly around the beginning of each gas year.
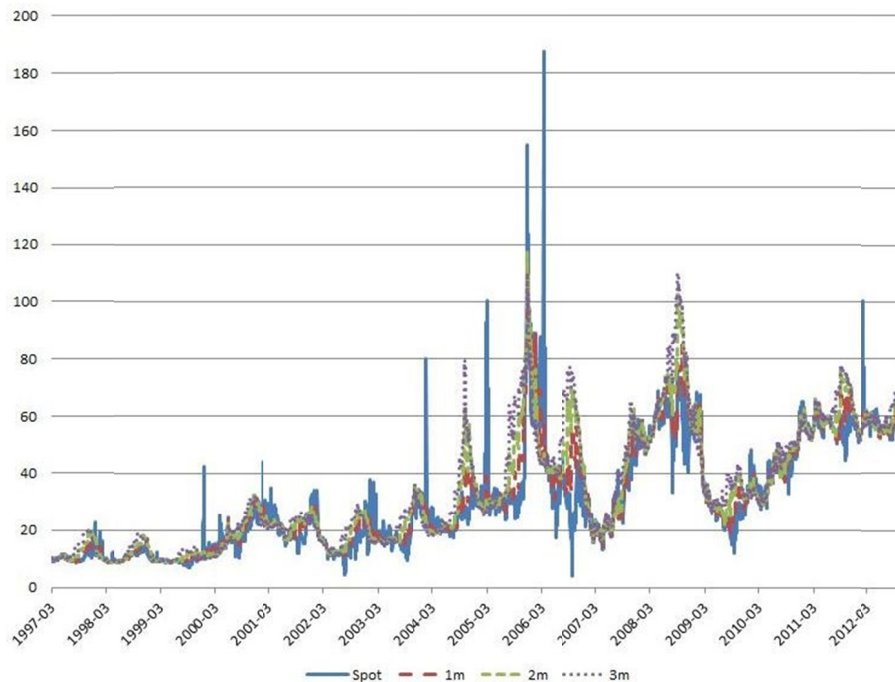
Figure 1. Evolution of spot and futures prices at NBP

Note. Evolution of spot and futures prices traded at NBP in pence per therm (*GBp/therm*). The time horizon considered ranges from March 1997 to November 2012.

Figure 1 also reveals that, on average, futures prices exceed spot prices, indicating that the market is in contango. For the 1 month futures rates, this fact is observed for 56.5% of the observations, while for futures contracts with 2 and 3 months maturities, the corresponding values are 63.2% and 63.6%. Nevertheless, for all maturities also shorter time periods of normal backwardation are observable (Note 13). Obviously and also mentioned in several articles as for example Asche et al. (2009) or Koenig (2012), the NBP price series seems to be subject to several structural breaks (Note 14). Thus, for our upfront test whether the UH holds for NBP, we divide our sample into three sub-samples. The first sample ranges from 1997 to July 9th, 2004 (Phase 1), the date at which BactonBalgzand Company was established aiming to design, construct, operate and exploit a pipeline that connects Balgzand (in the Netherlands) to Bacton (UK) (see operator website for further details). The second sample covers the period in which spot and futures prices drift far apart or show extraordinary peaks which were formerly ascribed to shortages in the Norwegian gas fields. Moreover, within this time period there is a shift from rather low prices for natural gas with low volatilities to higher price levels that are also subject to higher fluctuations. For the analysis of the predictive performance of the 1 to 3 months futures we exclude this part of the series (Note 15). The third sample (Phase 2) becomes evident with the established connection that directly allows a transfer of natural gas from the Netherlands to UK via the BactonBalgzand pipeline which started operation with the December 1st, in 2006.

3.2 Henry Hub—Preliminary Analysis

Figure 2 shows the evolution of spot and futures prices traded at HH with maturities from one to three months. In late 2000, the impact of the Californian Energy Crisis is observable as this event led to a spike in natural gas prices. Similar to NBP the impact of the financial crisis in 2008 lead to a decrease in natural gas prices at HH. Overall, we see a close relationship between spot and futures prices, which implies that the prices at HH are rather integrated (across maturities). In line with the NBP, futures prices exceed spot prices, on average. However, also times of normal backwardation are observable. In general, we find rather similar results for Europe and the USA. For 37.1% of the observations the spot prices exceed the futures prices, when considering the 1 month futures. The corresponding values for the 2 and 3 months futures contracts are 23.4% and 18.0%, respectively.
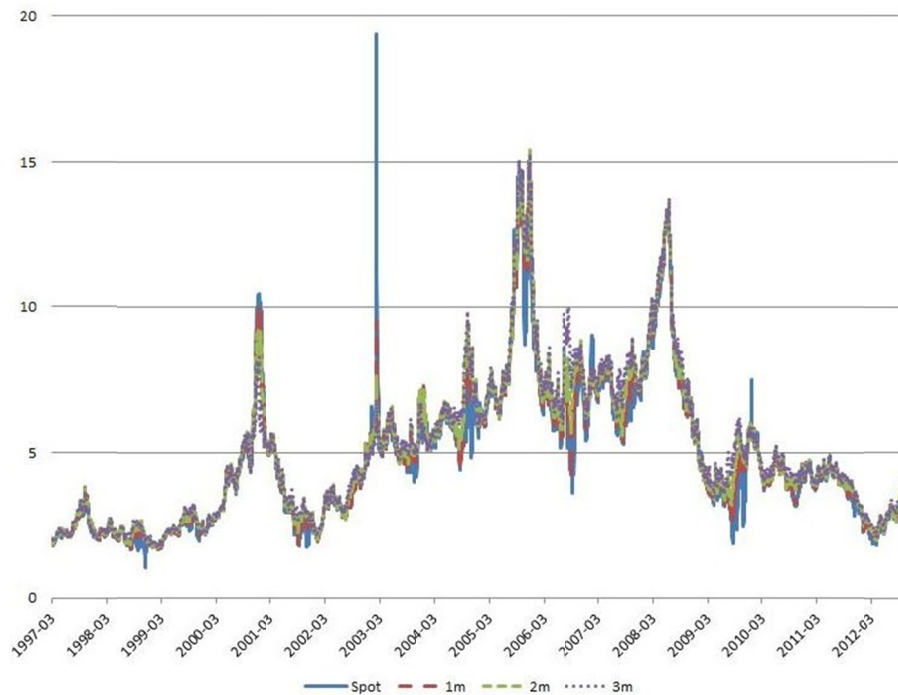
Figure 2. Evolution of spot and futures prices at HH

Note. Evolution of spot and futures prices traded at HH in USD per million British thermal units (*USD/mmBtu*). The time horizon considered ranges from March 1997 to November 2012.

3.3 Time Series Properties and Descriptive Results

In a next step, we determine the descriptive statistics (see Table 1 and 2) for all considered spot and futures prices observable at NBP and HH (Note 16).

Table 1. Descriptive statistics—NBP

NBP	*Spot*	*F*1*m*	*F*2*m*	*F*3*m*
Nrobs	3879	3879	3879	3879
Mean	3.2088	3.2492	3.2952	3.3303
Median	3.2288	3.2580	3.2970	3.3320
Maximum	5.2338	4.7560	4.7660	4.7040
Minimum	1.4351	2.1440	2.1350	2.1310
Std. Dev.	0.6288	0.6209	0.6350	0.6437
Skewness	0.0185	-0.0407	-0.0619	-0.0738
Kurtosis	2.0522	1.9334	1.9348	1.9531
Excess Kurtosis	-0.9478	-1.0666	-1.0652	-1.0469
Jarque-Bera	145.4025	184.9494	185.8637	180.6689

Note. This table provides the descriptives for log spot prices, as well as log prices for one to three month futures contracts traded at NBP. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

Table 2. Descriptive statistics—HH

HH	Spot	F1m	F2m	F3m
Nrobs	3900	3900	3900	3900
Mean	1.4511	1.4715	1.4983	1.5185
Median	1.4667	1.4789	1.4899	1.5088
Maximum	2.9642	2.7329	2.7361	2.7270
Minimum	0.0296	0.4874	0.5074	0.5283
Std. Dev.	0.4929	0.4877	0.4855	0.4839
Skewness	0.0849	0.0811	0.0656	0.0628
Kurtosis	2.2641	2.2130	2.2181	2.2265
Excess Kurtosis	-0.7359	-0.7870	-0.7819	-0.7735
Jarque-Bera	92.6840	104.9178	102.1367	99.7889

Note. This table provides the descriptives for log spot prices, as well as log prices for one to three month futures contracts traded at HH. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

As illustrated by Table 1 and 2, we see that as time to maturity increases, the mean increases whereas the volatility evolves rather stable. The minimum and maximum values for Europe are higher than for the US, which also results in higher volatility for the European market. We observe a negative excess kurtosis in both series considered, indicating a platykurtotic price series distribution. In addition, the descriptive statistics show that the skewness is positive for HH and mostly negative for NBP (HH shows longer right and NBP longer left tails). This indicates that the probability of observing extreme price levels is higher than in comparison to the normal distribution. This finding is also confirmed by the Jarque-Bera test statistic.

Many financial time series exhibit a non-stationary behavior and, thus, unit root or stationarity tests have to be applied. Thus, testing for the order of integration is the crucial step to determine the appropriate econometric model for the considered financial time series. In line with Siliverstovs et al. (2005) we implement the Augmented Dickey-Fuller test (ADF), the Phillips-Perron test (PP) as well as the Kwiatkowski-Phillips-SchmidtShin test (KPSS) to determine the order of integration of the considered natural gas price series at HH and NBP (Note 17). Moreover, the joint implementation of these tests is done to draw conclusions on the robustness of our results (Note 18). Table 3 reports the result of the ADF, PP and KPSS test in levels and first differences (Note 19).

Table 3. ADF, PP and KPSS tests on natural gas price series

Hub	Series	Stationarity Testing					
		Levels			First Differences		
		ADF	PP	KPSS	ADF	PP	KPSS
NBP	F_{1m}	-2.3937	-1.9639	8.5402***	-10.4261***	-59.1404***	0.0176
	F_{2m}	-2.7143	-1.9124	8.4198***	-8.9062***	-59.5193***	0.0285
	F_{3m}	-2.7370	-1.9452	8.3237***	-9.0151***	-57.3551***	0.0302
	Spot	-2.3899	-3.0348**	8.5402***	-19.9486***	-82.1516***	0.0176
HH	F_{1m}	-2.4883	-2.4795	3.6672***	-65.8247***	-65.8699***	0.0999
	F_{2m}	-2.3484	-2.3718	3.8586***	-65.7597***	-65.7544***	0.1152
	F_{3m}	-2.2793	-2.2617	4.1274***	-43.6987***	-64.7378***	0.1378
	Spot	-2.2608	-2.7854	3.5621***	-14.1182***	-61.7484***	0.0836

Note. This table provides the results from the ADF, PP and KPSS test. The 5% significance level is indicated by **; *** indicate the 1% significance level. The ADF, PP and KPSS test in levels and first differences is performed with an intercept but not with a time trend.

The null hypothesis that the series have a unit root (ADF and PP test) cannot be rejected for any price series in levels. Taking first differences leads to a rejection of the null hypothesis for all price series, which argues in favor of an integration of order 1. A similar picture occurs when we consider the KPSS test, where the null hypothesis that the gas price series are integrated with order 0 (stationarity) is rejected for all maturities and hubs. Hence, also for the KPSS test the corresponding gas price series are difference stationary and, thus, integrated with order 1 in levels. This finding is in line with Pindyck (1993) and other seminal papers and, thus, constitute a

stylized fact of energy spot and futures data.

4. Empirical Results on the Unbiasedness of Futures Prices

The results of the unit root and stationary tests (as shown in Table 3) suggest that the UH in Equation (8) cannot be tested by applying a standard OLS-regression (Note 20). Thus, we have to follow the test procedure introduced in Section (2). First, a test on the number of cointegration relations is performed to determine whether natural gas price series are cointegrated or not (Note 21).In case price series share a common stochastic trend i.e. are cointegrated with at most one cointegrating relation ($r \leq 1$), we proceed with step 2 in which we obtain the coefficients $\delta$and $\gamma$from the corresponding normalized cointegrating vector (β). In a final step, we impose restrictions on the cointegrating vector and the remaining variables of the VEC model to provide statistical inference on the question whether the spot and futures contracts traded at NBP or HH are unbiased in the short-and long-run.

The results of the cointegration test (based on the Johansen procedure) are reported in Tables (4) to (6) (Note 22).

Table 4. Cointegration rank test for NBP—Phase 1

Series	Lag Length	Hypothesized No. of CE(s) $H_0^a: r = 0$	Max-Eigenvalue Statistic	Hypothesized No. of CE(s) $H_0^b: r \leq 1$	Max-Eigenvalue Statistic
$Spot_{1m}|F_{1m}$	1	None	36.5843***	At most 1	6.1150
$Spot_{2m}|F_{2m}$	1	None	20.0661**	At most 1	5.5593
$Spot_{3m}|F_{3m}$	1	None	16.6240***	At most 1	3.8290

Note. ** indicate the rejection of the null hypothesis at the 5% significance level, and *** at the 1% significance level. H_0^a reports the max-eigenvalue test statistic for the null hypothesis of no cointegration (r = 0). H_0^b reports the max-eigenvalue test statistic for the null hypothesis of at most one cointegration ($r \leq 1$). The optimal lag length is chosen by the SC.

Table 5. Cointegration rank test for NBP—Phase 2

Series	Lag Length	Hypothesized No. of CE(s) $H_0^a: r = 0$	Max-Eigenvalue Statistic	Hypothesized No. of CE(s) $H_0^b: r \leq 1$	Max-Eigenvalue Statistic
$Spot_{1m}|F_{1m}$	1	None	22.6635***	At most 1	3.8368
$Spot_{2m}|F_{2m}$	2	None	33.1495***	At most 1	9.6418**
$Spot_{3m}|F_{3m}$	2	None	19.4293**	At most 1	13.1901***
$Spot_{2m}|F_{2m}$	6	None	16.8216***	At most 1	6.4186
$Spot_{3m}|F_{3m}$	5	None	17.6015***	At most 1	2.5219

Note. ** indicate the rejection of the null hypothesis at the 5% significance level, and *** at the 1% significance level. H_0^a reports the max-eigenvalue test statistic for the null hypothesis of no cointegration (r = 0). H_0^b reports the max-eigenvalue test statistic for the null hypothesis of at most one cointegration ($r \leq 1$). The optimal lag length is chosen by the SC.

As indicated earlier, testing for cointegration at NBP initially requires the split of the spot and futures price series into Phase 1 and Phase 2 (Note 23). The analysis of the Phase 1 indicates that the null hypothesis of no cointegration ($H_0^a: r = 0$) can be rejected for all price series, i.e., s_t and $f_{t-\tau}$ are cointegrated with rank 1. For Phase 2 a slightly different picture occurs. Here only the 1 month relation between $s_t$and $f_{t-\tau}$ is cointegrated at most with rank 1, whereas for longer maturities the null hypothesis of no cointegration and the null hypothesis of cointegration with at most rank 1 is rejected (Note 24). Kellard et al. (1999) identify a similar cointegration relation for live hogs. They state that the SC does on occasion, produce an overly parsimonious approximation of the underlying data process and, therefore, the analysis should be carried out in line with the lag length chosen by the unit root test. For 2 and 3 months maturities at NBP the results of the VEC model are sensitive to the chosen lag length. With a lag length of 6 or 5 the null hypothesis of cointegration with at most rank 1 is accepted for the 2 and 3 months spot-futuresrelation. Hence, we base our analysis of the cointegration relation between the spot and futures price series for Phase 2 on the chosen lag length of the ADF test (see Kellard et al., 1999).

Table 6. Cointegration rank test for HH

Series	Lag Length	Hypothesized No. of CE(s) $H_0^a : r = 0$	Max-Eigenvalue Statistic	Hypothesized No. of CE(s) $H_0^b : r \leq 1$	Max-Eigenvalue Statistic
$Spot_{1m}|F_{1m}$	1	None	88.0227***	At most 1	5.3632
$Spot_{2m}|F_{2m}$	1	None	41.3472***	At most 1	5.2043
$Spot_{3m}|F_{3m}$	1	None	38.9961***	At most 1	4.5332

Note. ** indicate the rejection of the null hypothesis at the 5% significance level, and *** at the 1% significance level. H_0^a reports the max-eigenvalue test statistic for the null hypothesis of no cointegration (r = 0). H_0^b reports the max-eigenvalue test statistic for the null hypothesis of at most one cointegration ($r \leq 1$). The optimal lag length is chosen by the SC.

For all natural gas price series at Henry Hub the analysis of the results indicates that the null hypothesis of no cointegration ($H_0^a : r = 0$) can be rejected. Furthermore, the test statistics suggest that s_t and $f_{t-\tau}$ are cointegrated at most with rank 1. This implies that the two variables never drift far apart, which is a necessary condition for the UH to hold in the long-run. However, the cointegration between spot and futures prices becomes weaker as the time to maturity increases (Note 25).

Based on the former results, we are able to investigate whether the long-run hypothesis of unbiasedness ($\delta = 0$ and $\gamma = 1$) and risk neutrality holds. The results of this test on basis of a likelihood ratio test statistic are reported in Tables (7) to (9).

Table 7. Restriction test on Johansen's cointegrating coefficients for NBP—Phase 1

Series	δ	γ	$H_0^c : \delta = 0$ and $\gamma = 1$	$H_0^d : \gamma = 1$
$Spot_{1m}|F_{1m}$	0.0218	-0.9910	5.4172	0.0236
	(0.1422)	(0.0515)	[0.0666]	[0.8779]
$Spot_{2m}|F_{2m}$	0.3797	-1.1107	5.7089	0.9578
	(0.2552)	(0.0920)	[0.0576]	[0.3277]
$Spot_{3m}|F_{3m}$	0.7239	-1.2406	3.6053	2.9824
	(0.3450)	(0.1241)	[0.0575]	[0.0842]

Note. δ and γ are the normalized cointegrating parameters from the Vector Error Correction model (standard errors are reported in parenthesis). The Likelihood Ratio (LR) test statistics for testing the joint hypothesis of $\delta = 0$ and $\gamma = 1 (H_0^c)$, and $\gamma = 1$ (H_0^d) have a chi-square distribution with two or one degrees of freedom, respectively. The p-values of the LR test are reported in square brackets.

Table 8. Restriction test on Johansen's cointegrating coefficients for NBP—Phase 2

Series		δ	γ	$H_0^c : \delta = 0$ and $\gamma = 1$	$H_0^d : \gamma = 1$
$Spot_{1m}|F_{1m}$		0.4520	-1.1061	3.8134	1.4635
		(0.3061)	(0.0808)	[0.1486]	[0.2264]
$Spot_{2m}|F_{2m}$	(6 lags)	0.7007	-1.1610	14.3920	9.9362
		(0.5813)	(0.1528)	[0.0008]	[0.0016]
$Spot_{3m}|F_{3m}$	(5 lags)	2.8897	-1.7086	15.1239	10.1709
		(1.0033)	(0.2636)	[0.0005]	[0.0014]

Note. δ and γ are the normalized cointegrating parameters from the Vector Error Correction model (standard errors are reported in parenthesis). The Likelihood Ratio (LR) test statistics for testing the joint hypothesis of $\delta = 0$ and $\gamma = 1 (H_0^c)$, and $\gamma = 1$ (H_0^d) have a chi-square distribution with two or one degrees of freedom, respectively. The p-values of the LR test are reported in square brackets.

For NBP, the null hypothesis (H_0^d) that $\gamma = 1$ is accepted for the 1 and 2 months maturity (sub-sample prior the break in 2004) and after the second break for the 1 month futures contract, whereas the joint hypothesis of unbiasedness and risk neutrality ($\delta = 0$ and $\gamma = 1$, H_0^c) is rejected for all futures-spot-relations in Phase 1 and accepted for the 1 month futures contract in Phase 2. However, the small number of observations and the rather low p-value of the according hypothesis test (H_0^c) for Phase 2, does not allow to draw a final conclusion on the overall absence of a risk premium or the unbiasedness of the 1 month futures contract. In addition, these results indicate that 1 and 2 months futures contracts during Phase 1 are biased predictors, while for 3 months futures (Phase 1) as well as 2 and 3 months (Phase 2) contracts the restrictions on the slope coefficient (H_0^d) is rejected such that no long-run biasedness relationship can be identified. Overall, this finding is in line with Hobaeck-Haff et al. (2008) who identified a time-varying risk premium in the UK natural gas forward markets.

Table 9. Restriction test on Johansen's cointegrating coefficients for HH

Series	δ	γ	$H_0^c: \delta = 0$ and $\gamma = 1$	$H_0^d: \gamma = 1$
$Spot1m|F1m$	0.0600	-1.0172	19.3037	1.2106
	(0.0235)	(0.0152)	[0.0001]	[0.2712]
$Spot2m|F2m$	0.0343	-0.9835	11.2039	0.2123
	(0.0530)	(0.0335)	[0.0037]	[0.6450]
$Spot3m|F3m$	0.0292	-0.9723	16.7756	0.7614
	(0.0475)	(0.0297)	[0.0002]	[0.3829]

Note. δ and γ are the normalized cointegrating parameters from the Vector Error Correction model (standard errors are reported in parenthesis). The Likelihood Ratio (LR) test statistics for testing the joint hypothesis of $\delta = 0$ and $\gamma = 1 (H_0^c)$, and $\gamma = 1$ (H_0^d) have a chi-square distribution with two or one degrees of freedom, respectively. The p-values of the LR test are reported in square brackets.

The null hypothesis (H_0^d) that $\gamma = 1$ is cannot be rejected for HH at the 1% level. Moreover, testing for $\delta = 0$ and $\gamma = 1 (H_0^c)$, simultaneously, indicates the rejection of the null hypothesis for all maturities considered. This finding implies that futures prices are solely biased predictors of the future spot price.

The characterized long-run equilibrium at NBP and HH enables us to identify the short-run unbiasedness for the considered futures and spot prices by a quasi-VECM setting (see Kellard et al., 1999). We present the results of the short-term test on a pricing bias and inefficiencies in Tables (10) and (13) (Note 26).

Table 10. Quasi-vector error correction model for NBP

	Phase 1		Phase 2
	$Spot_{1m}|F_{1m}$	$Spot_{2m}|F_{2m}$	$Spot_{1m}|F_{1m}$
α	-0.8679	-0.5436	-0.1569
	(-5.0665)	(-4.0622)	(-1.4522)
λ_1	0.8905	0.6758	0.3343
	(8.3567)	(3.5781)	(2.4717)
λ_2	-0.2052	-0.0334	0.4200
	(-1.2724)	(-0.2969)	(3.0008)
λ_3	-0.1440	0.1128	0.1781
	(-1.3309)	(1.0512)	(1.9517)
ρ_1	0.0171	-0.1581	-0.2317
	(0.1621)	(-3.3532)	(-2.9673)
ρ_2	-0.0437	-0.1627	-0.2404
	(-0.4721)	(-1.2741)	(-2.5573)
ρ_3	0.0280	-0.1433	-0.0801
	(0.3077)	(-1.4284)	(-1.1050)
R^2	0.5039	0.4973	0.1758
DW stats	1.8896	2.1117	2.0582
ARCH(8)	0.9710	0.1589	0.9550
Q(8)	0.6131	0.7580	0.9880

Note. α is the speed of adjustment back to the long-run equilibrium level. λ_1 is the short-run relationship on the first difference of the futures price. λ_k and ρ_j defines the lags ($k = j > 1$) on the difference operators (Δ) of the corresponding spot and futures price. The t-statistic of α and λ_1 are reported in parenthesis. Q() represents the Ljung-Box test statistic for serial correlation in the residuals. ARCH() denotes the statistical test for heteroskedasticity in residuals.

The test results for NBP presented in Table (10) clearly indicate that the adjustment term α is significant (at least for the 1 and 2 months relationship in Phase1) for the considered maturities and exert the expected negative sign. This is in line with the findings from the cointegration test in Table (4) and highlights a mechanism of mean-reversion in the futures-spot-relationship for the considered maturities. In addition, we observe that certain lags of λ_k and ρ_j are significant which clearly indicates that past information is not completely reflected by current price differentials (Δ) and that these lags should be incorporated in order to improve the model. Thus, the forecasting of futures spot prices is improved by using past futures and spot price information. The estimation results of λ_1 highlight a significant short-run relationship with the according expected positive sign. A positive magnitude of λ_1 highlights that spot markets are systematically led by the futures market (and vice versa for a negative sign). This lead-lag relationship between futures and spot markets could be due to the fact that trading in futures contracts is less expensive than compared to the spot markets and, thus, price discovery and processing of information takes place faster in the futures market. Furthermore, we applied a Ljung-Box test to account for

serial correlation patterns and an ARCH LM test to control for potential heteroskedasticity in the residuals. As our results suggest, the selected model captures the dynamics of the changes the spot price (Δs_t) for all considered short-run relationships quite well.

Table 11. Restriction test on the quasi-vector error correction model for NBP

Series		$H_0^e: \alpha = -1$	$H_0^{f1}: \alpha\gamma = \lambda_1$ $H_0^{f2}: \lambda_1 = 1$ $H_0^{f3}: \lambda_1 = 0$	$H_0^g: \lambda_k = \rho_j = 0$	$H_0^h: \alpha = -1,$ $\alpha\gamma = \lambda_1$ and $\lambda_k = \rho_j = 0$
$Spot_{1m}\|F_{1m}$	(Phase 1)	0.4407	0.7758 0.3039 0.0000	0.0351	0.0005
$Spot_{2m}\|F_{2m}$	(Phase 1)	0.0006	0.7031 0.0860 0.0003	0.0000	0.0000
$Spot_{1m}\|F_{1m}$(Phase 2)		0.0000	0.2347 0.0000 0.0134	0.0000	0.0000

Note. The p-values of the Wald test statistics for testing the hypothesis of $\alpha = -1$ (H_0^e), $\alpha\gamma = \lambda_1$ (H_0^{f1}), $\lambda_1 = 1$ (H_0^{f2}), $\lambda_1 = 0$ (H_0^{f3}), $\lambda_k = \rho_j = 0$ (H_0^g), and $\alpha = -1, \alpha\gamma = \lambda_1$ and $\lambda_k = \rho_j = 0$ (H_0^h) have a chi-square distribution with *m* degrees of freedom. One asterisk indicates the acceptance of the null hypothesis (H_0^e, H_0^{f1}, H_0^{f2}, H_0^{f3}, H_0^g, H_0^h) at the 10% significance level, whereas two and three asterisks indicate the acceptance at the 5% and 1% significance level.

The test results of H_0^h (hypothesis of efficiency and risk neutrality) are reported in Table (11). For all considered contracts and Phases the null hypothesis of efficiency and risk neutrality is strongly rejected. This finding suggests that markets participants require a risk premium in these markets (see El HediArouria et al., 2013).

Table 12. Quasi-vector error correction model for HH

		$Spot_{1m}\|F_{1m}$	$Spot_{2m}\|F_{2m}$	$Spot_{3m}\|F_{3m}$
α		-0.4634	-0.8507	-0.4421
		(-1.6919)	(-2.6000)	(-1.0853)
λ_1		0.9131	0.8288	0.1312
		(6.0573)	(3.0556)	(0.3621)
λ_2		-0.2173	-0.3464	0.0925
		(-0.6648)	(-1.1524)	(0.2961)
λ_3		0.03786	-0.0736	-0.0647
		(0.1161)	(-0.2426)	(-0.3401)
λ_4		0.0343	-	-
		(0.2338)		
λ_5		0.0785	-	-
		(0.3271)		
λ_6		0.1291	-	-
		(0.5408)		
λ_7		-0.0226	-	-
		(-0.0928)		
λ_8		0.1198	-	-
		(1.2378)		
λ_9		-0.2595	-	-
		(-1.1856)		
ρ_1		-0.4602	-0.0143	0.2306
		(-1.3173)	(-0.0513)	(1.2324)
ρ_2		0.0422	0.3788	-0.0182
		(0.1479)	(1.5081)	(-0.0902)
ρ_3		0.1710	0.1068	-0.1402
		(0.5547)	(0.5098)	(-0.7991)
ρ_4		-0.0091	-	-
		(-0.0653)		
ρ_5		-0.1176	-	-
		(-0.7195)		

ρ_6	-0.1079	-	-
	(-0.5616)		
ρ_7	0.0628	-	-
	(0.4661)		
ρ_8	-0.0600	-	-
	(-0.7503)		
ρ_9	-0.0064	-	-
	(-0.0403)		
R^2	0.2263	0.1403	0.0863
DW stats	1.9917	2.0336	1.9504
ARCH(10)	0.5663	0.9649	0.3580
Q(10)	0.9960	0.8780	0.8340

Note. α is the speed of adjustment back to the long-run equilibrium level. λ_1 is the short-run relationship on the first difference of the futures price. λ_k and ρ_j defines the lags ($k = j > 1$) on the difference operators (Δ) of the corresponding spot and futures price. The t-statistic of α and λ_1 are reported in parenthesis. Q() represents the Ljung-Box test statistic for serial correlation in the residuals. ARCH() denotes the statistical test for heteroskedasticity in residuals.

Table (12) presents the test results of the short-run quasi-VECM for HH. As in the case of the NBP the speed of adjustment parameter α is significant (at least for the 1 and 2 months relationship) and shows the expected negative sign. This result confirms the finding of the cointegration test in Table (6). In line with NBP, the results of λ_1 strongly indicate that again spot markets are systematically led by the futures market. However, for the 3 months futures-spot-relation we observe that α and λ_1 are insignificant. The test statistics of the ARCH LM and the Ljung-Box test highlight that the chosen model setup (lags of λ_k and ρ_j) fits the dynamics of the changes the spot price (Δs_t) for all considered short-run relationships.

Table 13. Restriction test on vector error correction model for HH

Series	$H_0^e: \alpha = -1$	$H_0^{f1}: \alpha\gamma = \lambda_1$ $H_0^{f2}: \lambda_1 = 1$ $H_0^{f3}: \lambda_1 = 0$	$H_0^g: \lambda_k = \rho_j = 0$	$H_0^h: \alpha = -1,$ $\alpha\gamma = \lambda_1 \ and \ \lambda_k = \rho_j = 0$
$Spot_{1m}|F_{1m}$	0.0501	0.0034 0.5643 0.0000	0.0000	0.0000
$Spot_{2m}|F_{2m}$	0.6482	0.9770 0.5280 0.0022	0.1045	0.2303
$Spot_{3m}|F_{3m}$	0.1708	0.4096 0.0165 0.7173	0.0000	0.0000

Note. The p-values of the Wald test statistics for testing the hypothesis of $\alpha = -1$ (H_0^e), $\alpha\gamma = \lambda_1$ (H_0^{f1}), $\lambda_1 = 1$ (H_0^{f2}), $\lambda_1 = 0$ (H_0^{f3}), $\lambda_k = \rho_j = 0$ (H_0^g), and $\alpha = -1, \alpha\gamma = \lambda_1 \ and \ \lambda_k = \rho_j = 0$ (H_0^h) have a chi-square distribution with *m* degrees of freedom. One asterisk indicates the acceptance of the null hypothesis (H_0^e, H_0^{f1}, H_0^{f2}, H_0^{f3}, H_0^g, H_0^h) at the 10% significance level, whereas two and three asterisks indicate the acceptance at the 5% and 1% significance level.

Table (13) reports the results of H_0^h. For the 1 and 3 months short-run relationship our findings are in line with the long-run estimates, i.e., the null hypothesis of unbiasedness and risk neutrality is strongly rejected. In line with NBP, this signals that markets participants ask for a risk premium. However, for the 2 month futures-spot-relation the hypothesis of efficiency and risk neutrality is accepted (p-value = 0.2303).

Concluding, our findings for the long- and short-run UH for the considered futures and spot prices at NBP and HH reveal several interesting facts. First, futures prices and the subsequent spot prices can be considered as cointegrated which reveals that both follow a common stochastic trend, and, thus are in a long-run equilibrium relationship. Second, futures prices are biased predictors in the long- and short-run and, thus, the joint hypothesis of unbiasedness and risk neutrality has to be rejected. This finding is in line with several seminal studies

Movassagh & Modjtahedi (2005); Modjtahedi & Movassagh (2005); Wei & Zhu (2006); Cartea & Williams (2008); Hobaeck-Haff et al. (2008). Third, we highlight that past information is relevant to forecast future prices, and, thus, the information incorporated in futures contracts can be used to speculate on future spot price levels.

5. Predicting Future Spot Prices

Based on the former results we implement a mechanical predictor to determine the future spot price and according risk premium. For this, we allow the market participants to use all information provided within the market at a certain date which implies that we allow the parameters to vary over time.

In the following section we proxy the risk premium via the Kalman filter technique, which enables us to replace the full rational expectations assumption by the incomplete rational expectations setup. In a second step, the forecast error—the differences between the expected and realized future spot price—is determined and tested for its influencing factors and correlations with other market variables (Note 27).

5.1 Future Spot Price Modeling

In order to predict the uncertain future spot price and the unknown risk premium, we follow the approach mentioned in Bigman et al. (1983), Fama & French (1987), Deaves & Krinsky (1992), Wei & Zhu (2006), and Weron (2008) who expect that futures prices ($F_{t,T}$) at time t for delivery at time T, plus a risk premium equal the expected spot prices at time T (given the information available at time t) (Note 28). This relation is given by the following equation:

$$F_{t,T} + RP_{t,T} = E_t[S_T] \tag{11}$$

Reshuffling terms and solving for the risk premium yields:

$$RP_{t,T} = E_t[S_T] - F_{t,T} \tag{12}$$

In order to extract the risk premium and use it to forecast the future spot prices, we follow the method applied in Hsieh and Kulatilaka (1982), Pindyck (1999), but also in Wei & Zhu (2006). In a first step, we estimate Equation (13) in log form:

$$p_T = a_1 f_{t,T} + a_2 rp_{t,T} + u_T, \tag{13}$$

where $rp_{t,T}$ is the expected risk premium at time t for delivery at T. Moreover, p_T denotes the spot price at time $t = T$ for a certain futures maturity, and $f_{t,T}$ is the futures price.

As the future spot price is unknown, the expected risk premium cannot be approximated directly. For this reason, we fit a state-space model using Kalman filter techniques (for an outstanding discussion of this approach see for example Harvey, 1993) to estimate the unobservable risk premium which captures also the dynamic behavior of the data set included (Note 29).

The state-space model of risk premium is given by the following system of equations (Note 30):

$$p_T = a_1 f_{t,T} + a_2 rp_{t,T} + \omega_T. \tag{14}$$

This equation is also known as the measurement equation, which represents the relation between price series and state variables. Here p_T denotes the spot price at time $t = T$, and $f_{t,T}$ is the futures price. The transition equation characterizes the dynamics of the state variables and is given by (Note 31):

$$rp_{t,T} = a_3 rp_{t-1,T} + \varepsilon_T, \tag{15}$$

where $rp_{t-1,T}$ is the expected risk premium at time $t - 1$.

Given the observable futures prices and the according expected risk premium yields to a forecast of the future spot price given by:

$$E_t[S_T] = F_{t,T} + rp_{t,T}. \tag{16}$$

5.2 Forecast Error

The difference between the realized spot price (S_T) and the expected spot price ($E_t[S_T]$) is called the forecast error ($FE_{t,T}$) which is given by:

$$S_T - E_t[S_T] = FE_{t,T}. \tag{17}$$

The forecast error incorporates the market participants' unmet expectations (Note 32). Thus, the forecast error deems as a proxy for the dynamics and the level of efficiency of a certain market place (Note 33). Based on the forecast error we set up the following linear OLS-regression that allows us to determine the driving factors of the forecast error given as:

$$FE_t = \alpha_0 + \alpha_1 AUTDum_t + \alpha_2 WINDum_t + \alpha_3 SADum_t + \alpha_4 CRET_t + \alpha_5 VOIL_t + \alpha_6 CRB_t + \alpha_7 LIQ_t +$$
$$\alpha_8 VOLUME_t + \alpha_9 SPRET_t + \alpha_{10} R_{m,t} + \varepsilon_t. \tag{18}$$

These variables are exhaustively discussed and analyzed due to their impact on commodity prices and volatility. For expository convenience, we summarize the expected effective direction for this model in the following table which guides the interested reader to certain references. The variables itself are described in detail below.

Table 14. Expected impact of the driving factors on the forecast error

Influencing Factor	Abbreviation	Sign	Literature
Seasonal Dummies	$AUTDum_t$; $WINDum_t$	−;+	Fama and French (1987), Cartea and Williams (2008)
Storage Ann. Dummy	$SADum_t$	−	Linn and Zhu (2004), Mu (2007)
Crude Oil Return	$CRET_t$	+	Brown and Yücel (2008), Däuper (2004)
Crude Oil Volatility	$VOIL_t$	+	Brown and Yücel (2008), Däuper (2004)
Commodity Index	CRB_t	+	McKenzie et al. (2004)
Liquidity	LIQ_t	+	Sarr and Lybeck (2002)
Volume	$VOLUME_t$	+	Sarr and Lybeck (2002)
Financial Market Return	$SPRET_t$	+	Fama and French (1987), Bailey and Chan (1993)
Risk-free Rate	$R_{m,t}$	−	Pindyck (2004)

Note. This table provides an overview of the relevant literature on commodity driving factors.

5.3 Financial Market Factors

Several articles deal with financial market factors that drive the dynamics of storable commodities (see for example Deaton & Laroque, 1992, 1996; Routledge et al., 2000; Wei & Zhu, 2006). Therefore, we include the financial market return ($SPRET_t$) and risk-free rate ($R_{m,t}$) of the respective country as explanatory variables in the equation to explain the forecast error. $SPRET_t$ covers the stock markets returns from the leading (national) blue chip indices of the respective countries. We refer to the FTSE 100 index for the UK and to the NYSE US 100 index for the US. The FTSE 100 is a share index and covers the stocks of the 100 companies with the highest market capitalization listed on the London Stock Exchange. The incorporation of the FTSE 100 index is crucial, due to the fact that five out of the ten largest FTSE 100 companies are related to the sectors oil, gas, and mining. The NYSE US 100 index tracks the top 100 companies traded on the New York Stock Exchange and is an important indicator for American and world economic trends. The index is diversified in 10 industry sectors and covers twelve companies related to the sectors oil and gas. As shown in Fama & French (1987) and Bailey & Chan (1993), who incorporate the systematic risk of financial markets to measure the effect on the convenience yield, we expect a similar behavior of this variable on the forecast error. Hence, as we expect the natural gas spot prices to be positively related to the financial market return and because the risk premium can be seen as insurance premium paid in order to avoid future uncertainty, this implies that also the risk premium as well as the forecast error are expected to be positively related to the security market index.

The risk-free rate is given by the zero yield curve. Sadorsky (2002) finds that the risk-free rate has a positive and significant impact on gasoline returns. In contrast to this, Pindyck (2004) shows that the risk-free rate exerts a positive and significant impact on crude oil returns but does not affect natural gas returns. Nevertheless, although the evidence is not unambiguous, we expect the risk-free rate to affect the forecast error in a negative way.

5.4 Commodity Spillovers

The RJ/CRB index (CRB_t) is a commodity futures index provided by Reuters which encompasses 19 different futures that are traded at international commodity forward exchanges (Note 34). As shown in Gorton & Rouwenhorst (2004) or McKenzie et al. (2004), the index serves as an important indicator for the future development of inflation and cost trends for the manufacturing industry. We expect the RJ/CRB index to have a positive impact on the risk premium, due to the fact that higher index values indicate an increase in commodity prices. As a result, the forecast error should be affected by the commodity futures index in a similar way as the risk premium itself. If thus, the unfulfilled expectations of the market participants traces back to an unforeseen change in the overall commodity industry, this variable should show a significant sign (Note 35).

Additional, crude oil returns (denoted by $CRET_t$) are included, as crude oil provides the closest substitute to

natural gas. Däuper (2004) and Brown & Yücel (2008) pointed out that natural gas prices do not only react to changes in inventory levels and the weather (natural gas market analysts stress that these factors are the main drivers), but also react sensitively to changes in close substitutes, as for example oil prices or petroleum prices. This may be due to the coupling of oil and gas prices in several retail contracts. As a consequence, we assume that crude oil returns have a positive impact on the forecast error. In addition to the crude oil return, the variance of oil ($VOIL_t$) may also play a certain role for the forecast error. Therefore, we include $VOIL_t$, representing the recursive sample volatility of crude oil, and expect a positive impact on the forecast error which is also in line with Däuper (2004) and Brown & Yücel (2008).

5.5 Market Liquidity and Trading Volume

The liquidity of the market is captured by LIQ_t. Following Sarr & Lybeck (2002), liquidity measures can be differentiated into four different groups—a) transaction cost measures, b) volume-based measures, c) equilibrium price-based measures and d) marketimpact measures. Driven by data availability constraints, we focus on the volume-based measure, which provides insights whether the markets' liquidity affects the forecast error. In case we observe forecast errors which are unequal to zero, the market is said not to be efficient. Hence, we expect that the higher the level of market liquidity, i.e. the smaller the spread, the lower the forecast error which speaks in favor of a positive sign.

For the liquidity measure, we refer to the Lui-Heubel liquidity ratio, which is calculated by:

$$L_{LH} = \frac{(P_{max} - P_{min})/P_{min}}{V/S}, \tag{19}$$

with P_{max} and P_{min} giving the maximum and minimum prices of a certain date and V and S giving the volume variables determining the traded volume and the amount outstanding. Here, V/S can simply be seen as the turnover of the certain product.

Among others, Sarr & Lybeck (2002) concluded that a single measure alone is not able to tell the whole story on the liquidity of a certain market. Thus, we also include $VOLUME_t$ to unfold the influence of liquidity on the forecast errors in the OLS-regression. Our expectations regarding the impact of $VOLUME_t$ are in line with the sign of the LuiHeubel liquidity ratio, i.e., we expect a positive sign.

5.6 Seasonalities and Storage Effects

Additional to these explanatory variables we include a storage announcement dummy ($SADum_t$) which is set equal to 1 on the day storage information is released, and 0 otherwise. In line with Linn & Zhu (2004), we expect the forecast error to be negatively affected by the announcement of storage levels, i.e. more information in the market is said to reduce uncertainty and, therefore, also the according forecast error.

Finally, as natural gas is a commodity that is clearly driven by seasonal variations in demand, we incorporate two seasonal dummies, one for the autumn and one for the winter months (Note 36). We expect the seasonal dummy variable to exert a positive (negative) impact on the forecast error as in autumn (winter) storage of natural gas becomes less (more) important than in other seasons. This is for example shown in Fama & French (1987) or Cartea & Williams (2008), who ascribe this finding to the existing seasonalities in production and demand (Note 37).

6. Empirical Results on Spot Price Predictions

In the following section, we provide the results of the expected future spot price and the forecast error for each hub.

6.1 Risk Premium Results

For NBP, the expected (or also titled ex-ante) risk premium for all maturities shows positive and negative values as suggested by the minimum and maximum values given in Table (15). With increasing maturity the average (negative) risk premium implied in the market increases in absolute terms.

Table 15. Descriptive statistics—ex-ante risk premium NBP

NBP	$RP_t^{ante,1m}$	$RP_t^{ante,2m}$	$RP_t^{ante,3m}$
Nrobs	3163	3163	3163
Mean	-0.0461	-0.0829	-0.0938
Median	-0.0329	-0.0491	-0.0728
Maximum	0.7841	0.6119	0.8678
Minimum	-0.7427	-1.0264	-0.9689
Std. Dev.	0.1665	0.2268	0.3022
Skewness	-0.0490	-0.6106	-0.0220
Kurtosis	6.3577	4.2846	3.7508
Excess Kurtosis	3.3577	1.2846	0.7508
Jarque-Bera	1487.1190	414.0236	74.5344

Note. This table provides the descriptives for the ex-ante risk premium for one to three month maturities. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

For HH, the descriptives of the forecasted expected risk premium are given in Table (16). We find that the ex-ante risk premium for HH exhibits positive and negative values, similar to the findings for Europe.

Table 16. Descriptive statistics—ex-ante risk premium HH

HH	$RP_t^{ante,1m}$	$RP_t^{ante,2m}$	$RP_t^{ante,3m}$
Nrobs	3203	3203	3203
Mean	-0.0220	-0.0511	-0.0661
Median	-0.0121	-0.0565	-0.0607
Maximum	0.6175	0.7903	0.4772
Minimum	-0.5319	-0.5664	-0.5493
Std. Dev.	0.1221	0.1785	0.1609
Skewness	0.1339	0.9679	0.0643
Kurtosis	6.8220	6.6298	2.8346
Excess Kurtosis	3.8220	3.6298	-0.1654
Jarque-Bera	1958.980	2258.476	5.8585

Note. This table provides the descriptives for the ex-ante risk premium for one to three month maturities. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

Concluding, the average risk premium estimates for Europe and HH are negative for all maturities (this is in line with Movassagh and Modjtahedi, 2005). Regarding the term structure of NBP, we observe that, as the time to maturity increases, the volatility (in absolute terms) increases. For HH, we see that, as the time to maturity increases, the volatility (in absolute terms) evolves rather stable whereas the mean increases. Overall, our findings for NBP are not in line with the results derived by Cartea & Williams (2008) for the UK natural gas forward market. In addition, these findings once more emphasis that futures prices are biased predictors of the future spot prices.

6.2 Explaining the Forecast Error—Results and Descriptives

In line with our results from Section (4), we observe non-zero forecast errors which are shown together with the descriptive statistics in Tables (17) and (18).

Table 17. Descriptive statistics—forecast error NBP

NBP	FE_{1m}	FE_{2m}	FE_{3m}
Nrobs	3163	3163	3163
Mean	-0.0028	-0.0021	-0.0014
Median	-0.0011	-0.0013	-0.0008
Maximum	1.1128	1.2175	1.2547
Minimum	-0.6727	-0.7424	-0.7871
Std. Dev.	0.0633	0.0566	0.0571
Skewness	0.6814	2.1433	3.3712
Kurtosis	60.7291	102.7638	129.1560
Excess Kurtosis	57.7291	99.7638	126.1560
Jarque-Bera	439459.7	1314120	2103498

Note. This table provides the descriptives for the forecast error for one to three month maturities. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

Table 18. Descriptive statistics—forecast error HH

HH	FE_{1m}	FE_{2m}	FE_{3m}
Nrobs	3203	3203	3203
Mean	-0.0021	-0.0021	-0.0020
Median	-0.0014	-0.0016	-0.0016
Maximum	0.6541	0.8419	0.5168
Minimum	-0.6456	-0.7037	-0.4251
Std. Dev.	0.0555	0.0540	0.0428
Skewness	0.3673	1.5389	0.3726
Kurtosis	35.8394	66.9438	36.9871
Excess Kurtosis	32.8394	63.9438	33.9871
Jarque-Bera	143996.5	546949.6	154234.5

Note. This table provides the descriptives for the forecast error for one to three month maturities. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

We observe that the volatility of the forecast error at NBP lies well below the according variability in the forecast error values of HH which is also true for the implied mean.

Moreover, the volatilities for all contracts and across all hubs are above the means (in absolute terms). As mentioned earlier, market participants are not able to build perfect expectations, i.e., differently spoken, this means that not time per se seems to exert an impact on the forecast power of the model introduced. Tables (17) and (18) further imply the existence of excess kurtosis across all hubs and maturities.

Overall, the price evolution of the realized spot price and the expected spot price is rather close, which implies that investors' ex-ante forecasts of the risk premium are close to the realized market price of risk implied in a certain market. The forecast error resulting from the mechanical predictor offered by the Kalman filter setting shows that the predicted values fluctuate around the realized expected spot prices, such that no clear statement can be given whether there is a significant over or underestimation of the expected spot price. Therefore, in the next step, we have a closer look on possible driving factors on the forecast error.

6.3 Explanatory Variables

Tables (19) and (20) report the outcomes of the OLS-regression which provide insight into factors that influence the differences between expected spot price ($E_t[S_T]$) and realized spot price (S_T) obtained for the different trading places (Note 38).

Table 19. OLS regression results—NBP

NBP	FE_{1m}		FE_{2m}		FE_{3m}	
	Coeff	Prob	Coeff	Prob	Coeff	Prob
Constant	0.0414	0.1387	0.0204	0.4354	0.0415*	0.0510
$AUTDum_t$	-0.0023	0.3624	0.0030	0.2334	0.0130***	0.0001
$WINDum_t$	0.0076**	0.0167	0.0083***	0.0013	0.0136***	0.0000
$SEADum_t$	0.0023	0.4912	0.0017	0.6211	0.0003	0.9267
$CRET_t$	-0.0216	0.1239	-0.0056	0.6094	-0.0144	0.1782
$VOIL_t$	-1.3828**	0.0329	-0.5500	0.3456	-1.2173**	0.0133
CRB_t	-0.0368**	0.0331	-0.0463***	0.0007	-0.0251	0.2505
LIQ_t	-0.0319**	0.0102	-0.0245***	0.0001	-0.0015	0.6852
$VOLUME_t$	0.0041**	0.0112	0.0010	0.3984	0.0015	0.1221
$SPRET_t$	-0.0018	0.8552	-0.0113	0.3113	-0.0012	0.9021
$R_{m,t}$	-0.0041	0.1239	-0.0025	0.3742	-0.0059**	0.0392
Adjusted R^2	0.0142		0.0126		0.0139	
Durbin-Watson Stats	2.0191		2.0016		1.9948	

Note. ***indicates significance at 1%, ** at 5%, and * at 10% level.

Table 20. OLS regression results—HH

HH	FE_{1m}		FE_{2m}		FE_{3m}	
	Coeff	Prob	Coeff	Prob	Coeff	Prob
Constant	0.0516**	0.0131	0.0258	0.1509	0.0206*	0.0883
$AUTDum_t$	-0.0017	0.4992	0.0006	0.8174	-0.0014	0.4680
$WINDum_t$	0.0022	0.4055	0.0042*	0.0797	0.0028	0.1413
$SEADum_t$	0.0008	0.8196	-0.0025	0.3501	-0.0009	0.6970
$CRET_t$	-0.0006	0.9571	-0.0096	0.4926	0.0037	0.6578
$VOIL_t$	-0.5421	0.2063	-1.4190***	0.0007	-0.4793	0.1458
CRB_t	-0.2112***	0.0000	-0.2097***	0.0000	-0.1993***	0.0000
LIQ_t	-0.0174	0.2080	0.0108*	0.0603	0.0023	0.6167
$VOLUME_t$	-0.0027*	0.0879	0.0047**	0.0107	-1.38E-05	0.9928
$SPRET_t$	0.0137	0.1109	0.0215**	0.0331	0.0151**	0.0265
$R_{m,t}$	-0.0026*	0.0932	-0.0031**	0.0394	-0.0019	0.1011
Adjusted R^2	0.0742		0.0620		0.0866	
Durbin-Watson Stats	2.0013		1.9852		1.9761	

Note. ***indicates significance at 1%, ** at 5%, and * at 10% level.

For NBP, we observe that for the forecast errors resulting from the different futures' maturities at least one of the two included seasonal dummies exerts a significant impact (for the three months maturity both show a significant impact). From the market variables that assist the analysis of the impact of the overall economic situation, i.e. the CRB index, $SPRET_t$ and the risk-free rate, also at least one variable turns out to significantly affect the forecast error. While for shorter maturities the liquidity measure affects the forecast error, for longer maturities (three months) crude oil becomes important. However, the signs of the significant variables show an effective direction opposed to our expectations (Note 39). This indicates that the information processing when expectations on risk premium are build is not perfect or unbiased, which thus, results in a forecast error. A similar line of reasoning holds for the variables that show the expected effective direction. For these, the information processing works in the correct direction, however, the information processing does not sufficiently transform the information given into an appropriate price signal.

For HH, we observe a slightly different picture. The seasonal dummies only affect the two months forecast error.

For all forecast errors, the CRB index shows a significant impact with effective direction opposed to our expectations. All other variables that turn out to be significant show the expected sign (besides $Volume_t$ for the one month forecast error). Also in line with NBP, at least one market variable affects the forecast error. However, additional to the CRB index, the risk-free rate and the market return are significant. Similar to the results for NBP for the shorter maturities we observe a significant impact of the liquidity measures.

7. Concluding Remarks

In this paper, we analyzed the US and UK natural gas futures market, represented by one, two, and three months contracts traded at NBP and HH. In a first step, we investigated whether futures prices are unbiased predictors of future spot prices. For this, we applied the Johansen Full Information Maximum Likelihood approach to test for a common stochastic trend in futures and spot prices. Following this, we implemented a test design to determine whether or not the considered markets can be viewed as unbiased in the short- and long-run. In a second step, we allowed market participants to form expectations on the future spot prices with respect to a mechanical predictor given by the Kalman filter procedure. Last, we analyzed the differences between realized and expected future spot price, also known as forecast error, and studied its relationship to other market variables.

Our results indicate that futures and spot prices are non-stationary for all considered contract maturities at NBP and HH. Nevertheless, they are cointegrated with order one which highlights that futures and spot prices share a common stochastic trend and are in a long-run equilibrium. Furthermore, the UH is rejected for both natural gas trading hubs. More specifically, results argue (at least in most cases) in favor of the existence of a risk premium. This is consistent with results, for example, from Movassagh & Modjtahedi (2005) and Hobaeck-Haff et al. (2008). Moreover, we illustrated that past information is relevant to forecast future prices and, thus, should be incorporated to increase the predictive performance of the model. Hence, market participants can use this information to build profitable trading strategies. Nevertheless, as mentioned previously with respect to the model chosen it cannot be concluded that one market is less efficient than the other (see Kellard et al., 1999).

In order to approximate for the future spot price and the according risk premium implied in the European and North American natural gas markets, we implemented a mechanical predictor via Kalman filtering procedures. Based on these results, we analyzed the difference between the forecasted and realized spot prices at the futures' maturity, i.e., the forecast errors which give insight into problems to predict future spot prices given the current futures price and information set. The strong predictive power of the forecast method chosen became immediately evident as the average of the forecast error is close to zero. Furthermore, this fact is a strong indication that futures contracts are good predictors for the futures spot price, even though they are biased. In addition, the OLS-regression on the forecast errors illustrated that some of the variables included explain parts of the forecast error, its bulk part remains unexplained which may be ascribed to biased expectations, incomplete or defective information or information processing or market frictions inherent in the trading places covered. As for example mentioned in Fama & French (1987), Fama & French (1988), or Neumann et al. (2008) who state that the resulting biased expectations may be related to the lack of efficiency caused by the restricted storage market.

What does that mean for market participants and their trading strategies?

In general, in an efficient market, expectations should be based on all necessary information provided in the market. In this case, the ex-ante risk premium should be in line with the realized premium and, thus, no forecast error should be observable. We find that futures and spot prices at both, the European and US- natural gas markets can be seen as biased. However, the forecast error is quite small and expected premia oscillates around the realized counterparts. These findings are consistent with previous literature and constitute that futures are biased predictors of the subsequent spot prices and the relevancy of past information. Additionally, our model also suggests in both markets forecast errors which imply additional costs for hedging and diversification (as indicated in Chang (1985), Fama & French (1987), Fama & French (1988), Roon et al. (2000), or Neumann et al. (2008)).

Overall, our findings give insights into the fact that the differences between the European and the American market are not tremendously large. However, in both markets, the information processing and the forecasting power leaves room for improvements. The fact that the European natural gas market is still emerging (i.e., regulations, storage access) may hinder the exploitation of clear trading strategies, such that the reduction of market frictions is the utmost important task for policy makers. In order to do so, a more opened and European wide legal framework is needed, especially, as a natural monopoly (infrastructure) is included. Also the existence of different, although connected national networks requires further assessment in cross border trading and reductions of cross border congestion and frictions. As also mentioned in Kremser & Rammerstorfer (2016), the market coupling and opening is an important basis for the price integration of different trading places and the

implementation of an arbitrage-free and efficient market.

References

Alvarez-Ramirez, J., Alvarez, J., & Rodriguez, E. (2008). Short-term predictability of crude oil markets: A detrended fluctuation analysis approach. *Energy Economics, 30*, 2645-2656. https://doi.org/10.1016/j.eneco.2008.05.006

Asche, F., Osmundsen, P., Sikveland, M., & Tveteras, R. (2009). Volatility and risk sharing in European gas markets. *Proceedings—32nd IAEE International Conference*.

Bailey, W., & Chan, K. C. (1993). Macroeconomic influences and the variability of the commodity futures basis. *The Journal of Finance, 48*, 555-573. https://doi.org/10.1111/j.1540-6261.1993.tb04727.x

Beck, S. E. (1994). Cointegration and market efficiency in commodity futures markets. *Applied Economics, 26*, 249-257. https://doi.org/10.1080/00036849400000006

Bigman, D., Goldfarb, D., & Schechtman, E. (1983). Futures market efficiency and the time content of the information sets. *The Journal of Futures Markets, 3*, 321-334. https://doi.org/10.1002/fut.3990030307

Bomhoff, E. (1992). Four econometric fashions and the kalman filter alternative—a simulation study. *Tilburg Center for Economic Research, Discussion Paper No. 9227*.

Brooks, C. (2008). *Introductory Econometrics for Finance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511841644

Brown, S., & Yücel, M. (2008). What drives natural gas prices? *The Energy Journal, 29*, 45-60. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol29-No2-3

Cartea, A., & Williams, T. (2008). Uk gas markets: The market price of risk and applications to multiple interruptible supply contracts. *Energy Economics, 30*, 829-846. https://doi.org/10.1016/j.eneco.2007.03.001

Chang, E. (1985). Return to speculators and the theory of normal backwardation. *The Journal of Finance, 40*, 193-208. https://doi.org/10.1111/j.1540-6261.1985.tb04944.x

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28*, 591-605. https://doi.org/10.2307/1910133

Cornell, B. (1977). Spot rates, forward rates and exchange market efficiency. *Journal of Financial Economics, 5*, 55-65. https://doi.org/10.1016/0304-405X(77)90029-0

Crowder, W. J., & Hamed, A. (1993). A cointegration test for oil futures market efficiency. *The Journal of Futures Markets, 13*, 933-941. https://doi.org/10.1002/fut.3990130810

Däuper, O. (2004). *Gaspreisbildung und Europäisches Kartellrecht*. C.H. Beck-Verlag, München.

Deaton, A., & Laroque, G. (1992). On the behaviour of commodity prices. *Review of Economic Studies, 59*, 1-23. https://doi.org/10.2307/2297923

Deaton, A., & Laroque, G. (1996). Competitive storage and commodity price dynamics. *Journal of Political Economy, 104*, 896-923. https://doi.org/10.1086/262046

Deaves, R., & Krinsky, I. (1992). Risk premiums and efficiency in the market for crude oil futures. *The Energy Journal, 13*, 93-117.

El HediArouria, M., Hammoudehb, S., Lahianic, A., & Ngu-yen, D. K. (2013). *On the short- and long-run efficiency of energy and precious metal markets*. Working Paper hal-00798036.

Elam, E., & Dixon, B. L. (1988). Examining the validity of a test of futures market efficiency. *The Journal of Futures Markets, 8*, 365-372. https://doi.org/10.1002/fut.3990080309

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*, 251-276. https://doi.org/10.2307/1913236

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*, 383-417. https://doi.org/10.2307/2325486

Fama, E., & French, K. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *The Journal of Business, 60*, 55-73. https://doi.org/10.1086/296385

Fama, E., & French, K. (1988). Business cycles and the behavior of metal prices. *The Journal of Finance, 63*, 1075-1093. https://doi.org/10.1111/j.1540-6261.1988.tb03957.x

Frenkel, J. A., & Mussa, M. L. (1980). *Efficiency of foreign exchange markets and measures of turbulence.* NBER Working Paper No. 476. https://doi.org/10.3386/w0476

Geman, H. (2005). *Commodities and Commodity Derivatives: Energy, Agriculturals and Shipping*. Wiley Finance.

Gorton, G. B., & Rouwenhorst, K. G. (2004). *Facts and fantasies about commodity futures*. NBER Working Paper No. 10595. https://doi.org/10.3386/w10595

Harvey, A. C. (1993). *Time Series Models*. Cambridge: The MIT Press.

Haugom, E., & Ullrich, C. J. (2012). Market efficiency and risk premia in short-term forward prices. *Energy Economics, 34*, 1931-1941. https://doi.org/10.1016/j.eneco.2012.08.003

Hobaeck-Haff, I., Lindqvist, O., & Loland, A. (2008). Risk premium in the UK natural gas forward market. *Energy Economics, 30*, 2420-2440. https://doi.org/10.1016/j.eneco.2007.12.002

Hsieh, D., & Kulatilaka, N. (1982). Rational expectations and risk premia in forward markets: Primary metals at the London metals exchange. *The Journal of Finance, 27*, 1199-1207. https://doi.org/10.1111/j.1540-6261.1982.tb03612.x

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*, 231-254. https://doi.org/10.1016/0165-1889(88)90041-3

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to demand for money. *Oxford Bulletin of Economics and Statistics, 52*, 169-210. https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x

Kaminsky, G., & Kumar, M. S. (1990). Efficiency in commodity futures markets. *International Monetary Fund Staff Papers, 34*, 670-699. https://doi.org/10.2307/3867269

Kellard, N., Newbold, P., Rayner, T., & Ennew, C. (1999). The relative efficiency of commodity futures markets. *The Journal of Futures Markets, 19*, 413-432. https://doi.org/10.1002/(SICI)1096-9934(199906)19:4%3C413::AID-FUT2%3E3.0.CO;2-F

Koenig, P. (2012). *The effect of lng on the relationship between uk and continental European natural gas markets*. Cambridge Working Paper 1253-EPRG Working Paper 1225.

Krehbiel, T., & Adkins, L. C. (1993). Cointegration tests of the unbiasedness expectations hypothesis in metal markets. *The Journal of Futures Markets, 13*, 753-763. https://doi.org/10.1002/fut.3990130704

Kremser, T., & Rammerstorfer, M. (2016). The convenience yield implied in the European natural gas markets: the impact of storage and weather. *Journal of Energy Markets*, 27-58. https://doi.org/10.21314/jem.2016.147

Lai, K. S., & Lai, M. (1991). A cointegration test for market efficiency. *The Journal of Futures Markets, 11*, 567-575. https://doi.org/10.1002/fut.3990110505

Linn, S. C., & Zhu, Z. (2004). Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *The Journal of Futures Markets, 24*, 283-313. https://doi.org/10.1002/fut.10115

Maslyuk, S., & Smyth, R. (2008). Unit root properties of crude oil spot and futures prices. *Energy Policy, 36*, 2591-2600. https://doi.org/10.1016/j.enpol.2008.03.018

McKenzie, A. M., & Holt, M. T. (2002). Market efficiency in agricultural futures markets. *Applied Economics, 34*, 1519-1532. https://doi.org/10.1080/00036840110102761

McKenzie, A. M., Thomsen, M. R., & Dixon, B. L. (2004). The performance of event study approaches using daily commodity futures returns. *The Journal of Futures Markets, 24*, 533-555. https://doi.org/10.1002/fut.10126

Modjtahedi, B., & Movassagh, N. (2005). Natural-gas futures: Bias, predictive performance, and the theory of storage. *Energy Economics, 27*, 617-637. https://doi.org/10.1016/j.eneco.2005.04.005

Moosa, I. A., & Al-Loughani, N. E. (1994). Unbiasedness and time varying risk premia in the crude oil futures market. *Energy Economics, 2*, 99-105. https://doi.org/10.1016/0140-9883(94)90003-5

Movassagh, N., & Modjtahedi, B. (2005). Bias and backwardation in natural gas futures prices. *The Journal of Futures Markets, 25*, 281-308. https://doi.org/10.1002/fut.20151

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics, 29*, 46-63. https://doi.org/10.1016/j.eneco.2006.04.003

Neumann, A., Rammerstorfer, M., & Stronzik, M. (2008). Does the European natural gas market pass the competitive benchmark of the theory of storage? Indirect tests for three major trading points. *Energy Policy, 37*, 5432-5439.

Pindyck, R. S. (1993). The present value model of rational commodity pricing. *Economic Journal, 103*, 511-530. https://doi.org/10.2307/2234529

Pindyck, R. S. (1999). *The long-run evolution of energy prices.* CEEPR Working Paper No. 99-001 WP. https://doi.org/10.5547/issn0195-6574-ej-vol20-no2-1

Pindyck, R. S. (2001). The dynamics of commodity spot and futures markets: A primer. *The Energy Journal, 22*, 1-30. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol22-No3-1

Pindyck, R. S. (2004). Volatility and commodity price dynamics. *The Journal of Futures Markets, 24*, 1029-1047. https://doi.org/10.1002/fut.20120

Roon, F., Nijam, T., & Veld, C. (2000). Hedging pressure effects in futures markets. *The Journal of Finance, 55*, 1437-1456. https://doi.org/10.1111/0022-1082.00253

Routledge, B. R., Seppi, D. J., & Spatt, C. S. (2000). Equilibrium forward curves for commodities. *The Journal of Finance, 55*, 1297-1338. https://doi.org/10.1111/0022-1082.00248

Sadorsky, P. (2002). Time-varying risk premiums in petroleum futures prices. *Energy Economics, 24*, 539-556. https://doi.org/10.1016/S0140-9883(02)00062-2

Sarr, A., & Lybeck, T. (2002). *Measuring liquidity in financial markets.* IMF Working Paper WP/02/232.

Shambora, W., & Rossiter, R. (2007). Are there exploitable inefficiencies in the futures market for oil? *Energy Economics, 29*, 18-27. https://doi.org/10.1016/j.eneco.2005.09.004

Shea, G. (1992). Benchmarking the expectations hypothesis of the interest rate term structure: An analysis of cointegrating vectors. *Journal of Business and Economic Statistics, 10*, 347-366. https://doi.org/10.1080/07350015.1992.10509910

Siliverstovs, B., L'Hegaret, G., Neumann, A., & Von Hirschhausen, C. (2005). International market integration for natural gas? a cointegration analysis of prices in europe, north america and japan. *Energy Economics, 27*, 603-615. https://doi.org/10.1016/j.eneco.2005.03.002

Switzer, L. N., & El-Khoury, M. (2007). Extreme volatility, speculative efficiency, and hedging effectiveness on the oil futures markets. *The Journal of Futures Markets, 27*, 61-84. https://doi.org/10.1002/fut.20235

Taylor, M. P. (1995). The economics of exchange rates. *Journal of Economic Literature, 33*, 13-47.

Wei, S. Z. C., & Zhu, Z. (2006). Commodity convenience yield and risk premium determination: The case of the U.S. natural gas market. *Energy Economics, 28*, 523-534. https://doi.org/10.1016/j.eneco.2005.10.002

Weron, R. (2008). Market price of risk implied by asian-style electricity options and futures. *Energy Economics, 30*, 1098-1115. https://doi.org/10.1016/j.eneco.2007.05.004

Notes

Note 1. The informational content of natural gas futures contracts on the subsequent price of the spot product has many important practical implications, as not only traders in futures markets study the price evolution, dynamics, and forecasting power of futures contracts on future levels of the spot price, but also producers of goods which base their production planning on this information and trying to minimize their risk against adverse price fluctuations. Hence, a wide range of economic agents use the information conveyed through futures contracts.

Note 2. Thus, the price process of F_{t-1} is a Q-martingale and on average the price of the futures contract is constant. The martingale property of F_{t-1} implies that $F_{t-1} = E_Q[F_t|\Omega_{t-1}]$ which can be reformulated to $F_{t-1} = E_Q[S_t|\Omega_{t-1}]$ when the market is arbitrage-free.

Note 3. In line with Moosa & Al-Loughani (1994), irrationality of market participants is only empirically testable if a survey is conducted among market participants. Hence, if the UH does not hold, a common theoretical assumption is to expect that market participants are rational but require a risk premium.

Note 4. Cornell (1977) and Frenkel & Mussa (1980) examined the UH for certain currency relations, whereas Crowder & Hamed (1993) and Moosa & Al-Loughani (1994) tested the UH for oil futures contracts. A similar test design is given in Hsieh & Kulatilaka (1982) who analyze the UH on forwards contracts for primary metals with respect to rational expectations based on full information. Whereas Kellard et al. (1999) studied the relative degree of inefficiency of certain commodities. Moreover, McKenzie & Holt (2002) and Haugom & Ullrich (2012) test for market efficiency and the presence of a time-varying risk premium in agricultural and the PJM electricity market.

Note 5. With this distinction between unbiasedness and efficiency, several papers try to disentangle the effects of both hypotheses.

Note 6. The OLS-coefficients δ and γ are tested by applying the Wald test and restricting the coefficients to $\delta = 0$ and $\gamma = 1$.

Note 7. In case of cointegration, the restrictions imposed on the coefficients ($\delta = 0$ and $\gamma = 1$) can be tested using a likelihood ratio test. This approach traces back to Engle and Granger (1987). Also the Johansen procedure can be used (see Johansen, 1988; Johansen & Juselius, 1990; Johansen, 1996).

Note 8. According to the representation theorem of Granger, if the rank (r) of $\prod$ is $r < m$ then $\prod$ has a reduced rank and is equal to $\alpha\beta'$, each of dimension $(m \; x \; r)$ and $(r \; x \; m)$. Here, r represents the order of cointegration, β corresponds to the cointegrating vector, and α is the matrix of adjustment coefficients.

Note 9. This is also in line with McKenzie & Holt (2002) who tested for market efficiency in agricultural futures markets.

Note 10. For $\lambda_1 = 1$ the futures price is an unbiased predictor and the change in the futures price is immediately reflected in current spot price changes.

Note 11. The futures contract price series for 1-month for example shows the rolling prices for a futures contract with the maturity of one month. After expiration, this contract is replaced by the next futures contract with the same maturity, which starts operation directly after the former futures contract has expired. As a result, only at initiation date, maturity is at full length.

Note 12. The gas year starts on October 1st, and ends on September 30th.

Note 13. According to Pindyck (2001) normal backwardation means that $F_{t,T} < S_t$, and contango indicates that $F_{t,T} > S_t$.

Note 14. These points are identified as significant breaks with respect to the Chow Breakpoint test (see Chow, 1960).

Note 15. However, for the predictions with respect to the Kalman filter we allow the information set to include all past information.

Note 16. The descriptive statistics of the Phase 1 and 2 for NBP are given in Table (21) and (22) in the Appendix.

Note 17. The ADF and PP test are based on the null hypothesis that the considered time series contain a unit root (non-stationarity). Whereas the null hypothesis of the KPSS test states that the time series are (trend-) stationary. A time series which can be considered as stationary has a constant mean, variance and auto-covariance for each lag.

Note 18. As discussed by Brooks (2008), the reason to implement a unit root test (e.g., ADF or PP) and a stationarity test (e.g., KPSS) is because unit root tests can have a low explanatory power if the process is

stationary but with a root close to the non-stationarity boundary. Hence, if the unit root test rejects and the stationarity test accepts the null hypothesis (or vice versa) we can conclude that the considered time series are stationary or non-stationary.

Note 19. The lag length for the ADF test is selected by the Akaike information criterion (AIC), whereas for the PP and KPSS test we used the Parzen kernel to determine the probability density function and the Newey-West bandwidth for lag selection.

Note 20. Furthermore, the attempt to apply the "first-difference" or "percentage change" form of Equation (8) results in the removal of the long-run equilibrium relationship between the spot and futures prices. Thus, this approach cannot be applied here without appropriate justification.

Note 21. A cointegration relation between spot and futures prices series is only observable if these series are non-stationary in their levels, stationary after taking first differences and there exist a specific linear combination in levels which makes $u_t = s_t - \delta - \gamma f_{t-\tau}$ stationary. Thus, the cointegration regression can be specified as follows: $s_t = \delta + \gamma f_{t-\tau} + u_t$.

Note 22. In line with Kellard et al. (1999) we choose the optimal number of lags for the Johansen procedure according to the Schwarz Criterion (SC). The results of the Johansen procedure were not sensitive to the number of lags used in the VAR (maximum number of 10 lags), expect for the 2 and 3 months futures spot-relation at NBP during Phase 2.

Note 23. The Cointegration Rank Test based on the Trace Statistic confirms our findings for the Maximum Eigenvalue Test Statistic for both hubs. Additionally, we implement the Engle-Granger approach (see Engle & Granger, 1987) to test for cointegration. This approach requires a unit root test on the residuals from Equation (5). Overall, we find that residuals are stationary and, thus, are in line with the results for both hubs from the Johansen procedure.

Note 24. The rejection of H_0^a and H_0^b indicates that these series have two cointegration relationships and, thus, are not cointegrated. In such a case, a VAR model in levels would fit the data best.

Note 25. This is, for example, in line with Shea (1992) who tested the expectations hypothesis on US zero coupon yields and found lower cointegration relation for rates with a longer maturity.

Note 26. The optimal lag length for the short-run test of market unbiasedness was selected through the general to-specific testing in which the initial setting for the number of lags included in the model was equal to 10.

Note 27. For this we have to change our notation as we have now a forward looking approach in opposite to the ex-post approach mentioned above.

Note 28. A similar model is used by Haugom & Ullrich (2012) who refer to the PJM market in the US.

Note 29. Bomhoff (1992) demonstrated in a simulation study that the Kalman filter allows for non-stationarity in price series data and that it is not subjected to the fact whether non-stationary financial price series are cointegrated or not.

Note 30. With $E(\omega_T) = 0$ and $Var(\omega_T) = H_t$.

Note 31. Please note that $E(\varepsilon_T) = 0$ and $Var(\varepsilon_T) = Q_t$. In addition, it is assumed that the error terms (ω_T and ε_T) are uncorrelated.

Note 32. For example market restrictions (i.e., storage access and short-sale constraints) which hinders market participants to implement arbitrage-exploiting strategies.

Note 33. In the following we refer to $FE_{t,T}$ as FE_t.

Note 34. According to McKenzie et al. (2004), the RJ/CRB index always keeps an investment in close to maturity futures and, thus, measures the return from investing in nearby commodity futures and rolling them forward each month.

Note 35. In the following, this line of reasoning holds for each of the considered explanatory variables.

Note 36. The dummy variable is 1 for the months September, October, November (autumn), and December, January, and February (winter).

Note 37. Apart from the risk-free rates and the RJ/CRB index, all data is taken from Bloomberg. Data for the risk-free rate and the RJ/CRB Index is provided by Thompson Reuters Datastream.

Note 38. The OLS-regression may require in some cases the incorporation of AR and MA terms which are not given in the tables below.

Note 39. Besides $Volume_t$ for the one month maturity and the risk-free rate for the forecast error resulting with respect to the forecast of the spot rate in three months.

Appendix

Table 21. Descriptive statistics—NBP for the Phase 1

NBP	Spot	F_{1m}	F_{2m}	F_{3m}
Nrobs	1860	1860	1860	1860
Mean	2.7011	2.7329	2.7614	2.7864
Median	2.6804	2.7695	2.7950	2.8160
Maximum	4.3820	3.5720	3.6140	3.5950
Minimum	1.5151	2.1440	2.1350	2.1310
Std. Dev.	0.3848	0.3692	0.3715	0.3717
Skewness	0.2593	0.1256	0.0228	-0.0921
Kurtosis	2.1981	1.8904	1.8505	1.8109
Excess Kurtosis	-0.8019	-1.1096	-1.1495	-1.1891
Jarque-Bera	70.680	100.3095	102.5586	112.2029

Note. This table provides the descriptives for log spot prices, as well as log prices for one to three month futures contracts traded at NBP Pre-Break-Phase. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

Table 22. Descriptive statistics—NBP for the Phase 2

NBP	Spot	F_{1m}	F_{2m}	F_{3m}
Nrobs	1408	1408	1408	1408
Mean	3.7295	3.7530	3.7965	3.8284
Median	3.8749	3.9165	3.9340	3.9225
Maximum	4.6102	4.4460	4.6300	4.7040
Minimum	2.4765	2.6760	2.7680	2.7580
Std. Dev.	0.3909	0.3925	0.3944	0.3949
Skewness	-0.7184	-0.6609	-0.5334	-0.4961
Kurtosis	2.4176	2.3336	2.4626	2.8655
Excess Kurtosis	-0.5824	-0.6664	-0.5374	-0.1345
Jarque-Bera	141.0152	128.5615	83.69904	58.80417

Note. This table provides the descriptives for log spot prices, as well as log prices for one to three month futures contracts traded at NBP for the Post-Break-Phase. Here, the number of observations (Nrobs), Mean, Median, Maximum, Minimum, Standard Deviation (Std. Dev.), Skewness, Kurtosis, Excess Kurtosis, and the Jarque-Bera test statistic are highlighted.

Table 23. ADF, PP and KPSS tests on natural gas price series at NBP

			Stationarity Testing				
Phase	Series		Levels		First Differences		
		ADF	PP	KPSS	ADF	PP	KPSS
Phase 1	F_{1m}	-2.5895	-1.9433	3.7182***	-7.0334***	-39.6015***	0.0396
	F_{2m}	-2.9871**	-1.9525	3.6271***	-6.1435***	-40.0243***	0.0456
	F_{3m}	-3.0225**	-1.8049	3.5164***	-5.5867***	-39.9266***	0.0436
	Spot	-2.2796	-4.0586***	3.6572***	-14.4730***	-64.0289***	0.0330
Phase 2	F_{1m}	-1.8162	-1.7654	1.2681***	-13.2658***	-35.4385***	0.0773
	F_{2m}	-1.6705	-1.5938	1.1806***	-23.0102***	-36.1216***	0.0912
	F_{3m}	-1.5454	-1.6548	1.1251***	-26.7844***	-33.6064***	0.0780
	Spot	-1.8053	-2.1624	1.4139***	-8.4297***	-45.9085***	0.0656

Note. This table provides the results from the ADF, PP and KPSS test. The 5% significance level is indicated by **; *** indicate the 1% significance level. The ADF, PP and KPSS test in levels and first differences is performed with an intercept but not with a time trend.

Copyrights