

NHH

Norwegian School of Economics

Bergen, Fall 2013



Seasonality in Natural Gas Prices

An empirical study of Henry Hub Natural Gas Futures Prices

By: Baste Fladmark & Guro Berg Grimstad

Supervisor: Jørgen Haug

Master thesis in Financial Economics and Economic Analysis

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

This page intentionally left blank

Abstract

In this thesis we investigate whether seasonality is a significant factor in natural gas futures prices. We test for seasonality by estimating the two-factor model of Schwartz & Smith (2000), using Kalman filtering techniques in Matlab¹. Next, we extend the model with a trigonometric seasonality function, following Sørensen (2002), to see if the new factor is significant and leads to better estimation of other parameters in the model².

Our results indicate that Model 1 suffers from an omitted parameter bias, caused by the lack of a seasonal factor. After including seasonality in Model 2, the model improves significantly; leading us to conclude that seasonality is present in natural gas prices. This seasonality causes prices to be higher in winter months and lower in summer months.

¹ We refer to this model as Model 1.

² We refer to this model as Model 2.

This page intentionally left blank

Preface

This thesis is written as a concluding part of our Master of Science in Economics and Business Administration at the Norwegian School of Economics (NHH). The master thesis is written in conjunction with our majors in Financial Economics and Economic Analysis.

Working with this thesis has been a very rewarding experience. We both have a keen interest in mathematics, time-series analysis and financial derivatives, and have enjoyed exploring these fields further. Understanding and applying the Kalman filter on futures prices has been the most challenging part of this process. The work has been very educational, and we are left with a greater insight in how one can use multifactor models to study commodity price processes.

We would like to thank our advisor, Jørgen Haug, who not only taught the course derivatives pricing and risk management, which really opened our eyes for pricing of financial instruments, but whom also has given us invaluable advice and insightful feedback when writing this thesis. We would also like to thank Maren Berg Grimstad for proofreading the thesis and providing useful comments on our work.

Contents

- ABSTRACT..... 3**
- PREFACE..... 5**
- CONTENTS..... 6**
- FIGURES..... 8**
- TABLES..... 9**
- 1. INTRODUCTION 11**
- 2. THEORY 12**
 - 2.1 MARKET BACKGROUND AND DYNAMICS 12
 - 2.2 HOW THE TRADING IS CONDUCTED..... 14
 - 2.3 FORWARD AND FUTURES CONTRACTS 15
- 3. DATA..... 17**
 - 3.1 PRELIMINARY LOOK AT THE DATA..... 18
 - 3.2 AUGMENTED DICKEY-FULLER TEST 21
- 4. MODEL DESCRIPTION..... 23**
 - 4.1 WHY THE SCHWARTZ & SMITH MODEL..... 23
 - 4.2 OVERVIEW – MODEL 1 23
 - 4.3 DERIVING THE DISTRIBUTION OF THE SPOT PRICE 24
 - 4.4 VALUING FUTURES CONTRACTS USING THE MODEL..... 25
 - 4.5 A STATE-SPACE FORMULATION OF THE SCHWARTZ & SMITH – MODEL 27
- 5. ESTIMATION OF OUR STATE SPACE MODEL 30**
 - 5.1 A COMBINATION OF MAXIMUM LIKELIHOOD AND THE KALMAN FILTER..... 30
 - 5.2 KALMAN FILTERING 31
 - 5.3 KALMAN FILTERING ROUTINE IN MATLAB 34
 - 5.3.1 *Obtaining standard errors of parameter estimates by inverting the Hessian* 35
 - 5.3.2 *Forcing measurement errors to be zero* 35

6.	A REPLICATION OF THE SCHWARTZ & SMITH – MODEL.....	36
7.	RESULTS FROM MODEL 1.....	39
7.1	ESTIMATED STATE VARIABLES – MODEL 1.....	41
7.2	RESIDUAL ANALYSIS – MODEL 1.....	41
8.	EXTENDING THE MODEL FURTHER – ALLOWING FOR SEASONALITY.....	44
8.1	EXTENDING WITH DETERMINISTIC SEASONALITY.....	44
8.1.1	<i>Dynamics and futures price.....</i>	<i>45</i>
8.1.2	<i>Extra parameters and new measurement equation.....</i>	<i>46</i>
9.	RESULTS FROM MODEL 2.....	47
9.1	ESTIMATED STATE VARIABLES – MODEL 2.....	50
9.2	RESIDUAL ANALYSIS MODEL 2.....	50
10.	CONCLUDING REMARKS.....	54
11.	ASSUMPTIONS, LIMITATIONS AND IMPROVEMENTS.....	56
	BIBLIOGRAPHY.....	58
12.	APPENDICES.....	60
12.1	APPENDIX A.....	60
12.2	APPENDIX B.....	61

Figures

Figure 2.1 Production, consumption, net imports and net storage withdrawal of natural gas in the United States, 2001-2013 12

Figure 2.2 Natural gas consumption in the United States in 2012..... 13

Figure 2.3 NG1 contracts traded by year, 1990*-2012 *Annual volume for 1990 begins in April 15

Figure 3.1 Log of weekly future prices for NG1, NG10 and NG22 19

Figure 3.2 Average log prices for all contracts, conditional on month..... 21

Figure 6.1 Estimated Spot and Equilibrium Prices for the Futures Data (Schwartz & Smith, 2000)..... 38

Figure 6.2 Estimated Spot and Equilibrium Prices for the Futures Data 38

Figure 7.1 Plot of estimated state variables from Model 1 41

Figure 7.2 ACF-plot Model 1, sample autocorrelation for lags up to two years..... 43

Figure 9.1 Plot of the estimated seasonality factor 49

Figure 9.2 Plot of estimated state variables from Model 2 50

Figure 9.3 ACF-plot Model 2, sample autocorrelation for lags up to two years..... 52

Tables

Table 3.1 Average daily trading volume (1990-2013) and missing dates for selected contracts	17
Table 3.2 Monthly average prices relative to yearly average prices	20
Table 3.3 Summary statistics of log prices for all contracts, conditional on delivery month ..	20
Table 3.4 Test results from Augumented Dickey-Fuller and Ljung-Box tests	22
Table 5.1 Bounds on random initial values	34
Table 6.1 Parameter results from replicating Schwartz & Smith (2000) data	36
Table 6.2 Errors in the model fit to the logarithm of futures prices	37
Table 7.1 Parameter results from Model 1	39
Table 7.2 Summary statistics of residuals from Model 1	42
Table 7.3 Mean residual value by delivery month, Model 1	42
Table 7.4 Ljung-Box Q-test statistics for residual autocorrelation	43
Table 9.1 Parameter results from Model 2	47
Table 9.2 Summary statistics of residuals from Model 2	51
Table 9.3 Mean residual value by delivery month, Model 2	51
Table 9.4 Ljung-Box Q-test statistics for residual autocorrelation	52

This page intentionally left blank

1. Introduction

In order to try to capture all of the dynamics of a commodity price process, various authors have introduced several multifactor models since the 1990s. Important contributions are for instance the two-factor model by Gibson & Schwartz (1990) the three-factor model by Schwartz (1997), and the three-factor maximal model by Casassus & Collin-Dufresne (2005). Other authors, such as Sørensen (2002) and Lucia & Schwartz (2002), have extended these models in order to capture the seasonal trait that seems to be evident in some commodities.

In this thesis, we estimate the Schwartz & Smith (2000) model both with and without an extended seasonality function, following Sørensen (2002), to investigate whether seasonality is a significant factor in natural gas prices.

Our thesis is structured as follows. First, we provide some qualitative insights into the natural gas market as well as a description of our data set. We then do a simple time series analysis on natural gas spot prices in order to figure out how we might approach modeling these prices. Most non-stationary traits in our data set indicate that the two-factor model of Schwartz & Smith (2000) is a good fit. An exception is what we believe is seasonality in prices. By estimating the original model of Schwartz & Smith (2000) as well as extending it with a seasonal factor, we study how state variables, parameter estimates and residuals are affected, making us able to infer whether seasonality is present in our data set or not.

2. Theory

2.1 Market background and dynamics

Natural gas is one of the most important commodities for producing heat and electricity in American homes and companies. Domestic production contributes to over 95% of the natural gas consumed in the U.S (U.S. Energy Information Administration, 2013a), causing natural gas prices to be driven primarily by domestic supply and demand. Seasonality seems to be present in both the demand and supply side in such a way that prices are higher in the winter and lower in the summer. This could affect natural gas prices causing them to follow a seasonal pattern. Before we introduce more quantitative models to investigate this trait, we first discuss the supply and demand side of natural gas to try and get a qualitative sense of why seasonality might be present.

Figure 2.1 shows the total production, consumption, net imports and net storage withdrawal of natural gas in the United States from January 2001 to August 2013 (U.S. Energy Information Administration, 2013a).

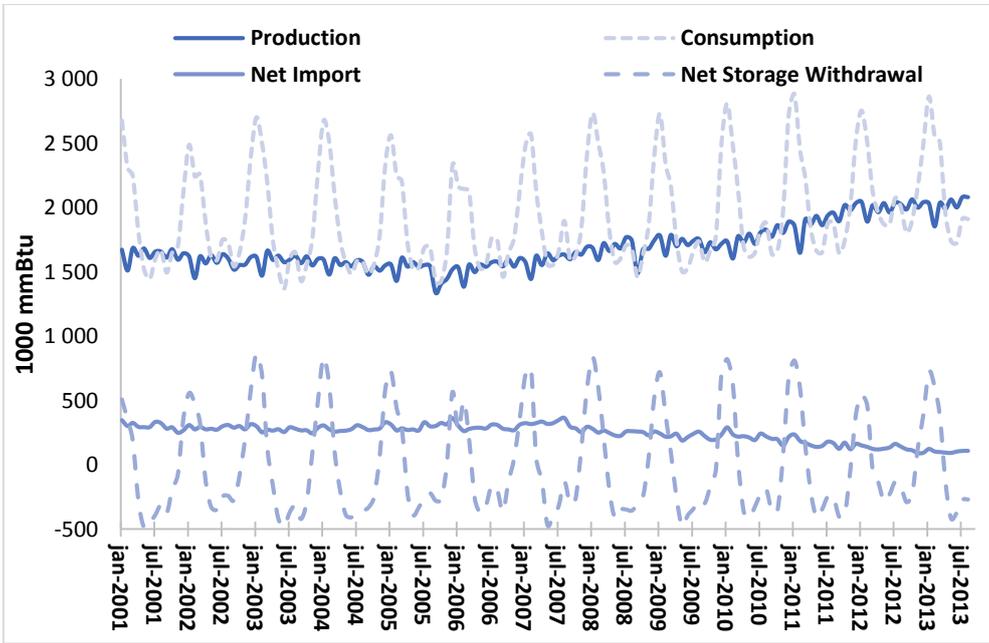


Figure 2.1 Production, consumption, net imports and net storage withdrawal of natural gas in the United States, 2001-2013

Looking at the curves for consumption and net storage withdrawal, we see a clear seasonal pattern. Because of physical limitations on how much gas that can be transported through high-pressure pipelines in one period, gas producers put a fraction of the gas they produce in storage

in low-demand periods of the year (Augustine, et al., 2006). These storage facilities are located all over the US. This makes them able to deliver natural gas locally when demand increases above the pipeline’s capacity. In the months of April to November, gas consumption is at a steady low, with some minor peaks in July and August. In this period, net withdrawals are negative meaning that more gas is stored than what is being used. In the colder winter months, however, gas consumption reaches its yearly maximum. On the storage side, this is shown by positive net withdrawals. Usually, stored natural gas cannot cover the entire excess demand in the winter, and since supply from production is constant, this is likely to put upward pressure on prices in this period (Augustine, et al., 2006). With excess supply in summer months, and lower consumption, prices should be lower. Seasonality is therefore likely to be evident in natural gas prices.

Figure 2.2 shows the major sectors on the demand side in the natural gas market. Looking closer at some of these, we wish to explain why consumption tends to vary with season (U.S. Energy Information Administration, 2013a).

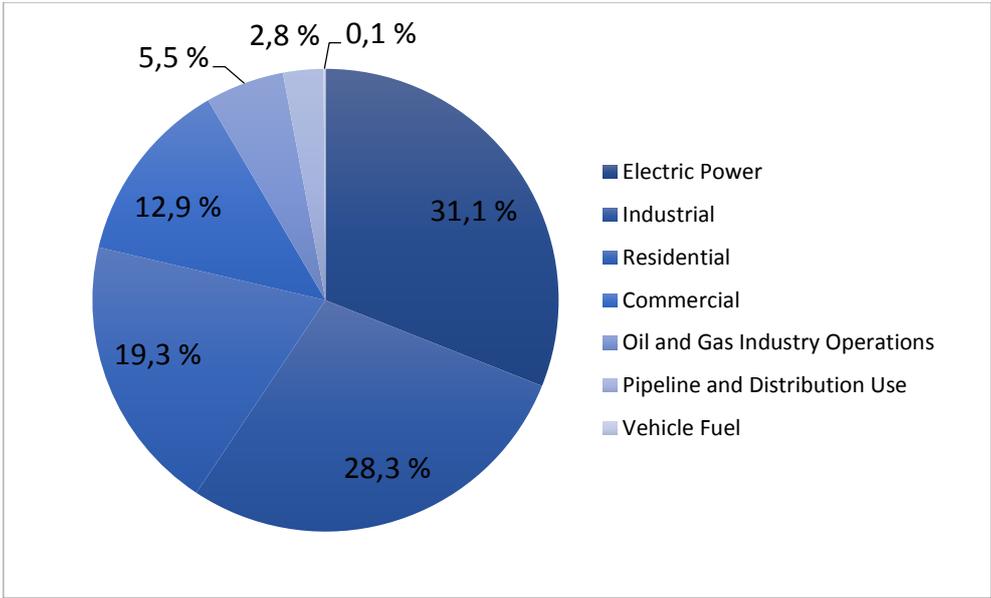


Figure 2.2 Natural gas consumption in the United States in 2012

The industrial sector uses natural gas mainly in production and manufacturing, causing consumption in this sector to be quite constant during the year. On the other hand, the commercial and residential sectors use natural gas for heating, causing their consumption to spike during the cold winter months, while dropping to a low in the warmer summer season. Lastly, there is the electric power sector, where the primary use of natural gas is for air conditioning. This causes the sector’s consumption to be relatively flat over the year, with a

small spike in the warmer summer months of July and August. Over all, demand tends to go up in winter months and down the rest of the year, except for the hottest parts of the summer – a clear seasonal pattern in demand.

The short-run price elasticity of demand is almost inelastic (Bernstein & Madlener, 2011), especially in the commercial and residential sector. Combined with the seasonal patterns on the demand side, this should put an upward pressure on natural gas prices in the winter months and the opposite in the summer.

Other factors also contribute to changes in natural gas prices, and are important to be aware of when we later will model its price dynamics. Unstable weather and unforeseen temperature changes affect prices especially in the short term, while severe weather phenomena like hurricanes and earthquakes can affect both short-term and medium-term prices. These effects affect prices through changes in both demand and supply. The price of substitute commodities like coal and oil, affect prices both in the short and longer term, mostly caused by reduced demand. Fluctuations in the national economy affect both demand and supply, moving long-term gas prices. Breakthroughs in production technology, affects long-term supply, causing downward pressure on prices. The shale gas revolution caused by new technology development is an example of such a disruptive technology.

2.2 How the trading is conducted

Along with the variety of factors that influence the prices of natural gas, the natural gas market is highly competitive, consisting of thousands of producers that sell their gas either to local distribution companies, to marketers, or directly to the customers. The main market center in the U.S. is the Henry Hub (HH) in Louisiana, which is connected to 16 different inter- and intrastate pipelines and is the highest-volume trading point in all of North America. Henry Hub is used as the delivery point for the New York Mercantile Exchange's (NYMEX) natural gas futures contract, and is a pricing reference point for virtually the entire North American natural gas market (Augustine, et al., 2006).

The natural gas futures contracts (NG) traded on NYMEX can be traded for 72 consecutive months starting with the next calendar month. Each contract is for 10,000 million British

thermal units³ (mmBtu) of Natural Gas, and prices are quoted in dollars and cents per mmBtu (CME Group, 2013).

Figure 2.3 shows number of natural gas contracts with 1 month to maturity (NG1) traded on the NYMEX each year between 1990 and 2012 reflecting the markets increasing popularity. The average daily contract volume for NG traded on NYMEX was 390,000 in 2013, making it the most liquid natural gas contract in the world (CME Group, 2013).

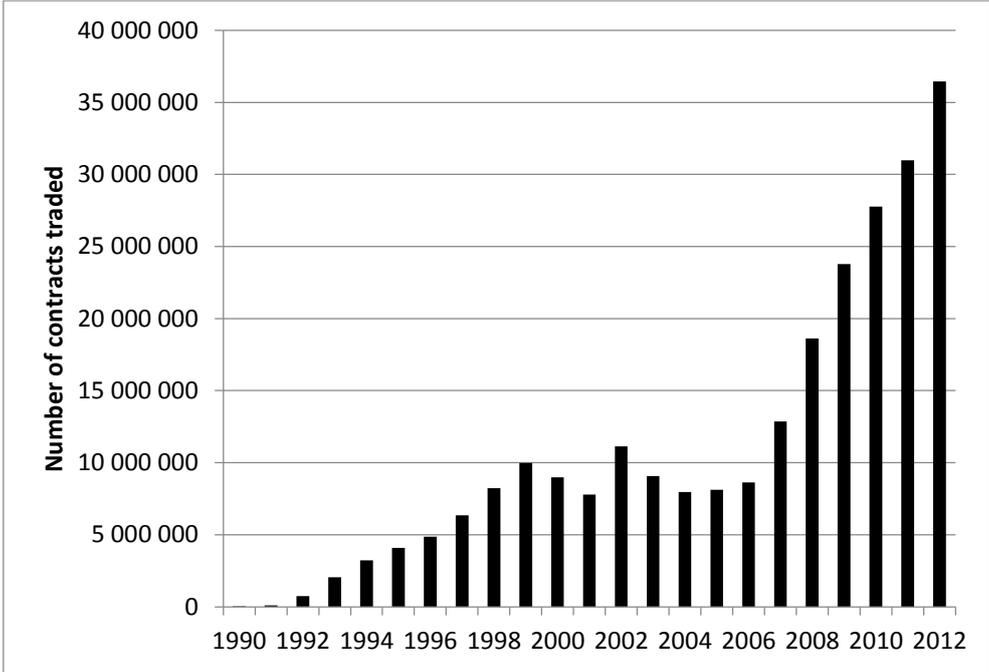


Figure 2.3 NG1 contracts traded by year, 1990*-2012
*Annual volume for 1990 begins in April

2.3 Forward and Futures contracts

Since our data analysis uses futures prices when modeling natural gas prices, and these are highly connected with forward prices, we now wish to highlight some important aspects of forward and futures contracts. A forward contract is an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price (Hull, 2013). Forward contracts are traded in over-the-counter (OTC) markets. The person buying a forward contract holds a long position, while the person selling a forward contract holds a short position.

³ One British thermal unit (Btu) refers to the amount of energy needed to cool or heat one pound of water by one degree Fahrenheit, approximately equal to 1,055 joules.

A futures contract is similar to a forward contract, but differs because they are traded on an organized exchange, with standardized contract terms determined by that particular exchange. Further, a futures contract settles at the end of each trading day, through a margin account. When an investor buys a futures contract, he has to make a deposit, known as initial margin, to this margin account. The size of the deposit is determined by the exchange the trade is conducted at. At the end of each trading day, the margin account adjusts according to the investor's gain or loss. The seller of the futures contract also has a margin account, which changes proportionally to the buyers account. As the futures price usually varies over time, one of the parties involved will have a cumulative loss, while the other has a cumulative gain at the end of each trading day. In order to reduce the cumulative loss, one of the investors can close out his position, by entering into the opposite trade as the original agreement (Hull, 2013). The ability to close out a position causes most futures contract to never end in delivery of the underlying asset. This margin account increases liquidity in futures compared to the forward market.

In the next section, we describe the data set we use to look for seasonality in prices. Next, we will perform a simple time series analysis on the data in order to identify sources of non-stationary factors in prices. It is important to have a clear picture of non-stationarity in prices, when we later model natural gas prices to see if we can infer anything about seasonality.

3. Data

Our thesis' quantitative analysis uses weekly Friday observations of NYMEX Natural Gas Futures, gathered from the open source data provider Quandl (2013). Contracts exist for all months of the year and are listed with maturity of 1 to 72 months. Prices can be collected as far back as April 6th, 1990.

Contract	Average Volume	Dates missing 1990-2013		Dates missing 2005-2013	
NG1	47434	60	5 %	13	1 %
NG2	22639	60	5 %	13	1 %
NG3	11457	60	5 %	13	1 %
NG4	6484	61	5 %	14	1 %
NG5	4369	61	5 %	14	1 %
NG6	3227	62	5 %	14	1 %
NG7	2442	62	5 %	15	1 %
NG8	1893	62	5 %	15	1 %
NG9	1562	64	5 %	16	1 %
NG10	1270	63	5 %	16	1 %
NG11	1003	68	6 %	16	1 %
NG12	870	93	8 %	16	1 %
NG13	707	169	14 %	19	2 %
NG14	503	177	14 %	23	2 %
NG15	397	187	15 %	29	2 %
NG16	290	194	16 %	31	3 %
NG17	238	218	18 %	48	4 %
NG18	215	235	19 %	50	4 %
NG19	187	432	35 %	55	4 %
NG20	158	442	36 %	55	4 %
NG21	118	451	37 %	59	5 %
NG22	112	463	38 %	61	5 %
NG23	97	474	39 %	63	5 %
NG24	82	482	39 %	61	5 %
NG30	31	555	45 %	63	5 %
NG36	21	653	53 %	66	5 %
NG42	11	865	70 %	67	5 %
NG48	7	873	71 %	69	6 %
NG54	5	870	71 %	66	5 %
NG60	4	877	71 %	74	6 %
NG66	1	909	74 %	105	9 %
NG72	0	943	77 %	139	11 %

Table 3.1 Average daily trading volume (1990-2013) and missing dates for selected contracts

Table 3.1 shows the average daily trading volume from 1990 to 2013 for each contract with maturity from 1 to 24 months, and contracts with maturity 30, 36, 42, 48, 54, 60, 66 and 72

months. In addition, it shows the number and percentage of missing data for each contract from April 6th, 1990 to November 1st, 2013 and September 30th, 2005 to November 1st, 2013. The average daily trading volume decreases and number of missing dates increases when maturity increases.

In order to secure a high trading volume, but still include contracts with long maturities necessary for our model, we chose to include contracts with maturity from 1 to 24 months. Looking closer at our data set, we saw that several of the contracts had large gaps before the fall of 2005 (U.S. Department of Energy, 2009a). In order to reduce the risk of these gaps causing problems for our model, we chose to include prices from September 30th, 2005 to November 1st, 2013.

For the dates that were missing⁴, we approximated the values using a linear interpolation⁵ in order to run the model more smoothly. When we later applied our model, we saw that by reducing our number of contracts to every third, we could reduce the running time of our model substantially without affecting the results significantly.

This gave us a final dataset consisting of 423 weekly Friday settle prices ranging from September 30th, 2005 to November 1st, 2013 for eight futures contract with maturity 1, 4, 7, 10, 13, 16, 19 and 22 months.

3.1 Preliminary look at the data

Before modeling prices and doing formal testing, we take a qualitative look at our data to see if it indicates drift, trends, seasonality or other non-stationarities that might be important for our model. The more our model is able to describe the dynamics of natural gas prices, the better we are able to isolate an eventual effect from a seasonal factor.

Figure 3.1 shows the log of weekly futures prices with 1, 10 and 22 months to maturity. At first glance, there seems to be a slight downward trend in the prices from 2008 to 2013. The trend appears to be stochastic, but it could also be a deterministic drift. If prices are efficient, and follow a random walk, such a trend or drift should not be evident. Still, it is hard to judge if

⁴ Appendix A shows the exact dates that were missing.

⁵ We approximated the missing prices using the following linear interpolation (example shows formula used if one date is missing): $x_t = \frac{x_{t+1} + x_{t-1}}{2}$.

such traits are inherent in prices, since longer data samples might generate other results. Formal tests for a random walk, random walk with drift and trend-stationarity should therefore be performed.

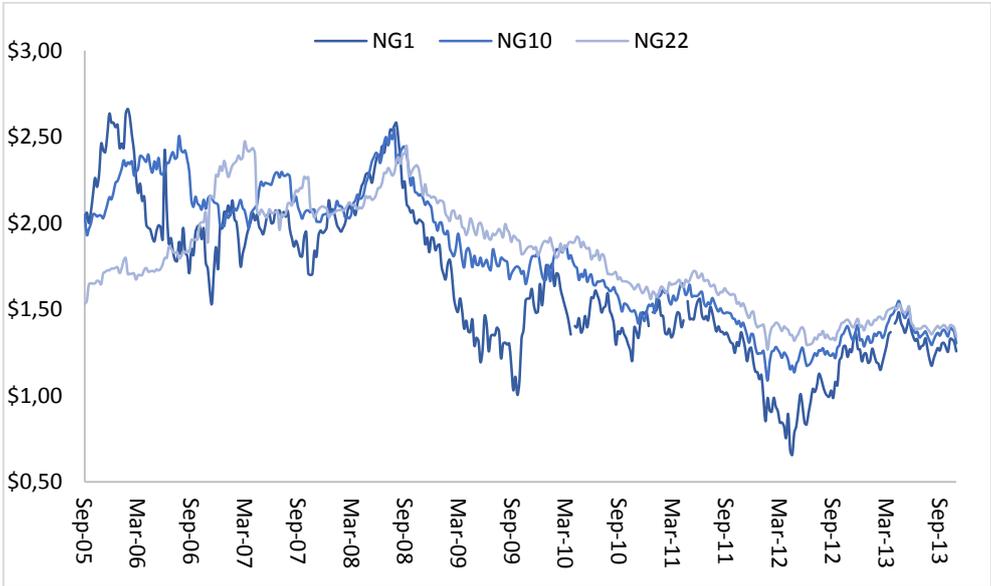


Figure 3.1 Log of weekly future prices for NG1, NG10 and NG22

Further, the 1-month contract seems to fluctuate more than the 22-month contract, which is consistent with the Samuelson hypothesis (1965), arguing that the futures price volatility increases as the futures contract approaches maturity. Time-varying volatility of some sort might therefore be important when modeling natural gas prices.

At first glance, it is hard to deduct any traits of seasonality in Figure 3.1. To get a closer look at this, we calculated average prices of each contract conditional on delivery month. Table 3.2 shows the monthly average prices divided by the yearly average. A percentage higher than 100% indicates that prices are higher in this month than the yearly average. As Table 3.2 shows, most prices seem to be highest when contracts mature in November to March and lowest when maturing in summer months. The one-month contract is the exception, showing less signs of seasonality.

Month	NG1	NG4	NG7	NG10	NG13	NG16	NG19	NG22
Jan	100 %	101 %	108 %	105 %	100 %	103 %	105 %	105 %
Feb	102 %	102 %	106 %	108 %	104 %	103 %	106 %	107 %
Mar	101 %	103 %	102 %	105 %	103 %	103 %	103 %	105 %
Apr	100 %	93 %	96 %	99 %	98 %	94 %	96 %	97 %
Mai	102 %	98 %	95 %	99 %	99 %	97 %	95 %	98 %
Jun	101 %	98 %	98 %	97 %	99 %	97 %	98 %	97 %
Jul	104 %	98 %	93 %	96 %	99 %	98 %	95 %	96 %
Aug	100 %	100 %	97 %	96 %	100 %	101 %	98 %	96 %
Sep	95 %	99 %	97 %	98 %	98 %	99 %	98 %	99 %
Okt	94 %	102 %	97 %	94 %	97 %	100 %	99 %	96 %
Nov	97 %	102 %	104 %	100 %	99 %	103 %	104 %	101 %
Des	106 %	103 %	105 %	103 %	104 %	103 %	105 %	104 %

Table 3.2 Monthly average prices relative to yearly average prices

To look for seasonality in all contracts combined, we also calculated the combined average log price for all contracts conditional on delivery month, shown in Table 3.3.

Month	Mean	Mean %	Median	Min	Max	St.dev
Jan	1,68	105 %	1,90	1,00	1,98	0,39
Feb	1,67	104 %	1,89	1,02	1,96	0,38
Mar	1,65	103 %	1,84	1,02	1,94	0,37
Apr	1,56	97 %	1,76	0,93	1,83	0,36
Mai	1,56	97 %	1,75	0,95	1,81	0,35
Jun	1,56	98 %	1,74	0,97	1,83	0,35
Jul	1,57	98 %	1,75	0,93	1,84	0,36
Aug	1,57	98 %	1,75	0,96	1,84	0,35
Sep	1,57	98 %	1,75	0,97	1,85	0,35
Okt	1,57	98 %	1,76	0,94	1,85	0,36
Nov	1,62	101 %	1,82	0,99	1,90	0,36
Des	1,66	104 %	1,87	1,03	1,94	0,37

Table 3.3 Summary statistics of log prices for all contracts, conditional on delivery month

The percentages in the third column are calculated by dividing the mean for each month by the yearly average. This once again shows that prices seem to follow a seasonal pattern, being higher than the yearly average in the winter and lower in the summer. Figure 3.2 provides a plot of the average prices found in Table 3.3 in order to show the pattern more visually.

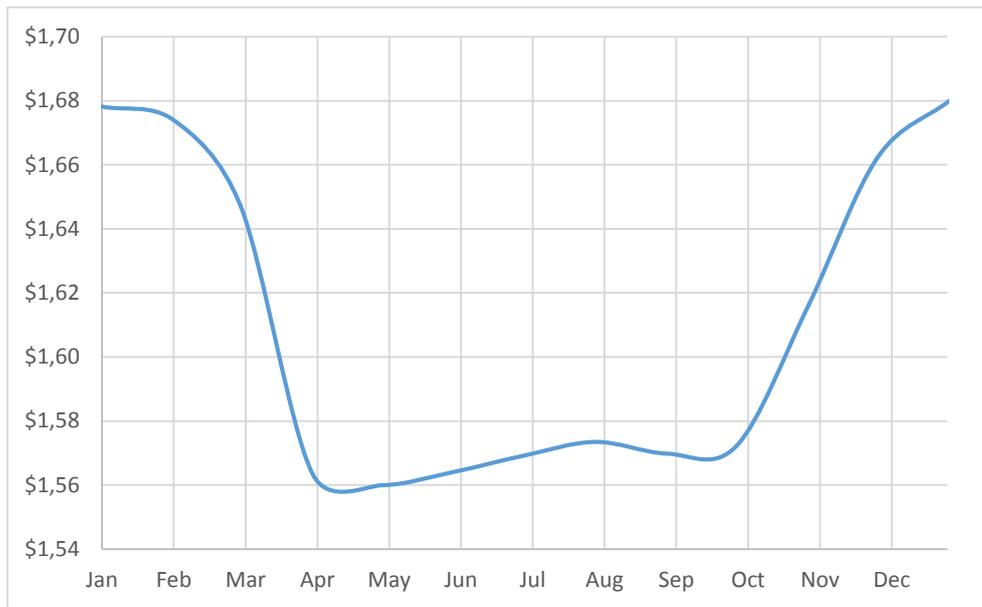


Figure 3.2 Average log prices for all contracts, conditional on month

3.2 Augmented Dickey-Fuller test

To formally test for stationarity in our data set, we use the augmented Dickey-Fuller (ADF) test. This involves three different regression equations to test for the presence of a unit root (Enders, 2009). Figuring out which type of stationarity that is inherent in prices is important for modeling purposes. If prices turn out to be non-stationary, then explanatory factors, such as seasonality, might be significant. If prices follow a random walk or a random walk with trend, an AR(1) process or its continuous analog Geometric Brownian Motion can be used to model their dynamics. If prices follow a random walk with trend, they will be mean reverting, opening the possibility for the Ornstein-Uhlenbeck process as a good fit.

Table 3.4 shows the result we obtained using Matlab’s “adftest” to run the three ADF tests on a time series of weekly log spot prices⁶. For the first and second ADF, we fail to reject the null hypothesis that the time series is non-stationary. For the third ADF however, we reject the null hypothesis in favor of the time series being a trend-stationary process.

⁶ We did similar tests on of the futures contracts’ prices in our data series, obtaining similar results. Since futures prices depend on spot prices, and the results are similar, we chose to only include the spot price results in this section.

Test	P-value	ϕ (S.E.)	c (S.E.)	δ (S.E.)	BIC	Adj R^2	Ljung-Box Q of residuals (C.V.) ⁷
AR ⁸	18.06%	0.9963 (0.0029)			-757.48	95.35%	91.50 (31.41)
ARD ⁹	8.39%	0.9713 (0.0108)	0.0424 (0.0177)		-757.17	95.19%	87.31 (31.41)
TS ¹⁰	2.88%	0.1283 (0.0385)	0.9376 (0.0172)	-0.0002 (0.0001)	-757.45	95.24%	83.11 (31.41)

The P-value is given from the F-test. The null hypothesis is rejected for P-value > 5%

Table 3.4 Test results from Augmented Dickey-Fuller and Ljung-Box tests

It is difficult to conclude which results are most valid, since both the adjusted R^2 , BIC¹¹ and parameter significance were quite similar in all three tests. A usual practice in such cases is to select the most parsimonious model as a description of the time series. Either way this leads us to conclude that natural gas prices are indeed non-stationary. At the same time, the series might inhibit mean-reverting tendencies. The three tests therefore indicate that both a Geometric Brownian Motion and an Ornstein-Uhlenbeck process might be used to model the time series' dynamics.

If the time series can be fully explained using the processes in the tests, residuals should be generated from a white noise process, being independent and having a constant mean and variance. A white noise would therefore indicate that other factors are not present in the data series. This would reject the possibility of seasonality in the data. We therefore performed a Ljung-Box test for autocorrelation in residuals, shown in the rightmost column of Table 3.4.

We were able to reject a null hypothesis of no autocorrelation when performing the Ljung-Box-test on the residuals from all three ADF tests. This leads us to conclude that the time series is not solely generated from a Geometric Brownian motion or an Ornstein Uhlenbeck process. This opens the possibility for other factors, like seasonality, to be present in prices.

⁷ Critical value in Ljung-Box is based on a significance level of 95%

⁸ $H_0: y_t = y_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2), H_A: y_t = \phi y_{t-1} + \epsilon_t, \phi < 1$

⁹ $H_0: y_t = y_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2), H_A: y_t = c + \phi y_{t-1} + \epsilon_t, \phi < 1$

¹⁰ $H_0: y_t = c + y_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2), H_A: y_t = c + \delta t + \phi y_{t-1} + \epsilon_t, \phi < 1$

¹¹ Bayesian Information Criteria/Schwartz criterion

4. Model Description

4.1 Why the Schwartz & Smith model

In order to investigate further if seasonality is a significant feature in natural gas futures prices, we wish to use a model that can capture and isolate important features of the natural gas market. As mentioned earlier, both long-term changes like disruption in production technology, and short-term temporary shocks caused by weather and other temporary disturbances, affects supply and demand in the natural gas market. Further, our time series analysis show signs of both mean reversion and a random walk in spot prices.

The two-factor model of Schwartz & Smith (2000) therefore seems to fit the natural gas market fairly well, with its short-term Ornstein Uhlenbeck- and long-term Geometric Brownian motion dynamics. In addition, this model is simple enough to be extended with a deterministic seasonality function to try and isolate the effect of seasonal variation. By estimating the model both with and without a seasonal factor, we can get a sense of whether or not seasonality contributes to the non-stationarity in natural gas futures prices.

4.2 Overview – Model 1

The model presented by Schwartz & Smith (2000) is a state-space model, which decompose the log spot price on a commodity into two unobservable stochastic variables, χ and ξ , each one of them evaluated at time t :

$$\ln(S_t) = \chi_t + \xi_t$$

Changes in the short-term factor χ_t represent temporary changes in price, which can be caused by several factors. This could for instance be difficulty in delivery of the commodity, extreme weather conditions, or unforeseen changes in demand. In essence, these changes are short-term and temporary, contrary to the ξ -factor.

Temporary short-term changes are assumed to revert back to a mean over time. χ_t therefore follows an Ornstein-Uhlenbeck process with the following dynamics:

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dz_\chi,$$

where the parameter κ is a mean-reversion coefficient that reflects how fast prices reverts back to approach the mean level of the factor.

Changes in the equilibrium price level ξ_t are affected by overall macroeconomic quantities, technology factors etc. In natural gas prices interest rates, inflation, and of course the shale-gas revolution can be thought of as important examples.

ξ_t therefore follows a Geometric Brownian Motion process:

$$d\xi_t = \mu_\xi dt + \sigma_\xi dz_\xi$$

This process has a constant drift term, and a dispersion term containing a Wiener process. The two Wiener processes dz_ξ and dz_χ are correlated with $dz_\xi dz_\chi = \rho_{\xi\chi} dt$.

Estimating the model based on futures contract prices with different maturities makes us able to estimate the two factors. The model is set up such that changes in contracts with long-term maturities give us information about changes in the equilibrium price factor, while changes in the difference between near- and long-term futures prices give information about the short-term factor. (Schwartz & Smith, 2000)

4.3 Deriving the distribution of the spot price

As we are modeling futures prices, and the spot price is an important part of the futures curve, we have to obtain the distribution of the spot price in our model. Given initial values of χ_0 and ξ_0 , Schwartz & Smith (2000) show that the two state variables are jointly normally distributed with the following expectation and covariance:

$$E[\chi_t, \xi_t] = [e^{-\kappa t} \chi_0, \xi_0 + \mu_\xi t]$$

$$Cov[\chi_t, \xi_t] = \begin{bmatrix} \frac{(1 - e^{-2\kappa t})\sigma_\chi^2}{2\kappa} & \frac{(1 - e^{-\kappa t})\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} \\ \frac{(1 - e^{-\kappa t})\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} & \sigma_\xi^2 t \end{bmatrix}$$

Given the same initial values, they continue to show that the log of the future spot price is normally distributed with:

$$E[\ln(S_t)] = e^{-\kappa t} \chi_0 + \xi_0 + \mu_\xi t$$

$$Var[\ln(S_t)] = \frac{(1 - e^{-2\kappa t})\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 t + \frac{2(1 - e^{-\kappa t})\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}$$

From this, Schwartz & Smith (2000) infer that the spot price is log-normally distributed with:

$$E[S_t] = e^{E[\ln(S_t)] + \frac{1}{2}Var[\ln(S_t)]} \quad 12$$

By taking the logarithm of both sides, they obtain:

$$\begin{aligned} \ln[E[S_t]] &= E[\ln(S_t)] + \frac{1}{2}Var[\ln(S_t)] \\ &= e^{-\kappa t}\chi_0 + \xi_0 + \mu_\xi t + \frac{1}{2}\left(\frac{(1 - e^{-2\kappa t})\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 t + \frac{2(1 - e^{-\kappa t})\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) \end{aligned}$$

This equation shows every parameter in our model's contribution to the spot price. If the state variables and model parameters are estimated with confidence, it will indicate that our model is able to capture the non-stationary effects of natural gas prices. If this model can explain all variation in our natural gas prices, it would reject the possibility of a seasonal effect in prices. If it cannot explain all the variation, a seasonal factor might be present.

4.4 Valuing futures contracts using the model

We have now specified the dynamics and distribution of our short-term/long-term model. Since we estimate the model parameters based on futures data, we need a general expression of futures prices given that our observed prices follow the model's distribution. According to Hull (2013, pp. 111-112), "when the short-term risk free interest rate is constant, the forward price for a contract with a certain delivery date is in theory the same as the futures price for a contract with the same delivery date". Consequently, we will derive the general expression of the futures price using the price of a forward contract.

The price c_t of any derivative $h(T)$ is given by the following equation:

$$c_t = e^{-r(T-t)}E_t^Q[h(T)]$$

¹² This can be derived using the fact that if $X = \ln(S_t)$ is $N(\mu, \sigma^2)$, then

$$Y = e^X = e^{\ln(S_t)} = S_t \text{ is } N\left(\mu + \frac{1}{2}\sigma^2, e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)\right)$$

If we consider a long forward contract maturing at date T with delivery price $F_{t,T}$, the value of this contract at maturity is $[S_T - F_{t,T}]$, where S_T is the spot price of the underlying asset at maturity. Since one does not pay anything up front to enter a forward contract, we can write the equation above as

$$0 = e^{-r(T-t)} E_t^Q [S_T - F_{t,T}].$$

As the forward price is known at time t , this transforms to

$$0 = e^{-r(T-t)} E_t^Q [S_T] - e^{-r(T-t)} F_{t,T},$$

which reduces further to

$$F_{t,T} = E_t^Q [S_T],$$

showing that forward prices (and subsequently futures prices) are equal to the expected future spot price under the risk-neutral process.

Schwartz & Smith (2000) use this result and the risk neutral process of S_T to derive the following futures price¹³:

$$\begin{aligned} \ln(F_{0,T}) &= \ln(E^Q[S_T]) \\ &= E^Q[\ln(S_T)] + \frac{1}{2} Var^Q[\ln(S_T)] \\ &= e^{-\kappa T} \chi_0 + \xi_0 + A(T), \end{aligned}$$

where

$$A(T) = \mu_{\xi}^*(T) - \frac{(1-e^{-\kappa T})\lambda_{\chi}}{\kappa} + \frac{1}{2} \left(\frac{(1-e^{-2\kappa T})\sigma_{\chi}^2}{2\kappa} + \sigma_{\xi}^2(T) + \frac{2(1-e^{-\kappa T})\rho_{\xi\chi}\sigma_{\chi}\sigma_{\xi}}{\kappa} \right) \quad 14$$

¹³ See Appendix B for the full derivation

¹⁴ From now on we denote the risk-corrected drift of the equilibrium level as $\mu_{\xi}^* \equiv \mu_{\xi} - \lambda_{\xi}$.

We now have the distributions and SDEs of our model of commodity future prices, in addition to an analytical solution of what we should expect the futures price to be given that the observed futures prices evolve according to our model.

The next step is to estimate the original model using observed futures prices, by applying an iterative method called the Kalman filter. In order to use this method, we have to write our model in so-called state-space form. We do this in the next section.

4.5 A state-space formulation of the Schwartz & Smith – model

State-space models allow one to model and observe time series as being explained by a vector of state variables, each following a stochastic process. The state variables can be both observed and unobserved. One first selects which factors that are driving the phenomena in our model - here a long-term and short-term factor. Then one specifies how they are combined in order to yield the economic quantity the phenomenon is measured by. This gives us a measurement equation that is affected by some kind of noise. Next one also has to have some insight into how these state-variables evolve over time, which is stated in the transition equation – also affected by a noise term.

Writing our model in state-space form therefore involves using two equations – the measurement- and transition equation.

Measurement equation

The measurement equation of our two-factor model can be written as:

$$y_t = d_t + F_t'x_t + v_t, t = 1, \dots, n_T,$$

where

$$y_t \equiv [\ln F_{T_1}, \dots, \ln F_{T_n}]$$

is an $n \times 1$ vector of observed log futures prices with time to maturity T_1, T_2, \dots, T_n ;

$$d_t \equiv [A(T_1), \dots, A(T_n)] \text{ is an } n \times 1 \text{ vector;}$$

$$F_t \equiv [e^{-\kappa T_1} \mathbf{1}, \dots, e^{-\kappa T_n} \mathbf{1}] \text{ is an } n \times 2 \text{ matrix;}$$

$$x_t = [\chi_t, \xi_t] \text{ is a } 2 \times 1 \text{ vector; and}$$

v_t is an $n \times 1$ vector of serially uncorrelated, normally distributed disturbances with

$$E[v_t] = 0 \text{ and } \text{Cov}[v_t] = V$$

The measurement equation describes the relation between the futures prices we observe, y_t , and what we should expect the futures prices to be given our analytically solved futures price, given by the terms $d_t + F_t'x_t$.¹⁵ We also assume that the observed futures prices, y_t , are measured with error, reflected in the measurement error term v_t . In this case, this error term can be thought of as noise in the observed futures prices, caused for instance by low trading volume, transaction costs, mistyped data etc.

¹⁵ One can easily observe that the right hand side of the measurement equation equals the analytical futures price from the last section, only written in matrix form for several maturities.

Transition equation

The transition equation of the Schwartz & Smith (2000)-model is defined as:

$$x_t = c + Gx_{t-1} + w_t, t = 1, \dots, n_T$$

where

$x_t \equiv [\chi_t, \xi_t]$ is a 2×1 vector of the state variables;

$c \equiv [0, \mu_\xi \Delta t]$ is a 2×1 vector;

$G \equiv \begin{bmatrix} e^{-\kappa \Delta t} & 0 \\ 0 & 1 \end{bmatrix}$ is a 2×2 matrix;

w_t is a 2×1 vector of serially uncorrelated, normally distributed disturbances with

$$E[w_t] = 0 \text{ and } Var[w_t] = W \equiv Cov[(\chi_{\Delta t}, \xi_{\Delta t})]$$

$\Delta t \equiv$ the length of the time steps; and

$n_T \equiv$ the number of time periods in the data set.

This transition equation describes the evolution of the state variables using their previously defined expectations. The term w_t is assumed to reflect that also the state variables are subject to some kind of noise. We observe that $c + Gx_{t-1}$ is equal to the expectation of (χ_t, ξ_t) , whereas the term w_t makes sure that the covariance structure of the state variables are equal to what we defined earlier.

Since both the measurement equation and transition equation is affected by noise, this gives us a challenge when estimating the system's parameters later on.

5. Estimation of our state space model

The parameters in both the measurement and the transition equation are unknown and have to be estimated. The only concrete information we have starting out is the futures prices for each contract and the two state variables given by our model. Since our model is based on unobservable variables, we cannot infer parameter values using ordinary methods, such as least squares estimation¹⁶. This means that we have to use another method. The solution is a likelihood-based inference called the Kalman filter. The method is named after Rudolf E. Kálmán, and is described in detail in (Kalman, 1960).

5.1 A combination of maximum likelihood and the Kalman Filter

The Kalman filter is an iterative procedure that allows us to construct a likelihood function associated with our state-space model. It is designed to produce estimates of unobserved variables using observed data that is assumed to contain noise. Before deriving the Kalman filter, we need to construct a likelihood function, to be able to estimate parameters.

Assume that we observe a time series $\{y_t\}_{t=1}^T$, of futures prices that we want to describe by the measurement- and transition equation we previously defined. Assume that we have values of the model's parameter set, $\theta = \{\kappa, \sigma_\chi, \lambda_\chi, \mu_\xi, \sigma_\xi, \mu_\xi^*, \rho_{\xi\chi}, \sigma_H\}$. Let the sample density associated with the state-space model containing the parameters θ be denoted by $f(y_1, y_2, \dots, y_T | \theta)$. Since all observations have the same distribution and are independent, we can write this density function as

$$f(y^T | \theta) = f(y_1, \theta) f(y_2 | y_1, \theta) f(y_3 | y_2, y_1, \theta) \dots f(y_T | y_{T-1}, \dots, y_1, \theta) \quad 17$$

This sample density is often represented as

$$L(\theta; y^T) = f(y^T | \theta) = \prod_{t=1}^T f(y_t | y^{t-1}, \theta).$$

¹⁶ The only way one could use least squares estimation, is by fitting the parameters of the futures price formula such that it fits observations of several futures curves. This would lead to estimates of the risk-neutral parameters of our model, but we would not obtain the values of the state variables or the parameters related to the physical process. We need both the state variables and the physical drift parameter when we look for seasonality traits later.

¹⁷ $y^T = \{y_1, y_2, \dots, y_T\} \forall T \geq 1, y^0 = \emptyset$

By holding y^T fixed and varying the parameters θ , we can maximize the likelihood of obtaining the “correct” parameter estimates. This is usually referred to as maximum likelihood estimation.

In order to construct the likelihood function, we thus need to derive the densities:

$$f(y_t|y^{t-1}, \theta), \quad t = 1, 2, \dots, T$$

Since the system is linear and its errors Gaussian, these densities can be obtained using the Kalman filter.

5.2 Kalman filtering

Kunst (2007) describes Kalman filtering as an iterative procedure involving the following four steps, which we have modified to fit our model:

1. Initialization step

The first step when running the Kalman filter is to provide some initial values. The algorithm needs starting values of the transition equation $x_0 = [\chi_0, \xi_0]$ and an estimate of its covariance matrix C_0 . Using these initial values, the Kalman filtering process can start with the prediction step.

2. Prediction step

We first estimate the mean and covariance matrix of (χ_t, ξ_t) , conditional on what we know at period $t-1$ (starting at $t-1 = 0$):

$$E[(\chi_t, \xi_t)|(\chi_{t-1}, \xi_{t-1})] = a_t \equiv c + Gm_{t-1}$$

$$Cov[(\chi_t, \xi_t)|(\chi_{t-1}, \xi_{t-1})] = R_t \equiv G_t C_{t-1} G_t' + W$$

These conditional expectations of the state variables are based on the transition equation, which we previously defined in the state-space formulation of our model. We define m_t under step 3.

Using the observed futures prices in our data set, y_t , we can construct the forecast error at time t by subtracting the estimated value:

$$u_t = y_t - E(y_t|y_{t-1}) = y_t - (F_t' a_t + d_t)$$

We observe that the latter term is the conditional expectation of the measurement equation.

Since these forecast errors are Gaussian, it follows that $u_t \sim N(0, V + F_t' R_t F_t)$. Furthermore, since we can write $y_t = u_t + E(y_t | y_{t-1})$, it follows that $f(y_t | y^{t-1}; \theta) = f(u_t; \theta)$. We have thus shown that the distribution of the forecast errors u_t is equal to the distribution of the conditional expectation of the futures prices given all information up to time t-1 and the parameter set θ .

Given a_t and R_t , we can now compute $f(y_t | y^{t-1}, \delta)$ from the normal density function:

$$f(y_t | y^{t-1}, \theta) = f(u_t; \theta) = \frac{1}{\sqrt{(2\pi) * |V + F_t' R_t F_t|}} e^{-\frac{u_t' (|V + F_t' R_t F_t|)^{-1} u_t}{2}}$$

Consequently, to compute an estimate of the next set of futures prices, $f(y_{t+1} | y^t, \theta)$, we need the expectation and covariance structure of our state variables conditional on all information up until time t. This is given by:

$$m_t = E[(\chi_t, \xi_t) | (\chi_t, \xi_t)]$$

$$C_t = Var[(\chi_t, \xi_t) | (\chi_t, \xi_t)]$$

3. Correction step

Observing the real futures prices y_t , we can update the predictions a_t and R_t according to the Kalman (1960) formulae:

$$m_t = E[(\chi_t, \xi_t) | (\chi_t, \xi_t)] \equiv a_t + A_t(y_t - f_t)$$

$$C_t = Var[(\chi_t, \xi_t) | (\chi_t, \xi_t)] \equiv R_t - A_t Q_t A_t',$$

where

$$f_t = E[y_t | y^{t-1}, \theta] \equiv d_t + F_t' a_t$$

$$Q_t = Cov[y_t | y^{t-1}, \theta] \equiv F_t' R_t F_t + V,$$

and

$$A_t \equiv R_t F_t Q_t^{-1}$$

f_t and Q_t are the expectation and covariance of the period-t log futures prices given information up to time t-1. The matrix A_t corrects the predicted state variables, a_t , based on the difference between the log prices observed at time t, y_t , and the predicted time-t price vector f_t .

The corrected prediction m_t , is thus a linear combination of the state variable's expectation conditional on time $t - 1$, a_t , and the current prediction error $y_t - f_t$. The larger the variances in the estimates of the state variables, the larger A_t becomes, and more weight is focused on the observed difference between the real and estimated futures price. A_t is chosen such that it minimizes the state variables' prediction error variance. m_t is thus the best estimate we have of the state variables' values, based on information up until time t.

4. Likelihood construction

By going through the prediction and correction step for each observation in the data set, one obtains a value of $f(y_t|y^{t-1}, \theta)$ in each step. We are thus able to construct the likelihood function as:

$$L(y^T, \theta) = \prod_{t=1}^T f(y_t|y^{t-1}, \theta) = \prod_{t=1}^T \frac{1}{\sqrt{(2\pi) * |V + F_t'R_tF_t|}} e^{-\frac{u_t'(|V + F_t'R_tF_t|)^{-1}u_t}{2}}$$

The best estimate for the parameter set θ is found by maximizing this function with respect to the model's parameter set. A useful result here is that if one maximizes the logarithm of a function with respect to a parameter set, it has the same solution as if one maximizes the function itself, before taking the logarithm. We therefore want to maximize the following log-likelihood function, in order to obtain our parameter set θ :

$$\ln L(y^T, \theta) = -\frac{T * n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln|V + F_t'R_tF_t| - \frac{1}{2} \sum_{t=1}^T u_t'|V + F_t'R_tF_t|^{-1}u_t$$

where n is the number of futures contracts used.

5.3 Kalman Filtering routine in Matlab

The lecture by Fusai (2012), as well as the Kalman filtering explanation in Sørensen (2002) and Schwartz & Smith (2000), made it easier for us to understand the Matlab syntax and Kalman filtering algorithm.

To estimate our model, we created a Matlab code that performs Kalman filtering through the same steps described in the previous section, based on our data set of futures prices. The code iterates successively between each observation of the futures curve, doing the prediction and estimation step, giving value to the log-likelihood function.

To find the optimal parameter values, we used Matlab `fmincon`¹⁸ function to optimize the log-likelihood function by changing the parameter set θ around its pre-set initial values. The function will only find a local maximum around an initial set of parameters. We therefore selected initial parameters randomly, and made Matlab find the local maximum of the log-likelihood around these values. After doing this several times, we used the parameter set with the highest log likelihood as the best estimate of our parameter set.

The process was very time consuming and took up a lot of processing power¹⁹. To try to limit the amount of computer time, we put the following constraints on which random values the initial values could take²⁰:

Parameter	Boundary
κ	[0, 4]
σ_{χ}	[0, 1]
λ_{χ}	[-1, 1]
μ_{ξ}	[-1, 1]
σ_{ξ}	[0, 1]
μ_{ξ}^*	[-1, 1]
$\rho_{\chi\xi}$	[-1, 1]

Table 5.1 Bounds on random initial values

¹⁸ This function actually finds a minimum of a function, but finding the minimum of $-f(x)$ is the same as finding the maximum of $f(x)$.

¹⁹ For instance, running one optimization routine with one set of initial values, takes about 25 seconds using data for the original Schwartz & Smith (2000) article on a double four-core processor and 8 gigabytes of RAM. With millions of combinations of initial values, checking all of them would take a substantial amount of time.

²⁰ The values of σ_{χ} , σ_{ξ} and $\rho_{\chi\xi}$ are always between 0 and 1. The drift coefficients μ_{ξ} , μ_{ξ}^* and the drift reduction term λ_{χ} usually have percentage values on an annual basis. They will therefore seldom have a value greater than 1 in absolute terms. We seldom found a κ greater than 2 when trying out different initial values and optimizing parameter sets. We therefore set the range of κ to between 0 and 4.

5.3.1 Obtaining standard errors of parameter estimates by inverting the Hessian

An advantage of using Matlab's `fmincon` function is that it can output the Hessian matrix from each optimization routine. The Hessian matrix contains the second derivatives of the Lagrangian expression of a constrained optimization problem. A very useful result is that by taking the inverse of the Hessian matrix, the elements on the diagonal form the squared standard errors of the parameters in the optimization problem. We are therefore able to measure the confidence of our parameter estimates θ :

$$\theta_{SE} = \sqrt{\text{diag}(H^{-1})}$$

5.3.2 Forcing measurement errors to be zero

In Schwartz & Smith (2000), they assume that the covariance matrix for the measurement errors (V) is diagonal with elements $(s_1^2, s_2^2, \dots, s_n^2)$. In their estimation routine, the authors have forced the model to have close to zero measurement error in the 13-month contract, by forcing s_4^2 to be equal to zero. This choice was done in order to yield the highest maximum-likelihood value. With N futures contracts in the data set, this will lead to $N!$ different combinations of measurement errors that can be set equal to zero. Adding this to the possible millions of combinations in our random parameter set would demand a huge amount of computer power.

When we implemented our models, we instead did a simplified analysis. We set a constant starting value for $\theta, \chi_0, \xi_0, m_0$ and C_0 , varying only the measurement errors forced to zero. We then compared the values of the optimized likelihood functions. The combination yielding the highest log-likelihood determined which contracts we forced to have zero measurement errors.

6. A replication of the Schwartz & Smith – model

In order for us to be certain that our code is working correctly, we used it to estimate the parameter set in the original Schwartz & Smith (2000) article based on the same data. We downloaded weekly NYMEX oil futures prices ranging from 1/2/90 to 2/17/95 with the same maturities of 1, 5, 9, 13 and 17 months from Quandl (2013). After running the optimization routine 2500 times with randomized initial values, the following parameter estimates had the highest log-likelihood value:

Parameters	Optimized values	S&S (2000) values
κ	1.5040 (0.0403)	1.49 (0.03)
σ_χ	0.3223 (0.0155)	0.286 (0.01)
λ_χ	0.0604 (0.0304)	0.157 (0.144)
μ	-0.0324 (0.0628)	-0.0125 (0.0728)
σ_ξ	0.1642 (0.0074)	0.145 (0.005)
μ_ξ^*	0.0085 (0.0020)	0.015 (0.0013)
$\rho_{\chi\xi}$	0.4292 (0.0623)	0.30 (0.0044)
s_1	0.0427 (0.0002)	0.042 (0.002)
s_2	0.0053 (0.0000)	0.006 (0.001)
s_3	0.0033 (0.0000)	0.003 (0.000)
s_4	0 (0)	0 (0.000)
s_5	0.0039 (0.0000)	0.004 (0.000)
Ln L	4040	5140

Standard errors in parenthesis

Table 6.1 Parameter results from replicating Schwartz & Smith (2000) data

We observe that most of the parameters have similar values as the ones in Schwartz & Smith (2000), but there are some differences in the reliability of the estimates. The estimate of λ_χ is much more reliable from our estimation, whereas μ_ξ^* , which contains information about λ_ξ , is equally unreliable as in Schwartz & Smith (2000).

In the article they refer to this as being caused by the fact that price expectations are unobserved quantities. They are therefore difficult to estimate, and require a lot more data to be estimated precisely.

The estimate of $\rho_{\chi\xi}$ is less reliable. The parameter value of the correlation coefficient is still within two standard deviations away from the estimate in Schwartz & Smith (2000).

We believe that if we had more time, we could run the optimization routine for a longer period of time, and thus get more accurate parameters. We do however see that the model errors fit the data in pretty much the same way as in the article:

	Mean error	Standard deviation of error	Mean absolute error
1 Month	-0.0061 (-0.0053)	0.0413 (0.0414)	0.0306 (0.0314)
5 Month	0.0004 (0.0005)	0.0035 (0.0044)	0.0027 (0.0035)
9 Month	-0.0002 (-0.0002)	0.0029 (0.0025)	0.0023 (0.0020)
13 Month	0 (0)	0 (0)	0 (0)
17 Month	0 (0)	0.0038 (0.0035)	0.0030 (0.0028)

Article values in parenthesis

Table 6.2 Errors in the model fit to the logarithm of futures prices

We observe that our log-likelihood is different from the result in in Schwartz & Smith (2000), which might be caused by differences in the data set. In Schwartz & Smith (2000), the authors used 259 sets of futures prices ranging from 1/2/1990 to 2/17/95. In this analysis we used 268 sets of futures prices, indicating that more prices from the period are available today. If some prices in the data had errors that were corrected after 2000, this would affect our estimations. In addition, a difference in length in the time series would affect the value of the log likelihood. There is no summary statistics of the data in Schwartz & Smith (2000), so a further comparison of the two data sets is difficult.

In Figure 6.1 and Figure 6.2 we show the estimated spot and equilibrium prices given by our estimates, compared to the original figure from Schwartz & Smith (2000). The estimated spot price and equilibrium price are given by $e^{\chi_t + \xi_t}$ and e^{ξ_t} respectively.

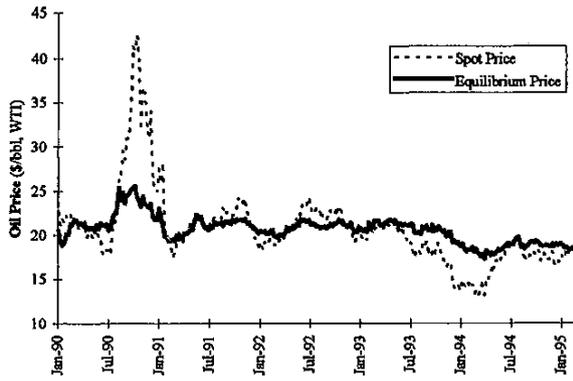


Figure 6.1 Estimated Spot and Equilibrium Prices for the Futures Data (Schwartz & Smith, 2000)

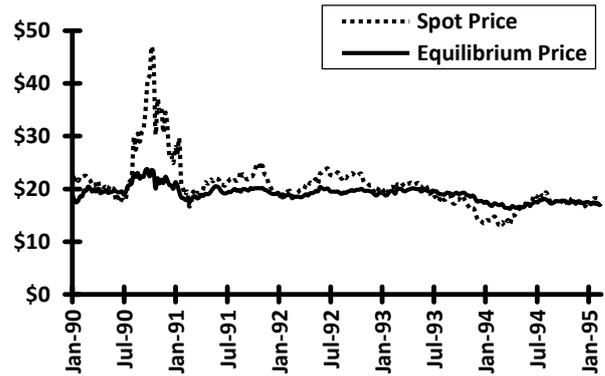


Figure 6.2 Estimated Spot and Equilibrium Prices for the Futures Data

Based on these results, we conclude that our estimation of the state parameters and parameter set θ are similar to those in the original article of Schwartz & Smith (2000). This is a good indication that our estimation routine in Matlab is working correctly.

7. Results from Model 1

After running the Kalman filter optimizing procedure 5000 times, using random initial values of the parameter set θ , $[\chi_0, \xi_0]$ and C_0 we got the following results:

Parameters	Optimized values
κ	0.6308 (0.0113)
σ_χ	0.6664 (0.0113)
λ_χ	-0.2288 (0.1024)
μ_ξ	-0.1530 (0.0906)
σ_ξ	0.3046 (0.0120)
μ_ξ^*	-0.1436 (0.1546)
$\rho_{\chi\xi}$	-0.7120 (0.0234)
s_{1M}	0.1262 (0.0011)
s_{4M}	0 (0)
s_{7M}	0.0822 (0.0005)
s_{10M}	0.0613 (0.0002)
s_{13M}	0.0653 (0.0003)
s_{16M}	0.0809 (0.0005)
s_{19M}	0.0834 (0.0005)
s_{22M}	0.0108 (0.0001)
Log Likelihood	4309.2

Standard errors in parenthesis

Table 7.1 Parameter results from Model 1

If seasonality is indeed present in natural gas prices, this could cause an omitted variable bias in the model affecting its parameter values. It is thus useful to analyze and compare our results both before and after we introduce the seasonality factor.

The short-term and long-term volatilities, σ_χ and σ_ξ , are estimated with confidence, and the larger value of σ_χ vs σ_ξ , seems to fit well with the Samuelson hypothesis, stating that the volatility of futures contracts increase when approaching maturity. We do observe that the volatilities are quite high, especially compared to the estimated oil price volatilities in Schwartz & Smith (2000) of 28.6% and 14.5% respectively. Natural gas prices may be twice as volatile as oil prices, but one should also consider the possibility that seasonality could increase volatility through large variation in prices over the course of the year.

The positive and significant value of the mean reversion parameter, κ , indicates that the short-term factor, χ_t , exhibits mean reversion in its dynamics. Consider now that seasonality is present, and leads to a period of high prices above the long-term mean in winter months. Reverting to the annual mean would take a longer time if prices stay high the entire winter. If seasonality was not present, the mean reversion should be faster. Thus the value of κ might be downward biased in this model if seasonality is present in natural gas prices.

Looking at the estimates of the drift rates, μ_ξ and μ_ξ^* , we see that these are not estimated significantly different from zero. If seasonality is causing vast swings in prices, this could make estimation of drift rates difficult. If investors anticipate a seasonal effect on futures prices, this could also affect the risk premium they require for natural gas. This might be the reason for the barely significant short run risk premium λ_χ , and the insignificant λ_ξ inherent in $\mu_\xi^* = \mu_\xi - \lambda_\xi$. Other reasons for the insignificance here might be that risk premia are unobserved quantities and often require a large set of data to be estimated precisely (Schwartz & Smith, 2000).

When looking at the standard errors of the measurement equation, s_i , we observe that these are significantly different from zero. If seasonality is present, and the model is not able to capture it, these estimations should be affected.

7.1 Estimated state variables – Model 1

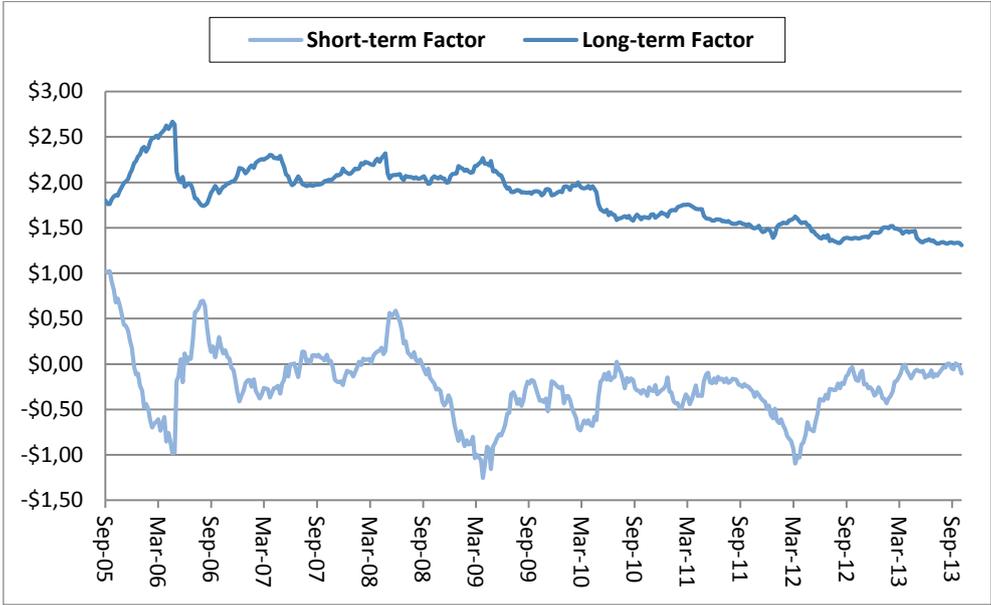


Figure 7.1 Plot of estimated state variables from Model 1

Looking at Figure 7.1, the long-term state variable, ξ_t , seems to be fluctuating in a wave pattern with a slightly downward trend. The fluctuations have an annual wavelength, with higher peaks in winter months. This fits well with our seasonality hypothesis. The short-term variable, χ_t , is more volatile than the long-term variable, with several large spikes. It also seems to revert to a mean around zero (as one would expect). Although it is hard to say anything certain about seasonality only from looking at this plot, it appears seasonality is affecting long-run prices more than short-run prices, which are more affected by temporary shocks. This seasonality is captured by the long-term factor ξ_t , which might cause the difficulty in estimating parameters.

7.2 Residual analysis – Model 1

If our model is able to capture all of the non-stationary properties of natural gas futures prices, its residuals should be a stationary white noise process. In other words, the residuals should be independent with a zero mean and a constant variance (Enders, 2009). If the residuals are not a white noise, it will be an indication that other factors, for instance seasonality, are present in the prices. We therefore start by looking at the mean and standard deviations of our different futures contracts' residuals:

	NG1	NG4	NG7	NG10	NG13	NG16	NG19	NG22
Mean	-1.67%	0	0.19%	-0.03%	-0.32%	-0.15%	-0.16%	0.02%
	(0.61%)	(0)	(0.40%)	(0.30%)	(0.32%)	(0.39%)	(0.40%)	(0.02%)
StDev	12.53%	0	8.20%	6.12%	6.54%	8.07%	8.23%	0.50%

Standard errors of estimate in parenthesis

Table 7.2 Summary statistics of residuals from Model 1

From the summary statistics, we observe that all contracts, except for the 1M contract, have a mean significantly centered on zero. The standard deviations are interesting when we compare them to the ones from Model 2 later.

To get a sense of how the residuals vary across the year, we estimated mean residual values for each contract conditional on delivery month. The results are shown in Table 7.3. It seems that residuals for contracts delivered in the winter months of December, January, February and March are generally positive. This means that our model is underestimating prices in these months. In the warmer months of July to October, our model seems to overestimate prices.

	1M	7M	13M	19M
January	7,4 %	13,6 %	9,5 %	13,2 %
February	5,8 %	7,7 %	6,4 %	12,1 %
March	4,2 %	-0,6 %	1,0 %	6,8 %
April	4,2 %	-7,5 %	-7,9 %	-1,7 %
May	6,0 %	-7,3 %	-7,6 %	-3,1 %
June	4,6 %	-4,5 %	-5,2 %	-2,4 %
July	-0,4 %	-1,5 %	0,4 %	-2,0 %
August	-6,5 %	-2,4 %	-0,6 %	-4,7 %
September	-17,1 %	-3,9 %	-3,8 %	-8,7 %
October	-19,6 %	-3,8 %	-3,3 %	-9,4 %
November	-8,6 %	3,0 %	0,7 %	-4,7 %
December	1,7 %	10,3 %	6,2 %	2,4 %

The 16M, 19M and 22M show similar traits, whereas the 4M is set to zero measurement error

Table 7.3 Mean residual value by delivery month, Model 1

In contracts delivered in April to June, the futures contracts seems to be overestimated in long-term futures prices, whereas the 1M contract seems to be underestimated. A reason for this might be that weather conditions are somewhat predictable in the short term, whereas in longer time horizons, investors cannot get a sense of how the weather is going to turn out. Long-term investors would therefore use less volatile long-term trends in temperatures in their estimates, whereas short-term investors would incorporate volatile short-term weather predictions, if they believe seasonality is present. Seasonality traits would, in this case, be more evident in long-term than short-term contracts. Another explanation might be that the short-term shocks

captured by χ_t are distorting the seasonal effect on short-run prices. Since short-term shocks are less present in the long term, seasonality might be more apparent in longer contracts.

Overall, an underestimation in winter months and overestimation in summer months seem to infer that a seasonal effect is present, and that our model is not able to capture this.

To formally test for time-independent residuals, a requirement for a white noise process, we perform a Ljung-Box-Q test. With a critical value of 31.41, we can reject the null hypothesis in favor of auto-correlated residuals, for all relevant futures contracts²¹:

1M	4M	7M	10M	13M	16M	19M	22M
1762	25	1814	2417	1759	2689	2080	370

Critical value of Q is 31.41

Table 7.4 Ljung-Box Q-test statistics for residual autocorrelation

This autocorrelation is also evident in the sample ACF of the residuals. We include the plot since it shows tendencies of seasonality in the residuals - the residuals seem to be positively correlated with residuals from some parts of the year, and negatively correlated with residuals from other parts.

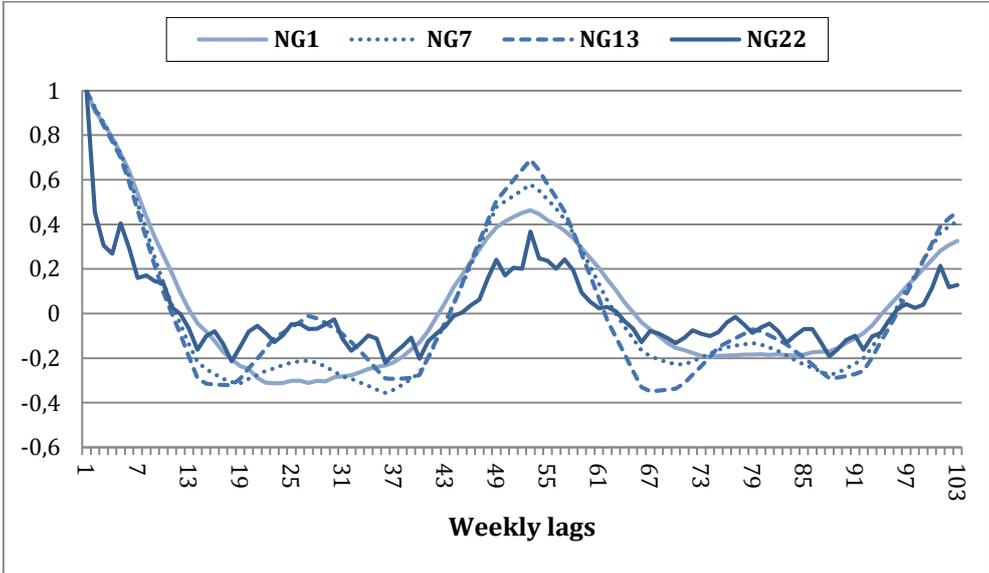


Figure 7.2 ACF-plot Model 1, sample autocorrelation for lags up to two years

The preceding results leads us to conclude that the residuals of Model 1 are not from a white noise process, and that other factors, like seasonality, are likely to be present in the true data generating process.

²¹ Since the 4M contract is set to have a zero measurement error, its residuals will always be a white noise.

8. Extending the model further – allowing for seasonality

We now extend our model with a deterministic seasonality factor to see how parameter estimates, state variables and residuals are affected. This will give us an indication as to if seasonality is indeed present in our data.

8.1 Extending with deterministic seasonality

Sørensen (2002) extends the two-factor model of Schwartz & Smith (2000) with the following trigonometric seasonality function to better capture the seasonality in corn-, soybean- and wheat futures:

$$g(t) = \sum_{i=1}^I (\gamma_i \cos(2\pi it) + \gamma_i^* \sin(2\pi it))$$

We believe this model extension should be able to capture seasonal features in our data if they are present. This is because trigonometric functions have a wavelike pattern similar to what we observe in seasonal patterns.

This type of seasonality modeling was initially proposed by Hannan, Terrell and Tuckwell (1970) as an alternative to seasonality functions using standard dummy variables. The parameter I is chosen freely, but for simplicity, we choose to follow Sørensen (2002) and use $I=2$.²²

After adding the seasonal component, the log spot price is given by:

$$\ln(S_t) = g(t) + \chi_t + \xi_t$$

²² In Sørensen (2002) they tried different values of K , and obtained the optimal Akaike Information Criterion (AIC) with $K=2$. In our thesis we only wish to show that seasonality is present in the time series, and to save time, we have used the same K value without a similar analysis when extending our model. Other values of K might prove to be a better description of the seasonality traits of the series.

We can already observe that if the seasonality component is significant, it could lead to other parameter values of our state variables.²³

8.1.1 Dynamics and futures price

The added seasonality component does not affect the dynamics of the state variables. The basic idea when using this model is that “*g(t) captures price movements that are entirely related to season*” (Sørensen, 2002, p. 6) while the two other state variables have the same interpretation as before. This makes us able to model seasonality in prices separately from the long-term and short-term factors. This will make us able to see if the extension is significant.

Sørensen (2002) states the analytical solution to the futures price at time t , with maturity on date τ as:

$$F_{t,\tau} = e^{g(\tau)+A(\tau-t)+e^{-\kappa(\tau-t)}\chi_t+\xi_t}.$$

Here, $A(\tau - t)$ is defined exactly as in the Schwartz & Smith (2000) model, only substituting T with $(\tau - t)$, t is here given as the value of the trading date of the futures contract²⁴. Each successive date is then equal to $t + \Delta$, where Δ equals the time increment. τ relates to the maturity date of the contract²⁵. The $g(t)$ function is constructed such that every $g(t)$ value from $t = 0$ to $t = 1$ is repeated for every $\Delta = 1$ ²⁶. The constant A depends the distance between today's date and delivery, $(\tau - t)$. All of this means that we have to both take into account the time of year trading is conducted, which time of year the contract matures as well as time to maturity, when estimating this extended model.

²³ A ln spot price of 5\$ might for instance be explained by $\chi_t + \xi_t = 5$ in Model 1. If $g(t) = 1$ for the same date in Model 2, this would imply a value of $\chi_t + \xi_t = 4$, in order to yield the same ln spot price.

²⁴ The first observation in the sample has a value according to which day in the year the data series starts. If the first date is the 3rd of January, $t=3/365$.

²⁵ If the contract matures on January 28th the same year, τ would equal $28/365$, whereas if it matures on January the 28th the next year, it would equal $(28+365)/365$ etc.

²⁶ This means that each January value will be the same for subsequent January months forward in time.

8.1.2 Extra parameters and new measurement equation

The only other thing that is changed in our estimation procedure is the inclusion of four new parameters $(\gamma_1, \gamma_1^*, \gamma_2, \gamma_2^*)$, as well as a modified measurement equation taking the new seasonal factor into account:

$$y_t = g(\tau) + d_t + F_t' x_t + v_t,$$

where

$$g(\tau) = \gamma_1 \cos(2\pi\tau) + \gamma_1^* \sin(2\pi\tau) + \gamma_2^* \cos(4\pi\tau) + \gamma_2^* \sin(4\pi\tau),$$

and

d_t, F_t, x_t and v_t are defined as in Model 1.

In the next section, we will show the results from this estimation, and how the new seasonality factor has affected parameters, state variables and residuals.

9. Results from Model 2

After running the Kalman filter optimizing procedure 1400 times using random initial values of the parameter set θ, χ_0, ξ_0 and C_0 we got the following results:

Parameters	Model 1	Model 2
κ	0.6308 (0.0113)	1.4093 (0.0202)
σ_χ	0.6664 (0.0113)	0.5440 (0.0257)
λ_χ	-0.2288 (0.1024)	-0.4103 (0.0702)
μ_ξ	-0.1530 (0.0906)	-0.1081 (0.0521)
σ_ξ	0.3046 (0.0120)	0.1461 (0.0084)
μ_ξ^*	-0.1436 (0.1546)	-0.0266 (0.0047)
$\rho_{\chi\xi}$	-0.7120 (0.0234)	0.1123 (0.0347)
s_{1M}	0.1262 (0.0011)	0 (0)
s_{4M}	0 (0)	0.0640 (0.0003)
s_{7M}	0.0822 (0.0005)	0.0568 (0.0002)
s_{10M}	0.0613 (0.0002)	0.0388 (0.0001)
s_{13M}	0.0653 (0.0003)	0.0320 (0.0001)
γ_1		0.0606 (0.0010)
γ_1^*		-0.0025 (0.0010)
γ_2		0.0272 (0.0010)
γ_2^*		0.0030 (0.0010)
Log Likelihood	4309.2	5511.3

Standard errors in parenthesis

Model 1 results are included for comparison

Table 9.1 Parameter results from Model 2

We observe that all of our parameters are significantly different from zero on a 95 % confidence level. This implies that the new seasonality factor have improved the model's ability to estimate parameter values.

The degree of mean reversion, κ , is higher when including the seasonality term in our model. This fits well with the hypothesis of possible underestimation of κ caused by seasonality, as described in the results of Model 1.

Looking at the volatility parameters, σ_χ and σ_ξ , we observe that both estimates are lower compared to the estimates in Model 1. This fits well with our a priori assumption that seasonality might cause large swings in prices over the course of the year. When correcting for seasonality through $g(t)$, the underlying volatility appears to have become much lower. The Samuelson hypothesis still holds with a higher volatility in the short-term factor, σ_χ , than in the long-term factor's volatility, σ_ξ . This indicates that our model is still capturing important aspects of our futures prices' dynamics.

In Model 1, neither μ_ξ nor μ_ξ^* were significantly different from zero. We now observe that the extended model is able to estimate both parameters significantly on a 95% level. Correcting for seasonality may thus have made it easier to infer the underlying growth rate of our state variables.

The two risk premium coefficients, λ_ξ and λ_χ , are also captured with better significance in this version of the model. This is another indication of improvement from the new model. Since we are able to extract risk-premiums and risk-neutral growth rate in the new model it seems that investors in natural gas futures consider seasonality, when determining the value of futures contracts. If for instance a long-term price in natural gas is 4\$, and an investor knows that this long-term price will tend to be 3\$ in the summer and 6\$ in the winter, its futures price will incorporate this, even though the risk premium for the investor can be constant over the course of the year. If we do not consider this in our model, this might lead to difficulty in estimating the risk premium. This may be what we observed in Model 1, which might explain why, when accounting for seasonality in Model 2, our estimates improved.

Looking at the standard deviations of the errors in the measurement equation, s_i , we observe that they are smaller compared to those in Model 2, except for the 22-month contract. This might indicate that Model 2 is more able to fit the cross-section of futures prices than Model 1.

Another indication of Model 2's better ability to fit the data is the log likelihood value of 5511.3. Compared to Model 1, the addition of the seasonality factor yields a log likelihood that is 1292.1

higher²⁷. The inclusion of seasonality has therefore significantly increased the likelihood of obtaining the true parameters of our model.

Turning to the deterministic seasonal factor, we see that all its parameters are estimated significantly different from zero. Looking at the plotted seasonality function shown in Figure 9.1, we observe that prices are higher from the middle of August to the start of April when they turn lower. The significant seasonal factor thus seems to exhibit the kind of price variations we would expect a priori.

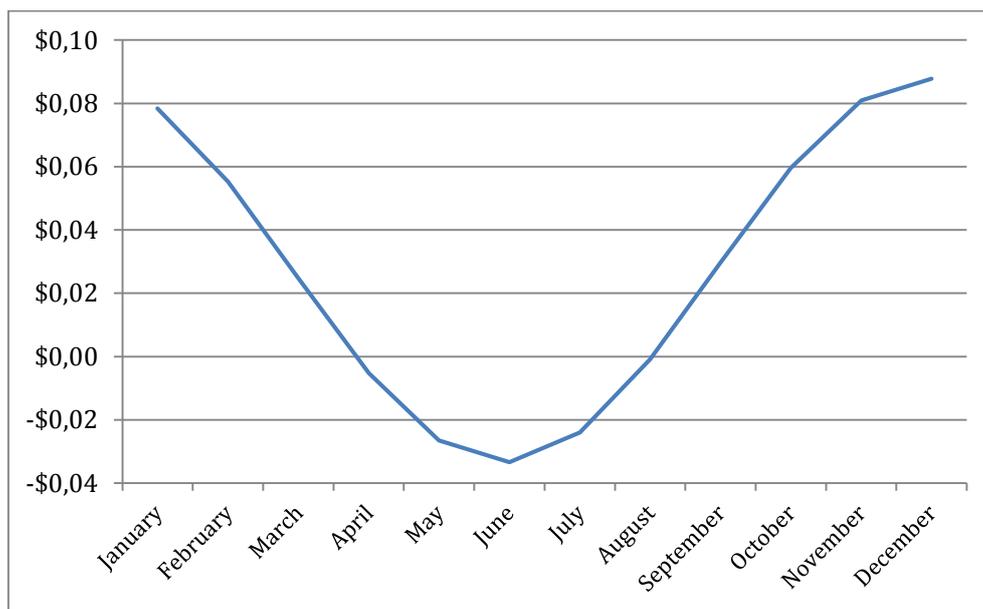


Figure 9.1 Plot of the estimated seasonality factor

²⁷ According to Schwartz & Smith (2000) the relevant test statistic for this comparison is the chi-squared distribution with 3 degrees of freedom with a 99th percentile equal to 11.34.

9.1 Estimated state variables – Model 2

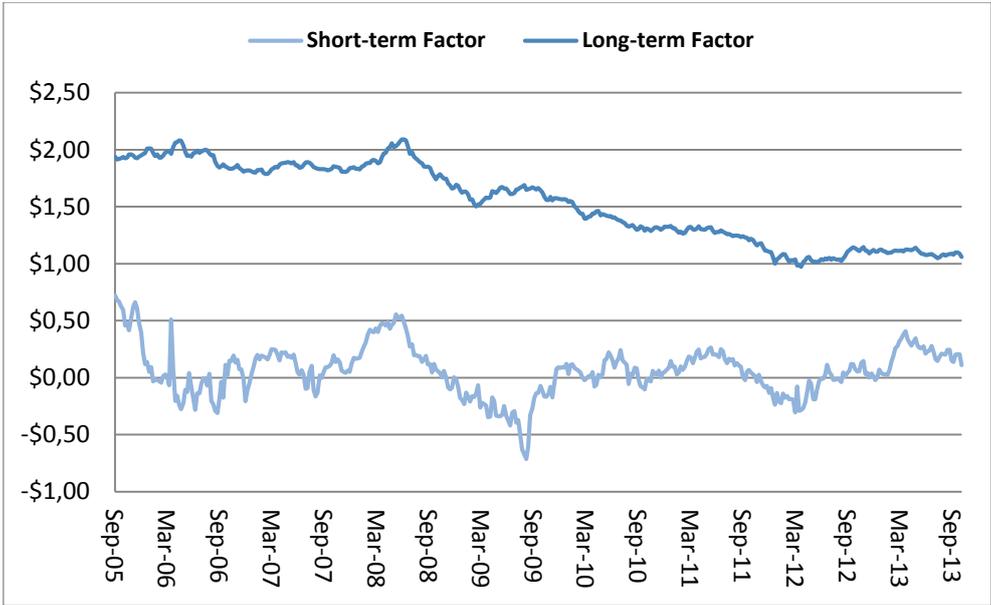


Figure 9.2 Plot of estimated state variables from Model 2

Looking at the long-term state variable, ξ_t , in Figure 9.2, the downward trend is still evident, but the seasonality traits have disappeared. It appears that the seasonal factor has captured this effect, leading to less variance in ξ_t , as observed in the lower σ_ξ . The short-term factor is still reverting to a mean of zero, with several short-term spikes. Compared to Model 1, the magnitude of these spikes seems to have been reduced, and they seem to be lasting for shorter time periods. This is reflected in our parameters through a lower σ_χ and a higher κ . The new seasonality factor thus seems to have removed the seasonality traits in the long-term factor, as well as reduced volatility in both.

9.2 Residual analysis Model 2

Our previous results indicate that the deterministic seasonality function has improved our model’s description of the natural gas market through better parameter estimates and log likelihood value. We now look at the residuals of the model, to assess how they are affected by the new factor. As we previously mentioned, if the model is capturing the most important features of the market, its residuals should be a strict white noise.

From the summary statistics shown in Table 9.2, we observe that all residuals have a mean significantly centered on zero. Their volatility is significantly smaller in this model compared

to the residuals in Model 1, which we see as a sign of the model's improvement from the seasonality factor²⁸.

Contract	1M	4M	7M	10M	13M	16M	19M	22M
Mean	0 (0)	0.4% (0.31%)	0.31% (0.27%)	0.14% (0.19%)	0.09% (0.15%)	0.06% (0.12%)	0.11% (0.16%)	0.23% (0.23%)
Std. dev.	0	6.34%	5.61%	3.81%	3.03%	2.52%	3.24%	4.79%
Std. dev. Model 1	12.53%	0	8.20%	6.12%	6.54%	8.07%	8.23%	0.50%

Standard errors of estimate in parenthesis
Model 1 std.dev. results are included for comparison

Table 9.2 Summary statistics of residuals from Model 2

Looking at the mean residual values by delivery months, we observe that the seasonality patterns are much less evident in this model. The values seem to be randomly distributed around zero, and vary randomly across months from contract to contract. The residuals are, in addition, a lot smaller than in Model 1 indicating that the new seasonality factor has improved overall results. Our model no longer seems to be underestimating winter prices and overestimating summer prices as was evident as in Model 1. Since this variation is significantly captured by our new factor, we believe this is a strong indication of a presence of seasonality in our data set.

	4M	7M	13M	19M
1	2,5 %	0,7 %	-0,3 %	0,3 %
2	-0,4 %	-0,8 %	-0,3 %	0,5 %
3	-0,1 %	1,3 %	1,0 %	0,6 %
4	2,3 %	1,7 %	-1,4 %	-1,9 %
5	4,6 %	1,6 %	0,2 %	-0,5 %
6	2,6 %	0,7 %	-0,5 %	-0,6 %
7	-0,8 %	1,1 %	-0,1 %	0,1 %
8	-3,4 %	0,5 %	-0,2 %	-0,1 %
9	-2,9 %	-0,6 %	-0,3 %	-0,1 %
10	-1,0 %	-2,3 %	0,7 %	1,3 %
11	-0,7 %	-1,5 %	0,2 %	1,3 %
12	2,2 %	-0,1 %	-0,8 %	-0,1 %

The 16M, 19M and 22M show similar traits, whereas the 1M is set to zero measurement error

Table 9.3 Mean residual value by delivery month, Model 2

Table 9.4 show the results from the Ljung-Box Q-test for autocorrelation. They indicate that there still is autocorrelation in all relevant contracts' residuals²⁹. Therefore, the extension of the

²⁸ The volatilities in the 1M and 4M contracts are not comparable, since these are set to zero measurement-error when estimating model 2 and model 1 respectively.

²⁹ Except for the 1M contract with a pre-set zero measurement error

model did not lead to white noise residuals, meaning other significant factors may still be present in the data. Looking at the relevant values of the test statistics³⁰, we observe that these are lower compared to the results in Model 1. This indicates that the residuals in Model 2 are closer to a white noise than the one's in Model 1.

1M	4M	7M	10M	13M	16M	19M	22M
22.29	1201	1977	1502	980	917	920	2023

Critical value of Q is 31.41

Table 9.4 Ljung-Box Q-test statistics for residual autocorrelation

Autocorrelation is also evident from the ACF plot, as shown in Figure 9.3, but we observe that both the seasonality and variance of the autocorrelation function have been reduced significantly, compared to those in Model 1.

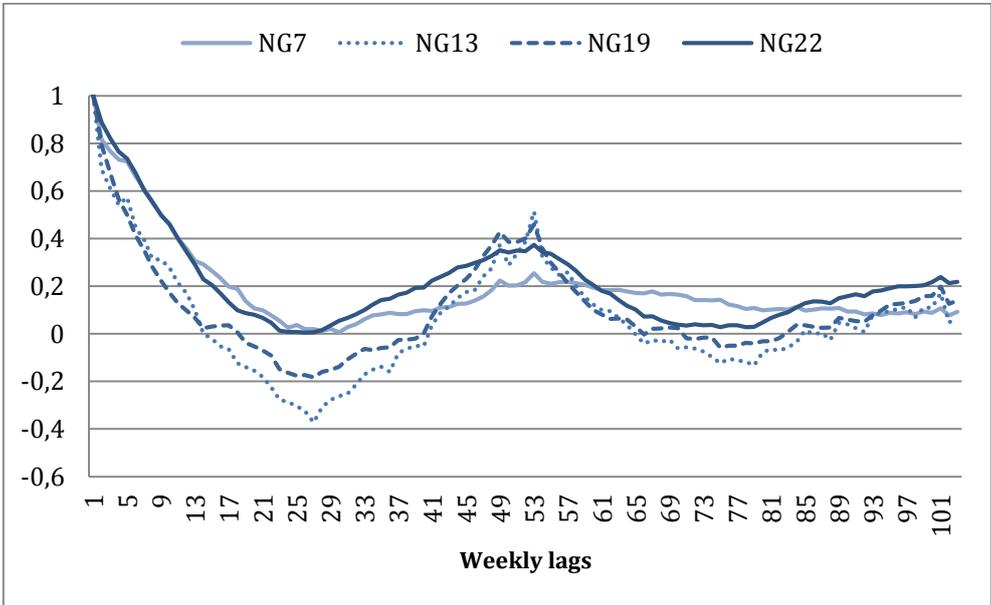


Figure 9.3 ACF-plot Model 2, sample autocorrelation for lags up to two years

There are still some seasonality traits left in the ACF, but these are of smaller magnitude. This might suggest that our seasonality factor is not able to fully capture the seasonality in natural gas prices. Explanations for this might be that we have chosen a deterministic seasonal factor. If there is a trend or time variation in seasonality, our deterministic seasonality function could be too small or too high in some of the years. This will reduce the seasonality in residuals, but there will still be some seasonality left. Another explanation might be that our data set is too small to be able to capture the seasonal factor properly. Sørensen (2002) used a data sample of

³⁰ Remember, in this model, we forced the 1M contract to have zero measurement errors, whereas in model 1, this was done in the 4M contract.

25 years of weekly data when estimating the seasonal factors for other commodities. This is a larger data set compared to ours, with only eight years of weekly data. A third hypothesis explaining this could be that seasonality is present only in long-term contracts affecting the estimation of the seasonality function.

Over all, the residuals do not seem to be generated from a white noise process. This leads to the possibility that other factors are present in the prices. At the same time, the significant seasonality function has reduced the mean, variance and autocorrelation in residuals indicating an improvement from the new seasonality function.

10. Concluding remarks

In this thesis we have estimated the Schwartz & Smith (2000) model both with and without a deterministic seasonal factor to see whether seasonality is a significant factor in natural gas prices.

In our initial time series analysis of our futures data, we find traits of seasonality in futures contracts with maturities longer than 4 months, whereas these traits are less evident in the 1-month contracts. The seasonal traits are characterized by prices being higher for contracts delivered in winter months and lower for contracts delivered in summer months. The time series also indicates that spot prices are non-stationary, either following a random walk, a random walk with drift or a trend-stationary process. At the same time, the residuals from these processes indicate that other factors are likely to be present in the series, such as seasonality.

The estimated parameters in Model 1 show signs of an omitted parameter bias. The presence of seasonality seems to have led to an underestimation of the mean reversion parameter, κ , and overestimation of the volatility parameters σ_χ and σ_ξ in Model 1. After including the seasonal factor in Model 2 this effect is removed. Another argument for an omitted parameter bias in Model 1 is the insignificant drift parameters, the barely significant short-term risk premium and insignificant long-term risk premium. When introducing the seasonality factor in Model 2, all parameters are estimated significantly, indicating that the included seasonality factor has improved the model. The significant seasonality factor clearly shows a tendency of prices being higher in winter months and lower in the summer. Comparing standard measurement errors and the log-likelihood of both models, this shows that the added seasonality factor has improved the model's ability to capture the dynamics in the data, indicating that seasonality is present and significant in natural gas prices.

The estimated state variables in Model 1 show seasonal tendencies in the long-term factor, whereas this is less evident in the short-term factor. After including the seasonality factor in Model 2, these seasonality traits are more or less eliminated in the state variables, showing that the new factor has captured this effect.

The residuals in Model 1 show signs of seasonality when estimating mean values conditional on delivery month. In Model 2 this is no longer evident. The residuals in Model 2 have lower variance as well, indicating a better fit to the data. We lastly find that the residuals in both

models are auto correlated, indicating that neither model fully explains variation in natural gas prices. Even though other factors may still be present in natural gas prices, seasonality seems to have made a significant improvement to our model.

From our results, we conclude that seasonality seems to be a significant factor in natural gas prices.

11. Assumptions, limitations and improvements

Although our results show clear tendencies of seasonality in natural gas prices, we wish to highlight some limitations to our approach, as well as suggest topics further study.

While natural gas futures have been traded on NYMEX as far back as 1990, we only used data from September 2005 to November 2013. Although gathering data for a longer time period might let us say more about the development in the state variables and the seasonal factor, small volumes throughout the 1990s and large gaps in prices went into our choice of time period.

For the few missing dates in our dataset, we chose to use linear interpolation in order to fill in the short gaps. This could have led to us missing some short-term shocks, as for instance hurricane Katrina led to a shutdown in trading of Natural Gas futures on NYMEX (U.S. Department of Energy, 2009a). However, we have tried to correct for this by looking for sources of the missing prices, and for our data set, the missing prices seem to come mainly from holiday trading stops, rather than factors that would cause the prices to fluctuate.

In our model we use a deterministic seasonal factor to describe monthly price variations over the year. More advanced functions like a stochastic seasonal factor, might be better at explaining seasonal variation, but will require extra parameters to be estimated and probably a longer period of data. Our estimated seasonality function implicitly assumes that the seasonality effect on spot prices is equal across all maturities. This might be a simplifying assumption, and further studies should formally investigate if this is indeed the case.

Observable variables like weather reports, interest rates and inventory levels might provide a better picture of why winter prices are higher than summer prices in our data set. Including these in our model could make us able to pin down the exact reasons for seasonality in prices, as well as enable better forecasting of the magnitude of seasonality in prices. In our thesis we only show that seasonality traits exist, but do not quantitatively tie them down to explanatory factors. This would be an interesting topic for further study.

In our model, we use constant risk premiums with no assumptions of their dynamics. A constant risk premium makes an implicit assumption that its correlation between other assets in the

economy is constant. This might be a too simplifying assumption. Different papers³¹ indicate that time-dependent risk premiums may affect the degree of mean reversion, because of negative correlation between the spot price and risk premium. Looking at interest rates, one might also be tempted to think about a possible level-dependent correlation between risk premia. If interest rates are low, people in the economy will demand more goods, driving the overall prices up. This might reduce risk premiums if demand for commodities increases because of this. The opposite applies if interest rates are high. For further study one could incorporate interest rate dependent, or spot-price dependent risk premia. One would then have to specify the dynamics of the two risk-premia parameters as well as their relationship to other dependent variables. This would lead to a more complicated model. In our thesis, however, we only wish to show that seasonality is present, so a simplified risk premium might not affect the results too much.

More advanced features could also be added to the stochastic processes of our state variables³². For instance jump features or volatility clustering might be present, distorting parameter values in our model, making it harder to isolate the effects of seasonality. At the same time, our model is good enough to indicate seasonality, which is the main purpose of our thesis.

When running the Kalman filter optimizing procedure in Matlab, the objective was to find the initial values that would produce the highest maximum likelihood. In order to find the global maximum however, the procedure would have to run through all possible initial values. We only ran our two models' optimizing procedures for 2500 and 1400 initial values respectively, but we wish to highlight that this procedure optimized for local maxima around these starting values. Although running the procedure for a longer time might have marginally increased the maximum likelihood function, we possessed neither the computer power nor the time to do this.

Since we were using Matlab code for model estimation and time series analysis, we wish to highlight the possibility of errors in code arising from mistypings' etc. The replicated results in section 6 do not indicate that such errors exist, but we cannot offer any guarantees.

³¹ *"However, allowing for time-varying risk premia is important since, as argued by Fama & French (1987), negative correlation between risk premia and spot prices may generate mean reversion in spot prices"* (Casassus & Collin-Dufresne, 2005, p. 2285).

³² See for instance "Maximal Convenience Yield Model" by Cassasus & Dufresne (2005), a three-factor model using both observed and unobserved variables in a very complicated way to model commodity prices.

Bibliography

Augustine, C. W., Broxson, B. & Peterson, S., 2006. *Understanding Natural Gas Markets*. [Online]

Available at: http://www.api.org/~media/Files/Oil-and-Natural-Gas/Natural_Gas/UNDERSTANDING_NATURAL_GAS_MARKETS.pdf

[Accessed 25 November 2013].

Back, J., Prokopczuk, M. & Rudolf Markus, 2013. Seasonality and the valuation of commodity options. *Journal of Banking & Finance*, Volume 37, pp. 273-290.

Bernstein, R. & Madlener, R., 2011. *E.ON Energy Research Center*. [Online]

Available at: http://www.eonerc.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaaacivgc

[Accessed 23 November 2013].

Casassus, J. & Collin-Dufresne, P., 2005. Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates. *The Journal of Finance*, 60(5), pp. 2283-2331.

CME Group, 2013. *Henry Hub Natural Gas Futures - Dynamic Trading Opportunities*. [Online]

Available at: http://www.cmegroup.com/education/files/PM310_Natural_Gas_Futures.pdf

[Accessed 25 November 2013].

Enders, W., 2009. *Applied Econometric Time Series*. 3rd ed. New Jersey: Wiley Publishing.

Fama, E. F. & French, K. R., 1987. Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *The Journal of Business*, 60(1), pp. 55-73.

Fusai, G., 2012. *Risk Management of a Composite Commodity Portfolio Using Monte Carlo Simulations and MATLAB*. London, MATLAB Computational Finance Conference.

Gibson, R. & Schwartz, E., 1990. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, pp. 959-976.

Hannan, E. J., Terrell, R. D. & Tuckwell, N. E., 1970. The Seasonal Adjustment of Economic Time Series. *International Economic Review*, 11(1), pp. 24-52.

Hull, J. C., 2013. *Options, Futures and Other Derivatives*. Harlow: Pearson Education Limited.

Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1), pp. 35-45.

Kunst, R., 2007. *State Space Models and the Kalman Filter*. [Online]
Available at: <http://homepage.univie.ac.at/robert.kunst/statespace.pdf>
[Accessed 15 November 2013].

Lucia, J. J. & Schwartz, E., 2002. Electricity prices and power derivatives: evidence from the nordic power exchange. *Review of Derivatives Research*, 5(1), pp. 5-50.

Quandl, 2013. *Quandl - Natural Gas Futures*. [Online]
Available at: <http://www.quandl.com/futures/natural-gas-futures>
[Accessed 25 November 2013].

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

Schwartz, E., 1997. The stochastic behavior of commodity prices: implications for valuation and hedging. *Journal of Finance*, pp. 923-973.

Schwartz, E. & Smith, J. E., 2000. Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, 46(7), pp. 893-911.

Sørensen, C., 2002. Modeling Seasonality in Agricultural Commodity Futures. *The Journal of Futures Markets*, 22(5), pp. 393-426.

U.S. Department of Energy, 2009a. *Comparing the Impacts of the 2005 and 2008 Hurricanes on U.S. Energy Infrastructure*. [Online]
Available at: <http://www.oe.netl.doe.gov/docs/HurricaneComp0508r2.pdf>
[Accessed 11 November 2013].

U.S. Energy Information Administration, 2013a. *Natural Gas Monthly*. [Online]
Available at: http://www.eia.gov/naturalgas/monthly/pdf/ngm_all.pdf
[Accessed 25 November 2013].

12. Appendices

12.1 Appendix A

The following dates were missing for all eight futures: 25.11.05, 06.04.07, 21.03.08, 04.07.08, 10.04.09, 03.07.09, 23.10.09, 25.12.09, 01.01.10, 02.04.10, 24.12.10, 22.04.11, 29.03.13

For the remaining missing dates, they were only missing for some or one of the futures contracts

NG4: 14.04.06

NG7: 14.04.06, 24.11.06,

NG10: 14.04.06, 25.08.06, 24.11.06,

NG13: 30.12.05, 06.01.06, 14.04.06, 25.08.06, 13.10.06, 24.11.06,

NG16: 14.10.05, 28.10.05, 18.11.05, 02.12.05, 09.12.05, 16.12.05, 30.12.05, 14.04.06, 25.08.06, 01.09.06, 13.10.06, 03.11.06, 24.11.06, 01.12.06, 05.01.07, 09.02.07, 09.03.07,

NG19: 30.09.05, 28.10.05, 04.11.05, 11.11.05, 18.11.05, 02.12.05, 09.12.05, 16.12.05, 30.12.05, 14.04.06, 28.04.06, 18.08.06, 25.08.06, 01.09.06, 29.09.06, 13.10.06, 20.10.06, 27.10.06, 03.11.06, 10.11.06, 17.11.06, 24.11.06, 08.12.06, 22.12.06, 05.01.07, 09.02.07, 23.02.07, 02.03.07, 09.03.07, 16.03.07, 23.03.07, 20.04.07, 27.04.07, 26.08.11, 02.09.11, 09.09.11, 16.09.11, 23.09.11, 30.09.11

NG22: 30.09.05, 21.10.05, 28.10.05, 11.11.05, 18.11.05, 02.12.05, 09.12.05, 16.12.05, 23.12.05, 30.12.05, 14.04.06, 18.08.06, 25.08.06, 01.09.06, 08.09.06, 22.09.06, 29.09.06, 13.10.06, 20.10.06, 03.11.06, 10.11.06, 17.11.06, 24.11.06, 01.12.06, 08.12.06, 15.12.06, 22.12.06, 05.01.07, 23.02.07, 09.03.07, 23.03.07, 30.07.07, 20.04.07, 27.04.07, 04.05.07, 18.05.07, 01.06.7, 08.06.07, 15.06.07, 26.08.11, 02.09.11, 09.09.11, 16.09.11, 23.09.11, 30.09.11

12.2 Appendix B

To derive the analytical futures price of our model, we have to use the risk-neutral processes of the state variables. Given initial values χ_0 and ξ_0 , S&S (2000) show that the risk-neutral processes of χ_t and ξ_t are jointly normally distributed with mean vector and covariance matrix:

$$E^Q[(\chi_t, \xi_t)] = [e^{-\kappa t} \chi_0 - \frac{(1 - e^{-\kappa t}) \lambda_\chi}{\kappa}, \xi_0 + \mu_\xi^* t]$$

$$Cov^Q[(\chi_t, \xi_t)] = Cov[(\chi_t, \xi_t)]$$

They continue to show that under Q the logarithm of the spot price is normally distributed with:

$$E^Q[\ln(S_t)] = e^{-\kappa t} \chi_0 + \xi_0 - \frac{(1 - e^{-\kappa t}) \lambda_\chi}{\kappa} + \mu_\xi^* t,$$

$$Var^Q[\ln(S_t)] = Var[\ln(S_t)].$$

Using the fact that the futures price is equal to the expected future spot price under Q, we can solve for the analytical futures price:

$$\begin{aligned} \ln(F_{t,T}) &= \ln(E^Q[S_T]) = E^Q[\ln(S_T)] + \frac{1}{2} Var^Q[\ln(S_T)] \\ &= (e^{-\kappa(T-t)} \chi_0 + \xi_0 - \frac{(1 - e^{-\kappa(T-t)}) \lambda_\chi}{\kappa} + \mu_\xi^*(T-t)) \\ &\quad + \frac{1}{2} \left(\frac{(1 - e^{-2\kappa(T-t)}) \sigma_\chi^2}{2\kappa} + \sigma_\xi^2(T-t) + \frac{2(1 - e^{-\kappa(T-t)}) \rho_{\xi\chi} \sigma_\chi \sigma_\xi}{\kappa} \right) \\ &= e^{-\kappa T} \chi_0 + \xi_0 + A(T-t) \end{aligned}$$

where

$$A(T-t) = \mu_\xi^*(T-t) - \frac{(1 - e^{-\kappa(T-t)}) \lambda_\chi}{\kappa} + \frac{1}{2} \left(\frac{(1 - e^{-2\kappa(T-t)}) \sigma_\chi^2}{2\kappa} + \sigma_\xi^2(T-t) + \frac{2(1 - e^{-\kappa(T-t)}) \rho_{\xi\chi} \sigma_\chi \sigma_\xi}{\kappa} \right)$$