NTNU – Trondheim
Norwegian University of
Science and Technology

Price-Volatility Modeling in the US Natural Gas Market

Torstein Rikter-Svendsen
Cecilie Nilsen Kielland
Bjørn Heineman

Industrial Economics and Technology Management
Submission date: June 2012
Supervisor: Sjur Westgaard, IØT
Co-supervisor: Asgeir Tomasgård, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Preface

This master thesis was written during the spring of 2012 at the Norwegian University of Science and Technology (NTNU), Department of Industrial Economics and Technology Management within the field of Financial Engineering, specifically Empirical Finance.

The challenging work with the thesis has enabled us to investigate an interesting market by applying our knowledge in financial theory, mathematics and statistic, thereby considerably increasing our understanding of the natural gas market, and volatility modeling and forecasting.

We would like to thank our academic supervisors, Sjur Westgaard and Asgeir Tomasgard, for feedback and insightful contributions during the process. We also wish to express our gratitude towards Aiste Luksyte at Timberlake Consultants for valuable insight and guidance with programming issues.

For the empirical tests and modeling the 6th version of OxMetrics™ was utilized.

The authors alone are responsible for all content and any errors.

Trondheim, June 8th 2012

NTNU
Det skapende universitet

MASTERKONTRAKT

- uttak av masteroppgave

1. Studentens personalia

Etternavn, fornavn **Heineman, Bjørn**	Fødselsdato **07. mai 1986**
E-post **heineman@stud.ntnu.no**	Telefon **48194078**

2. Studieopplysninger

Fakultet **Fakultet for Samfunnsvitenskap og teknologiledelse**	
Institutt **Institutt for industriell økonomi og teknologiledelse**	
Studieprogram **Industriell økonomi og teknologiledelse**	Hovedprofil **Investering, finans og økonomistyring**

3. Masteroppgave

Oppstartsdato **16. jan 2012**	Innleveringsfrist **11. jun 2012**
Oppgavens (foreløpige) tittel **Price-Volatility Modeling in the US Natural Gas Market**	
Oppgavetekst/Problembeskrivelse Purpose The aim of this master thesis is to develop and analyze volatility models in the US natural gas market, and to forecast volatility. Main tasks 1. Establish factors affecting the price-volatility in the US natural gas market 2. Data collection 3. Create and implement price-volatility models, and to forecast volatility 4. Overall assessment of the insights gained through the model analysis	
Hovedveileder ved institutt **Førsteamanuensis Sjur Westgaard**	Medveileder(e) ved institutt **Asgeir Tomasgård**
Merknader **1 uke ekstra p.g.a påske.**	

4. Underskrift

Student: Jeg erklærer herved at jeg har satt meg inn i gjeldende bestemmelser for mastergradsstudiet og at jeg oppfyller kravene for adgang til å påbegynne oppgaven, herunder eventuelle praksiskrav.

Partene er gjort kjent med avtalens vilkår, samt kapitlene i studiehåndboken om generelle regler og aktuell studieplan for masterstudiet.

Trondheim 14/01-12
Sted og dato

Student

Hovedveileder

Originalen lagres i NTNUs elektroniske arkiv. Kopi av avtalen sendes til instituttet og studenten.

NTNU
Det skapende universitet

MASTERKONTRAKT
- uttak av masteroppgave

1. Studentens personalia

Etternavn, fornavn	Fødselsdato
Kielland, Cecilie Nilsen	**06. jun 1987**
E-post	Telefon
cecilien@stud.ntnu.no	**98827239**

2. Studieopplysninger

Fakultet	
Fakultet for Samfunnsvitenskap og teknologiledelse	
Institutt	
Institutt for industriell økonomi og teknologiledelse	
Studieprogram	Hovedprofil
Industriell økonomi og teknologiledelse	**Investering, finans og økonomistyring**

3. Masteroppgave

Oppstartsdato	Innleveringsfrist
16. jan 2012	**11. jun 2012**
Oppgavens (foreløpige) tittel	
Price-Volatility Modeling in the US Natural Gas Market	
Oppgavetekst/Problembeskrivelse Purpose The aim of this master thesis is to develop and analyze volatility models in the US natural gas market, and to forecast volatility. Main tasks 1. Establish factors affecting the price-volatility in the US natural gas market 2. Data collection 3. Create and implement price-volatility models, and to forecast volatility 4. Overall assessment of the insights gained through the model analysis	
Hovedveileder ved institutt	Medveileder(e) ved institutt
Førsteamanuensis Sjur Westgaard	**Asgeir Tomasgård**
Merknader	
1 uke ekstra p.g.a påske.	

4. Underskrift

Student: Jeg erklærer herved at jeg har satt meg inn i gjeldende bestemmelser for mastergradsstudiet og at jeg oppfyller kravene for adgang til å påbegynne oppgaven, herunder eventuelle praksiskrav.

Partene er gjort kjent med avtalens vilkår, samt kapitlene i studiehåndboken om generelle regler og aktuell studieplan for masterstudiet.

Trondheim 16.01.2012
Sted og dato

Student

Hovedveileder

Originalen lagres i NTNUs elektroniske arkiv. Kopi av avtalen sendes til instituttet og studenten.

NTNU
Det skapende universitet

MASTERKONTRAKT
- uttak av masteroppgave

1. Studentens personalia

Etternavn, fornavn **Rikter-Svendsen, Torstein**	Fødselsdato **19. apr 1986**
E-post **torsteri@stud.ntnu.no**	Telefon **93421701**

2. Studieopplysninger

Fakultet **Fakultet for Samfunnsvitenskap og teknologiledelse**	
Institutt **Institutt for industriell økonomi og teknologiledelse**	
Studieprogram **Industriell økonomi og teknologiledelse**	Hovedprofil **Investering, finans og økonomistyring**

3. Masteroppgave

Oppstartsdato **16. jan 2012**	Innleveringsfrist **11. jun 2012**
Oppgavens (foreløpige) tittel **Price-Volatility Modeling in the US Natural Gas Market**	
Oppgavetekst/Problembeskrivelse Purpose The aim of this master thesis is to develop and analyze volatility models in the US natural gas market, and to forecast volatility. Main tasks 1. Establish factors affecting the price-volatility in the US natural gas market 2. Data collection 3. Create and implement price-volatility models, and to forecast volatility 4. Overall assessment of the insights gained through the model analysis	
Hovedveileder ved institutt **Førsteamanuensis Sjur Westgaard**	Medveileder(e) ved institutt **Asgeir Tomasgård**
Merknader **1 uke ekstra p.g.a påske.**	

4. Underskrift

Student: Jeg erklærer herved at jeg har satt meg inn i gjeldende bestemmelser for mastergradsstudiet og at jeg oppfyller kravene for adgang til å påbegynne oppgaven, herunder eventuelle praksiskrav.

Partene er gjort kjent med avtalens vilkår, samt kapitlene i studiehåndboken om generelle regler og aktuell studieplan for masterstudiet.

Trondheim 16. jan 2012
Sted og dato

Student

Hovedveileder

Originalen lagres i NTNUs elektroniske arkiv. Kopi av avtalen sendes til instituttet og studenten.

NTNU
Det skapende universitet

SAMARBEIDSKONTRAKT

1. Studenter i samarbeidsgruppen

Etternavn, fornavn	Fødselsdato
Heineman, Bjørn	**07. mai 1986**
Etternavn, fornavn **Kielland, Cecilie Nilsen**	Fødselsdato **06. jun 1987**
Etternavn, fornavn **Rikter-Svendsen, Torstein**	Fødselsdato **19. apr 1986**

2. Hovedveileder

Etternavn, fornavn	Institutt
Westgaard, Sjur	**Institutt for industriell økonomi og teknologiledelse**

3. Masteroppgave

Oppgavens (foreløpige) tittel
Price-Volatility Modeling in the US Natural Gas Market

4. Bedømmelse

Kandidatene skal ha *individuell* bedømmelse ☐
Kandidatene skal ha *felles* bedømmelse ☒

Trondheim 14/01-12
Sted og dato

Bjørn Heineman

Hovedveileder

Cecilie Nilsen Kielland

Torstein Rikter-Svendsen

Originalen oppbevares på instituttet.

Abstract

Understanding price-volatility in the natural gas market is important as it affects new investments and the behavior of market participants. In this paper the volatility of US natural gas prices is investigated using daily Henry Hub futures data for the period 1996 to 2011. The purpose is to determine the best conditional volatility model for forecasting and modeling, and to investigate the fundamental drivers of volatility. Several models are applied and compared on the basis of explanatory power, post-estimation tests, as well as in- and out-of-sample one-day-ahead forecasting capabilities evaluated using the Dynamic Quantile Test and Kupiec LR test. Based on these evaluation criteria EGARCH is found superior to GARCH, GJR, IGARCH, RiskMetrics and APARCH. Additionally, EGARCH is found satisfactory when evaluating 5, 10 and 20-day-ahead forecasts using the Kupiec LR test on Monte Carlo simulated VaR levels. To investigate the drivers of volatility, proxies for each determinant are included in the conditional volatility models and in an OLS framework. Economic activity, seasonality and daily effects are found to be statistical significant, with the daily effects having the largest influence, while oil volatility, changes in temperature, production and storage levels are insignificant.

From the results it can be concluded that if the aim of the conditional volatility modeling is short-term forecasting, the determinants should be excluded as they do not improve forecasting accuracy. Conversely, if the aim is to explain the causes of volatility, the in-sample evaluation indicate that the inclusion of determinants is a reasonable approach, and a good foundation for scenario analyses. Our findings are useful for producers, traders, risk managers and other market participants as they provide an accurate measure of price risk, and can be used to understand the causes of volatility.

Sammendrag

En grunnleggende forståelse av pris-volatilitet i naturgassmarkedet er viktig fordi det påvirker nye investeringer og adferden til markedsaktørene. I denne oppgaven er volatiliteten til amerikanske gasspriser undersøkt ved anvendelse av daglige Henry Hub futures data for perioden 1996 til 2011. Hensikten er å finne den beste betingede volatilitetsmodellen til modellering og forecasting, og å analysere volatilitetens grunnleggende drivere. Flere modeller anvendes og sammenlignes på grunnlag av forklaringskraft, tester, samt deres prestasjoner på prognoser in -og out-of-sample én dag frem, evaluert med Kupiec LR og Dynamic Quantile Test. Basert på disse evalueringskriteriene konkluderes det med at EGARCH er bedre enn GARCH, GJR, IGARCH, RiskMetrics og APARCH. I tillegg er EGARCH funnet tilfredsstillende ved evaluering av 5, 10 og 20 dagers prognoser med Kupiec LR testen på Monte Carlo simulerte VaR nivåer. For å undersøke driverne av volatilitet er determinantene kvantifisert og inkludert i de betingede volatilitetsmodellene og i et OLS rammeverk. Økonomisk aktivitet, sesongvariasjoner og daglig effekter er funnet å være statistisk signifikante, der de daglige effektene har størst påvirkning, mens oljevolatilitet, endringer i temperatur, produksjon og lagring er statistisk ikke-signifikante.

Fra resultatene kan det konkluderes at dersom målet med modelleringen er kortsiktige prognoser, bør determinanter ekskluderes ettersom de ikke bedrer prognosenes nøyaktighet. Derimot, dersom målet er å forklare årsakene til volatilitet, tyder in-sample evalueringen på at inkludering av determinanter er en fornuftig tilnærming, og et godt grunnlag for scenarioanalyser. Våre funn kan være nyttige for produsenter, tradere, risikostyrere og andre aktører ettersom de gir et nøyaktig mål på risiko, og kan brukes til å forstå årsakene til volatilitet.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

"The US gas rollercoaster" is what one of the leading economist on gas, Jonathan Stern at the Oxford Institute of Energy, has nicknamed the development of US natural gas prices (Foss 2007). Prior to 2000 the prices were regarded as relatively stable, but during the last decade prices have varied between $2 and $15 per MMBtu[1], with an increasing trend until 2008. In 2009 the market experienced a plunge in prices due to the economic recession and the addition of shale gas. This has made the market evolve from a seller's market driven by tight supply to a buyer's market. However, since the liberalization of the natural gas market, the short-term volatility has been high regardless of the price level. Second to electricity the natural gas market is the most volatile energy market in the US (Henning, Sloan et al. 2003).

High price volatility creates uncertainty which market participants, from risk managers, production planners to end-users, are forced to deal with. Furthermore, assessment of volatility dynamics is of interest to non-industrial market participants who may take a speculative position in the market, such as banks, trading companies and institutional investors. The US government is concerned with price volatility in terms of security of supply since volatility affects the willingness to invest in the natural gas value chain and the potential effect it may have on the economy. In 2009, 7.5 % of US GDP was spent on energy, of which 11 % on natural gas.[2]

A comprehensive understanding of market dynamics, modeling and forecasting in the natural gas market holds a significant role on strategies towards attaining future positions in the market. The aim of this paper is to model volatility in the US natural gas market, and to forecast volatility in the short-term, defined as 1, 5, 10 and 20 days ahead. This has several applications, such as pricing of derivatives, portfolio selection and Value at Risk (VaR). A vast amount of financial literature on the modeling and

[1] Million British thermal unit
[2] Own calculations based on numbers from http://www.eia.gov/

forecasting of volatility has emerged together with several extensions of conditional volatility models[3] of which six will be adopted in this study. The family of GARCH models is preferred over other techniques such as moving average, linear regression and exponential smoothing (Sadorsky 2006a). Although implied volatility is commonly perceived as a better measure of volatility (Duffie and Gray 1995), this is not considered due to the unavailability of options data.[4]

According to Engle (2001) 'the goal of volatility analysis must ultimately be to explain the causes of volatility'. This study follows up on this statement by combining a fundamental understanding of volatility and its drivers through modeling and forecasting. In studies by Mastrangelo (2007) and Serletis and Shahmoradi (2006) an OLS analysis is applied in order to explain the determinants of volatility in the natural gas markets, while other studies use determinants in the conditional variance equation (Pindyck 2004; Mu 2007). In order to strengthen the final conclusions in this paper, both methodologies are applied in order to assess the volatility fundamentals.

This paper is believed to be an improvement to the existing body of literature based on four main aspects. Firstly, it contains updated analyses of volatility in the natural gas market, which should be of interest to market participants. Secondly, the inclusion of a larger number of determinants compared to existing literature is believed to strengthen the understanding of the market dynamics. Thirdly, several models are assessed in order to find the model best suited for forecasting and modeling in the US natural gas market. Existing literature investigating fundamentals of volatility usually employ only one specific model, whereas this paper uses several different models. To the best of our knowledge there is no study on the US natural gas market which addresses both the causes of volatility and evaluates the model best suited for modeling and forecasting. Lastly, research on modeling and forecasting issues in energy commodity markets is in large limited to oil-related commodities; thus, this paper is believed to fill a gap in the literature.

This paper is structured as follows: the next chapter introduces the natural gas market and gives a description of the value chain and market participants. In chapter 3 a short overview of relevant literature on volatility is given, and based on a description of several fundamental drivers of volatility, hypotheses on their expected influence is presented. Chapter 4 includes data and characteristics for natural gas returns, while chapter 5 presents the methodology used in this paper. In chapter 6, results are evaluated through statistical tests, and the results and their implications are further discussed in chapter 7. Chapter 8 contains an elaborate discussion of the determinants' impact on natural gas volatility. Finally, concluding remarks are presented in chapter 9.

[3] See Bollerslev, Chou et al. (1992) for a survey of the GARCH literature and Poon and Granger (2005) on related forecasting
[4] Implied volatility can be obtained through purchase of options data on NYMEX

Chapter 2

Market Overview

Today, the US is the world's largest market and consumer of natural gas. According to the US Energy Information Administration (EIA), natural gas accounts for about 25 % of primary energy used in the US by source (EIA 2011a), which in 2010 corresponded to a total demand of nearly 24 trillion cubic feet (Tcf). After Russia, the US is the second largest producer of natural gas with over 21 Tcf of dry gas supply (EIA 2011b). The yearly supply gap is in large covered by imports from Canada and Mexico, in addition to liquefied natural gas (LNG) from the world market. Unlike the oil market, the domination of pipeline infrastructures has essentially divided the natural gas market into major regions. Pipeline infrastructure is regarded as a capital intensive and inflexible way of transporting gas from producers to consumers (Söderholm 2000); however, the addition of LNG is believed to move the regional markets closer (Shively and Ferrare 2007).

Since the liberalization, starting with the Natural Gas Wellhead Decontrol Act in 1989, the market has been subject to several changes.[5] In the years following 2000, a future of tight gas supply was anticipated. Increasing prices and decreasing gas resources were expected to be balanced by increased imports of LNG (Henning, Sloan et al. 2003; Youngquist and Duncan 2003). However, instead of becoming a substantial importer of LNG, the successful use of horizontal drilling in conjunction with hydraulic fracturing, starting in the period 2004-2006, has greatly increased the profitability and prospects of recovering natural gas from low-permeability geological formations, especially shale gas (EIA 2011d). As indicated in Figure 1, the amount of unconventional gas[6] has increased twelve fold from 2000 to 2010, off-setting the effect of decreasing conventional supplies. Thus, at times the US

[5] See Smead (2010) and Henning, Sloan et.al (2003) for a detailed account

[6] Unconventional gas is a term generally connected to gas reservoirs that require treatment to stimulate gas production such as shale gas, tight gas and coalbed methane

production of natural gas is expected to exceed consumption, with the potential of making the US a net exporter (EIA 2012).

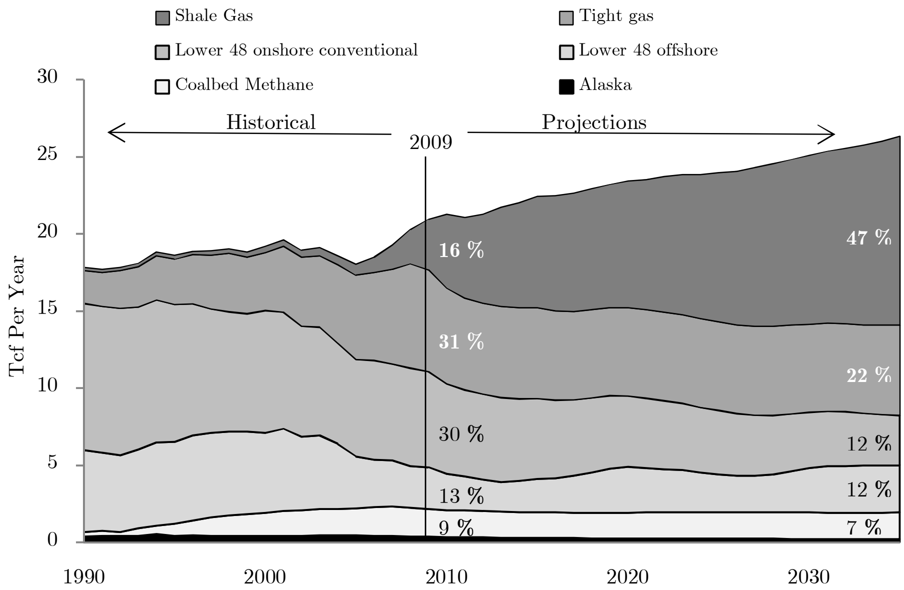

Figure 1: Projections for US natural gas produciton towards 2035 indicates a larger domestic supply due to unconventional gas (EIA)

2.1 Value Chain and Market Participants

When assessing the volatility in the natural gas market, it is important to understand the different roles the market participants hold, and how they react to and influence price volatility. Similar to other fossil fuels, the natural gas value chain can be divided into upstream, midstream and downstream activities. Some of the market participants are associated with several of the activities in the value chain, e.g. financial services, marketers, integrated energy companies and storage providers (Shively and Ferrare 2007).

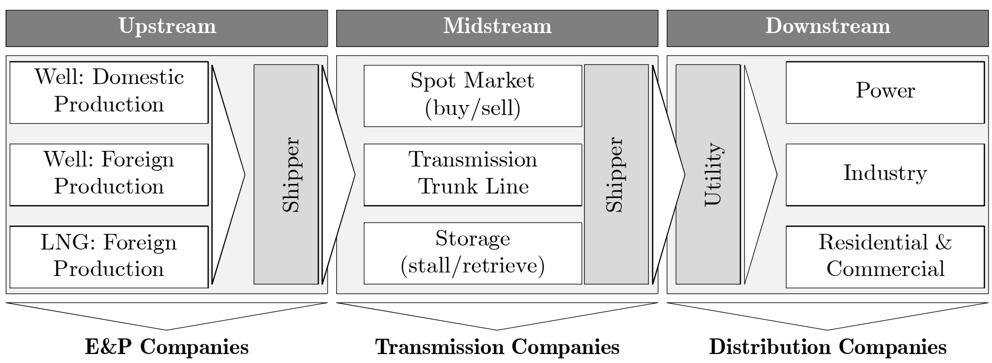

Figure 2: The natural gas value chain can be divided into three major activities, similar to many other fossil fuels

2.1.1 Upstream

Upstream activities consist of exploration, drilling and production from gas reservoirs, traditionally conducted by the large exploration and production firms.[7] The increase in unconventional gas production has, on the contrary, been driven by smaller companies.

Large upfront investments are required prior to production which makes risk management and financial services vital. Capital at a reasonable rate and hedging against long periods of low prices are important factors for successful investments. Once a field is up and running, the operational expenditures are so small compared to the capital expenditures that the benefits of a consistent level of gas produced and transported is regarded as higher than the drawback of low prices. This usually leaves the only interruptions in production to be related to shut-downs due to maintenance, or shut-ins due to extreme weather such as hurricanes.

After extraction from the reservoir, raw gas is connected to the transmission system through the gathering system.[8] The gas is either sold directly from the production company to end users, through aggregators who pool gas supplies and sell them in blocks, or through marketers in the midstream sector.

2.1.2 Midstream

Midstream activities are generally associated with the transmission of gas through pipelines and include transportation, storage and trading. The transmission system is responsible for moving the gas from the producer or aggregator to consumers, covering large distances, where marketers often act as a link between the two. In addition, marketers arrange transportation and storage, and sometimes provide services such as risk management and financing.[9] The shipper is responsible for the actual transmission, and may be any market participant holding a contract to transport gas in a pipeline.

Since the gas produced in the upstream sector is kept at a consistent level, it is the storage facilities in the midstream sector that are able to dampen price volatility by serving as a physical hedge in periods of high or low demand. Long-term storage is met by storage providers operating underground facilities, and when this is not available LNG is used. Short-term storage, often known as balancing or parking, is provided by pipelines, hub operators and local distribution companies (LDCs).

---

[7] Since gas and oil have often been found together large oil corporations such as BP, Chevron, ConocoPhillips, ExxonMobile and Shell are also large gas producers

[8] A gathering system usually consist of a system of small pipes which delivers the gas to a processing plant (less frequently the gas is processed at the wellhead). At the processing plant the gas (methane) is separated from impurities and valuable byproducts known as natural gas liquids (NGL), consisting of ethane, propane, butane and pentane.

[9] Today marketers are a combination of the big oil companies that market gas directly, financial houses that do physical marketing (e.g. Goldman Sachs, Merrill Lynch, and UBS), utility-based trading subsidiaries and smaller regional marketers (Shively and Ferrare 2007)

There exist several points of trade in the US (Figure 3). These are usually physical locations, called hubs, where multiple pipelines intersect. Henry Hub in Louisiana, the nexus of 16 intra- and interstate natural gas pipeline systems, is the hub with the largest traded volume, and is considered the price setter for other regions (Walls 1995). The pipelines from Henry Hub are connected to markets throughout the US East Coast, the Gulf Coast, the Midwest, and up to the Canadian border.

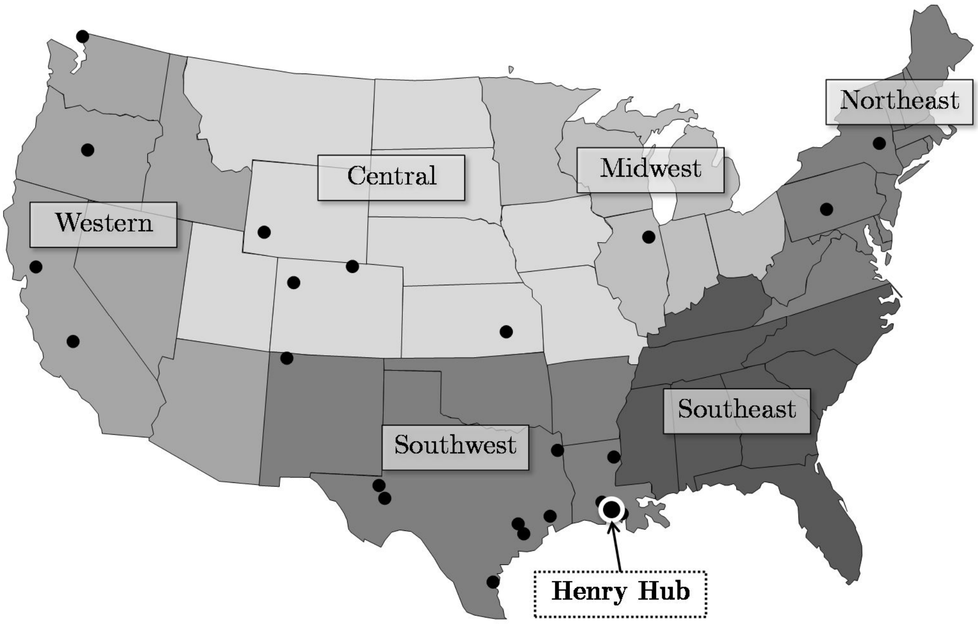

Figure 3: Of the US natural gas pipeline hubs, the Henry Hub is the largest and considered the most important

The market for Henry Hub futures contracts opened on NYMEX April 3rd 1990 and is regarded as the leading benchmark for the North American natural gas market. It is the second most-actively traded futures contract in the world based on a physical commodity, and far more liquid than the spot. In 2011 and 2012 it has been consistently traded with a daily volume above 300,000 contracts (CMEGroup 2012a). Compared to the spot price it is also more reliable as spot prices are not recorded at a centralized exchange, but by agencies basing their price estimates on polls from traders. Thus, only Henry Hub futures are considered in this paper.

The Henry Hub natural gas futures contract is a binding legal obligation to make or take delivery in a particular future month. The underlying asset of one contract is 10,000 MMBtu of pipeline-quality natural gas delivered at the Henry Hub in Louisiana. The contracts are traded up to three business days prior to the first day of the delivery month, and offered up to 12 years ahead (CMEGroup 2012b).

2.1.3 Downstream

Downstream refers to the distribution and marketing to wholesalers and retailers, and consumption by end users. In 2011 the electrical power generation was the largest end-user of natural gas, accounting for about 31 % of total consumption, followed by the industrial sector consuming around 28 %. Residential and commercial customers are responsible for 13 and 14 % of the consumption, respectively (EIA 2012b).

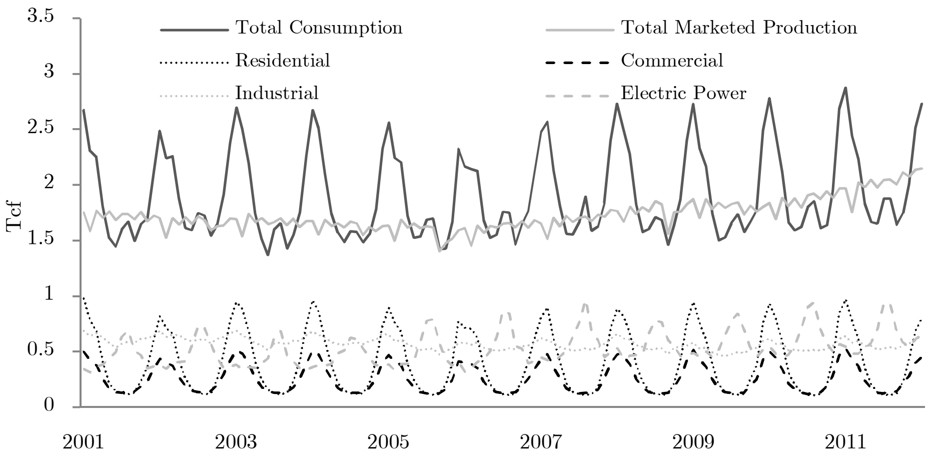

Figure 4: US consumption by source display a pattern of seasonality in demand (EIA)

Residential gas consumption exhibit inelasticity to prices in the short term and follows the seasons as natural gas is in large consumed for heating. As indicated in Figure 4, most consumption occurs in the period November through March. The consumption is sensitive to weather, and in the long run it is partly based on past prices as this affects the investments into new energy equipment. The same holds for commercial customers who's consumption is generally driven by weather and business, but their seasonal consumption is less dramatic.

Larger consumers in the industrial and electric generation sector tend to turn to marketers or producers directly to buy gas. Price sensitivity is reflected in both the day-to-day business, and in long term investment decisions. Industrial gas use tends to be more volatile in the short term than the residential and commercial sector due to the close link to the overall business market. In addition, a share of the industrial and electrical generation sectors have a dual-fuel capability, which makes them able to respond to changes in price by switching fuel. In 2002, around 20 % of US industrial gas consumption could be switched to other fuels (EIA 2002). An alternative to switching is the option of shutting down production when prices are high, or in electrical power generation decide which plants to operate (Brown and Yücel 2008).

Electrical generation by natural gas provides the majority of the marginal power generation capacity in the US, meeting the seasonal peak demand. Natural gas tends to be a high-cost fuel as opposed to most base load electricity generation delivered by coal and nuclear. Thus, contrary to residential and commercial customers, gas consumption is highest in the summer due to the high demand created by air conditioning. Natural gas power plants tend to have lower up-front capital cost, which implies that the economics of natural gas power plants are dependent on future gas prices. High volatility makes gas investments less attractive relative to coal and other alternatives with more stable fuel prices (Henning, Sloan et al. 2003).

Chapter 3

Volatility in the US Natural Gas Market

3.1 Literature review

The literature review is divided in two parts; the first section gives a short overview of research related to modeling and forecasting in energy commodity markets, and the second section presents research on fundamentals of natural gas volatility.

3.1.1 Modeling and Forecasting

Research addressing different class of models and forecasting of futures in the US natural gas market is to our best knowledge limited. Focusing on daily spot prices of petroleum commodities, Hung et al. (2008) use their findings from three models from the GARCH family to suggest that heavily-tailed distribution is more suitable for energy commodities due to the leptokurtic features exhibited in asset returns. Aloui (2010) confirm these results by computing the VaR of four major oil and gasoline commodities using three different GARCH models. In addition, evidence of asymmetry and long memory was reported. Focusing on petroleum futures, Sadorsky (2006a) concludes that different models suit different markets, and that TGARCH (Glosten, Jagannathan et al. 1993; Zakoian 1994) fits better than less sophisticated models.

Yaffy, Heddy et al. (2008) do a comprehensive study on forecasting and the use of different GARCH models on natural gas futures, concluding that Risk Metrics (Morgan 1996) and Asymmetric Power GARCH (APARCH) (Ding, Granger et al. 1993) are preferred; however, the paper is limited to the UK natural gas market. Satisfactory results from the APARCH model is also found by Giot and Laurent (2003) who test VaR models for several commodity spot prices including Brent and WTI [10] for one-day-ahead forecasts. Gogas and Serletis (2010) on the other hand, find that for three different forecasting horizons[11], and for both static and dynamic forecasts, an EGARCH specification is preferred in the natural gas market.

3.1.2 Fundamentals

There are several papers on the US natural gas market addressing fundamentals of volatility. A common weakness with these studies is that they assume that the models used are suitable, without evaluating the models specifications and performance such as validity of error terms and model fit. This may provide an incorrect relationship between the natural gas volatility and the fundamentals.

EIA (2007) analyze the volatility in spot prices for the period 1994-2006, concluding that there is no long-term trend in volatility, and there exist seasonal patterns and a strong correlation with storage dynamics. However, the application of an OLS framework does not account for non-linearity, and is a weakness of their study. Serletis and Shahmoradi (2006) also employ an OLS model in their investigation of determinants of volatility for the Henry Hub futures contracts. The difference is that they first use a regular GARCH(2,1) model, following the methodology of Liew and Brooks (1998), to provide an estimate for the conditional volatility, which is then used as a dependent variable in an OLS regression. They find significant evidence of seasonality, yearly and open interest effects in both returns and volatility; however, they fail to report test results for their models.

Mu (2007) models daily Henry Hub futures using vanilla GARCH models[12] to show that weather shocks have a significant effect on the conditional volatility. By including exogenous variables in the conditional volatility model it is found that persistence is reduced by 40 %, corroborating the importance of fundamental factors driving volatility. Mu (2007) is one of few studies including several drivers in the conditional volatility model; however, since only one specific model is used, efforts should have been made to evaluate the validity and performance of this model. In addition, the market dependencies does not appear to be fully investigated, which is evident when only the winter season is included. The weather effect in the natural gas market is also assessed by Fleming, Kirby et al. (2006) who find a strong link between public information flow and volatility by comparing the trading versus non-trading period variance ratios.

Ewing, Malik et al. (2002) investigates the volatility transmission between two indexes based on major companies in the oil and gas sectors. It is argued that the behavior of natural gas volatility differs from that of oil; the natural gas

[10] West Texas Intermediate (WTI), also known as Texas light sweet. Grade of crude oil used as benchmark in oil pricing

[11] One, two and four weeks ahead

[12] Vanilla GARCH refers to a GARCH(1,1) model with a Gaussian error distribution

sector is more persistent and directly affected by shocks such as events or 'news' in its own sector and indirectly by the oil sector. This may imply that the inclusion of market information when modeling and forecasting natural gas volatility is valuable. Larger persistency is interpreted by Ewing, Malik et al. (2002) as a potential economic benefit as market participants would have longer time to react to a potential shock, given that it is properly understood. Pindyck (2004) contradicts parts of Ewing, Malik et al.'s (2002) conclusions by arguing that the shocks to volatility in both oil and gas commodities are transitory. However, the different results could easily be related to the fact that Ewing, Malik et al. (2002) analyzes times series based on the stock market while Pindyck (2004) use price series based on futures.

Pindyck (2004) estimate volatility through weekly sample standard deviations of returns and by using weekly and daily GARCH models; however, the conclusions drawn in the paper does not differ between the models applied. A statistically significant positive trend in volatility for natural gas is found, but it is not regarded as significant in economic terms. In addition, no increase in volatility was found due to the Enron collapse. The insignificance of the Enron collapse on volatility is also confirmed by Murry and Zhu (2004). This result is rational as Enron held the position of a marketer in the gas market; the disappearance of a marketer should not affect the physical demand and supply balance in the gas market (Shively and Ferrare 2007).

Since the market liberalization the natural gas price has gone from being less volatile than the oil market, to more. Susmel and Thompson (1997) find an increase in investments in storage facilities due to the increase in volatility following the liberalization, arguing that the regulatory changes taking place during the sample period is why a regime switching GARCH (SWARCH) model outperforms a GARCH model. Susmel and Thompson (1997) also confirm seasonality in price and variance in relation to storage levels. Although several factors influencing volatility are investigated, and considerations of model fit is undertaken, an improvement would be to give a fundamental analysis as to how these dynamics function, not just confirm the existence of volatility drivers.

3.2 Determinants of volatility

The fundamental drivers of volatility can be divided into supply and demand which gives a clear structure in the analysis; however, it should be noted that distinguishing between different factors is complex. Factors such as seasonality and day-of-the-week effects are hard to define as either supply or demand, and are thus not categorized. Demand factors (section 3.2.1-3) considered are temperature, economic activity and substitutes, while supply factors (section 3.2.4-5) are storage dynamics and changes in production.

3.2.1 Substitutes

For most end-users of natural gas flexibility is very limited in the short-term as they are not capable of switching to an alternative fuel on short notice. However, both the industrial and electrical generation sector holds fuel switching potential in order to react to fluctuating prices. In the electricity producing sector, gas competes primarily against coal, heavy fuel oil, nuclear power and renewables. The main competing fuels in the industry are heavy fuel oil, coal and electricity, whereas heating oil and electricity are the principal competitors in the commercial and residential sectors.

In the long term, demand for gas is sensitive to the price of gas relative to other fuels (IEA 2009). This implies a correlation between relevant fuels in terms of both prices and volatility. In periods where electricity demand is relatively low, coal prices function as a floor for gas prices. An example is the sharp decline in natural gas prices in 2008 and 2009. This led to significant substitution from coal to gas in the power sector, resulting in a boost in gas demand, preventing the gas price to fall below coal prices (IEA 2009).

Most literature focuses on the relationship between oil and gas, and this is also the focus of this section. Villar and Joutz (2006) find Henry Hub futures prices to be co-integrated[13] to WTI with a long term relationship. In the period 1989-2005 permanent and temporary shocks to the WTI price are shown to be transmitted to the Henry Hub price. Hartley, Medlock et al. (2008) investigate an overlapping period, confirming Villar and Joutz (2006) results. The same holds for Pindyck (2004) and Ewing, Malik et al. (2002), concluding that oil volatility has spillover effects on gas volatility, but not the other way around. Brown and Yücel (2008) also find a relationship between the gas and oil price, but argue that the gas price shows a tendency to move more independent since the number of fuel switching facilities has been considerably reduced over the last 15 years. This is supported by Bachmeier and Griffin (2006) who only find a weak relationship between oil, gas and coal.

With recent oil and gas prices having record high and low levels, respectively, some suggest that the changes in the North American gas market dynamics may be proof of a permanent rapture between the two prices (IEA 2009). Ramberg and Parsons (2010) addresses this view and provide evidence that the relationship is capable of shifting, implying that despite large temporary deviations there exists evidence of a co-integrating relationship and that it is only the equilibrium between the two prices that is changing.

Due to the spillover effects from the oil market, the following relationship is expected:

H1: *An increase in oil volatility causes higher volatility*

[13] Co-integration is a measure of long term dependency between asset prices; see Alexander (2001a) for details

3.2.2 Temperature

The end-user demand is largely driven by seasonality and changes in temperature. These factors are known to be one of the strongest short-term influences on volatility (Mastrangelo 2007). Weather shocks due to unexpected cold or warm weather increases demand which then causes prices to spike and increase volatility significantly (Mu 2007). Days with unexpected cold weather in the winter season is especially susceptible to volatility for two reasons. Firstly, peak demand is met by storage levels that may be fluctuating. Secondly, during the winter the transmission system may already be operating at maximum capacity. When this is the case, a balanced market can only be achieved by increasing prices to reduce demand. Similarly, if a warm summer is experienced more gas than expected is consumed, increasing prices.

H2: *Temperature deviating from the average causes higher volatility*

3.2.3 Economic Activity

Economic activity influences the natural gas market in large through its commercial and industrial end-users. A recession would reduce the demand for goods and services, leading to a decrease in industrial and commercial production rates, which in turn reduces gas demand. Residential consumers on the other hand are less flexible – if it is cold, heating is needed no matter the state of the economy. However, in periods of economic growth, increase in personal disposable income may lead to increase in residential demand (Mastrangelo 2007). This factor may also be related to the supply side, as periods of economic downturn reduce the willingness to take on risky drilling and production projects, as well as investments in natural gas facilities.

There are several ways to measure economic activity, such as GDP, industrial production, treasury bills or the stock market. Serletis and Shahmoradi (2005) suggests that natural gas prices are cyclical and can be expressed though industrial production. However, as this paper investigates daily futures prices of natural gas, a daily proxy of economic activity is required. It is a common perception that the stock market contains important information related to economic activity (Fama 1981). The volatility of the S&P 500 represents the systematic risk in the market. If the economy is stable, this also implies that the volatility of the S&P 500 is stable. Similarly, when there are recessions and upturns in the overall economy, this will lead to higher volatility in the S&P 500. This is proven by Hamilton and Lin (1996) and Schwert (1989), who find economic recession to be the single largest factor accounting for volatility in stock returns. To the best of our knowledge there is no paper investigating the impact of economic activity on natural gas volatility. Thus, assuming the volatility of the S&P 500 to reflect changes in the business cycle, it should have spillover effects on the gas volatility implying the following relationship:

H3: *An increase in S&P 500 volatility causes higher volatility*

3.2.4 Storage

Storage levels receive considerable amounts of attention because of the physical hedge it provides during periods of high and low demand. The seasonal pattern displayed by the storage level is divided into two periods due to the consumption pattern. The period April to October is defined as the injection season while November to March is called the withdrawal season. On average storage reservoirs supply around 20 % of the gas consumed and up to 50 % of consumption on days with peak demand during the withdrawal season (EIA 1995).

According to Alterman (2012), September and October exhibit an especially high volatility since it is the end of the injection season, and the gas available to meet winter demand is from that point on finite while the weather conditions for the upcoming winter are uncertain. If an especially warm summer occurs, the injection season will be shortened, which may cause unsatisfactory storage levels for the winter season, in turn causing high volatility during the winter. Thus, one should be aware that the storage factor may be driven by temperature, and as the withdrawal season starts in November, storage owners might be reluctant to withdraw large amounts due to uncertainty related to demand later in the winter (Mastrangelo 2007). Independent of seasonality, the storage level gives an indication of the storage facilities ability to balance the market.

H4: *Storage levels deviating from the average cause higher volatility*

Every Thursday the market receives information on inventory levels for working gas in storage across the US through the weekly update storage report published by EIA.[14] This report is viewed as a good indicator of the supply and demand balance in the market. Following the theory of storage (Deaton and Laroque 1992; Pindyck 1994; Deaton and Laroque 1996) the commodity price is inversely related to storage level as an increase in the gas price is connected to either a decrease in supply or increase in demand, or both. This suggests that an unexpected change in the storage level may create a shift in the perceived balance of the market, creating uncertainty which is subject to different interpretation by different market participants (Linn and Zhu 2004). Linn and Zhu (2004) investigate intraday volatility and find that the volatility is significantly impacted by the release of the weekly update storage report. This impact is also confirmed by the findings of Murry and Zu (2004) and Mu (2007), which corroborates that news has a substantial effect on volatility.

H5*: The release of the weekly update storage report causes higher volatility*

[14] Before April 2002 the report was compiled by the American Gas Association (AGA) and released every Monday

3.2.5 Production

Production is normally only disrupted by maintenance work, which is put to periods where production levels are higher than consumption, causing minimal impact on volatility. Other factors causing changes to production levels can be divided into short-term and long-term factors.

Disruptions in natural gas production from outside influences, such as periods with shut-ins due to hurricanes and other severe weather, have traditionally caused high volatility and can be regarded as short-term factors. In 2005, hurricanes along the US Gulf Coast caused shut ins of ~4 % of US total production between August 2005 and June 2006 (EIA 2011c). Alterman (2012) identify six periods of high volatility due to hurricanes since 1997 suggesting that their impact on volatility has been considerably reduced since around 2005. [15] This is attributed to milder winters and new onshore production from shale gas, which has replaced three times the sustained loss of offshore supply (Smead 2010).

The impact of depleting gas resources or additional supply through new production can be regarded as long-term factors. Depleting resources would increase volatility as one must rely on imports to satisfy demand (Henning, Sloan et al. 2003). The addition from domestic unconventional gas on volatility has, to our best knowledge, not been properly investigated. More supply means a consistent rise in the average gas level in storage. Some suggests that this may contribute to enhanced price stability and a low volatility environment in the US natural gas market (Brown and Krupnick 2010; Alterman 2012). However, with the plunge in prices in 2009 partly due to the recession, together with limited data on shale gas production, the only factor included in this study is the change in production as a whole.

H6: *Production levels deviating from the average cause higher volatility*

3.2.6 Seasonality and the Monday Effect

The impact of seasonality is mentioned in several of the supply and demand factors. Demand varies with the season, and may exceed the infrastructures ability to deliver gas in extreme cold weather during the winter. Since demand usually exceeds production during winter it is left to storage facilities and imports to balance the market. Warm summers reduce the storage levels for the winter, which imply that spring may be the period with lowest volatility. Interruptions in the gas production also follow the season to some extent; maintenance is planned for periods of low demand and shut-ins occur in the hurricane season.[16] Alterman (2012) state that winter and fall have the highest volatility, supported by results from Mastrangelo (2007) and Serletis and

[15] Weaknesses in Alterman's (2012) conclusions should be noted, as they are only based on a qualitative observations

[16] Hurricane season in the Atlantic begins in June 1st and ends November 30th. The Eastern Pacific hurricane season begins May 15th and also ends November 30th (http://www.nhc.noaa.gov).

Shahmoradi (2006); however, their results disagree on which season has the highest volatility of the two. It is believed that the winter has a larger impact on volatility due to the inelasticity of supply and demand.

***H7:** Winter exhibits the highest volatility, followed by fall, summer and spring*

Evidence of a Monday effect in the natural gas market is found by Serletis and Shahmoradi (2006), Fleming, Kirby et al (2006) and Mu (2007). The Monday effect is the increase in volatility due to the information acquired during the weekend. Because of the market's inability to react to the events during the weekend, Monday will accumulate the entire effect of the news from late Friday to early Monday. Fleming, Kirby et al (2006) and Mu (2007) argue that this effect is due to news about weather information generated during the weekend; however, the attribution of the Monday effect as a result of weather is out of the scope of this paper.

***H8:** Mondays exhibit higher volatility*

Table 1: Summary of the hypothesized impact of determinants on natural gas volatility

Determinant	Hypotheses
Substitutes	+
Temperature	+
Economic activity	+
Storage	+
Update	+
Production	+
Spring	-
Summer	-
Winter	+
Monday effect	+

Chapter 4

Data and Descriptive Statistics

The dataset used in this study include natural gas production and storage levels, publication dates for the weekly gas storage report, S&P 500 index levels, temperature data, and one- and two months ahead daily closing futures prices for natural gas at Henry Hub and WTI crude oil at NYMEX. All data related to energy commodity prices, production and storage levels was obtained from EIA[17]. Time-series for closing prices for the S&P 500 was obtained from π-trading[18], and the temperature data for the ten largest US cities[19] was gathered from the National Climatic Data Center (NCDC)[20].

The times series used are summarized in Table 2. For daily observations where one or more of the below time series are missing a data point, all observations for this data point are excluded. This filtration leads to the exclusion of 3 days, leaving a total of 3,893 observations ranging May 13th 1996 to November 30th 2011.

Table 2: Length, number of observations, frequency and source of the time series used

Time-series	Start date	End date	# obs	Freq	Source
NYMEX HH Futures Prices	20.12.1993	31.01.2012	4539	Daily	EIA
Crude Oil Futures Price	30.03.1983	07.02.2012	7241	Daily	EIA
S&P 500 Index	01.04.1960	02.10.2012	13119	Daily	π-trading
Natural Gas Production	30.01.1981	30.11.2011	8044	Daily	EIA
Natural Gas Storage Level	31.12.1993	03.02.2012	4721	Daily	EIA
Natural Gas Storage Report	13.05.1996	30.11.2011	3893	Daily	EIA
Temperature Data	01.01.1993	31.12.2011	6939	Daily	NCDC

17 http://www.eia.gov
18 http://pitrading.com/free_market_data.htm
19 New York, Los Angeles, Chicago, Dallas, Philadelphia, Houston, Washington, Miami, Atlanta and Boston
20 http://gis.ncdc.noaa.gov/map/cdo/

4.1 Natural Gas Prices

As mentioned previously (section 2.1.2) futures are preferred over spot prices due to liquidity and the decentralized spot price. Because futures contracts have an expiry date, these must be rolled over to create a continuous time series. Following Pindyck (2004), this was done by converting daily futures prices to daily spot prices:

$$P_t = F1_t \left(\frac{F1_t}{F2_t} \right)^{\frac{n_{0t}}{n_t}} \qquad \text{Eq. 4.1}$$

where

P_t : spot price on day t
$F1_t$: price of the nearest futures contract
$F2_t$: price of the next-to-nearest futures contract
n_{0t} : number of days from time t to expiration of the first contract
n_1 : number of days between the expiration dates for the first and second contract

Comparison of the created spot price and the actual spot price indicate that the method used is a reasonable approximation, with the two series having a 99 % correlation (see Appendix I for derivation of the method).

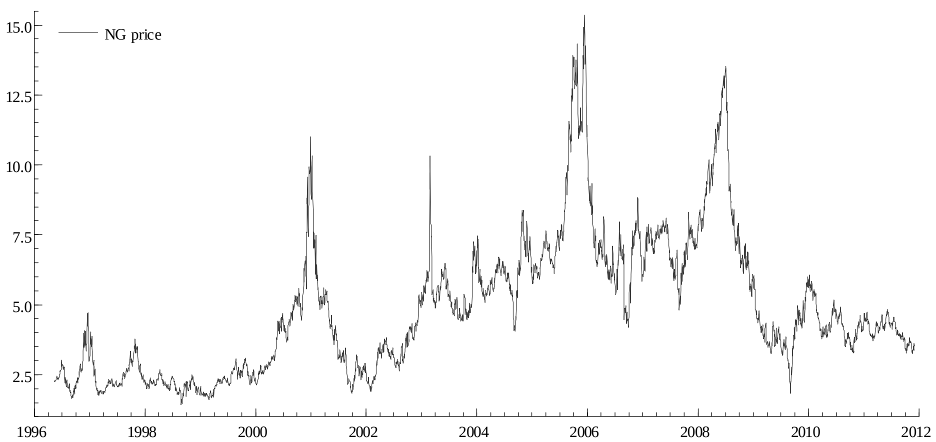

Figure 5: Natural gas created spot series ($/MMBtu) when futures are rolled over using Pindyck's formula

As can be seen in Figure 5, the natural gas price has been fluctuating since the beginning of the sample period. Some noteworthy episodes[21] that have impacted prices are: the Californian energy crisis 2000-2001, the hurricanes in 2005 and the collapse of prices following the summer of 2008 arguably due to the recession and increase in unconventional gas.

[21] For a more detailed description of different episodes causing large changes in prices and high volatility see Alterman (2012), Smead (2010) and Henning, Sloan et.al (2003)

4.2 Returns Characteristics

The daily logarithmic return was calculated using the following formula[22]:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100 \qquad \text{Eq. 4.2}$$

Looking at Table 3, we see that the minimum and maximum return over the period is quite large, consistent with the volatile nature of natural gas prices previously discussed. With an average daily return of 0.01 %, corresponding to an annualized return of 2.5 %[23], the mean return is positive, caused by a risk premium. The standard deviation for daily natural gas returns is 4.01 %, suggesting that the risk is high compared to the return. In comparison, the S&P 500 has an annualized mean return of 4.3 % and a daily standard deviation of 1.34%. This indicates that the natural gas market has a high risk–reward relationship. Natural gas market participants are therefore expected to search for hedging strategies rather than arbitrage possibilities. This highlights the need for analyses of volatility in the natural gas market.

Table 3: Descriptive statistics for the natural gas returns imply a high risk-reward relationship

				Quantiles							Excess
	Obs	Mean	Min	1 %	5 %	95 %	99 %	Max	SD	Skewness	Kurtosis
NG ret	3893	0.01	-33.5	-10.1	-6.10	6.27	10.7	32.9	4.01	0.05	6.70

From Figure 6 there appears to be volatility clustering in the natural gas returns. This is confirmed from the natural gas squared returns (Figure 7), which clearly shows the presence of volatility clustering in the returns series. This indicates that a conditional volatility model is appropriate to model the price-volatility in the natural gas market. To verify this hypothesis further analysis of the autocorrelations in the squared returns is conducted.

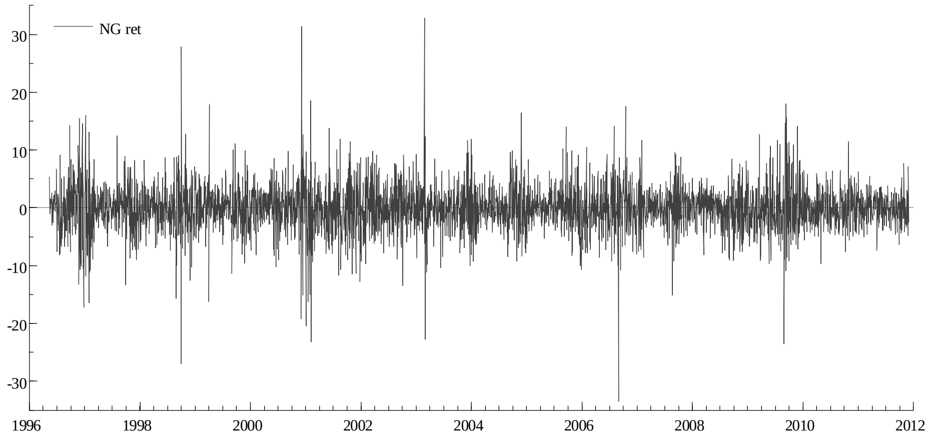

Figure 6: Returns of natural gas created spot, with signs of volatility clustering

[22] This is multiplied by 100 to avoid convergence issues
[23] Using 250 days in a year

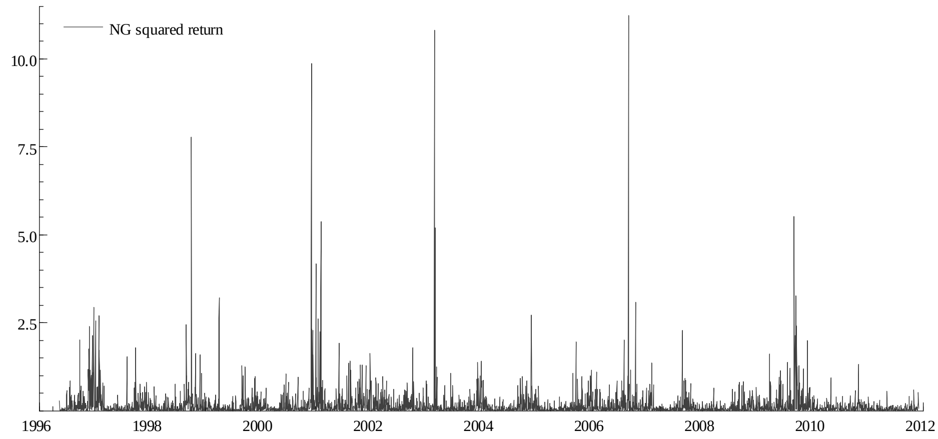

Figure 7: Squared returns of natural gas created spot, confirming volatility clustering

The tests performed on the returns and squared returns are summarized in Table 4, which will be discussed throughout this chapter.

Table 4: Tests performed on returns and squared returns suggest conditional volatility and non-normality

Test	Statistic	P-value
Normality test		
Skewness	0.046114	[0.2400]
Excess Kurtosis	6.7049	[0.0000]**
Jarque-Bera	7293.5	[0.0000]**
Q-statistics on raw data		
Q(5)	32.5996	[0.0000]**
Q(10)	39.4036	[0.0000]**
Q(20)	50.5426	[0.0002]**
Q(50)	88.4439	[0.0007]**
Q-statistics on squared data		
Q(5)	299.071	[0.0000]**
Q(10)	404.846	[0.0000]**
Q(20)	539.996	[0.0000]**
Q(50)	723.36	[0.0000]**
LM ARCH test		
ARCH 1-2	81.698	[0.0000]**
ARCH 1-5	47.766	[0.0000]**
ARCH 1-10	26.68	[0.0000]**
ADF test		
ADF	-35.8573	[0.0000]**

4.2.1 Autocorrelation in Returns

In an arbitrage-free market the returns should not be correlated. Thus, one should not find any specific pattern in the autocorrelations for the different lags.

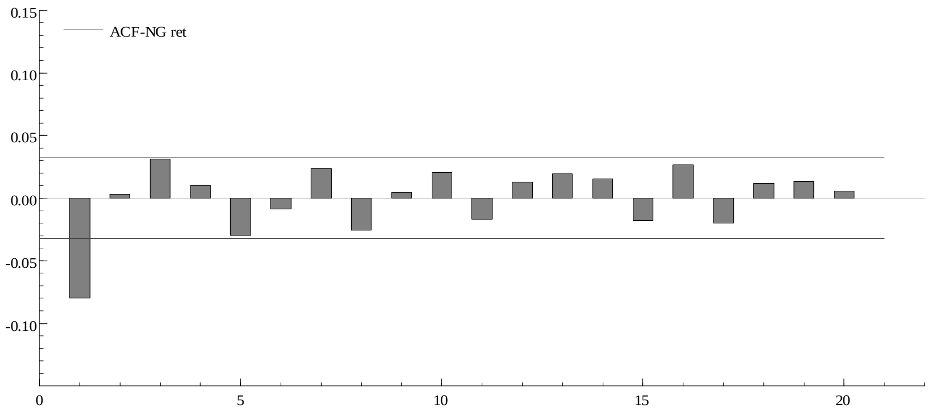

Figure 8: Autocorrelogram for natural gas returns with lags 1-20 display signs of an AR(1)-term

In Figure 8 the straight lines represent the 95 % confidence interval, whereas the solid bars represent the autocorrelation between returns for lags 1-20. The figure indicates no specific pattern or structure; however, there is autocorrelation in returns for one lag. Ideally, such a property in returns should not be encountered and implies a need for an AR(1)-term in the mean equation to account for the autocorrelation. This is also supported by the Box-Pierce test (Table 4) which confirms the presence of autocorrelation in returns.

4.2.2 Autocorrelation in Squared Returns

A popular stylized fact about returns is the existence of a positive dependence between squared returns on nearby days. For natural gas returns this is apparent from Figure 9, which shows significant autocorrelation for every lag. This is also verified by the Box-Pierce statistics for squared returns (Table 4). Since the sample returns exhibits volatility clustering and autocorrelations for squared returns, the use of conditional volatility models to model the price-volatility can be justified. GARCH models are a popular choice, and based on the figure below a large persistence parameter can be expected.[24]

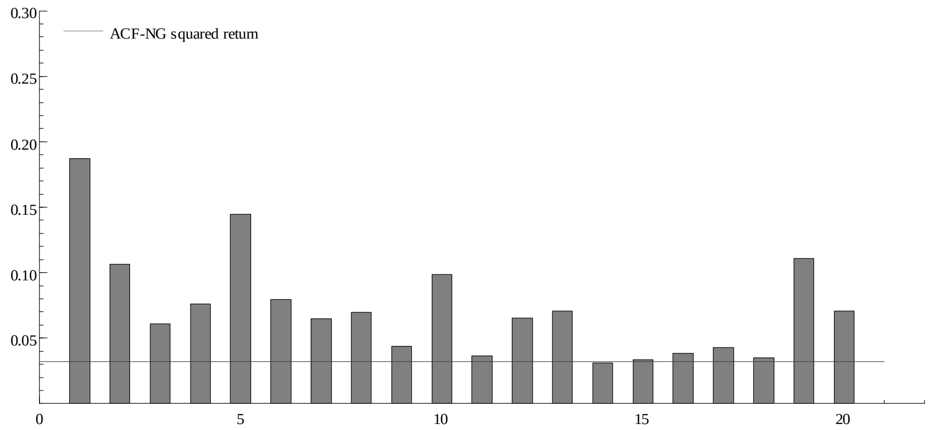

Figure 9: Autocorrelogram for natural gas squared returns with lags 1-20, confirming conditional volatility in natural gas returns

[24] This indicates that we can expect $\alpha + \beta$ to be close to 1 in the GARCH models

4.2.3 Distribution of Returns

Another popular stylized fact for returns is that they are usually not normally distributed; they exhibit high peaks and fat tails. The Jarque-Bera test for normality (Table 4) rejects the hypothesis of normally distributed returns. In addition, leptokurtic[25] properties can be observed from the same table, and from Figure 10. Volatility clustering explains why the distributions of daily returns are not normal. Since the returns exhibits conditional volatility, the complete sample is obtained from a mixture distribution. When this is the case, the kurtosis exceeds the normal kurtosis of three, and fat tails will occur.

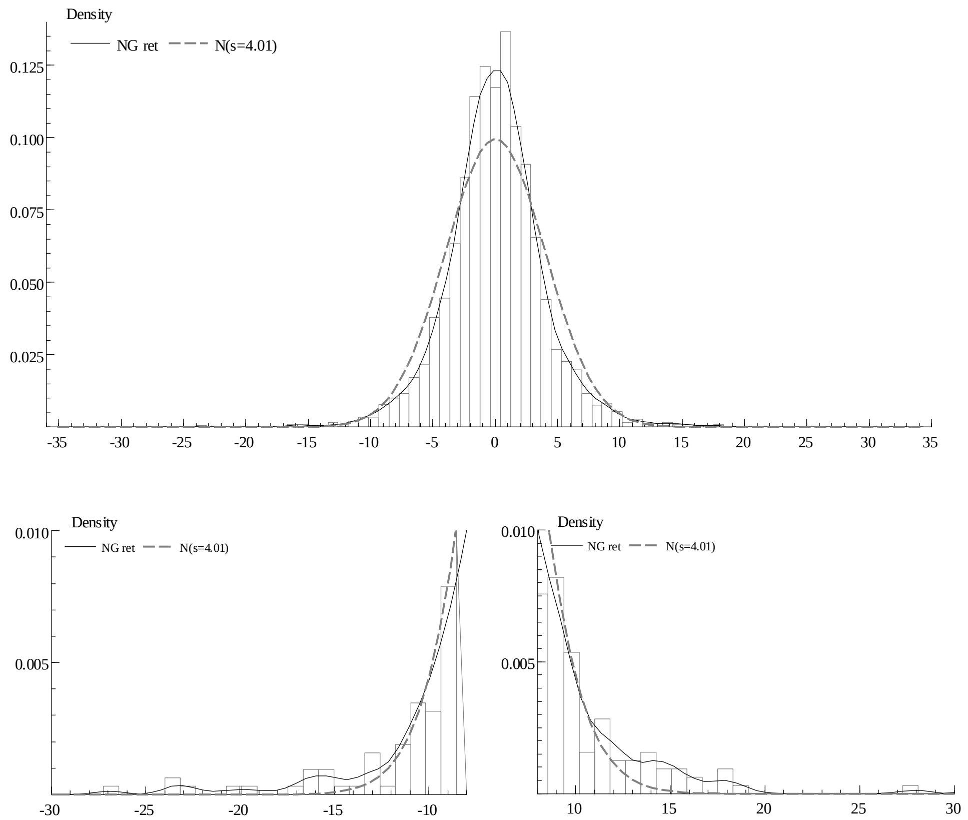

Figure 10: The comparison with the Gaussian distribution suggest leptokurtic properties in the returns distribution

Verification that the returns data is leptokurtic can also be seen from the QQ-plot in Figure 11. In addition, the figure does not show signs of one tail being heavier than the other, implying that the data is symmetrically distributed. This is confirmed by the statistically insignificant skewness parameter the Jarque-Bera test.

[25] Distributions with positive excess kurtosis are called leptokurtic distributions, whereas distributions with excess negative kurtosis are called platykurtic distributions

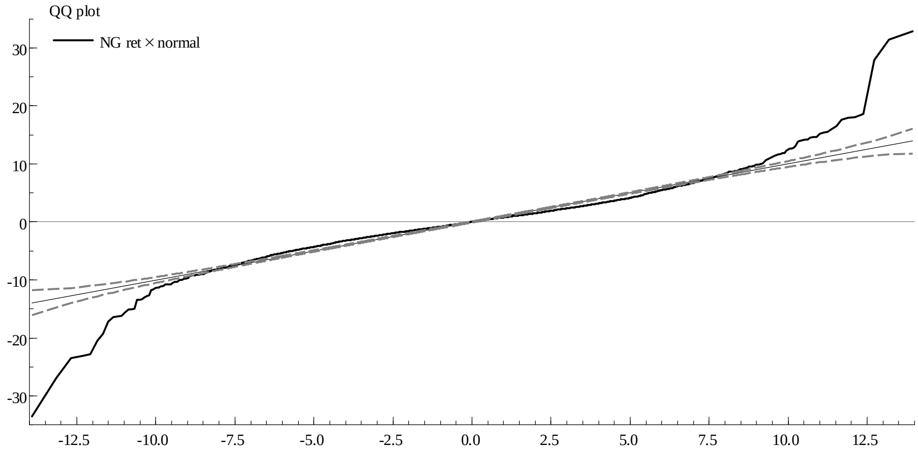

Figure 11: The QQ-plot demonstrate fat-tailed properties in returns and a symmetrical distribution

To model the price-volatility in the natural gas market an assumption for the error distribution is needed. The existence of leptokurtic properties suggests that a student-t or a skewed student-t error distribution is preferred over the normal distribution. In addition, it is reasonable to expect both the student-t and skewed student-t model will perform satisfactory since the returns data appears to be symmetrically distributed.

4.2.4 Stationarity

If the returns series of natural gas exhibits unit roots, certain features changes with time, implying that the returns series is not stationary. The Augmented Dickey-Fuller (ADF) test with two lags is applied to test for stationarity (Table 4). The large negative ADF statistic proves that natural gas returns are indeed stationary, with constant mean, variance and covariance for each lag.

4.2.5 Summary and Implications for Modeling

The annualized standard deviation in Figure 12 shows that the volatility is far from constant. The existence of autocorrelation in the squared returns, together with volatility clustering, implies that the returns of natural gas exhibit conditional heteroskedasticity, which suggest that the family of GARCH models are suitable to model the price-volatility in the US natural gas market. This is supported by the LM ARCH test (Table 4) which rejects the null of no ARCH effects at every lag, and at every level of significance.

The returns distribution shows substantial leptokurtic properties which implies that the student-t distribution is suitable to model the errors. Although a skewed student-t error distribution is preferred in many academic applications, the QQ-plot shows no signs of asymmetry, implying that a skewed student-t error distribution may not be required.

Due to the significant autocorrelation in returns for lag 1, the inclusion of an AR(1)-term in the mean equation will be investigated further.

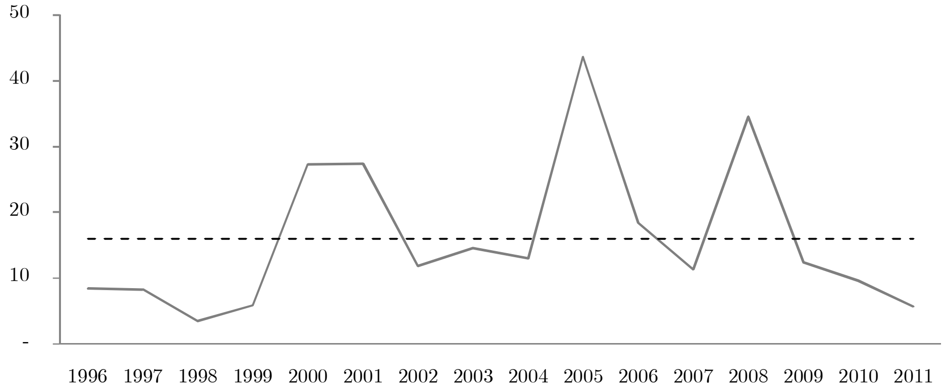

Figure 12: Annualized standard deviation for returns in 1996-2011 indicate that volatility is conditional (dotted line represents the average)

4.3 Summary of Determinants

In order to test and measure the hypothesized impact of the factors influencing volatility presented in chapter 3, proxies for the determinants were created. Further details about the determinants and the construction of their respective proxies can be found in Appendix III.

- Substitutes: oil volatility (Oil vol) is the conditional volatility estimates from a vanilla GARCH model based on returns from WTI
- Temperature: the squared Degree Day difference (DD diff sq) is defined as the squared difference between the degree days[26] for day t and the average degree days for the entire period on this day of the year
- Economic activity: S&P 500 volatility (S&P vol) is the conditional volatility estimates from a vanilla GARCH model based returns from the S&P 500
- Storage: the absolute stock difference (Stock diff (abs)) is the absolute value of the difference between day t's production level and the two year historical average for the corresponding week
- Storage: the weekly storage update report (Update) is the publishing date for this report
- Production: the absolute production difference (Prod diff (abs)) is the absolute value of the difference between day t's storage level and the five year historical average for the corresponding week
- Seasonality: Spring is defined as March to May, Summer as June to August and Winter as December to February
- Monday effect: (Monday) is a dummy for each Monday in the data set

[26] Degree days are the sum of heating degree days (HDD) and cooling degree days (CDD). $HDD = \max\{0, 65 - X\}, CDD = \max\{0, X - 65\}$, where X is the average temperatur for a given day

The summary statistics presented in Table 5 indicate that most of the determinants possess leptokurtic properties and skewness. Comparing the returns for oil and the S&P 500 with the natural gas returns, the absolute values of the minimum and maximum returns are highest for natural gas, followed by oil and then the S&P 500 returns. The mean returns for the S&P 500 and oil are approximately two and four times the mean return of natural gas, respectively. The absolute stock and production differences have low means and standard deviations, together with low levels of skewness and excess kurtosis. The remaining determinants all have a very high excess kurtosis and skewness, together with high means and maximum values for returns.

Table 5: Descriptive statistics show that most of the determinants have skewed and leptokurtic properties

	# obs	Minimum	Mean	Maximum	SD	Skewness	Excess kurtosis
Oil ret	3893	-17.25	0.040	18.646	2.582	-0.048	3.914
SP ret	3893	-9.470	0.017	10.957	1.334	-0.217	6.938
Stock diff (abs)	3893	0.000	0.148	0.649	0.137	1.315	1.153
Prod diff (abs)	3893	0.000	0.066	0.284	0.056	1.089	0.500
DD diff sq	3893	0.000	10.25	195.2	17.29	3.477	16.71
Oil vol	3893	2.489	6.527	53.29	5.124	4.557	26.19
SP vol	3893	0.297	1.791	27.97	2.548	5.522	38.27

Chapter 5

Methodology

In this paper, OxMetrics™ have been used in all model estimations. The quasi-Newton method of Broyden, Fletcher, Goldfarb and Shannon (BFGS) was used when solving the conditional volatility models with the maximum likelihood approach.

This paper investigate two important aspects of natural gas volatility; what factors impact volatility and which model is best suited for modeling and forecasting purposes. In order to provide an answer to our hypotheses on volatility drivers, both an OLS and GARCH models are used. The latter methodology also allows for an evaluation of the potential model improvement from including additional market information.

5.1 Sample Size

The in-sample period is defined from May 14th 1996 to December 9th 2008, consisting of 3,142 observations which are regarded as sufficient for historical VaR estimates as accurate results can be expected above the 99th percentile. The out-of-sample period is defined from December 10th 2008 to November 30th 2011, consisting of 750 observations. This may appear as large for out-of-sample forecasts, but is believed to be required to achieve a sufficient degree of robustness when performing the out-of-sample VaR forecasts.

The sample size chosen is believed to be of importance to the accuracy of forecasts and the coefficients (Angelidis, Benos et al. 2004). As an in-sample period ranging from 1996-2008 is used, the coefficients in the GARCH models will in large be influenced by past observations. Alexander (2008b) argues that the longer the sample period, the more questionable is this assumption 'because a long historical period is likely to cover several different market regimes in which the market behavior would be very different from today'. Thus, the Nyblom test for stability is used in the evaluation of the GARCH models to assess if the sample size is a source of bias in the coefficients.

5.2 OLS

As the conditional volatility is unobservable, an estimate for the volatility is needed in order to perform the determinant analysis in an OLS framework. Following Serletis and Shahmoradi (2006), the time series for the conditional volatility of natural gas $\hat{\sigma}_t^2$ is estimated with a vanilla GARCH model and used as a dependent variable in the OLS model. The methodology is applied on the in-sample period only.

The following forms the basis of the OLS model, with proxies for the volatility drivers as independent variables:

$$\begin{aligned}\hat{\sigma}_t^2 = \quad & \alpha + \delta_1 Oil\,vol_{t-1} + \delta_2\ S\&P\,vol_{t-1} + \delta_3 Stock\ diff\ (abs)_{t-1} \\ & +\delta_4\ Prod\ diff\ (abs)_{t-1} + \delta_5 DD\ diff\ sq_{t-1} + \delta_6 Spring_t \\ & +\delta_7 Summer_t + \delta_8 Winter_t + \delta_9 Update_t + \delta_{10} Monday_t\end{aligned} \qquad \text{Eq. 5.1}$$

5.3 Conditional Volatility Models

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is based on Bollerslev's (1986) generalization of Engle's (1982) ARCH model. This framework has led to an abundance of extensions of the model of which the RiskMetrics™ (Morgan 1996), IGARCH (Bollerslev 1986), EGARCH (Nelson 1991), APARCH (Ding, Granger et al. 1993) and GJR (Glosten, Jagannathan et al. 1993) models are used in this paper in addition to GARCH. In this section only the GARCH model, with and without determinants in the variance and mean equations, is presented as the methodology for including determinants is very similar across all models. For details regarding the ARCH framework and the extensions of Bollerslev's (1986) GARCH model applied in this paper see Appendix II.

The GARCH model has the following mean and variance equations:

$$r_t = \mu_t + \varepsilon_t \qquad \text{Eq. 5.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \text{Eq. 5.3}$$

The descriptive statistics (section 4.2.3) indicate that the error terms are not correctly specified using the normal distribution, consistent with stylized facts. Thus, the models will only be evaluated by using the student-t and skewed student-t error distribution.

To choose the appropriate autoregressive lags for the conditional volatility and the innovation term[27], information criteria for the lags ranging from 1 to 5 is evaluated for all six models without determinants (Appendix IV). This

[27] The innovation term is defined as the shock, more specific the error term in the mean equation that is included in the conditional volatility equation

procedure for determination of optimal values of p and q is similar to that of Pantula, Gonzales-Farias et al. (1994) and Liew and Brooks (1998).

The descriptive statistics (section 4.2.1) also provide evidence of the potential need for an AR(1)-term in the mean equation. Hence, this term is tested for by including it in all six models in order to identify a potential increase in explanatory power (see Appendix IV). The same methodology is used to test for an ARCH-in-mean effect. Similar procedures for determining the inclusion of AR-terms and ARCH-in-mean terms can be found in Angelidis, Benos et al. (2004) and Gogas and Serletis (2010).

The inclusion of only one lag for the conditional volatility and the innovation term, and an autoregressive term in the mean equation, provides the following GARCH specification:

$$r_t = \mu + \theta r_{t-1} + \varepsilon_t \qquad \text{Eq. 5.4}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad \text{Eq. 5.5}$$

When determinants are included, the lagged oil and S&P 500 returns are included in the mean equation to remove what the authors of this paper view as known market dependencies. Including these factors in the mean equation ultimately changes the natural gas volatility estimates due to changes in the innovation term; however, as these dependencies are expected to be known, these returns do not cause volatility.

The conditional volatility models including determinants in the mean and variance equation can be expressed as follows:

$$r_t = \mu + \theta_1 r_{t-1} + \theta_2 r_{Oil,t-1} + \theta_3 r_{SP,t-1} + \varepsilon_t \qquad \text{Eq. 5.6}$$

$$\sigma_t^2 = W + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad \text{Eq. 5.7}$$

where:

$$\begin{aligned} W = \; & \omega + \delta_1 Oil\,vol_{t-1} + \delta_2\, S\&P\,vol_{t-1} + \delta_3 Stock\,diff\,(abs)_{t-1} \\ & + \delta_4\, Prod\,diff\,(abs)_{t-1} + \delta_5 DD\,diff\,sq_{t-1} + \delta_6 Spring_t \\ & + \delta_7 Summer_t + \delta_8 Winter_t + \delta_9 Update_t + \delta_{10} Monday_t \end{aligned} \qquad \text{Eq. 5.8}$$

5.4 Evaluating Forecasting Accuracy

To validate the conditional volatility models, one-day-ahead in- and out-of-sample VaR forecasts are performed. The models are tested with a VaR level α which varies between the 25th, 10th, 5th, 2.5th and 1st percentiles. To assess the models' performance, the failure rates for the returns are calculated. The failure rate can be defined as the number of times the absolute value of the returns exceed the forecasted VaR. If the models are specified in a correct manner, the empirical failure rate should be equal to the pre-specified VaR level. While the failure rates for long positions are equal to the percentage of negative returns which are smaller than the one-day-ahead VaR, the failure rates for short positions are defined as the percentage of positive returns larger than the one-day-ahead VaR. This makes it possible to test these empirical failure rates, defined as $\hat{f}$, against the VaR level α. The tests used to assess the failure rates against the VaR level in both the in- and out-of-sample one-day-ahead model evaluation are the Dynamic Quantile Test (DQT) and the Kupiec LR test (Appendix V). For the out-of-sample model evaluation with 5, 10 and 20-day-ahead forecasts, only the Kupiec LR test is applied.

5.4.1 In-Sample Evaluation with One-Day Forecasts

When using one-day-ahead VaR forecasts to evaluate the models performance in-sample, the model parameters are estimated using the entire in-sample data. These parameters are then used to make the one-day-ahead VaR forecasts for all the in-sample data points. The evaluation of the forecasts is made in accordance with the above methodology, and is a method of back-testing the models performance.

When evaluating the models in-sample there are, in addition to VaR validation through back-testing, other criteria that can be investigated. Although it is regarded as common practice to choose the model with the highest log-likelihood value, this can be misleading as the value will be higher the more coefficients are included in the model. A way to mitigate this problem is to evaluate the information criteria (Appendix V). These criteria incorporate a trade-off between including additional variables and increasing the log-likelihood value. If the inclusion of more parameters does not increase the information in the model, the information criteria will increase. Consequently one wishes to minimize the information criteria to assess the models, and thereby choose between the models based on the information criteria with the lowest values.

Another way to compare the in-sample models is by evaluating the post-estimation tests, which consist mainly of testing the validity of the error terms. The post-estimation tests consists of the Portmanteau test on the error and squared error terms, the sign bias test, the Nyblom stability test and the adjusted Pearson goodness-of-fit test (Appendix V).

5.4.2 Out-of-Sample Evaluation with One-Day Forecasts

While in-sample evaluation, or back-testing, is performed on data used to estimate the model, out-of-sample evaluation is made on the out-of-sample observations. By validating the model out-of-sample, the models performance is assessed on observations not included in the model estimation, which makes the evaluation of the model's performance more realistic. In evaluating the out-of-sample model performance, the same failure rates for short and long positions defined previously are used, together with the DQT and Kupiec LR test.

OxMetrics™ provides a code for out-of-sample model validation with one-day-ahead VaR forecasts; however, this code uses an expanding window with re-estimation of the model every 50 days. It is believed that a fixed length rolling window with daily re-estimations is more suitable, therefore a modified code in OxMetrics was made to implement these changes. The steps in the out-of-sample one-day-ahead VaR forecasts are as follows: the respective conditional heteroskedasticity model is constructed on the in-sample data, i.e. until time t. Then the one-day-ahead VaR prediction for day t+1 is calculated. This VaR prediction is compared to the observed return for short and long positions, and the result is stored. For the next iteration the parameters in the model are re-estimated using the same amount of observations, less the first observation in the in-sample period, and including the observation for day t+1 in the out-of-sample data. This model produces new VaR predictions for day t+2 and stores them. The process is repeated until the model reaches the end of the out-of-sample period. The models' forecasting capabilities are then evaluated by assessing the failure rates in the respective percentiles.

5.4.3 Out-of-Sample Evaluation with 5, 10 and 20 Day Forecasts

To make a final assessment of the value of modeling and forecasting with and without determinants, a Monte Carlo simulation is performed. Due to the complexity in these forecasts through approximately 18 million random draws, it is critical to choose one of the conditional volatility models before performing the model evaluation. The model believed to be the best, both in- and out-of-sample, is therefore chosen before conducting these forecasts.

To assess the accuracy of the out-of-sample forecasts 5, 10 and 20 days-ahead (h = 5, 10 and 20) VaR-levels are Monte Carlo simulated. These VaR levels are compared to the actual h-day returns using the Kupiec LR test. Following Alexander (2008a), the simulated h-day VaR predictions are generated in the following manner:

1. The parameters in the conditional volatility model are estimated for the sample ending at time t
2. The return for day t+1 is simulated using a random draw of z from the predefined distribution:

$$r_{t+1} = \mu_{t+1} + z_{t+1}\sigma_{t+1} \qquad \text{Eq. 5.9}$$

3. The conditional variance for day t+2 is then calculated using the previously estimated conditional volatility model:

$$\sigma_{t+2}^2 = W + \alpha\left(z_{t+1}\sigma_{t+1}\right)^2 + \beta\sigma_{t+1}^2 \qquad \text{Eq. 5.10}$$

4. Steps 2 and 3 continue until return for day t+h has been simulated
5. The simulation procedure is carried out 10,000 times and the h-period return is calculated for each of the simulations

6. From the distribution of the 10,000 h-period returns, the required VaR levels are calculated
7. The conditional volatility model is rolled forward h days and steps 1-6 are repeated
8. Steps 1-7 are repeated until the end of the out-of-sample period is reached
9. The time series of VaR-levels is then compared to the actual h-day returns series using the Kupiec LR test

Due to the limited amount of out-of-sample observations available, not all of the percentiles used in the in- and out-of-sample evaluations are tested for. As the 20 day forecasts only allow for 37 observations in the out-of-sample period, the Kupiec LR test is only applied to the 25th and the 10th percentile. Similarly, the 10 day forecasts test the 25th, 10th and 5th percentiles, while the 5 day forecasts test the 25th, 10th, 5th and 2.5th percentiles. This yields an expected number of observations in the smallest percentile for each forecast length to be about 3-4 observations, regarded as sufficient to attain reasonable estimates for the failure rates. The procedure described above, less step 7-9, can be utilized to provide long term forecasts for returns and volatility.

The out-of-sample evaluation with 5, 10 and 20 day forecasts is also performed when determinants are included, using the following expression for W in Eq. 5.7:

$$W = \omega + ln\left(1 + \sum_{i=1}^{10} \delta_i X_i\right) \qquad \text{Eq. 5.11}$$

where X_i are the determinants used throughout this paper. This requires that the determinants, except the dummy variables, are simulated before included in the conditional volatility equation. This is done by first assuming a Gaussian error distribution for the determinants. For a detailed description on the simulation process see Appendix VII.

The oil and S&P 500 volatility estimates are calculated using vanilla GARCH models due to the presence of volatility clustering in the squared returns (Appendix III). These models are in turn used to obtain volatility forecasts. Random z-values are drawn from the Gaussian distribution, which are used to calculate the daily oil and S&P 500 return from the mean equation of the conditional volatility model. Estimates for the volatility for day h_{t+1} are obtained by using the innovation term in the volatility equation. This procedure is repeated until estimates for the oil and S&P 500 volatility and returns for an entire year are produced.

Since the stock and production difference are dependent on what has been stored and produced in recent time, an AR(1)-specification should be used.[28] First, the parameters and residuals in the regression equation are estimated,

[28] The dependencies on past information can be seen in Appendix III

and the distribution for the residuals is obtained by using the residuals mean and standard deviation from the regression equation in the Gaussian distribution. Random z-values drawn from this distribution are used to simulate the innovation terms, which are then inserted into the regression equation to simulate the determinants.

Although winter appears to exhibit larger differences in degree days (Appendix III), an i.i.d. property is assumed due to the random nature of temperature; hence the variable can be drawn from an identical distribution each day without considering past information. Sample mean and standard deviation are estimated, and randomly drawn z-values from the Gaussian distribution are included in the regression equation to find estimates for the degree day difference. The squared value are then stored for use in the natural gas volatility simulation as a representation of the squared degree day difference.

Chapter 6

Results

This section consists of two main parts. First, the results and an evaluation of the OLS regression are presented, followed by the results and evaluation of the conditional volatility models.

6.1 OLS Results

The purpose of this analysis is to establish a quantitative foundation when discussing the market factors. It will be used in the discussion of the determinants' influence on natural gas volatility in chapter 8, together with the results from the conditional volatility models. The results from this regression are presented in the table below.

Table 6: Results from the OLS regression on the in-sample period show a high level of statistical significance for most of the determinants

Proxy	**OLS** *Coefficient*	*t-value*
Constant	20.7613***	(24.2)
Oil vol (lag 1)	-0.3006***	(-2.75)
SP vol (lag 1)	0.363***	(2.6)
DD diff sq (lag 1)	0.055***	(3.33)
Abs stock diff (lag 1)	27.002***	(11.8)
Abs prod diff (lag 1)	-6.3918	(-1.29)
Spring	-14.7604***	(-16.2)
Summer	-10.4926***	(-13.3)
Winter	1.2145	(1.42)
Update	-0.0326	(-0.05)
Monday	-0.5286	(-0.73)

*The t-values are given in parentheses. *** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

6.2 OLS Model Evaluation

The correlation matrix in Table 7 can be used to assess potential multicollinearity issues between the independent variables.

Table 7: Correlation matrix (excluding dummy variables), indicating possible multicollinearity

	Oil vol	SP vol	DD diff sq	Abs stock diff	Abs prod diff
Oil vol	1	0.5981	0.0485	0.1334	0.0915
SP vol	0.5981	1	-0.0325	-0.0855	0.0682
DD diff sq	0.0485	-0.0325	1	0.0496	-0.0412
Abs stock diff	0.1334	-0.0855	0.0496	1	-0.1121
Abs prod diff	0.0915	0.0682	-0.0412	-0.1121	1

Multicollinearity is highest between the volatilities for oil and S&P 500. The severity of potential multicollinearity is evaluated by quantification through the variance inflation factor (VIF)[29], found in Table 8. This indicates that multicollinearity is not sufficiently high to require a re-estimation of the model.

Table 8: The VIF suggest that multicollinearity does not pose as a severe problem

Determinant	R-squared	VIF	Severity
Oil vol	0.42	1.71	Low
S&P vol	0.41	1.68	Low
DD diff sq	0.16	1.19	Low
Abs stock diff	0.33	1.49	Low
Abs prod diff	0.03	1.03	Low

Table 9 reveals that the independent variables are jointly significant. Furthermore, the tests for normality, homoscedasticity and functional form fail, implying that the assumptions for BLUE[30] do not hold. As a consequence, the OLS estimation is flawed; however, since this analysis is only regarded as a supplement to the determinant analysis in the conditional volatility models, further investigation is not warranted.

Table 9: Regression statistics from the OLS estimation rejecting the BLUE assumptions

Descriptive statistics and tests		
R^2	0.1566	
Adj.R^2	0.1539	
# of parameters	11	
Joint significance	58.13	[0.0000]**
Normality test	11054	[0.0000]**
Hetero test	10.25	[0.0000]**
Hetero-X test	8.28	[0.0000]**
RESET23 test	12.51	[0.0000]**

*The p-values are given in brackets. *** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

[29] VIF is calculated on the in-sample period only and should be below 4
[30] Best Linear Unbiased Estimator

6.3 Conditional Volatility Model Results

In accordance with the methodology, several initial evaluations were conducted concerning model specifications in order to reduce the amount of alternative models. These evaluations showed that the preferred lags for all models without determinants were one for both the innovation term and the lagged volatility.[31] The AR(1)-term is always significant and improves the information criteria, implying that the model is better when an AR(1)-term is included. Although the ARCH-in-mean effect is significant in all six models, the information criteria showed mixed results, hence the term was excluded. Due to the leptokurtic properties observed in the returns, a student-t error distribution is assumed, and used for all models except for the EGARCH which only converge with the skewed student-t error distribution. For further details on these preliminary results, see Appendix IV.

Table 10: The in-sample estimations without determinants display a high level of statistical significance in the model-specific coefficients across most models

	RiskMetrics	GARCH	EGARCH	GJR	APARCH	IGARCH
μ	0.0712	0.0712	0.1045*	0.0827	0.0963*	0.0739
	(1.16)	(1.31)	(1.84)	(1.52)	(1.79)	(1.37)
AR(1)	-0.0496***	-0.0476***	-0.0458**	-0.049***	-0.0455***	-0.0478***
	(-2.71)	(-2.72)	(-2.15)	(-2.79)	(-2.58)	(-2.81)
ω		0.3894***	2.7219***	0.3905***	0.0782***	0.2512***
		(4.16)	(9)	(4.07)	(2.94)	(4)
α	0.06	0.0888***	0.1383	0.1067***	0.0911***	0.1039***
		(7.79)	(0.39)	(6.13)	(9.33)	(8.99)
β	0.94	0.8907***	0.9745***	0.8901***	0.9103***	0.8961
		(71.4)	(181.7)	(69.09)	(98.28)	
γ				-0.0341*	-0.1865**	
				(-1.67)	(-2.28)	
θ_1			0.0251**			
			(2.02)			
θ_2			0.1562***			
			(3.03)			
δ					0.9402***	
					(5.71)	
v	6.3727***	5.8762***	6.0556***	5.8981***	5.9715***	5.2491***
	(11.85)	(9.08)	(9.18)	(9.1)	(9.16)	(9.97)
Skewness			0.0058			
			(0.23)			

*The t-values are given in parentheses. *** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

Table 10 shows that most coefficients are statistical significant at a 5 % level. The exceptions are the constant in the mean equation, the alpha and skewness coefficient in EGARCH, and the asymmetry coefficient in GJR.

[31] (p,q)=(1,4) also performed well, but since this did not add any additional information to the models (p,q)=(1,1) were preferred as this is easier to estimate. Additionally, the EGARCH-model only converges for a few combinations

Table 11: The in-sample estimations with determinants display a varying degree of statistical significance for the determinants

	RiskMetrics	GARCH	EGARCH	GJR	IGARCH
μ	0.0788 (1.43)	0.0849 (1.54)	0.1236** (2.15)	0.0926* (1.68)	0.0864 (1.57)
Oil ret (lag 1)	-0.0628** (-2.56)	-0.0625** (-2.53)	-0.0681*** (-2.87)	-0.0626** (-2.54)	-0.061** (-2.48)
S&P ret (lag 1)	0.0725 (1.5)	0.0668 (1.39)	0.089* (1.88)	0.0674 (1.41)	0.0696 (1.45)
AR(1)	-0.033* (-1.85)	-0.0312* (-1.68)	-0.0303 (-1.57)	-0.0323* (-1.73)	-0.0329* (-1.85)
ω		0.3368 (0.83)	2.8816*** (10.23)	0.3118 (0.76)	-0.2383 (-0.83)
Oil vol (lag 1)	-0.0199 (-1.47)	-0.0242 (-1.08)	-0.0237 (-1.37)	-0.0218 (-0.97)	-0.0197 (-1.06)
S&P vol (lag 1)	0.0383** (2)	0.0547* (1.73)	0.0266 (1.14)	0.0533* (1.7)	0.0422* (1.65)
Abs stock diff (lag 1)	0.1668 (0.98)	0.5117 (1.5)	0.8891 (1.61)	0.5009 (1.46)	0.1559 (0.62)
Abs prod diff (lag 1)	-0.2203 (-0.46)	0.1066 (0.13)	0.1802 (0.17)	-0.0511 (-0.06)	-0.0629 (-0.09)
DD diff sq (lag 1)	-0.003 (-0.79)	-0.0032 (-0.58)	-0.0019 (-1.07)	-0.0027 (-0.48)	-0.0038 (-0.79)
Spring	-0.3904*** (-3.73)	-0.9094*** (-3.31)	-0.7836*** (-4.07)	-0.8826*** (-3.17)	-0.4096*** (-2.67)
Summer	-0.1873 (-1.63)	-0.5974** (-2.39)	-0.3455** (-2.3)	-0.5512** (-2.19)	-0.2162 (-1.36)
Winter	-0.3897*** (-2.91)	-0.4965** (-2.2)	-0.041 (-0.27)	-0.4746** (-2.12)	-0.3153* (-1.7)
Update	1.0764 (1.46)	1.9955** (2.33)	0.2669*** (3.69)	2.0317** (2.38)	2.0113** (2.16)
Monday	1.4859** (2.05)	2.8703*** (3.43)	0.4565*** (5.73)	2.7179*** (3.2)	2.3287*** (2.74)
α	0.06	0.0687*** (5.42)	-0.154 (-0.59)	0.0822*** (4.29)	0.0938*** (6.5)
β	0.94	0.8823*** (37.64)	0.9715*** (132.9)	0.8827*** (35.95)	0.9062
γ				-0.0247 (-1.17)	
θ_1			0.0364** (2.54)		
θ_2			0.1706*** (3.12)		
v	4.8001*** (9.5)	6.1575*** (8.5)	6.9631*** (7.78)	6.1667*** (8.54)	4.9426*** (9.2)
Skewness			0.0192 (0.74)		

*The t-values are given in parentheses. *** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

Table 11 shows that the AR(1)-term in the mean equation exhibits weaker statistical significance compared to the models without determinants. This may be due to the inclusion of lagged returns for the oil and S&P 500 in the mean equation. The S&P 500 lagged return shows none or poor levels of statistical significance, while the lagged oil return is statistical significant in all models at the 5 % level. The remaining model-specific coefficients show a high level of statistical significance. The exceptions are alpha and skewness coefficients in EGARCH, and the asymmetry coefficient in GJR.

6.4 Conditional Volatility Model Evaluation

Evaluation of the models was done considering four different aspects. First, an evaluation of the models based on the explanatory power was performed followed by post-estimation tests to assess the validity of the error terms. Finally, in-sample and out-of-sample forecasting capabilities were evaluated.

6.4.1 Explanatory power

Based on the information criteria and the log-likelihood values for the models without determinants, presented in Table 12, the EGARCH skewed student-t model can be regarded as the best model as it has the highest log-likelihood values and lowest information criteria. APARCH is the next best performer, indicating that the asymmetric models outperform the symmetric models. In addition, the integrated models have the poorest performance, suggesting that the returns are stationary, in line with the ADF test (Table 4).

To assess whether the student-t distribution is preferred over the skewed student-t distribution, a comparison of the six models for both distributions was undertaken based on log-likelihood values and information criteria. Table 12 shows that there is very little difference between the log-likelihood values for the different distributions; however, the AIC and BIC favors the student-t distribution for all models. In addition, the skewness coefficient is always insignificant when using the skewed student-t distribution. This implies that the student-t is the preferred distribution, except for EGARCH due to convergence issues.

Table 12: Comparing the in-sample models without determinants suggest that EGARCH holds the largest explanatory power, and that the student-t error distribution is preferred over the skewed student-t distribution

	RiskMetrics	GARCH	EGARCH	GJR	APARCH	IGARCH
Log-L	-8533	-8502.7	no conv.	-8501.2	-8489.7	-8506.5
AIC	5.43347	5.41612	no conv.	5.41577	5.40908	5.41791
BIC	5.439252	5.427681	no conv.	5.429252	5.424493	5.427546
Log-L from skewed	-8533	-8502.7	-8486.3	-8501.2	-8489.6	-8506.5
AIC from skewed	5.4341	5.41676	5.40756	5.41639	5.4097	5.41854
BIC from skewed	5.44181	5.43024	5.42489	5.4318	5.42704	5.43009

When including determinants, APARCH does not converge for either of the distributions, and EGARCH still only converge for the skewed student-t distribution. Once again the EGARCH is the best performer based on log-likelihood values and information criteria. The comparison between the different distributions yield the same results as before: the student-t distribution marginally outperforms the skewed student-t distribution.

Table 13: Comparing the in-sample models with determinants suggest that EGARCH holds the largest explanatory power

	RiskMetrics	GARCH	EGARCH	GJR	IGARCH
Log-L	-8492.3	-8471.6	no conv.	-8470.7	-8486.3
AIC	5.41521	5.40392	no conv.	5.40404	5.41264
BIC	5.44411	5.4386	no conv.	5.44064	5.44539
Log-L from skewed	-8492.1	-8471.5	-8437.5	-8470.7	-8486
AIC from skewed	5.41575	5.40454	5.38417	5.40463	5.41309
BIC from skewed	5.44657	5.44114	5.42462	5.44316	5.44777

6.4.2 Post-Estimation Tests

The results from the residual based test without determinants are presented in Table 14. The Portmanteau tests[32] are performed using 15 and 50 lags, where the amounts of lags used are based on Engle (2001) and common procedure in econometrical papers. The test shows that there is no residual serial correlation in the error or squared error terms, implying that the error distribution is correctly specified. According to Serletis and Gogas (1999) insignificant Q-statistics for the error and squared error terms indicates that the models captures much of the persistence in the volatility. Furthermore, there is no evidence of volatility clustering in the error terms as the squared error terms are not serially correlated.

The sign bias test is used to evaluate if there are any asymmetric effects in the error terms. As seen in Table 14, this is not the case for most models without determinants, except for EGARCH where the asymmetric effects in the error terms are significant at the 5 % level.

For the Nyblom stability test the joint statistic is reported. These values are evaluated against asymptotic critical values at a 5 % confidence level (Appendix V). The Nyblom stability test shows that the stability of the parameters cannot be rejected, indicating that the parameters are stable.

In line with Malo and Kanto (2005) the Adjusted Pearson goodness-of-fit test is used to check if the error distribution is correctly specified. The number of cells to use in this test is chosen on the basis of König and Gaab (1982) who state that the appropriate number of cells equals the number of observations to the power 0.4, which in this case equals ~25 cells. The null hypothesis of correctly specified error terms is rejected at a 5 % confidence level in

[32] Also referred to as the Q-statistics and Box-Pierce statistics

RiskMetrics, GARCH, GJR and IGARCH. For EGARCH and APARCH the distribution of the error terms is correctly specified.

On the basis of the mentioned tests, it is concluded that the models without determinants will work satisfactory in the estimations; however, due to the adjusted Pearson goodness-of-fit test the EGARCH and APARCH are preferred. It should be noted that both show signs of asymmetry in the error terms; the former at a 5 % level of significance and the latter at a 10 % level.

Table 14: Post-estimation tests for the models without determinants favors EGARCH and APARCH

	RiskMetrics	GARCH	EGARCH	GJR	APARCH	IGARCH
Q(15)	[0.806]	[0.8]	[0.706]	[0.789]	[0.696]	[0.804]
Q(50)	[0.735]	[0.634]	[0.487]	[0.59]	[0.471]	[0.678]
Q-sq(15)	[0.319]	[0.537]	[0.1]*	[0.47]	[0.108]	[0.502]
Q-sq(50)	[0.754]	[0.494]	[0.176]	[0.358]	[0.166]	[0.504]
Sign-Bias Test	[0.298]	[0.163]	[0.036]**	[0.244]	[0.071]*	[0.368]
Nyblom test	0.476	1.412	1.754	1.397	1.212	1.238
Adj Pearson (25)	[0.032]**	[0.03]**	[0.538]	[0.046]**	[0.474]	[0.021]**

**** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

Evaluating the validity of the error terms for the model with determinants in Table 15, all models are correctly specified according to the Portmanteau tests, which exclude volatility clustering in the error terms. In addition, the sign bias test shows no indication of asymmetric effects in the error terms at 5 % significance, which is an improvement for the EGARCH model when the determinants were not included. The Nyblom stability test statistics indicate that the coefficients for all models are stable through time. In contrast to the models without determinants, the adjusted Pearson goodness-of-fit test shows that the error distribution is correctly specified for all models, which is better than for the models without determinants. This indicates that the addition of determinants to the conditional variance equations acts as an improvement in terms of validity in the error terms.

Table 15: Post-estimation tests for models with determinants indicate that all models are correctly specified

	RiskMetrics	GARCH	EGARCH	GJR	IGARCH
Q(15)	[0.734]	[0.652]	[0.689]	[0.634]	[0.742]
Q(50)	[0.701]	[0.519]	[0.483]	[0.488]	[0.692]
Q-sq(15)	[0.134]	[0.223]	[0.181]	[0.189]	[0.29]
Q-sq(50)	[0.829]	[0.4]	[0.101]	[0.318]	[0.632]
Sign-Bias Test	[0.078]*	[0.072]*	[0.186]	[0.09]*	[0.328]
Nyblom test	3.385	3.904	4.381	3.944	3.327
Adj Pearson (25)	[0.624]	[0.454]	[0.281]	[0.176]	[0.122]

**** indicates a significance level of 1 %, ** indicates a significance level of 5 % and * indicates a significance level of 10 %*

6.4.3 In-Sample Evaluation with One-Day Forecasts

In terms of the models' forecasting capabilities, the statistics from the DQT and the Kupiec LR test are assessed for all six models with and without determinants for in-sample VaR forecasting one-day-ahead. If the models are able to describe the data in a satisfactory manner, the p-values should be larger than 0.05 and the null hypothesis that the empirical failure rate $\hat{f}$ is equal to the VaR level α is not rejected. The entire test results are reported in Appendix VI.

Table 16 shows that all the models without determinants perform satisfactory in the one-day-ahead VaR forecasts; however, IGARCH and RiskMetrics underperform the other four models. The former fails in the 1st percentile for long positions in both tests, whereas the latter fail in one long position, and in several short positions.

Table 16: DQT and Kupiec LR test for the models in-sample without determinants disfavor the integrated models

	RiskMetrics	GARCH	EGARCH	GJR	APARCH	IGARCH
Kupiec LR	3/10	0/10	0/10	0/10	0/10	1/10
DQT	2/10	0/10	0/10	0/10	0/10	1/10
Total	**5/20**	**0/20**	**0/20**	**0/20**	**0/20**	**2/20**

The in-sample forecasting results with determinants are similar to those without determinants. The table below shows that EGARCH, GJR and GARCH are very good in one-day-ahead VaR forecasts, whereas the GARCH model only fails in the 5th percentile for long positions in the DQT. RiskMetrics and IGARCH fail several times, mostly for the long positions. Summarized, the inclusion of determinants makes RiskMetrics slightly better, and IGARCH and GARCH slightly worse. The remaining models are unaffected.

Table 17: DQT and Kupiec LR test for the models in-sample with determinants disfavors the integrated models

	RiskMetrics	GARCH	EGARCH	GJR	IGARCH
Kupiec LR	3/10	0/10	0/10	0/10	3/10
DQT	1/10	1/10	0/10	0/10	2/10
Total	**4/20**	**1/20**	**0/20**	**0/20**	**5/20**

6.4.4 Out-of-Sample Evaluation with One-Day Forecasts

In this section the models are evaluated based on their out-of-sample VaR forecasting capabilities (for detailed test summaries see Appendix VI). When determinants are not included EGARCH only fail for two of the test results reported, making it the best performing model. The model is good in describing all quantiles in the Dynamic Quantile Test, but in the Kupiec LR test it fails in the long positions for 25th and 2.5th percentile. This model is closely followed by the RiskMetrics student-t model, which fails in the long positions in the 25th percentile for both tests, and the long position in the 10th percentile in the DQT test. The remaining models fail mostly for long positions.

Table 18: DQT and Kupiec LR test for the models out-of-sample without determinants favors EGARCH and RiskMetrics

	RiskMetrics	GARCH	EGARCH	GJR	APARCH	IGARCH
Kupiec LR	1/10	4/10	2/10	4/10	3/10	5/10
DQT	2/10	2/10	0/10	3/10	2/10	3/10
Total	**3/20**	**6/20**	**2/20**	**7/20**	**5/20**	**8/20**

In the out-of-sample forecasts with determinants EGARCH fail for long positions in the 25th and 10th percentile for the Kupiec LR test and DQT, respectively. GARCH and GJR fail for long positions in the 25th, 5th and 2.5th percentiles for the Kupiec LR test[33], and the integrated models fail in several of the percentiles, long positions in particular.

Table 19: DQT and Kupiec LR test for the models out-of-sample with determinants disfavor the integrated models

	RiskMetrics	GARCH	EGARCH	GJR	IGARCH
Kupiec LR	4/10	3/10	1/10	3/10	6/10
DQT	1/10	0/10	1/10	0/10	5/10
Total	**5/20**	**3/20**	**2/20**	**3/20**	**11/20**

6.4.5 Out-of-Sample Evaluation with 5, 10 and 20 Day Forecasts

The previous sections indicate that the EGARCH model is the best performing model based on explanatory power, post-estimation tests and in- and out-of- sample forecasting both with and without determinants. As a consequence, only the EGARCH model will be used to evaluate the forecasting capabilities for 5, 10 and 20 day forecasts.

Table 20, 21 and 22 show the results from the rolling 5, 10 and 20 day forecasts performed using a Monte Carlo simulation. The Kupiec LR tests show that none of the forecasted VaR-levels have failure rates significantly different from the theoretical failure rates. This implies that EGARCH produces correct forecasts of the simulated VaR-levels for all quantiles and forecasting horizons.

The accuracy of these results is questioned as the out-of-sample sample size only allows for 150, 75 and 37 observations for 5, 10 and 20 day-ahead forecasts, respectively. This implies that the allowable deviation between the theoretical and the empirical failure rate is quite large. The empirical failure rate for the 10th percentile for the 20-day forecasts could be as high as 27 % before rejecting the null of correctly forecasted VaR-level. As a result, inspection of the simulated VaR-levels has been carried out, which shows that

[33] APARCH does not converge when including determinants in the out-of-sample VaR forecasts

the short positions' VaR-levels are overestimated for the 25th and 10th percentile, and for most forecast lengths, implying that the actual return does not exceed the simulated VaR-levels enough times. The trend for the long position simulations is opposite, where the simulated VaR-levels are underestimated, except for the 5th and 2.5th percentile. In addition, the underestimation of the long positions' central quantiles, and the overestimation of the short positions' central quantiles, suggests that the models VaR estimates are skewed compared to the actual VaR-levels.

When comparing the theoretical and empirical failure rates, it is evident that there are mixed results. When including determinants the forecasting performance is better in long positions for the 5-day-ahead forecast and in two percentiles in the 10-day-ahead forecast. When determinants are not included the forecasting performance is better in short positions for the 5-day-ahead forecast, and for long positions in the 20-day-ahead forecast. Consequently, comparing the failure rates proves inconclusive, which is a common problem in econometrics as one GARCH model rarely performs well on all criteria evaluated. Still, one can conclude that EGARCH performs satisfactory with and without determinants for 5, 10 and 20 day-ahead forecasts.

Table 20: The Kupiec LR test for 5 day forecasted VaR-levels show that EGARCH performs best with determinants (best performance in bold)

			5 day forecast					
			Without determinants			***With determinants***		
			Failure rate	Kupiec LR	p-value	Failure rate	Kupiec LR	p-value
VaR-level	Long	2.5 %	1.3 %	0.437	0.51	**2.0 %**	**0.072**	**0.79**
		5 %	3.3 %	0.429	0.51	**5.3 %**	**0.015**	**0.90**
		10 %	12.0 %	0.274	0.60	12.0 %	0.274	0.60
		25 %	32.7 %	1.924	0.17	**32.0 %**	**1.611**	**0.20**
	Short	25 %	**20.0 %**	**0.912**	**0.34**	18.0 %	1.829	0.18
		10 %	**8.0 %**	**0.309**	**0.58**	7.3 %	0.562	0.45
		5 %	4.0 %	0.147	0.70	4.0 %	0.147	0.70
		2.5 %	2.7 %	0.007	0.93	2.7 %	0.007	0.93

Table 21: Kupiec LR test for 10 day forecasted VaR-levels show that EGARCH performs best with determinants (best performance in bold)

			10 day forecast					
			Without determinants			***With determinants***		
			Failure rate	Kupiec LR	p-value	Failure rate	Kupiec LR	p-value
VaR-level	Long	5 %	5.3 %	0.007	0.93	5.3 %	0.007	0.93
		10 %	10.7 %	0.016	0.90	9.3 %	0.016	0.90
		25 %	24.0 %	0.018	0.89	**25.3 %**	**0.002**	**0.97**
	Short	25 %	24.0 %	0.018	0.89	24.0 %	0.018	0.89
		10 %	10.7 %	0.016	0.90	10.7 %	0.016	0.90
		5 %	10.7 %	1.686	0.19	**8.0 %**	**0.526**	**0.47**

Table 22: Kupiec LR test for 20 day forecasted VaR-levels show that EGARCH performs best without determinants (best performance in bold)

			20 day forecast					
			Without determinants			***With determinants***		
			Failure rate	Kupiec LR	p-value	Failure rate	Kupiec LR	p-value
VaR-level	Long	10 %	**10.8 %**	**0.011**	**0.91**	13.5 %	0.201	0.65
		25 %	**27.0 %**	**0.035**	**0.85**	32.4 %	0.447	0.50
	Short	25 %	16.2 %	0.726	0.39	16.2 %	0.726	0.39
		10 %	5.4 %	0.445	0.50	5.4 %	0.445	0.50

Chapter 7

Model Discussion

7.1 Model Specification

Based on descriptive statistics (section 4.2), a GARCH specification to describe the natural gas volatility appears appropriate due to the inherent nature of the returns and squared returns. This is also evident from the post-estimation tests, where the results from the Portmanteau and Adjusted Pearson goodness-of-fit tests imply that the distributions for the error terms are correctly specified, which in turn suggests that the models captures much of the persistence in the volatility. As the AR(1)-term show statistical significant results in most models, it is also concluded that the inclusion of this term improves the models due to the models' increased explanatory power.

The results reveal that an asymmetric response is present in the natural gas returns, implying that positive and negative returns have different impact on the natural gas volatility. For the models without determinants, APARCH, GJR and EGARCH show statistical significant asymmetry coefficients. The signs of the asymmetry coefficients, together with the alpha coefficients in the EGARCH and APARCH models, imply that positive returns is source for a higher volatility. This result is anticipated, as natural gas markets usually exhibit an inverse leverage effect due to the convexity of marginal costs in natural gas production (Bermejo-Aparicio, Moreno et al. 2007). For the models with determinants, there are only two asymmetrical models to compare. These display mixed results for the asymmetry effect, implying that some of this effect is captured by the determinants.

When constructing conditional volatility models on commodity markets, the Nyblom test for parameter stability often fails. Among others, Naryan and Naryan (2007) encounter inconsistency in evidences of asymmetry and persistency when examining sub-samples in their data. When the Nyblom test fails, the coefficients are unstable and the model may be miss-specified especially in regards to sample size. This can have significant effects on the

VaR forecasting accuracy, which will be especially apparent on the out-of-sample VaR forecasts. As the Nyblom stability test show that the parameters are stable in the in-sample models with and without determinants, it is concluded that the sample size is not problematic, although quite large. This is evident in the out-of-sample forecasts, which performs quite well. An improvement could be to consider the individual statistics for the Nyblom stability tests.

7.2 Model Selection

When considering the models without determinants, EGARCH is the best performer based on explanatory power, closely followed by APARCH. These two models are also the only models that pass the Adjusted Pearson goodness-of-fit test. However, both show signs of asymmetry in the error terms, whereas the former at a higher level of significance than the latter, which is not present in the other models. In terms of in-sample forecasting capabilities, all models perform excellent except for RiskMetrics and IGARCH. For the out-of-sample VaR forecasts, EGARCH is best whereas RiskMetrics is a close second.

The models show similar results in terms of significance of the coefficients when determinants are included. The explanatory power is highest for EGARCH, followed by GJR and GARCH, and the residual based tests indicate that all models are correctly specified. EGARCH and GJR are also the best models for in-sample VaR forecasts, closely followed by GARCH. In the out-of-sample VaR forecasts EGARCH is again the best model, followed by GARCH and GJR.

As the EGARCH model does not impose any restrictions on the coefficients, and because it describes the asymmetry effect, it has been preferred by several papers studying energy markets (Narayan and Narayan 2007; Alizadeh and Talley 2009). Consistent with this paper's results, Serletis, Apostolos and Gogas (1999) find that EGARCH is superior to GARCH in terms of log-likelihood when evaluating three time series on North American natural gas volatility. Similarly, Gogas and Serletis (2010) conclude that an EGARCH specification with GED distributed error terms is the model best suited for forecasting in daily commodities futures markets. In volatility forecasting for crude oil futures, Marzo and Zagaglia (2010) conclude that EGARCH is the best model for this market, and is preferred on forecasts, with both a student-t and a GED error distribution. Yaffe, Heddy et al. (2008) on the other hand find APARCH and RiskMetrics to outperform EGARCH, but since the paper investigate the UK natural gas market and use different evaluation criteria, the conflicting results are not considered further.

A drawback with EGARCH when used in this paper is that the alpha coefficient is not significant for the in-sample estimations. In addition, the use of a skewed student-t error distribution is not favorable as it includes an unnecessary skewness parameter, which is insignificant. Also, signs of asymmetric effects in the error terms in the EGARCH model without determinants suggest that the model does not capture all the asymmetric response on volatility. Still, it is concluded that the EGARCH model is the overall best model to describe the natural gas price-volatility in the US natural gas market due to its superior properties in terms of explanatory power, validity in error terms and forecasting capabilities, in addition to support from previous empirical research.

7.3 The Value of Including Determinants

To conclude whether or not the inclusion of determinants adds explanatory value to the models, the models with and without determinants are compared using both log-likelihood values and information criteria. When determinants are included the log-likelihood values are significantly higher in all models[34]; however, as the inclusion of more parameters usually increases the log-likelihood value, the information criteria should also be investigated. While the AIC are always better with determinants, this is not the case for the BIC. All models except EGARCH show that the BIC is worse when the determinants are included. Due to conflicting results in the information criteria, inclusion of determinants based on explanatory power is inconclusive. An advantage with including determinants is that it allows for investigation of how different factors affect the natural gas volatility, without a reduction in the models' explanatory power. However, a drawback with the inclusion is the increased complexity of model estimations, and potential convergence issues as illustrated by the APARCH model.

When the determinants were included in the in-sample models, the validity of the error terms was improved. This is apparent from the adjusted Pearson goodness-of-fit test, where the inclusion of determinants made the distribution in the error terms valid for all models. Additionally, the errors' asymmetric effect in the EGARCH model was not present when the determinants were included.

In the in-sample VaR forecasts, the inclusion of determinants makes RiskMetrics slightly better, and IGARCH and GARCH worse. The remaining models however, are unaffected. In the out-of-sample VaR forecasts RiskMetrics and IGARCH perform worse when the determinants are included, GARCH and GJR perform better, and the EGARCH model is unaffected. This implies that the integrated models do not handle the inclusion of determinants in a satisfactory manner.

The long term forecasts showed that the EGARCH models with and without determinants performs satisfactory as they pass the Kupiec LR test for all percentiles and for all forecasting horizons. However, as noted in section 6.4.5,

[34] Using $2(L_1 - L_0) \sim \chi^2(m)$ where m is the number of added determinants

a visual inspection was necessary due to the small amount of observations, which proved to be non-conclusive. This implies that including determinants in long term forecasts is not preferred due to the added model complexity.

A common critique of conditional volatility models is that they have a tendency to overestimate the persistency (Lamoureux and Lastrapes 1990). Mu (2007) finds that the observed volatility persistence is reduced when including determinants into the conditional variance equation, with a reduction in half-life by 43 %. This is also evident in our models as seen in Table 23. The half-life, which is only dependent on the persistency parameter, is reduced by ~59 % when determinants are included. This implies that the inclusion of determinants has a significant effect on the speed of which the volatility converges to its mean-reverting level. The unconditional variance, or long term volatility, is dependent on the expectation of the determinants when included, in addition to the persistency parameter.[35] This inclusion reduces the unconditional variance by ~14 %, implying that the determinants reduce the estimated mean reverting level for natural gas (for further details see Appendix VIII).

As stated in Mu (2007) 'recent literature on volatility persistence suggests that the persistence in the conditional variance may be generated by an exogenous driving variable that is itself serially correlated.' This implies that since the persistence is partly generated by the determinants, the persistence parameter is reduced to what is believed to be a more accurate measure of the persistency parameter when determinants are included. Taking this into account, and that one of the purposes of modeling is to remove some of the persistency, it is believed that the inclusion of determinants improves the specification of the models.

Table 23: The persistence, unconditional volatility and half-life are lower when determinants are included

	GARCH	GJR
Without determinants		
Persistence (φ)	0.9794	0.9798
Half-life (days)	33.36	34.01
Unconditional variance	18.94	19.36
Annualized standard deviation	68.81	69.57
With determinants		
Persistence (φ)	0.9510	0.9525
Half-life (days)	13.8	14.25
Unconditional variance	16.30	16.47
Annualized standard deviation	63.84	64.16

The above considerations show that the answer to whether or not the inclusion of determinants adds value to the models depends on the purpose of the models. Looking at Figure 13, the inclusion of determinants causes more

[35] For the volatility of the oil and S&P 500, the long term volatility from the vanilla GARCH models is used as the expected values

noise in the conditional volatility estimates, which is believed to be triggered mainly by the daily effects as these proved to have the largest effects on the natural gas volatility. However, this is not believed to be problematic, and may be a more correct specification of the volatility. In regards to the validity of the error terms, adding determinants improves some of the models. Neither the explanatory power of the models, nor the forecasting capabilities, is significantly affected by the inclusion of determinants. It is therefore suggested that determinants should not be included for forecasting purposes as this makes the model estimations more complex, and do not significantly improve the forecasting capabilities. When determinants are included these can provide additional information of the effects of external factors on natural gas volatility and makes it possible to run scenario analyses, which should be of value to market participants. In addition, it is believed that the models with determinants give a more accurate measure of the persistency parameter.

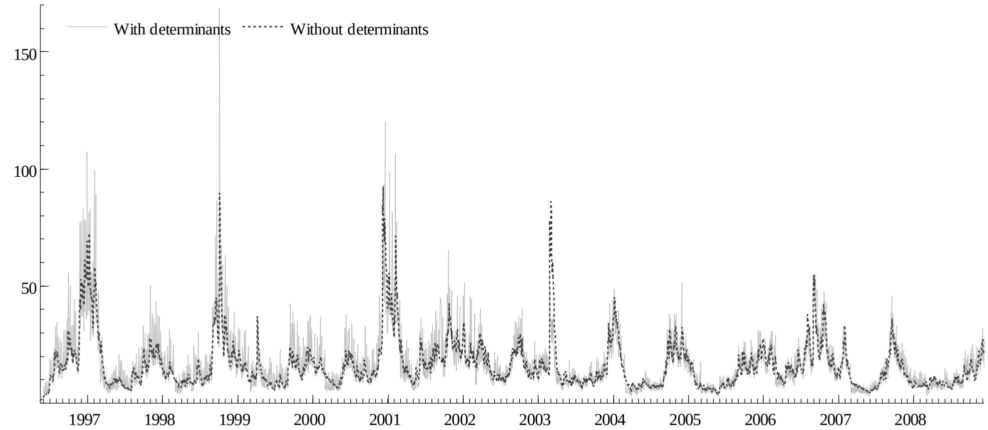

Figure 13: The EGARCH model with determinants display more noise in the conditional volatility estimates than the EGARCH model without determinants

Chapter 8

Determinants Discussion

Both the OLS analysis and the conditional volatility models are used in combination with empirical evidence to describe the determinants' effect on the natural gas volatility. There are several factors where the models disagree on significance. Emphasis is put on the conditional volatility models as they give a more correct picture of the actual relationship due to the non-linear nature of natural gas volatility; however, we believe the OLS analysis may give additional insights.

Table 24: Comparison between the hypothesized and actual effect of the determinants imply mixed results depending on the models used[36]

Determinant	Proxy	Hypotheses	OLS		GARCH	
		Sign	*Sign*	*Significance*	*Sign*	*Significance*
Substitutes	Oil vol	+	-	Yes	-	No
Temperature	DD diff sq	+	+	Yes	-	No
Economic activity	S&P vol	+	+	Yes	+	Yes (4)
Storage	Abs stock diff	+	+	Yes	+	No
Storage / news	Update	+	-	No	+	Yes (4)
Production	Abs prod diff	+	-	No	+/-	No
Seasonality	Spring	-	-	Yes	-	Yes (5)
Seasonality	Summer	-	-	Yes	-	Yes (3)
Seasonality	Winter	+	+	No	-	Yes (4)
Monday effect	Monday	+	-	No	+	Yes (5)

8.1 Substitutes

Oil volatility is only significant for the OLS analysis, not for any of the six conditional volatility models. This result is similar to Mu (2007) who also includes the volatility of oil into the conditional volatility equation of natural gas. In addition, the sign of the oil volatility is negative in all models, contrary to our original hypothesis.

[36] Significance at 10 % level in both OLS and conditional volatility models. The number in parentheses shows how many of the conditional volatility models that gave significant results for the respective determinant.

An elaborate discussion of why the relationship is different than expected could be attempted, but since the results contradicts former empirical evidence (Ewing, Malik et al. 2002; Pindyck 2004) it seems more reasonable to concur with Villar and Joutz (2006) in that the OLS approach is 'faulty and subject to spurious results'. Division of the in-sample data into two samples yields different signs for the oil volatility in the OLS, further undermining the result. This could indicate that Ramberg and Parsons' (2010) argument of the existence of a "shifting relationship" between oil and gas is true, but our results and methods are not sufficient to draw conclusions on that matter.

The lagged oil return in the mean equation is negative and significant. The autocorrelation figure in section 4.2.1 displays a negative correlation between natural gas return and natural gas returns of lag 1. Therefore, the negative coefficient for lagged oil return appears to replace the AR(1)-term in the mean equation, which is supported by the AR(1)-term's insignificant coefficient. As a result, the correlation between lagged oil return and lagged natural gas return is positive, consistent with Mu (2007) and in line with empirical research on the relationship between oil and gas prices (Villar and Joutz 2006; Brown and Yücel 2008; Ramberg and Parsons 2010), indicating a relationship between these two markets. To explain and describe the relationship between natural gas and oil, if any, a co-integration or multivariate conditional volatility analysis might be better suited.

8.2 Temperature

The proxy for temperature is only significant in the OLS analysis, where it is in line with our hypothesis. This contradicts Mastrangelo's (2007) OLS analysis on volatility, where a ratio for heating degree days proves insignificant. Since the temperature proxy is insignificant in all the conditional volatility models, our hypothesis is not accepted. However, it should be noted that Mu (2007) gets significant results when weather shocks are included in the conditional variance equation of a vanilla GARCH model. A potential reason for the difference in results compared to earlier studies might be attributed to the use of a different proxy to measure unexpected changes in temperature, or influence from other determinants on the weather proxy.

8.3 Economic Activity

In accordance with our original hypothesis, the results from both the OLS regression and most of the conditional volatility models indicate a positive dependence between the volatility from the S&P 500 and the natural gas volatility. This is expected, as the S&P 500 Index represents the systematic risk in the market, and changes in this will spill over on other markets. Thus, economic activity influences the dynamics of the natural gas market. This implies that including the volatility of the S&P 500 in the conditional volatility equation, rather than in the mean equation, is a correct specification. To our knowledge, there are no papers that use this as a proxy for economic activity in energy commodity markets; as a result our findings are not comparable with other empirical evidence.

Similar to Mu (2007), the results for the S&P 500 returns included in the mean equation proves to be statistical insignificant.

8.4 Storage

When investigating storage effects, two proxies were employed; the difference in stock from average and the release of the weekly update storage report. The stock difference has a positive sign in all models, in line with our hypothesis, but is only significant in the OLS. Mastrangelo (2007) also investigate the storage level's effect on natural gas volatility in an OLS framework, and find that their proxy is positive and statistically significant. This result however, cannot be fully compared since they employ weekly storage levels.

With the weekly update storage report, the market learn about the status of the storage levels, and their reaction to these news are expected to increase volatility as the trade of futures reacts either to an increase or a decrease in storage levels. The publishing of the storage report is statistical significant in four out of five conditional volatility models, supporting earlier empirical evidence (Murry and Zhu 2004; Mu 2007), while in the OLS it is statistical insignificant with a negative coefficient. It should be noted that the coefficient for the weekly update storage report has one of the largest coefficients in the conditional volatility models.

Imperfect market information may explain why only one of the storage proxies are statistical significant. If this market does not convey perfect information, the market participants do not have information of what the current stock levels are; hence they may not react on deviations from the historical stock levels. Contrary, as they acquire this information when the weekly update storage report is released, this is a better proxy of the impact of storage on natural gas volatility. Since the storage difference proxy is insignificant in the conditional volatility models, it could thus imply that the market does only consider weekly storage levels when trading.

8.5 Production

The production in the short term is mainly affected by seasonality and extreme weather conditions, whereas the former has been eliminated from the proxy for production. This proxy is not significant in any of the models, which implies that disruptions in production do not affect volatility. This contradicts Alterman (2012) who identifies periods of high volatility due to hurricanes. However, it should be noted that Alterman (2012) only performs a visual inspection of volatility graphs to establish the relationship between volatility and hurricanes, and does not seek to identify if this relationship is causal. As severe weather, and especially hurricanes, can be forecasted, market participants may expect a change in production before the change occurs. Consequently, this proxy may lag the actual change in expectation. An improvement could be to use the same proxy a week ahead to describe the expected production, or to simply use a dummy for severe weather. Another explanation can also be found in Alterman (2012) who suggests that the impact of disruption in production have had reduced effect on volatility since 2005, which implies that production may not have a significant impact on volatility since the proxy only considers changes due to disruptions.

8.6 Seasonality and the Monday Effect

Consistent with empirical evidence (Serletis and Shahmoradi 2006; Mastrangelo 2007; Suenaga, Smith et al. 2008), both the OLS and all conditional volatility models show statistical significant seasonality effects. All models produce significant results for spring, and both summer and winter show significance across most of the models. The coefficients show that spring has the lowest volatility, and that all seasons have lower volatility than fall. This supports the hypotheses that volatility is lowest during spring and summer; however, it contradicts our hypothesis that winter has the highest volatility.

The results from the conditional volatility models are not supported by the OLS as no significant difference between fall and winter is exhibited. In a similar analysis, Serletis and Shahmoradi (2006) use months as proxies for seasonality, finding December and January to hold most volatility. However, Mastrangelo (2007) find October and November to have significantly higher volatility than January when using an OLS framework. This suggests that the volatility in these seasons may vary according to the model and sample used.

Fall is dominated by inflexibility in storage facilities as market participants attempt to buy gas and fill their storage targets for the upcoming winter. In addition, the storage levels may be low in the beginning of fall due to unusually warm summer seasons. Strain on storage levels can also be caused by an early winter season which may heighten demand even further. This increased demand in combination with the lingering hurricane season makes the fall highly uncertain, which may be why this is the most volatile season.

Across all the conditional volatility models, the coefficient for the Monday effect is positive, significant and has the highest value of all the determinants, consistent with our hypothesis and existing literature (Fleming, Kirby et al. 2006; Serletis and Shahmoradi 2006; Mu 2007). This implies that information generated during the weekend is reflected in higher price changes in the gas market on Mondays.

Chapter 9

Conclusion and Further Research

9.1 Summary and Concluding Remarks

Since its liberalization, the US natural gas market has undergone large changes where volatility has played, and still plays, a significant role both in terms of future investments and the need to hedge against daily risk. Considering that the natural gas market is the second most volatile energy market in the US, and the second largest futures market in the world, there is limited research on volatility in this market substantiating the need for further analysis.

The purpose of this paper is to find the model that best describes the price-volatility in the US natural gas market, to establish causes of volatility and to forecast in the short term. Six different conditional volatility models were applied on natural gas returns based on Henry Hub futures. Factors affecting the price-volatility were investigated by including determinants in the conditional volatility equation. First, the conditional volatility models were estimated in-sample and compared on the basis of their explanatory power and the post-estimation tests. Secondly, in and out-of-sample forecasting was performed using the DQT and Kupiec LR test to evaluate the models forecasting accuracy. The out-of-sample forecasts were performed using a rolling window with daily re-estimations, which is believed to be more accurate than an expanding window with intermittent re-estimations. Finally, simulation of 5, 10 and 20-day-ahead VaR levels was carried out for the best performing model, both with and without simulated determinants. This allows for testing of the forecasting capabilities using the Kupiec LR test. To the best of our knowledge such an extensive forecasting evaluation, and conditional volatility forecasts with determinants, have not been documented in this market.

Based on explanatory power, validity of error terms and forecasting capabilities in- and out-of-sample, it is concluded that the EGARCH[37] model is superior to GARCH, GJR, IGARCH, RiskMetrics and APARCH[38] to describe the price-volatility in the US natural gas market, independent of the inclusion of determinants. In the 5, 10 and 20-day-ahead forecasts with simulated VaR levels EGARCH does not fail for any percentiles in either long or short positions. Of the remaining models, APARCH is the overall second best model; however, it does not converge when determinants are included. GJR and GARCH vary in performance depending on the criteria considered, and IGARCH and RiskMetrics are the weakest models, independent of the inclusion of determinants. However, in the out-of-sample forecasting without determinants RiskMetrics is the best performer after EGARCH. Market participants, who usually apply RiskMetrics for hedging purposes, are therefore advised to use EGARCH, as this outperforms the other models in terms of forecasting and modeling.

The proxies for economic activity, seasonality and daily effects were the determinants that proved to be statistical significant in the conditional volatility models. Economic activity has a positive effect, implying that the natural gas volatility is positively dependent on systematic risk. Fall exhibits the highest volatility, followed by winter, then summer and spring. Monday and the weekly update storage report increase the natural gas price-volatility, and are shown to have the largest influence on volatility.

Two main conclusions can be drawn when comparing the results with and without determinants; if the aim of the model is short term forecasting, the determinants should be excluded as they do not improve the forecasting accuracy. Conversely, if the aim is to explain the causes of volatility, the in-sample evaluation indicates that the inclusion of determinants is a reasonable approach. This type of determinant analyses is also a good foundation for scenario analyses, and should be of value to market participants.

There are four main aspects that are believed to make this paper an improvement to the existing body of literature. Firstly, it contains updated analyses of volatility in the US natural gas market. Secondly, the inclusion of a larger number of determinants compared to existing literature is believed to strengthen the understanding of the market dynamics. Thirdly, several models are assessed in order to find the best suited model for volatility in the US natural gas market. Lastly, research on modeling and forecasting issues in energy commodity markets is in large limited to oil-related commodities; thus, this paper is believed to fill a gap in the literature.

[37] AR(1)-EGARCH(1,1) with a skewed error distribution

[38] GARCH, GJR, IGARCH, RiskMetrics and APARCH with the student-t error distribution and AR(1)-term in the mean equation

9.2 Further Research

For further research, conditional VaR[39] could be used in addition to VaR for a more robust forecasting evaluation. To increase the accuracy of the forecasts, CAViaR model (Engle and Manganelli 1999) could be applied to directly model the tails of the distribution. The application of other members of the GARCH family might improve the analyses, such as: Fractionally Integrated GARCH to account for long memory effects[40], Switching GARCH to model different regimes, or multivariate GARCH to investigate the relation between different US natural gas markets, or between the US and other regional natural gas markets. Implied volatility is often argued to be a better measure of volatility; therefore, it could be an improvement to include this in the analyses.

The determinants chosen in this paper are important fundamentals of the gas market; however, exploration of different proxies for the determinants, or inclusion of other determinants, may improve the model performance. Based on the knowledge of how the fundamental drivers affect the natural gas volatility, long term forecasts through simulation, or scenario analyses, can be performed. In addition, our methodology could be used to study different markets or future contracts with varying time to maturity.

[39] Also called the expected shortfall
[40] This can be evaluated with the long memory tests by Geweke and Porter-Hudak (1983) or Robinson and Henry (1998)

Bibliography

Abascal, F., R. Zardoya, et al. (2005). "ProtTest: selection of best-fit models of protein evolution." Bioinformatics **21**(9): 2104-2105.

Akaike, H. (1974). "A new look at the statistical model identification." IEEE Transactions on Automatic Control **19**(6): 716-723.

Alexander, C. (2008a). Market Risk Analysis II: Practical Financial Econometrics. Chichester, West Sussex, J. Wiley & Sons

Alexander, C. (2008b). Market Risk Analysis IV: Value-at-Risk Models. Chichester, West Sussex, J. Wiley & Sons

Alizadeh, A. H. and W. K. Talley (2009). "Dynamics of the Forward Curve and Volatility of Energy Futures Prices." SSRN eLibrary: 1-34.

Aloui, C. and S. Mabrouk (2010). "Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models." Energy Policy **38**(5): 2326-2339.

Alterman, S. (2012). "Natural Gas Price Volatility in the UK and North America." unpublished work: 1-46.

Angelidis, T., A. Benos, et al. (2004). "The use of GARCH models in VaR estimation." Statistical Methodology **1**(1–2): 105-128.

Ates, A. and H. K. Wang (2007). "Price Dynamics in Energy Spot and Futures Markets: The Role of Inventory and Weather." unpublished work: 1-42.

Bachmeier, L. J. and J. M. Griffin (2006). "Testing for Market Integration: Crude Oil, Coal, and Natural Gas

" The Energy Journal **27**(2): 55-72.

Bera, A. K. and M. L. Higgins (1993). "ARCH MODELS: PROPERTIES, ESTIMATION AND TESTING." Journal of Economic Surveys **7**(4): 305-366.

Bermejo-Aparicio, R., M. Moreno, et al. (2007). "European Natural Gas Spot Markets: Volatility Transmission and Jumps Modeling." unpublished work: 1-61.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity." Journal of Econometrics **31**(3): 307-327.

Box, G. E. P. and D. A. Pierce (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." Journal of the American Statistical Association **65**(332): 1509-1526.

Brooks, C. and G. Persand (2002). "Model choice and value-at-risk performance." Financial Analysts Journal **58**(5): 87-99.

Brown, S. P. and A. Krupnick (2010). "Abundant Shale Gas Resources: Long-Term Implications for U.S. Natural Gas Markets." SSRN eLibrary: 1-39.

Brown, S. P. A. and M. K. Yücel (2008). "What Drives Natural Gas Prices?" The Energy Journal **29**(2): 45-60.

CMEGroup (2012a). "CME Group Announces Trading Volume Record in Henry Hub Natural Gas Futures." Retrieved 27.04, 2012, from http://investor.cmegroup.com/investor-relations/releasedetail.cfm?ReleaseID=642001.

CMEGroup (2012b). "Henry Hub Natural Gas Futures." Retrieved 27.04, 2012, from http://www.cmegroup.com/trading/energy/natural-gas/natural-gas_contract_specifications.html.

Deaton, A. and G. Laroque (1992). "On the Behaviour of Commodity Prices." The Review of Economic Studies **59**(1): 1-23.

Deaton, A. and G. Laroque (1996). "Competitive Storage and Commodity Price Dynamics." Journal of Political Economy **104**(5): 896-923.

Dickey, D. A. and W. A. Fuller (1981). "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." Econometrica **49**(4): 1057-1072.

Ding, Z., C. W. J. Granger, et al. (1993). "A long memory property of stock market returns and a new model." Journal of Empirical Finance **1**(1): 83-106.

Duffie, D. and S. F. Gray (1995). Volatility in energy prices. Managing Energy Price Risk. R. Jameson. London, Risk Publications.

EIA (1995). "The value of underground storage in today's natural gas industry." Retrieved 08.03, 2012, from ftp://ftp.eia.doe.gov/natgas/059195.pdf.

EIA (2002). "Manufacturing Fuel - Switching Capability." Retrieved 05.03, 2012, from http://www.eia.gov/emeu/mecs/special_topics/energy_use_manufacturing/energyuse98_02/fuel_switch.html.

EIA (2011a). "Annual Energy Review 2010." Retrieved 02.02, 2012, from http://205.254.135.7/totalenergy/data/annual/pdf/aer.pdf.

EIA (2011b). "Natural Gas Annual 2010." Retrieved 05.04, 2012, from http://205.254.135.7/naturalgas/annual/pdf/nga10.pdf.

EIA (2011c). "Factors Affecting Natural Gas Prices." Retrieved 18.04, 2012, from http://www.eia.gov/energyexplained/index.cfm?page=natural_gas_factors_affecting_prices.

EIA (2011d). "Review of Emerging Resources: U.S. Shale Gas and Shale Oil Plays." Retrieved 15.03, 2012, from http://www.eia.gov/analysis/studies/usshalegas/.

EIA (2012). "Annual Energy Outlook 2012 Early Release Overview." Retrieved 04.04, 2012, from http://www.eia.gov/forecasts/aeo/er/pdf/0383er%282012%29.pdf.

EIA (2012b). "Natural Gas Consumption by End Use." Retrieved 07.05, 2012, from http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm.

Engle, R. (2001). "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." The Journal of Economic Perspectives **15**(4): 157-168.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." Econometrica **50**(4): 987-1007.

Engle, R. F. and S. Manganelli (1999). "CAViaR: Conditional Value at Risk by Quantile Regression." National Bureau of Economic Research Working Paper Series(7341): 9-12.

Engle, R. F., V. K. Ng, et al. (1993). "Measuring and Testing the Impact of News on Volatility." The Journal of Finance **48**(5): 1749-1778.

Ewing, B. T., F. Malik, et al. (2002). "Volatility transmission in the oil and natural gas markets." Energy Economics **24**(6): 525-538.

Fama, E. F. (1981). "Stock Returns, Real Activity, Inflation, and Money." The American Economic Review **71**(4): 545-565.

Fleming, J., C. Kirby, et al. (2006). "Information, Trading, and Volatility: Evidence from Weather-Sensitive Markets." The Journal of Finance **61**(6): 2899-2930.

Foss, M. M. (2007). "United States Gas Prices to 2015." unpublished work: 1-36.

Geweke, J. (1986). "Modeling the persistence of conditional variances: A comment. ." Econometric Review **5**: 57-61.

Giot, P. and S. Laurent (2003). "Market risk in commodity markets: a VaR approach." Energy Economics **25**(5): 435-457.

Glosten, L. R., R. Jagannathan, et al. (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." The Journal of Finance **48**(5): 1779-1801.

Gogas, P. and A. Serletis (2010). "Forecast evaluation in daily commodities futures." International Journal of Financial Markets and Derivatives **1**(2): 155-168.

Hamilton, J. D. and G. Lin (1996). "Stock market volatility and the business cycle." Journal of Applied Econometrics **11**(5): 573-593.

Hansen, B. E. (1990). "Lagrange Multiplier Test for Parameter Instability in Nonlinear Models." unpublished work: 1-30.

Hartley, P. R., K. B. Medlock, et al. (2008). "The Relationship of Natural Gas to Oil Prices." The Energy Journal **29**(3): 47-66.

Henning, B., M. Sloan, et al. (2003). Natural Gas and Energy Price Volatility prepared for the Oak Ridge National Laboratory by the American Gas Foundation. Arlington, Virginia**:** 1-154.

Hung, J.-C., M.-C. Lee, et al. (2008). "Estimation of value-at-risk for energy commodities via fat-tailed GARCH models." Energy Economics **30**(3): 1173-1191.

IEA (2009). "World Energy Outlook 2009." Retrieved 28.01, 2012, from http://www.worldenergyoutlook.org/media/weowebsite/2009/WEO2009.pdf.

Jarque, C. M. and A. K. Bera (1987). "A Test for Normality of Observations and Regression Residuals." International Statistical Review **55**(2): 163-172.

Kupiec, P. H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models." Journal of Derivatives **3**(2): 73-84.

König, H., W. T. A. T. o. S. Gaab, vol. 2 of Inference, et al. (1982). The Advanced Theory of Statistics, vol 2 of Inference and Relationship. New York, Haffner.

Lamoureux, C. G. and W. D. Lastrapes (1990). "Persistence in Variance, Structural Change, and the GARCH Model." Journal of Business & Economic Statistics **8**(2): 225-234.

Liew, K. Y. and R. Brooks (1998). "Returns and volatility in the Kuala Lumpur crude." Journal of Futures Markets **18**(8): 985-999.

Linn, S. C. and Z. Zhu (2004). "Natural gas prices and the gas storage report: Public news and volatility in energy futures markets." Journal of Futures Markets **24**(3): 283-313.

Malo, P. and A. w.-. Kanto (2005). "Evaluating multivariate GARCH models in the Nordic Electricity Markets." Helsinki School of Economics working paper w-382: 1-38.

Marzo, M. and P. Zagaglia (2010). "Volatility forecasting for crude oil futures." Applied Economics Letters **17**(16): 1587-1599.

Mastrangelo, E. (2007). "An Analysis of Price Volatility in Natural Gas Markets." Retrieved 04.02, 2012, from ftp://ftp.eia.doe.gov/features/ngprivolatility.pdf.

Morgan, J. P. (1996). RiskMetrics-Technical Document (4th edition), J.P. Morgan, New York.

Mu, X. (2007). "Weather, storage, and natural gas price dynamics: Fundamentals and volatility." Energy Economics **29**(1): 46-63.

Murry, D. and Z. Zhu (2004). "EnronOnline and Informational Efficiency in the U.S. Natural Gas Market." Energy Journal **25**(2): 57-74.

Narayan, P. K. and S. Narayan (2007). "Modelling oil price volatility." Energy Policy **35**(12): 6549-6553.

Nelson, D. B. (1990). "Stationarity and Persistence in the GARCH(1,1) Model." Econometric Theory **6**: 318-334.

Nelson, D. B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." Econometrica **59**(2): 347-370.

Nyblom, J. (1989). "Testing for the Constancy of Parameters Over Time." Journal of the American Statistical Association **84**(405): 223-230.

Pearson, K. (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." Philosophical Magazine **50**(50): 157-172.

Pentula, S. (1986). "Modeling the persistence of conditional variances: A comment." Econometric Review **5**: 71-74.

Pindyck, R. (2004). "Volatility in natural gas and oil markets." Journal of Energy and Development **30**(1): 1-19.

Pindyck, R. S. (1994). "Inventories and the Short-Run Dynamics of Commodity Prices." The RAND Journal of Economics **25**(1): 141-159.

Ramberg, D. J. and J. E. Parsons (2010). "The Weak Tie Between Natural Gas and Oil Prices." SSRN eLibrary: 1-99.

Sadorsky, P. (2006a). "Modeling and forecasting petroleum futures volatility." Energy Economics **28**(4): 467-488.

Schwert, G. (1989). "Why Does Stock Market Volatility Change over Time?" Journal of Finance **44**(1115-53).

Schwert, G. (1990). "Stock volatility and the crash of '87." Review of Financial Studies **3**(1): 77-102.

Serletis, A. and P. Gogas (1999). "The North American natural gas liquids markets are chaotic." Energy Journal **20**(1): 83-103.

Serletis, A. and R. Rangel-Ruiz (2004). "Testing for common features in North American energy markets." Energy Economics **26**(3): 401-414.

Serletis, A. and A. Shahmoradi (2005). "Business cycles and natural gas prices." OPEC Review **29**(1): 75-84.

Serletis, A. and A. Shahmoradi (2006). "Returns and volatility in the NYMEX Henry Hub natural gas futures market." OPEC Review **30**(3): 171-186.

Shively, B. and J. Ferrare (2007). Understanding Today's Natural Gas Business. Laporte, Colorado, Energy Dynamics Corp.

Smead, R. G. (2010). "Price Instability in the U.S. Natural Gas Industry, Historical Perspective and Overview Prepared for the Task Force on Natural Gas Market Stability." unpublished work: 1-45.

Suenaga, H., A. Smith, et al. (2008). "Volatility dynamics of NYMEX natural gas futures prices." Journal of Futures Markets **28**(5): 438-463.

Susmel, R. and A. Thompson (1997). "Volatility, storage and convenience: Evidence from natural gas markets." Journal of Futures Markets **17**(1): 17-43.

Söderholm, P. (2000). "Fuel flexibility in the West European power sector." Resources Policy **26**(3): 157-170.

Taylor, S. J. (1986). Modelling Financial Time Series. New York, J. Wiley &Sons.

Taylor, S. J. (2005). Asset price dynamics, volatility, and prediction. Princeton, Princeton University Press.

Tse, Y. K. (2002). "Residual-Based Diagnostics for Conditional Heteroscedasticity Models." The Econometrics Journal **5**: 358-373.

Villar, J. and F. Joutz (2006). "The Relationship Between Crude Oil and Natural Gas Prices." Retrieved 27.02, 2012, from: www.eia.doe.gov/pub/oil_gas/natural_gas/feature_articles/2006/reloilgaspri/reloilgaspri.pdf.

Walls, W. D. (1995). "An Econometric Analysis of the Market for Natural Gas Futures." The Energy Journal **16**(1): 71-84.

Yaffe, R. A., M. A. Heddy, et al. (2008). "Forecasting Volatility and Value at Risk of United Kingdom Natural Gas futures." unpublished work: 1-12.

Youngquist, W. and R. Duncan (2003). "North American Natural Gas: Data Show Supply Problems." Natural Resources Research **12**(4): 229-240.

Zakoian, J.-M. (1994). "Threshold heteroskedastic models." Journal of Economic Dynamics and Control **18**(5): 931-955.

Appendix I Pindyck's (2004) Formula

The formula given is the following:

$$P_t = F1_t \left(\frac{F1_t}{F2_t}\right)^{\frac{n_{0t}}{n_1}} \qquad \text{Eq. I.1}$$

P_t : the spot price on day t
$F1_t$: the price of the nearest futures contract
$F2_t$: the price of the next-to-nearest futures contract
n_{0t} : the number of days from time t to expiration of the first contract
n_1 : the number of days between the expiration dates for the first and second contracts

Derivation:

$$F1_t = P_t \times e^{r_1 \times n_{0t}} \qquad \text{Eq. I.2}$$

$$F2_t = P_t \times e^{r_2 \times (n_{0t} + n_1)} \qquad \text{Eq. I.3}$$

where r_1 is the return between time t and the expiration for the nearest futures contract, and r_2 is the return between time t and the expiration for the next-to-nearest futures contract. Assuming a perfect market:

$$r_1 = r_2 = r$$

From Eq. I.2:

$$P_t = F1_t \times e^{-(r \times n_{0t})}$$

Insert this into Eq. I.3:

$$F2_t = F1_t \times e^{-(r \times n_{0t})} \times e^{r(n_{0t} + n_1)}$$

$$\Rightarrow F2_t = F1_t \times e^{r(n_1)} \qquad \text{Eq. I.4}$$

Using Eq. I.2 together with Eq. I.4 can eliminate the return r:

$$\frac{n_{0t}}{n_1} \times ln\left(\frac{F2_t}{F1_t}\right) = ln\left(\frac{F1_t}{P_t}\right)$$

$$ln\left(\frac{F2_t}{F1_t}\right)^{\frac{n_{0t}}{n_1}} = ln\left(\frac{F1_t}{P_t}\right)$$

$$\Rightarrow P_t = F1_t \times \left(\frac{F1_t}{F2_t}\right)^{\frac{n_{0t}}{n_1}} \qquad \text{Eq. I.5}$$

From Figure 14, the actual spot price and the spot price created using Pindyck's formula can be observed. These two series have a high degree of correlation which implies that the risk premium in the futures market is low. Also, it implies that the rate of return over the period for the two futures is approximately equal. This confirms one of the assumptions used in the formula, namely that the discount rate, and hence the risk, is expected to be the same in the two time periods n_{0t} and n_1.

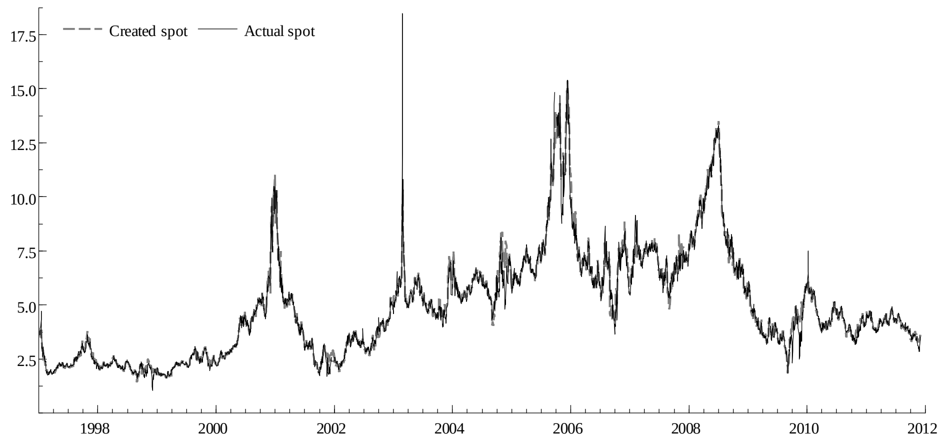

Figure 14: Created spot series vs. actual spot prices[41] ($/MMBtu)

Appendix II Methodology

II.1 OLS methodology

The pooled Ordinary Least Squares (OLS) model provides a linear regression model by minimizing the errors. The model is simple to understand and to implement. The time series for the conditional volatility is estimated using a GARCH(1,1) model with a Gaussian error distribution. The following will form the basis of the OLS model:

$$\hat{\sigma}_t^2 = \alpha + \sum_{i=1}^{10} \beta_i X_i \qquad \text{Eq. II.1}$$

where the summation term includes the determinants. Details on the determinants can be found in Appendix III.

[41] Dates not included due to constraints in natural gas spot price series: 13.jan.94-06.jan.97, 10.mar.00 and 23.sep.05-06.oct.05. Data not included due to constraints in natural gas created spot price series: 08.mar.04.

II.2 Conditional Volatility Models

In this section we present the models used in this paper to model the conditional volatility.

II.2.1 The ARCH model

The ARCH[42] model by Engle (1982) can be used whenever one suspects a time-varying variance, which can be detected by volatility clustering among other things. The parameters in the model are estimated by maximizing the log-likelihood value. It uses information set I_{t-1} whose content is known:

$$I_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\} \qquad \text{Eq. II.2}$$

The returns are defined using the information set:

$$r_t \mid I_{t-1} \sim D(\mu_t, \sigma_t^2) \qquad \text{Eq. II.3}$$

The conditional mean and variance are defined as:

$$\mu_t = E[r_t \mid I_{t-1}] \qquad \text{Eq. II.4}$$

$$\sigma_t^2 = Var(r_t \mid I_{t-1}) \qquad \text{Eq. II.5}$$

The distribution D(·) can vary depending on the distribution used for the error terms. The error term, or innovation, is defined as:

$$\varepsilon_t \mid I_{t-1} \sim D(0, \sigma_t^2)$$
$$\varepsilon_t = r_t - \mu \qquad \text{Eq. II.6}$$

In all GARCH models we can express the error term with the standardized variables z_t and the standard deviation:

$$z_t = \frac{r_t - \mu}{\sigma_t} = \frac{\varepsilon_t}{\sigma_t}$$
$$\Rightarrow \varepsilon_t = \sigma_t z_t \qquad \text{Eq. II.7}$$

The standardized variables are i.i.d.[43] and are considered to have zero mean and unit variance:

$$z_t \mid I_{t-1} \sim D(0,1)$$

With no ARMA process in the mean, the ARCH model is defined as:

$$r_t = \mu + \varepsilon_t \qquad \text{Eq. II.8}$$

[42] Auto Regressive Conditional Heteroskedasticity
[43] Independent and identically distributed

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad \text{Eq. II.9}$$

The coefficients are all assumed to be non-negative:

$$\omega > 0, \alpha_i \geq 0$$

Brooks and Persand (2002) points out that ARCH rarely have been used in recent time, and have been replaced by the Generalized ARCH (GARCH) model.

II.2.2 GARCH

This Generalized ARCH model comes from the approach based on Bollerslev's (1986) generalization of Engle's (1982) ARCH model. It contains an autoregressive term of the conditional volatility itself, and is defined as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \text{Eq. II.10}$$

To ensure that the conditional variance is positive, we require the following in the maximum likelihood estimation of the parameters:

$$\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$$

Also, to ensure a covariance stationary process we have:

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \qquad \text{Eq. II.11}$$

II.2.3 GJR-GARCH

This model was proposed by Glosten, Jagannathan et.al (1993), and its generalized version is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \text{Eq. II.12}$$

where

$$S_t^- = \begin{cases} 1 & if \quad \varepsilon_{t-1} < 0 \\ 0 & if \quad \varepsilon_{t-1} \geq 0 \end{cases}$$

γ determines the sign and size of the asymmetry effect. In equity returns this is typically positive, whereas it is usually negative for commodity markets. The second regularity condition is given by $\alpha + \beta + 0.5\gamma < 1$.

II.2.4 IGARCH

The GARCH models are said to be second order stationary provided that $\alpha + \beta < 1$. Often this sum is very close to unity, and the Integrated GARCH (IGARCH) model by Engle and Bollerslev (1986) has the assumption of $\alpha + \beta = 1$. This implies that the IGARCH model is not covariance stationary.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}(1-\beta_i)\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 \qquad \text{Eq. II.13}$$

Even though this model is not covariance stationary, it is still strictly stationary with a well-defined non-degenerate limiting distribution (Nelson 1990).

II.2.5 RiskMetrics

This model was invented by the risk management group at J.P. Morgan in 1994, and is a standard model in the market risk measurement due to its simplicity. The model is a variety of the IGARCH model, where the ARCH and GARCH coefficients are fixed to 0.06 and 0.94, respectively (Morgan 1996). Formally, the RiskMetrics model is defined as:

$$\sigma_t^2 = \omega + (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2 \qquad \text{Eq. II.14}$$

where $\omega = 0$ and $\lambda = 0.94$ when using daily data. This model does not require iteratively solving for the parameters using the max-likelihood estimation.

II.2.6 EGARCH

Due to the complexity in the EGARCH model, (p,q)=(1,1) is used in the model descriptions, which is also what is used in this paper. The Exponential GARCH model by Nelson (1991) does not require the strict restrictions of the coefficients being all positive, and formulates the conditional variance equation in terms of the log of the variance. While the log of the variance might be negative, the variance will always be positive. This model also uses the standardized innovations defined by Eq. II.7. The model specification for EGARCH is as follows:

$$\ln(\sigma_t^2) = \omega + \alpha g(z_{t-1}) + \beta \ln(\sigma_{t-1}^2) \qquad \text{Eq. II.15}$$

The function g(z) is defined as:

$$g(z_t) = \gamma_1 z_t + \gamma_2(|z_t| - E[|z_t|]) \qquad \text{Eq. II.16}$$

γ_1 defines the sign effect, while γ_2 defines the magnitude of the asymmetry effect. If $\gamma_1 < 0$, negative shocks will have a larger impact on the future volatility than positive shocks of the same magnitude. The expectance of the normalized innovation depends on the unconditional density assumption made on the error distributions.

II.2.7 APARCH

Due to the complexity in the APARCH model, (p,q)=(1,1) is used in the model descriptions, which is also what is used in this paper. This model was introduced by Ding, Granger, and Engle (1993), and can be expressed as:

$$\sigma_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta \qquad \text{Eq. II.17}$$

where $\delta > 0$ and $-1 < \gamma < 1$. The δ parameter plays the role of a Box-Cox transformation of the conditional standard deviation, and γ reflects the asymmetry effect.

An interesting feature with the APARCH model is that it has many other ARCH extensions, some of which are listed below:

1. If $\delta = 2$, $\gamma = 0$ and $\beta = 0$ we get Engle's (1982) ARCH model
2. If $\delta = 2$ and $\gamma = 0$ we get the GARCH model of Bollerslev (1986)
3. If $\delta = 1$ and $\gamma = 0$ we get the GARCH model of Taylor (1986) and Schwert (1990)
4. If $\delta = 2$ we get the GJR-GARCH model of Glosten, Jagannathan et. al (1993)
5. If $\delta = 1$ we get the TARCH model of Zakoian (1994)
6. If $\beta = 0$ and $\gamma = 0$ we get the NARCH model of Bera and Higgins (1993)
7. If $\delta \to 0$ and we get the Log-ARCH of Geweke (1986) and Pentula (1986)

II.2.8 The Autoregressive Process

This process is simply that a variable depends on its previous values. An Autoregressive (AR) model of order m, denoted as AR(m), can be expressed as the following with returns as the variable of interest:

$$r_t = \mu_t + \sum_{i=1}^{m} r_{t-i} + \varepsilon_t \qquad \text{Eq. II.18}$$

II.2.9 Autocorrelation

The formal definition of autocorrelation is that the returns are not independent of previous returns. The autocorrelation function (ACF) is defined as follows:

$$\varphi_k = \frac{\operatorname{cov}(r_t, r_{t-k})}{\sigma_{r,t} \times \sigma_{r,t-k}} \qquad \text{Eq. II.19}$$

Appendix III Determinants

III.1 Definition of the determinants

The volatility of oil (Oil Vol) and the S&P 500 index (S&P vol) are initially unobservable. To create proxies for these determinants we have used vanilla GARCH models to create an estimate for the volatility for each of the determinants, which are used throughout the paper. As this paper focuses on natural gas futures, Pindyck's (2004) formula is also applied to the oil futures to create a spot price, yielding comparable time series of oil and natural gas.

According to hypothesis 5 and 7 in section 3.2 the storage and production proxies are to be created in such a manner that a deviation (negative or positive) from the historical average is to increase the size of the variables. As a consequence we need to create a proxy for these determinants that can be readily measured, is unaffected by seasonality, and increases when deviating from the historical average.

The production proxy (Prod diff (abs)) is generated by using the average of the weekly averages for the subsequent five years, and then taking the absolute value of the difference between this and the respective day's actual production level.

$$Prod\ \ diff\ \ (abs)_t = \left| Prod_t - \frac{\sum_{i=1}^{2}\sum_{j=1}^{5} Prod_{t-i*365-j}}{25} \right| \qquad \text{Eq. III.1}$$

For the storage proxy (Stock diff (abs)), the same method is used, except that only the two last years are used to create the historical average level of storage due to data limitations. As a result, increased values of these determinants imply a deviation from the historical level, which will serve as a proxy for the deviation from expectation.

$$Stock\ \ diff\ \ (abs)_t = \left| Stock_t - \frac{\sum_{i=1}^{5}\sum_{j=1}^{5} Stock_{t-i*365-j}}{25} \right| \qquad \text{Eq. III.2}$$

Similar to the above proxies, the proxy for temperature has to be unaffected by seasonality, and increases when deviating from the historical average. To understand the squared Degree Day difference (DD diff sq) proxy we need to first define the Heating Degree Days (HDD) and Cooling Degree Days (CDD). The HDD is defined as the number of degrees above 65°F, otherwise zero, for the respective day it is measured. Similarly, the CDD is defined as the number of degrees below 65°F. Since weather is very variable across geographical locations in the US we have chosen the 10 largest cities, and then made an estimate for the HDD and CDD based on the population-weighted average for the cities to make a weather index for the entire US. The DD is the sum of HDD and CDD for the respective date. For each day of the year, an average DD is found, and the DD difference is then defined as the DD for the respective day less the average DD for the entire period on this day of the year. This implies that if the DD diff sq is high, the temperature has deviated from normal levels.

$$DD_t = \left[\left(HDD_t + CDD_t \right) - \left(\overline{HDD_t} + \overline{CDD_t} \right) \right]^2 \qquad \text{Eq. III.3}$$

Where $\overline{HDD}$ and $\overline{CDD}$ is the average heating degree days and cooling degree days for that particular day of the year.

The dummies in the analysis are relatively self-explanatory. The seasonal dummies are 1 in their respective seasons, 0 otherwise. Winter is defined as December through February, Spring as March through May, and Summer as June through August. The Update dummy is 1 for the date of publication for the weekly natural gas storage report, 0 otherwise. This is often on Thursdays, but varies on several occasions.

Summary statistics for the determinants are included in Table 25.

Table 25: Descriptive statistics for the determinants used in this paper

	# obs	Minimum	Mean	Maximum	St.dev.	Skewness	Excess kurtosis
Oil ret	3893	-17.25	0.040	18.646	2.582	-0.048	3.914
SP ret	3893	-9.470	0.017	10.957	1.334	-0.217	6.938
Stock diff (abs)	3893	0.000	0.148	0.649	0.137	1.315	1.153
Prod diff (abs)	3893	0.000	0.066	0.284	0.056	1.089	0.500
DD diff sq	3893	0.000	10.25	195.2	17.29	3.477	16.71
Oil vol	3893	2.489	6.527	53.29	5.124	4.557	26.19
SP vol	3893	0.297	1.791	27.97	2.548	5.522	38.27

III.2 Time series of determinants

Figure 15 and 16 show the presence of volatility clustering in returns on both the S&P 500 and oil price. Therefore, the application of a vanilla GARCH model is appropriate to model the volatility of these two prices series. Figure 17 to Figure 21 show the times series of determinants (not including dummy variables). The stock and production difference clearly show signs of an AR-process, while the temperature proxy seems to be a normally distributed variable.

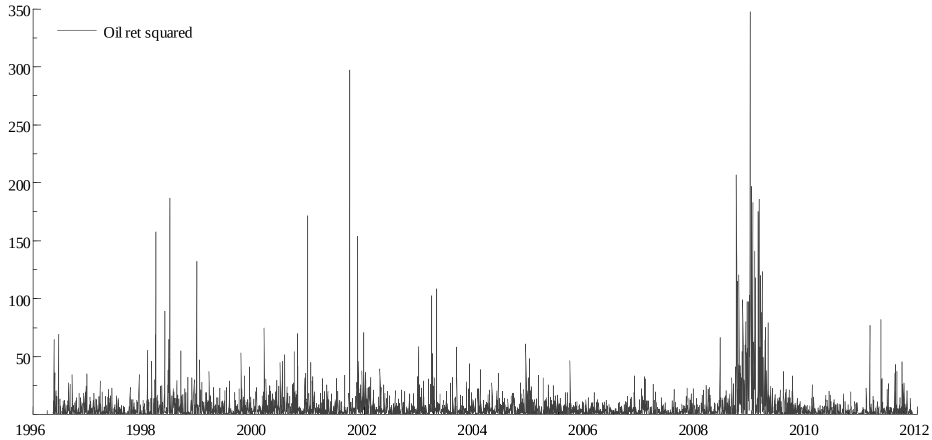

Figure 15: Time series of squared oil returns

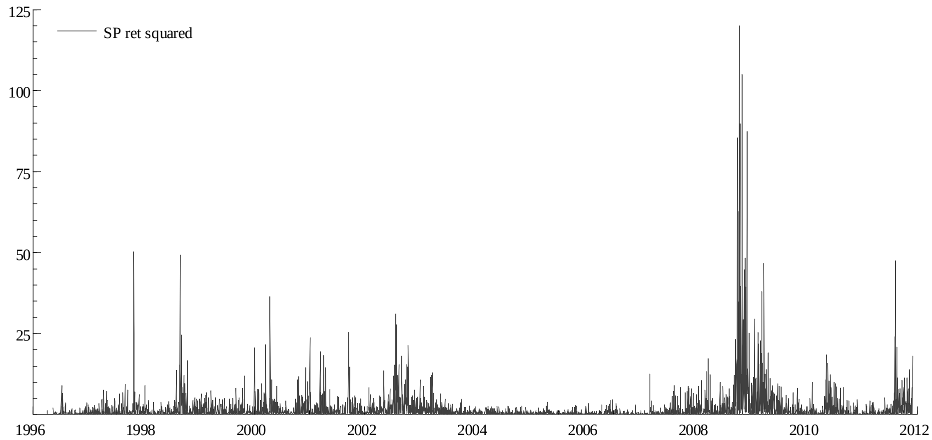

Figure 16: Time series of squared S&P 500 returns

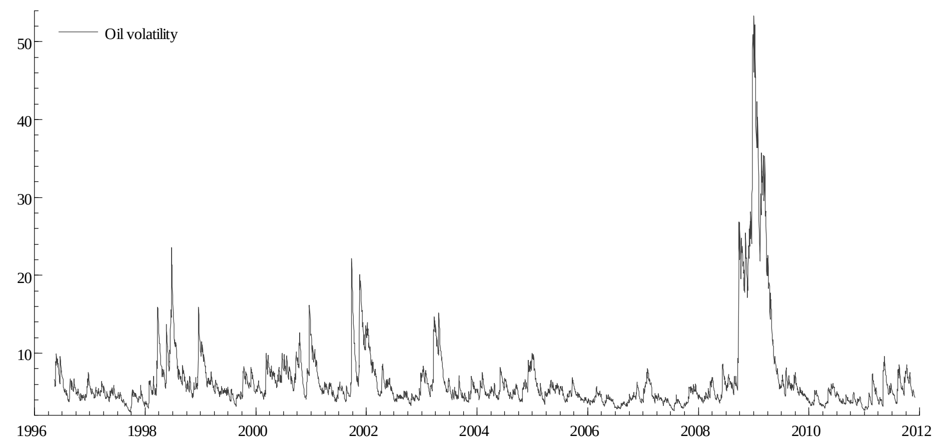

Figure 17: Time series of estimated oil volatility

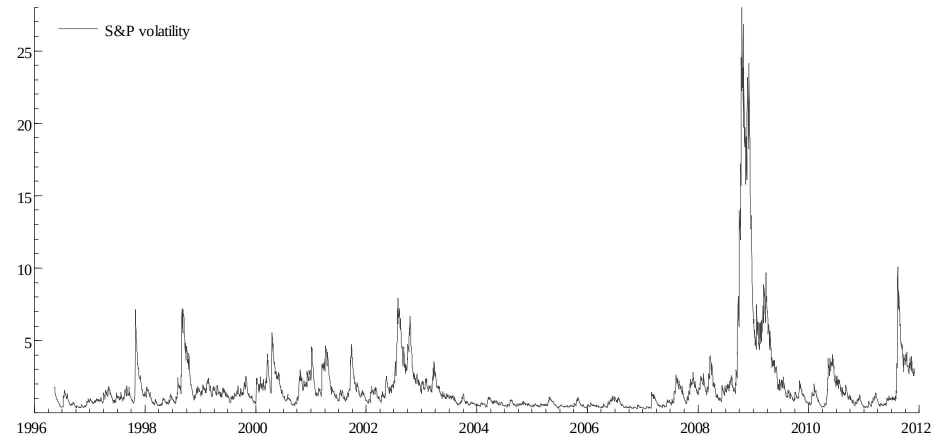

Figure 18: Time series of estimated S&P 500 volatility

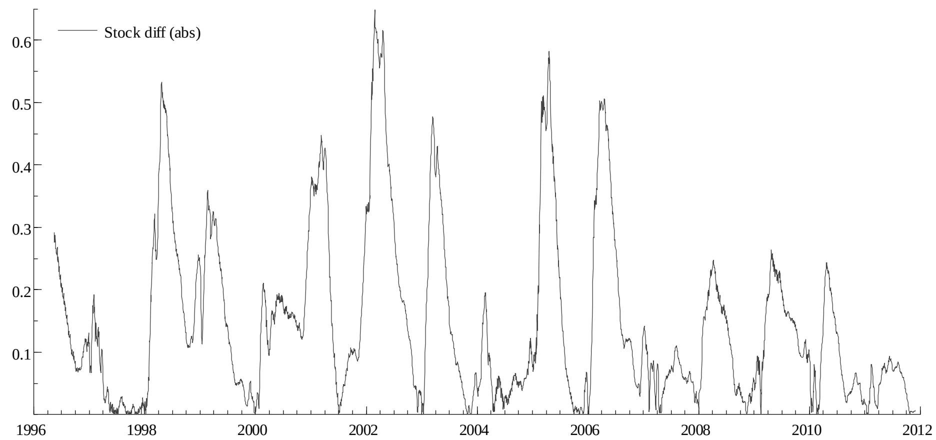

Figure 19: Time series of Stock diff (abs)

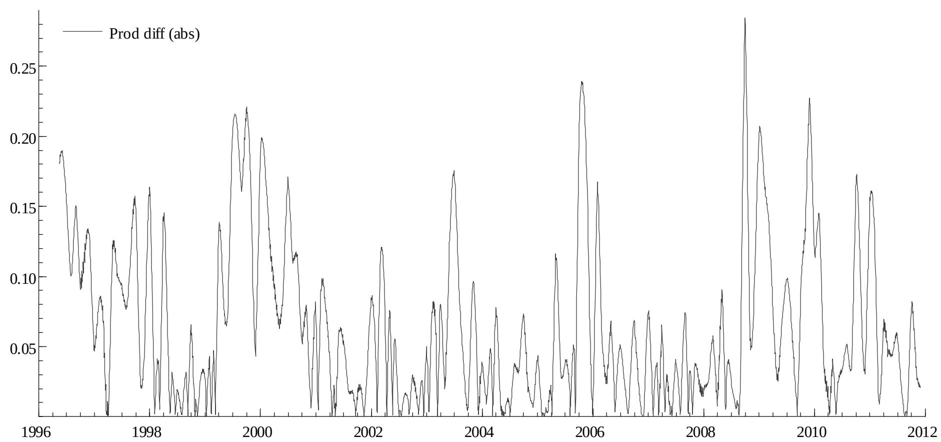

Figure 20: Time series of Prod diff (abs)

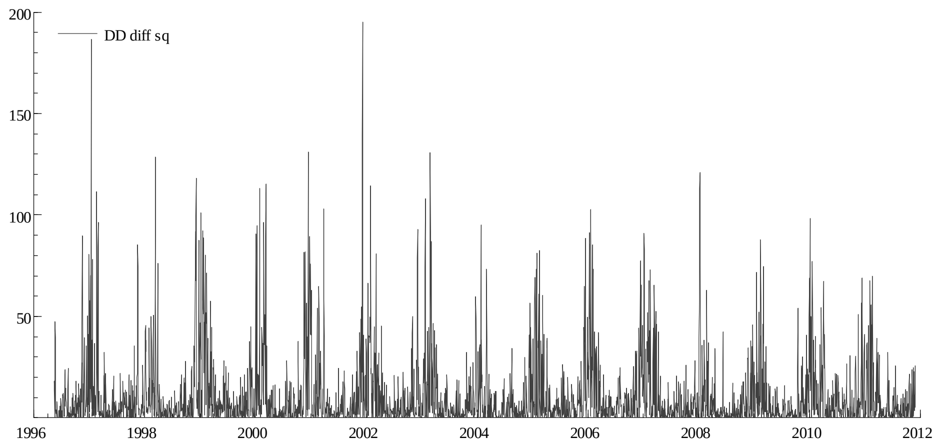

Figure 21: Time series of DD diff sq

Appendix IV Model Selection

IV.1 Determination of the autoregressive lags in the conditional volatility equation

By altering p and q for all the six models, we can see that adding more parameters does not add information or explanatory power. The estimation performed is on the in-sample period with no AR or ARCH-in-mean terms with student-t error distribution. As seen from the tables below, using (p,q)=(1,1) is the best combination; however, (p,q)=(4,1) also give good results. Since the better of the two combinations will depend on whether or not we use AIC or BIC[44], we conclude that the two are as good combinations of the autoregressive lags. Since (p,q)=(1,1) is far more easy to understand and estimate, we will use this combination throughout the paper.

[44] To be minimized

Table 26: Information criteria with varying p and q (Darkest=best, Lightest=worst)

GARCH(p,q)

p \ q		1	2	3	4	5
1	AIC	5.4178	5.4184	5.4189	5.4184	5.4177
	BIC	5.4274	5.4299	5.4324	5.4338	5.4351
2	AIC	5.4186	5.4187	5.4193	5.4186	5.4185
	BIC	5.4302	5.4322	5.4347	5.4359	5.4378
3	AIC	5.4191	5.4186	5.4165	5.4192	5.4188
	BIC	5.4326	5.4340	5.4339	5.4385	5.4400
4	AIC	5.4165	5.4172	5.4176	5.4204	n/a
	BIC	5.4319	5.4345	5.4369	5.4416	n/a
5	AIC	n/a	5.4179	n/a	n/a	5.4178
	BIC	n/a	5.4372	n/a	n/a	5.4428

EGARCH(p,q)

p \ q		1	2	3	4	5
1	AIC	n/a	n/a	n/a	5.4098	n/a
	BIC	n/a	n/a	n/a	5.4291	n/a
2	AIC	n/a	n/a	5.4810	n/a	n/a
	BIC	n/a	n/a	5.5002	n/a	n/a
3	AIC	n/a	n/a	5.4081	n/a	n/a
	BIC	n/a	n/a	5.4293	n/a	n/a
4	AIC	n/a	n/a	n/a	n/a	n/a
	BIC	n/a	n/a	n/a	n/a	n/a
5	AIC	n/a	n/a	n/a	n/a	n/a
	BIC	n/a	n/a	n/a	n/a	n/a

GJR(p,q)

p \ q		1	2	3	4	5
1	AIC	5.4175	5.4188	5.4199	5.4201	n/a
	BIC	5.4291	5.4342	5.4391	5.4432	n/a
2	AIC	5.4184	5.4190	5.4201	5.4198	n/a
	BIC	5.4319	5.4363	5.4413	5.4448	n/a
3	AIC	5.4189	5.4189	5.4176	5.4179	n/a
	BIC	5.4343	5.4381	5.4407	5.4449	n/a
4	AIC	5.4165	5.4198	5.4163	5.4176	n/a
	BIC	5.4338	5.4410	5.4413	5.4465	n/a
5	AIC	5.4191	5.4210	5.4171	n/a	n/a
	BIC	5.4383	5.4441	5.4441	n/a	n/a

APARCH(p,q)

p \ q		1	2	3	4	5
1	AIC	5.4107	5.4119	5.4133	n/a	n/a
	BIC	5.4242	5.4293	5.4345	n/a	n/a
2	AIC	5.4111	5.4122	5.4135	5.4147	n/a
	BIC	5.4265	5.4315	5.4366	5.4417	n/a
3	AIC	5.4115	5.4087	n/a	n/a	n/a
	BIC	5.4289	5.4337	n/a	n/a	n/a
4	AIC	5.4104	n/a	n/a	n/a	n/a
	BIC	5.4297	n/a	n/a	n/a	n/a
5	AIC	5.4112	5.4123	n/a	n/a	n/a
	BIC	5.4324	5.4373	n/a	n/a	n/a

p \ q		1	2	3	4	5
		IGARCH(p,q)				
1	AIC	5.4196	5.4201	5.4207	5.4203	5.4195
	BIC	5.4273	5.4298	5.4323	5.4338	5.4349
2	AIC	5.4204	5.4209	5.4215	5.4204	5.4203
	BIC	5.4300	5.4324	5.4349	5.4358	5.4376
3	AIC	5.4211	5.4207	5.4206	5.4210	5.4207
	BIC	5.4327	5.4341	5.4360	5.4383	5.4399
4	AIC	5.4179	5.4185	5.4192	5.4206	n/a
	BIC	5.4314	5.4339	5.4365	5.4398	n/a
5	AIC	5.4204	5.4185	5.4201	n/a	n/a
	BIC	5.4358	5.4397	5.4393	n/a	n/a

IV.2 Determination of AR(1) Process and ARCH-in-mean

By including an AR(1)-term in all six models, it can be seen from the tables below that the AR(1)-term is always significant. Also, with an AR(1)-term in the mean equation the highest log-likelihood values are received. Additionally, by adding an AR(1)-term the AIC is significantly improved, whereas the BIC is approximately the same. Therefore an AR(1)-term will be included in the GARCH models since this increases the explanatory power of the models. When adding the ARCH-in-mean term, this term is always significant. Slightly higher log-likelihood value compared to the models without the ARCH-in-mean term is also found. While the AIC is approximately the same with and without this term, the BIC is lower, which implies that adding this term does not increase the explanatory power by much. Therefore the ARCH-in-mean term will not be included in the GARCH models. All estimations were performed on the in-sample period with p and q set to 1, and with student-t distributed error terms.

Table 27: Information criteria and log likelihood values with and without AR and ARCH-in-mean terms

		RiskMetrics(1,1)	**GARCH(1,1)**	**EGARCH(1,1)**
AR(0) No ARCH-in-mean	Log-likelihood	-8537	-8506.3	no conv.
	AIC	5.43539	5.41776	no conv.
	BIC	5.43924	5.42739	no conv.
AR(1) term	t-prob AR(1)	0.0069	0.0066	no conv.
	Log-likelihood	-8533	-8502.7	no conv.
	AIC	5.43347	5.41612	no conv.
	BIC	5.43925	5.42768	no conv.
ARCH-in-mean	t-prob (var)	0.0137	0.0143	no conv.
	Log-likelihood	-8534.4	-8503.8	no conv.
	AIC	5.4344	5.41682	no conv.
	BIC	5.44017	5.42838	no conv.

		GJR(1,1)	APARCH(1,1)	IGARCH(1,1)
AR(0) No ARCH-in-mean	Log-likelihood	-8504.9	-8493.2	-8510.2
	AIC	5.41753	5.41071	5.41958
	BIC	5.42909	5.42419	5.42728
AR(1) term	t-prob AR(1)	0.0053	0.0098	0.005
	Log-likelihood	-8501.2	-8489.7	-8506.5
	AIC	5.41577	5.40908	5.41791
	BIC	5.42925	5.42449	5.42755
ARCH-in-mean	t-prob (var)	0.0156	0.9068	0.009
	Log-likelihood	-8502.5	-8489.9	-8507.6
	AIC	5.41663	5.40925	5.41861
	BIC	5.43011	5.42467	5.42824

Appendix V Statistical Tests

V.1 Kupiec LR test

The models in this paper are evaluated by comparing the one-day-ahead forecasts VaR-predictions for long and short positions with VaR-levels ranging from 1[st] to 25[th] percentiles according to Kupiec (1995)). This test is designed to test if the observed failure rate ($\hat{f}$) is different from the theoretical failure rate (α) in a VaR-prediction. The test statistic is defined as,

$$LR = -2\log\left(\frac{\alpha^N (1-\alpha)^{T-N}}{\hat{f}^N \left(1-\hat{f}\right)^{T-N}}\right) \qquad \text{Eq. V.1}$$

where T is the number of predictions, N is the number of violations of the VaR-limits, f is the failure rate and α is the theoretical failure rate, and has the following hypothesis,

$$H_o: \quad f = \alpha$$
$$H_1: \quad f \neq \alpha$$

Under the null hypothesis the test statistic is asymptotically distributed as χ^2.

V.2 Dynamic Quantile test

The dynamic quantile test by Engle and Manganelli (1999) is designed to assess whether the VaR-violations are serially correlated. This is equivalent to testing whether the sequence $\{I(y_t < -VaR_t)\}_{t=1}^{T} \equiv \{I_t\}_{t=1}^{T}$ is i.i.d. This is done by defining the following variable,

$$Hit(\alpha)_t \equiv I\left(y_t < VaR_t(\alpha)\right) - \alpha \qquad \text{Eq. V.2}$$

$$Hit(1-\alpha)_t \equiv I\left(y_t > VaR_t(1-\alpha)\right) - \alpha \qquad \text{Eq. V.3}$$

$Hit(\alpha)$ assumes the value $1-\alpha$ when $y_t < VaR(\alpha)$ and $-\alpha$ otherwise. The expectation of $Hit(\alpha)$ is therefore by design 0. Similarly, since the sequence $\{I_t\}_{t=1}^{T}$ is i.i.d, the expectation of $Hit(\alpha)$ given $t-i$ is 0 for $t < i$. Accoring to Engle and Manganelli (1999), a convenient way of constructing a test for the statement above is to perform the following OLS regression,

$$Hit_t = X\delta + u_t \quad where \quad u_t = \begin{cases} -\alpha & prob(1-\alpha) \\ 1-\alpha & prob(\alpha) \end{cases} \qquad \text{Eq. V.4}$$

In this regression X is a T x k matrix where the first column is a column of ones, the next p columns are$Hit_{t-1}, \dots, Hit_{t-p}$,and the remaining columns are the remaining independent variables included in the VaR estimate. A good VaR-model will produce a regression result where the explanatory power of the regression is very close to 0. As a result the test statistic becomes,

$$\frac{\hat{\delta}_{OLS}{}' X' X \hat{\delta}_{OLS}}{\alpha(1-\alpha)} \sim \chi^2(k)$$

V.3 Sign Bias test

The sign bias (SBT), negative sign bias (NSBT) and positive sign bias tests (NSBT) introduced by Engel and Ng (1993) are designed to test whether it is possible to predict the squared normalized residual by some variable observed in the past which are not included in the volatility model being used. If such a variable exist, then the current variance model is miss-specified. To perform this test the following variables are created,

SBT: S_{t-1}^{-}

NSBT: $S_{t-1}^{-}\hat{\varepsilon}_{t-1}$

PSBT: $S_{t-1}^{+}\hat{\varepsilon}_{t-1}$

Engel and Ng (1993) suggest running the following regressions to test whether the model is correctly specified,

SBT: $\hat{\varepsilon}_t^2 = a + bS_{t-1}^{-} + u_t$ Eq. V.5

NSBT: $\hat{\varepsilon}_t^2 = a + bS_{t-1}^{-}\hat{\varepsilon}_{t-1} + u_t$ Eq. V.6

PSBT: $\hat{\varepsilon}_t^2 = a + bS_{t-1}^{+}\hat{\varepsilon}_{t-1} + u_t$ Eq. V.7

and testing the significance of a and b using a t-test. Alternatively, the test can be run jointly on the three effects,

$$\hat{\varepsilon}_t^2 = a + bS_{t-1}^{-} + cS_{t-1}^{-}\hat{\varepsilon}_{t-1} + dS_{t-1}^{+}\hat{\varepsilon}_{t-1} + u_t \qquad \text{Eq. V.8}$$

In this case, the joint test for $H_0 : a = b = c = d = 0$ is reported.

V.4 Nyblom Stability Test

The statistic introduced by Nyblom (1989) is appropriate to test if the parameters in a model are consistent. The statistic tests if the parameters follow a martingale process by calculating the following cumulative moments,

$$S_t = \sum_{i=1}^{t} \frac{\partial}{\partial \theta} l_i\left(\hat{\theta}\right) \qquad \text{Eq. V.9}$$

for all t. The test statistic is given by,

$$L = \frac{1}{n}\sum_{t=1}^{n} S'_t \hat{V}^{-1} S_t \qquad \text{Eq. V.10}$$

L is asymptotically distributed and depends only on the number of parameters in θ . The statistic tests the null hypothesis that the vector θ is stable against the alternative that the entire vector is unstable. The test can be modified to test individual elements of the vector, and is given by,

$$L_k = \frac{1}{n}\sum_{t=1}^{n} \frac{S_{kt}^2}{\hat{V}_{kk}} \qquad \text{Eq. V.11}$$

For further details see Nyblom (1989)

Table 28: Critical values for the Nyblom stability test (Hansen 1990):

	Significance level					
Degrees of Fredoom (k_1)	*1 %*	*2.5 %*	*5 %*	*7.5 %*	*10 %*	*20 %*
1	.748	.593	.005	.398	.353	.243
2	1.07	.898	.749	.670	.610	.469
3	1.35	1.16	1.01	.913	.846	.679
4	1.60	1.39	1.24	1.14	1.07	.883
5	1.88	1.63	1.47	1.36	1.28	1.08
6	2.12	1.89	1.68	1.58	1.49	1.28
7	2.35	2.10	1.90	1.78	1.69	1.46
8	2.59	2.33	2.11	1.99	1.89	1.66
9	2.82	2.55	2.32	2.19	2.10	1.85
10	3.05	2.76	2.54	2.40	2.29	2.03
11	3.27	2.99	2.75	2.60	2.49	2.22
12	3.51	3.18	2.96	2.81	2.69	2.41
13	3.69	3.39	3.15	3.00	2.89	2.59
14	3.90	3.60	3.34	3.19	3.08	2.77
15	4.07	3.81	3.54	3.38	3.26	2.95
16	4.30	4.01	3.75	3.58	3.46	3.14
17	4.51	4.21	3.95	3.77	3.64	3.32
18	4.73	4.40	4.14	3.96	3.83	3.50
19	4.92	4.60	4.33	4.16	4.03	3.69
20	5.13	4.79	4.52	4.36	4.22	3.86

V.5 Box-Pierce

Box and Pierce (1970) introduce a statistic for testing if there exists residual autocorrelation in time-series models. The test statistic is defined as follows,

$$Q(m) = T^2 \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l} \qquad \text{Eq. V.12}$$

Where ρ_t is the serial correlation with lag l, m is the number of serial correlations being tested and T is the length of the sample being tested. Under the null of no serial correlation, the test statistic is $\chi^2(m)$-distributed. When evaluating the squared residuals of a GARCH(p,q) model, the test statistic becomes $\chi^2(m-p-q)$--distributed.

V.6 LM ARCH test

Engle (1982) proposed the LM ARCH statistic to test for the presence of ARCH effects in a time-series. Using a simple OLS regression,

$$\hat{y}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \hat{y}_{t-i}^2 \qquad \text{Eq. V.13}$$

estimates for $\hat{\alpha}_i$ are produced. Under the null hypothesis we have $\hat{\alpha}_i = 0$ for $i = 1, \dots, q$ and the test statistic becomes TR^2 which follows a χ^2-distribution with q degrees of freedom. This test can be modified to test the residuals of a regression by running the following regression,

$$\hat{\varepsilon}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i \hat{\varepsilon}_{t-i}^2 \qquad \text{Eq. V.14}$$

V.7 The Jarque-Bera Test for Normality

This test, introduced by Jarque and Bera (1987) tests the null hypothesis that the observed data comes from a normal distribution. This is a goodness-of-fit test that tests if the kurtosis and skewness could have been drawn from a normal distribution. The test statistic is defined as follows,

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4} \right) \qquad \text{Eq. V.15}$$

where n is the number of observations, S is the sample skewness and K is the sample kurtosis. The test statistics used has an asymptotic $\chi^2(2)$ distribution.

V.8 Information Criteria

The information criteria (Akaike 1974; Abascal, Zardoya et al. 2005) are a measure of the goodness-of-fit of the conditional variance model. It takes into account both the value of the maximized log-likelihood function and the number of parameters that must be estimated in the model by imposing a penalty for increasing the number of parameters. The four information criteria calculated by OxMetrics™ are the following,

$$AIC = -2\frac{LogL}{n} + 2\frac{k}{n} \qquad \text{Eq. V.16}$$

$$BIC = -2\frac{LogL}{n} + 2\frac{\log(k)}{n} \qquad \text{Eq. V.17}$$

$$Harmann - Quinn = -2\frac{LogL}{n} + 2\frac{k\log[\log(n)]}{n} \qquad \text{Eq. V.18}$$

$$Shibata = -2\frac{LogL}{n} + 2\left(\frac{n+2k}{n}\right) \qquad \text{Eq. V.19}$$

V.9 Adjusted Pearson Goodness-of-Fit Test

In his seminal paper, Pearson (1900) proposed a test to compare an empirical distribution to a theoretical one. The adjusted Pearson goodness-of-fit test compares the theoretical distribution of innovations to the distribution of innovations used in the conditional heteroskedasticity model. The test divides the theoretical distribution into r categories, where

$$p_1 = p_2 = \ldots = p_r = p$$

The statistic is computed as follows,

$$P(g) = \sum_{i=1}^{r} \frac{(n_i - En_i)^2}{En_i} \qquad \text{Eq. V.20}$$

where n_i is the number of empirical innovations in category i, En_i is the number of theoretical innovations in category i. According to Palm and Vlaar (1997) the asymptotic distribution of the test statistic under the null hypothesis of a correct distribution is bounded between $\chi^2(r-1)$ and $\chi^2(r-k-1)$ where k is the number of estimated parameters.

V.10Unit Root Test

Dickey and Fuller (1981) introduced a test for unit roots. The test statistic is given by,

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \qquad \text{Eq. V.21}$$

Where $\hat{\gamma}$ is determined by the following regression,

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p-1} + \varepsilon_t \qquad \text{Eq. V.22}$$

The test statistic is then compared to Dickey-Fuller table. The intuition behind the test is that if the time-series y is stationary it tends to return to some constant level. As a consequence large values of y tend to be followed by smaller ones and similarly for small values. Therefore, the size of y will be a predictor of the next periods change and will therefore have a negative coefficient.

V.11Residual Based Diagnostic

Tse (2002) presents a test based on the following regression,

$$\hat{\eta}_t^2 - 1 = \delta_1 \hat{\eta}_{t-1}^2 + \delta_2 \hat{\eta}_{t-2}^2 + \ldots + \delta_M \hat{\eta}_{t-M}^2 + \varepsilon_t \qquad \text{Eq. V.23}$$

where $\hat{\eta}_t$ is the standardized residuals. Because the regressors are inferred, not observed, under the null hypothesis of model adequacy, $H_0 : \delta_i = 0$ for $i = 1,\ldots,M$ the RBD statistic is $\chi^2(M)$ distributed.

Appendix VI Forecasts in- and out-of-sample

Table 29: In-sample back testing without determinants

		RiskMetrics				GARCH			
		Long		***Short***		***Long***		***Short***	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.931	0.335	0.021	0.885	0.051	0.821	1.026	0.311
	10 %	4.632	0.031	2.299	0.129	0.487	0.485	0.723	0.395
	5 %	1.445	0.229	4.278	0.039	0.252	0.615	0.522	0.470
	2.5 %	2.935	0.087	10.823	0.001	0.004	0.950	0.003	0.959
	1 %	2.692	0.101	1.305	0.253	0.659	0.417	0.390	0.532
DQT	25 %	4.980	0.546	4.749	0.576	5.006	0.543	5.172	0.522
	10 %	7.078	0.314	9.317	0.157	5.006	0.543	2.264	0.894
	5 %	5.425	0.491	10.258	0.114	6.716	0.348	3.431	0.753
	2.5 %	9.408	0.152	14.708	0.023	11.703	0.069	4.816	0.568
	1 %	8.588	0.198	17.716	0.007	4.265	0.641	1.743	0.942

		EGARCH				GJR			
		Long		***Short***		***Long***		***Short***	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.308	0.579	1.026	0.311	0.051	0.821	1.390	0.238
	10 %	0.665	0.415	0.096	0.757	0.764	0.382	0.829	0.363
	5 %	0.000	0.993	0.000	0.993	0.030	0.863	0.231	0.631
	2.5 %	0.027	0.869	0.276	0.600	0.004	0.950	0.153	0.695
	1 %	0.011	0.917	0.390	0.532	0.193	0.660	0.390	0.532
DQT	25 %	4.428	0.619	6.770	0.343	3.515	0.742	5.405	0.493
	10 %	2.106	0.910	3.831	0.700	4.139	0.658	3.109	0.795
	5 %	5.171	0.522	2.029	0.917	4.833	0.565	2.974	0.812
	2.5 %	10.091	0.121	4.946	0.551	11.703	0.069	2.869	0.825
	1 %	2.752	0.839	1.743	0.942	5.168	0.522	1.743	0.942

		APARCH				IGARCH			
		Long		***Short***		***Long***		***Short***	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.355	0.551	1.810	0.179	0.004	0.951	1.390	0.238
	10 %	0.337	0.562	0.096	0.757	0.000	0.991	0.943	0.332
	5 %	0.008	0.928	0.005	0.941	2.824	0.093	0.000	0.993
	2.5 %	0.027	0.869	0.086	0.770	0.276	0.600	0.769	0.381
	1 %	0.011	0.917	0.390	0.532	8.023	0.005	1.927	0.165
DQT	25 %	3.281	0.773	7.746	0.257	6.032	0.420	4.338	0.631
	10 %	1.515	0.958	4.156	0.656	4.515	0.607	2.235	0.897
	5 %	5.390	0.495	1.829	0.935	11.371	0.078	3.911	0.689
	2.5 %	10.091	0.121	5.923	0.432	9.188	0.163	6.263	0.394
	1 %	3.891	0.691	1.743	0.942	12.777	0.047	3.276	0.774

Table 30: In-sample back testing with determinants

		RiskMetrics				GARCH			
		Long		*Short*		*Long*		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.011	0.918	1.700	0.192	0.709	0.400	0.865	0.352
	10 %	1.471	0.225	1.622	0.203	0.337	0.562	0.036	0.849
	5 %	4.826	0.028	0.698	0.404	0.343	0.558	0.412	0.521
	2.5 %	2.948	0.086	0.412	0.521	0.004	0.950	0.027	0.869
	1 %	9.321	0.002	5.775	0.016	0.193	0.660	0.659	0.417
DQT	25 %	3.957	0.683	6.108	0.411	4.187	0.651	5.480	0.484
	10 %	2.781	0.836	8.863	0.181	4.648	0.590	3.125	0.793
	5 %	11.295	0.080	3.812	0.702	13.939	0.030	3.949	0.684
	2.5 %	8.378	0.212	5.188	0.520	5.700	0.458	5.626	0.466
	1 %	15.361	0.018	8.768	0.187	5.168	0.522	1.956	0.924

		EGARCH				GJR			
		Long		*Short*		*Long*		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.517	0.472	1.026	0.311	0.854	0.355	1.390	0.238
	10 %	1.361	0.243	0.000	0.991	0.118	0.731	0.137	0.712
	5 %	0.522	0.470	0.005	0.941	0.114	0.736	0.315	0.575
	2.5 %	0.576	0.448	0.027	0.869	0.032	0.859	0.027	0.869
	1 %	0.390	0.532	0.644	0.422	1.003	0.317	0.390	0.532
DQT	25 %	3.956	0.683	7.971	0.240	4.688	0.584	4.778	0.573
	10 %	2.227	0.898	2.221	0.898	5.051	0.537	2.777	0.836
	5 %	6.467	0.373	1.785	0.938	9.117	0.167	3.942	0.684
	2.5 %	9.584	0.143	3.536	0.739	8.382	0.211	5.102	0.531
	1 %	1.743	0.942	8.458	0.206	4.899	0.557	1.743	0.942

		IGARCH			
		Long		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	0.264	0.607	1.390	0.238
	10 %	0.624	0.430	0.624	0.430
	5 %	8.913	0.003	0.252	0.615
	2.5 %	2.545	0.111	1.521	0.217
	1 %	10.745	0.001	5.775	0.016
DQT	25 %	5.181	0.521	5.943	0.430
	10 %	6.714	0.348	4.224	0.646
	5 %	19.446	0.003	2.460	0.873
	2.5 %	9.329	0.156	6.821	0.338
	1 %	18.433	0.005	8.768	0.187

Table 31: Out-of-sample forecasting without determinants

		RiskMetrics				GARCH			
		Long		*Short*		*Long*		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	13.430	0.000	0.044	0.833	7.694	0.006	0.794	0.373
	10 %	3.576	0.059	2.040	0.153	0.233	0.629	0.015	0.903
	5 %	0.354	0.552	1.126	0.289	4.140	0.042	1.690	0.194
	2.5 %	1.351	0.245	0.084	0.772	9.894	0.002	1.351	0.245
	1 %	0.954	0.329	0.325	0.568	5.754	0.016	1.988	0.159
DQT	25 %	18.460	0.005	2.886	0.823	12.163	0.058	2.943	0.816
	10 %	15.444	0.017	5.651	0.463	7.931	0.243	4.575	0.599
	5 %	1.337	0.970	3.451	0.751	7.215	0.301	10.203	0.116
	2.5 %	3.134	0.792	2.837	0.829	20.272	0.002	3.018	0.807
	1 %	1.434	0.964	0.589	0.997	15.194	0.019	3.163	0.788

		EGARCH				GJR			
		Long		*Short*		*Long*		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	6.397	0.011	1.128	0.288	7.694	0.006	0.794	0.373
	10 %	0.015	0.903	0.378	0.539	0.707	0.401	0.135	0.713
	5 %	2.192	0.139	2.192	0.139	4.945	0.026	1.690	0.194
	2.5 %	6.418	0.011	0.825	0.364	9.894	0.002	2.023	0.155
	1 %	1.988	0.159	0.954	0.329	5.754	0.016	1.988	0.159
DQT	25 %	9.222	0.161	3.036	0.804	10.493	0.105	3.271	0.774
	10 %	8.203	0.224	6.627	0.357	13.446	0.036	6.010	0.422
	5 %	4.633	0.592	11.820	0.066	8.351	0.213	10.319	0.112
	2.5 %	11.291	0.080	3.940	0.685	20.272	0.002	6.471	0.372
	1 %	3.191	0.784	1.399	0.966	15.194	0.019	3.163	0.788

		APARCH				IGARCH			
		Long		*Short*		*Long*		*Short*	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	5.990	0.014	1.128	0.288	7.694	0.006	0.649	0.420
	10 %	0.241	0.624	0.747	0.387	0.015	0.903	0.000	1.000
	5 %	2.766	0.096	4.140	0.042	4.945	0.026	2.766	0.096
	2.5 %	9.894	0.002	2.023	0.155	12.048	0.001	5.032	0.025
	1 %	3.529	0.060	1.988	0.159	9.027	0.003	3.529	0.060
DQT	25 %	10.059	0.122	3.879	0.693	13.046	0.042	3.757	0.709
	10 %	7.582	0.270	7.233	0.300	9.161	0.165	4.301	0.636
	5 %	5.501	0.481	13.846	0.031	7.546	0.273	7.707	0.260
	2.5 %	20.272	0.002	6.471	0.372	27.580	0.000	8.425	0.209
	1 %	6.840	0.336	3.163	0.788	42.314	0.000	6.819	0.338

Table 32: Out-of-sample forecasting with determinants

		RiskMetrics				GARCH			
		Long		Short		Long		Short	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	6.397	0.011	0.794	0.373	4.495	0.034	0.649	0.420
	10 %	0.015	0.903	0.000	1.000	0.000	1.000	0.015	0.903
	5 %	10.282	0.001	1.257	0.262	5.832	0.016	0.064	0.800
	2.5 %	3.849	0.050	3.849	0.050	6.418	0.011	0.435	0.510
	1 %	3.529	0.060	3.529	0.060	3.529	0.060	0.954	0.329
DQT	25 %	12.353	0.055	2.461	0.873	11.308	0.079	1.902	0.929
	10 %	6.621	0.357	4.466	0.614	10.047	0.123	6.083	0.414
	5 %	18.472	0.005	5.206	0.518	9.731	0.136	8.456	0.207
	2.5 %	10.662	0.099	6.358	0.384	11.291	0.080	4.826	0.566
	1 %	6.840	0.336	6.819	0.338	6.840	0.336	1.399	0.966

		EGARCH				GJR			
		Long		Short		Long		Short	
	Quantile	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	6.397	0.011	2.226	0.136	5.216	0.022	0.794	0.373
	10 %	0.363	0.547	1.877	0.171	0.132	0.717	0.060	0.807
	5 %	0.892	0.345	1.257	0.262	6.806	0.009	0.354	0.552
	2.5 %	2.851	0.091	0.825	0.364	6.418	0.011	0.825	0.364
	1 %	0.033	0.856	1.988	0.159	3.529	0.060	1.988	0.159
DQT	25 %	10.536	0.104	4.006	0.676	11.421	0.076	2.163	0.904
	10 %	13.173	0.040	7.835	0.250	7.092	0.312	7.223	0.301
	5 %	4.487	0.611	6.455	0.374	12.025	0.061	4.245	0.644
	2.5 %	4.940	0.552	7.057	0.316	11.291	0.080	6.094	0.413
	1 %	0.491	0.998	3.163	0.788	6.840	0.336	3.163	0.788

		IGARCH			
		Long		Short	
	Quantile	Statistic	P-value	Statistic	P-value
Kupiec LR	25 %	5.990	0.014	0.519	0.471
	10 %	0.135	0.713	0.015	0.903
	5 %	9.027	0.003	1.257	0.262
	2.5 %	8.029	0.005	6.418	0.011
	1 %	5.754	0.016	5.754	0.016
DQT	25 %	12.931	0.044	2.676	0.848
	10 %	11.519	0.074	4.490	0.611
	5 %	15.953	0.014	4.305	0.635
	2.5 %	15.075	0.020	11.222	0.082
	1 %	15.194	0.019	15.180	0.019

Appendix VII Monte Carlo Simulation

VII.1 Simulating Natural gas Returns and Volatility

The mean equation in the AR(1)-EGARCH(1,1) skewed student-t model is given by:

$$r_t = \mu + \theta_1 r_{t-1} + \theta_2 r_{S\&P,t-1} + \theta_2 r_{Oil,t-1} + \varepsilon_t \qquad \text{Eq. VII.1}$$
$$where \quad \varepsilon_t = z_t \sqrt{h_t} \quad z_t \mid I_{t-1} \sim D(0,1)$$

Using lag operators, the variance equation is given by:

$$\ln(h_t) = W_t + \left(1-\beta(L)\right)^{-1}\left[1+\alpha(L)\right] g(z_{t-1}) \qquad \text{Eq. VII.2}$$
$$where \quad g(z_t) = \gamma_1 z_t + \gamma_2 \left(\left|z_t\right| - E\left[\left| z_t \right|\right]\right)$$

which implies

$$\Rightarrow \ln(h_t)\left(1-\beta(L)\right) = W_t\left(1-\beta(L)\right) + \left[1+\alpha(L)\right] g(z_{t-1})$$
$$\Rightarrow \ln(h_t) - \beta(L)\ln(h_t) = W_t - \beta(L)W_t + \left[1+\alpha(L)\right] g(z_{t-1})$$
$$\Rightarrow \ln(h_t) = W_t + \beta(L)\left(\ln(h_t) - W_t\right) + \left[1+\alpha(L)\right] g(z_{t-1})$$
$$with \quad p = q = 1$$
$$\Rightarrow \boxed{\begin{array}{l} \ln(h_t) = W_t + \beta\left(\ln h_{t-1} - W_{t-1}\right) + g(z_{t-1}) + \alpha g(z_{t-2}) \\ \quad where \quad W_t = \omega + \ln\left(1 + \sum_{i=1}^{n} \delta_i X_{i,t}\right) \end{array}} \qquad \text{Eq. VII.3}$$

The last equation represents the determinants. When simulating without determinants these are simply excluded from the above equations. Together, the mean and variance equation described above constitute the basis for the natural gas volatility simulation in this paper. For each period a z-value is drawn from the student-t distribution to generate the innovation term for natural gas. In combination with last period's natural gas return and returns for Oil and S&P 500, this period's return is generated. The next period's variance is then calculated from the previous two z-values together with the variables calculated in the three following sections using the variance equation derived above.

Because the conditional variance model includes five variables in the variance equation we need to provide simulated forecasts for these variables as well. The dummies are however predetermined.

VII.2 Simulating Returns and Volatility for Oil and S&P 500

The oil and S&P 500 volatility estimates were calculated using vanilla GARCH models due to the presence of volatility clustering in the squared returns (see Appendix III). These models were in turn used to obtain volatility forecasts. Random z-values were drawn from the Gaussian distribution, and the following relationship was used to generate the innovation terms:

$$\varepsilon_t = z_t\sqrt{h_t} \qquad z_t \left| I_{t-1} \sim N(0,1)\right.$$

These innovation terms were used to calculate the daily oil and S&P 500 return from the mean equation of the conditional volatility model:

$$r_t = \mu + \varepsilon_t$$

Estimates for the volatility for day h_{t+1} were obtained by using the innovation term in the volatility equation:

$$h_{t+1} = \omega + \alpha\varepsilon_t^2 + \beta h_t$$

This procedure was repeated until estimates for the oil and S&P 500 volatility and returns for an entire year were produced.

VII.3 Simulating the Production and Stock Difference

Stock and production difference are not independent variables as they are dependent on what has been stored and produced in recent time. The dependencies on past information can be seen from Appendix III. Therefore, these variables were simulated using an AR(1)-process. First, the following regression were used to obtain estimates for the AR(1)-parameter and the regression residuals:

$$y_t = c + \theta y_{t-1} + \varepsilon_t$$

The distribution for the error terms was obtained by using the mean and standard deviation from the above regression in the Gaussian distribution. Randomly drawn z-values drawn from this distribution were used in the following equation:

$$\varepsilon_t = \mu + \sigma z_t \qquad z_t \mid I_{t-1} \sim N(0,1)$$

When these error terms were simulated, y_t could be calculated using equation $y_t = c + \theta y_{t-1} + \varepsilon_t$, which were repeated 253 times to simulate an entire year. After simulating the entire process, the absolute values of each y_t were stored for use in the natural gas volatility simulation.

VII.4 Simulating the Degree Days Difference

In order to simulate the degree day difference, an i.i.d. property was assumed due to the random nature of weather, although winter seems to exhibit larger differences in degree days (Appendix III). This implies that the variable can be drawn from an identical distribution each day without considering past information. Sample mean and standard deviation were estimated and randomly drawn z-values drawn from the Gaussian distribution were included in the following equation:

$$y_t = \mu + \sigma z_t \qquad z_t \sim N(0,1)$$

The squared value of y were then stored for use in the natural gas volatility simulation as a representation of the squared degree day difference.

Appendix VIII Half-life, Unconditional Volatility and Persistence

In GARCH(1,1) the long term volatility, or unconditional volatility, can be calculated as follows. From GARCH (1,1) we have:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \qquad \text{Eq. VIII.1}$$

Also, according to Alexander (2008a), the long term volatility is equal to the expected value of the conditional volatility at any lags l:

$$h_{LT} = E[h_t] = E[h_{t-l}]$$

For the error term, we can use the following expression:

$$E[\varepsilon_{t-1}^2] = Var(\varepsilon_{t-1}^2) - E[\varepsilon_{t-1}^2]^2 = h_{LT}$$

By inserting this into Eq. VIII.1, we get the following:

$$h_{LT} = \omega + \alpha h_{LT} + \beta h_{LT}$$

$$\Rightarrow h_{LT} = \frac{\omega}{1-\alpha-\beta} = \frac{\omega}{1-\phi} \qquad \text{Eq. VIII.2}$$

where ϕ is the persistence parameter. The half-life can be defined as:

$$\phi^H = 1/2$$

$$\Rightarrow H = \frac{\ln(1/2)}{\ln(\phi)} \qquad \text{Eq. VIII.3}$$

In the GJR(1,1) model, the result is almost the same as in GARCH(1,1), the only difference being that we must account for the expectance of the asymmetry coefficient. According to Taylor (2005) the asymmetry coefficient has an expectation of one half, and the persistence parameter is therefore defined as:

$$\phi = \alpha + 1/2\,\alpha^{-} + \beta \qquad \text{Eq. VIII.4}$$

When determinants are included, the persistence parameter will only change because the empirical values of α, α^{-} and β are different from the log-likelihood estimations. However, the expectations of the determinants are not directly included in the persistence parameter, in contrast to the unconditional volatility. The formal definition for the unconditional volatility when determinants are included in the GARCH(1,1) model is:

$$\begin{aligned} h_t &= \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \sum_i \beta_i X_i \\ h_{LT} &= \omega + \alpha h_{LT} + \beta h_{LT} + \sum_i \beta_i E[X_i] \\ \Rightarrow h_{LT} &= \frac{\omega + \sum_i \beta_i E[X_i]}{1-\alpha-\beta} = \frac{\omega + \sum_i \beta_i E[X_i]}{1-\phi} \end{aligned} \qquad \text{Eq. VIII.5}$$

where X_i represents the determinants. For the expectations of the oil and S&P 500 volatility we have used the long term volatility generated from their respective persistence parameters. For the remaining determinants the average value is used for the representation of the expected values.