University of Twente

EY Advisory

Incorporating Seasonality and Volatility Updating in Gas Storage Valuation for the Purpose of Validation

---

A thesis submitted in partial
fulfillment of the requirements

for the degree of

Master of Science
in
Financial Engineering and Management

by

N.A.J. Roelofs

July 2015

Incorporating Seasonality and Volatility Updating in Gas Storage Valuation for the Purpose of Validation

A master thesis submitted to the Faculty of Management and Governance of the University of Twente

by

N.A.J. Roelofs
July 2015

Supervisory committee

Dr. Berend Roorda
University of Twente

Ir. Henk Kroon
University of Twente

Floris van de Loo
EY Amsterdam

Rob Balk
EY Amsterdam

Abstract

Deregulation and growing demand for gas resulted in a global, transparent and growing gas market. This is followed by growing investments in gas storages. In recent years valuation techniques are developed to price these storages. In response to these valuations, prices have to be audited for regulatory purpose. This study supports audit work on the valuation of gas storages.

A valuation method often used is the spot approach of Boogert and De Jong (2008). This method uses simulated spot price paths based on an Ornstein-Uhlenbeck process and Least Squares Monte Carlo to value gas storages. By an econometric analysis on the spot price, existence of seasonality and volatility updating is shown. We present two techniques to incorporate these price characteristics in the spot approach. The objective of this study is therefore: Investigate and incorporate seasonality and volatility updating in gas storage valuation for the purpose of validation.

The first technique relates to the equilibrium level parameter of the Ornstein-Uhlenbeck process. We take seasonality into account by smoothing a forward curve with daily granularity to form a time-dependent equilibrium level. The second technique is the incorporation of volatility updating in the simulation of future gas prices. By doing so we take account for the presence of volatility clustering in gas spot prices. We simulate volatility updating by a GARCH model.

The impact of both techniques on the distribution of values of the Least Squares Monte Carlo method is shown. Incorporating forward curve information to represent seasonality result in slightly higher values but gain is obtained by the fact that the equilibrium level is in line with market expectations. Using a GARCH model to incorporate volatility updating results in much higher values. This is because more extreme price movements are present in the price paths simulated by a GARCH model. These extreme price movements are profitable for gas storage holders. We also show the impact of these two extensions on the 'bad' left tail of the distribution of values. We conclude that the 5^{th} percentile point, to represent the left tail, is stable from a mean reverting rate of 0.02 and higher. A lower rate will cause a much fatter left tail which corresponds to more risk.

We conclude that the two presented techniques to extend the spot approach are appropriate for the purpose of model validation. If these techniques are used for having a claim on the one and 'true' value, they should be handled with care. This is because of the underlying assumptions and the impact of the mean reverting rate as shown by our sensitivity analysis.

Preface

Writing this thesis, conducting an internship and being a student was a very rich experience. I would never forget the wonderful time as a student. I will certainly bring the things I have learned, the joy I have had and the friends I have made with me for the rest of my life.

As for this master assignment I would like to thank all my future colleagues of the FS Risk department of EY. They were all very helpful and made my time as an intern interesting and fun. I got the opportunity to work on some side-assignments as well and was allowed to participate in every team-meeting, even in the ski-trip to Austria. A special thanks to Rob Balk, one of my EY supervisors. He was my counsellor and was always available for questions or a discussion.

I am also very thankful for the valuable feedback of Berend Roorda and Henk Kroon, my first and second UT supervisor. Besides supervising my master assignment I would like to thank both of them for the lectures they provided and the advice they gave me. It was very pleasant and helpful that, as a student, you can always just walk by for any questions.

Last but certainly not least, I want thank Rosa, my family and friends for their support and for being close to me.

Acronyms

ANOVA	Analysis of variance
ADF	Augmented Dickey-Fuller test
EMH	Efficient-market hypothesis
ES	Expected shortfall
GARCH	generalized autoregressive conditional heteroskedasticity
GBM	Geometric Brownian motion
GOU	Geometric Ornstein Uhlenbeck
KPSS	Kwiatowski-Phillips-Schmidt-Shin test
LSMC	Least Squares Monte Carlo
LNOU	Log-Normal Ornstein-Uhlenbeck
MC	Monte Carlo
OU	Ornstein-Uhlenbeck
PP	Phillips-Perron test
Sd	Standard deviation

Contents

1 Introduction

Firstly, a quick overview of the gas market is given. Second, the rationale of gas storages and validation is explained. Furthermore, the methods currently used to value gas storages are presented. At last, the outline and the sub-questions of this study are pointed out.

1.1 Quick overview of the current gas market

Natural gas is a hot topic in national and international political and economic news. Gas extraction at one of Europe's biggest gas fields (Groningen, the Netherlands) causes a lot of discussion. Proponents of this extraction highlight the positive impact on the Dutch economy, whereas opponents focus on damaged houses as a result of the extraction. Meanwhile, Russia is continuously threatening gas-dependent Europe to cut off its supply of natural gas to enforce political power. On the other side of the planet, the US are reshaping their energy-import dependent economy into a self-sustaining one through the development of shale gas. Currently, they are exporting energy in the form of Liquefied Natural Gas. These examples show the gas market is changing continuously.

Gas market conditions are also influenced by deregulation in the US and Europe. Under this pressure, the natural gas storage service is unbundled from the sales and infrastructure service. This creates a totally separated service of gas storage, making it possible to adjust storage trading decisions to price conditions (Boogert & De Jong, 2008). Similarly, the EU aims to integrate the European energy market in order to provide customers with more choice and competitive pricing (GasTerra, 2014). The total demand for natural gas is growing each year according to the International Energy Agency (2011). It estimates the demand for gas will grow more than 50% by 2035, providing over 25% of world energy consumption. Cumulative investments in this sector amounts to around 8 trillion dollars.

These developments result in a global, transparent, and growing gas market and require appropriate valuation techniques to analyse investments. One form of investment in the gas sector is gas storages. These storages create economic value for its owner by benefitting from changes in price, generally caused by seasonal effects.

1.2 The rationale of gas storages and validation

Traditionally, the existence of gas storages can be linked to the winter-summer spread of gas prices. During the year production capacity is constant, whereas demand fluctuates due to the heating of houses in the winter. Storage value is created by injecting gas during the summer and selling gas in the winter, when demand and prices are high. To clarify, the potential of these gas storages to benefit from changes in gas prices is valued, not the current gas in storage.

Besides the annual seasonality in natural gas prices, other characteristics can be found from the field of commodity pricing in general (Back & Prokopczuk, 2013; Bessembinder, Coughenour, Seguin, & Smoller, 1995; Pindyck, 2001; E. Schwartz & Smith, 2000). Other characteristics mentioned by these authors are mean-reversion, jumps or price spikes and volatility clustering.

The economic intuition behind mean reversion can be explained by the law of supply and demand made famous by Adam Smith in his book The wealth of Nations (Smith & Nicholson, 1887).

If prices exceeds marginal costs, new investments and new producers will enter the market, which leads to higher supply over time. The higher supply, and constant demand, will push prices back down, intensify competition, and will shrink margins. Producers will leave the unprofitable market, and this process continues. The characteristic of jumps or price spikes is related to the existence of sudden shortages in the market while the commodity is irreplaceable for its users in the short-term. In this scenario buyers have to pay the highly increased price for only a very short period.

Gas storage holders take these characteristics into account to determine their operating strategy. Here, the operating strategy means for a gas storage holder how much gas to inject/buy or withdrawal/sell at each point in time and at each volume level to maximize its storage value. This means the valuation of gas storages itself is influenced by these gas price characteristics.

Now the rationale for gas storages is known, there is also a rationale for validation. Owners of gas storages need to value their gas storages for accountancy purposes. This value is for example presented at the asset-side of the company's balance sheet. In turn, this balance sheet (together with other financial statements) needs, for regulatory purposes, to be checked by an assurance firm. So, the rationale of gas storages also creates from regulatory perspective a rationale for its value validation by accountants.

Accountants do in general not have the same sector related knowledge as their clients since it is not their core business. The validation of asset prices is therefore done in two ways. The first possibility is to only validate the model used by their clients to find the asset-value. Another, more extensive, way is to find the asset-value themselves and to check if the corresponding clients-price is reasonable. This study will attribute to the ability of accountants to validate the gas storage value in both ways. First, it investigates the need to incorporate seasonality and volatility updating in gas storage valuation so the clients model can be challenged. Secondly, this study shows how seasonality and volatility updating can be incorporated in gas storage valuation to determine whether the client's value is reasonable. Finding an own value is influenced by a lot of uncertainty and assumptions so the outcome should be handled with care.

Storage valuation and therefore validation is not restricted to the gas market. It plays an important role as well in other commodity markets. Take for instance the electricity, oil and other soft commodity markets. The principle here is that there should exist a reasonable liquid spot market and the spot price should exhibit mean-reversion.

1.3 Methods to price gas storages

In the liberalized and more transparent gas market, the objective of gas storage owners is to maximize the value of the gas storage. In practice four methods are used to price a gas storage (Breslin, Clewlow, Elbert, Kwok, & Strickland, 2008):

- The Intrinsic approach
- The rolling intrinsic approach
- The basket of spread options
- The spot approach

These four methods all uses market information on gas-contracts to find the value of the possibility to store natural gas. For example, spot prices, forward prices and gas option prices are used.

The intrinsic approach

The intrinsic approach derives a value from seasonal or time spread in the price of gas. It assumes the value of the storage is given by the optimal set of long and short positions of forward and/or future [1] contracts over the period of the storage. This initial position is entered at the first day and held to maturity. Since this position is fixed it can be seen as a static hedging strategy.

The rolling intrinsic approach

This approach adjusts the position in long and short forward contracts over time when additional value can be created. It recognises the changing value in the intrinsic spread as the forward curve evolves. Over time unprofitable positions are liquidated and new contracts are entered to lock in higher overall value. How much additional money can be created depends on the movements of the forward curve and especially on switches in the curve (Boogert & de Jong, 2011).

[1] In commodity pricing literature the terms 'forward' and 'futures' are used interchangeably. Bloomberg uses the term forward contracts for the OTC market and futures for the ICE market. Despite the difference in naming they both represent a contract between two parties to buy/sell an asset at a specified future time at a price agreed upon today. For the remainder of this paper the term 'forward' is used.

The basket of spread options

A third approach is the basket of spread options. A spread option is an option on the difference between two prices with a positive strike price (Lai, Margot, & Secomandi, 2010). This strategy looks at the storage as a long position in an optimal basket of calendar spread options. A Monte Carlo simulation can be used to calculate the expected value and to obtain a distribution of values. The rolling intrinsic and the basket of spread approaches capture additional value when prices evolve. However, they are still suboptimal, since they take no account of potential future trades (Breslin et al., 2008).

The spot approach

In this study the spot approach is chosen. This approach creates value by trading only in the spot. This technique regards storage as an American-style option with constraints determined by the storage characteristics like inject/withdrawal rate and capacity (Gray & Khandelwal, 2004). Gas spot prices are highly volatile and exhibits mean-reversion (Boogert & de Jong, 2011).

Focus of this paper

The spot optimisation approach is the focus of this study because it is the most flexible method to deal with the complex decisions a storage holder has to make. Besides, according to Maragos and Ronn (2002) the spot approach generally results in higher returns compared to the rolling intrinsic method. The spot approach is therefore often used in the field of gas storage valuation. In general two perspectives on the spot approach can be separated: stochastic control and Monte Carlo. In the perspective of stochastic control there exist a direct link between the price process and the optimal operating strategy. In this study the perspective of Monte Carlo is chosen because the two are separated here. This means we can experiment with different price processes. Another advantage is that Monte Carlo is able to deal with complex gas storage characteristics.

When following the Monte Carlo perspective the spot approach is divided into two parts: pricing and optimizing. The pricing part deals with the price dynamics and the simulation of future spot prices. The optimizing part determines the optimal operating strategy given these price paths. In Figure 1 the research structure is presented in which the pricing and optimizing part are clearly separated.

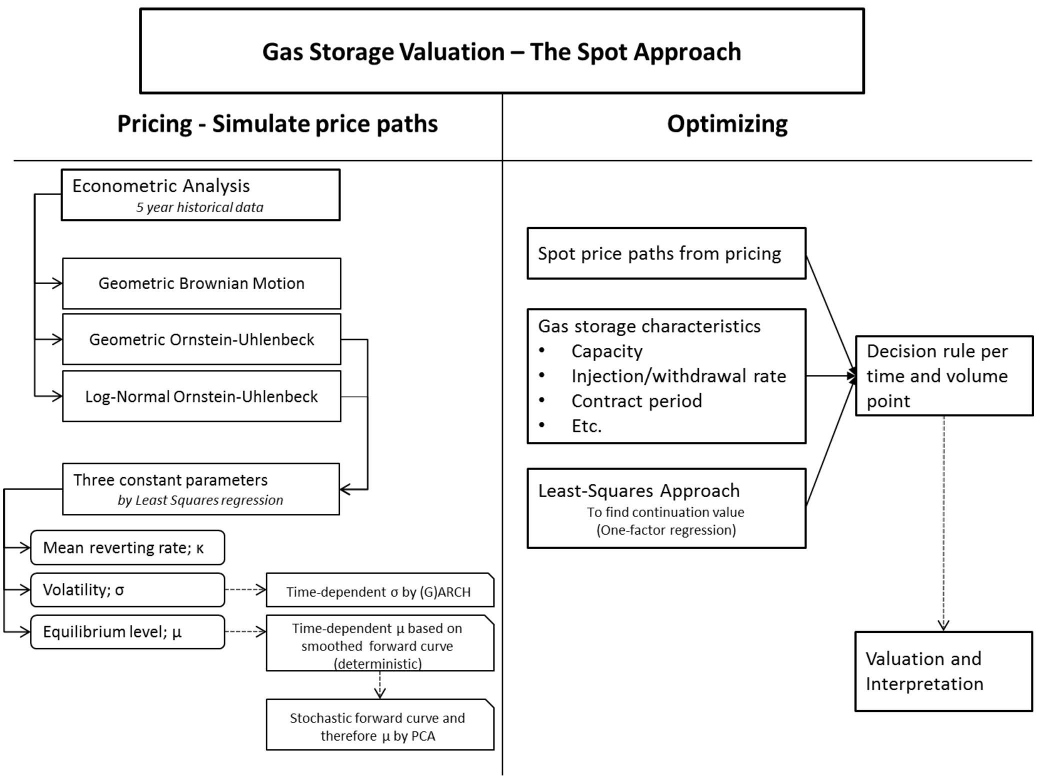

FIGURE 1: Research Structure

In the remainder of this study we will point out what the focus of each (sub) section is and how it relates to the overall research structure.

1.4 Research Outline

In this study the spot approach methodology presented by Boogert and De Jong (2008) to value gas is extended to take into account seasonality and volatility updating of the underlying gas spot price. The option pricing technique of Least Squares Monte Carlo (LSCM) of Longstaff and Schwartz (2001) is adapted. The LSMC method is particularly suited because it can handle complex constraints of the physical storage and it separates the pricing process of the underlying asset from the optimizing part of operating decisions. The perspective of validation is chosen to support audit work on energy companies in general.

The objective of this research is as follows: Investigate and incorporate seasonality and volatility updating in gas storage valuation using the spot approach for the purpose of validation. So the main research question is: How to incorporate seasonality and volatility updating in gas storage valuation for the purpose of validation?

The research is divided into three sub-questions:

1. How can spot price paths be simulated?
    a. How to describe the behaviour of gas spot prices?
    b. How can existing stochastic processes be used?
    c. How to incorporate seasonality in the simulation of gas spot prices?
    d. How to incorporate volatility updating in the simulation of gas spot prices?
2. What is the principle to find the optimal operating strategy for a gas storage using the spot approach?
3. How to obtain a value and interpret the value of gas storages using the spot approach?

As can be seen from the questions above, the separating of the spot approach into a pricing and optimizing part is also present in the sub-questions. The first sub-question covers the pricing part. The second sub-question handles the optimizing part. The third sub-question uses the answers on the first two sub-questions to find a value of gas storages and a way to interpret this value. An overview of the research structure is given in appendix A.

The remainder of this paper is organized as follows: Chapter 2 discusses the context of this study. Chapter 3 provides theory related to the research questions. Chapter 4 uses this theory to answer the sub-questions. Chapter 5 addresses a sensitivity analysis. Chapter 6 concludes and presents suggestions for further research.

2 Context Analysis

In this chapter the context is analysed in which this research is done. First, the research context is discussed. Sections 2.2 and 2.3 outlines the assumptions which are made to execute and demarcate this study. Section 2.4 explains which datasets are used in this study. The ethical perspective is discussed in Section 2.5.

2.1 Research Context

In this section we will point out the context of each sub-question and how it relates to the purpose of validation. After that, the pricing model and the corresponding parameters are stated. At last, a distribution of gas storage value(s) is presented which acts as a baseline.

The focus of this study is to incorporate seasonality and volatility updating in the price process which relates to sub-question 1 and show the effect of these pricing characteristics on the distribution of values in sub-question 3. The second sub-question deals with the algorithm of Boogert and De Jong (2008). This algorithm finds the operating strategy of the gas storage and a gas storage value per price path. This algorithm is programmed in the software environment of R. The corresponding code is presented in Appendix D. In this way it is determined whether or not seasonality and volatility updating should be taken into account in the validation of gas storage value(s).

To answer sub-question 1 a pricing model is needed to generate price paths. In this study the pricing model is an Ornstein-Uhlenbeck process. The log spot price follows (E. S. Schwartz, 1997):

$$dX = \kappa\left[\mu - \frac{\sigma^2}{2\kappa} - X\right]dt + \sigma dz \tag{1}$$

Here, X represents the log spot price, μ the equilibrium level, κ the mean reverting rate, σ the volatility, and *dz* the increment of a Wiener process. Hence, the equilibrium level, volatility and mean reverting rate are the three input parameters of this pricing model. By an econometric analysis on the gas spot price, two price characteristics are shown that relate to two of these parameters.

The first price characteristic is about seasonality in gas spot prices. A time-dependent equilibrium level is implemented in the Ornstein-Uhlenbeck process to represent seasonality. This means that the equilibrium level parameter μ, is different per time step in the simulation of price paths. This equilibrium level will in general be higher in winter-periods and lower in summer-periods due to the traditionally winter-summer spread. The starting point of this gas price characteristic is to construct a smoothed forward curve with daily granularity from monthly forward contracts. After that, a day-week profile is added to the smoothed curve. At last, an attempt is made to simulate future forward curves to construct a stochastic equilibrium level.

The second price characteristics is about volatility updating or clustering in gas spot prices. In the methodology of Boogert and De Jong (2008) volatility is treated as a constant parameter whereas this study shows that volatility is not stationary. This implies that the second moment evolution of a price process should not be neglected. This means that in the above price model the volatility parameter is set by a volatility model so that the volatility per price step is different. Volatility clustering will be implemented in the simulation of future time series by the use of a generalized autoregressive conditional heteroskedasticity model (GARCH).

The price paths from the pricing part are an input of the optimizing algorithm in sub-question 2. This algorithm generates a gas storage value per price path. Sub-question 3 uses these values per price path to construct a distribution of values. To further explain this last step, a baseline distribution generated with all constant input parameters is presented in Figure 2. The corresponding input parameters are given in Table 1 and the interpretation of the distribution is stated in Table 2.

Mean reverting rate; k	0.0137
Equilibrium price; μ	56.8330
Volatility; σ	3,45%
Storage value	32.12 million

TABLE 1: Gas Storage Input Parameters and Value

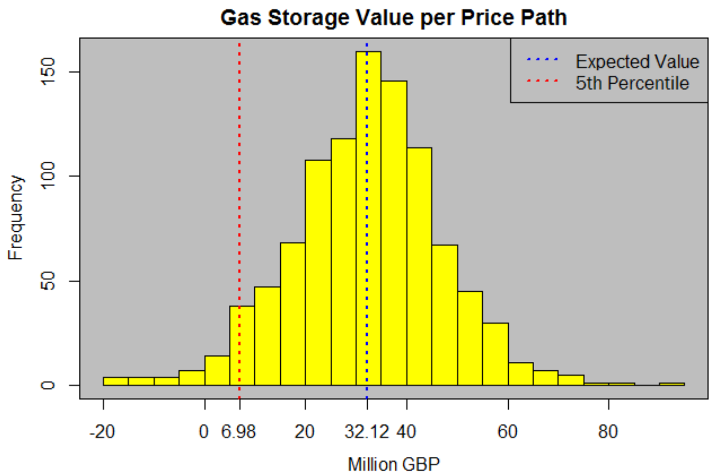

FIGURE 2: Gas Storage Value Distribution with Constant Parameters

5th Percentile	6.98 million
Expected Shortfall	-0.64 million
Standard deviation	14.76 million

TABLE 2: Interpretation of Value Distribution

In the figure above, a histogram of the value per price path is presented. In the remainder of this study the effect of changing the equilibrium level and volatility parameters on this histogram or distribution of values is shown. Since we use risk-neutral pricing the probabilities in Figure 2 are risk-neutral probabilities, not to be confused with real-world probabilities. Moreover, we follow the fundamental theorem of asset pricing which implies that the gas storage value is the expected value of discounted future payoffs under the risk-neutral measure. These assumptions are further discussed in Section 2.2.

2.2 Gas Storage Valuation and Option Theory

The valuation of gas storages can be looked upon from an option valuation perspective. In a very simple form, there sometimes exists an analytical, closed form solution to value an option. Under the Black-Scholes assumptions (Black & Scholes, 1973), this solution delivers the 'fair' price of the option with respect to the various input parameters (Richardson, 2009a). However, not all required assumptions are satisfied here to find a closed form solution. Firstly, markets are not complete. Secondly, a gas storage is as a very complex and path-dependent option. Hence, we have to make some simplifying assumptions. These assumptions are explained below.

Fundamental theorem of asset pricing

In this study the fundamental theory of asset pricing is followed which plays an important role in the modern theory of mathematical finance (Björk, 2004). The theorem provides two conditions which are pointed out shortly. The first condition is that markets are free of arbitrage. An arbitrage opportunity is a zero-cost strategy which has a nonnegative pay-off in all states and at least one positive pay-off in at least one state. The second condition is that markets are complete. This indicates that every portfolio of assets can be hedged perfectly at every point in time with a portfolio consisting of different assets. Under these assumptions and the risk-neutral measure, a derivative's price is the expected value of discounted future payoffs.

Risk-Neutral valuation

Now the two conditions of the fundamental theory of asset pricing are assumed we can use risk-neutral valuation. This valuation technique is widely used in quantitative finance to compute derivative prices. Usually, probabilities on events are expressed in terms of the so-called "real world" probabilities. However, computing a price of a financial asset requires discounting future cash flows. The problem here is that each financial asset should be discounted by its associated risk-profile. This requires an adjustment which is different per risk profile. Using risk-neutral valuation, all investments would return the risk-free interest rate. This means that to find the present value an investment all cash flows should be discounted by the risk-free interest rate.

In the commodity markets, risk-neutral valuation implies that the expected future spot price is represented by the forward curve (E. S. Schwartz, 1997).

$$F(0,T) = E_0[S(T)] \tag{2}$$

Here, F represents the price of a forward contract at time t=0 for delivery at t=T, and $E_0[S(T)]$ represents the expectation of the spot price at time t=0 for t=T. From this it can be concluded that the risk-free interest rate is already taken into account by forward prices. When we relate this to our pricing model, the equilibrium level parameter of the Ornstein-Uhlenbeck process should be set according to the forward curve.

The present value of any cash flow can be obtained by discounting its expected value at the risk-free rate. The risk-neutral valuation method simplifies the analysis of derivatives considerably. Now the expected return of each investment is the risk-free interest rate, *r.* By risk-neutral valuation the real world probabilities are adjusted given a certain model and the corresponding real world 'prices' so that the return on investment is the risk-free interest rate. It cannot be emphasized enough that risk-neutral probabilities are therefore no claim on real world probabilities. Besides, the assumptions underlying risk-neutral valuation do not hold in the real world. Risks cannot be hedged away in full and markets are not complete.

Monte Carlo Simulation

A closed form solution cannot be found because gas storages are path-dependent. We therefore have to use a numerical method to approximate the value. The chosen numerical method is Monte Carlo simulation. This is a process by which a large number of price paths are generated that evolve according to a particular stochastic process. The option payoff for each path is determined and the 'fair' value is calculated by taking the average of these 'fair' values.

Note that other numerical methods such as finite difference and binominal trees could also be used for approximating the value of gas storages. However, they are not as flexible to deal with the large complexity of gas storages (Felix & Weber, 2008). Besides, the price process is separated from the optimizing part so it is possible to experiment easily with different stochastic price processes. This is why in the remainder of this study Monte Carlo simulation is used to approximate the value of gas storages. A disadvantage of Monte Carlo Simulation is that it is relatively time-consuming.

Since risk-neutral pricing is used, the average of the simulated price paths at each point in time should approximate the equilibrium level. This equilibrium level is, in turn, set by the forward curve in which the risk-free interest rate is taken into account.

Efficient-Market Hypothesis

Another assumption is to follow the efficient-market hypothesis. In finance, the assumption that the market always incorporate and reflect all relevant information is called the efficient-market hypothesis (EMH). Related to the gas market, all information about for example weather expectations, demand shifts, and expectations about development of new technologies are already incorporated in the gas spot and forward prices. Consequently and in line with the fundamental theorem of asset pricing, an investor cannot consistently achieve returns in excess of average market returns on a risk-adjusted basis because assets always trade at their fair value. Following this theory, the only way an investor can possibly obtain higher expected returns is by taking more risk. The efficient market hypothesis is seen as a cornerstone of modern finance but it is also controversial and often disputed by practitioners. It is assumed in this thesis that markets are efficient. Together with the EMH it is also assumed that the market will not be influenced by our trades.

2.3 Data

This sub-section explains how the data used in this study are obtained. It also points out the restrictions associated with the data.

In the pricing part, market data is used to calibrate parameters of the stochastic processes and to analyse the natural gas market. Three datasets are retrieved from Bloomberg (2014) on three different markets:

- The US Henry Hub
- The UK NBP
- The Dutch TTF

These three markets differ in terms of geographic area (affects seasonal patterns), liquidity, and availability of data. Each dataset consist of the daily last traded spot price over the years 2010-2014. Data about available forward or future contracts are also retrieved from Bloomberg but only for a period of one year: 2014.

The period covered by these data sets, five years for spot prices and one year for forward contracts, are similar to periods used in comparable studies (Bjerksund, Stensland, & Vagstad, 2011; Zhao & van Wijnbergen, 2013). At a 95% confidence level, no significant difference is found between the datasets on the 2001-2014 period and the 2010-2014 period by performing an analysis of variance. The five and one year periods are therefore considered to be appropriate and reasonable.

In section 4 the day-week profile of gas spot prices is analysed. In order to find the gas prices related to the corresponding weekdays the following conversion is applied. Five years of day ahead and weekend ahead prices are retrieved from Bloomberg (2014). Here the day ahead price on Monday indicates the spot price of Tuesday, the day ahead price on Tuesday indicates the spot price of Wednesday and so on. However, the day ahead price on Friday stands for the spot price of the Monday of the next week. The spot price of Saturday and Sunday is given by the weekend ahead price on Friday and are therefore the same.

This study demonstrates its methodology on the UK NBP gas market because it is the most liquid gas market in Europe. Additional techniques to enable the methodology to work for the other gas markets are presented when needed. For example, the correlation between NBP forward contracts and TTF forward contracts is analysed in Section 4.1.3. Natural gas trades at the UK NBP market in Pence per therm whereas at the other two markets other units are used. We therefore consider the unit of gas as irrelevant for the remainder of this study. This is because the purpose of this study is to present techniques that can be used for validation, not the actual pricing of gas storages.

2.4 Ethical perspective

Investing in the commodity sector is for the most part uncontroversial. Under the condition that these investments are in companies or services that create value for customers and society. This is because commodity prices are in the long run driven by investments in production technology which lead to stable or declining real prices over time (Rohrbein, 2011). Nevertheless, there is controversy about speculation on commodity prices. A debate is going on whether or not this can be seen as ethical.

Historically, commodity contracts like futures and forwards are used by producers to manage the volatility of earnings in the future. This makes it easier to manage operations and plan investments. Nowadays this is still the case, but there are also other motives. Asset managers use these commodity contracts to gain exposure to commodities for three reasons. Diversification benefits and protection against inflation are seen as legit whereas pure speculation is seen as unethical and is therefore the focus of the debate. It is stated that speculation causes higher volatility and higher commodity prices in general to the disadvantage of import dependent third-world countries (Rohrbein, 2011). An argument in defence of speculators is that they do not take physical delivery of commodities and can therefore not affect the physical market. At the other side, the argument is stated that speculation definitely moves prices making physical delivery more dependent on the, very volatile, spot market.

This debate is particularly relevant for agricultural commodities. However, energy commodities also influences food prices by their effect on production and shipping costs. Besides, oil and gas prices will have an impact on everyone's purchasing power. This makes the ethical discussion also relevant for the gas market: The pursuit of profit in the gas sector versus the purchasing power of customers. One can say that investing in gas storages is unethical because investors tend to benefit from ordinary customers. On the other hand, it is argued that by investing in gas storages the spread between winter and summer gas prices will become smaller.

In this study the above ethical discussion is taken into account. Nevertheless, no point of view is chosen in this discussion. This is because the purpose of this study is not to support holders of gas storages to maximise their profit. The purpose is to validate gas storage value(s) and thereby support audit work.

3 Literature Study

This section discusses the literature per research question. Section 3.1 introduces and discusses the theory needed for the pricing part. This includes spot price analysis, potential stochastic processes, the analysis of the forward curve, and generalized autoregressive conditional heteroskedasticity (GARCH). Section 3.2 states the theory used for finding the optimal operating strategy, the optimizing part. Also, an example is given to demonstrate the corresponding algorithm. At last, section 3.3 describes the theory used for valuing gas storages and how to interpret this value.

Literature presented here per sub-section corresponds to the same sub-section in Section 4. In this way, relevant theory can directly be linked to its application.

3.1 Simulating Future Gas Prices – Pricing

In this sub-section all the relevant theory needed for sub-question 1 is presented, i.e.:

Sub-question 1: How to simulate gas spot price paths?

Theory to analyse and describe gas spot prices is presented first. Secondly, three stochastic models to simulate gas spot prices are introduced. After that, methods to analyse the forward curve are mentioned. At last, theory to incorporate volatility updating in the simulation of gas spot prices is discussed. In the figure below the focus of this section is encircled. The outcome on this sub-question will be the general input for the optimizing part.

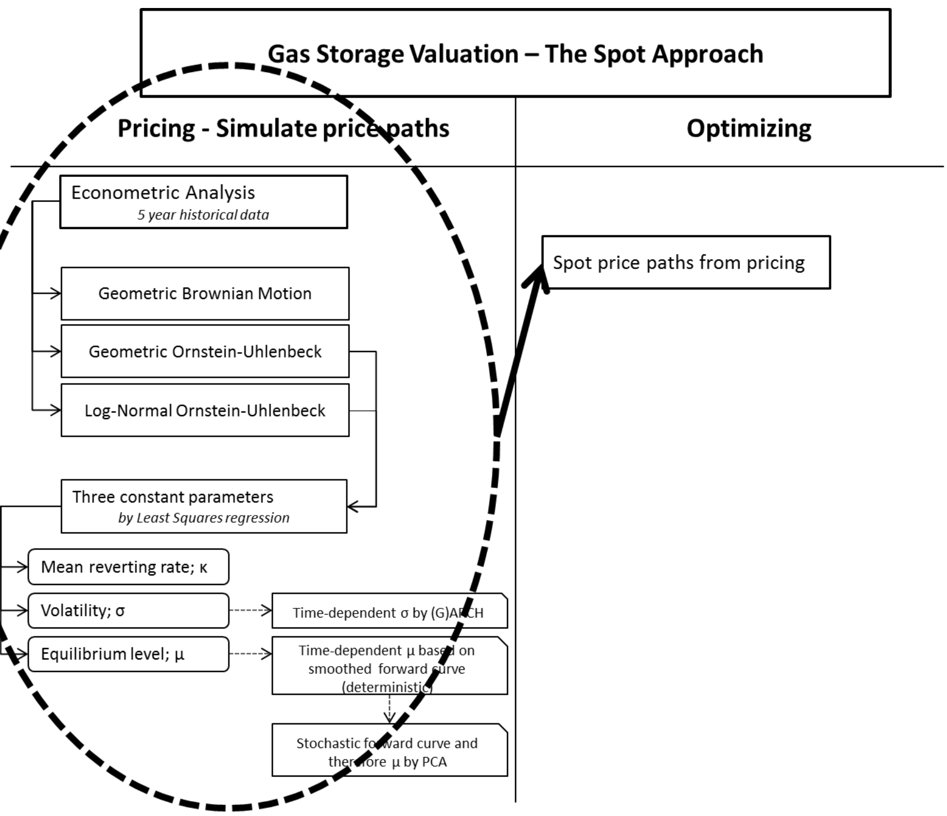

FIGURE 3: From Overall Research Structure; Section 3.1

3.1.1 Spot Price Analysis

This section relates to the econometric analysis. The focus is encircled in Figure 4.

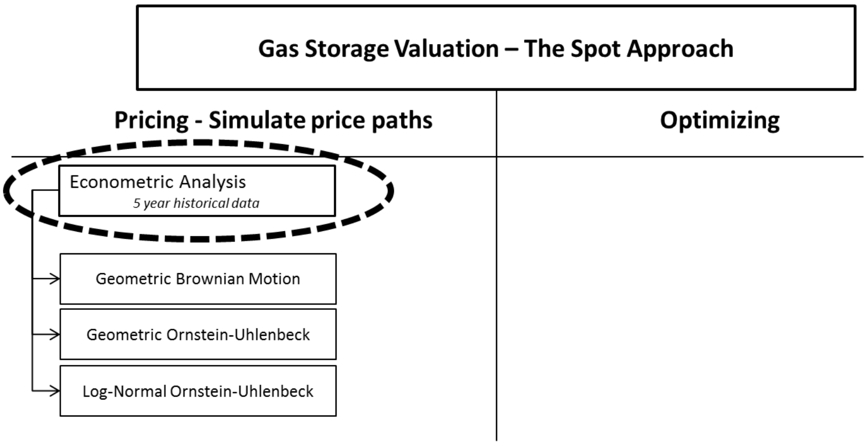

FIGURE 4: From Overall Research Structure; Section 3.1.1

Descriptive Statistics

For analysing the spot price and its behaviour, descriptive statistics are provided of 5 year historical data for the three datasets: UK NBP, US HH, and Dutch TTF. This is done so general understanding of gas spot price is obtained. The spot price process itself, together with the 'simple' return and the log returns are visualized. Visualization of the NBP gas spot price and its returns are given in section 4.1.1. The HH and TTF are annualized in respectively Appendix B and C.

Traditionally the 'simple' returns are denoted with a capital R and calculated as follows:

$$R_t = \frac{(S_t - S_{t-1})}{S_{t-1}} = \frac{S_t}{S_{t-1}} - 1 \qquad (3)$$

In the equation above S_t represents the price at time *t* and S_{t-1} the price at time *t-1.*

In finance the use of log returns is popular because it is often assumed that asset prices follow a lognormal distribution. An advantage of this assumption is that a lognormal distribution has a lower bound of zero and that the distribution is skewed to the right so it has a long right tail. Another advantage is that it takes account for continuous compounding. Log returns are denoted with a lowercase r and calculated as follows:

$$r_t = \log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1}) \qquad (4)$$

Again, S_t represents the price at time *t* and S_{t-1} the price at time *t-1*. Besides, log indicates here the natural logarithm.

Below it is explained how the skewness and kurtosis are calculated.

Skewness is a measure of the degree of asymmetry of a distribution. The skewness of a normal distribution is zero. A positive skewness indicates the variable is skewed to the right meaning the right tail is longer as compared to the left tail (Kholopova, 2006). The skewness is calculated as follows:

$$Skewness\ (S) = \frac{1}{(n-1) * \sigma^3} \sum_{k=1}^{n} (S_k - Mean)^3 \qquad (5)$$

Here, n represents the number of data-points, σ the standard deviation, and S_k the kth data-point.

Kurtosis is a measure of the ‘fatness’ of tails of the probability distribution. It shows whether the distribution is peaked or flat relative to the normal distribution. The kurtosis of a normal distribution is three. High kurtosis indicates a peak near the mean and a heavy tail (Kholopova, 2006). The kurtosis is calculated as follows:

$$Kurtosis\ (K) = \frac{1}{(n-1) * \sigma^4} \sum_{k=1}^{n} (S_k - Mean)^4 \quad (6)$$

Quantile-Quantile plots

An appropriate way to verify the skewness and the kurtosis according its ability to detect normality is to construct a Q-Q plot. This quantile-quantile (QQ) plot is a graphical technique to determine whether or not two data sets come from populations with a common distribution. To test for normality, the normal QQ plot graphically compares the distribution of a given variable to the normal distribution which is represented by a straight line (Kholopova, 2006). It plots the quantiles of the first data set against the quantiles of the fitted normally distributed data set. In section 4.1.1 the QQ plot of the simple returns and the log returns of gas spot prices are given.

To complement a QQ plot the Jarque-Bera and Shapiro-Wilk statistical tests are conducted to test for normality.

The Jarque-Bera test is a goodness-of-fit test that uses the above Skewness and Kurtosis to test if the data are sample drawn from a normal population. The null hypothesis of normality, e.g. skweness equal to zero and skewness equal to three, is tested. The Jarque-bera test statistic is calculated as follows (Jarque, 2011):

$$JB = \frac{n}{6}\left[S^2 + \frac{1}{4}(K-3)^2\right] \quad (7)$$

In addition to the Jarque-Bera test, the Shapiro-Wilk test is performed. This test also utilizes the null hypothesis to check whether the sample is drawn from a normal population. This test uses ordered data (from small to large) to calculate the test statistic. The corresponding equation is as follows (Shapiro & Wilk, 1965):

$$W = \frac{\left(\sum_{i=1}^{n} \alpha_i x_i\right)^2}{\sum_{i=1}^{n}(x_i - Mean)} \quad (8)$$

Here, x_i is the i-th smallest number in the sample and the constants α_i are given by

$$(\alpha_1, \ldots., \alpha_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \quad (9)$$

Where, in turn,

$$m = (m_1, \ldots, m_n)^T \quad (10)$$

m_i are the expected values of the order statistics of independent and identically distributed random variables samples from a standard normal distribution, V is the covariance matrix of those order statistics. Again, the null hypothesis is rejected when the statistics are below a certain threshold or p- value.

Seasonality

The existence of seasonality in the gas spot prices and in commodity prices in general is extensively described in commodity pricing literature (Back & Prokopczuk, 2013; Back, Prokopczuk, & Rudolf, 2013; Bernard, Khalaf, Kichian, & McMahon, 2008; Mirantes, Poblacion, & Serna, 2013). At first, there seems to be some sort of winter-summer seasonality present in gas spot prices. To prove the existence of seasonality the data points are separated by months and analysed for significant differences. This is done in section 4.1.1. A way to test for significant differences over the year is to perform an Analysis of Variance (ANOVA) of the data-points per month over the five-year period. These 12 datasets are visualized by boxplots and in addition the mutual relationships are tested by using the Tukey's Honest Significant Difference test. This test is used to find means that are significantly different from each other (Tukey, 1949).

Volatility clustering

Findings in finance and especially the commodity pricing literature show that the second moment evolution of a price process should not be neglected. Pindyck (2003) was one of the first to predict commodity prices and volatility with GARCH (Generalized Autoregressive conditional Heteroskedascity) models to incorporate clustering in volatility. In response to Pindyck (2003) several authors found that a GARCH model outperforms more sophisticated models in predicting volatility of commodity returns (Bates, 2003; Hansen & Lunde, 2005; Sadorsky, 2006). These models are originally introduced in the eighties of the last century (Bollerslev, 1986; Engle, 1982).

Before implementing a GARCH model, the existence of volatility clustering is shown by a unit root tests and a stationarity test: respectively the Phillips-Perron (PP) test the Augmented Dickey-Fuller (DF) test, and the Kwiatowski-Phillips-Schmidt-Shin (KPSS) test. The PP and DF tests have as null hypothesis the presence of a unit root whereas the KPSS test has as null hypothesis a stationary process. The KPSS test is used to complement the first test and should give similar results.

Both tests are programmed in the statistical software and environment of *R* based on Hamilton (1994), and Zivot (2006). The results are given in section 4.1.1.

Besides testing for a unit root the DF statistical test also tests for no mean reversion. As null hypothesis the DF has a mean reverting process. If the p-value is below the critical threshold for the spot price, it cannot be concluded that there is no mean reversion in the spot price. The same can be done for the return process of the spot price (Kim & Park, 2013).

3.1.2 Stochastic Processes

In this sub-section three stochastic processes are introduced which can be used in the Monte Carlo simulation of future price paths. These price paths are in turn used as an input in the optimizing part of this study. This can be seen from the research structure in Appendix A. First, a stochastic process itself is introduced. The three stochastic processes are as follows:

- The Geometric Brownian Motion (GBM)
- The Geometric Ornstein-Uhlenbeck (GOU)
- The Log-Normal Ornstein-Uhlenbeck (LNOU)

A variable whose value changes over time in an uncertain way can be represented by a stochastic process. Stochastic processes are often used to represent the evolution of some system of random values over time. In the field of mathematical finance these models are used to simulate prices of for example shares, bonds, and commodities. In many cases, the starting point of the process is known and many directions are possible in which the process may evolve.

The objective of this section is to provide stochastic price processes that can be used for the simulation of future gas prices. Ideally, the stochastic process represents the characteristics of gas spot prices known from literature and our econometric analysis. However, these characteristics may vary over time and may be commodity specific (Back & Prokopczuk, 2013).

In this section the three stochastic processes are described as a price model for gas spot prices. At first, the corresponding parameters of these processes are treated as constants. Later, these models are extended by implementing time-varying equilibrium level and volatility updating.

The decision which stochastic process to use in the simulation of gas spot prices depends on the econometric analysis of the gas spot prices over five year of historical data together with the possibilities to incorporate gas price characteristics like time-varying equilibrium level and volatility updating. In Figure 5 the focus of this section in relation to the overall research structure is given.

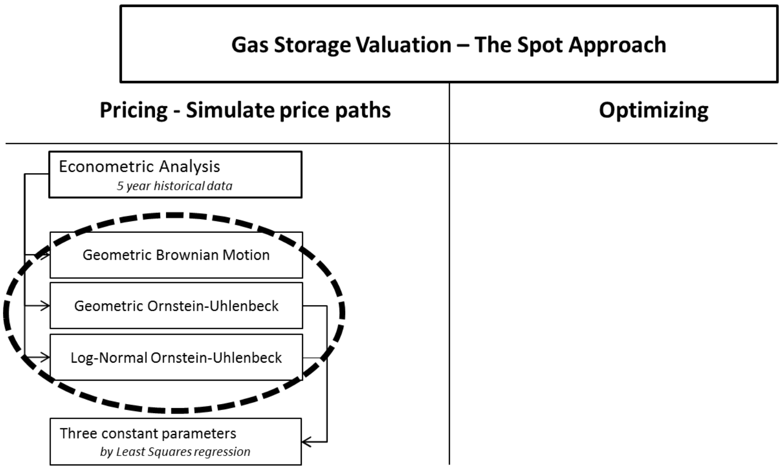

FIGURE 5: From Overall Research Structure; Section 3.1.2

Geometric Brownian Motion

This is a standard stochastic process often assumed for asset prices where the logarithm of the underlying follows a generalized Wiener process (J. C. Hull, 2006, p. 270). The process is defined by the following stochastic differential equation:

$$dS = \mu S dt + \sigma S dz \qquad (11)$$

Here μ represents the drift, dt the change in time, σ the volatility, and dz the increment of a Wiener process

The conditional distribution of S at time t is (Dixit & Pindyck, 1994):

$$E[S_{t+w}|S_t] = S_t e^{\mu w} \qquad (14)$$

$$Var(S_{t+w}|S_t) = S_t^2 \, e^{2\mu w}(e^{\sigma^2 w} - 1) \qquad (15)$$

Following the last condition, the variance of the process increases over time without bound.

Geometric Ornstein-Uhlenbeck

The Ornstein-Uhlenbeck stochastic process is used a lot in the field of commodity pricing since the process tends to drift towards an equilibrium level, taking account for mean-reverting behaviour of many commodity prices (Bessembinder et al., 1995). The standard Ornstein-Uhlenbeck process (Uhlenbeck & Ornstein, 1930) is:

$$dS = k[\mu - S]dt + \sigma dz \qquad (16)$$

Where $k \geq 0$ measures the speed of mean reversion, μ is the equilibrium price level, σ the volatility, and dz is the increment of a Wiener process. This process is stationary, Gaussian and Markovian.

Many variations of this stochastic process are since then developed to make the volatility price-dependent. The following process is suggested by Dixit and Pindyck (1994) and is known as the Geometric Ornstein-Uhlenbeck process.

$$dS = k[\mu - S]dt + \sigma S dz \qquad (17)$$

Following this process, the percentage change per time unit is normally distributed.

The conditional distribution is as follows:

$$E[S_{t+w}|S_t] = \mu + e^{-kw}(S_t - \mu) \qquad (19)$$

$$\mathrm{Var}(S_{t+w}|S_t) = \frac{\sigma^2 S_t}{2k}(1 - e^{-2kw}) \tag{20}$$

Log-Normal Ornstein-Uhlenbeck

This process is introduced by E. S. Schwartz (1997) and is often used in commodity pricing literature. It describes a stochastic mean reverting process to a constant mean. The difference with the Geometric Ornstein-Uhlenbeck process is that ito's lemma is applied to the commodity spot price. The (natural) log price is characterized by an Ornstein-Uhlenbeck stochastic process instead of the spot price itself. The commodity spot price follows the stochastic process:

$$dS = \kappa[\mu - \ln S]\, Sdt + \sigma S dz \tag{21}$$

Applying Ito's Lemma: $X(T) = \ln S(T)$ and $S(T) = exp\{X(T)\}$ the log price follows an Ornstein-Uhlenbeck stochastic process:

$$dX = \kappa[\alpha - X]dt + \sigma dz \tag{22}$$

Where

$$\alpha = \mu - \frac{\sigma^2}{2k} \tag{23}$$

Again, the magnitude of the speed of adjustment is represented by $k > 0$ to the long rung mean log price, α. σ is the volatility, and dz is the increment of a Wiener process. The conditional distribution of X at time T under the equivalent martingale measure is normal:

$$E_0[X(T)] = e^{-kT}X(0) + (1 - e^{-kT})\alpha^* \tag{24}$$

$$\mathrm{Var}_0[X(T)] = \frac{\sigma^2}{2k}(1 - e^{-2kT}) \tag{25}$$

Here, α^*= α-λ and λ is the market price of risk.

Following this process, the first difference of the log price is normally distributed instead of the percentage change in the Geometric Ornstein-Uhlenbeck process.

3.1.3 Forward Curve Analysis

It might be confusing to analyse the forward curve whereas the spot approach is the focus of this study. Nevertheless, the forward curve gives valuable information about the expected value of future spot prices. For example, the equilibrium level of an Ornstein-Uhlenbeck process to simulate spot prices can be determined by its forward curve instead of calibration on historical data. In this way, the equilibrium level is an implied parameter and consistent with market data.

First, general understanding how forward curve information can be used, together with risk-neutral valuation is explained. After that, smoothing of market curves is discussed because commodity forward contracts are only settled against an average spot price over the settlement period. At last, the theory of Principal Component Analysis (PCA) is stated. PCA is applied in order to simulate future forward curves and potentially make the equilibrium level stochastic.

From the overall research structure the focus of this section is encircled in the figure below. Smoothing the forward curve at the valuation date is done to create a deterministic equilibrium level. After that, an attempt is made to make the equilibrium level stochastic given one year of forward curve data.

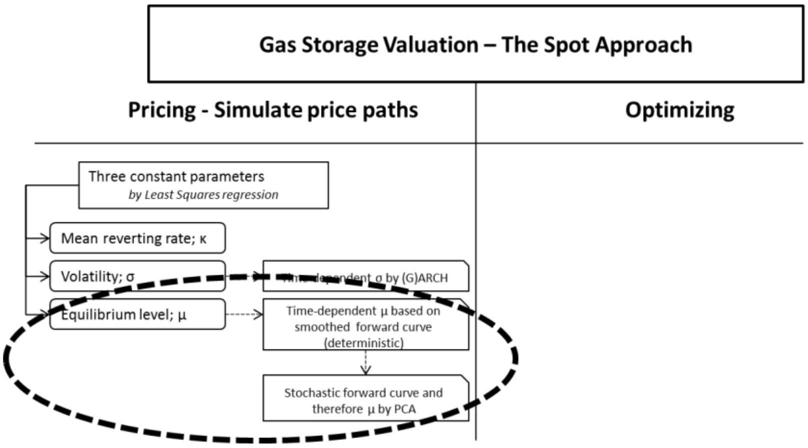

FIGURE 6: From Overall Research Structure; Section 3.1.3

Forward Contracts and Risk-Neutral Valuation

In section 3.1.2 the spot price representation is given of the Geometric Brownian Motion, Geometric Ornstein-Uhlenbeck and the Log-Normal Ornstein-Uhlenbeck to simulate future spot prices. To incorporate seasonality in these simulations forward curve information is used. Because the last two processes are mean reverting, forward contracts are used to determine a time-dependent equilibrium level. Since the Geometric Brownian Motion is not mean reverting this it is not relevant for this section. The dynamics of the spot price process of both Ornstein-Ulenbeck processes are under the assumption of risk-neutral valuation. Following this approach current forward prices represent the expected future spot price (E. S. Schwartz, 1997).

For the Geometric Ornstein-Uhlenbeck process forward prices can be linked directly to future spot prices.

$$F(0,T) = E_0[S(T)] \tag{26}$$

And this will set the time-dependent equilibrium level at time T:

$$\mu(T) = F(0,T) = E_0[S(T)] \tag{27}$$

Additional steps are needed to set the time-dependent equilibrium level for the Log-Normal Ornstein-Uhlenbeck process. This is because it considers the price to be log normal distributed. From the properties of the log-normal distribution, we have:

$$F(0,T) = E_0[S(T)] \tag{28}$$

$$F(0,T) = E_0[exp\{X(T)\} \tag{29}$$

$$F(0,T) = exp\left\{E_0[X(T)] + \frac{1}{2}var_0[X(T)]\right\} \tag{30}$$

$$\mu(T) = F(0,T) = exp\left\{E_0[X(T)] + \frac{1}{2}var_0[X(T)]\right\} \tag{31}$$

Information on forward and future contracts can be retrieved from data sources like Bloomberg. In commodity markets typical forward contracts are often settled against an average spot price during the settlement period. The delivery rate in the settlement period is constant and the contracts are called average-based forwards. These average contracts will create a very blocky forward curve with jumps between contract periods. Using this forward curve as the prediction of future spot prices is not in line with the smoothed behaviour of these prices. To overcome this problem, the forward curve known from market should be made smoothed and still be consistent with market information. By doing so, a forward curve with daily granularity is created. This process is known from electricity pricing literature (Koekebakker & Ollmar, 2005) and visualized in Figure 7.

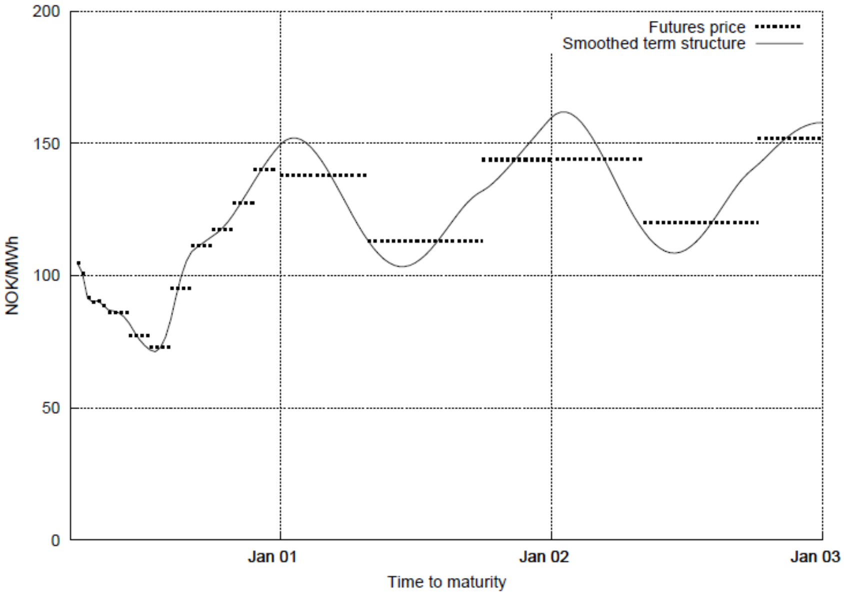

Figure 7: Smoothed Forward curve

Smoothing the Curve

The smoothing of the forward curve can be done in several ways and in different levels. One of the most commonly known theories to do this is the "moving average" principle. Other methods are for example the "Kalman filter", "exponential smoothing" and the fitting of polynomials to the data. The consideration which method is most appropriate is considered out of scope for this study. To create a smooth forward curve from average-based commodity contracts the method to fit a spline is used as discussed by Benth, Koekkebakker, and Ollmar (2007).

Smoothing splines provide a flexible way of estimating the underlying regression function (Tibs, 2014). A spline of order k is a piecewise polynomial function of degree k, that is continuous and has continuous derivatives of order 1,...k-1, at its knot points. Formally, a function f is a kth order spline with knot points at $t_1 < \cdots < t_m$ if

- f is a polynomial of degree k on each of the intervals (-∞, t_1], $[t_1, t_2]$, ... $[t_m, \infty)$, and
- $f^{(j)}$, the jth derivative of f, is continuous at $t_1, \dots t_m$, for each $j = 0,1 \dots k-1$.

The continuity in all of the lower order derivatives makes splines very smooth, often the location of the knots cannot be detected visually.

This study focuses on cubic smoothing splines, k=3. Functions of the form $\sum_{j=1}^{n} \beta_j g_j$ where $g_1, \dots g_n$are the truncated power basis functions for natural cubic splines with knots at $x_1, \dots x_n$. Specifically, the coefficients are chosen to minimize

$$(y - G\beta)_2^2 + \lambda\beta^T \Omega\beta \qquad (32)$$

Where G is the basis function defined as:

$$G_{i,j} = g_j(x_i), \;\; i,j = 1, \dots n, \qquad (33)$$

And Ω is the penalty function:

$$\Omega_{i,j} = \int g_i''(t) g_j''(t) dt, \quad i,j = 1, \dots n, \qquad (34)$$

Given the optimal coefficients β^, the smoothing spline estimate at x is:

$$r^{\wedge}\ (x) = \sum_{j=1}^{n} \beta^{\wedge}{}_j g_j(x) \tag{35}$$

The regularization term $\lambda\beta^T \Omega\beta$ has a tuning parameter, the smoothing parameter: λ. The higher the value of λ, the more shrinkage of the spline. There are two 'philosophical' approaches to choose the smoothing parameter. The first approach is a subjective choice. The second one is an automatic method, so the parameter is chosen by data. This method is called cross-validation (Wu, 2004). In this study the first approach is chosen because the automatic method is considered out of scope. Calculations related to this section are performed in the software environment of R.

Forward Markets with Less Liquidity

As stated in Section 2.4 about the availability of data, markets differ in the diversity of forward contracts they offer. For example there are 24 monthly forward contracts available on the US Henry Hub whereas the Dutch TTF market only quotes 4 monthly contracts. The most liquid natural gas market in Europe is the NBP in the UK. The International Commodity Exchange (ICE) provides quotes for 12 monthly forward contracts on this market.

For simulation purposes as explained in 3.1.2, prices of monthly forward contracts are needed to represent the summer-winter seasonality in the simulation of future spot prices. The minimal number of forward contracts needed to represent this seasonality is considered to be 12. In that case, 8 monthly forward contracts are lacking of the TTF market. A method to overcome this problem is to construct monthly forward contracts out of quarterly and seasonally contracts. If there are still monthly contracts missing a ratio analysis is applied. An approximation of the ratios between TTF monthly forward contracts and quarterly contract is found from the UK NBP market because these markets seems to be highly correlated. In Section 4.1.3 the correlation between the Dutch TTF market and the UK NBP market is examined. After determining the correlation between these markets in Chapter 4, the method is further discussed.

Principal Component Analysis

Principal Component Analysis (PCA) is a data reduction technique and is concerned with the identification of structure within a set of interrelated variables. Its aim is to determine the relevant factors or principal components which explains as much of the total variation in the data as possible (Koekebakker & Ollmar, 2005). PCA is a widely used method for simplifying complex data structures. Its application can be found in many fields, such as simulation and image analysis. Instead of attempting to describe 100% of the variance within a dataset, the idea is to filter out the most important factors and use them to simulate the market (Bjerksund et al., 2011).

For the purpose of this study, PCA is used to simulate forward curves conditional on the forward curve of the previous day. By simulating forward curves the equilibrium level parameter which is set by forward prices is made stochastic. Results are presented and discussed in Section 4.1.3. In this section it is concluded that the added value of a stochastic equilibrium level parameter is very low for the purpose of validation. Nevertheless the PCA on forward curves is presented for future research.

From a more mathematical perspective the origins of PCA lie in multivariate data analysis. PCA has been called one of the most important results from applied linear algebra and its most common use is as the first step in analysing large data sets (Richardson, 2009b). PCA uses a vector space transform to reduce the dimensionality of data sets. In this way, by using mathematical projection, the original set of data can often be interpreted in just a few variables, the so called principal components.

In this study, the application of PCA to gas forward curves is in line with the study of Koekebakker and Ollmar (2005). From this paper four different steps are separated:

1. Construct for every day a forward curve with daily granularity
2. Find a return function for each day as a function of time to maturity
3. Perform principal component analysis to find typical curve movements

4. Determine the volatility functions from the factor loadings in the PCA

The first step about smoothing the curve is already discussed in Section 3.1.3, the application of this theory is stated in Section 4.1.3. In step two a time homogenous model is assumed to obtain manageable input estimations:

$$\sigma_i(t,T) = \sigma_i(T-t) \tag{36}$$

This means that the loadings from the PCA depends only on time to delivery. The return function depends on the stochastic process that chosen to simulate gas prices. When the Log-Normal Ornstein-Uhlenbeck stochastic process (E. S. Schwartz, 1997) is followed the return function is as follows:

$$ln[curve2(t1)/curve1(t2)] \tag{37}$$

This means that the first day on a curve is compared to the second day of the curve of the previous day, corresponding to the same 'delivery' date. This return function is applied to all curves and to maturities (one year of forward curves with each 365 'delivery' points). Step three consists of the actual execution of the PCA. This is done by calculating the "eigenvalues" and corresponding "eigenvectors" of the covariance matrix and to sort them in decreasing order. Market practice, and frequently used in finance literature is to use the number of factors that are needed to explain 95% of the variance. At last, the volatility functions follow from the factor loadings of the eigenvectors.

To perform the simulations of the forward curves the following equation is applied (Bjerksund et al., 2011):

$$F(t+\Delta t,T) = F(t,T)exp\left[\sum_{i=1}^{N}(\sigma_i(T-t)\sqrt{\Delta t} * \varepsilon_i - \frac{1}{2}\sigma_i^2(T-t) * \Delta t)\right] \tag{38}$$

This is a discrete time representation because it is used to actually simulate the forward curve at time $t+\Delta t$ on the forward curve at time *t*. Furthermore, *F(t,T)* represents the forward price at time *t* for delivery at time *T*, ε_i are i independent standard normal distributed numbers, and $\sigma_i(T-t)$ are the factor loadings.

3.1.4 GARCH

The generalized autoregressive conditional heteroskedasticity (GARCH) technique is already mentioned shortly in Section 3.1.1. The general principle is to include the second moment evolution of a price process in the prediction of future prices. Whereas Section 3.1.1 focuses on the need for a GARCH model, its background and the way to implement it to a stochastic process is discussed here. The relation of this section to the overall research structure is presented in Figure 8.

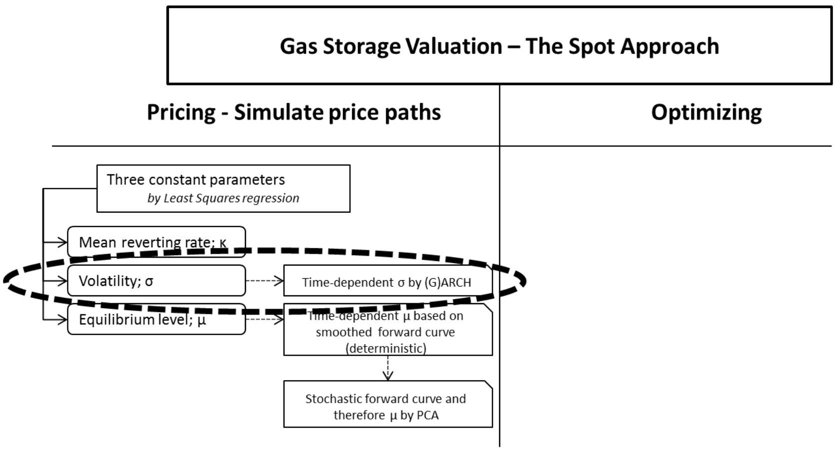

FIGURE 8: From Overall Research Structure; Section 3.1.4

At first, together with the other parameters, the volatility parameter is treated as a constant. This sub-section will present theory to incorporate volatility updating in the simulation of gas spot prices. Price path simulations using this model are presented in Chapter 4.

One of the founders of the volatility-updating principle is Engle (1982). He introduced an autoregressive conditional heteroskedasticity (ARCH) model of asset price changes. Four years later, Bollerslev (1986) developed upon this ARCH model a generalized version: GARCH. Following this generalization, the conditional variance is modelled as an autoregressive moving average (ARMA) process. The basic form of a GARCH (p,q) model is:

$$S_t = E[S_t|S_{t-1}] + \sigma_t^2 \in_t \tag{39}$$

Here, the price *S* at time *t* is based on the expectation of price *S*, conditional on the price at *t-1*, the volatility σ, and error term ∈.

The conditional expectation term can be represented by any stochastic process, including the Geometric Brownian Motion, Geometric Ornstein-Uhlenbeck process, and the Log-Normal Geometric Ornstein-Uhlenbeck process.
In the GARCH model the error part is $\in_t \overset{iid.}{\sim} N(0,1)$ and

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{40}$$

In the above equation, α_i represents q autoregressive terms and β_i represent p moving average terms.

The GARCH (1,1) model, so with one autoregressive term and one moving average term, is often used as a volatility model. Hansen and Lunde (2005) compare this volatility model with 330 other volatility models in their extensive study. They conclude as follows: “we cannot reject that none of the competing models are better than the GARCH (1,1) model.” Determining which volatility model is most suited to simulate future gas prices is considered out of scope for this study. In this study, the volatility is therefore modelled following the GARCH (1,1) model.

As stated in Section 2.3, we use 5 year of historical data to calibrate the parameters of the Ornstein-Uhlenbeck process and the parameters of the GARCH volatility model. The parameters of the OU process are calibrated by the use of linear regression. The parameters of the GARCH volatility model are found by a maximum likelihood analysis. At a 95% confidence level, no significant difference is found between the datasets on the 2001-2014 period and the 2010-2014 period by performing an analysis of variance. We consider therefore this dataset as appropriate.

3.2 Finding the optimal operating strategy – Optimizing

In this sub-section all the relevant theory needed for sub-question 2 is stated, i.e.:

Sub-question 2: What is the principle to find the optimal operating strategy for a gas storage using the spot approach?

First, the Least Square Monte Carlo Method is introduced. This method is used to value American options by using regression techniques. Secondly, the LSMC method is implemented in a dynamic programming problem. By working backwards, the continuation value of each decision at each point in time can be found. At last, an example is given to clarify the corresponding algorithm.

From the overall research structure this section is related to the optimizing part. The pricing part at the left is an input of finding the optimal operating strategy. In turn, the result of this sub-question is an input of sub-question three: valuation and interpretation. The focus of this section and its relationship to the overall structure is presented in Figure 9.

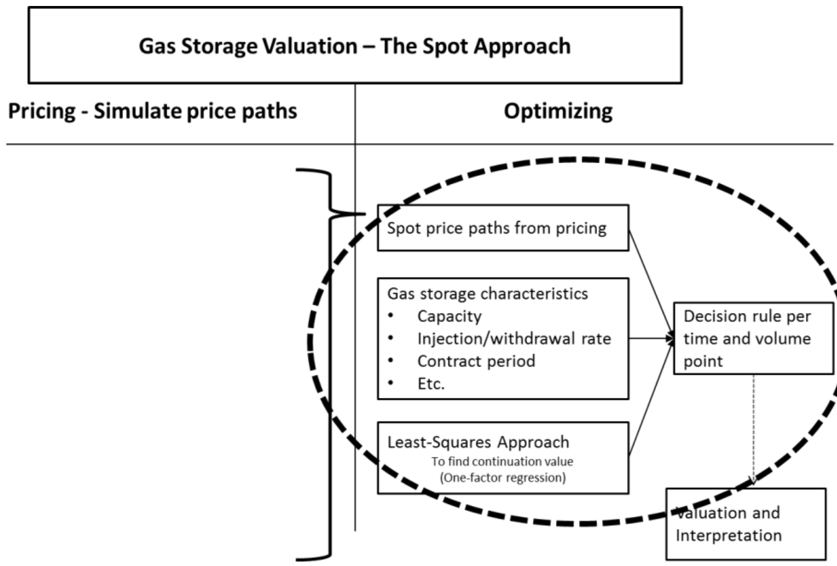

FIGURE 9: From Overall Research Structure; Section 3.2

3.2.1 Least Square Monte Carlo Method

Gas storage valuation can be looked upon from an option valuation perspective. At each point in time the storage holder basically has the option to exercise the option (withdrawal gas, inject gas or do nothing). A starting point is to set the option as a simple European call/put option. In this way a closed form solution can easily be derived using the classical option pricing model of Black and Scholes (1973). However, a European option does not account for path-dependency and does not come near the real flexibility of a storage. It has only one decision at maturity versus multiple at each time point. This makes a gas storage similar to an American option, or in discrete time, a Bermuda option (Zhao & van Wijnbergen, 2013).

Now the storage valuation presents itself as a dynamic programming problem to find the decision rule at each point in time: inject, do nothing, or withdraw. This choice depends on the immediate payoff and the continuation value of each decision. The continuation value in turn, depends on decisions made in the future. Consequently, this problem can only be solved working backwards.

Because the flexibility of a storage contract is too complex to have a closed form solution, advanced numerical techniques are required to solve the dynamic programming problem. Such a technique is the Least Square Monte Carlo method (Longstaff & Schwartz, 2001). This is a simple yet powerful approach for approximating by simulation the value of American and exotic options. The key of this approach is to use least squares regression to estimate the conditional expected payoff from continuation. The Longstaff and Schwartz (2001) method estimates the conditional expected option value by simulating a lot of price paths and carrying out a regression analysis on the resulting option values. This results in an approximation for the continuation values of the different decisions at each point in time.

The LSMC method is in general a relatively simple and easy technique to price complex options but it also has some critics. One of the biggest critics is that it is difficult to assess exactly how accurate the method is (Johnson, 2012). Other critiques are that the corresponding algorithm depends on the number of sample paths, the number of basis functions for the regression, the type of basis functions and the number of observation dates. Here, the number of sample paths and the number of basis functions are restricted to computational time and power. Besides, the method's performance is mixed and can often incur unknown approximation errors (Johnson, 2012). Also, the various polynomial fits and number of basis functions have different effects on performance (Moreno & Navas, 2003). Nevertheless, the method will provide good estimates for option values. Estimates are considered appropriate for this study because the perspective of validation is chosen.

3.2.2 Optimizing the Operation of the Gas Storage

At this section the storage contract and its characteristics are presented analytically from the perspective of the holder of the contract. The holder faces multiple decisions at every point in time. The value of the storage is the expected value of the accumulated future payoffs when following the optimal strategy. The algorithm to find these optimizing decisions is also given (Boogert & De Jong, 2008). This algorithm is programmed in the programming environment of R. The corresponding code is presented in Appendix D.

The contract is signed at time $t = 0$ and settled at time $t = T + 1$. Each day the storage holder faces the decision to inject, do nothing, or withdraw gas, under volumetric limitations.

The accumulated volume in storage at the start of day t is denoted by $v(t)$:

$$v(t) = v(0) + \sum_{i=1}^{t} \Delta(i-1) \tag{41}$$

The corresponding payoff at day t is $h(S(t), \Delta v(t))$ for $t = 0, \dots, T$:

$$h\big(S(t), \Delta v(t)\big) = \begin{cases} -c\big(S(t)\big)\Delta v(t) & inject\ at\ day\ t \\ 0 & do\ nothing\ at\ day\ t \\ -p\big(S(t)\big)\Delta v(t) & withdraw\ at\ day\ t \end{cases} \tag{42}$$

Here $S(t)$ represent the gas spot price at time t, $\Delta v(t)$ the change in volume level at time t, $c\big(S(t)\big)$ the cost of injection, and $p\big(S(t)\big)$ the profit of withdrawal. In Section 2.3 the market is assumed complete and frictionless. So in line with this assumption there are no transaction costs and the bid-ask spreads zero.

The volume metric and injection/withdrawal limitations are presented as follows:

$$\begin{aligned} v^{min} \leq v(t) \leq v^{max} \\ i^{min} \leq \Delta v(t) \leq i^{max} \end{aligned} \tag{43}$$

The scope of this research is to treat these limitations as constants whereas in reality they may depend on time and current volume. $i^{min}$is usually negative and i^{max} is usually positive.

As stated above, the value of the storage is the expected value of the future payoffs when following the optimal strategy π:

$$\sup_{\pi} \boldsymbol{E}\left[\sum_{t=0}^{T} e^{-\delta t} h\big(S(t), \Delta v(t)\big) + \ e^{-\delta(T+1)} q(S(T+1), v(T+1))\right] \tag{44}$$

This expectation is assumed to be under a risk-neutral measure. Here, $q(S(T+1), v(T+1))$ represents a penalty function which depend on (the lack of) remaining gas in storage at T+1. $e^{-\delta t}$ the discount-rate, and as stated above, $h(S(t), \Delta v(t))$ is the payoff at day t.

The continuation value, the value that is assigned to the contract after taking a specific action Δv, at time t, volume level v and spot price s, is defined as follows:

$$C(t, S(t), v(t), \Delta v) = \boldsymbol{E}\left[e^{-\delta} U(t+1, S(t+1), v(t) + \Delta v)\right] \quad \textbf{(45)}$$

Were $U(t, S(t), v(t))$ represents the following dynamic program:

$$U(T+1, S(T+1), \mathrm{v}(\mathrm{T}+1)) = \mathrm{q}(\mathrm{S}(\mathrm{T}+1), \mathrm{v}(\mathrm{T}+1)) \quad (46)$$

$$U(t, S(t), v(t)) = max\{h(S(t), \Delta v), C(t, S(t), v(t), \Delta v)\} \quad (47)$$

To find the solution for this dynamic problem regression-based simulation will be used. The above introduced Least Squares Monte Carlo method is well suited for this part because it can handle a variety of constraints. This regression must be calculated for every point in time, discrete volume level, and for each price path; $b = 1, \dots, M$. mathematically, this results in the following estimation:

$$C^b(t, S^b(t), v(t+1) \approx e^{-\delta} \, Y^b(t+1, S^b(t+1), v(t+1)) \quad (48)$$

Where $Y^b(t+1, S^b(t+1), v(t+1)$ denotes the accumulated value of future realized cash flows in path b following optimal decisions being at volume level *v(t+1)* and price $S^b(t+1)$.

All of the above can be summarized by the following pricing algorithm (Boogert & De Jong, 2008):

1. *Simulate M independent price paths* $S^b(1), \dots, S^b(T+1)$ *for* $b = 1 \dots M$ *starting at given* $S(0)$
2. *Assign a value to the contract at maturity, e.g. a penalty for (lack of) remaining gas in storage*
3. *Apply backward induction for* $t = T, \dots, 1$ *for each t, step over N allowed volume levels* $v(t; n)$
    a) *Run an OLS regression to find an approximation of the continuation value*
    b) *Combine the different continuation values into a decision rule*
    c) *Implement the decision rule to calculate the accumulated future cash flows*
4. *The storage value is the average accumulated future cash flow over all price paths*

3.2.3 An Example

At this section an example for step 2 – 4 of the algorithm above is given. Step 1 relates to the simulation of price paths (Section 3.1). As stated in the section before, the dynamic problem is solved by working backwards. To visualize this process a volume-time grid is presented when needed.

Step 2. Assign a value to the contract at maturity

Remaining (or lack of) gas in a storage represents value. To stimulate the ending volume level to be in line with the starting volume level a penalty function is introduced. This penalty is for example the last gas price times 1.1 per volume point below the starting volume level and the last gas price times 0.91 (1/1.1) per volume point above the starting volume level. At the figure below this step is visualized for an optimizing process of discretized volume points: $v = 0, \dots, 4$ discretized volume change: $\Delta v = -1, 0, 1$ and T=4. The P represents a penalty at maturity. The starting volume is 2 since no penalty is assigned to that volume level. At the right side, the price paths (b=1,..,5) are included creating a 3D-grid.

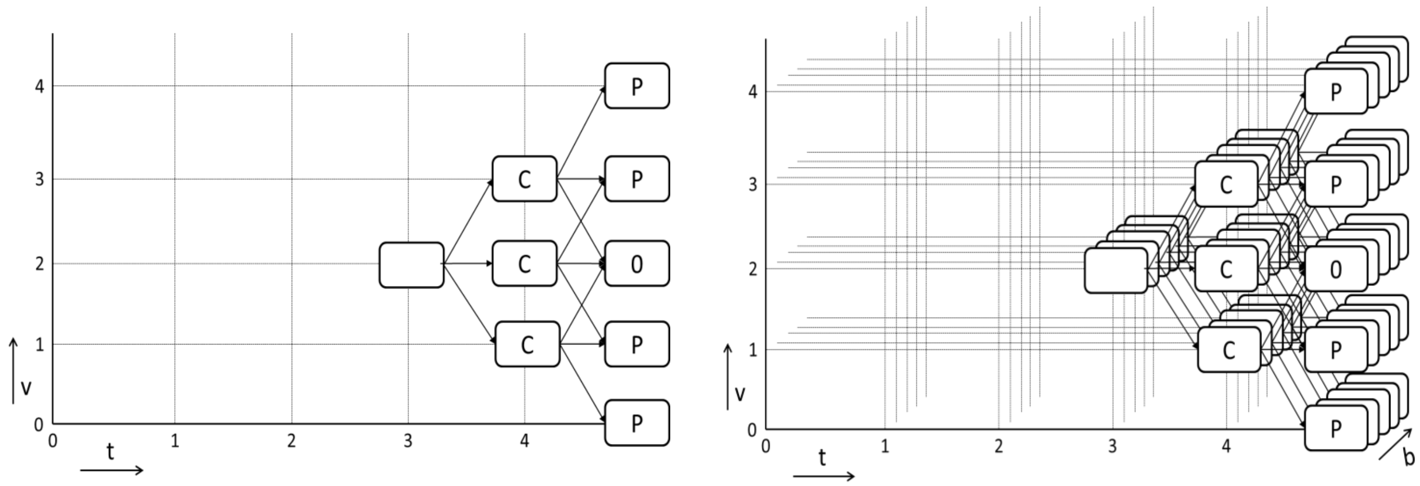

FIGURE 10: Time-Volume Grid for Assigning Value at Maturity

Step 3.a. Run an OLS regression to find an approximation of the continuation value

Maybe the most important step of this dynamic programming problem. The theory of Least Squares Monte Carlo is used to find the continuation value given a certain point in time and volume level. Important to notice is that in the following formula:

$$C^b(t, S^b(t), v(t+1)) \approx e^{-\delta}\, Y^b(t+1, S^b(t+1), v(t+1)) \tag{49}$$

$v(t+1)$ is used instead of *v(t)* . This is done so the transition in volume between *t* and *t+1* is irrelevant, decreasing the dimensionality of the problem. Back to our example, we regress Y, the accumulated future value (in this case the discounted penalties), on X, the price of gas in each price path. In Table 3 the input for the regression is given.

Parameters step 3.a	**b**	**X**	**Y**
t=4	1	21.18913	$e^{-r\Delta t} * \frac{1}{p} * 21.18913$
v=3	2	21.32970	$e^{-r\Delta t} * \frac{1}{p} * 21.32970$
p=1.1	3	19.19480	$e^{-r\Delta t} * \frac{1}{p} * 19.19480$
r=0.04	4	21.16185	$e^{-r\Delta t} * \frac{1}{p} * 21.16185$
$\Delta t = 0.25\ year$	5	21.91155	$e^{-r\Delta t} * \frac{1}{p} * 21.91155$

TABLE 3: Input for regression

Ordinary least square regression is used as the form of regression. This is in line with the numerical example presented by (Longstaff and Schwartz (2001)). When following the example, Y is regressed on a constant, *X*, and X^2, the result is as follows:

$$Y = -8.008 * 10^{-13} + 9.087 * 10^{-1} * X - 1.895 * 10^{-15} * X^2 \tag{50}$$

From this formula the continuation value under these parameters for the first time path, so b=1, is equal to 19.25368. In the above equation Y seems to be a linear function of X. This is because the regression is between the last time step of X and the penalty function Y. Here, the penalty is a function itself of the price at the last time step. At other points in time the regression does not result in a (almost) linear function, this is only the case at the regression between the last price and the penalty function.

Step 3.b. Combine the different continuation values into a decision rule.

Because in this example the change in volume per time-step is only one, three decisions are possible *Inject 1, do nothing, and withdraw 1.* This decision depends on the cash flow that can be generated at that particular point in time and the corresponding future payoffs of continuation. In general, an action with a high immediate payoff has a lower expected future payoff, and vice versa. The immediate or direct payoff is determined by the current price of gas. In our example the decision to inject gas has an immediate payoff of -1 multiplied by 21.18913 (the gas price at the first price path) whereas withdrawal results in a direct payoff of 1 multiplied by 21.18913. The direct payoff combined with the continuation value of each decision can be compared. The option that results in the highest value will determine the decision. For a visualization of this decision see Figure 11.

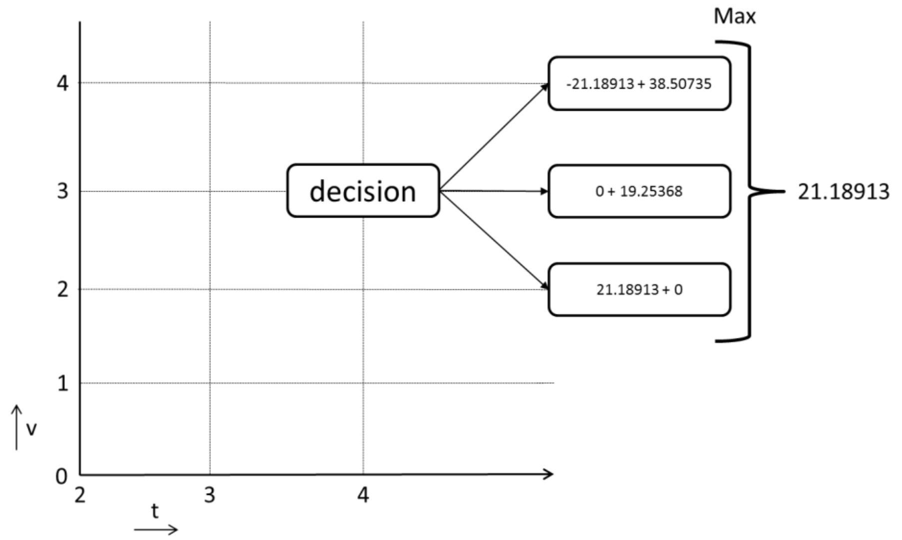

FIGURE 11: Time-Volume Grid to Determine Decision Rule

As can be seen in the figure above the decision that results in the highest expected payoff is to withdraw gas at t=4 for price path b=1.

3.c. Implement the decision rule to calculate the accumulated future cash flows

Now the accumulated future cash flow for this point can be calculated following the decision rule of withdrawal. $acf(t = 3, v = 3, b = 1) = 21.18913 + e^{-0.03*0.25} * 0 = 21.18913$. A commonly mistake here is to confuse the accumulated future cash flow with the continuation value. The continuation value is only used to determine the decision rule whereas the accumulated future cash flow is determined by the 'real' cash flows of that particular decision. To avoid misunderstanding, in the described example the continuation value and the accumulated future cash flow are the same due to the penalty function at the last time step.

3.3 Valuation and Interpretation

In this sub-section all the relevant theory needed for sub-question 3 is stated, i.e.:

Sub-question 3: How to obtain a value and interpret the value of gas storages?

This last sub-question combines the first two to find a value per simulated price path generating a distribution of values. In this section the theory is stated how this distribution of values can be interpreted. This section relates to Section 4.3 in which the theory is applied. In Figure 12 the perspective of this section to the overall structure can be seen.

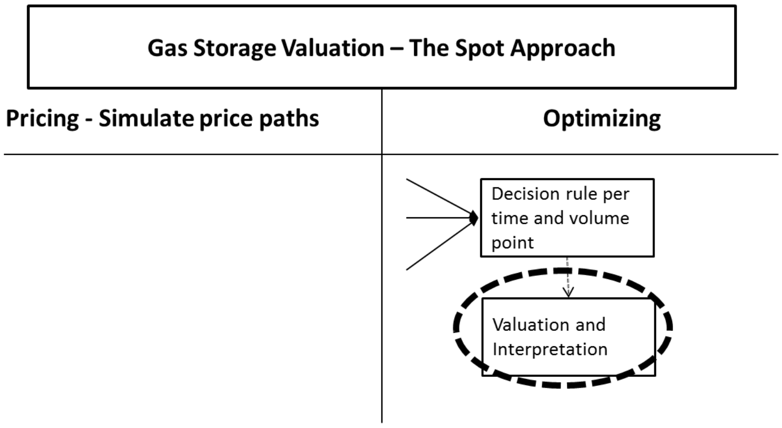

FIGURE 12: From Overall Research Structure; Section 3.3

Recall that the focus of this study is not to value gas storages in the sense to find a claim on the 'real' value of gas storages. Instead, the focus is to support audit work in validating the (client's) gas storage value. Nevertheless, this validating is done by finding a value, or a range of values to enable comparison with the client's value. So, also in the process of validating, values have to be obtained. The big difference is in the claim on the 'real' value. It is up to the auditors themselves to discuss whether a value is reasonable or not. This study presents the following methods to interpret the value distribution: Expected value, standard deviation, 5th percentile point, and Expected Shortfall. The first two are traditionally used to describe a distribution. The last two are often used for risk management purposes to discuss the tails of a distribution.

Expected Value

As stated in Section 3.2, the algorithm of Boogert and De Jong (2008) concludes by averaging all the accumulated discounted future cash flows over all price paths. This expected value is under risk-neutral valuation the gas storage value. However, by itself this number states nothing about the range of the value distribution.

Standard Deviation

The standard deviation is along with the expected value traditionally used to describe a distribution. This measure is used to describe the amount of variation in a set of data. In relation to this study we can use this measure to quantify the wideness of the distribution of values. A very small standard deviation indicates that all values are very close to the mean whereas a high number indicates that the values are spread over a wide range.

5th percentile point

Stating the 5th percentile point of the distribution of values is an attempt to provide a single number that gives intuition about the 'bad' left tail. This can be related to the Value at Risk method that summarizes the total risk in a portfolio and is therefore often used for risk measurement purposes (Hull, 2012). However, the Value at Risk uses real-world probabilities to summarize the total risk of a portfolio whereas our distribution of values represent risk-neutral probabilities. Nevertheless, the 5th percentile point is used to provide intuition about the 'bad' left tail of the distribution.

A disadvantage of the 5th percentile point is that it does not state anything of the behaviour or the corresponding tail itself. For example, it is irrelevant how worse the values below the 5th percentile point are. A way to describe the part of the distribution below this threshold is the expected shortfall.

Expected Shortfall

The expected shortfall is an attempt to provide a single number that summarizes the part of the probability distribution that is below the 5th percentile point. Again, this technique is used in general with real-world probabilities (Hull, 2012). Nevertheless, we use this technique to provide intuition about the behaviour of the 'bad' left tail when incorporating seasonality and volatility updating. A disadvantage is that the expected shortfall is a bit more difficult to understand from an intuitively perspective.

4 Simulation and Results

This chapter addresses the simulation of price paths and the effect of these simulations on gas storage value. The simulation of future gas prices is addressed in Section 4.1. The algorithm presented by Boogert and De Jong (2008) is applied in Section 4.2 to find a 'value' per price path. This distribution of values is then interpreted in Section 4.3 and can therefore be seen as the results of this study.

Each sub-question will be answered by applying the theory provided in Chapter 3. The chapter is constructed so that the theory of Section 3.1 is applied in Section 4.1, Section 3.2 relates to Section 4.2 and so on. From the overall research structure, first the pricing part is discussed which relates to sub-question 1. After that, the optimizing part is presented which relates to sub-question 2. This chapter concludes by answering sub-question 3 about valuation and interpretation.

4.1 Simulating Future Gas Prices – Pricing

In this section the first sub-question is answered, i.e.:

1. *How can spot price paths be simulated?*

The relation of this sub-question to the overall structure is for convenience presented in Figure 13.

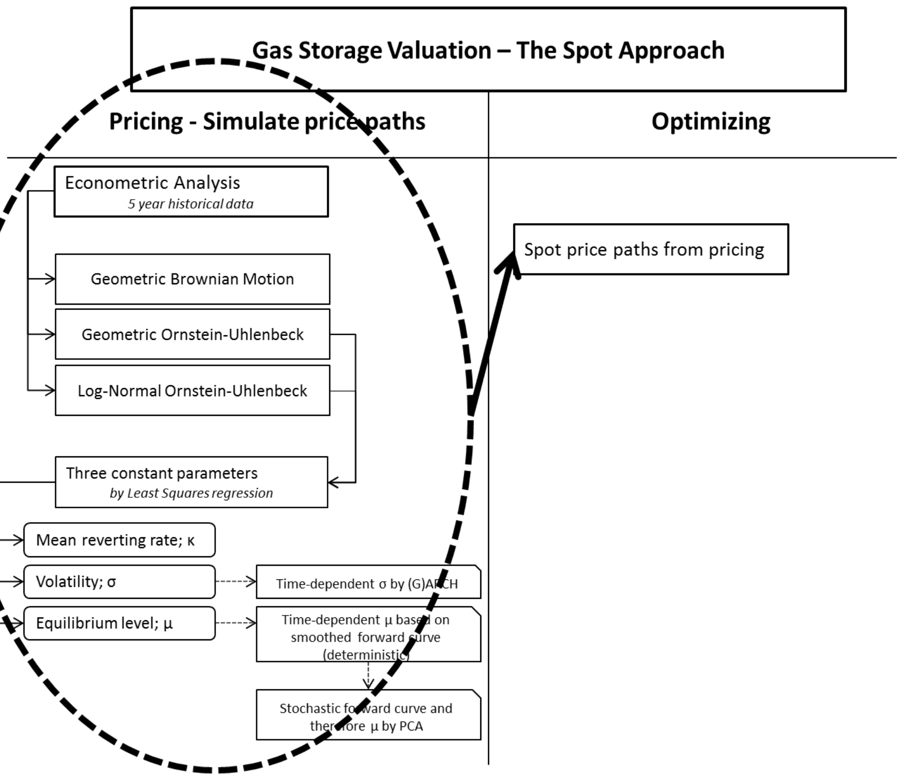

FIGURE 13: From Overall Research Structure; Section 4.1

As made clear in Chapter 2, all the information needed to simulate gas spot prices following a Monte Carlo simulation is presented in this section. First, the spot price is analysed. Secondly, three stochastic processes are discussed that can be used in the simulation. Section 4.1.2 also explains how parameters of the corresponding stochastic processes can be found by calibration and regression. Furthermore, the gas forward curve and gas spot volatility is analysed to incorporate seasonality and volatility updating. The findings on this sub-question are used in the simulation of future gas prices.

4.1.1 Spot Price Analysis

The focus of this sub-section is to perform an econometric analysis on the gas spot price. This focus in relation with the overall research structure is presented in Figure 14. Here, the existence of seasonality and volatility clustering is shown. The gas spot price itself together with the simple returns and the log returns are analysed. The simple returns relate to the Geometric Ornstein-Uhlenbeck process whereas the log returns relate to the Log-Normal Ornstein-Uhlenbeck process.

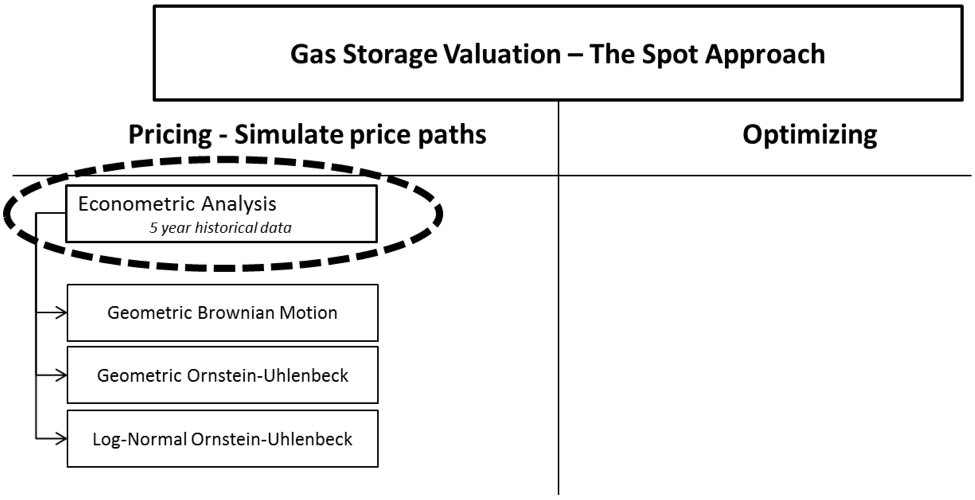

FIGURE 14: From Overall Research Structure; Section 4.1.1

Descriptive Statistics

As stated in the literature review, the NBP spot price together with the simple return and the log return are analysed. Analysis of the HH and TTF can be found in respectively Appendix B and C. The analysis of the NBP spot, simple return and log return is presented in Figure 15. Descriptive statistics are presented in Table 4.

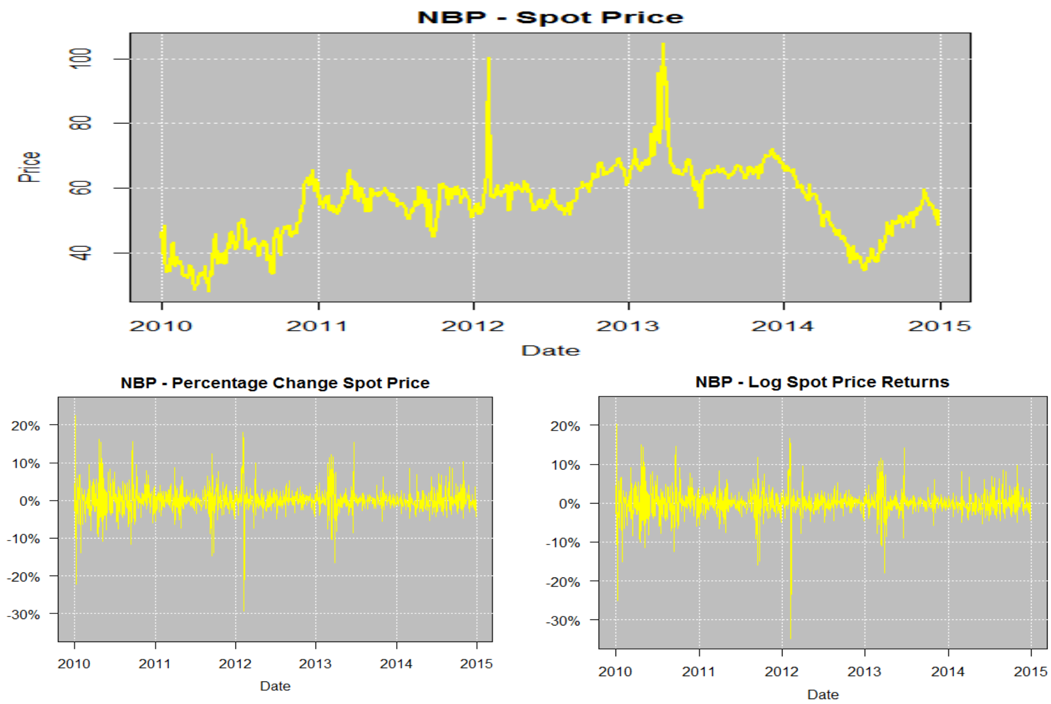

FIGURE 15: NBP - Spot Price and its returns 2010-2014

	National Balancing Point (NBP)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Mean	55.30	7.01e-4	7.09e-05
Min	27.95	-0.2935	-0.35
Max	105.00	0.22	0.21
Quantile:			
• 1%	31.89	-0.09	-0.10
• 5%	36.19	-0.04	-0.05
• 10%	38.75	-0.03	-0.03
• 90%	67.00	0.03	0.03
• 95%	69.31	0.05	0.05
• 99%	84.97	0.01	0.10
Standard Deviation	11.01	0.03	0.03
Skewness	0.04	-0.21	-1.00
Kurtosis	3.88	14.60	18.64
Number of observations	1258	1257	1257

TABLE 4: NBP - Descriptive Statistics 2010-2014

At first sight the NBP spot price seems to move roughly in the range 40 – 70 over 5 years of historical data, except for two so called price spikes: the first one in early 2012 and the second one in early 2013. The two return graphs show very identical behaviour. As in line with theory, the log returns are always smaller than the simple returns. This difference becomes larger when prices fluctuate more. Furthermore, three remarkable features are distinguished.

Seasonality in prices through the year is visible. Traditionally, this comes from the heating of houses in the winter which creates a winter-summer spread. So, gas prices will follow the outside temperature inversely. This can be clearly seen at the NBP – Spot Price in the year 2014. Here the beginning and end of the year (winters), have higher prices than the mid-2014 period (summer). This seasonality is further analysed in the remainder of this section.

When evaluating the skweness and kurtosis statistics, the spot price itself seems to be normally distributed whereas the two return process show different results. In the GOU and LNOU process the returns are assumed to be normally distributed. We present an analysis of the normality of the real returns to determine whether these processes are appropriate to simulate future gas prices. The normality of these return processes are further investigated by the use of QQ plots and relevant statistical tests in the remainder of this section.

The third mentionable feature is that the analysis on the returns suggests changes in variance. For example high volatility at the beginning of both 2012 and 2013 and far less volatility at the remaining of the year. This volatility clustering is also further investigated in the remainder of this section.

Q-Q plots

The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed. In a QQ plot the data is plotted against a theoretical normal distribution, represented by a straight line. The yellow points represent points from our return data sets and the normal distribution is represented by the straight black line.

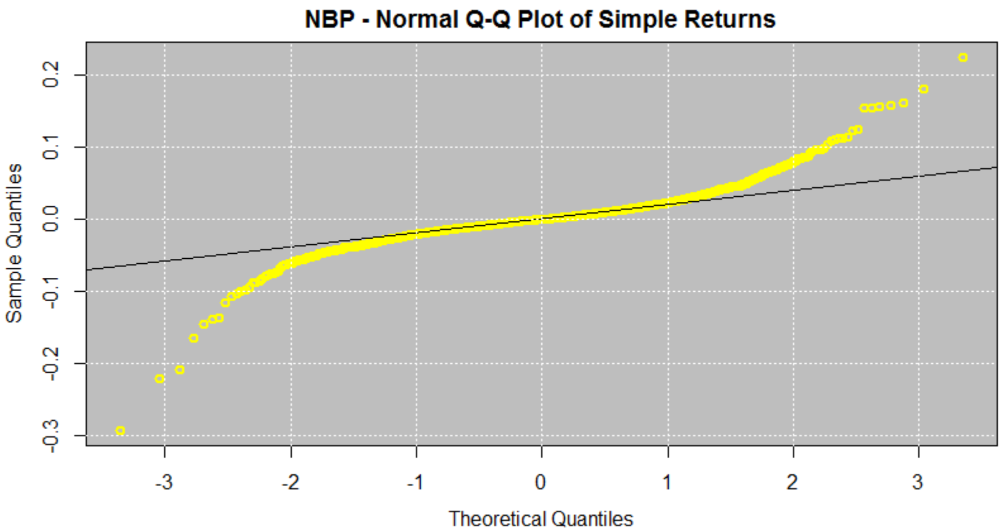

FIGURE 16: NBP - Normal Q-Q Plot of Simple Returns

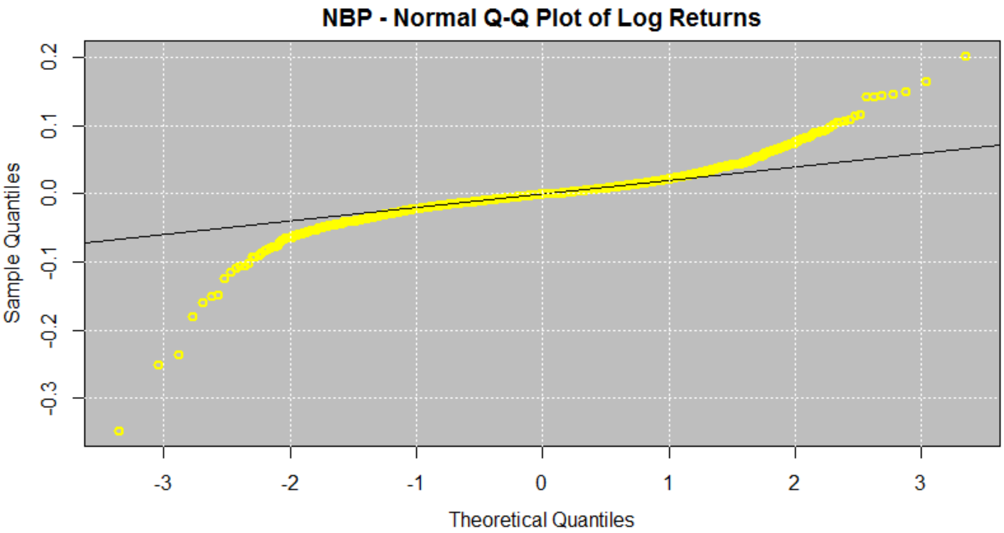

FIGURE 17: NBP – Normal Q-Q Plot of Log Returns

Again, both return plots show very similar results. As expected, the log returns are always a bit lower than the 'simple' returns. Both graphs reveal that the returns have fat tails and normality should be challenged further by statistical tests: the Shapiro-Wilk and Jarque-Bera normality tests. The corresponding p-values of these tests are presented in Table 5.

p-values	National Balancing Point (NBP)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Shapiro-Wilk	3.7e-16	2.2e-16	2.2e-16
Jarque-Bera	1.3e-09	2.2e-16	2.2e-16

TABLE 5: P-values Normality Tests

Because all p-values are below 0.05 the null hypothesis of normality must be rejected. These results are in line with the scores on skewness and kurtosis, and the two QQ plots.

Seasonality

Seasonality in spot prices is extensively described in commodity pricing literature. Besides, it can clearly be seen in the year 2014 with high winter prices and a low summer price. In this sub-section the difference between winter prices and summer prices is statistically proven by the use of boxplots and an analysis of variance (ANOVA). The data set needed to construct the boxplot below is created out of the 5 years of historical NBP spot prices. Out of one data set, twelve data sets are constructed to represent the spot prices of each month. To clarify, January 2010, January 2011, ..., January 2014 are all January prices. As can be seen in the boxplot below, months at the beginning (January, February) and at the end (November, December) of the year have higher prices than mid-year months (June, July, August).

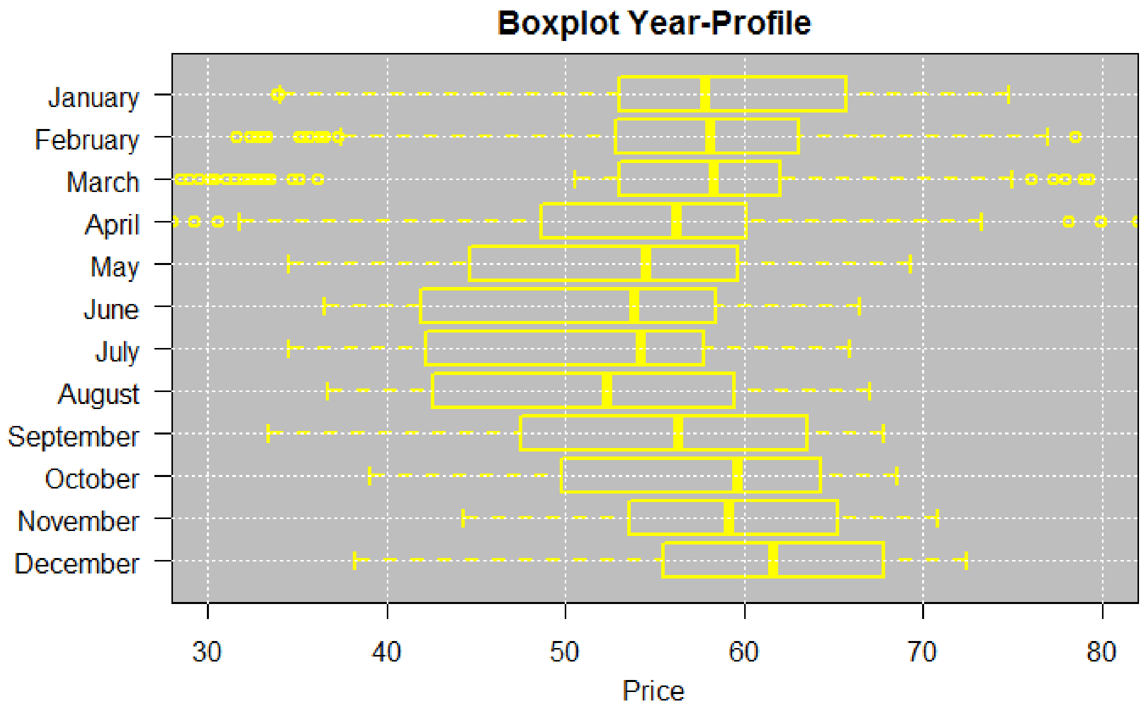

FIGURE 18: Boxplot Year-Profile

To further analyse the difference between months an analysis of variance (ANOVA) is performed. The results of this analysis are presented in Table 6.

Analysis of Variance Month-Year profile

	Df	Sum Squares	Mean Squares	F-Value	P-Value
Month	11	18364	1669.5	15.14	2e-16
Residuals	1763	194457	110.3		

TABLE 6: Analysis of Variance Month-Year Profile

The very low p-value of 2e-16 indicates that there is at least one combination of months which is significantly different from each other. Out of 66 combinations between months, 30[2] combinations turn out to be significantly different from each other with a 95% confidence level.

2 The Tukey Honest Significant Difference test is conducted for this conclusion.

Volatility Clustering

Here the investigation about volatility clustering is continued. Since Figure 15 shows clear signs of volatility clustering three statistical tests are conducted to test this observation. Volatility clustering in the spot price itself will result in a price process that is not stationary and has a unit root. As explained in Section 3.1.1 the Phillips-Perron (PP) test and Augmented Dickey-Fuller test have as null hypothesis the presence of a unit root whereas the Kwiatowski-Phillips-Schmidt-Shin (KPSS) test has as null hypothesis a stationary process. So in short, the first two tests and the last test 'try' to reject the opposite: the presence of a unit root in a process versus a stationary process.

P-values	National Balancing Point (NBP)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Phillips-Perron test	0.21	0.01	0.01
Augmented Dickey-Fuller test	0.33	0.01	0.01
KPSS test	0.01	0.1	0.1

TABLE 7: Statistical Tests Volatility Clustering

From the test results in Table 7, we conclude that the spot price itself is a non-stationary process, so with the presence of a unit root whereas both returns are significant stationary. This is in line with our thoughts at first sight as stated in the section about descriptive statistics.

4.1.2 Stochastic Processes

In this section the three stochastic processes presented in the literature review are evaluated on how suitable and appropriate they are in the simulation of future gas spot prices. The evaluation is twofold. First it discusses the advantages and disadvantages of the stochastic process in general and in relationship with the econometric analysis in section 4.1.1. Secondly, the way to calibrate the corresponding parameters of each process over a historical data set is presented. At the end of this section it is discussed which stochastic process to use in the remainder of this study.

How this section is related to the overall structure is presented in Figure 19.

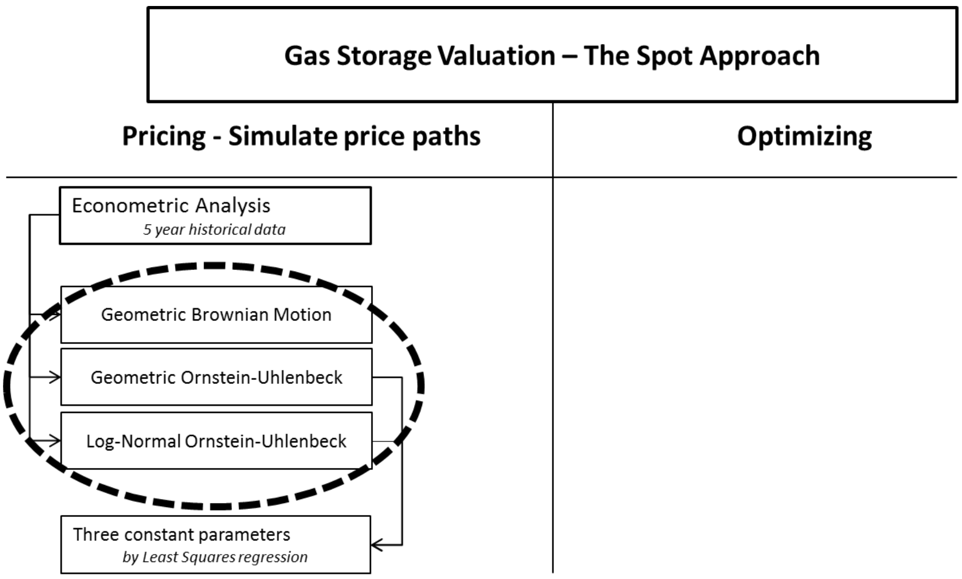

FIGURE 19: From Overall Research Structure; Section 4.1.2

Evaluation and Comparison of Stochastic Processes

The three stochastic processes are stated here shortly. After that, the Geometric Brownian Motion is evaluated and compared to the two Ornstein-Uhlenbeck processes. At the end of this section there is a discussion which stochastic process will be used in the remainder of this study.

Recall, from section 3.1.2, that the stochastic differential equation of the Geometric Brownian Motion (GBM) is as follows:

$$dS = \mu S dt + \sigma S dz \tag{11}$$

Here μ represents the drift, *dt* the change in time, σ the volatility, and *dz* is the increment of a Wiener process.

The Geometric Ornstein-Uhlenbeck (GOU) process is as follows:

$$dS = k[\mu - S]dt + \sigma S dz \tag{17}$$

Where $k \geq 0$ measures the speed of mean reversion, μ is the equilibrium price level and, σ the volatility, and dz is the increment of a Wiener process.

At last, the equation of the Log-Normal Ornstein-Uhlenbeck process can be found below.

$$dS = \kappa[\mu - \ln S]\, S dt + \sigma S dz \tag{21}$$

And after applying Ito's Lemma the log price follows an Ornstein-Uhlenbeck stochastic process:

$$dX = \kappa[\alpha - X]dt + \sigma dz \tag{22}$$

Where

$$\alpha = \mu - \frac{\sigma^2}{2k} \tag{23}$$

Again, the magnitude of the speed of adjustment is represented by $k > 0$ to the long rung mean log price, α. σ is the volatility, and dz is the increment of a Wiener process.

Evaluation and Comparison of the GBM to OU

The biggest advantage of the GBM process is its ease to understand and to implement in the simulation of price paths. Therefore it is used a lot in the field of mathematical finance. Nevertheless, the drawbacks of this model for pricing gas spot prices are stated below and are related to the conclusions drawn by the spot price analysis in Section 4.1.1.

In the econometric analysis performed in Section 4.1.1, two properties of gas spot prices that are demonstrated cannot be captured by the GBM. These properties are as follows:

- Seasonality; By the use of boxplots and ANOVA the existence of a seasonal pattern in gas spot prices is proven. Spot prices in the winter are significantly lower than spot prices in the summer. In the GOU and LNOU this property can be captured by a time-dependent equilibrium level.
- Mean reversion; Besides the clear presence of mean revering behaviour in commodity pricing known from literature, the spot price is tested for no-mean reversion by the Augmented Dickey Fuller test. The result on this test was by far not significant. From an intuitive perspective, the gas demand side is of cyclical nature and there is long term mean reversion due to the cost of new production capacity.

Because of these two gas price properties the GOU and LNOU processes are favoured above the GBM. Therefore, the calibration of parameters of the GBM is no longer relevant.

Calibration of the Geometric Ornstein-Uhlenbeck Parameters

The calibration of Geometric Ornstein-Uhlenbeck (GOU) parameters is performed by Least Square Regression in R over a historical data set of 5 years closing spot prices (the same set as used in the spot price analysis). A study of Insley and Rollins (2005) is followed to perform this calibration. Recall that there are three parameters in the GOU process: the mean reverting rate k, the equilibrium level μ, and the standard deviation σ. To perform the Regression in R, the GOU equation needs to be rewritten as follows:

$$dS = k[\mu - S]dt + \sigma S dz \tag{17}$$

$$dS = -kS + k\mu + \sigma\, S\, \varepsilon \tag{52}$$

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}} = -k + k\mu\frac{1}{S} + \sigma\,\varepsilon \tag{53}$$

$$r = \alpha + \beta\frac{1}{S} + \varepsilon \tag{54}$$

From the last formula the corresponding parameters can be estimated in the software environment of R using linear regression. The minimized R^2 in the NBP case is 0.0061. Here the mean reverting rate is determined by $-\alpha$, the equilibrium level by $\frac{\beta}{-\alpha}$, and the volatility by ε. The GOU parameters calibrated over 5 year of respectively NBP, HH, and TTF gas spot prices are presented in Table 8.

	NBP	HH	TTF
Mean reverting rate; k	0.0118	0.0086	0.0117
Equilibrium price; μ	55.8843	3.6805	23.0855
Standard deviation; σ	3.41%	3.62%	3.31%
Nr. Of observations	1258	1253	1260

TABLE 8: Parameter Estimation Geometric Ornstein-Uhlenbeck

Calibration of Log-Normal Ornstein Uhlenbeck Parameters

The calibration of Log-Normal Ornstein Uhlenbeck parameters is done in a very similar way. The only difference is that regression is applied on the log price instead of the 'normal' spot price. Besides, log returns are used instead of 'simple' returns. Following the LNOU the commodity log spot price follows:

$$dX = \kappa[\alpha - X]dt + \sigma dz \tag{22}$$

Where

$$\alpha = \mu - \frac{\sigma^2}{2k} \tag{23}$$

Again, the equation needs to be rewritten to perform the regression.

$$dX = \kappa[\varphi - X]dt + \sigma\,\varepsilon \tag{55}$$

$$dX = \kappa\varphi - \kappa\, X + \sigma\,\varepsilon \tag{56}$$

$$R = \alpha + \beta X + \varepsilon \tag{57}$$

The corresponding parameters are estimated in the software program of R using linear regression. In the NBP case, the minimized R^2 is 0.0062. The mean reverting rate is set by $-\beta$, φ by $\frac{\alpha}{-\beta}$, and the volatility by ε. In turn, the equilibrium level can be calculated by $\varphi + \frac{\sigma^2}{2k}$. The results of this calibration can be found in Table 9.

	NBP	HH	TTF
Mean reverting rate; k	0.0137	0.0139	0.0147
Equilibrium price; μ	56.8330	3.7641	23.3384
Standard deviation; σ	3.45%	3.51%	3.33%
Nr. Of observations	1258	1253	1260

TABLE 9: Parameter Estimation Log-Normal Ornstein-Uhlenbeck

Discussion

So far, three stochastic processes are evaluated and the calibration process for the relevant processes to find the corresponding parameters is explained. According to this evaluation, the Geometric Brownian motion is considered not appropriate for this study because it cannot handle seasonality and mean reversion. By analysing the behaviour of the daily return function it can be concluded that both returns are not normally distributed whereas price simulation by both Ornstein-Uhlenbeck processes would result in normally distributed returns. Especially the underestimating of fat tails by a normal distribution will cause conflicts. These fat tails represent price spikes and clustering of volatility. GARCH is introduced to take account for this volatility clustering, this is further explained in Section 4.1.4.

Comparing the parameter estimation by the GOU and the LNOU process very similar results are obtained. As expected, calibration by a LNOU results in a slightly higher equilibrium price level because it needs to compensate for the lower log returns. This makes sense because the two processes are after all both OU processes. In continuous time the results would be exactly the same. The mean reverting rate is surprisingly very low as compared to rate used by Boogert and De Jong (2008). These authors use in their study a mean reverting rate of 0.05. Therefore, in the remainder of this study the rate of 0.0137 as well as the suggested rate of 0.05 is used.

All in all, the discussion should not be about which OU process to choose, it should be about the assumptions underlying these processes. The Geometric Ornstein-Uhlenbeck process is particularly suited when gas spot prices are assumed to be normally distributed. On the other hand, the Log-Normal Ornstein-Uhlenbeck process suits when gas prices are assumed to be lognormal distributed. One of the advantages of the lognormal distribution is that it has a lower bound of zero. Intuitively, this is in line with gas price behaviour because one can say that a negative price for gas is strange. However, the assumption that prices cannot be negative is challenged in other markets like interest rates which also had the persuasion that prices cannot become negative. Besides, both processes have a return function that is normally distributed which underestimates the possibilities of extreme movements in price. This makes both stochastic processes, under current settings, unsuitable for risk management and pricing purposes. Nevertheless, it can be used for validation because it gives an indication of gas storage value and can provide a range of reasonable outcomes. For the remainder of this study the LNOU process is used to simulate future gas spot prices because of its advantage that no negative gas prices can be obtained. But as stated before, this is not a claim on real probabilities and prices, it is just a way to get an indication of gas storage value which can be used for validation purposes.

4.1.3 Forward Curve Analysis

The focus of this section is to analyse whether market information on gas forward contracts can be used to set the equilibrium level parameter of an OU process. How this focus relates to the overall research is presented in Figure 20. At first, it is presented how information about forward contracts can be observed from the market. Secondly, a way to smooth this ‘blocky’ forward curve is presented. Thirdly, a day-week profile is estimated. At last, one year of forward curve behaviour is analysed by principal component analysis to simulate future forward curves.

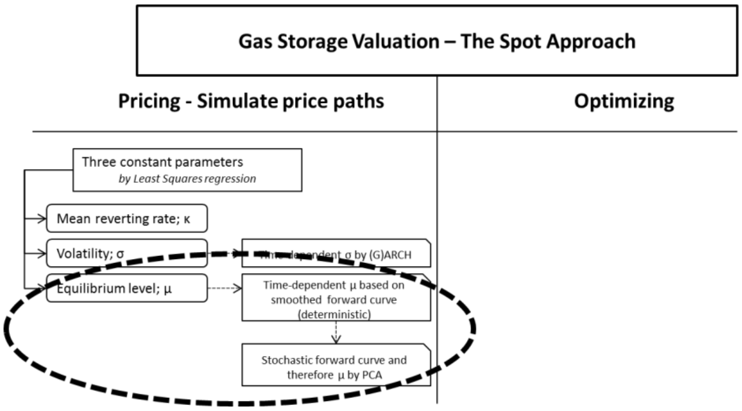

FIGURE 20: From Overall Research Structure; Section 4.1.3

Monthly Forward Contracts

In the literature review it is explained that the gas forward market consists of multiple average-based forward contracts. These contracts represent a constant delivery rate over the settlement period against an average price. An example of market information on forward contracts is presented in Figure 21. In this figure, 12 monthly forward contracts, and four quarterly contracts are visualized, as of 30-12-2014[3] on the NBP Gas market. These contracts are all represented by straight lines over the settlement period. To clarify, a buyer will receive a constant amount of gas over the settlement period for a constant price. Also in the forward market, a clear seasonal pattern is visible.

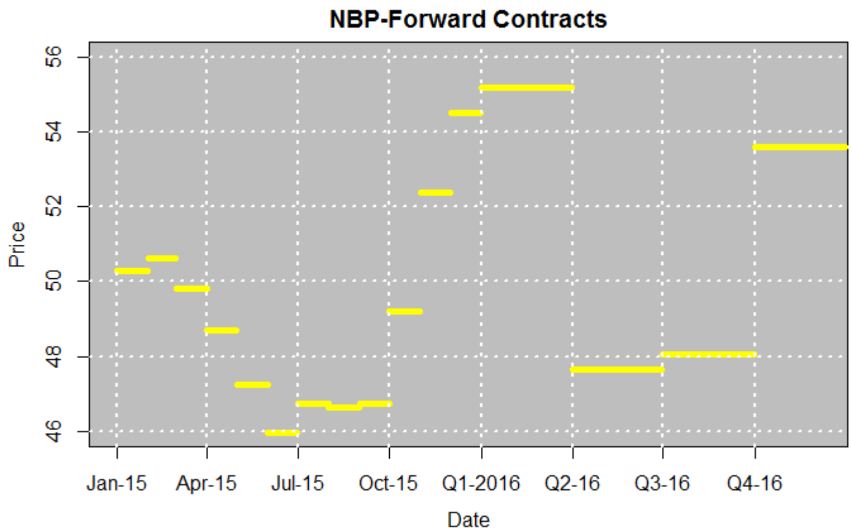

FIGURE 21: NBP-Forward Contracts at 30-12-2014

Forward curve information can be used in the simulation of future gas spot prices. The purpose of this section is to use forward curve information to set the equilibrium level of the Ornstein-Uhlenbeck processes. However, there is one big flaw in using this 'blocky' forward curve directly in the simulation of spot prices. Between the end of each contract period and the beginning of the next forward contract there will be a market inconsistent shock in the equilibrium level. This shock can cause, for example, a one-day equilibrium level shift of 12 percent (between Q1-16 and Q2-16) in the figure above. These shocks are caused by the averaging property of forward contracts. Spot prices will move far more gradually when compared to this blocky curve. Therefore, this blocky curve is smoothed in section 4.1.3.

[3] 30-12-2014 was the last trading day of 2014.

The NBP gas market is the most liquid market in Europe. It trades up till 12 monthly forward contract, 12 quarterly, and 6 seasonal forward contracts whereas the Dutch TTF market trades 'only' 4 monthly, 5 quarterly, and 4 seasonal forward contracts. As stated in Section 2.4, methodology is presented to overcome this lower level of market liquidity. One such methodology is the 'ratio-analysis' between months and quarters. This analysis is based on the correlation between the Dutch TTF gas market and the UK NBP market. This method is pointed out in the next section.

Ratio Analysis

This section presents a technique that makes this study also relevant for less liquid gas markets like the Dutch TTF. The NBP market is taken as a reference point with 12 monthly traded forward contracts and the Dutch TTF market is used as an example for a less liquid market. The gap between the Dutch and UK gas market is 8 months according to liquidity on forward contracts. Information of the NBP market can be used to fill the Dutch gap because these markets seems to behave accordingly. To actually test this, the correlation between the NBP and TTF market is analysed on two levels: one month ahead forward contracts and one quarter ahead forward contracts. The 21-day correlation between the two markets on these two contracts is calculated over a historical data set of 5 years. These two correlations are presented in Figure 22.

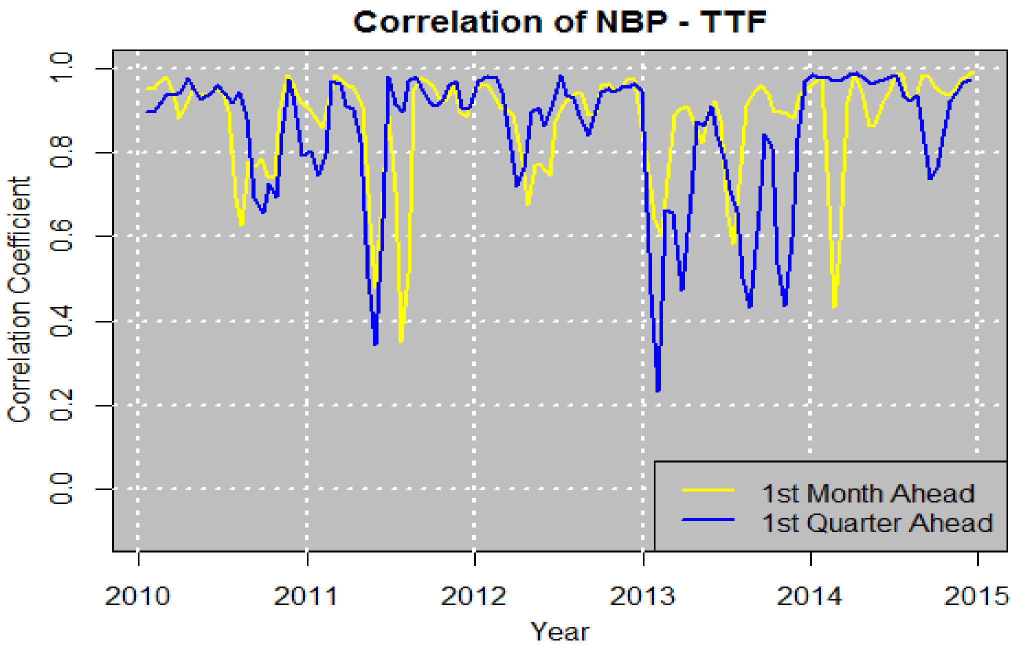

FIGURE 22: Correlation of NBP - TTF

A correlation coefficient of 1 represents perfect positive correlation whereas -1 indicates perfect negative correlation. A correlation coefficient in the middle of this continuum, a score of 0, is an indication of no correlation. Above 0.3 refers to low positive correlation and higher than 0.8 is considered as high positive correlation.

From the graph above we concluded that the Dutch and UK gas markets are highly correlated because around half of the time the coefficient is around or above 0.8 and almost always higher than 0.3. Intuitively this makes sense because there exist physical pipelines between the UK and the Netherlands. If prices would somehow move apart, operators of these pipelines will create value by transporting gas from one country to another. Since the UK and Dutch markets are highly correlated, information of NBP gas prices is used when TTF prices are not available. How this actually works is discussed in the remainder of this section.

Each day on the TTF market 4 monthly contracts, 5 quarterly, 4 seasons, and 5 yearly contracts are quoted. So 8 more monthly contracts and 3 more quarterly contracts are needed to have the same dynamics as the NBP market

(12 monthly contracts for the first year and four quarterly contracts for the second year). In this analysis q_i and m_i stand for the i[th] quarter and month respectively. d_i stands for the number of days of the i-th month.

Four different situations can be distinguished when applying this ratio analysis.

1. *Only one of three monthly contracts in a quarter is unknown.*

This is the simplest situation because no information of NBP ratios is needed to find the value of the missing month. The situation is visualized in the following figure:

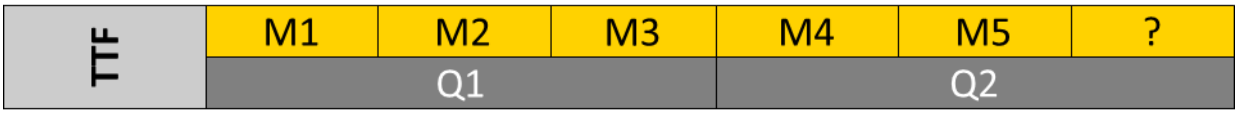

FIGURE 23: Ratio Analysis Scenario 1

This corresponding value can be found by the following formula:

$$m_6 = \frac{(d_4 + d_5 + d_6) * q_2 - d_4 * m_4 - d_5 * m_5}{d_3} \qquad (58)$$

2. *Two out of three monthly contracts are unknown.*

In this situation one NBP ratio is needed as an approximation for the relationship between two TTF months. Otherwise, the formula is unsolvable.

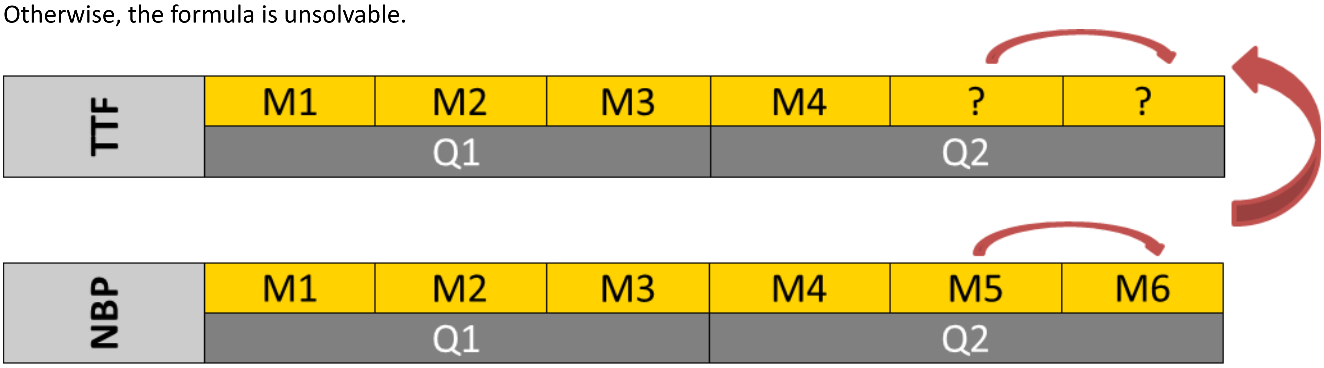

FIGURE 24: Ratio Analysis Scenario 2

This situation needs two formulas to find m_5 and m_6: one to set the ratio between m_5 and m_6, and one to apply this ratio with respect to q_2 and m_4.

$$d_5 * m_5 + d_6 * m_6 = (d_4 + d_5 + d_6) * q_2 - d_4 * m_4 \qquad (59)$$

$$\frac{m_5}{m_6} = \frac{m_5(NBP)}{m_6(NBP)} \qquad (60)$$

3. *Three out of three monthly contracts are unknown.*

This case requires three formulas and two approximations of ratios. The situation:

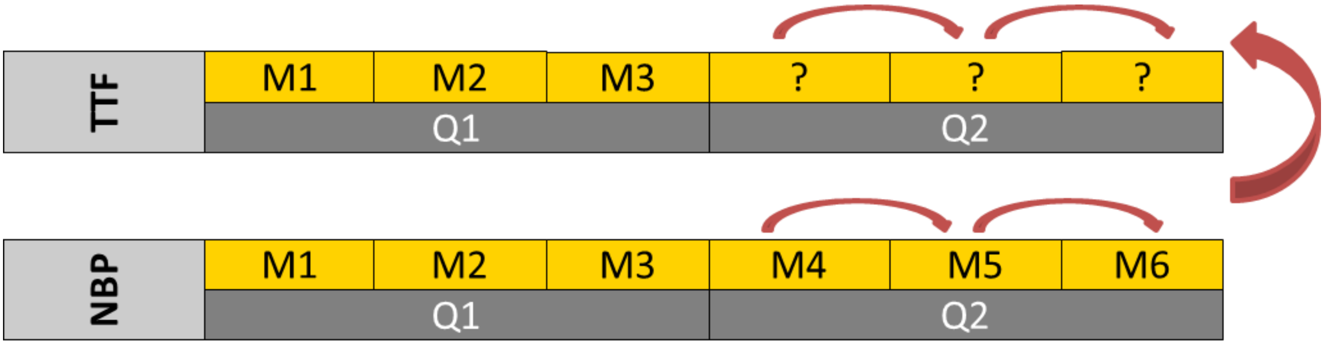

FIGURE 25: Ratio Analysis Scenario 3

The required formulas:

$$d_4 * m_4 + d_5 * m_5 + d_6 * m_6 = (d_4 + d_5 + d_6) * q_2 \tag{61}$$

$$\frac{m_4}{m_5} = \frac{m_4(NBP)}{m_5(NBP)} \tag{62}$$

$$\frac{m_5}{m_6} = \frac{m_5(NBP)}{m_6(NBP)} \tag{63}$$

4. *Two quarterly contracts are unknown whereas the corresponding seasonal contract is known.*

This situation is very similar to situation 2 and 3. Instead, quarters are unknown and seasonal contract is known as compared to unknown months and a known quarter. The case:

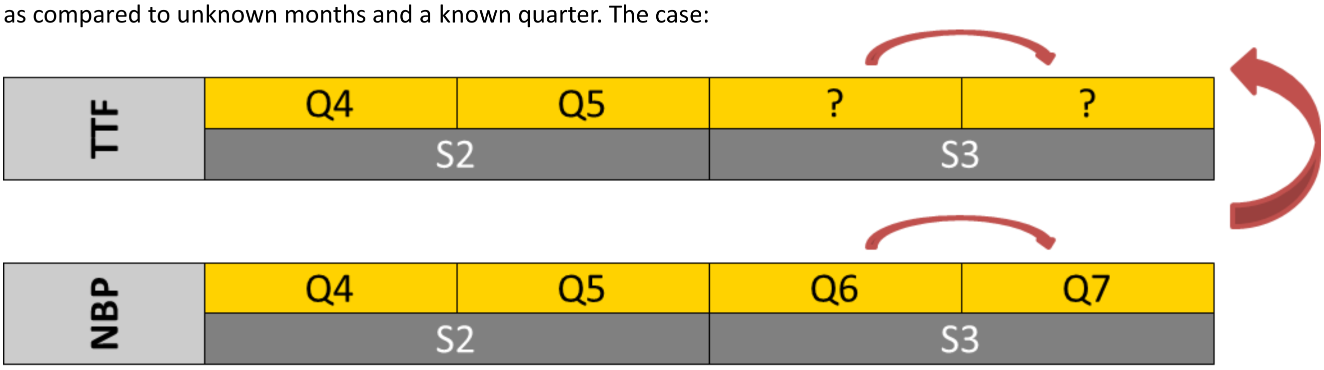

FIGURE 26: Ratio Analysis Scenario 4

These formulas are also very similar to situation 3:

$$(d_{16} + d_{17} + d_{18}) * \ q_5 + (d_{19} + d_{20} + d_{21}) * q_6 = (d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21}) * \ S_3 \tag{64}$$

$$\frac{q_6}{q_7} = \frac{q_6(NBP)}{q_7(NBP)} \tag{65}$$

This method can be applied to other gas markets as well when these markets are considered to be positively correlated to each other. Not only additional monthly forward contracts can be established, this methodology can also be applied on seasons on quarters, years on seasons and so on.

Smooth Forward Curve

The need for smoothing the 'blocky' forward curve is discussed in section 4.1.3. In this section the theory of smoothing a forward curve as mentioned in the literature review is applied. Actual smoothing is programmed in the software environment of R. The smoothing is performed by using cubic splines where the smoothing parameter can be set by hand. As mentioned before, the higher this parameter, the more shrinkage of the spline. In this study the smoothing parameter is set in a subjective way. This parameter is set to 0.75 because it shows a gradually curve without strange, unexpected movements. We consider setting the smoothing parameter by hand appropriate for this study. This is because of the purpose of validation in contrast to pricing. The result of a cubic spline on top of NBP monthly and quarterly forward contracts at 30-12-14 is given in Figure 27.

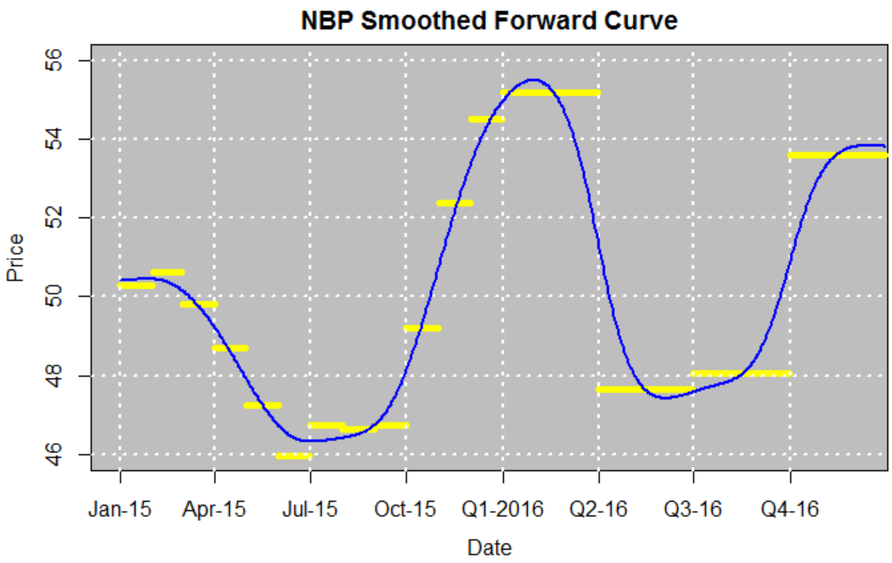

FIGURE 27: NBP - Smoothed Forward Curve

The smoothed forward curves goes through the middle of almost every contract and the direction of the spline at the edges of the contracts depends on adjacent contracts.

Day-Week Profile

The seasonality through the year is not the only type of seasonality discovered in the behavior of gas spot prices. From electricty pricing literature (Hildmann, Herzog, Stokic, Cornel, & Andersson, 2011) it is known that seasonality may also exist throughout the week. However, we do not know if this is also the case in gas pricing. Therefore, potential seasonality through the week is tested in this section.

Seasonality through the year is proven by comparing months of spot prices piecewise with each other, see Section 4.1.1. To test for seasonality in the week a quite similar methodology is used. However, not 'absolute' prices are used. Instead, the ratio of the day-price to the average price of the week is used. To clarify an example of the last week of 2014 is presented below.

Ratio of Tuesday 25th of December:

$$\frac{\text{Tuesday's (25th) spot price}}{\{\text{Monday's (24th)price} + \text{Tuesday's (25th) price} + , \ldots , + \text{Sunday's (30th) price}\}/7} \quad (66)$$

Now, the ratio's over 5 year historical data are grouped in seven week-days and the analysis is continued in the same way as in the analysis on the year-seasonality. This change in procedure is needed so the day-ratio's can be applied directly to the smoothed forward curve. In this way, week seasonality is added to the curve without changing the averge price of the week. The results are presented in Figure 28, Table 10, and Table 11.

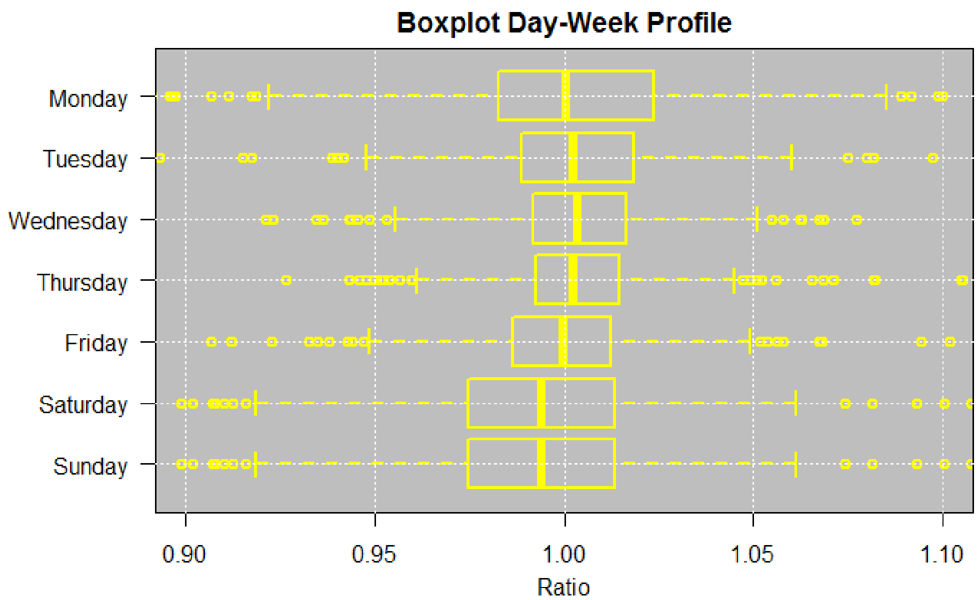

FIGURE 28: Boxplot Day-Week Profile

In Figure 28 boxplots of the day-ratios are shown. A difference between midweek-days and weekend-days can be distinguished. Whether or not this difference is significant is determined by ANOVA. The two weekend-days have the same boxplot since both prices are set by the weekend ahead prices on Fridays.

	Df	Sum Squares	Mean Squares	F-Value	P-Value
Month	6	0.0463	0.0077	4.653	0.0001
Residuals	1762	2.9236	0.0017		

TABLE 10: Analysis of Variance Day-Week Profile

The ANOVA shows, with a p-value of 0.0001, that there is at least one piecewise combination between the week-days that is significant different (even with confidence levels up to 99.9%). By performing a Tukey Honest Significance test the combinations are tested individually. The results show that under a 95% confidence level all weekend-days differ significantly with Tuesday, Wednesday and Thursday. This is in line with the results obtained by the boxplots because Tuesday, Wednesday, and Thursday are the days with the highest mean and the weekend-days have the lowest means. The p-values of these combinations can be found in the table below.

95% Confidence level	**P-value**
Saturday - Tuesday	0.0223
Saturday – Wednesday	0.0045
Saturday – Thursday	0.0194
Sunday - Tuesday	0.0223
Sunday - Wednesday	0.0045
Sunday - Thursday	0.0194

TABLE 11: Significant Different Days

Now it has been shown that besides seasonality through the year, there is also seasonality through the week, the ratios can be applied to the smoothed forward curve. The day-week ratios of NBP spot prices, calculated over 5 year historical data are presented in Table 12.

Day	Ratio
Monday	1.0000
Tuesday	1.0045
Wednesday	1.0059
Thursday	1.0044
Friday	1.0003
Saturday	0.9927
Sunday	0.9927

TABLE 12: Day-Week Ratios

Now these ratios can be applied to the smoothed forward curve, again on available forward contracts at 30-12-2014. In the graph below, the 'blocky' forward curve retrieved from market data is presented by the yellow flat lines, the smoothed forward curve is in blue and the added day-week profile is visualized by the red line. As expected, the seasonality through the week is clearly visible by the small up and down movements of the curve. Furthermore, the red line follows the smoothed blue curve perfectly.

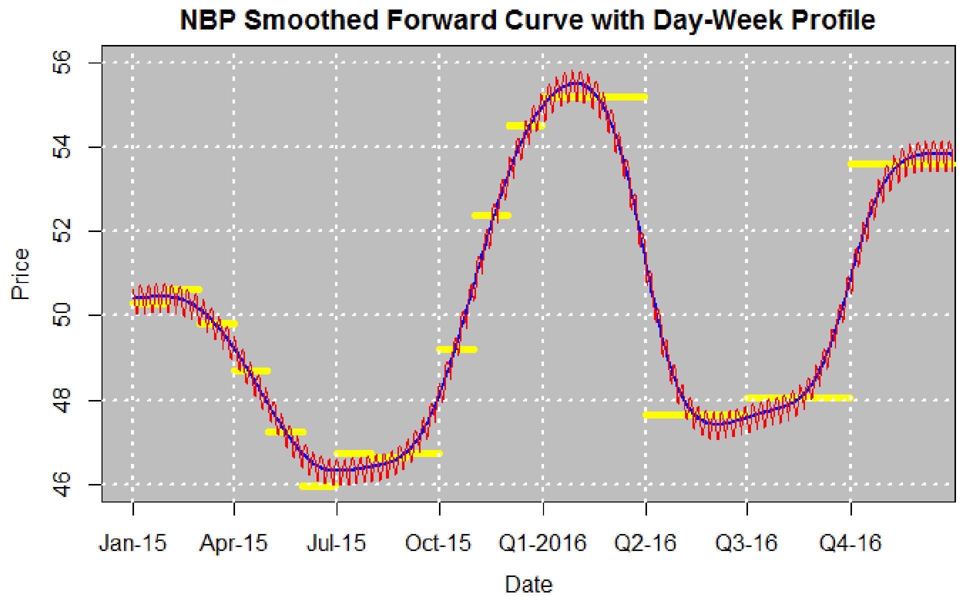

FIGURE 29: NBP – Smoothed Forward Curve with Day-Week Profile

The seasonality through the year was the motivation to analyse the forward curve. This analysis is used to form a time-dependent deterministic forward curve which can be used as the equilibrium level parameter in an Ornstein-Uhlenbeck process. The forward curve analysis includes constructing a smoothed forward curve by fitting a cubic spline and adding a day-week profile on top of this curve. The three related forward curves at 30-12-2014 are presented in Figure 29. In the next sub-section an attempt is made to construct a stochastic equilibrium level by simulating forward curves over time.

Time-Dependent Equilibrium Level and Monte Carlo Simulation

In the former sub-section it is shown how a time-dependent equilibrium level can be implemented in the simulation of gas spot prices. Here, we will visualise this effect on the simulation of 2 years of gas spot prices. For this simulation the parameters are set following the calibration in Section 4.1.2 but the constant equilibrium level is substituted by a time-dependent one. In this case the forward curve including day-week profile as presented in Figure 29. As stated in section 4.1.2 two mean reverting rates are used: the calibrated rate of 0.0137 and the rate used by Boogert and De Jong (2008), 0.05. The results are given in Figure 30 and 31, and discussed below.

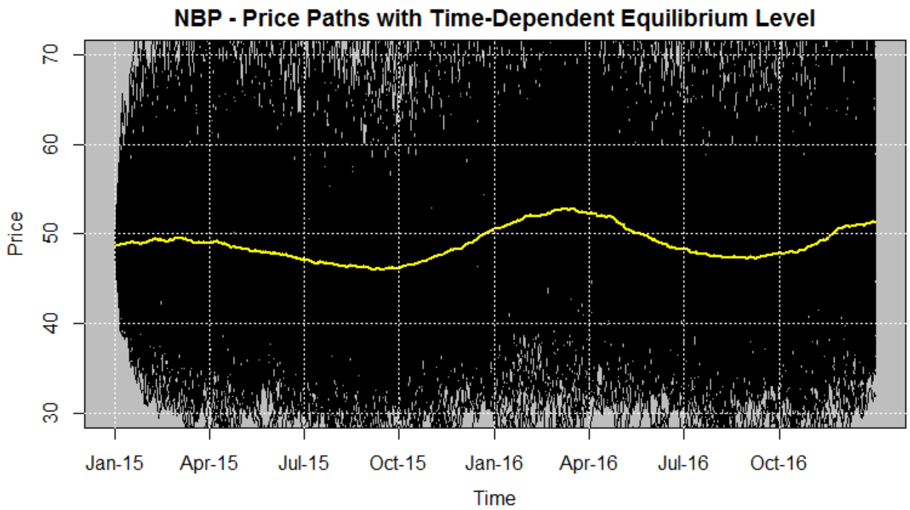

FIGURE 30: Price Paths with Time-Dependent Equilibrium Level; k=0.0137

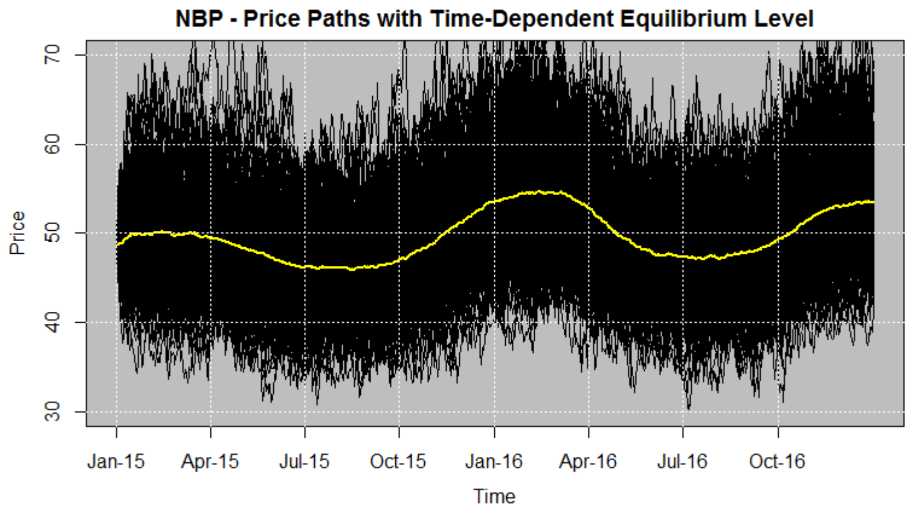

FIGURE 31: Price Paths with Time-Dependent Equilibrium Level; k=0.05

In the figures above the yellow curve represents the average of the simulated price paths. For this Monte Carlo simulation the amount of 1000 price paths is chosen. If one would use Monte Carlo simulation for the purpose to have a claim on the ‘real’ price, it can be discussed that 1000 runs are insufficient. Nevertheless, the amount of runs are considered reasonable in this study because of computational speed restrictions and the purpose of validating instead of pricing.

In Figure 30, the average line follows the forward curve in a somewhat squeezed way. Besides, the price paths move in a more ‘unrestricted’ way around the equilibrium level as compared to Figure 31. This is of course caused by the difference in mean revering rate. If this rate is higher, price paths are more near the equilibrium level. The day-week profile is therefore not visible in the average line of Figure 30 whereas it is for a bit detectable in Figure 31.

From this we can conclude that using a forward curve at the date of valuation is an appropriate way to take account for, market consistent, seasonality in the gas spot prices. However, it depends on the mean reverting rate in what order the time-dependent equilibrium level is followed by the price paths.

We should take additional care when interpreting the results because we changed one historical calibrated, constant, parameter into a forward-looking, time-dependent, parameter. By doing so we end up with a mixture of two calibrated parameters (volatility and mean reverting rate) and one market consistent, ‘implied’, parameter.

One can state that by changing one parameter, the other two have to be recalibrated. Nevertheless, for the scope of this study this mixture of parameters is considered appropriate.

Principal Component Analysis

The goal of this section is to simulate forward curves in the future based on the most recent known forward curve (for example the curve at 30-12-2015). Principal component analysis is used to find the 'principal components' that are needed to describe at least 95 % of the variance in the forward curve. As discussed in the literature review the principals components are used in the following way:

$$F(t+\Delta t, T) = F(t,T) exp\left[\sum_{i=1}^{N}(\sigma_i(T-t)\sqrt{\Delta t} * \varepsilon_i - \frac{1}{2}\sigma_i^2(T-t) * \Delta t)\right] \quad (38)$$

This is a discrete time representation because it is used to simulate the forward curve at time $t + \Delta t$ on the forward curve at time t. Furthermore, *F(t,T)* represents the forward price at time t for delivery at time *T*, $\varepsilon_i's$ are i independent standard normal distributed numbers, and $\sigma_i(T-t)$ are the factor loadings.

The exponential expression in the above formula indicates that log returns are used. This makes this process relevant for the Log-Normal Ornstein-Uhlenbeck. If the Geometric Ornstein Uhlenbeck is followed, the 'simple' returns have to be calculated. As mentioned in the literature review the log returns are calculated between the two contracts corresponding to the same 'delivery' date on the forward curve:

$$ln[curve2(t1)/curve1(t2)] \quad (37)$$

The steps performed in Section 4.1.3 to transform a 'blocky' forward curve retrieved from the market into a smoothed curve with day-week profile are visualized in Figure 29. Here one year of forward curves, over the year 2014, are transformed from using only the first year of the curve (up till 365 days to maturity). In Figure 32 the curves 'flow' from left to right, increasing in days to maturity. The top left 3D plot represents the forward contracts retrieved from the market, in the top right plot the forward curves are smoothed. To finalize, the day-week profile is added in the lower 3D plot.

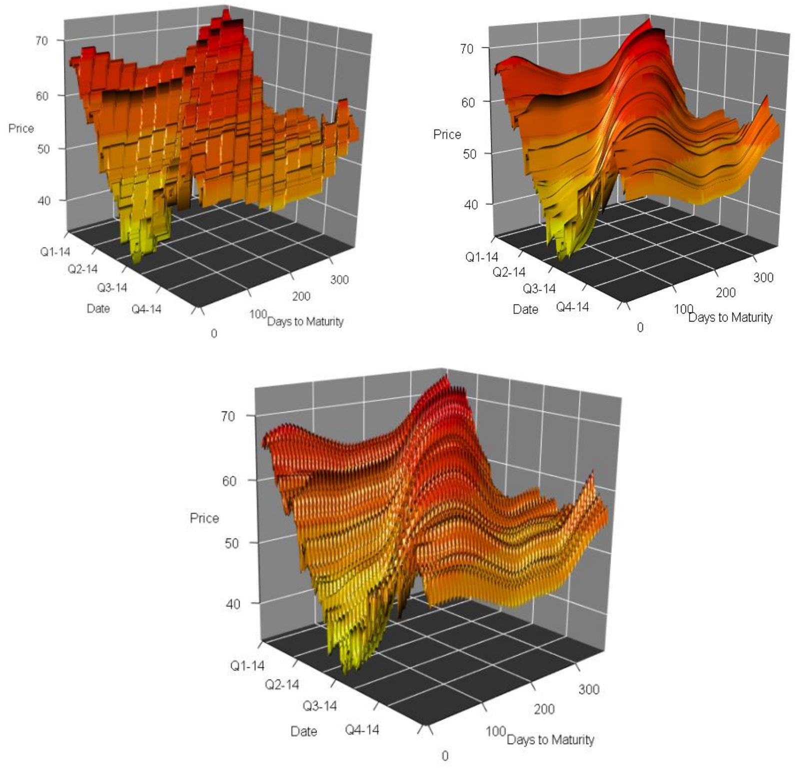

FIGURE 32: 3D Plots of 2014 Forward Curves

The seasonality through the year can easily be seen in these 3D plots. There are higher prices related to winter periods and lower prices in the summer. This is visualized in the 3D plots by the higher diagonal line from the corner in the back, to the front-corner representing the winter periods. See for instance that, the 360 days to maturity is high at January curves, 180 days to maturity is high at June curves and at the December curves the short period to maturity is high, all representing the same winter period.

From this last 3D plot, the log returns are calculated. Since we are interested in how the returns related to each other a covariance matrix is calculated. This 365 times 365 matrix is visualized in Figure 33. As expected the covariance show a positive relationship between all days to maturity. This can intuitively be explained by parallel shifts in forward curves which causes the most of the variation. Furthermore, the returns are more related at the short end of the curve than at the end.

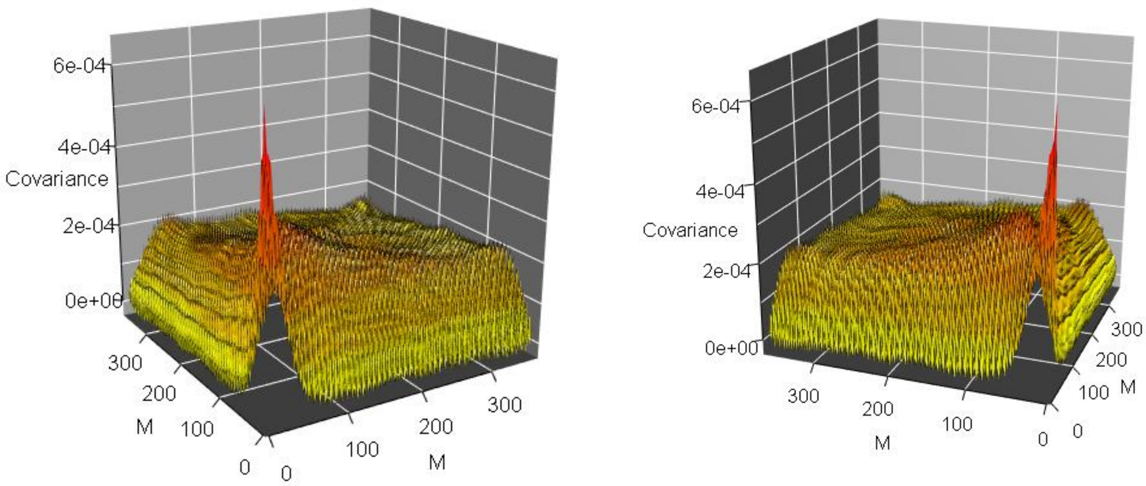

FIGURE 33: Covariance Matrix of Log Return (from different angles)

The above 3D plots represent the Covariance matrix from two different angels. As can be seen from the peak, the covariance between returns on the short end of the forward curve is much higher than between other parts of the forward curve.

The next, and final step of the principal component analysis is to actually determine the components that explain most of the variance. As in line with the literature review, the eigenvalues and corresponding eigenvectors are calculated. After that, these values and vectors are sorted. The bar-plot and table below show the cumulative attribution of the first 10 components.

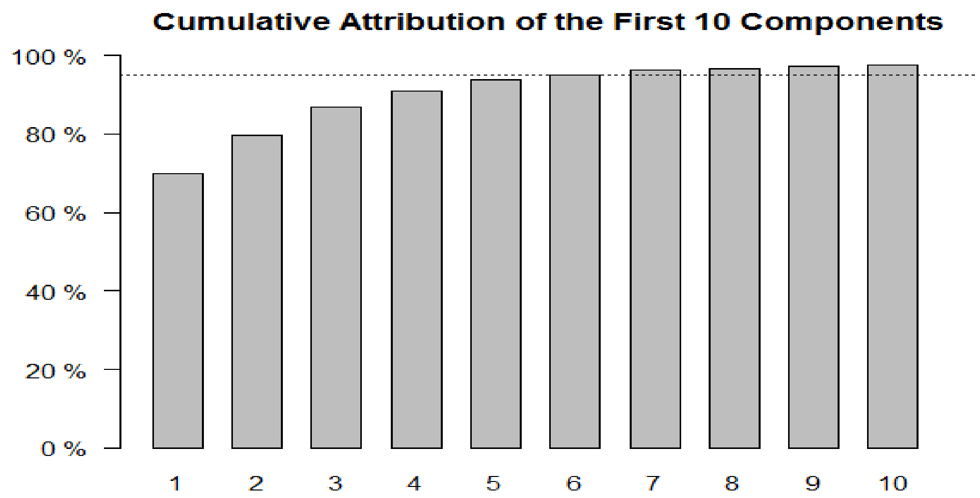

FIGURE 34: Bar-plot Cumulative Attribution of the First 10 Components

	1	2	3	4	5	6	7	8	9	10
Cum. Attr.	69.91%	79.67%	86.95%	90.80%	93.73%	95.11%	96.12%	96.69%	97.19%	97.59%

TABLE 13: Cumulative Attribution of the First 10 Components

To explain 95% of the variance, at least 6 components are needed. Whereas the first component explains already almost 70% of the variance, the 6th component only adds around 1.5%. The actual simulation of up to one year of forward curves is done in R. The code is presented in Appendix D.

As stated before, the forward curve is the expectation of future spot prices. Therefore, forward curves can be used as a time-dependent equilibrium level for the simulation of gas spot prices. In the sub-section above, this equilibrium level is set by the forward curve on valuation/validation date. However, forward curves change over time. In this sub-section an attempt is made to take this stochastic behaviour of the forward curves into account for setting the equilibrium level of gas spot prices. So we are not interested in the forward curves themselves, but in equilibrium levels they set. The corresponding process per simulated equilibrium level is as follows. Firstly, 365 forward curves are simulated based on Equation 38 and the relevant components found by PCA. Now we have, similar to Figure 29, up till one year of forward curves, each with 1 till 365 days to maturity. Secondly, we take from each forward curve the point on the curve that corresponds with 1 day to maturity. This point represents the next day's equilibrium level. For example, the equilibrium level of the 1st of July is determined by the point that represents 1 day to maturity on the forward curve of the 30th of June. At last, we can repeat this process for the amount of simulations needed.

An example of 5 simulated equilibrium levels is presented in Figure 37. In this figure the 'starting' forward curve of 30-12-2014 and the average of the simulated equilibrium levels is also presented. As can be seen, the equilibrium paths can differ a lot and the day-week profile is relatively large as compared to the 'real' forward curve at 30-12-2014. Only five 'paths' are shown, otherwise the individual paths cannot be separated. The average of the equilibrium levels is as expected very near to the 'real' curve on 30-12-2014. However, we cannot obtain a claim on this observation since the very small amount of simulations.

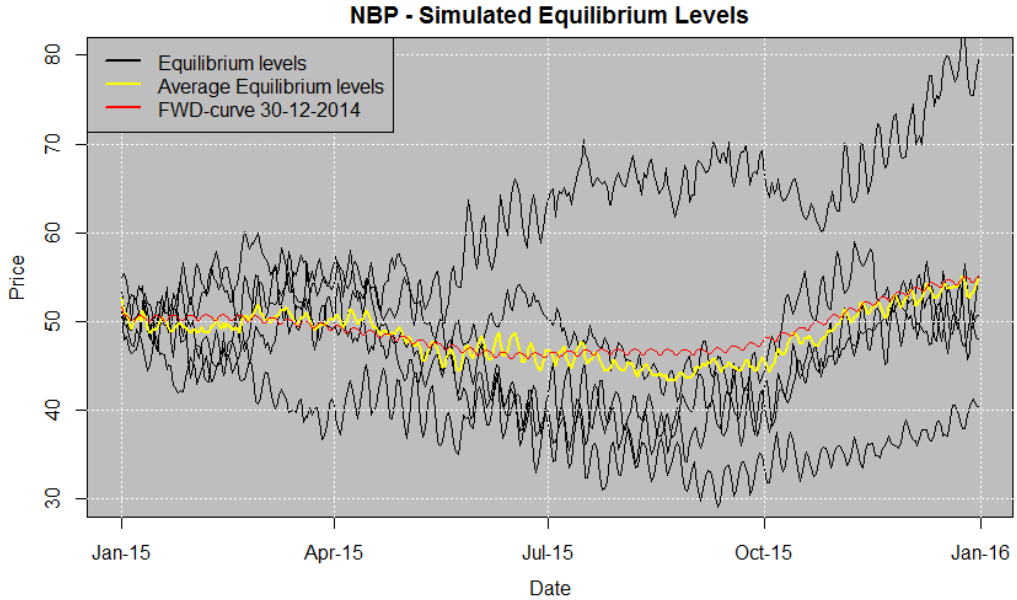

FIGURE 37: NBP – Simulated Equilibrium Levels

When we take a look at the simulations presented in Figure 37, the relatively large fluctuation within the week of each 'path' is a point of concern. As stated before, these equilibrium 'paths' are an input parameter in the Ornstein-Uhlenbeck process for the simulation of gas spot prices. Due to the large fluctuation within a week, it may happen that the equilibrium level moves up with 10% and returns in the same week. This is not in line with our findings regarding the spot price. One of the reasons of the large within week fluctuation we can think of, is that only the 1 day to maturity prices on the forward curves are used in the simulation. On this, very short end the forward curve is most volatile.

Besides the large within week fluctuation of the equilibrium levels, we conclude that there may be discrepancy between the movements of the spot price, and the forward curves over time. In the figure above, the equilibrium level is independent of spot price movements whereas the curve should fit with the spot price at the short end fluently. For example, the one-day to maturity price on the forward curve may move up with 5 points whereas the simulated spot price can move in the opposite direction.

All in all, we consider the advancements made by a stochastic equilibrium level in comparison to a deterministic equilibrium level insufficient regarding the additional points of concern. Especially because of the large within week fluctuations, we will not use the stochastic equilibrium levels in the remainder of this study.

4.1.4 GARCH

In the section before, methodology is presented to transform the equilibrium level from a constant parameter to a time-dependent parameter. The focus of this section is also to transform a constant parameter: the volatility parameter. How this is related to the overall structure of this study is presented in Figure 38.

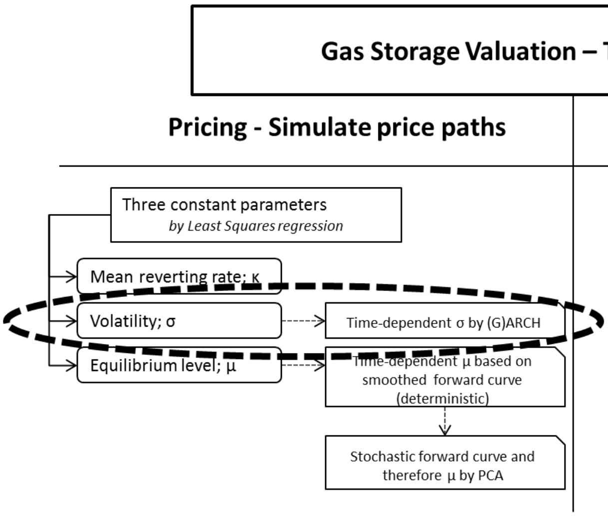

FIGURE 38: From Overall Research Structure; Section 4.1.4

The need for a non-constant volatility parameter is pointed out in section 4.1.1. The return functions of the spot price show obvious signs of volatility clustering which are also demonstrated by statistical tests. GARCH models are used to extent stochastic processes to capture the non-constant behaviour of its underlier. As mentioned in the literature review, in a GARCH model the conditional expectation term can be represented by any stochastic process, only the volatility part is transformed from a constant parameter to one that can handle volatility updating. Recall that the conditional variance is modelled by an autoregressive moving average process. The basic form of a GARCH (p,q) model is:

$$S_t = E[S_t | S_{t-1}] + \sigma_t^2 \in_t \tag{39}$$

Here, the price *S* at time *t* is based on the expectation of price *S*, conditional on the price at *t-1*, the volatility σ, and error term $\in$.

In the GARCH model the error part is $\in_t \overset{iid.}{\sim} N(0,1)$ and

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{40}$$

In line with Hansen and Lunde (2005) a GARCH (1,1) is chosen to set the volatility of the Ornstein-Uhlenbeck process. This results in the following volatility:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-i}^2 + \beta \sigma_{t-i}^2 \tag{67}$$

A maximum likelihood analysis over 5 year historical data is performed to find the corresponding parameters. The related code used for this calibration is presented in Appendix D.

Volatility Updating and Monte Carlo Simulation

In the sub-section before it is showed how volatility updating can be used in the simulation of gas spot prices. In this sub-section we will provide two figures to visualise this effect on the simulation of gas spot prices. Again the parameters are set following the calibration in section 4.1.2, except for the volatility. The volatility is set according to Equation 67. Also two mean reverting rates are used: the calibrated rate of 0.0137 and the rate used by Boogert and De Jong (2008), 0.05. The results are shown in Figure 34 and Figure 35, and discussed below.

In comparison with the price paths in Figure 30 and 31 the figures below have a much larger price range on the y-axis. As expected this is due to extreme price movements caused by the clustering of volatility. Again, 1000 price paths are simulated. As can be seen the price spikes are sometimes even higher than the range of the y-axis. This maximum is chosen so the 'main' part (price range of 30-70) of the price paths is still observable. If we take a look at the historical price path of the NBP gas spot two price spikes can be distinguished. Both of these price spikes reach price levels of 100 whereas the simulated price paths reach levels of 300-400. This is not surprisingly because the GARCH parameters are calibrated over the historical price paths which includes price spikes up till 100. So even higher price spikes can be expected when simulating 1000 paths.

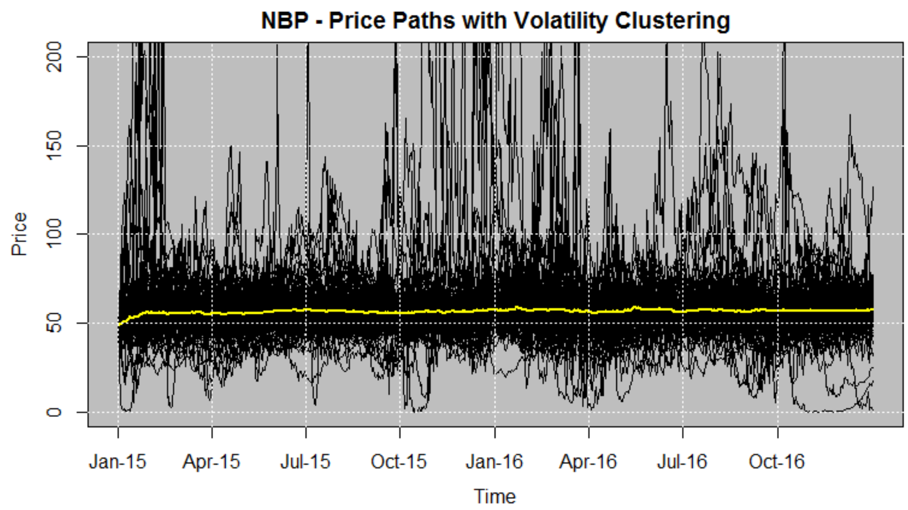

FIGURE 39: Price Paths with Volatility Clustering; k=0.0137

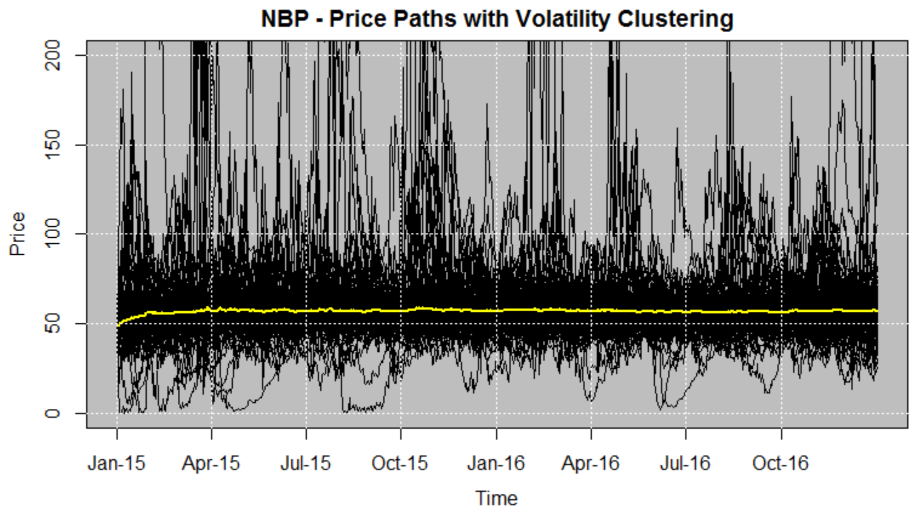

FIGURE 40: Price Paths with Volatility Clustering; k=0.05

The difference between Figure 39 and Figure 40 seems to be very small. However, this can be caused by the misleading greater price range of both figures. If we take a close look at the 'main' part of the price paths we can still see a difference in the price paths. The price paths with a mean reverting rate of 0.05 tend to move somewhat closer to the equilibrium level as compared to the price paths in Figure 39.

For the simulation of price paths under volatility updating we used a GARCH (1,1) model. Hansen and Lunde (2005) did not find much evidence that another model outperformed a (1,1) model. We follow this conclusion that from a general perspective the GARCH (1,1) model is appropriate. Nevertheless, we do not claim that a (1,1) model is most suited for the simulation of gas spot prices. Other models perform potentially better. However, it is considered out of scope for this study to find the best suited. Besides, the gain of a 'better' GARCH model is considered very small in the context of validation.

4.2 Finding the optimal operating strategy – Optimizing

In this section the second sub-question is answered, i.e.:

2. *What is the principle to find the optimal operating strategy for a gas storage when following the spot approach?*

The relation of this sub-question to the overall structure is for convenience presented in Figure 41.

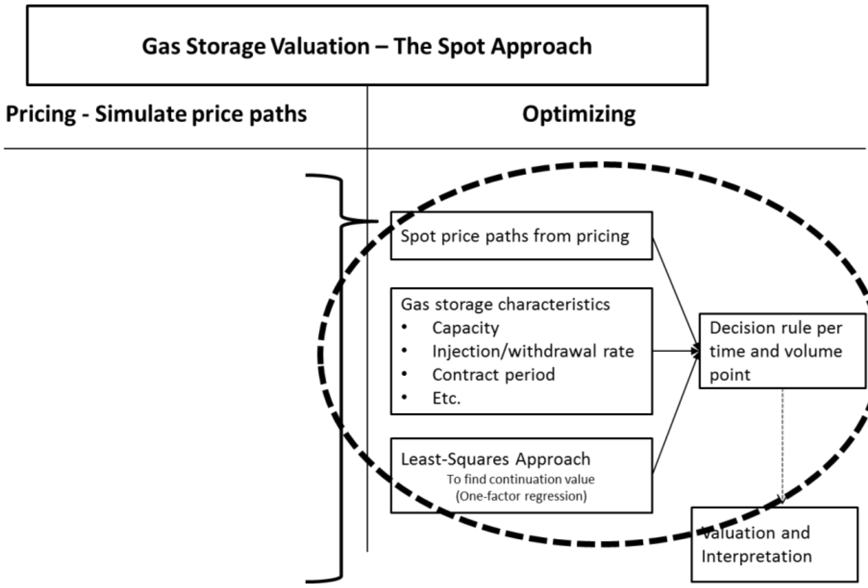

FIGURE 41: From Overall Research Structure; Section 4.2

The theory used in this section is given in the literature review, section 3.2. The general principle to find the optimal operating strategy is already discussed. As also mentioned in section 3.2, part of finding the optimal strategy is determining the decision rule per point in time and per volume point. If these decision rules are found, the optimal strategy can be found by going through all time steps and related volume levels from the starting volume level.

From the picture above, it can be distinguished that two sources of data and one regression technique are needed for determining the decision rules and to find the corresponding optimal strategy. The first data source is about spot price paths that can be used in the Monte Carlo simulation. These spot price paths are the result of section 4.1. The second source of data is about the physical characteristics of the gas storage itself. For example, the capacity and injection/withdrawal rate of the gas storage. These characteristics determine for example how fast the storage can inject/reject gas, how much gas it can handle in total, and therefore how much value can be created. The mentioned regression technique is the Least-Square Monte Carlo technique. This technique is used to find the continuation value in the optimizing algorithm as discussed in section 3.2.

Since the constructing of gas spot price paths and the Least-Square technique are already discussed, the gas storages characteristics remains to be pointed out in this section. These characteristics are different from one storage to another. That is why we use a fictional gas storage to analyse and evaluate the techniques used in this study. An attempt is made to stick as much as possible to the gas storage characteristics as presented in the paper of Bjerksund et al. (2011). As stated before, the focus of this study is not to value gas storages, its focus is to validate gas storage value(s). Besides, the effect of a time-dependent equilibrium level and volatility updating in the simulation of gas spot prices on the probability distribution of gas storages is analysed. The reference point in Section 4.3 is just the distribution of values with all constant parameters. From this point, time-dependent equilibrium level and volatility updating is incorporated in the simulation of gas spot prices and the effect on the value distribution is discussed.

The gas storage characteristics are programmed in R as input variables. By doing so, these characteristics can easily be adjusted for valuing other storage facilities. The programming code is presented in Appendix D.

The gas storage characteristics are presented in the Table below. The corresponding gas market is UK NBP. Therefore, the unit of gas is therm and it trades in GBP. Corresponding spot and forward prices relate to pence per therm, i.e. GBP per therm.

Initial storage	125 million therms
Terminal storage	125 million therms
Max storage	250 million therms
Injection	2.5 million therms per day
Deletion	2.5 million therms per day

TABLE 14: Gas Storage Characteristics

The contract period is one year. Furthermore it is assumed that there are no transaction costs, and the interest rate is set to zero (by chance this is actually the case in May 2015)[4].

According to the gas storage characteristics one restriction is programmed for computational speed purposes. The maximum injection rate is 2.5 million therms per day and the deletion rate is also 2.5 million therms per day. This results in many different possible changes in volume level per day, so much that the time needed to perform the optimizing strategy is not considered reasonable. To overcome this problem the change in volume level is restricted and discretised to only three possibilities: inject the maximum of 2.5 million therms, do nothing, or deplete 2.5 million therms. This restriction has significant, positive, impact on computational time but it has a negative impact on the gas storage value. Intuitively, this can be understood when other possible volume changes become available additional value can be created when these changes lead to higher payoffs.

[4] In May 2015 the Euribor 3 months is -0.01%, Euribor 6 months is 0.058%, and the Euribor 12 months is 0.167%. Therefore, it is reasonable to set the interest rate at 0.

4.3 Valuation and Interpretation – The Results

In this section the third sub-question is answered, i.e.:

3. *How to obtain a value and interpret the value of gas storages using the spot approach?*

The relation of this sub-question to the overall structure is for convenience presented in Figure 42.

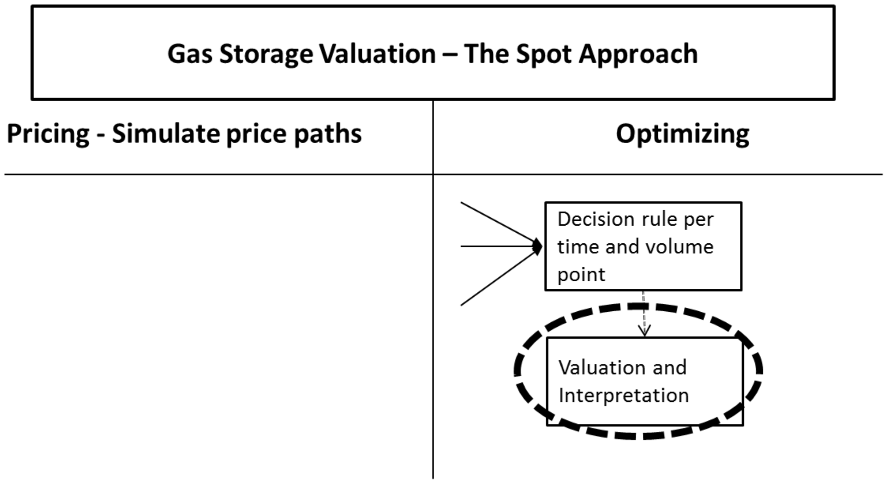

FIGURE 42: From Overall Research Structure; Section 4.3

Recall that the focus of this study is to support the validation of gas storage value(s). In Section 4.1, it is showed that client's valuation models can be challenged to incorporate seasonality and volatility updating in their valuation. This section combines the outcome on sub-question one and sub-question two in order to find a gas storage value. This value is not a claim on the 'real' value but it can be used to determine whether or not the client's value can be considered as reasonable.

By presenting a histogram of the values per price path the effect of a time-dependent equilibrium level (Section 4.1.3) and volatility updating (Section 4.1.4) is shown. For further interpretation the expected value, standard deviation, 5th percentile point, and Expected Shortfall of each distribution are given. As in line with Section 4.1.3 and 4.1.4, 1000 runs per simulation are used. According to computational restrictions and the purpose of validation this is considered appropriate. Recall that the distribution of values represent the risk-neutral probabilities and can therefore not be interpreted as real-world probabilities.

We will first present the distribution when three constant parameters are used as the input of the Ornstein-Uhlenbeck process to simulate future gas spot prices. This distribution is the same as the example used in Section 2.2. So without a time-dependent equilibrium level and volatility updating. This can be used as the baseline of this study. These three constant parameters are calibrated over 5 year of historical data (Section 4.1.2) and are therefore all three backward looking (no market or implied information is taken into account). As stated in Section 4.1.2, both the calibrated mean reverting rate of 0.0137 and rate of 0.05 suggested by Boogert and De Jong (2008) are used. In risk-neutral pricing the equilibrium level is set by the forward curve. We assume in this baseline simulation that the forward curve is the same as the calibrated equilibrium level.

Secondly, the constant equilibrium level is changed to a time-dependent (deterministic) equilibrium level in the Monte Carlo simulation of future gas spot prices. The resulting distribution of gas storage values (one per price path) is visualized in Figure 44. Thirdly, volatility-updating is implemented in the simulation instead of a constant volatility parameter. The other two parameters are constants. The result of this process is shown in Figure 45. At last, both the time-dependent equilibrium level and volatility updating are incorporated in the Monte Carlo simulation of future gas spot prices. The remaining mean reverting parameter is still constant. The result is presented in Figure 46.

Simulation with constant parameters

The three parameters of the Ornstein-Uhlebeck process are here treated as constants. Table 15 presents the input parameters and the expected value. Figure 43 shows the two corresponding histograms, and Table 16 states the 5th percentile point, Expected Shortfall (ES), and standard deviation (sd).

Mean reverting rate; k	0.0137		Mean reverting rate; k	0.05
Equilibrium price; μ	56.8330		Equilibrium price; μ	56.8330
Volatility; σ	3,45%		Volatility; σ	3,45%
Storage value	32.12 million		Storage value	34.38 million

TABLE 15: Gas Storage Input Parameters and Value

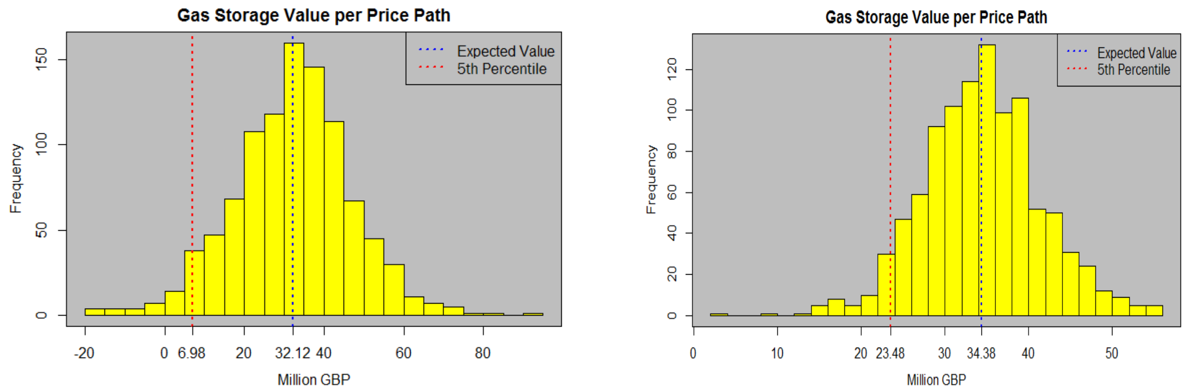

FIGURE 43: Gas Storage Value Distribution with Constant Parameters

5th percentile point	6.98 million		5th percentile point	23.49 million
Expected Shortfall	-0.64 million		Expected Shortfall	19.47 million
Standard deviation	14.76 million		Standard deviation	7.05 million

TABLE 16: Interpretation of Value Distribution

The above results can be seen as the baseline of this study. These results are obtained by using three constant parameters in the Ornstein-Uhlenbeck process. In Figure 43 a histogram of the values per price path is presented. The left figure presents the results with a mean reverting rate of 0.0137, whereas the right figure presents the results with a mean reverting rate of 0.05. The underlying assumption of the OU process is that the returns are normally distributed. Both figures also seems to represent a normal distribution. The expected value of both simulations are quite similar: 32 million versus 34 million. On the other hand, the numbers that give an indication of how the values are distributed give different results. As can be expected, the values of the simulation with the higher mean revering rate are distributed much closer to the mean. This is indicated by a higher 5th percentile point, and ES, and a lower sd. A mean reverting rate of 0.0137 results in twice the sd, a much lower 5th percentile point and a negative ES.

A remarkable effect in relationship with option pricing theory can be obtained from above results. When comparing the storage values in Figure 43, the distribution obtained by a higher mean reverting rate results in a somewhat higher storage value. An higher mean reverting rate would in turn imply a less volatile gas spot price. We see this as quite remarkable because according to option pricing theory, higher volatility in the underlying would result in an higher option price. Using the spot approach with Least Squares Monte Carlo this is not the case.

Discussion: What is a reasonable mean reverting rate?

In the simulations presented in the former sub-section, two mean reverting rates are used. One is obtained by calibration on historical data and one is suggested by literature. In the following simulations we will discuss the effect of changing the equilibrium level and the volatility parameters. Changing these parameters would also have an effect on the mean reverting rate. However, it is hard to determine exactly what this effect would be since we mix-up parameters that are historical (backward-looking) parameters and market consistent (forward-looking) parameters. Further research is needed to quantify this effect. For this study, we regard to this problem from an intuitively perspective.

We implemented a time-dependent equilibrium level to not only represent market expectations but also seasonality in the gas spot price. If seasonality is taken into account in the calibration on historical data, the difference between the observed spot prices and the equilibrium level would be smaller as compared to a constant equilibrium level. This would imply that the mean reverting rate is higher than the number found by calibration on a constant level. We can therefore conclude that the equilibrium level should be higher than 0.0137 for the NBP price. As an upper bound of the mean reverting rate we take the rate of 0.05 as suggested by literature. We consider the mean reverting rate to be somewhere in between. The following simulations are therefore also performed with both mean reverting rates.

Simulation with time-dependent equilibrium level

In the following two simulations, the constant equilibrium level is changed to a time-dependent equilibrium level which is in line with market expectations. The corresponding method is pointed out in Section 4.1.3. Again, the results at the left side relate to a mean reverting rate of 0.0137 and at the right side the results are presented using a mean reverting rate of 0.05.

Mean reverting rate; k	0.0137
Equilibrium price; μ	time-dependent
Volatility; σ	3,45%
Storage value	28.12 million

Mean reverting rate; k	0.05
Equilibrium price; μ	time-dependent
Volatility; σ	3,45%
Storage value	31.56 million

TABLE 17: Gas Storage Input Parameters and Value

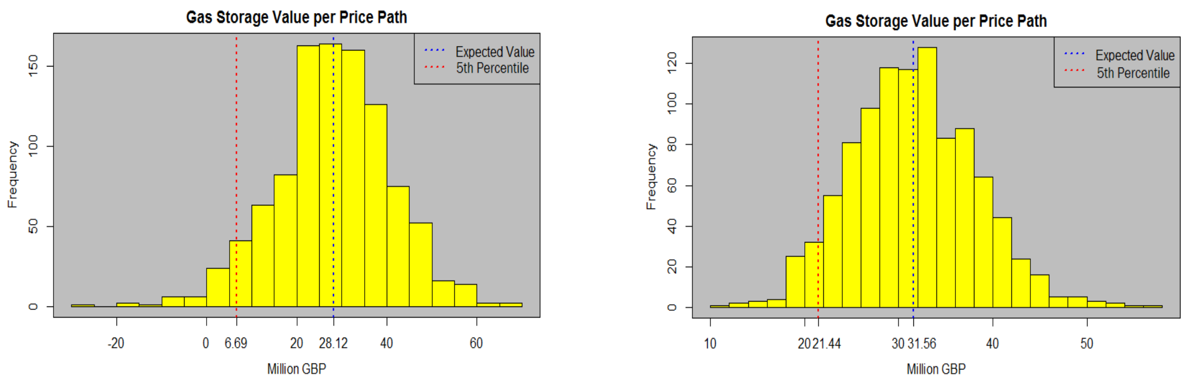

FIGURE 44: Gas Storage Value Distributions with Time-Dependent Equilibrium Level

5th percentile point	6.69 million
Expected Shortfall	-0.52 million
Standard deviation	12.72 million

5th percentile point	21.44 million
Expected Shortfall	18.73 million
Standard deviation	6.59 million

TABLE 18: Interpretation of Value Distribution

The time-dependent equilibrium level used in the above simulations is the same as the forward curve in Figure 29. Instead of a constant level of 56.83, the equilibrium level fluctuates roughly between 46 and 56. By using this forward curve as an equilibrium level, the seasonality throughout the year and throughout the week is incorporated in the simulation of future gas spot prices. Besides, the equilibrium level is in line with market expectations and can therefore be seen as implied.

The, on average, lower equilibrium level results in a somewhat lower expected value. The numbers to describe the tail give very similar results as compared to the simulation with a constant equilibrium level. Remarkable is that incorporating seasonality does not seems to create additional storage value whereas this is traditionally seen as the rationale behind gas storages. This can however not be claimed because the average of the time-dependent equilibrium level is also lower than the one calibrated on historical data (48.9960 versus 56.8330). This is further studied by comparing the results of the time-dependent simulations with simulations in which the equilibrium level is the average of the forward curve in Figure 29. In this way, the equilibrium level is again a constant parameter but instead it is in line with market expectations. The results of these simulations are presented in Table 19:

Mean reverting rate; k	0.0137
Equilibrium price; μ	48.9960
Volatility; σ	3,45%
Storage value	27.76 million

Mean reverting rate; k	0.05
Equilibrium price; μ	48.9960
Volatility; σ	3,45%
Storage value	29.58 million

TABLE 19: Gas Storage Value with Constant Equilibrium Level in Line with Market Expectation

The time-dependent simulations result in a slightly higher storage value. We can therefore conclude that taking account for seasonality will result in a higher storage value. However, the difference is small and it can be discussed whether the time-dependent equilibrium level will lead to significantly different conclusions according to the purpose of validation. Nevertheless, the discussed method is a very easy way to incorporate seasonality and market expectations in the simulation of gas spot prices. We can therefore conclude that using forward curve information in the simulation of future gas spot prices to incorporate seasonality and market expectations is appropriate for validation purposes.

Simulation with volatility updating

The following two simulations take account for volatility updating in the underlying gas spot price. The remaining two parameters are taken as constants as discussed in Section 4.1.2. The updating of volatility is performed by the use of GARCH as explained in Section 4.1.4.

Mean reverting rate; k	0.0137
Equilibrium price; μ	56.8330
Volatility; σ	Volatility updating
Storage value	41.84 million

Mean reverting rate; k	0.05
Equilibrium price; μ	56.8330
Volatility; σ	Volatility updating
Storage value	40.62 million

TABLE 20: Gas Storage Input Parameters and Value

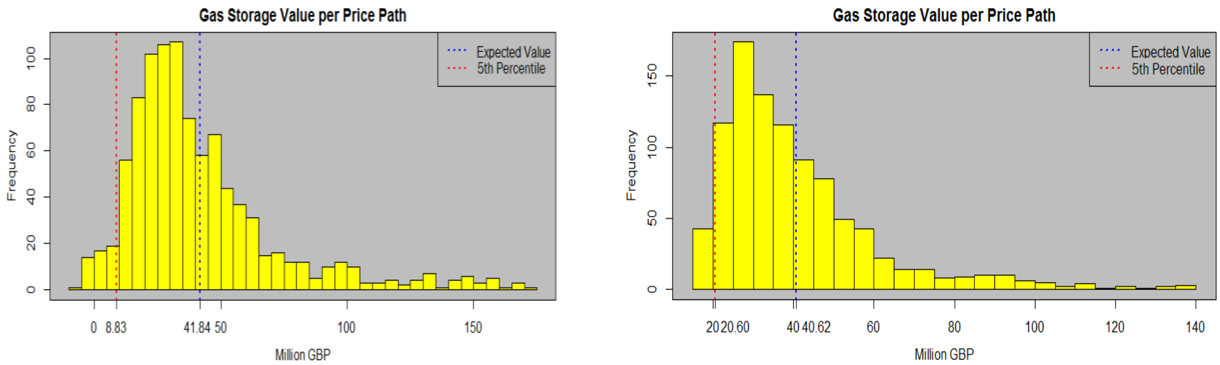

FIGURE 45: Gas Storage Value Distributions with Volatility Updating

5th percentile point	8.83 million
Expected Shortfall	2.61 million
Standard deviation	30.61 million

5th percentile point	20.60 million
Expected Shortfall	18.64 million
Standard deviation	19.68 million

TABLE 21: Gas Storage Input Parameters and Value

When analyzing the above results two remarkable insights are obtained. Firstly, the simulation with a lower mean reverting rate results in a higher expected value whereas our baseline simulation with three constant parameters resulted in the opposite.

Secondly, the 5th percentile point and ES are higher with a mean reverting rate of 0.0137 as compared to the simulation with three constant parameters. As can be seen in Figures 34 and 35, the updating of volatility results in large up and down movements in the simulated future gas spot prices. As in line with expectations this leads to a higher standard deviation in the distribution value of gas storages in comparison with the simulations with constant volatility. A remarkable finding however is that a higher standard deviation in the distribution of gas storage value(s) does not directly relate to a lower 5th percentile point or a lower ES. Instead, for the lower mean reverting rate the 5th percentile point goes up from 7 to 9 million. At the right side, simulations with a higher mean reverting rate, the 5th percentile point is somewhat lower in comparison with constant volatility.

From this we conclude that the optimizing strategy, as discussed in Section 3.2 and Section 4.2, is able to detect up and down price spikes and use it in its advantage. By doing so, the effect of selling gas when prices are high and buying gas when prices are low is higher in these two simulations than in the simulations with constant volatility. Therefore, we can state that when the mean reverting rate is lower the price spikes (up and down) can be exploited more intensively.

Simulation with time-dependent equilibrium level and volatility updating

In this last simulation we incorporate both a time-dependent equilibrium level and volatility updating in the simulation of future gas spot prices. Again, we visualize the effect of these two price characteristics on the distribution of gas storage value(s).

Mean reverting rate; k	0.0137
Equilibrium price; μ	time-dependent
Volatility; σ	Volatility updating
Storage value	39.34 million

Mean reverting rate; k	0.05
Equilibrium price; μ	time-dependent
Volatility; σ	Volatility updating
Storage value	37.97 million

TABLE 22: Gas Storage Input Parameters and Value

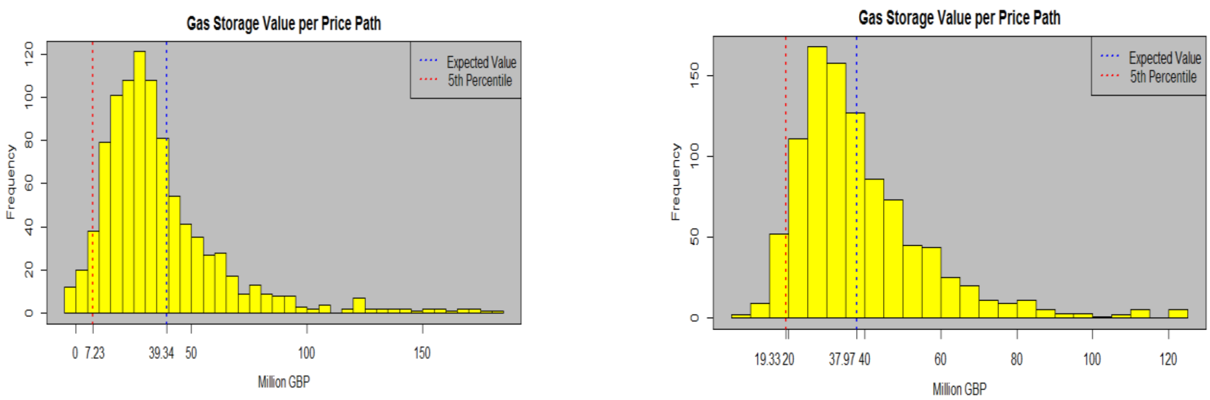

FIGURE 46: Value Distributions with Time-Dependent Equilibrium Level and Volatility Updating

5th percentile point	7.23 million
Expected Shortfall	2.79 million
Standard deviation	27.62 million

5th percentile point	19.33 million
Expected Shortfall	16.71 million
Standard deviation	17.61 million

TABLE 23: Interpretation of Value Distribution

The remarkable insights of the simulation before are also present in the outcome of the above two simulations. Since the time-dependent equilibrium level is on average lower than the constant level, all the values are somewhat lower than the simulations with only volatility updating. The somewhat higher risk associated with the lower mean reverting rate is here rewarded by a higher storage as compared to the simulation with a mean reverting rate of 0.05. The difference is however much smaller.

In these last two simulations both price seasonality and volatility updating are incorporated in the simulation of future gas spot prices. As is discussed before, a reasonable mean reverting rate would be between the calibrated rate of 0.0137 and the suggested rate by literature of 0.05. The effect of this mean reverting rate on the storage value becomes smaller by incorporating volatility updating and a time-dependent equilibrium level. The sensitivity of this rate is further analysed in Section 5.

All in all we conclude that the following aspects can be incorporated in the validation process of gas storages value(s) when using the spot approach:

- Seasonality in the gas spot price
- Market expectations of future gas spot prices
- Volatility clustering in the gas spot price

The first two aspects are taken into account for by the time-dependent equilibrium level. The volatility clustering is implemented in the simulation of gas spot prices by GARCH. By taking account of volatility clustering we also deal with the existence of price spikes in gas prices. The optimizing part of the spot approach is able to take these (extreme) price movements into account as can be seen by the big, positive, right tail in the distribution of gas storage value(s).

5 Sensitivity Analysis

In this section we perform two sensitivity analyses. Both are related to an in-model parameter of the Ornstein-Uhlenbeck process: the mean reverting rate. By using the spot approach to 'value' gas storages the behaviour of the underlying gas spot price is important. For simulating this spot price an Ornstein-Uhlenbeck process is used with three parameters. Two of these parameters are adjusted to represent seasonality and volatility clustering in spot prices. The third one, the mean reverting rate, is in this study treated as a constant. As stated in Section 4.1.2 and 4.3, it is hard to determine what the mean reverting rate should be when changing the other two parameters of the Ornstein-Uhlenbeck process. We therefore used two mean reverting rates: the calibrated rate of 0.0137 found by calibration on historical data and the rate of 0.05 suggested by Boogert and De Jong (2008). We stated that the relevant rate would be somewhere in between those two boundary-rates. The sensitivity of this mean reverting rate to the gas storage value and its tail is therefore analysed in this section. We present two plots with the gas storage value (its expected value) and the 5th percentile point at both axes. Each plot relates to one price characteristic that is used in the simulation of future gas prices. The first characteristic is the time-dependent equilibrium level to take account for seasonality and market expectations. The second characteristic is volatility updating to take account for the clustering of volatility.

The Sensitivity of the Mean Reverting Rate and Time-Dependent Equilibrium Level

As stated above, the sensitivity of the mean reverting rate on the gas storage value and the 5th percentile point is calculated. At this sub-section the smoothed forward curve is used to represent a time-dependent equilibrium level to take account for the seasonality of gas spot prices in a way which is consistent with market expectations. The corresponding plot is presented below.

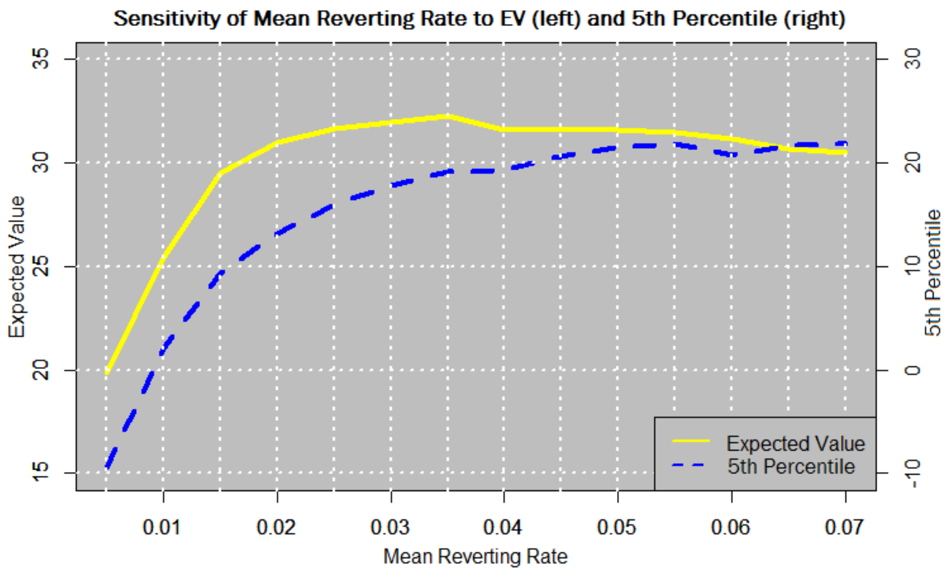

FIGURE 47: Sensitivity of the Mean Reverting Rate – Equilibrium Level

In the above figure the range of the mean reverting rate is between 0.005 and 0.07 with steps of 0.005. The graphs should be interpreted with care since different axes are used. The expected value corresponds with the left axis whereas the 5th percentile point relates to the right one.

As can be seen in Figure 47, the expected value has an upward slope until it reaches its maximum around 0.035. After that, the expected value slightly decreases. This makes sense because the ability of the gas storage to profit from price changes shrinks when the mean reverting rate becomes larger. The 5th percentile point behaves exactly as expected, it is positively related to the mean reverting rate. A higher mean reverting rate indicates that price paths do not fluctuate much, so all the gas storage values are more near to the mean.

The Sensitivity of the Mean Reverting rate and Volatility Updating

At this sub-section the sensitivity of the mean reverting rate is tested when volatility updating is used in the simulation of gas spot prices. Again, the mean reverting rate ranges from 0.005 to 0.07 with steps of 0.005 and two different axis are used. The left axis corresponds with the expected value and the right axes relates to the 5^{th} percentile point.

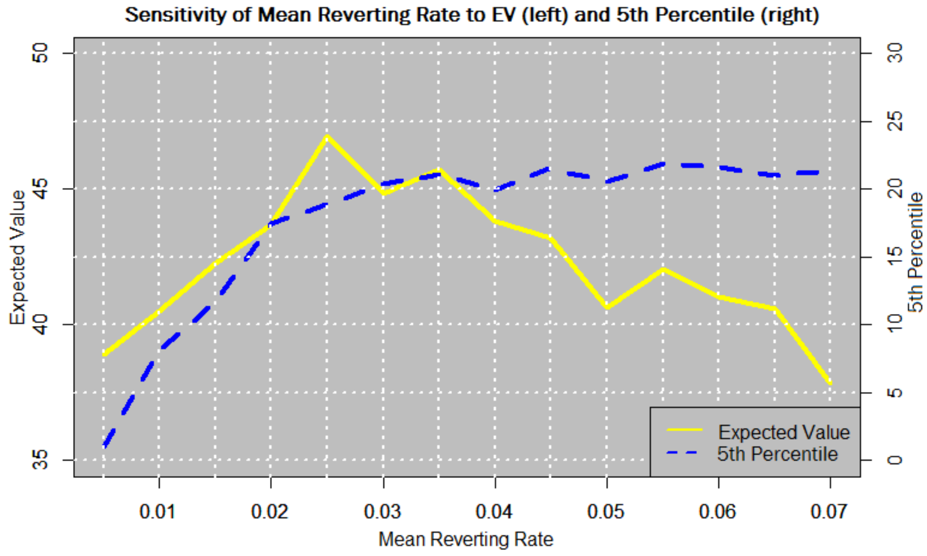

FIGURE 48: Sensitivity of the Mean Reverting Rate – Volatility Updating

In Figure 48 the graph that represents the expected value is not as stable as it is in Figure 47. It has the same upward slope as in Figure 47, but after it reaches it maximum around 0.03, it decreases a lot faster. On the other hand, the 5^{th} percentile point moves in a very identical way. This quite unstable behaviour may be due to the limited amount of runs per simulation. However, if we imagine a more smoothed curve, the difference in value would still be around 10% in the reasonable range (0.0137 – 0.05).

Conclusion

We conclude that the sensitivity of the mean reverting rate to the 5^{th} percentile point is quite low in the reasonable range. However, when taking volatility updating into account in the simulation of future gas prices, the mean reverting rate can have an impact on the expected value of the storage. Our findings indicate a difference up to 10% of the storage value. Since the mean reverting rate was not the focus of our study we cannot state this by heart. We can only state that in the validation of gas storage value(s) the sensitivity of the mean reverting rate should be taken into account. How to set an appropriate mean reverting rate is therefore a good subject for future research.

6 Conclusions and recommendations

In this study we presented two extensions to the method of Boogert and De Jong (2008) to value gas storages using the spot approach. This is done from a validation perspective to support audit work on assets like gas storages. The first extension is to incorporate seasonality in the simulation of future gas prices. The second extension relates to the clustering of volatility present in gas spot prices. The main research question is therefore as follows:

How to incorporate seasonality and volatility updating in gas storage valuation for the purpose of validation?

We will first discuss our findings followed by suggestions for future research.

6.1 Conclusions

We focused on the pricing side of the spot approach and found two extensions to the method of Boogert and De Jong (2008) for the purpose of validation. The corresponding optimizing algorithm is used to show the effect of these extensions on the value distribution of gas storages. By performing an econometric analysis on historical spot prices the existence of seasonality and the clustering of volatility is shown. Each extension relates to a parameter of the mean reverting Ornstein-Uhlenbeck process (E. S. Schwartz, 1997).

The first extension is the implementation of a time-dependent equilibrium level to take account for the seasonality of the gas spot price. The forward curve at validation date is used to set an equilibrium level that is consistent with market expectations. Smoothing a spline through this forward curve is needed to transform this curve of average-based forward prices to a curve with daily granularity. On top of that, a day-week profile is added to also take account for the significant seasonality through the week.

The implementations of a GARCH model to represents clustering in volatility is the second extension. As shown by the econometric analysis, returns on the gas spot prices are not normally distributed and subjected to fat tails. Since gas storage holders profit from extreme movements in price a simulation conditional on normally distributed returns underestimates gas storage value. We showed that by using a GARCH (1,1) model extreme price movements can be taken into account in the valuation and validation of gas storages.

To analyse the impact of above improvements on the valuation and validation of gas storages a baseline is created. As baseline we used a distribution of values when simulating future gas pricing with three constant parameters in the Ornstein-Uhlenbeck process. These constants are found by calibration on historical data.

A time-dependent equilibrium level results in a higher storage value as compared to a constant equilibrium level. However, this difference can be quite small. Nevertheless, this is a welcome extension because in this way market expectations are taken into account in the valuation, and validation. If market expectations show a much larger seasonal effect, this effect is automatically accounted for.

When a GARCH model is used to take account for volatility updating much higher storage values are obtained. This is because the optimizing algorithm is able to use extreme price movements in its advantage. Intuitively this makes sense because storage holders profit from large price movements. The effect of volatility updating is best shown by big, positive, right tails in the corresponding value distribution. We conclude that this is an appropriate improvement because the baseline method underestimates extreme price movements and therefore gas storage value.

All in all we conclude that the following aspects can be incorporated in the validation process of gas storages value(s) when using the spot approach:

- Seasonality in the gas spot price
- Market expectations of future gas spot prices
- Volatility clustering in the gas spot price

These aspects can be used for model validation. If these techniques are used for having a claim on the one and 'true' value, they should be handled with care. This is because of the underlying assumptions and the impact of the mean reverting rate. Nevertheless, it can be used for obtaining an intuition about reasonable values.

The parameter that is untouched in this study is the mean reverting rate. We cannot state by heart how to set an appropriate rate since it was not the focus of this study. Nevertheless, our sensitivity analysis showed that the mean reverting rate can have an impact on the expected value and the 5th percentile point. Validation on gas storages based on this study should therefore take account for this potential impact. We find the mean reverting rate an interesting topic for further research.

6.2 Further Research

While conducting this study we found interesting topics related to gas storage valuation and validation that are not covered by this study. The first one relates to the mean reverting rate which is one of the three parameters in the Ornstein-Uhlebeck process. The discussed extensions in this study corresponds to the other two parameters. However, the impact of changing these parameters on the mean reverting rate is unknown. Further research is needed to determine how to set an appropriate mean reverting rate conditional on the changes in the other parameters.

Secondly, there are possibilities to make the volatility parameter market consistent. In this study the volatility is calibrated on historical data whereas the equilibrium level parameter is 'looking-forward' by using market consistent, forward prices. We suggest further research on this topic to use for example implied volatility in simulation of future gas prices.

Another possibility is to use simulated forward curves by principal component analysis for setting the time-dependent equilibrium level. A first suggestion is already made in this study from witch further research can be continued. Although, we consider the potential advancements for the purpose of validation quite small.

Fourthly, we suggest further research on the discounting rate to be used in this study. We used the risk-free interest rate as the discounting factor because we used risk neutral valuation. Following this technique we assumed that discounting by the risk-free rate is incorporated in forward prices. However, in literature it is discussed that the market price of risk is incorporated in forward prices (E. Schwartz & Smith, 2000). Further research on this topic will therefore help in setting an appropriate discounting factor.

The last suggestions for further research relates to the GARCH model. We used the (1,1) model to simulate volatility updating and found satisfying results for the purpose of validation. However, we did not determine if this model is the most suited one. Furthermore, by using GARCH the chance of a positive price spike is of equal magnitude as a negative price spike. This seems not to be in line with our empirical analysis because only positive price spikes are detected. A potential way to improve GARCH volatility forecasts is by a regime-switching model. Research on this topic is relevant since the impact of volatility updating on gas storage value(s) is high.

Works Cited

Back, J., & Prokopczuk, M. (2013). Commodity Price Dynamics And Derivative Valuation: A Review. *International Journal of Theoretical and Applied Finance, 16*(06).

Back, J., Prokopczuk, M., & Rudolf, M. (2013). Seasonality and the valuation of commodity options. *Journal of Banking & Finance, 37*(2), 273-290. doi: 10.1016/j.jbankfin.2012.08.025

Bates, D. S. (2003). Empirical option pricing: A retrospection. *Journal of Econometrics, 116*(1), 387-404.

Benth, F. E., Koekkebakker, S., & Ollmar, F. (2007). Extracting and applying smooth forward curves from average-based commodity contracts with seasonal variation. *The Journal of Derivatives, 15*(1), 52-66.

Bernard, J. T., Khalaf, L., Kichian, M., & McMahon, S. (2008). Forecasting commodity prices: GARCH, jumps, and mean reversion. *Journal of Forecasting, 27*(4), 279-291.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., & Smoller, M. M. (1995). Mean reversion in equilibrium asset prices: evidence from the futures term structure. *The Journal of Finance, 50*(1), 361-375.

Bjerksund, P., Stensland, G., & Vagstad, F. (2011). Gas Storage Valuation: Price Modelling v. Optimization Methods. *Energy Journal, 32*(1), 203-227.

Björk, T. (2004). *Arbitrage theory in continuous time*: Oxford university press.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, 637-654.

Bloomberg, L. P. (2014). *Last Traded Spot Prices 2010 to 2014*. Bloomberg Database.

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

Boogert, A., & De Jong, C. (2008). Gas storage valuation using a Monte Carlo method. *Journal of Derivatives, 15*(3), 81-98.

Boogert, A., & de Jong, C. (2011). Gas storage valuation using a multifactor price process. *The Journal of Energy Markets, 4*(4), 29-52.

Breslin, J., Clewlow, L., Elbert, T., Kwok, C., & Strickland, C. (2008). Gas Storage: Overview & Static Valuation. *Lacima Group*.

Dixit, A. K., & Pindyck, R. S. (1994). Investment Under Uncertainty. *Princeton, NJ: Prince-ton University Press*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Felix, B. J., & Weber, C. (2008). *Gas storage valuation: comparison of recombining trees and least squares Monte-Carlo simulation.* Paper presented at the Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International.

GasTerra (Producer). (2014, 9-3-2015). Retrieved from http://www.gasterra.nl/en/csr/gas

Gray, J., & Khandelwal, P. (2004). Realistic Natural Gas Storage Models II: trading strategies. *commodities Now, 1-4*.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2): Princeton university press Princeton.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics, 20*(7), 873-889.

Hildmann, M., Herzog, F., Stokic, D., Cornel, J., & Andersson, G. (2011). *Robust calculation and parameter estimation of the hourly price forward curve.* Paper presented at the 17th Power Systems Computation Conference (PSCC).

Hull. (2012). *Risk Management and Financial Institutions,+ Web Site* (Vol. 733): John Wiley & Sons.

Hull, J. C. (2006). *Options, futures, and other derivatives*: Pearson Education India.

Insley, M., & Rollins, K. (2005). On solving the multirotational timber harvesting problem with stochastic prices: a linear complementarity formulation. *American Journal of Agricultural Economics, 87*(3), 735-755.

International Energy Agency. (2011). World Engergy Outlook - Are we entering a golden age of gas? Retrieved 11-03-2015, 2015, from http://www.worldenergyoutlook.org/goldenageofgas/

Jarque, C. M. (2011). Jarque-Bera test *International Encyclopedia of Statistical Science* (pp. 701-702): Springer.

Johnson, P. V. (Producer). (2012, 1-6-2015). Least Squares Monte Carlo. Retrieved from http://www.maths.manchester.ac.uk/~pjohnson/resources/math60082/lecture-monte-carlo-ls.pdf

Kholopova, M. (2006). Estimating a two-factor model for the forward curve of electricity.

Kim, J., & Park, J. Y. (2013). Mean reversion and unit root properties of diffusion models: Working paper, Indiana University.

Koekebakker, S., & Ollmar, F. (2005). Forward curve dynamics in the Nordic electricity market. *Managerial Finance, 31*(6), 73-94.

Lai, G., Margot, F., & Secomandi, N. (2010). An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research, 58*(3), 564-582. doi: 10.1287/opre.1090.0768

Longstaff, F. A., & Schwartz, E. (2001). Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies, 14*(1), 113-147.

Maragos, S., & Ronn, E. (2002). Valuation of the operational flexibility of natural gas storage reservoirs. *Real Options and Energy Management Using Options Methodology to Enhance Capital Budgeting Decisions*, 431-456.

Mirantes, A. G., Poblacion, J., & Serna, G. (2013). The stochastic seasonal behavior of energy commodity convenience yields. *Energy Economics, 40*, 155-166. doi: 10.1016/j.eneco.2013.06.011

Moreno, M., & Navas, J. F. (2003). On the robustness of least-squares Monte Carlo (LSM) for pricing American derivatives. *Review of Derivatives Research, 6*(2), 107-128.

Pindyck, R. S. (2001). The dynamics of commodity spot and futures markets: a primer. *The Energy Journal*, 1-29.

Pindyck, R. S. (2003). Volatility in natural gas and oil markets.

Richardson, M. (2009a). Numerical methods for option pricing. *Unpublished notes*.

Richardson, M. (Producer). (2009b, 13-5-2015). Principal Component Analysis Retrieved from http://people.maths.ox.ac.uk/richardsonm/SignalProcPCA.pdf

Rohrbein, N. (Producer). (2011, 2-6-2015). Commodities: Grub first, then ethics. Retrieved from http://www.ipe.com/commodities-grub-first-then-ethics/40378.fullarticle

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

Schwartz, E., & Smith, J. (2000). Short-term variations and long-term dynamics in commodity prices. *Management Science, 46*(7), 893-911.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance, 52*(3), 923-973.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 591-611.

Smith, A., & Nicholson, J. S. (1887). *An Inquiry Into the Nature and Causes of the Wealth of Nations*: T. Nelson and Sons.

Tibs, R. (Producer). (2014, 2-4-2015). Smoothing Splines. Retrieved from http://www.stat.cmu.edu/~ryantibs/advmethods/notes/smoothspline.pdf

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99-114.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical review, 36*(5), 823.

Wu, T. T. (Producer). (2004, 2-4-2015). Introduction to Smoothing Splines. Retrieved from www.ms.uky.edu/~mai/biostat277/smsp.ppt

Zhao, L., & van Wijnbergen, S. (2013). A real option perspective on valuing gas fields: Tinbergen Institute Discussion Paper.

Zivot, E. (Producer). (2006, 12-5-2015). Time Series Econometrics. Retrieved from http://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf

A. Research Structure

The overall structure of this research is visually presented in the figure below.

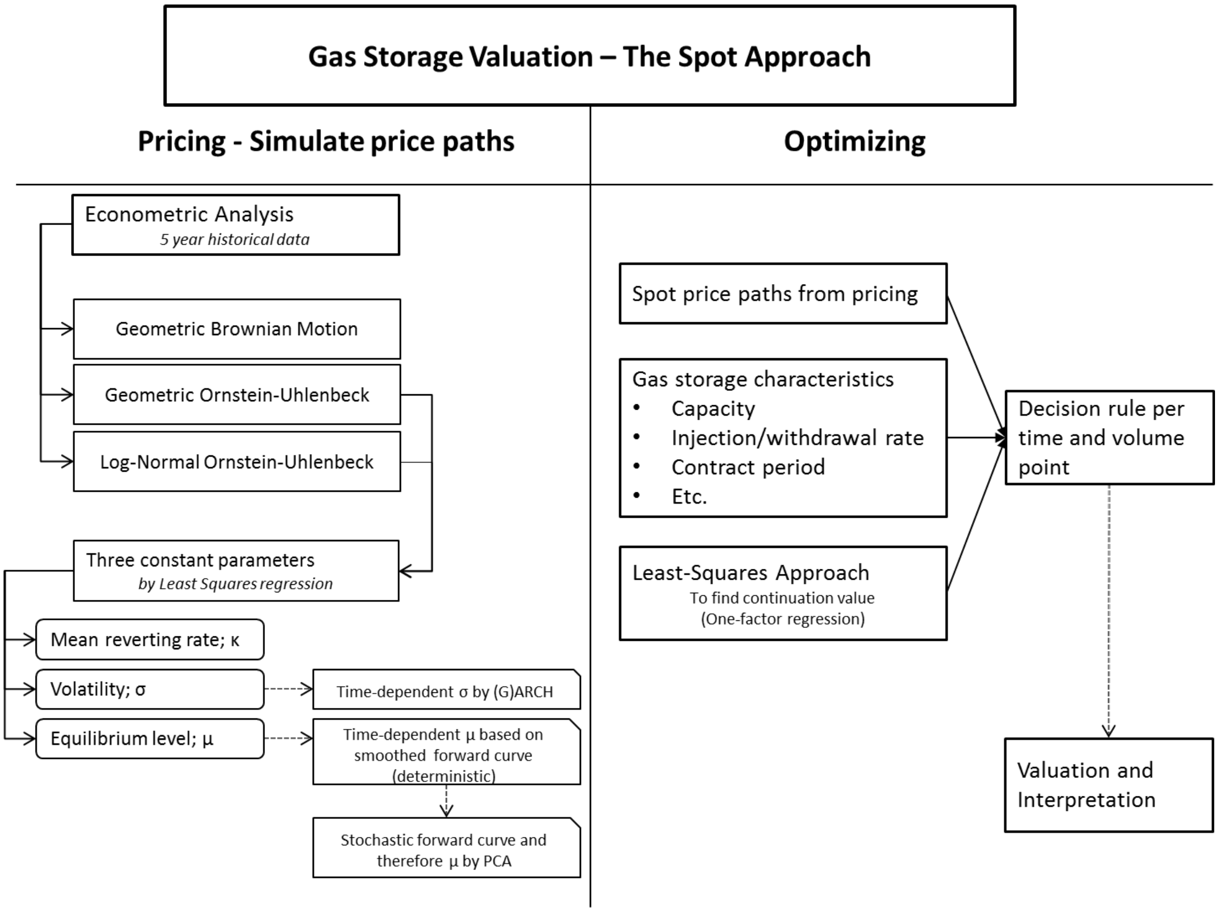

FIGURE 49: Overall Research Structure

B. Econometric Analysis HH Spot Price

The econometric analysis performed on the NBP spot price is also applied on the US Henry Hub spot price. The results can be seen in the following figures and tables.

FIGURE 50: HH - Spot Price and its Returns

	Henry Hub (HH)			
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))	
Mean	3.8405	6.4245e-05	-0.0006	
Min	1.8198	-0.2359	-0.2691	
Max	7.9247	0.4604	0.3787	
Quantile:				
• 1%	1.9836	-0.0783	-0.0815	
• 5%	2.3788	-0.0449	-0.0459	
• 10%	2.7357	-0.0326	-0.0331	
• 90%	4.6840	0.0322	0.0317	
• 95%	5.0136	0.0445	0.0436	
• 99%	6.1815	0.0884	0.0847	
Standard Deviation	0.8220	0.0362	0.0353	
Skewness	0.4040	2.4974	0.9983	
Kurtosis	5.1939	40.3301	27.7459	
Number of observations		1253	1252	1252

TABLE 24: HH – Descriptive Statistics 2010 - 2014

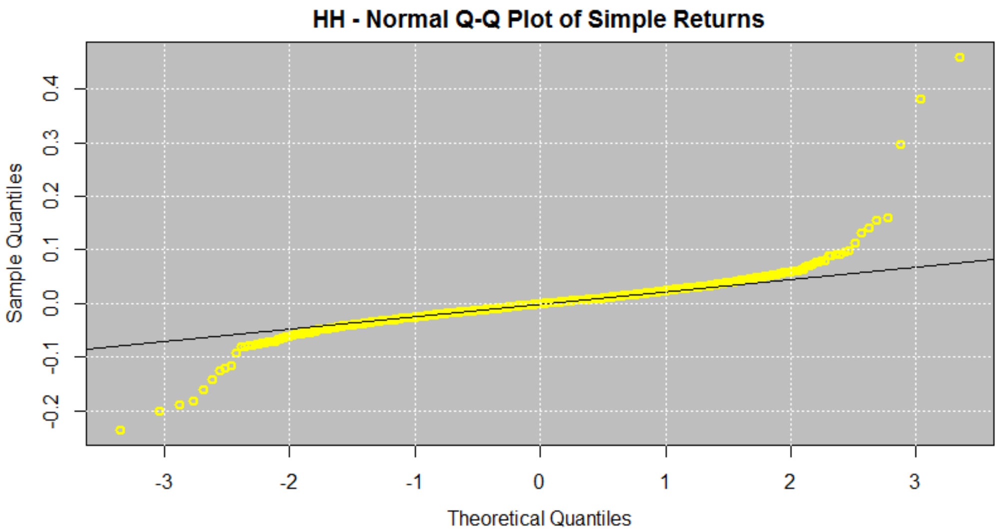

FIGURE 51: HH - Normal Q-Q Plot of Simple Returns

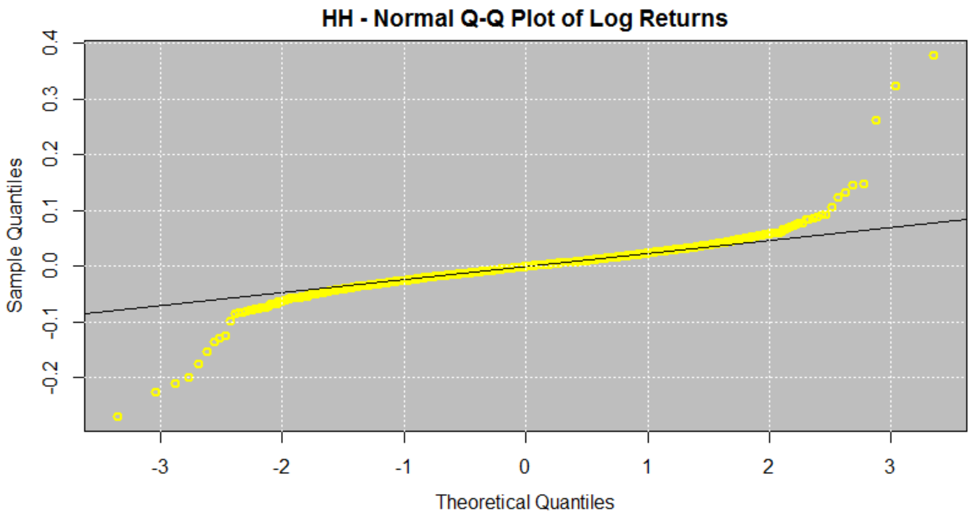

FIGURE 52: HH – Normal Q-Q Plot of Log Returns

	Henry Hub (HH)		
	Spot Price (S)	‘simple’ returns (%ΔS)	Log returns (Δln(S))
Shapiro-Wilk	2.2e-16	2.2e-16	2.2e-16
Jarque-Bera	2.2e-16	2.2e-16	2.2e-16

TABLE 25: HH - P-values Normality Tests

	Henry Hub (HH)		
	Spot Price (S)	‘simple’ returns (%ΔS)	Log returns (Δln(S))
Phillips-Perron test	0.0261	0.01	0.01
KPSS test	0.0132	0.1	0.1

TABLE 26: HH - Statistical Tests Volatility Clustering

C. Econometric Analysis TTF Spot Price

The econometric analysis performed on the NBP spot price is also applied on the Dutch Title Transfer Facility spot price. The results can be seen in the following figures and tables.

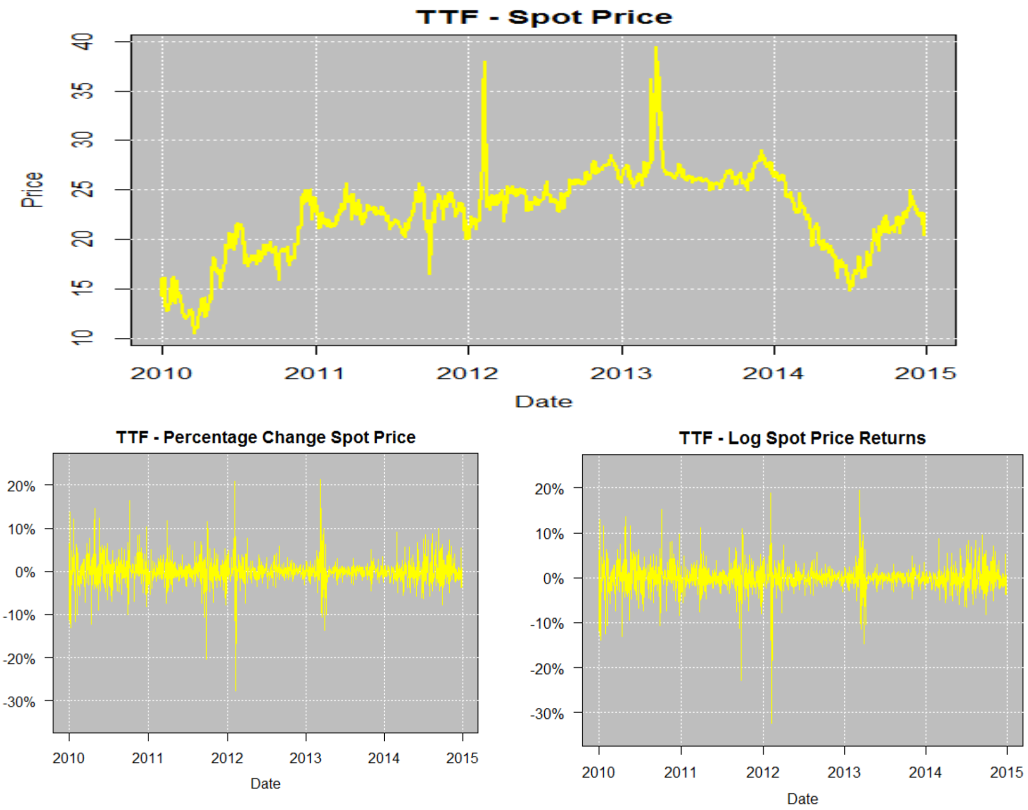

FIGURE 53: TTF - Spot Price and its Returns

	Title transfer Facility (TTF)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Mean	22.6115	0.0007	0.0002
Min	10.50	-0.2763	-0.3234
Max	39.50	0.2138	0.1938
Quantile:			
• 1%	12.2000	-0.0999	-0.1052
• 5%	14.1975	-0.0460	-0.0471
• 10%	16.9500	-0.0290	-0.0294
• 90%	26.9820	0.0336	0.0330
• 95%	27.6000	0.0518	0.0505
• 99%	32.8525	0.0996	0.0949
Standard Deviation	4.1568	0.0332	0.0334
Skewness	-0.3670	-0.1005	-0.7613
Kurtosis	3.7640	13.0515	15.8343
Number of observations	1260	1259	1259

TABLE 27: TTF – Descriptive Statistics 2010 - 2014

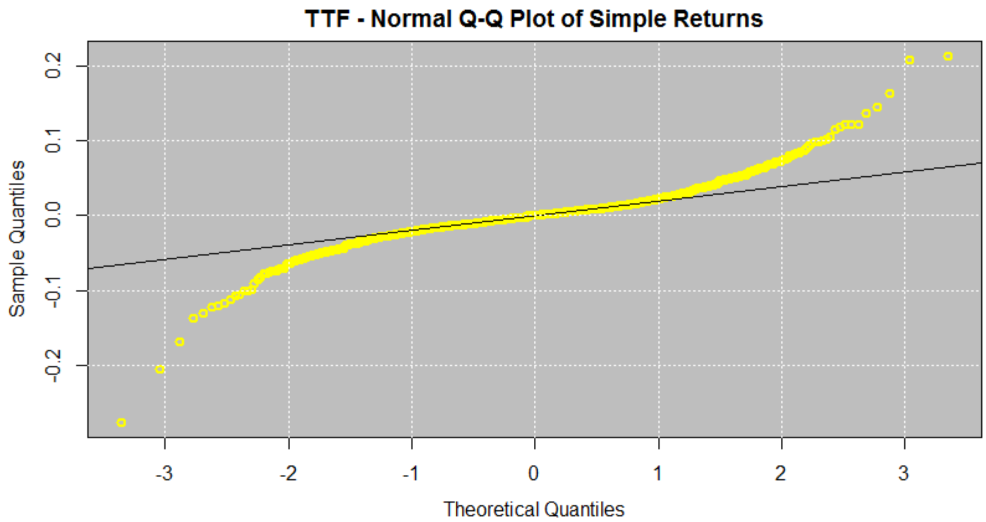

FIGURE 54: TTF - Normal Q-Q Plot of Simple Returns

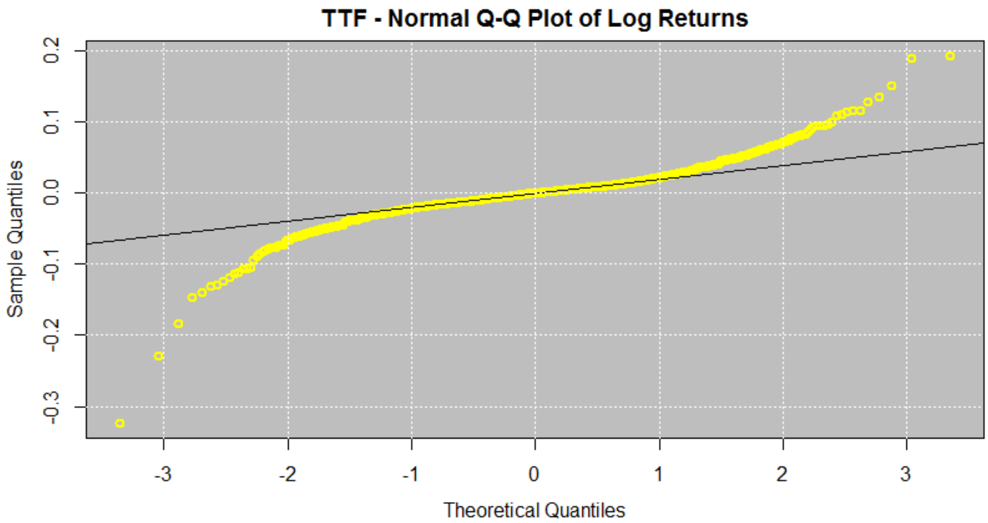

FIGURE 55: TTF – Normal Q-Q Plot of Log Returns

	Title transfer Facility (TTF)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Shapiro-Wilk	2.2e-16	2.2e-16	2.2e-16
Jarque-Bera	1.6e-13	2.2e-16	2.2e-16

TABLE 28: TTF - P-values Normality Tests

	Title transfer Facility (TTF)		
	Spot Price (S)	'simple' returns (%ΔS)	Log returns (Δln(S))
Phillips-Perron test	0.1873	0.01	0.01
KPSS test	0.01	0.1	0.1

TABLE 29: TTF - Statistical Tests Volatility Clustering

D. Programming Code

This appendix presents the relevant R code per section.

Section 4.1.1 Spot Price Analysis

```
#Month-Year Profile; Section 4.1.1

############# Now a boxplot of the months
Data=read.csv("", header=T, sep=";",dec=",")

######## Sort
Data$Month <-factor(Data$Month, levels=c("December","November","October","September",
"August","July","June","May","April","March","February","January"))

######## Set margins of par
par(mar=c(3.8,5.5,2,2))
par(mgp=c(2.2,1,0))

########### EY Colors
EYY=c(255,210,0)
EYG=c(128,128,128)
#BB=rbind(rep(EYY,length(Data[,2])))
###########

####### create a boxplot of the year-profile
BB=rbind(EYY,EYY,EYY,EYY,EYY,EYY,EYY,EYY,EYY,EYY,EYY,EYY)
boxplot(Data$Price ~ Data$Month, horizontal=TRUE,las=1, ylim=c(30,80))
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "gray")
boxplot(Data$Price ~ Data$Month, horizontal=TRUE,las=1,main="Boxplot Year-Profile", col=EYG,
border=BB,add=TRUE,xaxt='n',yaxt='n',lwd=2,xlab="Price")
abline(v=c(30,40,50,60,70,80), lwd=1, col="white", lty="dotted")
abline(h=c(1,2,3,4,5,6,7,8,9,10,11,12),lwd=1, col="white",lty="dotted")

####### create an ANOVA table
ratio.aov<-aov(Data$Price~Data$Month)
summary(ratio.aov)

#difference exists because P-value is very small
###### Conduct a Turkey's multiple comparison procedure to find the differences
TukeyHSD(ratio.aov)

#Statistical tests to test for unit root and/or stationarity.
#section 4.1.1.

#Load data
MR=HH[,2]
MR=NBP[,2]
MR=TTF[,2]

MR=log(MR)
MR=diff(MR)

#MEAN REVERSION
library(tseries)
library(fUnitRoots)

MR=DST

#Augemted Dickey-Fuller (ADF) to test mean reversion in a timeserie
adf.test(MR)  #Here a constant and a linear trend is used(!), whereas only a constant is
needed.
#May be a good proxi for the lag
adfTest(MR, type="c", lags=10)   #set type to "nc" ,"c", or "ct", for no constant, constant,
and constant & trend
#KPSS Test for Stationarity
kpss.test(MR, null="Level", lshort=F)   #null hypothesis should be level or 'trend', in this
case level.
#Phillips-Perron Test for Unit Roots
PP.test(MR, lshort=F)
```

Section 4.1.2 Stochastic Processes

```
#Log-Normal Ornstein Uhlenbeck parameter estimation and simulation
#Section 4.1.2

#Read in Data
prices=read.csv("", header=T, sep=";",dec=",")
NBP=prices[1:1258,3:4]

#Load data
price.HH=ts(HH[,2])
price.NBP=ts(NBP[,2])
price.TTF=ts(TTF[,2])

ts.plot(price.HH,price.NBP,price.TTF, lwd=1, col=c("blue","red","orange"))

######### Set which price process to simulate ######
price=price.NBP
#########

######### Make it log prices
price=log(price)
#########

######### Estimation of OU parameters using linear model##### !!!!! Following regression on
lnprices after Ito's Lemma Schwartz (1997)
D=diff(price)
S=price[-length(price)]
list=summary(lm(D~S))        #Simple linear regression
Eta=-list$coefficients[2]
Sigma=list$sigma
alpha=list$coefficients[1]/Eta
Mu=alpha+(Sigma^2/(2*Eta))
#########

##### SET T #####
T=365

###### Simulation of the OU process following Schwartz (1997) #######
OU.sim <- function(t=T+1, mu=Mu, eta=Eta, sigma =Sigma){
  P_0=price[length(price)]
  P=rep(P_0,t)
  for(i in 2:t){
    P[i]=P[i-1] + eta * (mu - (sigma^2/(2*eta)) - P[i-1]) + sigma * rnorm(1)
  }
  return(P)
}
#####

############ function for multiple runs of OU.Sim into a matrix #########
OU.Simulations=function(runs){
  X=matrix(1:((T+1)*runs),nrow=(T+1),ncol=runs)
for(i in 1:runs){
  X[,i]=exp(OU.sim())
}
X
return(X)
}
#############

X=OU.Simulations(3) #5 simulations of OU.Sim()
ts.plot(X)

############# Plot ###########

Y=matrix(nrow=T+1, ncol=ncol(X)+1)
Y[,1:(ncol(X))]=X
Y[,(ncol(X))+1]=rowMeans(X) # calc average and add it to matrix
ts.plot(Y,xlab="Time", ylab="Price",
col=c(rep("black",ncol(X)),"blue"),lwd=c(rep(1,ncol(X)),2))
###########
```

Section 4.1.3 Forward Curve Analysis

```
#Smoothing forward curve by a spline + Day-week Profile
#Section 4.1.3

Data=read.csv("", header=T, sep=";",dec=",")
library(timeDate)
library(timeSeries)
library(zoo)

########### Set timeSeries
fwd=timeSeries(Data[,2])
plot(fwd)
###########

########### EY Colors
EYY=c(255,210,0)
EYG=c(128,128,128)
BB=rbind(rep(EYY,length(Data[,2])))
###########
######## Set margins of par
par(mar=c(4,4,2,2))
par(mgp=c(2.5,1,0))

########### Plot
Data.ts=ts(Data[,2])
plot(Data.ts,type="p",xaxt="n",pch=20,col=EYY,main="NBP-Forward
Contracts",xlab="Date",ylab="Price",ylim=c(46,56),xlim=c(1,705))
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "gray")
points(Data.ts,col="yellow",pch=20)
axis(1, at=c(1,91,182,274,366,457,548,640),labels=c("Jan-15","Apr-15","Jul-15","Oct-15","Q1-
2016","Q2-16","Q3-16","Q4-16"))
abline(v=c(1,91,182,274,366,457,548,640), lwd=2, col="white", lty="dotted")
abline(h=c(46,48,50,52,54,56),lwd=2, col="white",lty="dotted")
##########

#######smoothSpline
fwd.Spline<- smoothSpline(fwd, spar=0.75)
plot(fwd.Spline, main="forward curve - Spline Smoothed")
fwd.m<-as.matrix(fwd.Spline[,2])
#####

######## day-week ratios
Mo=1.0000
Tu=1.0045
We=1.0059
Th=1.0044
Fr=1.0003
Sa=0.9927
Su=0.9927

##### loop to add day-week profile to fwd SET PERIOD 1 OR 2 YEAR
fwd.dw<-fwd.m
for (i in 1:731){
  if ((i+3)%%7==1) {      # 1-jan is on a Thursday
      fwd.dw[i]=fwd.dw[i]*Mo
  }
  else if ((i+3)%%7==2){
      fwd.dw[i]=fwd.dw[i]*Tu
  }
  else if ((i+3)%%7==3){
    fwd.dw[i]=fwd.dw[i]*We
  }
  else if ((i+3)%%7==4){
    fwd.dw[i]=fwd.dw[i]*Th
  }
  else if ((i+3)%%7==5){
    fwd.dw[i]=fwd.dw[i]*Fr
  }
  else if ((i+3)%%7==6){
    fwd.dw[i]=fwd.dw[i]*Sa
  }
  else if ((i+3)%%7==0){
    fwd.dw[i]=fwd.dw[i]*Su
  }
}
```

```
######

###### Plot all curves
fwd.ts=as.timeSeries(Data[,2])
fwd.m.ts=as.timeSeries(fwd.m)
fwd.dw.ts=as.timeSeries(fwd.dw)
#####

#####show plotjes in [3,1] frame
par(mfrow=c(3,1))
plot(fwd.ts)
plot(fwd.m.ts)
plot(fwd.dw.ts)
par(mfrow=c(1,1))
#######

####### graphs in same plot
plot(fwd.ts, ylim=c(45,56))
lines(fwd.m.ts,type="l", col="blue")
lines(fwd.dw.ts,type="l",col="red")
#####

########### Plot
Data.ts=ts(Data[,2])
plot(Data.ts,type="p",xaxt="n",pch=20,col=EYY,main="NBP Smoothed Forward Curve with Day-Week
Profile",xlab="Date",ylab="Price",ylim=c(46,56),xlim=c(1,705))
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "gray")
points(Data.ts,col="yellow",pch=20)
axis(1, at=c(1,91,182,274,366,457,548,640),labels=c("Jan-15","Apr-15","Jul-15","Oct-15","Q1-
2016","Q2-16","Q3-16","Q4-16"))
abline(v=c(1,91,182,274,366,457,548,640), lwd=2, col="white", lty="dotted")
abline(h=c(46,48,50,52,54,56),lwd=2, col="white",lty="dotted")
lines(fwd.m.ts,type="l", lwd=2, col="blue")
lines(fwd.dw.ts,type="l",col="red")
##########


#PCA
#Section 4.1.3
library(rgl)

#Load Data
Data=read.csv("",header=T,sep=";",dec=",")

#########LAY-OUT
EYY=c(255,210,0)
EYG=c(128,128,128)
colvec<-rbind("yellow","blue")
#
# Set par
par(mar=c(3.8,3.8,2,2))
par(mgp=c(2.2,1,0))
#########

######### Annualized Volatility Term Structure
ldr=matrix(1:(256*365),c(256,365))

for (i in 1:365){
  for (j in 1:256){
    ldr[j,i]=(Data[i+1,j*4]) # Don't take spot into account
  }
}

yvol=array(1:365)
for (i in 1:365){
  yvol[i]=sd(ldr[,i])*252^0.5
}

######### PLOT
yvol.ts=as.ts(yvol)
plot(yvol.ts, type="l", col=EYY, main="Annualized Volatility Term Structure",xlab="Days to
Maturity", ylab="Volatility")
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "gray")
abline(h=c(0.2, 0.25, 0.3, 0.35, 0.4), col="white", lwd=2, lty="dotted")
abline(v=c(0,100, 200, 300), lwd=2, col="white", lty="dotted")
points(yvol.ts,col="yellow",type="l",lwd=2)
#########
```

```
##### Create TMatrix
TMatrix=matrix(1:(365*257),c(365,257))
for (j in 1:257){
  for (i in 1:365){
    TMatrix[i,j]=Data[i,j*4-2]
  }
}
#######

####### 3D PLOT TMatrix
persp3d(1:365,1:257,TMatrix,col="yellow")
persp3d(1:365,1:256,ldr,col="yellow")
Cov=cov((ldr))
persp3d(1:365,1:365,Cor,col="yellow")
persp3d(1:365,1:365,Cov,col="yellow")
######

###############################  3D PLOT
nbcol = 6
color = rev(rainbow(nbcol, v= 1, start = 0/6, end = 1/6))
zcol  = cut(Cov, nbcol)
persp3d(1:365,1:365,Cov, col=color[zcol],ticktype="detailed",axes=FALSE,
xlab="M",ylab="M",zlab="",box=TRUE)
# Use custom labels
axis3d(edge= 'y--', at =c(0,100,200,300))
axis3d(edge= 'x--', at = c(0,100,200,300))
axis3d(edge= 'z-+', at = c(0,0.0002,0.0004,0.0006))
mtext3d("Covariance", edge='z-+', line=2) # put in z-axis label by hand
# Set-up background (BOX)
bbox3d(col="grey",labels=NULL, xlen=0,ylen=0,zlen=0, nticks=0,xat=500,yat=500,zat=500)
# Add grid
grid3d("x+",col="white", lwd=2)
grid3d("y+",col="white", lwd=2,lty="100")
grid3d("z-",col="white", lwd=2,lty="dotted")
##EXPORT
#GIF & PNG
movie3d(spin3d(axis = c(0,0,1), rpm = 1),duration=10,  type = "GIF", convert=FALSE)
###############################

# Set par
par(mar=c(3,4,3,1))
par(mgp=c(3,1,0))
#########

################# cov by EIG - Relevant Factors to explain 95% of the variance
eig=eigen(Cov,only.values=FALSE)
m=365

bp=array(1:m)
bp=eig$values/sum(eig$values)
barplot(bp)

bp.cum=array(1:m)

for (i in 1:m){
  if(i==1){
    bp.cum[1]=bp[1]
  }
  else if(i>1){
    bp.cum[i]=bp.cum[i-1]+bp[i]
  }
}
bp.cum=bp.cum[1:10]
barplot(bp.cum,xlim=c(1,9),ylim=c(0,1),yaxt="n", main="Cumulative Attribution of the First 10
Components", width=0.5, xpd=TRUE,
space=c(2,rep(0.6,9)),names.arg=c("1","2","3","4","5","6","7","8","9","10"))
axis(2,at=c(0,0.2,0.4,0.6,0.8,1),labels=c("0 %","20 %","40 %","60 %","80 %","100 %"),las=1)

abline(h=0.95,lty="dotted")
bp.cum[1:10]
###################### 6 components are needed


#Simulate 365 forward curves to create stochastic equilibrium level
#Section 4.1.3
```

```
##############Load
#incl spot incl d-profile option B
Data=read.csv("", header=T, sep=";",dec=",")

#ldr
ldr=matrix(1:(256*365),c(256,365))
for (i in 1:365){
  for (j in 1:256){
    ldr[j,i]=(Data[i+1,j*4])
  }
}
#TMatrix
TMatrix=matrix(1:(365*257),c(365,257))
for (j in 1:257){
  for (i in 1:365){
    TMatrix[i,j]=Data[i,j*4-2]
  }
}
#Cov
Cov=cov((ldr))
#eig
eig=eigen(Cov,only.values=FALSE)
#vectors
temp=eig$vectors[,1:6]
#######################

################ SIM fwd curves t+dt conditional on fwd curve t

#Create 365*366, in first column fwd curve at t=0
FWDCurves=matrix(1:(365*366),c(365,366))
FWDCurves[,1]=TMatrix[,1]

dt=1
for (F in 1:365){
  e1=rnorm(1)
  e2=rnorm(1)
  e3=rnorm(1)
  e4=rnorm(1)
  e5=rnorm(1)
  e6=rnorm(1)
  e7=rnorm(1)
  TArrayNew=array(1:365)
  for (T in 1:365){
    Tot=0
    for (i in 1:6){
      if (i==1){
        Tot=temp[T,i]*sqrt(eig$value[i]*dt)*e1-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==2){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e2-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==3){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e3-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==4){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e4-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==5){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e5-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==6){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e6-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
      if (i==7){
        Tot=Tot+temp[T,i]*sqrt(eig$value[i]*dt)*e7-0.5*(temp[T,i]^2)*eig$value[i]*dt
      }
    }
    TArrayNew[T]=FWDCurves[T,F]*exp(Tot)
  }
  FWDCurves[,F+1]=TArrayNew
}

EQ=array(1:365)
for (T in 1:365){
  EQ[T]=FWDCurves[T,T+1]
}
```

```
plot(EQ,type="l",col="blue")
lines(FWDCurves[,1],type="l",col="red")
########################
#slow
```

Section 4.1.4 GARCH

```
#Estimation of GARCH for c OU process
#Section 4.1.4

#######load libraries
library(fGarch)
library(rugarch)

######## Import data
prices=read.csv("", header=T, sep=";",dec=",")
NBP=prices[1:1258,3:4]
########
#Load data
price.NBP=ts(NBP[,2])
ts.plot(price.NBP, lwd=2, col="red")

######### Set which price process to simulate ######
price=price.NBP
#########

######## function to calculate percentage price change (GOU) ####
pct.diff=function(price){
  PCT =rep(0,length(price))
  d=diff(price)
  for(t in 1:length(price)){
    PCT[t]=d[t]/price[t]
  }
  PCT=PCT[-length(PCT)]
  return(PCT)
}
#########

######## function to calculate log returns (LNOU) ####
pct.diff=function(price){
  PCT =rep(0,length(price))

  for(t in 1:length(price)){
    PCT[t]=log(price[t+1]/price[t])
  }
  PCT=PCT[-length(PCT)]
  return(PCT)
}
#########

######### Return function
R=pct.diff(price)
##########################

######### ugarchfit #smaller unconditional variance than garchFit
UR=ugarchfit(ugarchspec(variance.model=
list(model="sGARCH",garchOrder=c(1,1)),mean.model=list(armaOrder=c(1,1), include.mean=FALSE),
distribution.model="norm"),R, fit.control = list(stationarity = 1))
UR@fit$coef
omega<-UR@fit$coef[3]
alpha<-UR@fit$coef[4]
beta<-UR@fit$coef[5]
UnVar=omega/(1-alpha-beta)
sqrt(UnVar)
Omega=omega
Alpha=alpha
Beta=beta

############### LNOU + GARCH
T=731
#process parameters
eta = 0.0137 #eta = 0 is equivalent to Geometric Brownian Motion
mu = log(56.8330) #the mean of the process
Sigma=0.0345
```

```
#dubble for-loop
F=100 # amount of simulations
X=matrix(nrow=T+1,ncol=F)
for(j in 1:F){

#GARCH volatility model
specs = garchSpec(model = list(omega = Omega, alpha = Alpha, beta = Beta))
sigma = garchSim(spec = specs, n = T+1)

P_0 = log(48.65) #starting price, known
P = rep(P_0,T+1)

for(i in 1:T+1){
  P[i] = P[i-1] + eta * (mu - (Sigma^2/(2*eta)) - P[i-1]) + sigma[i]
}
X[,j]=exp(P)
}
#
ts.plot(X)

############# Plot ###########

Y=matrix(nrow=T+1, ncol=ncol(X)+1)
Y[,1:(ncol(X))]=X
Y[,(ncol(X))+1]=rowMeans(X) # calc average and add it to matrix
ts.plot(Y,xlab="Time", ylab="Price",
col=c(rep("black",ncol(X)),"blue"),lwd=c(rep(1,ncol(X)),2))
###########

#prices and returns
par(mfrow=c(1,1))
plot(P,type="l",xlab="Time", main="", ylab = "Price")
plot(diff(P), type = 'l', ylab = 'Price Changes', xlab = 'Time')
#################

########### Plot
plot(Y[,1],type="l",xlab="Time",ylab="Price",main="NBP - Price Paths with Volatility
Clustering",ylim=c(0,200),xaxt="n")
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "gray")
for (i in 201:400){
  points(as.ts(Y[,i]),type="l")
}
points(as.ts(Y[,501]),col=EYY,type="l",lwd=2)

axis(1, at=c(1,91,182,274,366,457,548,640),labels=c("Jan-15","Apr-15","Jul-15","Oct-15","Jan-
16","Apr-16","Jul-16","Oct-16"))
abline(v=c(1,91,182,274,366,457,548,640), lwd=1, col="white", lty="dotted")
abline(h=c(0,50,100,150,200),lwd=1, col="white",lty="dotted")
##########
```

Section 4.2 and 4.3

```
#LSMC Running Script together with related functions
#Following Boogert & De jong (2006) and Longstaff & Schwartz (2001)
#Section 4.2

rm(list=setdiff(ls(), c("X","Y")))

#Start simple, add complexity
#Time-Volume grid
#l volume points: 1 represents 0
#Injection and Withdrawal rate: 1 per t.
#Start volume: 51
#End volume: 51
#X: spotprice paths from OU constant Mu 20-02 with T=11, T[1]=P_0, s0 10 timesteps
#Continuation value at T+1 (=11) =penalty function, (lack of)remaining gas in storage
#Matrix V represents all possible volume levels at each t, continuation value must be
implemented

#Set parameters
#Number of independent price pahts
m=ncol(X)
#Length of T
T=nrow(X)-1
#Total possible volume levels (including zero)
```

```
l=101
#penalty function: 1.5* average last price * volumedifference.
p=1.1
#start volume
v.s=51
#end volume
v.e=51
#injection/withdrawal rate 3 possible options per point; withdraw 1, do nothing, inject 1
dv=1
#discount rate
r=0

### Create 3D matrix to represent accumulated future cashflow at each volume, time, and for
each price path
afc=array(rep(0,(l*m*(T+1))), dim=c(l,m,T+1))
### Create 3D matrix to represent continuation values at each volume, time, and price path
cv=array(rep(0,(l*m*T)),dim=c(l,m,T))
### Create 4D matrix to represent decision rule per volume, time and price path
dec=array(rep(0,(l*m*T*3)),dim=c(l,m,T,3)) #For now, three possible decisions per point

############## Load in functions ####################

################################### For-loop to apply backward induction for t=T...,1.
afc[,,T+1]=P.matrix(p)
for (i in T:1){
    cv[,,i]=Calc.cv(i)
    dec[,,i,]=Det.dec(i)
    afc[,,i]=Calc.afc(i)
  }
##################################

GSV=mean(afc[v.s+1,1:m,1])
GSV
EV=(GSV*2.5)/100 #Transform for sake of unity of NBP market
EV
S.H.GSV=sort((afc[v.s+1,1:m,1]*2.5)/100)
plot(S.H.GSV,type="l")
VaR=mean(S.H.GSV[50:51]) #VaR 95%
ES=mean(S.H.GSV[1:50]) #ES 95%
sd(S.H.GSV)

############## histogram of Gas Storage Value per Price Path
### EY Colors
EYY=c(255,210,0)
EYG=c(128,128,128)
BB=rbind(rep(EYY,4))
###
###Set margins of par
par(mar=c(4,4,2,2))
par(mgp=c(2.5,1,0))
###

# histogram Section 4.3
H.GSV=(afc[v.s+1,1:m,1]*2.5)/100
H.GSV=hist(H.GSV,breaks=50,main="Gas Storage Value per Price Path",xlab="Million GBP")
rect(par("usr")[1],par("usr")[3],par("usr")[2],par("usr")[4],col = "Gray")
lines(H.GSV,col="yellow",main="Histogram of values")
abline(v=c(VaR,EV),lwd=2,col=c("red","blue"),lty="dotted")
axis(1, at=VaR,labels="8.10")
axis(1, at=EV, labels="50.87")
legend(x="topright",c("Expected Value","VaR
95%"),lty=c("dotted","dotted"),lwd=c(2,2),col=rbind("blue","red"),bg="gray")
#
##############



#Supporitng functionf for LSMC
#Section 4.2

###################### Functions to create  Penalties at T+1 for each b and V
##################
###Function to set penalty values at T+1, for given b
penalty=function(p,b){
  X.b=X[T+1,b]
  V=matrix(data= NA,nrow=l,ncol=T+1)
  for(i in 1:l) {  #l represents total possible volume levels (including zero)
```

```
    if (i<(v.e+1)){
      V[i,ncol(V)]=((i-1)-v.e)*X.b*p
    }
    else if (i==(v.e+1)){
      V[i,ncol(V)]=0
    }
    else if (i>(v.e+1)){
      V[i,ncol(V)]=((i-1)-v.e)*X.b*(1/p)
    }
  }
  return(V[,T+1])
}
###

### Function to calculate new matrix that gives penalty for each v and b
P.matrix=function(p){
  M=matrix(data=NA, nrow=l,ncol=m)
  for(i in 1:ncol(X)){
    M[,i]=penalty(p,i)
  }
  return(M)
}
###
######################################################################

###################### Function to calculate continuation values at t=10 input is afc at t+1
###############
Calc.cv=function(t){
  C_t=matrix(data=NA,nrow=l,ncol=m)
  Xt=X[t+1,]              #X starts at 2 so this represents 10
  for (i in 1:l) {
    Y0=exp(-r*(T/252))*afc[i,(1:m),t+1]
    list=coef(lm(Y0~poly(Xt,2,raw=TRUE)))
    C_t[i,]=list[1]+list[2]*Xt+list[3]*Xt^2
  }
  return(C_t)
}
######################################################################

########################## Function to find decision rule, given the continuation values +
transfer costs #########
Det.dec=function(t){
  for (i in 1:m){      #price path
    for (j in 1:l){    #volume level
      if (j==1){
        #a=X[t+1,i]+cv[j-1,i,t]      #withdraw, not possible
        b=0+cv[j,i,t]                #Do nothing
        c=-X[t+1,i]+cv[j+1,i,t]      #Inject
        max=max(b,c)
        dec[j,i,t,1:3]=c(0,b==max,c==max)
      }
      else if (j>1 & j<l){
        a=X[t+1,i]+cv[j-1,i,t]       #withdraw
        b=0+cv[j,i,t]                #Do nothing
        c=-X[t+1,i]+cv[j+1,i,t]      #Inject
        max=max(a,b,c)
        dec[j,i,t,1:3]=c(a==max,b==max,c==max)
      }
      else if(j==l){
        a=X[t+1,i]+cv[j-1,i,t]       #withdraw
        b=0+cv[j,i,t]                #Do nothing
        #c=-X[t+1,m]+cv[j+1,i,t]     #Inject, not possible
        max=max(a,b)
        dec[j,i,t,1:3]=c(a==max,b==max,0)
      }
      }
  }
  return(dec[,,t,])
}
##########################################################################

########################## Function to find accumulated future cashflow, given the decision
rule #########
Calc.afc=function(t){
  for (i in 1:m){         #price path
    for (j in 1:l){       #volume level
      for (k in 1:3){     #decision rule
```

```
        if (dec[j,i,t,k]==1){
          afc[j,i,t]=-(k-2)*X[t+1,i]+exp(-r*(T/252))*afc[j+(k-2),i,t+1]
        }

      }
    }
  }
  return(afc[,,t])
}
########################################################################
```