NTNU
Handelshøyskolen

Fredrik Tangen
Kristine Vasseng

Determinants of implied volatility and implied skewness for WTI crude oil

Determinanter for implisitt volatilitet og implisitt skjevhet for WTI råolje

MASTEROPPGAVE - Økonomi og administrasjon/siviløkonom
Trondheim, dato: 24.05.2018

Hovedprofil: Finansiering og investering

Veileder: Sjur Westgaard

Preface

This thesis concludes our Master in Economics and Business Administration at NTNU Business School, with finance and investment as our specialisation. The work with this thesis have been demanding, but interesting. We have learnt a lot regarding handeling large data sets, coding in matlab, econometric methods and the oil industry.

We would like to thank our supervisor Sjur Westgaard for constructive comments, guidance and feedback throughout the process of writing this thesis. We would also like to thank Valeriy Kunst for providing us with data and for valuable guidance in matlab.

The contents of this master thesis reflect our own personal views, and potential errors are on the authors account.

Trondheim

24.05.2018

Fredrik Tangen

Kristine Vasseng

Abstract

We analyze empirically what drives changes in the volatility smile for WTI crude oil, by calculating at-the-money implied volatility and a proxy for implied skewness on nearby future options from 25.07.2006 to 03.03.2016. To our knowledge no previous research use a proxy for implied skewness when examining what drives changes in implied skewness for WTI crude oil. We examine if macroeconomic conditions, physical oil market fundamentals and financial variables can explain what drives changes in implied volatility and implied skewness. Our results for implied volatility supports the findings of Robe and Wallen (2016). The results for implied skewness shows that physical oil market fundamentals contributes significantly in explaining changes in implied skewness. Storage capacity in Cushing is proxied by the slope of the WTI crude oil term structure, and is significant when in state of backwardation. Higher degree of backwardation, more available storage capacity, has a positive effect on implied skewness. OPEC spare capacity is close to significant at a 10% level and is measured by a dummy variable. High OPEC spare capacity has a positive effect on implied skewness. In addition the control variable for time to maturity is significant and increasing time to maturity has a negative effect on implied skewness.

Sammendrag

I denne masteroppgaven analyserer vi empirisk hva som driver endringer i volatilitetssmilet for WTI råolje. Vi kalkulerer at-the-money implisitt volatilitet og en proxy for implisitt skjevhet på første posisjon opsjoner på futures fra 25.07.2006 til 03.03.2016. Til vår kjennskap har ikke en proxy for implisitt skjevhet tidligere blitt brukt til å forklare drivere av implisitt skjevhet for WTI råolje. Vi undersøker om makroøkonomiske faktorer, fysiske oljemarkedsfaktorer og finansielle variabler kan forklare hva som driver endringer i implisitt volatilitet og implisitt skjevhet. Våre resultater for implisitt volatilitet støtter funnene til Robe og Wallen (2016). Resultatene for implisitt skjevhet viser at fysiske oljemarkedsfaktorer har signifikant forklaringskraft på implisitt skjehvet. Lagringskapasitet i Cushing målt ved helningen til terminstrukturen til WTI råolje, og er signifikant ved backwardation. Høyere grad av backwardation, større tilgjengelig lagringskapasitet, har positiv effekt på implisitt skjehvet. OPEC ledig produksjonskapasitet er nær signifikant for et 10% signifikansnivå, og er målt ved en dummy variabel. Høyere ledig produksjonskapasitet hos OPEC har positiv effekt på implisitt skjevhet. I tillegg er kontrollvariabelen tid til forfall signifikant, og lengre tid til forfall har negativ effekt på implisitt skjevhet.

Contents

List of Figures

List of Tables

1 Introduction

Crude oil prices have great impact on the global economy, and have been subject to high volatility in the last decades. One example being the financial crisis in 2008 when the price fell 70 percent in only few moths, and then recovered before falling again from over $100 per barrel in July 2014 to a 12-year low of $26 in January 2016. However from 2012 until 2014 implied volatility for crude oil were very low. Crude oil is the most actively traded commodity, and futures and options on crude oil are highly liquid. Options on crude oil are used to hedge and trade economic agents' expected forward-looking views on uncertainty.

The objective of our thesis is to examine what drives changes in forward-looking volatility and skewness for WTI crude oil. We calculate volatility smiles from 25.07.2006 to 03.03.2016 for nearby future contracts. Previous literature has concluded that investors do not only care about volatility, but are interested in skewness as well (Mitton and Vorkink, 2007; Barberis and Huang, 2008). Understanding which factors influence the distribution of WTI crude oil prices would improve forecasts on both volatility and skewness, giving commercial and non-commercial market participants better insight when trading WTI crude oil. Both upside and downside extreme values are interesting, as different market participants are interested in hedging different sides. Buyers of crude oil would like to hedge upside-risk, while sellers would like to hedge downside-risk. Positive implied skewness means that it is expensive to hedge upside-risk, and negative implied skewness means that it is expensive to hedge downside-risk (Alexander, 2008*c*). Thus understanding what drives implied skewness is essential for risk-managers.

The literature that is close to ours is Robe and Wallen (2016). They analyze empirically what drives implied volatility for crude oil, using macroeconomic, physical and financial variables. First we replicate Robe and Wallen (2016) using a different time period, including a more volatile period after 2014. Secondly we calculate a proxy for implied skewness, and examine if the variables Robe and Wallen (2016) use for implied volatility could also explain what drives changes in implied skewness. To our knowledge this has not been done before.

Our findings for implied volatility support the results of Robe and Wallen (2016). We find that uncertainty in equity markets, measured by the VIX-index, has significant impact on WTI crude oil implied volatility. This contributes to understanding the low levels of implied volatility for WTI crude oil in 2012 to 2014, since it coincides with the low levels of the VIX-index. Physical oil market fundamentals also contributes in explaining WTI crude oil implied volatility, with significant explanatory variables like OPEC spare capacity and storage tension in Cushing, measured by a proxy.

Implied skewness have fewer significant explanatory variables than implied volatility, but our results suggest that physical oil market fundamentals is key to understand what drives changes in implied skewness. Storage-tension is significant with a positive effect when the market is in a state of backwardation, reflecting that low supply relative to demand leads to higher prices and more positive implied skewness in the WTI crude oil market. OPEC spare capacity has positive effect and is close to significant at a 10% level. This indicate that high OPEC spare capacity could reflect positive forward-looking skewness, because high OPEC spare capacity tend to occur when WTI crude oil prices are low. In addition the control variable for time to maturity is significant with a negative effect, meaning that longer time to maturity decreases implied skewness, resulting in a more left-skewed distribution.

1.1 Related Literature

Several studies investigates the information content of option-implied volatilities. Day and Lewis (1993), Szakmary et al. (2003) and Haugom et al. (2014) find that including implied volatility significantly improves volatility forecasts. Implied volatility for crude oil is also found to be a better fit for firms investment behaviour than historic volatility measures (Borenstein and Kellogg, 2014). Navatte and Villa (2000) and Finta and Ornelas (2018) conclude that implied skewness gives better information than historic skewness for forecasting skewness. Doran and Krieger (2010) finds that measures containing information from the volatility skew has predictive power over future direction of the underlying asset price. Thus it is undoubtly important to understand what drives changes in implied volatility and implied skewness, which is the goal of this thesis.

Previous research show that macroeconomic factors, crude oil fundamentals and financial variables have impact on WTI crude oil. Robe and Wallen (2016) complements these studies by using several variables in each category to explain what drives implied volatility for WTI crude oil. We contribute to the literature by examining if these variables also explain what drives implied skewness for WTI crude oil.

Mixon (2002) examines whether macroeconomic factors have a relationship with the implied volatility surface for equities, and finds a clear relationship, but with diminishing effect on longer maturities. A link between the implied volatility surface for the S&P 500 and key macroeconomic variables have also been found by Guo, Han and Zhao (2014). Mork, Olsen and Mysen (1994) finds that macroeconomic factors and oil prices correlate, indicating a relationship between macroeconomic factors and the movements in oil prices. In a similar vein we examine if macroeconomic factors can help explain changes in implied volatility and implied skewness for WTI crude oil. Our results shows that macroeconomic factors have no significant explanation power on either implied volatility or implied skewness.

Studies (Chevillon and Rifflart, 2009; Kaufmann, 2011) show that OPEC, and thereby physical market conditions, have influence on the crude oil price. Kaufmann (2011) argue that the crude oil price changes in 2007-2008 partly can be explained by changes in OPEC spare capacity. We include OPEC spare capacity without Saudi Arabia in our analysis, and find that it contributes in explaining changes in implied volatility. It is also close to a 10% significance level for implied skewness, indicating that it could help explain changes in implied skewness as well. Buyuksahin et al. (2013) test if the spread between WTI and Brent crude oil are related to storage conditions in Cushing. We complement Buyuksahin et al. (2013) and find that storage conditions are significant for both implied volatility and implied skewness, confirming that storage conditions have impact on the distribution of WTI crude oil.

Guo, Han and Zhao (2014) find evidence that there exist a relationship between the implied volatility surface and financial variables, first and foremost the VIX-index. Previous research have found spillover effects from implied volatility in the crude oil market, measured by the OVX-index, and implied volatility in the equity market (Maghyereh, Awartani and Bouri, 2016; Liu, Ji and Fan, 2013). This is confirmed by our study where we find that VIX has significant effect on implied volatility for WTI crude oil. VIX can therefore be viewed as a better measure for market sentiment than macroeconomic factors. We do not find VIX to be significant for implied skewness. This is expected, because larger fluctuations in the price does not necessarily impact the markets view about future skewness.

Würsig (2017) explain implied volatility, skewness and kurtosis for WTI crude oil with the same variables as Robe and Wallen (2016). He extracts the implied moments from risk-neutral density functions. Our study differ to Würsig's (2017) in the way that we use a proxy for implied skewness. The advantage of our method for calculating implied volatility and implied skewness is that it requires less computational power and is faster to implement than the method used by Würsig (2017). Comparing our results with Würsig (2017) we have similar results for implied volatility, but find more significant variables for implied skewness.

We complement the work of Ohnsorge, Stocker and Some (2016). They use the same proxy for implied skewness as we uses in this thesis. They use the proxy for implied skewness on S&P 500, term spreads and average of Brent and WTI crude oil forward prices to forecast global risk. Our focus is instead to identify determinants for what drives changes in implied skewness for WTI crude oil

1.2 Structure

The rest of the thesis is as follows. Chapter 2 and 3 details the theory and estimation of implied volatility and implied skewness. Chapter 4 presents the exogenous regressors. Chapter 5 presents descriptives and stationarity tests, while Chapter 6 presents the empirical results from the regression models. Chapter 7 concludes.

2 Theory

This chapter explains the theory behind implied volatility smiles, and how the information in volatility smiles can be used to retrieve at-the-money implied volatility and implied skewness via a proxy for implied skewness. We start this chapter with an explanation of volatility.

2.1 Volatility

According to Alexander (2008*b*) the precise definition of volatility for an asset is the spread in the stochastic process that is used to model the log returns. Volatility is often measured as standard deviation (σ), and is a good measure for risk when returns are normally distributed. However this is rarely the case, thus volatility does not give a full description of the risk that the investor takes. Despite this, volatility is the most commonly used measure for risk. According to Alexander (2008*b*) we can not observe volatility, we can only make estimates and forecast volatility.

> Estimating volatility according to the formula given by a model gives an estimate of volatility that is 'realized' by the process assumed in our model. But this realized volatility is still only ever an estimate of whatever volatility had been during the period used for the estimate (Alexander, 2008*b*, pp.94).

Nevertheless volatility is easy to compute and widely used to measure risk by both practitioners and theorists. Volatility is normally calculated on historical data, looking back at what happened in the past. In contrast implied volatility, which is used in this thesis, is a forward-looking measure, calculated using option theory.

2.2 Black-Scholes

The Black-Scholes formula gives the fair price of European options (Black and Scholes, 1973). The formula has several assumptions, which are required for the formula to hold. One of them is that asset prices (S) follows a Geometric Brownian Motion with constant drift, which we elaborate in Section 2.3. The key assumption that leads to the formula is that no one can make arbitrage profits by owning a portfolio that contains variable quantities of the asset and the option (Taylor, 2005). The formula assumes that investors are risk neutral, in other words the personal risk preferences of the investor is not consider when valuating the option.

The Black-Scholes formula for a call option is:

$$c_{BS},(S,T,X,r,q,\sigma) = Se^{-qT}N(d_1) - Xe^{-rT}N(d_2) \tag{2.1}$$

Where d_1 and d_2 is defined as:

$$d_1 = \frac{log(\frac{S}{X}) + (r - q + \frac{1}{2}\sigma^2)}{\sigma\sqrt{T}} \tag{2.2}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{2.3}$$

Equation 2.1 calculates the price of a standard European call option. Where c_{bs} is the call premium, S is current price on the underlying, T is time to maturity, X is strike price, r is the risk-free interest rate, q is convenience yield, σ is standard deviation and N is the cumulative standard normal distribution.

To calculate the price of a standard European put option we have to rewrite the equation or use the put-call parity. The put-call parity is a theoretical relationship between the prices of European call and put options with the same underlying asset, strike price and time to maturity (Hull, 2012). The relationship can be written as, where p is the put premium:

$$c + Xe^{-rT} = p + S \tag{2.4}$$

The put-call partiy can be rewritten to give the price of a standard European put-option:

$$p = c + Xe^{-rT} - S \tag{2.5}$$

2.3 Implied Volatility

All inputs in the Black-Scholes formula are directly observable in the market except for the volatility parameter σ. The volatility parameter can be backed out using the option price observed in the market. Thus we can use the Black-Scholes formula to calculate implied volatility. The implied volatility for a European call option, traded at the price c_{market}, is the $\sigma_{implied}$ that solves the following equation:

$$c_{market} = c_{BS}(S, T, X, r, q, \sigma_{implied}) \tag{2.6}$$

Implied volatility reveals the markets expectations for future volatility. Options with different strikes and equal time to maturity on the same underlying should have the same implied volatilities if the assumptions of the Black-Scholes model holds. Then the return of the underlying would follow a lognormal distribution. However this is not the case in reality, because traders do not act according to the assumptions of the Black-Scholes formula. We will therefore observe a surface of market implied volatilities, by strike and maturity of the option (Alexander, 2008*c*).

The implied volatility of options with the same maturity, but different strikes, forms a volatility smile when plotted as a function of the options strike price. Normally implied volatility is lowest when the options are at-the-money, and increasing with higher difference between strike price and the price of the underlying asset. For options with long maturity the volatility smile is less pronounced than for options with short maturity (Hull, 2012). Puts and calls for European options should have the same implied volatility because of put-call parity, but in practice they often differ because of transaction costs, bid-ask spreads, and so forth (McDonald, 2014).

The Black-Scholes model presented in Section 2.2 assumes that the price of the underlying asset evolves according to a geometric Brownian motion with constant drift, where μ is the drift, σ is the volatility of the process and w is a wiener process.

$$dS = \mu S dt + \sigma S dw \tag{2.7}$$

The geomtric Brownian motion assumes constant volatility, but implied volatility is not constant for different strikes (Alexander, 2008*c*). This implies that real-world option prices does not follow the assumptions of the Black-Scholes option pricing model. The existence of the volatility smile curve shows that market participants make more complex assumptions about the price process than the geometric Brownian motion. Therefore the volatility smile give the investors valuable information about how the market views future uncertainty.

2.3.1 Barone-Adesi & Whaley implied volatility

For standard European options there are no possibilities of early exercise. It is therefore possible to use the method explained Section 2.3 to back out the implied volatility of the option. For American options there is a possibility for early exercise, which makes the Black-Scholes method invalid. For American options there are no analytic solutions, like the Black-Sholes formula for European options. Some popular approaches are to use binomial trees or finite-difference methods, but these methods are computationally expensive. Instead we use the approximation derived by Barone-Adesi and Whaley (1987). The advantage of using this approximation is that it can be calculated without extensive data power, and still maintain high accuracy. One should be aware that the accuracy decreases with increasing time to maturity, so the method is best when used for short maturities (Barone-Adesi and Whaley, 1987).

This is initially a method for computing the price of American options. This is done by adjusting for the possibility of early exercise. After adjusting for the possibility of early exercise the method can be used to calculate implied volatility for American options. In the case of future options on commodities there are no dividends, and the price of an American call option will be equal to the price of an European option. For puts on the other hand, there is always a possibility of early exercise (Barone-Adesi and Whaley, 1987). Therefore, when using future options on commodities to calculate implied volatility, this method is only necessary for put options.

For American future options on commodities the price of a put option is calculated using the following quadratic approximation derived by Barone-Adesi and Whaley (1987):

$$P(S,T) = \begin{cases} p(S,T) + A_1(S/S^{**})^{q_1}, & \text{when } S > S^{**},\ and \\ X - S, & \text{when } S \leq S^{**} \end{cases} \tag{2.8}$$

Where:

$$A_1 = -(S^{**}/q_1)\{1 - \exp^{(b-r)T} N[-d_1(S^{**})]\} \tag{2.9}$$

A_1 is the premium for the possibility of early exercise, S^{**} is the critical commodity price where the option is exercised early, b is convenience yield, q_1 is a parameter derived in Barone-Adesi and Whaley (1987), and the rest is the same as for the black-scholes formula in Section 2.2. When already having the option price, implied volatility is calculated with the same technique as when using Black-Scholes, simply backing out the volatility that gives the same theoretical price as the observed price.

2.4 Proxy for Implied Skewness

The volatility smile contains information about the future price movements of the underlying assets. Many studies use the pioneering work done by Breeden and Litzenberger (1978) to model the Risk-Neutral distribution of the underlying asset, and then calculate higher moments. We will also focus on the information contained in the volatility smile, but we will use a proxy for implied skewness, instead of calculating the Risk-Neutral distribution. The advantage of this method is that it requires less computational power and is faster to implement. Mixon (2011) argues that week to week changes in the volatility smile can be a good proxy for changes in skewness.

The definition of positive skewness (right-skewed) is a distribution were the upper tail is heavier, thus has higher probability than the lower tail. Opposite we have negative skewness (left-skewed) when the lower tail is heavier than the upper tail (Alexander, 2008*a*). The normal distribution is symmetrically and has a skewness of zero. However the distribution for almost all financial assets are skewed.

Multiple skew measures have been researched in the literature. Differences in out-of-the-money put and call volatilities have been used, both based on percentage moneyness(Bates, 1991) and deltas (Hull, Nelken and White, 2005). However Mixon (2011) finds that these skew measures are dependent on the level of at-the-money volatility, and instead he argues for standardizing the measures by dividing on at-the-money volatility. We use the skew measure that Mixon (2011) prefers as a proxy for implied skewness, but reorganize the equation to become more intuitive and make the proxy move in the same way as the implied skewness of the underlying distribution. Equation 2.10 shows the proxy for implied skewness that we use in this thesis.

$$\text{Proxy for implied skewness} = \frac{\text{25-delta call volatility} - \text{25-delta put volatility}}{\text{50-delta volatility}} \tag{2.10}$$

According to Mixon (2011), the big advantage of this measure is that it has minimal dependence on the level of at-the-money volatility. When using deltas we also gain the advantage that implied volatility is closer to linear in delta space than in strike space. Using 25 delta implied volatilities also secure liquidity at the points where the proxy is calculated, instead of, for example, using the endpoints of the smile, which could have very low liquidity.

3 Estimating Implied Volatility and Skewness

In this chapter we clarify the estimation of implied volatility and implied skewness for our data. We have used nearby contracts when calculating implied volatility and implied skewness, to narrow the scope of our thesis. When applying the formulas described in Chapter 2, we have chosen to use MATLAB. We also give a brief introduction to our data. See Appendix A for more information about the data used in this chapter and a practical approach to calculating volatility smiles.

3.1 WTI Crude Oil

Our data consists of futures and future options on WTI Crude oil, which is the most actively traded commodity. Historic future prices can be seen in Figure 3.1. We see that the price have large fluctuations around well-known events, such as the financial crisis in 2008. We will look closer at historical events in Section 3.7 and 6.4.

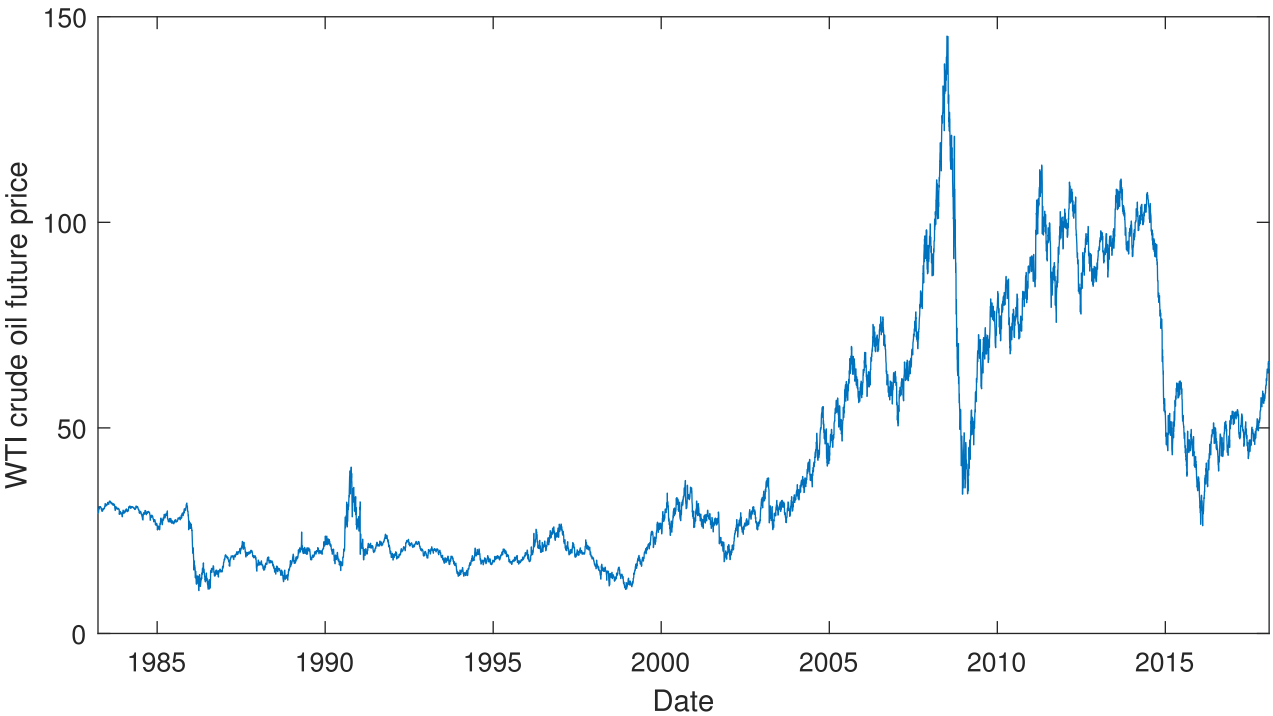

Figure 3.1: Historical WTI crude oil future price from 30.03.1983 to 02.02.2018. Source: Datastream.

3.2 Future Options

Options on futures gives the right to enter into a futures contract at a predetermined premium by a certain date (Hull, 2012). We use future options on WTI crude oil, which is American options and therefore the holder can exercise the contract at any time during the life of the option. If the investor exercise the call option he gets a long position in the underlying futures contract and a cash amount equal to the most recent settlement price of the future minus the strike price (Hull, 2012).

The main reason that futures on commodities and futures options on commodities are so popular is that they are more liquid and easier to trade than the underlying asset (Hull, 2012). For investors, both private and institutional, liquidity is very important, because it secures low bid-ask spreads and helps investors avoid being locked in a position. Hedging oil risk with options and future options have become very popular, and as Figure C.7 in the appendix shows, open interest in WTI crude oil futures continues to grow as it has become a more widespread method of hedging oil market risk. Another explanation being the increased investment activity from non-commercial participants, as described more in detail in Section 4.3.3.

3.3 Bounds for Future Options

Before calculating implied volatility we check if arbitrage conditions are satisfied. These bounds are derived from the put-call parity (Hull, 2012). Since WTI crude oil futures does not pay dividend we use the bounds for European options on calls. The price of put or call options can not be negative, so the bounds of the call options is, where F_0 is the future price:

$$c \geqslant (F_0 - X)e^{-rT} \tag{3.1}$$

When a call option is deep in the money, the corresponding put option is deep out of the money, then the price of the put p is very close to zero. Since the difference between call price c and its lower bounds equals p, the price of the call option must be very close to its lower bound Hull (2012). Put options may be exercised early, therefore we define the lower bound for an American put option as:

$$P \geqslant X - F_0 \tag{3.2}$$

3.4 Calculating Implied Volatility

When calculating one-month implied volatilities we use both put and calls, depending on strike prices relative to the price of the underlying future. Implied volatility for call options are calculated using the black-scholes method for extracting implied volatility explained in Section 2.2. For puts we have used the quadratic approximation derived by Barone-Adesi and Whaley (1987) explained in Section 2.3.1. For options with strikes below the underlying future price we have calculated implied volatility using put options, and call options when the strike price is above the underlying future price. This means that we have out-of-the-money options at both sides of the volatility smile. When the options are exactly at-the-money, meaning that the strike price and the price of the underlying asset is identical, we have calculated implied volatility for both put and call options, and used the arithmetic mean.

3.4.1 Interest Rate

To calculate implied volatilities we use the 1 month US treasury bill rate from Datastream. Because of our roll date on the implied volatilities, explained in Section 3.4.2, the actual time to maturity is between 1 and 2 months, but we still use the 1 month interest rate, because interpolating with interest rates with longer time to maturity would have minimal impact on the result. When calculating the implied volatility the interest rate is customized using datetime-objects in MATLAB, adjusting for actual time to maturity.

3.4.2 Monthly Rolling of Daily Implied Volatility

When future options and futures move close to the expiration date there is a possibility of low liquidity. When exploring our data we found that the last trade day of nearby options showed strange results, because of low liquidity. Robe and Wallen (2016) used open interest for futures to find that the 7th business day of the month should typically be used to roll between nearby and first-deferred contract. We choose to use the 7th business day of the month as roll date. In addition, when the 7th business day is later than the 10th of the month, we roll at the 10th to avoid possible problems with liquidity.

3.4.3 At-The-Money

At-the-money options are traditionally the most liquid options. In textbooks, at-the-money is defined as options with strike price equal to the price of the underlying asset (Hull, 2012). Our data consist, as explained earlier in this thesis, of future options. In reality the underlying future will seldom be priced equal to a possible strike, because strike prices come in increments of $0.50. This requires us to define at-the-money options in a different way. We use the definition from Xing, Zhang and Zhao (2010), where they define at-the-money options as options with a ratio of strike price to future price between 0,95 and 1,05. This gives us a wider range of at-the-money options. To solve the problem that we get multiple options at-the-money we simply calculate the arithmetic mean implied volatility of the options inside the span to get the at-the-money implied volatility. From Figure C.1 in the appendix we see that this measure for at-the-money implied volatility is equal to the 50-delta implied volatility used in the proxy for implied skewness in Equation 2.10. This supports that how we measure at-the-money implied volatility is satisfying.

3.5 Volatility Smile

Having calculated volatility smiles for all trade dates from 25.07.2006 until 03.03.2016 we end up with a total of 2421 smiles. Figure 3.2 show the volatility smile for 30.01.2012, with maturity date at 19.03.2012. At this date the price of the nearby WTI crude oil future was $98.78, and we see that the lowest implied volatility is at-the-money, as stated in Section 2.3. We see that for this date the implied volatilities for out-of-the-money puts are higher than for out-of-the-money calls. Thus we know that for this date the market believes it is more likely that the crude oil price will fall to lower levels, instead of increasing. This is known as a negative skew, since the log price density will be negatively skewed (Alexander, 2008*c*). This imply that it is more expensive to hedge downside risk than upside risk. In our thesis we seek to explain why the smile changes from one date to another, in other words what drives implied volatilities and implied skewness. In Appendix C.2 we have put more smiles to show how the volatility smile change over time.

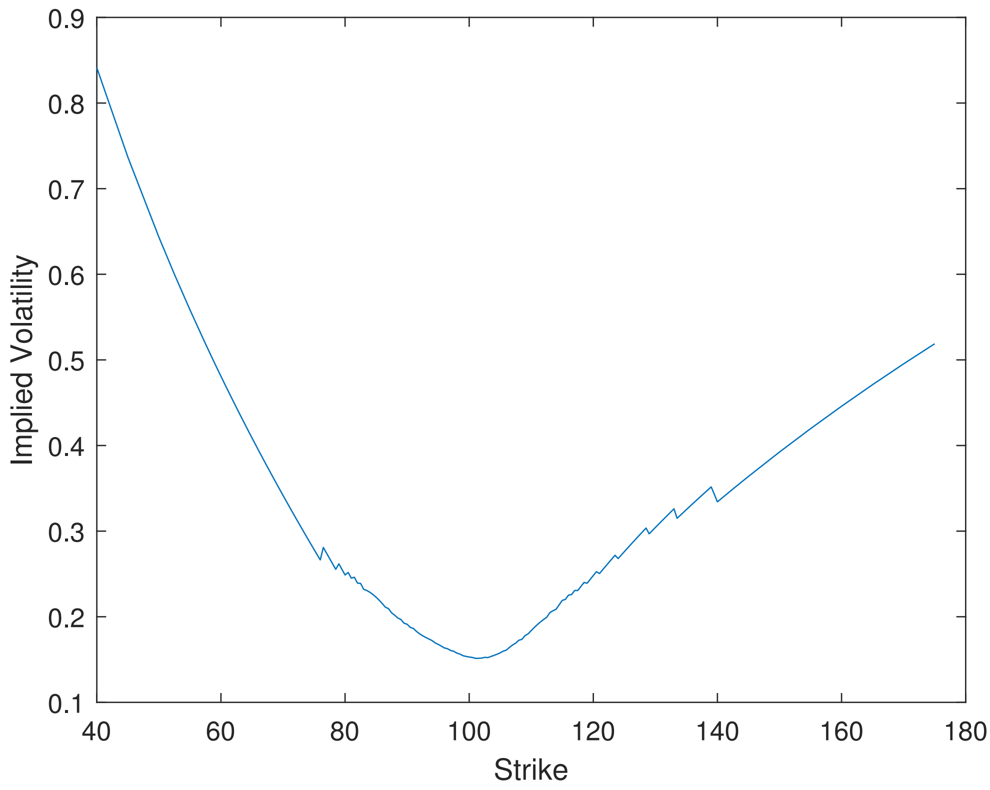

Figure 3.2: Implied volatility smile with trade date 30.01.2012 and maturity date 19.03.2012. The smile is calculated with futures and future options on WTI crude oil from CME.

3.6 Implied Skewness

From the volatility smiles in Figure 3.2 we can calculate the proxy for implied skewness explained in Section 2.4. In practice we would rarely see datapoints at exactly 25- or 50-delta, so we use the arithmetic mean inside a span of 0.03 around the deltas in Equation 2.10 to avoid problems with not being able to calculate the proxies. After calculating daily proxies for implied skewness for the whole data set we have 108 missing dates out of a total of 2421 dates. We correct for this with linear interpolation.

As described in Section 3.5 the smile has a negative skew. Thus we will expect our proxy for implied skewness for this date to be negative, and the calculations confirm this with a value of -0.0988.

3.7 Historical At-The-Money Implied Volatility

Figure 3.3 plots at-the-money implied volatility for WTI crude oil from 25.07.2006 to 03.03.2016. We see that implied volatility was relative stable from 2006 up to the financial crisis starting in the fall of 2008. The oil price increased in the same period until it peaked in 2008. One of the contributing factors for the growing oil price in this period was low OPEC spare capacity, which was a consequence of high oil demand (Brunetti et al., 2013). The oil price collapsed in the fall of 2008 after the financial crisis, and at the same time implied volatility raised to high levels. The financial crisis lead to a low demand for crude oil and OPEC spare capacity increased. After the financial crisis crude oil prices started to raise again and at the same time implied volatility was stable and back at the same levels as before the crisis. In August 2011 implied volatility peaked again following the U.S credit rating downgrade (*Russia oil row hits Europe supply*, 2007) and sovereign debt crisis in Europe, which lead to uncertainty about the level of global demand for crude oil. From the fall of 2012 to the fall of 2014 we observe historically low implied volatilities. Robe and Wallen (2016) argue that this could be a reflection of low uncertainty in financial markets, captured by a historically low VIX-index, or high crude oil supplies in this period.

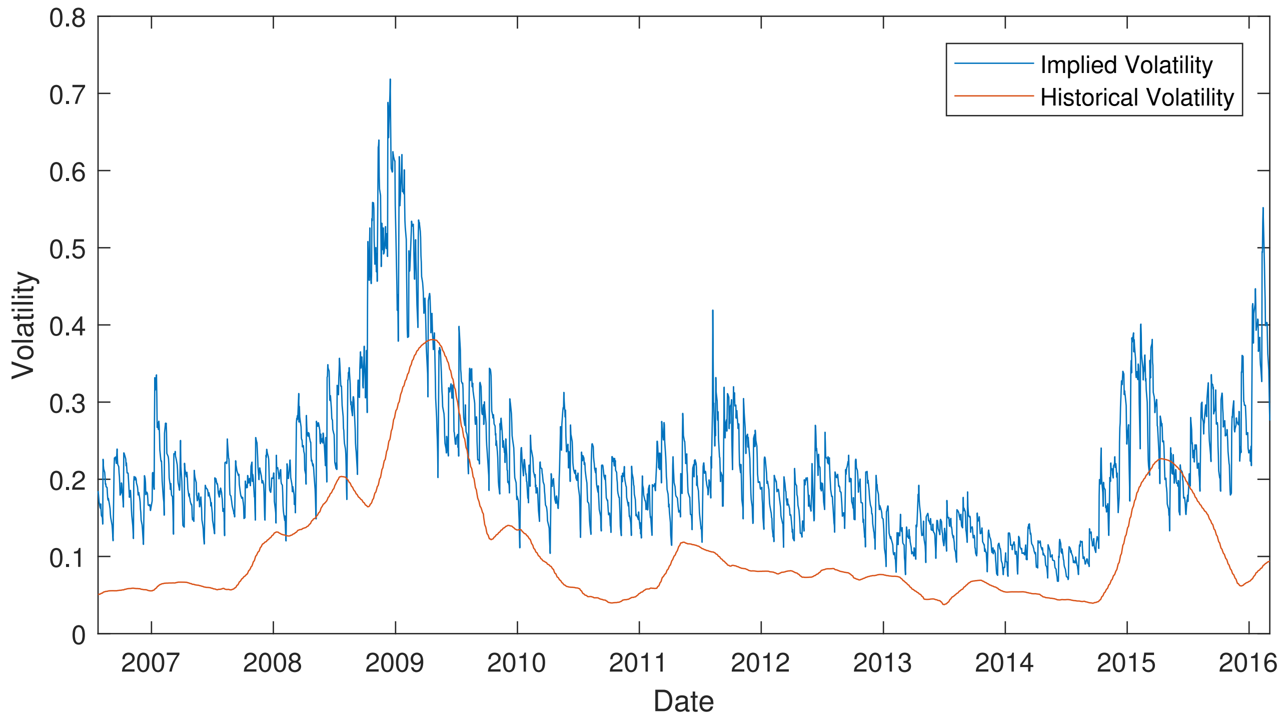

Figure 3.3: Daily volatility from 2006 to 2016. The blue line shows implied volatility calculated from WTI crude oil future options, and the red line shows 1-year historical rolled volatility using WTI crude oil continuous future price as underlying.

In the fall of 2014 crude oil prices dropped dramatically, which lead implied volatility to rise. What caused the drop in oil prices in the fall of 2014 is still an open question and the severity of the fall was surprising also for industry experts (Baumeister and Kilian, 2016, p.133).

Baumeister and Kilian (2016) found that over half of the price fall for Brent crude oil from June 2014 to December 2014 was predictable using publicly available information as of June 2014. They found that both demand shocks prior to July 2014, reflecting unexpected weakening in the global economy, and supply shocks caused by increased supply and changes in expected crude oil production helps explain parts of the drop in crude oil price.

Comparing our at-the-money implied volatilities with the ones in Robe and Wallen (2016) we find that our results are very similar, but systematically lower. The red line in Figure 3.3 shows the historical 1-year rolled volatility for continuous WTI crude oil futures. The historical rolled volatility is typically below the blue line representing implied volatility, but they follow each other closely and show a correlation-coefficient of 0.6747. The difference in implied and observed historical volatility represent a volatility premium taken when the options are issued.

3.8 Historical Implied Skewness

From Figure 3.4 we see that implied skewness fluctuate around a negative skewness of -0.1, having low volatility up to 2011, where it enters a more volatile period lasting to 2015. We believe one reason explaining why implied skewness is more volatile from 2011 is that the market have higher uncertainty of what the future oil price should be. From 2011 until 2014 the WTI crude oil future price fluctuates around $80-$100 and the implied skewness indicates that the investors expectations shifts from positive to negative skewness in line with the shifts in the WTI crude oil price. After 2015 implied skewness is less volatile. This might reflect that the market has adjusted to the new lower crude oil price after the dramatic fall in crude oil price in the Autumn of 2014, and that the market does not believe that the crude oil price is likely to keep falling in the future. Figure 3.4 also shows that implied skewness is less volatile than the historical 1-year rolled skewness. We believe one reason for this is that historical skewness also includes unexpected shocks that hit the market.

Alexander (2008*c*) argue that one reason options on equity indexes has a negative skew is that out-of-the-money puts on equity indexes is a good hedge for the market and therefore many buy them despite of a high premium. We believe that this reasoning can be transferred to the crude oil market, because crude oil is the most liquid commodity and a lot of producers uses the market to hedge downside risk in the crude oil price. The high demand for out-of-the money puts pushes up the price (Alexander, 2008*c*, p.236) and may explain why the observed implied skewness is negative on average.

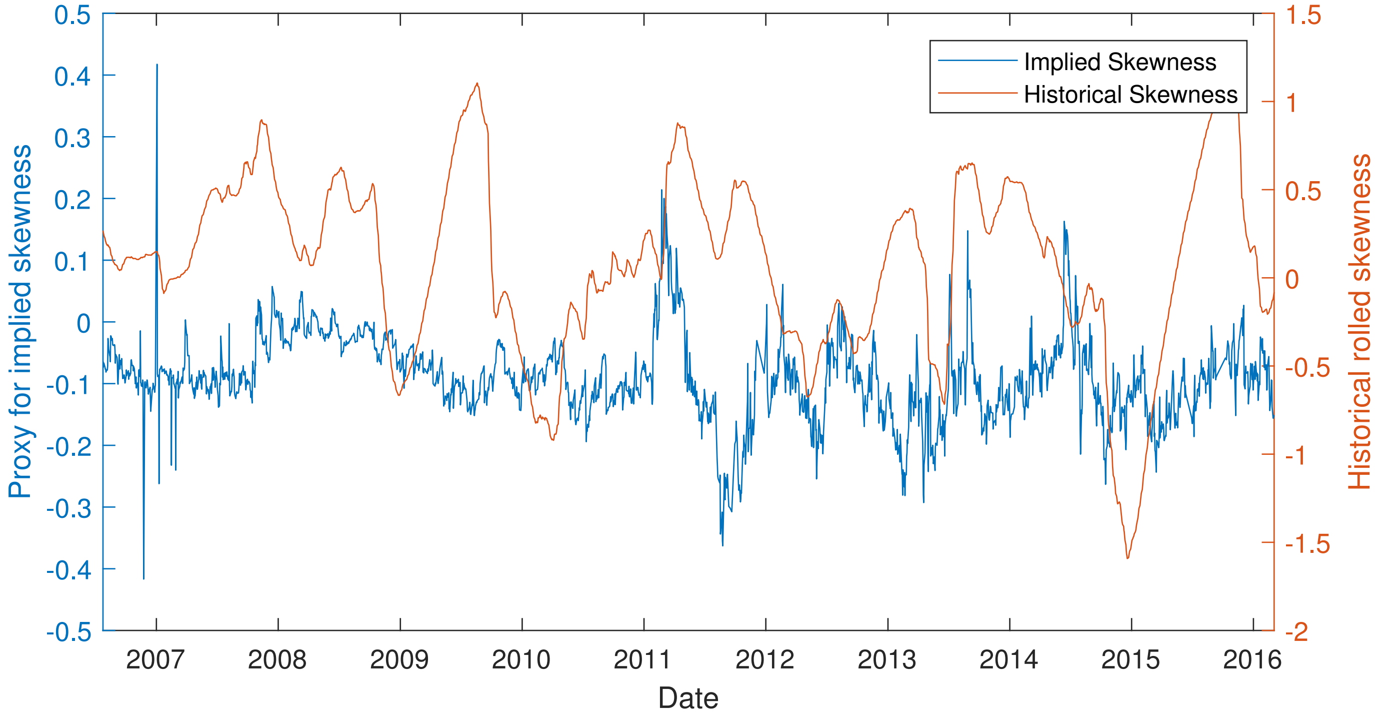

Figure 3.4: Daily skewness from 2006 to 2016. The blue line shows implied skewness calculated using WTI crude oil future options and the proxy for implied skewness (left hand scale). The red line shows 1-year historical rolled skewness on continuous WTI crude oil futures (right hand scale).

4 Explanatory Factors

The variables presented in this chapter are consider by Robe and Wallen (2016) to be variables that contain information about crude oil implied volatility. We expand their analysis by examining if the same variables can be used to explain implied skewness. We follow their classification for the variables and split them into three groups: macroeconomic fundamentals, physical-market conditions and financial variables. Some of the variables in this chapter is only available at a weekly frequency, therefore the regression in Chapter 6 is based on weekly data. For more information on the data used in this chapter, see Appendix A.

4.1 Macroeconomic Fundamentals

Empirical evidence tells us that stock market volatility is countercyclical (Corradi et al., 2013) and we expect the same for crude oil volatility, ceteris paribus. For example when the world economy is in a recession we will expect a lower demand for crude oil, and thus lower crude oil price and hence the implied volatility for WTI crude oil is expected to increase. Kilian and Park (2009) explain that U.S. stock market resilience to higher crude oil prices can be explained by strong global demand for industrial commodities (Kilian and Park, 2009, pp.1268).

4.1.1 US Economy

Robe and Wallen (2016) use the variable REAL computed using the method developed by Kilian (2009) to measure world economy. REAL is difficult and time-consuming to compute, and in addition it is not significant in Robe and Wallen's (2016) regression, therefore we choose to drop this variable. Since we only have American data we instead use the ADS-index developed by Aruoba, Diebold and Scotti (2009), which Robe and Wallen (2016) use as a control variable. They use this variable to consider the possibility that the U.S economy is more important to explain implied volatility for WTI crude oil than the world economy. We use weekly changes in the daily index of U.S business activity.

Aruoba, Diebold and Scotti (2009) managed to compute a daily index even though many of the input variables are available at different frequencies, for example industrial production is available at a monthly frequency and employment at a weekly frequency. The index incorporate many variables and thus provides a continuously updated measurement of the US economy. ADS does not consist of any oil component and therefore we use contemporaneous changes in the ADS index in the regression analysis. If there is a big change in the U.S. economy we expect a change in the implied volatility (Robe and Wallen, 2016). When the index decreases we expect implied volatility to rise, and opposite when the index increases. Thus we expect that the U.S economy variable will have a negative sign for implied volatility. For implied skewness we know intuitively that it should be positive in bad states and negative in good states of the economy, thereby expecting a negative sign for the ADS-index on implied skewness.

4.2 Physical-Market Conditions

Physical-market conditions capture disruptions caused by the supply side. The surplus production capacity for OPEC gives an indicator for the suppliers possibility to react if demand increases. Also the production output for WTI crude oil and a proxy for physical storage-market conditions is included in this section.

4.2.1 OPEC Surplus Production Capacity

We expect that high OPEC crude oil surplus will, ceteris paribus, put a lower pressure on crude oil volatility. Intuitively this must be so, because high surplus makes it possible for suppliers to react if demand increases. Studies have justified this argument empirically. For example, Brunetti et al. (2013) argue that the increased demand for oil after 2003 drained OPEC spare production and lead to higher oil price volatility. Robe and Wallen (2016) argue that lower crude oil prices leads to higher surplus production capacity, reflecting weak macroeconomic environments like we observed after the Lehman crisis. We expect high OPEC spare capacity to occur when the crude oil price is low. Thus expecting a heavier upper tail for the crude oil distribution. In other words we expect a positive effect on implied skewness from OPEC spare capacity.

We follow Robe and Wallen (2016) and use the non-Saudi OPEC spare capacity as a variable to help explain implied volatility and implied skewness. They have three arguments for excluding Saudi Arabia. First, Büyükşahin and Robe (2011) argue that the clearest evidence of a major change in world energy market fundamentals is reflected in this variable (Büyükşahin and Robe, 2011, p.22). Second, Saudi oil is not light or sweet and therefore not a good substitute for other types of crude oil, because oil refineries can not easily shift between different types of crude oil. Third, there does not exist any publicly available estimate for Saudi surplus capacity (Buyuksahin et al., 2013).

Figure 4.1: OPEC Spare capacity without Saudi Arabia from 25.07.2006 to 03.03.2016. The y-axis shows barrels per day in thousands. Source: EIA.

We have used monthly data from U.S. Energy Information Administration (EIA) to construct a time series. Figure 4.1 show how OPEC spare capacity has developed over time. The spare capacity was low in the period from 2006-2008. In this period the crude oil spot price ranged between \$33 and \$145. Opposite results were found when OPEC spare capacity increased in 2009-2010, when the spot price was less volatile and fluctuated around \$75 (Brunetti et al., 2013). However we observe that this pattern does not hold after mid-2014 when crude oil prices dropped dramatically. OPEC spare capacity remains at low levels despite low crude oil price. Late 2014 OPEC decided to moderately increase their production to sustain their position in the crude oil market, expecting that U.S. production would decrease when the crude oil price dropped below their break-even points. The oil production did not start to fall before mid-2015, partly because the break-even points were not broadly understood by the industry. This contributed to the fall in crude oil price from \$108 in mid-2014 to \$32 January 2016 (Kleinberg et al., 2018).

Following Robe and Wallen (2016) who adopted the method from Brunetti et al. (2013) we design a dummy variable that takes the value 0 when spare capacity is low and 1 when spare capacity is high. We choose to set the separator between high and low spare capacity at 1 million barrels per day.

In addition, we include an interaction variable between the dummy variable for OPEC spare capacity, and the change in U.S. crude oil production described in Section 4.2.2. Decreased WTI crude oil production should increase forward-looking volatility when OPEC has little capacity to produce more crude oil. Contrary, if there are hikes in WTI crude oil production it should reduce forward-looking volatility. Ceteris paribus we expect a positive sign for the interaction term, because both crude oil spare production capacity and output changes are both negatively related to crude oil implied volatility (Robe and Wallen, 2016, p.328). We established earlier in this section that high OPEC spare capacity would indicate more positive skewness. For oil production we explain in Section 4.2.2 that we expect a positive effect on implied skewness. In sum we therefore expect a positive sign for the intereaction term on implied skewness.

4.2.2 North-American Crude Oil Production

The shale oil revolution was stimulated by high crude oil prices after 2003 that made shale oil technology cost competitive. Since then the cost of producing has gone down, leading to the shale oil revolution that took place in 2009 (Kilian, 2016, pp.185). An increase in local crude oil supply is expected to give lower implied volatility for WTI crude oil because of lower price pressure. However this may not be the case if WTI crude oil faces difficulties in reaching international markets (Robe and Wallen, 2016). For implied skewness we expect a positive sign, because more local supply would most likely lead to lower crude oil prices and lower returns. In other words an increase in local supply is expected to give higher positive forward-looking skewness (more right-skewed distribution). We include the weekly changes in US production, lagged one week to avoid endogeneity issues.

4.2.3 Cushing Storage Capacity and Utilization

WTI crude oil is priced at Cushing, Oklahoma, and is connected to the international market via pipelines to the Gulf coast. In early 2011, the production of crude oil in United states' midwest and Canada exceeded the pipeline capacity, leading to a substantial difference in crude oil prices in the United states midwest and "on the water" locations (Borenstein and Kellogg, 2014). This is in line with other studies done on the topic aswell (e.g. Buyuksahin et al., 2013; Fattouh, 2010; Pirrong, 2010), suggesting that infrastructure constraints in Cushing, Oklahoma, leads to differences in prices of WTI crude oil and other types of crude oil.

Until 2007 the logistical bottleneck was to get enough oil into Cushing which often resulted in WTI prices rising to high levels compared to other benchmarks. After 2007, when the crude oil production increased, the new bottleneck became storage capacity in Cushing (Buyuksahin et al., 2013). Traders might not want to hold the future contract to expiration because of high transportation and storage cost for WTI crude oil, and therefore Robe and Wallen (2016) believe that futures around expiration might have higher volatility.

We follow the method used in Robe and Wallen (2016) and estimate the Cushing storage conditions with the slope of the term structure of WTI futures prices. This proxy is based on the work by Fama and French (1987, 1988).

We choose to use the inventory proxy presented in Geman and Ohana (2009). The proxy is presented in Equation 4.1, and gives us the one-year adjusted spread. We have used the annual LIBOR interest rate to adjust for interest rate fluctuations. Figure 4.2 shows the term structure from 25.07.2006 to 03.03.2016. When the slope (blue line) is above the black reference-line the net cost of carry is positive (contango), and likewise we have negative net cost of carry (backwardation) when the slope is below the reference-line.

$$\text{Adjusted spread } = \frac{13\text{Future} - 1\text{Future} * (1 + \text{LIBOR})}{1\text{Future}} \tag{4.1}$$

After calculating the daily slope-variable from 25.07.2006 to 03.03.2016 in matlab, we had 46 dates with missing variables. We correct for this with linear interpolation.

Figure 4.2 shows that the term structure of WTI crude oil futures is mostly in backwardation until 2009, where it changes to contango. This means that until 2009 there is enough storage space, but during the financial crisis the demand for WTI crude oil fell, and storage capacity became a bottleneck. The same happened in the fall of 2014, but this time it was caused by increased supply because of the shale-oil boom.

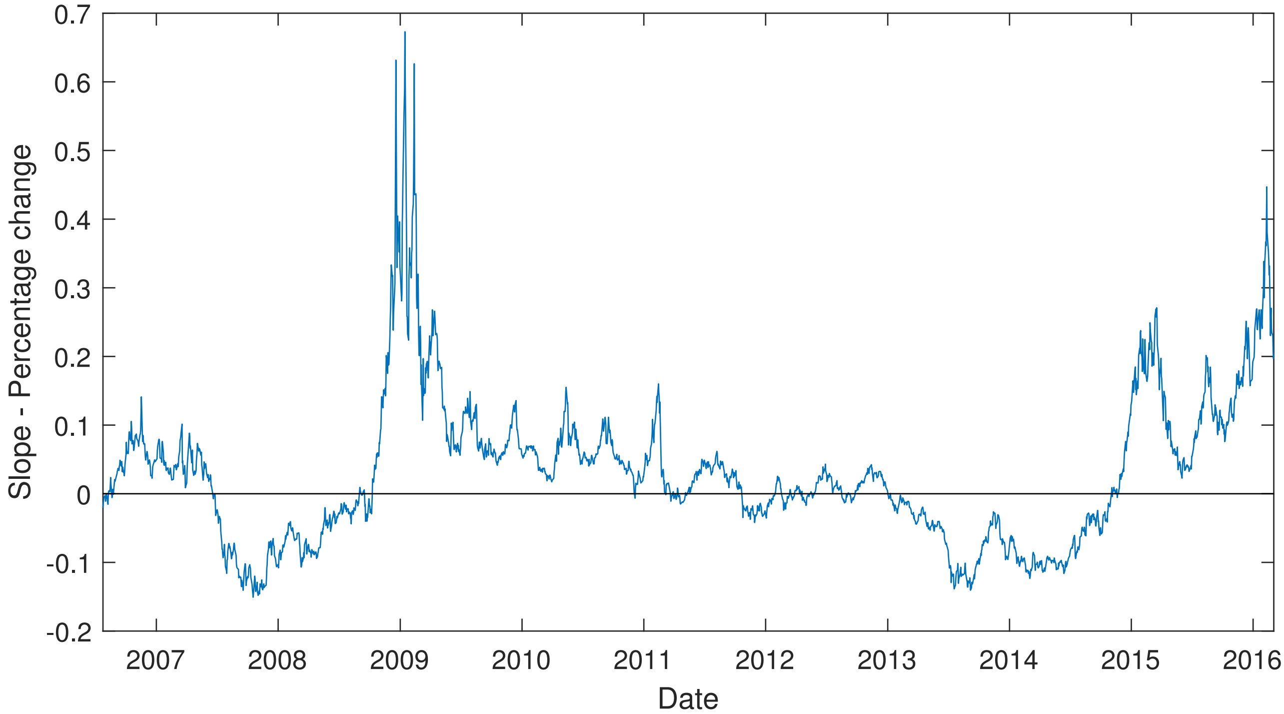

Figure 4.2: Slope of the WTI crude oil term structure from 25.07.2006 to 03.03.2016 calculated using Equation 4.1. When the slope is above the black reference-line we have a state of contango, and backwardation when the slope is below the reference line.

Robe and Wallen (2016) expect a positive relation between storage tension and WTI crude oil implied volatility, meaning that both high and low storage capacity leads to increased implied volatility. We use dummy-variables for the states of backwardation and contango, and multiply them with the level of the slope. In the regression we thereby expect a positive effect on implied volatility both in contango and backwardation, thus a positive sign for contango and a negative sign for backwardation is expected. A higher absolute level of the slope-variable would thereby mean higher implied volatility. Meaning that we have a convex(U-shaped) relationship between the forward curve and volatility (Kogan, Livdan and Yaron, 2009).

Contango indicates low storage space, hence more of the produced oil must reach the market, giving a higher possibility of lower prices. Therefore we expect a negative effect on implied skewness when in contango. Backwardation means enough available storage capacity in Cushing, indicating a bottleneck in getting enough oil into Cushing, meaning that the price will increase due to higher demand relative to supply. In other words backwardation is expected to give increased positive implied skewness, which means a negative sign for the coefficient. To avoid endogeneity issues we follow Robe and Wallen (2016) and use one-day lagged values for the slope variable in the regression.

4.3 Financial Variables

Robe and Wallen (2016) were the first to investigate if financial variables could help explain implied volatility in the oil market. Their hypothesis is that the financial market can have spillover effects to the commodity market. Xiong (2013) highlight that commodity markets' financialization has impacted commodity prices and returns. In this section we will present the financial variables used by Robe and Wallen (2016).

4.3.1 VIX

The VIX-index is an index issued by the Chicago Board Options Exchange. The index measure expectations of near-term volatility in the equity market. The equity market is proxied by the S&P 500 stock index. The VIX-index has been viewed by investors as a good measurement of investor sentiment and market volatility since its introduction in 1993 (*VIX*, 2018). The index is often cited as a measurement of fear in the market.

Gromb and Vayanos (2010) argue that there is a possibility of cross-asset arbitrage between the equity market and the energy market, more accurately the WTI crude oil market in our thesis. This indicates that there could be a relationship between volatility in the equity and oil market, and that the VIX-index can explain a portion of the implied volatility on WTI crude oil. We expect the VIX-index to have a positive relation with implied volatility on WTI crude oil. However higher volatility does not necessarily impact the skewness, so we do not expect the VIX-index to have significant explanation-power on implied skewness.

Figure 4.3 shows the historical price of the VIX-index in blue and the OVX-index in red. The OVX-index is a measure of volatility expectation for crude oil by using the same methodology as the VIX (*OVX*, 2018). Because the OVX-index was created in 2007, we don't have the same amount of data for the OVX as the VIX. From the graph we see that the two indexes move closesly together from 2008 to 2015, where they part, but still move in the same direction most of the time. This supports that there is a relationship between volatility in the equity and oil market.

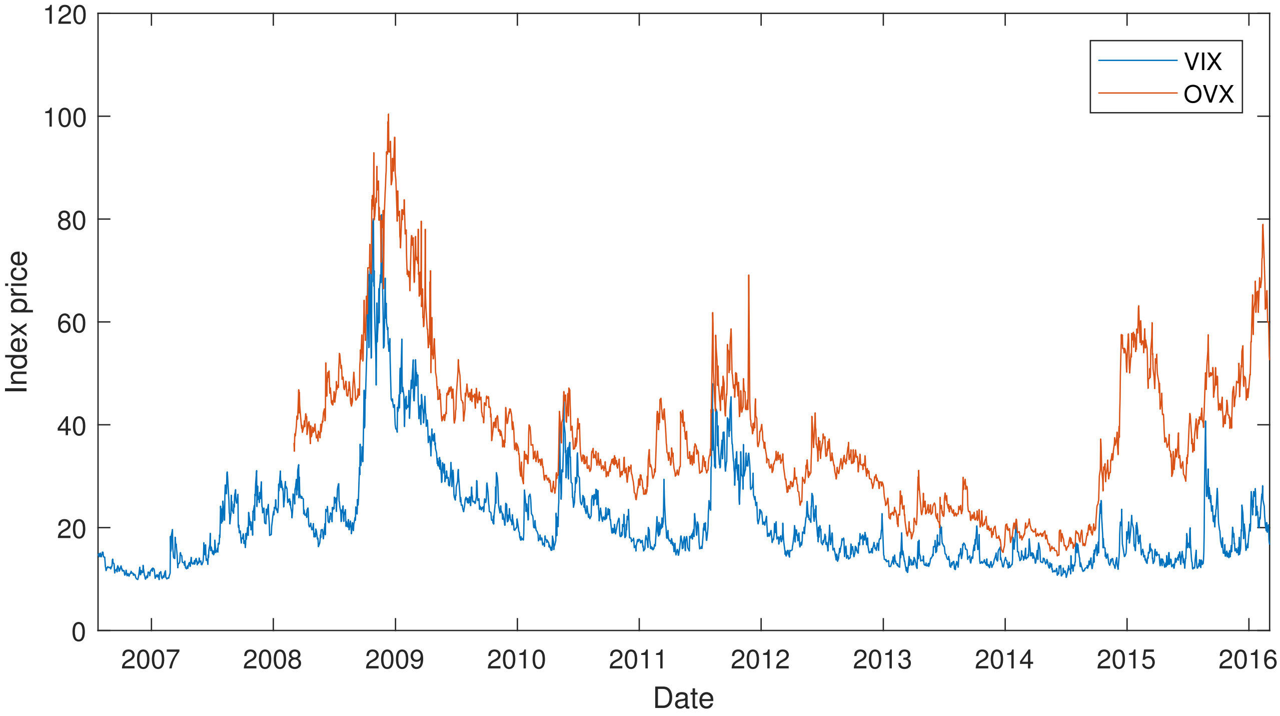

Figure 4.3: Historical values of the VIX and the OVX-index. The blue line show the VIX-index from 25.07.2006 to 03.03.2016. The the red line show the OVX-index from 03.03.2008 to 03.03.2016. Source: CBOE.

From the blue line in Figure 4.3 we see that the VIX-index spikes whenever there is a economic crisis or there is a global event that increases uncertainty in the financial markets. For example the index spiked after the Lehman-brother crisis in 2008 and the U.S credit rating downgrade in August 2011.

4.3.2 Paper Market Liquidity

Robe and Wallen (2016) argue that implied volatility for crude oil and market liquidity should be inversely related. They control for the trading volume for both the underlying future and for the options. Our dataset only contains information about the trading volume for the futures and therefore we only control for this. To capture paper-market liquidity we use Monday-to-Monday changes in WTI matching-maturity futures trading volume (Robe and Wallen, 2016).

The sequential information arrival hypothesis developed by Copeland (1976) says that individuals demand curve changes sequentially when new information is available to them. The theory predict a positive correlation between the absolute value of price changes and volume (Copeland, 1976, p.1167). Thus we could expect a positive correlation between implied volatility for WTI crude oil and future trading volume. An empirical study by Girma and Mougoué (2002) support the sequential information arrival hypothesis, they find that lagged volume and open interest has significant explanatory power for petroleum future spreads volatility. However the empirical research is indecisive, and new research by Abdullahi, Kouhy and Muhammad (2014) find no causal relationship between volume and return for WTI crude oil futures using daily data from 2008 to 2011, thereby rejecting the sequential information arrival hypothesis.

In a different vein an increase in activity in derivatives market is generally known to be connected with an increase of information arrival (Brunetti et al., 2016). When this information is incorporated into the current prices, forward-looking volatility might fall (Robe and Wallen, 2016). Therefore we can not predict the total effect of the futures volume on implied volatility, since it depends on which of the above effects that dominates. We do not expect that the liquidity have any affect on implied skewness, since it does not affect the probability for higher or lower crude oil prices.

4.3.3 Financial Traders' Positions in Crude Oil Paper Markets

Robe and Wallen (2016) argue that although there is not an agreement in the literature that commodity index traders impact commodity price levels, there is an agreement that commodity index traders are long-only and passive. Thus their position in WTI crude oil is not likely to hold predictive power for implied volatility for WTI crude oil. However it is more likely that more active market participant like hedge funds could have a predictive power for implied volatility (Robe and Wallen, 2016).

Hedge funds market share has increased significantly over time, from 2000 to 2006 hedge funds market share grew more than threefold (Brunetti, Büyükşahin and Harris, 2016, p.2). Brunetti, Büyükşahin and Harris (2016) find little evidence that financial speculators destable the market. On the contrary they find that hedge funds stabilize the market and provide valuable liquidity to the market. Alquist and Gervais (2013) find that changes in positions for non-commercials did not predict oil-price changes. They argue that the increased crude oil price from 2003-2008 can be explained by macroeconomic fundamentals and not financial speculation.

Following Robe and Wallen (2016) we include a proxy for hedge fund activity in our analysis. They use trader-posistion data for WTI futures published by the U.S Commodity Futures Trading Commission (CFTC) for the NYMEX's WTI futures markets and use the methodology from Working (1960) to calculate the Working T index. The CFTC's weekly Commitments of Traders report (COT) break down the total open interest between the positions of commercials hedgers and non-commercials. From 2009 the report also split non-commercials between hedge funds and other non-commercial traders. Even though this adjustment came in 2009, data containing this information is available from 2006 (Robe and Wallen, 2016, p.331), limiting our research to the period when the information is available.

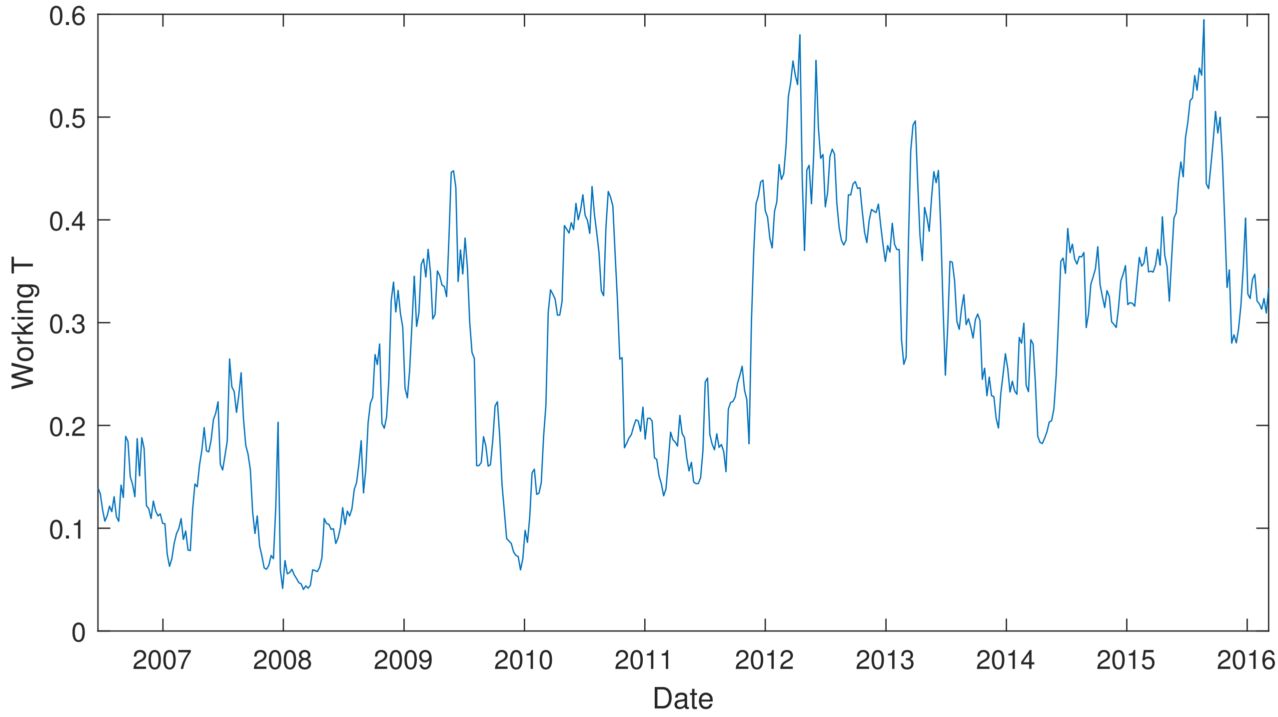

Figure 4.4: Historical development from 25.07.2006 to 03.03.2016 of Working T index, subtracted with 1. Source: Quandl.

Figure 4.4 show a long-term increase in the Working T index, and that the index is quite volatile. We expect a positive relationship with implied volatility for WTI crude oil, because hedge funds enter markets when they expect a price movement to occur. For implied skewness we do not expect the variable to be significant, since hedge funds take long positions as well as short positions in the market.

4.3.4 Time-To-Maturity

Following Robe and Wallen (2016) we include a variable for number of days left before expiration on the underlying future. Based on Samuelson (1965) we should expect an increase in the implied volatility when delivery date gets closer. Samuelson argues that more information is available to the investor closer to maturity, causing increased volatility (Bessembinder et al., 1996). Thus longer time to maturity would decrease implied volatility. For implied skewness we do not expect the variable to be significant, since time-to-maturity is not expected to impact the WTI crude oil future price. Table 4.1 summarise the expected signs for each independent variable for the regression models on both implied volatility and implied skewness.

Beta	Variable	Implied Volatility	Implied Skewness
β_1	lagged dependent variable	+	+
β_2	Overall financial uncertainty (Vix-index)	+	
β_3	U.S business cycle (ΔADS-index)	-	-
β_4	ΔOil output	-	+
β_5	Dummy for OPEC Spare capacity	-	+
β_6	ΔOil output * Spare	+	+
β_{7A}	Oil storage constraints	- (backwardation)	- (backwardation)
β_{7B}	Oil storage constraints	+ (contango)	- (contango)
β_8	Financial speculation (ΔWorking T index)	+	
β_9	Days to expiration (TTM)	-	
β_{10}	ΔFutures volume	+/-	

Table 4.1: Expected signs for the regression results on both implied volatility and implied skewness. Blank spaces implies that we do not expect the variable to have any impact.

5 Data

This chapter presents the descriptive statistics and stationarity tests for each variable used in our regression models.

5.1 Descriptives

Table 5.1 shows the descriptive statistics of our data at a weekly frequency from 25.07.2006 until 01.03.2016, giving a total of 490 observations. The mean for implied volatility is 0.22 and is lower than the mean calculated by Robe and Wallen (2016), as discussed in Section 3.7. They find a mean of 0.38 for implied volatility for 1 month WTI crude oil.

Implied skewness has a mean of -0.09 and support our assumption in Section 3.8 that the implied skewness fluctuates around -0.1. In average there is thereby more expensive to hedge downside risk than upside risk in the WTI crude oil market.

Variable	Mean	Median	Max	Min	StDev	Skewness	Kurtosis	Obs
Implied Volatility	0.2235	0.2045	0.7186	0.0700	0.1033	1.7128	6.8694	490
Implied Skewness	-0.0901	-0.0915	0.1610	-.3082	0.0654	0.0587	4.5339	490
VIX	0.2109	0.1822	0.8006	0.0997	0.1009	2.1886	9.1228	490
ΔADS	0.0032	-0.0001	0.5253	-0.4687	0.0973	0.1359	7.5272	490
ΔOil Output	0.0011	0.0009	0.1414	-.02216	0.0207	-1.8689	43.0554	490
Dummy Spare	0.2571	0	1	0	0.4375	1.1113	2.2350	490
Backwardation	-0.0231	0	0	-0.1416	0.0382	-1.5329	4.0113	490
Contango	0.0548	0.0227	0.5998	0	0.0838	2.5316	11.4487	490
ΔWorking T	0.0007	0.0001	0.1179	-0.1595	0.0313	-0.3560	6.0270	490
Time to maturity	55.9918	56	72	39	8.8162	-0.0004	1.8221	490
ΔFutures Volume	-0.0019	0.0310	1.3812	-1.5024	0.4973	-0.3086	2.9184	490
OVX	0.3838	0.3497	0.9893	0.1474	0.1572	1.2145	1.8621	413

Table 5.1: Summary statistics at a weekly freuqncy from 25.07.2006 to 01.03.2016 for the variables used in the regressions. For information about where data is collected from, see Appendix A.1.

The mean and median for contango is higher in absolute value than for backwardation. Historically the crude oil market has been dominated of backwardation. However Figure 4.2 show that for our time-period contango is more dominating and has higher peaks than backwardation. This may effect our regression results for storage tension in times of backwardation. The descriptive statistic for the other variables are similar with the findings of Robe and Wallen (2016).

5.2 Tests of Stationarity

For the regression-coefficients to be valid we check if the variables are stationary. When a time-series is stationary there is no trends or patterns that can lead to invalid results. A time-series is defined as stationary if:

$$E(y_t) = \mu \tag{5.1}$$

$$Var(y_t) = E[(y_t - \mu)(y_t - \mu)] = \sigma^2 \tag{5.2}$$

$$Cov(y_{t2}, y_{t1}) = E[(y_{t2} - \mu)(y_{t1} - \mu)] = \Omega_{t2-t1}, \qquad \forall\, t_2, t_1 \tag{5.3}$$

The first and second condition is that the time-series should have constant mean and variance over time, meaning that the data is independent of time. This means that most observations will be close to the mean, but how close depends on the variance. The third condition is that a joint distribution between adjacent variables in the data generating process is independent of time Alexander (2008*a*). Economic variables on level-form are normally not stationary. Normally we solve the problem by using first difference form, and at the same time making them stationary. For example Figure 4.4 shows that Working T is not stationary, and including it in the regression without using first difference changes could lead to severe consequences for the regression results.

One consequence of using a non-stationary variable in regression analysis is spurious regressions (Studenmund, 2016). Spurious regressions is regressions where a independent variable have the same underlying trend as the dependent variable, making it significant and normally giving the model high overall fit. While the model seem to give good results, there is no real underlying causal relationship between the variables.

Variable	Augmented Dickey-Fuller	Phillips-Peron
Implied Volatility	0.3489	0.0003***
Implied Skewness	0.0002***	0.0000***
VIX	0.0210**	0.0022***
ΔADS	0.0000***	0.0000***
ΔOil Output	0.0000***	0.0000***
Dummy Spare	0.4704	0.4629
Dummy Spare * ΔOil Output	0.0000***	0.0000***
Backwardation	0.1058	0.1125
Contango	0.0709*	0.0072***
ΔWorking T	0.0000***	0.0000***
Time to maturity	0.0000***	0.0000***
ΔFutures Volume	0.0000***	0.0000***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.2: Stationarity tests. ***, ** and * indicate the significance level at the 0.01, 0.05 and 0.1, respectively. We have used 4 lags for the Augmented Dickey-Fuller test, following Robe and Wallen (2016). For the Phillips-Perron test we have used the default lag-number set by STATA, as explained in Appendix B.2.

Table 5.2 shows the results of our stationarity-tests. We have conducted both a Augmented Dickey-Fuller and a Phillips-Peron test. The tests are explained in Appendix B.1 and B.2. For p-values we use the MacKinnon approximate p-value (MacKinnon, 1994).

The test results in Table 5.2 shows stationarity for most variables, except for the spare-dummy, which is not expected to be stationary, and backwardation. Backwardation is close to being significant at 10% significance level, and is most likely not stationary because of few periods of backwardation and thereby not enough datapoints. Implied volatility is stationary according to the Phillips-Perron test, but not according to the Augmented Dickey-Fuller test. We will threat it as stationary, as implied volatility is viewed as stationary by investors, even though it may not always be confirmed by unit root tests (Alexander, 2008*c*).

6 Empirical Analysis

This chapter presents the empirical results and tests for misspecification. In addition, we analyse large residuals from the regression model on implied volatility and explain that they represent unexpected shocks to the crude oil market. The last part of this chapter contains an robustness-check with an alternative measure of forward-looking volatility.

6.1 Methodology

The goal of this thesis is to examine what drives changes in forward-looking volatility and skewness for WTI crude oil. We measure the front month implied volatility and implied skewness for WTI crude oil on Tuesdays and regress each on: the one week lagged implied volatility or implied skewness, one-day lagged VIX (Monday), contemporaneous changes (Tuesday) in macroeconomic fundamentals for the U.S. measured by the ADS Index, oil market fundamentals measured by a SPARE dummy, the net cost of carry measured by the one-day lagged term structure SLOPE, the changes in North American production lagged one week, changes in the intensity of speculation measured by the Working T index lagged one week, a control variable for time to expiration and a liquidity control variable measured by changes in the relevant future volume. This leads to the two following regression equations:

$$
\begin{aligned}
\text{Implied Volatility}_t &= \alpha + \beta_1 \text{Implied Volatility}_{t-1} + \beta_2 \text{VIX}_{t-1} + \beta_3 \Delta \text{ADS}_{t-1} \\
&+ \beta_4 \Delta \text{Oil Output}_{t-1} + \beta_5 \text{SpareDummy}_t + \beta_6 (\Delta \text{Oil Output}_{t-1} * \text{Spare}_t) \\
&+ \beta_{7A} \text{SLOPE}_{t-1}\Big|^{\text{Backwardation}} + \beta_{7B} \text{SLOPE}_{t-1}\Big|^{\text{Contango}} + \beta_8 \Delta \text{WorkingT}_{t-1} \\
&+ \beta_9 \text{TTM}_t + \beta_{10} \Delta \text{Futures Volume}_{t-1} + \epsilon_t
\end{aligned}
\tag{6.1}
$$

$$
\begin{aligned}
\text{Implied Skewness}_t = \alpha + \beta_1 \text{Implied Skewness}_{t-1} + \beta_2 \text{VIX}_{t-1} + \beta_3 \Delta\text{ADS}_{t-1} \\
+ \beta_4 \Delta\text{Oil Output}_{t-1} + \beta_5 \text{SpareDummy}_t + \beta_6 (\Delta\text{Oil Output}_{t-1} * \text{Spare}_t) \\
+ \beta_{7A} \text{SLOPE}_{t-1}\Big|^{\text{Backwardation}} + \beta_{7B} \text{SLOPE}_{t-1}\Big|^{\text{Contango}} + \beta_8 \Delta\text{Workings}_{t-1} \\
+ \beta_9 \text{TTM}_t + \beta_{10} \Delta\text{Futures Volume}_{t-1} + \epsilon_t
\end{aligned}
\tag{6.2}
$$

6.2 Empirical Results

Variables	Implied Volatility	Implied Skewness
Lagged dependent	0.53437***	0.77395***
	(0.02348)	(0.02833)
VIX	0.29354***	-0.00442
(One-day lagged)	(0.01821)	(0.01990)
ADS	-0.00897	-0.02139
(One-day lagged change)	(0.01278)	(0.01886)
Oil Output	-0.04676	0.01988
(One-week lagged pct change)	(0.06336)	(0.09330)
Dummy SPARE	-0.01702***	0.00739
	(0.00311)	(0.00458)
Dummy Spare * Oil Output	0.05153	0.18186
(One-week lagged pct change)	(0.17739)	(0.26122)
Backwardation	0.16432***	-0.13231**
(One-day lagged slope)	(0.03683)	(0.05419)
Contango	0.26511***	0.01295
(One-day lagged slope)	(0.02216)	(0.02544)
Working T	-0.00553	0.04189
(One-week lagged change)	(0.03872)	(0.05713)
Time to maturity	0.00329***	-0.00107***
	(0.00014)	(0.00020)
Futures Volume	0.01131***	-0.00278
(One-day lagged change)	(0.00250)	(0.00362)
Constant	-0.14865***	0.03436***
	(0.00904)	(0.01281)
Observations	490	490
Adjusted R-squared	0.9335	0.6397

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.1: Regression results at a weekly frequency from 25.07.2006 to 01.03.2016 for implied volatility and implied skewness. ***, ** and * indicate the significance level at the 0.01, 0.05 and 0.1, respectively. Regression results are reported with standard error in parentheses. For information about where data is collected from, see Appendix A.1.

Table 6.1 shows regression results for both implied volatility and implied skewness. In Appendix A.4 we have included the regression results for more parsimonious regression models including only the significant variables. From Table A.4 in Appendix A.4 we see that the parsimonious regressions have the same sign and significance as the regressions presented in Table 6.1, thereby supporting that our regressions is robust.

6.3 Results Implied Volatility

In this section we present the regression result in Table 6.1 for implied volatility for each variable. Adjusted R-squared is 0,93, indicated that we have managed to capture the most important determinants of forward looking volatility on WTI crude oil.

6.3.1 Lagged Implied Volatility

Lagged implied volatility is positive and significant, meaning that if implied volatility a week ago was high, we expect high implied volatility today. This is natural since volatility tend to move in clusters.

6.3.2 VIX

Gromb and Vayanos (2010) argue that there could be a spillover effect between volatility in equity markets and energy markets. We therefore expect a positive relationship between VIX and WTI crude oil implied volatility. Table 6.1 shows that VIX has significant impact on implied volatility for WTI crude oil. When uncertainty in the equity market increases, implied volatility for WTI crude oil tend to rise aswell. Our findings align with the findings of Robe and Wallen (2016). They argue that this can help explain the very low levels of implied volatility in 2012-2014, since uncertainty in the equity market was very low for the same time period.

6.3.3 Business Cycles

We expect that an increase in U.S activity measured by the ADS index would have a inverse relation with WTI crude oil implied volatility. Similar to Robe and Wallen (2016) we find that the ADS-index is not statistically significant.

6.3.4 Oil Production Fundamentals

The market for WTI crude oil expect less volatility in the future when OPEC spare capacity is high, because OPEC can change their production if the demand for crude oil increases. OPEC spare capacity outside Saudi Arabia is significant and has the expected negative sign. Higher production surplus capacity is associated with lower implied volatility for crude oil. Our regression results confirm Robe and Wallen's (2016) conclusion that it is important to control for physical constraints on the production of crude oil.

Both U.S. oil output, and the interaction term between U.S. oil output and spare capacity are not statistical significant, corresponding with the results of Robe and Wallen (2016).

6.3.5 Oil Storage

From Table 4.1 and Section 4.2.3 we see that we expect a positive sign for contango. Contango is a proxy for low storage capacity in Cushing, which drives WTI crude oil prices down, because oil can not be stored and needs to enter the commodity market. This leads to increased implied volatility. Table 6.1 shows that contango is significant and positive in our regression, confirming the theory from Section 4.2.3. This is in line with the findings of Robe and Wallen (2016). However, 6 month implied volatility contango is not significant in their results, indicating that low storage capacity only boost implied volatility in the short term (Robe and Wallen, 2016).

Backwardation is a proxy for high storage capacity and represent the bottleneck of getting enough oil into Cushing, resulting in WTI crude oil prices rising to higher levels. Thus we expect a negative sign in the regression, meaning that a higher level of backwardation increases implied volatility. However our results in Table 6.1 shows a significant positive coefficient, which is not in line with our expectations. A possible explanation could be that we have few periods of backwardation, as shown in Figure 4.2. We believe that increasing the time-period to include more periods of backwardation could change the results.

Robe and Wallen (2016) does not find backwardation to be significant for 1 month implied volatility, but find it significant and positive for 6 month implied volatility. They calculate the slope-variable using the difference between nearby and first-deferred futures. We use the difference between nearby and 13 month futures, as shown in Equation 4.1, meaning that we calculate the slope-variable further out on the term structure than Robe and Wallen (2016). How far out on the term structure you calculate the slope-variable can have an impact on the result, but as far as we know, there have not been conducted any studies on the subject.

6.3.6 Speculative Activity in Oil Markets

To check if intensity of oil-market speculation have any effect on WTI crude oil implied volatility we use the Working T index. The expectation is that increased speculation should increase implied volatility. However both our results and the results of Robe and Wallen (2016) show that the Working T index is not significant, and do not help explain forward-looking volatility for WTI crude oil. Robe and Wallen (2016) point out that this might be because we only have public available information in the Working T index.

6.3.7 Paper Market Liquidity

We expect two opposite effects on implied volatility for WTI crude oil futures trading volume, and therefore no expected sign. An increase in future volume is significant positive related to WTI crude oil implied volatility, supporting the sequential information arrival hypothesis. Robe and Wallen (2016) also finds that future volume is significant and positive related to implied volatility. In addition they control for option volume in their regression, but did not find this variable to be significant.

6.3.8 Time-To-Maturity Effects

We expect that longer time to maturity should decrease implied volatility, because of the Samuelson (1965) effect. Our results in Table 6.1 shows the contradictory. We detected this pattern when looking at daily volatility smiles for each month, as the at-the-money implied volatility trends downward closer to maturity. A possible explanation for this is that liquidity in WTI crude oil futures decreases the last weeks before maturity, because most traders close out their positions or roll over to contracts with longer maturity, generating a kind for seasonality (Brunetti et al., 2013). This is supported by Ripple and Moosa (2009), which finds that trading volume and open interest dominate the Samuelson effect for WTI crude oil. On the other hand, Robe and Wallen (2016) finds the expected negative sign. Including open interest in the regression could give more insight into this matter.

6.4 Residual Analysis Implied Volatility Regression

Our model take into account macroeconomic fundamentals, physical-market conditions and financial variables. High residuals should thereby reflect unexpected shocks in the crude oil market, or reflect shocks on other omitted variables. Robe and Wallen (2016) study monthly Short Term Energy Outlook published by the EIA and identify declines in world production greater than 1 percent. They find that high residuals take place when there are unexpected oil market disruptions.

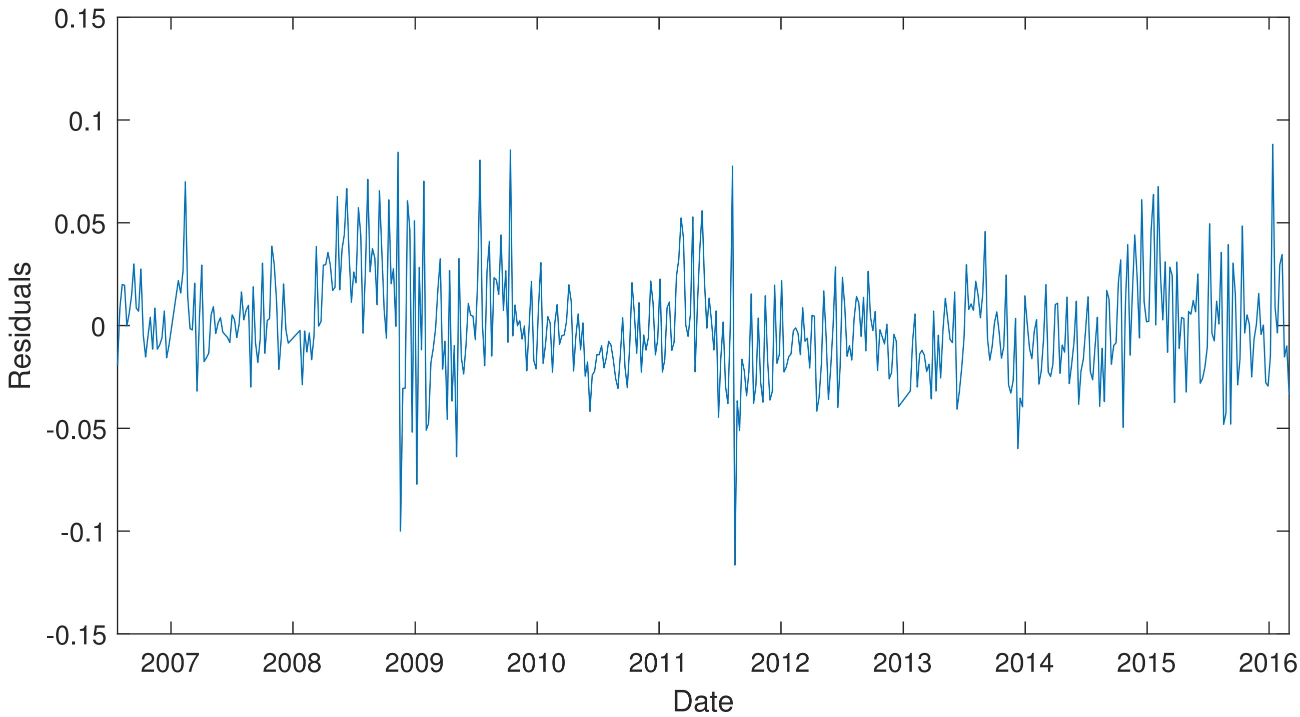

Figure 6.1: Timeline of the residuals from the regression on implied volatility from 25.07.2006 to 01.03.2016.

Figure 6.1 shows the residuals from the regression model on implied volatility. We find the same pattern as Robe and Wallen (2016) have in their analysis for the overlapping time period. At the 13th February 2007 we have a high residual. This might reflect that at 8th January 2007 Russia cut oil supplies to Poland, Germany and Ukraine, aimed to prevent Belarus illegally shipping off oil (*Russia oil row hits Europe supply*, 2007). Like Robe and Wallen (2016) we can clearly identify the period of the financial crisis that started in the fall of 2008 and lasted approximately a year. In August 2011 we also observe high residuals, which is likely caused by the US credit-rating downgrade that took place in August 2011, when Standard & Poor's downgraded the long term sovereign credit rating from AAA to AA+ (Peston, 2011).

After the US credit rating shock follows a long period with small residuals. From October 2014 to February 2015 we observe increased residuals, likely caused by the dramatic and unexpected drop in crude oil prices (Baumeister and Kilian, 2016). The last high residual in our sample is at 12th January 2016. This can be explained by the WTI crude oil price. From 1st January 2016 to 12th January 2016 the WTI crude oil price fell by about 18%, ending at $30.42. After studying the residuals from the regression on implied volatility we agree with Robe and Wallen's (2016) argument that high residuals can be explained by unexpected shocks.

6.5 Results Implied Skewness

As Table 6.1 shows, we do not find as many significant variables for implied skewness as we find for implied volatility. This is partly expected since higher moments of the distribution is harder to explain, because higher moments have more spikes. In addition several of the independent variables in Chapter 4 is not expected to impact implied skewness. The regression model for implied skewness shows an adjusted R-squared of 0.64, which is medium high. This shows that the regression model manage to describe parts of what drives changes in implied skewness. In the following analysis for implied skewness we choose to only focus on the significant variables.

6.5.1 Lagged Implied Skewness

Table 6.1 shows that implied skewness a week ago is highly significant in explaining implied skewness today. This is expected when we look at the Figure 3.4, which shows that implied skewness tends to move in periods between positive and negative. In other words implied skewness moves in clusters.

6.5.2 Oil Storage

We expect a negative effect for contango on implied skewness. Contango represent low storage space and thus high supply relative to demand, which gives a pressure on lower WTI crude oil price. Our regression does not confirm our expectation since contango is not significant.

For backwardation we expect a positive effect on implied skewness, indicating a negative sign in the regression. This is because higher levels of backwardation indicates a bottleneck of getting enough oil into Cushing, giving oil distributors limited possibility to answer demand-shocks. Therefore an increase in demand for WTI crude oil would result in higher WTI crude oil prices.

This is illustrated in Figure C.3 in the appendix, which shows an equilibrium in the WTI crude oil market, with the typical demand and supply curves found in this market. A hypothetical positive shift in the demand curve would lead to an increased WTI crude oil price. Thus higher backwardation should indicate more positive skewness, which means a negative sign for the coefficient. Our regression on implied skewness shows that backwardation is significant and has the expected negative sign, confirming our expectations.

For our time period we conclude that high storage capacity in Cushing helps explain more positive implied skewness, but we do not find evidence that low storage capacity in Cushing have impact on implied skewness.

6.5.3 Oil Production Fundamentals

The dummy variable that measures OPEC spare capacity without Saudi Arabia is expected to have a positive sign for implied volatility since high spare capacity is a reflection of lower WTI crude oil prices and therefore leads to more positive implied skewness. From the regression results in Table 6.1 we see that the spare dummy has the expected positive sign and a p-value of 0.107, which means that the variable is almost statistically significant at a ten percent level. Therefore we argue that this variable could be an important factor in explaining changes in implied skewness over time.

6.5.4 Time-To-Maturity Effects

We expect time to maturity to have no effect on implied skewness. However the regression results in Table 6.1 show that time to maturity is significant and negative. This means that closer to maturity we observe more positive skewness. To our knowledge there is no documented economic reason for this effect. Studies (Whaley, 1986; Hull and White, 1987) find that both the Black-Scholes and the Barone-Adesi and Whaley method have increasing overpricing of options with increasing time to maturity. This may impact the shape of the volatility smile. Therefore the negative effect on implied skewness from increased time to maturity could be a result of using different methods for each side of the volatility smile.

6.6 Testing for Misspecification

We have used VIF-criterias to check for multicollinearity. Table 6.2 shows VIF-values for both regressions. None of the VIF-indexes is above 5, which is a rule of thumb for when multicollinearity is severe according to Studenmund (2016). This indicate that we do not have any problems with multicollinearity, but we can not completely rule out the possibility of multicollinearity, because the test has no hard-and-fast decision rule. From Table A.3 in the appendix we see that none of the explanatory variables have correlations above 0.8, which is often viewed as a critical value for when multicollinearity is severe. This is another argument supporting that we have no multicollinearity in our regression models.

Residual tests for normality, heteroskedasticity, autocorrelation and linear form are found in Table 6.3. First, we see that both the regression models have residuals that do not follow the normal distribution. Secondly, we see that the regression model for implied volatility has a problem with heteroskedasticity. This can also be seen from the residual plot in Figure C.4 in the appendix. Even though we have heteroskedasticity the OLS results will be unbiased, but they might not be the most efficient of all linear unbiased estimators (Alexander, 2008*a*). For the regression model for implied skewness we do not have heteroskedasticity according to the Breusch-Pagan test.

Variable	Implied Volatility	Implied Skewness
Lagged Implied Volatility	4.05	
Lagged Skewness Proxy		1.08
VIX	2.32	1.28
ΔADS	1.06	1.07
ΔOil Output	1.19	1.19
Dummy Spare	1.28	1.28
Dummy Spare * ΔOil Output	1.16	1.16
Backwardation	1.37	1.36
Contango	2.37	1.44
ΔWorking T	1.01	1.02
Time to maturity	1.06	1.03
ΔFutures Volume	1.06	1.03

Table 6.2: VIF-values for the regression models on implied volatility and implied skewness.

Both regression models test positive for autocorrelation. The estimators are still unbiased, but here as well it might not be the most efficient unbiased estimators (Alexander, 2008*a*). When it comes to the Ramsey RESET test none of the models reject the null hypothesis for a 5%

significance level. However the regression model for implied volatility is significant for a 10% significance level. This means that there might be a non-linear combination of the explanatory variables that have predictive power on the dependent variable. For the regression model for implied skewness the p-value is 0.5402, and thereby no chance of rejecting the null hypothesis.

Test	Implied Volatility	Implied Skewness
Jarque-Bera	0.0000***	0.0000***
Breusch-Pagan	0.0000***	0.1129
Breusch-Godfrey	0.0002***	0.0009***
Ramsey RESET test	0.0671*	0.5402

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.3: Summary of misspecification-tests for residuals from regression models for implied volatility and implied skewness. ***, ** and * indicate the significance level at the 0.01, 0.05 and 0.1, respectively

Autorcorrelation, heteroskedasticity and non-normal distributed residuals violates the assumptions for OLS, and at the same time inspire further research. An extension of this master thesis could include other regressions-methods, such as generalized least squares to correct for autocorrelation, or weighted least squares to correct for heteroskedasticity (Studenmund, 2016). An alternative is to use White's robust standard errors instead of the standard errors of the OLS estimators, or to follow the Newey-West procedure (Alexander, 2008*a*).

6.7 Alternative Measure of Forward-Looking Volatility

As explained in Section 4.3.1 the OVX-index is calculated in the same way as the VIX-index, but for crude oil. As Robe and Wallen (2016) we test the robustness of our regression results by replacing implied volatility with the OVX-index in the regression model. Table 6.4 shows the results of the regression with the OVX-index as the dependent variable. Since the OVX-index was first introduced in 2007 we only have 413 observations in this regression, compared to 490 observations in the regression for implied volatility.

The results of the regression are highly similar to the initial regression presented in Table 6.1. These results supports that our regression with calculated implied volatility is robust. We have the same significant variables, the only difference being that time to maturity is only significant at a 5% significance level instead of 1%. It is interesting that this robustness check support the results for time to maturity in the initial regression in Table 6.1. Compared with Robe and Wallen's (2016) regression results on the OVX-index, we find similar results, except that they do not find time to maturity to be significant.

Variables	OVX-index
Lagged IV	0.70105***
	(0.04286)
VIX	0.49193***
(One-day lagged)	(0.03369)
ADS	-0.02393
(One-day lagged change)	(0.02302)
Oil Output	-0.07168
(One-week lagged pct change)	(0.11660)
Dummy Spare	-0.04854***
	(0.0.00557)
Dummy Spare * Oil Output	0.11250
(One-week lagged pct change)	(0.31057)
Backwardation	0.63298***
(One-day lagged slope)	(0.07746)
Contango	0.35720***
(One-day lagged slope)	(0.03950)
Working T	0.02688
(One-week lagged change)	(0.07077)
Time to maturity	0.00054**
	(0.0.00026)
Futures Volume	0.01479***
(One-day lagged change)	(0.00454)
Constant	0.09186***
	(0.01718)
Observations	413
Adjusted R-squared	0.913

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.4: Regression results at a weekly frequency from 04.03.2008 to 01.03.2016 with the OVX-index as the dependent variable. ***, ** and * indicate the significance level at the 0.01, 0.05 and 0.1, respectively. Regression results are reported with standard error in parentheses. For information about where data is collected from, see Appendix A.1.

7 Conclusion

In this thesis we calculate volatility smiles from 25.07.2006 to 03.03.2016 for WTI crude oil using nearby futures and future options. We use factors from Robe and Wallen (2016) to examine what drives changes in implied volatility and implied skewness. To our knowledge our study is the first that determines drivers of implied skewness using a proxy for implied skewness.

Our results for implied volatility support many of Robe and Wallen (2016) findings. We confirm that both physical oil market fundamentals for WTI crude oil and financial uncertainty in the equity market, measured by the VIX-index, contain information about implied volatility. This helps explain the record low levels of implied volatility for WTI crude oil, since it reflects the low levels of uncertainty in the financial market. The U.S. business cycle is not significant, showing that the VIX-index is a better measure for market sentiment.

Our results for implied skewness have fewer significant explanatory variables than implied volatility, but our results suggest that physical oil market fundamentals is key to understand what drives changes in implied skewness. Storage-tension is significant with a positive effect when the market is in a state of backwardation, meaning that low supply relative to demand results in more positive forward-looking skewness. OPEC spare capacity without Saudi-Arabia is almost significant. This reflect that high OPEC spare capacity tend to occur when WTI crude oil prices are low, which leads to more positive forward-looking skewness. In addition the control variable for time to maturity is significant with a negative effect, meaning that longer time to maturity decreases implied skewness, resulting in a more left-skewed distribution.

Our result have importance for investors and participant in oil-related industries that want to reduce WTI crude oil risk, because we find significant factors that explain changes in forward-looking volatility and skewness. These factors can be used to understand and predict changes in the distribution of WTI crude oil over time.

An interesting expansion of our thesis would be to test the prediction ability of implied volatility and implied skewness for our models. One way to do this is comparing predictions from our models with an ARIMA benchmark model. Forecasting distributions of WTI crude oil and predict VaR estimations could also be interesting future research. Another possible extension includes using a proxy for kurtosis and examine if the same variables explain kurtosis as well. Including realised volatility as an explanatory variabel in the model using intraday data could also be a fruitful extension.

We find a clear connection between WTI crude oil implied volatility and general equity market uncertainty measured by the VIX. Similar results could likely be found for other commodities. Further research is however necessary to make such a generalization. To strengthen our findings for the crude oil market, reasearch on other types of crude oil and options with longer maturities could also be a potential extension.

Bibliography

Abdullahi, S. A., Kouhy, R. and Muhammad, Z. (2014) Trading volume and return relationship in the crude oil futures markets, *Studies in Economics and Finance*, 31(4), pp. 426–438.

Alexander, C. (2008*a*) *Market risk analysis. : Volume I, : Quantitative methods in finance*, Chichester, England: John Wiley & Sons, Ltd.

Alexander, C. (2008*b*) *Market risk analysis. : Volume ii, : Practical financial econometrics*, Chichester, England: John Wiley & Sons, Ltd.

Alexander, C. (2008*c*) *Market risk analysis. : Volume III, : Pricing, hedging and trading financial instruments*, Chichester, England: John Wiley & Sons, Ltd.

Alquist, R. and Gervais, O. (2013) The role of financial speculation in driving the price of crude oil, *The Energy Journal*, 34(3), pp. 35–54.

Aruoba, S. B., Diebold, F. X. and Scotti, C. (2009) Real-time measurement of business conditions, *Journal of Business & Economic Statistics*, 27(4), pp. 417–427. doi: 10.1198/jbes.2009.07205.

Barberis, N. and Huang, M. (2008) Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review*, 98(5), pp. 2066–2100. doi: 10.1257/aer.98.5.2066.

Barone-Adesi, G. and Whaley, R. (1987) Efficient analytic approximation of american option values, *The Journal of Finance*, 42(2), pp. 301–320. doi: 10.1111/j.1540-6261.1987.tb02569.x.

Barsky, R. B. and Kilian, L. (2004) Oil and the macroeconomy since the 1970s, *Journal of Economic Perspectives*, 18(4), pp. 115–134.

Bates, D. (1991) The crash of 87 - was it expected - the evidence from options markets, *Journal Of Finance*, 46(3), pp. 1009–1044. doi: 10.1111/j.1540-6261.1991.tb03775.x.

Baumeister, C. and Kilian, L. (2016) Understanding the decline in the price of oil since june 2014, *Journal of the Association of Environmental and Resource Economists*, 3(1), pp. 131–158.

Bera, A. K. and Jarque, C. M. (1981) Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte carlo evidence, *Economics letters*, 7(4), pp. 313–318. doi: 10.1016/0165-1765(81)90035-5.

Bessembinder, H., Coughenour, J. F., Seguin, P. J. and Smoller, M. M. (1996) Is there a term structure of futures volatilities? reevaluating the samuelson hypothesis, *Journal of Derivatives*, 4(2), pp. 45–58. doi: 10.2469/dig.v27.n3.123.

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities, *The Journal of Political Economy*, 81(3), pp. 637–654. doi: 10.2307/1831029.

Borenstein, S. and Kellogg, R. (2014) The incidence of an oil glut: Who benefits from cheap crude oil in the midwest?, *The Energy Journal*, 35(1), pp. 15–33. doi: 10.5547/01956574.35.1.2.

Breeden, D. T. and Litzenberger, R. H. (1978) Prices of state-contingent claims implicit in option prices, *The Journal of Business*, 51(4), pp. 621–651.

Brunetti, C., Buyuksahin, B., Robe, M. A. and Soneson, K. R. (2013) Opec "fair price" pronouncements and the market price of crude oil, *The Energy Journal*, 34(4), pp. 79–108. doi: 10.5547/01956574.34.4.5.

Brunetti, C., Büyükşahin, B. and Harris, J. H. (2016) Speculators, prices, and market volatility, *Journal of Financial and Quantitative Analysis*, 51(5), pp. 1545–1574. doi: 10.1017/S0022109016000569.

Buyuksahin, B., Lee, T. K., Moser, J. T. and Robe, M. A. (2013) Physical markets, paper markets and the wti-brent spread, *The Energy Journal*, 34(3), pp. 129–151.

Büyükşahin, B. and Robe, M. A. (2011) Does paper oil matter?, *Working paper, CFTC and American University*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1855264 (accessed: 11.05.2018).

Chevillon, G. and Rifflart, C. (2009) Physical market determinants of the price of crude oil and the market premium, *Energy Economics*, 31(4), pp. 537–549. doi: 10.1016/j.eneco.2009.01.002.

Copeland, T. E. (1976) A model of asset trading under the assumption of sequential information arrival, *The Journal of Finance*, 31(4), pp. 1149–1168. doi: 10.1111/j.1540-6261.1976.tb01966.x.

Corradi, V., Distaso, W. and Mele, A. (2013) Macroeconomic determinants of stock volatility and volatility premiums, *Journal of Monetary Economics*, 60(2), pp. 203–220. doi: 10.1016/j.jmoneco.2012.10.019.

Day, T. E. and Lewis, C. M. (1993) Forecasting futures market volatility, *Journal of Derivatives*, 1(2), pp. 33–50. doi: 10.3905/jod.1993.407876.

Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American statistical association*, 74(366a), pp. 427–431. doi: 10.1080/01621459.1979.10482531.

Doran, J. S. and Krieger, K. (2010) Implications for asset returns in the implied volatility skew, *Financial Analysts Journal*, 66(1), pp. 65–76.

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

Fama, E. F. and French, K. R. (1988) Business cycles and the behavior of metals prices, *Journal of Finance*, 43(5), pp. 1075–1093. doi: 10.1111/j.1540-6261.1988.tb03957.x.

Fattouh, B. (2010) An anatomy of the oil pricing regime, *Oxford Energy Forum*, 82, pp. 5–8.

Finta, M. A. and Ornelas, J. R. H. (2018) Commodity return predictability: Evidence from implied variance, skewness and their risk premia, *SSRN Electronic Journal*. doi: 10.2139/ssrn.3134310.

Geman, H. and Ohana, S. (2009) Forward curves, scarcity and price volatility in oil and natural gas markets, *Energy Economics*, 31(4), pp. 576–585. doi: 10.1016/j.eneco.2009.01.014.

Girma, P. B. and Mougoué, M. (2002) An empirical examination of the relation between futures spreads volatility, volume, and open interest, *Journal of Futures Markets*, 22(11), pp. 1083–1102. doi: 10.1002/fut.10047.

Gromb, D. and Vayanos, D. (2010) Limits of arbitrage, *Annual Review of Financial Economics*, 2(1), pp. 251–275. doi: 10.1146/annurev-financial-073009-104107.

Guo, B., Han, Q. and Zhao, B. (2014) The nelson–siegel model of the term structure of option implied volatility and volatility components, *Journal of Futures Markets*, 34(8), pp. 788–806.

Haugom, E., Langeland, H., Molnár, P. and Westgaard, S. (2014) Forecasting volatility of the u.s. oil market, *Journal of Banking and Finance*, 47, pp. 1–14. doi: 10.1016/j.jbankfin.2014.05.026.

Hull, J. (2012) *Options, futures, and other derivatives*, 8th global edn, Boston, Massachusetts: Pearson.

Hull, J., Nelken, I. and White, A. (2005) Merton's model, credit risk, and volatility skews, *Journal of Credit Risk*, 1(1), pp. 3–27. doi: 10.21314/JCR.2005.004.

Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities, *The journal of finance*, 42(2), pp. 281–300. doi: 10.1111/j.1540-6261.1987.tb02568.x.

Kaufmann, R. K. (2011) The role of market fundamentals and speculation in recent price changes for crude oil, *Energy Policy*, 39(1), pp. 105–115. doi: 10.1016/j.enpol.2010.09.018.

Kilian, L. (2009) Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review*, 99(3), pp. 1053–1069. doi: 10.1257/aer.99.3.1053.

Kilian, L. (2016) The impact of the shale oil revolution on u.s. oil and gasoline prices, *Review of Environmental Economics and Policy*, 10(2), pp. 185–205. doi: 10.1093/reep/rew001.

Kilian, L. and Park, C. (2009) The impact of oil price shocks on the u.s. stock market, *International Economic Review*, 50(4), pp. 1267–1287. doi: 10.1111/j.1468-2354.2009.00568.x.

Kleinberg, R. L., Paltsev, S., Ebinger, C. K. E., Hobbs, D. A. and Boersma, T. (2018) Tight oil development economics: benchmarks, breakeven points, and inelasticities, *Energy Economics*, 70. doi: 10.1016/j.eneco.2017.11.018.

Kogan, L., Livdan, D. and Yaron, A. (2009) Oil futures prices in a production economy with investment constraints, *The Journal of Finance*, 64(3), pp. 1345–1375. doi: 10.1111/j.1540-6261.2009.01466.x.

Liu, M.-L., Ji, Q. and Fan, Y. (2013) How does oil market uncertainty interact with other markets? an empirical analysis of implied volatility index, *Energy*, 55, pp. 860–868. doi: 10.1016/j.energy.2013.04.037.

MacKinnon, J. G. (1994) Approximate asymptotic distribution functions for unit-root and cointegration tests, *Journal of Business & Economic Statistics*, 12(2), pp. 167–176.

Maghyereh, A. I., Awartani, B. and Bouri, E. (2016) The directional volatility connectedness between crude oil and equity markets: New evidence from implied volatility indexes, *Energy Economics*, 57, pp. 78–93. doi: 10.1016/j.eneco.2016.04.010.

McDonald, R. L. (2014) *Derivatives markets*, 3rd new international edn, Harlow: Pearson Education.

Mitton, T. and Vorkink, K. (2007) Equilibrium underdiversification and the preference for skewness, *The Review of Financial Studies*, 20(4), pp. 1255–1288. doi: 10.1093/revfin/hhm011.

Mixon, S. (2002) Factors explaining movements in the implied volatility surface, *Journal of Futures Markets*, 22(10), pp. 915–937. doi: 10.1002/fut.10039.

Mixon, S. (2011) What does implied volatility skew measure?, *Journal of Derivatives*, 18(4), pp. 9–25. doi: 10.3905/jod.2011.18.4.009.

Mork, K. A., Olsen, Ø. and Mysen, H. T. (1994) Macroeconomic responses to oil price increases and decreases in seven oecd countries, *The Energy Journal*, 15(4), pp. 19–35.

Navatte, P. and Villa, C. (2000) The information content of implied volatility, skewness and kurtosis: empirical evidence from long-term cac 40 options, *European Financial Management*, 6(1), pp. 41–56. doi: 10.1111/1468-036X.00110.

Newey, W. K. and West, K. D. (1987) Hypothesis testing with efficient method of moments estimation, *International Economic Review*, 28(3), pp. 777–787.

Ohnsorge, F., Stocker, M. and Some, M. Y. (2016) Quantifying uncertainties in global growth forecasts, *World Bank Policy Research Working Paper No. 7770*. doi: 10.1596/1813-9450-7770.

OVX (2018) Available at: http://www.cboe.com/products/vix-index-volatility/volatility-on-etfs/cboe-crude-oil-etf-volatility-index-ovx (accessed 01.03.2018).

Peston, R. (2011) US loses AAA credit rating after S&P downgrade, *BBC*. Available at: http://www.bbc.com/news/world-us-canada-14428930 (accessed: 10.04.2018).

Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in time series regression, *Biometrika*, 75(2), pp. 335–346. doi: 10.1093/biomet/75.2.335.

Pirrong, C. (2010) An evaluation of the performance of oil price benchmarks during the financial crisis, *Bauer College of Business Working Paper, University of Houston*, pp. 1–76. Available at: https://www.bauer.uh.edu/downloads/Pirrong_WTI_Report_091116_Final.pdf (accessed: 11.05.2018).

Ramsey, J. B. (1969) Tests for specification errors in classical linear least-squares regression analysis, *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), pp. 350–371.

Ripple, R. D. and Moosa, I. A. (2009) The effect of maturity, trading volume, and open interest on crude oil futures price range-based volatility, *Global Finance Journal*, 20(3), pp. 209–219. doi: 10.1016/j.gfj.2009.06.001.

Robe, M. A. and Wallen, J. (2016) Fundamentals, derivatives market information and oil price volatility, *The Journal of Futures Markets*, 36(4), pp. 317–344. doi: 10.1002/fut.21732.

Russia oil row hits Europe supply (2007) Available at: http://news.bbc.co.uk/2/hi/business/6240473.stm (accessed: 10.04.2018).

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

STATA manual (2018) Available at: https://www.stata.com/manuals13/tspperron.pdf (accessed: 10.04.2018.

Studenmund, A. H. (2016) *Using econometrics : a practical guide*, 7th edn, Harlow: Pearson Education.

Szakmary, A., Ors, E., Kim, J. K. and Davidson III, W. N. (2003) The predictive power of implied volatility: Evidence from 35 futures markets, *Journal of Banking & Finance*, 27(11), pp. 2151–2175. doi: 10.1016/S0378-4266(02)00323-0.

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

VIX (2018) Available at: http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index (accessed: 01.03.2018).

Whaley, R. E. (1986) Valuation of american futures options: Theory and empirical tests, *The Journal of Finance*, 41(1), pp. 127–150. doi: 10.1111/j.1540-6261.1986.tb04495.x.

Working, H. (1960) Speculation on hedging markets, *Food Research Instistute Studies*, 01(02), pp. 185–220.

Würsig, C. M. (2017) Analysis of Implied Volatility for Oil Futures Options, Master's thesis, Leibniz Universität Hannover.

Xing, Y., Zhang, X. and Zhao, R. (2010) What does the individual option volatility smirk tell us about future equity returns?, *Journal of Financial and Quantitative Analysis*, 45(3), pp. 641–662. doi: 10.1017/S0022109010000220.

Xiong, W. (2013) The financialization of commodity markets, *NBER Reporter, NBER, Cambridge, Mass*, 2. doi: 10.3386/w19642.

A Data

A.1 Databases

Our regression is based on data from 25.07.2006 to 03.03.2016. Data was made available from supervisor, except for continuous WTI crude oil futures, 1-month US treasury bill rate, annual LIBOR interest rate and the OVX-index. Table A.1 shows where the data is collected from:

Variables	Database
WTI crude oil future options	CME
WTI crude oil futures	CME
VIX	CBOE
OVX	CBOE
OPEC spare capacity	EIA
Oil Output	EIA
ADS	Quandl
Working T	Quandl
Continuous WTI crude oil futures	Datastream
1-month US treasury bill rate	Datastream
Annual LIBOR interest rate	Datastream

Table A.1: This table shows where all data is collected from. CME is Chicago Mercantile Exchange, CBOE is Chicago Board Options Exchange, and EIA is U.S. Energy Information Administration.

A.2 Practical Approach Calculating Volatility Smiles

The most time-consuming part of this thesis is the calculation of volatility smiles. We use MATLAB to do the calculation, because it could handle large data sets and have built in functions for finance. To calculate the volatility smiles we start out with csv-files with future options and futures for WTI crude oil. After importing the data we organize it so each future option is matched with the underlying future. Then we import interest rates and match them according to the trade dates. Having imported and organized all the data, we calculate implied volatility for each future option. After calculating implied volatilities we get one volatility smile for each time to maturity at each day. To narrow the scope of the thesis we decided only to use nearby contracts. Having the volatility smiles we retrieve at-the-money implied volatility and implied skewness.

Table A.2 show an overview of how many options we used to calculate daily volatility smiles. Total per year is the sum used each year, and daily average is the average number of options each volatility smile contain. Our sample is from 25.07.2006 to 03.03.2016, and that is the reason why 2006 and 2016 contain fewer options on total than the other years. The daily average is still similar to the other years.

Year	Total per year	Daily average
2006	12316	108.99
2007	29388	117.08
2008	44283	175.03
2009	36149	143.45
2010	37387	148.36
2011	41063	162.95
2012	44417	176.26
2013	40693	161.48
2014	35615	141.33
2015	36676	145.54
2016	5953	148.83

Table A.2: Number of options used per year. Daily average is calculated as the average number of options included in the volatility smile each day.

A.3 Correlation Matrix

Table A.3 contains the correlation matrix for our data.

	Implied Volatility	Implied Skewness	TTM	Dummy Spare	VIX	Con - tango	Back - wardation	Implied $Volatility_{t-1}$	Implied Skewness $_{t-1}$	ADS	Oil Output	Dummy Spare * Oil Output	Working T	Futures Volume
Implied Volatility	1.000													
Implied Skewness	0.084	1.000												
Time to maturity	0.268	-0.124	1.000											
Dummy Spare	0.195	0.096	-0.003	1.000										
VIX	0.756	0.044	0.008	0.258	1.000									
Contango	0.744	0.001	0.046	0.271	0.433	1.000								
Backwardation	0.381	-0.184	0.016	0.352	0.179	0.397	1.000							
Implied $Volatility_{t-1}$	0.889	0.094	-0.033	0.214	0.715	0.724	0.384	1.000						
Implied $Skewness_{t-1}$	0.126	0.788	0.021	0.097	0.058	0.016	-0.165	0.087	1.000					
ADS	- 0.007	-0.075	0.023	0.153	0.028	0.029	0.009	-0.014	-0.063	1.000				
Oil Output	0.045	-0.013	0.047	-0.005	0.064	0.017	-0.007	0.034	-0.018	0.158	1.000			
Dummy Spare * Oil Output	0.067	0.047	0.051	0.055	0.030	0.073	0.020	0.056	0.039	0.010	0.352	1.000		
Working T	0.006	0.054	0.011	-0.035	0.034	-0.041	0.006	0.005	0.050	0.032	-0.025	0.015	1.000	
Futures Volume	-0.027	0.016	-0.156	0.010	0.003	0.015	-0.010	-0.076	0.018	0.005	0.017	0.024	-0.031	1.000

Table A.3: Correlation Matrix for all the variables used in our regression models.

A.4 Parsimonious Regression Results

Table A.4 shows the regression results of more parsimonious regression models including only variables that contributes to determine implied volatility and implied skewness.

Variables	Implied Volatility	Implied Skewness
Lagged dependent	0.53552	0.77895
	(0.02337)	(0.02794)
VIX	0.29216	
(One-day lagged)	(0.01809)	
ADS		
(One-day lagged change)		
Oil Output		
(One-week lagged pct change)		
Dummy SPARE	-0.01724	0.00659
	(0.00306)	(0.00438)
Dummy Spare * Oil Output		
(One-week lagged pct change)		
Backwardation	0.16498	-0.11791
(One-day lagged slope)	(0.03667)	(0.05057)
Contango	0.26492	
(One-day lagged slope)	(0.02204)	
Working T		
(One-week lagged change)		
Time to maturity	0.00329	-0.00103
	(0.00014)	(0.00020)
Futures Volume	0.01131	
(One-day lagged change)	(0.00249)	
Constant	-0.14833	0.03327
	(0.00899)	(0.01200)
Observations	490	490
Adjusted R-squared	0.9338	0.6425

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Weekly regression results from 25.07.2006 to 01.03.2016 for implied volatility and implied skewness with only the significant variables from the oiginal regressions. ***, ** and * indicate the significance level at the 0.01, 0.05 and 0.1, respectively. Regression results are reported with standard error in parentheses. For information about where data is collected from, ses Appendix A.1.

B Statistical Tests

In this part of the appendix you will find information and theory about the statistical tests used in this thesis.

B.1 Augmented Dickey-Fuller Test

We use an augmented Dickey-Fuller test to test if all the time-series are stationary. A time series is stationary if the process is mean reverting, hence the series have a constant mean, variance and autovariance. The Dicky-Fuller test if the time-series has a unit root or not (Alexander, 2008*b*). The null hypothesis states that the process is non-stationary and alternative hypothesis states that the time-series is stationary. The Dicky Fuller test run the following regression:

$$\Delta y_t = \psi y_{t-1} + \sum_{n=1}^{p} \alpha_n y_{t-1} + u_t \tag{B.1}$$

ψ=1-ϕ, where ϕ is the unit root. p is the number of lags of the dependent variable included in the regression. And u_t is the residuals. The test statistic is:

$$\text{Test statistic} = \frac{\psi}{SE(\psi)} \tag{B.2}$$

$SE(\psi)$ is the standard error. We reject the null hypothesis if the test statistic is more negative than the critical value. Here it is important to be aware that we can not use the t-distribution, but have to use the Dickey and Fuller (1979) critical values.

B.2 Phillips-Perron Test

The Phillips-Perron test is a unit-root test in the same way as the Dickey-Fuller test explained in appendix B.1. The difference is that Phillips-Perron takes serial correlation into account. The test starts with the same regression as the Dickey-Fuller test, but where the Dickey-Fuller uses lagged first difference of the variable, the Phillips-Perron test uses Newey and West (1987) robust standard errors. The test involves fitting the following regression (Phillips and Perron, 1988):

$$y_i = \alpha + p y_{i-1} + \epsilon_i \tag{B.3}$$

After fitting the regression we calculate the two test statistics Z_ρ and Z_τ (*STATA manual*, 2018):

$$Z_\rho = n(\hat{\rho}_n - 1) - \frac{1}{2}\frac{n^2\hat{\sigma}^2}{S_n^2}(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \tag{B.4}$$

$$Z_\tau = \sqrt{\frac{\hat{\gamma}_{0,n}}{\hat{\lambda}_n^2}}\frac{\hat{\rho}_n - 1}{\hat{\sigma}} - \frac{1}{2}(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n})\frac{1}{\hat{\lambda}_n}\frac{n\hat{\sigma}}{S_n} \tag{B.5}$$

$$\hat{\gamma}_{j,n} = \frac{1}{n}\sum_{i=j+1}^{n}\hat{u}_i\hat{u}_{i-j} \tag{B.6}$$

$$\hat{\lambda}_n^2 = \hat{\gamma}_{0,n} + 2\sum_{j=1}^{q}(1 - \frac{j}{q+1})\hat{\gamma}_{j,n} \tag{B.7}$$

$$S_n^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{u}_i^2 \tag{B.8}$$

Where u_i is OLS residuals, k is the number of covariates in the regression, q is number of Newey-West lags included when calculating $\hat{\lambda}_n^2$, and $\hat{\sigma}$ is the standard error of $\hat{p}$ (*STATA manual*, 2018). To determine the number of Newey-West lags(q) we use the default set in *STATA manual* (2018):

$$4(T/100)^{2/9} \tag{B.9}$$

The test statistics is compared with the same critical values as the Dickey-Fuller test statistic, since they follow the same distribution.

B.3 VIF

VIF-indexes are calculated to detect multicollinearity, and is calculated in the following way according to Studenmund (2016). First you have to do a regression that has X_i as dependent variable and all the other explanatory variables as independent variables:

$$X_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_K X_K + v \tag{B.10}$$

When you have done this you find the VIX-index by using the following formula:

$$VIF(\hat{\beta}_i) = \frac{1}{1 - R_i^2} \tag{B.11}$$

The advantage of using VIF-indexes compared to bivariate correlation is that VIF-indexes takes all independent variables into account when checking for multicollinearity, while correlations only check pairwise.

B.4 Jarque-Bera Test

The Jarque-Bera test is a test for normality, and test wether a variable follow the normal distribution (Bera and Jarque, 1981). The test uses expected value and variance, denoted by μ and σ^2, and calculates the skewness and kurtosis, denoted by b_1 and b_2:

$$b_1 = \frac{E[\mu^3]}{(\sigma^2)^{\frac{3}{2}}} \tag{B.12}$$

$$b_2 = \frac{E[\mu^4]}{(\sigma^2)^2} \tag{B.13}$$

The Jarque-Bera test statistic is calculated as follow, where T is the sample size, and follow a chi-squared distribution with two degrees of freedom:

$$JB = \frac{T}{6}[b_1 + \frac{(b_2 - 3)^2}{4}] \sim \chi^2(2) \tag{B.14}$$

The null hypothesis is that the data follow a normal distribution. If the test statistic is greater than the critical value given by the chi-squared distribution, we reject the null hypothesis and conclude that the data does not follow the normal distribution.

B.5 Breusch-Pagan Test

The Breusch-Pagan test is a test for heteroskedasticity in a regression model. Heteroskedasticity means that the residuals does not have constant variance. The test is based on the residuals of the regression and is modelled in the following way (Studenmund, 2016):

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_K X_{Ki} + u_i \tag{B.15}$$

Having modelled the test, you test the overall significance with a chi-square test, testing if α_1 to α_K is equal to 0 or not. If all the alphas except α_0 is 0, then the variance is α_0, which is a constant. If you reject the null hypothesis, you have a problem with heteroskedasticity in your regression model.

B.6 Breusch-Godfrey Test

Breuch-Godfrey test is a test for autocorrelation in the regression model. This means that the residuals are correlated over time. The test is based on the residuals from the initial regression, and is modelled in the following way, with a new regression on lagged residuals:

$$u_t = \rho_0 + \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + v_t \tag{B.16}$$

The null hypothesis is that ρ_1 to ρ_p is equal to 0. If this is rejected we have a case of autocorrelation. Even with autocorrelation the OLS results will remain unbiased.

B.7 Ramsey RESET Test

The Ramsey regression equation specification error test (RESET) test examines if non-linear combinations of the fitted values explain the dependent variable (Ramsey, 1969). If a non-linear combination of the explanatory variables have any power on the dependent variable, the model might not be linear. The test is modelled in the following way, as an auxiliary regression with higher order terms of the predicted values together with the explanatory variables:

$$y_t = \alpha_1 + \alpha_2 \hat{y}_t^2 + \cdots + \alpha_p \hat{y}_p^2 + \sum \beta_i x_{it} + v_t \tag{B.17}$$

The test statistic is calculated as TR^2, and is chi-squared distributed with p-1 degrees of freedom:

$$TR^2 \sim \chi^2(p-1) \tag{B.18}$$

The null hypothesis is rejected if the test statistic is greater than the corresponding critical value in the χ^2 distribution. If the null hypothesis is rejected, the non-linear combinations of the explanatory variables has predictive power, and we may have a misspecified regression model.

C Additional Figures

C.1 At-The-Money Implied Volatility

Figure C.1 shows that our measure of at-the-money implied volatility defined in Section 3.7 moves close to 50-delta implied volatility. The high correlation between the two measures of at-the-money implied volatility supports that the measure we use in the regression model is a good measure.

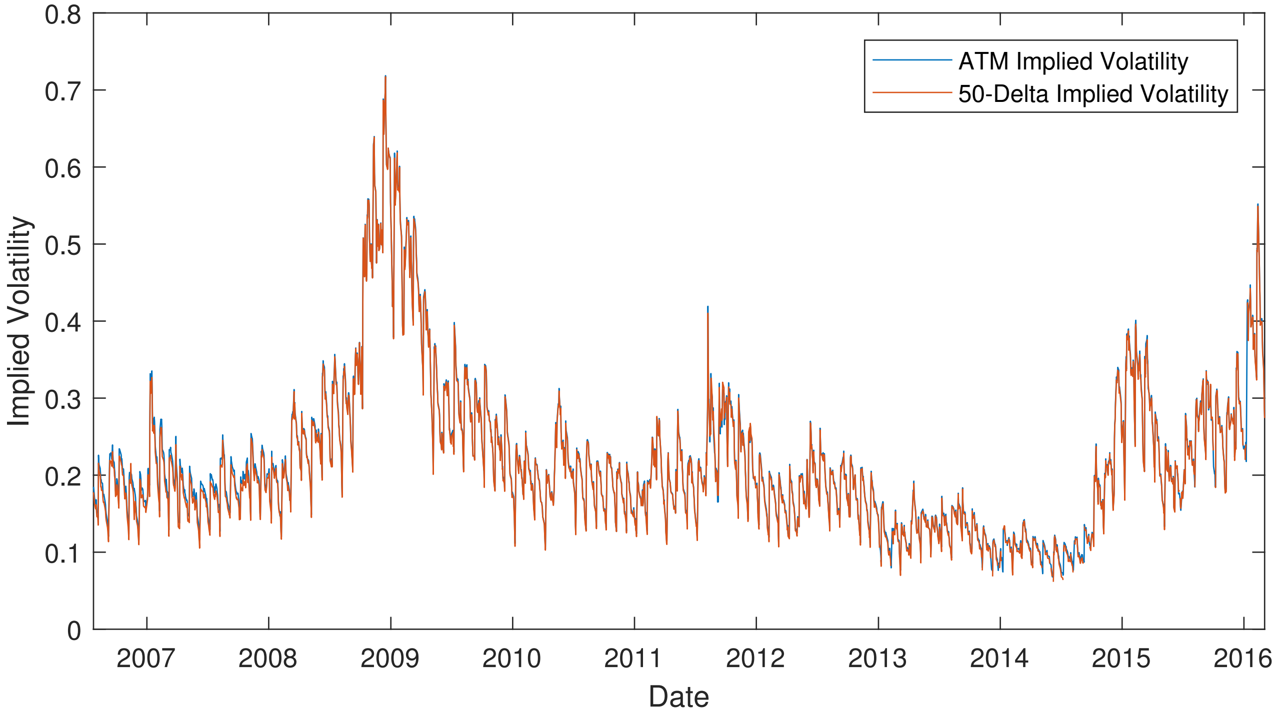

Figure C.1: Daily implied volatilities from 25.07.2006 to 03.03.2016. The blue line shows the at-the-money implied volatility calculated as the arithmetic mean of the implied volatilities with strikes in a 10% span around the future price. The red line shows the 50-delta implied volatility used in the proxy for implied skewness

C.2 Volatility Smiles

This section contains four volatility smiles in figure C.2, where the trade dates are one year apart, illustrating how implied volatility and implied skewness change over time.

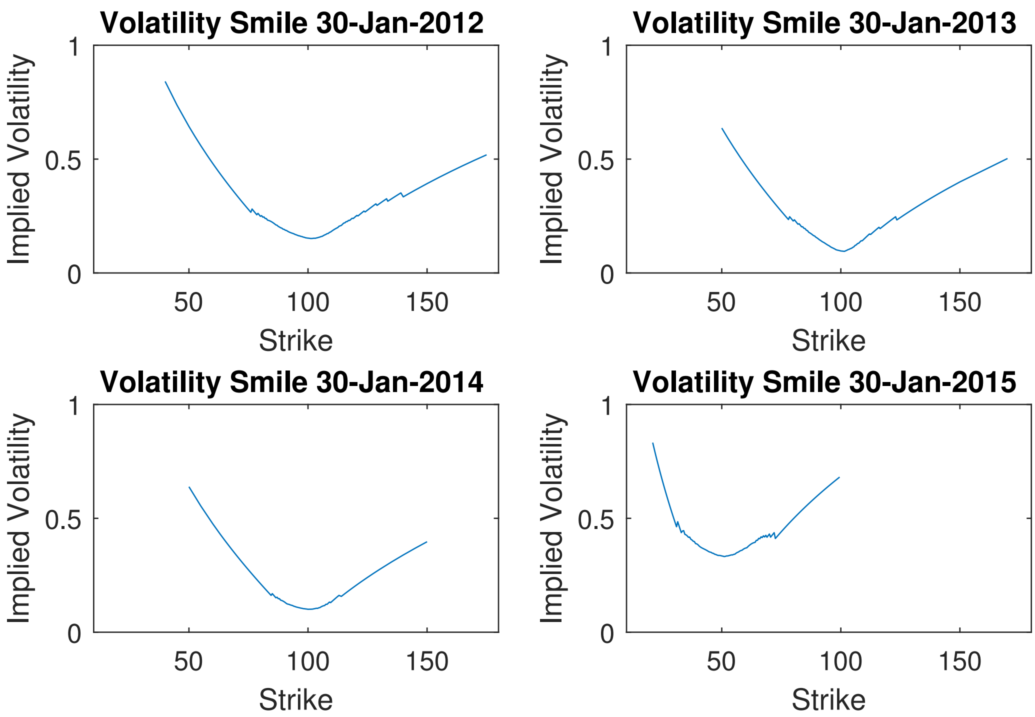

Figure C.2: Implied volatility smiles at the 30th of January for 2012, 2013, 2014 and 2015.

C.3 Demand and Supply

Figure C.3 contains a plot illustrating the demand and supply in the crude oil market, illustrating how a change in demand can give big price-changes.

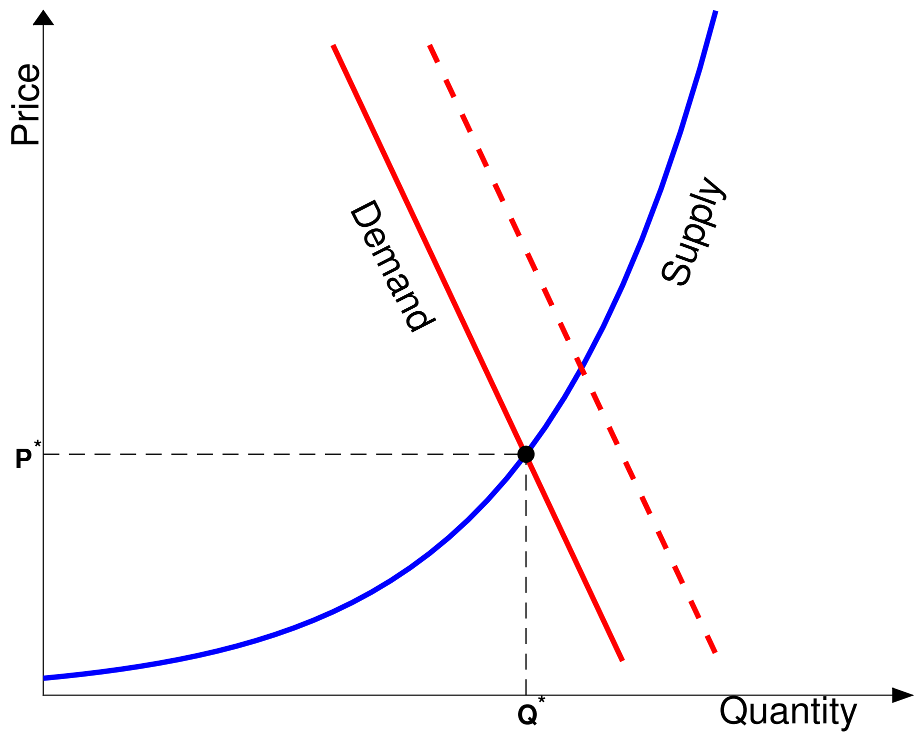

Figure C.3: The figure shows the typical demand and supply curves in the crude oil market in the short run, and a positive demand-shock for crude oil, giving an increased price. The reason why the supply-curve is so steep, is that the supply depends on pipelines and tanker capacity at least as much as to potential production capacity (Barsky and Kilian, 2004).

C.4 Residuals

This section contains plots with the residuals of the regressions. Figure C.4 is the residual plot from the regression on implied volatility, figure C.5 is the residual plot from the regression on implied skewness and figure C.6 is the residuals from the regression on implied skewness put on a timeline.

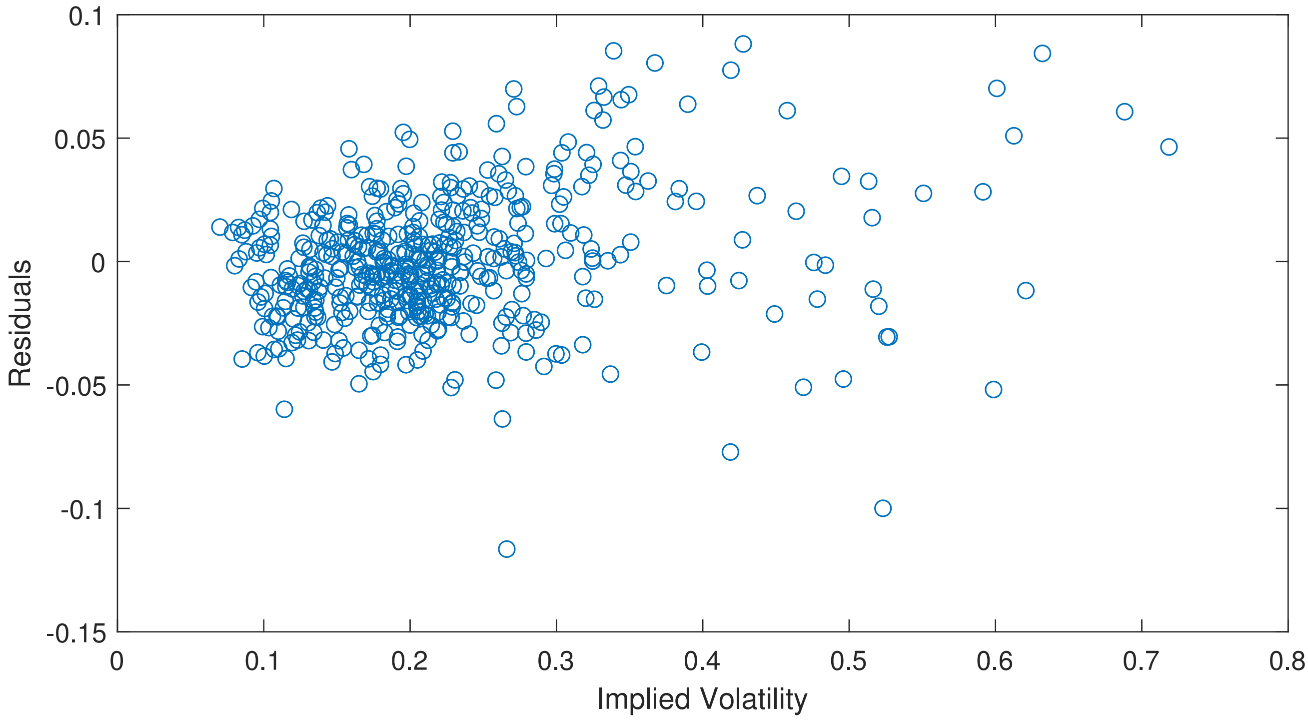

Figure C.4: Residual plot from the regression model on implied volatility.

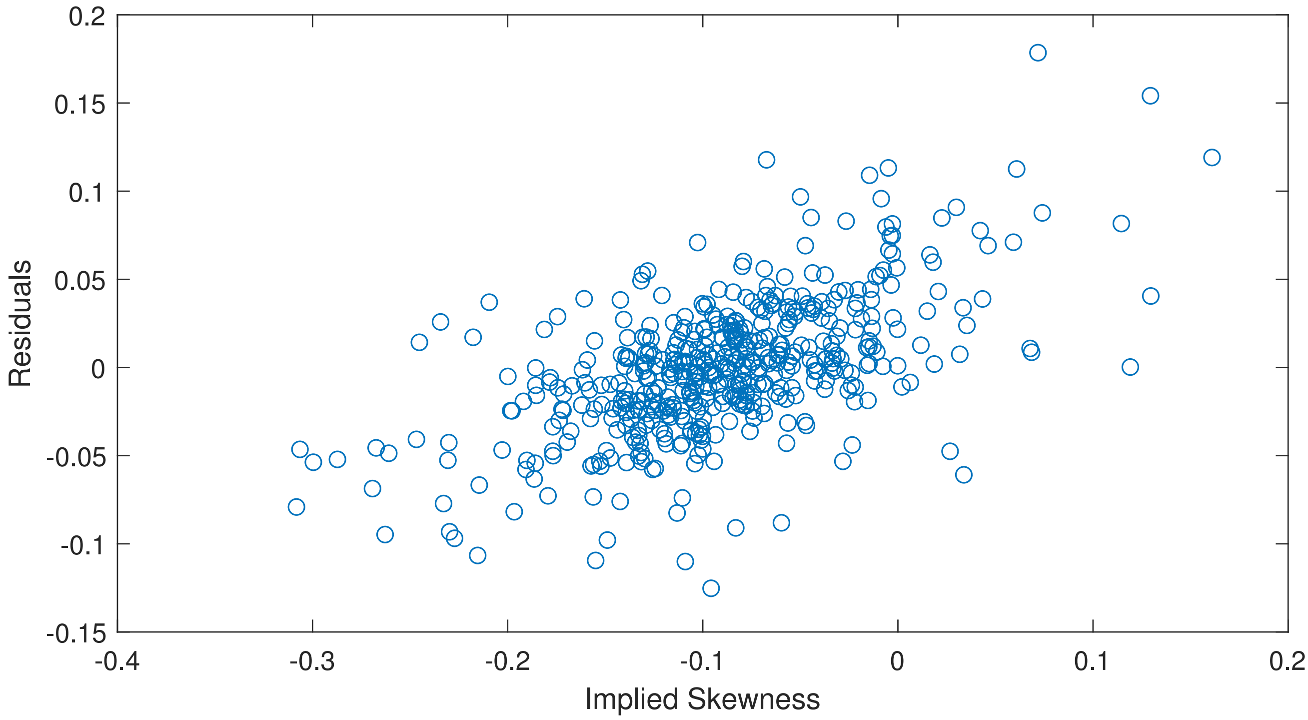

Figure C.5: Residual plot from the regression model on implied skewness.

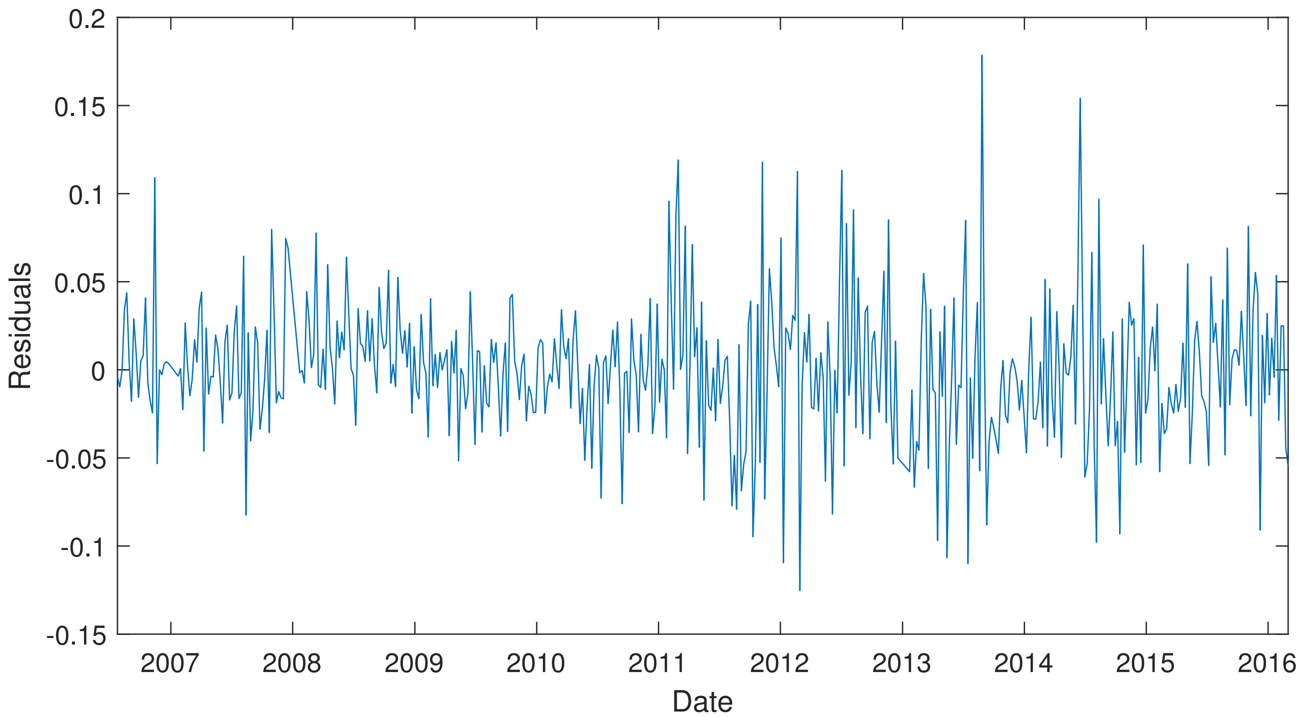

Figure C.6: Timeline of the residuals from the regression on implied skewness from 25.07.2006 to 01.03.2016.

C.5 Open Interest Crude Oil Futures

Figure C.7 shows how futures on crude oil has become more actively traded the last decades. Open interest is the total number of outstanding futures existing on a given day, delivered on a particular day. From 2000 to 2018 the open interest has increased fivefold.

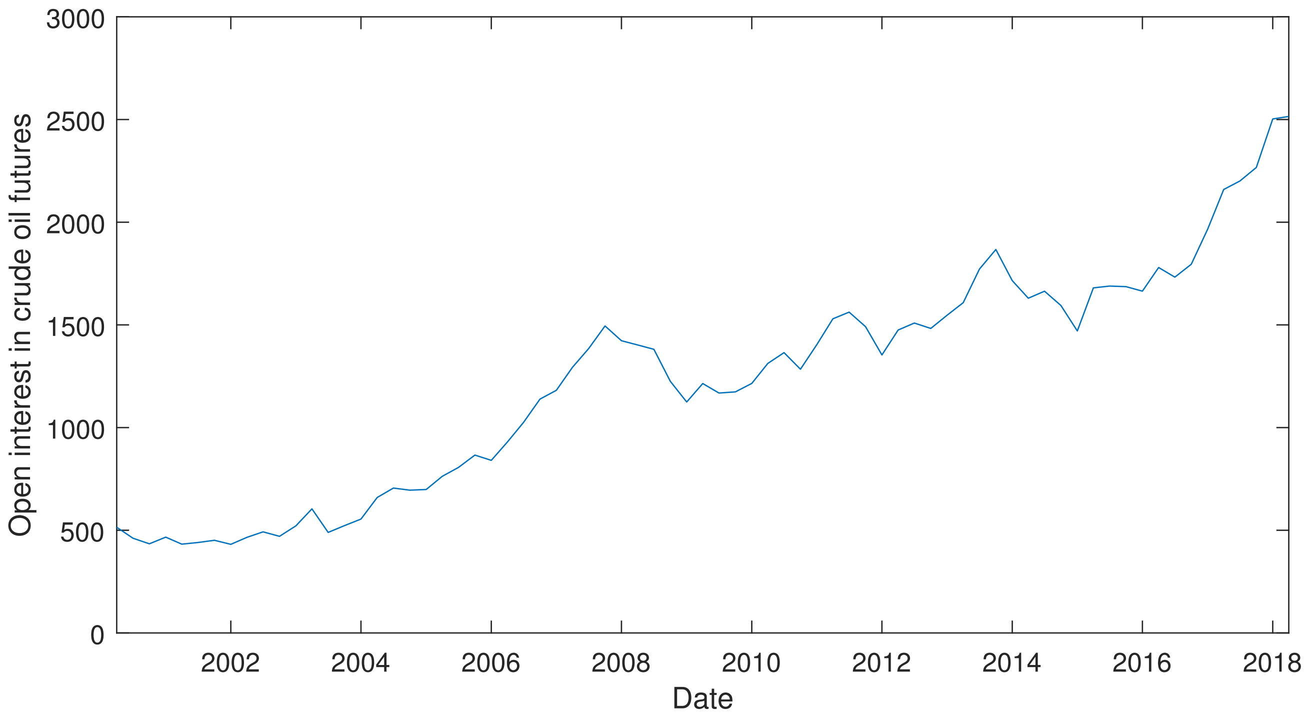

Figure C.7: Average open interest, in thousands, in crude oil futures on U.S exchanges from 1st quarter 2000 to 1st quarter 2018. The figure shows that open interest in WTI crude oil futures have increased steadily in the last 18 years. Source: EIA.