Option prices and implied volatility in the crude oil market

Soini, Vesa-Heikki; Lorentzen, Sindre

Published in:
Energy Economics

DOI:
10.1016/j.eneco.2019.07.011

Publication date:
2019

Document Version
Peer reviewed version, als known as post-print

Citation for published version (APA):
Soini, V.-H., & Lorentzen, S. (2019). Option prices and implied volatility in the crude oil market. *Energy Economics*, *83*, 515-539. https://doi.org/10.1016/j.eneco.2019.07.011

Option Prices and Implied Volatility in the Crude Oil Market

Vesa Soini[a,1,*], Sindre Lorentzen[b,2]

[a]*Finland - 00330 Helsinki*
[b] *Norway - 4036 Stavanger*

Abstract

This paper studies the determinants of WTI crude oil call option prices with a special emphasis on the relationship between implied volatility and moneyness. Our first-stage regression estimates a quadratic approximation of implied volatility as a function of moneyness, while our second-stage regression investigates correlations between the estimated parameters and a list of explanatory variables. The first-stage regressions show a positive coefficient on the quadratic term, suggesting that the market exhibits 'Implied Volatility Smile' and hence violates the Black-Scholes predictions. The main results of our paper concern the determinants of these violations. We find that the curvature of implied volatility as a function of moneyness is: (i) positively and significantly correlated with basis and hedging pressure of the underlying crude oil futures contract (ii) positively and significantly correlated with various measures of transaction costs on the options market. We explore various explanations for these results. The paper also contains a variety of robustness checks, mostly related to the assumed functional forms.

Keywords: Implied volatility, options, crude oil, hedging pressure.
JEL classification: G1, Q4

*Corresponding Author
[1]Hanken School of Economics, vesa.soini@gmail.com
[2]University of Stavanger, sindre.lorentzen@uis.no

1. Introduction

Oil is the most important commodity in the world. The effect of oil prices spills over to other industries in the economy, since oil is used as fuel for transportation and as an input to plastic production. In the modern era, an increasingly large amount of oil-related trading happens in the derivatives market. Despite the economic importance of trading in crude oil options, there is very little research on their pricing. This paper aims to fill that gap.

*Corresponding Author
[1] Hanken School of Economics, vesa.soini@gmail.com
[2] University of Stavanger, sindre.lorentzen@uis.no

Option Prices and Implied Volatility in the Crude Oil Market

Vesa Soini[a,1,*], Sindre Lorentzen[b,2]

[a]Finland - 00330 Helsinki
[b]Norway - 4036 Stavanger

Abstract

This paper studies the determinants of WTI crude oil call option prices with a special emphasis on the relationship between implied volatility and moneyness. Our first-stage regression estimates a quadratic approximation of implied volatility as a function of moneyness, while our second-stage regression investigates correlations between the estimated parameters and a list of explanatory variables. The first-stage regressions show a positive coefficient on the quadratic term, suggesting that the market exhibits 'Implied Volatility Smile' and hence violates the Black-Scholes predictions. The main results of our paper concern the determinants of these violations. We find that the curvature of implied volatility as a function of moneyness is: (i) positively and significantly correlated with basis and hedging pressure of the underlying crude oil futures contract (ii) positively and significantly correlated with various measures of transaction costs on the options market. We explore various explanations for these results. The paper also contains a variety of robustness checks, mostly related to the assumed functional forms.

Keywords: Implied volatility, options, crude oil, hedging pressure.
JEL classification: G1, Q4

1. Introduction

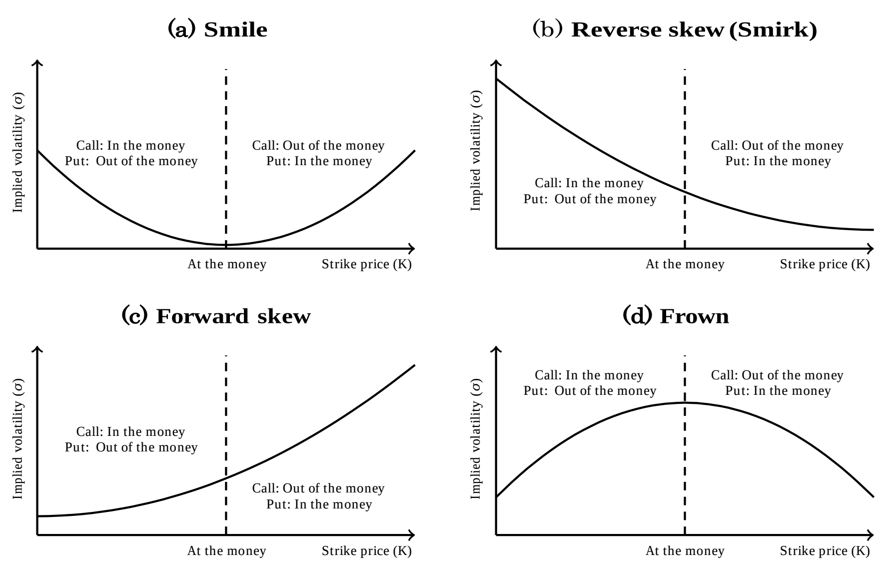
Oil is the most important commodity in the world. The effect of oil prices spills over to other industries in the economy, since oil is used as fuel for transportation and as an input to plastic production. In the modern era, an increasingly large amount of oil-related trading happens in the derivatives market. Despite the economic importance of trading in crude oil options, there is very little research on their pricing. This paper aims to fill that gap.

*Corresponding Author
[1]Hanken School of Economics, vesa.soini@gmail.com
[2]University of Stavanger, sindre.lorentzen@uis.no

Based on the Black-Scholes model and its assumptions, all options on the same underlying asset and with the same time to maturity should have the same implied volatility regardless of the strike price. However, in practice implied volatility is consistently found to change with the degree of moneyness, often in the form of 'Implied Volatility Smiles'. Since implied volatility is directly linked to pricing of options, deviations from constant implied volatility can tell us a lot about the determinants of options prices. Previous research has devoted very little attention to identifying the underlying drivers of the relationship between implied volatility and moneyness. Apart from the study of [1] on Spanish stock index options, the determinants of volatility smiles have been left unexplored in the literature.

While the presence of volatility smiles is a well-established stylized fact for some options, to our knowledge there is no prior research on volatility smiles in the crude oil market. Moreover, the relationship between implied volatility and moneyness differs significantly across various markets. In the U.S. equity market, implied volatility was relatively constant across various levels of moneyness until the Crash of 1987. Afterwards, however, implied volatility started to show a downward sloping pattern. Regarding other markets, currency and commodity markets typically exhibit a proper volatility smile, as shown in Panel (a) of Figure 1.

Figure 1: **Different type of relationships between implied volatility and moneyness**



The aim of this paper is to empirically investigate the relationship between implied volatility and moneyness to study the determinants of WTI crude oil options prices. We tackle this task by running a a two-stage regression analysis. In the first stage, we estimate a second-order Taylor approximation of implied volatility as a function of moneyness separately for each date and maturity class in the sample. In the second stage, we analyze correlations between the estimated parameters and several variables that we believe to be important for options prices.

As previous research has not studied volatility smiles for crude oil, we first test if volatility smiles are present in this market. While the use of quadratic approximation of the functional form may seem restrictive, we show that this specification fits the data surprisingly well. Estimating the model for each date and maturity class gives an average R^2 of 0.957. The average estimated parameters are 0.93 for the constant term, -1.44 for the linear term and 0.78 for the quadratic term. The fact that the last term is positive suggest that the crude oil options market tends to exhibit a volatility smile.

The main empirical result of the paper is that the curvature of implied volatility as a function of moneyness is positively and significantly correlated with the basis and hedging pressure of the underlying crude oil futures contract. That is, the implied volatility curves tend to be flatter when either the basis is low or commercial hedgers are net long. This result is new in the literature and not present in for example [1]. The result also suggests that the return distribution of the underlying has a key role in explaining volatility smiles, which contrasts with the net buying pressure explanation of [2] and the transaction cost explanation discussed in [3], [4] and [1].

We believe that the theoretical reason behind our main result is that crude oil futures return distribution tends to exhibit fatter tails and more skew during times of high basis and high hedging pressure. We demonstrate empirically that a high level of either of these variables is associated with a higher kurtosis in the historical return distribution. Therefore, our analysis suggests that the shape of implied volatility function is inherently affected by the return distribution of the underlying. Fatter tails during times of high basis can be attributed to the theory of storage [5, 6, 7]. According to this theory, a higher basis is associated with lower inventories and a greater likelihood of price spikes. This mechanism increases both skewness and kurtosis, and can therefore explain the volatility smile. [8] report empirical support for this mechanism and relate it to the convenience yield of the underlying futures contract.

Another important finding of this paper is that the curvature of implied volatility as a function of moneyness is positively correlated with the low-frequency [9] high-low spread estimator and negatively correlated with the traded volume of options. We regard these two variables as proxies for transaction costs. On the contrary, traded volume of the underlying futures contracts, Monday dummy or days to maturity do not have much explanatory power for the volatility smile.

The paper also includes various types of robustness checks. While the use of second-order Taylor approximation provides a good fit for the model, we realize that other functional forms may be even better. The difficulty is to find a good parameterization for the model. This would require a functional form which is parsimonious enough for estimation and would provide a good fit. We have experimented with a Multivariate Fractional Polynomials analysis which is an algorithm to find the best model fit (see [10] and [11]). As a result, we concluded that there is no other model specification which would provide a superior fit compared to our quadratic approximation for a majority of dates and maturity classes. As a consequence, we decided to carry on with the quadratic approximation which provides a good fit and has the additional benefit of easier interpretation of the estimated correlations.

2. Literature review

As pointed out, the Black-Scholes model implies that all option prices on the same underlying asset and with the same time to maturity should have the same implied volatility regardless of the strike price. Previous empirical research on various markets shows that this prediction is incorrect.

Researchers have put forth many potential explanations for the empirical violation of the Black-Scholes prediction. A popular approach is to relax the assumption of constant volatility and replace it with a volatility rate that changes either deterministically or stochastically. One strand of research uses an implied binomial tree framework [12, 13, 14, 15, 16] that can obtain a perfect fit with observed options prices. Nevertheless, [4] empirically test this framework and show that the inferred parameters are highly unstable over time.

In an early paper, [17] proposed a constant elasticity of variance model (an example of a "local variance model") that provides an additional degree of freedom and hence improves the model fit. However, [18] conclude that the constant elasticity of variance model does not improve over the Black-Scholes model out-of-the-sample.

Stochastic volatility models come in all shapes and sizes. The key point with this class of models is that if correlation between innovations to volatility and returns from the underlying asset is negative, the implied volatility function becomes downward sloping (a volatility skew). A seminal paper regarding stochastic volatility models is [19] and subsequently [20] and [21] have presented their versions. There is also a literature developing jump diffusion models [22, 23, 24, 25, 26, 27]. However, [24] and [28] show empirically that neither stochastic volatility models nor random jump models are alone sufficient to explain the empirical violations of the Black-Scholes model.

The explanations discussed thus far all aim to modify the return distribution of the underlying asset, but there are other explanations as well. Recently, [2] suggested that net buying pressure affects options prices and causes the observed patterns. That is, under the Black-Scholes assumptions, the supply curve for each option series should be a horizontal line, but in practice, there are plausibly limits to arbitrage (see [29], [30], [31]). Such limits to arbitrage make the supply curve for options upward sloping and additional demand will increase options price together with the implied volatility. [2] show empirical evidence for this channel for S&P 500 index as well as individual stock options.

A further explanation for the volatility smile is the existence of transaction costs. For example, papers by [3] and [4] imply that transaction costs and low liquidity could in fact be plausible causes for volatility smiles. Moreover, [1] find empirically that transaction costs, proxied by bid-ask spread, are important determinants for the volatility smile. However, they use the average bid-ask spread over the day in their analysis and do not control for the fact that transaction costs may depend on moneyness.

Related research in crude oil options is somewhat limited. [32] develop a method for estimating the implied probability density function for futures prices from American options prices and apply their method to crude oil options. [33] study changes in crude oil volatility after the introduction of NYMEX crude oil futures and the subsequent introduction of crude oil options and derivatives on other energy commodities. [34] use time-series econometrics to analyze the efficiency of WTI and Brent crude oil markets. [35] study whether crude oil futures prices are useful in forecasting spot price and document some stylized facts about the relationship between basis and futures prices. Forecasting of crude oil volatility is studied for example in [36] and [37]. [38] assesses factors that potentially influence the volatility of crude oil prices.

3. Data

The main data consists of settlement prices of WTI (West Texas Intermediate) crude oil call options traded at the New York Merchantile Exchange (NYMEX) between May 13, 2008 and May 31, 2016. We restrict attention to contract maturities ranging from 1-month to 12-months, as trading in longer maturity contracts is more rare which affects price formation. The underlying crude oil futures contracts are for delivery at Cushing, OK. The dataset is purchased from Commodity Research Bureau[3] (CRB). Crude oil options have maturities ranging from one month to 12 months. In practice, the contracts correspond roughly to how many months are left until maturity, plus around 20 days set by the regulations of NYMEX.

The Black-Scholes implied volatilities are computed separately for each observation. For this purpose we use prices of the underlying crude oil futures from the CRB. Moreover, computing the implied volatilities requires a proxy for the risk-free interest rate. For this purpose we use the market yield on U.S. Treasury securities at 1-year constant maturity which is quoted on investment basis[4].

Our raw dataset consists of 1 208 866 observations with positive open interest. However, we only analyze options with a positive trading volume on a given day, as prices may otherwise not be informative of the current market situation. Furthermore, there are some observations for which the empirical implied volatility equation is not satisfied for any positive value of implied volatility. We discard such observations. Overall, we are left with 305 212 observations with a positive trading volume and a well-defined measure of implied volatility.

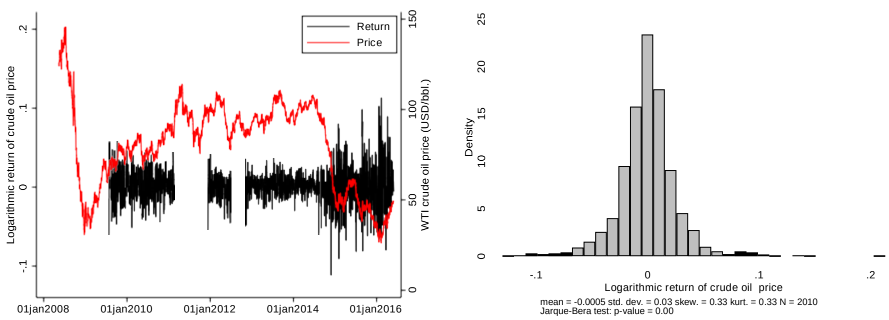
[3]See http://www.crbtrader.com/ for details on how to order the data.
[4]These data are available at the federal Reserve website (https://www.federalreserve.gov/releases/h15/data.htm).

3.1. *Descriptive statistics*

As a first look at the data, we plot the spot price and the corresponding returns in Figure 2. Since the graphs start on 13.05.2008, we observe a sharp decrease in price in the beginning of the sample period. Volatility of returns is higher in the beginning and at the end of the sample period. Mean daily return is -0.0005 with a standard deviation of 0.03 for crude oil, partially reflecting the downward trend in spot price during the sample period.

Figure 2: **WTI crude oil spot price (13.05.2008 - 31.05.2016)**

(a) **Price time-series development** (b) **Distribution**



Subfigure (a) shows temporal development of WTI crude oil spot price and logarithmic return between 13.05.2008 and 31.05.2016. Subfigure (b) shows the distribution of the logarithmic returns of crude oil price approximated through a histogram and an Epanechnikov kernel density plot.

Figure 3 shows the development of implied volatility over time[5]. The pattern has similarities with the volatility of returns, so that implied volatility is higher in the beginning and at the end of the sample. The higher IV in the beginning of the sample is likely to be related to the turmoil during the aftermath of the 2007-2008 crisis. Additional descriptive statistics can be found in Appendix E.

We consider seven potential determinants of volatility smiles. These include bid-ask spread, option trading volume, basis, hedging pressure (HP), future trading volume, dummy variable for Mondays and days-to-maturity (DTM). See Table 1 for a description of the explanatory variables.

According to one hypothesis, the deviation from the Black-Scholes postulated constant relationship between implied volatility and moneyness stems from the

[5]See Appendix A for further details on its calculation.

Figure 3: **Implied volatility (13.05.2008 - 31.05.2016)**

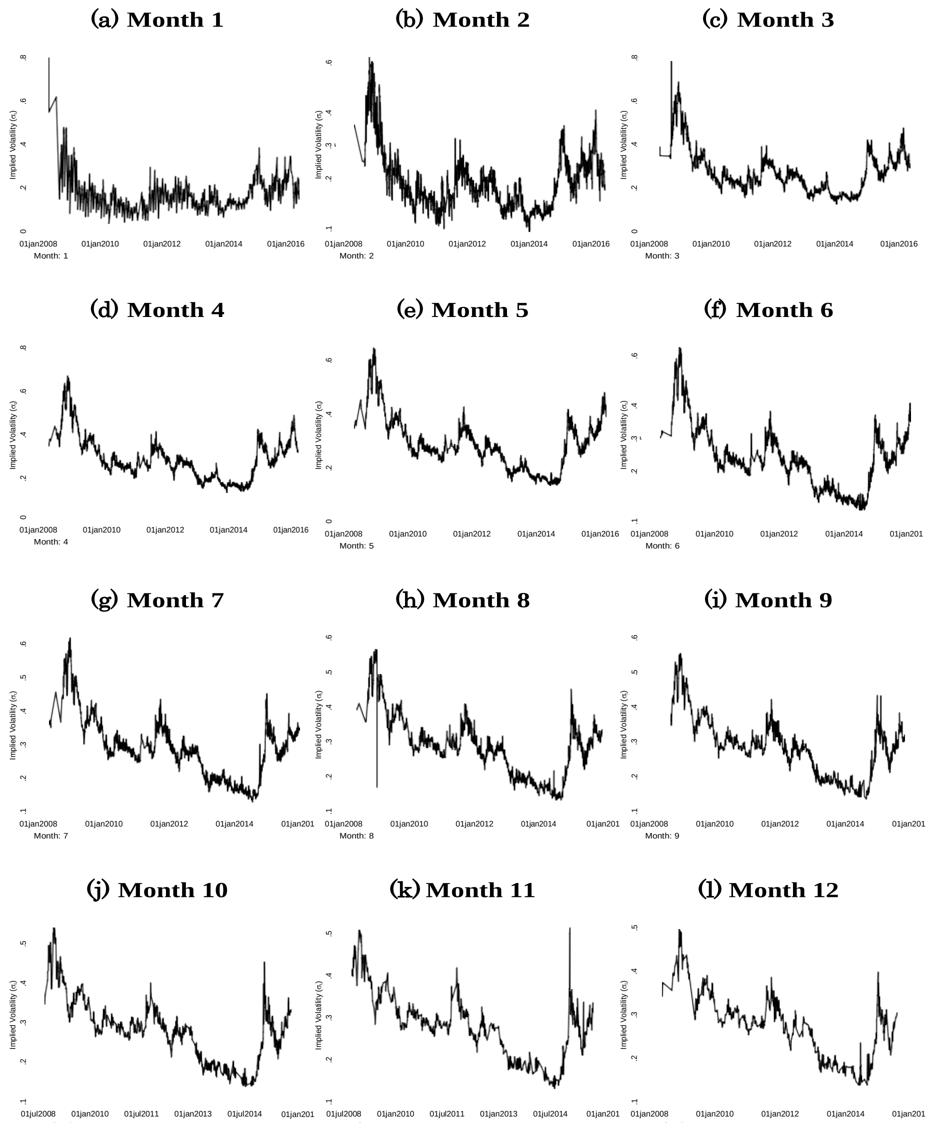


Figure shows temporal development of implied volatility between 13.05.2008 and 31.05.2016 for contracts of different maturities.

presence of transaction costs. Transaction costs include components such as bid-ask spread, opportunity cost and price impact. We utilize the [9] spread estimator, which captures the bid-ask spread and price impact component. Many spread estimators have been suggested throughout the literature. The [9] estimator is utilized in this paper as it is based on low frequency data and therefore

Table 1: **List of explanatory variables**

Variable	Description
Spread	Mean bid-ask spread for given date and maturity.
OptionVolume	Total trading volume of options for given date and maturity.
Basis	Percentage basis for the underlying crude oil future.
HP	Index of hedging pressure for the underlying crude oil future.
FutureVolume	Total trading volume of the underlying future for a given day and maturity.
Monday	Dummy variable for mondays
DTM	Days to maturity.

more feasible to implement. See Appendix B for further details. Ex ante, the presence of transaction costs is expected to cause an increase in implied volatility, i.e. the curvature of the volatility smile should increase with spread.

The size of the spread is driven by many components. For instance, spreads provide compensation for adverse selection to liquidity providers (dealers and limit orders). The probability of adverse selection is generally higher on Mondays because there has not been sufficient amount of time for the process of price discovery to eliminate information asymmetry accrued during the non-trading days of the week. In other words, a Monday dummy is expected to have the same effect on the curvature of the volatility smile as the spread.

Option trading volume can be considered as an indicator of liquidity as markets with more trading activity are generally more liquid. In contrast to spreads, which are measures of illiquidity, option trading volume is expected to have a negative effect on the curvature of the volatility smile.

Basis is the percentage difference between crude oil spot price and futures price, $(S(t) - F(t))/F(t)$. Even though basis is often defined as an absolute difference, we normalize by futures price in order to make the numbers comparable over time. The basis is an important variable. If spot price follows a Martingale process[6] for which $E(S(t+T)) = S(t)$, our definition of basis corresponds to expected hold-to-maturity returns from the underlying futures contract.

The index of hedging pressure aims to capture the relative numbers of commercial hedgers in the market. This variable is important because commercial hedgers are believed to pay a risk premium to speculators in order to fix the

[6] As pointed out by [35] and related literature, historically no-change forecasts have been better predictors of crude oil price than most forecasts based on futures price data in a Mean Squared Prediction Error (MPSE) sense.

price. This should have direct consequences on the distribution of futures returns. Appendix C provides further details.

Days-to-maturity (DTM) and futures trading volume are added as control variables. See Table 2 for summary statistics for these variables.

Table 2: **Summary statistics**

Variable	Mean	Std. dev.	Skew.	Kurt.	Min	Median	Max	N
Spread	0.04	0.06	4.49	40.41	0.00	0.01	1.75	305212
OptionVolume	300.85	772.63	6.91	92.84	1	38	28788	305212
Basis	-0.02	0.05	-1.88	10.31	-0.42	-0.01	0.16	305212
HP	-0.14	0.06	0.27	2.90	-0.29	-0.14	0.04	305212

Summary statistics of [9] spread estimator, option trading volume, basis (relative difference between spot price and futures price on WTI crude) and hedging pressure (ratio of net commercial hedgers to open interest) conditional on trading volume beging positive.

4. Results

The aim of this paper is to identify and study potential determinants of the volatility smile/skew related to WTI crude oil option contracts. It is important to keep in mind that the smile, or potentially a smirk, is not directly a variable - it is a functional form. Hence, our econometric approach is implemented through two stages. First, we begin by regressing implied volatility (σ_t) on moneyness (F/K). Second, based on the obtained regression model, we regress each of the obtained beta-coefficients from the first stage on a set of explanatory variables believed to be of importance. Stage 1 and stage 2 of this approach is presented in subsection 4.1 and 4.2 respectively. Subsections 4.3 and 4.4 further discusses the obtained results. In Subsection 4.6, we present additional robustness testing.

4.1. Stage 1

A variety of functional forms could feasibly be utilized to capture the relationship between implied volatility (σ_t) and moneyness $\left(\frac{F}{K}\right)$. One specification that has been used in the extant literature [1] and that appears to fit well with our data, is to include both a linear and squared term of moneyness, such that

$$\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right) + \beta_2 \left(\frac{F}{K}\right)^2 + \varepsilon_t. \tag{1}$$

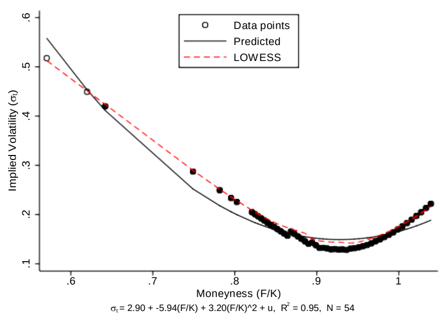
We chose to use the specification in (1) through an extensive experimentation with various functional forms. We utilized a Multivariate Fractional Polynomials approach to compute the optimal functional form, trading off model fit and parsimonious number of variables, for each date and maturity in the sample. The details of this approach can be found in Appendix D. However, the difficulty is that the optimal functional form tends to depend on the date and maturity. Therefore, there does not seem to be a particular functional form that would provide a superior fit compared to the specification in (1) for a reasonably large proportion of the sample. Moreover, the interpretation of the terms in (1) is straightforward and it can be treated as a second-order Taylor approximation for a more complex functional form.

To further justify the use of the specification in (1), we plot the relation between moneyness and implied volatility for selected dates in Figure 4. We observe that once we restrict attention to a particular date and maturity, the second-order polynomial captures the functional form surprisingly well.

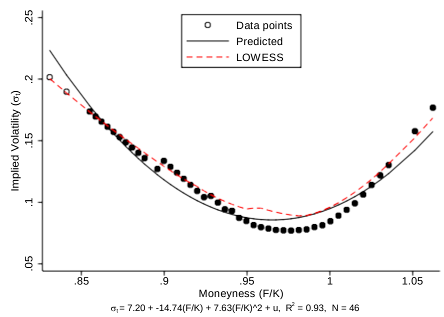
We regress implied volatility on moneyness utilizing the functional form specified in Equation (1) to our various sub-samples of data. Each sub-sample is constructed by pooling observations with the same valuation date and the same maturity date conditional on having a positive amount of trades. With these restrictions, a total of 12965 sub-samples are constructed. The sub-samples are further divided into twelve different categories based on how many months before the contracts expires. The regression results are reported in Table 3. For

Figure 4: **Relationship between implied volatility and moneyness**

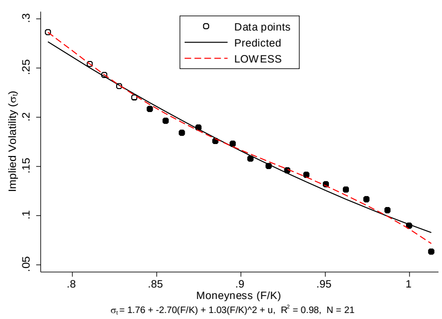
(a) **02.10.2014**



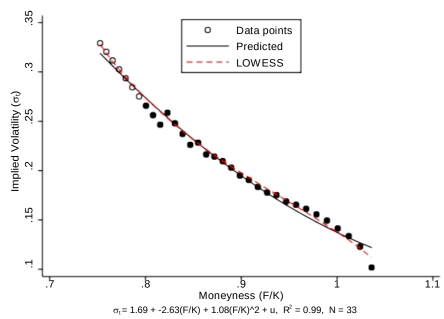
(b) **07.07.2014**



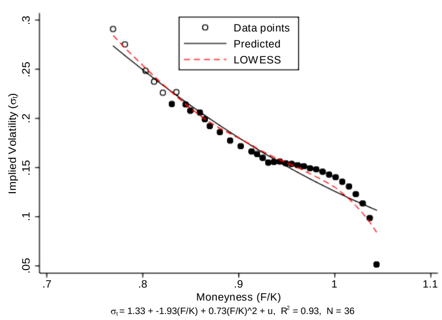
(c) **07.04.2016**



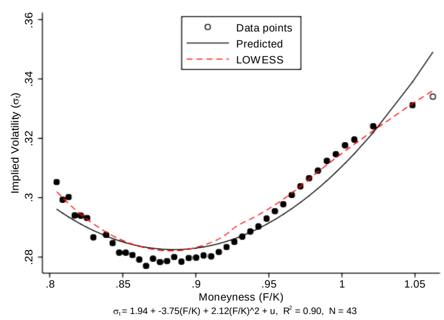
(d) **09.05.2016**



(e) **08.06.2010**



(f) **09.08.2011**



Examples of the relationship between implied volatility and moneyness selected randomly.

instance, there are 906 sub-samples of contracts with one month until expiration. Based on these 906 sub-samples, the average number of observations per sub-sample is 36.49 with a standard deviation of 14.57. Further, we obtain an average estimate of 1.22 for the constant term ($\bar{\beta}_0$), -1.94 for the linear term

($\bar{\beta}_1$) and 0.84 for the squared term ($\bar{\beta}_2$). The average explanatory power ($\bar{R}^2$) is 0.95. The distribution, approximated through histograms and (Epanechnikov) kernel density plots, of the β-coefficients across each category, are shown in Figures 5 - 7. Across the twelve categories, the constant term ($\bar{\beta}_0$) is ranging from 0.58 to 1.22, where the estimate seems to generally be higher the closer the contract is to maturity. The coefficient of the linear term ranges from −1.94 to 0.76 and between 0.48 to 0.90 coefficient of the squared term. The number of sub-samples and the average number observations in the sub-samples tend to decrease as time until expiration increases. Nevertheless, the average R^2 is quite high in all cases.

Table 3: **Stage 1 regression results**

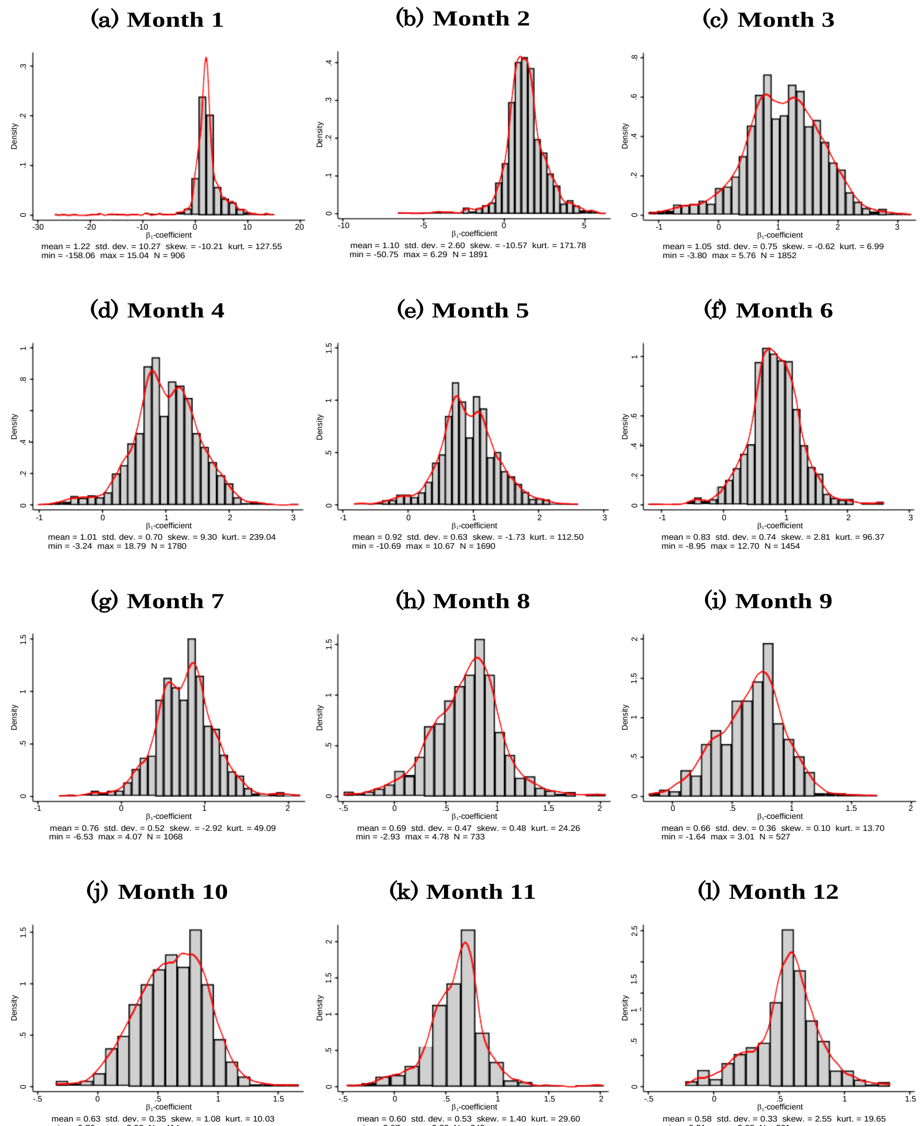
Month	$\bar{\beta}_0(\sigma_{\beta_0})$	$\bar{\beta}_1(\sigma_{\beta_1})$	$\bar{\beta}_2(\sigma_{\beta_2})$	$\bar{R}^2$	$\bar{N}(\sigma_N)$	**Sub-samples**
1	1.22(10.27)	-1.94(20.28)	0.84(10.02)	0.95	36.49(14.57)	906
2	1.10(2.60)	-1.72(5.17)	0.83(2.58)	0.93	47.07(16.34)	1891
3	1.05(0.75)	-1.68(1.69)	0.88(0.90)	0.92	37.97(16.52)	1852
4	1.01(0.70)	-1.63(1.55)	0.90(0.82)	0.95	24.37(13.33)	1780
5	0.92(0.63)	-1.46(1.40)	0.83(0.74)	0.97	14.38(9.82)	1690
6	0.83(0.74)	-1.29(1.63)	0.75(0.86)	0.98	9.63(6.87)	1454
7	0.76(0.52)	-1.13(1.26)	0.67(0.72)	0.98	8.35(6.41)	1068
8	0.69(0.47)	-1.02(1.09)	0.61(0.59)	0.98	7.47(5.57)	733
9	0.66(0.36)	-0.95(0.86)	0.58(0.48)	0.98	7.57(5.08)	527
10	0.63(0.35)	-0.91(0.84)	0.56(0.46)	0.98	8.00(5.39)	414
11	0.60(0.53)	-0.85(1.13)	0.52(0.59)	0.97	7.44(4.86)	349
12	0.58(0.33)	-0.76(0.76)	0.48(0.41)	0.99	6.77(4.43)	301

Empirical results from regressing implied volatility (σ_t) on both a linear and squared term of moneyness (F/K):

$$\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right) + \beta_2 \left(\frac{F}{K}\right)^2 + \varepsilon_t$$

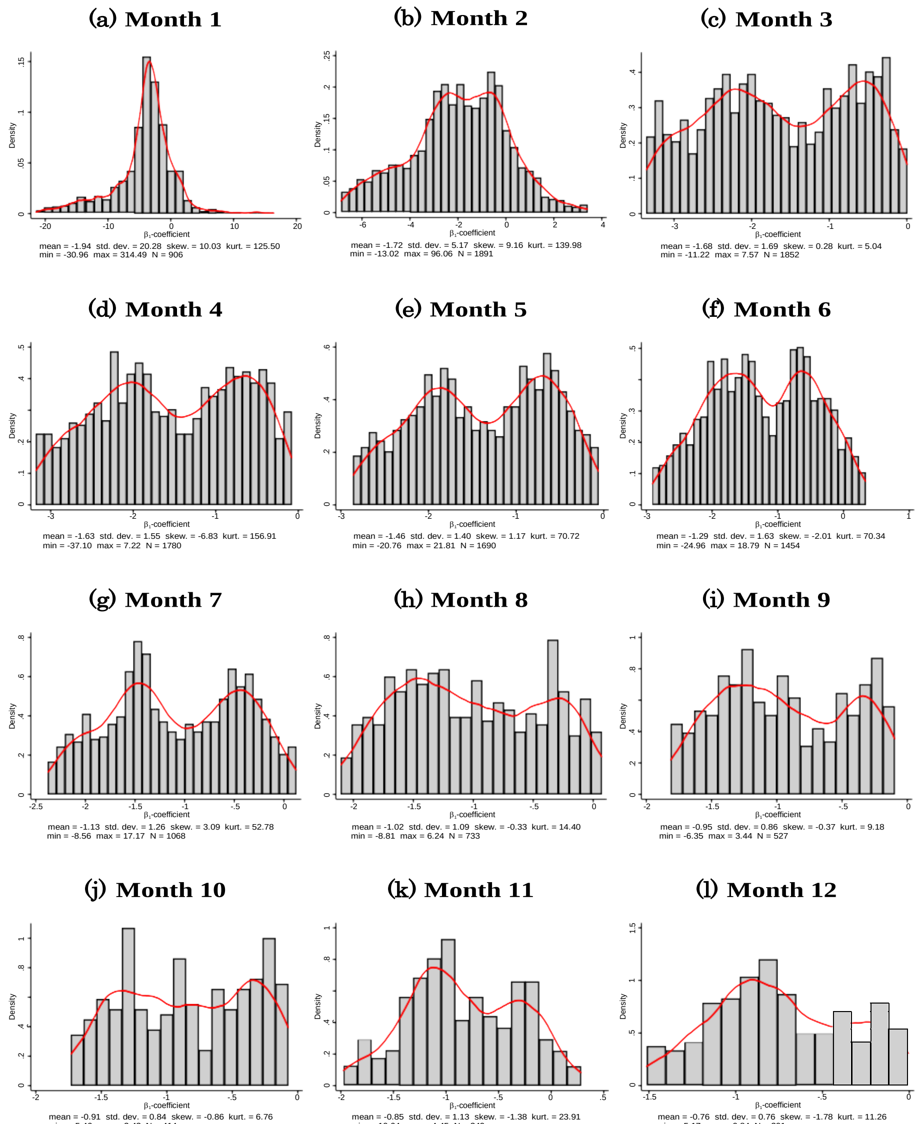
For sub-samples with the given number of months before maturity, the average values of β-coefficients, explanatory power (R^2), average number of observations (N) in the sub-samples and number of sub-samples are reported. Standard deviations for each statistic are reported in parenthesis.

Figure 5: **Distribution of β_0**



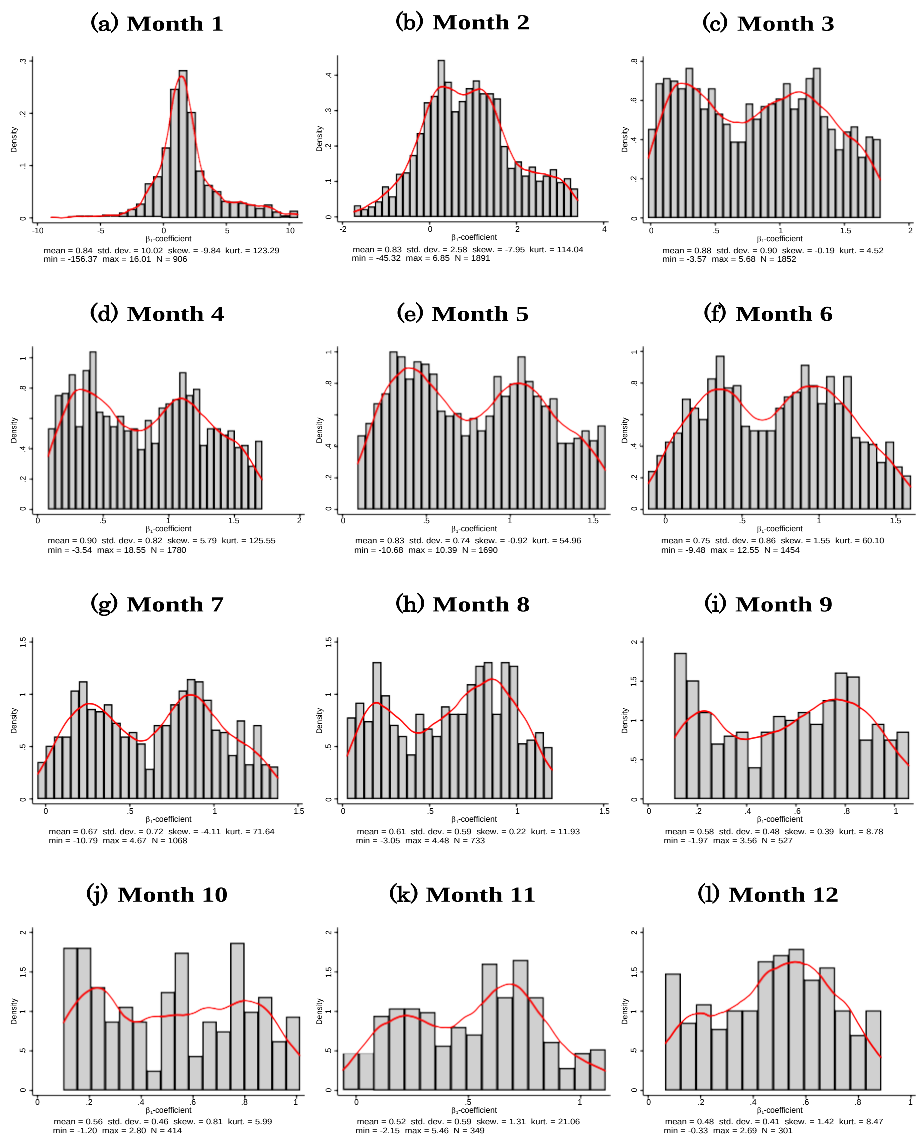
Distribution, proxied by a histogram and an Epanechnikov kernel density plot, of the constant term (β_0) from regressing implied volatility (σ_I) on both a linear and squared term of moneyness (F/K). Month indicates the approximate time to maturity.

Figure 6: **Distribution of β_1**



Distribution, proxied by a histogram and an Epanechnikov kernel density plot, of the linear term (β_1) from regressing implied volatility (σ_I) on both a linear and squared term of moneyness (F/K). Month indicates the approximate time to maturity.

Figure 7: **Distribution of β_2**



Distribution, proxied by a histogram and an Epanechnikov kernel density plot, of the squared term (β_2) from regressing implied volatility (σ_I) on both a linear and squared term of moneyness (F/K). Month indicates the approximate time to maturity.

4.2. Stage 2

To identify potential determinants of the volatility smile, we regress each of the estimated coefficients obtained from implementing Equation 1 on a set of proposed explanatory variables, see Equation 2.

$$\beta_{it} = \delta_0 + \sum_{j=1}^{N} \delta_j x_{jt} + \upsilon_{it} \quad \text{where } i \in \{0, 1, 2\} \tag{2}$$

Our list of explanatory variables is in Table 1. Regression results from implementing Equation 2 are provided in Table 4 - 6.

Table 4: **Stage-2 regression results (β_0)**

(a) Months 1 - 6:

Variable	1-month		2-month		3-month		4-month		5-month		6-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	7.04	(3.05E-01)	0.61	(4.47E-01)	0.80***	(2.02E-04)	0.35	(1.44E-01)	0.96***	(1.39E-04)	0.78*	(5.25E-02)
Spread	-26.55***	(6.25E-20)	0.07	(9.34E-01)	0.55	(1.35E-01)	2.05***	(2.76E-04)	1.18**	(3.44E-02)	1.47	(1.07E-01)
Option Volume	5.47***	(9.50E-32)	1.04***	(8.78E-72)	0.15***	(6.91E-24)	0.02	(1.51E-01)	0.01	(1.76E-01)	0.01	(5.66E-01)
Basis	-2.22	(8.97E-01)	11.47***	(6.80E-10)	4.06***	(1.97E-23)	3.14***	(8.24E-18)	1.80***	(8.24E-10)	1.02***	(6.35E-03)
HP	-2.32	(6.41E-01)	3.71***	(2.57E-05)	3.86***	(2.08E-48)	2.52***	(6.21E-19)	3.04***	(5.91E-32)	2.84***	(2.14E-15)
Future Volume	-2.59***	(5.40E-03)	-0.78***	(3.31E-09)	-0.18***	(8.40E-10)	0.00	(8.68E-01)	-0.01	(6.50E-01)	0.03	(2.81E-01)
Monday dummy	0.85	(2.75E-01)	-0.11	(4.48E-01)	-0.01	(7.68E-01)	-0.04	(3.17E-01)	0.02	(5.63E-01)	0.04	(4.47E-01)
DTM	-0.17**	(2.87E-02)	0.01*	(8.99E-02)	0.00	(1.67E-01)	0.00*	(5.98E-02)	0.00	(1.27E-01)	0.00	(3.57E-01)
Obs	906		1891		1852		1780		1690		1454	
R-squared	0.21		0.22		0.32		0.19		0.22		0.11	

(b) Months 7 - 12:

Variable	7-month		8-month		9-month		10-month		11-month		12-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	1.33***	(4.57E-04)	0.81*	(7.70E-02)	1.18***	(9.87E-03)	0.52	(3.35E-01)	0.18	(8.64E-01)	1.79**	(2.62E-02)
Spread	1.44*	(6.27E-02)	1.29	(1.21E-01)	1.68**	(3.71E-02)	0.75	(3.19E-01)	2.47	(3.27E-01)	1.94	(1.40E-01)
Option Volume	0.00	(7.79E-01)	-0.01	(3.54E-01)	0.01	(3.48E-01)	-0.02**	(1.09E-02)	0.00	(9.45E-01)	-0.04***	(3.37E-03)
Basis	1.26***	(1.23E-05)	1.14***	(1.10E-04)	0.40	(1.19E-01)	0.25	(2.70E-01)	0.64*	(9.77E-02)	0.34	(2.12E-01)
HP	1.62***	(7.62E-08)	1.61***	(3.80E-07)	1.71***	(4.89E-09)	2.28***	(3.58E-14)	1.21**	(3.51E-02)	0.48	(1.98E-01)
Future Volume	0.03	(1.38E-01)	-0.02	(2.82E-01)	0.00	(9.79E-01)	-0.04*	(7.21E-02)	0.07*	(9.01E-02)	-0.02	(3.86E-01)
Monday dummy	0.00	(9.95E-01)	-0.05	(2.44E-01)	0.02	(5.87E-01)	-0.02	(6.02E-01)	0.07	(3.65E-01)	0.02	(6.45E-01)
DTM	0.00**	(3.20E-02)	0.00	(5.65E-01)	0.00	(1.00E-01)	0.00	(7.76E-01)	0.00	(8.31E-01)	0.00	(1.12E-01)
Obs	1068		733		527		414		349		301	
R-squared	0.13		0.14		0.15		0.26		0.07		0.08	

Empirical results from regressing implied volatility β-coefficients on [9] spread estimator, option volume, basis, hedging pressure (HP), futures volume, Monday-dummy and DTM:

$$\begin{aligned}\beta_{it} &= \delta_0 + \delta_1 Spread + \delta_2 OptionVolume + \delta_3 Basis + \delta_4 HP + \\ &\quad \delta_5 FutureVolume + \delta_6 Monday + \delta_7 DTM + \varepsilon_{it} \\ &\quad \text{where } i \in \{0, 1, 2\}\end{aligned}$$

The dependent variable, β_0, is obtained from the following regression $\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right) + \beta_2 \left(\frac{F}{K}\right)^2 + \varepsilon_t$. One (*), two (**) and three (***) asterisks denotes significance levels of 10, 5 and 1%.

Table 5: **Stage-2 regression results (β_1)**

(a) Months 1 - 6:

Variable	1-month		2-month		3-month		4-month		5-month		6-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	-13.55	(3.17E-01)	-1.00	(5.26E-01)	-0.81^{*}	(7.52E-02)	0.05	(9.14E-01)	-1.30^{**}	(1.27E-02)	-0.94	(2.68E-01)
Spread	52.70^{***}	(4.06E-20)	-0.43	(7.92E-01)	-1.24	(1.06E-01)	-4.53^{***}	(1.04E-04)	-3.15^{***}	(6.22E-03)	-4.16^{**}	(2.85E-02)
Option Volume	-10.62^{***}	(8.08E-31)	-2.01^{***}	(2.21E-69)	-0.30^{***}	(7.35E-21)	-0.03	(2.81E-01)	-0.02	(3.15E-01)	-0.01	(7.11E-01)
Basis	-9.82	(7.71E-01)	-29.27^{***}	(1.17E-15)	-12.41^{***}	(2.20E-46)	-9.61^{***}	(2.33E-36)	-6.38^{***}	(2.59E-25)	-4.44^{***}	(1.43E-08)
HP	1.91	(8.46E-01)	-9.69^{***}	(2.11E-08)	-9.16^{***}	(2.89E-60)	-6.41^{***}	(1.47E-27)	-7.59^{***}	(5.21E-45)	-7.11^{***}	(2.74E-21)
Future Volume	5.16^{***}	(5.05E-03)	1.62^{***}	(4.50E-10)	0.35^{***}	(3.51E-08)	-0.02	(6.44E-01)	0.01	(7.95E-01)	-0.09^{*}	(9.02E-02)
Monday dummy	-1.64	(2.84E-01)	0.23	(3.98E-01)	0.03	(7.53E-01)	0.08	(3.19E-01)	-0.03	(7.00E-01)	-0.07	(4.75E-01)
DTM	0.33^{**}	(3.38E-02)	-0.02^{*}	(8.42E-02)	-0.01^{*}	(5.07E-02)	-0.01^{**}	(2.68E-02)	0.00	(2.10E-01)	0.00	(4.18E-01)
Obs	906		1891		1852		1780		1690		1454	
R-squared	0.21		0.24		0.40		0.30		0.34		0.20	

(b) Months 7 - 12:

Variable	7-month		8-month		9-month		10-month		11-month		12-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	-2.20^{**}	(1.06E-02)	-1.06	(2.73E-01)	-2.11^{**}	(3.88E-02)	-0.11	(9.28E-01)	-0.18	(9.35E-01)	-3.20^{*}	(5.90E-02)
Spread	-3.96^{**}	(2.43E-02)	-4.24^{**}	(1.54E-02)	-4.33^{**}	(1.66E-02)	-2.16	(1.89E-01)	-6.32	(2.20E-01)	-4.96^{*}	(7.43E-02)
Option Volume	0.00	(8.61E-01)	0.02	(3.17E-01)	-0.01	(4.44E-01)	0.06^{***}	(6.06E-03)	0.01	(8.62E-01)	0.09^{***}	(5.33E-04)
Basis	-4.64^{***}	(1.57E-12)	-3.81^{***}	(1.13E-09)	-2.51^{***}	(1.50E-05)	-1.66^{***}	(8.95E-04)	-2.70^{***}	(6.88E-04)	-2.11^{***}	(2.54E-04)
HP	-4.68^{***}	(9.12E-12)	-5.12^{***}	(4.60E-14)	-4.54^{***}	(5.75E-12)	-6.09^{***}	(5.68E-20)	-3.47^{***}	(3.15E-03)	-1.87^{**}	(1.89E-02)
Future Volume	-0.08^{*}	(5.52E-02)	0.04	(4.03E-01)	-0.01	(8.88E-01)	0.10^{**}	(4.48E-02)	-0.15^{*}	(8.01E-02)	0.03	(4.16E-01)
Monday dummy	-0.01	(9.35E-01)	0.12	(1.89E-01)	-0.03	(6.80E-01)	0.06	(5.18E-01)	-0.12	(4.20E-01)	-0.06	(5.71E-01)
DTM	0.01^{**}	(4.67E-02)	0.00	(5.64E-01)	0.01^{*}	(9.08E-02)	0.00	(9.16E-01)	0.00	(9.32E-01)	0.01	(1.21E-01)
Obs	1068		733		527		414		349		301	
R-squared	0.23		0.28		0.27		0.40		0.17		0.22	

Empirical results from regressing implied volatility β-coefficients on [9] spread estimator, option volume, basis, hedging pressure (HP), futures volume, Monday-dummy and DTM:

$$\begin{aligned} \beta_{it} &= \delta_0 + \delta_1 Spread + \delta_2 OptionVolume + \delta_3 Basis + \delta_4 HP + \\ &\quad \delta_5 FutureVolume + \delta_6 Monday + \delta_7 DTM + \varepsilon_{it} \\ &\quad \text{where } i \in \{0, 1, 2\} \end{aligned}$$

The dependent variable, β_1, is obtained from the following regression $\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right) + \beta_2 \left(\frac{F}{K}\right)^2 + \varepsilon_t$. One (*), two (**) and three (***) asterisks denotes significance levels of 10, 5 and 1%.

Table 6: **Stage-2 regression results (β_2)**

(a) Months 1 - 6:

Variable	1-month		2-month		3-month		4-month		5-month		6-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	6.42	(3.37E-01)	0.60	(4.41E-01)	0.33	(1.63E-01)	-0.05	(8.56E-01)	0.70^{***}	(9.71E-03)	0.57	(1.99E-01)
Spread	-26.14^{***}	(2.98E-20)	0.34	(6.68E-01)	0.62	(1.27E-01)	2.35^{***}	(9.62E-05)	1.71^{***}	(4.23E-03)	2.31^{**}	(2.04E-02)
Option Vol	5.16^{***}	(6.27E-30)	0.96^{***}	(3.29E-66)	0.15^{***}	(1.98E-19)	0.01	(3.66E-01)	0.01	(3.78E-01)	0.00	(7.01E-01)
Basis	12.51	(4.52E-01)	17.34^{***}	(9.33E-22)	7.47^{***}	(2.76E-59)	5.62^{***}	(1.43E-45)	3.83^{***}	(3.99E-33)	2.70^{***}	(5.05E-11)
HP	-0.02	(9.97E-01)	5.37^{***}	(3.28E-10)	4.67^{***}	(6.26E-57)	3.25^{***}	(1.35E-26)	3.89^{***}	(3.81E-44)	3.68^{***}	(8.80E-21)
Future Volume	-2.58^{***}	(4.62E-03)	-0.85^{***}	(3.22E-11)	-0.18^{***}	(2.55E-08)	0.02	(5.42E-01)	0.00	(9.05E-01)	0.06^{**}	(3.37E-02)
Monday dummy	0.80	(2.90E-01)	-0.12	(3.56E-01)	-0.01	(7.64E-01)	-0.04	(3.33E-01)	0.01	(7.62E-01)	0.04	(4.55E-01)
DTM	-0.15^{**}	(4.87E-02)	0.01^{**}	(3.94E-02)	0.01^{***}	(5.47E-03)	0.00^{***}	(9.28E-03)	0.00	(3.09E-01)	0.00	(4.00E-01)
Obs	906		1891		1852		1780		1690		1454	
R-squared	0.21		0.26		0.42		0.32		0.36		0.21	

(b) Months 7 - 12:

Variable	7-month		8-month		9-month		10-month		11-month		12-month	
	Coefficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value	Co efficient	p-value
Constant	1.27^{**}	(1.03E-02)	0.64	(2.23E-01)	1.31^{**}	(2.09E-02)	0.16	(8.06E-01)	0.20	(8.55E-01)	1.96^{**}	(2.96E-02)
Spread	2.28^{**}	(2.39E-02)	2.29^{**}	(1.53E-02)	2.44^{**}	(1.47E-02)	1.17	(1.91E-01)	3.53	(1.82E-01)	2.89^{**}	(4.95E-02)
Option Volume	0.00	(8.05E-01)	-0.01	(3.05E-01)	0.01	(5.10E-01)	-0.03^{***}	(2.63E-03)	0.00	(8.65E-01)	-0.05^{***}	(2.64E-04)
Basis	2.77^{***}	(2.07E-13)	2.26^{***}	(2.49E-11)	1.53^{***}	(2.05E-06)	1.06^{***}	(1.05E-04)	1.60^{***}	(9.54E-05)	1.30^{***}	(2.26E-05)
HP	2.46^{***}	(3.89E-10)	2.73^{***}	(8.48E-14)	2.33^{***}	(1.46E-10)	3.23^{***}	(3.79E-19)	1.82^{***}	(2.61E-03)	1.00^{**}	(1.78E-02)
Future Volume	0.05^{**}	(4.83E-02)	-0.01	(6.28E-01)	0.01	(7.12E-01)	-0.05^{*}	(5.83E-02)	0.09^{*}	(5.54E-02)	-0.02	(4.89E-01)
Monday dummy	0.01	(8.42E-01)	-0.06	(2.09E-01)	0.02	(6.62E-01)	-0.03	(5.91E-01)	0.06	(4.28E-01)	0.03	(5.35E-01)
DTM	0.00^{*}	(5.71E-02)	0.00	(5.61E-01)	0.00^{*}	(7.14E-02)	0.00	(9.43E-01)	0.00	(9.50E-01)	0.00^{*}	(7.98E-02)
Obs	1068		733		527		414		349		301	
R-squared	0.22		0.29		0.27		0.41		0.19		0.26	

Empirical results from regressing implied volatility β-coefficients on [9] spread estimator, option volume, basis, hedging pressure (HP), futures volume, Monday-dummy and DTM:

$$\begin{aligned}\beta_{it} &= \delta_0 + \delta_1 Spread + \delta_2 OptionVolume + \delta_3 Basis + \delta_4 HP + \\ &\quad \delta_5 FutureVolume + \delta_6 Monday + \delta_7 DTM + \varepsilon_{it} \\ &\quad \text{where } i \in \{0, 1, 2\}\end{aligned}$$

The dependent variable, β_2, is obtained from the following regression $\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right) + \beta_2 \left(\frac{F}{K}\right)^2 + \varepsilon_t$. One (*), two (**) and three (***) asterisks denotes significance levels of 10, 5 and 1%.

4.3. *Interpretation of Stage-2 Regression Results*

Interpretation of Stage-2 regression results is subject to two complexities. First, we are dealing with functional forms rather than individual observations. Second, the options contracts vary by maturity and there is no theoretical reason to believe that all maturities behave similarly. To address these complexities in more detail, we start with a couple of remarks.

Functional Form. We approximate the implied volatility function with a second-order polynomial. The approximation satisfies the following two rules:

- Given β_1 and β_2, an increase in β_0 shifts the whole implied volatility function upwards.
- Given β_0 and β_1, an increase in β_2 increases the curvature of the implied volatility function. If $\beta_2 > 0$ the function open upwards, otherwise it opens downwards.

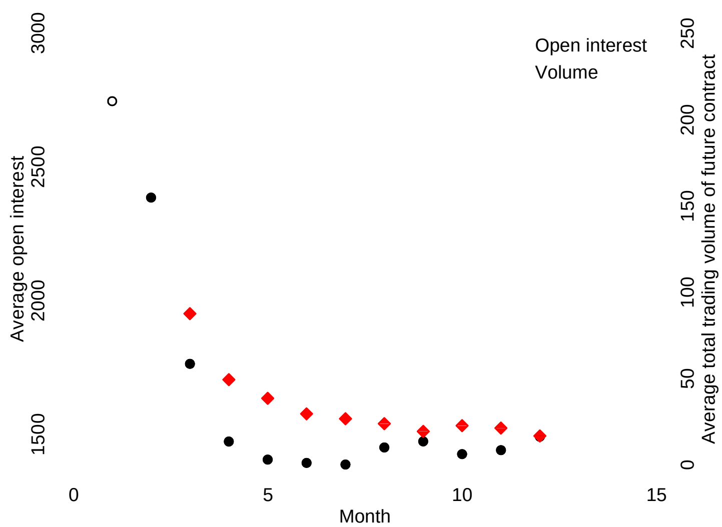
These two rules suggests that for our purposes, determinants of β_0 and β_2 are the most interesting. Changes in β_1 primarily shift the curve sideways[7].

Maturity of Option Contracts. We report the results separately for all maturities between 1-month and 12-months. The only data restriction is that we require a positive volume of trade for at least three levels of moneyness each day in order for an observation to be included. By investigating the results, we observe that the 1-month and 2-month contracts behave very differently from all other contracts. Moreover, the results seem very consistent across the longer maturity contracts. Therefore, our principal interest is to investigate results for maturities ranging from 3-month to 12-month contracts. We believe that the front contracts are mostly used for speculative trading whereas contracts with a longer maturity are used for hedging purposes. In fact, the traded volume and open interest tend to be much higher for the front contracts, as shown in Figure 8.

We now proceed with interpreting the results. It should be emphasized that statistical significance is subject to some caveats here. For one, the dependent variable itself is an estimate from the first-stage regression and therefore the usual OLS (Ordinary Least Squares) standard errors are not directly applicable. Secondly, recent research has shown that even the usual OLS standard

[7] To be more specific, a second-order equation is a parabola whose turning point in (x, y)-plane is given by $\left(\frac{-\beta_1}{2\beta_2}, \beta_0 - \frac{\beta_1^2}{4\beta_2}\right)$. Therefore, an increase in β_1 shifts the curve both vertically and horizontally. However, empirically our curves tend to have a turning point around moneyness level of one. Constraining $\frac{-\beta_1}{2\beta_2} = 1$ implies that there is a direct relationship between β_1 and β_2. Therefore, we believe that the precise value of β_1 does not offer many insights for our analysis and we choose to concentrate on analyzing the distribution of β_2.

Figure 8: **Open interest and volume for various contract maturities**



errors may be misleading in large samples[8]. However, despite these challenges we believe that the standard errors convey relevant information about the underlying correlation. We have added p-values to Tables 4-6 in case the reader wants to assess the results in more detail.

Table 4 shows that the coefficients for bid-ask spread are consistently positive for all maturities starting from 3 months. The coefficients on option volume are first positive but turn soon negative. Hence, higher transaction costs tend to be associated with a higher level of implied volatility.

The coefficients on basis and hedging pressure are consistently positive and statistically significant for maturities starting from 4 months. Therefore, a high basis or a net short position commercial hedgers is associated with a higher average level of implied volatility. The last three explanatory variables in Table 4 seem to be less correlated with implied volatility.

Table 6 contains the main results of this paper. We note that for most maturi-

[8] In particular, [39] argues that there is a need for a careful appraisal of the results from statistical models and that the sample size should be explicitly considered when assessing the relevance of estimated parameters. [40] shows that when assessing the statistical significance of parameter estimates in regression equations, relying on 'conventional' significance levels (e.g. 5%) implies to be wrong 30% of the time. As a response to this, [41] proposes that one should consider the sample size in each regression by setting a reference p-value of $\min\left(\frac{1}{T}, 1\%\right)$ where T is the sample size, and work out reference t-statistics in line with this. We thank an anonymous referee for pointing out this issue.

ties starting from 4 months, either bid-ask spread is positive or option volume is negative and statistically significant. Hence high transaction costs are associated with a higher degree of curvature in the implied volatility function. However, even more consistent and statistically significant pattern is observed with basis and hedging pressure. All maturities starting from 3 months are positive and numerically consistent with each other. Therefore, implied volatility function tends to smile more during times of a high basis or a net short position of commercial hedgers in the underlying futures contract. On the other hand, the last three variables in Table 6 do not show a very consistent pattern.

4.4. Empirical Kurtosis of the Underlying Futures Returns

The results in the previous section indicate that basis and hedging pressure are important determinants of volatility smiles. This is in contrast with [1] who estimate that transaction costs are the main culprits for volatility smiles. Furthermore, as discussed in the literature review, the Black-Scholes model has been extended in various ways to modify the return distribution of the underlying asset so that the predictions would be more in line with empirical observations.

Since the importance and role of the return distribution of the underlying asset seems to be controversial in the literature, we look further into the empirical return distribution in the hope of shedding more light on this important question. Since basis and hedging pressure seem to be of importance for crude oil options, we construct two categories based on the values of each of these variables. The first category includes observations below the 25th percentile and the second category is for observations above the 75th percentile. We plot the empirical return distributions for these categories.

Figure 9 shows the distribution for the basis categories for all maturity classes. The category with a higher basis has a lower standard deviation of returns for each maturity class. However, the main interest here is the fourth moment of the return distribution. In the Black-Scholes model the return distribution is log-Normal for which kurtosis equals three. As can be seen in Figure 10, the empirical kurtosis consistently exceeds 3. Even more importantly, kurtosis is much higher for high levels of basis which correspond to a bigger volatility smile, as concluded in the previous section.

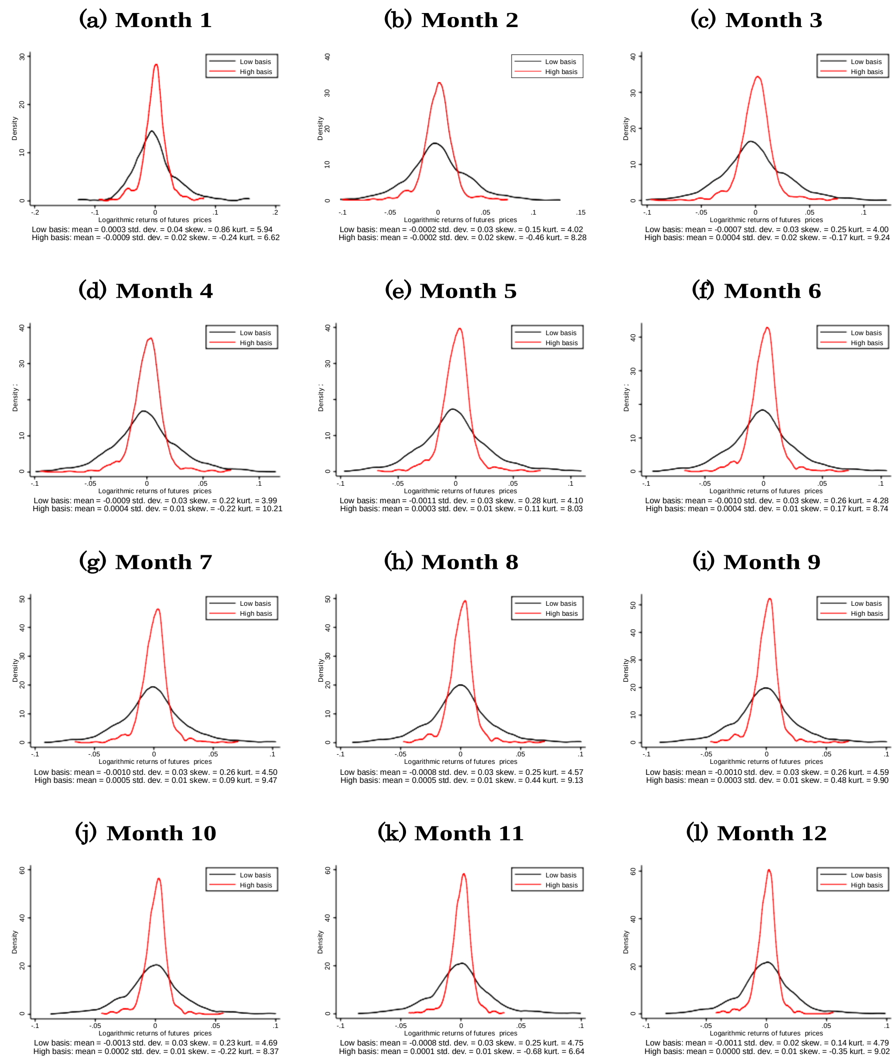
Figure 10 shows similar graphs using hedging pressure in forming the two categories. High hedging pressure correspond to a higher standard deviation of returns. The role of kurtosis is less obvious compared to Figure 10. For most maturity classes, especially for the short ones, kurtosis seems to be higher for the category with a lower hedging pressure. However, this is reversed for some contracts with a longer maturity.

Overall, since basis is the most convincing determinant of volatility smiles and its effect on the kurtosis of the underlying return distribution is very clear, we believe that we have found a channel which is plausible both in theory and practice.

4.5. Economic Intuition

Let us discuss the economic reasons why the underlying asset distribution may show more kurtosis during times of high basis. This feature follows from the theory of storage. Imagine a world where a representative producer receives a “harvest” each period and sells the harvest to the market. Since demand is downward sloping, a larger harvest corresponds to a lower price. Moreover, fluctuations in harvest over time cause price fluctuations. Let us further assume

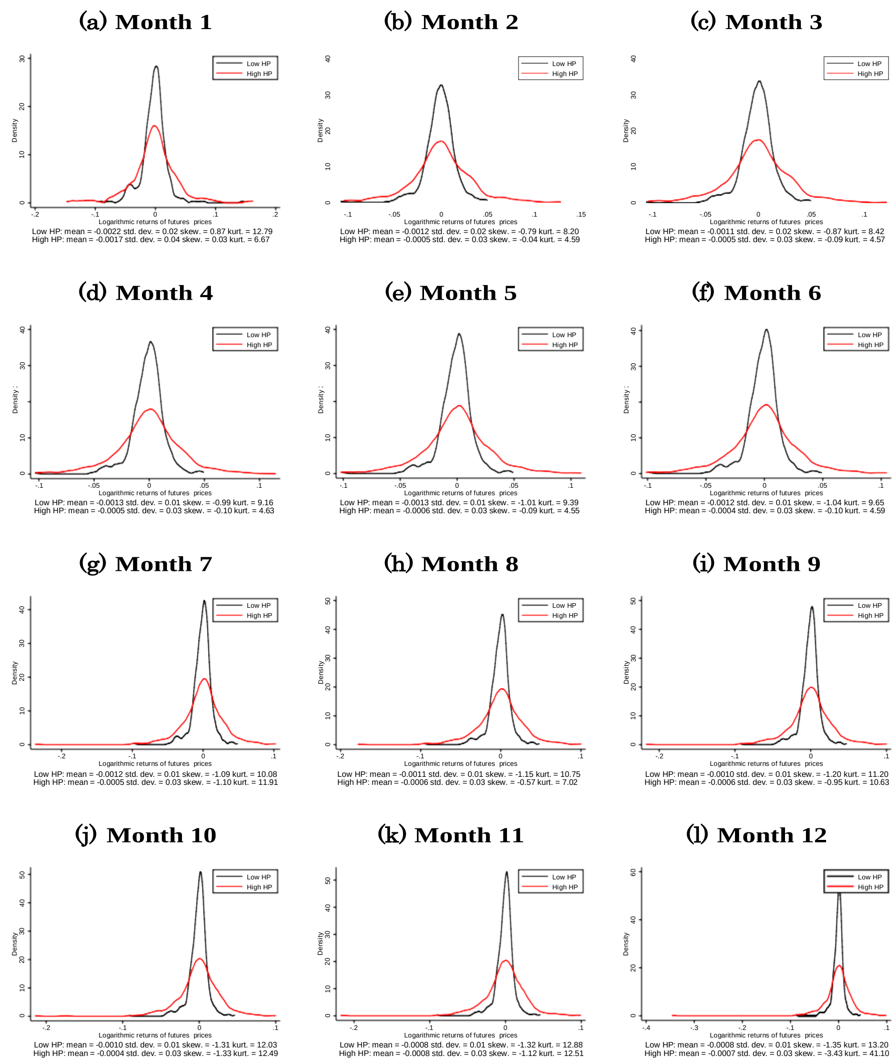
Figure 9: **Futures returns across two levels of basis**



that the price distribution over several periods follows a Normal distribution when there is no possibility to store. How does storage affect the long-term price distribution?

To answer this question, let us consider two regimes: (i) a regime with high inventory levels and infrequent stockouts (ii) a regime with low inventories and frequent stockouts. With high inventory levels, the effect of storage is to decrease kurtosis through general equilibrium price effects. Compared to the benchmark case of no storage, inventory holders tend to buy when the price is low, thereby

Figure 10: **Futures returns across two levels of HP**



"thinning out" the left end of the price distribution. Moreover, inventory holders tend to sell when the price is high, thereby thinning out the right end of the price distribution. As a result, kurtosis decreases as the inventory holder affect the price through their actions.

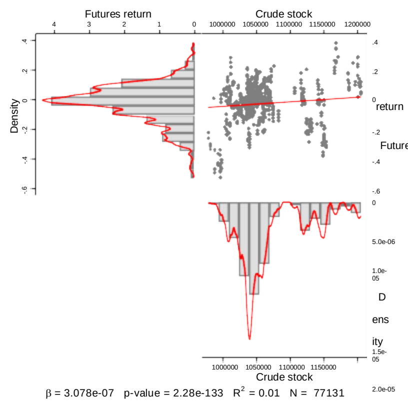
The effect of storage in the second regime is more complex. Inventory holders again buy when the price is low and thin out the left end of the distribution. However, in the case of a stockout, there is nothing more to sell and the price may peak at a very high level. This is the mechanism that creates price peaks, positive skewness and increased kurtosis according to the theory of storage. Compared to the first regime, where both left and right ends of the price distribution are being thinned out by inventory holders, the non-negativity constraint for inventories creates a fat right tail for the price distribution. This has a big impact on kurtosis.

We believe that the theory of storage is relevant for the crude oil market[9]. To study this further, we use data on U.S. crude oil inventories to see how they relate to realized returns from holding crude oil futures. Figure 11 shows the relationship between crude oil inventory level and daily futures return for various maturities. We observe that for shorter maturities, the relationship is positive while it quickly turns negative as the maturities become longer. As such, we take this as evidence that futures returns of longer maturities are higher when inventory levels are lower. We believe that these patterns relate to kurtosis and basis as discussed above.

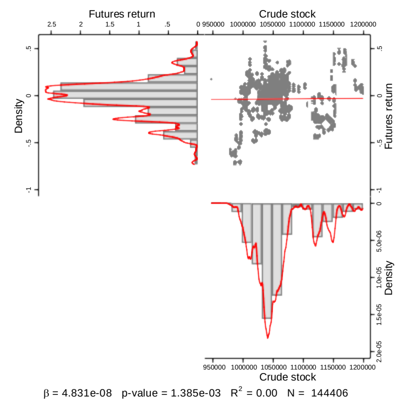
[9] See [42] for an analysis of the relationship between inventory levels and convenience yield in the natural gas market.

Figure 11: **Relationship between futures return and crude oil inventory level.**

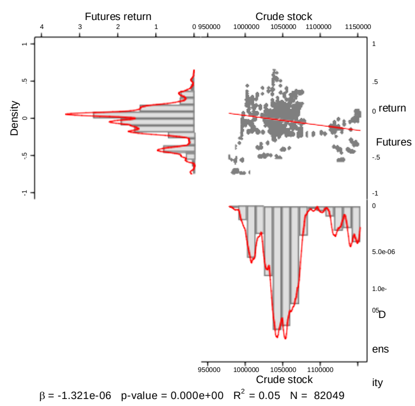
(a) **Month 1**



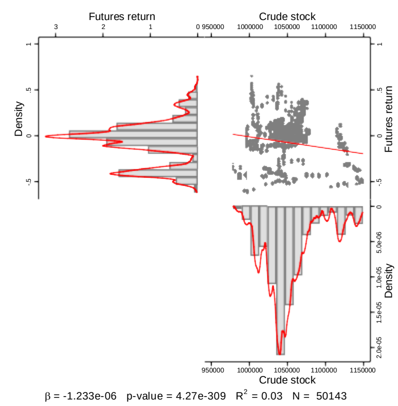
(b) **Month 4**



(c) **Month 7**



(d) **Month 10**



Scatter plot between logarithmic returns of futures prices and crude oil stock with a fitted OLS regression line. The statistical distribution, approximated through a histogram and an Epanechnikov kernel density plot, for each variable is also added.

4.6. Classification of smile

One pertinent question relates to the classification of the functional forms. In other words, are WTI option contracts predominantly smiling or smirking? Based on visual inspection of a subset of the 12965 stage 1 regressions, four categories appears to be present: smiles, smirks (reverse skew), forward skew and frowns. Based on the fitted functional forms, we classify the relationships as follows. Smiles are u-shaped functional forms. By ordering implied volatility in an increasing order based on moneyness, if the observation with the lowest level of moneyness (σ_1) is greater than the observation with the second lowest (σ_2) and the observation with second highest (σ_{N-1}) is lower than the highest (σ_N), then we are dealing with a smile. Frowns, as the name implies, is an upside down smile. Smirks, on the other hand, are monotonically decreasing and forward skews are monotonically increasing. See Equation 3 for a summary.

$$Type = \begin{cases} \text{Smile} & \text{if } \sigma_1 > \sigma_2 \text{ and } \sigma_{N-1} < \sigma_N \\ \text{Smirk} & \text{if } \sigma_1 > \sigma_2 \text{ and } \sigma_{N-1} > \sigma_N \\ \text{Frown} & \text{if } \sigma_1 < \sigma_2 \text{ and } \sigma_{N-1} > \sigma_N \\ \text{Forward skew} & \text{if } \sigma_1 < \sigma_2 \text{ and } \sigma_{N-1} < \sigma_N \end{cases} \quad (3)$$

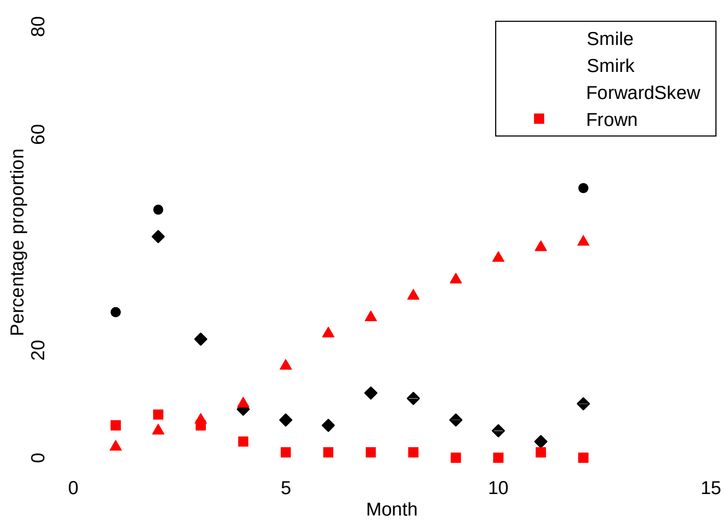
The shape of the functional form is informative of demand for options. For instance, a high demand for in-the-money and out-of-the-money call options leads to smiles. On the other hand, high demand for at-the-money calls results in frowns. High demand for in-the-money calls yields reverse skew (smirks), while high demand for out-of-the-money calls leads to forward skew. The presence of a smile indicates that the distribution of the underlying exhibits leptokurtosis, i.e. heavier tails compared to the assumed lognormal distribution of the Black-Scholes model. Smirk, on the other hand, implies a distribution of a heavier (leptokurtic) left-tail and less heavier (platykurtic) right-tail compared to log-normal distribution. Forward skew, as opposed to smirk, is characterized by a less heavier left-tail and heavier right-tail.

Table 7 and figure 12 shows the results from applying these classification heuristics. For contracts with one month until expiration, 27% are classified as smiles, 65% as smirks, 6% as frowns and 2% as forward skews. As the maturities become longer, smiles and forward skews tend to occur more frequently. On the contrary, smirks and frowns tend to occur less frequently. This pattern is interesting, and suggests that volatility smiles and forward skews are more relevant from the point of view of long-term investors.

Table 7: **Classification of implied volatility and moneyness**

Month	Smile	Smirk	Frown	Forward skew
1	241 (27%)	592 (65%)	55 (6%)	18 (2%)
2	879 (46%)	768 (41%)	151 (8%)	93 (5%)
3	1212 (65%)	404 (22%)	115 (6%)	121 (7%)
4	1384 (78%)	165 (9%)	46 (3%)	185 (10%)
5	1260 (75%)	120 (7%)	17 (1%)	293 (17%)
6	1009 (69%)	94 (6%)	16 (1%)	335 (23%)
7	657 (62%)	125 (12%)	11 (1%)	275 (26%)
8	429 (59%)	78 (11%)	9 (1%)	217 (30%)
9	316 (60%)	37 (7%)	1 (0%)	172 (33%)
10	241 (58%)	20 (5%)	1 (0%)	152 (37%)
11	200 (57%)	12 (3%)	2 (1%)	135 (39%)
12	152 (50%)	30 (10%)	0 (0%)	119 (40%)
Sum	7980(62%)	2445(19%)	424(3%)	2115(16%)

Figure 12: **Proportion of volatility-moneyness relationships**



Striving towards gaining an understanding for what causes the different smile-patterns, we apply multinomial logit regression to the classification reported in Table 7 (see Equation (4)). We use the same independent variables as in Stage 2 of the main regressions. Results from this approach are shown in Table 8. The coefficients can be challenging to interpret given the varying scale of the independent variables. Nevertheless, an increase in the high-low spread estimator is associated with an significantly increased probability of observing either a smile or smirk compared to the base category. Spread estimator, however, does not seem to help predicting forward skew. Option trading volume predicts increased probability of smile, but decreased probability of smirk. Basis seems to behave in a similar fashion as option trading volume and hedging pressure conforms to the spread estimator. Days-to-maturity increases probability of all classes compared to the base category, but more so for forward skew. While the rigorousness of this approach is not sufficient to claim causality, it does suggest that the shape of the relationship between implied volatility and moneyness tend to change with characteristics as transaction costs proxied through spread estimators and option trading volume.

$$p_{ij} = \frac{exp(X_i^t \beta_j)}{\sum_{l=1}^{m} exp(X_i^t \beta_l)}, \quad j = 1, \cdots, m \tag{4}$$

Table 8: **Multinomial logistic regression result**

Variable	Smile	Reverse skew	Forward skew
Constant	-0.13	-1.55*	-0.17
Spread	23.78*	22.86*	-29.47**
OptionVolume	0.25***	-0.21***	0.01
Basis	3.69**	-3.05*	10.39***
HP	16.12***	10.08***	1.71
FutureVolume	0.14	0.05	0.06
Monday	-0.07	0.01	0.01
DTM	0.014***	0.011***	0.02***

Reference category: Frown.

Table 8 suggests that basis and hedging pressure may affect the implied volatility function through different channels. A high basis increases the probability of forward skews while it decreases the probability of reverse skews. High hedging pressure, on the other hand, increases the probability of reverse skews but does not significantly affect the probability of forward skews. Altogether, this suggests that basis could be a more important determinant for the right side of the implied volatility function, while hedging pressure could be better at explaining the left side.

Let us elaborate further on the possibly different implications of basis and hedging pressure on volatility smiles. As discussed in section 4.5, the theory of storage provides a plausible explanation for the relationship between basis, skewness and kurtosis. These are likely to affect the right side of the implied volatility function. The patterns on the left side, however, may be more easily explained by the concept of 'crashophobia'.

Crashophobia refers to strong negative skewness in the expected crude oil price distribution. In other words, crashophobia happens when the probability of a large decrease in oil price exceeds the probability of a large increase. Typically, put options are used as hedging instruments to protect against large downward movements in oil price. A high demand by investors due to portfolio insurance strategies will increase the price of put options which is also reflected to the prices of call options through put-call parity. As a result, the left tail of the implied volatility distribution shifts up.

5. Conclusion

The main objective of this paper is twofold. First, we attempt to capture the relationship between implied volatility and moneyness (here defined as the ratio between futures prices and strike price) for WTI crude oil call options traded on NYMEX between May 13, 2008 and May 31, 2016. Second, conditional on a given functional form describing the relationship, we investigate several potential determinants of the β-coefficients.

We find that a second-order equation is sufficient to adequately capture the relationship. More elaborate model fitting with multivariate fractional polynomials yields a marginally better fit. The increase in explanatory power, however, does arguably not justify the decrease in parsimony. Inspection of the predicted values of implied volatility reveals that volatility smiles, reverse skew (smirk), forward skew and frowns are present to varying degrees depending on the time to maturity.

When regressing the β coefficients of the second-order equation on various candidate explanatory variables, we find a significant correlation between the shape of implied volatility functions and basis and hedging pressure. The volatility function tends to be flatter when either basis is low or commercial hedgers are net long. We believe that the reason for this pattern is that the underlying asset distribution exhibits more kurtosis during times of high basis or high hedging pressure.

Secondly, we find that the [9] spread estimator and option trading volume are also important determinants of the shape of the implied volatility function. Variables such as futures trading volume, days-to-maturity and weekend effect proxied through a dummy variable for Mondays appears to not have a significant effect.

References

[1] I. Pena, G. Rubio, G. Serna, Why do we smile? on the determinants of the implied volatility function, Journal of Banking & Finance 23 (8) (1999) 1151–1179.

[2] N. P. Bollen, R. E. Whaley, Does net buying pressure affect the shape of implied volatility functions?, The Journal of Finance 59 (2) (2004) 711–753.

[3] F. A. Longstaff, Option pricing and the martingale restriction, The Review of Financial Studies 8 (4) (1995) 1091–1124.

[4] B. Dumas, J. Fleming, R. E. Whaley, Implied volatility functions: Empirical tests, The Journal of Finance 53 (6) (1998) 2059–2106.

[5] A. Deaton, G. Laroque, On the behaviour of commodity prices, The Review of Economic Studies 59 (1) (1992) 1–23.

[6] A. Deaton, G. Laroque, Competitive storage and commodity price dynamics, Journal of Political Economy 104 (5) (1996) 896–923.

[7] H. Working, The theory of price of storage, The American Economic Review 39 (6) (1949) 1254–1262.

[8] P. Liu, K. Tang, The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield, Journal of Empirical Finance 18 (2) (2011) 211–224.

[9] F. Abdi, A. Ranaldo, A simple estimation of bid-ask spreads from daily close, high, and low prices, The Review of Financial Studies 30 (12) (2017) 4437–4480.

[10] P. Royston, D. G. Altman, Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling, Applied statistics (1994) 429–467.

[11] W. Sauerbrei, P. Royston, Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials, Journal of the Royal Statistical Society: Series A (Statistics in Society) 162 (1) (1999) 71–94.

[12] B. Dupire, et al., Pricing with a smile, Risk 7 (1) (1994) 18–20.

[13] E. Derman, I. Kani, Riding on a smile, Risk 7 (2) (1994) 32–39.

[14] M. Rubinstein, Implied binomial trees, The Journal of Finance 49 (3) (1994) 771–818.

[15] G. Brown, K. B. Toft, Constructing binomial trees from multiple implied probability distributions, The Journal of Derivatives 7 (2) (1999) 83–100.

[16] J. C. Jackwerth, M. Rubinstein, Recovering probability distributions from option prices, The Journal of Finance 51 (5) (1996) 1611–1631.

[17] J. C. Cox, S. A. Ross, The valuation of options for alternative stochastic processes, Journal of financial economics 3 (1-2) (1976) 145–166.

[18] D. C. Emanuel, J. D. MacBeth, Further results on the constant elasticity of variance call option pricing model, Journal of Financial and Quantitative Analysis 17 (4) (1982) 533–554.

[19] J. Hull, A. White, The pricing of options on assets with stochastic volatilities, The journal of finance 42 (2) (1987) 281–300.

[20] S. L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, The review of financial studies 6 (2) (1993) 327–343.

[21] E. M. Stein, J. C. Stein, Stock price distributions with stochastic volatility: an analytic approach, The review of financial studies 4 (4) (1991) 727–752.

[22] S. R. Das, S. Foresi, Exact solutions for bond and option prices with systematic jump risk, Review of derivatives research 1 (1) (1996) 7–24.

[23] D. S. Bates, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, The Review of Financial Studies 9 (1) (1996) 69–107.

[24] G. Bakshi, C. Cao, Z. Chen, Empirical performance of alternative option pricing models, The Journal of finance 52 (5) (1997) 2003–2049.

[25] Y. Aıt-Sahalia, Y. Wang, F. Yared, Do option markets correctly price the probabilities of movement of the underlying asset?, Journal of Econometrics 102 (1) (2001) 67–110.

[26] T. G. Andersen, L. Benzoni, J. Lund, Estimating jump-diffusions for equity returns, Journal of Finance 57 (3) (2002) 1239–1284.

[27] D. Duffe, J. Pan, K. Singleton, Transform analysis and option pricing for affine jump diffusion, Econometrica 68 (2000) 929–952.

[28] S. R. Das, R. K. Sundaram, Of smiles and smirks: A term structure perspective, Journal of Financial and Quantitative Analysis 34 (2) (1999) 211–239.

[29] A. Shleifer, R. W. Vishny, The limits of arbitrage, The Journal of Finance 52 (1) (1997) 35–55.

[30] J. Liu, F. A. Longstaff, Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities, The Review of Financial Studies 17 (3) (2003) 611–641.

[31] V. V. Acharya, L. A. Lochstoer, T. Ramadorai, Limits to arbitrage and hedging: Evidence from commodity markets, Journal of Financial Economics 109 (2) (2013) 441–465.

[32] W. R. Melick, C. P. Thomas, Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf crisis, Journal of Financial and Quantitative Analysis 32 (1) (1997) 91–115.

[33] J. Fleming, B. Ostdiek, The impact of energy derivatives on the crude oil market, Energy Economics 21 (2) (1999) 135–167.

[34] B. M. Tabak, D. O. Cajueiro, Are the crude oil markets becoming weakly efficient over time? a test for time-varying long-range dependence in prices and volatility, Energy Economics 29 (1) (2007) 28–36.

[35] R. Alquist, L. Kilian, What do we learn from the price of crude oil futures?, Journal of Applied Econometrics 25 (4) (2010) 539–573.

[36] S. H. Kang, S.-M. Kang, S.-M. Yoon, Forecasting volatility of crude oil markets, Energy Economics 31 (1) (2009) 119–125.

[37] P. Agnolucci, Volatility in crude oil futures: a comparison of the predictive ability of garch and implied volatility models, Energy Economics 31 (2) (2009) 316–321.

[38] X. Du, L. Y. Cindy, D. J. Hayes, Speculation and volatility spillover in the crude oil and agricultural commodity markets: A bayesian analysis, Energy Economics 33 (3) (2011) 497–503.

[39] J. P. Ioannidis, Why most published research findings are false, PLoS medicine 2 (8) (2005) e124.

[40] D. Colquhoun, An investigation of the false discovery rate and the misinterpretation of p-values, Royal Society open science 1 (3) (2014) 140216.

[41] H. Leeb, B. M. Pötscher, Model selection and inference: Facts and fiction, Econometric Theory 21 (1) (2005) 21–59.

[42] S. Z. C. Wei, Z. Zhu, Commodity convenience yield and risk premium determination: The case of the us natural gas market, Energy Economics 28 (4) (2006) 523–534.

[43] F. Black, The pricing of commodity contracts, Journal of financial economics 3 (1-2) (1976) 167–179.

[44] Z. Zhang, Multivariable fractional polynomial method for regression model, Annals of translational medicine 4 (9).

Appendix A. Implied volatility

Implied volatilities are computed from data using the [43] commodity option pricing formula for call options given by

$$c = e^{-rT}\left(F_0 N(d_1) - K N(d_2)\right) \tag{A.1}$$

where N refers to the c.d.f . of a standard Normal distribution and

$$d_1 = \frac{\ln\left(\frac{F_0}{K}\right) + \frac{\sigma^2}{2} T}{\sigma\sqrt{T}} \tag{A.2}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{A.3}$$

In the three equations above, c is the price of the call option, F_0 is the futures price at the time of purchase, K is the exercise price, r is the risk-free interest rate, T is time to maturity and σ is the implied volatility.

Appendix B. Bid-Ask Spread

A potential explanation for implied volatility smiles is the existence of transaction costs in the options market. Since our dataset does not include quote data, we utilize the low-frequency estimate developed by [9]. The advantage of this estimate is the fact that it utilizes a wider information set consisting of daily close, high and low prices. In particular, [9] define the mid-range as the average of daily high and low log-prices:

$$\eta_t = \frac{l_t + h_t}{2}$$

If c_t refers to daily closing price, the estimate for bid-ask spread s is given by

$$s = 2\sqrt{E\left((c_t - \eta_t)(c_t - \eta_{t+1})\right)} \tag{B.1}$$

As explained by [9], some observations for s^2 may be negative. We replace those observations by $s = 0$.

Appendix C. Index of Hedging Pressure

We construct the index of hedging pressure as follows:

$$HP = \frac{\text{Net Short Commercial Hedgers}}{\text{Open Interest}} \tag{C.1}$$

Investors in the futures market are categorized into reportable and non-reportable depending on the size of their holdings. Reportable investors are further divided into ‘Commercial Hedgers’ and ‘Speculators’. We use data on mean weekly amount of commercial hedgers divided by daily open interest to construct the index in equation C.1.

Appendix D. Multivariate Fractional Polynomials

Throughout the paper, we have modeled the implied volatility function as a second order equation, i.e., $\sigma_t = \beta_0 + \beta_1\left(\frac{F}{K}\right) + \beta_2\left(\frac{F}{K}\right)^2 + \varepsilon_t$. While this functional form have been found to be provide the best fit of the data in the extant literature [1], we apply more rigorous model specification procedure to ensure ensure robustness of our results. Specifically, we utilize multivariate fractional polynomials to model the relationship between implied volatility (σ_t) and moneyness for a given time to maturity. See [10] and [11]. The volatility smile is modeled as:

$$\sigma_t = \beta_0 + \sum_{m=1}^{M} \beta_m \left(\frac{F}{K}\right)^{p_m}, \tag{D.1}$$

where moneyness is defined as the ratio between futures price and strike price $\left(\frac{K}{F}\right)$ and $p \in \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. When p is zero, $moneyness^0$ is taken as $\ln(moneyness)$. A total of 44 model specifications are considered - 8 models with one term $\binom{8}{1}$ and 36 models with two terms $\left(\binom{8}{2} + \binom{8}{1}\right)$. In the case of a two-term model where p_1 is equal to p_2, the model is estimated as:

$$\sigma_t = \beta_0 + \beta_1 \left(\frac{F}{K}\right)^{p} + \beta_2 \left(\frac{F}{K}\right)^{p} \cdot \ln\left(\frac{F}{K}\right). \tag{D.2}$$

Model specification is based on a a three-step procedure comprised of an (1) inclusion test, (2) a non-linearity test and (3) a simplification test. First, to determine whether to include the proposed explanatory variable, the best-fitting second degree fractional polynomial (FP2) model is compared to a constant term only model (null model). Second, given that the independent variable is deemed useful, the best-fitting second-degree fractional polynomial is compared to the linear model. Third, given that the second degree fractional polynomial model is better that the linear model, it is compared to the best single term fractional polynomial model. See [44] for further information on the specification procedure. This procedure is depicted in Figure D.13 and the obtained results are shown in Table D.9.

Figure D.13: **Specification procedure**

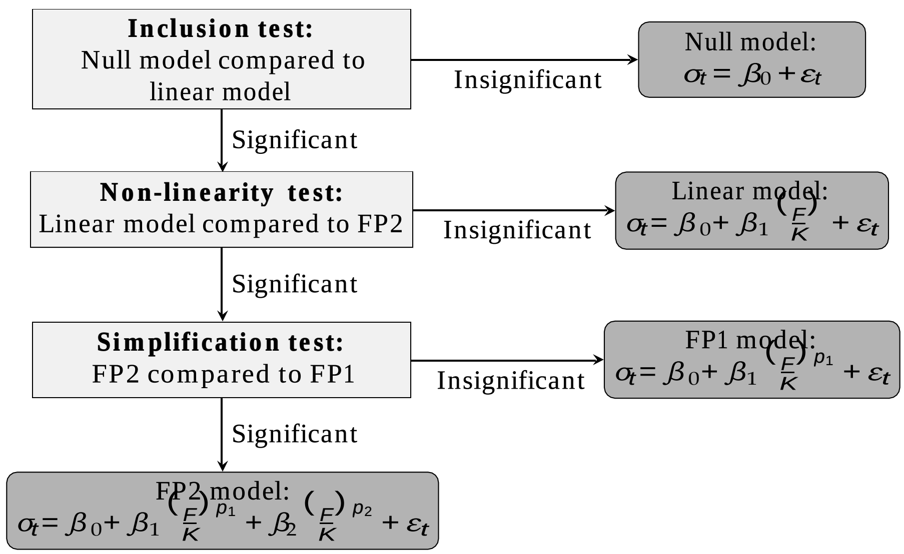


Illustration of the multivariate fractional polynomial model specification procedure. Based on three tests, there are four potential outcomes: (1) a null model with only a constant term, (2) a linear model, (3) a second-degree fractional polynomial model with 36 alternative specifications (FP2) and (4) a first-degree fractional polynomial model (FP1) with 8 alternative specifications.

Table D.9: **Multivariate fractional polynomial regression results**

	Specification	$\bar{\beta}_0(\sigma_{\beta_0})$	$\bar{\beta}_1(\sigma_{\beta_1})$	$\bar{\beta}_2(\sigma_{\beta_2})$	$\bar{R}^2$	N
1	p={1}	0.29(0.09)	-0.02(0.24)		0.78	4160
2	p={3,3}	0.19(0.07)	-0.40(0.44)	1.38(1.46)	0.97	2367
3	p={-2,-2}	0.29(0.09)	1.75(49.68)	-1.59(27.15)	0.94	1454
4	p={2,3}	0.21(0.06)	-1.77(0.95)	1.27(0.65)	0.99	652
5	p={2,2}	0.22(0.06)	-0.62(0.35)	1.46(0.68)	0.99	554
6	p={-2}	0.31(0.15)	0.00(0.98)		0.94	418
7	p={1,2}	0.24(0.07)	-2.33(1.98)	1.27(1.02)	0.99	401
8	p={1,1}	0.25(0.08)	-1.85(1.80)	2.05(1.87)	0.99	290
9	p={0.5,1}	0.26(0.08)	-6.09(4.19)	3.19(2.20)	0.99	243
10	p={-2,3}	0.39(0.11)	0.05(0.04)	-0.07(0.09)	0.97	227
11	p={0.5,0.5}	0.27(0.09)	-6.25(6.39)	3.28(3.28)	0.99	226
12	p={0.5,2}	0.24(0.07)	-2.53(1.78)	0.74(0.48)	0.99	184
13	p={0,0.5}	0.26(0.08)	-2.96(2.19)	6.20(4.52)	0.99	167
14	p={0,0}	0.28(0.07)	0.13(0.16)	0.63(0.60)	0.99	162
15	p={-2,-1}	0.29(0.10)	0.77(1.94)	-1.66(4.07)	0.98	148
16	p={0.5,3}	0.23(0.06)	-1.60(0.72)	0.35(0.13)	1.00	141
17	p={-0.5,0}	0.29(0.09)	5.38(9.30)	2.73(4.61)	0.99	139
18	p={3}	0.33(0.10)	-0.06(0.13)		0.95	137
19	p={-0.5,-0.5}	0.27(0.08)	-5.06(4.36)	-2.40(2.14)	0.99	129
20	p={-1,-1}	0.29(0.10)	-1.08(0.99)	-0.99(0.92)	0.98	124
21	p={-1,-0.5}	0.28(0.08)	2.27(4.60)	-4.76(9.42)	0.99	118
22	p={-1}	0.32(0.14)	0.35(0.27)		0.97	110
23	p={-2,-0.5}	0.30(0.07)	0.24(0.33)	-1.10(1.64)	0.98	78
24	p={1,3}	0.24(0.06)	-0.95(0.34)	0.40(0.12)	1.00	63
25	p={-0.5}	0.27(0.13)	1.06(0.74)		0.98	49
26	p={0}	0.25(0.11)	-0.61(0.37)		0.98	27
27	p={-0.5,3}	0.23(0.07)	0.87(0.44)	0.22(0.12)	1.00	22
28	p={-2,2}	0.30(0.07)	0.06(0.10)	0.08(0.24)	0.99	19
29	p={0,3}	0.25(0.07)	-0.50(0.31)	0.25(0.14)	1.00	19
30	p={2}	0.34(0.03)	-0.17(0.23)		0.97	18
31	p={-1,3}	0.29(0.08)	0.22(0.14)	0.07(0.14)	1.00	17
32	p={-2,0}	0.30(0.06)	0.07(0.17)	0.10(0.67)	0.98	13
33	p={-2,1}	0.27(0.08)	0.15(0.12)	0.46(0.51)	0.98	13
34	p={-2,0.5}	0.33(0.10)	-3.19(10.91)	0.33(1.39)	0.96	11
35	p={-0.5,2}	0.27(0.09)	0.76(0.46)	0.27(0.17)	0.99	11
36	p={-1,0.5}	0.29(0.05)	0.50(0.49)	1.24(1.22)	1.00	9
37	p={0,1}	0.37(0.04)	-0.54(0.11)	0.58(0.11)	1.00	8
38	p={0.5}	0.27(0.11)	-1.08(0.67)		0.99	7
39	p={-1,2}	0.32(0.05)	0.33(0.22)	0.22(0.11)	1.00	6
40	p={-1,0}	0.32(0.03)	0.52(0.33)	0.65(0.43)	1.00	6
41	p={-1,1}	0.30(0.04)	0.14(0.23)	0.25(0.24)	1.00	5
42	p={0,2}	0.30(0.05)	-0.41(0.21)	0.25(0.16)	1.00	5
43	p={-0.5,1}	0.25(0.08)	1.48(0.77)	0.92(0.47)	1.00	5
44	p={-0.5,0.5}	0.28(0.07)	2.21(1.83)	2.42(2.19)	1.00	3

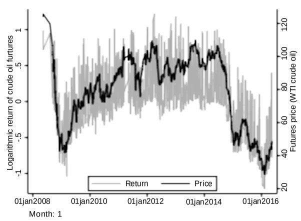
Empirical results from regressing implied volatility on moneyness: $\sigma_t = \beta_0 + \sum_{m=1}^{M} \beta_m \left(\frac{F}{K}\right)^{p_m}$.

Average values of the β-coefficient and explanatory power (R^2) are reported with their associated standard deviation in parenthesis. Specification lists the fractional polynomials powers and N denotes how many times the given functional form was selected out of 849 sub-samples.

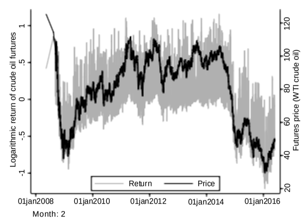
Appendix E. Additional descriptive statistics

Figure E.14: **WTI crude oil futures price (13.05.2008 - 31.05.2016)**

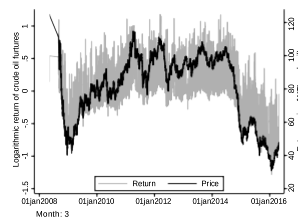
(a) **Month 1**



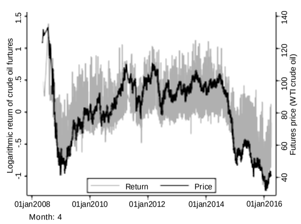
(b) **Month 2**



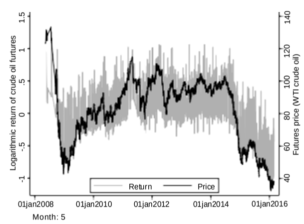
(c) **Month 3**



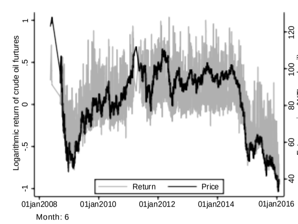
(d) **Month 4**



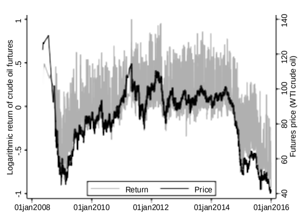
(e) **Month 5**



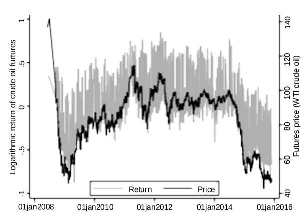
(f) **Month 6**



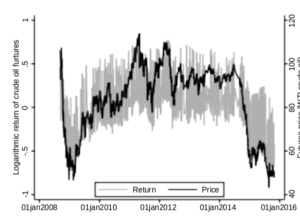
(g) **Month 7**



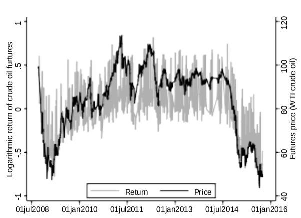
(h) **Month 8**



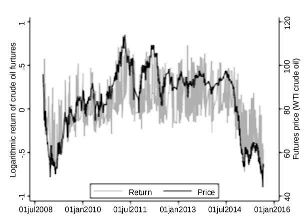
(i) **Month 9**



(j) **Month 10**



(k) **Month 11**



(l) **Month 12**

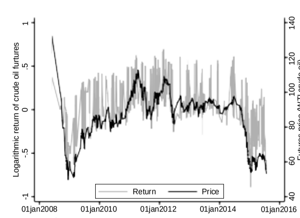


Figure shows temporal development of WTI crude oil futures price and logarithmic return between 13.05.2008 and 31.05.2016 for contracts of different maturities

Figure E.15: **Distribution of WTI crude oil futures return (13.05.2008 - 31.05.2016)**

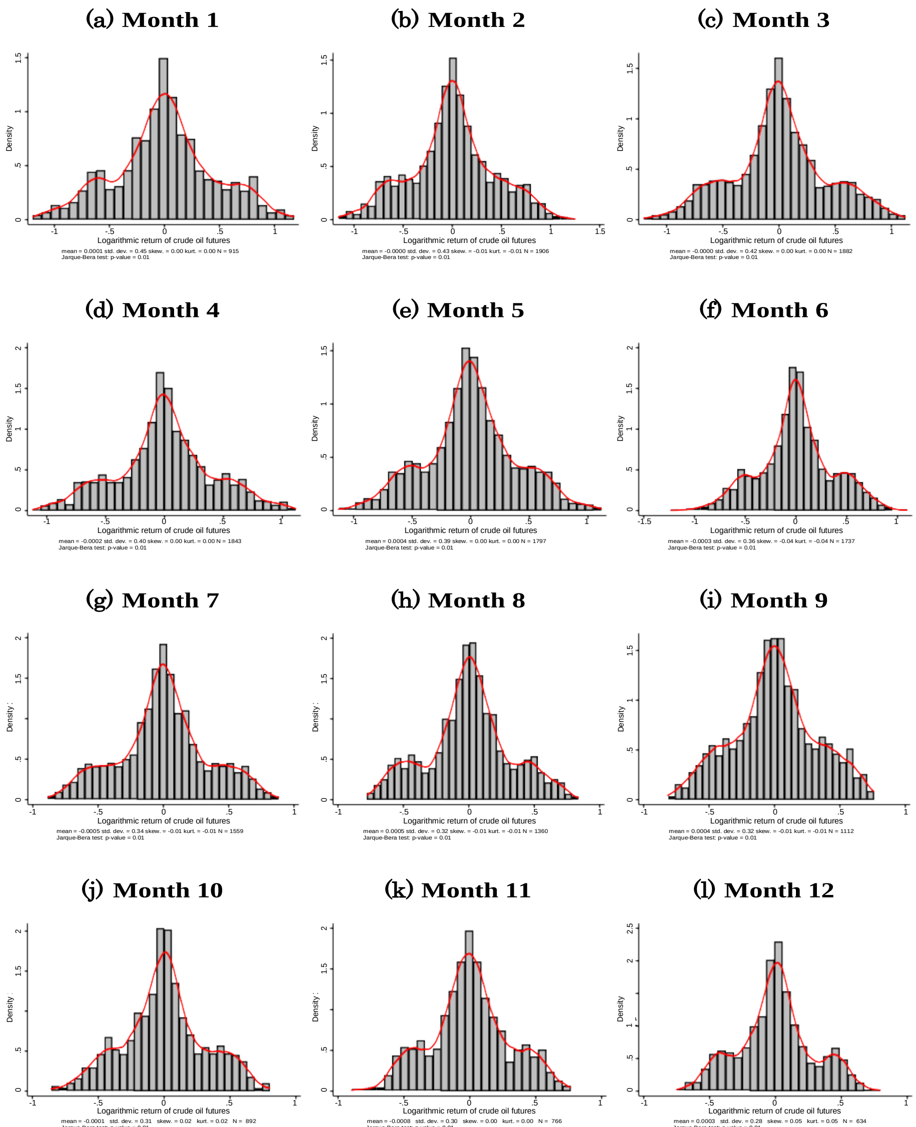


Figure shows the distributions of logarithmic returns between 13.05.2008 and 31.05.2016 for contracts of different maturities.

Figure E.16: **WTI crude oil option closing price (13.05.2008 - 31.05.2016)**

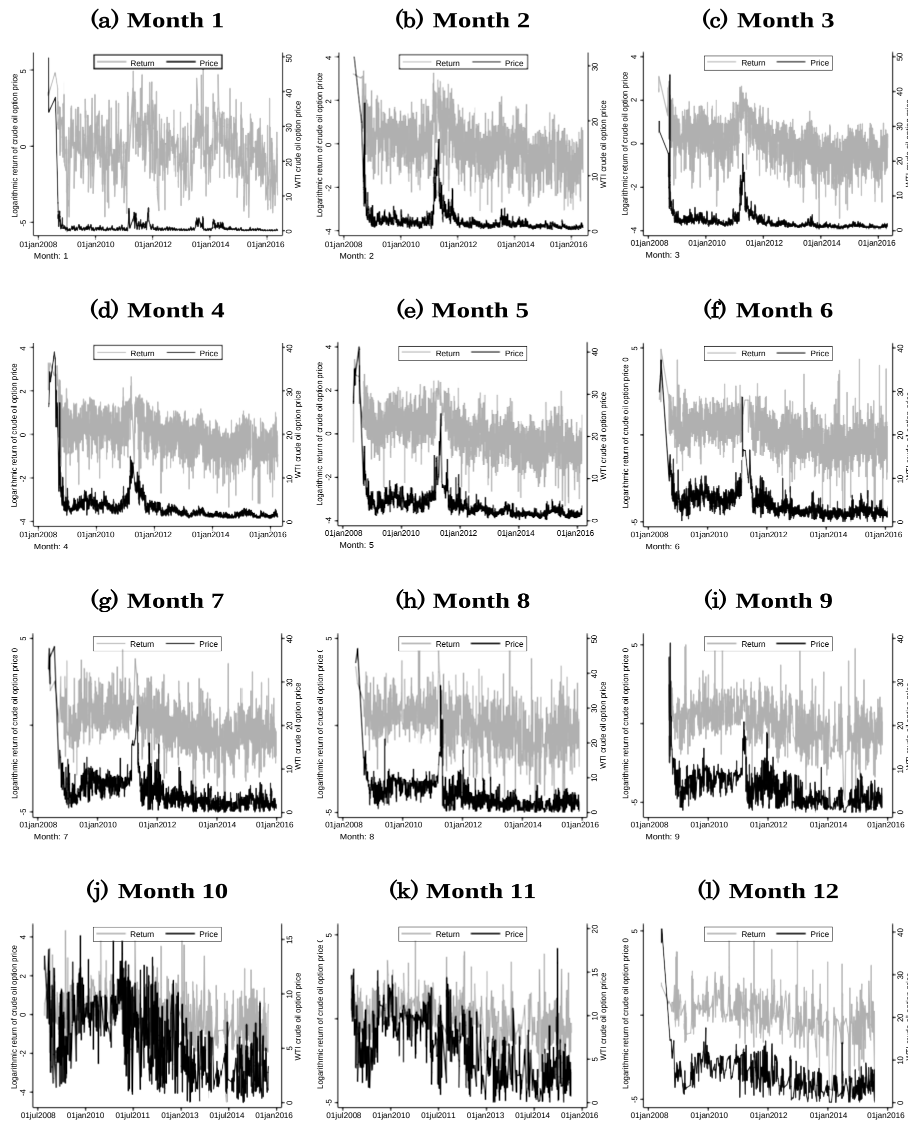


Figure shows temporal development of WTI crude oil options price and logarithmic return between 13.05.2008 and 31.05.2016 for contracts of different maturities.

Figure E.17: **Distribution of WTI crude oil option closing return (13.05.2008 - 31.05.2016)**

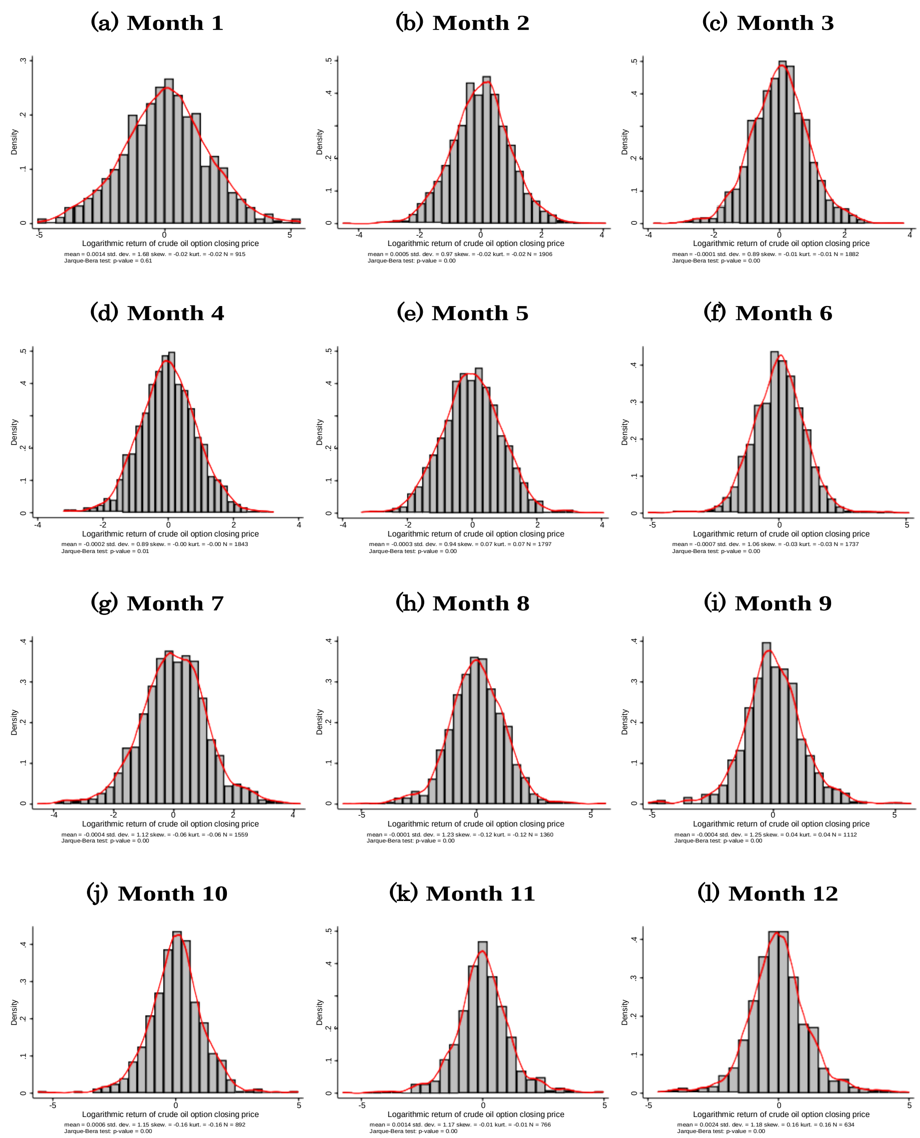


Figure shows the distributions of logarithmic options returns between 13.05.2008 and 31.05.2016 for contracts of different maturities.

*Research Highlights

Highlights

- This paper studies deviations from constant implied volatility across various levels of moneyness for crude oil options.
- We find that the implied volatility function exhibits a “smile” on 62% of the days in the sample.
- The occurrence of smiles is positively correlated with basis and hedging pressure of the underlying crude oil futures contract and positively correlated with various measures of transaction costs.
- We find that the underlying futures return distribution has fatter tails during the times of high basis or high hedging pressure.

***Abstract**

Option Prices and Implied Volatility in the Crude Oil Market

This paper studies the determinants of WTI crude oil call option prices with a special emphasis on the relationship between implied volatility and moneyness. Our first-stage regression estimates a quadratic approximation of implied volatility as a function of moneyness, while our second-stage regression investigates correlations between the estimated parameters and a list of explanatory variables. The first-stage regressions show a positive coefficient on the quadratic term, suggesting that the market exhibits 'Implied Volatility Smile' and hence violates the Black-Scholes predictions. The main results of our paper concern the determinants of these violations. We find that the curvature of implied volatility as a function of moneyness is: (i) positively and significantly correlated with basis and hedging pressure of the underlying crude oil futures contract (ii) positively and significantly correlated with various measures of transaction costs on the options market. We explore various explanations for these results. The paper also contains a variety of robustness checks, mostly related to the assumed functional forms.