The Implied Volatility Smirk in the Commodity Market

Xiaolan Jia
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: jiaxi125@student.otago.ac.nz

Xinfeng Ruan
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: xinfeng.ruan@otago.ac.nz

Jin E. Zhang
Department of Accountancy and Finance
Otago Business School, University of Otago
Dunedin 9054, New Zealand
Email: jin.zhang@otago.ac.nz

First Version: May 2019
This Version: 29 October 2019

The Implied Volatility Smirk in the Commodity Market

Abstract

This paper studies the implied volatility (IV) smirks in four commodity markets by adopting Zhang and Xiang's (2008) methodology. First, we document the term structure and dynamics of IV smirks. Overall, the commodity IV curves are negatively skewed with a positive curvature. Then we analyze the commodity and S&P 500 returns predictability based on in-sample and out-of-sample tests and find that the information embedded in IV smirks can significantly predict both monthly commodity and S&P 500 returns. For example, the in-sample and out-of-sample R^2 values of the crude oil IV slope for predicting the S&P 500 returns are 3.25% and 8.75%, respectively.

Keywords: Implied volatility smirks; commodity markets; return predictability; S&P 500

JEL Classifications: G13, G12.

1 Introduction

This paper investigates the term structure and time-series dynamics of implied volatility (IV) smirks based on commodity options, and studies the predictability of the excess returns on the underlying assets by using the information embedded in IV smirks. Recently, an extensive literature has focused on commodity markets, as trading in commodity production and the corresponding derivatives have increased massively along with increasing participation by investors. Studies on the commodity market based on options are limited, and most research focuses on the variance risk premium (VRP) investigating its predictive performance for commodity futures return. We first adopt and expand the approach developed by Zhang and Xiang (2008) to quantify the shape of the IV smirk and to examine the term structure and time-series dynamics of the quantified IV factors for commodity exchange-traded fund (ETF) options. The commodity ETFs are the United States Oil Fund (USO), the United States Natural Gas Fund (UNG), the SPDR Gold Trust (GLD) and the iShares Silver Trust (SLV), respectively. Then we investigate the excess return predictability on commodity ETFs by using the IV factors based on in-sample and out-of-sample regressions. Finally, we also analyze the S&P 500 excess return predictability by using the predictive variables of the commodity markets.

Commodities play a vital role in the economic development of a country, and derivatives based on commodities act as effective financial instruments with the main objective of minimizing the risks arising on account of price fluctuations. Commodity derivatives trading on exchanges around the world has shown rapid growth in recent years. In 2018, the number of commodity options and futures contracts was 5.92 billion. Options based on commodity markets have become increasingly popular, because they are a low-cost tool for hedging portfolio risks. Options markets also include information flow between the returns and the volatility of the underlying assets. We focus on the energy and precious metals markets since they are the most important commodity sectors and have the most liquid options markets.

Most recent studies related to commodity markets based on options concentrate on VRP. A large literature performs an empirical analysis on a negative relationship between VRP and expected futures returns in various markets, such as crude oil and natural gas (e.g., Trolle and Schwartz, 2010; Kang and Pan, 2015), corn (e.g., Wang et al., 2012) and energy and precious metals ETFs (e.g., Tee and Ting, 2017). Many studies have verified that VRP contains predicative information about commodity futures prices for most commodity markets (e.g., Kang and Pan, 2015; Fajardo, 2017). There is a handful of studies investigating the high-order risk-neutral moments in the commodity market. Ruan and Zhang (2018) study the return predictability for the crude oil market by using risk-neutral moments and differences in them. Chatrath et al. (2016) examine the information content of risk-neutral moments to explain crude oil futures returns.

To our knowledge, research about the IV smirk in commodity markets is limited. Soini and Lorentzen (2019) study the IV smile for crude oil by a second-order polynomial regression model and analyze correlations between the estimated coefficients and explanatory variables. However, we extend the research into four commodity markets (crude oil, natural gas, gold and silver, respectively) and provide a more comprehensive study on the term structures and dynamics of the IV shape. Moreover, we focus on the excess return predictability of estimated parameters rather than the determinants of these coefficients.

In this paper, we apply the methodology proposed by Zhang and Xiang (2008) to quantify the IV smirks of four commodity ETF options. Then we examine the term structure of the IV shape using maturity categories. We also investigate the dynamics of the quantified IV parameters to draw conclusions on how the commodity ETF options market behaves. In line with Ruan and Zhang (2018), we run predictive regressions to predict the monthly commodity ETF excess returns by using IV shape factors and the S&P 500 excess return predictability by using the predictive variables from the four commodity markets based on in-sample and out-of-sample tests.

The main findings can be summarized as follows: 1) Overall, the IV curves of the four commodity markets are negatively skewed with a positive curvature. 2) The information embedded in the commodity IV smirks can significantly predict both monthly commodity and S&P 500 returns based on in-sample and out-of-sample tests. For example, the in-sample and out-of-sample R^2 values of the crude oil IV slope for predicting S&P 500 returns are 3.25% and 8.75%, respectively. In addition, the information from the gold IV curves exhibits the best in-sample predictive performance. For example, the first differences of the risk-neutral fourth cumulant (DFC) from the gold market has the highest R^2 statistic value of 11.43% for GLD returns predictability and 16.53% for S&P 500 returns predictability.

The remainder of this paper is organized as follows: Section 2 summarizes the literature. Section 3 presents our data. Section 4 specifies the methodology for quantifying the IV curve and for predicting excess returns. Section 5 presents the empirical results on the term structures and dynamics of the IV smirks and return predictability for commodity ETFs and S&P 500, and Section 6 concludes.

2 Literature review

Our paper contributes to several strands of literature. First, we extend the study of the IV smirk. Many studies have documented the IV skewed to the left since the global market crash in 1987 (e.g., Rubinstein, 1994; Corrado and Su, 1997; Carr and Wu, 2003). There is also a handful of literature explaining the shape of the IV curve (e.g., Dennis and Mayhew, 2002; Pan, 2002; Bollen and Whaley, 2004). Based on different approaches, one strand of research examines the dynamics or term structures of the IV smirk of various options, such as the S&P 500 index option (e.g., Cont et al., 2002; Carr and Wu, 2003;

Christoffersen et al., 2009; Fajardo, 2017), the Financial Times Stock Exchange/Xinhua China 50 Index ETF (FXI) option (e.g., Li et al., 2019) and the S&P 500 Short-Term VIX Futures Index exchange-traded note (VXX) option (e.g., Gehricke and Zhang, 2019). However, related research about the IV smirk in commodity markets is somewhat limited. Soini and Lorentzen (2019) investigate the relationship between IV and moneyness by using a second-order polynomial and mainly study the potential determinants of the volatility smile in the crude oil market, but they do not provide a more elaborate discussion on the coefficients of volatility smiles. In this paper, we extend the research on IV curves to four commodity markets and use the information contained in IV smirks to see whether returns of the underlying assets can be predicted.

To investigate the information embedded in IV smirks, a large literature has examined the risk-neutral standard deviation, skewness and excess kurtosis using model-based as well as model-free approaches. Within the model-based methods, jump-diffusion models (e.g., Merton, 1976; Yan, 2011) and stochastic volatility models (e.g., Heston, 1993; Hull and White, 1987; Stein and Stein, 1991) are the most common approaches. There is also a vast strand of literature that extends the stochastic volatility model by incorporating different types of jumps (e.g., Bakshi et al., 1997; Pan, 2002; Eraker, 2004). In contrast to model-based approaches, the model-free methods can reduce the measurement error resulting from model misspecification and can calculate the risk-neutral moments from option prices easily. An extensive empirical literature documents the high-order risk-neutral moments using the model-free methodology suggested by Bakshi et al. (2003) (BKM) (e.g., Dennis and Mayhew, 2002; Neumann and Skiadopoulos, 2013; Bali and Murray, 2013).

However, as we know, options data do not contain a continuum of strike prices. Therefore, when we calculate risk-neutral moments by the BKM method, bias may be introduced into our estimate (e.g., Dennis and Mayhew, 2002; Neumann and Skiadopoulos, 2013). As Dennis and Mayhew (2002) point out, the discreteness of the strike price interval and asymmetry in the domain of integration are the two main causes. For some less active option markets with a smaller range of strike prices, the BKM method may not be suitable. To address this issue, we adopt the model-free approach proposed by Zhang and Xiang (2008) to examine the level, slope and curvature of the IV smirk, which are good proxies of the risk-neutral moments. This method uses a second-order polynomial function to quantify the IV curve, which cannot incur the bias mentioned before.

Second, we contribute to the literature on the commodity market based on options. Most studies focus on VRP to describe the commodity markets. For instance, Trolle and Schwartz (2010) investigate VRP for crude oil and natural gas by using a robust model-independent approach used by Carr and Wu (2009) and then compare results with results on the S&P 500 Index. Wang et al. (2012) find negative and time-varying VRP in the corn market by employing a synthesized model-free approach used by Carr and Wu (2009). Tee and Ting (2017) focus on VRP in four commodity ETFs of gold, silver, natural gas, and crude oil. Kang and Pan (2015) measure VRP using options and high-frequency futures data in the crude oil market, and find a negative relationship between VRP and expected futures returns. Different from the above studies regarding VRP, we newly use the quantified IV smirk proposed by Zhang and Xiang (2008) to describe commodity markets.

Third, this paper is related to the studies on the predictability of returns in commodity markets. Kang and Pan (2015) and Da Fonseca and Xu (2017) examine the return predictability of VRP on oil future returns in the crude oil market. Huang and Kilic (2019) investigate the ratio of gold to platinum price (GP) as a significant economic state variable to predict future stock returns. Christoffersen and Pan (2018) present oil volatility as a strong predictor of monthly returns and volatility of overall stock market. Ruan and Zhang (2018) study the United States Oil Fund (USO) return predictability using higher-order risk-neutral moments (RNMs) and differences in RNMs. Ruan and Zhang (2019) investigate the predictability of the six energy-related market returns by using the moment spreads. In this paper, we are the first to adopt quantified IV smirk factors to predict the returns of four commodity ETFs.

Finally, our empirical work also contributes to the existing literature on the S&P 500 return predictability. Welch and Goyal (2007) find that most macroeconomic predictors of the equity premium are not robust and have poor out-of-sample performance. Rapach et al. (2010) document that combining individual forecasts has much better out-of-sample forecasting accuracy. Bollerslev et al. (2014) verify the short-run predictability of VRP and further show that the aggregate worldwide VRP exhibits stronger predictability. Recently, Jondeau et al. (2019) have argued that the average skewness of a firm's returns has the best predictive performance for predicting market excess returns compared with the macroeconomic variables used by Welch and Goyal (2007) and the financial variables that capture the aggregate risk or fragility of financial market. There is limited research related to using predictive variables of commodity markets to predict S&P 500 index returns.[1] We fill this gap by examining the predictive power of the predictors from commodity markets for S&P 500 returns.

3 Data

3.1 Commodity ETFs

We obtain the commodity ETF data from two sources, Bloomberg and OptionMetrics, from 9 May 2007 to 29 December 2017. As the two commodity ETFs, UNG and SLV, have been split their shares, when we calculate the monthly ETF returns, it is necessary to use split-adjusted historical ETF prices in order to remove gaps caused by splits. The ETF data with adjustments for splits are obtained from Bloomberg. When we calculate forward

[1] Christoffersen and Pan (2018) and Huang and Kilic (2019) investigate the predictive power of the oil price volatility risk and the ratio of gold to platinum price for the overall stock market returns, respectively. They both use the returns on the Center for Research in Security Prices (CRSP) value-weighted index as U.S. stock returns.

prices, we need historical ETF prices without adjustments for splits. This is consistent with historical options data that are unadjusted. We hence obtain the unadjusted commodity ETF data from OptionMetrics Ivy DB. All underlying commodities are the ETFs quoted on the New York Stock Exchange (NYSE).

Table 1 reports summary statistics of the four commodity ETFs. GLD, with the earliest inception date, has the largest total assets of 30 billion dollars and average daily volume of 702 million dollars, as of 28 May 2019. In other words, GLD is the largest and most active commodity ETF. The second most liquid fund is USO with 276 million dollars, even though it does not have a tremendous amount of assets. SLV, the second largest of the ETFs, has less than half of the average daily volume of USO. The UNG ETF is the smallest fund compared with the others.

$<$ *Insert Table 1 about here* $>$

3.2 Commodity ETF returns

Based on Ruan and Zhang (2018), Jondeau et al. (2019) and others, the predictability of the future excess returns on commodity ETFs is measured at a monthly frequency. We define the excess return as follows:

$$ExR_t = \ln S_t - \ln S_{t-1} - r_t, \tag{1}$$

where S_t is the ETF price at the end of month t and r_t is the one-month U.S. Treasury yields obtained from the website of the U.S. Department of the Treasury.

Panel A of Table 2 presents the summary statistics of excess returns in four commodity markets. In general, the average monthly excess returns on the USO and UNG are negative, while those on the GLD and SLV are positive. USO and UNG track the prices of near-month futures contracts on WTI crude oil and natural gas, and the crude oil and natural gas markets have historically experienced long periods of contango. Therefore, when the USO and UNG have to roll their length in the futures they have to sell the cheaper spot contract and buy the more expensive second-month contract. Over the long term, because of the negative roll yield the fund investors experience losses. Gold and silver are considered to be safe haven assets, and their values are generally on the rise in the long run. GLD and SLV track the gold and silver spot price. Therefore, GLD and SLV investors may have positive mean returns. In terms of skewness, all ETFs are negative with an ascending order, which indicates their return distribution has a longer left tail in physical measure. USO has the most negative skewness (-0.7431), which points to the more extreme negative excess returns observed in the crude oil market. From Panel B, we find strong evidence that the excess returns on GLD are highly correlated to SLV excess returns at a value of 83% in precious metal markets, while in energy commodity markets, the correlation between USO and UNG is relatively low (i.e., 0.20).

< Insert Table 2 about here >

3.3 Commodity ETF options

The options data based on commodity ETFs (USO, UNG, GLD and SLV, respectively) are obtained from OptionMetrics Ivy DB. Because the four options have different inception dates (i.e., 9 May 2007, 9 May 2007, 3 June 2008 and 8 December 2008, respectively), the corresponding sample periods are from the start date of trading to 29 December 2017. All options on commodity ETFs are traded on the Chicago Board Options Exchange (CBOE).

Following Bakshi et al. (1997), Zhang and Xiang (2008) and Neumann and Skiadopoulos (2013), we apply several filters to the options data. First, we discard the options with less than seven days to expiration. Second, the options with zero bid price or zero open interest are also removed. Furthermore, we discard options for which the Ivy DB does not provide implied volatilities. Finally, we delete the maturities with less than five nonzero trading volumes on each day.

Table 3 reports the trading summary of the four commodity ETF options by maturity categories after cleaning the data. The statistical variables are mean number of strikes, number of observations, mean daily trading volume and mean daily open interest, respectively. Overall, the GLD options market has the largest value in terms of the four statistical variables, which indicates the GLD ETF has the most liquid options market. For USO and SLV options, on average, the mean number of strikes of USO are slightly larger than that of SLV, while the other three statistical variables of USO are smaller than that of SLV. For the options on UNG, which is the smallest ETF, they are the most inactive, with the mean daily number of strikes for each maturity of only 14. In addition, we find the number of strikes for all ETF options are smaller for a maturity of less than 30 days, when comparing with other maturity groups. That is because we have deleted the options data for less than seven days. Across the maturity group from 30-90 to >360, the trading volume and open interest of all options markets appear to be decreasing. For maturities less than 90 days, the trading volume and the open interest account for a huge proportion of the total value. That indicates the shorter the time to maturity, the higher the liquidity. Options investors prefer to trade options for short-term profits.

< Insert Table 3 about here >

4 Methodology

In this section, we introduce the methodology developed by Zhang and Xiang (2008) to quantify the IV curve by fitting a quadratic function, and then we describe the in-sample and out-of-sample prediction methodology, respectively.

4.1 Quantifying IV curve

Zhang and Xiang (2008) propose a simple and intuitive method to quantify the IV smirk. First, we define the moneyness of an option as follows:

$$\xi = \frac{\ln(K/F_{t,T})}{\bar{\sigma}\sqrt{\tau}}, \tag{2}$$

where K is the strike price, $F_{t,T}$ is the implied forward price for maturity T on day t, τ is the annualized time to maturity, and $\bar{\sigma}$ is an average IV. We obtain standardized options from OptionMetrics, and then calculate the mean of 30-day volatilities of at-the-money-forward call and put for each day as a proxy for $\bar{\sigma}$.

The definition of moneyness above is in line with the standard measure proposed by Carr and Wu (2003). The normalization with the square root of time to maturity can eliminate the effect of different maturities. The use of the constant $\bar{\sigma}$ is an industry convention and can make the transparent link between the slope and curvature of the IV smirk and the skewness and kurtosis of the risk-neutral distribution. More discussion about this can be found in Carr and Wu (2003) and Bedendo and Hodges (2009).

As USO, UNG, GLD and SLV do not pay a dividend, under no arbitrage, the forward price of a non-dividend-paying asset S_t is defined as

$$F_{t,T} = S_t e^{r(T-t)}, \tag{3}$$

where $F_{t,T}$ is the implied forward price for maturity T on day t, S_t is the price of the underlying ETF at time t and r is the risk-free rate obtained by linearly interpolating and extrapolating the U.S. Treasury yield rate at time t. The daily U.S. Treasury yield data are downloaded from the website of the U.S. Department of the Treasury. As discussed in subsection 3.1, S_t is the ETF price without adjustments for splits provided by Bloomberg.

Based on the moneyness defined in Equation (1), we can apply a second-order polynomial to describe the IV-moneyness function:

$$IV(\xi) = \alpha_0 + \alpha_1 \xi + \alpha_2 \xi^2, \tag{4}$$

where α_0, α_1 and α_2 are the coefficients of the above regression.

We can further convert these coefficients to a dimensionless form as follows

$$IV(\xi) = \gamma_0(1 + \gamma_1 \xi + \gamma_2 \xi^2), \tag{5}$$

where

$$\gamma_0 = \alpha_0, \quad \gamma_1 = \frac{\alpha_1}{\alpha_0}, \quad \gamma_2 = \frac{\alpha_2}{\alpha_0}.$$

The three factors are called the level, slope and curvature of the IV smirk, respectively. When $\xi = 0$, that is moneyness is equal to zero, γ_0 is taken to be the estimated ATM IV, which is slightly different to the ATM IV from the market data.

In line with Zhang and Xiang (2008), to obtain the three coefficients, we use a volume-weighted least squares method (VWLS) and minimize the volume-weighted mean square error

$$VWMSE = \frac{\sum_{\xi} Volume \times [IV_{market}(\xi) - IV(\xi)]^2}{\sum_{\xi} Volume}, \tag{6}$$

where $Volume$ is the trading volume and $IV_{market}(\xi)$ is the market IV, on a certain day for a given maturity. Considering liquidity and sensitivity to the model, we follow Carr and Wu (2003) and pick out-of-the-money (OTM) options to estimate the IV function. We choose call options whose strike prices are above the forward price, and choose put options whose strike prices are below the forward price.

The reason we adopt VWLS is that we focus our attention on more liquid options with a large trading volume that would contain more important information. However, if a small number of options contracts have relatively huge trading volume for a certain day and maturity, the fitted curve is likely to show a reverse trend. The unusual fitting may cause abnormal statistics. Therefore, we take ordinary least squares (OLS) as a complementary method to address this problem.[2] First, we calculate the R squared based on VWLS and OLS. The R squared, denoted by R^2, is

$$R^2 = 1 - \frac{\sum_i (IV_{market} - \hat{IV})^2}{\sum_i (IV_{market} - \overline{IV})^2}, \tag{7}$$

where $\hat{IV}$ is an estimate of IV and $\overline{IV}$ is sample mean within a maturity group for a given day. We then compare R^2_{VWLS} and R^2_{OLS} which present R^2 obtained by VWLS and OLS. In general, R^2_{OLS} is larger than R^2_{VWLS} but OLS cannot place emphasis on the more active options with a large trading volume we focus on. Therefore, as mentioned above, VWLS is our main method, while OLS is used to replace VWLS with extremely poor fitting. That is, only if R^2_{VWLS} is still smaller than R^2_{OLS} after adding a value bigger than zero do we use OLS to fit the function instead of VWLS. The value is set 1.[3] The sample using OLS fitting accounts for 9.8%, 16.18%, 5.86% and 3.39% of the whole sample for the crude oil, gas, gold and silver markets, respectively.

4.2 Measuring predictability

To examine whether the information contained in the IV smirks of commodity options can predict the excess returns on commodity ETFs, we run the monthly return predictability regression:

$$ExR_{t+1} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{8}$$

where ExR_{t+1} is the excess return in time $t+1$, X_t is a set of predictive variables in time t and ϵ_{t+1} is the residual.

[2]Our measure is different from Soini and Lorentzen (2019) who do regressing IV on moneyness (the future price divided by the strike price) by OLS.

[3]Under the case that the fitted IV curves do not match the market data very well, R^2_{VWLS} can take negative values. While R^2_{OLS} is normally in the 0–1 range, by setting the value of 1, we can use OLS fitting to replace most unusual VWLS fittings with a negative R squared.

We use nonoverlapping excess returns to estimate Equation (8) and report the slope coefficient estimate $\hat{\beta}$, adjusted R^2 statistics and the Newey and West (1987) t-statistics using optimal lag length.

Following Welch and Goyal (2007), Neumann and Skiadopoulos (2013) and Rapach et al. (2016), we also investigate the performance of predictive variables in terms of out-of-sample prediction. First, half of the total number of samples are chosen as the initial observations for first forecast. Then using a sequence of expanding windows, a series of out-of-sample excess return forecasts can be obtained. The out-of-sample R^2 is given by

$$R^2_{OS} = 1 - \frac{\sum_{t=n}^{N-1}(r_{t+1} - \hat{r}_{t+1|t})^2}{\sum_{t=n}^{N-1}(r_{t+1} - \overline{r}_{t+1|t})^2}, \tag{9}$$

where $\overline{r}_{t+1|t} = \frac{1}{t} \sum_{i=1}^{t} r_i$ means the historical average of excess returns up to the time t.

In line with Clark and West (2007), the adjusted mean squared prediction error statistic is defined by

$$f_{t+1} = (r_{t+1} - \overline{r}_{t+1|t})^2 - [(r_{t+1} - \hat{r}_{t+1|t})^2 - (\overline{r}_{t+1} - \hat{r}_{t+1|t})^2], \tag{10}$$

we regress this statistic on a constant and provide one-sided p-values for the R^2_{OS} statistic.

5 Empirical Results

5.1 Quantified IV curves

In this subsection, we examine and discuss the fitted IV curves of four commodity options markets. Figures 1–4 show the fitted IV curves, the market IVs and the trading volumes for all maturities on the randomly selected day (i.e., 27 December 2017) for the USO, UNG, GLD and SLV option markets, respectively. For all markets, the IV curves seem to be fitted well with the market IV.

The IV curve of USO options shows a smirk pattern (skewed to the left), which is in line with the findings of the S&P 500 smirk shape in Carr and Wu (2003). With a volatility smirk, the implied volatility rises faster at low strike prices than at high strike prices (e.g., Jiang and Tian, 2005). This reflects investors' fear of downward market price jumps. In contrast, the IV curve of SLV options displays another smirk pattern (skewed to the right), called a forward skew. It exhibits high volatility for high strike prices and low volatility for low strike prices. We also find that for shorter maturities, the UNG and GLD IV curves show the opposite smirk shape, which is a forward skew and smirk, respectively.

Summary results of all factors are reported in Table 4. The summary statistics are provided overall and by maturity categories (i.e., < 30, $30-90$, $90-180$, $180-360$,> 360). First, in Panel A, we can see that the USO IV curves are negatively skewed with a positive curvature, as the slope factor is negative and the curvature is positive on average and over all maturities. The standard deviation of the level decreases as maturity increases, which indicates the estimated ATM IV would mean-revert in the long run. The term structure of the slope is downward sloping, which means smirk slopes become steeper as maturity increases. For the curvature, it seems to be flat as maturity increases. The proportion of significant coefficients of the level is 99.38% overall, slightly higher than that of the slope (93.15%), while that of the curvature is only 88.46%. They tend to be upward sloping with smaller maturities. The mean adjusted R^2 is around 95.20% across all maturity categories.

Turning to the natural gas market, from Panel B, the IV curves are usually negatively sloped with an overall average slope of -0.0016, although the slopes are positive with a small absolute value when the time to maturity is less than 90 days. For the level factor, it seems to be higher than that of the crude oil market, and its standard deviation shows the same trend as that of the USO. Looking at the slope factor, it is positive for maturities less than 90 days and negative for other maturity groups. The curvature is usually positive. Compared with USO, except for level factor, the proportions of the fitted IV curves with significant slope and curvature coefficients are significantly lower. In particular, the slope coefficient is only significant for 62.54% of the IV curves overall. That could be due to the less liquidity with a low daily trading volume for UNG, as we discussed in Section 3.

From Panel C in Table 4, we can see that the IV curves for the GLD options market are negatively skewed with a positive curvature, although they become positively sloped for maturities more than 180 days. The term structures of the level and slope both show an upward trend across maturity categories, while the term structure of the curvature tends to be flat with positive convexity. Different from the above results shown in Panels A and B in Table 4, the standard deviation of the estimated ATM IV fluctuates around 0.0603 and does not show a mean-reverting manner. Overall, the proportions of significant coefficients of the three factors are 100%, 93.33% and 99.80%, respectively. The overall mean adjusted R^2 of 97.51% for GLD is considerably higher than that of USO and UNG.

In the SLV options market, overall, the IV curves exhibit the same trend as that of the three other markets, which is negatively skewed with a positive curvature. The term structures of the level and slope are upward sloping, and the term structure of the curvature shows a downward trend and tends to become less convex across the maturity categories. The standard deviation of level is downward sloping as maturities increase, meaning the ATM IV tends to mean revert consistent with the findings in the USO and UNG options markets. In terms of a significance test, the level (curvature) coefficients are significant for 99.55% (96.21%) of the fitted IV curves overall and the slope coefficient is only significant for 86.48% of the IV curves. The overall mean adjusted R^2 is 98.85%.

< *Insert Table 4 about here* >

To summarize, the IV curves of four commodity options are negatively skewed with a positive curvature as the overall average slope factor is negative and curvature factor is positive. Looking at the level factor first, except for GLD, the standard deviation of the level is downward sloping as maturities increase, which indicates the level of the three other options markets seems to mean revert. Second, for the energy options, the slope has a downward sloping term structure. By contrast, for the precious metal options, the slope has an opposite term structure. Finally, the IV curves of the GLD options have the largest proportion of significant coefficients of the three factors, while those of the UNG options have the smallest ones. The reason may be that the GLD options market is the most liquid and the UNG is the most inactive, as we discussed in Section 3. The high mean adjusted R^2 of the above 93.62% indicates our quantified IV curve fits the market IV very well.

5.2 Constant maturity IV curve dynamics

We discussed the term structures of the level, slope and curvature for maturity categories in subsection 5.1. In order to further examine the quantified IV curves, we document the time series dynamics of the three factors with 30-day and 180-day constant maturities. First, using interpolation and extrapolation, we obtain the 30-day and 180-day constant maturity level, slope and curvature factors, respectively. Then we present the time series of the three factors in Panels a, c and e in Figures 5–8 and the difference between the 30-day and 180-day constant maturity factors in Panels b, d and f in Figures 5–8. Finally, we show the comparison of the dynamics of the three factors for all four markets in Figure 9.

Figure 5 shows the time series of 30- and 180-day constant maturity IV curves in the crude oil market. We can see that there are large spikes over the sample period. During the financial crisis of 2008, the USO volatility reached a record high of close to 100%, and then was followed by an unprecedented decrease. After a long recovery, it rose to a relatively large level during the European debt crisis in 2011. From the end of 2014, the USO volatility experienced a rapid increase, reaching to the second-highest value in 2016. This is because the crude oil price went through a collapse continuing into early 2016. We also noticed that the difference between the 30-day and 180-day estimated ATM IV is huge when the volatility level is at a large spike. In terms of slope, it is usually negative and fluctuates a lot more at the economic events mentioned above. The 180-day slope, especially, seems to be more negative. The curvature factor is usually positive.

$<$ *Insert Figure 5 about here* $>$

With respect to the natural gas market, the 30- and 180-day level factors seem to fluctuate more violently than those of the crude oil market shown in Figure 6a. The volatility peak appeared in 2009 September due to continued demand depression from the 2008 global financial crisis. The time series of volatility in the natural gas market displays a major seasonal pattern. In other words, volatility spikes tend to coincide with the rapid increase in the gas price caused by the high natural gas demand in winter. For example, we can find there are two obvious gas volatility spikes in around January 2014 and January 2015, when the winters were both very cold in the United States. From Figure 6b, we can see that the difference between the 30-day and 180-day level exhibits several spikes during the time when the gas volatility level spikes. The slope seems to fluctuate around zero and the curvature is positive at most times.

$<$ *Insert Figure 6 about here* $>$

In general, the dynamics of the level factor in the gold market tends to have a downward sloping term structure during the sample period in Figure 7, with extremely low volatility in 2017. During the 2008 financial crisis, the gold volatility displayed the largest spike, and then it declined rapidly. After a fluctuation period, the volatility spiked at the end of 2011 as the gold price experienced a crash after its 2011 peak. The 2011 European debt crisis may be one of factors in the sharp price drop. In June 2013, the gold price jumped rapidly combined with a spike in volatility. However, during this time, the volatility for crude oil and natural gas fell, but it jumped significantly between 2014 and 2016. From Figure 7b, the level difference is extremely negative when gold volatility spikes, while it is usually positive at other times, which indicates the 180-day level is usually larger than the 30-day one, consistent with the finding that the level term structure of GLD is upward sloping in subsection 5.1. In addition, the curvature is always positive and the slope and curvature for the 180-day maturity tend to fluctuate more frequently than that of the 30-day maturity.

$<$ *Insert Figure 7 about here* $>$

With regard to another important precious metal product, the dynamics of the silver volatility level shows the same trend as that of the gold volatility, on average. The three largest volatility spikes correspond to the events mentioned above. In addition, from the middle of 2016 to the end of 2017, the silver volatility displays a downward-sloping term structure, with extremely low volatility in 2017. The difference between the 30-day and 180-day level is extremely negative when volatility spikes, while it is usually positive at other times, which is consistent with the result for gold. Silver curvature is also positive most of the time.

$<$ *Insert Figure 8 about here* $>$

In sum, for all four markets, the volatility levels exhibit large spikes during the period associated with the market-specific shocks as shown in Figure 9. The consistent results also can be seen in the dynamics of level difference. Specifically, there are more sharp spikes in the natural gas market because natural gas is the most volatile of the four markets, with seasonal winter volatility peaks. Gold, as a safe-haven asset, has the lowest volatility among the four markets. With respect to slope, it fluctuates dramatically at the economic or market-specific events mentioned above. Furthermore, curvatures from the four commodity markets are positive most of the time.

< Insert Figure 9 about here >

5.3 Predictive variables

In line with Dennis et al. (2006) and Xing et al. (2010), we investigate whether the risk-neutral information embedded in the IV smirks of commodity options can predict the excess returns of commodity ETFs, or even predict the returns of the S&P 500.

We employ IV factors at the end of each month (*Level*, *Slope* and *Curvature*), which are interpolated for days between the end of the current month and the end of the predicted month, as predictors for monthly excess return predictability on commodity ETFs.

In addition to the three variables, we also follow Ruan and Zhang (2018) and Zhang et al. (2019) to test the predictive power of the risk-neutral third and fourth cumulants. As shown by Zhang and Xiang (2008), the level, slope and curvature of the IV curve are related to the risk-neutral volatility, the skewness and the excess kurtosis, respectively. Following Ruan and Zhang (2018), we give the two cumulants as $TC{=}\gamma_1 \times \gamma_0^3$ and $FC{=}\gamma_2 \times \gamma_0^4$, where γ_0, γ_1 and γ_2 are interpolated level, slope and curvature, respectively. In line with Ang et al. (2006) and Chang et al. (2013), we also consider the first differences of these variables for predicting excess returns, that is $DLevel$, $DSlope$, $DCurv$, DTC and DFC.

Table 5 reports the summary statistics for 10 predictive variables of the four commodity markets. We find that all mean levels are significantly larger than the standard deviations of the excess returns in Panel A of Table 2. The reason is that they are measured by a different data frequency (i.e., annualized and monthly). However, in the same measurement, we find they are similar. For example, we convert the USO mean level (0.3446) into the standard deviation of monthly return: $0.3446/\sqrt{12} = 0.0995$, which is close to 0.0958.

The mean slopes for all markets are negative (i.e., -0.0589, -0.0011, -0.0174 and -0.0133, respectively), which indicates the excess return distributions for all four commodity markets have negative skewness in the risk-neutral probability measure. For the average curvature, all four markets have positive values at around 0.05. In addition, we provide the statistics for the first differences of the three factors in Table 5. $DLevel$, $DSlope$ and $DCurv$ have a negative mean in all markets except for UNG.

$<$ *Insert Table 5 about here* $>$

Turning to the risk-neutral cumulant predictors, we can see that the means of TC for USO, GLD and SLV are negative, while the mean of TC for UNG is positive. All FC predictors have positive means. In particular, UNG has the largest value of the mean FC, while GLD has the smallest value. Looking at the first differences of TC and FC, in general, UNG has the highest DTC and the lowest DFC.

The correlations among the predictors are reported in Table 6. Overall, for the crude oil and gold markets, *Level* and *Slope* have a slightly positive correlation, while *Level* and *Curv* have a negative correlation. For the natural gas market, both of the two pairs have a small positive correlation. In contrast, for the silver market, they have a negative correlation. According to the cumulants defined above, TC (FC) may be correlated to *Slope* (*Curv*) and *Level*. TC is highly negatively correlated to *Level* with a value of -0.8358 in the crude oil market, but highly positively correlated to *Slope* in other markets. For UNG, FC has a large correlation with *Level* and *Curv*, while for USO, GLD and SLV, FC only has a high correlation with *Level*. With respect to the correlation between TC and FC, it is not very large in absolute value in UNG, GLD and SLV, which indicates TC and FC could deliver different information about the excess returns. However, it is very high at a negative value of -0.8531 in USO. In addition, the correlations between the three factors and their first differences range from 0.2271 to 0.637 for all markets. The correlation between DTC (DFC) and TC (FC) is not very large, in absolute value, except SLV.

< Insert Table 6 about here >

5.4 Predictability of four commodity ETF returns

In this subsection, to verify whether the shape of the IV smirk contains the information related to the underlying ETFs' excess returns, we test the in-sample and out-of-sample predictive power of the 10 variables defined in subsection 5.3 for predicting monthly USO, UNG, GLD and SLV excess returns. Following the methodology provided in subsection 4.2, Table 7 reports the results of the predictability regressions for each of the predictors in four commodity markets.

< Insert Table 7 about here >

Based on in-sample regressions, we find that both FC and $DLevel$ can negatively predict USO excess returns significantly. This result is consistent with the findings in Ruan and Zhang (2018) that FC and the first difference of volatility have significant return predictability with negative slopes. In our test, FC has the higher significance level with a Newey and West (1987) t-statistic of -3.15. However, $DLevel$ has the larger adjusted R^2 statistic of 8.83, although its significance level is slightly lower than FC. For natural gas, FC performs the best among all predictors, with an adjusted R^2 statistic of 3.86% and a t-statistic of -2.40. The result is better than the one Ruan and Zhang (2019) report, that is, the R^2 statistic of the kurtosis spread (KTS) is only 1.76%.

Turning to the precious metal market, DFC exhibits the strongest predictive power for excess returns on GLD and SLV among all variables. The t-statistic of DFC for predicting the GLD (SLV) excess return is 10.76 (-5.08), with a statistical significance at the 1% (1%) level and an adjusted R^2 statistic of 11.43% (1.32%). For GLD, due to the high correlation between DTC and DFC shown in Panel C of Table 6, both DFC and DTC have high adjusted R^2 statistics for excess returns of the gold market. For SLV, FC and DTC also are good predictors for excess returns on the silver market.

To further investigate the forecasting results, we use out-of-sample predictions to examine whether our predictive regression model forecasts outperform the historical average forecasts. We find that some predictors from the energy market have good out-of-sample performance. For the crude oil market, $DLevel$ has good out-of-sample predictive power, with a R^2_{os} statistic of 4.06% and a statistical significance at the 10% level. This is consistent with Ruan and Zhang (2018), who present evidence that the first difference of volatility is also a good predictor, with a high R^2_{os} statistic. For the natural gas market, both $Level$ and FC can significantly predict UNG excess returns, with a R^2_{os} statistic of 3.91% and 3.75%, respectively. Moreover, we notice that all predictors from the gold market have no out-of-sample predictive power, and that FC from the silver market has some out-of-sample performance.

In summary, we can see that the information embedded in IV smirks can significantly predict monthly excess returns in four commodity markets. Based on in-sample tests, the information from gold IV curves has the best predictive performance. For example, DFC from GLD has an extremely large adjusted R^2 statistic value (11.43%). In addition, some predictors from the energy market and the silver market have good out-of-sample predictive power. In particular, the R^2_{os} of the natural gas IV level for predicting excess returns is 3.91%.

5.5 Predictability of S&P 500 returns

Since the financialization of the commodity market, a number of studies have investigated the links between commodities and stock markets (e.g., Creti et al., 2013; Chen, 2010; Ding et al., 2016). As shown in Chiang et al. (2015) and Christoffersen and Pan (2018), option-implied information based on commodities is significantly related to the expected stock return. Furthermore, Huang and Kilic (2019) suggest that commodity prices could reflect macroeconomic risk, so that the information embedded in commodity prices could predict U.S. stock market returns. Therefore, it is interesting to study the role of the information contained in the IV smirks of commodity options in forecasting stock returns.

We use S&P 500 index returns (from OptionMetrics Ivy DB) to present the U.S. stock returns and analyze the predictability of S&P 500 excess returns by using the predictive variables from four commodity markets at a monthly frequency. Following the methodology provided in subsection 4.2, the slope coefficients $\hat{\beta}$, their Newey and West (1987) t-statistics, in-sample adjusted R^2 statistics and out-of-sample adjusted R^2 statistics R^2_{os} are reported in Table 8.

< Insert Table 8 about here >

Panel A of Table 8 reveals that both *FC* and *DTC* from the crude oil market have strong predictive power for S&P 500 excess returns in the in-sample test. *FC* has the higher significance level, with a Newey and West (1987) t-statistic of -5.06 and an in-sample adjusted R^2 statistic of 4.88%, respectively, while *DTC* has the larger R^2 statistic value (6.46%). In terms of out-of-sample results, both *Slope* and *FC* outperform the other predictors for excess returns on the S&P 500. Their monthly R^2_{os} statistics are 8.75% and 8.77%, respectively, and significant, according to the Clark and West (2007) statistic.

Turning to the natural gas market, only *DSlope* shows some in-sample predictive performance for monthly S&P 500 excess returns. However, based on the out-of-sample R^2_{os} statistics, it exhibits significant predictive power, with a large R^2_{os} statistic of 8.01%. That means they outperform the historical average with respect to forecasting S&P 500 returns. As mentioned in Welch and Goyal (2007), it is not meaningful to discuss the statistically significant out-of-sample performance of *Slope*, *DLevel* and *DFC*, since the three predictors have no in-sample performance.

From Panel C of Table 8, we find that *TC*, *FC*, *DTC* and *DFC* from the gold market can strongly predict monthly S&P 500 excess returns. Specially, *DFC* has the largest t-statistic value (20.26) and adjusted R^2 statistic value (16.53%). Huang and Kilic (2019) empirically show that the ratio of gold to platinum prices (GP) is a strong predictor of stock returns. However, for a one-month horizon, the R^2 statistic of GP is just above 1%, far below our statistical value of 16.53%. With respect to the out-of-sample measure, all predictive variables from the gold market have no significant predictions of the future S&P 500 returns.

In the silver market, *DTC* is the best predictor for predicting S&P 500 excess returns with a R^2 statistic of 4.61 and a significance level of 1%. In addition, *TC*, *DFC* and *DLevel* also show good predictive power. However, in terms of the out-of-sample test, all predictors show poor predictive performance for excess returns on the S&P 500 index and cannot beat the historical average.

To summarize, we find that the information contained in IV smirks from commodity markets has good predictive power for monthly excess S&P 500 returns. For example, the in-sample and out-of-sample R^2 of the crude oil IV slope for predicting the S&P 500 returns are 3.25% and 8.75%, respectively. In addition, *DFC* from the gold market has the largest R^2 statistic value (16.53%). This result is superior to that of papers predicting S&P 500 returns by using 14 economic variables, volatility skew or other predictors (e.g., Welch and Goyal, 2007; Rapach et al., 2016; Huang and Kilic, 2019; Jondeau et al., 2019).[4] We also find that only some predictors from the energy market have significant out-of-sample predictive performance for S&P 500 returns.

6 Conclusion

In this paper, we newly apply the methodology proposed by Zhang and Xiang (2008) to quantify the IV smirks of four commodity ETF options. Then we examine the term structure and dynamics of IV shapes. In line with Ruan and Zhang (2018), we investigate the excess return predictability of the information embedded in the IV smirks at a monthly frequency based on in-sample and out-of-sample tests in four commodity markets. We also use the predictive variables from commodity markets to predict the S&P 500 excess return.

First, empirical evidence shows that the IV curves of four commodity markets are negatively skewed with a positive curvature, in general. Furthermore, the high mean adjusted R^2 of above 93.62% indicates our quantified IV curve fits the market IV very well. In terms of the dynamics of IV smirks, we find natural gas is the most volatile of the four markets, with seasonal winter volatility peaks, and that gold has the lowest volatility among four markets.

Second, the information embedded in IV smirks can significantly predict monthly excess returns on four commodity markets. Based on in-sample tests, the information from gold IV curves has the best predictive performance. For example, DFC from GLD has an extremely large adjusted R^2 statistic value (11.43%). In addition, some predictors from the energy market and the silver market have good out-of-sample predictive power.

Finally, the predictive variables from commodity markets also have good predictive power for monthly excess S&P 500 returns. For example, the in-sample and out-of-sample

[4]Huang and Kilic (2019) and Jondeau et al. (2019) adopt the returns on CRSP value-weighted index as stock market returns. This index is strongly correlated to S&P 500 index.

R^2 statistics of the crude oil IV slope for predicting the S&P 500 returns are 3.25% and 8.75%, respectively. Furthermore, DFC from the gold market has the largest R^2 statistic value (16.53%). This result is superior to that of papers predicting S&P 500 returns by using 14 economic variables, volatility skew or other predictors (e.g., Welch and Goyal, 2007; Rapach et al., 2016; Huang and Kilic, 2019; Jondeau et al., 2019).

References

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003–2049.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bali, Turan G, and Scott Murray, 2013, Does risk-neutral skewness predict the cross-section of equity option portfolio returns? *Journal of Financial and Quantitative Analysis* 48, 1145–1171.

Bedendo, Mascia, and Stewart D Hodges, 2009, The dynamics of the volatility skew: A kalman filter approach, *Journal of Banking and Finance* 33, 1156–1165.

Bollen, Nicolas PB, and Robert E Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance* 59, 711–753.

Bollerslev, Tim, James Marrone, Lai Xu, and Hao Zhou, 2014, Stock return predictability and variance risk premia: statistical inference and international evidence, *Journal of Financial and Quantitative Analysis* 49, 633–661.

Carr, Peter, and Liuren Wu, 2003, The finite moment log stable process and option pricing, *Journal of Finance* 58, 753–777.

Carr, Peter, and Liuren Wu, 2009, Stock options and credit default swaps: A joint framework for valuation and estimation, *Journal of Financial Econometrics* 8, 409–449.

Chang, Bo Young, Peter Christoffersen, and Kris Jacobs, 2013, Market skewness risk and the cross section of stock returns, *Journal of Financial Economics* 107, 46–68.

Chatrath, Arjun, Hong Miao, Sanjay Ramchander, and Tianyang Wang, 2016, An examination of the flow characteristics of crude oil: Evidence from risk-neutral moments, *Energy Economics* 54, 213–223.

Chen, Shiu-Sheng, 2010, Do higher oil prices push the stock market into bear territory? *Energy Economics* 32, 490–495.

Chiang, I-Hsuan Ethan, W Keener Hughen, and Jacob S Sagi, 2015, Estimating oil risk factors using information from equity and derivatives markets, *Journal of Finance* 70, 769–804.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* 55, 1914–1932.

Christoffersen, Peter, and Xuhui Nick Pan, 2018, Oil volatility risk and expected stock returns, *Journal of Banking and Finance* 95, 5–26.

Clark, Todd E, and Kenneth D West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Cont, Rama, José Da Fonseca, et al., 2002, Dynamics of implied volatility surfaces, *Quantitative Finance* 2, 45–60.

Corrado, Charles J, and Tie Su, 1997, Implied volatility skews and stock index skewness and kurtosis implied by S&P 500 index option prices, *Journal of Derivatives* 4, 8–19.

Creti, Anna, Marc Joëts, and Valérie Mignon, 2013, On the links between stock and commodity markets' volatility, *Energy Economics* 37, 16–28.

Da Fonseca, José, and Yahua Xu, 2017, Higher moment risk premiums for the crude oil market: A downside and upside conditional decomposition, *Energy Economics* 67, 410–422.

Dennis, Patrick, and Stewart Mayhew, 2002, Risk-neutral skewness: Evidence from stock options, *Journal of Financial and Quantitative Analysis* 37, 471–493.

Dennis, Patrick, Stewart Mayhew, and Chris Stivers, 2006, Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon, *Journal of Financial and Quantitative Analysis* 41, 381–406.

Ding, Haoyuan, Hyung-Gun Kim, and Sung Y Park, 2016, Crude oil and stock markets: Causal relationships in tails? *Energy Economics* 59, 58–69.

Eraker, Bjørn, 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices, *Journal of Finance* 59, 1367–1403.

Fajardo, José, 2017, A new factor to explain implied volatility smirk, *Applied Economics* 49, 4026–4034.

Gehricke, Sebastian A, and Jin E Zhang, 2019, The implied volatility smirk in the VXX options market, *Applied Economics* (forthcoming).

Heston, Steven L, 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.

Huang, Darien, and Mete Kilic, 2019, Gold, platinum, and expected stock returns, *Journal of Financial Economics* 132, 50–75.

Hull, John, and Alan White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281–300.

Jiang, George J, and Yisong S Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305–1342.

Jondeau, Eric, Qunzi Zhang, and Xiaoneng Zhu, 2019, Average skewness matters, *Journal of Financial Economics* 134, 29–47.

Kang, Sang Baum, and Xuhui Nick Pan, 2015, Commodity variance risk premia and expected futures returns: Evidence from the crude oil market, *Available at SSRN 2296932* .

Li, Jianhui, Sebastian A Gehricke, and Jin E Zhang, 2019, How do US options traders "smirk" on China? Evidence from FXI options, *Journal of Futures Markets* 39, 1450–1470.

Merton, Robert C, 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3, 125–144.

Neumann, Michael, and George Skiadopoulos, 2013, Predictable dynamics in higher-order risk-neutral moments: Evidence from the S&P 500 options, *Journal of Financial and Quantitative Analysis* 48, 947–977.

Newey, Whitney K, and KD West, 1987, A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix, *Econometrica* 55, 703–708.

Pan, Jun, 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Ruan, Xinfeng, and Jin E Zhang, 2018, Risk-neutral moments in the crude oil market, *Energy Economics* 72, 583–600.

Ruan, Xinfeng, and Jin E Zhang, 2019, Moment spreads in the energy market, *Energy Economics* 81, 598–609.

Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 49, 771–818.

Soini, Vesa, and Sindre Lorentzen, 2019, Option prices and implied volatility in the crude oil market, *Energy Economics* 83, 515–539.

Stein, Elias M, and Jeremy C Stein, 1991, Stock price distributions with stochastic volatility: an analytic approach, *Review of Financial Studies* 4, 727–752.

Tee, Chyng Wen, and Christopher Ting, 2017, Variance risk premiums of commodity ETFs, *Journal of Futures Markets* 37, 452–472.

Trolle, Anders B, and Eduardo S Schwartz, 2010, Variance risk premia in energy commodities, *Journal of Derivatives* 17, 15–32.

Wang, Zhiguang, Scott W Fausti, and Bashir A Qasmi, 2012, Variance risk premiums and predictive power of alternative forward variances in the corn market, *Journal of Futures Markets* 32, 587–608.

Welch, Ivo, and Amit Goyal, 2007, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao, 2010, What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis* 45, 641–662.

Yan, Shu, 2011, Jump risk, stock returns, and slope of implied volatility smile, *Journal of Financial Economics* 99, 216–233.

Zhang, Jin E, Eric C Chang, and Huimin Zhao, 2019, Market excess returns, variance and the third cumulant, *International Review of Finance* (forthcoming).

Zhang, Jin E, and Yi Xiang, 2008, The implied volatility smirk, *Quantitative Finance* 8, 263–284.

Table 1: Summary statistics for commodity ETFs

This table reports the summary statistics for the four commodity ETFs (i.e., USO, UNG, GLD and SLV) as of 28 May 2019.

Symbol	ETF Name	Issuer	Inception Date	Total Assets ($B)	Average Daily Volume ($M)
USO	United States Oil Fund	United States Commodity Funds	10 April 2006	1.48	276.32
UNG	United States Natural Gas Fund	United States Commodity Funds	18 April 2007	0.26	25.99
GLD	SPDR Gold Trust	State Street Global Advisors	18 November 2004	30.48	702.39
SLV	iShares Silver Trust	BlackRock	21 April 2006	4.51	96.47

Table 2: Descriptive statistics for excess returns

This table gives summary statistics and correlation coefficients for monthly excess returns on the four commodity ETFs. The USO, UNG, GLD and SLV ETF data are downloaded from Bloomberg for the period from 9 May 2007 to 29 December 2017. The monthly excess return is the return on ETFs in excess of the risk-free rate.

Panel A: Summary statistics						
Excess Returns	Mean	Std.dev.	Skewness	Kurtosis	Min	Max
USO	-0.0161	0.0958	-0.7431	4.5531	-0.3900	0.2387
UNG	-0.0366	0.1108	-0.6561	3.6966	-0.4063	0.2243
GLD	0.0006	0.0549	-0.2104	3.2786	-0.1772	0.1195
SLV	0.0014	0.0948	-0.0736	3.8264	-0.3361	0.2427

Panel B: Correlations				
	USO	UNG	GLD	SLV
USO	1.00			
UNG	0.20	1.00		
GLD	0.16	-0.03	1.00	
SLV	0.31	0.01	0.83	1.00

Table 3: Summary of the four commodity ETF options

This table shows the number of observations, the mean daily number of strikes, trading volume and open interest of the four commodity ETF options overall and for each maturity category after cleaning the options data.

	overall	< 30	$30 - 90$	$90 - 180$	$180 - 360$	> 360
Panel A: USO						
Number of observations	21,254	4,854	6,311	3,837	3,352	2,900
Mean number of strikes	26	13	25	36	35	28
Mean daily trading volume	59,765	26,705	24,886	7,372	3,988	2,077
Mean daily open interest	995,344	214,254	336,401	216,791	181,406	123,679
Panel B: UNG						
Number of observations	16,579	4,281	4,910	2,634	2,241	2,513
Mean number of strikes	14	9	11	19	19	21
Mean daily trading volume	21,091	8,495	8,176	3,778	2,217	1,366
Mean daily open interest	349,638	70,800	110,718	89,967	68,717	61,604
Panel C: GLD						
Number of observations	30,679	5,615	7,767	5,981	7,836	3,480
Mean number of strikes	54	40	55	64	59	46
Mean daily trading volume	106,929	43,721	42,850	12,357	8,576	3,678
Mean daily open interest	2,199,940	415,440	674,615	462,522	466,419	223,195
Panel D: SLV						
Number of observations	22,174	5,104	6,146	3,898	3,964	3,062
Mean number of strikes	22	11	20	28	30	30
Mean daily trading volume	61,220	23,203	23,116	8,001	6,282	4,611
Mean daily open interest	1,488,206	179,766	333,537	329,691	391,764	329,882

Table 4: Summary results of IV function estimation

The table shows the fitted results for the IV function: IV(ξ)=$\alpha_0+\alpha_1\xi+\alpha_2\xi^2$, where IV is the implied volatility and ξ is the standard moneyness of the option. α_0, α_1 and α_2 are the estimated coefficients and they can be converted to the dimensionless coefficients γ_0, γ_1 and γ_2, that is level, slope and curvature. The mean and standard deviation of coefficients are calculated overall and across five maturity groups. The percentage of the significant coefficients is the percentage of parameter estimates that are significant at the 5% level of significance.

	Overall	< 30	$30-90$	$90-180$	$180-360$	> 360
			Panel A: USO			
Mean						
α_0	0.3395	0.3371	0.3423	0.3361	0.3364	0.3455
α_1	-0.0258	-0.0189	-0.0238	-0.0264	-0.0299	-0.0362
α_2	0.0146	0.0146	0.0141	0.0146	0.0147	0.0157
γ_0	0.3395	0.3371	0.3423	0.3361	0.3364	0.3455
γ_1	-0.0766	-0.056	-0.0696	-0.079	-0.0899	-0.1078
γ_2	0.0436	0.0465	0.0431	0.043	0.0424	0.0422
SD						
α_0	0.1156	0.1376	0.1244	0.1061	0.0939	0.0866
α_1	0.0224	0.0153	0.0162	0.0162	0.0169	0.0428
α_2	0.0362	0.0077	0.0088	0.0098	0.0125	0.095
γ_0	0.1156	0.1376	0.1244	0.1061	0.0939	0.0866
γ_1	0.0617	0.0407	0.0411	0.0411	0.0443	0.1222
γ_2	0.104	0.0245	0.0246	0.023	0.0306	0.2741
% Significant Coefficients at 5% level						
α_0	99.38%	97.53%	99.87%	100.00%	100.00%	99.86%
α_1	93.15%	81.62%	95.23%	97.73%	98.15%	96.10%
α_2	88.46%	82.84%	91.49%	93.72%	89.80%	82.76%
Adjusted R^2						
Mean Adj R^2	95.20%	96.52%	97.31%	95.56%	92.97%	90.65%
			Panel B: UNG			
Mean						
α_0	0.4243	0.4178	0.4281	0.4341	0.4235	0.4184
α_1	-0.0003	0.0015	0.0005	-0.0018	-0.0046	0.0006
α_2	0.0178	0.0202	0.0176	0.0169	0.017	0.016
γ_0	0.4243	0.4178	0.4281	0.4341	0.4235	0.4184
γ_1	-0.0016	0.0029	0.0005	-0.0043	-0.0109	-0.0023
γ_2	0.042	0.0489	0.0411	0.0382	0.0393	0.0388
SD						
α_0	0.0929	0.1046	0.1001	0.091	0.0796	0.065
α_1	0.0245	0.0263	0.0243	0.0203	0.0231	0.0265
α_2	0.0172	0.0158	0.0147	0.0121	0.0202	0.0239
γ_0	0.0929	0.1046	0.1001	0.091	0.0796	0.065
γ_1	0.0529	0.055	0.051	0.0432	0.0494	0.0629
γ_2	0.0398	0.0358	0.0304	0.0237	0.0419	0.0648
% Significant Coefficients at 5% level						
α_0	96.57%	90.47%	96.76%	99.96%	100.00%	99.96%
α_1	62.54%	44.03%	58.17%	72.97%	74.56%	80.98%
α_2	74.82%	60.71%	73.28%	88.84%	84.61%	78.47%
Adjusted R^2						
Mean Adj R^2	93.62%	91.43%	96.68%	94.88%	91.81%	91.54%

Table 4: Summary results of IV function estimation (cont'd)

The table shows the fitted results for the IV function: IV(ξ)=$\alpha_0+\alpha_1\xi+\alpha_2\xi^2$, where IV is the implied volatility and ξ is the standard moneyness of the option. α_0, α_1 and α_2 are the estimated coefficients and they can be converted to the dimensionless coefficients γ_0, γ_1 and γ_2, that is level, slope and curvature. The mean and standard deviation of coefficients are calculated overall and across five maturity groups. The percentage of the significant coefficients is the percentage of parameter estimates that are significant at the 5% level of significance.

	Overall	< 30	$30-90$	$90-180$	$180-360$	> 360
			Panel C: GLD			
Mean						
α_0	0.1877	0.1685	0.1778	0.1867	0.1968	0.2218
α_1	-0.0005	-0.0038	-0.0016	-0.0003	0.0011	0.0036
α_2	0.0092	0.0084	0.0088	0.0092	0.0094	0.0109
γ_0	0.1877	0.1685	0.1778	0.1867	0.1968	0.2218
γ_1	-0.0063	-0.02	-0.0109	-0.0049	0.001	0.0078
γ_2	0.049	0.0507	0.05	0.0483	0.0477	0.0479
SD						
α_0	0.0603	0.0582	0.0633	0.0563	0.0528	0.0609
α_1	0.0118	0.0104	0.0105	0.0114	0.0119	0.0146
α_2	0.0054	0.0035	0.0042	0.0051	0.0056	0.009
γ_0	0.0603	0.0582	0.0633	0.0563	0.0528	0.0609
γ_1	0.0544	0.0533	0.054	0.0544	0.0528	0.0544
γ_2	0.0202	0.0141	0.0139	0.018	0.0232	0.0326
% Significant Coefficients at 5% level						
α_0	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
α_1	93.33%	93.16%	93.81%	93.53%	93.06%	92.79%
α_2	99.80%	99.88%	99.85%	99.92%	99.90%	99.17%
Adjusted R^2						
Mean Adj R^2	97.51%	99.08%	98.73%	96.95%	95.63%	97.44%
			Panel D: SLV			
Mean						
α_0	0.289	0.2724	0.2804	0.2918	0.296	0.3212
α_1	-0.0016	-0.0038	-0.0026	-0.0008	0.0012	-0.0009
α_2	0.0164	0.0167	0.016	0.0155	0.0154	0.0192
γ_0	0.289	0.2724	0.2804	0.2918	0.296	0.3212
γ_1	-0.0008	-0.0089	-0.0039	0.0021	0.0098	0.0017
γ_2	0.0558	0.062	0.0565	0.0517	0.0494	0.0571
SD						
α_0	0.0781	0.0791	0.0811	0.0787	0.0718	0.0655
α_1	0.0177	0.0153	0.0155	0.0163	0.0182	0.0248
α_2	0.0098	0.0078	0.0084	0.0079	0.0104	0.015
γ_0	0.0781	0.0791	0.0811	0.0787	0.0718	0.0655
γ_1	0.0564	0.0507	0.0518	0.0547	0.0587	0.0693
γ_2	0.0242	0.0213	0.0191	0.0193	0.0253	0.0359
% Significant Coefficients at 5% level						
α_0	99.55%	98.39%	99.77%	99.95%	100.00%	99.97%
α_1	86.48%	72.67%	85.75%	92.18%	95.21%	92.39%
α_2	96.21%	92.42%	96.81%	99.03%	97.91%	95.53%
Adjusted R^2						
Mean Adj R^2	98.85%	97.89%	98.63%	99.43%	99.01%	99.87%

Table 5: Summary statistics for predictors

The table displays summary statistics for 10 predictive variables in the four commodity markets, including mean, standard deviation (std.dev.), skewness, kurtosis, minimum and maximum. *Level*, *Slope* and *Curv* are IV factors at the end of each month, which are interpolated for days between the end of the current month and the end of the predicted month. TC and FC are the risk-neutral third and fourth cumulants denoted by $TC = Slope \times Level^3$ and $FC = Curv \times Level^4$, respectively. *DLevel*, *DSlope*, *DCurv*, *DTC* and *DFC* are the first differences of *Level*, *Slope*, *Curv*, *TC* and *FC*, respectively.

	Level	Slope	Curv	TC(10^2)	FC(10^3)	DLevel	DSlope	DCurv	DTC (10^2)	DFC (10^3)
Panel A: USO										
Mean	0.3446	-0.0589	0.0450	-0.3753	1.1300	-0.0005	-0.0005	-0.0003	-0.0139	0.0845
Std.dev.	0.1353	0.0392	0.0237	0.6369	2.5011	0.0599	0.0393	0.0302	0.2623	1.1695
Skewness	1.4259	0.3081	3.0002	-3.2242	4.8909	0.2352	0.0238	0.4420	-0.4326	9.7713
Kurtosis	5.7099	4.0940	23.0256	16.1351	29.3946	4.5494	4.1026	19.5896	11.1047	106.4524
Min	0.1394	-0.1761	0.0033	-4.3001	0.0076	-0.1660	-0.1209	-0.1679	-1.0673	-2.6256
Max	0.8673	0.0810	0.2161	0.5114	18.9453	0.2060	0.1323	0.1845	1.0773	12.6376
Panel B: UNG										
Mean	0.4294	-0.0011	0.0437	0.0550	2.1155	0.0003	0.0019	-0.0014	0.5576	-0.2453
Std.dev.	0.1021	0.0505	0.0287	0.7874	2.6497	0.0703	0.0628	0.0329	2.7430	2.9368
Skewness	0.5325	0.6761	1.4279	2.7573	2.1895	0.4939	0.1091	-0.5385	5.3651	-10.4958
Kurtosis	2.3969	6.5616	9.2051	15.1048	9.1020	5.3773	4.0461	4.7400	33.3679	115.5038
Min	0.2743	-0.1751	-0.0381	-1.5351	-1.8087	-0.2383	-0.1852	-0.1167	-5.3753	-32.3713
Max	0.6943	0.2061	0.1950	4.0664	14.8648	0.2450	0.2110	0.0992	18.1080	3.2347
Panel C: GLD										
Mean	0.1868	-0.0174	0.0524	-0.0049	0.1872	-0.0014	-0.0007	-0.0001	0.0385	-0.0157
Std.dev.	0.0775	0.0535	0.0134	0.0924	0.6464	0.0409	0.0450	0.0165	0.5721	0.1582
Skewness	2.3963	-0.5485	0.1787	0.8394	6.6105	1.5351	-0.4227	0.3751	9.8532	-9.9185
Kurtosis	10.2779	3.3344	2.9386	13.0871	51.5877	10.8191	4.7396	4.6415	103.1922	103.7123
Min	0.0929	-0.1980	0.0164	-0.3784	0.0036	-0.1140	-0.1668	-0.0424	-0.8711	-1.6578
Max	0.5734	0.0951	0.0902	0.4419	5.7047	0.2249	0.1389	0.0594	5.9690	0.2209
Panel D: SLV										
Mean	0.3023	-0.0133	0.0610	-0.0912	0.8915	-0.0039	-0.0001	-0.0006	-0.2231	-0.0029
Std.dev.	0.0957	0.0503	0.0203	0.3875	1.9068	0.0559	0.0425	0.0207	1.6528	0.3018
Skewness	1.2594	-0.2810	0.2305	-6.3290	5.6240	0.9557	-0.9287	0.0894	-7.9595	-4.9043
Kurtosis	5.5132	4.0116	6.2191	54.4299	37.8657	8.2967	7.5813	5.1622	69.9917	61.5434
Min	0.1572	-0.1804	-0.0188	-3.4541	-0.8980	-0.2065	-0.2119	-0.0664	-15.4175	-2.6364
Max	0.7114	0.0950	0.1356	0.5993	14.8599	0.2576	0.1054	0.0677	1.1853	1.5629

Table 6: Predictive variable correlations

This table reports correlation coefficients for 10 predictive variables in the four commodity ETF markets. *Level*, *Slope* and *Curv* are IV factors at the end of each month, which are interpolated for days between the end of the current month and the end of the predicted month. *TC* and *FC* are the risk-neutral third and fourth cumulants denoted by $TC = Slope \times Level^3$ and $FC = Curv \times Level^4$, respectively. *DLevel*, *DSlope*, *DCurv*, *DTC* and *DFC* are the first differences of *Level*, *Slope*, *Curv*, *TC* and *FC*, respectively.

	Level	Slope	Curv	TC	FC	DLevel	DSlope	DCurv	DTC	DFC
Panel A: USO										
Level	1									
Slope	0.0181	1								
Curv	-0.3119	0.2494	1							
TC	-0.8358	0.3066	0.2832	1						
FC	0.7925	0.0229	-0.0911	-0.8531	1					
DLevel	0.2271	0.0223	-0.2408	-0.1972	0.1887	1				
DSlope	0.0034	0.5064	0.0898	0.1551	-0.0194	0.0047	1			
DCurv	-0.0249	0.1155	0.6265	0.0551	0.0634	-0.2524	0.1954	1		
DTC	-0.0975	0.001	-0.2214	0.0717	-0.2191	0.0375	0.1486	-0.2109	1	
DFC	-0.0314	0.0689	0.6474	0.0189	0.0536	-0.2116	0.0284	0.5773	-0.3377	1
Panel B: UNG										
Level	1									
Slope	0.1415	1								
Curv	0.1267	-0.2924	1							
TC	0.2469	0.8669	-0.2947	1						
FC	0.7171	-0.1206	0.5549	-0.1064	1					
DLevel	0.3451	0.3261	-0.048	0.3471	0.3125	1				
DSlope	0.1408	0.637	-0.0374	0.5865	0.0358	0.4556	1			
DCurv	-0.1215	-0.2031	0.4786	-0.2606	0.1312	-0.2143	-0.2376	1		
DTC	0.2116	0.4285	-0.0247	0.4782	0.2266	0.1825	0.1613	-0.1819	1	
DFC	-0.2266	-0.1225	-0.0748	-0.2122	-0.4159	-0.3754	-0.2674	0.3564	-0.548	1
Panel C: GLD										
Level	1									
Slope	0.0776	1								
Curv	-0.0881	-0.0366	1							
TC	0.1852	0.5848	0.0186	1						
FC	0.808	0.1037	0.0126	0.0669	1					
DLevel	0.2676	-0.0549	-0.0405	-0.0944	0.2395	1				
DSlope	0.0168	0.4229	0.1556	0.313	-0.0171	-0.0411	1			
DCurv	-0.0236	0.1268	0.6232	-0.0246	0.0497	-0.099	0.2659	1		
DTC	0.2667	0.1091	-0.0747	0.4317	0.1949	0.515	0.1245	-0.0516	1	
DFC	-0.3077	-0.0749	0.1104	-0.4123	-0.2336	-0.5708	-0.0924	0.092	-0.9811	1
Panel D: SLV										
Level	1									
Slope	-0.2772	1								
Curv	-0.0257	-0.3507	1							
TC	-0.4212	0.5442	0.0076	1						
FC	0.7264	-0.1184	0.2545	-0.4096	1					
DLevel	0.2374	-0.1272	-0.0154	-0.3496	0.2632	1				
DSlope	-0.0672	0.4197	-0.0968	0.3177	-0.1127	-0.2034	1			
DCurv	-0.1195	-0.1512	0.4681	-0.0398	-0.0199	-0.0797	-0.2445	1		
DTC	-0.3988	0.2691	0.0022	0.8324	-0.5115	-0.4415	0.3635	-0.0468	1	
DFC	-0.3373	-0.0156	0.1628	0.6429	-0.4702	-0.2946	-0.0553	0.1838	0.5354	1

Table 7: Predictability of four commodity ETF returns

This table reports the monthly return predictability for USO, UNG, GLD and SLV ETFs based on in-sample and out-of-sample tests. The sample periods are 9 May 2007 to 29 December 2017 (Panels A and B), 3 June 2008 to 29 December 2017 (Panel C) and 8 December 2008 to 29 December 2017 (Panel D). The definitions of all predictors are the same as those in Table 4. The table also presents the estimated slope coefficients $\hat{\beta}$, their Newey and West (1987) t-statistics, in-sample adjusted R^2 statistics and out-of-sample adjusted R^2 statistics (R^2_{os}). *,** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	β	t	$R^2(\%)$	$R^2_{os}(\%)$	β	t	$R^2(\%)$	$R^2_{os}(\%)$
	Panel A: USO				*Panel B: UNG*			
Level	-0.14	(-1.14)	2.97	-9.78	-0.17**	(-2.04)	1.74	3.91**
Slope	-0.17	(-0.76)	-0.30	-1.35	0.32	-1.45	1.28	-2.55
Curv	-0.06	(-0.17)	-0.78	-2.02	-0.65**	(-2.31)	2.04	-0.57
TC	2.15	(0.93)	1.27	-6.00	0.69	(0.53)	-0.56	-1.71
FC	-11.13***	(-3.15)	7.72	-3.86	-8.99**	(-2.40)	3.86	3.75**
DLevel	-0.49**	(-2.53)	8.83	4.06*	-0.00	(-0.03)	-0.81	-1.03
DSlope	0.21	(0.95)	-0.07	0.21	0.07	(0.57)	-0.64	-0.68
DCurv	-0.15	(-1.00)	-0.59	-2.71	-0.26	(-1.12)	-0.20	1.1*
DTC	8.53	(1.37)	4.70	-21.09	-0.26	(-0.71)	-0.40	-6.92
DFC	-3.86	(-1.18)	-0.58	-40.29	4.16***	(4.76)	0.43	0.83
	Panel C: GLD				*Panel D: SLV*			
Level	0.04	(0.46)	-0.64	-0.05	0.09	(1.07)	-0.13	1.50
Slope	-0.03	(-0.34)	-0.81	-6.71	-0.01	(-0.06)	-0.94	-4.86
Curv	0.63**	(2.00)	1.51	-11.06	0.42	(1.26)	-0.15	-0.68
TC	-12.54	(-1.55)	3.60	-2.17	-2.64*	(-1.94)	0.23	0.54
FC	10.98	(1.48)	0.79	1.83**	6.42**	(2.50)	0.74	1.96*
DLevel	-0.31***	(-2.86)	4.58	-0.02	-0.06	(-0.26)	-0.84	0.08
DSlope	-0.11	(-0.99)	-0.15	-6.33	0.07	(0.31)	-0.85	-13.40
DCurv	0.22	(0.91)	-0.46	-18.44	0.24	(0.48)	-0.68	-0.13
DTC	-3.10***	(-12.05)	9.65	-5.30	-0.67**	(-2.15)	0.45	-1.90
DFC	121.24***	(10.76)	11.43	-1.32	-46.83***	(-5.08)	1.32	0.93

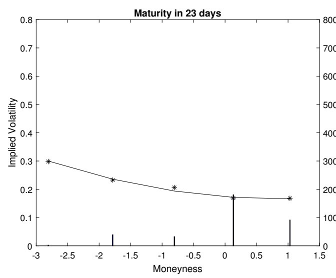
Table 8: Predictability of S&P500 returns

This table reports the monthly S&P500 return predictability by using the predictors from the four commodity markets based on in-sample and out-of-sample tests. The sample periods are 9 May 2007 to 29 December 2017 (Panels A and B), 3 June 2008 to 29 December 2017 (Panel C) and 8 December 2008 to 29 December 2017 (Panel D). The definitions of all predictors are the same as those in Table 4. The estimated slope coefficients $\hat{\beta}$, their Newey and West (1987) t-statistics, in-sample adjusted R^2 statistics and out-of-sample adjusted R^2 statistics (R^2_{os}) are also reported. *,** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

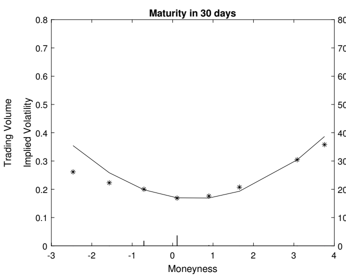
	β	t	$R^2(\%)$	$R^2_{os}(\%)$	β	t	$R^2(\%)$	$R^2_{os}(\%)$
	Panel A: USO-SPX				*Panel B: UNG-SPX*			
Level	-0.06	(-1.35)	2.53	6.13**	-0.03	(-0.48)	-0.40	0.54
Slope	-0.23**	(-2.17)	3.25	8.75***	0.10	(1.24)	0.41	4.34**
Curv	0.06	(0.30)	-0.71	-3.81	-0.01	(-0.05)	-0.80	-0.87
TC	0.36	(0.34)	-0.55	-0.39	-0.18	(-0.28)	-0.70	-8.80
FC	-4.32***	(-5.06)	4.88	8.77***	-0.14	(-0.07)	-0.80	-1.49
DLevel	-0.15*	(-1.72)	3.27	1.27*	0.07	(1.61)	0.54	5.38***
DSlope	-0.05	(-0.39)	-0.64	-1.29	0.11*	(1.84)	1.71	8.01**
DCurv	0.02	(0.19)	-0.79	-4.64	0.08	(0.77)	-0.47	0.60
DTC	4.63*	(1.80)	6.46	-24.52	-0.05	(-0.33)	-0.74	-17.00
DFC	-0.45	(-0.27)	-0.79	-117.01	0.58	(1.34)	-0.67	2.64***
	Panel C: GLD-SPX				*Panel D: SLV-SPX*			
Level	-0.14	(-1.41)	5.14	-17.56	0.03	(0.60)	-0.22	1.56*
Slope	-0.03	(-0.35)	-0.79	-24.06	-0.01	(-0.13)	-0.94	-2.54
Curv	0.14	(0.54)	-0.72	-6.02	-0.28	(-0.83)	1.17	-9.34
TC	-11.70**	(-2.39)	5.24	-10.09	-1.97***	(-2.81)	3.01	1.17
FC	-20.96***	(-4.44)	8.75	-0.45	-2.38	(-0.51)	0.46	-3.14
DLevel	-0.24	(-1.13)	4.06	-35.54	0.12*	(1.93)	2.45	-1.01
DSlope	0.00	(0.02)	-0.90	-14.56	-0.06	(-0.72)	-0.48	-1.33
DCurv	0.13	(0.60)	-0.66	-19.64	-0.13	(-0.80)	-0.43	-0.44
DTC	-3.08***	(-18.27)	15.47	1.51	-0.53***	(-6.44)	4.61	-1.50
DFC	114.92***	(20.26)	16.53	-0.84	-22.79*	(-1.82)	2.48	-8.38

Figure 1: USO IV smirks on 27 December 2017

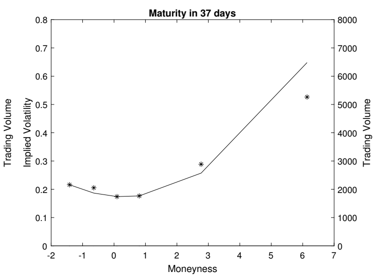
This figure illustrates USO market and fitted IV curves for 10 different time to maturity terms (23, 30, 37, 51, 79, 114, 205, 296, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.



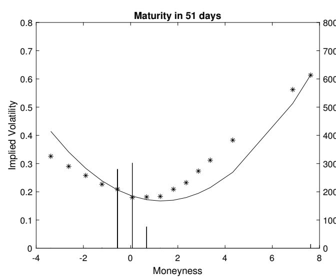
(a)



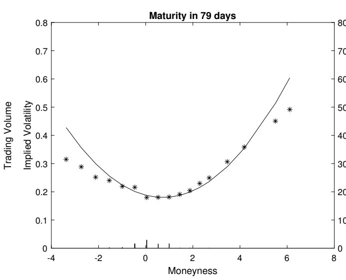
(b)



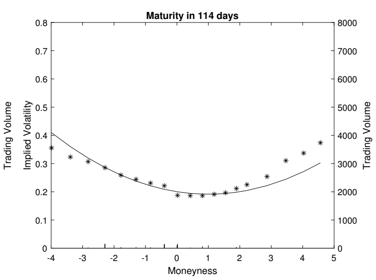
(c)



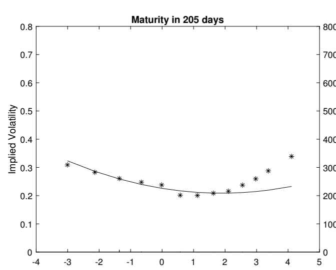
(d)



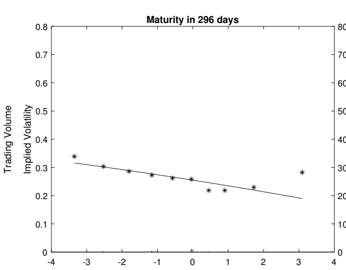
(e)



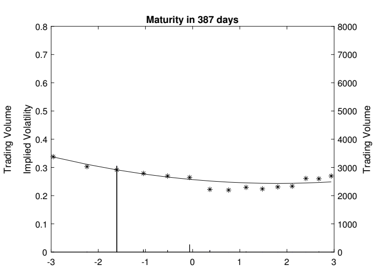
(f)



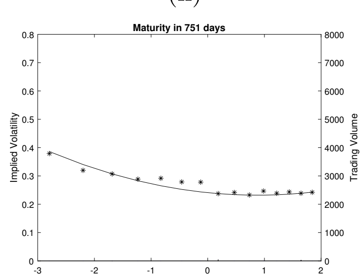
(g)



(h)



(i)



(j)

Figure 2: UNG IV smirks on 27 December 2017

This figure illustrates UNG market and fitted IV curves for 10 different time to maturity terms (9, 16, 23, 30, 37, 51, 114, 205, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.

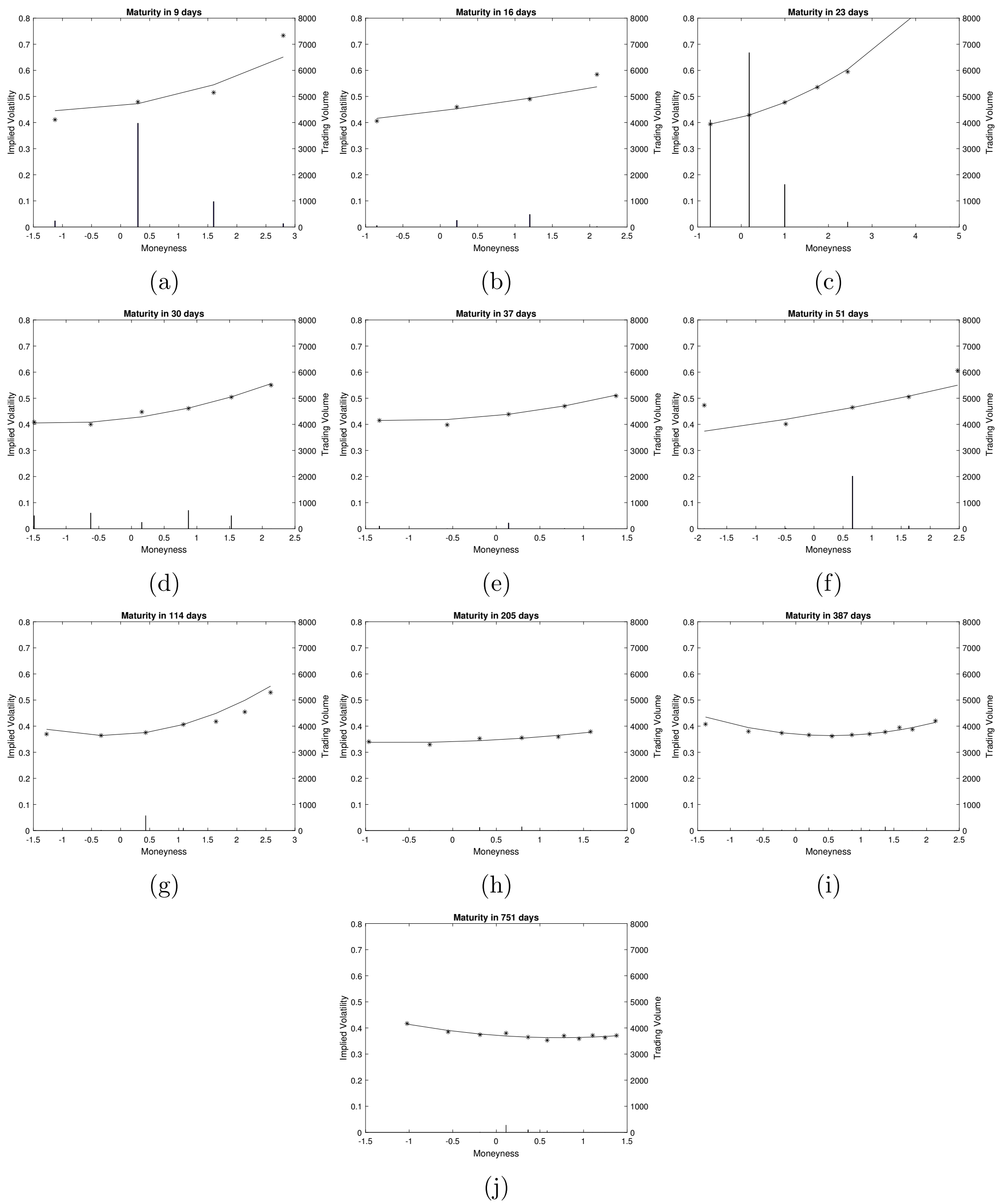
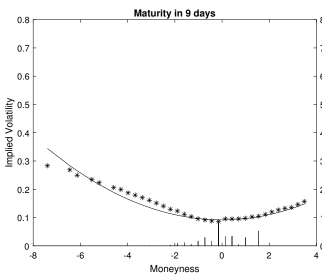
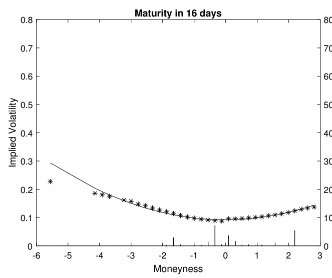


Figure 3: GLD IV smirks on 27 December 2017

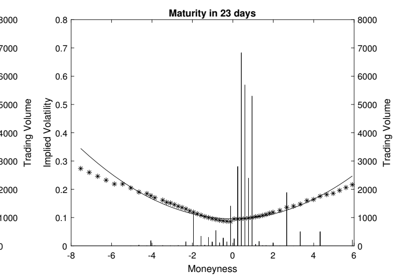
This figure illustrates GLD market and fitted IV curves for 14 different time to maturity terms (9, 16, 23, 30, 37, 51, 79, 92, 114, 170, 184, 268, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.



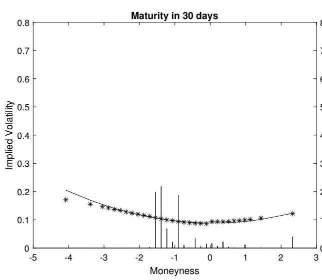
(a)



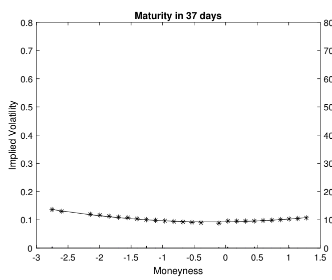
(b)



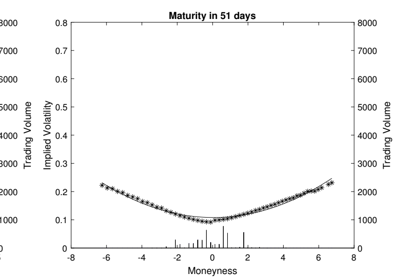
(c)



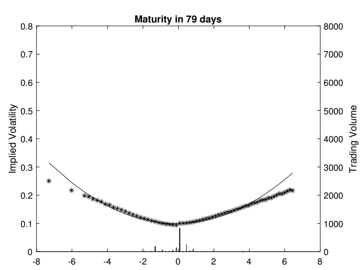
(d)



(e)



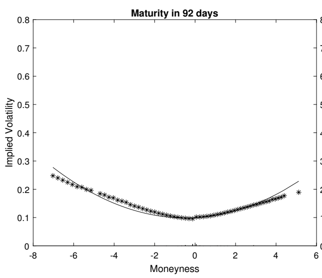
(f)



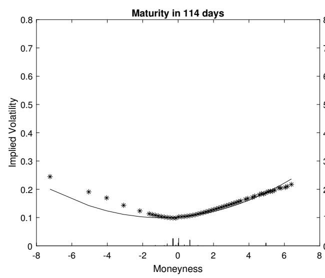
(g)

Figure 3: GLD IV smirks on 27 December 2017 (cont'd)

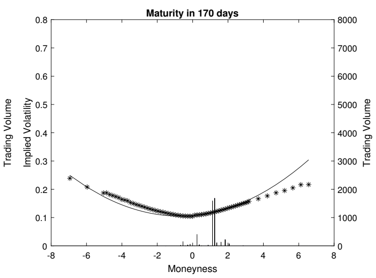
This figure illustrates GLD market and fitted IV curves for 14 different time to maturity terms (9, 16, 23, 30, 37, 51, 79, 92, 114, 170, 184, 268, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.



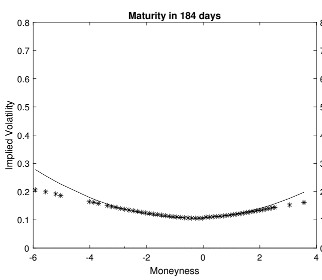
(a)



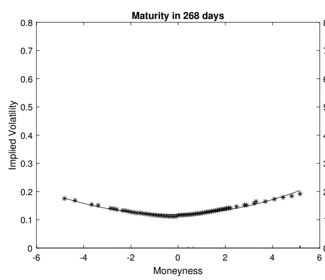
(b)



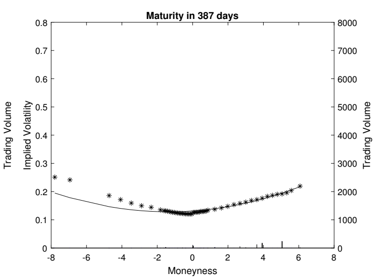
(c)



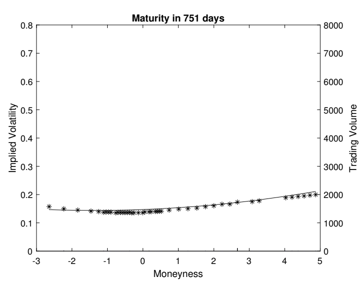
(d)



(e)



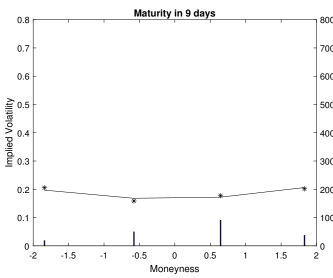
(f)



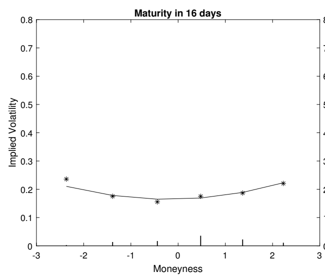
(g)

Figure 4: SLV IV smirks on 27 December 2017

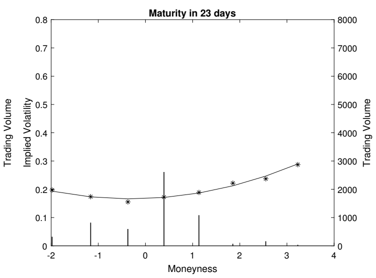
This figure illustrates SLV market and fitted IV curves for 15 different time to maturity terms (9, 16, 23, 30, 37, 51, 79, 92, 114, 184, 205, 275, 296, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.



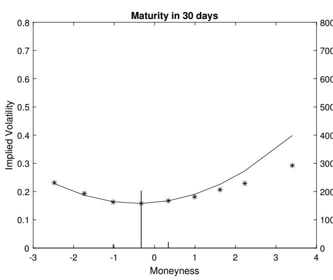
(a)



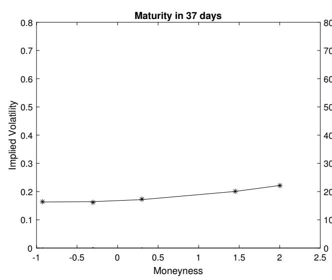
(b)



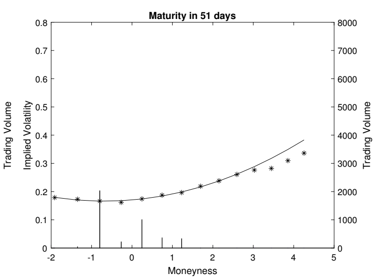
(c)



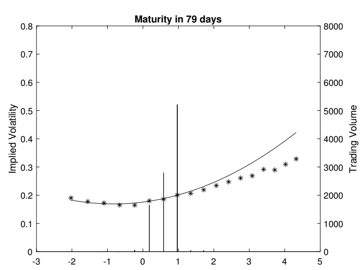
(d)



(e)



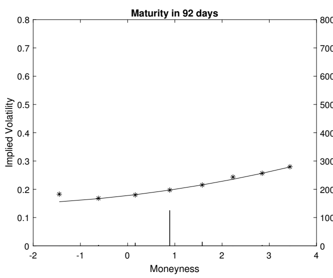
(f)



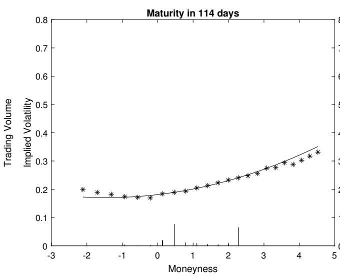
(g)

Figure 4: SLV IV smirks on 27 December 2017 (cont'd)

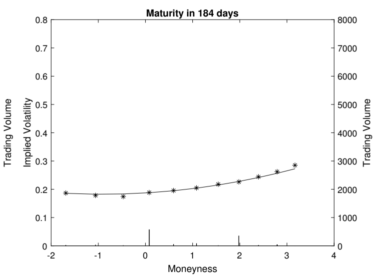
This figure illustrates SLV market and fitted IV curves for 15 different time to maturity terms (9, 16, 23, 30, 37, 51, 79, 92, 114, 184, 205, 275, 296, 387 and 751 days) on 27 December 2017. The stars in each graph are the market IVs, the solid lines are fitted IV curves and the bars are the trading volume.



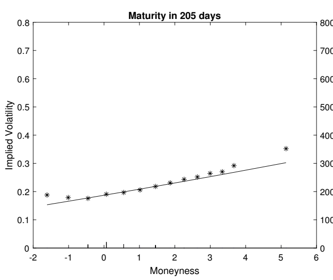
(a)



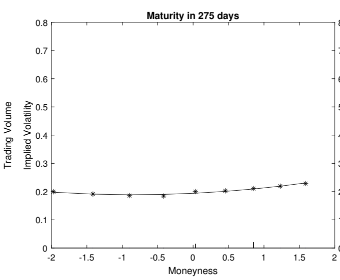
(b)



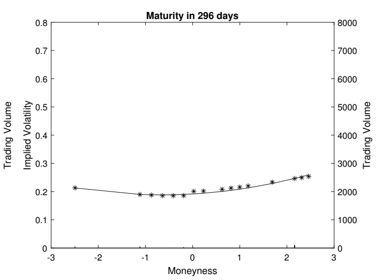
(c)



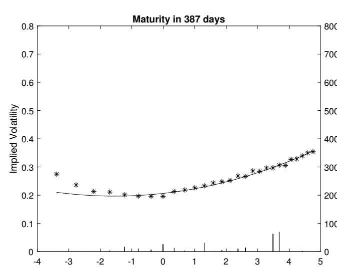
(d)



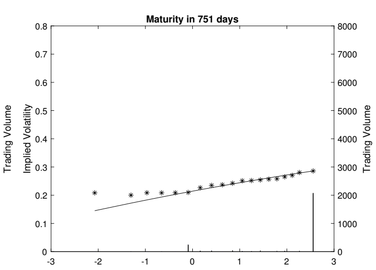
(e)



(f)



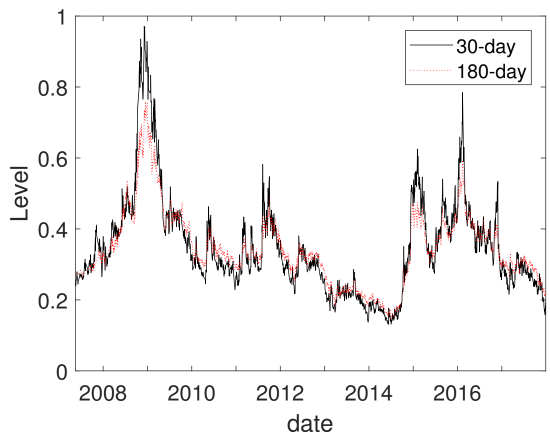
(g)



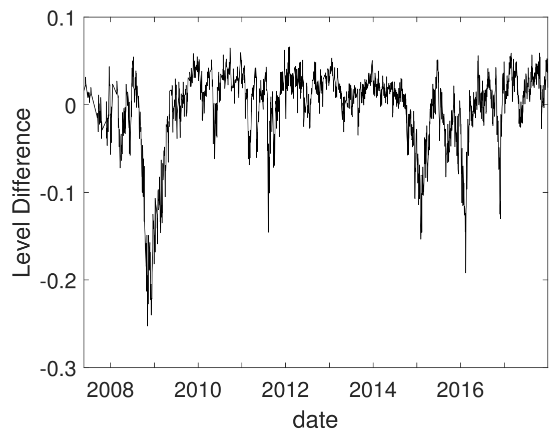
(h)

Figure 5: Constant maturity IV dynamics: crude oil

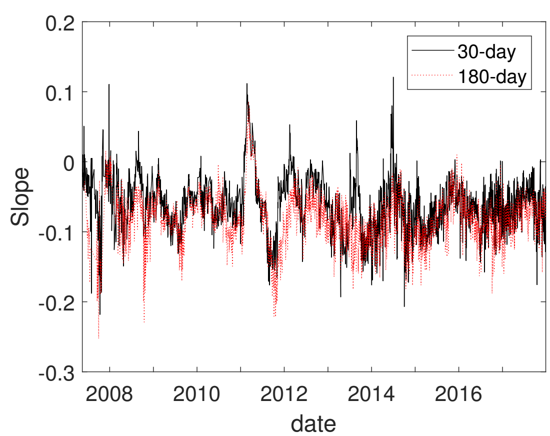
This figure shows dynamics of the 30- and 180-day constant maturity estimate ATM IV, slope and curvature and the difference of the 180-day and 30-day in the crude oil market.



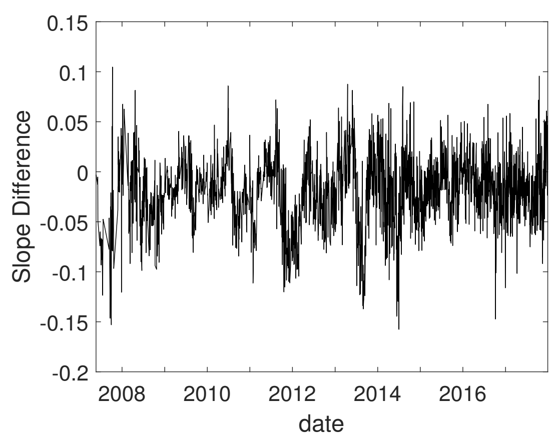
(a) IV level in two terms



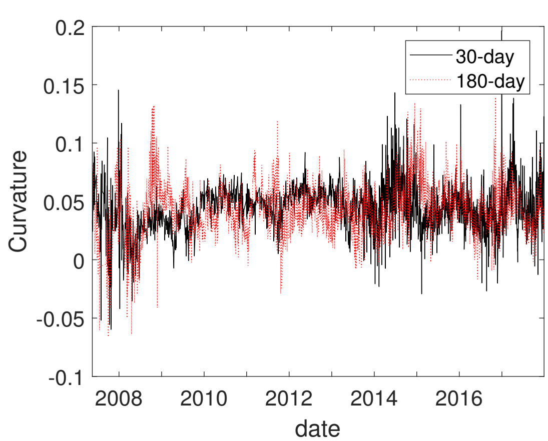
(b) Difference in IV level



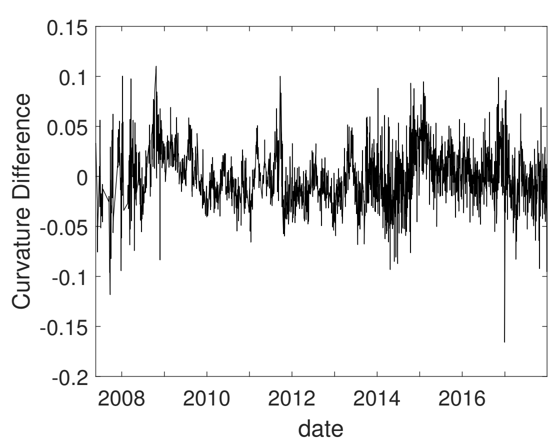
(c) IV slope in two terms



(d) Difference in IV slope



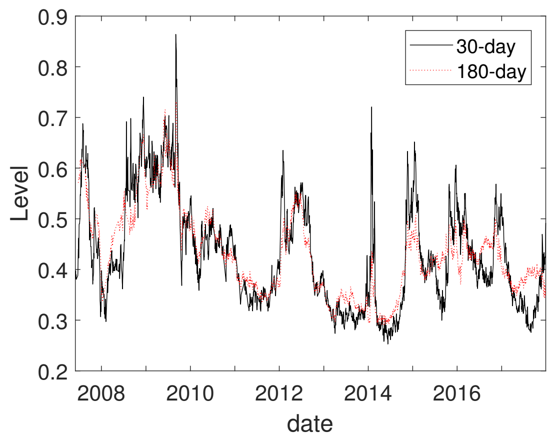
(e) IV Curvature in two terms



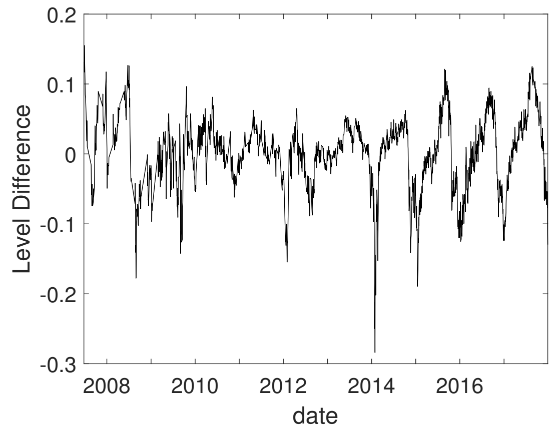
(f) Difference in IV curvature

Figure 6: Constant maturity IV dynamics: natural gas

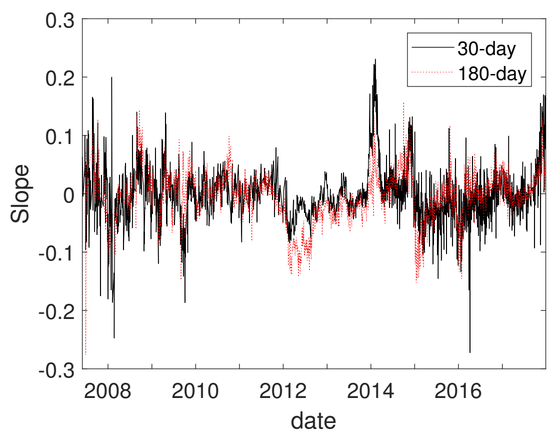
This figure shows dynamics of the 30- and 180-day constant maturity estimate ATM IV, slope and curvature and the difference of the 180-day and 30-day in the natural gas market.



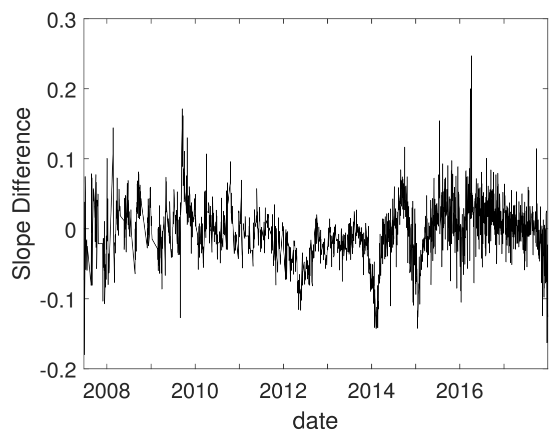
(a) IV level in two terms



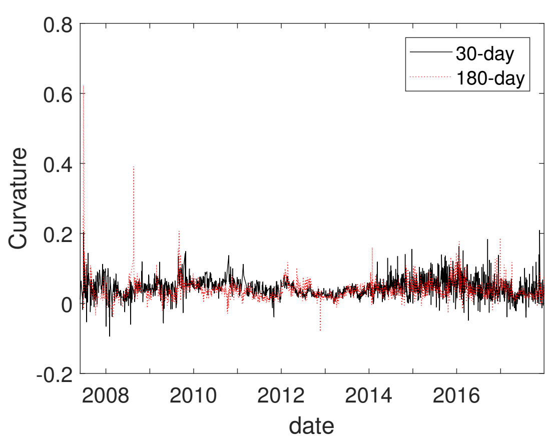
(b) Difference in IV level



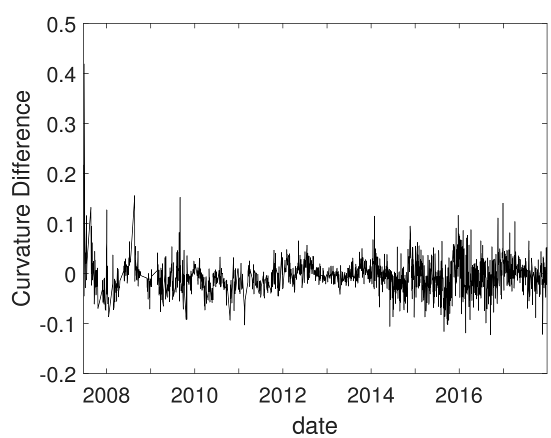
(c) IV slope in two terms



(d) Difference in IV slope



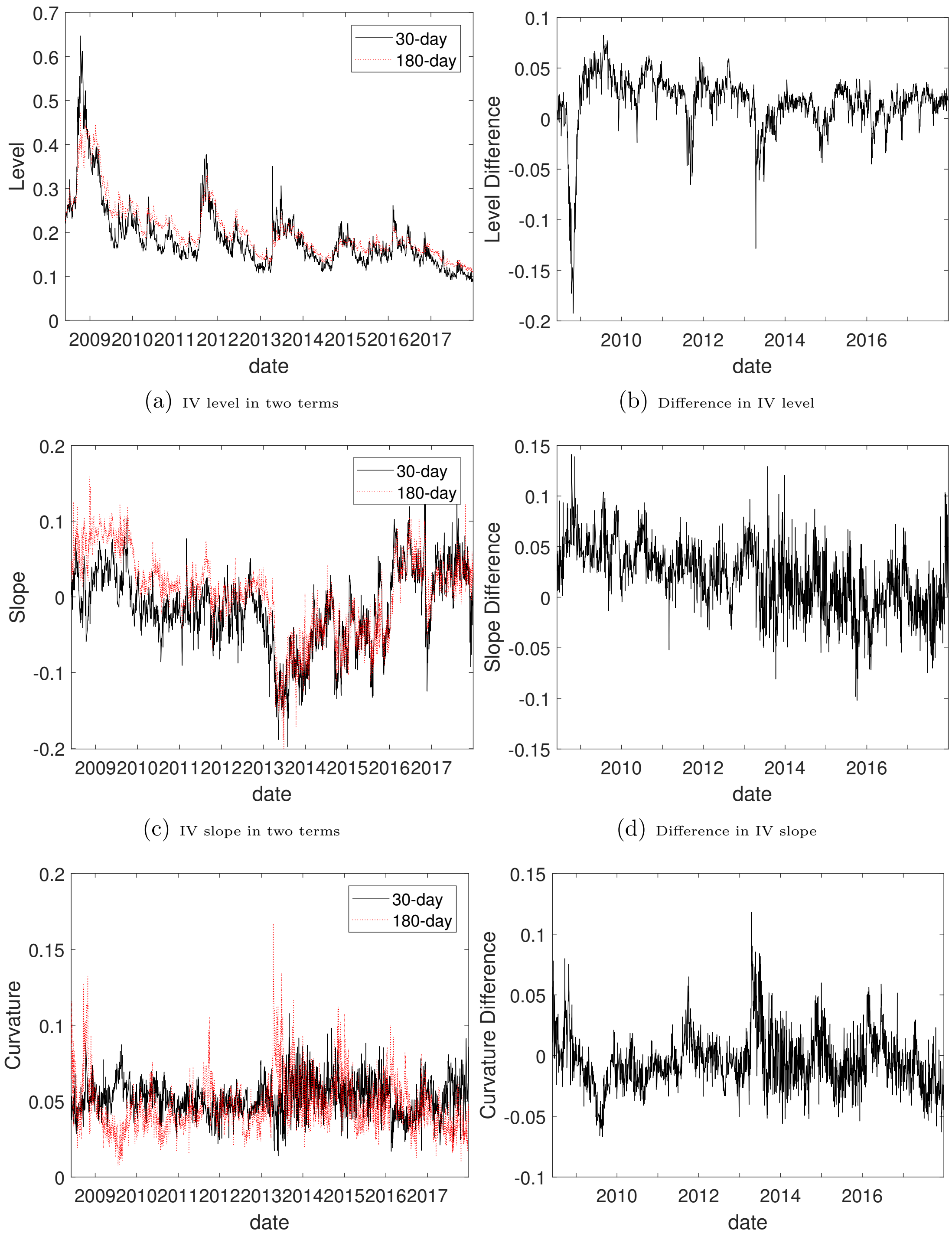
(e) IV Curvature in two terms



(f) Difference in IV curvature

Figure 7: Constant maturity IV dynamics: gold

This figure shows dynamics of the 30- and 180-day constant maturity estimate ATM IV, slope and curvature and the difference of the 180-day and 30-day in the gold market.



(a) IV level in two terms

(b) Difference in IV level

(c) IV slope in two terms

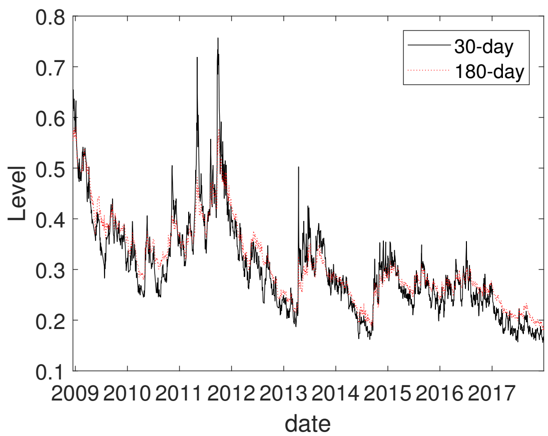
(d) Difference in IV slope

(e) IV Curvature in two terms

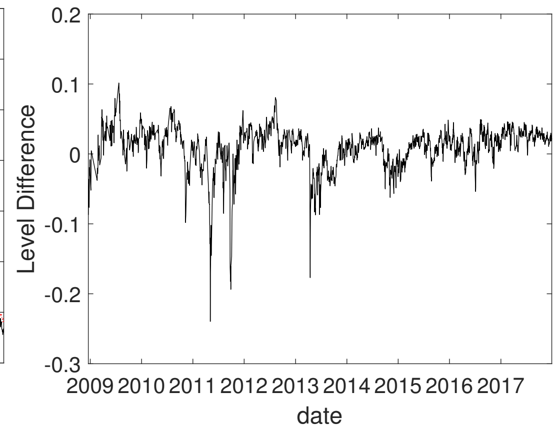
(f) Difference in IV curvature

Figure 8: Constant maturity IV dynamics: silver

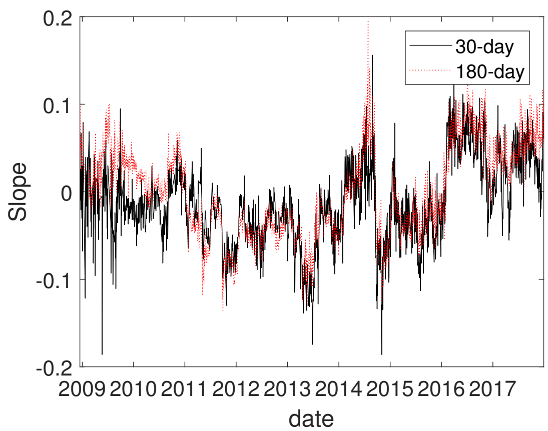
This figure shows dynamics of the 30- and 180-day constant maturity estimate ATM IV, slope and curvature and the difference of the 180-day and 30-day in the silver market.



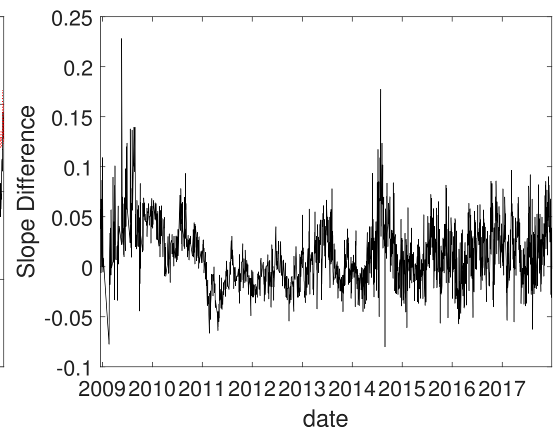
(a) IV level in two terms



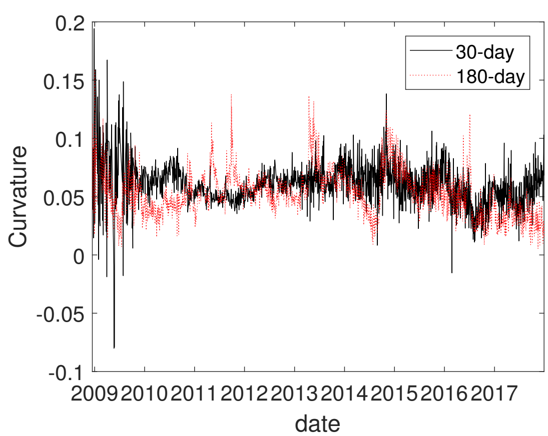
(b) Difference in IV level



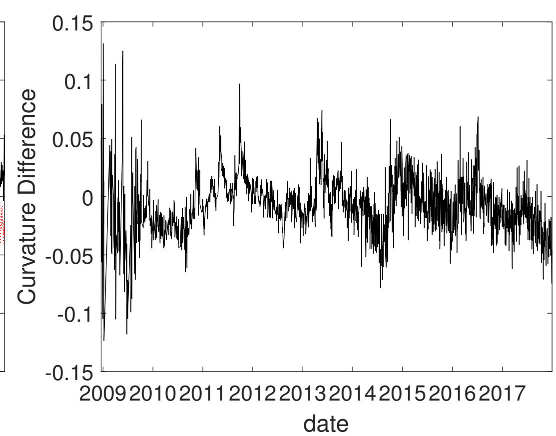
(c) IV slope in two terms



(d) Difference in IV slope



(e) IV Curvature in two terms



(f) Difference in IV curvature

Figure 9: Constant maturity IV dynamics in four commodity markets

This figure represents the 30-day constant maturity dynamics of the estimated ATM IV, the slope and the curvature for crude oil, natural gas, gold and silver markets, respectively.