Investor Beliefs and State Price Densities in the Crude Oil Market*

Xuhui (Nick) Pan
Freeman School of Business
Tulane University

August 28, 2015

Abstract

Standard asset pricing theory suggests that state price densities (SPDs or the pricing kernel) monotonically decrease with returns. We find that the SPDs implicit in the crude oil market display a time-varying U-shape pattern. This implies that investors assign high state prices to both negative and positive returns. We further use data of the crude oil derivatives market, where speculation and short sales are not regulated, to document how the SPDs are dependent on investor beliefs. Investors' preferences over return states are reflected in belief dispersions in options and the underlying futures contracts: Investors overall assign higher state prices to negative returns when there are higher demands for out-of-the-money put options, and when there are increased speculations during the post-financialization period. Higher state prices of negative (positive) returns are also associated with higher trading volume of out-of-the-money put (call) options.

JEL Classification: G12, G13

Keywords: State price density; Skewness; Investor belief; Crude oil; Speculation.

*We would like to thank Peter Christoffersen, Hitesh Doshi, Stephen Figlewski, Kris Jacobs, Sang Baum Kang, Scott Linn, Sergei Sarkissian, Kenneth Singleton, Lars Stentoft, Feng Zhao, and participants at the CICF (2015), EFA (2012), and NFA (2012) conferences and seminar participants at the Bank of Canada and Tulane University for their comments. Correspondence to: Xuhui (Nick) Pan, 7 McAlister Dr., New Orleans, LA 70118. Tel: (504) 314-7031; Fax: (504) 862-8327; E-mail: xpan@tulane.edu.

1 Introduction

State price densities (SPDs or the pricing kernel) contain information about investor preferences and essentially determine expected returns and risk premia. According to the no-arbitrage principle, we can price any asset as long as we know the SPDs and the final payoff of this asset. Although several studies have estimated the SPDs using equity market data,[1] few papers investigate other financial markets.[2] Given that (a) commodities have emerged as a fast growing asset class, (b) each market only contains part of the wealth of the aggregate economy and a subset of information about the aggregate pricing kernel, and (c) participants in the commodity market are often different from investors in other markets, a systematic study of the SPDs implicit in this market is warranted. Using the crude oil market, which is the largest and most liquid commodity derivatives market, this paper estimates the SPDs and backs out investor preferences towards different states of economy.

We estimate SPDs using crude oil futures and options data from 1990 to 2012, and we document time variation in the SPDs and their dynamic structure. We find that the SPDs in the crude oil market display a time varying asymmetric U-shape pattern: Investors assign high state prices (per unit probability) to both large negative and large positive returns. This implies that either extreme low or high oil futures prices correspond to bad states of the economy. The slope of the U-shaped SPDs in both the decreasing and increasing regions also varies depending on the market condition. In other words, investors' marginal value of payoffs at large negative and positive returns, exhibits significant variations across time. We also document that average returns on out-of-the-money (OTM) oil call and put options are negative, consistent with the SPDs having both decreasing and increasing regions (Bakshi, Madan, and Panayotov, 2010).

Another strand of the literature has argued that SPDs depend on differences in investor beliefs, and that this heterogeneity affects expected returns and the price of risk (e.g., Anderson, Ghysels, and Juergens, 2005; Beber, Buraschi, and Breedon, 2010). In particular, Bakshi, Madan, and Panayotov (2010) advocate that the pattern of the SPDs implied by the index option is compatible with theory when risk-averse investors have heterogeneous beliefs and are able to short sell equities. However, there is limited literature directly testing how heterogeneous beliefs affect the shape of the SPDs due to various reasons: Heterogeneity of investor beliefs is difficult to measure precisely; short selling is highly regulated in some

[1] An incomplete list includes Aït-Sahalia and Lo (2000), Jackwerth (2000), Rosenberg and Engle (2002), Chernov (2003), Chabi-Yo (2012), Christoffersen, Heston, and Jacobs (2013), Linn, Shive and Shumway (2014), Song and Xiu (2014).

[2] Notable exceptions include Beber and Brandt (2006) and Li and Zhao (2009) who study empirical SPDs in the fixed income market, and Kitsul and Wright (2013) estimate empirical pricing kernel using inflation options.

financial markets, while it is a critical element for the theory to explain the empirical SPDs; and accurate estimation of the dynamic structure of the SPDs can be data-demanding.

The crude oil market provides an excellent laboratory to examine the dependence of SPDs on investor beliefs for the following reasons. First, data on speculators' positions are available in this market. The level of speculation can be interpreted as a measure of heterogeneous beliefs because speculators usually bet on certain price movements, and disagreements among investors induce speculative trading (e.g., Scheinkman and Xiong, 2003). Second, there are no restrictions on short sales in this market. While some investors trade crude oil futures to hedge risks according to their real demand and supply of crude oil, others may take any positions simply based on their beliefs about futures prices. Therefore we can test whether investor beliefs affect the SPDs. Third, the crude oil derivatives market is fast growing, and historical data are available for more than twenty years. We have large cross-sections of futures and option prices, which allow us to accurately extract the SPDs. Moreover, data on trading volumes and open interests of both futures and options enable us to construct various measures of investor heterogeneous beliefs as suggested by the literature (e.g., Kandel and Pearson, 2005; Buraschi and Jiltsov, 2006).

We investigate whether investor beliefs about futures prices embedded in trading activities in crude oil futures and options affect the slope of the SPDs, one of the fundamental characteristics of the SPDs. The slope can be related to investors' risk aversion (Rosenberg and Engle, 2002). It also compares the marginal value of payoffs in different economic states, as measured by the level of futures returns. Since the physical densities of oil futures returns are relatively symmetric, the slope of the SPDs is primarily characterized by risk-neutral skewness. If risk-neutral skewness is more negative (positive), the decreasing (increasing) region of the SPDs has a steeper slope, and investors assign higher state prices to more negative (positive) returns. We first provide evidence that a direct measure of the slope of the SPDs, defined as the difference between the value of the SPDs of two return points, is affected by investor beliefs reflected in options and futures trades. However, this measure of slope can be arbitrary and the slope itself can be noisy around the distribution tails. We therefore focus on risk-neutral skewness, which provides a more reliable and comprehensive measure of variations in the marginal rate of substitution across states. We calculate risk-neutral skewness from two distinct approaches and regress it on investor beliefs. We also investigate whether the SPDs are affected by speculative activities in the crude oil market after the financialization of the commodity market. The selection of the sub-sample period is based on the evidence by Tang and Xiong (2012), who document the financialization of commodity markets beginning around 2004-2005 and how speculative investments affect commodity futures prices.[3]

Our empirical results indicate that the slopes of the U-Shaped SPDs are affected by various measures of investors beliefs. When there are more OTM put option demands, the risk-neutral distribution is more negatively skewed. As such, the decreasing side of the SPD has a steeper slope and investors assign higher state prices to more negative returns. Investor beliefs embedded in options trading volume also affect the shape of the SPDs. When there are higher OTM put (call) trading volume, the slope of the decreasing (increasing) side of the SPD is steeper, i.e. the marginal rate of substitution across negative (positive) oil futures returns is higher. We find that a high level of speculation after the financialization of the commodity market is associated with the steeper slope of decreasing region of the SPD and investors overall are worried more about large negative returns. Our findings lend support to the argument of Singleton (2014): flows from financial investors and speculative activities have significant effects on the crude oil futures market during the post-financialization period.

We also investigate the impact of investor beliefs of the equity market and the aggregate economy on the SPDs. Although investors' expectations about the equity market significantly affect index option prices and return distributions in the equity market (Han, 2008), bearish stock market beliefs do not imply more negative risk-neutral skewness of crude oil futures returns, and do not have significant impacts on the slope of the SPDs in the crude oil market. It is the belief dispersions from investors in the crude oil market, rather than investor beliefs about the equity market or the economy, that significantly affect state prices of returns in the crude oil market. This finding aligns with Goldstein, Li and Yang (2013), who argue that although information is integrated and fast moving, financial markets can be relatively segmented due to the specialization and friction of investments.

This paper is part of a growing list of recent studies that examine how the activities of investors in the commodity market, both hedgers and speculators, affect futures prices and returns. Hamilton and Wu (2014) document significant changes in oil futures risk premia due to active investments from financial institutions in recent years. Motivated by the coincident price rise and increased financial participation in the crude oil market, Büyükşahin and Harris (2011) analyze whether the crude oil price is driven by hedge funds and other speculators. Acharya, Lochstoer, and Ramadorai (2013) find that producers' hedging demand, captured by their default risk, predicts commodity returns. Hong and Yogo (2012) show that the high level of commodity market activity, measured by the high open-interest growth, predicts high commodity returns. Etula (2013) finds that the supply of speculator capital, captured by changes in broker-dealer balance sheets, predicts commodity returns, especially in energy

[3]More recent contributions on the financialization of commodity markets include Hamilton and Wu (2014), Cheng and Xiong (2014), Henderson, Pearson and Wang (2014) and others.

commodities. However, this paper examines how investor beliefs embedded in derivatives trading activities affect the SPDs, which determine not only the commodity futures prices and returns, but also the risk premia.

The rest of the paper proceeds as follows. We present the estimation of the risk-neutral densities, physical densities, and the SPDs in Section 2. Section 3 discusses how the SPDs are affected by investor beliefs. Section 4 concludes.

2 Risk-Neutral Densities and SPDs in the Crude Oil Market

In this section, we discuss the economic framework to obtain the SPDs in the crude oil market based on the no-arbitrage principle. We then discuss the estimation methodology of the risk-neutral densities, physical densities and the SPDs using futures and option prices. Next we describe futures and option data and present the estimated SPDs, as well as option implied moments.

2.1 Economic Framework

If we denote the SPD by ξ, based on the no-arbitrage principle, as long as we know the final payoff p_T of an asset, the price of the asset at time t can be obtained by

$$p_t = E[\xi_T p_T | \mathcal{F}_t], \tag{1}$$

where $\mathcal{F}_t$ denotes the investors' information set at time t. Consider a European call option written on a futures contract $F_{t,T}$ with the strike price K, which matures at time T.[4] Call option price is the final payoff discounted to time t,

$$\begin{aligned} C(F_{t,T}, K, t, T) &= E[\xi_T (F_T - K)^+ | \mathcal{F}_t] \\ &= \int_K^\infty \xi_T(x)(F_T(x) - K) P(F_T(x)|\mathcal{F}_t) dx, \end{aligned} \tag{2}$$

where we use $P(F_T(x)|\mathcal{F}_t)$ to denote the conditional physical density at time t. However, note that the true SPD or pricing kernel depends on many state variables and is unknown

[4] Crude oil options expire three business days prior to the expiration of the underlying futures contract. To simplify the notation, we do not explicitly distinguish between the futures maturity date T and the option maturity date T' in this paper.

to investors.[5]

In order to price derivatives, we usually rely on the price dynamics of underlying assets under the risk-neutral measure Q. Under this measure, option prices discounted at the riskless rate are martingales. At time t, we price the call option by

$$\begin{aligned} C(F_{t,T}, K, t, T) &= e^{-r(T-t)} E^Q[(F_T - K)^+ | \mathcal{F}_t] \\ &= e^{-r(T-t)} \int_K^{\infty} (F_T(x) - K) P^Q(F_T(x)|\mathcal{F}_t) dx, \end{aligned} \tag{3}$$

where $P^Q(F_T(x)|\mathcal{F}_t)$ is the conditional density of F_T under the risk-neutral measure. Based on this equation, we can price any option with a known final payoff once we have $P^Q(F_T(x)|\mathcal{F}_t)$. Breeden and Litzenberger (1978) have shown that $P^Q(F_T(x)|\mathcal{F}_t)$ can be obtained by taking the second order derivative of call prices with respect to the strike price K,

$$P^Q(F_T|\mathcal{F}_t) = e^{r(T-t)} \frac{\partial^2 C(F_{t,T}, K, t, T)}{\partial K^2} |_{K=F_T}. \tag{4}$$

Although it is not possible to obtain the SPD ξ that is defined over the aggregate economy, we estimate the SPD in the crude oil market, $\widetilde{\xi}$, and we focus on this specific but relatively segmented financial market. This allows us to infer relevant information about investors' preferences and expectations for the purpose of pricing crude oil derivatives; and how investors value certain economic states and foresee the probability of those states in the crude oil market. Combining equations (2) and (3), we can estimate the projected SPD in the crude oil market,

$$\widetilde{\xi}(F_T|\mathcal{F}_t) = e^{-r(T-t)} \frac{P^Q(F_T|\mathcal{F}_t)}{P(F_T|\mathcal{F}_t)}. \tag{5}$$

Defined as the Arrow-Debreu price of per unit of probability, SPDs reflect how investors evaluate possible states of nature and their expectations of the probability of those states happening. While many studies estimate the SPDs using index options (i.e., $\widetilde{\xi}$ in the equity market [e.g., Jackwerth, 2000]) and interest rate derivatives (i.e., $\widetilde{\xi}$ in the fixed income market [e.g., Li and Zhao, 2009]), this paper investigates the SPDs in the crude oil market. As shown in (5), SPDs depend on two components: risk-neutral densities and physical densities. We first discuss the estimation of the risk-neutral density.

[5] Since we estimate the SPDs using certain specific assets, which are only a subset of the aggregate wealth, we can only obtain the SPDs projected onto those assets. For example, the SPDs estimated from index options are the projection of ξ onto the index returns.

2.2 Estimation of the Risk-Neutral Density

We compute conditional estimates of the risk-neutral density using option prices. More specifically, we adapt the semi-parametric approach introduced by Aït-Sahalia and Lo (1998) and further developed by Christoffersen and Jacobs (2004). The semi-parametric approach is designed to utilize all available information implicit in the entire cross-section of option prices, while keeping the parametric assumptions to a minimum. On a given day, we first fit Black (1976) implied volatilities of the cross-sectional option data as a second order polynomial function of strike price and maturity. Then we construct a grid of strike prices and obtain at-the-money Black (1976) implied volatilities from the fitted polynomial function for each maturity $T-t$. With these implied volatilities, we back out call prices $\widehat{C}\left(F_t, K, t, T, \sigma(K, T)\right)$ on the desired grid of strike prices, and then calculate the risk-neutral density (4) for the futures price at the maturity date T. Lastly we compute the second order derivative of the fitted option price with respect to the strike price

$$\widehat{P}^Q(F_T|\mathcal{F}_t) = e^{r(T-t)} \cdot \frac{\partial^2 \widehat{C}\left(F_{t,T}, K, t, T, \sigma(K,T)\right)}{\partial K^2}|_{K=F_T}. \tag{6}$$

Let the return of longing a futures contract maturing at T be $R_{t,T} = \log(F_T/F_{t,T})$. We can obtain the density of futures return over the period of $T-t$ as

$$\begin{aligned} \widehat{P}^Q(R_{t,T}|\mathcal{F}_t) &= \frac{\partial}{\partial u}\Pr\left(\log(F_T/F_{t,T}) \leq u|\mathcal{F}_t\right) = \frac{\partial}{\partial u}\Pr\left(F_T \leq F_{t,T}\exp\left(u\right)|\mathcal{F}_t\right) \\ &= \widehat{P}^Q(F_{t,T}\exp\left(u\right)|\mathcal{F}_t) \cdot F_{t,T}\exp\left(u\right), \end{aligned} \tag{7}$$

where Pr(.) denotes the cumulative distribution function. We compute a fixed one-month (21 business-day or 30 calendar-day) horizon option-implied density by interpolating the term structure of density (7) on each day. Alternatively, we obtain the risk-neutral density of futures returns using the nonparametric approach from Aït-Sahalia and Duarte (2003) and Li and Zhao (2009), which we have not reported here. Results are qualitatively similar. Our estimation spans a long period while keeping the parametric assumptions to a minimum, compared with the existing literature in the commodity market (e.g., Melick and Thomas [1997] estimate the risk-neutral distribution from crude oil options around the time of the first Gulf war under restrictive lognormal assumptions).

2.3 Estimation of the Physical Density and the SPD

Once we have obtained the risk-neutral density $\widehat{P}^Q(R_{t,T}|\mathcal{F}_t)$ from option prices, the other component needed to compute the SPD is the physical density $\widehat{P}(R_{t,T}|\mathcal{F}_t)$. We estimate the physical density from historical futures prices. Estimation of the historical distribution needs to take two practical factors into account. First, one needs to use a time series of data as long as possible in order to increase precision of estimates. The longer the sample period is, the more efficient estimator we can obtain. Second, the estimation methodology needs to account for the potential time-varying nature of physical density, especially to allow for the presence of stochastic volatility in the crude oil market as documented by Trolle and Schwartz (2009). Bansal, Kiku, Shaliastovich, and Yaron (2014) further highlight the importance of time-varying volatility when estimating SPDs from financial market data.

We first calculate the time series of daily futures returns $\{R_{t,T}\}_{t=1}^{N}$ from 1990 to 2012 using futures prices. It is equivalent to the continuously compounded returns of holding the futures contract to maturity and realizing returns by closing out the position at the maturity date T. At each time t, we normalize the time series of returns with its sample mean $\overline{R}$ and conditional volatility σ_t, the estimation of which is described as below. This gives a time series of return innovation $\{z_{t,T}\}_{t=1}^{N}$, defined as $(\{R_{t,T}\}_{t=1}^{N} - \overline{R})/\sigma_t$. Then, similar to Jackwerth (2000), we estimate the density with a kernel function using the return innovation at t. The physical density of returns is then obtained by $\widehat{P}(R_{t,T}|\mathcal{F}_t) = \widehat{P}(\overline{R} + \sigma_t \cdot z_{t,T})$.

We utilize high-frequency intraday oil futures prices to estimate volatility σ_t. Andersen and Bollerslev (1998) and others have shown the superior property of volatility estimated from intraday high frequency data compared to daily data. We calculate daily volatility using the two-scale estimation approach, which Andersen, Bollerslev and Meddahi (2011) have shown to be robust to the impact of microstructure noise in the high-frequency data.

In order to match the forward looking horizon of option-implied risk neutral density, we compute expected one-month volatility using the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). We first estimate the regressions of

$$Vol_t = a + b_d Vol_{t-1:t} + b_w Vol_{t-5:t} + b_m Vol_{t-21:t} + e_t,$$

where $Vol_{t-1:t}$, $Vol_{t-5:t}$, and $Vol_{t-21:t}$ denote the most recent daily, weekly, and monthly volatility, respectively. Then we use the HAR regression to predict volatility for the next month $\sigma_t = E_t\left[Vol_{t+1:t+h}\right]$, with h equal to 21. The HAR regressions have been used in many studies including Busch, Christensen, and Nielsen (2011) to forecast volatilities in various financial markets. We estimate the HAR regression coefficients using a rolling window of 250 days. This estimation of volatility is free from the look-ahead bias, since at any time t we only use realized historical information.

Finally, we interpolate the physical density of returns onto the same spacing as the risk-neutral density so that the SPD in the crude oil market is obtained by

$$\widehat{\xi}(R_{t,T}|\mathcal{F}_t) = e^{-r(T-t)} \frac{\widehat{P}^Q(R_{t,T}|\mathcal{F}_t)}{\widehat{P}(R_{t,T}|\mathcal{F}_t)}. \tag{8}$$

Our estimated SPD in the crude oil market is the discounted ratio of option-implied density and physical density estimated from options and futures prices.

2.4 Futures and Option Data

Our crude oil futures and option data are from two sources. We obtain daily crude oil futures and option data from January 2, 1990 to December 31, 2012 from the Chicago Mercantile Exchange (CME Group, formerly NYMEX); and we get the high-frequency intraday oil futures prices from TickData. Crude oil traded on the CME is the largest and most liquid commodity. The range of maturities of futures, and the range of strike prices of options are also larger than other commodity derivatives. An advantage of the data is that crude oil futures and options have been traded on this exchange for more than 20 years, which allows us to study a long time series spanning recessions and many geopolitical events such as the gulf wars, the 9/11 terrorist attacks, the recent financial crisis, and especially the recent boom and bust of the commodity markets. This dataset also provides open interest and trading volume of both options and the underlying futures contracts.[6] Relative demand and trading volume in futures and options reveal investor beliefs and expectations (e.g., Buraschi and Jiltsov, 2006), and therefore are informative when studying the SPDs.

The calculation of the risk-neutral density in (7) is based on European options, but the crude oil option data are American type. We convert American option prices into European option prices following Trolle and Schwartz (2009) who use the methodology of Barone-Adesi and Whaley (1987). After obtaining European option prices, we exclude those observations with Black (1976) implied volatility less than 1% or greater than 200%; we exclude those options with prices less than \$0.05 and contracts violating standard no-arbitrage constraints. The empirical analysis is at the weekly frequency and uses OTM calls and puts. Using OTM options is due to two motivations: First it minimizes the effect of possible approximation errors in the early exercise premium; and second, OTM options are usually more liquid than in-the-money (ITM) options. Each week, we use Wednesday since it is the day least likely to be a holiday during a week. In addition, it is also less likely to be affected by day-of-the-week effects. This selection of data has been widely used in the literature (e.g., Bates, 1996,

[6] Crude oil futures (option) trading volume data are missing from December 15, 2006 (December 1, 2006) to May 21, 2007 due to technical reasons when the CME group converted data from the NYMEX database. But futures and option price and open interest data are available throughout the entire period.

2000; Heston and Nandi, 2000). Since the calculation of risk-neutral density is based on call prices, we utilize OTM puts and transform them into ITM calls. Together with observed OTM calls, call option data effectively span the entire moneyness to apply formula (4).

Panel A of Table 1 provides descriptive statistics for the futures data by maturity. Although the number of contracts and average prices are relatively constant across maturities, average open interest and trading volume decrease sharply beyond the two-month maturity. While open interest of six-month futures contracts is around 20% of one-month contracts, the trading volume of six-month contracts is only about 4% of one-month contracts. This shows that long maturity futures often lack liquidity, which is also true for options as reported in Panel B. Trading volume of all option contracts beyond six-month (or 180 calendar days) is only about 5% of one-month contracts. Panel C of Table 1 reports option data across moneyness. We observe that although deep OTM (ITM) options have large amount of contracts, at-the-money options are most heavily traded. Across moneyness, the average Black (1976) implied volatility displays a smile pattern with deep OTM (ITM) options having higher implied volatility than ATM options. Across maturities, short maturity options on average have a higher implied volatility than long maturity options as shown in Panel B.

2.5 Empirical SPDs in the Crude Oil Market

2.5.1 The U-shaped SPDs

We compute conditional risk-neutral densities and physical densities on each Wednesday using the approach we discussed in sections (2.2) and (2.3). In unreported results, we observed that while the risk-neutral density can be either negatively or positively skewed, the physical density is relatively more symmetric, and the risk-neutral density has fatter tails than the physical density. When we compute the time series of skewness and excess kurtosis, risk-neutral skewness (or kurtosis) dominates physical skewness (or kurtosis); the magnitudes of physical skewness and excess kurtosis are only a small fraction of the ones in risk-neutral distributions. Therefore, the shape of the SPD, which we calculate as the log ratio of the risk-neutral and physical densities, is mainly driven by the risk-neutral density. Figure 1 shows the estimated SPDs as a function of futures returns in the crude oil market on each Wednesday for one-month maturity for the years 1998 to 2012. The horizontal axis denotes returns, and the sample year is indicated in the title of each graph.

Figure 1 shows that the SPDs are nonlinear and in general display an asymmetric U-shape as a function of returns. At the aggregate level, investors in the crude oil market regard the states with extremely low returns or extremely high returns as bad states and assign a high value for payoffs received in those states. This might be due to the heterogeneous nature of investors in the crude oil derivative market: Investors (such as net long investors) who have net long futures positions, will bear losses in the case of futures price decreasing if their positions are not protected. They regard negative returns as bad states and highly value payoffs received in these states. Investors (such as net short investors) who hold net short futures positions, will suffer from increasing futures price and consider those states with positive returns as bad states. They assign a high value to payoffs received when oil futures returns are extremely high.

The U-shaped SPDs display dramatic variations across time. First, we observe that across time investors assign different values of state price (per unit probability) to the same level of returns. For example, for a given level of returns, the state price is higher in 2003 than in 2005, which means investors could have priced a higher payoff for exposure of the same level of negative or positive returns in 2003 than in 2005. The asymmetric U-shape is also wider in the year of 2008-2009 than in other years. Second, the U-shape has different level of dispersion across time. For a given range of returns, the change of state prices significantly differs. For example, consider the year of 2000 and 2007: while investors assign similar state price to certain returns and have constant preferences towards those returns during 2000, their value for the same level of returns varies much more during 2007. Third, the dispersion in both the decreasing and the increasing regions starts to increase around 2004 or 2005. The slopes in both the decreasing and increasing regions can become steeper or flatter depending on the state of nature. In other words, investors' marginal value of payoffs in difference states, when returns are negative or positive, exhibits significant variations across time.

2.5.2 Option Returns Support the U-shaped SPDs

Since the SPDs are computed as the ratio of risk-neutral and physical densities. They are not necessarily unbiased even though densities are estimated correctly. Therefore, we next use another model-free approach, which is returns of OTM puts and calls, to show the SPDs are non-monotonic and U-shaped. Generally investors in the crude oil futures market can hedge potential losses due to low (high) oil prices using OTM puts (calls). In particular, if net long investors are worried about extreme negative oil futures returns, they will demand OTM puts. OTM puts will be expensive and their returns will be negative and increase with strike prices since deep OTM puts (with low strike prices) provide protection for extremely bad states for long investors. If net short investors consider extremely positive returns as bad states, they will demand OTM calls. Returns of OTM calls will be negative and will decrease with strike prices since deep OTM calls (with high strike price) protect large losses due to high oil prices. Therefore, by analyzing returns of OTM puts and OTM calls, i.e.

those assets pay off when futures have large negative and positive returns, we can detect whether investors generally regard both extreme low and high oil futures prices as bad states of economy, which supports the SPDs with both decreasing and increasing regions.

We compute option returns as follows. On the third Wednesday of every month, we take a long position in available calls and puts with maturity as close to 30 calendar days as possible. We compute the hold-to-maturity returns of calls and puts as

$$\begin{aligned} r_{t,T}^{call} &= max(F_{t,T}e^{R_{t,T}} - K, 0)/c_{t,T} - 1, \\ r_{t,T}^{put} &= max(K - F_{t,T}e^{R_{t,T}}, 0)/p_{t,T} - 1, \end{aligned} \tag{9}$$

where $F_{t,T}$ is the price of underlying futures, $R_{t,T} = \log(F_T/F_{t,T})$ is the return of underlying futures contract over the period $T-t$, K is the strike price, $c_{t,T}$ and $p_{t,T}$ are prices of European style call and put options. We use observed option prices and we do not create artificial prices by interpolation or extrapolation. On each considered Wednesday, we assign available options to various bins according to their moneyness, defined as $X = K/F_{t,T}$, and returns are averaged within each bin. This procedure provides a non-overlapping return time series with various moneyness. In Table 2 we show the average return, its standard t-statistics, the 10^{th}, 50^{th} and 90^{th} percentile in each moneyness bin from January 1990 to December 2012, as well as average returns for two sub-sample periods: pre-financialization January 1990 to December 2004, and post-financialization January 2005 to December 2012.[7] We consider five bins of moneyness: $(0.5, 0.85]$, $(0.85, 0.95]$, $(0.95, 1.05]$, $(1.05, 1.15]$ and $(1.15, 1.5)$. OTM puts have moneyness of $(0.5, 0.85]$ and $(0.85, 0.95]$; OTM calls have moneyness of $(1.05, 1.15]$ and $(1.15, 1.5)$. We are mostly interested in returns of deep OTM puts ($X \in (0.5, 0.85]$) and deep OTM calls ($X \in (1.15, 1.5)$).

Panel A of Table 2 reports put option returns. For the period of 1990 to 2012, put options mostly have negative returns and returns increase across moneyness. Negative returns of OTM puts have higher magnitude than ITM puts and are statistically significant. When we compare the difference between OTM and ITM puts, returns from OTM puts are statistically lower and more negative. Call option returns are somewhat different from put returns as shown in Panel B of Table 2. There is no clear monotonic pattern of returns across moneyness. Although ITM and ATM call options can have positive returns, returns of deep OTM calls ($X \in (1.15, 1.5)$) are negative and statistically significant. In addition, returns from OTM calls are always lower than returns from ITM calls. Combining the negative returns of OTM puts and OTM calls in Panels A and B, we conclude that both deep OTM

[7] Our results remain if we directly work on American option prices or if we date the financialization back to 2004.

puts and deep OTM calls are more expensive than other options.

When we compare the pre- and post-2005 periods, we find that the negative return pattern of put options becomes weaker in the post-2005 period than in the pre-2005 period, since only deep OTM puts have significantly negative returns after 2005. Interestingly, call option returns *decrease* across moneyness after 2005, which is not compatible with the monotonically downward-sloping SPDs. OTM call options become more expensive in the post-2005 period as returns of OTM calls are more negative, which implies that investors become worried about positive oil futures returns. This could be because short position speculators (who do not have natural hedge and use options to hedge their futures positions) have exposure to the futures price risk, and demand OTM call options to hedge.[8]

We next conducted two more analyses. The first analysis is to check whether the expensiveness of OTM calls (and puts) is due to illiquidity premia. We do not find supporting evidence for this. In our sample period, both trading volume and the ratio of trading volume over open interests of OTM calls (puts) are higher than ITM calls (puts), and OTM calls and puts are more liquid than ITM calls and puts. Therefore, OTM calls and puts are actively traded, and the negative returns cannot be imputed to illiquidity. The second analysis is the return of the butterfly spread. On the third Wednesday of every month, we take a long position in the butterfly spread constructed from call options with maturity close to 30 calendar days. We buy an ITM call option and an OTM call option with strike prices K_1 and K_3, and we sell two ATM call options with a strike price of K_2 and $K_2 - K_1 = K_3 - K_2$. We find that average return of the butterfly spread is significantly negative when K_3 is above certain threshold. It is consistent with the above argument that OTM call options are more expensive relative to other calls, which supports the increasing SPDs in the region of large positive returns. The fact that returns of call options *decrease* with strike prices, along with negative returns of the butterfly spread, supports the non-monotonic SPDs (Chaudhuri and Schroder, 2015).[9]

In summary, we find that put option returns are negative and increase across the strike price. OTM call option returns are negative and decrease across strike prices when strike prices are high enough, especially after 2005. Negative returns of OTM puts and calls imply that investors are worried about both extreme negative and positive oil futures returns and consider both as bad states, which lends support to the U-shaped SPDs; and the different

[8] In general speculators have net long positions in aggregate to provide liquidity for commodity producers (hedgers). However, there are certain number of short position speculators, such as the Morningstar Short/Flat Commodity Index and the Morningstar Short-Only Commodity Index. In addition, some excessive long positions of financial traders need short positions from other traders, who may hedge themselves.

[9] Chaudhuri and Schroder (2015) show that the SPDs are monotonically downward-sloping *if and only if* returns of calls *increase* in the strike price.

pattern of put returns and call returns supports the asymmetry of the U-shaped SPDs. Our evidence of negative returns on OTM call options is consistent with the model of Bakshi, Madan and Panayotov (2010) where investors have heterogeneous beliefs and can take short positions.

2.6 Option Implied Risk-Neutral Moments

We are not only interested in the empirical shape of SPD in the crude oil market, but also how SPDs are affected by investors' heterogeneous beliefs. This is related to the economic question of how much more investors in the aggregate are willing to pay for securities in one state of economy over another. Since the shape of the SPDs is mainly driven by the properties of the risk-neutral distribution, we compute risk-neutral moments and link their time variations to investor beliefs and trading activities.[10] Risk-neutral moments are calculated from the estimated risk-neutral densities:

$$Var_{t,T} = E_t^Q\left[\left(R_{t,T} - E_t^Q[R_{t,T}]\right)^2\right] \tag{10}$$

$$Skew_{t,T} = \frac{E_t^Q\left[\left(R_{t,T} - E_t^Q[R_{t,T}]\right)^3\right]}{\left\{E_t^Q\left[\left(R_{t,T} - E_t^Q[R_{t,T}]\right)^2\right]\right\}^{3/2}} \tag{11}$$

where $E_t^Q[x] = \int_{-\infty}^{\infty} xP^Q(x)dx$ is the expected value under the risk-neutral measure and $P^Q(x)$ is the density estimated from option prices. Since risk-neutral moments are calculated using the option implied densities and available returns on each Wednesday, they are model free and conditional.

Figure 2 displays the time series of weekly option-implied variance (the upper panel) and skewness (the lower panel) from 1990 to 2012. From left to right, the first two vertical dotted lines denote two significant days of the first Gulf War: the Iraq invasion of Kuwait on August 2, 1990, and the liberation of Kuwait on January 17, 1991. Other vertical dotted lines denote the September 11, 2001 terrorist attacks; the second Gulf War on March 20, 2003; the week when the WTI crude oil spot price reached its highest level in history (July 23, 2008 [$145.31]); the week when the Lehman Brothers filed for Chapter 11 bankruptcy protection (September 15, 2008); the week when the crude oil spot price reached its lowest level during the recent crisis period (December 23, 2008 [$30.28]); the week when the Libyan

[10] Datta, Londono and Ross (2014) investigate how option-implied moments change around important events in the crude oil market.

Civil War began and oil and gas production in Libya fell by more than 60%; and the week when Standard & Poor's downgraded the U.S. long-term sovereign credit rating from AAA to AA+, respectively. We note that variance rises sharply on those days. Consistent with Robe and Wallen (2014), these sharp hikes in the upper panel of Figure 2 indicate that oil variance is affected by not only the oil market fundamentals, but also macroeconomic conditions. However, when we compare the upper and lower panels of Figure 2, it appears that skewness reacts more to oil market-specific information. For example, skewness significantly increases during the week of the second Gulf War in March 2003 and the week of Libyan War in February 2011, but it does not change much when the U.S. sovereign credit is downgraded.

The empirical results so far have shown that the SPDs in the crude oil market display an asymmetric U-shape as a function of returns, and exhibit remarkable variations across time. Returns of OTM options are consistent with this nonlinear pattern of the SPDs. We also find some preliminary evidence that risk-neural skewness of oil futures returns is more likely to be affected by oil market-specific information. We next investigate how the shape of the SPDs depend on investors' heterogeneous beliefs.

3 Investor Beliefs and the SPDs

The literature has shown that the SPDs depend on investor disagreements. Heterogeneity is represented not only in asset prices and returns, but also in the relative positions and trading volumes in equilibrium. Dependence of the SPDs on heterogeneous beliefs is present in the equity market (e.g., Anderson, Ghysels and Juergens, 2005; Buraschi and Jiltsov, 2006), as well as in the foreign currency market (Beber, Buraschi and Breedon, 2010). When agents have different beliefs about asset prices, they engage in trading either for speculation or hedging.

How do investor beliefs affect the SPDs in the crude oil market? In this section, we investigate how investor beliefs embedded in crude oil derivatives trading affect the slope of the SPDs, because it is one of the fundamental characteristics of the SPDs and can be directly related to investors' risk aversion. The steeper the slope of SPDs is, the higher state prices investors assign to economic states of more extreme returns. We first describe how investors' trading positions have evolved over the past twenty years in crude oil futures and options, and the measures of investor beliefs in both the crude oil futures and option markets. Subsequently, we document that the slopes of the SPDs in both decreasing and increasing regions are affected by investor beliefs. We further substitute the slope with risk-neutral skewness and investigate the impact of heterogeneous beliefs.

3.1 Market Participation and Measure of Beliefs

The CFTC classifies large traders in the crude oil derivatives market into commercial traders or hedgers and non-commercial traders or speculators.[11] Figure 3 shows long and short positions taken by hedgers and speculators in the futures market as well as their ratios, which we obtain from the CFTC futures-only Commitments of Traders (COT) report. Although participation in the futures market by both hedgers and speculators has experienced steady growth from 1990 onwards, the increases in positions have been faster since 2004-2005, as shown in the top panel of Figure 3. When we look at the ratio of long positions taken by speculators over hedgers, we see the ratio typically fluctuates around 0.2 before 2004-2005, but it has constant growth afterwards. The ratio of short positions also has steady growth, although not as substantial as the ratio of long positions. The rapid growth of positions from speculators and the historical boom and bust of futures price in 2008 (as shown in the bottom panel of Figure 3) have drawn significant attention from the academicians, practitioners and policy makers. Some literature has attributed position growth of speculators to commodity linked investments from financial institutions.

The crude oil option market has experienced even more dramatic growth. Figure 4 shows the number of OTM options and their open interests and trading volumes aggregated within each week. OTM calls are defined as call options with moneyness (K/F)>1.05, and OTM puts are put options with moneyness (K/F)<0.95. We only consider options with maturity less than 180 calendar days. We also report the weekly average annualized 30 calendar-day fixed maturity oil option-implied volatility, which is constructed using the methodology of CBOE's VIX. While there are an increasing number of OTM options traded in the market, the growth of the number of options after $2004-2005$, especially in the first half of 2008 is indeed dramatic. In particular, the number of OTM call option contracts jumps in 2008 as shown in the top panel of Figure 4. When oil prices are most volatile, and investors demand OTM options to hedge their positions. We observe that open interests of options have certain decrease after 2010, as shown in the mid panel. Given the active participation in the futures and option markets, we consider measures of investor beliefs in both markets and investigate how they affect the SPDs. We will focus on those variables measuring the level of belief heterogeneity, as well as those variables that directly relate to either net long or net short positions. Although these measures are different, they are all broadly related to investors beliefs, as suggested by the literature.

[11] We acknowledge that this classification of hedgers and speculators is not perfect and does not cover detailed information about all investors in the crude oil derivatives market. However, Büyükşahin and Robe (2014) confirm that the speculative activities inferred from the public CFTC position data are able to capture speculative activities inferred from non-public trader-level CFTC data well.

First we use the speculation index, which quantifies the level of speculation activity, as a measure of investor beliefs in the crude oil market. Several studies such as Scheinkman and Xiong (2003) and Xiong and Yan (2010) argue that agents with heterogeneous beliefs engage in speculative trading among each other. The speculation index is to gauge intensity of speculation relative to hedging (Working, 1960; Büyükşahin and Harris, 2011). If we denote SS (SL) as the speculative short (long) position, and HS (HL) as the hedging short (long) position, we define

$$\text{Speculation Index}_t = \begin{array}{l} 1 + \frac{SS_t}{HL_t + HS_t} \text{ if } HS_t \geq HL_t, \\ 1 + \frac{SL_t}{HL_t + HS_t} \text{ if } HS_t < HL_t. \end{array} \tag{12}$$

Since speculators take positions in crude oil futures by anticipating certain price movements, this speculation index contains information on belief differences among investors.[12] The speculation index has to be interpreted together with hedging activities in the futures market. It measures the extent by which speculative positions exceed the necessary level to offset hedging positions. Panel A of Figure 5 plots the speculation index from 1990 to 2012. We observe that there has been a high level of speculative activities in recent years. Before 2000s, the speculation index is below 1.05 meaning less than 5% of excessive speculation; however, it rises steadily over time to 1.2 in 2010 and drops to around 1.1 afterwards. As such, the premia that hedgers pay for insurance against futures price risk are highly affected by the active participation of speculators.

Besides positions taken by speculators, actual trading volume of futures may reflect the degree of heterogeneity and how investors speculate and share risks among each other. The literature (e.g., Kandel and Pearson, 1995) has documented the positive relationship between investor heterogeneous beliefs and volume of trade. Buraschi and Jiltsov (2006) show that, the trading volume of stocks and options is positively correlated with the dispersion in beliefs. Carlin, Longstaff, and Matoba (2014) show that increased disagreement is associated with higher trading volume. Therefore, we use the trading volume of futures as another measure of investor belief dispersion in the underlying futures market. Panel B plots the trading volume of 30-day futures contracts, which grows steadily during the entire period but becomes very volatile during the recent few years.

Next we discuss two measures of heterogeneous beliefs in the option market used by Han (2008). One is the open interest ratio of OTM puts to calls, which measures the relative demand for insurance against downside risks and reflects the hedging needs of heterogeneous agents. A higher open interest ratio of OTM puts to calls suggests investors are overall more

[12] Singleton (2014) discusses how disagreements among investors induce speculative activities, price drift, and high volatility in the crude oil market.

pessimistic, and they tend to demand put options either to get protection against future price drops or to pursue potential returns on put positions. Panel C shows that the open interest ratio of OTM puts to calls having several spikes during our sample period. The other measure is the trading volume of options, which is a proxy for the level of disagreement among options investors. Open interests and trading volume of options do not capture the same information since open interests are the outstanding positions investors take, while trading volume can be due to opening or closing a contract. However, to separate the factor that specifically affects net long and net short futures investors, we consider trading volume of OTM puts and OTM calls individually. Panels D and E shows that, the trading volume of OTM puts and OTM calls is relatively stable but starts to rise steadily from 2004-2005. Because the trading volume of futures and options from December 2006 to March 2007 is missing, we interpolate and obtain a complete time series in the subsequent regression analysis. We also adjust the time series of futures and option trading volume with a Hodrick-Prescott filter following the literature. Since some trades of the near maturity futures and option contracts are due to liquidation or rollover to the next maturity contract, trading volumes and open interests measured from the near contracts may have little information content. We therefore interpolate the closest-to-maturity contract and the next available contract, and obtain trading volume and open interests with a maturity of fixed 30 calendar days, which also matches our time-horizon of the SPDs and risk-neutral moments.

To what extent do investor beliefs of the economy and the equity market affect the SPDs in the crude oil market? We consider two types of proxies of beliefs to address this question: investor beliefs about the equity market and economy, as well as investors' expectations about the market volatility. The first measure is the bull-bear spread based on the survey of Investors Intelligence which has been used by Brow and Cliff (2004, 2005) and Han (2008). Every week, Investors Intelligence sends out 150 surveys to institutional investment advisors and collects their expectations of future market movements as bullish, bearish, or neutral. The bull-bear spread is then calculated as the percentage of bullish investors minus the percentage of bearish investors, and it is often used as a proxy for beliefs of institutional investors about the equity market. Secondly, we consider the consumer sentiment index from the University of Michigan as plotted in Panels G.[13] The literature has used this index to study how sentiment affects stock prices (e.g., Lemmon and Portniaquina, 2006; Stambaugh, Yu, and Yuan, 2012). The third proxy we use to measure investor beliefs is the CBOE's VIX index. VIX has become a benchmark for measuring investors' expectations of market volatility and investor sentiment as a fear indicator. As plotted in Panels H and F, VIX is in

[13]When we include the investor sentiment index in Baker and Wurgler (2006), which restricts our analysis up to 2010, our main results remain similar.

general negatively correlated with the bull-bear spread. The VIX index tends to be higher when more investors are less confident of the market.

3.2 Other Control Variables

We consider two sets of control variables in the regression analysis. One set includes oil market-specific variables that determine futures returns and therefore may affect the SPDs. We first follow Hong and Yogo (2012) and calculate the basis as

$$\text{Basis}_{t,T} = (\frac{F_{t,T}}{S_t})^{\frac{1}{T-t}} - 1, \tag{13}$$

where $F_{t,T}$ is the price of futures with maturity of $T - t$ at time t, S_t is the spot price at time t. Basis is the spread between futures prices and spot prices, and can be interpreted as the implied net convenience yield.

Next we include option implied volatility due to the following reasons. First, Trolle and Schwartz (2009) present strong evidence of stochastic volatility in the crude oil market; and stochastic volatility has to be incorporated into the model to price commodity derivatives. Therefore, volatility could affect futures prices and return distributions. Second, contrary to the traditional theory of storage, which claims volatility tends to be high when the futures price is in backwardation, Carlson, Khokher, and Titman (2007) and Kogan, Livdan, and Yaron (2008) document that the relationship between the futures term structure slope and volatility is non-monotonic. While the sign of the slope of the futures term structure implies investors' demand for the convenience yields, volatility captures complementary information about investors' exposure to risk.

We also consider the storage level of crude oil and historical returns of futures, both of which affect futures returns as documented in the literature (e.g., Bessembinder, 1992; Bessembinder and Chan, 1992; de Roon, Nijman, and Veld, 2000). We use the storage level defined as U.S. total stocks of crude oil, excluding strategic petroleum reserves, based on the report from Energy Information Administration (EIA). Following Acharya, Lochstoer, and Ramadorai (2013), a Hodrick-Prescott filter is applied to remove the trend, where the smoothing parameter is set to a number appropriate for weekly data. The historical return is the moving average of daily return of holding a long position of the futures contract over the previous one week, since our regression is weekly.

The other set of control variables captures the macroeconomic condition and investment opportunities. They include 1-month treasury-bill rate, the spread between yields on the 10-year treasury bond and the 3-month treasury bill, the difference in yields on Baa and Aaa corporate bonds, the Chicago Fed National Activity Index and the log growth in aggregate industrial production over the last 12 months. These variables are used by Han (2008) and others. We obtain the one-month treasury bill rate from the online data library of Kenneth French, and other macro variables are from the Federal Reserve Bank of Chicago and St. Louis.

Table 3 reports the summary statistics of risk-neutral skewness computed from (11), main regressors and control variables in the crude oil market. We apply necessary filters to some regressors to remove time trends. The last column of Table 3 show that the unit-root tests are rejected for all time series we use in regressions at the 5% significant level. Note that overall skewness is negative; investors on average demand more OTM puts than OTM calls; and historical annualized return of holding the one-month futures contract is 7.56% with a sharpe ratio of 0.47. Table 4 reports the correlation matrix of regressors, as well as control variables in the crude oil market. We observe that all correlations are rather low, except for oil volatility (IVOil) having a correlation of 0.65 with VIX, and OTM put/call ratio having a correlation of -0.41 with basis. Given the high correlation between IVOil and VIX, IVOil is orthogonalized with VIX before being included in the regression analysis. That is, we use the residual in the regression of $IVOil_t = a + bVIX_t + e_t$. We also notice that the trading volumes of futures, call and put options are positively correlated among each other.

3.3 Regression Results

We consider three levels of measures to examine how investor beliefs affect the slope of the SPD (which characterizes the marginal rate of substitution across states) in the crude oil market: 1) belief dispersion in the crude oil futures market; 2) belief dispersion in the crude oil option market; and 3) investor beliefs of the equity market and the economy. In our regression, we add control variables discussed above that may affect crude oil futures returns and reflect macroeconomic conditions. The full regression model is given by

$$\begin{aligned} \text{Slope of the SPD}_t \quad = \quad & \alpha_t + \beta_t \cdot \text{Beliefs in the crude oil futures market}_t \\ & + \gamma_t \cdot \text{Beliefs in the crude oil option market}_t \\ & + \delta_t \cdot \text{Beliefs in the equity market}_t \\ & + \eta_t \cdot \text{Control variables}_t. \end{aligned} \tag{14}$$

To assess the dependence of the slope of the SPD on investor beliefs, we run the regression (14) using different dependent variables. We first directly measure the slope of the SPD curve based on the value differences of the SPDs over two return points. Second, we substitute risk-neutral skewness for the direct slope measure, since the asymmetric U-shape pattern of the SPDs is mainly determined by the level of risk-neutral skewness. Before performing regressions, we standardize and winsorize all variables at the 1^{st} and 99^{th} percentile. To minimize the estimation noise due to limited option data, we only include those days with at least two OTM calls and two OTM puts in the regression sample to ensure the estimation of our dependent variable, risk-neutral skewness, is reliable.

3.3.1 Dependence of the Slope of the SPDs on Investor Beliefs

Since the SPDs in the crude oil have a time-varying U-shape, investors assign high state price to large negative and positive returns. The decreasing (increasing) region slope of the U-shaped SPDs captures investors' preferences towards different level of negative (positive) returns. The steeper the slope, the higher state price investors assign to extreme returns. On each Wednesday, we compute the slope as the difference between the value of the SPDs of two return points:

$$\begin{aligned} \text{Decreasing-Region Slope}_t &= \text{SPD}_t(R_2) \text{ - } \text{SPD}_t(R_1), \text{ where } R_2 < R_1 << 0; \\ \text{Increasing-Region Slope}_t &= \text{SPD}_t(R_2) \text{ - } \text{SPD}_t(R_1), \text{ where } R_2 > R_1 >> 0. \end{aligned}$$

This direct measure of slope shares the same idea with Xing, Zhang and Zhao (2010) who quantify the steepness of stock option smirks as the difference of implied volatilities of two moneyness points. To provide a broad and consistent picture of the slope, we consider two measures of the slope in both the decreasing- and increasing regions. The decreasing-region slopes are based on return intervals $(-0.20, -0.14)$ and $(-0.19, -0.13)$, and the increasing-region slopes are measured from return intervals of $(0.15, 0.18)$ and $(0.17, 0.20)$.

Table 5 presents regression results. It shows that futures trading volume negatively affects both the decreasing- and increasing-region slopes, which means state prices for both large negative and positive returns are lower when investors in the futures market trade more. The open interest ratios of OTM puts over OTM calls positively (negatively) affect the steepness of the decreasing- (increasing-) region slope. In other words, when investors demand more OTM puts, as suggested by higher OTM put open interests, they overall assign higher state price to negative returns and lower state price to positive returns. We also find that, if investors trade more OTM call options, the increasing region slope becomes steeper and investors assign higher state price to large positive returns. Although the results are not very strong, there is evidence that the level of speculation is positively (negatively) associated with the decreasing- (increasing-) slope, indicating that investors' marginal utility over negative returns increases (or they assign higher state price to negative returns) when there are more speculative activities. We also observe that coefficients of investor beliefs of the equity market and the economy are not robust.

Although we present the evidence that the slope of the SPD is affected by investor beliefs in the crude oil market, we have to discuss the decreasing- and increasing-region slope separately if we rely on this direct measure. In addition, there are multiple return intervals on each trading day and different choices of return intervals may change the magnitude of slope. Therefore, this measure of slope only represents investors' preference over certain return regions but not all possible economic states. Next, we argue that risk-neutral skewness is a more comprehensive measure of the slope and investors' preference, and our later analyses mainly focus on the risk neutral-skewness.

3.3.2 Option-Implied Skewness and Investor Beliefs

The asymmetric U-shape pattern of the SPDs arises from the differences between risk-neutral skewness and physical skewness. Because the skewness of the risk-neutral density has more pronounced time variations compared with the physical density, the slope of the U-shaped SPDs mostly depends on the skewness of the risk-neutral density. Furthermore, the risk-neutral skewness measure utilizes all data available with valid closing quotes. It provides a thorough picture of the shape of the SPDs and investors' overall attitude towards risk. Therefore, risk-neutral skewness is a more reliable and comprehensive measure of the slope (or the marginal rate of substitution across states). We next rely on risk-neutral skewness and investigate whether it is affected by investor beliefs. Risk-neutral skewness is calculated from risk-neutral densities and the formula (11). To assess the impact of speculative activities on the SPDs during the post-financialization period, we use a dummy variable that equals to one for the period of 2005 to 2012 and equals to zero for the period of 1990 to 2004. The coefficient of Speculation Index$*$Dummy will tell us whether speculation has different impacts on risk-neutral skewness or the SPDs during the two sub-period periods.

Table 6 contains the basic results of the dependence of risk-neutral skewness on belief measures in the oil futures and options market. In Panel A, the dependent variable is risk-neutral skewness. We include the lagged skewness to control for its positive autocorrelation in all specifications. Regression results show that the increasing level of speculation after 2005 is associated with a more negative risk-neutral skewness, i.e. the decreasing-region slope is steeper and investors' marginal utility over negative returns is higher. We observe that open interest ratio between OTM puts and calls is negatively related with skewness. A high open interest ratio of OTM puts to calls suggests investors are overall more concerned about potential large negative returns, and is therefore associated with high state prices of large negative returns. We also observe that OTM put volume (OTM call volume) is negatively (positively) associated with risk-neutral skewness, which means that higher OTM put volume (OTM call volumes) suggests a steeper decreasing- (increasing-) region slope of the SPDs, and investors are worried about large negative (positive) returns.

Panel B of Table 6 reports the regression results when we use the AR(1) residual of risk-neutral skewness as the dependent variable. In other words, we regress the residual of $Skew_t = a + b * Skew_{t-1} + \varepsilon_t$ on various measures of investor beliefs, which are also the residual from the AR(1) regressions. Results are consistent with Panel A. The higher the open interest ratio of OTM puts to calls and the higher the OTM put trading volume are, the more negative the risk-neutral skewness is and investors overall assign higher state prices to large negative returns. The higher the trading volume of OTM call options is, the more positive the risk-neutral skewness is and investors overall assign higher state prices to large positive returns. Different with Table 5, where we separately consider the decreasing-region and increasing-region slopes, the impact of futures trading volume on skewness is not significant in Table 6.

Two comparisons with the existing literature can be made here. First, the empirical evidence that the OTM puts to calls ratio affecting the risk-neutral skewness aligns with findings in the stock and index option literature (Dennis and Mayhew, 2002; Han, 2008). Since heterogeneous investors have different demands for OTM options due to their various expectations of market fundamentals, pessimists can share risks with others by buying insurance from optimists. The larger the difference in beliefs, the higher the demand for OTM puts which drives up the prices of options with low strike prices, and the distribution is therefore more negatively skewed. Second, consistent with Buraschi and Jiltsov (2006) we confirm that option trading volume can affect the slope of the SPDs as a proxy for dispersion of beliefs. OTM put (call) trading volume negatively (positively) affects risk-neutral skewness, meaning a higher level of belief dispersion of put (call) investors is related with steeper slope in the decreasing- (increasing-) region, or higher state price of more negative (positive) returns. This is also consistent with the expensiveness of OTM put and call options we document in Table 2; while net long investors futures investors use OTM puts to hedge large futures price drops, net short investors trade OTM calls to hedge their potential loss due to large futures price increases.

3.3.3 Robustness Check

To further illustrate the dependence of risk-neutral skewness on investor beliefs, we perform two types of robustness exercises. The first exercise is to include belief measures of the equity market and the economy, control variables in the crude oil market that can affect futures returns, as well as macroeconomic variables which represent economic conditions and investment opportunities, in the regression model (14). In the other robustness check, we replace the dependent variable by another model-free measure of risk-neutral skewness (Bakshi, Kapadia, and Madan 2003; BKM hereafter), and we check whether the results obtained from density-based skewness hold.

Table 7 reports regression results when we include other belief measures and control variables. For ease of comparison, we only report coefficients of the belief measures, i.e., β_t, γ_t, and δ_t in (14). The coefficients of the additional control variables are omitted for brevity. Our conclusions from Table 6 hold when we consider belief measures in the equity market and the economy in column (1). Belief measures in the equity market have no significant impacts on the risk-neutral skewness and the shape of the SPDs in the crude oil market. In column (2), we include control variables in the crude oil market, namely, basis, storage level, historical returns, and IVOil. In column (3), we further include macro control variables, which are 1-month treasury-bill rate, the spread between yields on the 10-year treasury bond and the 3-month treasury bill, the difference in yields on Baa and Aaa corporate bonds, the Chicago Fed National Activity Index and the log growth in aggregate industrial production over the last 12 months. These variables reflect the general economic condition and may affect the demand and supply of oil futures and therefore their expected returns. In column (4), we include both sets of control variables. Although adding control variables marginally increases R^2 in columns (2)-(4), the main results in column (1) and Table 6 still hold for both the significance level and the magnitude of coefficients of interest. We observe that the risk-neutral skewness is more negative (marginal value of payoffs at large negative returns are higher) when there are higher OTM put open interest over OTM calls and when investors trade more OTM puts; marginal value of payoffs at large positive returns are higher when there are higher level of belief dispersion in the call options. Lastly, a higher level of speculative activities after 2005 is related to more negative skewness and induces a higher probability of negative expected returns. It suggests that the increasing speculative activities after $2004-2005$ do have a significant impact on the futures market.

We next use another measure of risk-neutral skewness computed from OTM option prices (BKM, 2003). The numerical implementation of extracting risk-neutral moments follows Duan and Wei (2009). While our previous calculation of risk-neutral skewness in section (2.6) is based on the densities, risk-neutral skewness of BKM (2003) is calculated by only utilizing OTM option prices. Since these two methodologies of calculating moments are very different by nature, one cannot expect estimated moments to perfectly accord with each other. For example while the two measures of risk-neutral variance have a correlation of about 95%; the two measures of skewness have a correlation of 70.04% after 2005.

Table 8 reports the regression results using the alternative measure of risk-neutral skewness. All columns in Table 8 have the same model specification as the regressions in Table 7. The main conclusions still remain valid: Higher OTM put demand implies more negatively skewed risk-neutral densities. OTM put (call) trading volume is negatively (positively) associated with risk-neutral skewness and therefore the steepness of the decreasing- (increasing-) region slope of the SPDs. The significantly negative coefficients of Speculation Index$*$Dummy suggest that a higher level of speculation after the financialization around 2005 is associated with more negative skewness, indicating a higher marginal value of payoffs at large negative returns.

Overall, regression results indicate that the slope of the SPDs is dependent on investor beliefs embedded in futures and options trade; in aggregate investors assign higher state prices to more negative returns (i.e., the decreasing-region slope of the SPDs is steeper) when there are higher OTM put demand and higher OTM put trading volume. More speculative positions after the financialization around 2005 induce more negative risk-neutral skewness and reinforce this effect. Higher level of belief dispersion in the call options is associated with higher state prices in the increasing region, which means that the relative state price change with respect to positive returns becomes higher when investors overall trade more OTM calls. Investors' sentiment of the equity market such as bull-bear spread does not have a strong impact on the SPDs in the crude oil market. The findings are consistent and robust, regardless of whether we consider more control variables or use another measure of risk-neutral skewness. Our regression results based on risk-neutral skewness are also consistent with the results based on the direct measure of the slope of the SPDs using individual return intervals.

4 Conclusion

Using more than twenty years of futures and option data, we back out investor preferences towards different states of economy in the crude oil market. We characterize the time variation in the SPDs, which can be important in pricing contingent claims on crude oil. Moreover, we investigate the dependence of the SPDs on investor beliefs by utilizing the informative features of the crude oil derivatives market. Comparing with other financial markets, speculation data is available in this market and investors can take long or short positions based on their beliefs without being subject to regulations. We investigate how investor beliefs about oil futures prices affect the slope of the SPDs, which can be directly related to the marginal rate of substitution across states. Our regression analysis also attempts to examine the impact of speculations on the SPDs during the post-financialization period, when the level of speculative activities in the crude oil market is believed to be high.

We obtain three main results. First, we find that the SPDs in the crude oil market display a time varying U-shape pattern. This implies that investors with various hedging or speculative purposes regard both extreme low or high oil futures prices as bad states of economy. We show that returns on both OTM calls and puts are negative and are consistent with the U-shaped SPDs. Second, we document the dependence of the SPDs on investor beliefs. When there are higher demands for OTM puts, the decreasing-region slope of the SPDs is steeper, and investors assign higher state prices to more negative returns. The slope of the SPDs is also affected by belief dispersions measured by options trading volume. These findings are robust under different model specifications and when we use the alternative measure of risk-neutral skewness. Third, the increase in speculation after 2005 is associated with more negative risk-neutral skewness and higher marginal value of payoffs in those states of large negative returns.

The empirical findings in this paper suggest some extensions in the broader context of commodity markets. First, it would be interesting to develop theoretical commodity pricing models with heterogeneous beliefs. Prices are determined in an equilibrium where investors share risks and speculate based on their beliefs about the market. Sockin and Xiong (2014) spearhead this literature by analyzing how informational frictions, a source causing investors heterogenous beliefs, affect commodity prices. Second, our results highlight the importance of exploring the implications of heterogeneous beliefs for the relative demand and trading volume of commodity futures and options. When agents have different beliefs of commodity prices, they engage in trading derivatives either for hedging or speculation.

References

[1] Acharya, V., L. Lochstoer and T. Ramadorai, 2013. Limits to Arbitrage and Hedging: Evidence from Commodity Markets. *Journal of Financial Economics* 109, 441-465.

[2] Aït-Sahalia, Y. and A. Lo, 1998. Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *Journal of Finance* 53, 499-547.

[3] Aït-Sahalia, Y. and A. Lo, 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94, 9-51.

[4] Aït-Sahalia, Y. and J. Duarte, 2003. Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 116, 9-47.

[5] Andersen, T. and T. Bollerslev, 1998. Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *Journal of Finance* 53, 219-265.

[6] Andersen, T, T. Bollerslev, and N. Meddahi, 2011. Realized Volatility Forecasting and Market Microstructure Noise. *Journal of Econometrics* 160, 220-234

[7] Anderson, E., E. Ghysels, and J. Juergens, 2005. Do Heterogenous Beliefs Matter for Asset Pricing? *Review of Financial Studies* 18, 875-924.

[8] Baker, M. and J. Wurgler, 2006. Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61, 1645–1680.

[9] Bakshi, G., N. Kapadia, and D., Madan, 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. *Review of Financial Studies* 16, 101-143.

[10] Bakshi, G., D. Madan, and G. Panayotov, 2010. Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels. *Journal of Financial Economics* 97, 130-154.

[11] Bansal, R., D. Kiku, I. Shaliastovich and A.Yaron, 2014. Volatility, the Macroeconomy, and Asset Prices. *Journal of Finance* 69, 2471-2511.

[12] Barone-Adesi, G. and R. Whaley, 1987. Efficient Analytic Approximation of American Option Values. *Journal of Finance* 42, 301-320.

[13] Bates, D. 1996. Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies* 9, 69-107.

[14] Bates, D. 2000. Post-'87 Crash Fears in the S&P 500 Futures Option Market, *Journal of Econometrics* 94, 181-238.

[15] Beber, A. and M. Brandt, 2006. The effects of Macroeconomic News on Beliefs and Preferences: Evidence from the Options Market. *Journal of Monetary Economics* 53, 1997-2039.

[16] Beber, A., A. Buraschi, and F. Breedon, 2010. Differences in Beliefs and Currency Risk Premiums. *Journal of Financial Economics* 98, 415-438.

[17] Bessembinder, H. 1992. Systematic Risk, Hedging Pressure, and Risk Premia in Futures Markets. *Review of Financial Studies* 5, 637-667.

[18] Bessembinder, H. and K. Chan, 1992. Time Varying Risk Premia and Forecastable Returns in Futures Markets. *Journal of Financial Economics* 32, 169-193.

[19] Black, F. 1976. The Pricing of Commodity Contracts. *Journal of Financial Economics* 3, 167-79.

[20] Breeden, D. and R. Litzenberger, 1978. Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business* 51, 621-651.

[21] Brown, G. and M. Cliff, 2004. Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance* 11, 1-27.

[22] Brown, G. and M. Cliff, 2005. Investor Sentiment and Asset Valuation. *Journal of Business* 78, 405-440.

[23] Busch, T., B. Christensen and M. Nielsen, 2011. The Role of Implied Volatility in Forecasting Future Realized Volatility and Jumps in Foreign Exchange, Stock, and Bond Markets. J*ournal of Econometrics* 160, 48-57.

[24] Buraschi, A. and A. Jiltsov, 2006. Model Uncertainty and Option Markets with Heterogeneous Agents. *Journal of Finance* 61, 2841-2897.

[25] Büyükşahin, B. and J. Harris, 2011. Do Speculators Drive Crude Oil Futures Prices? *The Energy Journal* 32, 167-202.

[26] Büyükşahin, B. and M. Robe, 2014. Speculators, Commodities, and Cross-Market Linkages. *Journal of International Money and Finance* 42, 38–70.

[27] Carlin, B., F. Longstaff and K. Matoba, 2014. Disagreement and Asset Prices. *Journal of Financial Economics* 114, 226-238

[28] Carlson, M., Z., Khoker, and S. Titman, 2006. Equilibrium Exhaustible Resource Price Dynamics. *Journal of Finance* 42, 1663-1703.

[29] Chabi-Yo, F., 2012. Pricing Kernels with Stochastic Skewness and Volatility Risk. *Management Science* 58, 624-640.

[30] Chaudhuri, R., and M. Schroder, 2015. Monotonicity of the Stochastic Discount Factor and Expected Option Returns. *Review of Financial Studies,* forthcoming.

[31] Cheng, I.-H. and W. Xiong, 2014. The Financialization of Commodity Markets. *Annual Review of Financial Economics.*

[32] Chernov, M., 2003. Empirical Reverse Engineering of the Pricing Kernel, *Journal of Econometrics* 116, 329-364.

[33] Christoffersen, P. and K. Jacobs, 2004. Which GARCH Model for Option Valuation? *Management Science* 50, 1204-1221.

[34] Christoffersen, P., S. Heston, and K. Jacobs, 2013. Capturing Option Anomalies with a Variance-Dependent Pricing Kernel. *Review of Financial Studies* 26, 1962-2006.

[35] Corsi, F., 2009. A Simple Approximate Long Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7, 174–196.

[36] Datta, D., J. Londono and L. Ross, 2014. Generating Options-Implied Probability Densities to Understand Oil Market Events. Working Paper, Federal Reserve Board of Governors.

[37] Dennis, P. and S. Mayhew, 2002. Risk-Neutral Skewness: Evidence from Stock Options. *Journal of Financial and Quantitative Analysis* 37, 471-493.

[38] De Roon, F., T. Nijman, and C. Veld, 2000. Hedging Pressure Effects in Futures Markets. *Journal of Finance* 55, 1437-1456.

[39] Duan, J.-C. and J. Wei, 2009. Systematic Risk and the Price Structure of Individual Equity Options. *Review of Financial Studies* 22, 1981-2006.

[40] Etula, E. 2013. Broker-Dealer Risk Appetite and Commodity Returns. *Journal of Financial Econometrics* 11, 486-521.

[41] Goldstein, I., Y. Li, and L. Yang, 2013. Speculation and Hedging in Segmented Markets. *Review of Financial Studies* forthcoming.

[42] Hamilton, J. and J. Wu, 2014. Risk Premia in Crude Oil Futures Prices. *Journal of International Money and Finance* 42, 9-37.

[43] Han, B. 2008. Investor Sentiment and Option Prices. *Review of Financial Studies* 21, 387-414.

[44] Henderson, B., N. Pearson and L. Wang, 2014. New Evidence on the Financialization of Commodity Markets. Working paper, University of Illinois at Urbana-Champaign.

[45] Heston, S. and S. Nandi. 2000. A Closed-Form GARCH Option Pricing Model. *Review of Financial Studies* 13, 585-626.

[46] Hong, H. and M. Yogo, 2012. What Does Futures Market Interest Tell Us about the Macroeconomy and Asset Prices? *Journal of Financial Economics* 105, 473-490.

[47] Jackwerth, J. 2000. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies* 13, 433-451.

[48] Kandel, E. and N. Pearson, 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. *Journal of Political Economy* 103, 831-872.

[49] Kitsul, Y. and J. Wright, 2013. The Economics of Options-Implied Inflation Probability Density Functions. *Journal of Financial Economics* 110, 696–711

[50] Kogan, L., D. Livdan and A. Yaron, 2008. Oil Futures Prices in a Production Economy with Investment Constraints. *Journal of Finance* 64, 1345-1375.

[51] Lemmon, M. and E. Portniaquina, 2006. Consumer Confidence and Asset Prices: Some Empirical Evidence. *Review of Financial Studies* 19, 1499–1529.

[52] Li, H. and F. Zhao, 2009. Nonparametric Estimation of State-Price Densities Implicit in Interest Rate Cap Prices. *Review of Financial Studies* 22, 4335-4376.

[53] Linn, M, S. Shive and T. Shumway, 2014. Pricing Kernel Monotonicity and Conditional Information. Working Paper, University of Michigan.

[54] Melick, W., and C. Thomas, 1997. Recovering an Asset's Implied SPD from Option Prices: an Application to Crude Oil During the Gulf Crisis. *Journal of Financial and Quantitative Analysis* 32, 91-115.

[55] Newey, W. and K. West, 1987. A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.

[56] Robe, M. and J. Wallen, 2014. Fundamentals, Derivatives Market Information and Oil Price Volatility. Working paper, American University.

[57] Rosenberg, J. and R. Engle, 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64, 341-372.

[58] Scheinkman, J. and W. Xiong, 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111, 1183-1219

[59] Singleton, K., 2014. Investor Flows and the 2008 Boom/Bust in Oil Prices. *Management Science* 60, 300-318.

[60] Sockin, M. and W. Xiong, 2014. Informational Frictions and Commodity Markets. *Journal of Finance*, forthcoming.

[61] Song, Z. and D. Xiu, 2014. A Tale of Two Option Markets: Pricing Kernels and Volatility Risk. Working paper, University of Chicago.

[62] Stambaugh, R., J. Yu and Y. Yuan, 2012. The Short of It: Investor Sentiment and Anomalies. *Journal of Financial Economics* 104, 288–302.

[63] Tang, K. and W. Xiong, 2012. Index Investment and the Financialization of Commodities. *Financial Analysts Journal* 68, 54-74.

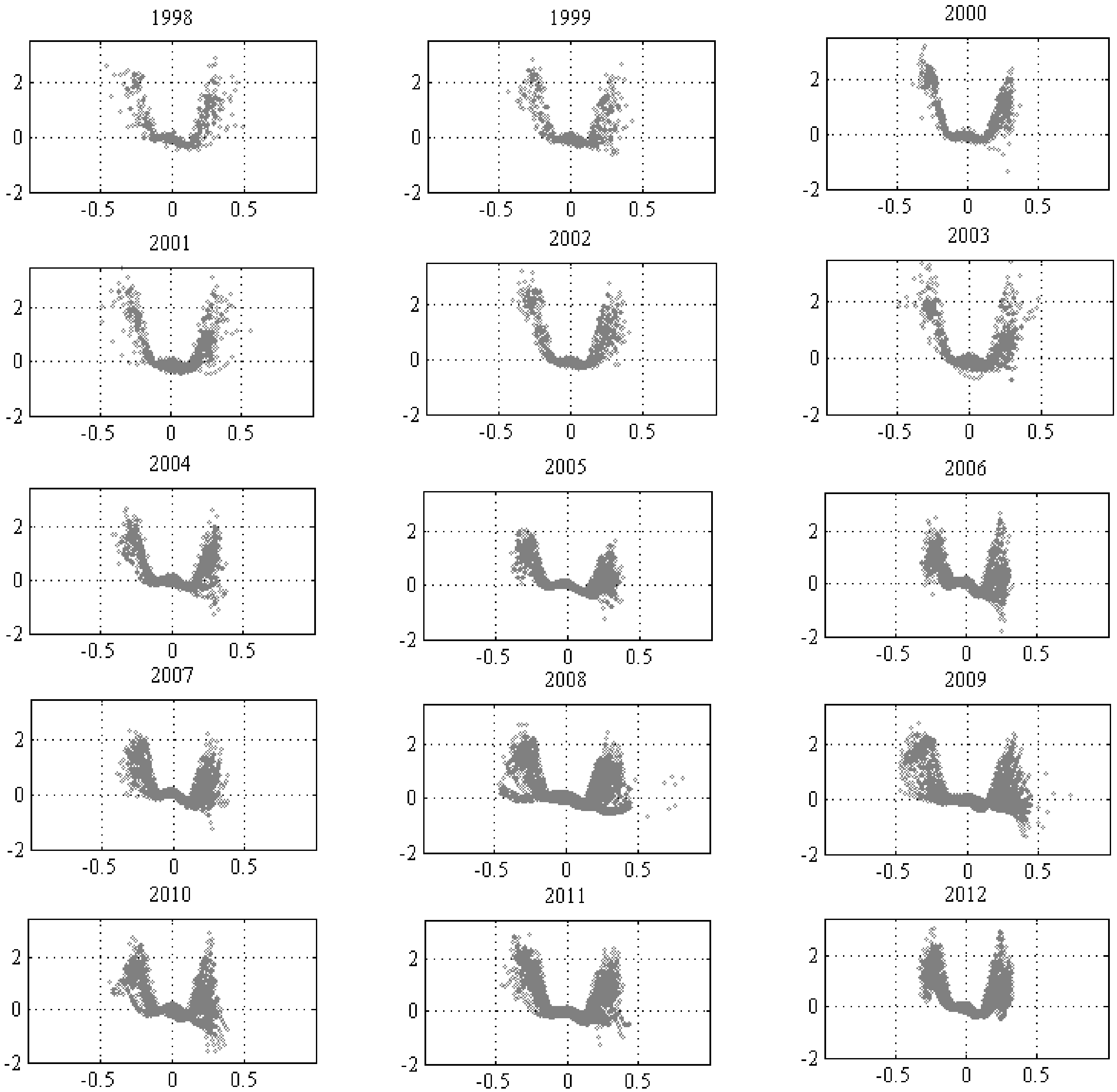
[64] Trolle, A. and E. Schwartz, 2009. Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives. *Review of Financial Studies* 22, 4423-4461.

[65] Working, H. 1960. Speculation on Hedging Markets. Stanford University Food Research Institute Studies 1, 185-220.

[66] Xing, Y., X. Zhang and R. Zhao, 2010. What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? *Journal of Financial and Quantitative Analysis* 45, 641-662.

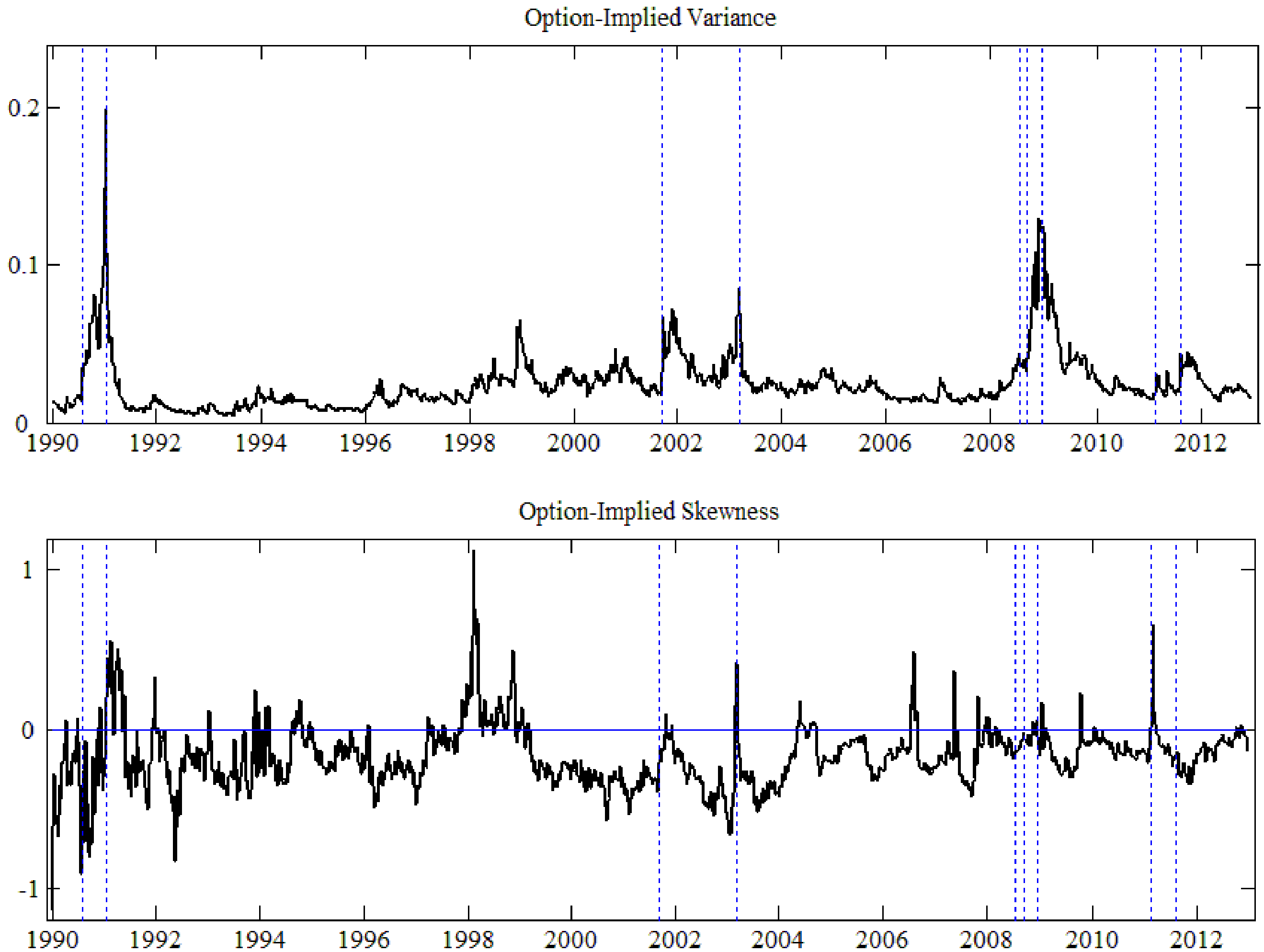
[67] Xiong, W. and H. Yan, 2010. Heterogeneous Expectations and Bond Markets. *Review of Financial Studies* 23, 1433-1466.

Figure 1: State Price Densities in the Crude Oil Market



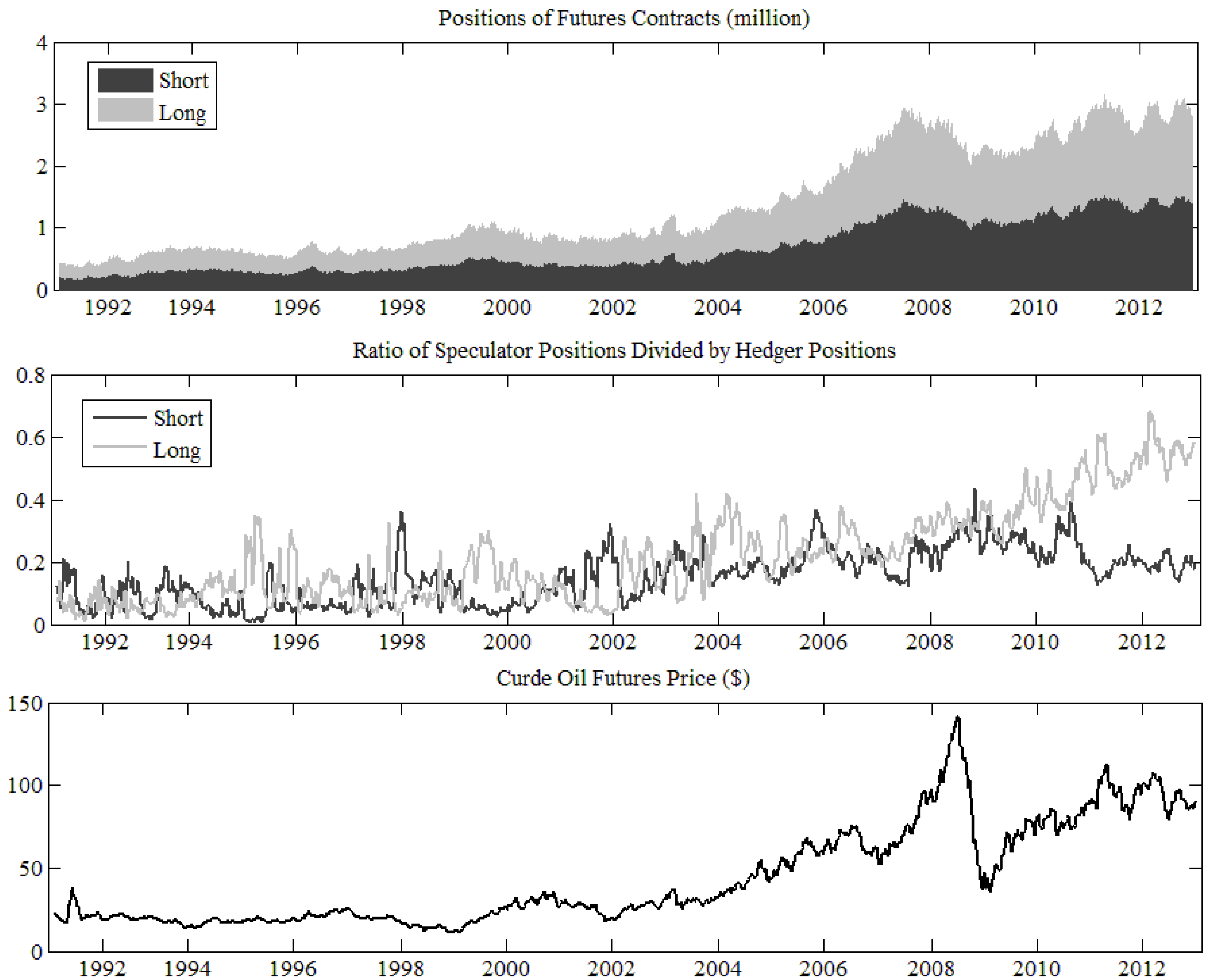
Notes: This figure shows the estimates of the state price densities (SPDs) in the crude oil market in one-month horizon. For each year from 1998 to 2012 we plot the log ratio of the risk-neutral densities and physical densities of futures returns on each Wednesday. The horizontal axis is futures returns defined as $\log(F_T/F_{t,T})$.

Figure 2: Option-Implied Variance and Skewness



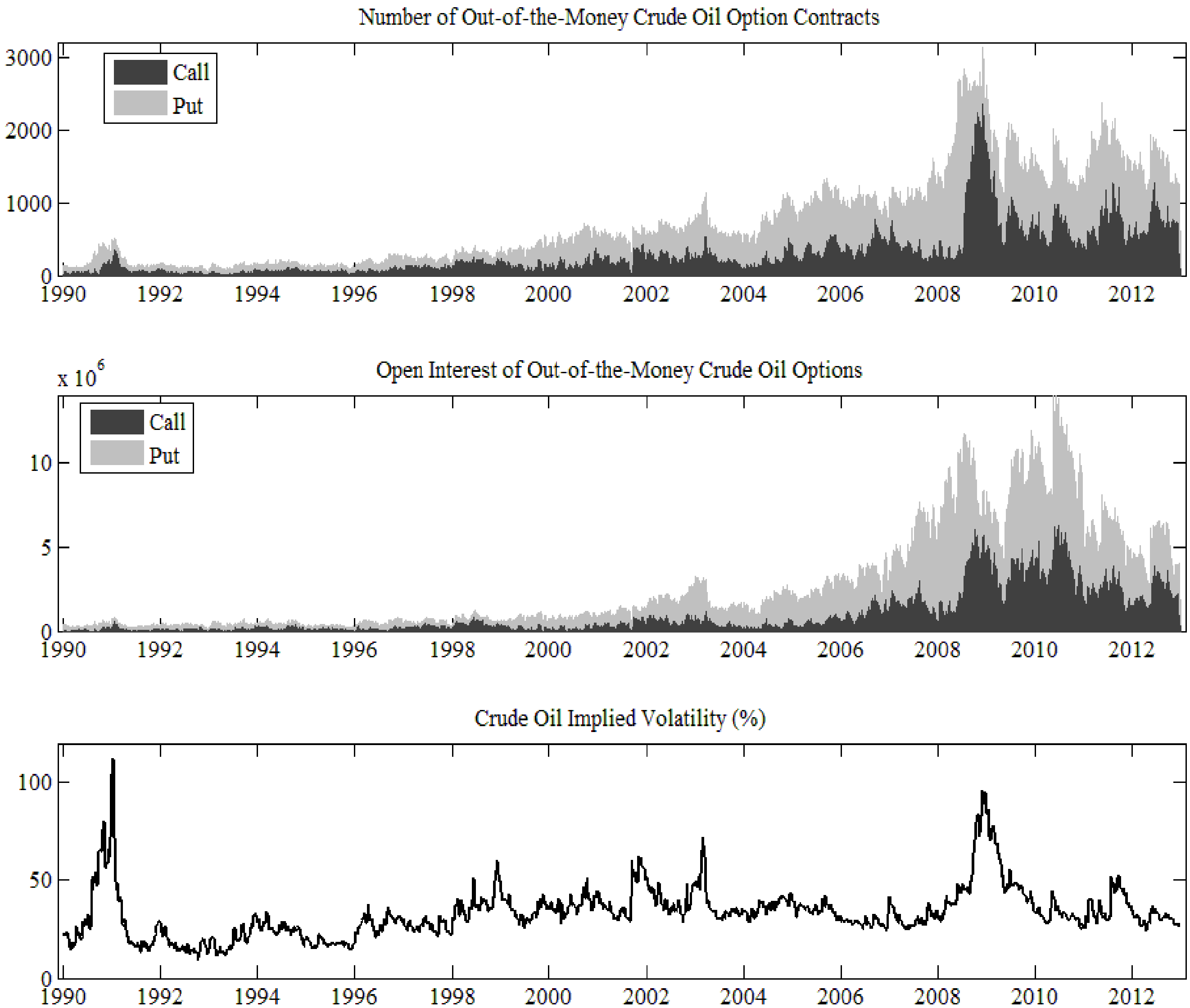
Notes: This figure shows weekly option-implied variance and skewness of one-month maturity crude oil futures returns from January 1990 to December 2012. We first obtain risk-neutral densities on each Wednesday using the semi-parametric method (Ait-Sahalia and Lo, 1998, 2000). Variance and skewness are then calculated based on estimated densities. From left to right, the first two vertical dotted lines denote two significant weeks of the first Gulf War: the Iraq invasion of Kuwait on August 2, 1990, and the liberation of Kuwait on January 17, 1991. Other vertical dotted lines denote the September 11, 2001 terrorist attacks; the second Gulf War on March 20, 2003; the week when WTI crude oil spot price reached its highest level in history (July 23, 2008 [$145.31]); the week when Lehman Brothers filed for Chapter 11 bankruptcy protection (September 15, 2008); the week when WTI crude oil spot price reached its lowest level during the recent crisis periods (December 23, 2008 [$30.28]); the Libyan Civil War began on February 15, 2011 and oil and gas production in Libya fell by more than 60%; and Standard & Poor's downgraded the U.S. long-term sovereign credit rating from AAA to AA+, respectively.

Figure 3: Crude Oil Futures Positions Taken by Hedgers and Speculators



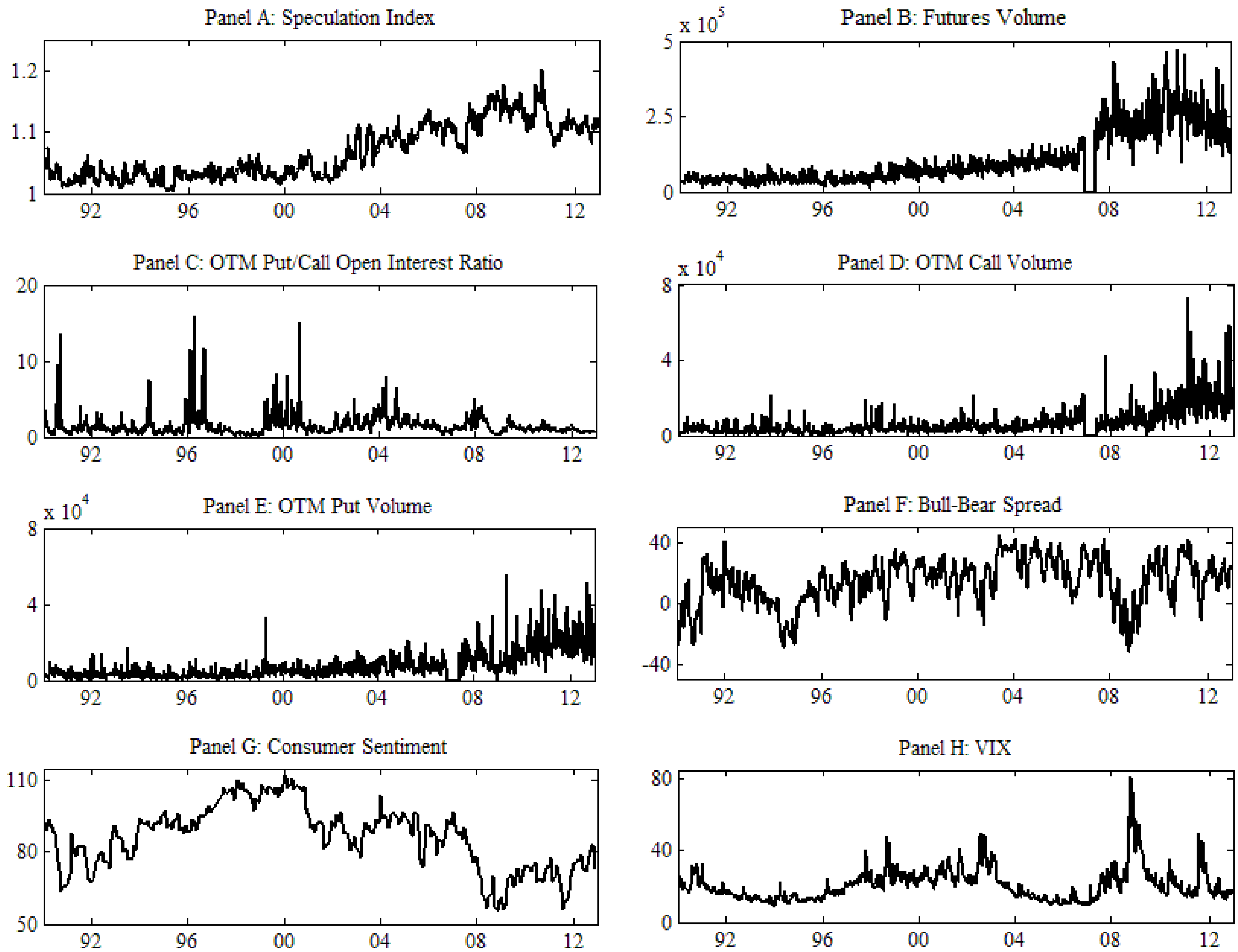
Notes: This figure shows long and short futures positions taken by hedgers and speculators from 1990 to 2012. Positions of hedgers and speculators are from the U.S. Commodity Futures Trading Commission (CFTC) futures-only Commitments of Traders (COT) report. As defined in the report, hedgers (commercials) are those investors who have direct exposure to the underlying crude oil commodity and use futures for hedging purposes; and speculators (non-commercials) are those investors who are not directly engaged in crude oil business activities and use derivatives markets for the purpose of financial profits. Before September 30, 1992, the CFTC publishes reports twice every month; then reports are available every week. The top panel displays the aggregate long (short) futures positions taken by large traders; the mid panel shows the ratio of speculator positions over hedger positions; the bottom panel reports the closest-to-maturity futures price averaged within the CFTC reporting week.

Figure 4: Crude Oil Options Contracts, Open Interests, Trading Volume and Implied Volatility



Notes: This table reports the number of out-of-the-money (OTM) crude oil option contracts and their open interests aggregated within a week. While OTM calls are defined as call options with moneyness (K/F) >1.05, OTM puts are put options with moneyness (K/F)<0.95. We only report options with the maturity less than 180 calendar days. We also report 30-day oil option-implied volatility averaged within each week in the bottom panel. The data period covers January 2, 1990 to December 31, 2012.

Figure 5: Measures of Investor Beliefs



Notes: This figure shows measures of investor beliefs in the crude oil market, the equity market and the economy. Speculation index measures the extent by which speculative positions exceed the necessary level to offset hedging positions. Data is collected from the CFTC trader report. Futures volume is based on the futures contracts maturing in 30 days. OTM put/call open interest ratio is open interest of OTM puts divided by open interest of OTM calls. Both open interest and trading volume of options are measured with a maturity of 30 days. Futures (options) trading volume data are not available from December 15, 2006 to May 21, 2007 (December 1, 2006 to May 21, 2007) when the CME group converted data from the NYMEX database. But the futures and option price and open interest data are available throughout the period. Bull-Bear spread is defined as the percentage of bullish investors minus the percentage of bearish investors based on the survey of Investors Intelligence. Consumer sentiment index is from Surveys of Consumers from the University of Michigan. VIX is the CBOE's Volatility Index.

Table 1: Crude Oil Futures and Options Data

Panel A: Futures Data by Maturity (months)

	1	2	3	6	12	24	All
Num. of Contracts	5759	5778	5780	5781	5774	4414	33286
Average Price ($)	42.673	42.773	42.833	42.827	42.527	44.669	42.984
Average Open Interest	141954	120005	62132	28914	15209	5701	64596
Average Volume	113774	62385	22319	4790	1316	317	35492
Average Volume/OI Ratio	0.80	0.52	0.36	0.17	0.09	0.06	0.55

Panel B: Option Data by Maturity (calendar days)

	(0,30]	(30,60]	(60,90]	(90,120]	(120,180]	(180,∞)	All
Num. of Contracts	57032	69285	70475	65641	100265	398810	761508
Average IV	0.459	0.415	0.395	0.380	0.360	0.303	0.348
Average Price ($)	5.799	5.532	5.690	6.127	6.334	10.768	8.466
Average Open Interest	4207	3579	2919	2585	2197	1672	2299
Average Volume	390	207	105	70	48	20	81
Average Volume/OI Ratio	0.09	0.06	0.04	0.03	0.02	0.01	0.04

Panel C: Option Data by Moneyness (K/F)

	(0,0.85]	(0.85,0.90]	(0.90,0.95]	(0.95,1.05]	(1.05,1.10]	(1.10,1.15]	(1.15,∞)	All
Num. of Contracts	151885	62582	76275	165759	56497	43542	204968	761508
Average IV	0.369	0.329	0.316	0.307	0.317	0.327	0.395	0.348
Average Price ($)	6.300	7.045	6.625	6.316	6.426	6.724	13.861	8.466
Average Open Interest	2685	2409	2356	2227	2234	2261	2041	2299
Average Volume	51	83	97	146	99	84	38	81
Average Volume/OI Ratio	0.02	0.03	0.04	0.07	0.04	0.04	0.02	0.04

Notes: This table reports descriptive statistics of crude oil futures and options data from January 2, 1990 to December 31, 2012. Panel A presents daily futures contracts across maturities. Panel B and C report options data across maturities and moneyness on each Wednesday, where moneyness is defined as the strike price K divided by the underlying futures price F.

Table 2: Average Returns of Crude Oil Futures Options

Panel A: Put Option Returns

	1	2	3	4	5	OTM - ITM		
Moneyness	(0.5, 0.85]	(0.85, 0.95]	(0.95, 1.05]	(1.05, 1.15]	(1.15, 1.5)	(1-5)	(1-4)	(2-5)
1990-2012								
Ret	-0.8808***	-0.3091**	-0.1303	-0.1111**	-0.0565	-0.7406***	-0.6909***	-0.1709
t-stat	(-11.19)	(-2.04)	(-1.48)	(-2.11)	(-1.47)	(-5.02)	(-7.08)	(-0.71)
10^{th} Percentile	-1.0000	-1.0000	-1.0000	-1.0000	-0.5796	-1.3442	-1.8509	-1.1504
50^{th} Percentile	-1.0000	-1.0000	-0.9158	-0.2356	-0.0360	-0.9148	-0.7116	-0.7091
90^{th} Percentile	-1.0000	0.6479	1.8261	1.0348	0.5340	-0.3212	0.0000	0.2909
1990-2004								
Ret	-0.9590***	-0.4476***	-0.1792	-0.1488**	-0.1090**	-0.8462***	-0.6638***	-0.4841***
t-stat	(-42.24)	(-3.08)	(-1.65)	(-2.35)	(-2.23)	(-13.41)	(-8.32)	(-3.71)
2005-2012								
Ret	-0.8120***	-0.0709	-0.0386	-0.0454	0.0173	-0.6625***	-0.7125***	0.2507
t-stat	(-5.54)	(-0.22)	(-0.26)	(-0.49)	(0.28)	(-2.62)	(-4.35)	(0.47)

Panel B: Call Option Returns

	1	2	3	4	5	OTM - ITM		
Moneyness	(0.5, 0.85]	(0.85, 0.95]	(0.95, 1.05]	(1.05, 1.15]	(1.15, 1.5)	(5-1)	(5-2)	(4-1)
1990-2012								
Ret	0.0041	0.1470**	0.2737**	0.0985	-0.7740***	-0.8850***	-0.8904***	-0.4040**
t-stat	(0.07)	(2.38)	(2.41)	(0.37)	(-6.92)	(-12.59)	(-8.36)	(-2.54)
10^{th} Percentile	-0.7589	-1.0000	-1.0000	-1.0000	-1.0000	-1.5367	-2.0506	-1.2320
50^{th} Percentile	0.0001	0.0267	-0.6946	-1.0000	-1.0000	-0.9568	-0.9230	-0.8352
90^{th} Percentile	0.6544	1.2286	2.8071	2.2012	-1.0000	-0.2020	0.0000	-0.0143
1990-2004								
Ret	0.0898	0.2437***	0.4292***	0.4108	-0.6184***	-0.9097***	-0.8376***	-0.2960
t-stat	(1.09)	(2.85)	(2.71)	(1.02)	(-2.86)	(-7.01)	(-4.37)	(-1.13)
2005-2012								
Ret	-0.0781	-0.0046	-0.0177	-0.4447***	-0.9347***	-0.8623***	-0.9473***	-0.5076***
t-stat	(-1.06)	(-0.06)	(-0.13)	(-2.64)	(-22.92)	(-13.33)	(-11.78)	(-2.72)

Notes: This table reports the average returns of options across moneyness. Moneyness is defined as the strike price K divided by the futures price F. On the third Wednesday of each month, we calculate hold-to-maturity returns of available options, with the maturity as close to 30 calendar days as possible. Without interpolation or extrapolation, we assign available options to various bins according to their moneyness, and returns are averaged within each bin. For each moneyness interval, we show the average return, its standard error, the 10th, 50th and 90th percentile from 1990 to 2012, as well as average returns for two subsample periods: 1990 to 2004, 2005 to 2012. We also report the return differences between out-of-the-money (OTM) and in-the-moneyness (ITM) options. Returns of call and put are calculated as $r_{t,T}^{call} = \max(F_{t,T}e^R - K, 0)/c_{t,T} - 1$ and $r_{t,T}^{put} = \max(K - F_{t,T}e^R, 0)/p_{t,T} - 1$, where $F_{t,T}$ is the price of underlying futures, $R = \log(F_T/F_{t,T})$ is the return of the underlying futures contract over the period $T - t$, K is the strike price, $c_{t,T}$ and $p_{t,T}$ are prices of European style call and put options. T-statistics are reported in the parentheses. ***, **, and * represent the significant level of 1%, 5% and 10%, respectively.

Table 3: Summary Statistics of Regression Variables

Variable	Mean	Std Dev	Percentile			P value (ADF test)
			25%	50%	75%	
Risk-Neutral Skewness	-0.1732	0.1932	-0.2888	-0.1824	-0.0782	0.0010
Speculation Index	1.0660	0.0437	1.0286	1.0490	1.1043	0.6714
Speculation Index (Adjusted)	0.0000	0.0138	-0.0090	-0.0010	0.0085	0.0010
Futures Volume	105960	88317	42701	71235	145395	0.0010
Futures Volume (Adjusted)	0	31696	-12620	-1279	12191	0.0010
OTM Put/Call	1.5149	1.3863	0.7737	1.1617	1.7859	0.0010
OTM Call Volume	6530	7047	2427	4428	7943	0.0010
OTM Call Volume (Adjusted)	0	4687	-1989	-599	1121	0.0010
OTM Put Volume	7506	7476	2642	4997	9561	0.0010
OTM Put Volume (Adjusted)	0	4611	-2145	-733	1401	0.0010
Bull-Bear Spread	14.2516	15.2359	4.7000	16.2000	25.3000	0.0010
Consumer Sentiment	86.3806	13.2932	76.2000	88.4500	94.7000	0.4058
Consumer Sentiment (Adjusted)	-0.0688	4.2139	-2.5000	-0.2000	2.4000	0.0010
VIX (%)	20.9520	8.9127	14.4000	19.2650	24.8650	0.0120
Basis	0.0162	0.1998	-0.0845	0.0199	0.0851	0.0010
Storage Level	325329	25734	305639	327121	343221	0.6515
Storage Level (Adjusted)	0	10667	-7495	-631	7880	0.0010
IVOil (%)	33.3345	13.1565	26.2556	31.7878	37.9442	0.0456
Historical Return (%)	0.0296	1.0030	-0.5686	0.0791	0.6241	0.0010

Notes: This table reports the summary statistics of variables used in the regression analysis (14). Variables are measured as closely as possible to each Wednesday from January 1990 to December 2012. Skewness of one-month futures returns is calculated from the risk-neutral densities implied by option prices, as defined by equation (11). Speculation Index measures the extent by which speculative positions exceed the necessary level to offset hedging positions, and a Hodrick-Prescott filter is applied to remove the trend. OTM Put/Call OI is the ratio of total open interests of OTM puts over OTM calls with a maturity of 30 days. Futures Volume, OTM Call Volume and OTM Put Volume are the trading volume of crude oil futures and OTM options with a maturity of 30 days, and they are adjusted for a deterministic time trend. Bull-bear Spread is the proportion of bullish investors minus bearish investors based on the survey conducted by Investors Intelligence. Consumer Sentiment is the Index of Consumer Sentiment from the University Michigan; and we take the first difference to remove the autocorrelation. VIX is the CBOE's Volatility Index. Basis is the discounted spread between futures and spot prices. Storage Level is U.S. total stocks of crude oil, excluding SPR, based on the report from Energy Information Administration (EIA), and a Hodrick-Prescott filter is applied to remove the trend. IVOil is the 30-day oil volatility of futures returns implied by option prices. Historical Return is the moving average of daily return of holding a long position of one-month futures contract during the previous week.

Table 4: Correlation Matrix of Regression Variables

	Speculation Index	Futures Volume	OTM Put/Call	OTM Call Volume	OTM Put Volume	Bull-Bear Spread	Consumer Sentiment	VIX	Basis	Storage Level	IVOil
Futures Volume	0.0448										
OTM Put/Call OI	0.0705	0.0677									
OTM Call Volume	-0.0416	0.3350	-0.1171								
OTM Put Volume	-0.0063	0.4504	0.1181	0.5279							
Bull-Bear Spread	-0.1196	0.0267	0.0791	-0.0176	0.0045						
Consumer Sentiment	-0.0893	-0.0571	-0.1153	-0.0034	-0.0365	0.1569					
VIX	0.1072	-0.0134	-0.0470	0.0400	-0.0233	-0.2685	-0.1258				
Basis	-0.0533	0.0021	-0.4104	0.0490	-0.0535	0.0053	0.0721	0.0465			
Storage Level	-0.0991	0.0593	-0.0119	-0.0225	0.0786	0.0137	-0.0414	-0.0633	0.2882		
IVOil	0.0389	0.0094	-0.0028	0.1055	0.0229	-0.0854	-0.0272	0.6533	0.1276	-0.0904	
Historical Return	-0.0213	-0.0430	0.3547	-0.1238	-0.0083	0.0817	0.0157	-0.0574	-0.1349	0.0646	-0.0281

Notes: This table reports the correlation matrix of variables used in the regression (14). Variables are measured as closely as possible to each Wednesday from January 1990 to December 2012. Speculation Index measures the extent by which speculative positions exceed the necessary level to offset hedging positions, and a Hodrick-Prescott filter is applied to remove the trend. OTM Put/Call OI is the ratio of total open interests of OTM puts over OTM calls with a maturity of 30 days. Futures Volume, OTM Call Volume and OTM Put Volume are the total trading volume of crude oil futures and OTM options with a maturity of 30 days, and they are adjusted for a deterministic time trend. Bull-bear Spread is the proportion of bullish investors minus bearish investors based on the survey conducted by Investors Intelligence. Consumer Sentiment is the Index of Consumer Sentiment from the University Michigan; and we take the first difference to remove the autocorrelation. VIX is the CBOE's Volatility Index. Basis is the discounted spread between futures and spot prices. Storage Level is U.S. total stocks of crude oil, excluding SPR, based on the report from Energy Information Administration (EIA), and a Hodrick-Prescott filter is applied to remove the trend. IVOil is the 30-day oil volatility of futures returns implied by option prices. Historical Return is the moving average of daily return of holding a long position of one-month futures contract during the last week.

.

Table 5: Investor Beliefs and the Slope of the SPDs in the Crude Oil Market

	Decreasing-Region Slope (a)		**Decreasing-Region Slope (b)**		**Increasing-Region Slope (a)**		**Increasing-Region Slope (b)**	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Speculation Index	0.0518	0.0767**	0.0280	0.0503	-0.0236	-0.0120	-0.0225	-0.00925
	(1.32)	(2.03)	(0.72)	(1.34)	(-0.57)	(-0.29)	(-0.53)	(-0.22)
Future Volume	-0.0249	-0.0295	-0.0527*	-0.0548*	-0.0771***	-0.0785***	-0.0518*	-0.0531**
	(-0.76)	(-0.97)	(-1.69)	(-1.96)	(-2.88)	(-3.05)	(-1.90)	(-1.97)
OTM Put/Call OI	0.0908**	0.0844*	0.0827*	0.0762	-0.0885*	-0.0903*	-0.0822*	-0.0791
	(1.98)	(1.90)	(1.73)	(1.62)	(-1.76)	(-1.79)	(-1.66)	(-1.61)
OTM Call Volume	-0.0174	-0.00172	-0.0239	-0.00780	0.0349	0.0417	0.0685**	0.0724**
	(-0.45)	(-0.05)	(-0.63)	(-0.23)	(1.05)	(1.34)	(2.07)	(2.30)
OTM Put Volume	-0.0007	-0.0094	0.0036	-0.0072	-0.0101	-0.0139	-0.0145	-0.0150
	(-0.02)	(-0.26)	(0.09)	(-0.20)	(-0.29)	(-0.41)	(-0.45)	(-0.47)
Bull-Bear Spread		0.0318		-0.0387		0.0152		0.0875*
		(0.80)		(-0.94)		(0.32)		(1.95)
Consumer Sentiment		0.0182		0.0318		0.0167		0.0335
		(0.43)		(0.76)		(0.40)		(0.81)
VIX		-0.1690***		-0.2120***		-0.0717		0.00587
		(-3.17)		(-4.16)		(-1.30)		(0.10)
Adjusted R^2	0.008	0.039	0.007	0.049	0.012	0.016	0.010	0.017

Notes: This table presents the results of the weekly regression that examines the dependence of the slope of the SPD on investor beliefs. The dependent variables in the left panel are the decreasing region (or negative return region) slopes measured using return intervals (-0.20, -0.14) and (-0.19, -0.13); the dependent variables in the right panel are the increasing region (or positive return region) slope measures using return intervals (0.15, 0.18) and (0.17, 0.20). We do not report the intercept term in regressions to save the space. Newey-west t-stats are reported in parentheses. ***, **, and * represent the significant level of 1%, 5% and 10%, respectively. The data period covers January 2, 1990 to December 31, 2012.

Table 6: Investor Beliefs and Risk-Neutral Skewness

Panel A: Skewness						
	(1)	(2)	(3)	(4)	(5)	(6)
Speculation Index	0.0206					0.0340
	(0.93)					(1.49)
Speculation Index * Dummy	-0.0323					-0.0515*
	(-1.17)					(-1.86)
Future Volume		0.0166				0.0182
		(1.26)				(1.23)
OTM Put/Call OI			-0.111***			-0.0995***
			(-4.91)			(-4.57)
OTM Call Volume				0.0673***		0.0774***
				(4.03)		(4.00)
OTM Put Volume					-0.0124	-0.0502***
					(-0.99)	(-3.15)
Adjusted R^2	0.731	0.731	0.741	0.735	0.731	0.746

Panel B: AR(1) Residual						
	(1)	(2)	(3)	(4)	(5)	(6)
Speculation Index	-0.0091					0.0010
	(-0.29)					(0.03)
Speculation Index * Dummy	-0.0544					-0.0638
	(-0.80)					(-0.98)
Future Volume		0.0190				0.0235
		(1.35)				(1.49)
OTM Put/Call OI			-0.199***			-0.182***
			(-6.32)			(-5.97)
OTM Call Volume				0.0689***		0.0678***
				(4.26)		(3.86)
OTM Put Volume					-0.0185	-0.0496***
					(-1.33)	(-3.13)
Adjusted R^2	0.017	0.017	0.091	0.033	0.017	0.104

Notes: This table presents the results of the weekly regression that examines the dependence of risk-neutral skewness on investor beliefs embedded in the crude oil futures and options markets. The dependent variable in Panel A is skewness based on the risk-neutral densities estimated from option prices, as defined in equation (11) and we include the lagged dependent variable in regressions. The dependent variable and regressors in Panel B are the AR(1) residual in the regression of $X_t = a + b*X_{t-1} + \varepsilon_t$. The dummy variable equals to one for the period of 2005 to 2012 and equals to zero during the period of 1990 to 2004. We do not report the intercept term in regressions to save the space. Newey-west t-stats with 12 lags are reported in parentheses. ***, **, and * represent the significant level of 1%, 5% and 10%, respectively. The data period covers January 2, 1990 to December 31, 2012.

Table 7: Investor Beliefs and Risk-Neutral Skewness

(Robustness to Control Variables)

	(1)	(2)	(3)	(4)
Speculation Index	0.0335	0.0362*	0.0265	0.0292
	(1.53)	(1.72)	(1.15)	(1.30)
Speculation Index* Dummy	-0.0514*	-0.0631**	-0.0469*	-0.0572*
	(-1.94)	(-2.30)	(-1.65)	(-1.94)
Future Volume	0.0181	0.0197	0.0140	0.0151
	(1.23)	(1.37)	(0.88)	(0.96)
OTM Put/Call OI	-0.0997***	-0.111***	-0.0561***	-0.0678***
	(-4.49)	(-4.97)	(-2.67)	(-3.26)
OTM Call Volume	0.0782***	0.0765***	0.0634***	0.0624***
	(4.03)	(3.85)	(3.42)	(3.27)
OTM Put Volume	-0.0512***	-0.0491***	-0.0475***	-0.0472***
	(-3.19)	(-3.10)	(-3.10)	(-3.07)
Bull-Bear Spread	-0.0135	-0.0120	-0.0138	-0.0130
	(-0.93)	(-0.77)	(-0.90)	(-0.80)
Consumer Sentiment	-0.0068	-0.0087	-0.0057	-0.0058
	(-0.55)	(-0.65)	(-0.43)	(-0.42)
VIX	-0.0157	-0.0293	-0.0209	-0.0227
	(-1.08)	(-1.36)	(-1.43)	(-1.08)
Adjusted R^2	0.745	0.747	0.759	0.759

Notes: This table reports the results of the weekly regression that examines the robustness of the relation between investor beliefs and risk-neutral skewness. The dependent variable is skewness based on the risk-neutral densities estimated from option prices, as defined in equation (11). All regressions include the lagged dependent variable. The dummy variable equals to one for the period of 2005 to 2012 and equals to zero during the period of 1990 to 2004. Column (2) includes control variables in the crude oil market, which are the basis, storage level, historical returns, and oil option-implied volatility (IVOil). Column (3) includes macroeconomic control variables: the 1-month treasury-bill rate, the spread between yields on the 10-year treasury bond and the 3-month treasury bill, the difference in yields on Baa and Aaa corporate bonds, the Chicago Fed National Activity Index and the log growth in aggregate industrial production over the last 12 months. Column (4) includes both sets of control variables in columns (2) and (3). We do not report the intercept term in regressions to save the space. Newey-west t-stats with 12 lags are reported in parentheses. ***, **, and * represent the significant level of 1%, 5% and 10%, respectively. The data period covers January 2, 1990 to December 31, 2012.

Table 8: Investor Beliefs and Risk-Neutral Skewness

(Robustness to the Alternative Measure of Skewness)

	(1)	(2)	(3)	(4)
Speculation Index	0.0237	0.0266	0.0235	0.0273
	(0.96)	(1.03)	(0.92)	(1.05)
Speculation Index* Dummy	-0.0670**	-0.0822***	-0.0798**	-0.100***
	(-2.30)	(-2.59)	(-2.55)	(-2.86)
Future Volume	0.00818	0.0108	0.00975	0.0125
	(0.52)	(0.69)	(0.60)	(0.76)
OTM Put/Call OI	-0.0376*	-0.0473**	-0.0524**	-0.0712***
	(-1.87)	(-2.53)	(-2.19)	(-3.02)
OTM Call Volume	0.0699***	0.0681***	0.0524**	0.0485**
	(3.45)	(3.22)	(2.34)	(2.05)
OTM Put Volume	-0.0748***	-0.0717***	-0.0725***	-0.0727***
	(-3.79)	(-3.68)	(-3.69)	(-3.69)
Bull Spread	0.0322	0.0318	0.0288	0.0267
	(1.38)	(1.35)	(1.19)	(1.11)
Consumer Sentiment	-0.0174	-0.0223	-0.0232	-0.0235
	(-1.00)	(-1.17)	(-1.35)	(-1.29)
VIX	0.0323	0.0207	0.0359*	0.0542*
	(1.58)	(0.87)	(1.70)	(1.86)
Adjusted R^2	0.598	0.602	0.610	0.616

Notes: This table reports the results of the weekly regression that examines the robustness of the dependence of risk-neutral skewness on investor beliefs using another measure of risk-neutral skewness. The dependent variable, risk-neutral skewness, is obtained by using the methodology of Bakshi, Kapadia, and Madan (2003). All regressions include the lagged dependent variable. Column (2) includes control variables in the crude oil market, which are the basis, storage level, historical returns, and oil option-implied volatility (IVOil). Column (3) includes macroeconomic control variables: the 1-month treasury-bill rate, the spread between yields on the 10-year treasury bond and the 3-month treasury bill, the difference in yields on Baa and Aaa corporate bonds, the Chicago Fed National Activity Index and the log growth in aggregate industrial production over the last 12 months. Column (4) includes both sets of control variables in columns (2) and (3). We do not report the intercept term in regressions to save the space. Newey-west t-stats with 12 lags are reported in parentheses. ***, **, and * represent the significant level of 1%, 5% and 10%, respectively. The data period covers January 2, 1990 to December 31, 2012.