

The Missing Tail Risk in Option Prices

Jason P. Brown, Nida Çakır Melek,
Johannes Matschke, and Sai A. Sattiraju

March 2023

RWP 23-02

<http://doi.org/10.18651/RWP2023-02>

FEDERAL RESERVE BANK *of* KANSAS CITY



The Missing Tail Risk in Option Prices*

Jason P. Brown[†]

Federal Reserve Bank of Kansas City

Nida Çakır Melek[‡]

Federal Reserve Bank of Kansas City

Johannes Matschke[§]

Federal Reserve Bank of Kansas City

Sai A. Sattiraju[¶]

Federal Reserve Bank of Kansas City

March 29, 2023

Abstract

This paper contributes to the literature on deviations from rational expectations in financial markets and to the literature on evaluating density forecasts. We first develop a novel statistic to evaluate the overall accuracy of distributional forecasts, and find two methods that yield accurate distributional forecasts. We then propose another statistic to examine the relative accuracy over the entire distribution range. Our results indicate more oil price realizations in the left tail than predicted. We argue that this finding points to a persistent behavioral forecasting bias and a departure from the rational expectations hypothesis. Investors hence underestimate left tail risk and under-insure against very low oil prices.

Keywords: Option Pricing; Density Forecasts; Behavioral Finance; Under-insurance

JEL: C52, C58, G12, G17, G41, Q47

*We thank Bahattin Buyuksahin, Lutz Kilian, Emanuel Moench, Alessandro Rebucci, Michel Robe and seminar participants at the Federal Reserve Bank of Kansas City for valuable feedback. The views expressed in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

[†]E-mail: Jason.Brown@kc.frb.org

[‡]E-mail: Nida.CakirMelek@kc.frb.org

[§]E-mail: Johannes.Matschke@kc.frb.org

[¶]E-mail: Sai.Sattiraju@kc.frb.org

1. INTRODUCTION

Financial markets exhibit patterns that are inconsistent with the rational expectations theory, which assumes that investors have complete knowledge of the fundamental structure of the economy and make rational decisions based on all available information. The implication of these assumptions is that returns should not be predictable. Yet, a large empirical literature provides evidence that asset returns are predictable and that trading activity is excessive beyond what is justified by the arrival of new information (Daniel et al., 1998; Barber and Odean, 2001; Daniel and Hirshleifer, 2015). A related literature considers behavioral theories to explain such deviations from rational expectations. Common theories center around overconfidence (De Bondt and Thaler, 1995), learning (DeCanio, 1979), ambiguity (Ju and Miao, 2012), or more recently, the extrapolation of news (Bordalo et al., 2022). The empirical support for these theories is largely based on asset prices, such as stock prices, which are indicative of investor beliefs about the first moment. In this paper, we use option prices, which allow us to extract the entire belief distribution, and develop two statistics to gain new insights on possible deviations from rational expectations. Our analysis suggests that investors in the oil market primarily under-predict left tail events, which is at odds with the rational expectations-based theories of price formation.

This paper makes two contributions. The first is methodological. We develop two statistics that allow us to assess 1) the overall accuracy of a distributional forecast, and 2) the relative accuracy over the entire distribution range (i.e. its percentiles). Both statistics are simple to use and they can be applied in *any* context. Using the overall accuracy measure, we show that two methods that are used to extract beliefs based on option prices, namely Binomial Trees (Cox et al., 1979) and a non-parametric approach a la Swanson (2006), yield quite accurate distributional forecasts. On the other hand, the popular approach by Shimko (1993), applying Black and Scholes (1973) and Breeden and Litzenberger (1978), leads to distributions that have less support in (crude oil) options data. The second contribution builds on the relative accuracy measure. We find that investors consistently under-predict and thus under-insure against left tail events in the crude oil market, such as the decline in oil prices during the Global Financial Crisis, during 2014-2015, or during the onset of the COVID-19 pandemic. We argue that these anomalies are behavioral and hence a deviation from rational expectations. This may have important implications: the investment portfolio could become excessively volatile, or businesses in the oil industry could expose themselves to excess risk during downturns.

We estimate daily density forecasts, the so called 'risk-neutral density' (RND),

based on 21 years of daily transaction data for WTI crude American-style options.¹ We choose the crude oil option market, as it is one of the most liquid and largest markets with around 1 million transactions on average per month for options that expire within 30 days alone. Oil is also very different from many assets in the sense that it is an input to production and consumption goods, whose price is often determined by physical supply and demand factors. For example, oil options are shown to contain information about economic conditions and various security returns that is distinct from the information contained in the stock index (Christoffersen et al., 2022). That is, the oil market is clearly important for the overall economy and may also contain information on investor beliefs.

Any attempt to isolate investor beliefs as an explanation for missing tail risk requires ruling out possible confounders. In our context, factors that can lead to deviations between the forecasted and the actual probability distributions include financial market frictions, statistical methods used to estimate density forecasts, inaccurate beliefs, and preferences (risk attitudes). As such, ex-ante, it is not obvious to interpret the forecasting errors as reflecting misguided beliefs. Given that the oil option market is very liquid with almost zero bid-ask spreads, one can eliminate financial market frictions as a potential explanation. Regarding the statistical procedures, we resort to three aforementioned methods: Binomial Trees, Swanson (2006)'s non-parametric approach, and a semi-parametric approach a la Shimko (1993). We analyze the accuracy of each, which itself is a contribution to the literature. Each method involves a set of orthogonal assumptions that pose more (or less) structure on the density forecast. We show, based on our novel statistics, that the Binomial Tree and the non-parametric Swanson (2006) approach yield distributions that are almost equivalently accurate, both in terms of the overall distribution and over specific percentiles.

More specifically, our first statistic measuring 'Total Accuracy' ($TAcc$) indicates that both the Binomial Tree and the non-parametric Swanson (2006) approach are quite accurate ('correctly calibrated'). Intuitively, this statistic compares the realization of the underlying at option maturity relative to the forecasted distribution over time. 'Total Accuracy' is then determined by the deviation between realized percentiles and an appropriate benchmark, such that a value of 1 reflects a correctly calibrated forecast. Our $TAcc$ measure ranges between 0.8 - 0.9 across various forecasting horizons for

¹RNDs are used in many different contexts in the literature: for prediction, for examining major economic events, or explaining risk premium and crash risk. See, for example, Melick and Thomas (1997), Bliss and Panigirtzoglou (2004), Jiang and Tian (2005), Christoffersen and Mazzotta (2005), Carr and Wu (2007), Kitsul and Wright (2013), Datta et al. (2016), among others.

distributions based on the Binomial Tree and the Swanson method. It, however, drops to 0.6 - 0.7 for distributions based on the popular Shimko procedure. We, therefore, argue that option implied distributions should not necessarily be interpreted at face value. The statistical method may confound the distributional forecast and a proper evaluation on a case-by-case basis appears warranted.

With that, we focus on the Binomial Tree and the Swanson method and further examine distributional forecasts. Specifically, we assess how accurately different parts of the distribution are forecasted by introducing our second novel statistic, 'Relative Accuracy' ($RAcc$). This measure calculates the relative sampling frequency for each percentile in the distribution, with a value above 1 indicating oversampling and vice versa. Even though investors forecast the overall price distribution quite well, not all parts of the distribution are as accurately forecasted. Our 'Relative Accuracy' measure reveals a striking anomaly concerning left tail events that is robust across forecasting horizons. In particular, we observe significantly more realizations of oil prices that are within the first percentile of the forecasted distribution than predicted.

What can account for the anomaly in the left tail? Given that we obtain very similar and overall accurate densities based on the two orthogonal approaches, we interpret related densities as reflecting beliefs and preferences. And, to gain further insights, we need to distinguish beliefs from preferences. This is because the missing left tail could be a consequence of misspecified beliefs, or driven by a particular set of preferences. For example, risk loving investors, or investors that do not want to insure against tail events because it does not affect their businesses, are less likely to purchase insurance, which would drive down demand for options that insure against extreme events. In that case, estimated densities would appear as if they under-predict tail events, which could drive the observed anomaly in the left tail.

We argue that the anomaly in the left tail is driven by beliefs rather than preferences. There are several arguments that support our behavioral interpretation. First, we only observe an anomaly in the left, but not in the right tail. Commercial investors such as oil producers, however, suffer from lower profit margins when oil prices are very low. These businesses, hence, have an incentive to insure against very low oil prices and economic downturns, which should drive up demand for options. Everything else equal, this implies more, rather than less, probability mass in the left tail. In addition, the oil market also consists of sizeable non-commercial investors, such as money managers ([Rouwenhorst and Tang, 2012](#); [Kang et al., 2020](#)). These institutional investors are generally perceived to be risk averse (or at least not risk loving), implying that they would want to insure against downturns and hence very low oil prices

(Mehra and Prescott, 1985; Basak and Pavlova, 2013; Kojien, 2014). Second, based on returns of portfolios that insure against risk in oil markets, Dew-Becker et al. (2021) find that investors are willing to pay a premium (accept a negative return) to insure themselves against realized volatility. Building on this finding, we show that the realized left tail events in our sample indeed coincide with high realized volatility, suggesting that investors would want to specifically insure against these events. Yet, we observe the opposite.

We are not the first to find mispricing in option markets, but to our knowledge, we are the first documenting deviations from rational expectations in the oil option market. Constantinides et al. (2009), based on a set of equilibrium conditions, argue that S&P 500 option prices do not satisfy no-arbitrage opportunities. Furthermore, investors can increase expected utility by engaging in a zero-net-cost trade, where an investor uses an options strategy to lock in a gain by buying an out-of-the-money (OTM) put and selling a equally-priced OTM call, such as a strangle. Baele et al. (2019) show that an asset pricing model augmented by prospect theory can match observed S&P 500 index put and call returns well. And, they generate a variance premium similar to what they observe in the data.² Relative to these papers, we directly analyze the forecasted distribution with regard to possible deviations from rational expectations. Our findings indicate that investors consistently underestimate the likelihood of very low prices and hence economic downturns.

The rest of the paper is organized as follows. Section 2 presents the data we use in the analysis. In Section 3, we discuss our approach to estimate investor beliefs from option prices. Section 4 introduces two novel measures to assess the accuracy of the estimated distributions. In Section 5, we present and discuss our results. Section 6 concludes.

2. DATA

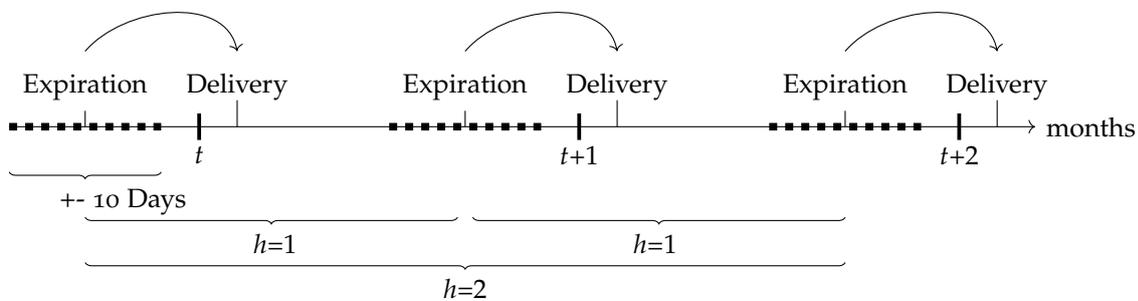
Our data is accessed through the Chicago Mercantile Exchange (CME) and presents end of day quotes and daily trading volumes for American Oil options (both puts and calls) on underlying West Texas Intermediate (WTI) futures from January 2001 until June 2022. The CME facilitates transactions of both American style and European style options. The difference is that American, unlike European, options can be exercised at any date prior to maturity. The bulk of option related transactions involve American

²Prospect theory, developed by Daniel Kahneman and Amos Tversky, is a theory of behavioral economics, centering around the empirical observation that agents asymmetrically feel losses greater than that of an equivalent gain.

style options, which we therefore focus on in our analysis.³ Each American option allows the investor to purchase (call option) or sell (put option) an underlying oil future contract at a given strike price any date prior to expiration. These oil options expire between the 15th-20th of each month, usually 6 business days prior to the expiration of the underlying future contract due around the 25th. The oil future contract is set for delivery in the following month. Given that we forecast the distribution of the underlying at expiration of the option, with the underlying future set for delivery soon after, we essentially forecast the oil spot price.

We forecast the oil price distribution over the next $h = \{1, 2, \dots, 12\}$ months. Because of fixed expiration dates, the contract maturity declines over the course of a month. For example, a contract due in June 2022 has a maturity of roughly 30 days (or 1 month) by mid May 2022, which subsequently declines to zero by mid June. To better utilize the available data and also avoid overlapping windows, we introduce the following timing convention: For each month to maturity, we focus on transactions around a plus/minus 10 calendar day window, indicated by the dotted intervals in Figure 1. For example, an option that expires in one month includes contracts within a 20-40 calendar day expiration window. It is worth mentioning that only about 2/3 of all calendar days are actual trading days. We thus effectively analyze data from roughly 13-14 dates per month.

Figure 1: Timing: Oil Option Market

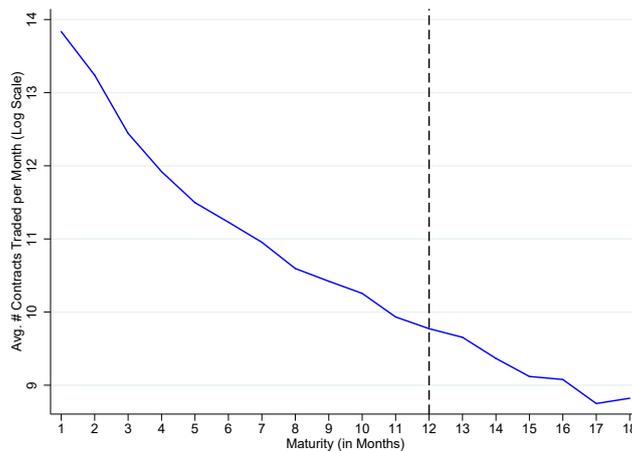


The oil option market is very liquid as shown by Figure 2. The chart presents the average monthly transactions (in log scale) for American options that are set to expire between months $h - 1$ and h . For example, on average, we observe $\exp(13.84) \approx 1$ million transactions per month for options that expire within 30 days. Because each contract is in units of 1,000 barrels, transactions for options expiring in less than a

³For recent trading volumes of American style oil options, see <https://www.cmegroup.com/markets/energy/crude-oil/light-sweet-crude.volume.options.html#optionProductId=190>. For European style options, see <https://www.cmegroup.com/markets/energy/crude-oil/light-sweet-crude.volume.options.html#optionProductId=1353>.

month are “worth” 4,000 million barrels of oil. The transaction volume decreases with maturity, and, consequently, we choose to focus on options that expire within the first $h=12$ months.⁴ This allows us to work with very liquid options, which ensures that observed option prices reflect investor beliefs and preferences about developments in the oil market rather than financial market frictions, which can manifest in bid-ask spreads. For this reason, we also discard option prices for strikes with less than 10 trades per day.⁵

Figure 2: Oil Option Transaction Volume



Notes: The blue line presents the average number of transactions per month in log scale as a function of the contract maturity in months (x -axis). Each contract is in units of 1,000 barrels. Monthly transactions are aggregated over American call and put options as well as all available strikes.

3. FROM OPTION PRICES TO DENSITY FORECASTS

This section lays out our approach to estimate density forecasts from option prices. Options provide investors insurance against certain market outcomes. If a call (put) option is exercised, investors purchase (sell) the underlying asset at a specific strike price, which provides a positive payoff when the market value of the underlying asset is above (below) the strike. Otherwise the option is not exercised. If exercised, the payoff of a call option is linearly increasing in the market value of the underlying, which provides insurance against high market prices. Similarly, put options provide insurance against low market prices.

In Section 3.1, we review a few general principles about American style options. Section 3.2 lays out how we extract investor beliefs and preferences from observed

⁴Options that expire in 1 year still record about 17,600 transactions per month.

⁵Our results are robust to various alternative minimum trade requirements.

option prices. Section 3.3 provides an initial evaluation of the extracted densities with visuals.

3.1. Option Pricing

We analyze American options that expire in $h \leq 12$ months (including the ± 10 day window). The market value of the underlying asset, in our case the oil future, at time τ is denoted by x_τ . Future realizations of x_τ are uncertain and investors form beliefs represented by density $g_t(x_\tau)$. The distribution is time varying and depends on the duration between t and τ . Uncertainty accumulates over time and may be more pronounced during periods of financial or macroeconomic turmoil. The valuation of an option also depends on the marginal utility of the investor, $u'(x_\tau)$, which can be a function of the underlying. If $u'(x_\tau)$ is large, investors value additional funds and hence are willing to pay a higher price for an option that makes a payment in that state. This leads us to the definition of a stochastic discount factor (SDF), which is increasing in $u'(x_\tau)$. Investors further discount the future by the discount rate $\beta < 1$, which is inversely related to the steady state monthly real risk-free rate r . The value of an American call option with maturity h and strike k can then be defined as

$$v_t(h, k) = \max_{\tau \in [t; t+h]} \left\{ \underbrace{\beta^{\tau-t}}_{\text{Discount: } e^{-r(\tau-t)}} \int_0^\infty \underbrace{\frac{u'(x_\tau)}{u'(x_t)}}_{\text{SDF: } \lambda(x_\tau)} \underbrace{\max(x_\tau - k, 0)}_{\text{Payoff}} g_t(x_\tau) dx_\tau \right\}, \quad (1)$$

such that the value of an American call equals the maximum, discounted, marginal utility weighted expected payoff over the lifetime of the option.⁶ The first max operator arises from American options, which unlike European options, can be exercised at any date until maturity.

Financial Market Frictions

Our goal is to make inferences about beliefs $g_t(x_\tau)$. Yet, we only observe the risk-free rate and the payoff conditional on x_τ . However, one can infer the valuation of options based on observed market prices. In any financial market, the price of an asset equals the valuation plus some error ϵ_t , which represents price distortions from financial market imperfections, such as bid-ask spreads or transaction costs. Therefore,

$$\underbrace{c_t(h, k)}_{\text{Price}} = \underbrace{v_t(h, k)}_{\text{Valuation}} + \underbrace{\epsilon_t}_{\text{Frictions}}. \quad (2)$$

⁶Put options are analogously defined with $\max(k - x_\tau, 0)$ representing the payoff.

As long as markets are liquid, the error term ϵ_t is small, implying

$$c_t(h, k) \approx v_t(h, k).$$

Conversely, if markets are illiquid or distorted by financial frictions, prices may not reflect the value assigned by investors. This explains why we focus on highly liquid options like the American oil options.

Risk-Neutral Probabilities

The SDF tends to be related to the realization of the underlying asset. For example, in the case of oil, generally speaking, investors are either commercial investors, such as airlines or oil producers, who have 'skin in the game' with profits directly depending on oil prices, or non-commercial investors, whose profits are tied to economic conditions, which are associated with oil prices.

Combining the SDF with the stochastic belief distribution, we can define a "risk-neutral" distribution (RND), $\tilde{g}_t(x_\tau) \equiv \lambda(x_\tau)g_t(x_\tau)$, that re-weights the physical states according to their desirability. Note that the risk-neutral distribution bears its name from the assumption that a risk-neutral distribution coincides with beliefs only if the investor is risk-neutral. All of our results regarding the forecasted distribution are related to the risk neutral distribution rather than beliefs per se. We then resort to results established in the literature to analyze if deviations of $\tilde{g}_t(x_\tau)$ from actual realizations are due to preferences, as encapsulated in the SDF ($\lambda(x_\tau)$), or due to misspecified beliefs, as represented by $g_t(x_\tau)$.

Statistical Error

Given above arguments, the price of an American oil option approximately equals the maximum, discounted, expected payoff over the lifetime of the option, where expectations are formed with respect to the risk neutral distribution. For a call option, this implies

$$c_t(h, k) \approx \max_{\tau \in [t, t+h]} \left\{ e^{-r(\tau-t)} \int_0^\infty \max(x_\tau - k, 0) \tilde{g}_t(x_\tau) dx_\tau \right\}. \quad (3)$$

Apart from $\tilde{g}_t(x_\tau)$, every element in this equation is directly observable or can be inferred conditional on x_τ . In order to extract the RND of the underlying at maturity,

$\tilde{g}_t(x_{t+h})$, we resort to the existing option pricing literature.⁷ Overall, option pricing methods differ by whether they are suitable for American or European options and by the parametric assumptions they impose. Given the myriad of approaches and given that the underlying option pricing problem for oil futures is relatively simple, we prefer to use methods that perform well in typical environments and are easy to implement.

Each method introduces its own statistical error, so the choice of the option pricing model is important. Suppose we use method z to extract $\tilde{g}_t(x_{t+h})$. The estimated density function $\widehat{\tilde{g}_t(x_{t+h})}_z$ can then be described by

$$\underbrace{\widehat{\tilde{g}_t(x_{t+h})}_z}_{\text{Calculated Distribution}} = \underbrace{\tilde{g}_t(x_{t+h})}_{\text{RND}} + \underbrace{z_t(x_{t+h})}_{\text{Error Function}},$$

where $z_t(x_{t+h})$ refers to the error function associated with method z , which may vary over time. An ideal approach would be one that guarantees $z_t(x_{t+h}) \approx 0$. However, because the RND is latent, it is not straightforward to distinguish $\tilde{g}_t(x_{t+h})$ from $\widehat{\tilde{g}_t(x_{t+h})}_z$. To tackle this issue, we proceed as follows. We adopt three popular approaches that are quite distinct in terms of their assumptions to extract the RND: a parametric approach based on Binomial Trees (Cox et al., 1979), a semi-parametric procedure based on Shimko (1993) which adopts a variant of the Black-Scholes option pricing model, and a non-parametric method (Swanson, 2006). We then assess their overall accuracy. We find that Binomial Trees and the Swanson method yield almost identical results with quite accurate distributional forecasts. The Shimko method, on the other hand, yields density forecasts that are less accurate. Given that we obtain almost identical results based on two well-performing approaches, we argue that our results are not an artifact of the statistical method adopted. Moreover, with Binomial Trees being particularly designed to analyze American style options, we focus on this method in the remainder of this section. The other two approaches are discussed in the appendix.

3.2. The Binomial Tree

The idea of a Binomial Tree is that future (unknown) price developments of the underlying can be approximated by a series of discrete upward (u) and downward (d) jumps. Starting from the current price of the underlying at time t and calibrating

⁷This literature is extensive, hence a thorough review is beyond the scope of this paper. Important contributions include, but are by far not limited to, Black and Scholes (1973), Cox et al. (1979), and Longstaff and Schwartz (2001).

the tree to observed option prices, one can obtain a lattice that approximates the distribution $\tilde{g}_t(x_{t+h})$ based on the likelihood of observing a specific number of upward and downward jumps. Figure 3 illustrates the first two periods of a Binomial Tree. Following Cox et al. (1979) for the specification of the up and down movements, one can characterize the market value of the underlying as

$$\begin{aligned} x_{t+1} &= x_t * \underbrace{\exp\{\sigma\sqrt{\Delta t}\}}_u && \text{with probability } p \\ x_{t+1} &= x_t * \underbrace{\exp\{-\sigma\sqrt{\Delta t}\}}_d && \text{with probability } 1 - p, \end{aligned}$$

where the parameter σ determines the dispersion of the estimated RND.⁸ The variable Δt captures the length of each discrete time interval: with longer intervals, the jumps have to increase in order to generate the same uncertainty about the underlying asset at maturity. We fix the total number of intervals (or periods) at N to ensure that the RNDs are approximated by the same number of nodes regardless of the horizon h . As a consequence, all forecasted distributions are equivalently “smooth”.

The probability of an upward movement, p , is given by

$$p = \frac{e^{r\Delta t} - d}{u - d},$$

which equalizes the expected return for each Δt with the expected return of the continuous counterpart over the same period.⁹ Furthermore, given a real risk-free rate r , the probability of an upward movement depends on only one unknown parameter: σ .

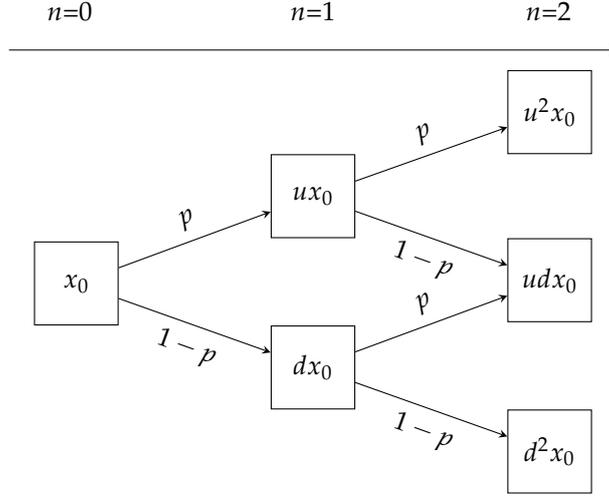
A few remarks about this approach are in order. First, due to the exponential function in the jump process, the value of the underlying asset is strictly greater than zero at each node, which is a realistic feature as oil prices are generally positive.¹⁰ Second, the tree is recombinant, i.e. if the underlying moves up and then down (u,d), the value (price) will be the same as if it had moved down and then up (d,u). This property reduces the number of tree nodes, and thus the computational burden. Third, the discrete distribution converges to a log normal distribution if the number of periods

⁸More specifically, σ characterizes the standard deviation of the continuously compounded return of the underlying future contract.

⁹Formally, p is the solution to $px_tu + (1-p)x_td = x_t \exp(r\Delta t)$, where the risk-free rate r is the expected return of the underlying asset in a risk-neutral world.

¹⁰The oil price turned negative on April 20, 2020 during the early stages of the COVID-19 pandemic when concerns of physical storage for crude reaching full capacity implied acute over supply. However, oil prices returned to positive values the next day.

Figure 3: Binomial Tree: Illustration for 2 Periods



is large. Therefore, the crucial assumption is that oil futures returns approximate a normal distribution, an assumption that is also imposed by [Black and Scholes \(1973\)](#). How consequential is this assumption? [Figure A1](#) plots returns from oil futures, both at daily and quarterly frequency. While returns are not normally distributed, the deviations are rather small. Overall, they appear to have more mass around the center, thus approximating a t-distribution. The more subtle and also more important question is if this implicit distributional assumption relates to our main finding that investors consistently underestimate tail risk. We argue this is not the case, primarily because we derive similar results based on the non-parametric [Swanson \(2006\)](#) approach.

We are now able to price an American call or put option. Let n denote the discrete period of the Binomial Tree, $n = \{0, 1, \dots, N\}$. We denote Binomial Tree episodes by n rather than time t , as the time interval of a period in the Binomial Tree does not coincide with calendar dates. However, both counters are linked: period N is equivalent to time $t + h$. In other words, the last set of nodes in the Binomial Tree represents an estimate for $\tilde{g}_t(x_{t+h})$. As such, $n = 0$ refers to “today,” i.e. time t . At the terminal nodes of the tree, the payoff and hence the price of a call option at maturity is simply

$$c_{N,i,j}(0,k) = \max(u^i d^j x_0 - k, 0) \quad \text{s.t. } i+j=N \text{ and } i, j \geq 0, \quad (4)$$

where i (j) denotes the number of up (down) movements. Therefore, i and j jointly characterize the specific node of the distribution $\tilde{g}_t(x_{t+h})$. Iterating back, in any given period $n < N$, the value of an option depends on the maximum of exercising the call option and the continuation value, that is

$$c_{n,i,j}(N-n,k) = \max \left(\underbrace{\left(u^i d^j x_0 - k \right)}_{\text{Early Exercise}}, e^{-r\Delta t} \underbrace{\left(p c_{n+1,i+1,j}(N-n+1,k) + (1-p) c_{n+1,i,j+1}(N-n+1,k) \right)}_{\text{Continuation Value}} \right) \quad (5)$$

such that

$$i+j=n \text{ and } i,j \geq 0.$$

The continuation value represents the discounted expected payoff from holding the option until the next period. With probability p , the price of the underlying jumps up and with probability $1-p$, the price jumps down. With backward induction, we can then recursively calculate the Binomial Tree implied option price at time t , or $n=0$, as a function of σ , the only unknown. We choose σ to minimize the difference between the model implied option price and the observed option price in the data:

$$\min_{\sigma} \left\{ c_{0,0,0}(h,k;\sigma) - c_t(h,k)|_{data} \right\}, \quad (6)$$

for each strike, maturity, trade date, and put or call option.

We then introduce the subscript l to uniquely identify options based on their strike price and option type, i.e. $l \in \{k, put=\{0,1\}, call=\{0,1\}\}$. For given h and t , we obtain an estimate $\hat{\sigma}_l$ from the minimization problem described by equation (6) and iterate the tree forward to estimate $\tilde{g}_t(\widehat{x_{t+h}})_l$. In other words, we obtain one distribution for each l at every forecasting horizon and date. To obtain our object of interest $\tilde{g}_t(\widehat{x_{t+h}})$, we average these individual distributions based on a weighting scheme that allocates more weight to triples l with a higher trading volume. As shown in Figure A2, strikes closer to the value of the underlying are more frequently traded. More liquid options in turn might contain a more precise signal about future beliefs. But even without this argument, it is intuitive to allocate the same weight to each individual transaction, which requires a weighting scheme based on volume per strike. The description of our algorithm can be summarized as follows:

Binomial Tree Algorithm: Let the time interval between two nodes be $\Delta t = h/N$ months, where $N = 500$, and assume an annual real risk-free interest rate of 0.02 (monthly rate, r , is therefore 0.02/12). Then, for each l identifying the strike and option type, horizon h , and trade date t ,

1. given a target σ , the model implied option price at time t is recursively calculated according to equations (4) and (5),

2. $\hat{\sigma}_l$ is obtained by solving the minimization problem (6),
3. given $\hat{\sigma}_l$, forward iteration of the Binomial Tree from time t until $t+h$ yields one forecasted density $\widehat{\tilde{g}_t(x_{t+h})}_l$,
4. these individual $\widehat{\tilde{g}_t(x_{t+h})}_l$'s are weighted by the number of transactions for each l relative to the total number of transactions to obtain $\widehat{\tilde{g}_t(x_{t+h})}$. That is, let $\#l_{t,h}$ be the number of transactions related to l for given t and h , then

$$\widehat{\tilde{g}_t(x_{t+h})} = \sum_l \widehat{\tilde{g}_t(x_{t+h})}_l \frac{\#l_{t,h}}{\sum_l \#l_{t,h}} \quad \forall t, h.$$

Note that each individual distribution $\widehat{\tilde{g}_t(x_{t+h})}_l$ is by construction discrete with unique x_{t+h} values. So, to average discrete distributions, we first generate identical bins (with a size of one U.S. dollar each) across individual forecasts.

It is worth emphasizing that our results are robust to various weighting schemes.¹¹ In the next subsection, we show that the estimated distributions provide a reasonable characterization of the oil market.

3.3. An Initial Evaluation of Oil Density Forecasts

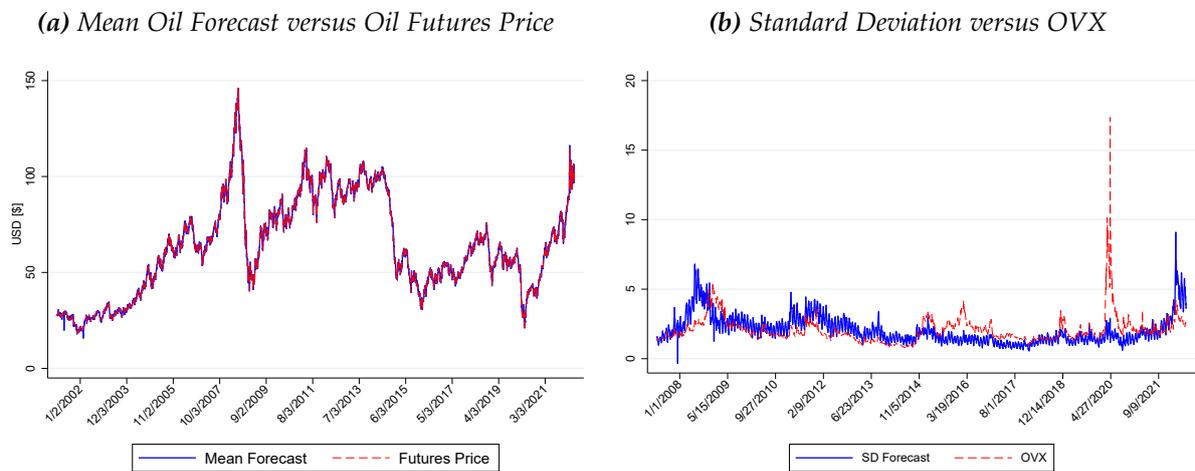
We begin by examining the first and second order moments generated by our risk-neutral distributions (RNDs) for an initial assessment of their overall accuracy. Figure 4, Panel (a) presents the mean of the RNDs based on the WTI crude oil options (solid blue line) along with the price of the WTI futures from 2001 to mid-2022 (red dashed line). These two series are expected to overlap given that futures prices reflect the price of oil that the buyer and the seller both agree upon delivery, hence providing direct information about investors' expectations. As can be seen, oil prices have posted large fluctuations ranging between around \$20 a barrel and \$140 a barrel over this period. For example, from the end of 2001 to mid-2007, oil prices tripled with a steady increase. Prices then steepened sharply and reached an all time high of over \$140 a barrel in the summer of 2008, only to be followed by a dramatic collapse.¹² The

¹¹In particular, we obtain identical results when we increase the required transactions per strike and day (from 10 in our baseline), or when we only derive forecasting distributions based on at-the-money options.

¹²Remarkably stable global oil production in the 2000s combined with strongly growing global oil demand, largely driven by China, contributed to the price appreciation during 2007-2008 (Hamilton, 2009). These gains were erased in less than six months with the global financial crisis weighing on demand just as additional supplies were ramping up, and the market experienced a historic price collapse.

first moment of our RNDs captures these extreme price movements. The mean is almost perfectly correlated with the underlying futures price over the entire sample (corr=0.99), validating the reliability of our estimates.

Figure 4: Validity of Density Forecasts



Notes: Panel (a): The solid blue line displays the mean of the extracted distributions based on oil options expiring within 2-3 months. The dashed red line presents the price of the underlying future expiring at the same time. Sample: Daily Observations, January 2001-June 2022. Panel (b): The solid blue line displays the standard deviations of the distributions based on oil options expiring within 20-40 days. The dashed red line represents the level of the OVX. The OVX series is only available since May 2007 and is a direct measure of implied volatility over the next 30 days. Both series are normalized. Sample: Daily observations, May 2007 - June 2022.

A more critical validation requires looking into higher moments, such as volatility. One way to directly measure oil market volatility based on option prices is the CBOE's Crude Oil ETF Volatility Index (OVX). The index characterizes the implied 30 day volatility, just as the VIX for the S&P 500. The red dashed line in Figure 4, Panel (b) presents the historical OVX. Not surprisingly, the OVX reached a peak during the financial crisis in line with the severe price movements highlighted above. The implied volatility was even higher during the onset of the pandemic with the OVX reaching unprecedented levels, largely driven by the economic contraction caused by COVID-19 and a sudden increase in oil supply following the suspension of agreed production cuts in early 2020 among the Organization of the Petroleum Exporting Countries (OPEC) and partner countries. Overall, this option-implied parameter co-moves reasonably well with the standard deviation of our extracted RNDs (right panel, blue line), though the standard deviation misses the spike in April 2020. In fact, the correlation between the two series is 0.31.¹³

¹³The OVX is based on a different methodology. Moreover, its calculation is based on changes in oil prices, whereas our RNDs forecast the level of the oil price. These factors contribute to the less than perfect fit.

Figure 5: Density Forecasts: 10th, 50th, 90th Percentiles and Futures Price at Maturity

(a) 1-Month Horizon

(b) 6-Month Horizon



Notes: Charts present the 10th, 50th, 90th percentiles of the forecasted distribution, as well as the oil futures price at different maturities: panel (a) plots statistics for the 1-month horizon, panel (b) for the 6-month horizon. In each month, we choose the trading date that is closest to a maturity in 1 or 6 months. Shaded areas represent the 10 - 90 percentile range. Sample: Monthly observations, January 2001 - June 2022.

Finally, we look into the behaviour of our RNDs at different maturities and over time. As uncertainty increases with the forecasting horizon, one may expect the dispersion of the RNDs to increase at more distant maturities (see, for example, [Chen et al., 2018](#)). In this context, Figure 5 presents the 10th, 50th, and 90th percentiles of our estimated RNDs at two maturities, 1-month and 6-month, over the entire sample. Similar charts for all 12 forecasting horizons are available in Figure A3 in the appendix. The blue lines show the median forecasts and shaded areas indicate the 10-90 percentile range. In the panels, we also present the realizations of the 1-month ahead and 6-month ahead oil futures prices (red dashed lines). Comparing the left and right panels of Figure 5 reveals two notable features. First, estimated distributions widen with the horizon. Second, although most realizations are within the 10-90 percentile range, there are deviations particularly at the lower end, hinting towards an under-prediction of left tail events. To explore these findings in more depth, we develop a systematic way to evaluate the accuracy of our estimated distributions in Section 4.

4. NEW MEASURES OF DENSITY FORECAST ACCURACY

Our framework presented in Section 3 highlights several factors that could lead to deviations between the forecasted distribution and the target distribution. We quantify these (potential) deviations via our ‘Total Accuracy’ and ‘Relative Accuracy’ statistics.

The main difference between these statistics and existing work (see, for example, [Kolmogorov, 1933](#); [Berkowitz, 2001](#); [Bai and Ng, 2005](#); [Corradi and Swanson, 2006](#)) is that we do not resort to a limited set of characteristics, such as the first few moments or the maximum distance between actual and target distribution, to determine the accuracy of a distribution. Our statistics also do not involve formal hypothesis testing and hence do not require a limiting distribution or assumptions about the correlation structure.

A density forecast is accurate ('correctly calibrated') if its distribution closely matches the true uncertainty of the underlying for the same horizon h and the same time period t . A correctly calibrated density forecast, therefore, does not necessarily imply an accurate mean forecast of the underlying, which is difficult to come by. For example, it has been thus far challenging to provide accurate point forecasts in the oil market, even for shorter horizons (see, for example, [Alquist et al., 2013](#), [Calomiris et al., 2021](#), [Baumeister et al., 2022](#)). Using our new measures, we show that our density forecasts are relatively accurate, but they have a large variance, making it difficult to derive precise point forecasts.

In a nutshell, our assessment of density forecasts is based on two steps. First, we calculate the percentile of the underlying at maturity relative to the forecasted distribution, which is often referred to as a probability integral transformation. Second, we examine if these realized percentiles for many different dates follow a uniform distribution on the $[0, 1]$ interval, which would be the (desired) outcome if the density was correctly calibrated ([Dawid, 1984](#); [Diebold et al., 1998](#)). The intuition is as follows. The $t+h$ forecast performed at time t contains information that is relevant at time t . Over the period of h months, however, new information arrives. A negative demand shock, for example, can reduce the oil price, resulting in a price at time $t+h$ that is at the lower end of the forecasted distribution. An individual realization is therefore not particularly informative about the accuracy of the distribution. However, if upside or downside risks are appropriately forecasted, then as multiple realizations at different periods in time are collected, positive and negative shocks cancel out. Because percentiles of *any* distribution are uniformly distributed, one can contrast the distribution of realized percentiles with a uniform distribution. The distance then determines the precision of the forecast.

We conceptualize these thoughts in Section [4.1](#) and introduce two new statistics that allow us to determine the overall forecasting accuracy and the accuracy over certain percentile intervals in Section [4.2](#).

4.1. Methodology

Let us denote the CDF of $\tilde{g}_t(x_{t+h})$ by $\tilde{G}_t(x_{t+h})$. By construction, the percentiles of a CDF would follow a uniform distribution on the interval $[0, 1]$, thus,

$$\tilde{G}_t(x_{t+h}) \sim U_{[0,1]} \quad \forall t, h. \quad (7)$$

For every time period t and forecasting horizon h , the associated percentile of *one* realization of the oil future price at $t+h$, referred to as \hat{x}_{t+h} , is $\tilde{G}_t(\hat{x}_{t+h})$. If the forecasted distribution was accurate, then \hat{x}_{t+h} would be drawn from $\tilde{g}_t(x_{t+h})$, implying $\tilde{G}_t(\hat{x}_{t+h}) \sim U_{[0,1]}$. In other words, unless \hat{x}_{t+h} was drawn from a different distribution, the percentiles of realized observations would follow a uniform distribution.

An individual observation is insufficient to obtain a distribution over the realized percentiles. However we can exploit that equation (7) holds for all t . Formally, one can stack the CDFs for a given horizon h in the vector $\tilde{\mathbf{G}}(x_h) \equiv \{\tilde{G}_t(x_{t+h})\}_{\forall t}$, where x_h refers to the underlying x , h months ahead for all t . Each element in $\tilde{\mathbf{G}}(x_h)$ by construction follows a uniform distribution:

$$\tilde{\mathbf{G}}(x_h) \sim U_{[0,1]} \quad \forall h. \quad (8)$$

The crucial point is that we have about 13 to 14 observations for $\tilde{\mathbf{G}}(\hat{x}_h)$ per month, which based on 21 years of data, implies about 3300 realizations for 1-month forecasts. The sample size declines with the forecast horizon as shown in Table 1 due to sample boundaries and fewer days with trading activity. Yet, the sample size remains large, which is helpful for two reasons. First, our results rely on a Law of Large Numbers. Without sufficient observations, the CDF of the realized percentiles ($\tilde{\mathbf{G}}(\hat{x}_h)$) does not converge to a uniform distribution, even when the forecasted distribution is correctly calibrated. Second, realized percentiles for multi-period ahead forecasts are serially correlated. For example, realizations from forecasts made in January and February 2020 for April and May 2020, respectively, would be both in the left tail, simply because the COVID-19 shock was not yet realized at the beginning of 2020. Such persistence shrinks the effective information in each realization, requiring more data that covers time periods with positive *and* negative shocks to the price of oil.

Finally, it is important to point out that the aggregation of CDFs across time is not without a caveat. If realized percentiles $\tilde{\mathbf{G}}(\hat{x}_h)$ deviate from a uniform distribution, one can conclude that at least some density forecasts $\tilde{g}_t(x_{t+h})$ are inaccurate. However, we cannot identify individual periods during which density forecasts are inaccurate. That is, our results in subsequent sections represent averages over time.

Table 1: Sample Size by Horizon

$h =$	1	2	3	4	5	6	7	8	9	10	11	12
Obs.	3746	3732	3702	3666	3545	3248	2784	2322	1859	1526	1243	1006

Notes: The table presents the number of realizations for each horizon h .

4.2. Accuracy Measures

This subsection introduces two novel statistics, which we refer to as ‘Total Accuracy’ ($TAcc(h)$) and ‘Relative Accuracy’ ($RAcc(\alpha, h)$). The former provides insights into the overall accuracy of a forecasted distribution, while the latter sheds light on the accuracy over individual percentiles. Importantly, these statistics are applicable to any density forecast and are not limited to option implied distributions.

Let’s begin with defining the realized percentile at maturity by

$$\hat{\alpha}_{t,h} \equiv \tilde{G}_t(\hat{x}_{t+h}).$$

The empirical CDF of $\hat{\alpha}_{t,h}$ over t , $F_{\hat{\alpha}_h}(\alpha)$, can then be obtained as

$$F_{\hat{\alpha}_h}(\alpha) \equiv \frac{\sum_t \mathbb{1}\{\hat{\alpha}_{t,h} \leq \alpha\}}{T},$$

where T denotes the total number of realizations over time (see Table 1). The numerator counts the number of $\hat{\alpha}_{t,h}$ ’s that are less than or equal to α for a given horizon h . Based on previous arguments, the empirical CDF, $F_{\hat{\alpha}_h}(\alpha)$, follows a uniform distribution by the Law of Large Numbers provided that \hat{x}_{t+h} is sampled from $\tilde{g}_t(x_{t+h})$. Given that, we define our first statistic ‘Total Accuracy’ as follows:

$$TAcc(h) \equiv 1 - \frac{\int_0^1 (|F_{\hat{\alpha}_h}(\alpha) - F_{\alpha_h}(\alpha)|) d\alpha}{\int_0^1 F_{\alpha_h}(\alpha) d\alpha}. \quad (9)$$

The statistic is normalized to range between 0 and 1, where 1 indicates correct calibration. The numerator of the fraction calculates the absolute difference between the empirical CDF and the theoretical uniform benchmark, $F_{\alpha_h}(\alpha)$, and integrates the difference over all percentiles. The denominator is simply a normalizing object: It captures the maximum possible difference between the empirical CDF and the benchmark, which materializes when $F_{\hat{\alpha}_h}(\alpha) = 0 \forall \alpha$, as $F_{\hat{\alpha}_h}(\alpha)$ is by construction non-decreasing in α .

The second statistic, ‘Relative Accuracy’, corresponds to the PDF of the realized percentiles and captures the slope of the empirical CDF. More formally,

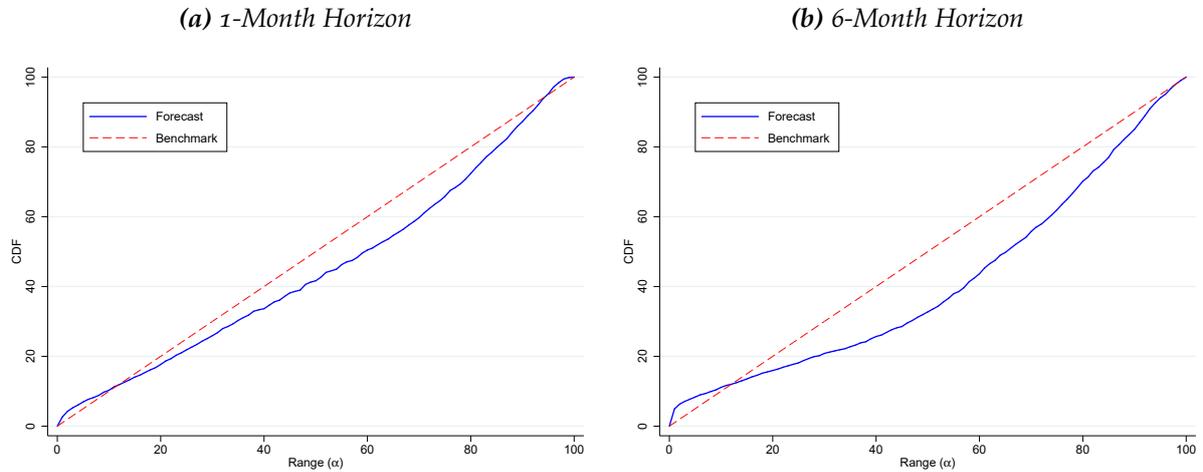
$$RAcc(\alpha, h) \equiv \frac{\partial F_{\hat{\alpha}_h}(\alpha)}{\partial \alpha}. \quad (10)$$

If \hat{x}_{t+h} was sampled from $\tilde{g}_t(x_{t+h})$, the slope of $F_{\hat{\alpha}_h}(\alpha)$ would be 1. Therefore, the relevant threshold for this statistic is 1. A value above 1 indicates oversampling relative the specific interval of the RND and a value below 1 suggests fewer realizations.

Note that both statistics do not follow a well defined distribution. We therefore determine confidence bands based on a simple block bootstrap algorithm that generates 1000 new samples by drawing from the original sample with replacement.¹⁴

We illustrate both statistics in Figure 6 for 1- and 6-month horizons. The solid blue lines show the empirical CDF, $F_{\hat{\alpha}_h}(\alpha)$, as a function of α . The dashed red lines indicate the hypothetical uniform distribution benchmark for a correctly calibrated density, $F_{\alpha_h}(\alpha)$. The numerator in the $TAcc(h)$ statistic simply integrates over the absolute difference between the red and blue line and the denominator normalizes the numerator by the maximum possible deviation between both lines. The $RAcc(\alpha, h)$ statistic represents the slope of the solid blue lines.

Figure 6: Empirical CDF versus Benchmark



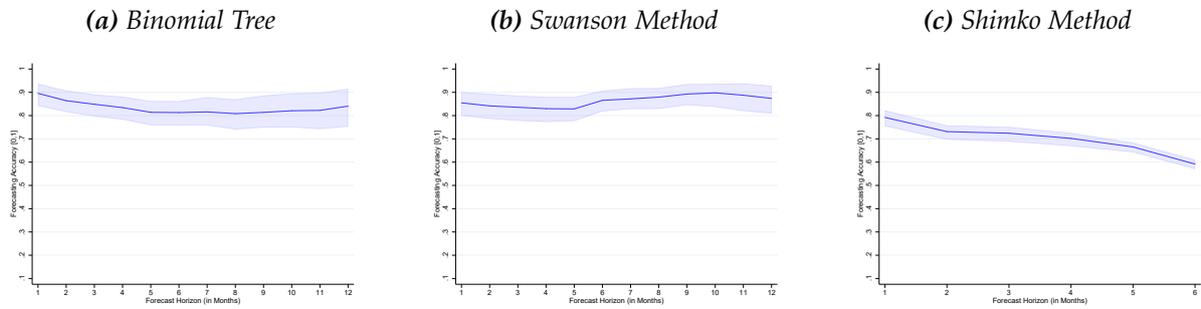
Notes: Horizontal axis: percentiles of distribution. Vertical axis: CDF. If the actual forecast (solid blue line) is below the benchmark (dashed red line), there are fewer realizations than predicted based on the distributional forecast for a given percentile. Panel (a) (Panel (b)) presents the accuracy of the distributional forecast for the 1 (6) month horizon. Sample: January 2001 - June 2022.

¹⁴Each block is of length $N^{1/3}$, where N is the total number of observations in the original data set.

5. RESULTS: THE MISSING TAIL RISK

We first assess the density forecasts based on our statistics developed in the previous section. Figure 7 shows the $TAcc$ of the density forecasts over different forecast horizons based on the Binomial Tree, the Swanson, and Shimko methods. In each panel, the solid lines present the values of the $TAcc$ measure and the shaded areas represent 90% confidence intervals. Panels (a) and (b) show that the Binomial Tree and Swanson method perform quite well yielding a relatively stable total accuracy measure that ranges between 0.8 and 0.9 across forecast horizons up to a year. In contrast, the Shimko approach performs significantly worse, with the $TAcc$ declining from 0.8 to below 0.6 as the forecast horizon increases from 1 month to 6 months.¹⁵

Figure 7: Overall Accuracy of Distributional Forecasts



Notes: These charts plots the accuracy of the distributional forecast based on Binomial Trees (Panel (a)), the Swanson method (Panel (b)), or the Shimko method (Panel (c)) over various forecast horizons (horizontal axis). The goodness of fit is characterized by the $TAcc(h)$ measure as defined in equation (9). A value of 1 refers to a “perfect” distributional forecast (‘correct calibration’). Shaded areas represent 90% confidence intervals based on 1000 block bootstrap replications. Sample: January 2001 - June 2022.

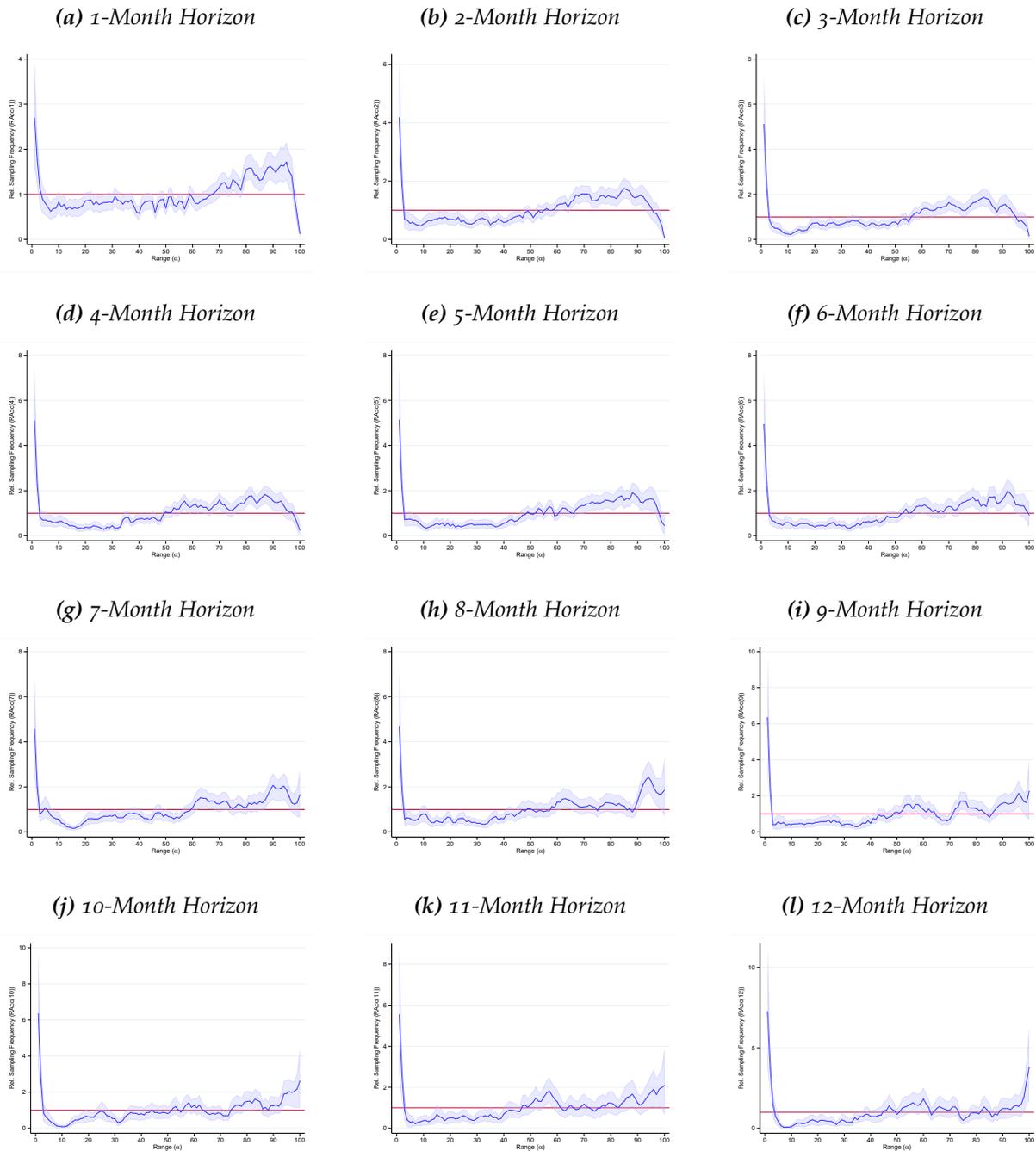
We draw two insights from these findings. First, the comparison of different approaches highlights the importance of an appropriate statistical procedure. As a result, we advocate to carefully assess a density forecast before interpreting it at face value. This can be accomplished via our $TAcc$ measure that can be universally applied. Second, the 90% confidence intervals for the Binomial Tree and the Swanson methods imply relatively accurate forecasts that overlap with each other despite the orthogonal assumptions underpinning each of these methods.

Next, we analyze the relative accuracy of our distributional forecasts. Figure 8 shows the relative forecasting accuracy ($RAcc$) measures over the entire distribution based on the Binomial Tree method across different forecast horizons.¹⁶ As apparent, the $RAcc$ measures in the right tail are generally close to 1 apart from the 1-month

¹⁵The numerical algorithm did not converge for forecasting horizons beyond 6 months.

¹⁶We smooth the $RAcc$ statistic to improve the exposition. Our results hold under the unfiltered $RAcc$ statistic too, see Figures A5 and A6.

Figure 8: Binomial Tree: Forecasting Accuracy over the Distribution



Notes: The charts portray the relative sampling frequency per percentile. A value above 1 implies oversampling. Horizontal axis: Theoretical percentile of distribution (α). Vertical Axis: Relative sampling frequency for a given percentile ($RAcc(\alpha, h)$). $RAcc(\alpha, h)$ is smoothed over three percentiles. The shaded area represents 90% confidence intervals based on 1000 block bootstrap replications. Sample: January 2001 - June 2022.

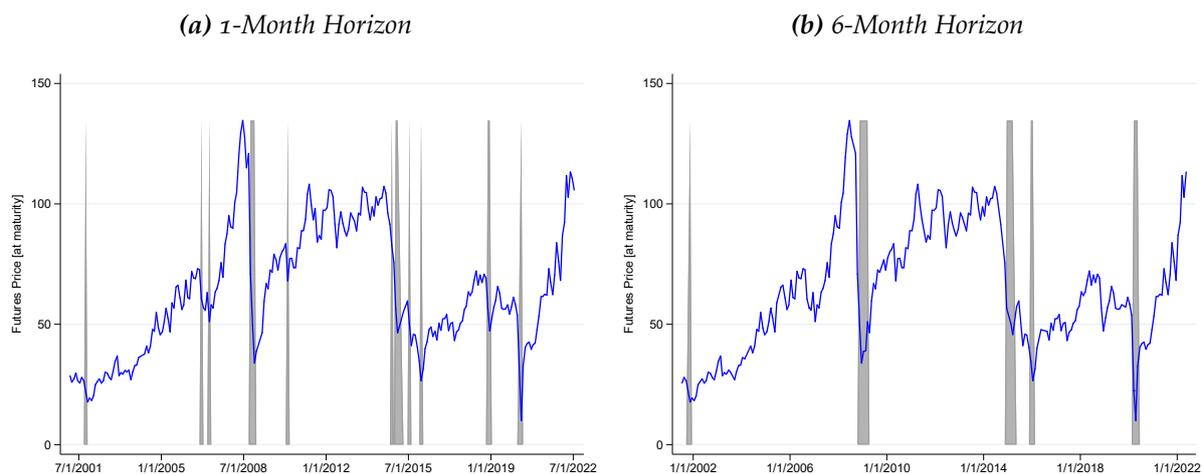
and 12-month horizons. Importantly, there is no systematic pattern regarding the over- or under-sampling of oil prices in the right tail. This changes in the left tail: At every horizon, the smoothed $RAcc$ for the 1st percentile is well above 1, indicating oversampling of left tail events relative to the specific interval of the forecasted

distribution. This feature carries over when we use the Swanson method as apparent from Figure A4. We subsequently zoom in on the anomaly in the left tail.

A Closer Look at the Left Tail

What type of events can characterize the left tail of the distribution over our sample period? Figure 9 plots the realized futures price (blue line) along with the left tail events that represent the 1st percentile realizations (shaded regions). These *extreme* events tend to cluster around either economic downturns or oil supply surprises. For example, a number of 1% left tail events occur during the Global Financial Crisis in 2008 and the pandemic recession in 2020, when global demand was hit hard. The episodes during 2014-2016 also coincide with a plunge in oil prices, partly driven by oil supply surprises including the U.S. shale oil boom. The key takeaway is that we observe more of these 1% events than expected based on our distributional forecasts.

Figure 9: Futures Price and Left Tail Events



Notes: The chart plots the futures price (blue line) and the incidence of left tail events defined as future price realizations within the 1st percentile of the forecasted distribution (grey shaded areas) for $h=1,6$. Monthly Observations. Sample: January 2001 - June 2022.

Discussion: What can explain the anomaly?

In section 3.1, we showed that the price of an option depends on investor beliefs, preferences, financial market frictions, and statistical methods used in estimating density forecasts. This means that any of these factors can play a role in driving the missing left tail of oil options.

To begin with, we can rule out two of these possibilities: financial frictions and statistical methods. As shown in Figure 2, oil options are very liquid, translating

into small bid-ask spreads.¹⁷ This limits the possibility that our results are driven by financial frictions. Furthermore, our results are not likely to be driven by the statistical method. This is because we observe similar results based on the Binomial Tree and the Swanson method, despite orthogonal modeling assumptions. In addition, we observe trading activity for options that insure against extreme left tail events (as shown in Figure A8), highlighting that our finding in the tail is not an extrapolated artifact of the estimation method but instead is driven by actual data.

This leaves us with an explanation based on misguided beliefs and/or preferences. We resort to the existing literature to argue that distorted beliefs are more likely the reason. Therefore, we interpret this robust anomaly as investors under-insuring themselves against oil price risk. In other words, the options are mispriced with investors being exposed to excessive risk during downturns when oil prices decline significantly. Standard finance models would interpret this persistent anomaly as investors “leaving money on the table”.

The first piece of evidence comes from [Christoffersen et al. \(2022\)](#), who found, based on the [Black \(1976\)](#) option price model, that very low and very high oil prices are associated with a high marginal utility. The argument is that large increases or large decreases in oil prices are indicative of deteriorating economic conditions, which in turn imply a high marginal utility. The ratio of the marginal utilities (SDF) would then be a u-shaped function of the oil price. A similar argument is also made by [Jacobs and Li \(2023\)](#). They find that call and put option returns in the oil market are on average negative, and more so for out of the money options in the tails of the oil price distribution. A negative return particularly in the tails implies that investors are willing to pay a premium for insurance against extreme outcomes, which again suggests a u-shaped SDF. Taking these findings at face value would imply that investors increase demand for options that insure against left or right tail events. A higher option price would then be associated with a higher probability mass in the tails of our risk-neutral density forecasts, which is clearly not what we observe.

One can infer further insights from studies examining the behavior of oil market participants. The literature examining the trading behavior of market participants in commodity futures tend to group traders as *commercials* and *noncommercials*, which are often viewed as *hedgers* and *speculators*, respectively.¹⁸ [Rouwenhorst and Tang](#)

¹⁷This is also confirmed by [Christoffersen et al. \(2022\)](#) and [Jacobs and Li \(2023\)](#), who show that crude oil options have very large open interest and trading volume.

¹⁸This grouping partly stems from the definition provided by the Commodity Futures Trading Commission. Note that although traders in commodity futures markets have been analyzed, see for example [Kang et al. \(2020\)](#), we are unaware of a comprehensive analysis of participants in the oil option market.

(2012) show for crude oil futures that open interest represented by *non-commercials* has gradually increased since the early 2000's, converging toward the share of *commercials*.¹⁹ Relatedly, the authors argue that commercial investors adjust their positions in response to price fluctuations relatively quickly. Assuming similar patterns hold for the oil option market, this has the following implications. First, the meaningful share of non-commercial investors suggests that investor characteristics are likely similar to other asset markets, such as equities where investors are risk averse (e.g., [Mehra and Prescott, 1985](#)). Taken together, these suggest that non-commercial investors in the oil market might have a high marginal utility when the oil price is significantly lower (recession fears). Second, if non-commercial investors can quickly adjust their portfolio of oil options, just like they do for futures, then they are likely to be marginal and drive equilibrium prices.²⁰

Preferences of commercial traders are more difficult to gauge. For example, jet fuel expenses represent about 20% of total operating expenses for airlines ([Rampini et al., 2014](#)). Oil price fluctuations hence represent a key source of cash flow risk. This may suggest that airlines hedge against high oil prices, but not necessarily low oil prices. But this is far from clear: [Rampini et al. \(2014\)](#) find that financial constraints (and not fuel prices per se) are an important determinant of risk management for commercial airlines. Airlines may therefore hedge less during economic downturns or when fuel prices skyrocket, as these episodes correlate with financial constraints and overall low net worth. Oil producers, on the other hand, could have a higher incentive to hedge against low oil prices, as their profits directly increase with oil prices. Overall, there is no clear evidence against a u-shaped SDF among commercial traders.

The next piece of evidence favoring misguided beliefs rather than preferences follows from [Dew-Becker et al. \(2021\)](#). The authors compute the return of straddles that fully hedge against risk (long call and put with same strike) and show that investors in multiple option markets (including crude oil) are willing to accept a negative risk premium to limit their exposure to realized volatility (negative gamma), which often coincides with economic downturns. Motivated by their findings, we also compute realized volatility, defined as the squared realized oil futures returns over the

¹⁹In more detail, the share of trader activity for non-commercials increased from below 20% to above 40% in a decade, while commercials' share dropped from about 80% to around 50% (non-reportables make up the rest). In a more recent paper, [Kang et al. \(2020\)](#) consider net trading (defined as the net purchase of futures contracts normalized by the open interest) as a measure to characterize the positions and trading behavior of market participants. They show that the absolute values of the net position changes of commercial and noncommercial traders average 1.72% and 1.42% of total open interest in the oil market, respectively. These overall suggest that non-commercials are as active as commercials.

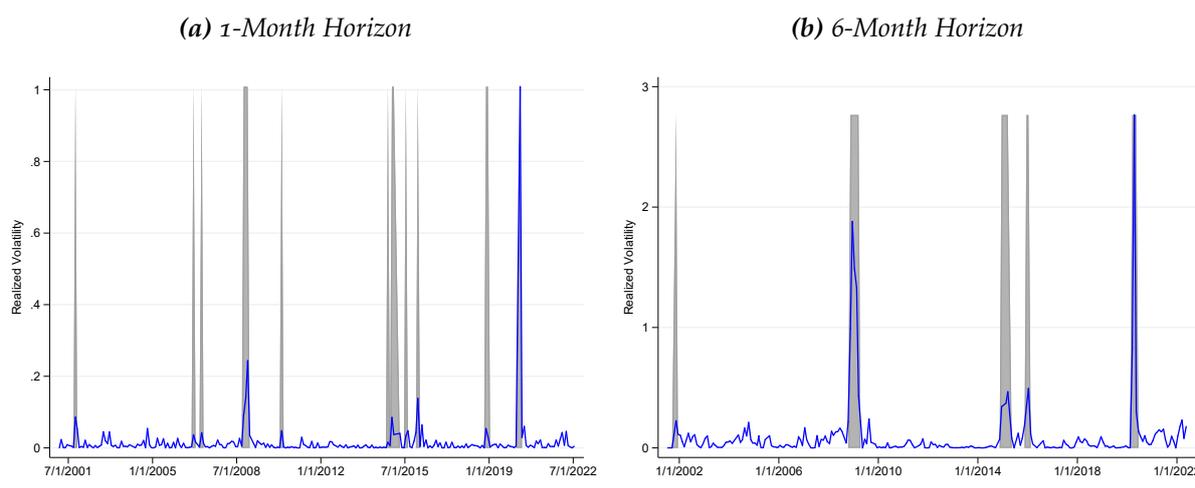
²⁰Supporting this argument is [Kang et al. \(2020\)](#)'s finding that non-commercials are on average momentum traders.

forecasting horizon, i.e.

$$\text{Realized Volatility} = (\ln(x_{t+h}) - \ln(x_t))^2.$$

The blue lines in Figure 10 plot this measure of realized volatility in oil options along with left tail events (shaded regions) for 1- and 6-month forecasting horizons. Similar graphs for the remaining horizons are available in Figure A9 in the appendix. We find that periods of high realized volatility are all associated with left tail events. This in turn suggests that investors are willing to pay a premium to insure against events that drive the aforementioned anomaly in the left tail. In other words, investors have a high marginal utility and hence a high SDF during left tail events, which should increase the risk neutral probability mass in the left tail, contrary to what we observe. As a result, we argue that the under-prediction of left tail events is likely due to a behavioral bias.

Figure 10: Realized Volatility and Left Tail Events



Notes: The charts present the realized volatility defined as squared ex-post returns along with the futures price realizations within the 1st percentile of the forecasted distribution (grey shaded areas) for $h=1,6$. Monthly observations over January 2001 - June 2022.

6. CONCLUSION

This paper contributes to the literature on deviations from rational expectations in financial markets and to the literature on evaluating density forecasts. In that context, we address two questions: How accurate are distributional forecasts based on oil option prices? Do we observe any behavior that suggests deviations from rational expectations?

To answer these questions, we propose two new statistics that 1) analyze the overall accuracy of a distributional forecast and 2) the relative accuracy over the range of the distribution. Both measures are easy to implement and can be broadly applied. Based on these novel statistics, we find an anomaly in the left tail of the distribution that is likely driven by a behavioral bias: Investors consistently underestimate the likelihood of very low oil prices and hence economic downturns. This type of underinsurance may have important implications. The investment portfolio could become excessively volatile. Similarly, businesses in the industry could expose themselves to excess risk during downturns.

We also highlight the importance of statistical methods. One needs to be careful when interpreting option based density forecasts at face value due to potential statistical errors in the estimation stage. We, therefore, suggest to compare different forecasting methods in terms of their accuracy and avoid premature conclusions regarding investor beliefs.

A. APPENDIX: METHODS

Swanson (2006) Approach

We subsequently detail a non-parametric approach to estimate density forecasts based on Swanson (2006). The crucial assumption of this method is that options are not exercised until maturity. This is equivalent to assuming that early exercise is never optimal. While this is true for American call options on non-dividend paying stocks, this is not true for put options, or call and put options on underlying futures, like the oil futures in this study. However, assuming that American options are not exercised prematurely, does not seem to have much of an effect on density forecasts (Chaudhury and Wei, 1994; Soderlind and Svensson, 1997).

The starting point for this non-parametric approach is equation (3). We rewrite this equation for $\tau=t+h$ below:

$$c_t(h, k) \approx e^{-rh} \int_0^\infty \max(x_{t+h} - k, 0) \tilde{g}_t(x_{t+h}) dx_{t+h}.$$

The equation prices a call option. For a put option, the arguments in the max operator are reversed. The risk neutral distribution $\tilde{g}_t(x_{t+h})$ is subsequently discretized into M bins $\{x_{1,t+h}, \dots, x_{M,t+h}\}$. We pick 100 bins ($M=100$) ranging from \$0 to \$200. Each realization $x_{i,t+h}$ materializes with the risk neutral probability $\tilde{p}_{i,t}$. Therefore,

$$c_t(k, h) \approx e^{-rh} \sum_{i=1}^M \tilde{p}_{i,t} \max(x_{i,t+h} - k, 0). \quad (\text{A.1})$$

Each day, we observe option prices for multiple strikes at a given horizon. We denote these strikes by $\{k_1, \dots, k_N\}$ where N refers to the total number of traded strikes and stack equation (A.1) across strikes. This provides a system of equations,

$$\mathbf{c}_t(h) \approx \mathbf{A}_t(h) \tilde{\mathbf{p}}_t,$$

where

$$\mathbf{c}_t(h) = \begin{bmatrix} c_t(k_1, h) \\ \vdots \\ c_t(k_N, h) \end{bmatrix}; \mathbf{A}_t(h) = e^{-rh} \begin{bmatrix} \max(x_{1,t+h} - k_1, 0) & \cdots & \max(x_{M,t+h} - k_1, 0) \\ \vdots & \ddots & \vdots \\ \max(x_{1,t+h} - k_N, 0) & \cdots & \max(x_{M,t+h} - k_N, 0) \end{bmatrix}; \tilde{\mathbf{p}}_t = \begin{bmatrix} \tilde{p}_{1,t} \\ \vdots \\ \tilde{p}_{M,t} \end{bmatrix}.$$

We can derive a similar system for put options and stack both systems, hence

$$\underbrace{\begin{bmatrix} c_t(h)^{call} \\ c_t(h)^{put} \end{bmatrix}}_{\underline{c}_t(h)} \approx \underbrace{\begin{bmatrix} \mathbf{A}_t(h)^{call} \\ \mathbf{A}_t(h)^{put} \end{bmatrix}}_{\underline{\mathbf{A}}_t(h)} \tilde{\mathbf{p}}_t,$$

which leads to the following constrained optimization problem:

$$\tilde{\mathbf{p}}_t^* \equiv \min_{\tilde{\mathbf{p}}_t} \{ \underline{\mathbf{A}}_t(h) \tilde{\mathbf{p}}_t - \underline{c}_t(h) \} \quad \text{subject to} \quad \sum_{i=1}^M \tilde{p}_{i,t} = 1 \quad \& \quad \tilde{p}_{i,t} \geq 0 \quad \forall i \quad (\text{A.2})$$

We solve this optimization problem separately for every horizon and trade date with a numerical nonlinear least-square solver in Matlab (“lsqnonlin”).

Shimko (1993) Approach

This procedure relies on a well-known result in finance by [Breedon and Litzenberger \(1978\)](#) which states the density forecast is proportional to the second derivative of the pricing function (3), that is,

$$\tilde{g}_t(x_{t+h} = k) \approx e^{rh} \frac{\partial^2 c_t(h, k)}{\partial^2 k}. \quad (\text{A.3})$$

We only observe the price of an option at a finite number of strikes, hence it is necessary to interpolate the cost function and then take a numerical second derivative. We follow the widely used approach by [Shimko \(1993\)](#), who acknowledges the sensitivity to the interpolation method, and suggests to interpolate implied volatility rather than the cost function directly. We obtain implied volatility based on the formula of [Black \(1976\)](#), which itself is a variant of the famous Black Scholes model adjusted for options with an underlying futures contract. The formula for a call option is

$$c_t(h, k) = e^{-rh} (x_t \Phi(d_1) - k \Phi(d_2)),$$

and for a put option,

$$c_t(h, k) = e^{-rh} (k \Phi(-d_2) - x_t \Phi(-d_1)).$$

Φ is the CDF of a standard normal distribution. d_1 and d_2 are defined as

$$d_1 = \frac{\ln(x_t/k) + (\sigma^2/2)T}{\sigma\sqrt{T}},$$

and

$$d_2 = d_1 - \sigma\sqrt{T},$$

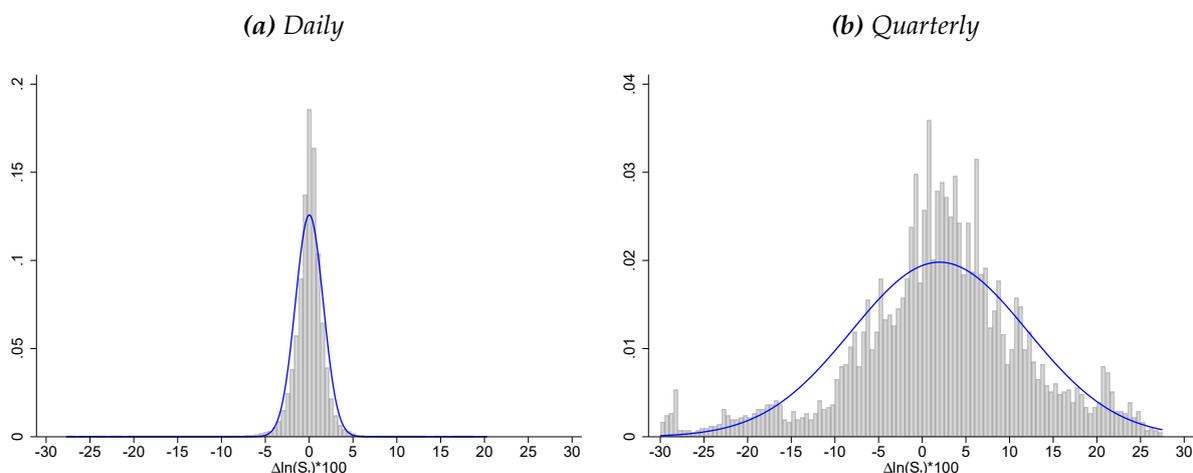
where T represents the days until maturity.

We solve for the implied volatility σ using numerical methods in Matlab (“fsolve”). For numerical reasons we drop all options with prices less than $< \$0.1$. We also drop spurious observations that do not satisfy basic arbitrage conditions, that is, a positive (negative) first derivative of the cost function with respect to the strike for a call (put) option, and options with a negative second derivative. We then fit a smoothing spline with smoothing parameter 0.99 through the implied volatility and strike price space for all calls and puts jointly. We average the implied volatility of calls and puts with the same strike.²¹ We then back out the option price based on Black (1976) for all interpolated implied volatilities (a retransformation). Last but not least, we take a numerical second derivative of the interpolated option price function (“diff” in Matlab), drop negative probabilities (prices with negative second derivative), and re-normalize the density to one. This provides one density forecast for each trade date t and forecast horizon h .

²¹We also analyzed calls and puts separately, however obtained a worse fit based on our $TAcc$ statistic.

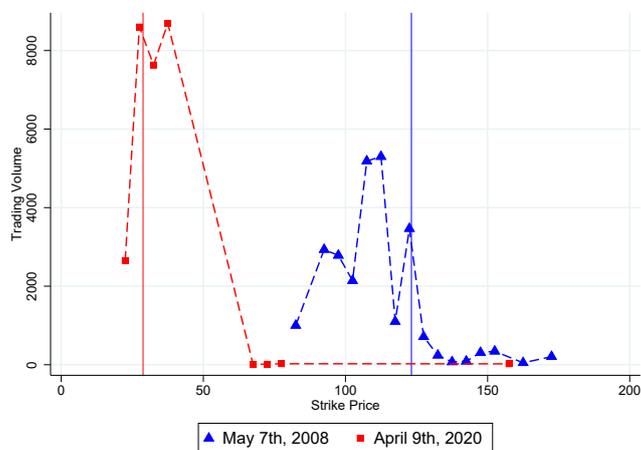
B. APPENDIX: ADDITIONAL FIGURES

Figure A1: Future Returns



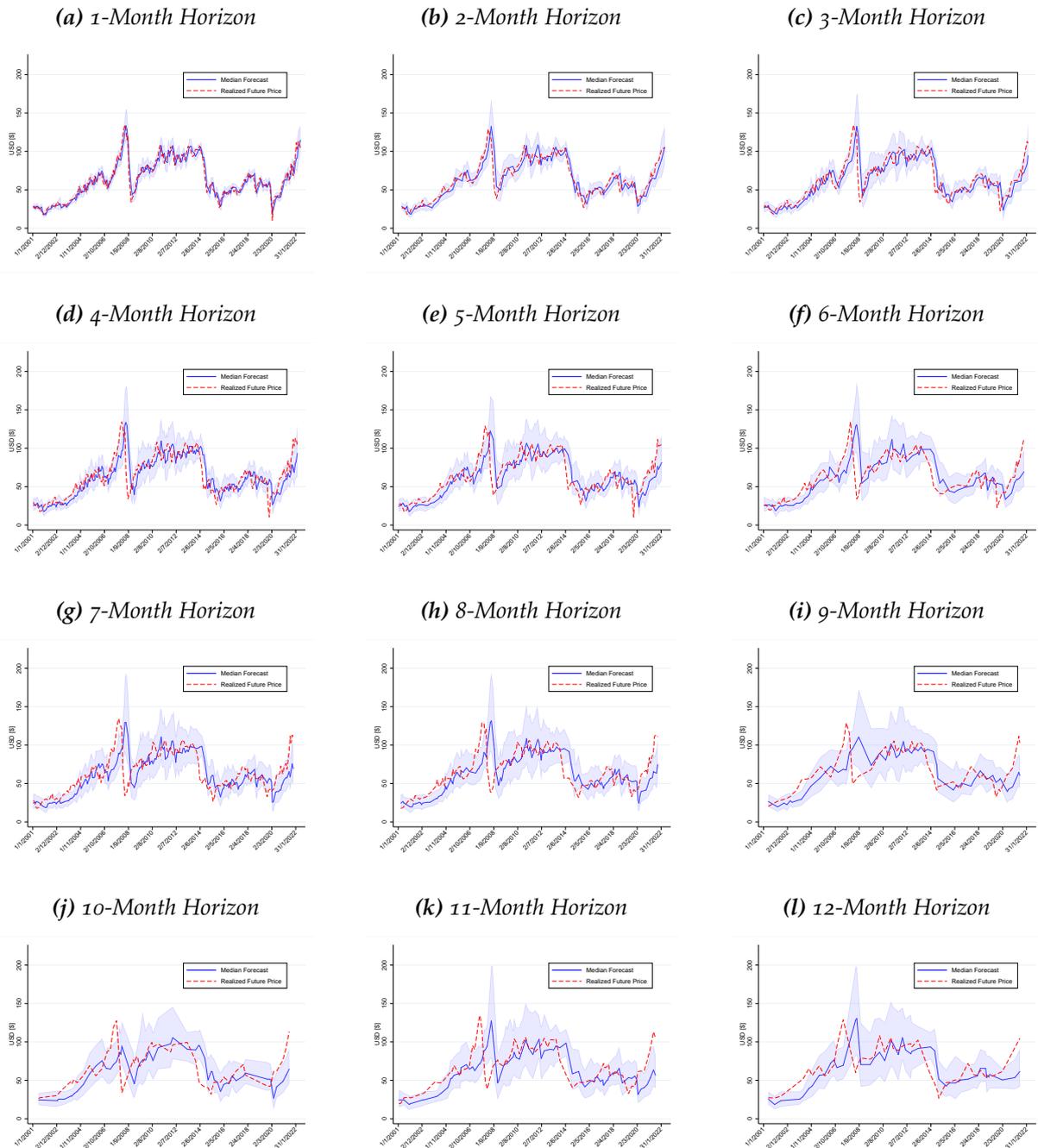
Notes: The charts show the daily (Panel (a)) or quarterly (Panel(b)) returns $\Delta \ln(x_t)$ (in %) of oil future contracts. The return is calculated as the log difference between consecutive days (or quarters). We focus on futures that expire in more or equal to 60 days to avoid effects from the convergence of future prices towards the spot price. The blue lines represent normal distributions with mean and standard deviation fitted to the data.

Figure A2: Trading Volume by Strikes



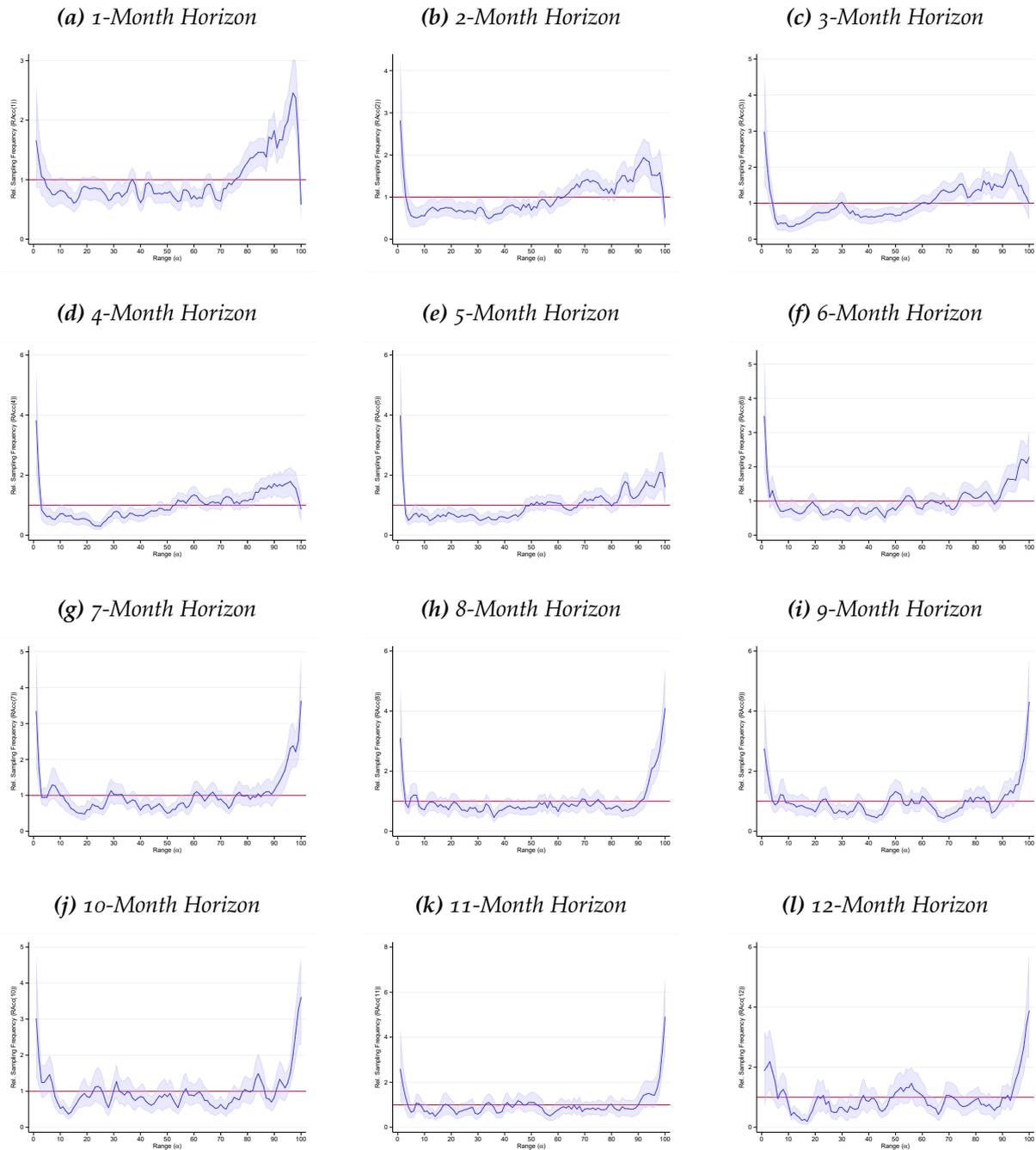
Notes: The chart plots the trading volume for 1-month options at two dates relative to the strike price. The specific dates are indicative of very high and low oil prices. Trading volumes are combined put and call volumes. Strike prices are further discretized into bins of 5 US dollar to smooth the distribution. The vertical lines portray the price of the underlying future.

Figure A3: Binomial Tree: Density Forecasts Over Time



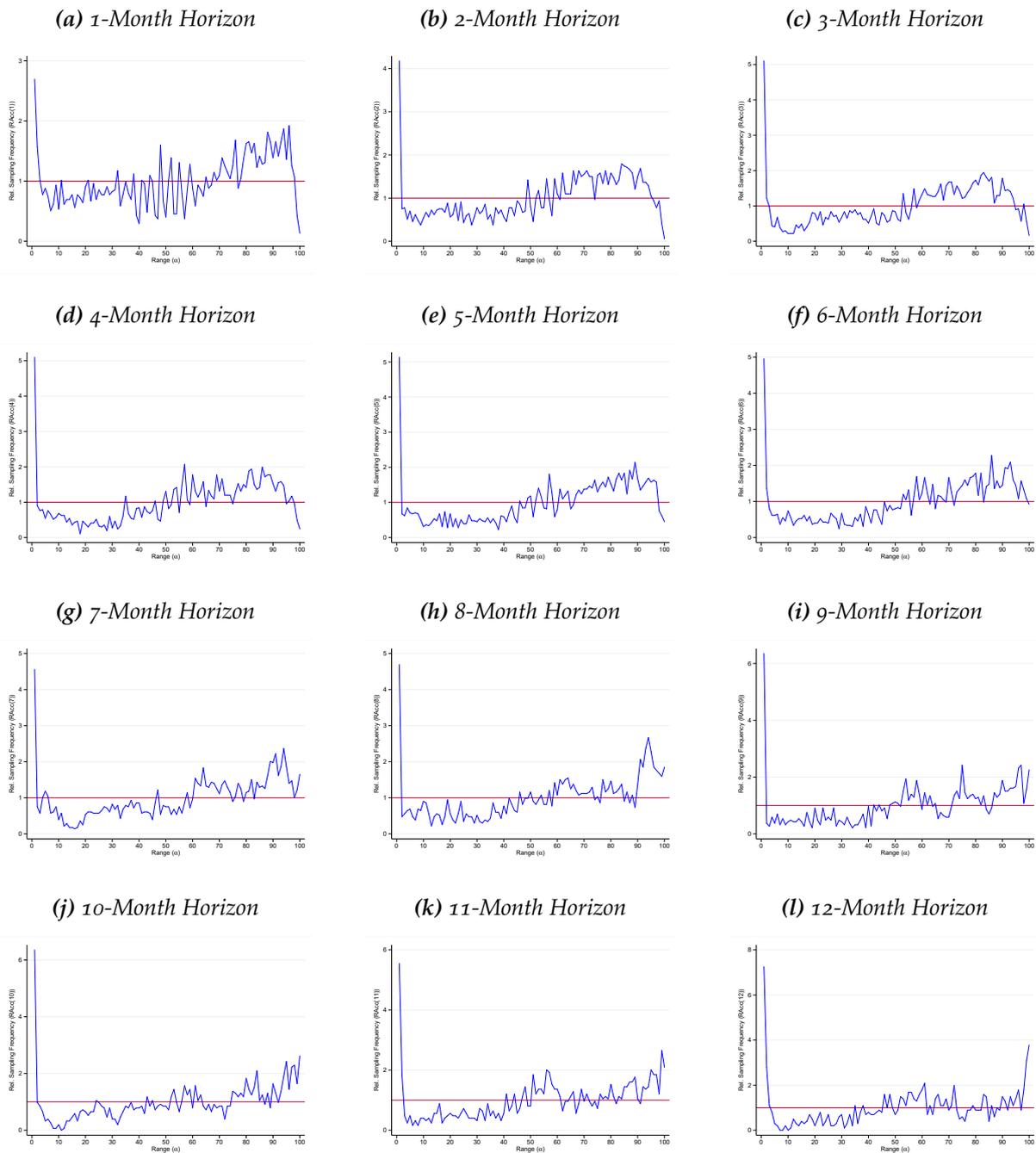
Notes: Each chart plots the 10th, 50th, 90th percentile of the forecasted distribution, as well as the oil future price at maturity over time. We choose trading dates that are closest to a maturity in exactly h -months. Shaded areas represent the 90th-10th percentile range. Sample: Monthly Observations, January 2001 - June 2022.

Figure A4: Swanson Method: Forecasting Accuracy over the Distribution



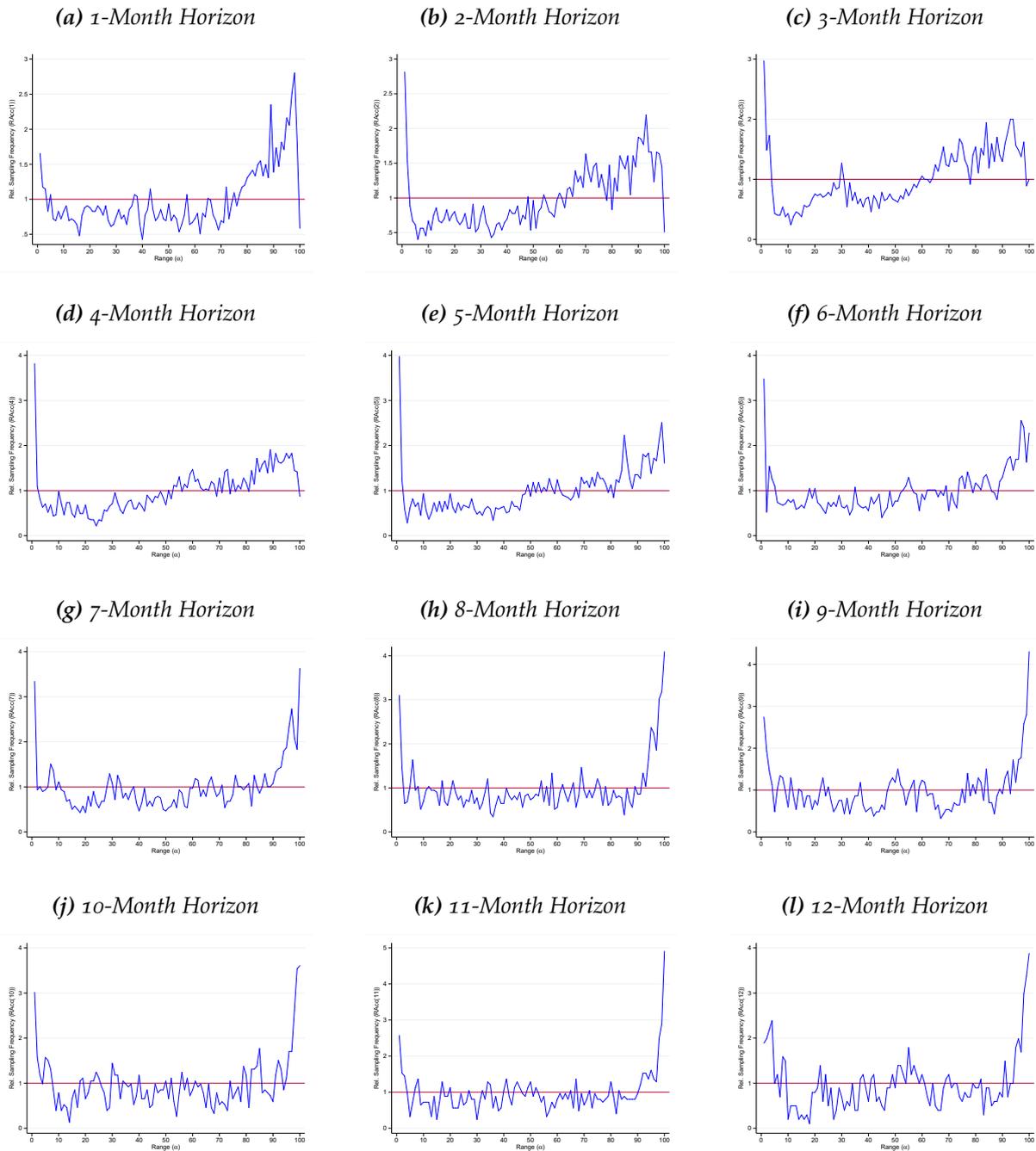
Notes: The charts portray the relative sampling frequency per percentile. A value above 1 implies oversampling. Horizontal axis: Theoretical percentile of distribution (α). Vertical Axis: Relative sampling frequency for a given percentile ($RAcc(\alpha, h)$). $RAcc(\alpha, h)$ is smoothed over three percentiles. The shaded area represents 90% confidence intervals based on 1000 block bootstrap replications. Sample: January 2001 - June 2022.

Figure A5: Binomial Tree: Forecasting Accuracy over the Distribution - Not Smoothed



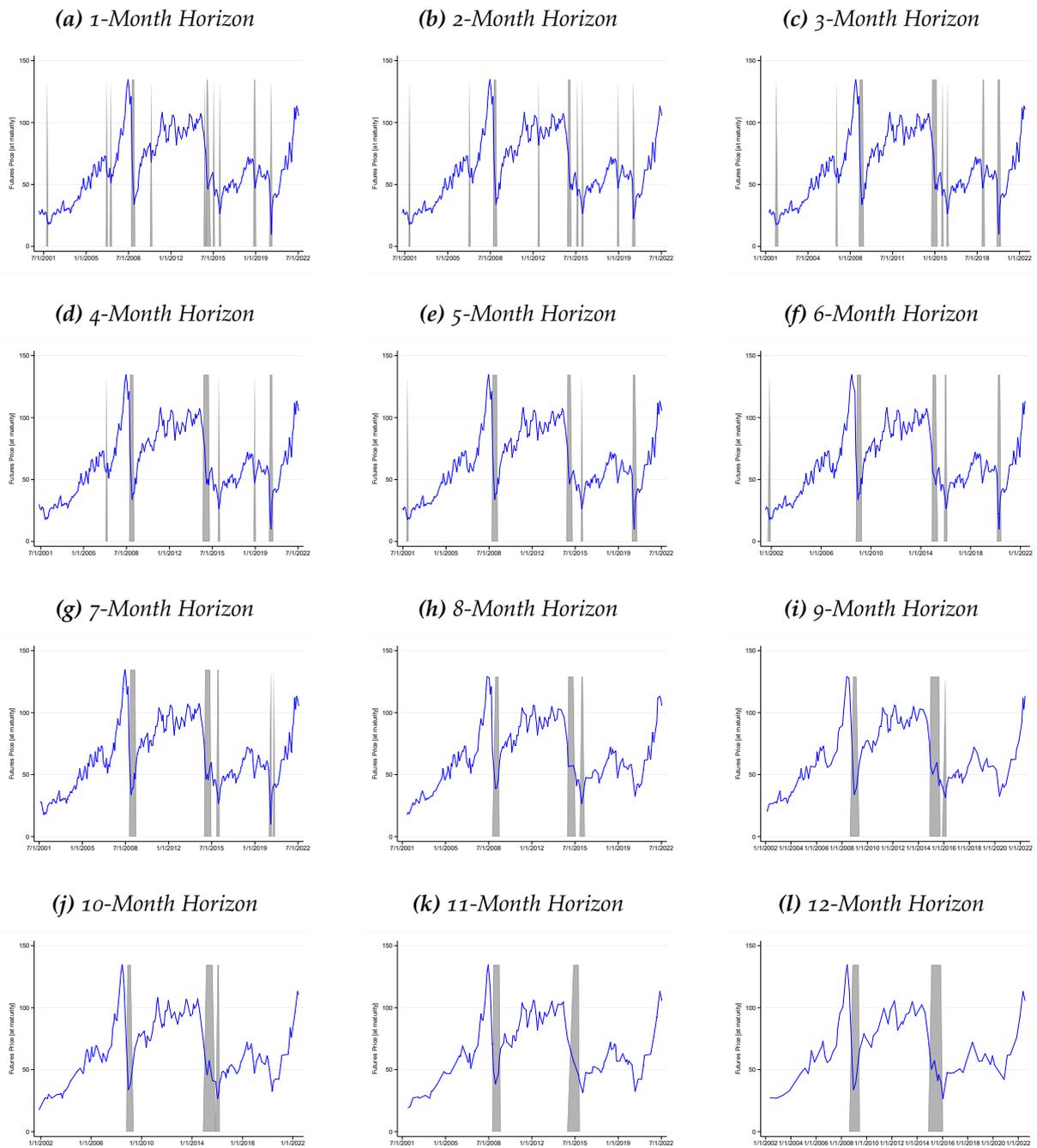
Notes: The charts portray the relative sampling frequency per percentile. A value above 1 implies oversampling. Horizontal axis: Theoretical percentile of distribution (α). Vertical Axis: Relative sampling frequency for a given percentile ($RAcc(\alpha, h)$). Sample: January 2001 - June 2022.

Figure A6: Swanson Method: Forecasting Accuracy over the Distribution - Not Smoothed



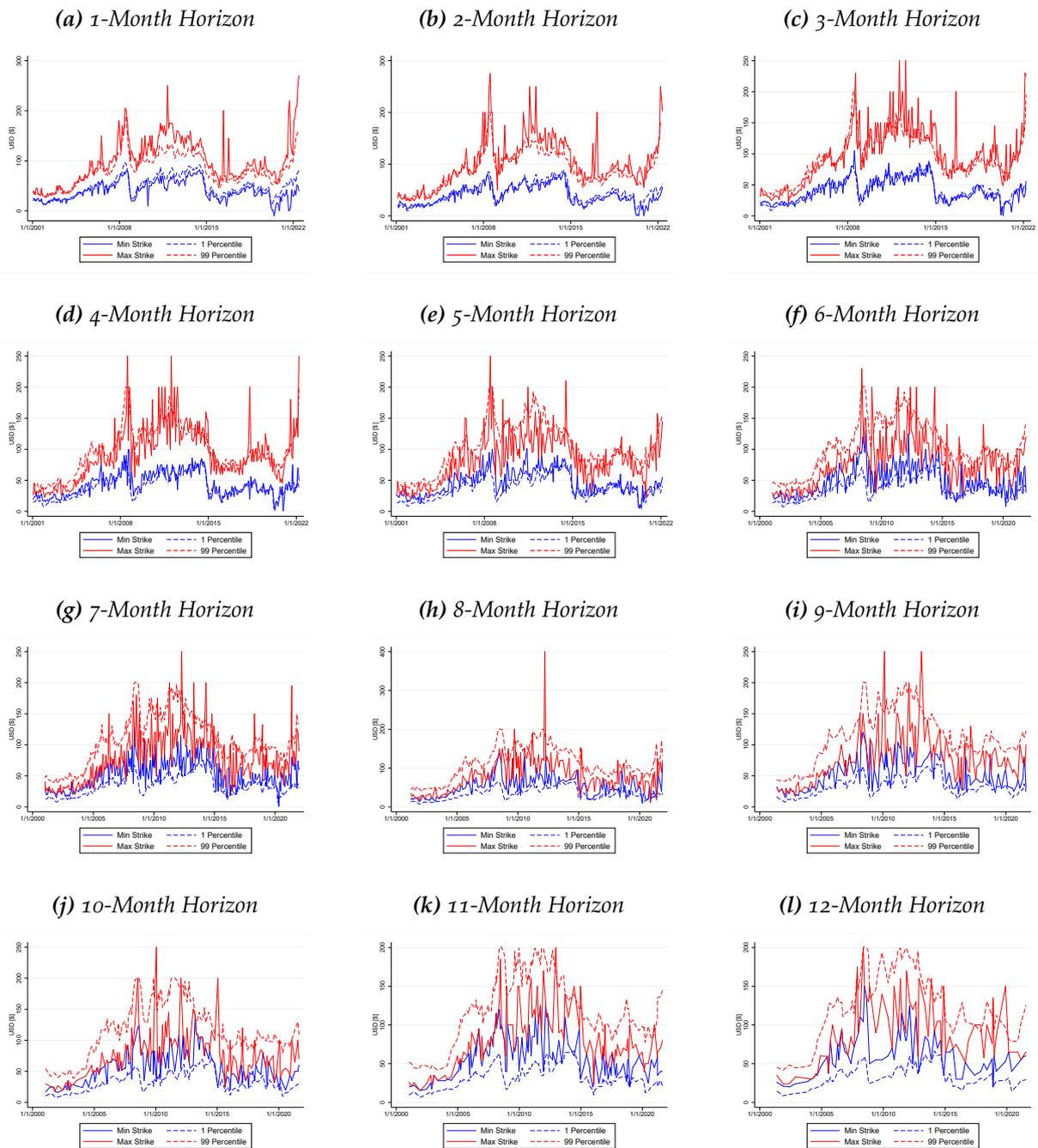
Notes: The charts portray the relative sampling frequency per percentile. A value above 1 implies oversampling. Horizontal axis: Theoretical percentile of distribution (α). Vertical Axis: Relative sampling frequency for a given percentile ($RAcc(\alpha, h)$). Sample: January 2001 - June 2022.

Figure A7: Futures Price and Left Tail Events - All Horizons



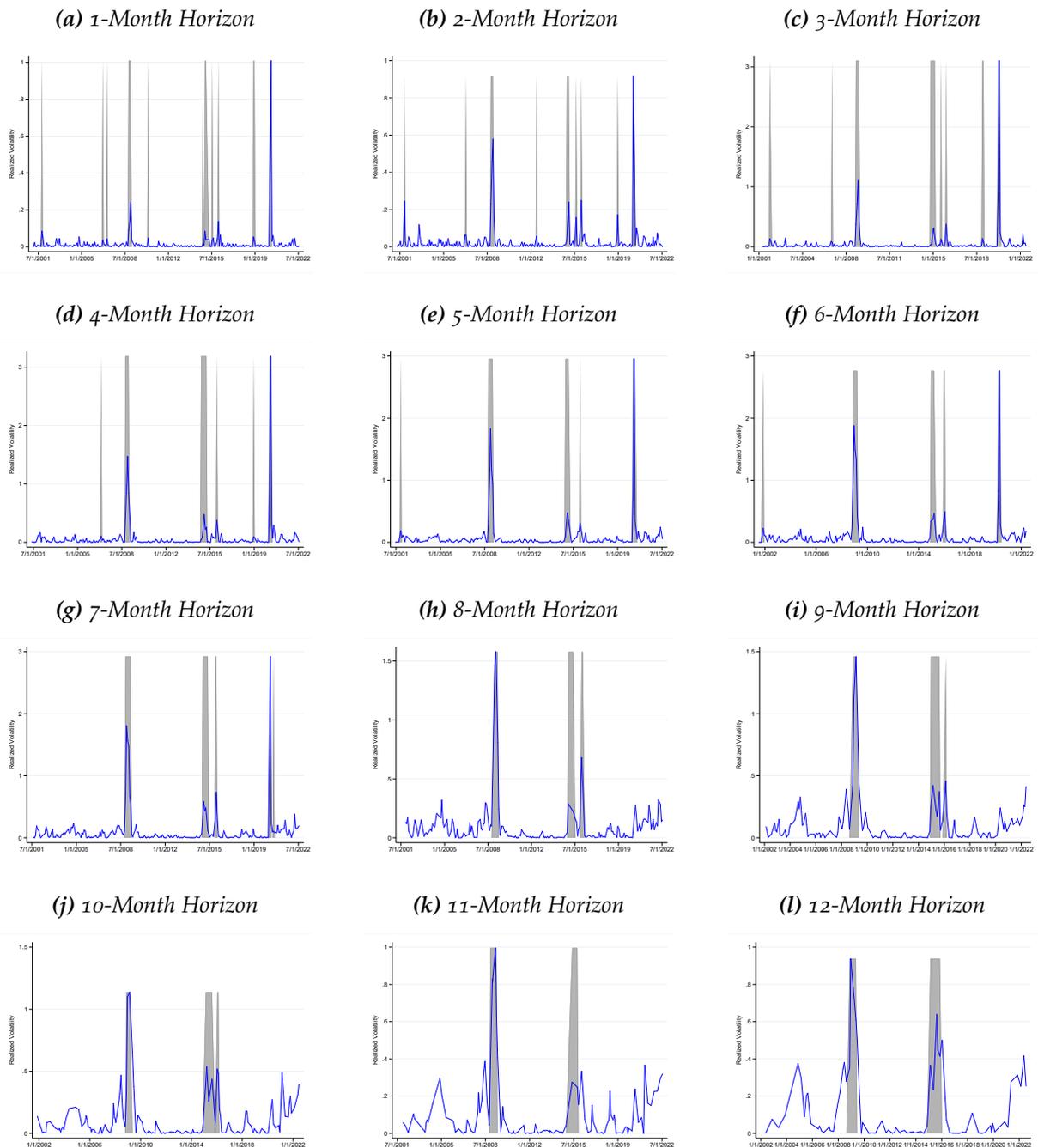
Notes: The charts plot the futures price (blue line) and the incidence of left tail events defined as future price realizations within the 1st percentile of the forecasted distribution (grey shaded areas) for various horizons. Monthly Observations. Sample: January 2001 - June 2022.

Figure A8: Minimum and Maximum Strikes and the Left/Right Tail of the Forecasted Distribution



Notes: The charts display the minimum and maximum strike over time along with the 1st and 99th percentile of the forecasted belief distribution. Monthly Observations. Sample: January 2001 - June 2022.

Figure A9: Realized Volatility and Left Tail Events - All Horizons



Notes: The charts portray the realized volatility defined as squared ex-post returns along with the futures price realizations within the 1st percentile of the forecasted distribution (grey shaded areas) for various horizons. The return is defined as the log difference between the futures price at maturity and at settlement. Monthly Observations. Sample: January 2001 - June 2022.

REFERENCES

- Alquist, R., Kilian, L., Vigfusson, R., 2013. Forecasting the price of oil, Elsevier. volume 2. chapter Chapter 8, pp. 427–507.
- Baele, L., Driessen, J., Ebert, S., Londono, J.M., Spalt, O.G., 2019. Cumulative prospect theory, option returns, and the variance premium. *The Review of Financial Studies* 32, 3667–3723.
- Bai, J., Ng, S., 2005. Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business & Economic Statistics* 23, 49–60.
- Barber, B.M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economic* 116, 261–92.
- Basak, S., Pavlova, A., 2013. Asset prices and institutional investors. *American Economic Review* 103, 1728–58.
- Baumeister, C., Korobilis, D., Lee, T.K., 2022. Energy markets and global economic conditions. *Review of Economics and Statistics* 104, 828–844.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business Economic Statistics* 19, 465–74.
- Black, F., 1976. The pricing of commodity contracts. *Journal of Financial Economics* 3, 167–179.
- Black, F., Scholes, M.S., 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 637–654.
- Bliss, R.R., Panigirtzoglou, N., 2004. Option-implied risk aversion estimates. *Journal of Finance* 59, 407–446.
- Bordalo, P., Gennaioli, N., Shleifer, A., 2022. Overreaction and diagnostic expectations in macroeconomics. *Journal of Economic Perspectives* 36, 223–44.
- Breeden, D.T., Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. *The Journal of Business* 51, 621–51.
- Calomiris, C.W., Çakır Melek, N., Mamaysky, H., 2021. Predicting the oil market. NBER Working Paper 29379 .
- Carr, P., Wu, L., 2007. Stochastic skew in currency options. *Journal of Financial Economics* 86, 213–247.

- Chaudhury, M.M., Wei, J., 1994. Upper bounds for american futures options: A note. *Journal of Futures Markets* 14, 111–116.
- Chen, R.R., Pei-lin, H., Huang, J., 2018. Crash risk and risk neutral densities. *Journal of Empirical Finance* 47, 162–189.
- Christoffersen, P., Jacobs, K., Pan, X.N., 2022. The State Price Density Implied by Crude Oil Futures and Option Prices. *Review of Financial Studies* 35, 1064–1103.
- Christoffersen, P.F., Mazzotta, S., 2005. The accuracy of density forecasts from foreign exchange options. *Journal of Financial Economics* 3, 578–605.
- Constantinides, G.M., Jackwerth, J.C., Perrakis, S., 2009. Mispricing of sp 500 index options. *The Review of Financial Studies* 22, 1247–1277.
- Corradi, V., Swanson, N., 2006. Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779–806.
- Cox, J.C., Ross, S., Rubinstein, M., 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7, 229–263.
- Daniel, K.D., Hirshleifer, D., 2015. Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives* 29, 61–88.
- Daniel, K.D., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. *Journal of Finance* 53, 1839–86.
- Datta, D.D., Londono, J.M., Ross, L.J., 2016. Generating options-implied probability densities to understand oil market events. *Energy Economics* 64, 440–457.
- Dawid, A.P., 1984. Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A* 147, 278–290.
- De Bondt, W.F., Thaler, R.H., 1995. Chapter 13 financial decision-making in markets and firms: A behavioral perspective, in: *Finance*. Elsevier. volume 9 of *Handbooks in Operations Research and Management Science*, pp. 385–410.
- DeCanio, S., 1979. Rational expectations and learning from experience. *Quarterly Journal of Economics* 93, 47–57.
- Dew-Becker, I., Giglio, S., Kelly, B., 2021. Hedging macroeconomic and financial uncertainty and volatility. *Journal of Financial Economics* 142, 23–45.

- Diebold, F., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–83.
- Hamilton, J., 2009. Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity* .
- Jacobs, K., Li, B., 2023. Option returns, risk premiums, and demand pressure in energy markets. *Journal of Banking and Finance* 146.
- Jiang, J.G., Tian, Y.S., 2005. The model-free implied volatility and its information content. *The Review of Financial Studies* 18, 1305–1342.
- Ju, N., Miao, J., 2012. Ambiguity, learning, and asset returns. *Econometrica* 80, 559–591.
- Kang, W., Rouwenhorst, G., Tang, K., 2020. A tale of two premiums: The role of hedgers and speculators in commodity futures markets. *Journal of Finance* 75, 377–417.
- Kitsul, Y., Wright, J.H., 2013. The economics of options-implied inflation probability density functions. *Journal of Financial Economics* 110, 696–711.
- Koijen, R.S., 2014. The Cross-Section of Managerial Ability, Incentives, and Risk Preferences. *Journal of Finance* 69, 1051–1098.
- Kolmogorov, A., 1933. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italiano degli Attuari* 4, 83–91.
- Longstaff, F., Schwartz, E.S., 2001. Valuing american options by simulation: A simple least-squares approach. *Review of Financial Studies* 14, 113–47.
- Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145–161.
- Melick, W.R., Thomas, C.P., 1997. Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis* 32, 91–115.
- Rampini, A.A., Sufi, A., Viswanathan, S., 2014. Dynamic riskmanagement. *Journal of Financial Economics* 111.
- Rouwenhorst, K.G., Tang, K., 2012. Commodity investing. *Annual Review of Financial Economics* 4, 447–467.

Shimko, D., 1993. Bounds of probability. *Risk* 6, 33–47.

Soderlind, P., Svensson, L., 1997. New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics* 40, 383–429.

Swanson, E., 2006. Have increases in federal reserve transparency improved private sector interest rate forecasts? *Journal of Money, Credit and Banking* 38, 791–819.