Predicting intraday crude oil returns with higher order risk-neutral moments

Patrick Wong*

January 26, 2022

Abstract

High frequency crude oil option data is used to extract the risk-neutral semi-moments from the crude oil market. These risk-neutral moments include the variance, third central moment and the recently developed tail risk variation measures. The returns of second and third semi-moments, and to a lesser extent the returns of the tail risk measures, are found to explain and predict returns in the crude oil and S&P 500 futures at high frequency.

JEL Classification: G12; G13; C58
Keywords: High frequency option data, higher risk-neutral moments, crude oil, prediction.

*UNSW Sydney, Business School, School of Risk and Actuarial Studies, Sydney, NSW, 2052, Australia. Email: p.a.wong@unsw.edu.au

Predicting intraday crude oil returns with higher order risk-neutral moments

*

January 26, 2022

Abstract

High frequency crude oil option data is used to extract the risk-neutral semi-moments from the crude oil market. These risk-neutral moments include the variance, third central moment and the recently developed tail risk variation measures. The returns of second and third semi-moments, and to a lesser extent the returns of the tail risk measures, are found to explain and predict returns in the crude oil and S&P 500 futures at high frequency.

JEL Classification: G12; G13; C58
Keywords: High frequency option data, higher risk-neutral moments, crude oil, prediction.

*

1 Introduction

Crude oil is the world's largest and most actively traded commodity. Crude oil has both a deep spot market and an exchange available to trade derivatives on them. The derivatives on crude oil are available to be traded 23-hours a day, 6-days a week on the New York Mechantile Exchange (NYMEX), and accounted for over 50% of the total volume traded in energy contracts in 2015 (Kyriakou et al., 2016). Further to this, crude oil prices are commonly monitored by multiple stakeholders, including consumers and producers of oil, investors, and policy makers. As such, it is necessary to have a deep understanding behind the underlying dynamics in the crude oil market.

The underlying dynamics of crude oil can be explained through the use of risk-neutral moments. These risk-neutral moments can be extracted through options, and have been extensively studied in the literature. For example, the ubiquitous Chicago Board of Exchange (CBOE) Volatility Index (VIX) is based on the risk-neutral second moment that is extracted from options, and is directly used by many as a measure of volatility in the market. Carr and Wu (2009) use options at a daily frequency to study the variance risk premium in the U.S. equities markets through the means of a variance swap contract, which is the difference between the risk-neutral second moment and the realised variance. Building on their methodology, Kozhan et al. (2013), also use options at a daily frequency, extend the analysis to include the third moment, and find that the second and third risk-neutral moments can both explain the S&P 500 excess returns. Similar studies on the risk-neutral moments extracted from option prices have also been conducted in the crude oil market. For example, Prokopczuk et al. (2017) study the variance risk present in the commodity markets and Ruan and Zhang (2018) look at the second and third risk-neutral moments in the crude oil market.

An area coming more recently into focus involves the tails of the return distribution. There is increasing evidence that the majority of the predictability in the variance risk premium, which involves the risk-neutral second order moment, is actually from the tails of the variance, in particular the left tail. This is demonstrated in Bollerslev et al. (2015) for the S&P 500 and Andersen et al. (2021) for the Nikkei 225. In addition, Andersen et al. (2015b) and Andersen et al. (2020), using daily option data and parametric models that explicitly account for the tail risk component, find that when the variance risk premium is stripped of the left tail variation, it has insignificant forecasting power on the U.S. and European indices. Beckmeyer et al. (2019) use high frequency S&P 500 option data to show that the tail variation, extracted from S&P 500 options, also predicts returns in the period just before the Federal Open Market Committee (FOMC) announcements. A remarkable aspect of the work by Beckmeyer et al. (2019) is their use of high frequency options data.

Research on high frequency option data is an area that is relatively (to the daily frequency) unexplored. There are a couple of reasons for this. For example, high frequency, intraday stock returns are subject to considerably more noise than is typically found at a lower daily frequency. Adding to this, options have two further dimensions, namely the strike price and time-to-maturity. Thus, moving even from a single asset to the options written on that asset increases the dimensionality of the problem significantly. In saying that, the use of high frequency option data is not a foreign concept in the literature. For example, Andersen et al. (2015a) utilise tick-by-tick S&P 500 option data to construct a VIX measure that is more robust to idiosyncratic changes arising from the time variant strike range. Griffin and Shams (2018) show using high frequency S&P 500 option data that the VIX is susceptible to manipulation.

Finally, the aforementioned work of Beckmeyer et al. (2019) highlights strong intraday option movements just prior to the FOMC announcements. In all of these cases, the insights gained from these works would not be possible without the use of high frequency option data. High frequency crude oil option data is an area that has not been studied. By using the available high frequency crude options we aim to glean any insights that are not present at lower frequencies.

Our work contributes to the literature in two different ways. Firstly, we estimate the risk-neutral higher order moments at high frequency. Thanks to the dataset we employ, we are able to decompose the higher order moments into what are commonly called semi-moments (semivariance and semi-skew). By decomposing the higher moments into the semi-moments, we can quantify the contribution to the overall higher order moment from either the 'left' component, the put options, or the 'right' component, the call options. Our dataset even allows us to quantify how the tail measures evolve at the higher frequencies, which has not been done before. Secondly, we explore the explanatory and predictive power of the semi-moments in *explaining* the high frequency log returns of the crude oil and S&P 500 futures. We are, to the best of our knowledge, the first to try and explain high frequency crude oil returns using high frequency higher order semi-moments. Our results indicate the higher order semi-moments extracted from the crude oil options yield reasonable explanatory and predictive power for both the crude oil and S&P 500 futures returns.

The remainder of this article is presented as follows. In Section 2 we present the methodology behind deriving our higher order risk-neutral moments, which include the risk-neutral variance, third moment and tail variations. In Section 3 we describe the dataset used for the analysis and provide a brief description on the risk-neutral moments derived for our dataset. Section 4 studies the explanatory and predictive power of the semi-moments on the crude oil and S&P 500 futures returns through the use of the risk measures. We conclude in Section 5.

2 Methodology

2.1 Variance

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space with the filtration $\mathbb{F} = \mathcal{F}_{t\in[0,T]}$ on which we will define any price process and $\mathbb{P}$ the historical probability measure. Assume now that the futures price f_t evolves according to the following jump diffusion

$$\frac{df_t}{f_t} = \mu_t dt + \sqrt{V_t} dW_t + \int_{\mathbb{R}} (e^x - 1)\tilde{\mu}_J^{\mathbb{P}}(dx, dt), \tag{1}$$

where W_t is a Wiener process under $\mathbb{P}$, the drift μ_t and instantaneous variance V_t are assumed to have cádlág paths, but are left unspecified otherwise, x is the size of the jump in the log-price, $\tilde{\mu}_J^{\mathbb{P}}(dx, dt) \equiv \mu(dx, dt) - \nu_t^{\mathbb{P}}(dx)dt$ is a martingale measure under $\mathbb{P}$, where $\mu(dx, dt)$ is a counting measure for jumps in f_t and $\nu_t^{\mathbb{P}}(dx)dt$ is its corresponding jump compensator.

The variability of the price over the period $[t, t+\tau]$ is measured by the quadratic variation which is defined as

$$QV_{t,\tau} = \int_t^{t+\tau} V_s\, ds + \int_t^{t+\tau} \int_{\mathbb{R}} x^2\, \mu(dx, ds). \tag{2}$$

The quantity Eq.(2) is in essence the variance in the price which is contributed by both the diffusive movements in the futures price and the jumps in the futures price. One way to estimate future expected variance is to construct a financial instrument that has payout Eq.(2) with no intermediary cash flows. If we are able to do so, then the expected variance of f_t over the period $[t, t+\tau]$ would be in theory $\mathbb{E}_t^{\mathbb{Q}}[QV_{t,\tau}]$, where $\mathbb{Q}$ is the equivalent risk-neutral measure.

Annualising this security we have

$$var_{t,\tau} = \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} V_s\,ds + \int_t^{t+\tau}\int_{\mathbb{R}} x^2\,\mu(dx,ds)\right]. \tag{3}$$

It is well understood in the literature that a trading strategy with payoff Eq.(2) can be approximately replicated with a portfolio of European call and put options and risk-free interest bearing instruments. Assume there exist European call / put options with time to maturity τ and strike price K with time t price denoted respectively by $C_{t,\tau}(K)$ / $P_{t,\tau}(K)$, and that these options are written on the futures contract f_t which expires after the expiration of the options. Further assume that these options trade on a continuum of strike prices. Then

$$var_{t,\tau} \approx \frac{2e^{r_{t,\tau}}}{\tau}\left[\int_0^{F_{t,\tau}} \frac{P_{t,\tau}(K)}{K^2}dK + \int_{F_{t,\tau}}^{\infty}\frac{C_{t,\tau}(K)}{K^2}dK\right] = \frac{2e^{r_{t,\tau}\tau}}{\tau}\int_0^{\infty}\frac{M_t(\tau,K)}{K^2}dK, \tag{4}$$

where $r_{t,\tau}$ is the risk free interest over $[t, t+\tau]$, $M_{t,\tau}(K) = \min\{C_{t,\tau}(K), P_{t,\tau}(K)\}$, and $F_{t,\tau}$ is the time t forward price of the underlying futures contract with maturity $t+\tau$.[1] A call (put) option is labelled out-of-the-money (OTM) for a specific strike K if $K > F_{t,\tau}$ $(K < F_{t,\tau})$, at-the-money when $K = F_{t,\tau}$, which is when $C_{t,\tau}(F) = P_{t,\tau}(F)$, and in-the-money (ITM) if $K < F_{t,\tau}$ $(K > F_{t,\tau})$. Since the forward price $F_{t,\tau}$ can be implied from the put-call parity then Eq.(4) in turn only requires OTM options to be computed. Going forward, we set

$$var^L_{t,\tau} = \frac{2e^{r_{t,\tau}}}{\tau}\int_0^{F_{t,\tau}}\frac{P_{t,\tau}(K)}{K^2}dK, \qquad var^R_{t,\tau} = \frac{2e^{r_{t,\tau}}}{\tau}\int_{F_{t,\tau}}^{\infty}\frac{C_{t,\tau}(K)}{K^2}dK, \tag{5}$$

to represent the 'left' and 'right' semi-moment for the second order risk-neutral moment.[2] An

[1] We acknowledge there is an approximation error in Eq.(4). As noted in Andersen et al. (2021) the r.h.s. of Eq.(4) is equal to $\frac{1}{\tau}\int_t^{t+\tau}\mathbb{E}_t^{\mathbb{Q}}[V_s]\,ds + \frac{2}{\tau}\int_t^{t+\tau}(e^x - 1 - x)\mathbb{E}_t^{\mathbb{Q}}[\nu_s^{\mathbb{Q}}(dx)]$, which is only equal to $\frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}[QV_{t,\tau}]$ up to a third order term (in an expansion around zero). However, as outlined in Carr and Wu (2009), Eq.(4) already provides a very good approximation to the future return variation even when jumps are present, and thus we ignore the approximation error going forward.

[2] The 'left' and 'right' are used here in relation to the forward price F. The 'left' semi-moment are the options with strike prices left of the forward price F (the put options), and the 'right' semi-moment are the options with strike prices right of the forward price F (the call options).

important note to recognise is that since

$$var_{t,\tau} = var^L_{t,\tau} + var^R_{t,\tau}, \tag{6}$$

then the aggregated risk-neutral second moment $var_{t,\tau}$ could be thought of as a portfolio that is comprised of two components; one component is long put options $var^L_{t,\tau}$, and the other is long call options $var^R_{t,\tau}$, and thus the return of $var_{t,\tau}$ in Eq.(6) is the sum of the returns of $var^L_{t,\tau}$ and $var^R_{t,\tau}$.

One of the key problems with Eq.(4) is the requirement of having accurate pricing information for the options on a continuum of strike prices. This assumption is obviously not met as options do not trade on a continuum of strikes, but rather on a finite set of actively traded strike prices. This requires the necessary truncation of the strike price range in the calculation of Eq.(4) and leads naturally to the approximation

$$\widehat{var}_{t,\tau} = \frac{2e^{r_{t,\tau}}}{\tau}\sum_{j=1}^{N}\frac{\Delta K_{t,j}}{K^2_{t,j}}M_{t,\tau}(K_{t,j}) \tag{7}$$

$$= \frac{2e^{r_{t,\tau}}}{\tau}\sum_{j=1}^{F}\frac{\Delta K_{t,j}}{K^2_{t,j}}P_{t,\tau}(K_{t,j}) + \frac{2e^{r_{t,\tau}}}{\tau}\sum_{j=F+1}^{N}\frac{\Delta K_{t,j}}{K^2_{t,j}}C_{t,\tau}(K_{t,j}) \tag{8}$$

$$= \widehat{var}^L_{t,\tau} + \widehat{var}^R_{t,\tau}. \tag{9}$$

Here $K_{t,j}$ are the strike prices where we have available information, and $K_{t,F}$ is the first available strike price less than the forward price $F_{t,\tau}$, i.e., $0 < K_{t,1} < ... < K_{t,F} < F_{t,\tau} < K_{t,F+1} < ... < K_{t,N}$ and $\Delta K_{t,j} = (K_{t,j+1} - K_{t,j-1})/2$.[3]

The CBOE VIX for the S&P 500 equity index is defined as the square root of the expected variation (in percentage) over the fixed horizon of 30 calendar days $\tau_M = \frac{30}{365}$. The VIX methodology

[3]At the boundaries, in-line with the CBOE definition we define $\Delta K_{t,1} = K_{t,2} - K_{t,1}$ and $\Delta K_{t,N} = K_{t,N} - K_{t,N-1}$.

uses two options sets with time to maturities τ_1, τ_2 which are qualified as the near-term and far-term, except those with less than 7 calendar days to expiry such that $\frac{7}{365} < \tau_1 < \tau_M = \frac{30}{365} < \tau_2$ and linearly combines the variances $\widehat{var}_{t,\tau_1}$ and $\widehat{var}_{t,\tau_2}$ to obtain an estimate for the expected variance at τ_M. Formally, the CBOE defines the VIX as

$$\mathrm{VIX}_t = 100\sqrt{\left[\tau_1 \widehat{var}_{t,\tau_1} w_1 + \tau_2 \widehat{var}_{t,\tau_2} w_2\right] \frac{1}{\tau_M}}, \tag{10}$$

where $w_1 = \frac{\tau_2 - \tau_M}{\tau_2 - \tau_1}$ and $w_2 = \frac{\tau_M - \tau_1}{\tau_2 - \tau_1}$ such that $w_1 + w_2 = 1$. Going forward, similar to the VIX methodology, using an annualised 30-day measure for var_{t,τ_M}, var^L_{t,τ_M} and var^R_{t,τ_M} , we set

$$\widehat{var}_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{var}_{t,\tau_1} w_1 + \tau_2 \widehat{var}_{t,\tau_2} w_2\right], \tag{11}$$

$$\widehat{var}^L_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{var}^L_{t,\tau_1} w_1 + \tau_2 \widehat{var}^L_{t,\tau_2} w_2\right], \tag{12}$$

$$\widehat{var}^R_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{var}^R_{t,\tau_1} w_1 + \tau_2 \widehat{var}^R_{t,\tau_2} w_2\right]. \tag{13}$$

The quantities var^L_{t,τ_M} and var^R_{t,τ_M} are the contribution to the expected 30-day variance from the put and call options respectively.

As an aside, from the necessary nature of truncating the integral in Eq.(4), prior work has demonstrated this introduces idiosyncratic changes in the VIX due to the sudden inclusion of OTM options at the boundaries that were previously truncated from either inactivity, or due to a zero bid price. As highlighted in Andersen et al. (2015a), this can create non-trivial short-lived deviations that can appear as a 'jump' in volatility, but which actually arise from the inclusion of previously dormant options in the VIX calculation. Andersen et al. (2015a) suggest instead to use a temporally and economically coherent index by using a truncation method that is not so sensitive to these boundary options, but rather one that captures the main information of the VIX and truncates these liquidity-sensitive options. However the truncation method used in Andersen et al. (2015a) is a bit troublesome as it requires both the OTM and ITM options to be actively traded. In our case the ITM options were not liquid enough to enable analysis using this method. Furthermore, the problem of the inclusion with the previously dormant options is partly mitigated with a flexible enough choice of the rolling window in the 'previous tick' method we describe later. Finally, the crude oil options that we use in our dataset are more liquid than the S&P 500 options used in Andersen et al. (2015a). By not actively truncating the tail options it allows us to explore the tail distribution which has been coming more into focus in the recent literature. As such, going forward we knowingly use these tail options, even though they can create idiosyncratic changes in the second moment, since we intend to exploit these tail options explicitly later on.

2.2 Third centralised moment

Higher risk-neutral moments of the future's return distribution have also been explored previously in the literature. Kozhan et al. (2013) exploit the risk-neutral third moment to study the risk-neutral skew of the S&P 500 market. They demonstrate that the following portfolio of OTM European call and put options locally approximates the third central moment of returns:[4]

$$\kappa_{t,\tau} = \frac{6e^{r_{t,\tau}}}{\tau}\left[\int_{F_{t,\tau}}^{\infty} \frac{K - F_{t,\tau}}{K^2 F_{t,\tau}} C_{t,\tau}(K)dK - \int_0^{F_{t,\tau}} \frac{F_{t,\tau} - K}{K^2 F_{t,\tau}} P_{t,\tau}(K)dK\right]. \tag{14}$$

We will also denote the following left and right side of the third moment

$$\kappa^R_{t,\tau} = \frac{6e^{r_{t,\tau}}}{\tau}\int_{F_{t,\tau}}^{\infty} \frac{K - F_{t,\tau}}{K^2 F_{t,\tau}} C_{t,\tau}(K)dK, \quad \kappa^L_{t,\tau} = \frac{6e^{r_{t,\tau}}}{\tau}\int_0^{F_{t,\tau}} \frac{F_{t,\tau} - K}{K^2 F_{t,\tau}} P_{t,\tau}(K)dK, \tag{15}$$

which are positive by construction, so that

$$\kappa_{t,\tau} = \kappa^R_{t,\tau} - \kappa^L_{t,\tau}. \tag{16}$$

[4] Note that the second moment defined in Eq.(4) and third moment in Eq.(14) differ from those defined in Bakshi et al. (2003). Section 1.3 of Kozhan et al. (2013) explains the advantages of these choices.

Hence, using the similar arguments that were used for $var_{t,\tau}$, $\kappa_{t,\tau}$ could be thought of as a portfolio that is comprised of two components; one component that is long call options $\kappa^R_{t,\tau}$ and the other short put options $\kappa^L_{t,\tau}$. Then the return of $\kappa_{t,\tau}$ is given by the difference of the returns from $\kappa^R_{t,\tau}$ and $\kappa^L_{t,\tau}$.

Similar to the variance, the integral in Eq.(14) requires truncation. This leads us to the following estimators for the third central moment[5]

$$\widehat{\kappa}_{t,\tau} = \frac{6e^{r_{t,\tau}}}{\tau}\sum_{j=f+1}^{N}\Delta K_{t,j}\frac{K_{t,j}-F_{t,\tau}}{K^2_{t,j}F_{t,\tau}}C_{t,\tau}(K_{t,j}) - \frac{6e^{r_{t,\tau}}}{\tau}\sum_{j=1}^{f}\Delta K_j\frac{F_{t,\tau}-K_{t,j}}{K^2_{t,j}F_{t,\tau}}P_{t,\tau}(K_{t,j}) \tag{17}$$

$$= \widehat{\kappa}^R_{t,\tau} - \widehat{\kappa}^L_{t,\tau}. \tag{18}$$

As an aside, if we normalise the third moment $\kappa_{t,\tau}$ by $var^{3/2}_{t,\tau}$, then we have the implied skewness measure $skew_{t,\tau}$ defined as

$$skew_{t,\tau} = \frac{\kappa_{t,\tau}}{var^{3/2}_{t,\tau}}, \tag{19}$$

where the relevant skewness estimator of Kozhan et al. (2013) would be given by

$$\widehat{skew}_{t,\tau} = \frac{\widehat{\kappa}^R_{t,\tau}}{\widehat{var}^{3/2}_{t,\tau}} - \frac{\widehat{\kappa}^L_{t,\tau}}{\widehat{var}^{3/2}_{t,\tau}}. \tag{20}$$

In this article we only study the third moment. Using a similar approach to the CBOE's VIX methodology, we look at the third central moment of returns along with their 'left' and 'right' components over thirty calendar days, (i.e., $\kappa_{t,\tau_M}, \kappa^L_{t,\tau_M}$ and κ^R_{t,τ_M}) and do this through linear

[5]For both the second moment Eq.(4) and the third moment Eq.(14), the integrals are discretised and no interpolation between the discrete market quotes are used. There is an abundant literature dealing with alternative approximations of those integrals, see e.g., Byun and Kim (2016) and Hollstein and Prokopczuk (2016). However, section A.2.3 of Kozhan et al. (2013) shows that more refined approximations of those integrals, in particular using splines, do not significantly change the results.

interpolation of the near and far terms of the option maturities. As such we set our estimators as

$$\widehat{\kappa}_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{\kappa}_{t,\tau_1} w_1 + \tau_2 \widehat{\kappa}_{t,\tau_2} w_2\right], \tag{21}$$

$$\widehat{\kappa}^L_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{\kappa}^L_{t,\tau_1} w_1 + \tau_2 \widehat{\kappa}^L_{t,\tau_2} w_2\right], \tag{22}$$

$$\widehat{\kappa}^R_{t,\tau_M} = \frac{1}{\tau_M}\left[\tau_1 \widehat{\kappa}^R_{t,\tau_1} w_1 + \tau_2 \widehat{\kappa}^R_{t,\tau_2} w_2\right], \tag{23}$$

where again $w_1 = \frac{\tau_2 - \tau_M}{\tau_2 - \tau_1}$ and $w_2 = \frac{\tau_M - \tau_1}{\tau_2 - \tau_1}$ such that $w_1 + w_2 = 1$.

2.3 Tail risk

An increasingly important area of study is the risk of extreme tail movements. This area has recently been put under focus as the tail risk appears to have strong predictability power, and perhaps even consumes most of the predictability power that is available from the variance risk premium. We aim to utilise these tail risk measures for the crude oil market, and we are, to the best of our knowledge, the first to study the tail risk variations in the crude oil markets.

The following subsection definitions and results are based on the work of Bollerslev et al. (2015). We define the left and right risk-neutral jump variation over the period $[t, t+\tau]$ by

$$LJV^{\mathbb{Q}}_{t,\tau} = \int_t^{t+\tau} \int_{x<-q_t} x^2 \nu^{\mathbb{Q}}_u(dx)du, \qquad RJV^{\mathbb{Q}}_{t,\tau} = \int_t^{t+\tau} \int_{x>q_t} x^2 \nu^{\mathbb{Q}}_u(dx)du, \tag{24}$$

where $q_t > 0$ is a chosen time-varying cutoff for the log-jump size. We define the risk measures LJV and RJV as the expectation of the quantities in Eq.(24)

$$LJV_{t,\tau} = \frac{1}{\tau}\mathbb{E}^{\mathbb{Q}}[LJV^{\mathbb{Q}}_{t,\tau}], \qquad RJV_{t,\tau} = \frac{1}{\tau}\mathbb{E}^{\mathbb{Q}}[RJV^{\mathbb{Q}}_{t,\tau}]. \tag{25}$$

The tail variations LJV and RJV represent the contribution to the variance var that is only from extreme (tail) price jump risk. As shown in Bollerslev and Todorov (2014), we can obtain estimators for $\mathbb{E}_t^{\mathbb{Q}}[LJV_{t,\tau}^{\mathbb{Q}}]$ and $\mathbb{E}_t^{\mathbb{Q}}[RJV_{t,\tau}^{\mathbb{Q}}]$ solely from option data. Furthermore, taking the difference between LJV and RJV leads us to

$$LJV_{t,\tau} - RJV_{t,\tau} = \frac{1}{\tau}\left(\mathbb{E}_t^{\mathbb{Q}}[LJV_{t,\tau}^{\mathbb{Q}}] - \mathbb{E}_t^{\mathbb{Q}}[RJV_{t,\tau}^{\mathbb{Q}}]\right). \tag{26}$$

This difference represents the excess risk investors place that is attributed to the possibility of abrupt negative price movements over the possibility of abrupt positive price movements. It can be interpreted as the pricing of negative tail events, or as argued in Bollerslev et al. (2015) as a proxy of fear in the markets. Similar to before, the difference could be thought of as a portfolio comrpised of the two components that are long in $LJV_{t,\tau}$ and short in $RJV_{t,\tau}$, and hence the returns of $LJV_{t,\tau} - RJV_{t,\tau}$ in Eq.(26) is given by the difference in returns from $LJV_{t,\tau}$ and $RJV_{t,\tau}$.

The estimation of $\mathbb{E}^{\mathbb{Q}}[LJV_{t,\tau}^{\mathbb{Q}}]$ and $\mathbb{E}^{\mathbb{Q}}[RJV_{t,\tau}^{\mathbb{Q}}]$ is done following Bollerslev and Todorov (2014). They impose a general specification on the (extreme) risk-neutral jump intensity process, from which they are able to estimate the $\mathbb{Q}$ jump tail measures based on this specification. Suppose that the extreme jumps follow the specification

$$\nu_t^{\mathbb{Q}}(dx) = \left(\phi_t^{+} \times e^{-\alpha_t^{+} x}\mathbb{1}_{\{x>0\}} + \phi_t^{-} \times e^{-\alpha_t^{-}|x|}\mathbb{1}_{\{x<0\}}\right). \tag{27}$$

This specification allows for the left ($^{-}$) and right ($^{+}$) jump tails to differ, and is very flexible in that it allows for time-varying level shifts and shape governed by the parameters $\phi_t^{\pm}$ and $\alpha_t^{\pm}$ respectively, compared to the usual parametric option pricing model which typically fixes the shape of the jump.

Now let $O_{t,\tau}(k)$ be the price of an OTM option on the futures f_t, with log-moneyness[6] k and time to maturity τ. With specification Eq.(27), and using the results derived in Bollerslev and

[6] Log-moneyness for an option with strike price K and forward price F is defined as $k = \ln(K/F)$.

Todorov (2011) and Bollerslev and Todorov (2014), we have the following approximations for short dated options

$$O_{t,\tau}(k) \approx \begin{cases} \tau e^{-r_{t,\tau}} F_{t,\tau} \phi_t^- \frac{e^{k(1+\alpha_t^-)}}{\alpha_t^-(\alpha_t^-+1)}, & k < 0, \\ \tau e^{-r_{t,\tau}} F_{t,\tau} \phi_t^+ \frac{e^{k(1-\alpha_t^+)}}{\alpha_t^+(\alpha_t^+-1)}, & k > 0. \end{cases} \tag{28}$$

Utilising the approximations in Eq.(28), we obtain estimates of $\alpha_t^{\pm}$ and $\phi_t^{\pm}$ through the following optimisation problems

$$\widehat{\alpha}_t^{\pm} = \underset{\alpha^{\pm}}{\arg\min} \frac{1}{N_t^{\pm}} \sum_{j=1}^{N_t^{\pm}} \left| \ln\left(\frac{O_{t,\tau}(k_{t,j})}{O_{t,\tau}(k_{t,j-1})}\right)(k_{t,j} - k_{t,j-1})^{-1} - \left(1 \pm (-\alpha^{\pm})\right) \right|, \tag{29}$$

$$\widehat{\phi}_t^{\pm} = \underset{\phi^{\pm}}{\arg\min} \frac{1}{N_t^{\pm}} \sum_{j=1}^{N_t^{\pm}} \left| \ln\left(\frac{e^{r_{t,\tau}} O_{t,\tau}(k_{t,j})}{\tau F_{t,\tau}}\right) - \left(1 - \mp \widehat{\alpha}_t^{\pm}\right) k_{t,j} + \ln\left(\widehat{\alpha}_t^{\pm} \mp 1\right) + \ln\left(\widehat{\alpha}_t^{\pm}\right) - \ln(\phi^{\pm}) \right|, \tag{30}$$

where $N_t^{\pm}$ is the total number of calls (puts) used in the estimation with log-moneyness $0 < k_{t,1} < \cdots k_{t,N_t^+}$ $(0 < -k_{t,1} < \cdots < -k_{t,N_t^-})$.

Then we have

$$\mathbb{E}_t^{\mathbb{Q}}[LJV_{t,\tau}^{\mathbb{Q}}] = \tau \phi_t^- e^{-\alpha_t^-|q_t|}(\alpha_t^- q_t(\alpha_t^- q_t + 2) + 2)/(\alpha_t^-)^3, \tag{31}$$

$$\mathbb{E}_t^{\mathbb{Q}}[RJV_{t,\tau}^{\mathbb{Q}}] = \tau \phi_t^+ e^{-\alpha_t^+|q_t|}(\alpha_t^+ q_t(\alpha_t^+ q_t + 2) + 2)/(\alpha_t^+)^3. \tag{32}$$

For the estimation procedure of the tail parameters $\alpha_t^{\pm}$ and $\phi_t^{\pm}$ we rely on the use of deep OTM options. This is done through filtering out options that are relatively close to ATM options. Following Bollerslev et al. (2015) and Andersen et al. (2021) we only use put options with log-moneyness less than -2.5 times the maturity-normalised ATM Black-Scholes implied volatility (BSIV) for the left-tail parameters and call options with log-moneyness in excess of the maturity-normalised BSIV. The choice for a more lenient cut-off for the call options arises from the fact that the call options in the S&P 500 options are far less liquid in comparison to their put option counterparts. Although this is not as necessary for the crude oil options due to being similar levels of liquidity for the call and put options, we do this so our methodology falls in line with the previous works of Bollerslev et al. (2015) and Andersen et al. (2021).[7]

To minimise the effect of microstructure noise in our estimates of $\alpha_t^{\pm}$ and $\phi_t^{\pm}$, we only allow $\alpha_t^{\pm}$ to change daily and for $\phi_t^{\pm}$ to change only every five minutes. Previous works of the authors Bollerslev et al. (2015) and Andersen et al. (2021) work with option data at a daily frequency where they only allow $\alpha_t^{\pm}$ to change weekly and $\phi_t^{\pm}$ to change daily.[8] By utilising data available at higher frequencies we can allow the tail risk parameters to change more frequently than would otherwise be possible at lower frequencies.

We mention as an aside that the tail risk measures can be easily worked under the framework of risk premiums. Bollerslev et al. (2015) and Andersen et al. (2021) work explicitly with risk premiums instead of only the risk neutral tail variation measures. One key caveat to mention is they do this through the assumption that the realised jump tail risk is symmetric, which in turn leads to the difference cancelling out the historical measure component. This by consequence means their risk premium is equivalent to the difference in Eq.(26). Since we do not use risk premiums in this article, and further to this, use an assumption that essentially cancels out the realised component, we do not delve into the historical component.

[7] We did find however having a symmetric cut-off for the call options did not materially change our results.

[8] It is important to note that typical parametric models in the literature, including the affine jump diffusion models of Duffie et al. (2000), impose a constant tail shape parameter $\alpha_t^+ = \alpha_t^- = \alpha$ and that the scale parameters describe both tails identically $\phi_t^+ = \phi_t^-$. Allowing the parameters to change even weekly affords us a very flexible model even when compared to the most advanced parametric models.

3 Data

Our dataset is comprised of tick-by-tick quotes of the monthly West Texas Intermediate (WTI) crude oil options that trade on the New York Mercantile Exchange (NYMEX), their underlying monthly WTI crude oil futures, and the S&P 500 futures. The sample covers the period of 1 January 2016 to 31 December 2019, and we restrict the hours of our analysis to 09:30 - 16:00 Eastern Time (ET) Monday to Friday, the primary trading hours of the New York Stock Exchange. For the futures, we are in particular interested in the shortest term maturity available with more than seven calendar days to maturity. Figure 1 shows the evolution of the price of the lead-term futures of WTI and S&P 500 index. The four years in our sample contain the crude oil 2014-2016 supply glut noted by the sharp drop in the crude oil price at the beginning of our sample and the 2018-2019 drop which was spurred by fears of an escalating trade war between the United States and China.

The WTI options are American style and are written on a WTI futures contract that is deliverable in the next calendar month after expiration of the crude oil option. While the theory developed in the previous section is derived with European style options, we instead work with American options on crude oil for the following reasons. First, the American options listed on the NYMEX are by far the most liquid and actively traded crude oil option. To illustrate the difference in liquidity between the American and European options, for our data set there was an average daily volume of 74,920 American style contracts traded with an average daily open interest of 680,670 contracts per day. In comparison, the European style options (which are also listed on the NYMEX) only had an average daily volume of 258 contracts traded and an average daily open interest of 12,760 contracts. Since the underlying contract sizes are the same for both the American and European options, the American style options would embed far more timely and material information regarding the crude oil prices while their European counterpart's quotes would be stale. Second, while it is true that American style options have a pricing premium when compared to their European style counterparts, this premium arises solely from their early exercise rights. Since we only use OTM options for all the aforementioned metrics, then the benefit of the early exercise premium will be at a minimum.[9] Furthermore, while there are crude oil options that expire weekly traded on the NYMEX which could then be used to supplement our analysis, we also found them to have insufficient liquidity to work with at high frequency.

[Insert Figure 1 here.]

We use all available tick-by-tick quotes except for those with a time-to-maturity of less than 7 days and days that have insufficient quote activity. We end up with 985 trading days in our sample after discarding 38 trading days due to insufficient quotes on those days. For our dataset, we have an extensive amount of quotes to work with. On average there are 4.19 (4.10) million OTM put (call) option quotes per day, and we have in total 4.12 (4.04) billion OTM put (call) option quotes for our entire sample.[10]

Quote activity varies significantly within the trading day and across moneyness. To highlight quote activity within the trading day, Figure 2 plots the percentage of the total daily OTM

[9] It is customary in the literature to convert American options prices to European prices by using the Barone-Adesi and Whaley (1987) approximation. It would be of interest to use it and assess the impact of using directly American options in place of the converted European options. We leave this very computer intensive question for future work. We would like to thank the anonymous referee for bringing this to our attention.

[10] To give an insight into how actively quoted these options are, the equivalent period for the S&P 500 options on the CBOE has an average of 0.90 (0.41) million OTM put (call) option quotes per day and in total 876 (406) million OTM put (call) option quotes for the entire sample.

option quotes recorded every five minutes for our dataset. Quote activity is slightly elevated at the open, with approximately around 2% of the daily quotes arriving every five minutes until 11:30, from which there is a lull in activity until approximately 14:00 in the afternoon. Quote activity spikes at 14:30, which is when Trade at Settlement (TAS) finishes for the day in the crude oil futures.[11] During the five minutes of 14:25 - 14:30 approximately 3.28% of the daily quotes are recorded, more than double the daily average of 1.28% being recorded every five minutes. After TAS, the average quote activity is very subdued where every five minutes only approximately 0.44% of the daily quotes arrive, close to one-third of the daily average. Figure 3 provides insight into quote activity across the options moneyness. Here quote activity is highly concentrated around the ATM options for both the near term and far term options. For the near (far) term options 52.9% (44.4%) of the total OTM option quotes are from options with moneyness between (0.9, 1.1) and 79.3% (78.2%) of the quotes between the moneyness range (0.75, 1.25). The near term options quotes are more concentrated than their far term counterpart, likely arising from the lower likelihood of the OTM options becoming ITM as time-to-maturity is lower.

[Insert Figure 2 here.]

[Insert Figure 3 here.]

For our subsequent analysis we begin by splitting the option and futures quotes into 15 second intervals using the 'previous tick method', similar to the method employed in Andersen et al. (2015a). The last available quote is used if there is no arrival of a quote in the 15 second interval. Similar to Andersen et al. (2015a) we limit how far back we use the last available quote

[11] TAS is a mechanism employed by NYMEX that allows parties to the futures contract to execute the futures contract within a spread around the current daily settlement price.

to prevent staleness in our measures. We further employ some light data cleaning filters to remove any erroneous quotes and outliers. The details of the data cleaning procedure are listed in Appendix A. After applying the filters, we have for every 15 second interval an average of 30 (34) strikes of OTM put (call) quotes for the near term maturity and an average of 42 (52) strikes of OTM put (call) quotes for the far term maturity. Lastly, we aggregate our 15 second intervals into five minute intervals in order to limit the effect of microstructure noise. We then compute our various risk measures based on the five minute series.

The higher order risk-neutral moments and their respective semi-moments are displayed in Figures 4 and 5 respectively, and Table I displays summary statistics for them. The variance changes substantially over time, as evidenced by the elevated state in 2016 to then a subdued state in 2017 to late 2018, and the 75th percentile being 0.1574 but the maximum is 0.7586. Standard deviation is large as well, being 0.0891, while the mean is 0.1303. Globally, the third central moment κ is negative on average with a mean of -0.0079 and the 75th percentile being negative at -0.0018. The third central moment κ is also negatively skewed, and contains leptokurtic tails with a kurtosis of 8.1555. The difference in the left and right jump variations is positive on average with a mean of 0.0315 and the 25th percentile at 0.0140. The difference is positively skewed and similar to the third moment, and contains heavy tails with a kurtosis of 9.6055.

[Insert Figure 4 here.]

[Insert Figure 5 here.]

[Insert Table I here.]

The left and right semi-moments are very similar in shape. The left semi-moment is larger in scale than the corresponding right semi-moment for all the variables, indicating the left semi-moment is the main contributor to the aggregated higher order risk-neutral moments. This is more or less confirmed in Table I, where each of the left semi-moments have higher means, quantiles and standard deviations than their corresponding right semi-moments. The tail measure LJV mean of 0.0052 is larger than the RJV mean of 0.0026 by a factor of two. This is in contrast to the results found in Bollerslev et al. (2015) and Andersen et al. (2021) where they found respectively that for the S&P 500 and Japanese equity markets the left tail variation LJV is about ten times larger than the right tail variation RJV, which allows them to safely ignore the right tail for their analyses. Our results indicate that the right tail is as important as the left tail in the crude oil markets, and cannot be ignored. It also suggests the crude oil market is more concerned (relatively speaking) about tail price increases compared to the S&P 500 market, which can be potentially explained by the fact that consumers of crude oil care about price increases almost as much as price decreases, where the S&P 500 market tends to be more concerned with extreme price decreases than increases. It is also interesting that the kurtosis for the right semi-moment is larger than its left's counterpart, indicating the right semi-moment is more heavy tailed than the left semi-moment.

4 Regressions

In this section we aim to explain and predict the high frequency, intraday crude oil and S&P 500 futures returns through the use of our available high frequency risk-neutral semi-moments. The literature has demonstrated that the use of risk-neutral moments (through risk premiums) are able to partly explain and predict the excess returns of equity indexes at lower frequencies, such as daily and monthly (see for example, Carr and Wu (2009), Kozhan et al. (2013) and Kilic and Shaliastovich (2019)). These studies were conducted using daily data, which are then used to explain or predict the futures excess returns over a monthly horizon.

Here we adjust the methodology of the previous works in the literature in order to explain and predict high frequency futures returns. The use of the risk premium framework is untenable in our situation. The realised component requires the entire month of returns, which is unavailable when we are trying to explain the very same intraday returns. Since the risk neutral semi-moments are available to us at these high frequencies, we instead focus our efforts on extracting the information embedded in these moments. As a means to this end, we utilise the commonly used regression:

$$\frac{1}{h_2} Y^j_{t,t+h_2} = \gamma_0(h_1,h_2) + \gamma_1(h_1,h_2) D(h_1)^j_t + u^j_{t,t+h_2}, \quad t = 1, ..., T^j, \tag{33}$$

where j is either the crude oil or S&P 500 futures, t refers to the specific 5-minute period we are inspecting, h is the number of five-minute periods (i.e., $h_1 = 1$ means 5-minutes), $Y_{t,t+h} = \ln(f_{t+h}) - \ln(f_t)$ are the log returns of the relevant futures over the period $[t, t+h]$, $D_t(h_1)$ is the relevant predictor in our analysis, which may depend on the overlapping return period h_1 that overlaps with our prediction interval (i.e., $0 < h_1 \leq h_2$).[12] We restrict our regressions to intraday predictions to remove any potential overnight effects. To account for overlap, we rely on the robust Newey and West (1987) t-statistic with a lag of $2h_1$, consistent with the approach of Bollerslev et al. (2015) and Andersen et al. (2021). We first look at contemporaneous regressions by setting $h_1 = h_2$, and then subsequently predictive regressions by setting $h_1 < h_2$.

[12] The norm in the literature is to work with excess returns, which is the excess of the log return less the risk-free rate. As our horizon is over short periods (i.e., up to an hour) the risk-free return over such a small period is effectively zero. In either case, the log returns on the futures are equivalent to the excess returns on the spot (as the initial cost of the futures is zero).

An important remark, through setting $h_1 < h_2$ we enter into a predictive situation. This is due to the fact that one variable is known, but the other is not. Our situation is different to the norm where the predictive regression uses non-overlapping periods. Our results found that having non-overlapping periods leads to extremely poor results. However, it is also necessary to stress here when h_1 is much smaller than h_2 (which is what we do when we set $h_1 = 1$ (5 minutes) and $h_2 = 12$ (one hour)) then the challenge of predicting accurately still remains.

4.1 Contemporaneous regressions on crude oil futures

An important note to recognise regarding the studies from Carr and Wu (2009), Kozhan et al. (2013) and even Bollerslev et al. (2015) and Andersen et al. (2021) is that the authors use a *return* to explain and/or predict the returns of the futures. Carr and Wu (2009) and Kozhan et al. (2013) use precisely a return through the construction of synthetic swaps (see for example Eq.(9) from Carr and Wu (2009), Eq.(29)-(30) from Kozhan et al. (2013) and Eq.(15)-(16) from Kilic and Shaliastovich (2019) for further details) and the tail risk premiums are similar to a return in the sense that they 'long' the risk-neutral option portfolio and 'short' the realised component. Since our dataset affords us the ability to calculate the higher order risk-neutral moments at high frequencies, it is natural to consider explaining the returns of the high frequency futures with the *returns* on our high frequency risk-neutral moments.[13]

In what follows we conduct regressions based on the returns of the higher order risk-neutral

[13] We also considered using the *level* higher order risk neutral moments. The level higher order moments had no predictive power and the coefficients from the univariate regression were not statistically significant at the 10% level on a forecasting horizon of 15, 30 and 60 minutes. For brevity, we omit these results, but they are available upon request.

semi-moments. This is done primarily to isolate the effects from either the put or call options and to see whether the left or right semi-moments of the higher order moments are more important in explaining the futures returns at these higher frequencies. Moreover, by using the semi-moments we are able to utilise the deeper information available within them, over the aggregated risk-neutral moments. To demonstrate the benefit of the semi-moments, we also create a synthetic series of the aggregated higher order moment returns for comparison. Since the risk-neutral third central moment and difference between the left and right jump variations can be positive or negative, as seen in Figure 4, creating a series of log returns with them directly is not possible. As such, we create these returns by utilising the portfolio component argument laid forth in Section 2. We label these synthetic series of returns as var', κ', and $(LJV - RJV)'$, and they are created by appropriately summing or differencing the returns of the individual semi-moments (refer to Eq.(6), Eq.(16) and Eq.(26) respectively).

Tables II and III conduct contemporaneous regressions at 5 and 30 minutes on the crude oil futures returns respectively.[14] The results are quite compelling in that the higher order risk-neutral semi-moments are able to reasonably explain the crude oil futures returns at all the time intervals considered. Globally speaking, the semi-moments for var and third moment κ have more explanatory power over the futures returns than the tail risk variation measures. At both time intervals, the coefficients of the semi-moments are always negative and highly significant for all variables. The negative sign of the left and right semi-moments of the variance aligns closely with the leverage effect observed in the crude oil markets (Kang et al. (2020)). When the variance increases (decreases) then both the crude oil futures decrease (increase) on average.

[14]We also conducted these contemporaneous regressions at other time intervals of interest. The results at these time intervals are similar to the results presented, and hence we omit these regressions.

The negative signs on the jump variations align with the previous findings in Bollerslev et al. (2015) and Andersen et al. (2021) which means futures are priced higher when the perceived risk of a tail movement in either direction is lower. Furthermore, in all but one case (at 5-minutes for the third moment) the coefficients of the left semi-moment of the risk measure are larger (in absolute terms) then the corresponding right semi-moment of the risk measure, suggesting the left semi-moment carries more information than its corresponding right semi-moment.

[Insert Table II here.]

[Insert Table III here.]

What is also interesting is the difference in explanatory power between the synthetic aggregated series var', κ', and $(LJV - RJV)'$ and their semi-moments. In each case, the synthetic series did not perform as well as the equivalent semi-moment regression, clearly underlining the benefit of decomposing the aggregated risk-neutral moments into their semi-moments. The difference in performance is potentially attributed to the fact that by splitting the moments into their respective semi-moments, it allows the more prevalent information in the regression to be highlighted rather than being lost due to aggregation.

Through the utilisation of our high frequency dataset, we were able to show that the information embedded in high frequency options contains strong explanatory power over the futures returns. Through creating returns with the high frequency higher order risk-neutral moments, we demonstrated that these returns were able to partly explain the crude oil futures returns, and it does highlight that even at intraday frequencies, there is some sort of leverage effect present.[15]

[15]The leverage effect is the description given to the phenomena observed in the equity markets where volatility

4.2 Predictive regressions on crude oil futures

In this part we focus our efforts on *predicting* the crude oil futures returns. We set up the regressions so that we have predictive regressions through the setting of $h_1 < h_2$. Specifically, the following regressions are set up so that we vary $h_1 = 1, ..., 6$ between 5-minutes to 30-minutes, and we set $h_2 = 12$ to be one hour. Since the overlapping period does not cover the full interval (unlike the contemporaneous regressions) the regressions by construction are predictive. The following analysis only uses the higher order semi-moments. We conducted similar predictive regressions with the synthetic aggregated series var', κ' and $(LJV - RJV)'$, but we found that, similar to the contemporaneous regressions, the semi-moments perform uniformly better, and hence we omit them.

Table IV contains predictive regression for the hourly returns of the crude oil futures matched with the log returns of the semi-higher order risk neutral moments. The top panel contains the regression from the returns of the left and right jump variations from Eq.(31)-(32). The coefficients estimated are all negative with significance at the 1% level. The negative coefficient implies that when there is a decrease (increase) in the expected jump variation, crude oil returns are expected to increase (decrease). This result is consistent with the equity markets when studied at a daily frequency (see Bollerslev et al. (2015) and Andersen et al. (2021)) and is consistent with the leverage effect. The coefficient for the log returns on the LJV is larger (in

in a share tends to increase inversely to its share price with the common interpretation that a company's leverage increases as a proportion of the company's equity (see Black (1976) and Aït-Sahalia et al. (2013) for further details). Different economic interpretations of the leverage effect present in the commodity markets have been explained through different theories, and for more details we refer the interested reader to Ng and Pirrong (1994), Basak and Pavlova (2016), Chiarella et al. (2016), Baur and Dimpfl (2018), Kang et al. (2020) and their references within for further details. We would like to thank the anonymous referee for bringing this to our attention.

absolute terms) and is more significant than the log returns of the RJV. This indicates that, similar to the equity markets, the left tail of extreme returns are more significant than the right tails (i.e., market participants demand greater compensation for extreme drops in the price, but not as much for extreme increases). These results are quite impressive considering they are in line with the predictive regressions found in Andersen et al. (2021) when the forecasting period is 3-months or less, even though high frequency returns are much more noisy and difficult to predict than monthly returns. Further to this, the coefficients remain statistically significant at the 1% level across all periods considered here, where the coefficients in Andersen et al. (2021) are significant at the 5% level typically. As expected, predictive power decreases as h_1 decreases, but nevertheless, the coefficients estimated still remain highly significant at the 1% level even when the overlapping period is only 5-minutes.

[Insert Table IV here.]

The middle and bottom panel of Table IV contain predictive regressions for the hourly returns of the crude oil futures predicted by the semi-moments of the variance var and third moment κ respectively. The coefficients of the semi-moments for the variance and third moment are both highly significant and deeply negative for all periods considered. The predictive power of the variance is very impressive, the adjusted R^2 is 12.1% when h_1 is set to 30-minutes and retains an adjusted R^2 of 4% even when the overlapping period is only 10-minutes, clearly suggesting the returns of the variance process contain a significant amount of information in predicting the crude oil returns. These results are strong considering their adjusted R^2 are high, and are also in line with the results in Andersen et al. (2021), even in the presence of noisier high frequency data. Similar to the jump variations, the signs of the coefficients are negative, implying that the crude oil futures tend to increase (decrease) when the variance decreases (increases), echoing the notion that market participants invest when perceived risk is lower. The coefficient of the left variance semi-moment is larger (in absolute terms) then the coefficient of the right variance semi-moment, suggesting put options changes have more information embedded in them compared to their respective call options. The returns of the left and right third moment are also highly significant at the 1% level across all levels and are able to obtain an adjusted R^2 of 11.8% when h_1 is 30-minutes and retain an adjusted R^2 of 5.9% and 3.9% at 15 and 10 minutes respectively. Their coefficients are negative at all the time periods considered, but since the third moment is defined as the difference between the right and left semi-moment, global interpretation is not possible. What is interesting however is that for longer overlapping periods the left third moment is larger than its corresponding right semi-moment, but for shorter overlapping periods this is reversed and the right third moment is larger than its corresponding left semi-moment. We are not sure what causes this phenomena, but note it as a potential open question for the literature.

4.3 Cross-market analysis between crude oil and S&P 500

We now turn our attention to explaining and predicting the S&P 500 futures using high frequency *crude oil* risk-neutral semi-moments. There is increasing evidence that the commodity and equity markets are becoming more and more interlinked with the increasing financialisation of the commodity markets (Basak and Pavlova, 2016). Thus, an interesting question to observe is whether the information embedded in the crude oil options can be used to deepen our understanding of the S&P 500 market. Here we use the same regression structure in Eq.(33), but with the S&P 500 futures. We demonstrate that the crude oil options are able to partly explain and predict the S&P 500 futures, albeit to a lesser degree than the crude oil futures.

4.3.1 Contemporaneous regressions

Here we set out with the goal of explaining the high frequency S&P 500 futures returns using the returns of the high frequency crude oil semi-moments.[16] Similar to the tables presented in Section 4.1, Table V reports the results for the 60-minute contemporaneous regressions on the S&P 500 futures using the returns of the high frequency risk-neutral semi-moments and the synthetic aggregated series var', κ' and $(LJV - RJV)'$.[17] There are clear similarities between the dynamics of the S&P 500 futures returns and the crude oil futures returns. The results are impressive, with the adjusted R^2 being 6.1% and 6.2% for the semi-moments of the variance and third moment. Here, the coefficients on each semi-moment are negative, similar to the results from the previous section. Further to this, the coefficient of each left semi-moment is larger (in absolute terms) than the coefficient of the right semi-moment. This indeed indicates there is a systematic information differential present in the put options and call options in the crude oil options even when used in a cross-market setting. On top of this, each of the coefficients of the left semi-moments are significant at the 1% level, and so are the right semi-moments with the exception of κ^R that is significant at 10% only. Similar to the results found in the crude oil futures, the variance and third moment contain more explanatory power over the S&P 500 futures than the tail risk measures. Finally, even across different markets, the benefit of utilising the semi-moments over the regular higher order moments is apparent with the synthetic aggregated series failing to perform as well as the decomposed semi-moments.

[Insert Table V here.]

[16] We conducted similar regressions with the level risk measures. We found the level risk measures provided no predictability and all the coefficients were insignificant at the 10% level, and thus we do not report them.

[17] The regressions were conducted at other time intervals of interest, however the results are quite similar and are not reported here.

4.3.2 Predictive regressions

Finally, we also study the predictability of the S&P 500 futures returns using the returns of the high frequency semi-moments. Similar to the analysis in Section 4.2, we set $h_1 < h_2$, fix $h_2 = 12$, one hour and vary the overlapping period $h_1 = 1, ..., 6$ between 5-minutes to 30-minutes.

The results of our regressions on the returns of the semi-moments are outlined in Table VI, which contains the regressions on the tail variations, variance and third moment. The coefficients of the tail variation measures are significant at the 1% level, except when h_1 is 5-minutes which is significant at the 5% level. The signs are negative and the coefficient for the left jump variation is larger (in absolute terms) than the right jump variation, similar to the crude oil market.

[Insert Table VI here.]

The tail variations afford little to no predictability over the S&P 500 futures returns, but nevertheless, the results are still respectable given the cross-market nature, as they are significant at the 1% level at all overlapping periods (except at 5-minutes) and are more or less in line with the results found in Andersen et al. (2021) when they forecast less than 3-months.

The returns of the left and right variance moments in the middle panel of Table VI are very impressive, given we are able to achieve an adjusted R^2 of up to 3%. The coefficients for the variance are significant at the 1% level across all time intervals considered, and again the left variance's coefficient is larger than the right variance's coefficient. What is also interesting is the negative signs in the variance suggests there is a 'cross-market leverage' effect happening. When the variance in the crude oil increases, the expected return for the S&P 500 futures decrease.

Finally, the regressions of the third moment are reported in the bottom panel of Table VI. The coefficients for the third moment are all negative, similar to the findings in Table IV, and the third moment achieves an adjusted R^2 up to 2.6%. The scale of the third moment displays the similar pattern found in the crude oil futures, in that when the overlapping period is small, the right third moment's coefficient is larger in scale than the left third moment's coefficient and is statistically significant at the 1% level. However, when the overlapping period is larger, the left third moment's coefficient is larger. We are not quite sure what is creating this dynamic, but it is interesting to document that this phenomenon translates across into the S&P 500 futures returns.

Our results clearly indicate that the returns of the high frequency crude oil semi-moments are able to partly predict the returns on the S&P 500 futures returns. The second and third semi-moments perform better than the tail risk measures, similar to the crude oil market, and the signs of the coefficient are similar to the results found in Section 4.2. Further to this, we document that the left semi-moments in the variance and tails contain more predictive power than their right semi-moment counterparts, and that the third moment exhibits the similar pattern to the findings in the crude oil market in that the right third moment contains more predictive information at smaller overlapping periods, which is clearly the challenging case. Overall, the quality of our results is in line, if we consider the significance of the coefficients and the level of the R^2, with the literature, although in our case we deal with the far more difficult objective of predicting returns at high frequency.

5 Conclusion

In this article we investigate the unexplored dataset of high frequency crude oil options. The use of these options allows us to extract the higher order risk neutral semi-moments, including the tail variation measures used in Bollerslev et al. (2015) and Andersen et al. (2021), at a high frequency - a first for the crude oil market. We document that the second moment and third moment have more explanatory power over both the tail risk measures, in slight contrast to the results found at the daily frequency, but nevertheless, the tail risk measures' coefficients are highly significant at all time intervals and contain some explanatory and predictive power over the crude oil futures and S&P 500 futures. We decompose these moments into semi-moments and find that the 'left' component, the put options, are larger than their 'right' counterpart, which suggests the put options are the main contributor to the overall higher order moment. We also find that the semi-moments are able to both explain, and predict, the crude oil and S&P 500 futures high frequency returns. We demonstrate the benefit of using semi-moments by comparing our regressions to regressions on synthetic moments designed to reproduce the aggregated risk-neutral moments; the semi-moments unambiguously perform better. Our results also indicate that mostly the left semi-moment contains more predictive information over its right counterpart for each moment. Overall, our results show that high frequency options, despite being challenging to handle, provide relevant information for the difficult problem of explaining and predicting returns at high frequency.

References

T. G. Andersen, O. Bondarenko, and M. T. Gonzalez-Perez. Exploring return dynamics via corridor implied volatility. *The Review of Financial Studies*, 28(10):2902–2945, 06 2015a. doi: 10.1093/rfs/hhv033.

T. G. Andersen, N. Fusari, and V. Todorov. The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584, 2015b. doi: 10.1016/j.jfineco.2015.06.005.

T. G. Andersen, N. Fusari, and V. Todorov. The pricing of tail risk and the equity premium: Evidence from international option markets. *Journal of Business & Economic Statistics*, 38(3):662–678, 2020. doi: 10.1080/07350015.2018.1564318.

T. G. Andersen, V. Todorov, and M. Ubukata. Tail risk and return predictability for the Japanese equity market. *Journal of Econometrics*, 222(1):344–363, 2021. doi: 10.1016/j.jeconom.2020.07.005.

Y. Aït-Sahalia, J. Fan, and Y. Li. The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109(1):224–249, 2013. doi: 10.1016/j.jfineco.2013.02.018.

G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1):101–143, 2003. doi: 10.1093/rfs/16.1.0101.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32, 2009. doi: 10.1111/j.1368-423X.2008.00275.x.

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

S. Basak and A. Pavlova. A model of financialization of commodities. *The Journal of Finance*, 71(4):1511–1556, 2016. doi: 10.1111/jofi.12408.

D. G. Baur and T. Dimpfl. The asymmetric return-volatility relationship of commodity prices. *Energy Economics*, 76:378–387, 2018. doi: https://doi.org/10.1016/j.eneco.2018.10.022.

H. Beckmeyer, N. Branger, and T. Grünthaler. The Fed call: FOMC announcements and stock market uncertainty. *Available at SSRN*, 2019. doi: 10.2139/ssrn.3379561.

F. Black. Studies of stock market volatility changes. *1976 Proceedings of the American Statistical Association Bisiness and Economic Statistics Section*, pages 177–181, 1976.

T. Bollerslev and V. Todorov. Tails, fears, and risk premia. *The Journal of Finance*, 66(6):2165–2211, 2011. doi: 10.1111/j.1540-6261.2011.01695.x.

T. Bollerslev and V. Todorov. Time-varying jump tails. *Journal of Econometrics*, 183(2):168 – 180, 2014. doi: 10.1016/j.jeconom.2014.05.007.

T. Bollerslev, V. Todorov, and L. Xu. Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134, 2015. doi: 10.1016/j.jfineco.2015.02.010.

S.-J. Byun and D.-H. Kim. Gambling preference and individual equity option returns. *Journal of Financial Economics*, 122(1):155–174, 2016. doi: 10.1016/j.jfineco.2016.06.004.

P. Carr and L. Wu. Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341, 2009. doi: 10.1093/rfs/hhn038.

C. Chiarella, B. Kang, C. S. Nikitopoulos, and T.-D. Tô. The return–volatility relation in commodity futures markets. *Journal of Futures Markets*, 36(2):127–152, 2016. doi: 10.1002/fut.21717.

D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000. doi: 10.1111/1468-0262.00164.

J. M. Griffin and A. Shams. Manipulation in the VIX? *The Review of Financial Studies*, 31(4):1377–1417, 2018. doi: 10.1093/rfs/hhx085.

F. Hollstein and M. Prokopczuk. Estimating beta. *Journal of Financial and Quantitative Analysis*, 51(4):1437–1466, 2016. doi: 10.1017/S0022109016000508.

B. Kang, C. S. Nikitopoulos, and M. Prokopczuk. Economic determinants of oil futures volatility: A term structure perspective. *Energy Economics*, 88:104743, 2020. doi: 10.1016/j.eneco.2020.104743.

M. Kilic and I. Shaliastovich. Good and bad variance premia and expected returns. *Management Science*, 65(6): 2522–2544, 2019. doi: 10.1287/mnsc.2017.2890.

R. Kozhan, A. Neuberger, and P. Schneider. The skew risk premium in the equity index market. *The Review of Financial Studies*, 26(9):2174–2203, 2013. doi: 10.1093/rfs/hht039.

I. Kyriakou, P. K. Pouliasis, and N. C. Papapostolou. Jumps and stochastic volatility in crude oil prices and advances in average option pricing. *Quantitative Finance*, 16(12):1859–1873, 2016. doi: 10.1080/14697688.2016.1211798.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

V. K. Ng and S. C. Pirrong. Fundamentals and volatility: Storage, spreads, and the dynamics of metals prices. *The Journal of Business*, 67(2):203–230, 1994.

M. Prokopczuk, L. Symeonidis, and C. Wese Simen. Variance risk in commodity markets. *Journal of Banking & Finance*, 81:136–149, 2017. doi: 10.1016/j.jbankfin.2017.05.003.

X. Ruan and J. E. Zhang. Risk-neutral moments in the crude oil market. *Energy Economics*, 72:583–600, 2018. doi: 10.1016/j.eneco.2018.04.026.

A Data cleaning

The following filters are applied to the individual option quotes for the entire dataset.

(F1) When multiple quotes have the same timestamp, we replace all these with a single entry with the median bid and median ask price. (Barndorff-Nielsen et al., 2009)

(F2) Delete entries for which the spread is negative. (Barndorff-Nielsen et al., 2009)

(F3) Delete entries for which the spread is more that 50 times the median spread on that day. (Barndorff-Nielsen et al., 2009)

(F4) Delete entries for which the mid-quote deviated by more than 10 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). (Barndorff-Nielsen et al., 2009)

(F5) Delete entries in which prices exceed 9 times the rolling standard deviation over the past 2 minutes and either (1) 75% of this price movement is reversed within one minute or (2) 80% of this price movement is reversed within two minutes. (The Bounceback Filter from Andersen et al. (2015a))

(F6) The last available quote in the past five minutes is used. When there is no quote for more than five minutes the option is removed from the calculations until a new quote becomes available. (Andersen et al., 2015a)

These filters are mild and are there to primarily guard against errant option quotes which would influence the final risk measure calculation. Filter (F1) is used to combine multiple quotes with the same timestamp. (F2) removes serious quote errors. (F3) and (F4) are used to remove excessive outliers that can arise out of a mistake in the dissemination of the quote data. Here we use 50 rolling observations similar to Barndorff-Nielsen et al. (2009). (F5) is the Bounceback filter from Andersen et al. (2015a) which is also used to remove outliers that end up reversing themselves within two minutes. This reversal is typically created from an error in the outlier quote that is then subsequently corrected. (F6) is used to prevent staleness in the option quotes.

B Tables

Table I: Summary statistics for the higher order risk-neutral moments and their semi-moments

	Mean	Std.dev.	Skewness	Kurtosis	Min	Q1	Q2	Q3	Max
var	0.1303	0.0891	2.5777	8.7882	0.0281	0.0751	0.0984	0.1574	0.7586
κ	−0.0079	0.0094	−2.4164	8.1555	−0.0839	−0.0099	−0.0053	−0.0018	0.0275
$LJV - RJV$	0.0315	0.0263	2.4457	9.6055	−0.0542	0.0140	0.0250	0.0400	0.2475
var^L	0.0062	0.0043	2.5024	8.3239	0.0013	0.0034	0.0047	0.0075	0.0372
var^R	0.0045	0.0030	2.6564	9.3372	0.0010	0.0027	0.0035	0.0054	0.0278
κ^L	0.0016	0.0018	3.3733	15.0954	0.0001	0.0006	0.0010	0.0019	0.0162
κ^R	0.0009	0.0011	3.8950	19.7896	0.0001	0.0004	0.0006	0.0010	0.011
LJV	0.0052	0.0036	2.5046	8.2085	0.0014	0.0028	0.0040	0.0062	0.0299
RJV	0.0026	0.0020	3.4481	16.0279	0.0006	0.0015	0.0019	0.0028	0.0201

Note. The table presents the mean, standard deviation, skewness, kurtosis and the quantiles for the higher order risk-neutral moments and their semi-moments.

Table II: Regression results for 5-minute crude oil futures returns using *contemporaneous returns* of the risk measures.

	(1)	(2)	(3)	(4)	(5)	(6)
Const.	11.5628***	11.5628***	11.5629***	11.5629***	11.5629***	11.5629***
	(2.13×10^6)	(2.07×10^6)	(2.05×10^6)	(1.84×10^6)	(1.88×10^6)	(1.88×10^6)
var^L	−0.0588***					
	(−37.81)					
var^R	−0.0443***					
	(−38.52)					
var'		−0.0345***				
		(−42.03)				
κ^L			−0.0088***			
			(−5.13)			
κ^R			−0.0577***			
			(−20.49)			
κ'				−0.0203***		
				(−15.33)		
LJV					−0.0008***	
					(−7.89)	
RJV					−0.0005***	
					(−7.37)	
$(LJV - RJV)'$						0.0002***
						(3.02)
Adj. R^2	19.4	12.9	19.6	2.3	0.4	0.0

Note. This table reports contemporaneous regressions for the 5-minute log returns on the crude oil and S&P 500 futures using the *returns* of the higher order risk-neutral semi-moments and the synthetic aggregated series var', κ' and $(LJV - RJV)'$. The constant and slope coefficients are reported with the robust t-statistic with a lag of 2 in parenthesis below. Adjusted R^2 is reported in percentage form. The symbol *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$.

Table III: Regression results for 60-minute crude oil futures returns using *contemporaneous returns* of the risk measures.

	(1)	(2)	(3)	(4)	(5)	(6)
Const.	9.0778*	9.0778***	9.0779***	9.0779***	9.0779***	9.0780***
	(1.29×10^5)	(1.28×10^5)	(1.35×10^5)	(1.19×10^5)	(1.22×10^5)	(1.16×10^5)
var^L	−0.0631***					
	(−21.47)					
var^R	−0.0450***					
	(−19.08)					
var'		−0.0505***				
		(−20.15)				
κ^L			−0.0606***			
			(−17.81)			
κ^R			−0.0136***			
			(−3.00)			
κ'				0.0296***		
				(8.10)		
LJV					−0.0164***	
					(−10.51)	
RJV					−0.0024***	
					(−6.60)	
$(LJV - RJV)'$						−0.0024***
						(−5.99)
Adj. R^2	23.8	21.6	23.5	2.3	5.8	0.3

Note. This table reports contemporaneous regressions for the 60-minute log returns on the crude oil and S&P 500 futures using the *returns* of the higher order risk-neutral semi-moments and the synthetic aggregated series var', κ' and $(LJV - RJV)'$. The constant and slope coefficients are reported with the robust t-statistic with a lag of 24 in parenthesis below. Adjusted R^2 is reported in percentage form. The symbol *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$.

Table IV: Regression results for forecasting 60-minute crude oil futures returns using the returns of the pairs LJV and RJV, var^L and var^R, and κ^L and κ^R.

	30-mins	25-mins	20-mins	15-mins	10-mins	5-mins
Const.	9.0780^{***}	9.0780^{***}	9.0780^{***}	9.0780^{***}	9.0780^{***}	9.0780^{***}
	(1.27×10^5)	(1.27×10^5)	(1.26×10^5)	(1.26×10^5)	(1.25×10^5)	(1.25×10^5)
LJV	-0.0095^{***}	-0.0077^{***}	-0.0063^{***}	-0.0049^{***}	-0.0027^{***}	-0.0011^{***}
	(-16.15)	(-15.21)	(-14.12)	(-12.86)	(-10.29)	(-6.85)
RJV	-0.0013^{***}	-0.0018^{***}	-0.0011^{***}	-0.001^{***}	-0.0012^{***}	-0.0004^{***}
	(-6.65)	(-6.85)	(-6.76)	(-6.50)	(-7.22)	(-4.76)
Adj. R^2	1.8	1.2	0.8	0.5	0.2	0.0
Const.	9.0779^{***}	9.0779^{***}	9.0779^{***}	9.0779^{***}	9.0779^{***}	9.0779^{***}
	(1.36×10^5)	(1.36×10^5)	(1.33×10^5)	(1.31×10^5)	(1.29×10^5)	(1.27×10^5)
var^L	-0.0620^{***}	-0.0620^{***}	-0.0616^{***}	-0.0616^{***}	-0.0618^{***}	-0.0608^{***}
	(-24.20)	(-23.59)	(-22.73)	(-21.82)	(-20.78)	(-19.01)
var^R	-0.0456^{***}	-0.0460^{***}	-0.0459^{***}	-0.0461^{***}	-0.0465^{***}	-0.0457^{***}
	(-23.22)	(-22.96)	(-22.37)	(-21.61)	(-20.73)	(-19.14)
Adj. R^2	12.1	10.1	8.0	6.0	4.0	1.8
Const.	9.0779^{***}	9.0779^{***}	9.0779^{***}	9.0780^{***}	9.0779^{***}	9.0779^{***}
	(1.27×10^5)	(1.26×10^5)	(1.25×10^5)	(1.26×10^5)	(1.29×10^5)	(1.43×10^5)
κ^L	-0.0470^{***}	-0.0428^{***}	-0.0380^{***}	-0.0321^{***}	-0.0239^{***}	-0.0138^{***}
	(-13.83)	(-12.59)	(-11.19)	(-9.688)	(-7.45)	(-4.44)
κ^R	-0.0293^{***}	-0.0338^{***}	-0.0385^{***}	-0.0450^{***}	-0.0531^{***}	-0.0603^{***}
	(-7.11)	(-8.50)	(-10.04)	(-9.69)	(-15.70)	(-16.41)
Adj. R^2	11.8	9.8	7.7	5.9	3.9	2.0

Note. This table reports predictive regressions for the 60-minute log returns on the *crude oil* futures using the *returns* of the higher order risk-neutral moments. The top panel displays the results for the left and right components of the tail jump variation LJV and RJV. The middle panel displays the results for the left and right semi-moments of the variance var^L and var^R. The bottom panel displays the returns for the left and right semi-moments of the third central moment κ^L and κ^R. The constant and slope coefficients are reported with the robust t-statistic with a lag of $2h_1$ in parenthesis below. Adjusted R^2 is reported in percentage form. The symbol *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$.

Table V: Regression results for 60-minute S&P 500 futures returns using *returns* of the risk measures.

	(1)	(2)	(3)	(4)	(5)	(6)
Const.	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***
	(2.96×10^5)	(2.96×10^5)	(3.00×10^5)	(2.89×10^5)	(2.92×10^5)	(2.88×10^5)
var^L	−0.0127***					
	(−11.16)					
var^R	−0.0090***					
	(−10.63)					
var'		−0.0102***				
		(−10.86)				
κ^L			−0.0131***			
			(−9.54)			
κ^R			−0.0019*			
			(−1.90)			
κ'				0.0068***		
				(5.78)		
LJV					−0.0033***	
					(−7.44)	
RJV					−0.0004***	
					(−3.48)	
$(LJV - RJV)'$						−0.0006***
						(−4.94)
Adj. R^2	6.1	5.5	6.2	0.8	1.5	0.1

Note. This table reports contemporaneous regressions for the 60-minute log returns on the S&P 500 futures using the *returns* of the higher order risk-neutral semi-moments and the synthetic aggregated series var', κ' and $(LJV - RJV)'$. The constant and slope coefficients are reported with the robust t-statistic with a lag of 24 in parenthesis below. Adjusted R^2 is reported in percentage form. The symbol *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$.

Table VI: Regression results for forecasting 60-min S&P 500 futures returns using the returns of the pairs LJV and RJV, var^L and var^R, and κ^L and κ^R.

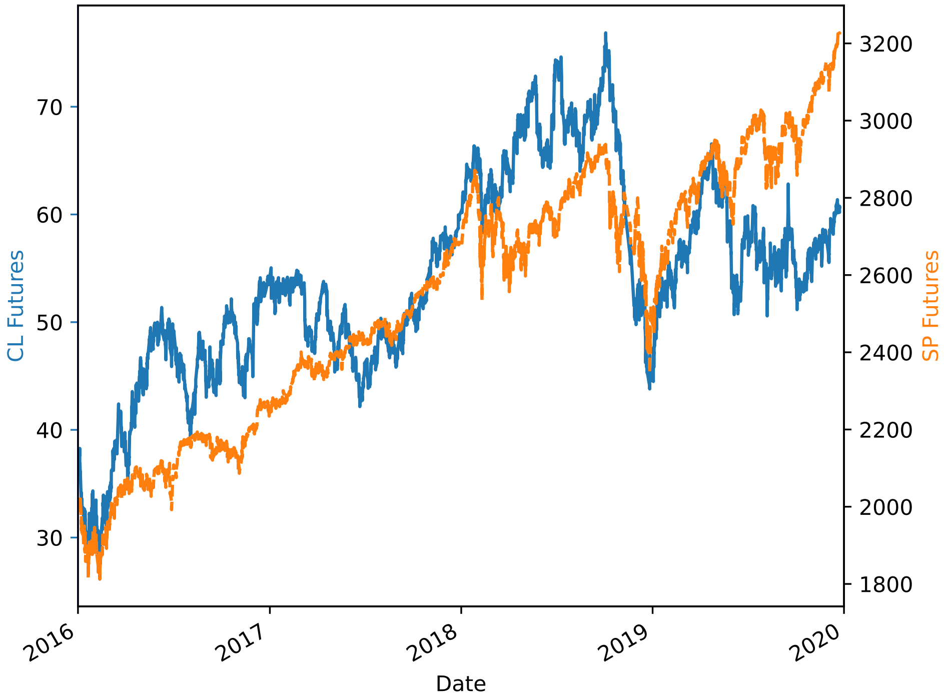
	30-mins	25-mins	20-mins	15-mins	10-mins	5-mins
Const.	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***
	(3.33×10^5)	(3.33×10^5)	(3.33×10^5)	(3.32×10^5)	(3.32×10^5)	(3.32×10^5)
LJV	−0.0018***	−0.0014***	−0.0012***	−0.0009***	−0.0005***	−0.0001**
	(−9.68)	(−9.18)	(−8.57)	(−7.76)	(−6.04)	(−2.46)
RJV	−0.0003***	−0.0003***	−0.0002***	−0.0002***	−0.0003***	$-5.98 \times 10^{-5**}$
	(−3.85)	(−3.80)	(−3.98)	(−3.84)	(−5.12)	(−2.40)
Adj. R^2	0.4	0.3	0.2	0.1	0.1	0.0
Const.	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***
	(3.38×10^5)	(3.37×10^5)	(3.36×10^5)	(3.35×10^5)	(3.34×10^5)	(3.33×10^5)
var^L	−0.0119***	−0.0116***	−0.0111***	−0.0108***	−0.0103***	−0.0094***
	(−12.76)	(−12.18)	(−11.48)	(−10.76)	(−9.80)	(−8.68)
var^R	−0.0089***	−0.0088***	−0.0085***	−0.0084***	−0.0081***	−0.0072***
	(−12.66)	(−12.19)	(−11.65)	(−11.10)	(−10.16)	(−8.84)
Adj. R^2	3.0	2.4	1.8	1.2	0.8	0.3

	30-mins	25-mins	20-mins	15-mins	10-mins	5-mins
Const.	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***	9.0780***
	(3.00×10^5)	(3.02×10^5)	(3.06×10^5)	(3.12×10^5)	(3.24×10^5)	(3.64×10^5)
κ^L	−0.0095***	−0.0086***	−0.0074***	−0.0060***	−0.0043***	−0.0022**
	(−7.58)	(−6.78)	(−5.83)	(−4.92)	(−3.63)	(−2.17)
κ^R	−0.0047***	−0.0054***	−0.0062***	−0.0073***	−0.0085***	−0.0091***
	(−4.37)	(−5.01)	(−5.74)	(−6.84)	(−8.48)	(−9.63)
Adj. R^2	2.6	2.1	1.6	1.1	0.7	0.3

Note. This table reports predictive regressions for the 60-minute log returns on the S&P 500 futures using the *returns* of the higher order risk-neutral moments. The top panel displays the results for the left and right components of the tail jump variation LJV and RJV. The middle panel displays the results for the left and right semi-moments of the variance var^L and var^R. The bottom panel displays the returns for the left and right semi-moments of the third central moment κ^L and κ^R. The constant and slope coefficients are reported with the robust t-statistic with a lag of $2h_1$ in parenthesis below. Adjusted R^2 is reported in percentage form. The symbol *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$.

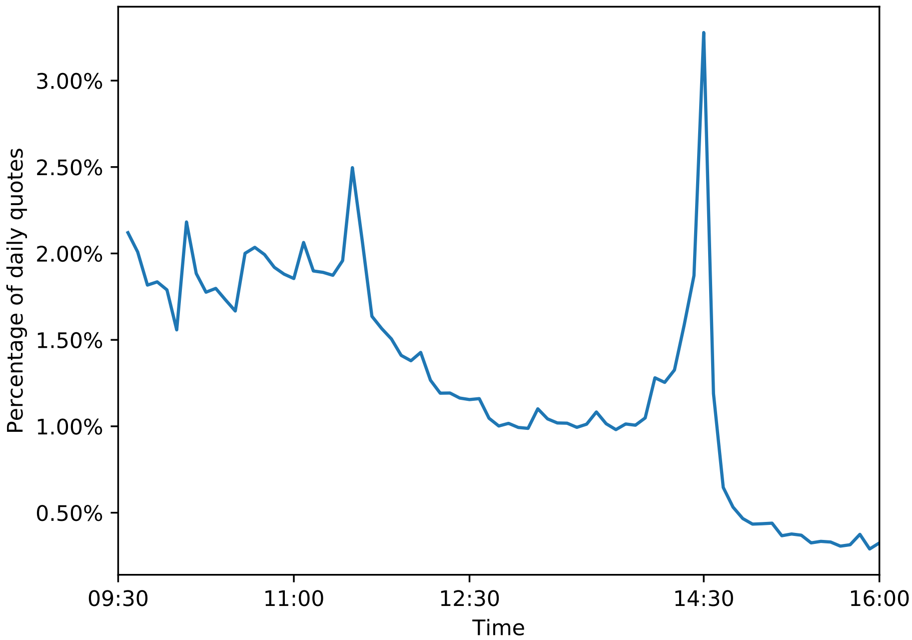
C Figures

Figure 1: Price of the crude oil futures and the S&P 500 futures.



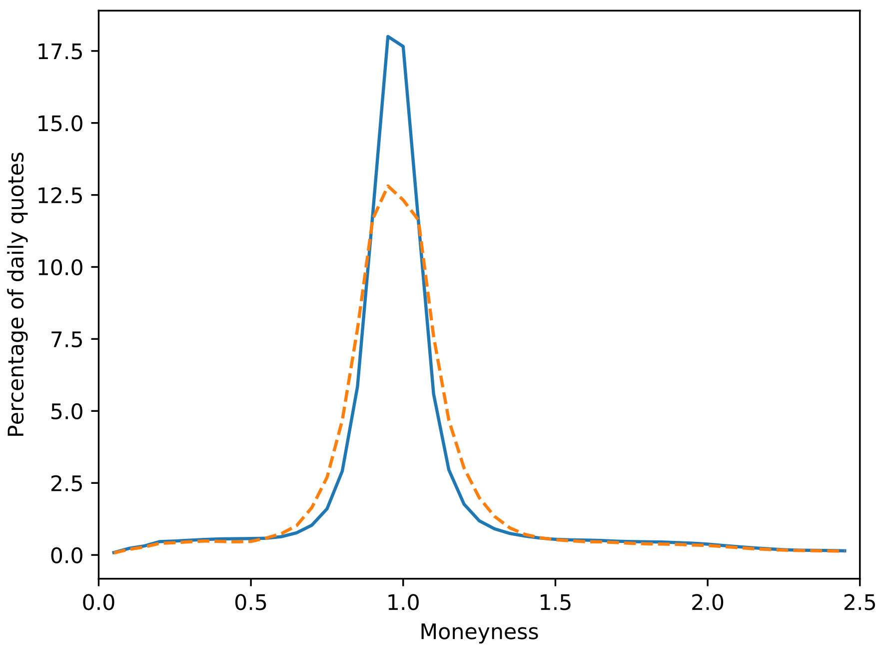
Note. Left y-axis is the price of the crude oil (CL) lead-month futures (blue solid). Right y-axis is the price of the S&P 500 (SP) lead month-futures (orange dashed). Prices are plotted over the sample period Jan 1, 2016 to Dec 31, 2019.

Figure 2: Average OTM quote activity within the trading day



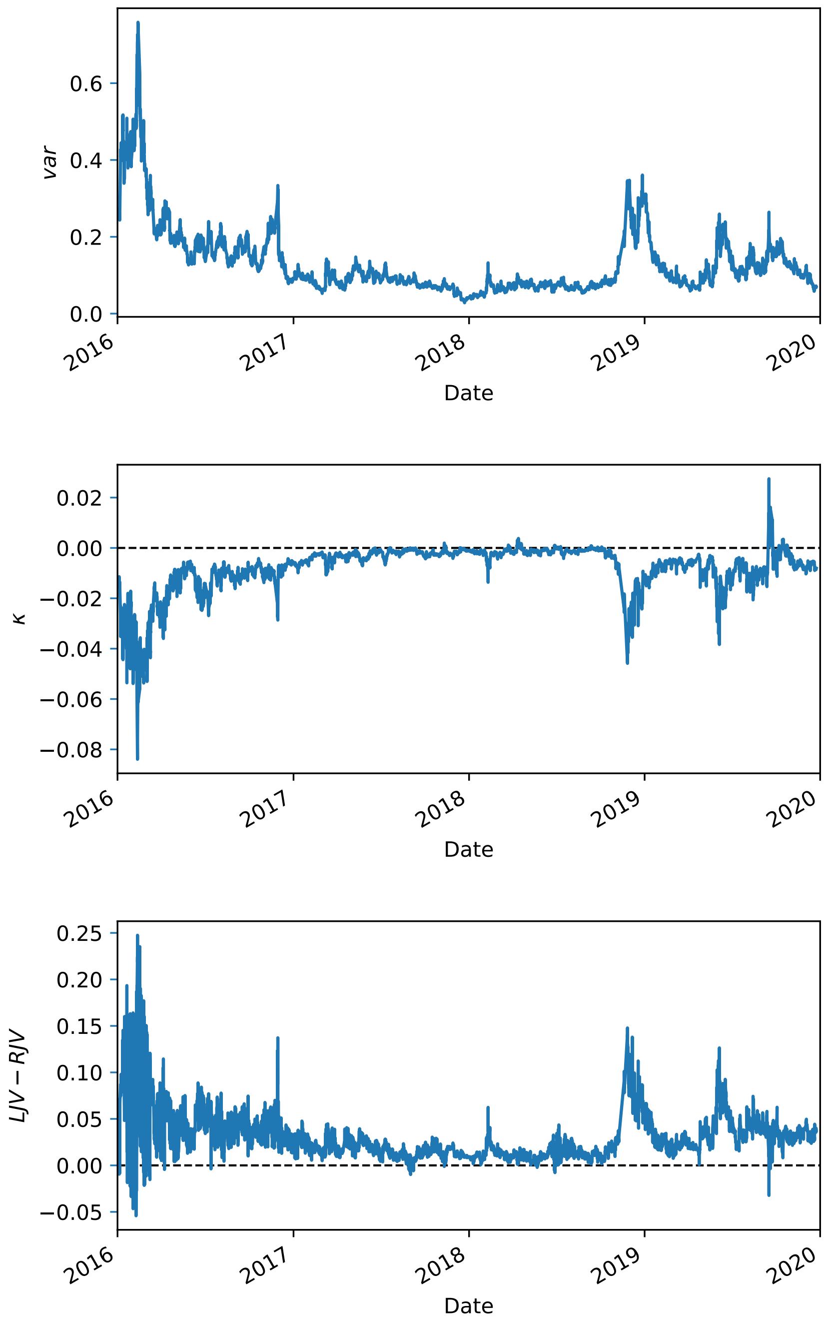
Note. Percentage of the daily OTM quotes arriving in 5 minute intervals. For example, the value of 3.28% at 14:30 means 3.28% of the daily OTM quotes arrived between 14:25-14:30.

Figure 3: Average OTM quote activity across moneyness



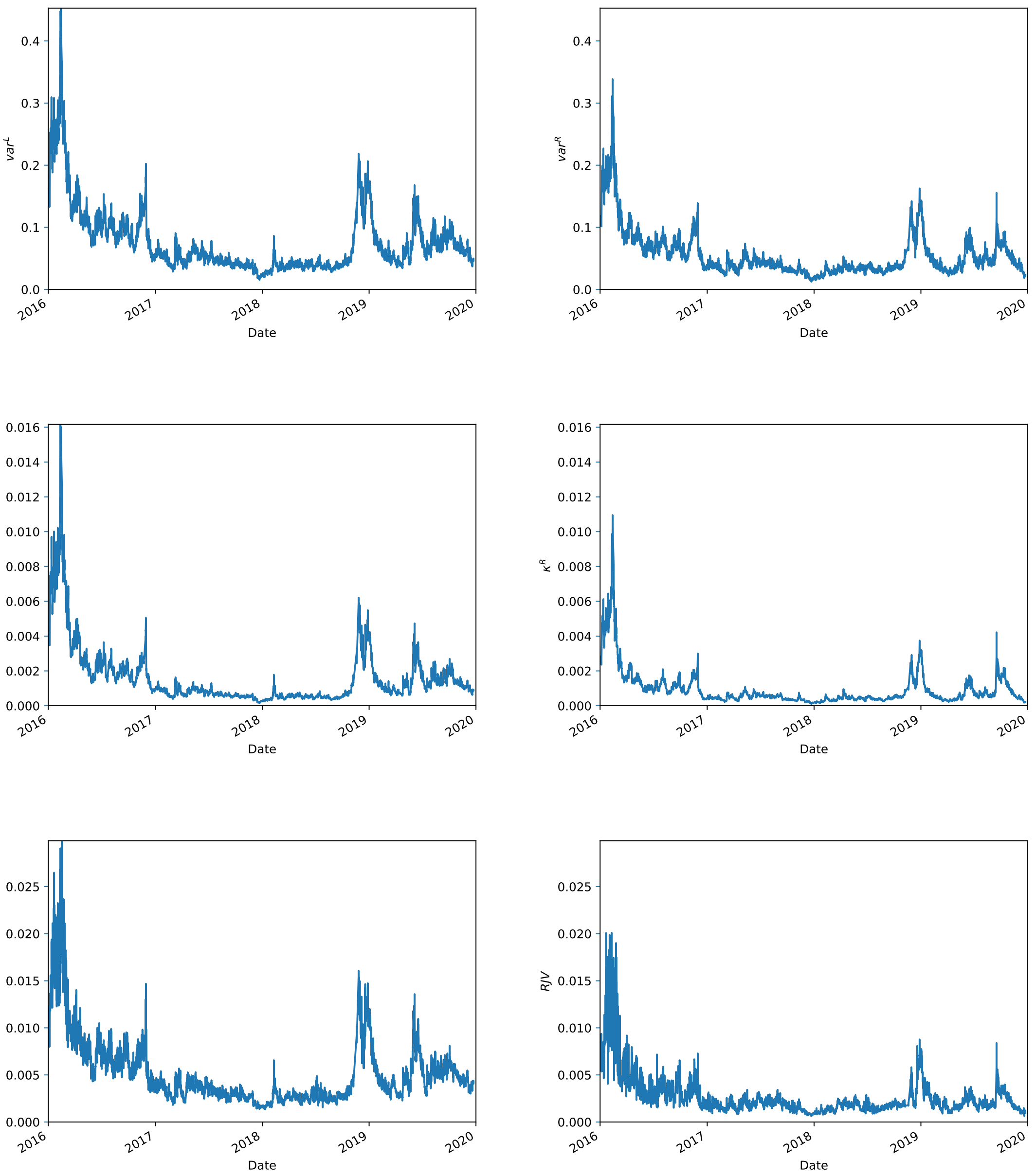
Note. Percentage of daily quotes across moneyness for the near term (solid blue) and far term (dashed orange). Moneyness of the option is defined as K/F, where K is the strike price of the option and F is the forward price of the underlying asset. Moneyness is grouped by values of 0.05. For example, the value of 17.6% at 1 for the near term means 17.6% of the daily quotes are from the options with a moneyness between (0.95, 1).

Figure 4: Risk measures



Note. Each panel displays the higher order risk-neutral moments over the sample period Jan 1, 2016 to Dec 31, 2019. The top panel displays the 30-day expected variance var from Eq.(11). The middle panel displays the third central moment κ from Eq.(21). The bottom panel displays the difference between the left jump variation and right jump variation from Eq.(26).

Figure 5: Higher order semi-moments



Note. Each panel displays the higher order risk-neutral semi-moment over the sample period of Jan 1, 2016 to Dec 31, 2019. The top panel displays the left and right semi-moments of the variance var^L and var^R from Eq.(12)-(13). The middle panel displays the left and right semi-moments of the third moment κ^L and κ^R from Eq.(22)-(23). The bottom panel displays the left and right jump variations LJV and RJV from Eq.(31)-(32).