Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests

Riza Demirer[a], Konstantinos Gkillas[b], Rangan Gupta[c], Christian Pierdzioch[d]

September 2019

Abstract

We analyze the predictive power of time-varying risk aversion for the realized volatility of crude oil returns based on high-frequency data. While the popular linear heterogeneous autoregressive realized volatility (HAR-RV) model fails to recognize the predictive power of risk aversion over crude oil volatility, we find that risk aversion indeed improves forecast accuracy at all forecast horizons when we compute forecasts by means of random forests. The predictive power of risk aversion is robust to various covariates including realized skewness and realized kurtosis, various measures of jump intensity and leverage. The findings highlight the importance of accounting for nonlinearity in the data-generating process for forecast accuracy as well as the predictive power of non-cashflow factors over commodity-market uncertainty with significant implications for the pricing and forecasting in these markets.

JEL classification: G17; Q02; Q47

Keywords: Oil price; Realized volatility; Risk aversion; Random forests

[a] Department of Economics and Finance, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1102, USA; E-mail address: rdemire@siue.edu.

[b] Department of Business Administration, University of Patras – University Campus, Rio, P.O. Box 1391, 26500 Patras, Greece; Email address: gillask@upatras.gr.

[c] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; E-mail address: rangan.gupta@up.ac.za.

[d] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: c.pierdzioch@hsu-hh.de.

1 Introduction

The recent financialization of the commodity market has resulted in increased participation of hedge funds, pension funds, and insurance companies in oil-based financial products (e.g., Hamilton and Wu 2015). While several studies in the literature argue that supply and demand fundamentals serve as the main driver of commodity prices (e.g., Irwin and Sanders 2011), others suggest otherwise, arguing that speculative activities have played a significant role in risk premiums and volatility in the commodity market (e.g., Hong and Yogo 2012, Singleton 2014). Indeed, recent evidence suggests that investors' risk preferences towards the oil price contain a business-cycle component (Christoffersen and Pan 2017) and that risk premiums embedded in crude oil prices are determined by the risk aversion of market participants (Li 2018). Further, motivated by the evidence in Robe and Wallen (2016) that implied volatilities extracted from crude oil options are driven partly by the VIX, considered as a proxy for risk aversion and macro uncertainty, this paper provides novel insight to the predictive value captured by changes in risk preferences over crude oil-returns volatility within an innovative forecasting setting by utilizing the measure of time-varying risk aversion recently developed by Bekaert et al. (2017). By doing so, this study contributes to the literature on the predictive power of behavioral factors over commodity-market dynamics from a novel context.

Considering that crude oil is arguably the world's most strategic commodity, accurate forecasts of oil-return volatility are of paramount importance to not only oil traders and corporate decision makers, but also to policymakers as oil market uncertainty can negatively impact economic activity and contribute to aggregate macroeconomic uncertainty (Elder and Serletis 2010, Aye et al. 2014, van Eyden 2019). Naturally, a large literature exists on the forecastability of oil-returns volatility using various types of univariate and multivariate Generalized

Autoregressive Conditional Heteroskedasticity (GARCH) models (see, for example, Sadorsky 2006, Sadorsky and McKenzie 2008, Agnolucci 2009, Kang et al. 2009, Wei et al. 2010, Nomikos and Pouliasis 2011, Arouri et al. 2012, Hou and Suardi 2012, Kang and Yoon 2013, Chkili et al. 2014, Efimova and Serletis 2014). In general, these studies find that univariate GARCH-type models are able to produce more accurate forecasts than its competitors. More recently, comparing the Markov-switching multifractal (MSM) model with a battery of GARCH models, Lux et al. (2016) show that the MSM model is the preferable framework in most cases across various forecasting horizons and sub-samples analyzed.

A common feature of the aforementioned studies, however, is that they examine oil returns at daily frequency without taking advantage of the predictive information contained in intraday dynamics. In fact, numerous works including Andersen et al. (2001, 2003, 2010), Oomen (2001), Hansen and Lunde (2005), Engle and Sun (2007), McAleer and Medeiros (2008) and Tay et al. (2009) suggest that the rich information contained in intraday data can be utilized to produce more accurate estimates and forecasts of daily return volatility. In light of this, Haugom et al. (2014), Sévi (2014), Prokopczuk et al. (2015), Degiannakis and Filis (2017), Liu et al. (2017), Chen et al. (2019) and Gkillas et al. (forthcoming) use variants of the Heterogeneous Autoregressive (HAR) model developed by Corsi (2009) to forecast the realized volatility (RV) of crude oil returns (i.e., the sum of non-overlapping squared high-frequency oil returns observed within a day, see Andersen and Bollerslev 1998).[1] The popularity of the HAR-RV model stems from its ability to capture important stylized facts of financial-market volatility such as long memory and multi-scaling behaviour. In fact, barring the recent

[1]Chatrath et al. (2015) and Phan et al. (2016) also forecast realized volatility of oil returns as derived from intraday data via regression and GARCH-based models instead of the HAR-RV framework. In addition, two other related papers are that of Asai et al. (2019, forthcoming) which looks at the role of jumps, leverage, spillovers and geopolitical risks in forecasting the co-volatility of the oil and gold markets based on intraday data.

studies of Degiannakis and Filis (2017), Liu et al. (2017), Chen et al. (2019) and Gkillas et al. (forthcoming), earlier works based on intraday data generally conclude that none of the alternative models is able to outperform the forecasting accuracy of a simple HAR-RV model, which uses only the information embedded in the realized volatility to produce forecasts. In contrast, Degiannakis and Filis (2017) show that it is possible to beat the HAR-RV model by incorporating information on the exogenous volatilities of four different asset classes (stocks, currencies, commodities and macroeconomic policy), while Gkillas et al. (forthcoming) show that extending the baseline linear HAR-RV model to incorporate an index of financial stress improves forecast accuracy once they account for the potential asymmetry of a forecaster's loss function. At the same time, Liu et al. (2017) and Chen et al. (2019) show that allowing for time-variation and accounting for asymmetric volatility jumps and co-jumps with the equity (S&P 500) market, respectively can outperform the benchmark HAR-RV model.

Against this backdrop, this paper extends the limited literature on forecasting realized oil-returns volatility (derived from intraday data) by incorporating the recently developed measure of time-varying risk aversion by Bekaert et al. (2017) into the popularly employed HAR-RV framework. This measure of risk-aversion distinguishes the time variation in economic uncertainty (the amount of risk) from time variation in risk aversion (the price of risk), and provides an unbiased representation of time-varying risk aversion in financial markets. It can be argued that the predictive value of risk aversion over crude oil-returns volatility stems from at least two distinct channels. Considering the evidence that investor sentiment has forecasting power over future macroeconomic conditions (e.g., Ludvigson 2004, Souleles 2004, Lemmon and Portniaguina 2006) and that real economic activity drives oil prices (e.g., Kilian and Hicks 2013), real economic activity is one possible channel that links risk aversion to oil-returns volatility.[2] A second channel that links risk aversion to oil-returns volatility can operate through the speculative activities in the marketplace. Li (2018) notes that crude oil-risk premiums are low when the level of speculation is high (which corresponds to periods when aggregate risk aversion of market participants is low). Therefore, it is possible that changes in risk aversion serve as a determinant of speculative behavior in the oil market and, thereby, capture predictive information over volatility in this market. To the best of our knowledge, ours is the first study to utilize the unbiased measure of risk aversion developed by Bekaert et al. (2017) in the context of forecasting crude oil-returns volatility.

On the technical side, our research adds to the large literature on HAR-RV modeling in empirical finance in that we use a machine-learning technique known as random forests to model the realized volatility of crude oil returns. In our context, the random forest technique offers two key advantages relative to the standard HAR-RV model as it captures possible nonlinear links between realized volatility and risk aversion in a data-driven way and allows, in a unified modeling framework, the incremental predictive value of risk aversion to be studied in the presence of several other covariates of realized volatility studied in the empirical-finance literature. Indeed, our findings show that the standard HAR-RV model misses the nonlinearity in the data and fails to recognize the predictive value of risk aversion in out-of-sample tests. In sharp contrast, using various variants of random forests (including quantile random forests and so called extreme random forests), we document strong evidence that risk aversion helps to improve out-of-sample accuracy of forecasts for realized oil volatility, even in the presence of the additional covariates including measures of good and bad realized volatility, real-

[2] Indeed, Qadan and Nama (2018) document evidence of a significant link between volatility in sentiment indices to volatility in oil prices, particularly after oil-based financial products started to attract greater attention from financial investors.

ized skewness and kurtosis, and various types of realized volatility jumps that are employed in the literature (see, e.g., Lyócsa and Molnár 2016, Mei et. al. 2017). To that end, this paper presents a technical contribution to the strand of the literature on volatility forecasting in the commodity market in addition to providing novel insight into the predictive role of risk aversion over volatility in the crude oil market.

The remainder of the paper is organized as follows. We describe our empirical methods in Section 2 and our data in Section 3. We summarize our empirical results in Section 4 and offer some concluding remarks in Section 5.

2 Methods

2.1 The HAR-RV Model

Following Andersen et al. (2012), daily realized volatility (RV) is measured by the median realized variance (MRV) as a jump-robust estimator of integrated variance, computed from intraday data as

$$MRV_t = \frac{\pi}{6-4\sqrt{3}+\pi}\frac{T}{T-2}\sum_{i=2}^{T-1} \mathrm{med}\left(|r_{t,i-1}|,|r_{t,i}|,|r_{t,i+1}|\right)^2, \tag{1}$$

where $r_{t,i}$ is the intraday return i within day t and $i = 1,..,T$ denotes the number of intraday observations within a day. As an estimator for realized volatility, MRV is considered to be robust to jumps and less biased in the presence market-microstructure noise.[3] The baseline model to forecast realized volatility follows the widely-studied HAR-RV model (Corsi, 2009) for h-days-ahead, described as

$$RV_{t+h} = \beta_0 + \beta_d\, RV_t + \beta_w\, RV_{w,t} + \beta_m\, RV_{m,t} + \varepsilon_{t+h}, \tag{2}$$

[3]Researchers often use the term volatility to denote the standard deviation of asset-price returns. To avoid any risk of confusion, we use the term realized volatility to denote the realized variance of oil-price returns, and explicitly denote the square root of the realized volatility as the realized standard deviation of oil-price returns.

where $h =$ forecast horizon, $RV_{w,t} =$ average RV from day $t-5$ to day $t-1$, and $RV_{m,t} =$ average RV from day $t-22$ to day $t-1$. In this specification, we define $RV_{t+h} = \text{mean}\{RVt+1,...,RVt+h\}$. Three forecast horizons are analyzed: a short daily forecasting horizon ($h = 1$), a medium weekly forecasting horizon ($h = 5$), and a long monthly forecasting horizon ($h = 22$). The main idea motivating the HAR-RV model is to use volatilities from different time resolutions to forecast realized volatility and, thereby, to capture the key element of the heterogeneous market hypothesis (Müller et al. 1997). In our context, this translates into a setting in which the market for oil is populated by various groups of market participants, with short- and long-term traders responding to information flows at different time horizons. To that end, the HAR-RV framework renders it possible to account for different groups of agents including managed money traders who are subject to funding constraints (Acharya et al. 2013) as well as commodity index traders (e.g., Hamilton and Wu 2014), who may be less focused on short-term interactions.

We use the standard HAR-RV model as a nucleus for building more complex models and first extend the baseline model to the HAR-RV-RA model by including risk aversion, RA_t, as

$$RV_{t+h}{=}\beta_0 + \beta_d\ RV_t + \beta_w\ RV_{w,t} + \beta_m\ RV_{m,t} + \theta\ RA_t + \varepsilon_{t+h}. \tag{3}$$

Note that the HAR-RV-RA model given in Equation (3) retains the linear structure of the baseline HAR-RV model. A drawback of this specification is that the HAR-RV-RA model may fail to recognize the contribution of risk aversion to the predictive value of the model if the relationship between risk aversion and future realized volatility is nonlinear.

2.2 The HAR-RV-RF Model

As noted earlier, one of the technical novelties of this study is that we use a machine-learning technique known as random forests to model the realized volatility of crude oil returns. Compared to the widely-employed HAR-RV model, which is based on the OLS methodology, the random forest technique adopts a data-driven approach to capture possible non-linearities in the relationship between the variable of interest and its predictors. Furthermore, this technique makes it possible to study the incremental predictive value of a given predictor (risk aversion in our case) in the presence of several other predictors, where the potential interactions of the predictors are automatically accounted for by the data-driven modeling philosophy that forms the foundation of random forests.

Following the notation also used by Hastie et al. (2009, Chapter 9), a regression tree, T, that features J terminal nodes recursively partitions the space of covariate variables, $\mathbf{x} = (x_1, x_2, ...)$, into l non-overlapping regions, R_l. The list of covariate variables comprises the standard HAR-RV predictors (that is, $RV_t, RV_{w,t}$, and $RV_{m,t}$), risk aversion (RA_t) and, in addition, the other variables that have been utilized in the strand of the literature on modeling and forecasting realized volatility (see Section 2.3). In addition to the advantages listed earlier, an important feature of regression trees is that they render it possible to study an arbitrarily large number of covariate variables via a simple greedy algorithm that determines the partitions in a top-down and binary way. At the top level of a regression tree, the algorithm chooses the first partition in such a way that the partitioning covariate variable, s, and the selected partitioning point, p, define the half-planes $R_1(s,p) = \{x_s | x_s \leq p\}$ and $R_2(s,p) = \{x_s | x_s > p\}$ that minimize the loss function:

$$\min_{s,p} \left\{ \min_{\bar{RV}_1} \sum_{x_s \in R_1(s,p)} (RV_i - \bar{RV}_1)^2 + \min_{\bar{RV}_2} \sum_{x_s \in R_2(s,p)} (RV_i - \bar{RV}_2)^2 \right\}, \tag{4}$$

where i denotes the $t+h$ data on realized volatility that belong to the half-planes, $\bar{RV}_k = \text{mean}\{RV_i \,|x_s \in R_k(s,p)\}, k = 1,2$ denotes the half-plane-specific mean of realized volatility, and the inner minimization is done by choosing the half-plane-specific means to minimize the region-specific squared error loss.[4] The outer minimization, in turn, requires a simple search over all possible combinations of s and p that results in the selection of the first optimal partitioning predictor, the first optimal partitioning point, and the two region-specific means of realized volatility. This search produces a slightly more complex new regression tree with two terminal nodes.

In the next step, the minimization setup described in Equation (4) is carried on to the new regression tree with two-terminal-nodes, this time separately for the two optimal top-level half-planes, $R_1(s,p)$ and $R_2(s,p)$. This minimization yields up to two second-level optimal partitioning covariates, two second-level optimal partitioning points, and four second-level region-specific means of realized volatility. This minimize-and-partition search process is repeated until the regression tree has a preset maximum number of terminal nodes or every terminal node has a minimum number of observations.[5] The final regression tree sends the covariates into the R_l optimal regions of the regression tree and then predicts the $i-$th observation of $t+h$ realized volatility by its region-specific mean

$$T\left(\mathbf{x}_i, \{R_l\}_1^L\right) = \sum_{l=1}^{L} \bar{RV}_l \mathbf{1}(\mathbf{x}_i \in R_l), \tag{5}$$

where $\mathbf{1}$ is the indicator function.

In this procedure, finer predictions for realized volatility can be computed via Equation (5) by growing a large enough regression tree. However, because a large

[4]For ease of notation, we do not use in Equation (4) an index for the forecast horizon, h and the time index, t.

[5]This procedure requires a preset maximum number of terminal nodes and a preset minimum number of observations per terminal node.

individual regression tree is data-sensitive due to its hierarchical structure, a more efficient way to compute granular predictions of realized volatility is to combine a large number of simple regression trees. One approach used in this regard is bootstrap aggregation (bagging, Breiman 1996) which proceeds in three steps. In the first step, a large number of bootstrap samples is obtained from the data. In the second step, a regression tree is estimated on every sample. In the third step, realized volatility is predicted by averaging over the bootstrapped regression trees. An alternative approach, the one that is adopted in this paper, is to grow random forests (Breiman 2001). As compared to bootstrap aggregation, random forests have the advantages that they mitigate the influence of individual covariates and that they decorrelate the predictions from individual trees and, thereby, stabilize the predictions of realized volatility. Growing a random forest resembles bootstrap aggregation in that it requires drawing a large number of samples from the data. In the next step, however, a random regression tree is estimated on every sample. A random regression tree differs from a standard regression tree in that for every partitioning step, only a random subset of the covariates is selected.

2.3 Other Covariates of Realized Volatility

As mentioned earlier, an advantage of random forests is that we can use an arbitrary number of covariates for tree-building, where the trees account in a natural way for potential interactions between the covariates. This feature of random forests is advantageous because earlier studies have used several covariates of realized volatility to extend the standard HAR-RV model. In our case, rather than considering various models in isolation, we use random forests as an integrated modeling platform that permits the data to decide which covariate to use for tree-building. This procedure also allows us to estimate HAR-RV-RF models with and without risk aversion in the list of covariates and study the incremental out-of-sample predictive value of risk aversion after controlling for the effects of other covariates employed in earlier studies. In our application, we analyze a large number of covariates for tree-building including (i) various types of jumps (i.e., total jumps, downside (bad) jumps, upside (good) jumps, asymmetric jumps, truncated large jumps, truncated small jumps) and (ii) realized skewness and kurtosis. Our methodology for the detection of jumps is based on a benchmark scheme developed by Huang and Tauchen (2005) and Barndorff-Nielsen and Shephard (2006).

2.3.1 Detecting jumps in the volatility process

The procedure to detect jumps in the volatility process is based on the decomposition of volatility into a jump and a continuous component. This procedure requires an estimator that excludes jumps. In one of the pioneering works in this regard, Barndorff-Nielsen and Shephard (2006) propose an accurate estimate of the integrated variance excluding jumps, named bipower variation (BPV), formulated as

$$BPV_t = \xi_1^{-2} \sum_{i=2}^{T} |r_{t,i}||r_{t,i-1}|, \tag{6}$$

where $\xi_1 = \sqrt{(2/\pi)} = E(|Z|)$ denotes the absolute value of mean of a standard normally distributed random variable Z. As in Andersen et al. (2007), total jumps are detected by the ratio-statistic $(ZJ_t^{(BPV)})$ (also called jump-ratio test), which is given by

$$ZJ_t^{(BPV)} = \frac{(RV_t - BPV_t)\, RV_t^{-1}}{\left(\left(\xi_1^{-4} + 2\xi_1^{-2} - 5\right) max\left\{1, TQ_t BPV_t^{-2}\right\}\right)^{1/2}} \to N(0,1), \tag{7}$$

where $RV_t = \sum_{i=1}^{T} r_{t,i}^2$ is the realized variance defined as a consistent estimator of the integrated variance plus the jump contribution, and the the realized tripower quarticity is given by $TQ_t = T\xi_{4/3}^{-3} \sum_{i=3}^{T} |r_{t,i}|^{4/3} |r_{t,i-1}|^{4/3} |r_{t,i-2}|^{4/3}$. TQ_t is an asymptotically unbiased estimator of integrated quarticity in the absence of microstructure noise and converges in probability to integrated quarticity. The $ZJ_t^{(BPV)}$

is useful as a pre-test, prior to detecting jumps under the null hypothesis of no jumps. Finally, total jumps are detected by applying the jump-detection scheme which follows from

$$RJ_t = I_{[ZJ_t^{(BPV)} > \Phi_a]} \, |RV_t - BPV_t|, \tag{8}$$

where $I_{\{.\}}$ denotes the indicator function for $ZJ_t^{(BPV)}$ exceeding a given critical value. Therefore, a jump occurs when the $ZJ_t^{(BPV)}$ exceeds the corresponding critical value of the normal distribution, denoted by Φ_a, at a a level of significance.

2.3.2 Upside, downside and asymmetric jumps

A growing number of studies suggest that accounting for possible asymmetries in the volatility and/or jump processes governing returns can improve forecasting models for future excess stock market returns (e.g., Guo et al. 2019, among others). In a realized estimation framework, Bollerslev et al. (forthcoming) use the realized up (good) and down (bad) semi-variances for this distinction. Good and bad realized volatility (semi-variance) can serve as measures of downside and upside risk and capture the sign asymmetry in the volatility process. Sévi (2014) and Chen et al. (2019) uses HAR-type models along with realized semivariances and signed jumps as in Patton and Sheppard (2011) to forecast crude oil volatility. In this paper, following Barndorff-Nielsen et al. (2010), downside and upside realized semi-variance (RV^B and RV^G) are computed as

$$RV_t^B = \sum_{i=1}^{T} r_{t,i}^2 \, I_{[(r_{t,i})<0]}, \tag{9}$$

$$RV_t^G = \sum_{i=1}^{T} r_{t,i}^2 \, I_{[(r_{t,i})>0]}, \tag{10}$$

where $I_{\{.\}}$ is the indicator function. We consider daily RV^B as "bad" realized volatility and daily RV^G as "good" realized volatility in order to capture the sign asymmetry of the volatility process. Understandably, $RV^S = RV^B + RV^G$.

Further, following Duong and Swanson (2011, 2015), we estimate downside (bad), upside (good) jumps formulated by power transformation instead of decomposition using truncation levels (see sub-section 2.3.3) as

$$RJ_t^- = I_{[ZJ_t^{(BPV)} > \Phi_a]} \sum_{i=1}^{T} |r_{t,i}|^q \; I_{[(r_{t,i})<0]}, \tag{11}$$

$$RJ_t^+ = I_{[ZJ_t^{(BPV)} > \Phi_a]} \sum_{i=1}^{T} |r_{t,i}|^q \; I_{[(r_{t,i})>0]}, \tag{12}$$

where q is the asymmetry parameter which affects the limiting behavior of the estimator. We are interested in the case where $q \geq 2$, since large values of q, are dominated by large jumps. For $q < 2$, the jump variations are not always guaranteed to be finite. The q is selected to be 2.5 per Duong and Swanson (2011) who found more significant HAR coefficients from a range of q from 2 to 6.

Finally, the asymmetric-jump component is computed as

$$RJA_t = I_{[ZJ_t^{(BPV)} > \Phi_a]} \; RJ_t^+ - RJ_t^- . \tag{13}$$

The RJA_t is a so-called signed jump too. A negative (positive) sign implies that the jump occurred in day t is associated with kind of events that primarily affected the bad (good) volatility, hence economic conclusions can be drawn.

2.3.3 Truncated large and small jumps

Following Duong and Swanson (2011), we also estimate the realized measure of truncated large $(RVLJ_t)$ and small jumps $(RVSJ_t)$ using a decomposition based on a fixed truncation level (γ) (a threshold method to select the appropriate value of γ, that can separate the $(RVLJ_t)$ from $RVSJ_t$). Noting the limited literature on forecasting oil volatility, particularly taking into account the forecasting ability of $RVLJ_t$ and $RVSJ_t$, Liu et al. (2018) show that HAR-type models including both $RVLJ_t$ and $RVSJ_t$ can provide significantly superior forecasting performance on oil futures data. In particular, including $RVSJ_t$ in a HAR model is found to significantly improve the forecast accuracy at the 1-day forecasting horizon, while including both $RVLJ_t$ and $RVSJ_t$ can achieve significantly higher forecast accuracy at longer forecasting horizons (e.g., weekly or monthly). In our application, $RVLJ_t$ and $RVSJ_t$ are computed as

$$RVLJ_t{=}min\{RJ_t, I_{[ZJ_t^{(BPV)}>\Phi_a]}\ (\sum_{i=1}^{T} r_{t,i}^2\ I_{|r_{t,i}|\geq\gamma})\}, \tag{14}$$

$$RVSJ_t{=}RJ_t - RVLJ_t, \tag{15}$$

where the selection of γ is data driven. Following Duong and Swanson (2011), we set $\gamma = 2$ in order to capture large size jumps. Note that RJ_t is already defined in Equation (8).

2.3.4 Realized skewness and kurtosis

As the final set of covariates, we compute realized skewness (*RSK*) and realized kurtosis (*RKU*) as measures of the higher-moments of the distribution of daily returns, computed from intra-day returns. Following Amaya et al. (2015), we consider *RSK* as a measure of the asymmetry of the daily return distribution and *RKU* as a measure that accounts for extreme occurrences. Following earlier studies, RSK_t, standardized by RV_t, and RKU_t are computed as

$$RSK_t{=}\frac{\sqrt{T}\sum_{i=1}^{T} r_{t,i}^3}{(\sum_{i=1}^{T} r_{t,i}^2)^{3/2}}, \tag{16}$$

$$RKU_t{=}\frac{T\sum_{i=1}^{T} r_{t,i}^4}{(\sum_{i=1}^{T} r_{t,i}^2)^2}. \tag{17}$$

The scaling of RSK_t and RKU_t by $\sqrt{T}$ and T, respectively, ensures that their magnitudes correspond to daily skewness and kurtosis.

3 Data

We use intraday data on West Texas Intermediate (WTI) oil futures that are traded on NYMEX over a 24 hour trading day (pit and electronic) to construct daily measures of standard realized volatility, and the corresponding good and bad variants, as well as the other intraday oil-based covariates described in Section 2.3. The futures price data, in continuous format, are obtained from `www.disktrading.com` and `www.kibot.com`. Close to expiration of a contract, the position is rolled over to the next available contract, provided that activity has increased. Daily returns are computed as the end of day (New York time) price difference (close to close). In the case of intraday returns, 1-minute prices are obtained via last-tick interpolation (if the price is not available at the 1-minute stamp, the previously available price is imputed), and 5-minute returns are then computed by taking the log-differences of these prices. Liu et al. (2015) note that 5-minute is a sampling frequency adequate for liquid assets like WTI futures. Thus, the sampling frequency used in our analysis is not too low to lead to poor data analysis and it is not too high to induce spurious jumps due to market frictions.

As far as the measure of risk aversion is concerned, we utilize the risk aversion index of Bekaert et al. (2017), which is available for download from: `https://www.nancyxu.net/risk-aversion-index`. This new measure of time-varying risk aversion can be calculated from observable financial information at high (daily) frequencies and relies on a set of six financial instruments, namely, the term spread, credit spread, a de-trended dividend yield, realized and risk-neutral equity return variance and realized corporate bond return variance. As discussed earlier, an important feature of this measure is that it distinguishes time variation in economic uncertainty (the amount of risk) from time variation in risk aversion (the price of risk) and, thus, provides an unbiased representation for time-varying risk aversion based on a utility function in the hyperbolic absolute risk aversion

(HARA) class. Based on the availability of the risk aversion data, the sample period is February 1997 to December 2016.

4 Empirical Findings

4.1 Results for the Standard HAR-RV Model

Table 1 presents the full-sample estimation results for the baseline HAR-RV model estimated by the ordinary-least-squares technique.[6] The findings show that the baseline HAR-RV model generally fits the data well with statistically significant estimated coefficients at all three forecasting horizons and the adjusted coefficient of determination taking on values in the range of 0.74−0.79. Interestingly, the estimated coefficient for risk aversion is significant with a positive value, indicating that higher level of risk aversion predicts a subsequent increase in realized volatility. Christoffersen and Pan (2017) note that an increase in economic activity lowers investors' relative risk aversion towards oil-price changes and large oil-price fluctuations are associated with deteriorating economic conditions. To that end, the observed positive predictive value of risk aversion over volatility may be driven by the real economic activity channel and possible comovement of risk aversion with changes in economic conditions. We also observe that the adjusted coefficient of determination increases only slightly for short forecast horizons when we add risk aversion to the model, while risk aversion makes a somewhat larger contribution to the overall fit of the model in the case of the medium and long forecast horizons.

– Please include Table 1 about here. –

[6] All estimation results documented in this research were computed using the R programming environment (R Core Team 2019). The Newey-West robust standard errors reported in Table 1 were computed using the R packages "sandwich" (Zeileis 2004).

Clearly, in-sample fit does not guarantee forecasting benefits in terms of out-of-sample performance. For this reason, we report in Table 2 the results of Diebold and Mariano (1995) tests to examine whether risk aversion helps to improve out-of-sample forecast accuracy. To this end, we first compute out-of-sample forecast errors for the three forecast horizons by means of a rolling-estimation window of length 2,350 observations, which corresponds to half of the sample size. We then derive the test results using the well-known modified Diebold-Mariano test proposed by Harvey et al. (1997).[7] We report p-values of the tests for (i) standard forecast errors, defined as actuals minus forecasts, and (ii) scaled forecast errors, computed by dividing the forecast error by actual realized volatility, in order to take into account heteroskedasticity (for such a heteroskedasticity adjustment, see e.g. Bollerslev and Ghysels 1996). Irrespective of whether we use scaled or unscaled forecast errors, the test results strongly suggest that risk aversion does not help to improve out-of-sample forecast accuracy within the linear HAR-RV framework. We find that the forecasts extracted from the standard HAR-RV model are significantly more accurate than the forecasts extracted from the HAR-RV-RA model, consistently across alternative loss functions based on the quadratic forecast errors and the absolute forecast errors as arguments. In short, our preliminary out-of-sample tests suggest that risk aversion does not provide any predictive value within the linear HAR-RV model.

– Please include Table 2 about here. –

4.2 Results for the HAR-RV-RF Model

Despite its popularity in the forecasting literature, as noted earlier, one of the shortcomings of the standard HAR-RV model is that it is based on a linear spec-

[7] We report p-values computed using the R package "forecast" (Hyndman 2017, Hyndman and Khandakar 2008).

ification which may lead to biased results, especially if nonlinearities are present in the links across the variables of interest. To that end, the findings reported in Table 2 may not fully reflect the potential impact of risk aversion on the accuracy of forecasts of realized volatility as changes in the level of risk aversion may have a differential impact on forecast accuracy.[8] For example, a visual examination of the daily risk aversion series reveals a sudden, large increase in the level of risk aversion during the financial crisis of 2007/2008, suggesting that risk aversion may be state dependent. Considering that the level of speculation is negatively correlated with the aggregate risk aversion of market participants (Li 2018) and that speculators require commodity positions in their portfolios for diversification purposes, particularly during periods of uncertainty (Hamilton and Wu 2014), one can argue that the informational content of high levels of risk aversion during times of crisis for subsequent realized volatility may differ from the informational content of low levels of risk aversion observed during normal times. To that end, the random forest technique offers an appropriate approach to account for such a possible nonlinearity in the predictive relationship between realized volatility and risk aversion.

As mentioned earlier, the HAR-RV-RF framework based on random forests does not only account for possible nonlinearity in the data-generating process, but also offers a natural modeling platform to capture the impact of several additional covariates of realized volatility other than risk aversion.[9] Following the previous works on realized volatility in different contexts, we use the following co-

[8]As a robustness check, we also estimate a HAR-RV-RA model using the natural logarithm of risk aversion. This model yields results (not reported, but available from the authors upon request) similar to those we report in Table 2.

[9]We use the R package "randomForest" to estimate random forests (see Liaw and Wiener, 2002). We set the number of trees at 500 and use #floor(total number of covariates/3) random covariates for splitting. The same values are also used to estimate the models described in Section 4.3.

variates in our forecasting model in addition to risk aversion (and the standard weekly and monthly realized volatility): measures of good and bad volatility, realized skewness and realized kurtosis, oil returns, a leverage effect variable defined as $\max(\text{returns}_t, 0)$, and various jump measures including asymmetric jumps, good/bad jumps and small/large jumps.

– Please include Figure 1 about here. –

Figure 1 depicts the marginal effect of risk aversion on realized volatility as measured by the so-called partial dependence function. The partial dependence function informs how the prediction of realized volatility changes when we vary risk aversion and hold all other predictors constant.[10] We obtain the marginal effect by combining 500 trees to estimate the HAR-RV-RF model over the whole sample where each tree has a maximum number of five terminal nodes. The marginal effect shows that higher values of risk aversion are associated with higher values of realized volatility, a pattern one would expect given that both realized volatility and risk aversion tend to be higher in times of financial stress than during tranquil times. Importantly, the estimated marginal effect exhibits a moderate slope for several lower decentiles of risk aversion, while the slope substantially rises for the largest decentiles. Clearly, the HAR-RV-RF model is able to successfully capture the nonlinearity resulting from this pattern in the data, whereas the linear HAR-RV-RA model overlooks this pattern and, therefore, draws an incomplete picture of the impact of risk aversion on forecast accuracy.

In the case of out-of-sample forecasts of realized volatility, we estimate the HAR-RV-RF model using the full list of covariates, with and without risk aversion, again using a rolling-estimation window length of 2,350 observations. This procedure

[10]For brevity, we only plot the marginal effect obtained for the short forecast horizon. The effects for the medium and long forecast horizons are similar (available upon request).

yields two time series of forecasts that we compare by means of the modified Diebold-Mariano test. Table 3 reports the test results for three different specifications of the hyperparameters of the HAR-RV-RF model by setting the maximum number of terminal tree nodes to 5 and 10, while the third specification fixes the minimum nodesize to 10 observations. As a standard procedure in the literature on random forests, we choose randomly one third of the covariates for tree building (that is, for splitting). The results reported in Table 3 are rather similar across the three specifications of the HAR-RV-RF model and lend strong support to the hypothesis that time varying risk aversion contains significant predictive value for subsequent realized volatility of crude oil returns. The tests yield significant results in case of the short and medium forecast horizon, with generally somewhat weaker significance levels observed in the case of unscaled forecast errors, and $L2$ loss. For the long forecast horizon, the test results are significant in two out of the three specifications (especially under $L1$ loss).

– Please include Table 4 about here. –

In order to further elaborate the predictive value of risk aversion, we report in Table 4 a ranking of the covariates according to their relative importance. Relative importance is measured by inspecting the increase in node purity (as measured by the residual sum of squares across all trees) that results from the inclusion of a variable to the list of predictors. The ranking of the covariates reported in Table 4 is based on the out-of-sample estimation results, that is, we report the mean rank of a covariate across all rolling-estimation windows. As one would have expected, daily realized volatility and its weekly and monthly counterparts (that is, the standard predictors used for estimating a HAR-RV model) are among the top-ranked predictors, followed by realized skewness and realized kurtosis (on the role of realized skewness and realized kurtosis for modeling and forecasting realized stock-market volatility, see Mei et al. 2017). We see that risk aversion has an average rank of about six, indicating its superior predictive ability compared to returns and leverage and, importantly, the various jump covariates.

At the same time, the low ranks observed for the jump covariates is not unexpected as we use a jump-robust estimator of realized volatility in our tests, following Andersen et al. (2012). Because the realized-volatility estimator employed is substantially less biased than other measures of realized volatility in the presence of jumps and/or market-microstructure noise, it is decoupled from the discontinuous and jump part of volatility (Giot et al. 2010). To that end, the low ranks obtained for the jump-based covariates indicate that random forests reliably identify the important covariates of realized volatility in our tests.

– Please include Table 5 about here. –

In Table 5, we summarize the pseudo R-squared of the HAR-RV-RF models averaged across all rolling-estimation windows. The pseudo R-squared uses the so called out-of-bag data (that is, the quasi-out-of-sample data not used for tree building when bootstrapping a random forest) to assess model performance (for details, see Liaw and Wiener 2002). We observe that the out-of-bag pseudo R-squared statistic of the HAR-RV-RF models is larger for the medium and long forecast horizons than for the short forecast horizon and when we increase the number of terminal nodes and fix the minimum terminal node size at 10 observations. In all cases, the out-of-bagpseudo R-squared values exceed the in-sample/full-sample adjusted R-squared values we report for the linear HAR-RV-RA models in Table 1.

– Please include Table 6 about here. –

In order to check the robustness of our findings with regard to the length of the rolling-estimation window, we summarize in Table 6 the results for alternative rolling-estimation windows. Specifically, we vary the length of the rolling estimation window in increments of 100 observations, starting with 1800 up to 2800, and estimate the models for every alternative window length. We see that the findings are robust to the size of the rolling-estimation window, suggesting that risk aversion carries significant predictive information over future realized volatility in the crude oil market, contributing to improved forecast accuracy.

– Please include Table 7 about here. –

Finally, we present as an alternative to the Diebold-Mariano test in Table 7 results for the test proposed by Clark and West (2007). We present results for three different specifications of the number of terminal nodes and the size of the terminal nodes, and for a rolling-estimation window of length 2350 observations. The key result is that all tests yield significant results, corroborating the results of the Diebold-Mariano tests.

4.3 Several Extensions

Having obtained strong results in favor of time-varying risk aversion as a predictor of crude oil-returns volatility, we next extend our analysis to study the quantiles of the conditional distribution of realized volatility in order to explore whether the predictive relationship observed is driven by the state of the market which can be represented via quantiles. To this end, we use quantile random forests and estimate HAR-RV-QRF models. A quantile random forest, proposed by Meinshausen (2006) as an important extension of the random-forest technique, does not only store the information on the mean of realized volatility at the leaves (as a conventional regression tree does), but rather keeps all observations of realized volatility in order to subsequently compute an estimate of its conditional distribution function. This distribution function is then used to inspect the incremental predictive value of risk aversion for realized volatility at the quantile level.[11] In a recent application of quantile random forests to stock-returns volatility forecasting, Gupta et al. (2019) provide a compact description of the technical details of quantile random forests.

– Please include Table 8 about here. –

Table 8 summarizes the results of Diebold-Mariano tests for the HAR-RV-QRF model, where we present results for two different specifications (maximum of five terminal nodes, Panel A; minimum ternominal node size fixed at ten observations, Panel B). Once again, we use the modified version of the Diebold-Mariano test along with scaled forecast errors to study the incremental value of risk aversion for forecast accuracy. In addition, we use the standard check function underlying analyses of quantile regressions in order to evaluate the forecast errors. In other words, we explicitly take into account the asymmetry of a forecaster's loss function based on the assumption that a forecaster targets quantiles of the conditional distribution of realized volatility rather than the median (on the role of asymmetry of a forecaster's loss function for forecasts of crude oil-returns volatiltiy; see Gkillas et al. (forthcoming)). The findings reported in Table 8 are generally in line with those obtained from the standard HAR-RV-RF model in that risk aversion helps to improve forecast accuracy at all three forecast horizons provided the quantile of the conditional distribution of realized volatility is not too small (especially in Panel B).

– Please include Table 9 about here. –

[11] We use the R package "quantregForest " (Meinshausen, 2016) to estimate the HAR-RV-QRF model.

As a second exension, we consider extreme random forests as developed by Geurts et al. (2006).[12] Extreme random forests further expand the advantages of standard random forecasts over bootstrap aggregation as described in Section 2.2 in that, given the list of randomly selected covariates that can be used for splitting at a node, a (uniform) random split is sampled for every covariate in this list (using the full sample rather than a bootstrap sample), and then the best split/covariate combination is selected among these randomly sampled splits. The results for the HAR-RV-ERF model, reported in Table 9, yield significant Diebold-Mariano test values for the short and medium forecast horizon and, thus, provide further support to the predictive power captured by risk aversion over realized volatility.

– Please include Table 10 about here. –

As a third extension, we repeat our analysis using an alternative measure of realized volatility rather than the *MRV* estimator used in previous tests in order to examine whether the results will hold when we a use volatility estimator that is less robust to jumps. For this purpose, we use $RV = RV_t^B + RV_t^G$ as an estimator of reailized volatility. Table 10 summarizes the results of Diebold-Mariano tests for this extension of the model. As can be seen in the table, all but one (for $h = 22$) test results are found to be significant, further supporting our earlier findings on risk aversion as a robust predictor of volatility.

Although not reported in the paper due to space considerations, we perform several other extensions for further robustness checks. In one extension, we study the predictive value of risk aversion for realized volatility at a monthly frequency and find that risk aversion does not help to improve forecast accuracy for monthly data, highlighting the benefits of using intraday data to study the predictive role of risk

[12] We use the R package "extraTrees" to estimate random forests (see Simm et al. 2014).

aversion for volatility forecasting. As an alternative model, we study a model that features a monthly variance risk premium as another covariate and, once again observe insignificant results for this variable. As yet another extension, we consider a model that features the first difference of risk aversion and obtain significant test results. Finally, using the speculative ratio of Chan et al. (2015), we compare the model that features risk aversion with a model that features a measure of speculative activity in the oil market.[13] The results show that adding risk aversion to the vector of predictors delivers significantly more accurate forecasts than adding the speculative ratio to the model. The results for these additional tests are available upon request.

5 Concluding Remarks

This paper introduces time-varying risk aversion as a predictor of realized volatility in the crude oil market. Using high-frequency data on WTI crude oil futures, we examine the predictive power of risk aversion in the presence of various covariates including realized skewness and realized kurtosis, various measures of jump intensity, and leverage that have often been employed in volatility-forecasting models. We show that nonlinearity in the predictive relationship between risk aversion and realized volatility plays an important role for the accuracy of forecasts of realized volatility. While the popular linear heterogeneous autoregressive realized volatility (HAR-RV) model fails to recognize the predictive power of risk aversion over crude oil-returns volatility, we find that risk aversion indeed improves forecast accuracy at all forecast horizons when we compute forecasts by means of random forests. While this finding for time-varying risk aversion

[13] The speculative ratio is computed as the trading volume divided by open interest for crude oil futures (data obtained from Commodity Systems Inc.).

is robust to a battery of extensions and variations in the model specification, we argue that the predictive value of time-varying risk aversion over realized crude oil-returns volatility is driven by the real economic activity channel and possible comovement of risk aversion with changes in economic conditions.

Overall, our study highlights the importance of nonlinearity in the detection of predictors to improve the accuracy of forecasting models as well as the predictive power of non-cashflow factors over commodity-market uncertainty. For future research, it will be interesting to use random forests and other machine-learning techniques in order to further explore the channels through which risk aversion serves as a predictor of realized volatility of crude oil returns and other commodities. Various alternative measures of speculative activity could also be used to further rule out the speculation channel in the risk aversion-volatility relationship.

References

Acharya, V.V. , Lochstoer, L.A. , and Ramadorai, T. (2013). Limits to arbitrage and hedging: evidence from commodity markets. *Journal of Financial Economics*, 109, 441−465.

Agnolucci, P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31, 316−321.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118: 135−167.

Andersen, T. G., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885−905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, 42−55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579−625.

Andersen, T. G., Bollerslev, T., and Diebold, F.X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89, 701−720.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. In L.P. Hansen and Y. Ait-Sahalia (Eds.), *Handbook of Financial Econometrics: Tools and Techniques*, 67−137, Elsevier, North-Holland.

Andersen, T.G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169, 75−93.

Arouri, M. E. H., Lahiani, A., Lévy, A., and Nguyen, D. K. (2012). Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Economics*, 34, 283−293.

Asai, M., Gupta, R., and McAleer, M. (2019). The impact of jumps and leverage in forecasting the co-volatility of oil and gold futures. *Energies*, 12, 3379.

Asai, M., Gupta, R., and McAleer, M. (Forthcoming) Forecasting volatility and co-volatility of crude oil and gold futures: effects of leverage, jumps, spillovers, and geopolitical risks. *International Journal of Forecasting*.

Aye, G.C., Dadam, V., Gupta, R., and Mamba, B. (2014). Oil price uncertainty and manufacturing production. *Energy Economics*, 43, 41−47.

Bekaert, G., Engstrom, E. C., and Xu, N. R. (2017). The time variation in risk appetite and uncertainty. Columbia Business School Research Paper No. 17-108.

Barndorff-Nielsen, O.E., and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4, 1−30.

Barndorff-Nielsen, O.E., Kinnebrouk, S., and Shephard, N. (2010). Measuring downside risk: realised semivariance. In T. Bollerslev, J. Russell and M. Watson (eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle*, 117-136. Oxford University Press.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123−140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5−32.

Bollerslev, T., and Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics*, 14, 139−151.

Bollerslev, T., Li, S. Z., and Zhao, B. (forthcoming). Good volatility, bad volatility and the cross-section of stock returns. *Journal of Financial and Quantitative Analysis*.

Chan, L. H. Nguyen, C. M, and Chan, K. C. (2015). A new approach to measure speculation in the oil futures market and some policy implications. *Energy Policy*, 86, 133−141.

Chatrath, A., Miao, H., Ramchander, S., and Wang, T. (2015). The forecasting efficacy of risk-neutral moments for crude oil volatility. *Journal of Forecasting*, 34, 177−190.

Chen, Y., Ma, F., and Zhang, Y. (2019). Good, bad cojumps and volatility forecasting: new evidence from crude oil and the U.S. stock markets. *Energy Economics*, 81, 52−62.

Chkili, W., Hammoudeh, S., and Nguyen, D. K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics*, 41, 1−18.

Clark, T.D., and West,K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138, 291−311.

Christoffersen, P., and Pan, X. (2017). Investor beliefs and state price densities in the crude oil market. Working paper. Tulane University.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174−196.

Degiannakis, S., and Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, 76, 28−49.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253−263.

Duong, D., and Swanson, N. R. (2011). Volatility in discrete and continuous-time models: A survey with new evidence on large and small jumps. *Missing Data Methods: Time-Series Methods and Applications*, 179−233. Emerald Group Publishing Limited.

Duong, D., and Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics*, 187, 606−621.

Efimova, O., and Serletis, A. (2014). Energy markets volatility modelling using GARCH. *Energy Economics*, 43, 264−273.

Elder, J., and Serletis, A. (2010). Oil price uncertainty. *Journal of Money, Credit and Banking*, 42, 1137−1159.

Engle, E., and Sun, Z. (2007). When is noise not noise? A microstructure estimate of realized volatility. NYU Working Paper No. FIN-07-047.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63, 3−42.

Giot, P., Laurent, S., and Petitjean, M. (2010). Trading activity, realized bolatility and jumps. *Journal of Empirical Finance*, 17, 168−175.

Gkillas, K., Gupta, R., and Pierdzioch, C. (Forthcoming). Forecasting realized oil-price volatility: The Role of financial stress and asymmetric loss. University of Pretoria.

Gupta, Rangan, Pierdzioch, C., Vivian, A. J., Wohar, M. E., (2019). The predictive value of inequality measures for stock returns: An analysis of long-span UK data using quantile random forests. *Finance Research Letters*, 29, 315-322.

Guo, H., Wang, K., and Zhou, H. (2019). Good Jumps, Bad Jumps, and Conditional Equity Premium. *Asian Finance Association (AsianFA) 2014 Conference Paper*, PBCSF-NIFR Research Paper No. 14-05.

Hamilton, J.D. , and Wu, J.C. (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance*, 42, 9−37.

Hamilton, J.D., and Wu, J.C. (2015). Effects of index-fund investing on commodity futures prices. *International Economic Review*, 56, 187−205.

Hansen, P. R., and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1,1)?. *Journal of Applied Econometrics*, 20, 873−889.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13, 281−291.

Hastie, T., Tibshirani, R., and Friedman, J. (2009) *The elements of statistical learning. Data mining, inference, and prediction*, Springer Science and Business Media, LLC.

Haugom, E., Langeland, H., Molnár, P., and Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking and Finance*, 47, 1−14.

Hong, H. , and Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, 105, 473−490 .

Hou, A., and Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. *Energy Economics*, 34, 618−626.

Hyndman, R.J. (2017). forecast: Forecasting functions for time series and linear models. R package version 8.0, URL: `http://github.com/robjhyndman/forecast`.

Hyndman, R.J., and Khandakar, Y (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26, 1−22.

Huang, X., and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3, 456−499.

Irwin, S. , and Sanders, D. (2011). Index funds, financialization, and commodity futures markets. *Appl. Econ. Perspect. Policy*, 33, 1−31.

Kang, S. H., Kang, S. M., and Yoon, S. M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31, 119−125.

Kang, S.-H., and Yoon, S.-M. (2013). Modelling and forecasting the volatility of petroleum futures prices. *Energy Economics*, 36, 354−362.

Kilian, L., and Hicks, B. (2013). Did unexpectedly strong economic growth cause the oil price shock of 2003−2008?. *Journal of Forecasting*, 32, 385−394.

Lemmon, M., and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, 19, 1499−1529.

Liu, J., Wei, Y., Ma, F., and Wahab, M.I.M. (2017). Forecasting the realized range-based volatility using dynamic model averaging approach. *Economic Modelling*, 61, 12−26.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic perspectives*, 18, 29−50.

Lux, T., Segnon, M., and Gupta, R. (2016). Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data. *Energy Economics*, 56, 117−133.

McAleer, M., and Medeiros, M. C. (2008). Realized volatility: a review. *Econometric Reviews*, 27, 10−45.

Li, B. (2018). Speculation, risk aversion, and risk premiums in the crude oil market. *Journal of Banking and Finance*, 95, 64−81.

Liaw, A., and Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2, 18−22.

Liu, J., Ma, F., Yang, K., and Zhang, Y. (2018). Forecasting the oil futures price volatility: large jumps and small jumps. *Energy Economics*, 72, 321−330.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187, 293−311.

Lyócsa, Š., and Molnár, P. (2016) Volatility forecasting of strategically linked commodity ETFs: gold-silver, *Quantitative Finance*, 16, 1809−1822.

Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help?. *Physica A*, 481, 153−159.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning*, 7, 983−999.

Meinshausen, N. (2016). quantregForest: Quantile regression forests. R package version 1.3-5. URL `https://CRAN.R-project.org/package=quantregForest`.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., and Pictet, O. V. (1997). Volatilities of different time resolutions − Analyzing the dynamics of market components. *Journal of Empirical Finance*, 4, 213−239.

Nomikos, N. K., and Pouliasis, P. K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics*, 33, 321−337.

Oomen, R. (2001). Using High Frequency Data to Calculate, Model and Forecast Realized Volatility. *Computing in Economics and Finance 2001 75*, Society for Computational Economics.

Patton, A. J., and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97, 683−697.

Phan, D. H. B., Sharma, S. S., and Narayan, P. K. (2016). Intraday volatility interaction between the crude oil and equity markets. *Journal of International Financial Markets, Institutions and Money* 40, 1−13.

Prokopczuk, M., Symeonidis, L., and Wese Simen, C. (2015). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets*, 36, 758−792.

Patton, A. J., and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97, 683−697.

Qadan, M., and Nama, H. (2018). Investor sentiment and the price of oil. *Energy Economics*, 69, 42−58.

R Core Team (2019). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/. R version 3.6.0.

Robe, M.A., and Wallen, J. (2016). Fundamentals, derivatives market information and oil price volatility. *Journal of Futures Markets*, 36, 317−344 .

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28, 467−488.

Sadorsky, P., and McKenzie, M. D. (2008). Power transformation models and volatility forecasting. *Journal of Forecasting*, 27, 587−606.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235, 643−659.

Simm J., de Abril I., and Sugiyama M. (2014). Tree-based ensemble multi-task learning method for classification and regression, *IEICE Transactions on In-*

formation and Systems, 97: 1677−1681. URL: `http://CRAN.R-project.org/package=extraTrees`.

Singleton, K.J. (2014). Investor flows and the 2008 boom/bust in oil prices. *Management Science*, 60, 300−318.

Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit and Banking*, 39−72.

Tay, A., Ting, C., Tse, Y. K., and Warachka, M. (2009). Using high-frequency transaction data to estimate the probability of informed trading. *Journal of Financial Econometrics*, 7, 288−311.

van Eyden, R., Difeto, M., Gupta, R., and Wohar, M.E. (2019). Oil price volatility and economic growth: Evidence from advanced OECD countries using over one century of data. *Applied Energy*, 233/234, 612−621.

Wang, Y., Ma, F., Wei, -Y., and Wu. C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking and Finance*, 64, 136−149.

Wei, Y., Wang, Y., and Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32, 1477−1484.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11, 1−17.

Table 1: Full-Sample Estimates (OLS)

			$h = 1$			
Model	Intercept	MRV	MRV_w	MRV_m	RA	Adj. R2
HAR-RV	3.2286	10.0487	5.2102	4.8466	–	0.7364
p-value	0.0012	0.0000	0.0001°	0.0001°	–	–
HAR-RV-RA	-1.9150	9.3804	5.3565	4.7750	2.9494	0.7389
p-value	0.0555	0.0001°	0.0001°	0.0001°	0.0032	–
			$h = 5$			
Model	Intercept	MRV	MRV_*w*	MRV_m	RA	Adj. R2
HAR-RV	2.4619	10.9463	3.7549	5.0119	–	0.7876
p-value	0.0138	0.0001°	0.0002	0.0001°	–	–
HAR-RV-RA	-1.7560	10.4366	3.7681	4.8818	2.6756	0.7919
p-value	0.0791	0.0001°	0.0002	0.0001°	0.0075	–
			$h = 22$			
Model	Intercept	MRV	MRV_w	MRV_m	RA	Adj. R2
HAR-RV	3.1752	9.3592	2.2664	2.9516	–	0.7386
p-value	0.0015	0.0001°	0.0234	0.0032	–	–
HAR-RV-RA	-1.8773	8.9473	2.8770	3.1814	5.2033	0.7547
p-value	0.0605	0.0001°	0.0040	0.0015	0.0001°	–

Note: Estimated coefficients are scaled by their estimated standard error. RA = Risj aversion. p-values are based on Newey-West standard errors. ° denotes a value smaller than 0.0001. Adj. R2 = adjusted coefficient of determination.

Table 2: Out-of-Sample Results (HAR-RV Model)

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.8096	0.6409	0.9626
$L2$	0.6228	0.6437	0.7545
	Scaled forecast error		
$L1$	1.0000	0.9999	0.9999
$L2$	1.0000	0.9999	0.9999

Note: p-values of Diebold-Mariano tests (HAR-RV vs. HAR-RV-RA forecasts) for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and is computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features the risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using a rolling-estimation window (2350 observations).

Table 3: Out-of-Sample Results (HAR-RV-RF Model)

Panel A: Maximum number of terminal nodes is 5

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0000	0.0000	0.0115
$L2$	0.0002	0.0144	0.1230
	Scaled forecast error		
$L1$	0.0000	0.0000	0.0001
$L2$	0.0000	0.0000	0.0011

Panel B: Maximum number of terminal nodes is 10

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0000	0.0000	0.0475
$L2$	0.0116	0.0253	0.0936
	Scaled forecast error		
$L1$	0.0000	0.0000	0.0003
$L2$	0.0000	0.0000	0.0009

Panel C: Minimum node size is 10

Loss function	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
$L1$	0.0010	0.0367	0.3717
$L2$	0.0040	0.0480	0.1417
	Scaled forecast error		
$L1$	0.0000	0.0000	0.4011
$L2$	0.0002	0.0000	0.2316

Note: p-values of Diebold-Mariano tests (HAR-RV-RF vs. HAR-RV-RF-RA forecasts) for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and is computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using random forests and a rolling-estimation window (2350 observations).

Table 4: Ranking of Covariates (HAR-RV-RF Model)

Covariate	$h=1$	$h=5$	$h=22$
MRV	1.83	2.20	3.13
MRV_w	2.65	1.82	1.19
MRV_m	4.94	3.88	3.09
Asymmetric jumps	13.33	13.85	13.50
Bad jumps	9.71	10.12	10.20
Good jumps	9.92	9.13	9.06
Large jumps	7.08	6.95	7.18
Small jumps	8.63	8.61	9.14
Returns	9.05	9.73	9.97
Bad realized volatility	2.00	2.91	2.98
Good realized volatility	3.81	4.19	4.73
Realized skewness	12.99	13.32	13.40
Realized kurtosis	12.84	12.28	11.96
Leverage	14.62	14.33	14.56
Risk aversion	6.61	6.67	5.90

Note: The mean rank of a covariate across all rolling-estimation windows (2350 observations) is computed based on its effect on node purity as measured by the residual sum of squares. The number of trees per random forest is 500, and the maximum number of terminal nodes is five.

Table 5: Pseudo R-Squared (HAR-RV-RF Model)

Specification	$h=1$	$h=5$	$h=22$
Max. no. of terminal nodes = 5	0.7468	0.7953	0.7636
Max. no. of terminal nodes = 10	0.7663	0.8183	0.7902
Min. terminal node size = 10	0.7737	0.8356	0.8302

Note: The pseudo R-squared is defined as $1 - MSE_{oob}/Var$, where MSE_{oob} denotes the mean-squared error that obtains when the estimated model is used to compute the out-of-bag data (that is, the data not used for tree building when bootstrapping a random forest) and Var is an estimate of the variance of the dependent variable. The number of trees was set to 500.

Table 6: Out-of-Sample Results (Alternative Rolling-Window Lengths)

Window	$h = 1$	$h = 5$	$h = 22$
	Unscaled forecast error		
1800	0.0000	0.0176	0.1120
1900	0.0049	0.0155	0.0883
2000	0.0014	0.0054	0.0911
2100	0.0000	0.0086	0.0878
2200	0.0001	0.0058	0.0965
2300	0.0000	0.0136	0.1155
2400	0.0005	0.0158	0.1198
2500	0.0001	0.0077	0.1227
2600	0.0001	0.0125	0.1044
2700	0.0277	0.1717	0.3586
2800	0.3130	0.2305	0.0167
	Scaled forecast error		
1800	0.0000	0.0000	0.0020
1900	0.0000	0.0000	0.0015
2000	0.0000	0.0000	0.0021
2100	0.0000	0.0000	0.0015
2200	0.0000	0.0000	0.0014
2300	0.0000	0.0000	0.0012
2400	0.0000	0.0000	0.0014
2500	0.0000	0.0000	0.0010
2600	0.0000	0.0000	0.0009
2700	0.0000	0.0000	0.0009
2800	0.0000	0.0000	0.0009

Note: p-values of Diebold-Mariano tests (HAR-RV-RF vs. HAR-RV-RF-RA forecasts) for three different forecast horizons. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the risk aversion are more accurate. Loss function: $L2$. The models were estimated using random forests consisting of 500 trees and a maximum of five terminal nodes. Window = length of the rolling-estimation window.

Table 7: Clark-West-Tests (HAR-RV-RF Model)

Specification	$h = 1$	$h = 5$	$h = 22$
Max. no. of terminal nodes = 5	2.6076	1.9376	1.6034
Max. no. of terminal nodes = 10	1.8527	1.7524	1.7491
Min. terminal node size = 10	2.6019	2.1078	1.8252

Note: t-statistics of the Clark-West tests are based on Newey-West standard erros. The number of trees was set to 500. Critical values (one-sided test) are: 1.282 (10%) and 1.645 (5%).

Table 8: Out-of-Sample Results (HAR-RV-QRF Model)

Panel A: Max. no. of terminal nodes = 5

Loss function	$L1$			$L2$		
Quantile	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
0.02	0.017	0.0324	0.0348	0.0346	0.0364	0.0673
0.1	0.000	0.0000	0.0000	0.0001	0.0003	0.0002
0.2	0.000	0.0000	0.0000	0.0000	0.0000	0.0002
0.3	0.000	0.0000	0.0000	0.0000	0.0000	0.0004
0.4	0.000	0.0000	0.0000	0.0000	0.0000	0.0010
0.5	0.000	0.0000	0.0000	0.0000	0.0000	0.0012
0.6	0.000	0.0000	0.0000	0.0000	0.0000	0.0021
0.7	0.000	0.0000	0.0000	0.0000	0.0000	0.0022
0.8	0.000	0.0000	0.0000	0.0000	0.0000	0.0013
0.9	0.000	0.0000	0.0000	0.0000	0.0000	0.0006
0.98	0.000	0.0000	0.0000	0.0000	0.0000	0.0010

Panel B: Min. terminal node size = 10

Loss function	$L1$			$L2$		
Quantile	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
0.02	0.2399	0.2636	0.2950	0.5462	0.5430	0.5382
0.10	0.2618	0.2650	0.2772	0.0895	0.0945	0.1153
0.20	0.0003	0.0002	0.0010	0.0070	0.0067	0.0134
0.30	0.0001	0.0000	0.0000	0.0010	0.0006	0.0013
0.40	0.0197	0.0260	0.0219	0.0212	0.0373	0.0338
0.50	0.0003	0.0005	0.0024	0.0001	0.0001	0.0002
0.60	0.0001	0.0001	0.0003	0.0099	0.0078	0.0072
0.70	0.0000	0.0000	0.0000	0.0190	0.0121	0.0104
0.80	0.0010	0.0011	0.0009	0.0032	0.0048	0.0061
0.90	0.1785	0.1761	0.1962	0.3848	0.3857	0.3671
0.98	0.2933	0.3221	0.3020	0.2224	0.2748	0.2764

Note: p-values of Diebold-Mariano tests (HAR-RV-QRF vs. HAR-RV-QRF-RA forecasts) for three different forecast horizons. The HAR-RV-QRF model features 500 trees and a maximum of five terminal nodes. The scaled forecast error accounts for heteroscedasticity in the data and is computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-QRF model that features the risk aversion are more accurate. Loss function: check function of the $L1$ (absolute error loss) or the $L2$ (squared error loss) type. The models were estimated using a rolling-estimation window (2350 observations).

Table 9: Out-of-Sample Results (HAR-RV-ERF Model)

Loss function	$h=1$	$h=5$	$h=22$
	Unscaled forecast error		
$L1$	0.0000	0.0330	0.2536
$L2$	0.0123	0.0302	0.0958
	Scaled forecast error		
$L1$	0.0000	0.0000	0.4463
$L2$	0.0000	0.0000	0.3143

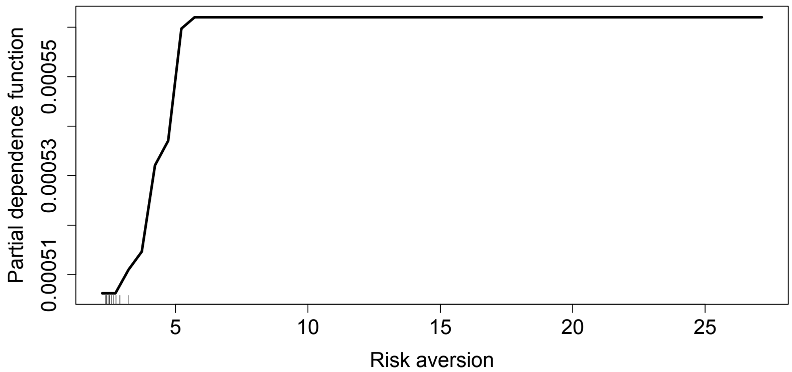
Note: p-values of Diebold-Mariano tests (HAR-RV-ERF vs. HAR-RV-ERF-RA forecasts) for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and is computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-ERF model that features the risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using extreme random forests (500 trees, minimum nodesize was set to ten) and a rolling-estimation window (2350 observations).

Table 10: Out-of-Sample Results (HAR-RV-RF Model, Alternative Measure of *RV*)

Specification	$h = 1$	$h = 5$	$h = 22$
Max. no. of terminal nodess = 5	0.0000	0.0000	0.0018
Max. no. of terminal nodess = 10	0.0000	0.0000	0.0010
Min. terminal node size = 10	0.0007	0.0001	0.2154

Note: p-values of Diebold-Mariano tests (HAR-RV-RF vs. HAR-RV-RF-RA forecasts) for three different forecast horizons. The scaled forecast error accounts for heteroscedasticity in the data and is computed as (actual - forecast)/actual. Null hypothesis: the series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV-RF model that features the risk aversion are more accurate. Loss function: *L*2. The models were estimated using random forests consisting of 500 trees and a maximum of five terminal nodes.

Figure 1: Marginal Effect (HAR-RV-RF Model)



Note: The full-sample estimated marginal effect of risk aversion on one-day-ahead realized volatility implied by a random-forest model consisting of 500 trees. The maximum number of terminal nodes is five. The rugs on the horizontal axis show the decentiles of risk aversion.