Forecasting realized volatility of crude oil futures prices based on machine learning

Received: 24 September 2022 | Revised: 25 January 2024 | Accepted: 25 January 2024
DOI: 10.1002/for.3077

RESEARCH ARTICLE

WILEY

Forecasting realized volatility of crude oil futures prices based on machine learning

Jiawen Luo[1] | **Tony Klein**[2,3] | **Thomas Walther**[4,5] | **Qiang Ji**[6]

[1]School of Business Administration, South China University of Technology, Guangzhou, China
[2]Faculty of Business and Economics, Technische Universität Chemnitz, Chemnitz, Germany
[3]Queen's Business School, Queen's University, Belfast, United Kingdom of Great Britain and Northern Ireland
[4]Utrecht School of Economics, University Utrech, Utrecht, the Netherlands
[5]Faculty of Business and Economics, Technische Universität Dresden, Dresden, Germany
[6]Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China

Correspondence
Thomas Walther, Utrecht School of Economics, University Utrecht, PO Box 80125, 3508TC Utrecht, the Netherlands.
Email: t.walther@uu.nl

Funding information
National Natural Foundation of China, Grant/Award Numbers: 72171088, 71803049; Natural Foundation of Guangdong Province, Grant/Award Number: 2023A1515012527; Guangdong Philosophy and Social Science Foundation, Grant/Award Number: GD23CGL01; Guangzhou Municipal Science and Technology Bureau, Grant/Award Number: 2023A04J1298

Abstract
Extending the popular HAR model with additional information channels to forecast realized volatility of WTI futures prices, we show that machine learning-generated forecasts provide better forecasting quality and that portfolios that are constructed with these forecasts outperform their competing models resulting in economic gains. Analyzing the selection process, we show that information channels vary across forecasting horizon. Variable selection produces clusters and provides evidence that there are structural changes with regard to the significance of information channels.

KEYWORDS
crude oil, exogenous predictors, forecasting, machine learning, realized volatility

1 | INTRODUCTION

As a tangible quantification of uncertainty, the measuring, understanding, and modeling of volatility is a cornerstone of comprehending commodity markets in their functionality, price discovery, and structure of connectedness to other markets. Commodity markets are inevitably linked to the economic activity and well-being of global and local economies as they embody the backbone for supply and demand of primary materials and feedstock for production, biomaterial for agricultural and food industries, and energy commodities among others. Energy commodity markets in particular play an important role in research as their understanding is of utmost importance to pricing, security, and planning of supply, as well as policy design and implications. The link between oil prices and economic growth is well documented and discussed in literature (Charfeddine et al., 2020; Hamilton, 2003; Kilian, 2009). In this study, we present novel evidence on the impact of exogenous information channels on the realized volatility of crude oil futures prices. Building on the popular heterogeneous autoregressive (HAR) model by Corsi (2009), we show that models incorporating variables that map the equity channel, sentiment, or foreign exchange (FX) market volatility result in more accurate predictions of realized volatility than the baseline HAR. In addition, we find that machine learning (ML) variants of the HAR model outperform baseline HAR models by selecting the most useful exogenous predictors from the set of different information channels; both in view of forecasting performance and economic gain. We also analyze in detail the time variation and importance of variables to predict oil price volatility. When comparing the variable selection across forecasting horizons, we find a positive relation between forecasting horizon and number of different information channels. Looking at the variable importance of those channels, we find particularly FX to adds to the information contained in past realized volatility.

We contribute to the literature on the connectedness of crude oil markets. It is argued that there exists an information transmission between the volatility of markets leading to co-movement of the global oil market (Kaufmann & Banerjee, 2014; Luo et al., 2023; Reboredo, 2011). In addition, WTI is taking a leading role as global benchmark, and other energy markets such as Brent crude oil follow these movements (Elder et al., 2014; Klein, 2018). Adding to the complexity of this research, studies show that exogenous factors and information channels that cannot be directly associated with energy markets are also affecting price and volatility behavior (e.g., Nguyen & Walther, 2020). Notably, other information channels, such as equity markets (Degiannakis & Filis, 2017; Haugom et al., 2014; Luo et al., 2020) or Google search volume (Afkhami et al., 2017), are shown to provide valuable information to forecast oil volatility with extended HAR models.

In contrast to the extent literature, we test the impact of volatility of different equity markets (e.g., Dow Jones Industrial, the CAC40, and the S&P 500, commodity futures markets(e.g., Natural Gas and Gold), FX markets, sentiment, volatility indicators (e.g., OVX and VIX), and, Google trends (e.g., oil price-related search terms) on realized volatility of the WTI crude oil. As we utilize an overall number of 28 exogenous channels from five categories resulting in a total of 84 exogenous HAR factors, model parameter parsimony is—evidently—forgone. We extend the baseline HAR model with each category of exogenous variables separately. Then, we introduce a "kitchen sink" approach to obtain the last HAR extension in which we simultaneously include all exogenous variables. Hence, the number of parameters in each HAR extension ranges from 16 to 31 for separated channels and 88 for the model incorporating all channels since each channel element is included in daily, weekly, and monthly averages. These highly parameterized models are predestined to be combined with ML methods to filter for the most relevant contributing factors as well as to overcome multicollinearity within and across information channels. Additionally, we are tracking the selection of variables across different forecasting horizons for inference on possibly time-varying impact of equity and sentiment channels on forecasts of realized volatility of crude oil futures prices.

In addition, we contribute to the literature on modeling and forecasting the (realized) volatility of oil prices. It is well documented that crude oil volatility behavior is characterized by its time-varying nature, including structural breaks or other time-varying effects (Fattouh, 2010; Fong & See, 2002; Klein & Walther, 2016; Nomikos & Pouliasis, 2011). This motivates the adaption of models covering these properties. In contrast, some studies show that more sophisticated HAR models based on stylized volatility facts do not necessarily outperform HAR version without extensions or exogenous factors (Degiannakis et al., 2022; Prokopczuk et al., 2016; Sévi, 2014). Related to our paper, Ma et al. (2018) and Zhang, Wei, et al. (2019) show that ML improves volatility forecasts for crude oil prices using a large set of exogenous predictors. We extend these studies by comparing various popular ML-based variable selection methods on large predictor sets for crude oil forecasts and their economic gains. These methods include the widely applied LASSO and Bayesian Model Averaging (BMA) as well as novel approaches to HAR models with Bootstrap Aggregation and Bagging (BAG) and Stochastic Search Variable Selection (SSVS). In this regard, we also contribute to the general literature of modeling and forecasting with HAR models.

The remainder of the paper is structured as follows. Section 2 introduces HAR models, their extensions, and the variable or adaptive selection methodologies of the study. Section 3 describes the data sets and presents some preliminary analysis of exogenous information channels. Section 4 reports and discusses the results of the empirical study, and Section 5 concludes.

2 | MODEL AND METHODOLOGY

2.1 | Heterogeneous autoregressive volatility models with adaptable predictors

In order to measure time-varying daily volatility based on the available high-frequency data sets, we calculate the realized volatility RV_t of day t as

$$RV_t = \sum_{i=1}^{m} r_{t,i}^2, \tag{1}$$

where m is the number of observed intraday returns $r_{t,i}$ with $i=1,\ldots,m$. This ith intraday return of day t is calculated as logarithmic difference of intraday prices as

$$r_{t,i} = 100 \times (\log P_{t,i} - \log P_{t,i-1}),$$

where $P_{t,i}$ and $P_{t,i-1}$ are two consecutive intraday prices. We choose to sample prices at 5-min intervals.[1]

We follow the approach of Corsi (2009) and exploit the autoregressive character of daily realized volatility by cascading its memory structure in a short-, medium-, and long-term component that yields the HAR model for realized volatility. Our baseline HAR model uses the specifications outlined in Corsi and Renò (2012). Here, we run a different regression model for each forecasting horizon h. This way, long-term forecasts do not have to rely on the relative weights for 1-day-ahead predictions for $h>1$ (Ederington & Guan, 2010). In particular, we define

$$\log RV_{t+h}^{(h)} = \frac{1}{h}\sum_{j=1}^{h} \log RV_{t+h-j+1} \quad \text{and} \tag{2}$$

$$\log RV_{t}^{(h)} = \frac{1}{h}\sum_{j=1}^{h} \log RV_{t-j+1}, \tag{3}$$

where $h \in \{1,\ldots,22\}$ denotes the days-ahead forecasting horizons. We distinguish $\log RV_{t+h}^{(h)}$, which is the average realized volatility for time $t+1$ to $t+h$, from $\log RV_t^{(h)}$, which is the average realized volatility for time $t-h+1$ to t. Lastly, $\log RV_t$ is the realized volatility at time t and equivalent to $\log RV_t^{(1)}$.

Our HAR baseline model reads as

$$\log RV_{t+h}^{(h)} = \beta_0 + \beta_1 \log RV_t + \beta_2 \log RV_t^{(5)} + \beta_3 \log RV_t^{(22)} + u_{t+h}^{(h)}, \tag{4}$$

where β_0, β_1, β_2, and β_3 are real-valued coefficients corresponding to the intercept and the short-, medium-, and long-term autoregressive impact of realized volatility and u is an i.i.d. error term with zero mean. In the standard HAR defined in Equation (4), RV_t then refers to the realized volatility of the previous day, $RV_t^{(5)}$ refers to the average realized volatility over the past 5 days, and $RV_t^{(22)}$ refers to the 22-day average.

As outlined in the introduction, there exist numerous extensions to the HAR model, in particular in view of modeling realized volatility of commodities.[2] In this research, we focus on an extension of the HAR model framework with exogenous factors following the notion of Zhang, Ma, and Wang (2019) and Luo et al. (2022).

We consider five classes of exogenous predictors that might impact the volatility behavior of crude oil futures prices. The first class includes equity market volatilities of eight major economies. The second class refers to market sentiment and uncertainty measures. The third and fourth classes of exogenous variables comprise realized volatility of related commodity futures—with a documented economic link to crude oil futures—and foreign currency (FX) markets, respectively. The last class includes a constructed sentiment index based on the popularity of different Google search queries. The details on these exogenous predictors are described in detail in the data description in Section 3.

We extend the basic HAR of Equation (4) with lagged daily, weekly, and monthly aggregates of exogenous variables $X_{k,t}$ as follows:

$$\begin{aligned}\log RV_{t+h}^{(h)} = {} & \beta_0 + \beta_1 \log RV_t + \beta_2 \log RV_t^{(5)} + \beta_3 \log RV_t^{(22)} \\ & + \sum_{k=1}^{N_C} \alpha_{1,k} X_{k,t}^{C} + \sum_{k=1}^{N_C} \alpha_{2,k} X_{k,t}^{C,(5)} \\ & + \sum_{k=1}^{N_C} \alpha_{3,k} X_{k,t}^{C,(22)} + u_{t+h}^{(h)}.\end{aligned} \tag{5}$$

where N_C refers to the number of exogenous predictors from category C and $\alpha_{1,k}$, $\alpha_{2,k}$, and $\alpha_{3,k}$ for $k=1,\ldots,N_1$ are the corresponding coefficients of the daily, weekly, and monthly averages or aggregates of the kth component of exogenous predictor category C.

Using model Equation (5) in combination with our predictor classes, we obtain five different models, which we label HAR-Equity, HAR-Sentiment, HAR-Commodity, HAR-FX, and HAR-Google. For HAR-Equity, we use realized volatilities sampled at a matching 5-min interval. We employ squared daily returns of commodity futures and FX rates as well as the US dollar index to proxy volatilities of commodity futures and FX markets. The weekly and monthly volatilities of commodity futures and FX markets are defined as $r_{i,t}^{2,(5)}$ and $r_{i,t}^{2,(22)}$, which are computed in an identical way as the weekly and monthly realized volatilities.

In addition to the aforementioned extensions of the baseline HAR model using different asset classes and information channels, we employ a model containing *all* exogenous factors of this study, which we label HAR-All. This HAR-All contains 84 exogenous parameters, three endogenous parameters, and the intercept. As we use different ML approaches for exogenous variable selection, the HAR-All poses an additional opportunity to compare these differing techniques in terms of forecasting performance and economic valuation. For further robustness of our findings, we also employ a model class using the information from the first principal component of each asset/information channel for each time horizon. This model class is labeled HAR-PCA.

In order to have a very parsimonious benchmark, we also include an autoregressive (AR) model with lag 1. Doing so allows us to assess to which extent additional and aggregated information adds predictive power over the information already present in the last observed realized volatility.

Finally, we also add an augmented AR(1) model, AR(1)-X, which includes the additional information sets similar to the extended HAR in Equation (5).

2.2 | HAR models with variable selection approaches

There exist some popular approaches in literature with regard to selecting predictors and shrinking variable dimensions. Our motivation for implementing a range of these methodologies is twofold. First, information processing is computationally costly, and the inclusion of insignificant predictors leads to an increasing need of information processing capacity. Investors and researchers alike prefer to focus on relevant information channels via variable selection schemes to simplify economic models (Gabaix, 2014; Luo & Young, 2016; Zhang, Ma, & Wang, 2019). Second, many exogenous and powerful predictors, such as macroeconomic and financial market indicators, are highly correlated, which can lead to an over-fitting problem and overshadowing of significant parameters through present multicollinearity. With variable selection methods, only the most relevant variables are selected as predictors, thus reducing the risk of over-fitting issues and improving the forecast accuracy (Campbell & Slack, 2008; Korobilis, 2017; Korobilis & Koop, 2020). Analyzing the systematic selection of variables also allows us to identify breakpoints offering additional insight on the variability of the statistical and economical benefit of including exogenous predictors in HAR frameworks.

As outlined in the introduction, we construct a set of competing HAR models based on several variable selection approaches, including the well-known LASSO, BMA, BAG, and SSVS. These approaches can be further categorized as Bayesian algorithms that include predictor selection with BMA and SSVS approaches and ML with LASSO and BAG falling in this category. We briefly introduce the selection approaches and use the abbreviation for the selection method as prefix to the respective HAR variant in what follows.

For the LASSO-HAR, we employ the R-package "glmnet" of Friedman et al. (2010) to compute and select the coefficients of the HAR models and its extensions introduced above. The coefficients are obtained by solving

$$\widehat{\beta}^{LASSO} = \arg\min_{\beta_0,\gamma}\left(\frac{1}{2T}\sum_{t=1}^{T}\left(y_{i,t}-\beta_0-Z_{t-1}\beta\right)^2+\lambda\sum_{j=1}^{3K}\left|\beta_j\right|\right),$$

where Z_{t-1} is a K-dimensional column vector of all the predictor variables in the different HAR models above and β is a K-dimensional row-vector with the respective coefficients. The non-negative regularization or penalty parameter λ is selected given the minimum mean cross-validated error. Eventually, $\widehat{\beta}^{LASSO}$ is the K-dimensional vector of the estimated coefficients from the LASSO regression.

The BMA tackles the problem of model uncertainty and model selection from a Bayesian perspective. Over a set of models $\mathcal{M}_i$ for $i=1,...,2^K$ (i.e., all combinations of K predictors), the BMA determines the posterior probability of each model. The approach then averages the models using their posterior probability. For the BMA-HAR model, we follow the setting of Fernández et al. (2001). We specify the g-prior with $g=1/K^2$. Markov-chain Monte Carlo methods are used for the estimation of the BMA-HAR models according to Madigan et al. (1995) and Dellaportas et al. (2002).

The BAG approach reduces the variance of forecasts by averaging the randomness of variable selection and has been applied widely in the areas of macro and financial forecasting (Audrino & Medeiros, 2011; Huang & Lee, 2010; Inoue & Kilian, 2008; Lee & Yang, 2006; Stock & Watson, 2012). With bagging, the bootstrap method is employed to generate a large amount of samples from the original data to re-evaluate the selection of predictors and to produce the bootstrap replications of forecasts and the forecast realized volatilities are obtained by averaging the forecasting values based on the bootstrap samples. Particularly, we use the same setting and critical values according to Inoue and Kilian (2008) and Ribeiro (2016). The optimal predictor set is data-dependent in the sense of the pre-tests in the extended linear HAR model above. The bagging approach is proceeded as follows:

- Arrange the set of predictors Z_t, $t=2,...,T$, in the form of a matrix M of dimension $(T-1)\times K$, where K is the number of predictors in Z_t.
- Create bootstrap samples of the form $\left\{Z'^{*}_{(i)2},...,Z'^{*}_{(i)T}\right\}$, for $i=1,...,B$, by drawing blocks of m rows of M with replacement, where the block size m is selected to capture the possible dependence in the error term of the realized volatility for oil futures. Based on Ribeiro (2016), we choose $B=100$ replications and a block size of $m=\lfloor T^{1/4}\rfloor$ for the moving block bootstrap procedure.
- Re-estimate the model with the B replicative bootstrap data of the selected predictors, and then use the estimated parameter as well as the original data of predictors to forecast the realized volatility.
- The final forecast is derived by averaging these bootstrapped forecasts.

For the SSVS-HAR, we use the full hierarchical specification of the SSVS prior according to George and McCulloch (1993) and in a slightly different context according to Korobilis (2013) as follows:

$$\begin{aligned}
&p(\beta_i|\xi_i)\sim(1-\xi_i)N\left(0,\tau_0^2\right)+\xi_i N\left(0,\tau_1^2\right),\\
&P(\xi_i|\pi)\sim Bernoulli(\pi_i).
\end{aligned}$$

These priors are specified with $\pi=0.5$, $\tau_0=0.001$, and $\tau_1=4$. Given these priors, the posterior results of the parameters β,ξ, and σ^2 are sampled with the Gibbs sampling method based on their conditional posterior distributions.

Although LASSO and the BAG approaches reduce dimensionality of the predictor set, the BMA and SSVS use a weighting of predictors to select and emphasize the most important variables. The above selection methods are applied to Equation (4) and the information class-model combinations of (5). This includes the baseline HAR model and necessarily the HAR-RV components in each HAR extension to allow for a deselection of endogenous variance measures.

2.3 | Forecast evaluation

We measure the out-of-sample performance for three different forecast horizons, that is, short-term (1 day ahead, $h=1$), medium-term (1 week ahead, $h=5$), and long-term (1 month ahead, $h=22$). From the full data set, we use 2/3 as in-sample or initial training set and use the remainder as out-of-sample set.

The 1-day-ahead forecasts are obtained by re-estimating the models on an expanding training set. For the 5- and 22-day-ahead forecasts, we follow Corsi (2009) and Corsi and Renò (2012) and estimate the aggregated or averaged realized volatility. Alternatively, one could also forecast h-steps ahead by iterating the model (Marcellino, 2006). However, this forecast method is seen to be less favorable given a propagation of forecast errors due to iterating 1-day-ahead forecasts (Ederington & Guan, 2010; Sizova, 2011).

In what follows, we evaluate the set of combinations of the HAR model with a variable selection approach and different sets of exogenous variables presented in Section 2, both with respect to forecast precision and economic significance.

2.3.1 | Statistical evaluation

We evaluate the forecast accuracy by the mean squared error (MSE) and mean absolute error (MAE) loss functions that read

$$MSE=T^{-1}\sum_{t=1}^{T}\left(RV_t-\widehat{RV}_t\right)^2,$$

$$MAE=T^{-1}\sum_{t=1}^{T}\left|RV_t-\widehat{RV}_t\right|,$$

where RV_t is the actual realized variance and $\widehat{RV}_t$ is the forecasted realized variance. Here, T refers to the out-of-sample number of observations. We then statistically test whether a model outperforms its peers, by employing the Model Confidence Set (MCS) by Hansen et al. (2011).[3]

In addition, we use the out-of-sample R^2_{OS} to measure the proportion of variance explained by the forecasts and to overcome the scale-depended drawbacks for the loss functions above (Blair et al., 2001; Campbell & Slack, 2008). The R^2_{OS} is computed as

$$R^2_{\text{OS}}=1-\frac{\sum_{t=1}^{T}\left(\widehat{RV}_t-RV_t\right)^2}{\sum_{t=1}^{T}\left(RV_t-\overline{RV}\right)^2},$$

and can be compared with the well-known in-sample R^2.

2.3.2 | Economic evaluation

In order to assess the economic relevance of the forecasts, we calculate the economic gain of the variable selection

TABLE 1 Summary statistics and preliminary analyses of all variables with $n = 2596$ observations spanning an observation period from January 5, 2010, to May 11, 2018.

	Mean	StD	Skewness	Kurtosis	LB $Q(5)$	LB $Q(10)$	LB $Q(22)$	ADF
Panel A: Oil price volatility								
WTI	1.6437	2.1386	5.9846	72.07	1328.90***	2154.52***	3902.56***	−22.79***
Panel B: Stock market volatility								
DJI	0.6906	1.7313	20.2846	638.60	677.82***	915.44***	1172.76***	−28.99***
CAC40	1.0420	1.4011	6.4217	68.54	3007.06***	4644.45***	6735.51***	−16.00***
FTSE	0.8330	1.7036	19.2349	584.26	629.27***	937.99***	1313.33***	−29.36***
DAX30	1.0383	1.4757	6.4259	68.48	3390.22***	5270.25***	8212.72***	−15.70***
HSI	0.5594	0.7255	8.4951	109.85	968.97***	1338.56***	1836.23***	−21.77***
NIK225	0.7241	1.4659	9.6691	136.76	830.84***	1125.37***	1358.30***	−25.09***
S&P500	0.6808	1.4255	11.5266	232.55	1476.86***	2132.20***	3010.69***	−23.51***
SSEC	1.3308	2.8446	7.5797	79.72	3647.71***	5241.38***	8480.71***	−18.81***
ESTOXX	1.2290	1.9335	12.4211	283.72	1456.65***	2210.27***	3167.91***	−22.29***
Panel C: Commodity market volatility								
Natural gas	7.3079	15.1738	6.7816	78.48	101.62***	150.39***	245.00***	−35.78***
Gold	1.0719	3.2044	15.7551	399.87	43.02***	66.20***	124.05***	−37.66***
Corn	3.0175	14.3666	34.4336	1416.36	1.42	3.87	5.37	−43.39***
Soybeans	1.9011	5.9385	16.4822	402.11	12.74**	25.52***	70.26***	−41.38***
Panel D: Currency market volatility								
USD Ind	0.2154	0.4035	5.0454	43.70	39.98***	75.60***	142.44***	−34.35***
EUR/USD	0.3441	0.6227	3.8660	23.73	53.34***	110.52***	221.31***	−34.23***
YEN/USD	0.3815	0.9172	6.8758	67.78	29.96***	52.48***	101.66***	−36.96***
YUAN/USD	0.0218	0.0992	20.0454	577.77	152.56***	181.70***	209.26***	−34.38***
RUB/USD	1.1339	7.4884	25.6058	751.43	872.65***	1046.20***	1174.03***	−23.73***
GBP/USD	0.3301	1.6039	37.3447	1585.73	101.01***	111.08***	115.10***	−37.47***
Panel E: Sentiment variables								
GFSI	86.6074	8.2043	0.3545	1.65	10590.41***	21029.28***	45525.06***	0.67
VIX	17.1377	5.9161	1.6643	6.36	8783.62***	15907.39***	29015.25***	−2.18**
OVX	33.3028	10.6372	0.8036	3.78	9888.71***	18927.99***	37680.58***	−1.29
EPU	105.9849	62.8497	1.5436	6.41	3500.21***	6109.56***	11052.43***	−10.19***
Panel F: Google search volume								
G1: "Oil Production"	16.8724	11.5640	1.9246	10.54	2415.77***	4227.39***	7041.84***	−13.36***
G2: "Financial Crisis"	28.0232	12.8161	1.1214	4.89	7446.00***	13463.94***	22717.45***	−4.96***
G3: "Oil Demand"	12.2505	13.1789	1.3274	5.76	770.13***	1373.19***	2242.12***	−22.37***
G4: "Oil Price"	14.8950	13.7390	2.0530	8.77	9068.64***	16746.01***	31571.58***	−4.26***
G5: "OPEC Conference"	2.8353	7.9100	4.6590	36.03	171.30***	287.13***	434.10***	−35.15***

**Statistical significance at 5% level.
***Statistical significance at 1% level.

approaches over the standard AR(1) models. Particularly, we consider an investor who has the mean–variance utility:

$$U(R_p) = E(R_p) - \frac{1}{2}\gamma Var(R_p)$$

where γ is the risk-aversion rate. The return of the investor's portfolio is denoted R_p, $E(R_p)$ is the expected portfolio return, and $Var(R_p)$ is the portfolio variance.

The risk-averse investor allocates a budget to a portfolio comprising a risky asset and a risk-free asset. Following Campbell and Thompson (2008), Ferreira and

Santa-Clara (2011), and Neely et al. (2014), the optimal portfolio weights, based on information at time t for allocations at time $t+1$, should be

$$\widehat{w}_t = \frac{1}{\gamma}\left(\frac{\widehat{r}_{t+1}}{\widehat{\sigma}^2_{t+1}}\right),$$

where $\widehat{r}_{t+1}$ is the forecasted excess return of the oil futures based on a moving average with a 1-year (256 trading days) rolling window. The forecasted variance $\widehat{\sigma}^2_{t+1}$ is derived from the various combinations of the AR and HAR models and variable selection methods presented in Section 2. In particular, we calculate the portfolio weights at the end of each month, using the predicted next month's variance from the various 22-day-ahead models. We repeat the process until the end of our out-of-sample period. Then, the portfolio return at time $t+1$ is given by

$$R_{p,t+1} = \widehat{w}_t r_{t+1} + R^f_{t+1},$$

where r_{t+1} and R^f_{t+1} are the excess return of the WTI futures and the risk-free return of a 5-year US treasury note, respectively.

To assess the economic value, we follow Fleming et al. (2001) and Fleming et al. (2003). Based on the quadratic utility

$$U(r_{p,t},\gamma) = (1+r_{p,t}) - \frac{\gamma}{2(1+\gamma)}(1+r_{p,t})^2$$

with risk aversion γ and portfolio excess return $r_{p,t}$, we determine the economic value with the constant Δ between two portfolios such that

$$\sum_{t=T_1+1}^{T} U(r_{p_1,t}) = \sum_{t=T_1+1}^{T} U(r_{p_2,t} - \Delta).$$

The greater Δ is, the more return a risk-averse investor is willing to sacrifice to switch from portfolio p_1, representing a benchmark AR(1) model, to p_2 being a HAR/variable selection combination. We follow the literature and use two levels of risk aversion rate for the investor, the mild risk aversion rate $\gamma = 1$ and the strong risk aversion rate $\gamma = 10$.

3 | SAMPLE AND DATA

In this study, we make use of high-frequency intraday data of WTI futures in 5-min resolution. This data is obtained from the Tick Data Database. We choose the most liquid crude oil futures contract instead of a fixed maturity to avoid any price and volatility distortion due to rollover processes.

Following Degiannakis and Filis (2017), we include five different classes of exogenous factors. The first group referring to equity market volatility comprises indices for the USA (S&P 500 and DJI), Germany (DAX30), France (CAC40), UK (FTSE), Hong Kong (HSI), Japan (Nikkei225), and China (SSEC), as well as the European Stock index (STOXX50). The second group includes

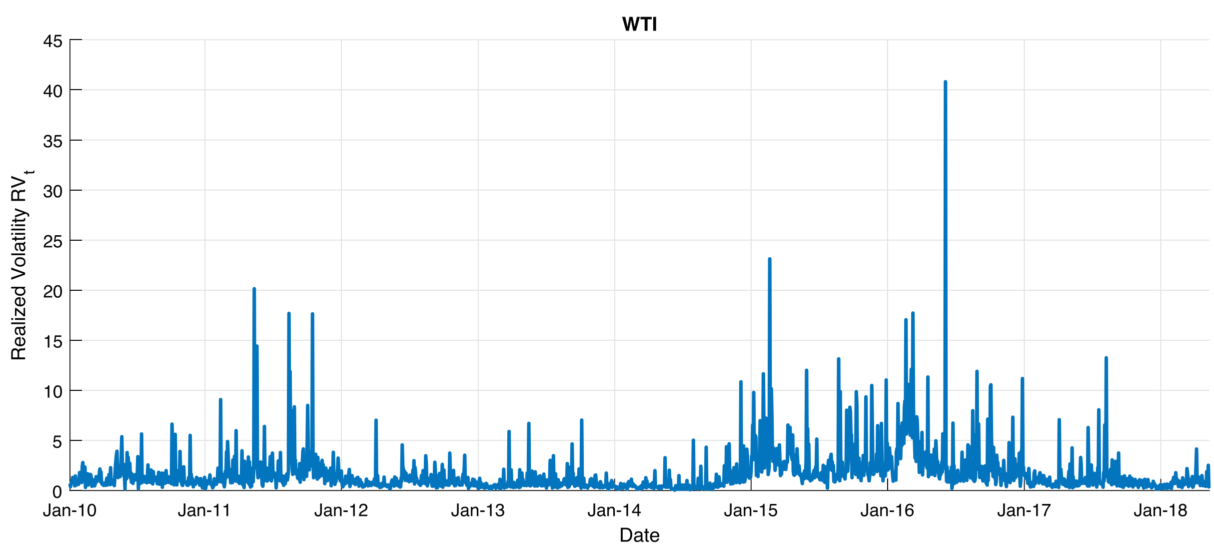

FIGURE 1 Realized volatility of WTI futures prices on daily resolution from January 5, 2010, to May 11, 2018, calculated based on Equation (1).

TABLE 2 Statistical evaluations of various volatility models for one-step forecasts.

Model	MSE		MAE		R^2_{OS}
AR(1)	3.3568	(0.1558)	1.1461	(0.0125)	0.5492
HAR	2.7912	(0.5952)	1.0142	(0.0938)	0.6251
LASSO-HAR	3.0235	(0.3811)	1.0271	(0.0938)	0.5939
BMA-HAR	2.9122	(0.3811)	1.0088	(0.2217)	0.6089
BAG-HAR	2.7914	(0.5952)	1.0142	(0.0938)	0.6251
SSVS-HAR	2.8062	(0.3811)	1.0075	(0.0938)	0.6231
AR(1)-Equity	3.8988	(0.1558)	1.1875	(0.0938)	0.4764
HAR-Equity	2.9388	(0.3811)	1.0413	(0.0938)	0.6053
LASSO-HAR-Equity	2.9864	(0.3811)	1.0215	(0.0938)	0.5989
BMA-HAR-Equity	5.5407	(0.3811)	1.0915	(0.0938)	0.2559
BAG-HAR-Equity	2.7979	(0.3811)	1.0019	(0.0988)	0.6242
SSVS-HAR-Equity	2.8050	(0.3811)	1.0078	(0.0938)	0.6233
AR(1)-Sentiment	3.3205	(0.1558)	1.1425	(0.0125)	0.5540
HAR-Sentiment	2.7944	(0.4838)	1.0040	(0.0938)	0.6247
LASSO-HAR-Sentiment	2.9876	(0.3811)	1.0217	(0.0938)	0.5987
BMA-HAR-Sentiment	2.7981	(0.5952)	1.0024	(0.0938)	0.6242
BAG-HAR-Sentiment	2.7868	(0.5952)	1.0080	(0.0938)	0.6257
SSVS-HAR-Sentiment	2.8070	(0.3811)	1.0073	(0.0938)	0.6230
AR(1)-Commodity	3.3389	(0.1558)	1.1389	(0.0091)	0.5516
HAR-Commodity	2.8099	(0.3811)	1.0211	(0.0938)	0.6226
LASSO-HAR-Commodity	3.0269	(0.3811)	1.0253	(0.0938)	0.5935
BMA-HAR-Commodity	2.7984	(0.4838)	1.0139	(0.0938)	0.6242
BAG-HAR-Commodity	2.7975	(0.3811)	1.0150	(0.0938)	0.6243
SSVS-HAR-Commodity	2.8054	(0.3811)	1.0078	(0.0938)	0.6232
AR(1)-FX	3.3670	(0.1558)	1.1546	(0.0325)	0.5478
HAR-FX	3.2294	(0.3811)	1.0521	(0.0938)	0.5663
LASSO-HAR-FX	2.9660	(0.3811)	1.0187	(0.0938)	0.6017
BMA-HAR-FX	2.7426	(0.5952)	1.0131	(0.0938)	0.6317
BAG-HAR-FX	2.7843	(0.5952)	1.0028	(0.0953)	0.6261
SSVS-HAR-FX	2.8035	(0.3811)	1.0075	(0.0938)	0.6235
AR(1)-Google	3.3051	(0.1558)	1.1427	(0.0325)	0.5561
HAR-Google	2.7542	(0.5952)	1.0023	(0.0953)	0.6301
LASSO-HAR-Google	2.9671	(0.3811)	1.0161	(0.0938)	0.6015
BMA-HAR-Google	2.7162	(0.5952)	0.9897	(0.5070)	0.6352
BAG-HAR-Google	**2.7075**	(1.0000)	0.9897	(0.5070)	**0.6364**
SSVS-HAR-Google	2.8041	(0.3811)	1.0064	(0.0938)	0.6234
AR(1)-All	3.0871	(0.3811)	1.0954	(0.0938)	0.5854
HAR-All	3.1935	(0.3811)	1.1077	(0.0938)	0.5711
LASSO-HAR-All	2.7787	(0.5952)	**0.9734**	(1.0000)	0.6268
BMA-HAR-All	2.7324	(0.5952)	1.0898	(0.0325)	0.6330
BAG-HAR-All	2.7337	(0.5952)	0.9822	(0.5070)	0.6328
SSVS-HAR-All	2.7986	(0.4838)	1.0066	(0.0938)	0.6241

TABLE 2 (Continued)

Model	MSE		MAE		R^2_{OS}
AR(1)-PCA	3.1650	(0.1558)	1.0782	(0.0344)	0.5749
HAR-PCA	2.8896	(0.3811)	1.0117	(0.0938)	0.6119
LASSO-HAR-PCA	2.9953	(0.3811)	1.0215	(0.0938)	0.5977
BMA-HAR-PCA	2.8386	(0.3811)	1.0033	(0.1035)	0.6188
BAG-HAR-PCA	2.7992	(0.3811)	1.0085	(0.0938)	0.6240
SSVS-HAR-PCA	2.8047	(0.3811)	1.0070	(0.0938)	0.6233

Note: This table presents the mean squared error (MSE), the mean average error (MAE) loss results, and the R^2_{OS} for short-term forecasts of WTI volatility. Lower values of MSE and MAE loss functions imply higher forecast precision. The corresponding MCS *p*-values are listed in parentheses on the right sides of the MAE and MSE results, with $p^{MCS} \geq 0.1$ implying an inclusion in the MCS at 10% confidence while $p^{MCS} \geq 0.25$ implies inclusion in the MCS at the 25% confidence level. The R^2_{OS} measures the proportion of variance explained by the forecasts: $R^2_{OS} = 1 - \frac{\sum_{t=T_1}^{t=T}\left(\widehat{RV}_t - RV_t\right)^2}{\sum_{t=T_1}^{t=T}\left(RV_t - \overline{RV}\right)^2}$. Higher values of R^2_{OS} suggest higher forecast precision.

sentiment and uncertainty measures of financial markets, such as the Global Economic Policy Uncertainty (GEPU) index (Baker et al., 2016), S&P500 implied volatility (VIX), oil price implied volatility (OVX), and the BofA Merrill Lynch Global Financial Stress Index (GFSI). The third and the fourth groups contain the volatility—measured as daily squared returns—of other commodity futures traded at the CME such as Natural Gas, Gold, Corn, and Soybeans, as well as FX rates with reference US Dollar against the Euro, Japanese Yen, Chinese Yuan, Russian Ruble, and the UK Pound. Realized volatilities of equity indices of nine major stock markets are obtained from the Oxford-Man *Realized Library* (Heber et al., 2009).[4] The remaining lower frequency data are acquired from Datastream. The sample spans the time period from January 5, 2010, to May 11, 2018. By data pre-processing and synchronization of the data sets, we obtain $n = 2596$ daily observations. In order to capture market attention with regard to oil, we follow Afkhami et al. (2017) and include Google search volumes for relevant search terms as exogenous variables. Table A1 provides an overview and describes the construction.

Summary statistics and preliminary analyses are given in Table 1. For realized volatilities of WTI futures prices, we find the mean and standard deviation to be higher than those of the financial market realized volatilities. This underlines the observation that prices in crude oil futures markets are much more volatile than equity indices. Most importantly, we observe highly significant autocorrelations of realized volatilities for WTI—and for all stock market indices to that end. This reinforces our decision to apply HAR frameworks as in the HAR model in both realized volatility as well as exogenous factors. The only exception to this present autocorrelation in realized volatility are Corn futures, where neither 5, 10, nor 22 lags yield sufficient evidence regarding dependencies between daily realized volatilities. Corn and Natural Gas have the highest mean and standard deviation across all volatility measures. This translates to very high intraday returns, possibly triggered by intraday jumps, which is also visible in Figure B2, plot (b). All volatility measures are found to be stationary as expected from the construction in Equation (1). The resolution of Google search volume is in index points. Hence, summary statistics cannot be compared directly. More importantly for our methodological framework, we find highly significant autocorrelations, which we utilize in the employed autoregressive structure of the HAR extensions.

The visualizations of the remainder of the variables are found in Appendix B. Figure 1 visualizes the daily realized volatility of WTI futures on our data range. Figure B1 analogously visualizes the volatilities of major stock market indices. We observe the clustering of highly volatile periods in oil markets for 2011 and 2015–2017. These volatility clusters are also present for equity indices indicating some degree of connectedness which is explicitly modeled in our model framework. Periods of high and low volatility are also depicted by specific sentiment indices such as the VIX for equity markets (S&P500) and the OVX (WTI) visualized in Figure B2a. Highly volatile in general, the commodity futures for Natural Gas, Gold, Corn, and Soybean—plotted in Figure B2b—also show clustering of high- and low-volatility regimes albeit of much higher magnitude and frequency than WTI and equity futures in our sample. Volatility of FX markets is depicted in Figure B3, and again, we find highly volatile markets in 2015–2017. Some important events, such as the sanctions of the USA against Russia and China as well as the oil price collapse 2014–2015, are visible as increasingly volatile rate movements in FX markets.

TABLE 3 Statistical evaluations of various volatility models for five-step forecasts.

Model	MSE		MAE		R^2_{OS}
AR(1)	1.3563	(0.0899)	1.0005	(0.0207)	0.6131
HAR	1.2262	(0.1700)	0.9738	(0.0207)	0.6502
LASSO-HAR	1.0817	(0.5343)	0.9077	(0.4274)	0.6914
BMA-HAR	1.2606	(0.1700)	0.9315	(0.0207)	0.6404
BAG-HAR	1.2276	(0.1352)	0.9744	(0.0207)	0.6498
SSVS-HAR	1.2408	(0.1284)	0.9560	(0.0207)	0.6460
AR(1)-Equity	2.6016	(0.0899)	1.0409	(0.0207)	0.2578
HAR-Equity	1.2215	(0.1700)	0.9654	(0.0207)	0.6515
LASSO-HAR-Equity	1.1642	(0.2327)	0.9438	(0.0207)	0.6679
BMA-HAR-Equity	1.3090	(0.1352)	0.9487	(0.0207)	0.6266
BAG-HAR-Equity	1.2593	(0.1284)	0.9918	(0.0182)	0.6408
SSVS-HAR-Equity	1.2350	(0.0899)	0.9553	(0.0207)	0.6477
AR(1)-Sentiment	1.3264	(0.0899)	0.9941	(0.0207)	0.6216
HAR-Sentiment	1.2425	(0.1352)	0.9835	(0.0207)	0.6455
LASSO-HAR-Sentiment	1.0913	(0.5343)	0.9107	(0.4274)	0.6887
BMA-HAR-Sentiment	1.2521	(0.0899)	0.9577	(0.0207)	0.6428
BAG-HAR-Sentiment	1.2337	(0.1352)	0.9781	(0.0207)	0.6480
SSVS-HAR-Sentiment	1.2411	(0.1284)	0.9561	(0.0207)	0.6459
AR(1)-Commodity	1.3505	(0.0899)	0.9982	(0.0207)	0.6147
HAR-Commodity	1.2310	(0.1352)	0.9788	(0.0207)	0.6488
LASSO-HAR-Commodity	1.0835	(0.5343)	0.9087	(0.4274)	0.6909
BMA-HAR-Commodity	1.2441	(0.1284)	0.9525	(0.0207)	0.6451
BAG-HAR-Commodity	1.2291	(0.1352)	0.9751	(0.0207)	0.6494
SSVS-HAR-Commodity	1.2355	(0.1284)	0.9557	(0.0207)	0.6475
AR(1)-FX	1.3719	(0.0899)	1.0043	(0.0207)	0.6086
HAR-FX	1.1413	(0.1700)	0.9484	(0.0207)	0.6744
LASSO-HAR-FX	**1.0624**	(1.0000)	**0.8946**	(1.0000)	**0.6969**
BMA-HAR-FX	1.1952	(0.1700)	0.9312	(0.0207)	0.6590
BAG-HAR-FX	1.2697	(0.0899)	1.0078	(0.0068)	0.6378
SSVS-HAR-FX	1.2347	(0.0899)	0.9541	(0.0207)	0.6478
AR(1)-Google	1.2448	(0.1700)	0.9864	(0.0207)	0.6449
HAR-Google	1.3549	(0.0899)	0.9978	(0.0207)	0.6135
LASSO-HAR-Google	1.1934	(0.1700)	0.9356	(0.0207)	0.6595
BMA-HAR-Google	1.3016	(0.0899)	0.9647	(0.0207)	0.6287
BAG-HAR-Google	1.3103	(0.0899)	0.9900	(0.0207)	0.6262
SSVS-HAR-Google	1.2424	(0.0899)	0.9563	(0.0207)	0.6456
AR(1)-All	1.4600	(0.0899)	0.9527	(0.0207)	0.5835
HAR-All	1.2649	(0.1284)	0.9780	(0.0207)	0.6392
LASSO-HAR-All	1.1733	(0.1700)	0.9527	(0.0207)	0.6653
BMA-HAR-All	1.2582	(0.1352)	0.9513	(0.0207)	0.6411
BAG-HAR-All	1.2898	(0.0899)	0.9944	(0.0182)	0.6320
SSVS-HAR-All	1.2140	(0.1284)	0.9559	(0.0207)	0.6537

TABLE 3 (Continued)

Model	MSE		MAE		R^2_{OS}
AR(1)-PCA	1.1660	(0.2327)	0.9162	(0.4274)	0.6674
HAR-PCA	1.2578	(0.1284)	0.9969	(0.0120)	0.6412
LASSO-HAR-PCA	1.0958	(0.2327)	0.9202	(0.0207)	0.6874
BMA-HAR-PCA	1.2570	(0.0899)	0.9642	(0.0207)	0.6414
BAG-HAR-PCA	1.2486	(0.1284)	0.9921	(0.0182)	0.6438
SSVS-HAR-PCA	1.2409	(0.1284)	0.9562	(0.0207)	0.6460

Note: This table presents the mean squared error (MSE), the mean average error (MAE) loss results, and the R^2_{OS} for medium-term forecasts of WTI volatility. Lower values of MSE and MAE loss functions imply higher forecast precision. The corresponding MCS *p*-values are listed in parentheses on the right sides of the MAE and MSE results, with $p^{MCS} \geq 0.1$ implying an inclusion in the MCS at 10% confidence while $p^{MCS} \geq 0.25$ implies inclusion in the MCS at the 25% confidence level. The R^2_{OS} measures the proportion of variance explained by the forecasts: $R^2_{OS} = 1 - \frac{\sum_{t=T_1}^{t=T}\left(\widehat{RV}_t - RV_t\right)^2}{\sum_{t=T_1}^{t=T}\left(RV_t - \overline{RV}\right)^2}$. Higher values of R^2_{OS} suggest higher forecast precision.

Interestingly, this also aligns with Google search queries, featured as indices in Figure B4, for "oil production" (G1), "oil demand" (G3), and "oil price" (G4), which show that there is an increase in general interest in those topics, which we utilize as additional attention measure.

4 | OUT-OF-SAMPLE RESULTS

4.1 | Forecasting accuracy

This section presents the results of the out-of-sample analysis. In particular, we forecast the WTI volatility 1, 5, and 22 days ahead using the above-mentioned models, methods, and predictors. We stress the fact that all information are known at the time to forecast the next periods' volatility (real time).

We start our analysis by assessing the statistical accuracy of the different models. The results for 1-, 5-, and 22-day-ahead forecasts are presented in Tables 2, 3, and 4.

For the 1-day-ahead forecasts, the best models are BAG-HAR-Google and LASSO-HAR-All in terms of MSE and MAE, respectively. The 25% MCS for the MSE includes all models but most of the AR models. Thus, the performance of any other model cannot be differentiated from the BAG-HAR-Google. For the MAE, the MCS are more exclusive. Only four models belong to the 25% MCS (BMA- and BAG-HAR-Google, LASSO- and BAG-HAR-All), and additional two models (BMA-HAR and BMA-HAR-PCA) are elements of 10% MCS.

Turning to the 5-day-ahead forecast, we find a similar pattern in variable selection. The 10% MCS for the MSE loss function is quite large. However, only four models are included in the 25% MCS: LASSO-HAR, LASSO-HAR-Sentiment, LASSO-HAR-Commodity, and LASSO-HAR-FX. The latter is also the best model in terms of the lowest MSE and MAE. For the MAE, only five models are part of the 10% and 25% MCS. Those models include the aforementioned LASSO models as well as AR(1)-All.

The statistical accuracy for 22-day-ahead forecasts is almost equal over the entire set of models. At least the MCS includes all of them for both loss functions. The MSE set even includes all HAR model variants in the more restrictive 25% set. For MAE, with some other models, all LASSO variants belong to the 25% MCS. We also find that the best models for both loss functions are the Google variants.

In contrast to Degiannakis and Filis (2017), we cannot see a clear outperformance of HAR models incorporating other asset class channels over the standard HAR models. For all three forecasting horizons, the HAR and the various HAR-X models are either together in the MCS or they are not. Thus, we conclude that the ML approaches, on top of the combination of HAR model and information class, are what make the difference.

Summarizing, we find an outperformance of the LASSO-HAR variants that are most of the time included in the even more restrictive 25% confidence sets. The very simple AR models are only included in two occasions in the most restrictive sets indicating the value of the long-term components of the HAR models. The LASSO-HAR-All model, selection out of a set of all information, is included in the 25% MCS for 1- and 22-day-ahead forecast and thus cannot be distinguished from the best performing models in terms of its forecast accuracy. We will use this model in the following section, to investigate the time-varying selection of variables.

TABLE 4 Statistical evaluations of various volatility models for 22-step forecasts.

Model	MSE		MAE		R^2_{OS}
AR(1)	0.9316	(0.2538)	0.9087	(0.0743)	0.6040
HAR	0.7596	(0.3160)	0.7691	(0.6133)	0.6771
LASSO-HAR	0.7127	(0.5850)	0.7536	(0.8506)	0.6970
BMA-HAR	0.7494	(0.5850)	0.7831	(0.8131)	0.6814
BAG-HAR	0.7612	(0.3160)	0.7701	(0.1741)	0.6764
SSVS-HAR	0.7413	(0.3160)	0.7851	(0.1741)	0.6849
AR(1)-Equity	3.3078	(0.0586)	1.0013	(0.1741)	−0.4061
HAR-Equity	0.7240	(0.5850)	0.7686	(0.1741)	0.6922
LASSO-HAR-Equity	0.7356	(0.3779)	0.7607	(0.8506)	0.6873
BMA-HAR-Equity	0.7413	(0.3779)	0.7835	(0.1741)	0.6849
BAG-HAR-Equity	0.7353	(0.3779)	0.7665	(0.6133)	0.6875
SSVS-HAR-Equity	0.7405	(0.3160)	0.7729	(0.1741)	0.6852
AR(1)-Sentiment	0.8986	(0.3160)	0.9041	(0.1575)	0.6180
HAR-Sentiment	0.7556	(0.3160)	0.7697	(0.6133)	0.6788
LASSO-HAR-Sentiment	0.7245	(0.5850)	0.7575	(0.8506)	0.6920
BMA-HAR-Sentiment	0.7318	(0.3779)	0.7813	(0.1741)	0.6889
BAG-HAR-Sentiment	0.7594	(0.3160)	0.7714	(0.1741)	0.6772
SSVS-HAR-Sentiment	0.7411	(0.3160)	0.7851	(0.1741)	0.6850
AR(1)-Commodity	0.9407	(0.2520)	0.9121	(0.0655)	0.6001
HAR-Commodity	0.7733	(0.3160)	0.7751	(0.1741)	0.6713
LASSO-HAR-Commodity	0.7406	(0.3160)	0.7626	(0.8131)	0.6852
BMA-HAR-Commodity	0.7503	(0.3160)	0.7824	(0.1741)	0.6811
BAG-HAR-Commodity	0.7743	(0.3160)	0.7750	(0.1741)	0.6709
SSVS-HAR-Commodity	0.7268	(0.3779)	0.7724	(0.6133)	0.6911
AR(1)-FX	0.9652	(0.0586)	0.8985	(0.1741)	0.5897
HAR-FX	0.7704	(0.3160)	0.7773	(0.1741)	0.6725
LASSO-HAR-FX	0.7235	(0.5850)	0.7575	(0.8506)	0.6925
BMA-HAR-FX	0.7451	(0.3160)	0.7900	(0.1741)	0.6833
BAG-HAR-FX	0.7722	(0.3160)	0.7754	(0.1741)	0.6718
SSVS-HAR-FX	0.7457	(0.3160)	0.7752	(0.1741)	0.6830
AR(1)-Google	0.8153	(0.3160)	0.8919	(0.1741)	0.6535
HAR-Google	0.6851	(0.9752)	**0.7476**	(1.0000)	0.7088
LASSO-HAR-Google	0.7228	(0.5850)	0.7570	(0.8506)	0.6928
BMA-HAR-Google	**0.6811**	(1.0000)	0.7595	(0.8506)	**0.7105**
BAG-HAR-Google	0.6885	(0.5850)	0.7492	(0.8506)	0.7074
SSVS-HAR-Google	0.7369	(0.3779)	0.7836	(0.1741)	0.6867
AR(1)-All	1.7713	(0.0586)	0.9155	(0.1741)	0.2471
HAR-All	0.7471	(0.3160)	0.7786	(0.1741)	0.6824
LASSO-HAR-All	0.7451	(0.3160)	0.7691	(0.6133)	0.6833
BMA-HAR-All	0.7245	(0.5850)	0.7766	(0.6133)	0.6920
BAG-HAR-All	0.7469	(0.3160)	0.7687	(0.6133)	0.6825
SSVS-HAR-All	0.7426	(0.3779)	0.7617	(0.8506)	0.6843

TABLE 4 (Continued)

Model	MSE		MAE		R^2_{OS}
AR(1)-PCA	0.8703	(0.3160)	0.8518	(0.1741)	0.6300
HAR-PCA	0.7714	(0.3160)	0.7781	(0.1741)	0.6721
LASSO-HAR-PCA	0.7410	(0.3160)	0.7640	(0.6133)	0.6850
BMA-HAR-PCA	0.7539	(0.3160)	0.7862	(0.1741)	0.6795
BAG-HAR-PCA	0.7842	(0.3160)	0.7827	(0.1741)	0.6667
SSVS-HAR-PCA	0.7418	(0.3160)	0.7855	(0.1741)	0.6847

Note: This table presents the mean squared error (MSE), the mean average error (MAE) loss results, and the R^2_{OS} for long-term forecasts of WTI volatility. Lower values of MSE and MAE loss functions imply higher forecast precision. The corresponding MCS *p*-values are listed in parentheses on the right sides of the MAE and MSE results, with $p^{MCS} \geq 0.1$ implying an inclusion in the MCS at 10% confidence while $p^{MCS} \geq 0.25$ implies inclusion in the MCS at the 25% confidence level. The R^2_{OS} measures the proportion of variance explained by the forecasts: $R^2_{OS} = 1 - \frac{\sum_{t=T_1}^{t=T} \left(\widehat{RV}_t - RV_t\right)^2}{\sum_{t=T_1}^{t=T} \left(RV_t - \overline{RV}\right)^2}$. Higher values of R^2_{OS} suggest higher forecast precision. The best model per measure is highlighted in bold face.

4.2 | Economic value

From a statistical point of view, the standard HAR models are outperformed by their ML-augmented counterparts. In the next step, we evaluate the economic value of the model augmentations. In particular, we focus on portfolio construction based on volatility forecasts produced by all predictive models. We assess the annualized return that a risk-averse investor would sacrifice in order to switch from a benchmark portfolio—AR(1) models without ML-driven variable selection—to any other portfolio, including those that are produced by the augmented models. The results for the different forecast horizons are presented in Table 5.

In contrast to the statistical superiority of the LASSO variants, we observe that this outperformance does not translate to the best economic value. We find that the BMA models are superior and allow a risk-averse investor with a quadratic utility function to improve their overall utility. In particular, the BMA-HAR-FX yields the highest annualized return. This model also results in the highest economic value compared with an AR(1) model. An investor with $\gamma = 1$ would be willing to switch from a standard AR(1) model to a BMA-HAR-FX model up to 5.28%. Even to switch from the standard HAR model, the opportunity to incorporate more information than just past realized volatility would be worth more than 1% ($5.28\% - 4.22\%$). In risk-adjusted terms, however, we find BAG models to have the highest Sharpe ratios. The BAG-HAR-All produces the highest Sharpe ratio in our out-of-sample exercise (0.0678). The LASSO variants, which yielded the highest statistical accuracy, perform similarly to the standard HAR models. The variable selection does not appear to be able to translate the higher accuracy to economic performance.

In terms of adding other information channels, we find economic value in incorporating such information in volatility models. Although we did not necessarily find an outperformance on a statistical accuracy, additional information result in higher economic value in the majority of cases. For example, the HAR model has a smaller economic value and Sharpe Ratio than its sentiment, commodity, and FX-augmented counterparts as well as the HAR-All and HAR-PCA. Similar patterns can be observed between the ML-estimated HAR models and their variants with additional information.

4.3 | Time-varying drivers

In order to understand the time variation of different information classes, we investigate the selection of variables over our entire out-of-sample period. In particular, we want to understand (1) which are the most important variables (and at which frequency-average) when making no prior choice, (2) how sticky are those variables, and (3) can other information classes replace their own realized volatility (i.e., the WTI RV at 1, 5, and 22-day average).

The top plot in Figure 2 displays the 1-day-ahead forecasts variable selection with the LASSO-HAR-All over the whole forecast period. All three RV horizons (red) are selected throughout the out-of-sample period. In addition, we find that the most consistent variables are the OVX (oil price uncertainty) at 5 and 22 days horizon and the Google search volume for "Oil Price" the day before.

The middle plot in Figure 2 shows the 5-day-ahead forecast variable selection. We find that the RV of the WTI the day before is never selected. Only the weekly and monthly RV appear to be relevant for the forecast.

TABLE 5 Economic evaluations.

	$\gamma = 1$				$\gamma = 10$			
	r_p	$\frac{r_p}{\sigma_p}$	Δ_1	Δ_2	r_p	$\frac{r_p}{\sigma_p}$	Δ_1	Δ_2
AR(1)	−1.7566	−0.0338	0.0000	0.0000	1.3663	0.2558	0.0000	0.0000
HAR	2.4220	0.0541	4.2157	4.2157	1.7842	0.3849	0.4402	0.4402
LASSO-HAR	1.9851	0.0443	3.7714	3.7714	1.7405	0.3757	0.3938	0.3938
BMA-HAR	1.3423	0.0243	3.1058	3.1058	1.6762	0.2960	0.3242	0.3242
BAG-HAR	2.4014	0.0536	4.1947	4.1947	1.7821	0.3843	0.4380	0.4380
SSVS-HAR	2.7588	0.0500	4.5450	4.5450	1.8179	0.3203	0.4745	0.4745
AR(1)-Equity	2.3220	0.0434	0.0000	4.1024	1.7742	0.3225	0.0000	0.4283
HAR-Equity	2.1395	0.0467	−0.1733	3.9270	1.7559	0.3706	−0.0180	0.4100
LASSO-HAR-Equity	2.0429	0.0455	−0.2693	3.8299	1.7463	0.3758	−0.0281	0.3999
BMA-HAR-Equity	2.0508	0.0364	−0.2768	3.8225	1.7471	0.3019	−0.0289	0.3990
BAG-HAR-Equity	2.3699	0.0528	0.0596	4.1625	1.7790	0.3827	0.0063	0.4346
SSVS-HAR-Equity	3.1705	0.0605	0.8574	4.9694	1.8590	0.3443	0.0895	0.5188
AR(1)-Sentiment	−1.3490	−0.0253	0.0000	0.4041	1.4071	0.2567	0.0000	0.0422
HAR-Sentiment	2.5591	0.0566	3.9464	4.3549	1.7979	0.3845	0.4120	0.4547
LASSO-HAR-Sentiment	2.1009	0.0469	3.4811	3.8890	1.7521	0.3781	0.3634	0.4060
BMA-HAR-Sentiment	2.7992	0.0506	4.1773	4.5861	1.8219	0.3207	0.4361	0.4788
BAG-HAR-Sentiment	2.5643	0.0567	3.9517	4.3601	1.7984	0.3846	0.4126	0.4552
SSVS-HAR-Sentiment	2.7580	0.0499	4.1355	4.5442	1.8178	0.3202	0.4317	0.4744
AR(1)-Commodity	−2.0925	−0.0410	0.0000	−0.3328	1.3327	0.2539	0.0000	−0.0347
HAR-Commodity	3.0198	0.0653	5.1612	4.8237	1.8440	0.3858	0.5389	0.5037
LASSO-HAR-Commodity	2.4507	0.0539	4.5811	4.2441	1.7871	0.3797	0.4783	0.4431
BMA-HAR-Commodity	3.2071	0.0572	5.3392	5.0016	1.8627	0.3232	0.5574	0.5222
BAG-HAR-Commodity	3.0205	0.0652	5.1617	4.8242	1.8440	0.3848	0.5389	0.5037
SSVS-HAR-Commodity	2.6341	0.0497	4.7581	4.4210	1.8054	0.3307	0.4967	0.4615
AR(1)-FX	−3.7883	−0.0749	0.0000	−2.0091	1.1632	0.2236	0.0000	−0.2096
HAR-FX	2.9859	0.0653	6.8364	4.7896	1.8406	0.3893	0.7138	0.5001
LASSO-HAR-FX	2.2795	0.0506	6.1131	4.0704	1.7699	0.3800	0.6382	0.4250
BMA-HAR-FX	3.4803	0.0596	7.3270	**5.2775**	1.8900	0.3151	0.7649	**0.5510**
BAG-HAR-FX	2.8013	0.0615	6.6471	4.6014	1.8221	0.3867	0.6940	0.4804
SSVS-HAR-FX	2.4326	0.0469	6.2610	4.2174	1.7853	0.3344	0.6536	0.4403
AR(1)-Google	−2.5096	−0.0461	0.0000	−0.7515	1.2910	0.2309	0.0000	−0.0784
HAR-Google	2.3959	0.0541	4.9514	4.1897	1.7816	0.3887	0.5170	0.4375
LASSO-HAR-Google	2.0954	0.0468	4.6446	3.8835	1.7515	0.3783	0.4849	0.4055
BMA-HAR-Google	2.7495	0.0507	5.2993	4.5369	1.8169	0.3257	0.5532	0.4736
BAG-HAR-Google	2.5362	0.0571	5.0944	4.3325	1.7956	0.3904	0.5319	0.4524
SSVS-HAR-Google	2.7582	0.0500	5.3068	4.5444	1.8178	0.3205	0.5540	0.4744
AR(1)-All	−0.5690	−0.0110	0.0000	1.1862	1.4851	0.2787	0.0000	0.1237
HAR-All	2.8910	0.0603	3.4924	4.6900	1.8311	0.3697	0.3647	0.4897
LASSO-HAR-All	3.0138	0.0660	3.6203	4.8182	1.8434	0.3899	0.3781	0.5031
BMA-HAR-All	3.2646	0.0596	3.8635	5.0622	1.8685	0.3315	0.4034	0.5285
BAG-HAR-All	3.1723	0.0678	3.7804	4.9789	1.8592	0.3847	0.3948	0.5199
SSVS-HAR-All	2.1335	0.0462	2.7254	3.9204	1.7553	0.3676	0.2846	0.4093

TABLE 5 (Continued)

	$\gamma = \mathbf{1}$				$\gamma = \mathbf{10}$			
	$\boldsymbol{r_p}$	$\frac{r_p}{\sigma_p}$	Δ_1	Δ_2	$\boldsymbol{r_p}$	$\frac{r_p}{\sigma_p}$	Δ_1	Δ_2
AR(1)-PCA	−1.4749	−0.0280	0.0000	0.2795	1.3945	0.2576	0.0000	0.0292
HAR-PCA	2.7687	0.0608	4.2853	4.5682	1.8189	0.3858	0.4474	0.4770
LASSO-HAR-PCA	2.2566	0.0502	3.7648	4.0472	1.7677	0.3798	0.3931	0.4226
BMA-HAR-PCA	3.0610	0.0550	4.5697	4.8527	1.8481	0.3228	0.4770	0.5066
BAG-HAR-PCA	2.7372	0.0601	4.2532	4.5360	1.8157	0.3851	0.4441	0.4736
SSVS-HAR-PCA	2.7722	0.0502	4.2758	4.5587	1.8192	0.3205	0.4464	0.4759

Note: Results for 22-step-ahead forecasts for the economic evaluation. We present the annualized average excess portfolio return r_p, the average portfolio Sharpe Ratio $\frac{r_p}{\sigma_p}$, and the economic value Δ of the portfolio over the standard AR model within the models class. Here, Δ_1 and Δ_2 refer to the economic value compared to AR(1) model with the asset class information and the standard AR(1) model, respectively. The model with the highest economic value is highlighted in bold face.

Other variables at daily horizon are selected, probably replacing the RVs as a more suitable variable for prediction. Although we do not find a persistent variable on the short-term, weekly, and monthly averages of the Chinese stock market (SSEC), sentiment (OVX, VIX, EPU), commodities (corn, soybean, natural gas), and FX (USD, Yen/USD, GBP/USD) are selected almost through-out the entire sample.

A look at the bottom plot of Figure 2 for 22 days ahead reveals that the LASSO-HAR-All uses many different variables throughout the sample at different time horizons, for example, NIKKEI and SSEC, VIX, OVX, natural gas, soybeans, and the Google search volumes for "Financial Crisis" and "OPEC Conference."

With increasing forecasting horizon, LASSO identifies an increasing number of exogenous predictors to be useful for volatility forecasting. This ML approach also shows that the selection of variables is varying over time. Apart from the more persistent variables mentioned above, several exogenous variables only play a role for a limited period of time and we observe clustering of variable selections. This culminates in the observation that for the 22-day forecasting horizon, there appears to be a structural break in variable selection at the end of 2016. Stock market volatility, which was significant until this date, is not being chosen by LASSO, whereas some commodity- and FX-related variables are integrated in the forecasts. As to why LASSO discards several exogenous predictors at the same time is not within the scope of this work, and we leave this question open for future research.

Lastly, we also consider ranking exogenous variables with regard to variable importance. We follow Narajewski and Ziel (2020) and calculate the average variable importance over the out-of-sample by

$$\overline{VI}_i = T^{-1} \sum_{t=1}^{T} VI_{t,i},$$

$$VI_{t,i} = \frac{\left|\widehat{\beta}_i^{(t)}\right|}{\sum_{i=1}^{N} \left|\widehat{\beta}_i^{(t)}\right|},$$

where T is the number of out-of-sample observations, N is the number of variables, and $\widehat{\beta}_i^{(t)}$ refers to the estimated coefficient of variable i for the out-of-sample at observation t. Note that for the LASSO estimation, all variables are standardized. Hence, we are able to compare the individual contribution based on these estimates. Table 6 lists the 10 most important variables per forecast horizon.

Although there exists some correlation between the rankings of selections and importance, there are also major differences. We find the RVs of the WTI itself always in the ranking of 22 days, 5 days, and 1 day. The most important variable over all forecast horizons is the 22-day RV of the WTI, which contributes about 50% to the forecasted RV (or more). We find the FX channel to have a fair share of importance, especially in comparison with the other information channels. Sentiment indices (VIX, OVX, and GFSI) are also among the top 10 for all forecasting horizons. In contrast to the frequency of selection, however, Google search volumes do not carry much variable importance for forecasts beyond 1 day. In fact, the Google attention measures are not even in the top 25.

This observation is also confirmed, when looking at the aggregated channel level. In Table 7, we report the sum of the variable importance per variable of an information channel (i.e., per variable and time aggregated). Especially for longer horizons, the contribution

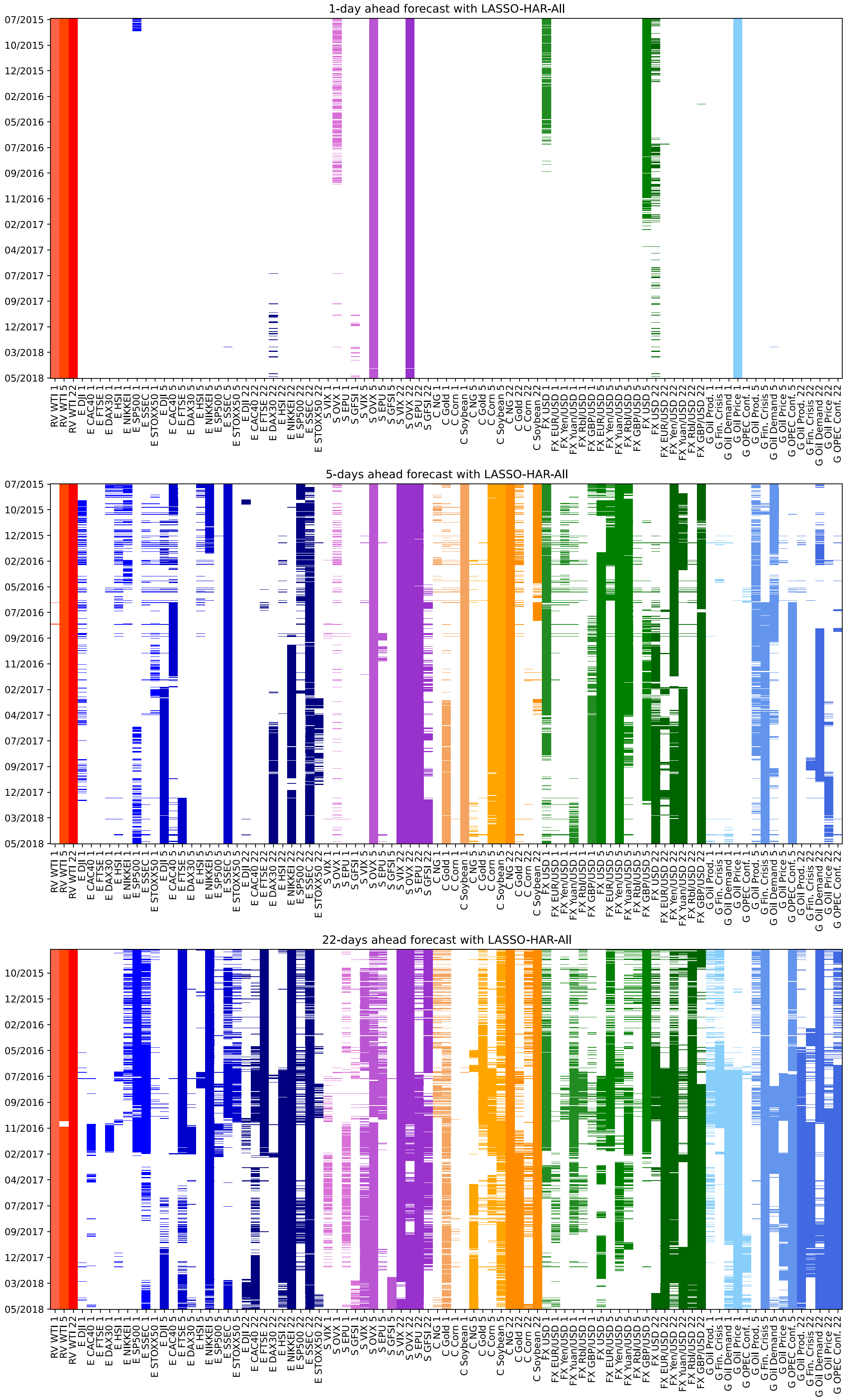

FIGURE 2 Variable selection over time with the LASSO-HAR-All.

TABLE 6 Ranked variable importance of individual variables (top 10) across the applied forecasting horizons.

Rank	1 day	VI	5 days	VI	22 days	VI
1	RV WTI 22	47.8%	RV WTI 22	53.6%	RV WTI 22	66.4%
2	RV WTI 5	30.9%	FX Yuan/USD 22	9.4%	FX EUR/USD 22	5.6%
3	RV WTI 1	8.9%	RV WTI 5	7.4%	FX Yuan/USD 22	4.1%
4	FX USD 5	5.1%	S OVX 22	5.2%	FX Yen/USD 22	2.8%
5	S OVX 22	3.5%	FX Yen/USD 5	4.6%	E NIKKEI 22	2.7%
6	S VIX 5	2.0%	FX Yuan/USD 5	4.3%	E HSI 22	2.6%
7	FX USD 22	0.7%	FX GBP/USD 22	3.0%	S GFSI 22	1.6%
8	FX USD 1	0.6%	FX USD 5	2.7%	RV WTI 5	1.5%
9	G Oil Price 1	0.5%	FX USD 22	2.2%	FX GBP/USD 22	1.3%
10	S OVX 1	<0.1%	FX Yen/USD 22	2.0%	C Soybean 22	1.3%

TABLE 7 Ranked variable importance of (aggregated) information channels across the applied forecasting horizons.

Rank	1 day	VI	5 days	VI	22 days	VI
1	RV	87.6%	RV	60.0%	RV	68.2%
2	FX	6.3%	FX	28.8%	FX	17.2%
3	Sentiment	5.6%	Sentiment	6.9%	Equity	8.6%
4	Google	0.5%	Equity	2.2%	Sentiment	3.5%
5	Equity	<0.1%	Commodity	1.5%	Commodity	2.6%
6	Commodity	0.0%	Google	3.0%	Google	1.1%

of FX is quite substantial. Again, we find little support for Google search volume. We also point out the contrast between the variable importance of the commodity channel and the frequency of selection for commodity variables. Finally, the variable aggregates over 22 days hold valuable information for all forecasting horizons.

We summarize that for all forecasting horizons, the realized volatilities of the WTI are the most important variables to predict future realized volatility of the WTI. However, the observed other information channels do contribute to the predictions and in this way either substitute or add information on top of the RVs. In particular, FX appears to be an important source of information.

5 | CONCLUSION

This paper demonstrates how extending existing models of realized volatility with additional information from other channels and recent ML techniques benefits the quality of forecasts and subsequent portfolio performance. We focus on modeling realized volatility of the most liquid WTI crude oil futures prices. Existing HAR models (Corsi, 2009) are first augmented with exogenous factors, which have been shown to improve the forecasting performance across different horizons (Luo et al., 2022; Ma et al., 2018). Motivated by Degiannakis and Filis (2017), we include several different information channels that include major stock markets, relevant FX market pairs, sentiment indices, other linked commodity markets, and Google search volumes for relevant search terms. We then extend these models with four machine-learning approaches that pick the most suitable factors for forecasting realized volatility over 1 day, 1 week, and 1 month. The model set is completed by a PCA variant and a model including all possible exogenous variables.

We present several novel findings. First and foremost, including ML to choose from a set of exogenous variables improves the quality of realized volatility forecasts. In particular, LASSO variants show significant improvements. Second, we find that the variable selection process depends strongly on the forecasting horizon. Although for short-term forecasts, endogenous factors dominate the selection of predictors, the number of predictors increases when increasing the forecasting horizon. Sentiment variables, such as the EPU or OVX, realized volatility of other stock markets and commodities as well as FX markets become increasingly important for longer horizons. We also show that a combination of short-, medium-, and long-term averages of text-based Google indicators are relevant exogenous factors that are included in the predictor set generated by ML algorithms. Third, we show that portfolios that are constructed with ML-HAR variants (particularly BMA) provide higher returns than the baseline AR and HAR models and its extensions with exogenous factors. Surprisingly, forecast accuracy does not necessarily imply the highest portfolio outperformance. Lastly, we show that the selection process with its dynamic implementation is time-varying with respect to the variable choices. Variable selection mostly clusters and these clusters differ across forecasting horizons. In particular, for the stock market channel, the results point toward structural changes, especially for longer time horizons. Although the most important variables remain the past (aggregates) of realized volatility of the WTI, a great deal of information is coming from the FX channel (especially for longer forecast horizons). In contrast, Google search volume and other commodity volatilities do not carry large fractions of variable importance for the forecast of WTI realized volatility.

Future research might focus on as to why these structural changes in the variable selection process exist and why these changes materialize differently across forecasting horizons. Having shown that superior fit does not necessarily translate to superior portfolio performance in our portfolio selection application, future work could, for example, address an exploitation of ML-generated forecasts of realized volatility for other commodity or asset classes or extend our framework to other ML models such as reinforcement learning (see, e.g., Lavko et al., 2023).

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Tony Klein, Jiawen Luo, and Thomas Walther. The first draft of the manuscript was written by Tony Klein and Thomas Walther, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We are thankful to Derek Bunn (editor) and an anonymous referee for their comments. An earlier version of this working paper was circulated under the title "Forecasting Realized Volatility of Crude Oil Futures Prices based on Variable Selection Approaches." Jiawen Luo is supported by the National Natural Foundation of China (grant number ofs 72171088 and 71803049), the Natural Foundation of Guangdong Province (grant number 2023A1515012527), the Guangdong Philosophy and Social Science Foundation (grant number GD23CGL01), the Guangzhou Municipal Science and Technology Bureau (grant number 2023A04J1298).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Thomas Walther https://orcid.org/0000-0003-4359-987X

ENDNOTES

[1] This choice of sampling frequency is widely adapted in literature as it poses a trade-off between possible microstructure noise, where sampling frequency is too high, and information loss, where sampling frequency is too low (Aït-Sahalia & Yu, 2009; Liu et al., 2015).

[2] Examples can be found in Sévi (2014), Klein and Todorova (2021), Luo et al. (2022), and Degiannakis et al. (2022) among others.

[3] From our full set of models $M_0 = \{M^i, i = 1, ..., k\}$, the procedure determines a set of models with superior forecast performance $\widehat{M}^*_{1-\alpha}$. Given a confidence level α, the MCS $\widehat{M}^*_{1-\alpha}$ includes all models from M_0, which are statistically indistinguishable from the best model in the set, that is, the model with the lowest MSE or MAE, respectively. We implement the procedure using the T_R statistic, the stationary bootstrap with 10,000 draws, and α levels of 0.1 and 0.25.

[4] Data are freely available at https://realized.oxford-man.ox.ac.uk/.

REFERENCES

Afkhami, M., Cormack, L., & Ghoddusi, H. (2017). Google search keywords that best predict energy price volatility. *Energy Economics*, *67*, 17–27. https://doi.org/10.1016/j.eneco.2017.07.014

Aït-Sahalia, Y., & Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, *3*, 422–457. https://doi.org/10.1214/08-AOAS200

Audrino, F., & Medeiros, M. C. (2011). Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, *26*, 999–1022. https://doi.org/10.1002/jae.1171

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, *131*, 1593–1636. https://doi.org/10.1093/qje/qjw024

Blair, B. J., Poon, S.-H., & Taylor, S. J. (2001). Forecasting S and P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, *105*, 5–26. https://doi.org/10.1016/S0304-4076(01)00068-9

Campbell, D., & Slack, R. (2008). Narrative reporting: analysts' perceptions of its value and relevance. http://nrl.northumbria.ac.uk/2583/

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*, 1509–1531. https://doi.org/10.1093/rfs/hhm055

Charfeddine, L., Klein, T., & Walther, T. (2020). Reviewing the oil price–GDP growth relationship: A replication study. *Energy Economics*, *88*, 104786. https://doi.org/10.1016/j.eneco.2020.104786

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196. https://doi.org/10.1093/jjfinec/nbp001

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*, 368–380. https://doi.org/10.1080/07350015.2012.663261

Degiannakis, S., & Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, *76*, 28–49. https://doi.org/10.1016/j.jimonfin.2017.05.006

Degiannakis, S., Filis, G., Klein, T., & Walther, T. (2022). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*, *38*, 74–96. https://doi.org/10.1016/j.ijforecast.2019.08.011

Dellaportas, P., Forster, J., & Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, *12*, 27–36. https://doi.org/10.1023/A:1013164120801

Ederington, L. H., & Guan, W. (2010). Longer-term time-series volatility forecasts. *Journal of Financial and Quantitative Analysis*, *45*, 1055–1076. https://doi.org/10.1017/S0022109010000372

Elder, J., Miao, H., & Ramchander, S. (2014). Price discovery in crude oil futures. *Energy Economics*, *46*, 18–27. https://doi.org/10.1016/j.eneco.2014.09.012

Fattouh, B. (2010). The dynamics of crude oil price differentials. *Energy Economics*, *32*, 334–342. https://doi.org/10.1016/j.eneco.2009.06.007

Fernández, C., Ley, E., & Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*, 381–427. https://doi.org/10.1016/S0304-4076(00)00076-2

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, *100*, 514–537. https://doi.org/10.1016/j.jfineco.2011.02.003

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, *56*, 329–352. https://doi.org/10.1111/0022-1082.00327

Fleming, J., Kirby, C., & Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, *67*, 473–509. https://doi.org/10.1016/S0304-405X(02)00259-3

Fong, W. M., & See, K. H. (2002). A Markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, *24*, 71–95. https://doi.org/10.1016/S0140-9883(01)00087-1

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22. https://doi.org/10.18637/jss.v033.i01

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, *129*, 1661–1710. https://doi.org/10.1093/qje/qju024

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889. https://doi.org/10.1080/01621459.1993.10476353

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*, 363–398. https://doi.org/10.1016/S0304-4076(02)00207-5

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*, 453–497. https://doi.org/10.3982/ECTA5771

Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the U.S. oil market. *Journal of Banking and Finance*, *47*, 1–14. https://doi.org/10.1016/j.jbankfin.2014.05.026

Heber, G., Lunde, A., Shephard, N., & Sheppard, K. K. (2009). Oxford-Man Institute's realized library Version 0.3.

Huang, H., & Lee, T.-H. (2010). To combine forecasts or to combine information? *Econometric Reviews*, *29*, 534–570. https://doi.org/10.1080/07474938.2010.481553

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of the American Statistical Association*, *103*, 511–522. https://doi.org/10.1198/016214507000000473

Kaufmann, R. K., & Banerjee, S. (2014). A unified world oil market: Regions in physical, economic, geographic, and political space. *Energy Policy*, *74*, 235–242. https://doi.org/10.1016/j.enpol.2014.08.028

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*, 1053–1069. https://doi.org/10.1257/aer.99.3.1053

Klein, T. (2018). Trends and contagion in WTI and Brent crude oil spot and futures markets—The role of OPEC in the last decade. *Energy Economics*, *75*, 636–646. https://doi.org/10.1016/j.eneco.2018.09.013

Klein, T., & Todorova, N. (2021). Night trading with futures in China: The case of aluminum and copper. *Resources Policy*, *73*, 102205. https://doi.org/10.1016/j.resourpol.2021.102205

Klein, T., & Walther, T. (2016). Oil price volatility forecast with mixture memory GARCH. *Energy Economics*, *58*, 46–58. https://doi.org/10.1016/j.eneco.2016.06.004

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, *28*, 204–230. https://doi.org/10.1002/jae.1271

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting*, *33*, 11–20. https://doi.org/10.1016/j.ijforecast.2016.07.005

Korobilis, D., & Koop, G. (2020). Bayesian dynamic variable selection in high dimensions. *Working Paper*, (pp. 1–51). URL: https://mpra.ub.uni-muenchen.de/100164/1/MPRA_paper_100164.pdf

Lavko, M., Klein, T., & Walther, T. (2023). Reinforcement learning and portfolio allocation: Challenging traditional allocation methods. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4346043

Lee, T.-H., & Yang, Y. (2006). Bagging binary and quantile predictors for time series. *Journal of Econometrics*, *135*, 465–497. https://doi.org/10.1016/j.jeconom.2005.07.017

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*, 293–311. https://doi.org/10.1016/j.jeconom.2015.02.008

Luo, J., Ji, Q., Klein, T., Todorova, N., & Zhang, D. (2020). On realized volatility of crude oil futures markets: Forecasting with exogenous predictors under structural breaks. *Energy Economics*, *89*, 104781. https://doi.org/10.1016/j.eneco.2020.104781

Luo, J., Klein, T., Ji, Q., & Hou, C. (2022). Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. *International Journal of Forecasting*, *38*, 51–73. https://doi.org/10.1016/j.ijforecast.2019.08.007

Luo, J., Marfatia, H. A., Ji, Q., & Klein, T. (2023). Co-volatility and asymmetric transmission of risks between the global oil and China's futures markets. *Energy Economics*, *117*, 106466. https://doi.org/10.1016/j.eneco.2022.106466

Luo, Y., & Young, E. R. (2016). Induced uncertainty, market price of risk, and the dynamics of consumption and wealth. *Journal of Economic Theory*, *163*, 1–41. https://doi.org/10.1016/j.jet.2016.01.003

Ma, F., Liu, J., Wahab, M., & Zhang, Y. (2018). Forecasting the aggregate oil price volatility in a data-rich environment. *Economic Modelling*, *72*, 320–332. https://doi.org/10.1016/j.econmod.2018.02.009

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, *63*, 215. https://doi.org/10.2307/1403615

Marcellino, M. (2006). Chapter 16 leading indicators. *Handbook of Economic Forecasting*, *1*, 879–960. https://doi.org/10.1016/S1574-0706(05)01016-5

Narajewski, M., & Ziel, F. (2020). Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, *19*, 100107. https://doi.org/10.1016/j.jcomm.2019.100107

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, *60*, 1772–1791. https://doi.org/10.1287/mnsc.2013.1838

Nguyen, D. K., & Walther, T. (2020). Modeling and forecasting commodity market volatility with long-term economic and financial variables. *Journal of Forecasting*, *39*, 126–142. https://doi.org/10.1002/for.2617

Nomikos, N. K., & Pouliasis, P. K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics*, *33*, 321–337. https://doi.org/10.1016/j.eneco.2010.11.013

Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2016). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets*, *36*, 758–792. https://doi.org/10.1002/fut.21759

Reboredo, J. C. (2011). How do crude oil prices co-move ? A copula approach. *Energy Economics*, *33*, 948–955. https://doi.org/10.1016/j.eneco.2011.04.006

Ribeiro, P. J. (2016). Revealing exchange rate fundamentals by bootstrap. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2839259

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, *235*, 643–659. https://doi.org/10.1016/j.ejor.2014.01.019

Sizova, N. (2011). Integrated variance forecasting: Model based vs. reduced form. *Journal of Econometrics*, *162*, 294–311. https://doi.org/10.1016/j.jeconom.2011.02.004

Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, *30*, 481–493. https://doi.org/10.1080/07350015.2012.715956

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? *Journal of Empirical Finance*, *54*, 97–117. https://doi.org/10.1016/j.jempfin.2019.08.007

Zhang, Y., Wei, Y., Zhang, Y., & Jin, D. (2019). Forecasting oil price volatility: Forecast combination versus shrinkage method. *Energy Economics*, *80*, 423–433. https://doi.org/10.1016/j.eneco.2019.01.010

AUTHOR BIOGRAPHIES

Jiawen Luo is an associate professor at the South China University of Technology. Her research interests include financial economics, financial risk analysis, energy finance, and so on. Her research are published in journals such as *International Journal of Forecasting, Journal of Forecasting, Journal of Futures Markets, Energy Economics, European Journal of Finance*, and so on.

Tony Klein is a Professor of Finance and holds the Chair of Finance and Banking at Technische Universität Chemnitz, Germany. He previously held positions at Queen's University Belfast, UK, and Universitat de Barcelona, Spain. His research interests include econometric modeling and machine learning within the fields of finance and economics, energy & carbon economics, and the financialization of commodity markets and connectedness to financial markets.

Thomas Walther is a tenured Assistant Professor of Finance with the Utrecht School of Economics, Utrecht University. Formerly, he was Assistant Professor for Energy Finance (Non-Tenure Track) at the School of Finance, University of St. Gallen. He is also a Research Fellow for Finance at the Faculty of Business and Economics, Technische Universität Dresden from which he received a B.Sc. (2010) and M.Sc. (2013) in Industrial Engineering and Management and a PhD (2017) in Business & Economics. His research focuses on commodity/energy/climate finance and financial econometrics with applications to risk management.

Qiang Ji is the executive director of the support center for discipline research in the Institutes of Science and Development, Chinese Academy of Sciences (ISDCAS). He presides the program supported by the Excellent Young Scientists Fund of the National Natural Science Foundation of China and is granted the Excellent Member by the Youth Innovation Promotion Association of CAS. He has been engaged in research work in the fields of energy strategic management, energy and climate finance, energy forecasting, and risk management for a long time. He is the co-founder of China Energy Finance Network and the vice president of the International Society for Energy Transition Studies and Society for the Studies of Climate Finance. He is a founding editor of the *Journal of Climate Finance* and serves as senior editor and associate editor for over 10 energy and financial journals, such as *Energy Economics, Finance Research Letters*, and *International Review of Financial Analysis*. He has published more than 150 SCI/SSCI indexed articles in peer-reviewed journals and has been named in the global "Highly Cited Researchers" lists in the year 2021–2023 released by Clarivate Analytics and "World's Top 2% Scientists" lists released by Stanford University in the year 2020–2023.

How to cite this article: Luo, J., Klein, T., Walther, T., & Ji, Q. (2024). Forecasting realized volatility of crude oil futures prices based on machine learning. *Journal of Forecasting*, 1–25. https://doi.org/10.1002/for.3077

APPENDIX A.

TABLE A1 Constructions of exogenous predictors.

Category	Data	Variable construction
Financial market volatility	DJI, CAC40, FTSE, DAX30, HSI, Nikki225, S&P500, SSEC, STOXX50	We obtain the realized volatilities of the 9 stock indices come from Oxford-Man Institute's "realized library"
Sentiment	VIX, OVX, US EPU, GFSI index	We obtain the four indices from Datastream database. The first difference of the variables is used as the predictor.
Commodity market	Natural gas, Gold, Corn, Soybean	The daily prices of the four commodity futures are obtained from Wind database. The return is computed by $r_t = 100 \cdot (\log P_t - \log P_{t-1})$. We use the square returns r_t^2 of the four commodity futures as the predictors.
Currency market	US dollar index, Euro/US exchange rate, Japanese Yan/US exchange rate, Chinese Yuan/US exchange rate, Russia Rouble/US exchange rate, UK pound/US exchange rate	The daily prices of US dollar index and the 5 foreign exchange rates are obtained from Wind database. The return is computed by $r_t = 100 \cdot (\log P_t - \log P_{t-1})$. We use the square returns r_t^2 of the US dollar index and the 5 foreign exchange rates as the predictors.
Google search volume	Oil production, Financial crisis, Oil demand, Oil price, OPEC conference	The Google search volume indices are downloaded from the Google index database. We employ the first difference of the Google search indices for the keywords as the predictors.

Note: VIX and OVX denote the implied volatility index for S&P 500 and WTI, US EPU is the economic policy uncertainty index for the USA, and GFSI index is the global financial stress index.

APPENDIX B.

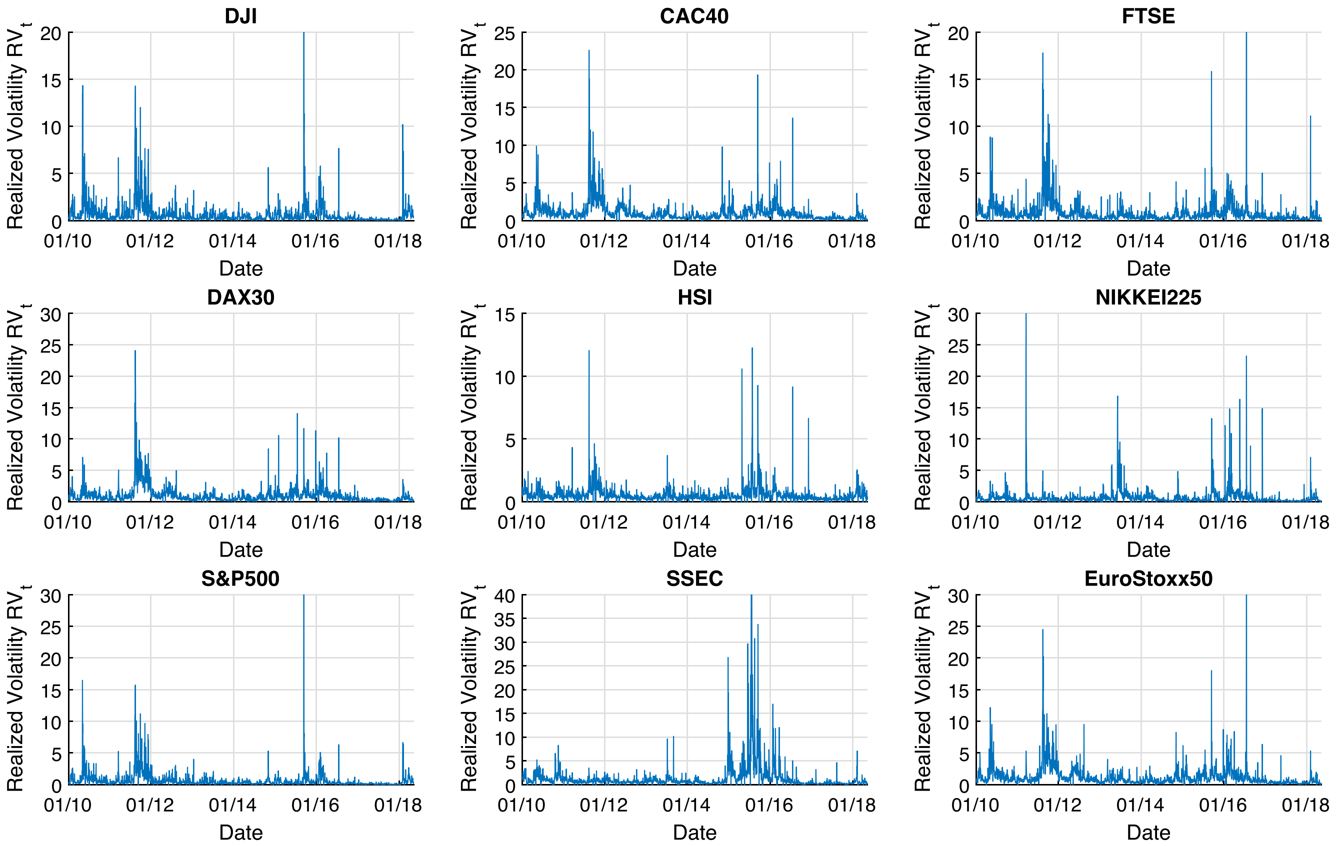

FIGURE B1 Realized volatility of stock indices on daily resolution from January 5, 2010, to May 11, 2018, obtained from the Oxford-Man Realized Library.

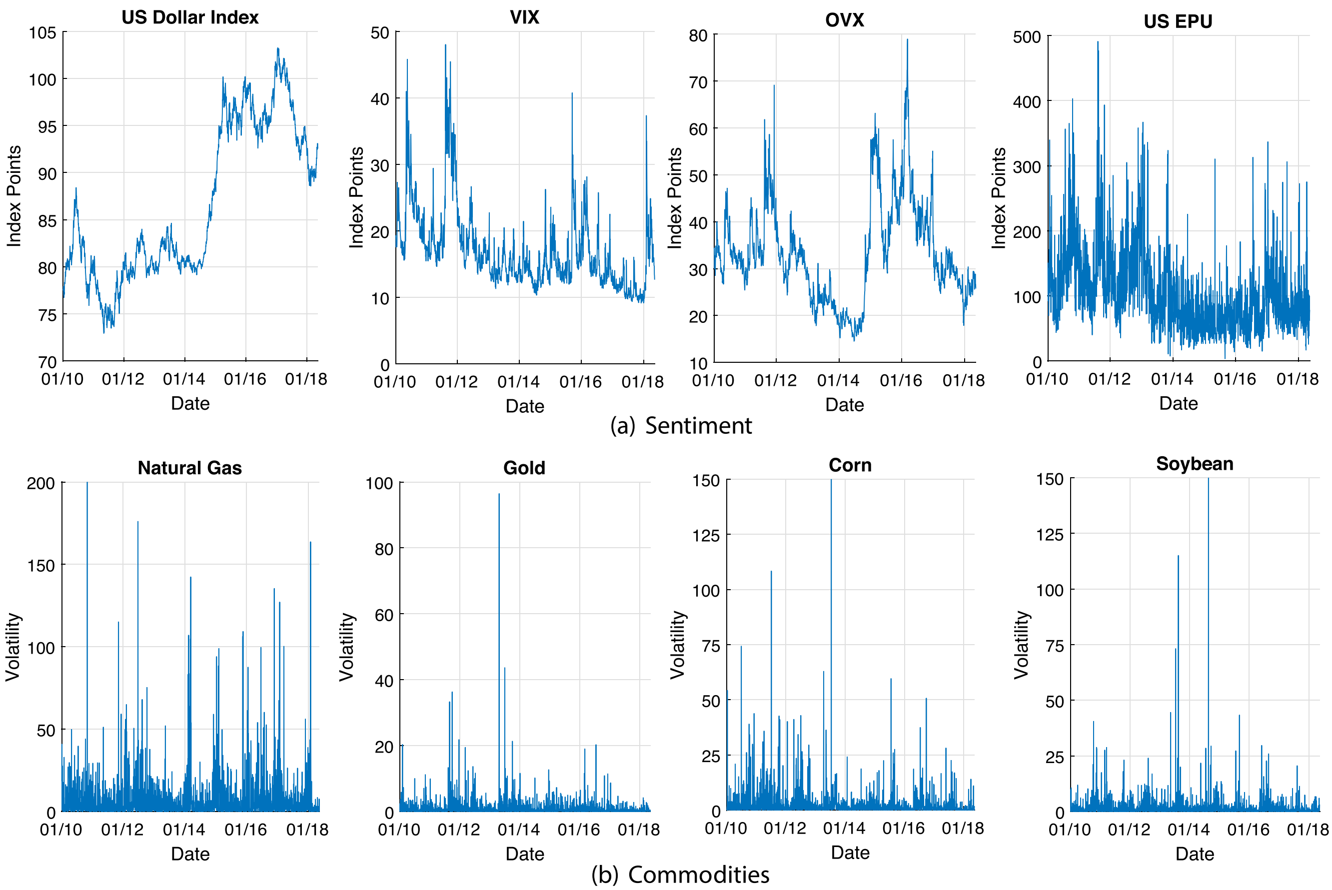

FIGURE B2 Sentiment indices as index points and realized volatility of commodity markets, proxied by squared daily returns, on daily resolution from January 5, 2010, to May 11, 2018, obtained from Datastream.

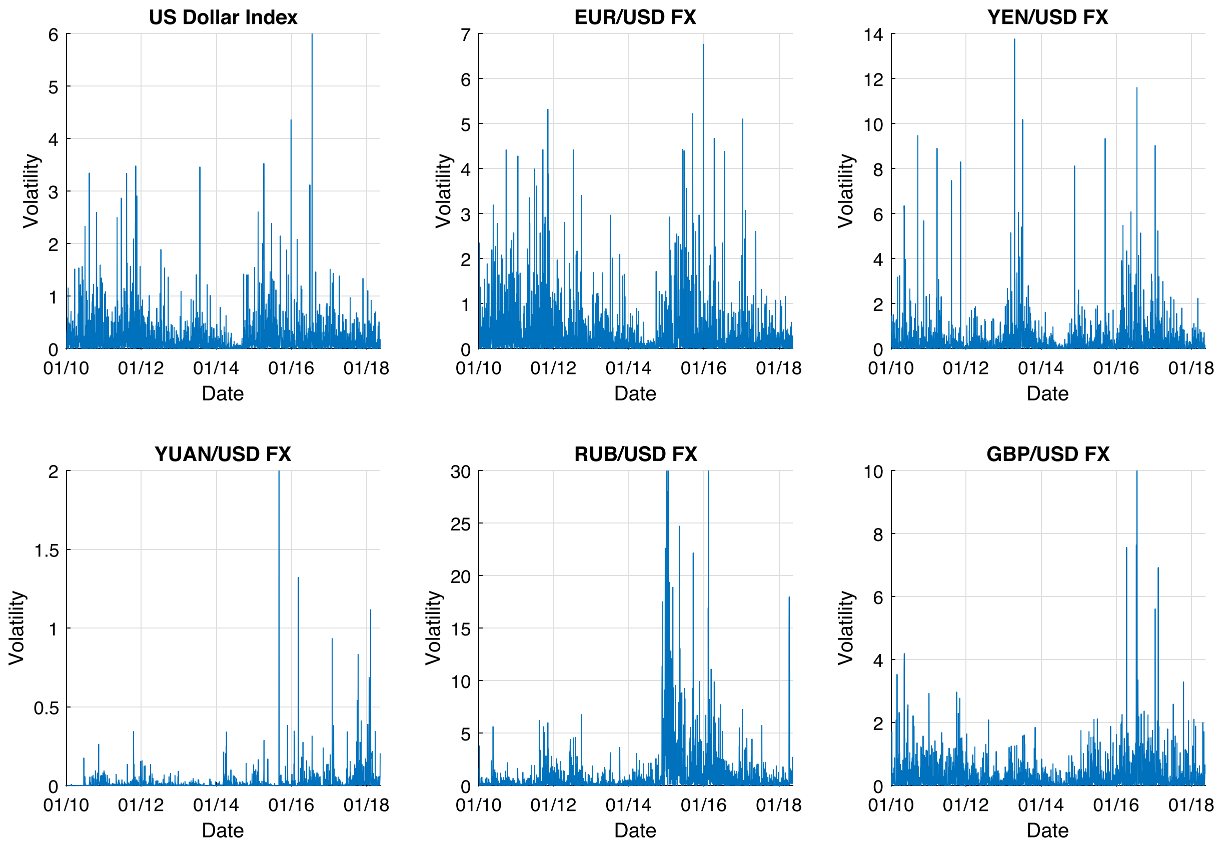

FIGURE B3 Realized volatility of FX markets, proxied by squared daily returns, on daily resolution from January 5, 2010, to May 11, 2018, obtained from Datastream.

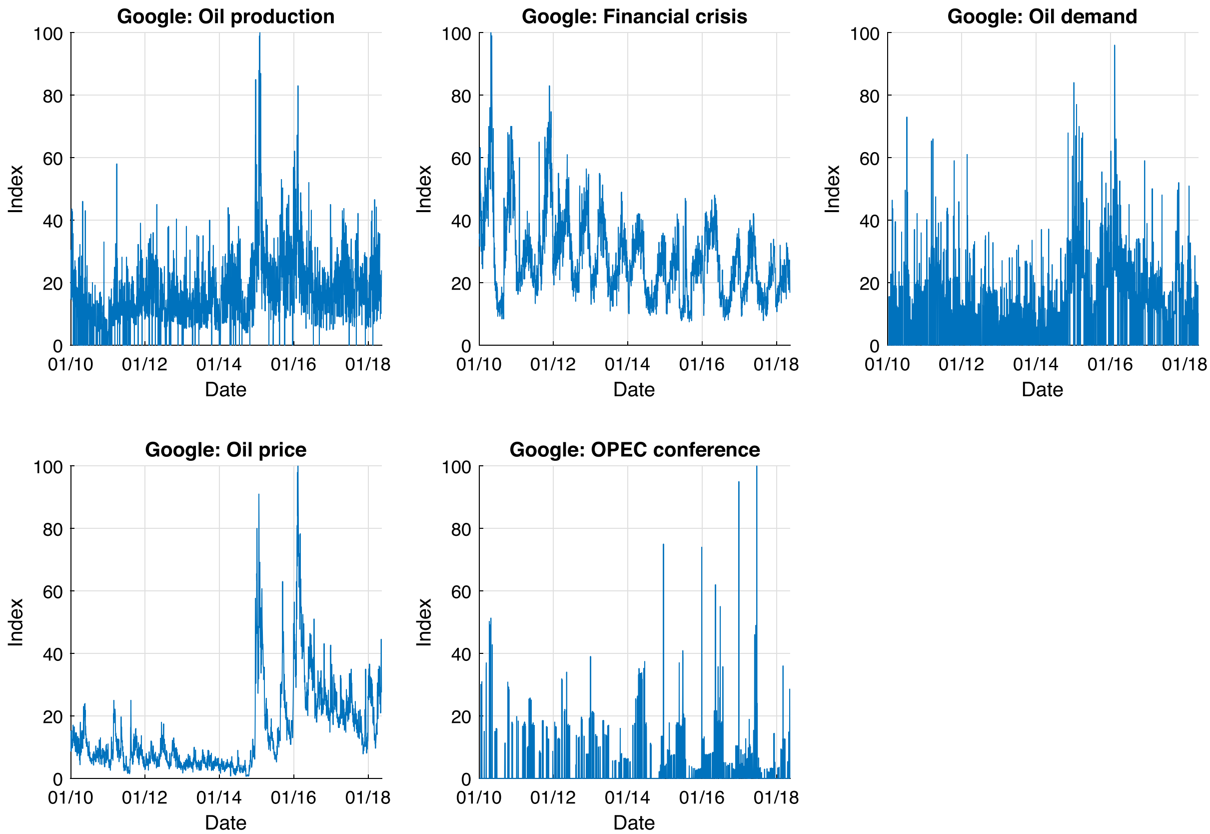

FIGURE B4 Google search volume indices on daily resolution from January 5, 2010, to May 11, 2018, obtained from Google.