Forecasting the volatility of U.S. oil and gas firms: beyond linearity and volatility factors[1]

Juan D. Díaz[2] Erwin Hansen[3] Gabriel Cabrera[4]
University of Chile University of Chile University of Chile

[Version as of February 2022]

Abstract

This paper forecasts the daily volatility of U.S. oil and gas firms using an extended version of the model introduced by Lyócsa and Todorova (2021). We show that forecast accuracy can be significantly improved by relaxing the linearity assumption and using several non-linear machine learning techniques. Furthermore, we demonstrate that besides Lyócsa and Todorova's (2021) volatility factors, economic variables are valuable predictors of firm volatility. Technical indicators also improve forecasting accuracy, albeit to a lesser extent. Finally, we also compare predictions using an economically motivated Value-at-Risk loss function as well as the more standard statistical loss functions.

Keywords: Oil & Gas industry, realized volatility, machine learning, HAR model, forecast combination.

[1] First draft

[2] Department of Management Control and Information Systems, Faculty of Economics and Business, University of Chile. Email: juadiaz@fen.uchile.cl. Address: Diagonal Paraguay 257, Of. 2001, Santiago, Chile.
[3] Department of Business Administration, Faculty of Economics and Business, University of Chile. Email: ehansen@fen.uchile.cl. Address: Diagonal Paraguay 257, Of. 1204, Santiago, Chile.
[4] Faculty of Economics and Business, University of Chile. Email: gcabrerag@fen.uchile.cl. Address: Diagonal Paraguay 257, Santiago, Chile.

1. Introduction

In a recent study, Lyócsa and Todorova (2021) investigate the volatility drivers of U.S. oil and gas firms listed on the S&P 500 index between 2007 and 2017. The authors focus on the role of volatility factors. More specifically, they evaluate the predictive power of each firm's own volatility, industry market volatility, local (U.S.) market volatility, world equity market volatility, oil price volatility and natural gas price volatility. The study concludes that the main drivers of firm volatility are: market volatility, each firm's own volatility and industry level volatility. Moreover, the authors find that by using a HAR model incorporating all the factors, they can make out-of-sample forecasts with four percent fewer errors than their benchmark model, i.e., Corsi's (2009) standard HAR model.

As is customary in the literature, Lyócsa and Todorova undertake their forecasting exercise using a linear framework and ordinary (weighted) least squares estimation methods. In this study, we relax this assumption and investigate to what extent more flexible, non-linear estimation methods improve the accuracy of U.S. oil and gas firm realized volatility forecasts. Specifically, we evaluate several machine learning (ML) methods that provide this flexibility, and which have also been shown to improve volatility forecasting accuracy at the firm level (see e.g., Christensen et al., 2021; Li and Tang, 2021; Rahimikia and Poon, 2020) and at the aggregate level (see, e.g., Mittnik et al., 2015; Bucci, 2020; Díaz et al., 2020). Thus, our first research question regards ML methods' ability to produce more accurate forecasts of U.S. oil and gas firm volatility than linear methods when using the volatility factors from Lyócsa and Todorova (2021).

Our second research question is whether there are any other variables apart from the ones incorporated into Lyócsa and Todorova's (2021) factors that can further improve our ability to forecast firms' realized volatility. Indeed, prior literature has shown that macroeconomic and financial variables are useful predictors of volatility (see, e.g., Christiansen et al., 2012; Engle et al., 2013; Nonejad, 2017; Christensen et al., 2021). Additionally, existing scholarship has shown that technical indicators such as moving average (MA) and momentum (MOM) indicators may improve forecasts of stock market volatility (see, e.g., Díaz et al., 2020; Liu and Pan, 2020; Liu and Tang, 2021) and oil price volatility (see Ma et al., 2018).[5] Hence, this paper assesses to what extent additional economic variables and technical indicators may yield more accurate forecasts of U.S. oil and gas firm realized volatility than just Lyócsa and Todorova's (2021) volatility factors. Finally, we investigate whether evaluating the models using an economic loss function based on value-at-risk (VaR) estimations (see González-Rivera et al., 2004) instead of the more traditional statistical loss function affects their ranking.

In recent years, ML models have been increasingly leveraged in economic and financial forecasting exercises.[6] For example, Gu et al. (2020) evaluated several ML techniques' ability to estimate stock risk premia in the U.S. and Bianchi et al. (2021) studied bond risk premia in the same country. In the same vein, recent research has evaluated the use of ML techniques in volatility forecasting exercises at the aggregate level (see e.g., Mittnik et al., 2015; Díaz et al. 2020; Bucci, 2020) and at the firm level (see e.g., Christensen et al, 2021;

[5] Neely et al. (2014) show that technical indicators improve forecasts of aggregate U.S. stock market returns more than a set of macroeconomic variables. In a similar exercise, Yin and Yang (2016) find that technical indicators possess in-sample and out-of-sample forecasting power for oil prices, which exceeds that of macroeconomic variables and state-of-the-art oil-macro variables.
[6] See, for example, Ghoddussi et al. (2019) for a survey of ML applications in energy economics and finance.

Li and Tang, 2021; Rahimikia and Poon, 2020). Given this evidence, we examine whether regularization methods (Lasso, Ridge and Elastic net) and tree-based methods (random forest and gradient boosting) can improve the accuracy of U.S. oil and gas firm realized volatility forecasts. Initially, we estimate the various ML models using the original set of predictors presented by Lyócsa and Todorova (2021). Subsequently, we expand this set by incorporating additional economic variables and technical indicators.

Our paper provides three main results. Firstly, we show that ML methods can yield more accurate forecasts of oil and gas firm volatility than a benchmark model (the best forecasting model presented by Lyócsa and Todorova, 2021) estimated using a (weighted) least squares method. As such, we show that ML methods are useful for forecasting the volatility of firms in the U.S. oil & gas industry. Estimated forecasting gains vary according to firm and ML method, but they are as high as 5%. Furthermore, when using the MCS test of Hansen et al. (2011) to identify the set of best forecasting models, we find that several ML methods almost always belong to this optimal set, whereas the benchmark model does not. We find that a forecast combination of the five ML methods is the top performing model of firm volatility, followed by the ridge method, which also performs well across forecasting exercises.

Secondly, we show that, besides the volatility factors identified by Lyócsa and Todorova (2021), other economic variables contain valuable information with which to forecast firm volatility. We find that the US 1-month treasury bill and the option-implied VIX index are particularly important in this regard. We also evaluate the value of several technical indicators when predicting firm volatility, but these variables do not have particularly high predictive power. Thirdly, we find that the ML methods' superior ability to forecast firm volatility is not only demonstrated when a statistical loss function is used to evaluate the

models, but also when an economic loss function based on the estimation of a Value-at-Risk metric (González-Rivera et al. 2004) is considered. An economically motivated metric for evaluating the forecasting accuracy of ML methods is of greater use to market practitioners.

The remainder of this paper is structured as follows. In section 2, we describe our dataset; in section 3, we describe the ML algorithms and the forecasting framework; we report our empirical results in section 4; and finally, we conclude in section 5.

2. Data

The database in the paper is an extended version of the data in Lyócsa and Todorova (2021), available online at the Energy Economics journal's webpage. The data contains information of fifteen S&P 500 companies of the sub industry oil & gas exploration & production. It is daily and covers the period January 2007 to December 2017. The main variable of interest in the paper is the realized volatility of each firm in the sample, computed using intraday prices. In particular, the realized volatility in the trading day t is given by $RV_{t,5m}^2 = \sum_{j=1}^{N_{t,5m}} r_{t,j,5m}^2$ where $r_{t,j,5m}^2$ is the continuous return within a 5-minute frequency and $N_{t,5m}$ is the number of return given the sample frequency. Following a suggestion by Patton and Sheppard (2009), the realized volatility at trading day t is computed as the average realized volatilities at the 5, 10, 15 and 30 minutes: $RV_t^2 = (RV_{t,5m}^2 + RV_{t,10m}^2 + RV_{t,15m}^2 + RV_{t,30m}^2)/4$. Finally, Lyócsa and Todorova (2021) adjusted firm realized volatility by considering overnight price variation in its definition as follows $RV_t = \ln(RV_t^2 + OV_t^2)$, where OV_t^2 is an overnight volatility estimate.

To estimate the HAR model of Corsi (2009), used as benchmark model by Lyócsa and Tedorova (2021), weekly and monthly realized volatilities are defined as $RV_t^W = (\frac{1}{5}) \sum_{t=0}^{4} RV_{t-i}$ and $RV_t^M = \left(\frac{1}{22}\right) \sum_{t=0}^{21} RV_{t-i}$.

Initially, we use as predictors the same six volatility factors used in Lyócsa and Todorova (2021). These predictors are the individual firm RV, the industry level RV factor, the US equity market RV estimated using S&P mini-index futures, a world equity market volatility factor estimated with MSCI world market index data, an oil volatility factor estimated using WTI future contracts, and finally, a natural gas volatility factor estimated using Henry Hub futures contracts. Further details on data filtering and variable definitions can be found in Lyócsa and Todorova (2021).

We extended the set of predictors by adding economic variables. In particular, we include the CBOE option-implied volatility (VIX) index, the US 1-month T-bill rate, the Aruoba, Diebold and Scotti (2009) business conditions (ADS) index, and the Economic Policy Uncertainty index from Baker et al. (2016). The selection of these four variables is motivated by recent evidence in Christensen et al. (2021) showing that they help to predict US stock's realized volatility at daily frequencies. All these variables are retrieved from the Federal Reserve Bank of St. Louis (FRED) online database but the ADS index that comes from the Federal Reserve Bank of Philadelphia's webpage.

Finally, we include a set of technical indicators for each firm in the sample. We considered two moving average (MA) indicators and two momentum (MOM) indicators. Each technical indicator defines a signal (S) that takes a value of 1 or 0 for long and short positions,

respectively. The specific definitions of this set of signals are as follows. A moving average signal is given by:

$$S_{i,t} = \begin{cases} 1 \; if \; MA_{s,t} \geq M_{l,t} \\ 0 \; if \; MA_{s,t} < MA_{l,t} \end{cases}$$

where

$$MA_{j,t} = \left(\frac{1}{j}\right) \sum_{i=0}^{j-1} P_{t-1}$$

for $j = s, l$; where P_t is the level of asset price, and $s(l)$ is the length of the short (long) MA ($s < l$). We considered MA rules with $s = 1$ and $l = 5, 22$.

The momentum strategy $MOM(m)$ defines the following signal:

$$S_{i,t} = \begin{cases} 1 \; if \; P_t \geq P_{t-m} \\ 0 \; if \; P_t < P_{t-m} \end{cases}.$$

We evaluated $m = 5, 22$. In the appendix, we provide a list of all the variables and their descriptive statistics.

3. Empirical Setting

This section presents the machine learning methods employed to forecast the realized volatility of oil and gas firms, including regularization methods and regression tree methods. Thereafter, we provide details of the forecasting exercise and the formal testing of differences in forecasting accuracy among ML methods and the benchmark model, which is the best performing model in Lyócsa et al. (2021) including eight factors (the AF-HAR model).

3.1. Machine Learning Methods

3.1.1. Regularization Methods

We consider three types of penalized linear regression: Ridge (Hoerl and Kennard, 1970); Lasso (Tibshirani, 1996); and Elastic Net (Zou and Hastie, 2005). These methods build on the classic linear regression, but the coefficient estimates are obtained by minimizing a penalized residual sum of squares. Specifically, consider the following model:

$$R = X\beta + \epsilon \qquad (1)$$

where R is a vector that includes the outcome variable of interest, X is a matrix comprising the predictors included in the model, β is a vector of coefficients to be estimated, and ϵ is the vector of error terms, each one independent and identically distributed (i.i.d.), with a mean of zero and constant variance. The Elastic Net method estimates β by solving the following optimization problem:

$$\min_{\beta} \|R - X\beta\|_2^2 + \lambda \cdot (\alpha \cdot \|\beta\|_1 + (1 - \alpha) \cdot \|\beta\|_2^2) \qquad (2)$$

for $\lambda > 0$ and $0 \leq \alpha \leq 1$, where $\|\cdot\|_1$ is the l^1 norm and where $\|\cdot\|_2$ is the Euclidean norm. In this setting, the tuning parameters λ and α are selected via cross-validation.

It is important to note that if we set $\alpha = 0$, then the optimization problem in (2) corresponds to that of a Ridge regression, while if we set $\alpha = 1$, it corresponds to that of a Lasso. Regardless of whether $\alpha = 0$ or $\alpha = 1$, the tuning parameter λ is selected via cross-validation.

3.1.2. Tree-based Methods

We consider two machine learning methods based on trees: the Random Forest (RF, Ho, 1995; Amit and Geman, 1997; Breiman, 2001) and Gradient Boosting Regression Tree (GBRT, Hastie et al., 2009, chapter 10).

Both methods are ensemble statistical methods that extend the single regression tree (Breiman et al., 1984) by combining predictions from multiple regression trees. Specifically, RFs exploit the idea of bootstrap aggregation (bagging, Breiman, 1996) by computing their final prediction as the average of those from several single regression trees. These single regression trees are adjusted with random subsets of training data and predictors. Meanwhile, based on the idea of boosting (Freund and Schapire, 1996; Schapire, 2003), GBRTs provide better prediction accuracy than single regression trees by sequentially decreasing the bias of simple regression trees with higher bias but low variance, where the contemporary ensemble regression tree model internalizes the knowledge acquired from previously fitted models.

In practice, these methods have tuning parameters that must be set. For the RF, we choose to randomly select eight predictors at each split; we employ 1,500 regression trees to be aggregated; and we estimate trees with five terminal nodes (their complexity parameter). For the GBRT, we determine the number of trees to be aggregated in the final model via cross-validation; we adjust regression trees with only one split; and we use a learning rate of 0.01.

3.2 Forecast Combination

Prior Literature has shown that forecast combination has been successfully produce more accurate forecasts (See, for example, Rapach et al. 2010 and references therein). Therefore,

we perform two exercises to combine forecasts as a predicting method. First, following Zhang et al. (2020), we combine the forecast of a particular ML model with the benchmark model. Zhang et al. (2020) show that in the case of aggregate stock return predictability, this approach outperforms forecasts from individual ML methods. The second exercise is to combine the forecast of all the ML methods. For both exercises, we use equal weights.

3.3. Forecasting Framework

Our forecasting procedure aims to resemble the one in Lyócsa and Todorova (2001) to compare as precisely as possible their best model, our benchmark, and forecast estimates coming from alternative ML methods and additional predicting variables. Specifically, we use a rolling window estimation procedure with an estimation window of 1,000 observations. Thus, we estimate each model in a particular window and forecast one period ahead. Then, we move the window one day forward, predict one day again, and so on. We repeat this procedure until the last observation in our sample is forecast.

As mentioned above, we use a sample period starting in January 2007 and ending in December 2017. To simulate a real out-of-sample forecasting exercise, we split the sample into training and testing subsamples and use the former to adjust the models and the latter to assess their predictive accuracy. We begin our empirical analysis by using a training subsample including the first 1,000 days from January 2007. The testing subsample begins the day 1001 up to December 2017.

The rolling window approach uses all the information in a window frame to train the models. Then, the estimated models are used to compute each new prediction in the testing sample.

For example, to forecast the first day in the testing sample, we begin by fitting the model using the initial training sample (i.e., all the information prior to that first day in the testing sample). Then, we rolled forward the training sample by one day to predict the second day in the testing sample, reestimated the models, and predicted for day two. We proceed in this manner until we obtain forecasts for all the days in the testing sample.

The accuracy of the forecasts is evaluated using three loss functions: the (mean) squared forecasts error, the Q-LIKE function (see Lyócsa and Todorova 2021, and Patton, 2011), and a Value-at-Risk (VaR) loss function (see González-Rivera et al., 2004). These functions are defined as follow:

$$MSFE = \frac{1}{M}\sum_{t=1}^{M}(RV_t - RV_t^F)^2 \ ,$$

$$QLIKE_t = \frac{RV_t}{RV_t^F} - ln\left(\frac{RV_t}{RV_t^F}\right) - 1 \ ,$$

$$l_t = (\tau - d_t^\tau)(y_t - VaR_t^\tau),$$

where VaR_t^τ is the τ-level predicted VaR_t with information up to time t, and $d_t^\tau = (y_t < VaR_t^\tau)$ is the τ-level quantile loss function. The predicted VaR_t is computed as $\Phi^{-1}(\alpha)RV_t^F$. As it is well known, the MSFE loss function is symmetric, meaning that both over and under estimation of the realized volatility are equally penalized, while the Q-LIKE loss function penalized more under estimation than over estimation. The VaR-type of loss function is also asymmetric and penalized more those realizations below the threshold of interest ($y_t < VaR_t^\tau$). The VaR lost function is also of interest as it provides an economic assessment of the realized volatility forecast as it is pointed out by González-Rivera et al. (2004) instead of a more standard purely statistical assessment of predictions. In other words, the VaR loss

functions evaluates the accuracy of alternative models by looking at to an economic metric that may be used by investors in the real-world application.

We evaluate and compare the forecasting performance of the alternative models using the model confidence set (MCS) test of Hansen et al. (2011). The test works sequentially by eliminating the worst performance model among a set of M models until identify those models for which the null hypothesis of equal forecast accuracy cannot be rejected at a significant level α. The MCS contains the best performing models from M at the $(1-\alpha)$ confidence level. We set $\alpha = 0.1$ and use the statistics T_R (see details in Hansen et al., 2011) with 5,000 bootstrap simulations.

4. Empirical Results

In this section, we report our empirical results. We start reporting a set of results using the same predictors of Lyócsa and Todorova (2021) to quantify forecasting gains associated with the use of ML methods rather than linear least squared methods. In this way, we assess the role of non-linear models in improving forecasting accuracy exclusively. After that, we expand the set of predictors and investigate their role in improving forecasting accuracy. We considered two sets of additional predictors: economic variables and technical indicators. See the data section above for a detailed list of these variables. Thus, after reporting baseline results using the predictors in Lyócsa and Todorova (2021), in the second set of results, we report results using, in addition, the economic variables and the technical indicators as predictors. In the appendix, we report additional results, including only the set of economic variables or technical indicators independently.

4.1 Beyond Linear Models

Table 1 reports the forecasting performance of a set of ML methods in predicting the realized volatility of fifteen firms in the oil & gas industry. Each column reports the results for a given firm. We use the Q-LIKE loss function to perform the comparison between models. The benchmark model is the best performing model in Lyócsa and Tedorova (2021). The symbol for this benchmark is "af-har" which means all factors evaluated by Lyócsa et al. (2021) plus the HAR specification, including the own lagged realized volatility and the weekly and monthly own realized volatility for each firm. Recall that the "af-har" model includes, in addition to the HAR model standard components, an industry, a US stock market, a world stock market, an oil factor, and a gas factor. As mentioned above, this benchmark outperforms the competing models in Lyócsa et al. (2021) as it provides forecasting accuracy gains of about 3% compared with the HAR model of Corsi (2009).

For each firm in the sample, Table 1 reports the percentage of relative gains/losses in forecasting accuracy compared with the benchmark model (af-har). A negative value indicates that the evaluated model produces a lower value of the Q-LIKE loss function, and therefore, it delivers superior performance than the benchmark model. Conversely, a positive value indicates a higher realization of the loss function and, thus, a worst relative performance against the benchmark model.

After reporting the benchmark in the first row of the table, we report the relative performance of the ML methods: ridge, lasso, elastic net, random forest, and gradient boosting methods. Below, we report the simple equally weighted combination between the benchmark and ML method and label it as "af-har +." Finally, the last row of the table reports the equally weighted forecast combination of the five ML methods ("pooled-avg.").

Table 1 shows a superior forecasting performance of some of the ML methods evaluated for all the oil & gas firms in the sample. When we look at the performance of the ML methods independently, we observe gains in forecasting accuracy for the regularization methods but not necessarily for the tree-based methods. For example, for the firm "apa," as compared with the benchmark model, the Q-LIKE loss function is reduced by 0.75% when a ridge estimator is used to forecast. Still, when the random forest is used instead, we observe that the estimated loss function increases by 1.25% as compared to the benchmark. In this case, the GBM procedure also produces more accurate forecasting estimates than the benchmark as the loss function drops by 0.57%. For most other firms in the sample, we observe a similar pattern where regularization methods outperform the benchmark. The ridge method performs best, followed by the elastic net and lasso methods. The tree-based methods tend to produce poorer forecasting results than the benchmark model. The random forest model produces the worst performance, and the GBM estimates show slightly better results considering that four out of fifteen firms can beat the benchmark.

We observe better results when the induvial ML methods are combined with the benchmark model. Indeed, in most of the cases, the forecast combinations reduce the loss function value estimate indicating more accurate predictions than the benchmark model. The gain in forecasting varies between values of 0.10% and 3.3% depending on the ML method considered and the firm. We find that the best performing ML methods are the tree-based regression methods. In particular, the GBM method that in half of the firms (8 out of 15) is the method that produces more accurate forecast when combined with the benchmark. The average gain across firms produced by the GBM method is 1.2% as compared with the benchmark model, followed by the tree-based method that produces an average gain of 0.7%.

There are firms in which these relative improvements vary between 2% and 3%, which is comparable in magnitude with the original gains reported in the forecasting exercise in Lyócsa et. al (2021).

Akin to the performance of the forecast combination of an individual ML method and the benchmark model, the performance of the forecast combination that consider the five ML methods simultaneously is even better. The average forecasting gain across firms is 1.5%, the highest among the alternatives considered before. The MCS test confirms the prior findings. First, we find that for any of the firms in the sample, the benchmark model (af-har) belongs to the optimal set of forecast models. This finding supports the idea that relaxing the linearity model implicit in the forecasting exercise in Lyócsa et al. (2021), significantly improves the accuracy of firm's realized volatility. Among the ML methods evaluated, we find that the ridge method is in the optimal set for 12 out 15 firms according to the MCS test, the forecast combination between the GBM methods and the benchmark model is in the optima set for 13 firms, and finally, the forecast combination including the five ML methods is the best performing model as it is included in the optimal set for the 15 firms in the sample.

Table 2 shows forecasting performance results using an economic loss function, a VaR-based loss function (Gonzalez-Rivera et al. 2014), rather than a statistical loss function such as the Q-LIKE used in Table 1. From a practical point of view, an economic loss function like this might be more informative for investors assessing firm's risk in a real-world application. In general terms, the results reported in Table 2 support Table 1. In most of the evaluated cases, we observe that ML methods improve the forecasting accuracy of a firm's realized volatility compared with the benchmark model (af-har). Moreover, the estimated gains are higher in percentage terms than those obtained using the Q-LIKE functions. For example, the relative

improvement can now be as high as 5% (firm "chk" using the GMB method). When we look at the performance of the individual ML method, as in Table 1, we find that the ridge method produces a very accurate forecast of the firm's realized volatility. However, now the tree-based methods perform very well also. When we look at the combinations between a single ML and the benchmark, tree-based models provide the more significant improvements in relative forecasting, with the superior performance of the GBM methods over the random forest method. Finally, the forecast combination that uses the five ML methods simultaneously also performs very well, being the best forecasting method among all the evaluated alternatives in several cases.

Importantly to support our first main conclusion, the MCS test shows that the benchmark model never belongs to the optimal set of forecast models, corroborating our prior results in Table 1. The MCS test shows that the forecasts from the ridge model are in the optimal set of predicting models for 13 firms, while the random forest is in the set for ten firms, and the GBM for all the 15 firms in the sample. When we look at combinations, the benchmark and the GBM model belong to the optimal set for 12 firms. Finally, the forecast combinations of the five methods are in the optimal set for all but one firm.

Figure 1 shows a heat map identifying the relative importance of the variables by ML method and firm. Dark (light) red color indicates more (less) importance in estimating the forecasting model. Similarly, to Lyócsa et al. (2021), we observe that among the volatility, the most important are the HAR factors (the lagged own volatility (rv_l1), the lagged weekly volatility (rv_l5), and the lagged monthly volatility (rv_l22)) and the industry volatility factor (irv_l1).

Overall, we have shown so far that ML methods, used either individually or combined, produce superior firm's realized volatility than the linear volatility factor model used as our

benchmark model. It is worth noting that our results are obtained using the same dataset that Lyócsa et al. (2021), so they are not explained for differences in data sources or sample coverage. Also, our results are robust to alternative loss functions (Q-LIKE and VaR-based). In terms of methods, we find the ridge method, the GBM, and forecast combinations produce superior forecasting than competing methods.

4.2 Additional Predictors

This section reports results using the same set of ML methods as the previous section. Still, instead of only using the AF-HAR model of Lyócsa and Todorova (2021) containing volatility factors, we incorporate an additional set of predictors. We include a set of four macro-finance variables motivated by recent evidence in Christiansen et al. (2021), showing that this set of variables contains predictive power on the realized volatility of US firms listed at a daily frequency. We use the Aruoba, Diebold, and Scotti (2009) business conditions index (ADS), the US 1-month T-bill rate, the Economic Policy Uncertainty (EPU) index from Baker et al. (2016), and CBOE implied volatility index (VIX). Besides, we incorporate a set of technical indicators for each firm in the sample. We consider moving average and momentum technical indicators. See in section 2 details on the way we build these technical indicators. Thus, we use three sets of predictors in our analysis, the "originals" variables (those in Lyócsa and Tedorova, 2021), the economic variables, and the technical indicators. In this section, we report results using the three sets together as predictors to save space. In the appendix, we provide different results in which the original set of variables is expanded with either the macro variables or the technical indicators.

Table 3 reports our results using the Q-LIKE loss function. The results indicate that the additional set of predictors improves forecasting accuracy for all the firms in the sample. The average improvement in the loss function, as compared with the benchmark model, is 2.28%. However, there are cases in which ML methods produce reductions in the loss function of around 5.5%. These magnitudes are significant if we compare them with the gains reported in Lyócsa and Todorova (2021) of about 4%. We observe that in this case, in which we have more potential predictors, the best performing model is the forecast combination of the 5 ML methods used throughout. For the fifteen firms in the sample, the reported average gain for this method is 3.9%, with a minimum gain of 1.30% (rrc) and a maximum gain of 5.68% (eog). Importantly, the MCS test show this forecast combination method is in the optimal set for the fifteen firms in the sample. At the same time, the benchmark is never included, confirming that ML methods provide additional gains in forecasting accuracy for a firm's volatility in the oil and gas industry.

Another ML method that works well is the ridge method, which is included in the optimal set of models for ten out of fifteen firms. In this set of results, we observe that except for the random forest and GBM methods, all the other ML estimates show some degree of predictability, not as high as the top performer, but they are included in the optimal set of models by the MCS test for between five to eight firms in the sample.

We find that economic variables are more important than technical indicators in improving the forecasting ability of estimated models. Figure 2 shows a heat map with the relative importance of the forecasting variables by the individual ML method and firm. Dark (light) red color indicates more (less) importance in estimating the forecasting model. Similarly, Lyócsa and Teodorova (2021) observe that the volatility factors are significant predictors of

a firm's volatility across methods and firms. We find that HAR factors are the most important variables: the lagged own volatility (rv_l1), the lagged weekly volatility (rv_l5), and the lagged monthly volatility (rv_l22). We also identify the industry volatility as an important driver of a firm's volatility.

Regarding the additional predictors we added, we observe that two variables appear particularly important: the US 1-month Treasury bill (dgs1mo_l1) when regularization methods are used and the VIX index when tree-based models are used. It is worth noting that the VIX index, which is built using option information, subsumes the stock market volatility factor identified as Lyócsa and Todorova (2021) as a significant predictor. This evidence is consistent with prior evidence showing the superior informational content of option-implied volatility over past volatility (used to compute realized volatility measures) in predicting volatility.

Table 4 repeats our prior exercises with an extended set of predictors using the VaR-based loss function. The broad picture reported in Table 3 does not change in qualitative terms. We continue observing that ML methods add value in forecasting a firm's volatility in the oil and gas industry; however, the benchmark model shows a better forecasting performance as it is included in the optimal set of models identified by the MCS test case of four firms. As before, the forecast combination using the five ML methods is the best performing model, followed by the ridge method. These two methods are included in the optimal set of models for all the firms in the sample. The tree-based methods show an excellent forecasting performance when using the VaR-based loss function, considering the random forest and GBM methods are included in the optimal set of models for thirteen and fifteen firms, respectively.

5. Conclusions

In a recent study, Lyócsa and Todorova (2021) evaluate several volatility factors' ability to predict U.S. oil and gas firms' realized volatility. Among the set of six volatility factors the authors evaluate, they conclude that the market equity volatility factor, each firms' own volatility and the industry factor are the best predictors. Furthermore, the authors show that a HAR model that has been extended with all the factors (AF-HAR) produces one-day ahead forecasting reductions of up to 4% as compared to a benchmark model (the standard HAR model from Corsi, 2009). The empirical analysis relies on a linear model estimated using ordinary (weighted) least squares.

In this paper, we relax the linearity assumption and investigate the forecasting performance of a set of more flexible non-linear ML methods. Furthermore, we extend the set of predictors by including economic variables and technical indicators in addition to the volatility factors from Lyócsa and Todorova (2021). Our results indicate that allowing for more flexible models reduces errors in U.S. oil and gas firm realized volatility forecasts by up to an additional 5% as compared with our benchmark model, the AF-HAR model from Lyócsa and Todorova (2021). Besides, we find that economic variables such as the US 1-month treasury bill and the option-implied VIX index contain useful information with which to forecast the realized volatility of firms in the U.S. oil & gas industry. The predictive power of the technical indicators is more limited than that of the economic variables we evaluated. Finally, we show that the superior forecasting performance of the ML models (in comparison to the AF-HAR model) is confirmed by a statistical loss-function based on Value-at-Risk estimates.

This paper contributes to the small body of literature examining the drivers of U.S. oil and gas firms' realized volatility by evaluating the performance of ML methods that have recently

been used to forecast equity volatility at the firm level (see e.g., Christensen et al., 2021; Li and Tang, 2021; Rahimikia and Poon, 2020) and at the aggregate level (see, e.g., Mittnik et al., 2015; Bucci, 2020; Díaz et al., 2020).

References

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation* 9(7), 1545-1588.

Antonakakis, N., Cunado, J., Filis, G., Gabauer, D., & De Gracia, F. P. (2018). Oil volatility, oil and gas firms and portfolio diversification. *Energy Economics*, *70*, 499-515.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417-427.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. The Review of Financial Studies, 34(2), 1046-1089.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, CA, USA.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24 (2): 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1): 5-32.

Bucci, A. (2020). Realized volatility forecasting with neural networks. Journal of Financial Econometrics, 18(3), 502-531.

Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. Journal of Applied Econometrics, 27(6), 956-977.

Christensen, K., Siggaard, M., & Veliyev, B. (2021). *A machine learning approach to volatility forecasting* (Vol. 3). Department of Economics and Business Economics, Aarhus University.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, *138*(1), 291-311.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174-196.

Díaz, J., Hansen, E., & Cabrera, G. (2020). Predictive Regressions for Aggregate Stock Market Volatility with Machine Learning. Available at SSRN 3824789.

Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics, 95(3), 776-797.

Feng, J., Wang, Y., & Yin, L. (2017). Oil volatility risk and stock market volatility predictability: Evidence from G7 countries. *Energy Economics*, *68*, 240-254.

Fragoso, T., Bertoli, W., & Louzada, F. (2018). Bayesian Model Averaging: A Systematic Review and Conceptual Classification. International Statistical Review 86 (1), 1-28.

Freund, Y., & Schapire, RE (1996). Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, July 3–6, 1996, Bari Italy. pp. 148–156. Morgan Kaufman, San Francisco, CA, USA.

Gambella, C., Ghaddar, B., & Naoum-Sawaya, J. (2020). Optimization problems for machine learning: A survey. *European Journal of Operational Research*.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727.

González-Rivera, G., Lee, T. H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of forecasting, 20(4), 629-645.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223-2273.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453-497.

Hastie T., Tibshirani R, & Friedman J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Ho, T. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition. 1: 278–282 vol.1.

Hoerl, A. & Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 12 (1): 55–67.

Huck, N. (2019). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, *278*(1), 330-342.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689-702.

Li, S. Z., & Tang, Y. (2021). Forecasting realized volatility: An automatic system using many features and many machine learning algorithms. Available at SSRN 3776915.

Liu, L., & Pan, Z. (2020). Forecasting stock market volatility: The role of technical variables. Economic Modelling, 84, 55-65.

Lyócsa, Š., & Todorova, N. (2021). What drives volatility of the US oil and gas firms?. *Energy Economics*, 105367.

Ma, F., Liu, J., Wahab, M. I. M., & Zhang, Y. (2018). Forecasting the aggregate oil price volatility in a data-rich environment. Economic Modelling, 72, 320-332.

Mittnik, S., Robinzonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of banking & Finance*, *58*, 1-14.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, *60*(7), 1772-1791

Nonejad, N. (2017). Forecasting aggregate stock market volatility using financial and macroeconomic predictors: Which models forecast best, when and why?. Journal of Empirical Finance, 42, 131-154.

Rahimikia, E., & Poon, S. H. (2020). Machine learning for realised volatility forecasting. *Available at SSRN 3707796*.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821-862.

Schapire, R. (2003). The boosting approach to machine learning – an overview. MSRI Workshop on Nonlinear Estimation and Classification, 2002 (eds D.D. Denison, M. H. Hansen, C. Holmes, B. Mallick & B. Yu). Springer, New York.

Tjaaland, S. H., Westgaard, S., Osmundsen, P., & Frydenberg, S. (2015). Oil and gas risk factor sensitivities for US energy companies. *The Journal of Energy and Development*, *41*(1/2), 135-173.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*. Wiley. 58 (1): 267–88.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455-1508.

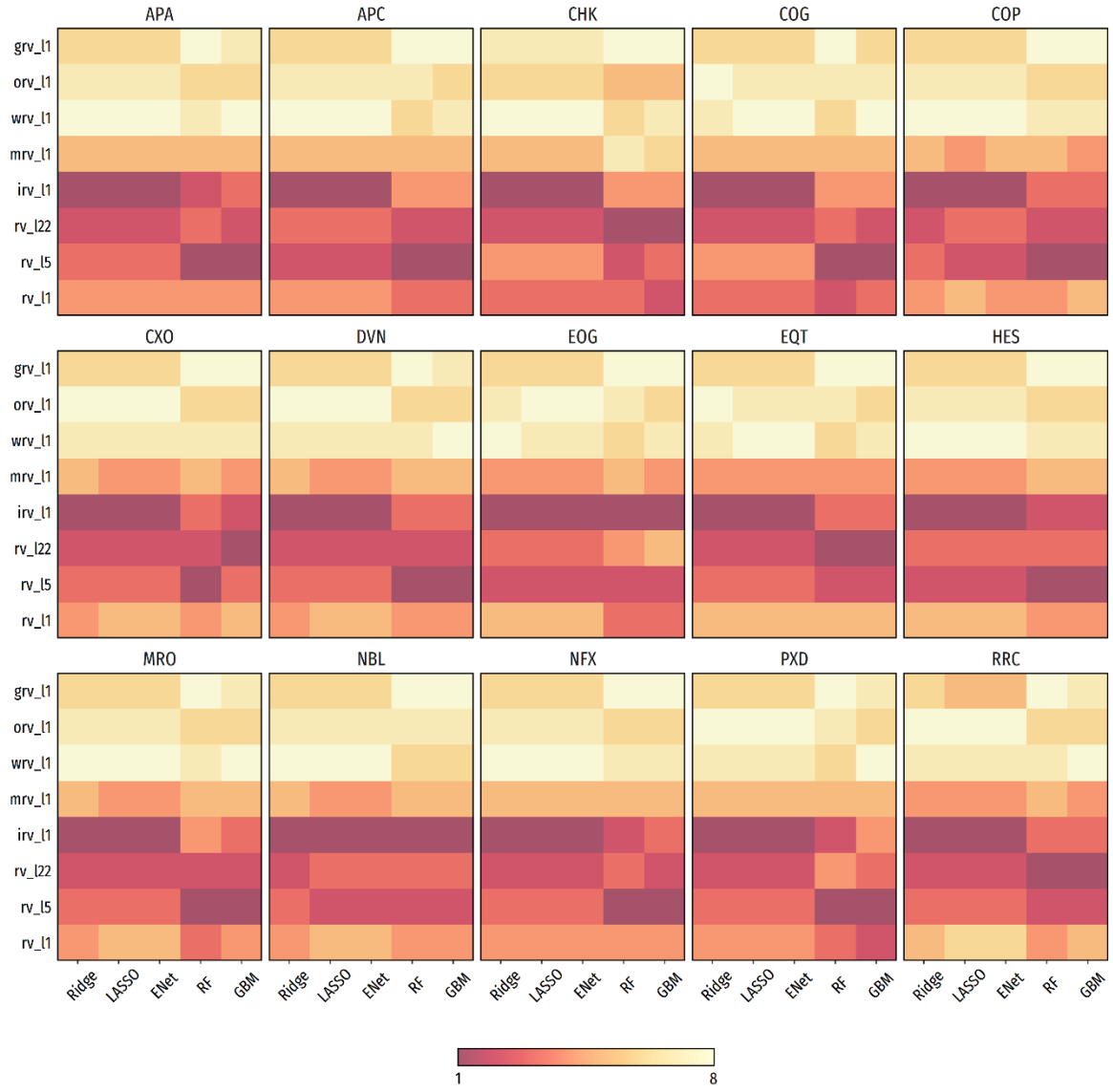
Yin, L., & Yang, Q. (2016). Predicting the oil prices: do technical indicators help?. Energy Economics, 56, 338-350.

Zhang, H., He, Q., Jacobsen, B., & Jiang, F. (2020). Forecasting stock returns with model uncertainty and parameter instability. *Journal of Applied Econometrics*.

Zhang, Y., Ma, F., & Liao, Y. (2020). Forecasting global equity market volatilities. *International Journal of Forecasting*, *36*(4), 1454-1475.

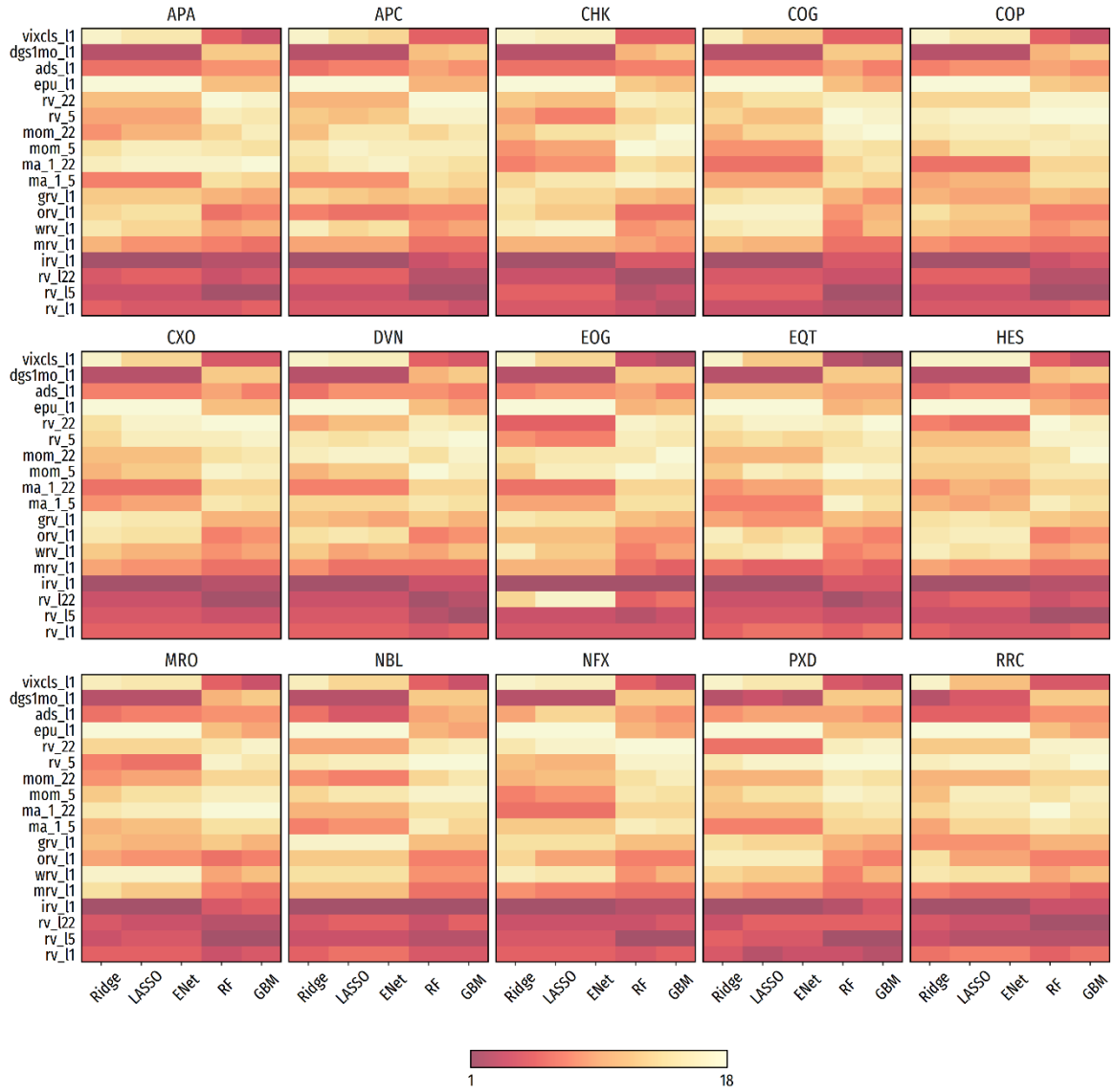
Zou, H. & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*. 67 (2): 301–320

Figure 1: Firm's volatility drivers (original volatility factors in Lyócsa and Todorova 2011)



Notes: The heat map shows the variable importance in predicting firm' realized volatility for the set of predictors used in Lyócsa et al. (2021) across different ML methods and for each firm (firm's nemo above each box). A darker color (1) indicates higher importance of a given predictor, and a lighter color (8) indicates lower importance in predicting. The set of predictors used are the own firm's lagged volatility (rv_l1), lagged weekly volatility (rv_l5), lagged monthly volatility (rv_l22); the industry volatility (irv_l1), the US stock market volatility (mrv_l1), the World stock market volatility (wrv_l1), the oil market volatility (oil_l1) and the gas market volatility (grv_l1).

Figure 2: Firm's volatility drivers (original volatility factors plus economic variables and technical indicators)



Notes: The heat map shows the variable importance in predicting firm' realized volatility for the set of predictors used in Lyócsa et al. (2021) plus economic variable and technical indicators predictors, across different ML methods and for each firm (firm's nemo above each box). A darker color (1) indicates higher importance of a given predictor, and a lighter color (18) indicates lower importance in predicting. The set of predictors used are the own firm's lagged volatility (rv_l1), lagged weekly volatility (rv_l5), lagged monthly volatility (rv_l22); the industry volatility (irv_l1), the US stock market volatility (mrv_l1), the World stock market volatility (wrv_l1), the oil market volatility (oil_l1), the gas market volatility (grv_l1), moving average indicators (ma_1_5 and ma_1_22), momentum indicators (mom_5 and mom_22), the economic policy

uncertainty index (epu_l1) of Bekaert et al. (2016), the Aruoba, Diebold and Scotti (2009) business conditions index (ads_l1), the US 1-month T-bill rate (dgs1mo_l1), and the CBOE volatility (VIX) index.

Table 1: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova, 2021; loss function: QLIKE)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.444	0.432	0.410	0.432	0.570	0.385	0.415	0.455	0.435	0.461	0.427	0.450	0.368	0.412	0.349
ridge	-0.75%	-0.86%*	-0.77%*	-0.36%	-1.01%*	-0.54%*	-0.79%*	-0.77%*	-0.73%*	-0.76%*	-0.65%*	-1.11%	-0.99%*	-0.72%*	-0.66%*
lasso	-0.15%	-0.67%*	-0.35%*	-0.08%	-0.22%	-0.36%*	-0.14%	-0.23%	-0.30%*	-0.52%*	-0.26%*	-1.01%	-0.35%	-0.24%	-0.32%*
enet	-0.14%	-0.79%*	-0.41%*	-0.22%	-0.46%	-0.31%*	-0.14%	-0.36%	-0.38%*	-0.51%*	-0.32%*	-0.95%	-0.52%	-0.38%	-0.45%*
rf	1.25%	2.71%	4.53%	0.45%	4.49%	4.47%	4.42%	3.47%	4.87%	5.97%	3.10%	-0.64%	4.51%	2.16%	7.56%
gbm	-0.57%	-1.03%*	2.62%	-0.98%*	0.65%	1.67%	2.27%	0.01%	3.39%	2.40%	2.56%	-2.49%*	1.09%	0.32%	4.21%
af-har + ridge	-0.44%	-0.50%*	-0.46%	-0.23%	-0.58%	-0.33%*	-0.48%	-0.45%	-0.44%*	-0.47%*	-0.42%*	-0.67%	-0.56%	-0.42%	-0.40%*
af-har + lasso	-0.10%	-0.38%	-0.20%	-0.08%	-0.15%	-0.20%*	-0.09%	-0.14%	-0.17%	-0.30%*	-0.16%*	-0.56%	-0.20%	-0.15%	-0.20%*
af-har + enet	-0.11%	-0.44%*	-0.24%	-0.15%	-0.26%	-0.19%*	-0.09%	-0.20%	-0.21%	-0.30%*	-0.20%*	-0.53%	-0.29%	-0.23%	-0.27%*
af-har + rf	-2.02%*	-1.20%	-0.17%	-2.33%*	-0.47%	-0.40%*	-0.18%	-0.85%	0.13%	0.48%	-1.18%*	-3.05%*	0.12%	-1.09%*	1.49%
af-har + gbm	-2.18%*	-2.17%*	-0.49%*	-2.20%*	-1.61%*	-0.98%*	-0.54%*	-1.72%*	0.14%	-0.39%*	-0.69%	-3.34%*	-0.86%*	-1.34%*	0.64%*
pooled-avg	-2.32%*	-2.19%*	-1.04%*	-2.39%*	-1.58%*	-1.18%*	-0.89%*	-1.69%*	-0.55%*	-0.72%*	-1.42%*	-3.57%*	-1.00%*	-1.57%*	0.24%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the Q-LIKE loss function for the benchmark model (af-har) and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table 2: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova, 2021; loss function: Value-at-Risk)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.866	0.947	2.240	1.243	0.637	1.057	1.000	0.834	0.730	0.890	1.278	0.940	1.303	1.003	1.230
ridge	-1.44%*	-1.33%*	-2.68%*	-0.98%*	-1.92%	-1.07%*	-1.83%*	-1.35%*	-1.46%*	-1.62%*	-2.15%*	-1.67%	-1.08%*	-1.25%*	-1.31%*
lasso	-0.37%	-0.43%	-0.63%	-0.51%	-0.46%	-0.49%	-0.23%	-0.32%	-0.44%	-0.64%	-0.71%	-0.80%	-0.31%	-0.63%	-0.69%
enet	-0.50%	-0.58%	-0.80%	-0.64%	-0.46%	-0.48%	-0.34%	-0.44%	-0.46%	-0.74%	-0.82%	-0.97%	-0.36%	-0.78%	-0.81%
rf	-1.92%*	-0.03%*	-3.08%*	-1.02%*	-2.61%*	-0.25%	-0.01%	0.78%	0.21%	-0.60%*	-1.20%*	-1.93%	-0.01%*	-1.18%*	-1.25%*
gbm	-2.66%*	-1.00%*	-5.39%*	-1.18%*	-3.88%*	-1.56%*	-2.30%*	-1.91%*	-1.32%*	-2.47%*	-3.40%*	-3.37%*	-1.77%*	-1.96%*	-2.66%*
af-har + ridge	-0.74%	-0.69%	-1.38%	-0.51%	-1.00%	-0.55%	-0.95%	-0.70%	-0.74%	-0.84%	-1.11%	-0.86%	-0.55%	-0.64%	-0.67%
af-har + lasso	-0.19%	-0.22%	-0.33%	-0.26%	-0.23%	-0.25%	-0.11%	-0.17%	-0.22%	-0.33%	-0.36%	-0.41%	-0.16%	-0.32%	-0.35%
af-har + enet	-0.26%	-0.30%	-0.42%	-0.32%	-0.23%	-0.25%	-0.17%	-0.23%	-0.23%	-0.39%	-0.42%	-0.50%	-0.18%	-0.40%	-0.41%
af-har + rf	-1.65%*	-0.65%	-2.48%	-1.17%*	-2.01%	-0.78%	-0.71%	-0.53%	-0.56%	-1.08%	-1.34%*	-1.91%	-0.78%*	-1.41%	-1.17%*
af-har + gbm	-1.90%*	-0.93%*	-3.72%*	-0.96%	-2.55%	-1.32%*	-1.74%*	-1.57%*	-1.18%*	-1.78%*	-2.31%*	-2.55%	-1.26%*	-1.67%*	-1.71%*
pooled-avg	-1.99%*	-1.19%*	-3.51%*	-1.40%*	-2.51%	-1.35%*	-1.62%*	-1.43%*	-1.34%*	-1.85%*	-2.33%*	-2.57%*	-1.24%*	-1.90%*	-1.79%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the VaR loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table 3: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus economic variables and technical indicators; loss function: QLIKE)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.444	0.432	0.410	0.432	0.570	0.385	0.415	0.455	0.435	0.461	0.427	0.450	0.368	0.412	0.349
ridge	-3.27%*	-3.61%	-1.02%*	-3.57%	-4.09%	-3.76%*	-1.58%*	-5.14%*	-1.07%*	-4.26%*	-2.02%	-3.05%	-2.42%*	-3.62%*	-1.17%*
lasso	-2.99%*	-3.69%	-0.89%*	-3.32%	-3.26%	-3.26%	-1.07%*	-4.85%*	-1.10%*	-3.34%	-1.75%	-2.89%*	-2.34%*	-3.18%	-0.94%*
enet	-3.02%*	-3.61%	-0.83%*	-3.27%	-3.58%	-3.26%	-1.19%*	-4.92%*	-1.27%*	-3.56%	-1.90%	-3.06%	-2.50%*	-3.23%	-0.83%*
rf	-2.12%	-1.90%	1.09%	-2.86%*	-1.83%	-0.32%	0.62%	-0.82%	3.64%	2.19%	-0.05%	-2.11%	1.61%	-0.99%	2.98%
gbm	-0.88%	-3.78%*	1.22%	-3.85%*	-3.47%	-1.33%	0.27%	-2.15%	2.01%	-0.01%	0.27%	-2.75%	0.08%	-1.94%	1.78%
af-har + ridge	-2.60%*	-2.82%	-1.15%*	-2.84%	-3.11%	-2.90%	-1.63%*	-3.82%	-1.19%*	-3.42%*	-2.02%	-2.55%	-2.02%*	-2.77%	-1.25%*
af-har + lasso	-2.46%*	-2.85%	-0.96%*	-2.73%	-2.73%	-2.61%	-1.33%	-3.65%	-1.18%*	-3.03%	-1.85%	-2.45%	-1.96%*	-2.49%	-1.02%*
af-har + enet	-2.46%*	-2.80%	-0.93%*	-2.69%	-2.86%	-2.60%	-1.39%*	-3.66%	-1.28%*	-3.11%	-1.91%	-2.51%	-2.03%*	-2.50%	-0.97%*
af-har + rf	-3.62%*	-3.56%	-1.69%*	-3.93%*	-3.65%	-2.46%	-2.11%*	-2.97%	-0.44%*	-1.43%	-2.78%*	-3.83%	-1.34%	-2.62%	-0.73%*
af-har + gbm	-2.79%*	-4.11%	-1.43%*	-4.12%*	-4.20%*	-2.79%	-2.09%*	-3.23%	-0.79%*	-2.00%	-2.57%*	-3.89%	-1.84%	-2.79%	-0.86%*
pooled-avg	-4.51%*	-5.30%*	-2.15%*	-5.37%*	-5.46%*	-4.53%*	-2.79%*	-5.68%*	-1.27%*	-3.82%*	-3.52%*	-5.37%*	-2.85%*	-4.40%*	-1.30%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the Q-LIKE loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table 4: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus economic variables and technical indicators; loss function: Value-at-Risk)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.866	0.947*	2.240*	1.243	0.637	1.057	1.000*	0.834	0.730	0.890	1.278*	0.940	1.303	1.003	1.230
ridge	-1.63%*	2.13%*	1.28%*	0.16%*	0.87%*	-2.19%*	2.78%*	-2.28%*	-0.87%*	-0.65%*	1.19%*	-2.18%*	-2.30%*	-2.82%*	-1.41%*
lasso	-1.18%*	2.56%	2.15%*	1.32%	2.02%	-1.49%	3.84%	-1.85%	-0.01%	0.88%	1.58%*	-1.12%	-1.98%*	-1.92%	-1.50%*
enet	-1.27%*	2.47%	1.99%*	1.30%	1.89%	-1.52%	3.66%*	-2.07%*	-0.09%	0.58%	1.39%*	-1.66%*	-2.03%*	-2.09%	-1.55%*
rf	-2.72%*	0.66%*	-1.45%*	-2.02%*	-2.58%*	-1.53%*	-0.84%*	-0.18%	-0.77%*	-1.14%*	-1.55%*	-1.42%	-1.29%*	-1.99%*	-1.80%*
gbm	-1.93%*	0.04%*	-5.13%*	-1.64%*	-1.97%*	-2.26%*	-1.37%*	-1.96%*	-1.05%*	-1.77%*	-2.24%*	-3.09%*	-2.16%*	-2.14%*	-2.96%*
af-har + ridge	-1.13%	0.70%*	0.40%*	-0.17%*	-0.06%	-1.36%*	1.13%*	-1.67%*	-0.61%*	-0.78%*	0.34%*	-1.44%	-1.34%*	-1.71%	-0.84%
af-har + lasso	-0.87%	0.91%*	0.89%*	0.37%	0.46%	-1.03%	1.64%	-1.45%	-0.18%	-0.06%	0.58%*	-1.02%	-1.18%	-1.24%	-0.84%
af-har + enet	-0.91%	0.88%*	0.80%*	0.37%	0.39%	-1.03%	1.55%*	-1.54%*	-0.24%	-0.17%	0.49%*	-1.23%	-1.20%	-1.33%	-0.87%
af-har + rf	-2.12%*	-0.25%*	-1.53%*	-1.62%*	-2.29%*	-1.30%*	-1.03%*	-0.95%	-1.05%*	-1.38%*	-1.51%*	-1.62%	-1.16%	-1.77%	-1.43%
af-har + gbm	-1.68%*	-0.49%*	-3.63%*	-1.26%*	-1.76%*	-1.69%*	-1.31%*	-1.66%*	-1.00%*	-1.53%*	-1.77%*	-2.42%*	-1.55%*	-1.81%	-1.85%*
pooled-avg	-2.54%*	0.96%*	-1.30%*	-0.75%*	-0.94%*	-2.42%*	0.83%*	-2.33%*	-1.05%*	-1.09%*	-0.65%*	-3.01%*	-2.44%*	-2.94%*	-2.29%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the VaR loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Appendix: Descriptive Statistics and Additional Results

Table A0: Descriptive Statistics

	Mean	SD	Skew	Kurt	AC1	AC5	AC22
Volatility Factors							
Industry	6.88	0.86	0.88	4.03	0.80	0.73	0.61
Market	4.83	1.35	-0.07	3.50	0.59	0.52	0.36
World	4.47	1.15	0.13	3.75	0.75	0.64	0.49
Oil	6.31	1.29	-0.50	3.80	0.42	0.39	0.29
Natgas	7.01	0.99	-0.24	6.12	0.33	0.31	0.15
Macro Variables							
ADS	-0.34	0.79	-2.35	9.23	1.00	0.99	0.86
DGS1MO	0.65	1.30	2.45	7.85	1.00	0.98	0.93
EPU	108.15	67.88	1.79	8.47	0.65	0.54	0.39
VIXCLS	19.91	9.67	2.34	10.29	0.98	0.93	0.81

Notes: The table reports descriptive statistics for the volatility factors (the same ones as in Lyócsa and Todorova, 2021) and macroeconomic variables. ADS is the Aruoba, Diebold and Scotti (2009) business conditions index CBOE, DGS1MO is the US 1-month T-bill rate, the EPU is the Economic Policy Uncertainty index from Baker et al. (2016), and VIXCLS is the implied volatility index VIX).

Table A1: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova, 2021; loss function: MSFE)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.394	0.396	0.464	0.450	0.428	0.375	0.375	0.420	0.382	0.423	0.410*	0.417	0.396	0.405	0.348
ridge	-0.70%	-0.76%	-0.62%*	-0.46%	-0.98%*	-0.62%*	-0.68%*	-0.65%	-0.70%*	-0.87%*	-0.55%*	-1.04%	-0.94%*	-0.77%*	-0.64%*
lasso	-0.09%	-0.63%	-0.41%*	-0.19%	-0.28%	-0.40%*	-0.07%	-0.25%	-0.32%*	-0.62%*	-0.25%*	-1.00%	-0.30%	-0.27%	-0.32%*
enet	-0.08%	-0.78%	-0.44%*	-0.29%	-0.46%	-0.35%*	-0.07%	-0.34%	-0.38%*	-0.63%*	-0.31%*	-0.90%	-0.46%	-0.41%	-0.43%*
rf	1.18%	2.46%	4.17%	0.84%	3.96%	3.40%	3.72%	2.62%	4.22%	4.62%	1.97%	-1.26%	3.95%	1.92%	6.93%
gbm	-0.25%	-0.34%	2.45%	-0.57%*	0.74%	1.11%	2.29%	-0.20%	2.88%	2.07%	2.25%	-2.55%*	1.02%	0.22%	4.32%
af-har + ridge	-0.41%	-0.45%	-0.39%*	-0.28%	-0.57%	-0.37%*	-0.43%*	-0.39%	-0.42%*	-0.52%*	-0.37%*	-0.63%	-0.53%	-0.44%	-0.39%*
af-har + lasso	-0.06%	-0.36%	-0.23%	-0.12%	-0.17%	-0.23%*	-0.05%	-0.15%	-0.17%	-0.34%	-0.15%*	-0.55%	-0.17%	-0.17%	-0.20%*
af-har + enet	-0.07%	-0.44%	-0.25%*	-0.18%	-0.26%	-0.21%*	-0.06%	-0.19%	-0.21%	-0.35%	-0.19%*	-0.50%	-0.26%	-0.24%	-0.25%*
af-har + rf	-1.92%*	-1.17%	-0.29%	-1.95%*	-0.67%	-0.96%*	-0.50%*	-1.11%	-0.08%	-0.07%	-1.60%*	-3.28%*	-0.08%	-1.16%*	1.29%
af-har + gbm	-1.95%*	-1.80%*	-0.65%*	-1.86%*	-1.59%*	-1.30%*	-0.51%*	-1.74%*	-0.06%	-0.54%*	-0.83%	-3.38%*	-0.85%*	-1.38%*	0.71%*
pooled-avg	-2.15%*	-1.99%*	-1.14%*	-2.12%*	-1.67%*	-1.58%*	-0.96%*	-1.77%*	-0.70%*	-1.04%*	-1.61%*	-3.64%*	-1.03%*	-1.63%*	0.20%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the MSFE loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table A2: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus economic variables; loss function: QLIKE)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.444	0.432	0.410	0.432	0.570	0.385	0.415	0.455	0.435	0.461	0.427	0.450	0.368	0.412	0.349
ridge	-3.70%*	-4.34%*	-1.60%*	-3.80%	-4.00%*	-4.18%*	-2.35%*	-4.60%*	-1.78%*	-3.09%*	-2.46%	-3.78%*	-2.60%*	-3.21%*	-2.12%*
lasso	-3.17%	-3.77%	-1.23%*	-3.54%	-3.34%	-3.73%	-1.84%*	-4.18%*	-1.28%	-2.56%	-2.00%	-3.50%	-2.20%*	-2.54%	-1.41%
enet	-3.19%	-3.79%	-1.31%*	-3.62%	-3.47%	-3.66%	-1.80%*	-4.33%*	-1.39%*	-2.72%*	-2.06%	-3.39%	-2.32%*	-2.55%	-1.51%
rf	-1.88%	-1.72%	2.31%	-1.98%	-0.77%	1.34%	1.26%	0.33%	5.20%	3.42%	0.45%	-1.16%	3.50%	-0.41%	4.20%
gbm	-0.95%	-3.85%	1.36%	-3.58%*	-3.09%	-1.16%	0.50%	-1.92%	2.07%	0.61%	0.19%	-2.85%	0.32%	-1.37%	1.82%
af-har + ridge	-2.44%	-2.89%	-1.20%*	-2.63%	-2.64%	-2.77%	-1.73%*	-3.08%	-1.35%*	-2.23%*	-1.93%	-2.58%	-1.76%	-2.12%	-1.57%*
af-har + lasso	-2.22%	-2.67%	-0.94%*	-2.59%	-2.39%	-2.59%	-1.47%*	-2.89%	-1.09%	-1.99%*	-1.70%	-2.55%	-1.58%	-1.80%	-1.17%
af-har + enet	-2.24%	-2.67%	-0.97%*	-2.62%	-2.44%	-2.55%	-1.44%	-2.97%	-1.14%*	-2.07%*	-1.72%	-2.47%	-1.63%	-1.80%	-1.23%
af-har + rf	-3.64%	-3.60%	-1.23%*	-3.68%	-3.39%	-1.84%	-1.98%*	-2.57%	0.15%	-0.83%	-2.71%*	-3.60%	-0.59%	-2.43%	-0.26%
af-har + gbm	-2.73%	-4.07%	-1.39%*	-3.91%	-4.00%	-2.67%	-2.00%*	-3.06%	-0.75%*	-1.59%*	-2.56%	-3.88%	-1.66%*	-2.39%	-0.85%
pooled-avg	-4.65%*	-5.56%*	-2.16%*	-5.31%*	-5.26%*	-4.46%*	-3.08%*	-5.03%*	-1.27%*	-2.74%*	-3.65%*	-5.52%*	-2.38%*	-3.92%*	-1.54%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the Q-LIKE loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table A3: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus economic variables; loss function: Value-at-Risk)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.866	0.947*	2.240*	1.243	0.637	1.057*	1.000*	0.834	0.730	0.890	1.278*	0.940	1.303	1.003	1.230
ridge	-1.30%*	2.38%	1.36%*	0.12%	1.62%*	-1.57%*	3.18%*	-1.35%*	-0.53%*	-0.14%*	1.36%*	-1.69%	-1.34%*	-2.76%*	-0.97%*
lasso	-0.74%	2.73%	2.63%	0.82%	2.95%	-0.95%	4.37%	-1.60%*	0.63%	0.96%	2.38%*	-1.25%	-0.93%*	-1.92%	-0.84%*
enet	-0.75%	2.51%*	2.51%	0.77%	2.87%	-1.04%	4.29%	-1.76%*	0.54%	0.76%	2.41%*	-1.22%	-0.94%*	-2.02%	-0.93%*
rf	-2.33%*	0.88%*	-0.35%	-1.98%*	-2.11%*	-0.99%*	0.17%*	0.06%	-0.10%*	-0.67%*	-0.92%*	-1.06%	-0.55%	-2.10%*	-1.41%*
gbm	-2.29%*	-0.03%*	-5.24%*	-2.04%*	-1.70%*	-2.06%*	-1.14%*	-2.05%*	-1.14%*	-1.84%*	-2.54%*	-3.24%*	-2.19%*	-2.21%*	-3.02%*
af-har + ridge	-0.84%	0.91%*	0.48%*	-0.16%	0.36%	-0.98%*	1.34%*	-1.12%	-0.44%*	-0.37%*	0.50%*	-1.17%	-0.82%*	-1.56%	-0.60%
af-har + lasso	-0.55%	1.03%	1.15%	0.16%	0.96%	-0.71%	1.91%	-1.23%*	0.15%	0.16%	1.00%*	-0.99%	-0.62%	-1.15%	-0.53%
af-har + enet	-0.57%	0.93%*	1.08%*	0.14%	0.93%	-0.75%	1.87%	-1.30%*	0.11%	0.07%	1.02%*	-0.96%	-0.63%	-1.20%	-0.57%
af-har + rf	-2.00%*	-0.18%*	-1.07%*	-1.66%*	-2.05%*	-1.15%*	-0.59%*	-0.88%	-0.71%*	-1.17%*	-1.25%*	-1.49%	-0.84%	-1.85%	-1.29%*
af-har + gbm	-1.79%*	-0.50%*	-3.69%*	-1.46%*	-1.63%*	-1.60%*	-1.22%*	-1.75%*	-1.04%*	-1.52%*	-1.88%*	-2.45%	-1.57%*	-1.81%	-1.90%*
pooled-avg	-2.24%*	1.04%*	-0.91%*	-0.98%*	-0.30%*	-2.00%*	1.34%*	-2.05%*	-0.66%*	-0.87%*	-0.24%*	-2.64%*	-1.69%*	-3.01%*	-1.93%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the VaR loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table A4: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus technical indicators; loss function: QLIKE)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.444	0.432	0.410*	0.432	0.570	0.385*	0.415*	0.455	0.435*	0.461	0.427*	0.450	0.368	0.412	0.349*
ridge	-1.04%	-0.69%	-0.22%*	-0.87%	-1.36%*	-0.30%*	-0.33%*	-2.24%*	0.01%*	-2.47%*	-0.65%*	-0.72%*	-0.97%*	-1.55%*	0.37%*
lasso	-0.73%	-0.75%	-0.47%*	-0.62%	-0.66%	-0.39%*	0.35%	-1.66%	0.10%*	-1.80%	-0.65%*	-0.70%*	-0.82%*	-1.22%*	0.19%*
enet	-0.73%	-0.76%	-0.67%*	-0.84%	-0.81%	-0.39%*	0.12%*	-1.81%	0.04%*	-2.07%*	-0.58%*	-0.99%*	-1.00%*	-1.32%*	0.26%*
rf	0.09%	0.82%	2.04%	-0.27%*	2.24%	1.96%	3.37%	0.69%	3.56%	4.38%	2.39%	-1.93%*	2.00%	0.60%	6.20%
gbm	-0.46%	-1.46%*	2.31%	-1.62%*	0.50%	1.18%	2.06%	-0.77%	3.08%	1.85%	2.75%	-2.89%*	0.60%	-0.19%	4.21%
af-har + ridge	-1.10%*	-0.79%	-0.45%*	-1.01%	-1.30%*	-0.66%*	-0.62%*	-1.85%	-0.32%*	-2.04%*	-0.86%*	-0.92%*	-1.01%*	-1.41%*	-0.11%*
af-har + lasso	-0.91%*	-0.76%	-0.51%*	-0.80%	-0.84%	-0.61%*	-0.20%	-1.49%	-0.16%*	-1.67%	-0.76%*	-0.76%*	-0.88%*	-1.15%*	-0.12%*
af-har + enet	-0.89%*	-0.76%	-0.62%*	-0.92%	-0.91%	-0.62%*	-0.31%	-1.56%	-0.21%*	-1.78%*	-0.72%*	-0.90%*	-0.97%*	-1.20%*	-0.07%*
af-har + rf	-2.27%*	-1.94%*	-1.09%*	-2.54%*	-1.39%*	-1.38%*	-0.47%*	-1.97%	-0.29%*	-0.28%	-1.31%*	-3.48%*	-0.93%*	-1.73%*	1.03%*
af-har + gbm	-2.21%*	-2.50%*	-0.63%*	-2.60%*	-1.78%*	-1.29%*	-0.65%*	-2.17%	-0.03%*	-0.79%	-0.67%*	-3.58%*	-1.21%*	-1.73%*	0.65%*
pooled-avg	-2.67%*	-2.49%*	-1.44%*	-2.98%*	-2.17%*	-1.58%*	-0.86%*	-3.25%*	-0.38%*	-2.12%*	-1.55%*	-3.77%*	-1.70%*	-2.40%*	0.54%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the Q-LIKE loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.

Table A5: Forecasting performance for firm's realized volatility using several ML methods

(Predictors in Lyócsa and Todorova (2021) plus technical indicators; loss function: Value-at-Risk)

	apa	**apc**	**chk**	**cog**	**cop**	**cxo**	**dvn**	**eog**	**eqt**	**hes**	**mro**	**nbl**	**nfx**	**pxd**	**rrc**
af-har	0.866	0.947*	2.240*	1.243	0.637	1.057	1.000*	0.834	0.730	0.890	1.278*	0.940	1.303	1.003	1.230
ridge	-1.59%*	-1.48%*	-2.92%*	-0.45%*	-2.46%*	-1.38%*	-1.89%*	-2.20%*	-1.68%*	-2.07%*	-2.34%*	-1.94%*	-1.50%*	-1.55%*	-1.76%*
lasso	-0.89%	-0.68%	-1.15%	0.05%	-1.06%	-1.11%*	-0.41%	-0.82%	-1.01%	-0.74%	-1.18%*	-0.81%	-0.85%	-0.92%	-1.31%
enet	-1.06%	-0.81%	-1.61%	0.02%	-1.11%	-1.23%*	-0.60%	-0.99%	-1.22%	-0.90%	-1.28%*	-1.17%	-0.91%	-1.19%	-1.38%
rf	-2.82%*	-0.61%*	-4.32%*	-1.27%*	-3.64%*	-0.80%*	-1.94%*	-0.14%	-0.88%*	-1.13%*	-2.02%*	-2.65%*	-1.14%*	-1.00%*	-2.26%*
gbm	-2.35%*	-0.86%*	-5.51%*	-0.86%*	-3.55%*	-1.70%*	-2.64%*	-1.79%*	-1.29%*	-2.08%*	-3.11%*	-3.12%*	-1.76%*	-1.67%*	-2.61%*
af-har + ridge	-0.99%	-0.85%*	-1.59%	-0.35%	-1.40%	-0.80%*	-1.03%	-1.26%	-0.91%	-1.26%	-1.28%*	-1.13%	-0.89%*	-0.91%	-0.96%
af-har + lasso	-0.60%	-0.43%	-0.65%	-0.10%	-0.66%	-0.63%*	-0.27%	-0.54%	-0.54%	-0.60%	-0.67%*	-0.58%	-0.54%	-0.56%	-0.70%
af-har + enet	-0.68%	-0.49%	-0.89%	-0.12%	-0.69%	-0.70%*	-0.37%	-0.62%	-0.65%	-0.67%	-0.72%*	-0.74%	-0.57%	-0.70%	-0.73%
af-har + rf	-2.07%*	-0.87%*	-3.05%	-1.24%*	-2.59%*	-0.97%	-1.63%*	-0.81%	-1.13%	-1.33%*	-1.72%*	-2.17%	-1.12%*	-1.33%	-1.62%*
af-har + gbm	-1.80%*	-0.86%*	-3.75%*	-0.84%*	-2.44%	-1.40%*	-1.88%*	-1.50%	-1.19%	-1.66%*	-2.20%*	-2.44%*	-1.30%*	-1.53%*	-1.68%*
pooled-avg	-2.38%*	-1.38%*	-4.01%*	-1.13%*	-3.13%*	-1.78%*	-2.13%*	-1.89%*	-1.81%*	-2.09%*	-2.63%*	-2.96%*	-1.74%*	-1.98%*	-2.26%*

Notes: The table reports forecasting performance values for ML models' one-day ahead predictions of the firm's realized volatility. It shows the value of the VaR loss function for the benchmark model (af-har), and the relative percentual change with respect to this benchmark model. The symbol * indicates that a model belongs to the superior set of forecasting models according to the MCS test of Hansen et al. (2011). The model labeled as "af-har + " refers to a forecast obtained by combining the benchmark model and the respective ML model. The combination forecast is equally weighted. "Pooled-avg" shows the performance of a model that (equally) combined the five ML methods considered.