Bonato, Matteo; Cepni, Oguzhan; Gupta, Rangan et al.

Book

Forecasting realized US stock market volatility : is there a role for economic policy uncertainty?

University of Pretoria
Department of Economics Working Paper Series

Forecasting Realized US Stock Market Volatility: Is there a Role for Economic Policy Uncertainty?

Matteo Bonato
University of Johannesburg and IPAG
Oguzhan Cepni
Copenhagen Business School and Ostim Technical University
Rangan Gupta
University of Pretoria
Christian Pierdzioch
Helmut Schmidt University
Working Paper: 2024-08
March 2024

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Forecasting Realized US Stock Market Volatility: Is there a Role for Economic Policy Uncertainty?

Submission: March 2024

Matteo Bonato[*], Oguzhan Cepni[†], Rangan Gupta[‡], Christian Pierdzioch[§]

Abstract

We compare the contribution of various popular economic policy uncertainty (EPU) measures with that of widely-studied realized moments (realized leverage, realized skewness, realized kurtosis, realized good and bad volatilities, realized jumps, and realized up and down tail risks) to the performance of out-of-sample forecasts of stock market volatility of the United States (US) over the sample period from 2011 to 2023. To this end, we construct optimal forecasting models by combining the popular heterogeneous autoregressive realized volatility (HAR-RV) model with optimal stepwise predictor selection algorithms and shrinkage estimators (lasso, elastic net, and ridge regression), where we control for macroeconomic factors and sentiment as well. We find that realized moments improve out-of-sample forecasting performance relative to the baseline HAR-RV model. EPU measures do not add to forecasting performance beyond realized moments, and even deteriorate forecasting performance as the length of the forecast horizon increases. The punchline is that realized moments rather than EPU measures matter for forecasting stock market volatility.

JEL Classifications: C22; C53; G10; G17; D80

Keywords: Stock market; Volatility; Forecasting; Moments; Economic policy uncertainty

[*]Department of Economics and Econometrics, University of Johannesburg, Auckland Park, South Africa; IPAG Business School, 184 Boulevard Saint-Germain, 75006 Paris, France. Email address: matteobonato@gmail.com.

[†]Copenhagen Business School, Department of Economics, Porcelænshaven 16A, Frederiksberg DK-2000, Denmark; Ostim Technical University, Ankara, Turkiye. Email address: oce.eco@cbs.dk.

[‡]Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa. Email address: rangan.gupta@up.ac.za.

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

Conflicts of interest: The authors declare no conflict of interest.

Funding statement: The authors declare that they did not receive any funding for this research.

1 Introduction

The present value model of asset prices (see, e.g., Shiller, 1981a; b) implies that asset market volatility depends on the variability of cash flows and the discount factor, while the general equilibrium models of Pástor and Veronesi (2012, 2013), which shed light on the role played by uncertainty about government policy, imply that policy changes raise the volatility of the stochastic discount factor. In consequence, risk premia go up and stock returns become more volatile. Given such a theoretical backdrop, some researchers (see, for example, Liu and Zhang (2015), Liu et al. (2017), Gong et al. (2022), Li et al. (2023), Salisu et al. (2023)) have utilized the newspapers-based index of economic policy uncertainty (EPU) constructed by Baker et al. (2016) to successfully forecast stock market volatility of the United States (US) by employing primarily variants of the generalized autoregressive conditional heteroskedasticity (GARCH) model of conditional volatility.[1]

McAleer and Medeiros (2008) have pointed out that intraday data containing rich information can lead to more accurate estimates and forecasts of daily asset returns volatility. Given this, we utilized the square root of the sum of non-overlapping squared high-frequency (5 minute-interval intraday) stock returns observed within a day (Andersen and Bollerslev (1998)) to compute our measure of realized volatility of US stock market volatility. As compared to the popular GARCH model, realized volatility has the advantage that it is an observable and, thereby, unconditional metric of volatility, which otherwise is a latent process. The characteristic feature of models belonging to the GARCH-family is that the

[1] In-sample predictability is reported by Pástor and Veronesi (2013) and Goodell et al. (2020).

conditional variance is a deterministic function of model parameters and past data. In other words, when using GARCH models, one obtains an estimate of volatility that is not unconditional (model-free), as it is in the case with realized volatility. We then forecast this model-free measure of realized volatility with extended variants of the heterogeneous autoregressive realized volatility (HAR-RV) model of Corsi (2009), which has become increasingly popular in empirical finance because of its ability to decode important features of financial market volatility, such as long-memory and multi-scaling behavior.[2]

An additional advantage of relying on intraday data is that we were able to compute realized moments, in our case realized leverage, realized skewness, realized kurtosis, realized good and bad volatilities, realized jumps, and realized up and down tail risks. Such realized moments have been shown to be playing major roles in driving realized volatility of various asset returns, including the US stock market (Mei et al., 2017; Zhang et al., 2021). As the results of our empirical research show, this is a crucial element when it comes to appropriately evaluating the role played by EPU in forecasting realized volatility, in the sense that the HAR-RV model that includes realized moments (with or without macroeconomic predictors and investor sentiment) tends to outperform the HAR-RV model with EPU as an additional predictor. In this regard, it is important to note that, unlike us, earlier researchers who have contributed to the literature cited above on the EPU-US stock returns volatility nexus thus far have, in gen-

[2]The key feature of the HAR-RV model is that it uses volatilities from different time resolutions to capture the main features of the data-generating process that drives the realized volatility. The HAR-RV model, thereby, captures the core idea motivating the heterogeneous market hypothesis (Müller et al., 1997), according to which different groups of market participants populate stock markets, with the members differing in respect of their sensitivity to information flows at various time horizons.

eral, ignored other control variables, in particular the role played by realized moments. Available empirical evidence, thereby, is likely to overweight the empirical role of uncertainty around government policy decisions as a predictor of stock market volatility. We were able to incorporate the information of a large number of predictors (up to 31 predictors, depending on the forecasting model and dataset that we study) in our augmented HAR-RV framework, because we estimated our forecasting models by means of optimal stepwise predictor selection algorithms and popular shrinkage estimators, i.e., lasso, elastic net, and ridge regression.

Understandably, our empirical research, being the first of its kind, is of paramount academic value. However, given that stock market volatility is a key input for portfolio and hedging decisions and the accuracy of volatility forecasts is critical for the effectiveness of portfolio and risk management strategies as well as the pricing of derivative securities (Poon and Granger, 2003; Rapach et al., 2008), our findings should be interesting for investors as well. In order to lay out our empirical findings, we organize the rest of the paper as follows. In Section 2, we provide a description of the data we use in our study, while we outline in Section 3 our forecasting models. In Section 4, we present our empirical results. In Section 5, we conclude.

2 Data

We used in our empirical reserarch 5-minute-interval intraday data on the S&P 500 index, with the data sourced from the Bloomberg terminal. The intraday dataset covers a 24-hour trading day and, thus, is ideally suited to computedaily measures realized moments, described in more detail at the end of the paper (Appendix A1).

As for the policy-related measures of uncertainty, we incorporated the information from the daily news-based EPU index of Baker et al. (2016), which is based on archives over 1000 newspapers available at Access World News' NewsBank service.[3] As explained in detail on the internet page where the data can be downloaded, the primary measure for the EPU index is the number of articles that contain at least one term from each of three sets of terms, namely, "economic" or "economy", "uncertain" or "uncertainty", and "legislation" or "deficit" or "regulation" or "congress" or "federal reserve" or "white house", corresponding to economy (E), policy (P), and uncertainty (U), respectively. More recently, Bergbrant and Bradley (2023) derive an alternative measure of EPU from the major US cable news networks (CNN, Fox News, and MSNBC) using the Stanford Cable TV News Analyzer, based on the same keywords used by Baker et al. (2016).[4] In fact, Bergbrant and Bradley (2022) provide five indexes: TV–EP, TV–PU, TV–EU and TV–EPU (which is what we used to correspond to the newspapers-reliant EPU), and TV–EPU–EXP, which additionally includes the terms "risk" and "risky" in the U component. We utilized the TV–EPU index along with the newspapers-based EPU index, as well as a measure of EPU from Twitter (Baker et al., 2021), given that Bergbrant and Bradley (2023) found that the three sources of EPU contain complementary information for volatility reactions of the US stock market.

[3]The data for this index is available for download from: https://policyuncertainty.com/us_monthly.html.

[4]The index can be downloaded from: https://policyuncertainty.com/cable_epu.html.

In this regard, it should be noted that Baker et al. (2021) first extract all messages (tweets) sent on Twitter that contain keywords related to uncertainty ('uncertain", "uncertainly", "uncertainties", "uncertainty") and the economy ("economic", "economical", "economically", "economics", "economies", "economist", "economists", "economy").[5] The authors, using the database of tweets, then construct four indexes, as described in detail on the corresponding internet page. The TEU-ENG index: informs about the total number of daily English-language tweets containing terms for the categories of both uncertainty and economy; the TEU–USA index: comprises the number of these tweets that originates from users in the US using a geo-tag-based classifier; the TEU–WGT index: modifies the TEU–USA index in that it weights each tweet by: (1+log(1+number of retweets)); TEU–SCA index: scales the number of tweets each day by the number of tweets on that day that contain the word "have", to control for changes in Twitter usage intensity over time. We utilized the TEU-SCA index among the set of uncertainty variables because it produced the highest Granger-causality effect test-statistic (of value 14.9295) on realized volatility at the 5% level of significance.[6]

In terms of the first-type of control variable involving the macroeconomy, we utilized the daily, real-time, real-activity index of Scotti (2016), which is a surprise index based on recent economic data surprises (defined by comparing the first release of the macroeconomic variable and its forecast given by the

[5]This Twitter economic uncertainty (TEU) index is downloadable from: https://policyuncertainty.com/twitter_uncert.html.

[6]The test-statistics for the null-hypothesis that TEU–ENG, TEU–USA and TEU–WGT "does not Granger cause $\tilde{RV}$" are 14.8843, 12.8068, and 11.3299, respectively, with first two being significant at the 5% and 10% levels of significance respectively, and the last one being insignificant even at the 10% level.

Bloomberg median expectation), associated with the real gross domestic product, industrial production, employees on nonagricultural payrolls or the unemployment rate, retail sales, the Institute for Supply Management manufacturing index (also known as the purchasing managers' index), and the personal income of the Bureau of Economic Analysis.[7]

As for as our alternative set of macroeconomic predictors, we employed the macroeconomic attention indexes (MAIs) of Fisher et al. (2022) to focus on different macroeconomic risks of the US.[8] The authors construct their indexes by considering eight macroeconomic news categories, which reflect risks stemming from unemployment, monetary policy, output growth, inflation, housing market, credit ratings, oil, and the US dollar. The authors then measure the attention of each category by constructing a word list to count the number of articles in every category. They construct the MAIs based on a text corpus of articles from the New York Times (NYT) and the Wall Street Journal (WSJ).

Finally, in order to capture high-frequency investor sentiment, we relied on the time-varying risk aversion measure of Bekaert et al. (2022), which is calculated from observable daily information on detrended earnings yield, corporate return spread, term spread, equity return realized variance, corporate bond return realized variance, and equity risk-neutral variance.[9]

Accounting for data availability based on the two alternative sources of macroe-

[7] The surprise index is available to download from the research page of the website of Dr. Chiara Scotti at https://sites.google.com/site/chiarascottifrb/research?authuser=0.

[8] The MAIs can be sourced from the data segment of the website of Professor Jinfei Sheng at: https://sites.google.com/site/shengjinfei/data?authuser=0.

[9] The data can be downloaded from the website of Professor Nancy R. Xu at: https://www.nancyxu.net/risk-aversion-index.

conomic indicators due to Scotti (2016) and Fisher et al. (2022), along with the other variables of interest, we compiled two data sets covering 1st June, 2011 to 30th April, 2021 and 1st June, 2011 to 31st December, 2020. We call them Dataset 1 and Dataset 2, while discussing our findings in Section 4.

3 Methods

3.1 Forecasting Models

In order to set the stage for our forecasting exercises, we started with the classical estimator of realized variance, i.e., the sum of squared intraday returns over a day (Andersen and Bollerslev, 1998), as given by:

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

where $r_{t,i}$ denotes the intraday $N \times 1$ S&P 500 returns vector, and $i = 1, ..., N$ denotes the number of intraday returns. In our empirical analysis, we mainly studied realized volatility, $\tilde{RV} = \sqrt{RV}$, to mitigate the impact of the usual large peaks in the realized variance, and the impact of the large shock due to the Covid-19 pandemic in particular.

The starting point of our empirical analysis was the HAR-RV model developed by Corsi (2009). This model can be specified by the following equation:

$$\tilde{RV}_{t+h} = \beta_0 + \beta_1 \tilde{RV}_t + \beta_2 \tilde{RV}_{w,t} + \beta_3 \tilde{RV}_{m,t} + u_{t+h}, \tag{2}$$

where estimation was done by the ordinary-least-squares technique, $\beta_j, j = 0, .., 3$ are the coefficients to be estimated, u_{t+h} denotes a disturbance term, and $\tilde{RV}_{t+h}$ is the average realized volatility over the forecast horizon, h. We analyzed a short, an intermediate, and a long forecast horizon. Specifically, we set $h = 1, 5, 22$. The predictors were the daily realized volatility, $\tilde{RV}_t$, the weekly realized volatility, $\tilde{RV}_{t,w}$, and the monthly realized volatility, $\tilde{RV}_{t,m}$. We defined the weekly realized volatility as the average realized volatility from period $t-5$ to period $t-1$, and the monthly realized volatility as the average realized volatility from period $t-22$ to period $t-1$.

Using Equation (2) as a core unified modeling platform, we added a vector, M_t, to study the contribution of various realized moments (realized leverage (i.e., the value of negative realized returns which occurs on a particular day and zero otherwise), realized skewness, realized kurtosis, realized good and bad volatilities, realized jumps, and realized up and down tail risks) to forecasting performance. We briefly summarize the computation of the realized moments at the end of the paper (Appendix; Section A1). This gave the following extended model, referred to as the HAR-RV-M model:

$$\tilde{RV}_{t+h} = \beta_0 + \beta_1 \tilde{RV}_t + \beta_2 \tilde{RV}_{w,t} + \beta_3 \tilde{RV}_{m,t} + \beta_4 M_t + u_{t+h}. \tag{3}$$

where β_4 is an appropriately dimensioned vector of coefficients. In order to inspect whether the EPU measures capturing government policy-related uncertainty, UN_t, add to forecasting performance, we estimated the following forecasting model:

$$\tilde{RV}_{t+h} = \beta_0 + \beta_1 \tilde{RV}_t + \beta_2 \tilde{RV}_{w,t} + \beta_3 \tilde{RV}_{m,t} + \beta_4 M_t + \beta_5 UN_t + u_{t+h}, \tag{4}$$

where β_5 again denotes an appropriately dimensioned coefficient vector.

We also controlled for macroeconomic factors, $Macro_t$, which gave the following forecasting model:

$$\tilde{RV}_{t+h} = \beta_0 + \beta_1 \tilde{RV}_t + \beta_2 \tilde{RV}_{w,t} + \beta_3 \tilde{RV}_{m,t} + \beta_4 M_t + \beta_6 Macro_t + u_{t+h}, \tag{5}$$

where β_5 denotes an appropriately dimensioned coefficient vector. Finally, we considered an all-in forecasting model of the following format:

$$\tilde{RV}_{t+h} = \beta_0 + \beta_1 \tilde{RV}_t + \beta_2 \tilde{RV}_{w,t} + \beta_3 \tilde{RV}_{m,t} + \beta_4 M_t + \beta_5 UN_t + \beta_6 Macro_t + u_{t+h}, \tag{6}$$

Rather than simply including all realized moments, uncertainty measures, and macroeconomic factors in a large forecasting model, we used two alternative algorithms to identify an "optimal" forecasting model, as described next in detail in Section 3.2.

3.2 Algorithms for Selecting Predictors

The first algorithm that we used in our empirical research is an optimal stepwise predictor selection algorithm (for a textbook exposition, see Hastie et al. (2009), Chapter 3). This algorithm can be implemented in different ways. One way is to opt for a forward approach. In order to describe the resulting optimal forward stepwise predictor selection algorithm, we use the forecasting model given in Equation (3) as an example, but emphasize that we continued in the same way for the forecasting models given in Equations (4)-(6).

Specifically, we started with HAR-RV model (and, hence, our forecasting models always included the predictors of the HAR-RV model), estimated by the ordinary-least-squared technique the forecasting models that incorporate only one of the realized moments in addition to the predictors mentioned in Equation (2), and stored the model for which we obtained the minimum residual sum of squares. We then started the next round of the algorithm with this model, estimated all forecasting models that include two realized moments (the one selected in the first step plus one additional realized moment), and again identified the forecasting model that minimizes the residual sum of squares. We continued this process, gradually adding realized moments, until we reached the complete forecasting model described in Equation (3). The result of application of this optimal forward stepwise predictor selection algorithm, thus, was a sequence of forecasting models with increasing complexity. In order to select the ultimate "optimal" forecasting model among the models in this sequence, we employed popular information criteria. To this end, we selected the forecasting model that (i) maximizes the adjusted R^2 statistic, (ii) minimizes the Bayesian Information Criterion (BIC), or (iii) minimizes Mallow's CP criterion.[10]

As an alternative to the optimal forward stepwise predictor selection algorithm, we considered a backward variant of the algorithm. This variant follows

[10] For our empirical research, we used the R language and environment for statistical computing (R Core Team, 2023) and the "leaps" add-on package by Lumley (2020), which is based on Fortran code by Alan Miller, to implement the optimal stepwise predictor selection algorithm.

the same procedure as the optimal forward stepwise predictor selection algorithm, but it starts from the full model featuring all realized moments and then iteratively removes realized moments from the forecasting model. In addition, we considered a hybrid approach, which combines elements of the forward and backward predictor selection algorithms. The hybrid approach adds predictors in a sequential way to the forecasting model as in the forecasting approach, but then can also remove predictors that do not contribute to the fit of the forecasting model anymore.

In order to assess the robustness of our empirical findings to the choice of the algorithm used for selecting predictors, and to identify parsimonious forecasting models, we considered three popular shrinkage estimators: the Lasso estimator, an elastic net, and a Ridge regression estimator. These three shrinkage estimators are special cases of the following penalized forecasting model (using again Equation (3) as an illustrative example):

$$\sum_{t=1}^{T}\left(\tilde{RV}_{t+h}-\beta_0-\beta_1\tilde{RV}_t-\beta_2\tilde{RV}_{w,t}-\beta_3\tilde{RV}_{m,t}-\beta_4 M_t\right)^2+\lambda\left(m||\beta||_1+(1-m)||\beta||_2^2/2\right), \tag{7}$$

where β (without an index) denotes the respective vector of coefficients to be estimated (the constant and the HAR-RV terms are not penalized), and $||.||$ is the usual norm notation. One obtains the Lasso estimator as a special case for $m=1$, an elastic net as an intermediate case with some mixing parameter $0 \leq m \leq 1$ (we set $m=0.5$ in our empirical research), and the Ridge regression estimator as another special case for $m=0$. Equation (7) shows that the basic idea motivating the Lasso estimator is to add to the standard quadratic loss function that forms the foundation of the ordinary-least-squares estimator estimator a penalty term that increases in the absolute value of the coefficients. Similarly, the Ridge regression estimator uses a quadratic penalty term, and an elastic net is a mixture of the Lasso and Ridge regression estimators.[11]

4 Empirical Results

We report the results for the optimal forward predictor selection algorithm in Table 1. We estimated the forecasting models using the first 1,000 observations of the data, and used the remaining data to produce out-of-sample forecasts. We report the results for the MAE and RMSE criteria and the four different forecast horizons under scrutiny for Dataset 1 (ending in 30th April, 2021) in Panel A and the results for Dataset 2 (end-point being 31st December, 2020) in Panel B. Four main results emerge. First, we observe for all three model selection criteria (that is, adjusted R2, BIC, and CP) that the HAR-RV-M model outperforms the HAR-RV model at the short and intermediate forecast horizons, while the two models exhibit a similar forecasting performance for the long forecast horizon. Second, the HAR-RV-M model in general dominates the HAR-RV-MACRO model, and this dominance is more pronounced under the RMSE than under the MAE criterion. Third, the HAR-RV-M-UN model does not add to forecasting performance relative to the HAR-RV-M model at short and intermediate the forecast horizons, and even performs worse than the latter at the long forecast horizon when we study the RMSE criterion. Hence, the uncertainty measures do not help to improve

[11] We used the R add-on package "grf" (Tibshirani et al., 2022) to implement the shrinkage estimators.

forecasting performance in any substantial way relative to realized moments. Fourth, the HAR-RV-M model outperforms the HAR-RV-MACRO-UN model, a result that basically is a synthesis of our other three results.

– Table 1 about here. –

We next repeated our out-of-sample forecasting experiment with the natural logarithm of the realized variance being now the variable to be predicted. Studying the natural logarithm of the realized variance is interesting because it not only mitigates the impact of large peaks but rather such a transformation also brings the data closer to normality. We summarize the results in Table 2. The results corroborate the four main results documented in Table 1. Specifically, the uncertainty measures to not go beyond realized moments in improving forecasting performance at the short and intermediate forecast horizons and even deteriorate forecast accuracy at the long forecast horizon, a result we observed for both Dataset 1 and Dataset 2.

– Table 2 about here. –

We summarize the results for the three shrinkage estimators (that is, the Lasso estimator, the elastic net, and the Ridge regression estimator) in Table 3. The picture that emerges closely resembles the results we obtained for the optimal predictor selection algorithms. The HAR-RV-M model outperforms the HAR-RV model at the short and intermediate forecast horizons. The HAR-RV-M model also performs better in general than the HAR-RV-M-MACRO model, especially when we consider the RMSE criterion. Moreover, the performance of the HAR-RV-M-UN model does not deviate much from the performance of the HAR-RV-M model at the short and intermediate forecast horizons, but tends to deteriorate relative to the performance of the latter in the forecast horizon when we consider the RMSE criterion. Finally, the HAR-RV-M-MACRO-UN model does not improve upon the HAR-RV-M model.

– Tables 3 and 4 about here. –

We report in Table 4 results of the Clark and West (2007; CW) test for the three shrinkage estimators, where we focus on a comparison of the HAR-RV vs. HAR-RV-M and the HAR-RV-M-UN vs. HAR-RV-M models. The null hypothesis stipulates an equal predictive performance of the two models, while the alternative hypothesis is that the rival model performs better than the benchmark model. Hence, the CW test is a one-sided test. The test results are statistically significant at the short and intermediate forecast horizons as far as the comparison of the HAR-RV vs. HAR-RV-M models is concerned. Moreover, the test results are statistically significant at the intermediate and long forecast horizons when we compare the HAR-RV-M-UN vs. HAR-RV-M models, in line with our other results.

– Table 5 about here. –

As another exercise, we studied a recursive estimation window. To this end, we used the first 1,000 observations of the data as an initialization period, and then expanded the estimation window recursively until we reached the end of the sample period. For every recursion step, we produced out-of-sample forecasts by means of the optimal forward predictor selection algorithm. As compared to the fixed estimation window, the results we report in Table 5 are qualitatively similar. In particular, adding the uncertainty measures to the array of predictors does not systematically improve forecasting performance relative to the HAR-RV-M model. The latter even performs somewhat better than the HAR-RV-M-UN model when we increase the length of the forecast horizon.

At the end of the paper (Appendix), we summarize the results of several robustness checks. In Table A1, we report the results we obtained when we changed from a forward to an optimal backward predictor selection algorithm, while we obtained the results Table A2 by means of an optimal hybrid predictor selection algorithm. Furthermore, we considered alternative datasets. To this end, we first conducted causality tests. A Granger causality analysis revealed that, in terms of the strength of causality (value of the test statistic) to $\tilde{RV}$, the ordering of the cable news networks-based EPUs is as follows: TV–EPU–EXP (19.3964), TV–EP (15.8508), TV–EU (14.1854), TV–EPU (11.4940), and TV–PU (6.6550), with the first three cases being significant at the 1% level, the fourth one significant at the 5% level, and the fifth case showing up as insignificant. Since the TV–EPU is not necessarily the strongest predictor of $\tilde{RV}$, to ensure that we do not underestimate the role of uncertainty emanating from the making of economic policy, in Panel A of Table A3, we summarize the results we obtained when we simultaneously incorporated the various EPUs derived from cable news channels, i.e., TV–EPU–EXP, TV–EP, TV–EU, TV–EPU, and TV–PU, over the longest period involved in our empirical study stretching from 1st July, 2010 to 30th November, 2023, which we call Dataset 3.[12] In Panel B of the same

[12]While TEU data are not available going back to July, 2010, the newspapers-based EPU was not included, as Granger causality to $\tilde{RV}$ revealed a test-statistic value of 3.2709, which was not significant even at the 10% level.

table, we present the results we obtained for an alternative dataset covering the relatively longer period of 1st June, 2011 to 20th April, 2023, i.e., Dataset 4, which we constructed by ignoring the macro variables and investor sentiment. Finally, as a final robustness check, we tabulate in Table A4 the results for the optimal forward predictor selection as applied to a rolling-estimation window. The results of all robustness tests corroborated our main findings that realized moments rather than uncertainty measures matter for forecasting stock market volatility.

5 Concluding Remarks

In recent research, researchers have derived propositions from theoretical models that uncertainty surrounding policy decisions of the government, i.e., economic policy uncertainty (EPU), drive stock market volatility, with some empirical studies depicting that indeed there are forecasting gains for US stock returns from utilizing the role of EPU. However, in this empirical research, utilizing intraday data, we have documented that realized moments rather than various popular EPU indexes matter for forecasting the realized volatility of US stock market returns. Using the well-known HAR-RV model as a unified modeling platform, we have obtained our main result based on a data-driven approach by applying optimal predictor selection algorithms and shrinkage estimators (lasso estimator, elastic net, ridge regression), with our findings being robust to several modifications of the forecasting setting involving a wide-array of macroeconomic and behavioral predictors.

A relevant question to ask at this stage would be: Why do our results differ from results documented in earlier literature that show an important predictive role of EPU for US stock returns volatility? The difference arises because the realized moments basically internalize the role of uncertainty in the stock price itself at each point in time, especially at a high-frequency (Bonato et al., 2023). This line of reasoning is vindicated by the fact that, several studies in this area (see, for example, Liu and Zhang (2015), Li et al. (2023), Salisu et al. (2023)) are based on GARCH-mixed data sampling (GARCH-MIDAS) models, whereby daily conditional volatility has short- and long-run components, and monthly EPU is designed to impact the latter. Hence, it is indeed possible that the impact of government policy-related uncertainties relate to the slow-moving component of volatility rather than the fast one, with this observation corroborating the fact that monthly EPU tends to predict monthly realized volatility, as in Gong et al. (2022), but daily EPU might not impact daily conditional GARCH-based volatility, especially in (GJR-GARCH (Glosten et al., 1993)) models that account for moments like leverage (see, Liu et al. (2017)), as we do in our HAR-RV framework.

Given these observations, as part of extensions to our current empirical analysis, it is interesting to forecast intraday data-based daily realized volatility using the information content of monthly EPU by estimating the HAR-RV model based on the reverse-MIDAS technique, developed by Foroni et al. (2018). Alternatively, staying within the realms of same- and/or mixed-frequency, one can possibly interrogate the role of various financial markets-related measures of uncertainty in forecasting stock market volatility.[13] In any case, based on our empirical findings, we conclude that in spite of theoretical predictions, on the practical-front, investors should closely track realized moments rather than EPU when they need to produce forecasts of the realized US volatility stock market volatility to be utilized as inputs in their portfolio allocation decisions. Finally, another significant avenue for future research is the extension of the current framework of this sudy to international markets. This comparative analysis could reveal whether the findings, particularly regarding the relative importance of EPU and realized moments, hold across different financial environments. By examining markets with varying characteristics, such as emerging versus developed markets or markets under different regulatory regimes, researchers could assess whether the the current results can be generalized in a broader context. Furthermore, this international perspective would allow for an in-depth understanding of how and when in time regional economic policies, market structures, and investor behaviors influence stock market volatility, by providing a richer global view of financial dynamics.

[13]This line of research is motivated further by the finding that the various versions of the Twitter-based economic uncertainty (TEU) translated to the equity market (TMU), namely, TMU–ENG, TMU–SCA, TMU–USA, and TMU–WGT, produce a statistically significant Granger causal impact on $\tilde{RV}$, with the respective test-statics given by 36.8574, 38.5873, 77.2646, and 86.7875 with all of them being significant at the 1% level.

References

Andersen, T.G., and Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885–905.

Baker, S.R., Bloom, N.A., and Davis, S.J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593–1636.

Baker, S.R., Bloom, N.A., Davis, S.J., and Renault, T. (2021). Twitter-derived measures of economic uncertainty. Available for download from: https://policyuncertainty.com/media/Twitter_Uncertainty_5_13_2021.pdf.

Bekaert, G., Engstrom, E., and Xu, N.R. (2022). The time variation in risk appetite and uncertainty. Management Science, 68(6), 3975–4004.

Bergbrant, M.C. and Bradley, D. (2022). Did they just say that? Using AI to extract economic policy uncertainty from cable news networks. Available for download at SSRN: https://ssrn.com/abstract=4059681 or http://dx.doi.org/10.2139/ssrn.4059681.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2023). Business applications and state-level stock market realized volatility: A forecasting experiment. Journal of Forecasting, 43(2), 456–472.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2024). Forecasting the realized volatility of agricultural commodity prices: Does sentiment matter? Journal of Forecasting. DOI: https://doi.org/10.1002/for.3106.

Clark, T.D., and 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.

Fisher, A.J., Martineau, C., and Sheng, J. (2022). Macroeconomic attention and announcement risk premia. Review of Financial Studies, 35(11), 5057–5093.

Foroni, C., Guérin, P., and Marcellino, M. (2018). Using low frequency information for predicting high frequency variables. International Journal of Forecasting, 34(4), 774–787.

Glosten, L.R., Jagannathan, R., and Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48, 1779–1801.

Gong, X., Zhang, W., Xu, W., and Li, Z. (2022). Uncertainty index and stock volatility prediction: evidence from international markets. Financial Innovation, 8(1), 57.

Goodell, J.W. , McGee, R.J., and McGroarty, F. (2020). Election uncertainty, economic policy uncertainty and financial market uncertainty: A prediction market analysis. Journal of Banking and Finance, 110(C), 105684.

Hastie, T., Tibshirani, R., and Friedman, J. (2009) The elements of statistical learning: Data mining, inference, and prediction, 2nd ed.: Springer: New York, NY, USA.

Li, D., Zhang, L., and Li, L. (2023). Forecasting stock volatility with economic policy uncertainty: A smooth transition GARCH-MIDAS model. International Review of Financial Analysis, 88(C), 102708.

Liu, Z., Ye, Y., Ma, F., and Liu, J. (2017). Can economic policy uncertainty help to forecast the volatility: A multifractal perspective. Physica A: Statistical Mechanics and Its Applications, 482(C), 181-188.

Liu, L., and Zhang, T. (2015). Economic policy uncertainty and stock market volatility. Finance Research Letters, 15(C), 99–105.

Lumley, T. (2020). leaps: Regression Subset Selection. R package version 3.1. Available for download from: https://CRAN.R-project.org/package=leaps.

McAleer, M., and Medeiros, M.C. (2008). Realized volatility: A review. Econometrics Review, 27(1–3), 10–45.

Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help?, Physica A: Statistical Mechanics and Its Applications, 481(C), 153–159.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., and Pictet, O.V. (1997). Volatilities of different time resolutions: Analyzing the dynamics of market components. Journal of Empirical Finance, 4(2–3), 213–239.

Pástor, Ľ., and Veronesi, P. (2012). Uncertainty about government policy and stock prices. Journal of Finance, 67(4), 1219-1264.

Pástor, Ľ., and Veronesi, P. (2013). Political uncertainty and risk premia. Journal of Financial Economics, 110(3), 520-545.

Poon, S-H., and Granger, C.W.J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 478–539.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available for download from: https://www.R-project.org/.

Rapach, D.E., Strauss, J.K., and Wohar, M.E. (2008). Forecasting stock return volatility in the presence of structural breaks, in Forecasting in the Presence of Structural Breaks and Model Uncertainty. In D.E. Rapach and M.E. Wohar (Eds.), Vol. 3 of Frontiers of Economics and Globalization, Bingley, United Kingdom: Emerald, 381–416.

Salisu, A.A., Demirer, R., and Gupta, R. (2023). Policy uncertainty and stock market volatility revisited: the predictive role of signal quality. Journal of Forecasting, 42(8), 2307–2321.

Scotti, C. (2016). Surprise and uncertainty Indexes: Real-time aggregation of real-activity macro surprises. Journal of Monetary Economics, 82(C), 1–19.

Shiller, R.J. (1981a). Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review, 71(3), 421–436.

Shiller, R.J. (l981b). The use of volatility measures in assessing market efficiency. Journal of Finance, 36(2), 291–304.

Tibshirani, J., Athey, S., Serdrup, E., and Wager, S. (2022). grf: Generalized Random Forests. R package version 2.2.1. Available for download from: https://CRAN.R-project.org/package=grf.

Zhang, Z., He, M., Zhang, Y., and Wang, Y. (2021). Realized skewness and the short-term predictability for aggregate stock market volatility. Economic Modelling, 103(C), 105614.

Table 1: Optimal forward predictor selection

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0253	1.0193	1.0082	0.9886	1.0475	1.0436	1.0312	1.0021
HAR-RV vs. HAR-RV-M / BIC	1.0318	1.0224	1.0091	0.9906	1.0590	1.0521	1.0302	1.0016
HAR-RV vs. HAR-RV-M / CP	1.0318	1.0250	1.0082	0.9874	1.0590	1.0548	1.0312	1.0021
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0355	1.0433	1.0741	1.0401	1.1917	1.1432	1.1519	1.0949
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0423	1.0331	1.0763	1.0177	1.2009	1.1403	1.1538	1.0768
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0405	1.0423	1.0754	1.0409	1.2029	1.1470	1.1549	1.0980
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9922	1.0008	1.0073	0.9913	1.0029	1.0059	1.0260	1.0213
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9986	1.0061	0.9921	1.0000	1.0066	1.0215	1.0195
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9970	1.0064	1.0052	0.9901	1.0073	1.0167	1.0225	1.0213
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0340	1.0449	1.0696	1.0362	1.1864	1.1493	1.1494	1.0700
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0423	1.0298	1.0654	1.0148	1.2009	1.1321	1.1423	1.0592
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0388	1.0374	1.0706	1.0329	1.1974	1.1416	1.1519	1.0705

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0253	1.0206	1.0107	0.9896	1.0498	1.0465	1.0330	1.0026
HAR-RV vs. HAR-RV-M / BIC	1.0317	1.0236	1.0114	0.9925	1.0614	1.0547	1.0318	1.0023
HAR-RV vs. HAR-RV-M / CP	1.0317	1.0263	1.0107	0.9887	1.0614	1.0579	1.0330	1.0026
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0673	1.0851	1.0488	1.0057	1.1823	1.1340	1.0799	1.0693
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0456	1.0536	1.0338	0.9987	1.2133	1.1429	1.0689	1.0725
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0565	1.0741	1.0538	1.0014	1.1822	1.1266	1.0864	1.0703
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9929	1.0017	1.0096	0.9892	1.0035	1.0065	1.0277	1.0215
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9994	1.0083	0.9920	1.0000	1.0073	1.0230	1.0203
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9973	1.0072	1.0076	0.9882	1.0077	1.0175	1.0242	1.0215
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0655	1.0866	1.0473	1.0090	1.1804	1.1419	1.0821	1.0565
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0456	1.0484	1.0254	1.0220	1.2133	1.1351	1.0676	1.0685
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0553	1.0832	1.0472	1.0058	1.1813	1.1355	1.0835	1.0560

The models were estimated using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table 2: Optimal forward predictor selection ($\ln(RV)$)

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0275	1.0284	1.0188	0.9937	1.0268	1.0327	1.0277	1.0069
HAR-RV vs. HAR-RV-M / BIC	1.0236	1.0337	1.0249	0.9972	1.0254	1.0380	1.0335	1.0093
HAR-RV vs. HAR-RV-M / CP	1.0275	1.0284	1.0210	0.9937	1.0268	1.0327	1.0288	1.0069
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0811	1.0930	1.1173	1.1435	1.2208	1.2537	1.3248	1.4067
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0496	1.0650	1.1386	1.1510	1.1769	1.2225	1.3715	1.4253
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0580	1.0893	1.1208	1.1435	1.1835	1.2500	1.3282	1.4067
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	1.0009	1.0072	1.0031	1.0198	1.0009	1.0051	1.0102	1.0397
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	1.0014	1.0006	1.0211	1.0000	1.0019	1.0063	1.0401
HAR-RV-M-UN vs. HAR-RV-M / CP	1.0009	1.0072	1.0072	1.0198	1.0009	1.0051	1.0091	1.0397
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0857	1.0990	1.1146	1.1554	1.2201	1.2532	1.3074	1.3208
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0496	1.0653	1.0575	1.1626	1.1769	1.2140	1.2218	1.3371
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0645	1.0894	1.1141	1.1554	1.1842	1.2410	1.3108	1.3208

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0273	1.0288	1.0198	0.9936	1.0277	1.0345	1.0294	1.0075
HAR-RV vs. HAR-RV-M / BIC	1.0237	1.0343	1.0264	0.9981	1.0266	1.0402	1.0358	1.0105
HAR-RV vs. HAR-RV-M / CP	1.0273	1.0288	1.0225	0.9936	1.0277	1.0345	1.0309	1.0075
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0739	1.0856	1.0855	1.0868	1.2041	1.2259	1.2538	1.2972
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0577	1.0762	1.0929	1.0893	1.2022	1.2508	1.2771	1.3219
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0739	1.0801	1.0861	1.0866	1.2175	1.2176	1.2604	1.2989
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	1.0009	1.0083	1.0045	1.0218	1.0012	1.0064	1.0124	1.0418
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	1.0017	1.0015	1.0237	1.0000	1.0022	1.0074	1.0427
HAR-RV-M-UN vs. HAR-RV-M / CP	1.0009	1.0083	1.0094	1.0218	1.0012	1.0064	1.0114	1.0418
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0803	1.0969	1.1035	1.1009	1.2061	1.2298	1.2729	1.2479
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0577	1.0764	1.0694	1.1138	1.2022	1.2415	1.2303	1.2796
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0857	1.0838	1.0940	1.1081	1.2284	1.2256	1.2654	1.2517

The models were estimated using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table 3: Results for the shrinkage estimators

Panel A: Lasso estimator/ Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0285	1.0206	1.0122	0.9921	1.0523	1.0451	1.0325	1.0024
HAR-RV-M-MACRO vs. HAR-RV-M	1.0411	1.0417	1.0415	1.0349	1.1995	1.1405	1.0860	1.0791
HAR-RV-M-UN vs. HAR-RV-M	0.9956	1.0021	1.0043	0.9925	1.0069	1.0094	1.0195	1.0191
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0355	1.0438	1.0357	1.0225	1.1831	1.1460	1.0865	1.0529

Panel B: Lasso estimator / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0264	1.0220	1.0148	0.9920	1.0512	1.0480	1.0339	1.0024
HAR-RV-M-MACRO vs. HAR-RV-M	1.0615	1.0763	1.0325	0.9951	1.1754	1.1214	1.0515	1.0489
HAR-RV-M-UN vs. HAR-RV-M	0.9929	1.0055	1.0086	0.9908	1.0010	1.0124	1.0220	1.0198
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0538	1.0379	1.0256	1.0002	1.1618	1.0716	1.0482	1.0372

Panel C: Elastic net / Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0262	1.0206	1.0135	0.9908	1.0484	1.0450	1.0322	1.0019
HAR-RV-M-MACRO vs. HAR-RV-M	1.0383	1.0416	1.0607	1.0212	1.1934	1.1399	1.1157	1.0560
HAR-RV-M-UN vs. HAR-RV-M	0.9934	1.0044	1.0064	0.9912	1.0032	1.0114	1.0197	1.0186
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0318	1.0437	1.0428	1.0276	1.1737	1.1453	1.0934	1.0589

Panel D: Elastic net / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0262	1.0238	1.0147	0.9923	1.0508	1.0503	1.0339	1.0025
HAR-RV-M-MACRO vs. HAR-RV-M	1.0604	1.0780	1.0244	0.9891	1.1720	1.1224	1.0379	1.0369
HAR-RV-M-UN vs. HAR-RV-M	0.9929	0.9964	1.0077	0.9912	1.0018	1.0012	1.0213	1.0194
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0553	1.0711	1.0251	0.9997	1.1617	1.1183	1.0451	1.0361

Panel E: Ridge / Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0266	1.0225	1.0127	0.9921	1.0482	1.0449	1.0329	1.0019
HAR-RV-M-MACRO vs. HAR-RV-M	1.0304	1.0370	1.0431	1.0260	1.1474	1.1154	1.0719	1.0427
HAR-RV-M-UN vs. HAR-RV-M	0.9940	1.0028	1.0093	0.9925	1.0029	1.0090	1.0212	1.0154
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0283	1.0280	1.0489	1.0354	1.1470	1.0855	1.0900	1.0629

Panel F: Ridge / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M	1.0268	1.0237	1.0145	0.9931	1.0500	1.0475	1.0341	1.0023
HAR-RV-M-MACRO vs. HAR-RV-M	1.0502	1.0710	1.0122	0.9857	1.1182	1.0923	1.0047	1.0173
HAR-RV-M-UN vs. HAR-RV-M	0.9948	1.0041	1.0123	0.9916	1.0031	1.0101	1.0228	1.0156
HAR-RV-M-MACRO-UN vs. HAR-RV-M	1.0567	1.0637	1.0387	1.0017	1.1415	1.0757	1.0374	1.0344

The models were estimated by a shrinkage estimator (Lasso estimator, elastic net, Ridge regression; 10-fold cross validation) using the first 1,000 observations of the data. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table 4: Test results for the shrinkage estimators

Panel A: Lasso estimator / Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0007	0.0012	0.0036	0.2074
HAR-RV-M-UN vs. HAR-RV-M	0.0767	0.0145	0.0106	0.0475

Panel B: Lasso estimator / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0009	0.0014	0.0039	0.1944
HAR-RV-M-UN vs. HAR-RV-M	0.2153	0.0065	0.0053	0.0461

Panel C: Elastic net / Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0007	0.0012	0.0037	0.2002
HAR-RV-M-UN vs. HAR-RV-M	0.1048	0.0073	0.0054	0.0477

Panel D: Elastic net / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0009	0.0015	0.0040	0.1975
HAR-RV-M-UN vs. HAR-RV-M	0.1718	0.1667	0.0069	0.0490

Panel E: Ridge / Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0007	0.0017	0.0030	0.2286
HAR-RV-M-UN vs. HAR-RV-M	0.1150	0.0071	0.0055	0.0545

Panel F: Ridge / Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	CW (pval) / h=1	CW (pval) / h=2	CW (pval) / h=5	CW (pval) / h=22
HAR-RV vs. HAR-RV-M	0.0009	0.0019	0.0034	0.2239
HAR-RV-M-UN vs. HAR-RV-M	0.1077	0.0059	0.0033	0.0584

The models were estimated by a shrinkage estimator (Lasso estimator, elastic net, Ridge regression; 10-fold cross validation) using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts and the Clark and West (2007; CW) test (p-values (pval) based on robust standard errors) was used to statistically compare the benchmark with the rival model. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. h = forecast horizon (in days).

Table 5: Optimal forward predictor selection (recursive-estimation window)

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0364	1.0303	1.0107	0.9940	1.0514	1.0546	1.0356	1.0031
HAR-RV vs. HAR-RV-M / BIC	1.0354	1.0262	1.0133	0.9959	1.0522	1.0478	1.0382	1.0031
HAR-RV vs. HAR-RV-M / CP	1.0368	1.0301	1.0105	0.9936	1.0522	1.0545	1.0349	1.0022
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0035	1.0000	1.0355	1.0535	1.0731	1.0382	1.1487	1.0833
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0029	0.9976	1.0323	1.0372	1.0790	1.0318	1.1496	1.0704
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0040	1.0004	1.0343	1.0513	1.0739	1.0374	1.1465	1.0820
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9997	1.0036	1.0128	1.0159	1.0024	1.0065	1.0151	1.0218
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0018	1.0012	1.0066	1.0052	1.0014	1.0034	1.0154	1.0157
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9996	1.0020	1.0122	1.0135	1.0018	1.0063	1.0149	1.0191
HAR-RVv-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0057	1.0024	1.0481	1.0654	1.0825	1.0478	1.2016	1.0893
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0029	1.0019	1.0470	1.0449	1.0790	1.0440	1.2076	1.0719
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0066	1.0033	1.0511	1.0654	1.0837	1.0486	1.2039	1.0871

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0369	1.0316	1.0126	0.9923	1.0533	1.0578	1.0368	1.0028
HAR-RV vs. HAR-RV-M / BIC	1.0359	1.0270	1.0156	0.9946	1.0542	1.0503	1.0397	1.0029
HAR-RV vs. HAR-RV-M / CP	1.0373	1.0313	1.0124	0.9918	1.0541	1.0577	1.0362	1.0019
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0179	1.0095	1.0447	1.0312	1.0846	1.0465	1.1854	1.0599
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0056	0.9967	1.0305	1.0301	1.0824	1.0314	1.1682	1.0544
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0126	1.0060	1.0421	1.0290	1.0806	1.0444	1.1812	1.0553
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9999	1.0040	1.0136	1.0168	1.0026	1.0069	1.0157	1.0221
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0020	1.0013	1.0075	1.0060	1.0015	1.0036	1.0162	1.0160
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9996	1.0023	1.0130	1.0142	1.0019	1.0066	1.0155	1.0194
HAR-RVv-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0201	1.0120	1.0587	1.0498	1.0938	1.0571	1.2322	1.0731
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0056	1.0035	1.0450	1.0395	1.0824	1.0437	1.2152	1.0603
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0152	1.0131	1.0564	1.0469	1.0893	1.0570	1.2310	1.0680

The models were estimated using the first 1,000 observations of the data as an initialization period. The estimation window was then recursively expanded until the end of the sample period were reached and for every recursion step out-of-sample forecasts were produced. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Appendix

A1 Realized Moments

The following brief description of how we calculated the various realized moments follows closely the description outlined in the recent paper by Bonato et al. (2024). For a a more detailed formal description of the derivation of the realized moments, we refer an interested reader to that paper, and links to the relevant literature.

To capture potential sign asymmetries in the realized-variance process, we estimated good and bad realized variance as follows:

$$RVB_t = \sum_{i=1}^{M} r_{t,i}^2 \ \mathbf{1}_{[(r_{t,i})<0]}, \tag{A1}$$

$$RVG_t = \sum_{i=1}^{M} r_{t,i}^2 \ \mathbf{1}_{[(r_{t,i})>0]}, \tag{A2}$$

where 1 denotes the indicator function.

We calculated realized skewness, RSK, and realized kurtosis, RKU, as follows:

$$RSK_t = \frac{\sqrt{M}\sum_{i=1}^{M} r_{(i,t)}^3}{RV_t^{3/2}}, \tag{A3}$$

$$RKU_t = \frac{M\sum_{i=1}^{M} r_{(i,t)}^4}{RV_t^2}. \tag{A4}$$

where we computed the sum over the intraday returns, $r_{i,t}, i = 1, ..., M$, as observed on day t.

Given that realized variance comprises both a discontinuous (jump) component and a permanent component, we obtained realized jumps as follows:

$$\lim_{M\to\infty} RV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=1}^{N_t} k_{t,j}^2, \tag{A5}$$

where N_t = number of jumps within day t, and $k_{t,j}$ = jump size. Hence, RV_t is a consistent estimator of the jump contribution plus the integrated variance $\int_{t-1}^{t} \sigma^2(s)ds$.

Next, we consider the daily realized bipolar variation, BV_t, given by

$$BV_t = \mu_1^{-2}\left(\frac{M}{M-1}\right)\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}| = \frac{\pi}{2}\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}|, \tag{A6}$$

where $\lim_{M\to\infty} BV_t = \int_{t-1}^{t} \sigma^2(s)ds$, and $\mu_a = E(|Z|^a), Z \sim N(0,1), a > 0$. A consistent estimator of the pure daily jump contribution is defined as:

$$J_t = RV_t - BV_t. \tag{A7}$$

where we implemented the following test of the statistical significance of the jump component:

$$JT_t = \frac{RV_t - BV_t}{(v_{bb} - v_{qq})\frac{1}{N} QP_t}, \tag{A8}$$

where $v_{bb} = \left(\frac{\pi}{2}\right) + \pi - 3$ and $v_{qq} = 2$, and QP_t is defined as the daily Tri-Power Quarticity:

$$TP_t = M\frac{M}{M-2}\left(\frac{\Gamma(0.5)}{2^{2/3}\Gamma(7/6)}\right)\sum_{i=3}^{M} |r_{t,i}|^{4/3}|r_{t,i-1}|^{4/3}|r_{t,i-2}|^{4/3}, \tag{A9}$$

which converges to $TP_t \to \int_{t-1}^{t} \sigma^4(s)ds$, even in the presence of jumps. For each t,

$JT_t \sim N(0,1)$ as $M \to \infty$.

A non-negative jump contribution obtains be redefining the jump measure as follows:

$$RJ_t = \max(RV_t - BV_t; 0). \tag{A10}$$

In order to obtain measures of tail risk, we constructed $X_{t,i}$, the set of reordered intraday returns $r_{t,i}$, such that $X_{t,i} \geq X_{t,j}$ for $i < j$ with $i, j = 1, \ldots, M$ where $M =$ number of observations per day. We computed the positive tail risk estimator as

$$H_t^{up} = \frac{1}{k} \sum_{j=1}^{k} \ln(X_{t,j}) - \ln(X_{t,k}) \tag{A11}$$

and the negative tail risk estimator as

$$H_t^{down} = \frac{1}{k} \sum_{j=n-k}^{M} \ln(X_{t,j}) - \ln(X_{t,M-k}) \tag{A12}$$

where $k =$ observation denoting the chosen α tail interval.

A2 Robustness Checks

Table A1: Optimal backward predictor selection

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0253	1.0200	1.0082	0.9886	1.0475	1.0480	1.0312	1.0021
HAR-RV vs. HAR-RV-M / BIC	1.0318	1.0224	1.0091	0.9874	1.0590	1.0521	1.0302	1.0039
HAR-RV vs. HAR-RV-M / CP	1.0318	1.0250	1.0082	0.9874	1.0590	1.0548	1.0312	1.0021
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0355	1.0440	1.0741	1.0401	1.1917	1.1480	1.1519	1.0949
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0423	1.0331	1.0763	1.0144	1.2009	1.1403	1.1538	1.0793
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0405	1.0423	1.0754	1.0409	1.2029	1.1470	1.1549	1.0980
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9922	1.0015	1.0073	0.9913	1.0029	1.0101	1.0260	1.0213
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9986	1.0061	0.9889	1.0000	1.0066	1.0215	1.0219
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9970	1.0064	1.0052	0.9901	1.0073	1.0167	1.0225	1.0213
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0340	1.0457	1.0696	1.0362	1.1864	1.1541	1.1494	1.0700
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0423	1.0298	1.0654	1.0115	1.2009	1.1321	1.1423	1.0617
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0388	1.0374	1.0706	1.0329	1.1974	1.1416	1.1519	1.0705

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0253	1.0213	1.0107	0.9896	1.0498	1.0511	1.0330	1.0026
HAR-RV vs. HAR-RV-M / BIC	1.0317	1.0236	1.0114	0.9890	1.0614	1.0547	1.0318	1.0045
HAR-RV vs. HAR-RV-M / CP	1.0317	1.0263	1.0107	0.9887	1.0614	1.0579	1.0330	1.0026
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0673	1.0858	1.0488	1.0057	1.1823	1.1390	1.0799	1.0693
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0456	1.0536	1.0338	0.9951	1.2133	1.1429	1.0689	1.0748
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0576	1.0741	1.0538	1.0014	1.1841	1.1266	1.0864	1.0703
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9929	1.0023	1.0096	0.9892	1.0035	1.0110	1.0277	1.0215
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9994	1.0083	0.9886	1.0000	1.0073	1.0230	1.0225
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9973	1.0072	1.0076	0.9882	1.0077	1.0175	1.0242	1.0215
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0655	1.0873	1.0473	1.0090	1.1804	1.1470	1.0821	1.0565
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0456	1.0484	1.0254	1.0184	1.2133	1.1351	1.0676	1.0709
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0565	1.0832	1.0472	1.0058	1.1827	1.1355	1.0835	1.0560

The models were estimated using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table A2: Optimal hybrid predictor selection

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0274	1.0200	1.0082	0.9886	1.0538	1.0480	1.0312	1.0021
HAR-RV vs. HAR-RV-M / BIC	1.0318	1.0224	1.0091	0.9874	1.0590	1.0521	1.0302	1.0039
HAR-RV vs. HAR-RV-M / CP	1.0318	1.0250	1.0082	0.9886	1.0590	1.0548	1.0312	1.0021
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0418	1.0440	1.0741	1.0401	1.2064	1.1480	1.1519	1.0949
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0423	1.0331	1.0763	1.0144	1.2009	1.1403	1.1538	1.0793
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0405	1.0423	1.0754	1.0421	1.2029	1.1470	1.1549	1.0981
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9928	1.0015	1.0073	0.9913	1.0024	1.0101	1.0260	1.0213
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9986	1.0061	0.9889	1.0000	1.0066	1.0215	1.0219
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9970	1.0064	1.0052	0.9913	1.0073	1.0167	1.0225	1.0213
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0395	1.0457	1.0696	1.0368	1.2005	1.1541	1.1494	1.0702
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0423	1.0298	1.0654	1.0115	1.2009	1.1321	1.1423	1.0617
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0388	1.0374	1.0706	1.0402	1.1974	1.1416	1.1519	1.0789

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0277	1.0213	1.0107	0.9896	1.0564	1.0511	1.0330	1.0026
HAR-RV vs. HAR-RV-M / BIC	1.0317	1.0236	1.0114	0.9890	1.0614	1.0547	1.0318	1.0045
HAR-RV vs. HAR-RV-M / CP	1.0317	1.0263	1.0107	0.9896	1.0614	1.0579	1.0330	1.0026
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0698	1.0858	1.0488	1.0057	1.1898	1.1390	1.0799	1.0693
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0456	1.0536	1.0338	0.9951	1.2133	1.1429	1.0689	1.0748
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0565	1.0741	1.0546	1.0023	1.1822	1.1266	1.0861	1.0703
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9935	1.0023	1.0096	0.9892	1.0029	1.0110	1.0277	1.0215
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9994	1.0083	0.9886	1.0000	1.0073	1.0230	1.0225
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9973	1.0072	1.0076	0.9892	1.0077	1.0175	1.0242	1.0215
HAR-RV-M-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0687	1.0873	1.0473	1.0090	1.1868	1.1470	1.0821	1.0565
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0456	1.0484	1.0254	1.0184	1.2133	1.1351	1.0676	1.0709
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0595	1.0862	1.0472	1.0067	1.1819	1.1487	1.0835	1.0560

The models were estimated using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table A3: Optimal forward predictor selection (alternative datasets)

Panel A: Dataset 3 (1st July, 2010 to 30th November, 2023)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0346	1.0256	1.0129	0.9941	1.0427	1.0394	1.0308	1.0014
HAR-RV vs. HAR-RV-M / BIC	1.0368	1.0311	1.0188	0.9967	1.0519	1.0508	1.0361	1.0030
HAR-RV vs. HAR-RV-M / CP	1.0346	1.0279	1.0129	0.9957	1.0427	1.0411	1.0308	1.0040
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9941	0.9958	1.0003	0.9961	1.0036	1.0134	1.0308	1.0137
HAR-RV-M-UN vs. HAR-RV-M / BIC	0.9942	0.9946	1.0074	0.9907	1.0030	1.0110	1.0321	1.0132
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9941	0.9961	1.0001	0.9977	1.0036	1.0133	1.0282	1.0165

Panel B: Dataset 4 (1st June, 2011 to 20th April, 2023)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0315	1.0177	1.0028	0.9937	1.0434	1.0363	1.0255	1.0035
HAR-RV vs. HAR-RV-M / BIC	1.0311	1.0209	1.0098	0.9977	1.0482	1.0453	1.0317	1.0037
HAR-RV vs. HAR-RV-M / CP	1.0340	1.0221	1.0028	0.9948	1.0501	1.0411	1.0255	1.0049
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	0.9894	0.9979	1.0019	1.0131	1.0013	1.0065	1.0223	1.0341
HAR-RV-M-UN vs. HAR-RV-M / BIC	1.0000	0.9957	1.0088	1.0139	1.0000	1.0035	1.0262	1.0323
HAR-RV-M-UN vs. HAR-RV-M / CP	0.9895	1.0008	1.0018	1.0142	1.0007	1.0065	1.0201	1.0355

The models were estimated using the first 1,000 observations of the data. The remaining data were then used to produce out-of-sample forecasts. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).

Table A4: Optimal forward predictor selection (rolling-estimation window)

Panel A: Dataset 1 (1st June, 2011 to 30th April, 2021)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0244	1.0243	1.0096	0.9985	1.0385	1.0544	1.0240	1.0042
HAR-RV vs. HAR-RV-M / BIC	1.0152	1.0228	1.0090	0.9997	1.0179	1.0525	1.0165	1.0003
HAR-RV vs. HAR-RV-M / CP	1.0234	1.0238	1.0105	0.9987	1.0365	1.0526	1.0226	1.0037
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0240	1.0074	1.0965	1.1436	1.1584	1.0804	1.6118	1.2043
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0220	1.0056	1.0926	1.0938	1.1556	1.0738	1.6146	1.1358
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0217	1.0049	1.1032	1.1436	1.1581	1.0755	1.6183	1.2038
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	1.0021	1.0026	1.0057	1.0095	1.0048	1.0043	1.0101	1.0013
HAR-RV-M-UN vs. HAR-RV-M / BIC	0.9999	1.0000	1.0032	0.9964	0.9998	1.0000	1.0063	0.9955
HAR-RV-M-UN vs. HAR-RV-M / CP	1.0010	1.0034	1.0068	1.0031	1.0040	1.0015	1.0105	0.9994
HAR-RVv-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0263	1.0065	1.1037	1.1181	1.1663	1.0886	1.6595	1.1453
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0218	1.0061	1.0967	1.0859	1.1577	1.0845	1.6671	1.0969
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0242	1.0040	1.1090	1.1247	1.1657	1.0828	1.6630	1.1513

Panel B: Dataset 2 (1st June, 2011 to 31st December, 2020)

Benchmark vs. rival model	MAE / h=1	MAE / h=2	MAE / h=5	MAE / h=22	RMSE / h=1	RMSE / h=2	RMSE / h=5	RMSE / h=22
HAR-RV vs. HAR-RV-M / Adj. R2	1.0272	1.0278	1.0114	0.9949	1.0410	1.0591	1.0246	1.0036
HAR-RV vs. HAR-RV-M / BIC	1.0170	1.0262	1.0105	0.9964	1.0190	1.0571	1.0167	0.9992
HAR-RV vs. HAR-RV-M / CP	1.0264	1.0273	1.0124	0.9950	1.0390	1.0571	1.0232	1.0030
HAR-RV-M-MACRO vs. HAR-RV-M / Adj. R2	1.0313	1.0161	1.1094	1.0981	1.1859	1.0914	1.6448	1.1399
HAR-RV-M-MACRO vs. HAR-RV-M / BIC	1.0238	1.0035	1.0900	1.0778	1.1725	1.0736	1.6213	1.1174
HAR-RV-M-MACRO vs. HAR-RV-M / CP	1.0284	1.0128	1.1059	1.0941	1.1828	1.0869	1.6511	1.1416
HAR-RV-M-UN vs. HAR-RV-M / Adj. R2	1.0023	1.0034	1.0061	1.0073	1.0051	1.0046	1.0106	1.0002
HAR-RV-M-UN vs. HAR-RV-M / BIC	0.9999	1.0000	1.0034	0.9979	0.9998	1.0000	1.0066	0.9958
HAR-RV-M-UN vs. HAR-RV-M / CP	1.0011	1.0036	1.0073	1.0006	1.0042	1.0015	1.0110	0.9983
HAR-RVv-MACRO-UN vs. HAR-RV-M / Adj. R2	1.0330	1.0147	1.1126	1.0780	1.1923	1.1003	1.6779	1.0982
HAR-RV-M-MACRO-UN vs. HAR-RV-M / BIC	1.0254	1.0071	1.0965	1.0648	1.1884	1.0863	1.6701	1.0774
HAR-RV-M-MACRO-UN vs. HAR-RV-M / CP	1.0303	1.0119	1.1124	1.0844	1.1906	1.0963	1.6819	1.1047

The models were estimated by the ordinary-least-squares technique a rolling-estimation window of length 1,000 observations and out-of-sample forecasts were produced. The HAR-RV model was estimated by the ordinary-least-squares technique. In all other models, the HAR-RV terms were forced to be included in the models. MAE = mean absolute error. RMSE = root-mean-squared error. The MAE and RMSE statistics are expressed as ratios with the statistic for a benchmark model forming the nominator and the statistic for a rival model forming the denominator. h = forecast horizon (in days).