

Forecasting the Realized Variance of Oil-Price Returns: A Disaggregated Analysis of the Role of Uncertainty and Geopolitical Risk

Rangan Gupta^a, Christian Pierdzioch^b

November 2021

Abstract

We contribute to the empirical literature on the predictability of oil-market volatility by comparing the predictive role of aggregate versus several disaggregated metrics of policy-related and equity-market uncertainties of the U.S. and geopolitical risks for forecasting the future realized volatility of oil-price (WTI) returns over the monthly period from 1985:01 to 2021:08. Using machine-learning techniques, we find that adding the disaggregated metrics to the array of predictors improves the accuracy of forecasts at intermediate and long forecast horizons, and mainly when we use random forests to estimate our forecasting model.

JEL classification: C22, C53, Q02

Keywords: Realized variance; Oil Price; Forecasting; Machine Learning; Uncertainty; Geopolitical Risk

^a Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa; Email address: rangan.gupta@up.ac.za.

^b Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: macroeconmoics@hsu-hh.de.

1 Introduction

Historically, movements in the volatility of the returns of the price of crude oil have been found to have predictive value for subsequent slowdowns in worldwide economic growth (van Eyden et al., 2019; Salisu et al., 2021a). At the same time, the recent trend towards financialization that has characterized the development of the oil market, especially post the Global Financial Crisis (GFC), has implied that the participation of hedge funds, pension funds, and insurance companies in the market has increased. In fact, oil is by now often viewed as an alternative investment in the portfolio decisions of financial institutions (Degiannakis and Filis, 2017; Bonato, 2019). Naturally, accurate forecasts of the future path of the volatility of oil-price returns are of paramount importance to policymakers in conceptualizing in due time macroeconomic policies to mitigate or even prevent economic recessions, and also for oil traders, because volatility is an important input to investment decisions and portfolio choices (Poon and Granger, 2003).

Heightened economic and financial uncertainties, following a series of crises and market jitters such as the GFC, the European sovereign debt crisis, the Brexit, and, of course, the ongoing COVID-19 pandemic, have led researchers to analyze in several significant recent empirical studies the role of uncertainty for forecasting the volatility of the oil-price movements (see, for example, Bonaccolto et al., (2018), Bouri et al., (2020), Li et al., (2020), Liang et al., (2020), Dutta et al., (2021), Pierdzioch and Gupta (2021)), which is not surprising given the importance crude oil plays as the most actively traded commodity. In terms of economic theory, the foundation for a positive relationship between uncertainty and volatility of oil can be derived from the ‘Theory of Storage’, which involves the commodities market in general (Bakas and Triantafyllou, 2018; 2020). This theory stipulates that increases in uncertainty tend to make the path of future aggregate demand of commodities, and as a result, also of aggregate production less predictable. Given this heightened unpredictability, risk averse commodity producers prefer to hold physical inventory that causes a rise in the convenience yield, which, in turn, leads to increased volatility of commodity prices.

In terms of the mushrooming existing literature on uncertainty and its usefulness as a predictor of the future volatility of oil-price returns, Bonaccolto et al. (2018) examine the role of newspapers-

based measures of economic-policy and equity-market uncertainty of the United States (U.S.) as predictors of the conditional quantiles of crude-oil returns and volatility by means of a non-parametric k -th order causality-in-quantiles framework. They find that these U.S.-based uncertainty indexes are relevant predominantly during periods of market distress, when oil risk is of paramount importance for market participants, where the effects are heterogeneous across different quantiles. Bouri et al. (2020) also use in their empirical analysis daily newspaper-based index of U.S. uncertainty. They, however, focus on uncertainty associated with infectious diseases (EMVID) for oil-market volatility, and report that adding EMVID to their forecasting models significantly ameliorates forecast accuracy across a range of short-, medium-, and long-run forecast horizons, using a heterogenous autoregressive model of (realized) volatility. A different avenue has been taken by Li et al., (2020), who apply a so called mixed-data-sampling generalized autoregressive conditional heteroscedastic (MIDAS-GARCH) model to shed light on the contribution of monetary-policy uncertainty besides overall economic-policy uncertainty of the U.S. to forecasting oil market volatility. Dutta et al., (2021), however, have shown, based on a quantiles-based modeling framework, that, unlike overall uncertainty of the U.S. related to policy decisions, equity-market volatility of the U.S. in general, and the same due to commodity market movements and crises, have a higher predictive value for oil-market volatility. In this regard, it is interesting to note that Liang et al., (2020) emphasize the importance of overall equity market volatility indices of the U.S. using a standard predictive regression model, model combination, and shrinkage techniques, while Li et al., (2020) do not find evidence of a metric of global uncertainty to be important for predicting oil-market volatility. Also taking a global perspective, Gupta and Pierdzioch (2021) compile a dataset for the group of G7 (Canada, France, Germany, Italy, Japan, the United Kingdom (UK) and the U.S.) countries and China to analyze, in an out-of-sample forecasting experiment, the usefulness of of uncertainty and its international spillovers as predictors for the volatility of West Texas Intermediate (WTI) and Brent crude oil price movements. They use shrinkage estimators for conducting their out-of-sample forecasting experiment and find that uncertainty and international spillovers improve the accuracy of forecasts of the realized volatility of oil returns at intermediate- and long- forecasting horizons.

Against the backdrop of this burgeoning empirical literature on uncertainty and its role for fore-

casting oil-market volatility, we contribute to this line of research by comparing the role of aggregate versus (ten and forty-four, respectively) disaggregated metrics of policy-related and equity-market uncertainties of the U.S. in predicting the future realized volatility of oil-price (WTI) returns over the monthly period from 1985:01 to 2021:08. Naturally, if disaggregated uncertainty measures, capturing the various multifaceted sources of aggregate policy and equity-market uncertainties, end up producing additional forecasting gains relative to their overall counterparts then this is a finding of considerable value to both investors and policymakers. While it is true, as observed from the concise review presented in the preceding paragraph, that few studies already have studied some of these disaggregated uncertainty measures, we provide a more comprehensive analysis by exhausting all the disaggregated metrics available. In doing so, unlike the existing studies, we use a statistical data-driven approach involving machine-learning methods, rather than randomly deciding on which disaggregate measure to use in the forecasting model of oil-market volatility. In particular, we use a two-step approach. In the first step, we use the least absolute shrinkage and selection operator (Lasso), proposed by Tibshirani (1996). The Lasso belongs to the spectrum of linear regression-analysis methods and performs both model selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. In a second step, we switch to a linear model and estimate random forests (Breiman, 2001), which, in turn, is a machine-learning technique tailored to operate in settings featuring a large array of predictors. In particular, random forests automatically capture the predictive value of any potential nonlinear links between the oil-market volatility and the (various decomposed) uncertainties (Van Robays, 2016) as well as any interaction effects between the predictors (Gabauer and Gupta, 2020). In addition to aggregate and disaggregate policy- and equity market-related uncertainties, we also investigate the role of overall geopolitical risks and decomposed values of the same into threats and acts, given that some studies have highlighted the role of such risks too in forecasting oil-market volatility (see for example, Liu et al., (2019), Asai et al., (2020), Mei et al., (2020)).

At this stage, it is important to highlight two further points. First, we forecast the monthly realized variance (RV) of returns of the WTI oil price, whereby, following Andersen and Bollerslev (1998), we measure RV in terms of the the sum of daily squared returns over a month, which

yields an observable and unconditional measure of volatility, which is otherwise a latent process. Conventionally (see the discussions in Chan and Grant (2016) and Lux et al., (2016) for further details), the time-varying volatility is modeled using various models belonging to the GARCH family, under which the conditional variance is a deterministic function of model parameters and past data. Alternatively, some researchers have considered in recent work stochastic volatility (SV) models, which model the volatility as a latent variable that follows a stochastic process. Irrespective of whether one uses GARCH or SV models, the underlying estimate of volatility is not model-free (or unconditional) as in the case of RV. Second, though oil is a globally traded commodity, data on the disaggregated policy and equity-market uncertainties are only available for the U.S. and, hence, we consider the WTI as our proxy for the global price of oil, with us able to do this given the dominance of the U.S. as a major player in the oil market both on the export and import fronts.

To the best of our knowledge, ours is the first empirical study to compare the role of aggregate and disaggregated measures of policy-, stock-market-, and geopolitical-events-related uncertainties and risks for forecasting the RV of oil-price returns, using linear and nonlinear machine-learning techniques, besides standard Ordinary Least Squares (OLS) applied to the smaller models (in terms of predictors) involving only the aggregate metrics. In the process, our paper adds to the already existing large and significant literature on the predictability of oil-returns volatility by considering the role of various decomposed uncertainties, with the literature having studied the predictive role of a wide array of macroeconomic, financial, behavioural, and climate patterns-related predictors by means of a large spectrum of linear and nonlinear univariate or multivariate models (see, for example, Asai et al., (2019), Bonato et al., (2020), Demirer et al., (2020, 2021), Gkillas et al., (2020), Bouri et al., (2021); Salisu et al., (2021b), Luo et al., (forthcoming), and the references cited within these papers).

We organize the remainder of our paper as follows: Section 2 contains a description of the data, while Section 3 is devoted to the methodologies. Then, we discuss in Section 4 the forecasting results. Section 5 concludes.

2 Data

As for the WTI crude oil price, we use the nominal daily data derived from Global Financial Data.¹ After having computed the daily log-returns (first-differences of the natural logarithm of the WTI oil price), we obtain the monthly realized variances, i.e., RV, by summing up the daily squared log-returns over a specific month. Figure 1 plots the resulting realized variance along with its autocorrelation function. Eyeballing the autocorrelation function suggests that an AR(1) model is a useful benchmark model against which we can assess the contribution of EMV, EPU, and GPR and their components to forecast performance.

– Figure 1 about here. –

Turning next to our uncertainty data, it is important to note that uncertainty is a latent variable and, thereby, requires ways to measure it. Gupta et al. (2018) document that three broad approaches have been used in the empirical literature to quantify uncertainty, apart from the various ones associated with financial markets (such as implied-volatility indices, realized volatility, idiosyncratic volatility of equity returns, corporate spreads). The three approaches can be described as follows: (1) A text-based approach uses the idea is to construct indices of uncertainty from searches of key words or terms related to (economic and policy) uncertainty in major newspapers or country-reports. (2) A second approach uses stochastic-volatility estimates computed by means of various small and large-scale structural models (related to macroeconomics and finance) to construct measures of uncertainty. (3) A third approach is to collect data on the dispersion of professional forecaster disagreements to obtain uncertainty estimates. For our metrics of aggregate and disaggregate uncertainties, we use the first approach as outlined by Baker et al., (2016, 2019). For the purpose of our analysis, their approach has several advantages. First, it does not require any complicated estimation of an econometric model to generate it in the first place and, thereby, does not depend on a specific model. Second, their the policy- and equity

¹<https://globalfinancialdata.com/>.

market-related uncertainty data are available publicly for download. In other words, the aggregate and the corresponding disaggregated measures of uncertainty that we use for our empirical analysis has only been thus far derived from a newspapers-based method, with the ultimate indexes being freely available. In this regard, it is worth emphasizing that the same two advantages also apply in case of the the geopolitical risks-related variables that we use as predictors.

The categorical data consists of a range of sub-indexes of the overall economic policy uncertainty (EPU) index, based solely on news data. The sub-indexes are derived using results from the Access World News database of over 2,000 U.S. newspapers. Every sub-index requires an article published in these newspapers to contain the terms: ‘uncertainty’ or ‘uncertain’, ‘economic’ or ‘economy’ and one or more of the terms: ‘congress’, ‘legislation’, ‘white house’, ‘regulation’, ‘federal reserve’, or ‘deficit’ as in Baker et al. (2016), as well as a set of categorical policy terms associated with Monetary Policy, Taxes, Fiscal Policy and Government Spending, Health Care, National Security, Entitlement Programs, Regulation, Financial Regulation, Trade Policy, and Sovereign Debt and Currency Crises.²

Baker et al. (2019) construct a newspaper-based equity market volatility (EMV) tracker that moves with the Volatility Index (VIX) of the Chicago Board Options Exchange (CBOE), and with the realized volatility of returns on the S&P 500. They further construct an array of 44 category-specific EMV trackers that render it possible to quantify the importance of each category in the level of U.S. stock-market volatility and its movements over time.³ In order to compute the overall EMV tracker, Baker et al. (2019) specify terms in the following three sets: E: economic, economy, financial; M: stock market, equity, equities, Standard & Poors (and variants), and; V: volatility, volatile, uncertain, uncertainty, risk, risky. They then obtain monthly counts of newspaper articles (as a percentage of all articles in the same paper and month) that

²The reader is referred to: http://policyuncertainty.com/categorical_terms.html for complete details on the term sets by category of EPU. The overall EPU and the nine categorical sub-indexes are available for download from: http://policyuncertainty.com/categorical_epu.html.

³The overall and category-specific data are publicly available at: http://policyuncertainty.com/EMV_monthly.html.

contain at least one term in each of E, M, and V from eleven major U.S. newspapers.⁴ Finally, Baker et al., (2019) average the scaled counts over newspapers by month, and rescale the resulting data in a multiplicative way so as to match the mean value of the VIX over the start- to the current-period of data availability.

In order to obtain the category-specific EMV trackers, Baker et al. (2019) calculate the share of EMV articles in each category and multiply by the contemporaneous EMV tracker value. The EMV tracker categories that they consider are the following: Policy-Related; Infectious Disease; Macroeconomic News and Outlook; Macro-Broad Quantity Indicators; Macro-Inflation EMV Indicator; Macro-Interest Rates; Macro-Other Financial Indicators; Macro-Labor Markets; Macro-Real Estate Markets; Macro-Trade; Macro-Business Investment and Sentiment; Macro-Consumer Spending and Sentiment; Commodity Markets; Financial Crises; Exchange Rates; Healthcare Matters; Litigation Matters; Competition Matters; Labor Disputes; Intellectual Property Matters; Fiscal Policy; Taxes; Government Spending; Deficits, and Debt; Entitlement and Welfare Programs; Monetary Policy Regulation; Financial Regulation; Competition Policy; Intellectual Property Policy; Labor Regulations; Immigration; Energy and Environmental Regulation; Lawsuit and Tort Reform, Supreme Court Decisions; Housing and Land Management; Other Regulation; National Security Policy; Government-Sponsored Enterprises; Trade Policy; Healthcare Policy; Food and Drug Policy; Transportation, Infrastructure, and Public Utilities; Elections and Political Governance; Agricultural Policy, and; Petroleum Markets.

Finally, the overall geopolitical risks (GPR), and its sub-indexes due to threats and attacks are based on the work of Caldara and Iacoviello (2019).⁵ The data are derived from an automated text search of the electronic archives of eleven national and international newspapers.⁶ Caldara and Iacoviello (2019) describe that they construct the index by counting the number of articles

⁴The newspapers are the following: The Boston Globe, Chicago Tribune, Dallas Morning News, Houston Chronicle, Los Angeles Times, Miami Herald, New York Times, San Francisco Chronicle, USA Today, Wall Street Journal, and Washington Post.

⁵The data can be downloaded from: <https://www.matteoiacoviello.com/gpr2019.htm>.

⁶The newspapers are the following: The Boston Globe, Chicago Tribune, The Daily Telegraph, Financial Times, The Globe and Mail, The Guardian, Los Angeles Times, The New York Times, The Times, The Wall Street Journal, and The Washington Post.

related to geopolitical risk in each newspaper for each month (as a share of the total number of news articles). The search identifies articles containing references to six groups of words: Group 1 comprises words associated with explicit mentions of geopolitical risk and mentions of military-related tensions involving large regions of the world and a U.S. involvement. Group 2 comprises words directly related to nuclear tensions. Groups 3 and 4 comprise mentions related to war threats and terrorist threats, respectively. Finally, Groups 5 and 6 reflect press coverage of actual adverse geopolitical events (as opposed to risks) which are likely to result in higher geopolitical uncertainty, such as terrorist acts or the beginning of a war. Based on the definitions of the six search groups, Caldara and Iacoviello (2019) further dissect the direct effect of adverse geopolitical events from the effect of pure geopolitical risks. In order to do so, they construct two indexes. The geopolitical threats (GPT) index only includes words belonging to search Groups 1 to 4. The geopolitical acts (GPA) index only includes words belonging to search Groups 5 and 6.⁷

– Figure 2 about here. –

Figure 2 plots the aggregate EMV, EPU, and GPR data. Based on data availability on RV and the predictors, our analysis covers the monthly period from 1985:01 to 2021:08.

3 Methods

An autoregressive model of order one (that is, an AR(1) model) is our benchmark forecasting model:

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \eta_{t+h}, \quad (1)$$

where $\beta_i, i = 0, 1$ are coefficients to be estimated by the ordinary-least-squares (OLS) technique, and η_{t+h} denotes a disturbance term. The parameter, h , denotes the forecast horizon. We consider

⁷The importance of disaggregating the overall GPR index when analyzing its impact on oil market volatility have been highlighted by recent studies (see for example, Demirer et al., (2018, 2019) and Lee et al., (2021)).

the following forecast horizons: $h = 1, 3, 6, 12, 24$, where we forecast the the average realized variance over the respective forecast horizon for $h > 1$. In addition, we structure the data matrix so that it has the same dimension for all forecast horizons.

Our first extension of the model given in Equation (1) features the aggregate EMV, EPU, and GLPR predictors and is given by

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 EMV_t + \beta_3 EPU_t + \beta_4 GPR_t + \eta_{t+h}. \quad (2)$$

Again, we use the OLS technique to estimate the coefficients of this forecasting model.

Next, we take a disaggregated perspective and use the various categories of EMV, EPU, and GPR (threats and actual geopolitical risks in case of the latter) for forecasting the realized variance.

Our forecasting model is of the format

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \sum_c^n \gamma_c x_{t,c} + \eta_{t+h}, \quad (3)$$

where γ_c are the coefficients corresponding to the n categories, $x_{t,c}$. Because of the large number, n , of categories, we do not estimate the model given in Equation (3) by the OLS technique, but rather apply the least absolute shrinkage and selection operator (Lasso), a machine-learning technique proposed by Tibshirani (1996). The Lasso estimates of the model coefficients are chosen to minimize the following expression (for a textbook exposition of the Lasso, see, e.g., Hastie et al., 2009):

$$\sum_{t=1}^N (RV_{t+h} - \beta_0 + \beta_1 RV_t + \sum_{c=1}^n \gamma_c x_{t,c})^2 + \lambda \left(|\beta_0| + |\beta_1| + \sum_{c=1}^n |\gamma_c| \right), \quad (4)$$

where N is the number of observations available for estimation of the model. Hence, the Lasso estimator is a shrinkage estimator that uses the L1 norm of the vector of coefficients to shrink the dimension of the estimated model. The magnitude of the shrinkage parameter, λ , governs the extent of the coefficient shrinkage. When λ is large enough, the Lasso estimator can even set to zero some of the coefficients, in which case the Lasso estimator can also be interpreted as a predictor-selection estimator.

The forecasting models that we specify in Equations (1), (2), and (4) all belong to the class of linear forecasting models. As indicated in the introduction, because there is evidence of the

various categories of uncertainty and geopolitical risks being interdependent (see also Caldara and Iacoviello, (2019) in this regard), and because the realized variance may be a nonlinear function of these predictors, it is warranted also to consider a nonlinear forecasting model. Given the large number of categories that we consider in our empirical analysis, we, therefore, also consider a forecasting model of the following format:

$$RV_{t+h} = f(RV_t, x_{t,1}, x_{t,2}, \dots, x_{t,n}) \quad (5)$$

We estimate the function, f , that maps the predictors into the forecast of the subsequently realized variance by a machine-learning technique known as random-forest technique (Breiman, 2001). A random forest consists of a large number of individual random regression trees and, thereby, belongs to the class of ensemble machine-learning techniques (see the textbook by Hastie et al., 2009). A regression tree consists of branches that recursively partition in a hierarchical and binary way the space of predictors into non-overlapping regions. Growing a large regression tree and, thus, forming a large number of such regions makes it possible to compute increasingly granular predictions of the realized variance. At the same time, however, the complex hierarchical structure of a regression tree eventually results in an overfitting and data-sensitivity problem. A random forest addresses this problem by combining a large number of random individual regression trees. To this end, one draws a large number of bootstrap samples from the data and estimates a random regression tree on every bootstrap sample. A random regression tree differs from a standard regression tree in that a researcher uses a random subset of the predictors to grow the nodes and branches of the tree. This additional element of randomness mitigates the potential effect of influential individual predictors on tree building (which may result, for example, from outliers in the data). Moreover, because a random forest consists of a large number of random regression trees, it reduces the correlation of predictions computed by means of the individual random regression trees. Finally, upon computing the average of the predictions across the individual regression trees helps to stabilize the random-forest-based forecasts of the realized variance of oil-price returns.

— Table 1 about here. —

To sum up, for our forecasting experiment, we study four different models. For convenience, Table 1 provides summary information of the estimated models. Model 1 is the baseline AR(1) model, Model 2 extends the baseline model to include the aggregate EMV, EPU, and GPR data, Model 3 uses the AR(1) term along with the disaggregate EMV, EPU, and GPR data and is estimated by means of the Lasso, and Model 4 uses the same predictors as Model 3 but is estimated using random forests.

In our empirical analysis, we use the R language and environment for statistical computing (R Core Team 2021) to set up our forecasting experiment. We use the R add-on package “glmnet” (Friedman et al., 2010) to estimate the Lasso model, where we use 10-fold cross-validation to determine the shrinkage parameter, where the optimal shrinkage parameter minimizes the mean cross-validated error. We use the R add-on package “grf” (Tibshirani et al., 2021) to estimate random forests. Every random forest consists of 2,000 individual random regression trees, and cross-validation is used to tune the tree parameters (number of randomly sampled predictors selected for tree building, minimum number of data at a terminal tree node, and maximum imbalance of a split at a node of a tree). For further details, we refer a reader to the documentation of these packages.

Because Figure 1 clearly shows that the realized variances of oil returns exhibited large fluctuations during our sample period, we use recursive- and rolling-estimation windows to estimate our forecasting models. We analyze five and ten year windows (that is, windows of 60 and 120 months), where “window” in the case of the recursive-estimation window means the training period used to initialize the estimations.

In order to evaluate forecast accuracy, we use a wide array of standard statistics. First, we compute the root-mean-square-forecast-error (RMSFE) for every model, and then form the ratio of the RMSFE of two models to compare these two models. Second, we use the mean-absolute forecast error (MAFE) to compare the forecasts across models. The MAFE uses the L1 loss function to evaluate forecasts errors (the RMSFE uses the L2 loss function) and, thereby, is less sensitive to large forecast errors than the RMSFE. Occasional large forecast errors are likely to results from the peaks of the realized variance documented in Figure 1. In order to assess whether our results are sensitive to large outliers, we also report results for \sqrt{RV} when we use statistical

tests to study forecast accuracy. Third, we use the Clark and West (2007) test to shed light on the statistical significance of differences of forecasts across models. The null hypothesis of the Clark-West test is that a benchmark and a rival model produce the same mean-squared prediction error (MSPE). The alternative hypothesis is that the rival model produces a smaller MSPE than the benchmark model. Fourth, because Model 2 is not a strictly nested version of Models 3 and 4, and the linear Model 3 is not a strictly nested version of the nonlinear Model 4, we also compare (when we study a rolling-estimation window) the accuracy of forecasts by means of the Diebold and Mariano (1995) test, where we use the modified version of the test as suggested by Harvey et al. (1997). The null hypothesis is that of equal forecast accuracy across the benchmark and rival models. The alternative hypothesis is that the forecasts implied by the rival model are more accurate than the forecasts implied by the benchmark model. We shall report results for both the L1 and the L2 loss function.

4 Empirical Results

Table 2 summarizes root-mean-square-forecast-error (RMSFE) ratios. A ratio larger than unity shows that the benchmark model produces a larger RMSFE than the rival model. We report results for a recursive- and a rolling-estimation window, two window lengths, and five different forecast horizons. For the recursive-estimation window, we find that Model 2 produces worse forecasts than Model 1 for forecast horizons up to six months, and produces only moderately seized forecasting gains for a twelve-months forecasting horizon. Similarly, Model 1 performs well when compared with Model 3, except for the two longest forecast horizons. The results further show that it is hardly possible to clearly rank Model 2 relative to Model 3. Model 3 outperforms Model 2 only for a forecast horizon of one and two years. In sharp contrast, Model 4 produces more accurate forecasts in terms of the RMSFE statistic than the other three models for a training window of 60 months, and for the longer training window of 120 months when we consider the long forecast horizons. When we switch to a rolling-estimation window, we find again that Model 2 does not systematically perform better than Model 1. Model 3, in turn, produces better forecasts in terms of the RMSFE statistic than Model 1 and 2 for more combinations

of rolling-estimation windows and forecast horizons than is the case for the recursive-estimation window, especially for the long forecast horizons. Model 4 outperforms the other three models for forecast horizons of $h \geq 6$, and for all forecast horizons when we consider the short training window.

– Table 2 about here. –

In sum, we find that adding the aggregate EMV, EPU, and GPR data as predictors to the forecast model does not reliably improve forecast performance in terms of the RMSFE statistic relative to the simple benchmark AR(1) model. Adding all disaggregated EMV, EPU, and GPR to the benchmark AR(1) model and, given the large number of additional predictors, using the Lasso shrinkage estimator helps to improve forecast accuracy relative to the AR(1) benchmark and the AR(1) cum aggregate EMV and GPR predictors model, especially when we study a long forecast horizon and/or use a rolling-estimation window. Using the disaggregated data in conjunction with random forests, in turn, produces the best results relative to the other models at forecast horizons of $h \geq 6$, and also for the short forecast horizons when we opt for the short training/rolling window.

– Table 3 about here. –

Because Figure 1 witnesses that the realized variance of oil-price returns exhibited several sudden peaks during the sample period under scrutiny, we report in Table 3 the results we obtain when we focus on the mean-absolute forecast error (MAFE). The overall picture does not change. Model 4 systematically outperforms the other models except when we study one of the short forecast horizons and use a training window or rolling-estimation window of ten years length. Model 2, in turn, has a hard time to beat Model 1, the AR(1) benchmark model, at the short forecast horizons and Model 3 only tends to produce better forecasts than Model 2 in terms of the MAFE statistic when we use a rolling-estimation window.

– Table 4 about here. –

Next, we report in Table 4 the results of the Clark-West test for an equal mean-squared prediction error (MSPE). The alternative hypothesis is that the rival model produces a smaller MSPE than the benchmark model. Model 2 performs significantly better than Model 1 when we study a rolling-estimation window and when we consider a recursive-estimation window, but only when we combine a short forecast horizon with a long training period. Model 3, in turn, significantly outperforms Model 1 for several model configurations in case of the three long forecast horizons when we use a rolling-estimation window. Model 3 also significantly outperforms Model 2 mainly when we study an intermediate forecast horizon and a recursive-estimation window, and in seven out of ten model configurations when we analyze a rolling-estimation window. Hence, we find, on balance, evidence that using disaggregated EMV, EPU, and gPR data in conjunction with the Lasso estimator often helps to improve forecast accuracy. Using random forests, however, even produces better forecasts in the majority of cases especially for the rolling-estimation window, where the overwhelming majority of test results for Model 4 is significant. For the recursive-estimation window, Model 4 performs better than Model 1 at the intermediate forecast horizons, and it performs better than Models 2 and 3 especially for the long forecast horizons and when we use the long training window.

– Table 5 about here. –

Table 5 summarizes the test results that we obtain when we forecast the square-root of the realized variance of oil-price returns, that is, the realized volatility. For the recursive-estimation window, we find significant test results in the vast majority of model configurations for forecast horizons of $h \geq 6$. Hence, Models 2 to 4 perform significantly better than the AR(1) benchmark model, while Model 3 produces significantly better results than Model 2, and Model 4 performs better in a statistically significant way than the other three models. This is also true for a rolling-estimation window, where, in addition, several test results are significant also for a forecast horizon of three months. For a forecast horizon of one months, the majority of tests yields insignificant results, but Model 4 tends to perform better than Models 2 and 3 (but not Model 1) in a statistically significant way when we assume a rolling-estimation window.

– Table 6 about here. –

Finally, we present results for the modified Diebold-Mariano test in Table 6 for the case of a rolling-estimation window and an absolute-error loss and a squared-error loss function (at this stage, it should be noted that Models 1 to 3 are not strictly nested versions of Model 4 due to the nonlinear nature of the latter). According to the test results, the forecasts computed by means of Model 2 are significantly better than the forecasts computed by means of Model 1 only for the longest forecast horizon that we consider, $h = 24$, and mainly for the absolute-error loss function. Moreover, the evidence that Model 3 yields systematically more accurate forecasts than Models 1 and 2 is concentrated, under the absolute-error loss function, at forecast horizons of $h \geq 6$. Finally, the accuracy of the forecasts extracted from Model 4 is significantly better than the accuracy of the forecasts implied by the other three models for the majority of model configurations for the two longest forecast horizons under consideration, $h = 12, 24$, for both loss functions (and for a few model configurations at $h = 6$).

5 Concluding Remarks

Our findings, derived using machine-learning techniques, show that EMV, EPU, and GPR are useful for predicting the realized variance (and the realized volatility) of oil-price returns when the forecast horizon is sufficiently long. However, our findings also show that taking a disaggregated perspective by considering the various categories of EMV, EPU, and GPR may help to further improve the accuracy of forecasts of the realized variance of oil-price returns. This finding obtains (i) mainly at intermediate and long forecast horizons, and, (ii) mainly when we use random forests to estimate our forecasting model. Hence, our findings suggest that a forecaster who needs forecasts at intermediate and long forecast horizons may find it useful to exploit the interplay of the disaggregated data and potential nonlinear links of the realized variance of oil-price returns to the disaggregated predictors when computing forecasts of the realized variance of oil-price returns.

Our findings clearly show that it is interesting to explore whether a disaggregated view of uncertainty and geopolitical risks also is useful for forecasting the realized variances of the movements of the prices of other fossil resources and perhaps also those of agricultural commodities.

In addition, the comparatively good performance of random forests in our forecasting experiment suggests that it is worthwhile to study the predictive performance of other widely-studied machine-learning techniques like, for example, support vector machines or boosted regression trees for the realized variance of the returns of the oil price and other commodities. We leave this for future research.

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.
- Asai, M. Gupta, R., and McAleer, M. (2020). Forecasting volatility and co-volatility of crude oil and gold futures: Effects of leverage, jumps, spillovers, and geopolitical risks. *International Journal of Forecasting*, 36(3), 933–948.
- Asai, M. Gupta, R., and McAleer, M. (2019). The impact of jumps and leverage in forecasting the co-volatility of oil and gold futures. *Energies*, 12, 3379.
- Bakas, D., and Triantafyllou, A. (2018). The impact of uncertainty shocks on the volatility of commodity prices. *Journal of International Money and Finance*, 87, 96–111.
- Bakas, D., and Triantafyllou, A. (2020). Commodity price volatility and the economic uncertainty of pandemics. *Economics Letters*, 193, 109283.
- Baker, S.R., Bloom, N.A., and Davis, S.J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4), 1593–1636.
- Baker, S.R., Bloom, N.A., Davis, S.J., and Kost, K.J. (2019). Policy news and stock market volatility. NBER Working Paper No. 25720.
- Bonaccolto, M., Caporin, M., and Gupta, R. (2018). The dynamic impact of uncertainty in causing and forecasting the distribution of oil returns and risk? *Physica A: Statistical Mechanics and its Applications*, 507, 446–469.
- Bonato, M. (2019). Realized correlations, betas and volatility spillover in the agricultural commodity market: What has changed? *Journal of International Financial Markets, Institutions and Money*, 62, 184–202.
- Bonato, M., Gkillas, K., Gupta, R., and Pierdzioch, C. (2020c). Investor happiness and predictability of the realized volatility of oil price. *Sustainability*, 12, 4309.

- Bouri, E., Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Infectious diseases, market uncertainty and realized volatility of oil. *Energies*, 13(16), 4090.
- Bouri, E., Gupta, R., Pierdzioch, C., and Salisu, A.A. (2021). El Niño and forecastability of oil-price realized volatility. *Theoretical and Applied Climatology*, 144, 1173–1180.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Caldara, D., and Iacoviello, M. (2019). Measuring geopolitical risk. Working Paper, Board of Governors of the Federal Reserve Board.
- Chan, J.C.C, and Grant, A. (2016). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics*, 54, 182–189.
- Clark, T. D., and West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138, 291–311.
- Degiannakis, S., and Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, 76, 28–49.
- Demirer, R., Gkillas, K., Gupta, R., and Pierdzioch, C. (2021). Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests. *Journal of the Operational Research Society*. DOI: <https://doi.org/10.1080/01605682.2021.1936668>.
- Demirer, R., Gupta, R., Ji, Q., and Tiwari, A.K. (2019). Geopolitical risks and the predictability of regional oil returns and volatility. *OPEC Energy Review*, 43(3), 342–361.
- Demirer, R., Gupta, R., Pierdzioch, C., and Shahzad, S.J.H. (2020). The predictive power of oil price shocks on realized volatility of oil: A note. *Resources Policy*, 69(C), 101856.
- Demirer, R., Gupta, R., Suleman, M.T., and Wohar, M.E. (2018). Time-varying rare disaster risks, oil returns and volatility. *Energy Economics*, 75(C), 239–248.
- Diebold, F.X., and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economics Statistics*, 13, 253–263.

- Dutta, A., Bouri, E., and Saeed, T. (2021). News-based equity market uncertainty and crude oil volatility. *Energy*, 222, 119930.
- Friedman, J. Hastie, T., and Tibshirani, R. (2010). Regularization paths for Generalized Linear Models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Gabauer, D., and Gupta, R. (2020). Spillovers across macroeconomic, financial and real estate uncertainties: A time-varying approach? *Structural Change and Economic Dynamics*, 52, 167–173.
- Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Forecasting realized oil-price volatility: The role of financial stress and asymmetric loss. *Journal of International Money and Finance*, 104(C), 102137.
- Gupta, R., Ma, J., Risse, M., and Wohar, M.E. (2018). Common business cycles and volatilities in U.S. states and MSAs: The role of economic uncertainty. *Journal of Macroeconomics*, 5, 317–337.
- Gupta, R., and Pierdzioch, C. (2021). Forecasting the volatility of crude oil: The role of uncertainty and spillovers. *Energies*, 14(14), 4173.
- Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281–291.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. 2nd ed., Springer: New York, NY, USA.
- Lee, C.C., Olasehinde-Williams, G., and Akadiri, S.S. (2021). Are geopolitical threats powerful enough to predict global oil price volatility?. *Environmental Science and Pollution Research*, 28, 28720–28731.
- Li, X., Wei, Y., Chen, X., Ma, F., Liang, C., and Chen, W. (2020). Which uncertainty is powerful to forecast crude oil market volatility? New evidence. *International Journal of Finance and Economics*. DOI: <https://doi.org/10.1002/ijfe.2371>.

- Liang, C., Wei, Y., Li, X., Zhang, X., and Zhang, Y. (2020). Uncertainty and crude oil market volatility: new evidence. *Applied Economics*, 52(27), 2945–2959.
- Liu, J., Ma, F., Tang, Y., and Zhang, Y. (2019). Geopolitical risk and oil volatility: A new insight. *Energy Economics*, 84, 104548.
- Luo, J., Demirer, R., Gupta, R., and Ji, Q. (Forthcoming). Forecasting oil and gold volatilities with sentiment indicators under structural breaks. *Energy Economics*.
- Lux, T., Segnon, M., and Gupta, R. (2016). Forecasting crude oil price volatility and value-at-risk: Evidence from historical and recent data. *Energy Economics*, 56, 117–133.
- Mei, D., Ma, F., Liao, Y., and Wang, L. (2020). Geopolitical risk uncertainty and oil future volatility: evidence from MIDAS models. *Energy Economics*, 86, 104624.
- 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 (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Salisu, A.A., Gupta, R., Bouri, E., and Ji, Q. (2021b). Mixed-frequency forecasting of crude oil volatility based on the information content of global economic conditions. *Journal of Forecasting*. DOI: <https://doi.org/10.1002/for.2800>.
- Salisu, A.A., Gupta, R., and Olaniran, A. (2021a). The effect of oil uncertainty shock on real GDP of 33 countries: a global VAR approach. *Applied Economics Letters*. DOI: <https://doi.org/10.1080/13504851.2021.1983134>.
- Tibshirani, J., Athey, S. E. Sverdrup, and Wager, S. (2021). grf: Generalized Random Forests. R package version 2.0.2. <https://CRAN.R-project.org/package=grf>.
- Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58: 267–288.

- Van Robays, I. (2016). Macroeconomic uncertainty and oil price volatility. *Oxford Bulletin of Economics and Statistics*, 78(5), 671–693.
- van Eyden R., Difeto, M., Gupta, R., and Wohar, M.E. (2019). Oil price volatility and economic growth: Evidence from advanced economies using more than a century of data. *Applied Energy*, 233, 612–621.

Statements and Declarations:**Funding:**

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions:

Both authors contributed equally to the study conception and design.

Table 1: Estimated Models

Model	Predictors	Estimator
Model 1	AR(1) benchmark model	OLS
Model 2	AR(1) plus aggregate EMV, EPU, and GPR predictors	OLS
Model 3	AR(1) plus disaggregate EMV, EPU, and GPR predictors	Lasso
Model 4	AR(1) plus disaggregate EMV, EPU, and GPR predictors	Random forest

The Lasso estimator uses 10-fold cross validation to determine the shrinkage parameter, where the optimal shrinkage parameter minimizes the mean cross-validated error. Random forests are built using 2,000 individual random regression trees. Cross-validation is used to tune the number of predictors are chosen at random for tree building, the minimum number of data at a terminal node, and the parameter that controls the maximum imbalance of a split.

Figure 1: Realized Variance and Autocorrelation Function

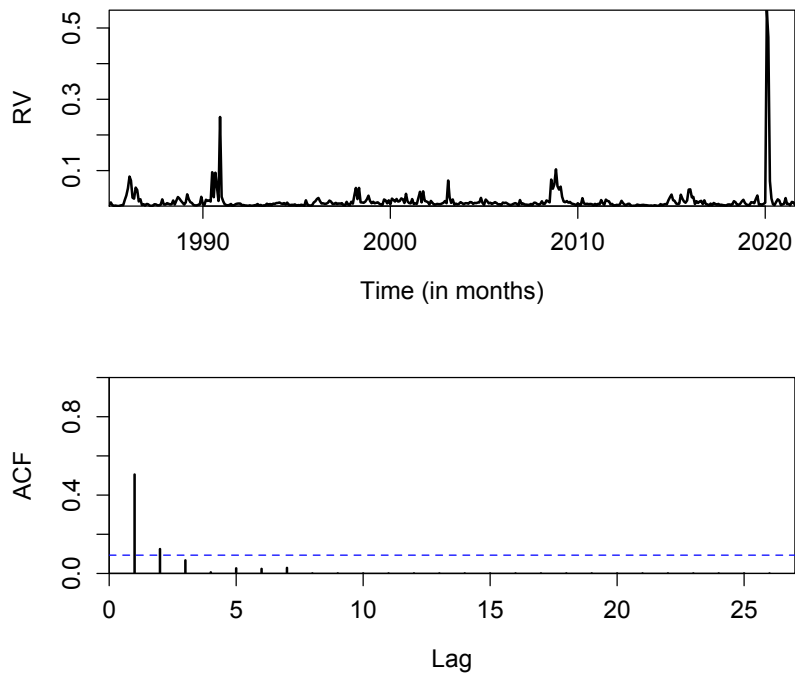


Figure 2: EMV and GPR Data

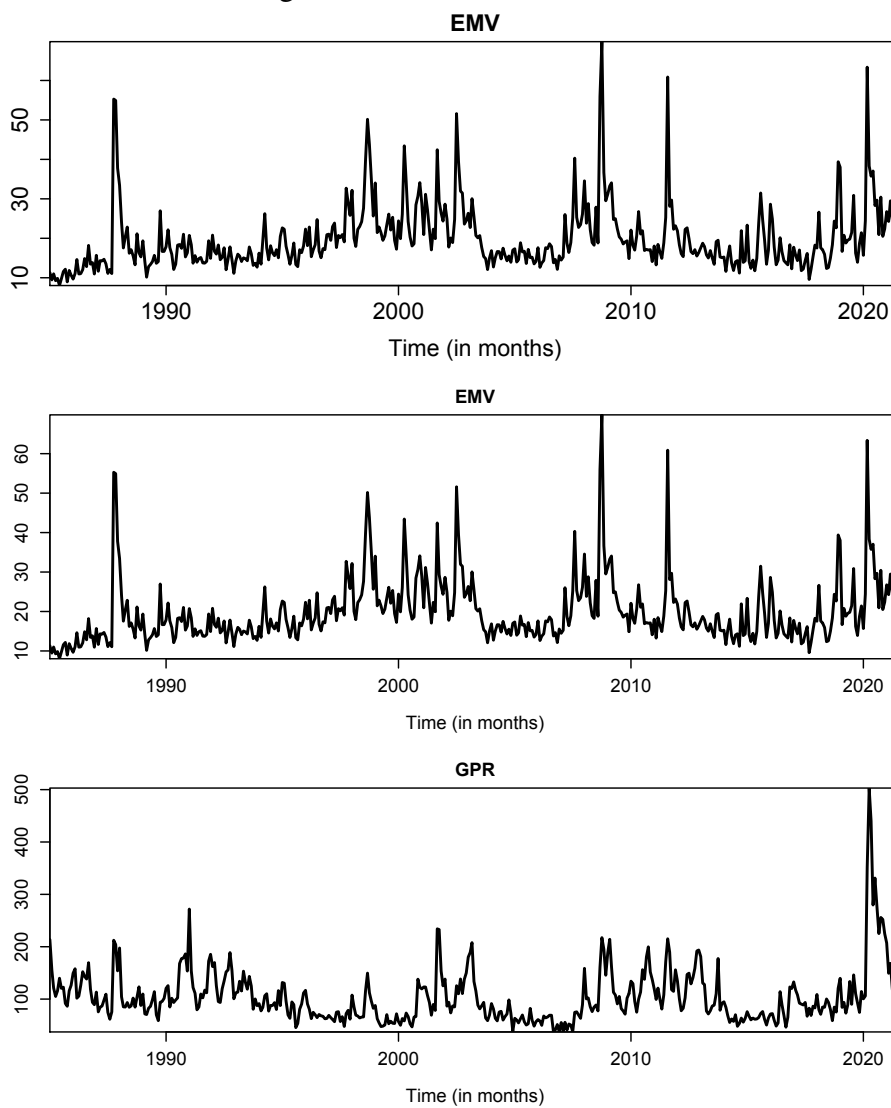


Table 2: RMSFE Ratios

Models	Window	h=1	h=3	h=6	h=12	h=24
<i>Recursive-estimation window</i>						
Model 1 vs. Model 2	60	0.9510	0.9672	0.9723	1.0083	1.0655
Model 1 vs. Model 2	120	0.9457	0.9641	0.9638	1.0085	1.0769
Model 1 vs. Model 3	60	1.0198	0.9566	0.9960	1.0648	1.2809
Model 1 vs. Model 3	120	0.9446	0.9539	0.8747	1.0623	1.3359
Model 2 vs. Model 3	60	1.0723	0.9890	1.0244	1.0560	1.2021
Model 2 vs. Model 3	120	0.9989	0.9894	0.9075	1.0534	1.2406
Model 1 vs. Model 4	60	1.0599	1.2019	1.1843	1.0883	1.4947
Model 1 vs. Model 4	120	0.9424	0.9417	0.9813	1.0950	1.5494
Model 2 vs. Model 4	60	1.1144	1.2427	1.2181	1.0793	1.4028
Model 2 vs. Model 4	120	0.9965	0.9767	1.0181	1.0857	1.4387
Model 3 vs. Model 4	60	1.0393	1.2564	1.1891	1.0221	1.1670
Model 3 vs. Model 4	120	0.9976	0.9872	1.1219	1.0307	1.1597
<i>Rolling-estimation window</i>						
Model 1 vs. Model 2	60	0.9680	1.0117	0.9955	0.9898	1.1832
Model 1 vs. Model 2	120	0.8480	0.9611	0.9947	1.0684	1.2942
Model 1 vs. Model 3	60	1.0431	0.8969	1.0737	1.1060	1.3008
Model 1 vs. Model 3	120	0.9130	1.0464	1.0142	1.1407	1.4493
Model 2 vs. Model 3	60	1.0775	0.8866	1.0785	1.1174	1.0994
Model 2 vs. Model 3	120	1.0767	1.0887	1.0195	1.0677	1.1198
Model 1 vs. Model 4	60	1.0435	1.2044	1.1999	1.1474	1.5299
Model 1 vs. Model 4	120	0.8942	0.9634	1.0226	1.1492	1.7800
Model 2 vs. Model 4	60	1.0780	1.1905	1.2053	1.1593	1.2930
Model 2 vs. Model 4	120	1.0544	1.0023	1.0281	1.0757	1.3754
Model 3 vs. Model 4	60	1.0004	1.3429	1.1176	1.0375	1.1761
Model 3 vs. Model 4	120	0.9794	0.9207	1.0084	1.0075	1.2282

The root-mean-square-forecast error (RMSFE) ratios are computed as the RMSFE of the forecast errors computed by means of the nominator model and the RMSFE of forecast errors computed by means of the denominator model. A RMSFE ratio larger than unity indicates that the denominator model produces more accurate forecasts in terms of the RMSFE criterion. The parameter h denotes the forecast horizon (in months).

Table 3: MAFE Ratios

Nominator vs. denominator model	Window	h=1	h=3	h=6	h=12	h=24
<i>Recursive-estimation window</i>						
Model 1 vs. Model 2	60	0.9491	0.9799	0.9989	0.9923	1.0395
Model 1 vs. Model 2	120	0.9291	0.9687	0.9718	0.9718	1.0448
Model 1 vs. Model 3	60	0.9058	0.9014	0.9472	1.0298	1.2271
Model 1 vs. Model 3	120	0.8979	0.9184	0.8391	0.9598	1.2030
Model 2 vs. Model 3	60	0.9543	0.9199	0.9482	1.0378	1.1805
Model 2 vs. Model 3	120	0.9664	0.9480	0.8635	0.9876	1.1515
Model 1 vs. Model 4	60	0.9626	1.0280	1.0987	1.1788	1.4908
Model 1 vs. Model 4	120	0.9217	0.9542	0.9881	1.1600	1.5068
Model 2 vs. Model 4	60	1.0142	1.0491	1.0998	1.1880	1.4342
Model 2 vs. Model 4	120	0.9920	0.9850	1.0168	1.1937	1.4422
Model 3 vs. Model 4	60	1.0627	1.1404	1.1599	1.1447	1.2149
Model 3 vs. Model 4	120	1.0266	1.0390	1.1775	1.2087	1.2525
<i>Rolling-estimation window</i>						
Model 1 vs. Model 2	60	0.9181	0.9753	0.9944	1.0148	1.2372
Model 1 vs. Model 2	120	0.7901	0.9442	0.9888	1.0643	1.2244
Model 1 vs. Model 3	60	0.9598	0.9871	1.1373	1.1272	1.4425
Model 1 vs. Model 3	120	0.8811	1.0104	0.9598	1.0796	1.3953
Model 2 vs. Model 3	60	1.0454	1.0121	1.1437	1.1107	1.1659
Model 2 vs. Model 3	120	1.1153	1.0702	0.9707	1.0144	1.1396
Model 1 vs. Model 4	60	0.9604	1.0385	1.1174	1.2505	1.6571
Model 1 vs. Model 4	120	0.8254	0.9444	1.0341	1.2474	1.7047
Model 2 vs. Model 4	60	1.0461	1.0648	1.1237	1.2322	1.3394
Model 2 vs. Model 4	120	1.0448	1.0002	1.0459	1.1721	1.3923
Model 3 vs. Model 4	60	1.0007	1.0521	0.9825	1.1094	1.1488
Model 3 vs. Model 4	120	0.9368	0.9346	1.0775	1.1555	1.2218

The mean-absolute-forecast error (MAFE) ratios are computed as the MAFE of the forecast errors computed by means of the nominator model and the MAFE of forecast errors computed by means of the denominator model. A MAFE ratio larger than unity indicates that the denominator model produces more accurate forecasts in terms of the MAFE criterion. The parameter h denotes the forecast horizon (in months).

Table 4: Baseline Test Results

Benchmark vs. rival model	Window	h=1	h=3	h=6	h=12	h=24
<i>Recursive-estimation window</i>						
Model 1 vs. Model 2	60	0.6935	0.5197	0.4857	0.4712	0.4391
Model 1 vs. Model 2	120	0.0842	0.0509	0.0882	0.3120	0.1444
Model 1 vs. Model 3	60	0.2231	0.2434	0.0203	0.3416	0.3062
Model 1 vs. Model 3	120	0.8563	0.1148	0.1174	0.2082	0.3789
Model 2 vs. Model 3	60	0.1169	0.0205	0.0074	0.2441	0.1246
Model 2 vs. Model 3	120	0.0199	0.0165	0.0497	0.4250	0.2008
Model 1 vs. Model 4	60	0.1346	0.1577	0.1075	0.0002	0.1427
Model 1 vs. Model 4	120	0.1541	0.1471	0.0337	0.0605	0.2544
Model 2 vs. Model 4	60	0.1100	0.1486	0.1011	0.0000	0.0365
Model 2 vs. Model 4	120	0.0019	0.0473	0.0275	0.0299	0.1326
Model 3 vs. Model 4	60	0.0084	0.1027	0.0101	0.0082	0.0002
Model 3 vs. Model 4	120	0.0004	0.0015	0.0011	0.0190	0.0039
<i>Rolling-estimation window</i>						
Model 1 vs. Model 2	60	0.0469	0.0507	0.0077	0.0443	0.0068
Model 1 vs. Model 2	120	0.1346	0.0086	0.0053	0.0001	0.4706
Model 1 vs. Model 3	60	0.1360	0.4150	0.0012	0.1968	0.0160
Model 1 vs. Model 3	120	0.8476	0.0793	0.0531	0.0329	0.1973
Model 2 vs. Model 3	60	0.0513	0.4479	0.0002	0.1538	0.0003
Model 2 vs. Model 3	120	0.0072	0.0172	0.0253	0.1579	0.0110
Model 1 vs. Model 4	60	0.0893	0.1234	0.0676	0.0010	0.0214
Model 1 vs. Model 4	120	0.1340	0.0207	0.0008	0.0177	0.2152
Model 2 vs. Model 4	60	0.0372	0.1045	0.0466	0.0118	0.0024
Model 2 vs. Model 4	120	0.0069	0.0027	0.0017	0.0004	0.0068
Model 3 vs. Model 4	60	0.0012	0.1243	0.0995	0.0092	0.0173
Model 3 vs. Model 4	120	0.0632	0.0021	0.0005	0.0098	0.1504

Results (p-values; robust heteroskedasticity and autocorrelation consistent standard errors) of the Clark-West tests for an equal adjusted mean-squared prediction error (MSPE). The alternative hypothesis is that the rival model produces a smaller MSPE than the benchmark model. The parameter h denotes the forecast horizon (in months).

Table 5: Test Results (\sqrt{RV})

Benchmark vs. rival model	Window	h=1	h=3	h=6	h=12	h=24
<i>Recursive-estimation window</i>						
Model 1 vs. Model 2	60	0.9011	0.7817	0.2519	0.0006	0.0001
Model 1 vs. Model 2	120	0.8647	0.1791	0.0268	0.0005	0.0001
Model 1 vs. Model 3	60	0.9253	0.1127	0.0173	0.0004	0.0000
Model 1 vs. Model 3	120	0.2531	0.1241	0.0825	0.0109	0.0065
Model 2 vs. Model 3	60	0.1877	0.0622	0.0058	0.0002	0.0000
Model 2 vs. Model 3	120	0.0593	0.1132	0.0609	0.0105	0.0049
Model 1 vs. Model 4	60	0.1773	0.1831	0.0485	0.0000	0.0000
Model 1 vs. Model 4	120	0.4821	0.3011	0.0083	0.0091	0.0007
Model 2 vs. Model 4	60	0.1504	0.1613	0.0787	0.0000	0.0000
Model 2 vs. Model 4	120	0.3351	0.3923	0.0745	0.0007	0.0001
Model 3 vs. Model 4	60	0.1096	0.0151	0.0007	0.0005	0.0001
Model 3 vs. Model 4	120	0.1775	0.0081	0.0048	0.0022	0.0002
<i>Rolling-estimation window</i>						
Model 1 vs. Model 2	60	0.7008	0.0984	0.0055	0.0005	0.0000
Model 1 vs. Model 2	120	0.2579	0.0111	0.0051	0.0007	0.0009
Model 1 vs. Model 3	60	0.5569	0.0526	0.0000	0.0001	0.0000
Model 1 vs. Model 3	120	0.7923	0.0260	0.0286	0.0010	0.0008
Model 2 vs. Model 3	60	0.0476	0.0131	0.0000	0.0000	0.0000
Model 2 vs. Model 3	120	0.2360	0.0077	0.0131	0.0001	0.0000
Model 1 vs. Model 4	60	0.1420	0.0737	0.0105	0.0000	0.0000
Model 1 vs. Model 4	120	0.4277	0.0072	0.0000	0.0000	0.0016
Model 2 vs. Model 4	60	0.0611	0.0710	0.0109	0.0177	0.0007
Model 2 vs. Model 4	120	0.2906	0.0721	0.0014	0.0000	0.0000
Model 3 vs. Model 4	60	0.0244	0.0862	0.0289	0.0015	0.0000
Model 3 vs. Model 4	120	0.0072	0.0024	0.0228	0.0010	0.0003

Results (p-values; robust heteroskedasticity and autocorrelation consistent standard errors) of the Clark-West tests for an equal adjusted mean-squared prediction error (MSPE). The alternative hypothesis is that the rival model produces a smaller MSPE than the benchmark model. The parameter h denotes the forecast horizon (in months).

Table 6: Diebold-Mariano Test

Benchmark vs. rival model	Window	h=1	h=3	h=6	h=12	h=24
<i>Absolute-error loss</i>						
Model 1 vs. Model 2	60	0.9981	0.7268	0.5562	0.3889	0.0018
Model 1 vs. Model 2	120	1.0000	0.9142	0.5827	0.1205	0.0595
Model 1 vs. Model 3	60	0.8407	0.7056	0.0168	0.0512	0.0001
Model 1 vs. Model 3	120	1.0000	0.4309	0.6912	0.2090	0.0683
Model 2 vs. Model 3	60	0.1553	0.3832	0.0188	0.0639	0.0036
Model 2 vs. Model 3	120	0.0052	0.1137	0.6439	0.4137	0.1376
Model 1 vs. Model 4	60	0.8080	0.2727	0.0037	0.0000	0.0003
Model 1 vs. Model 4	120	1.0000	0.8600	0.2899	0.0001	0.0231
Model 2 vs. Model 4	60	0.1419	0.1734	0.0008	0.0001	0.0014
Model 2 vs. Model 4	120	0.1678	0.4979	0.1529	0.0000	0.0090
Model 3 vs. Model 4	60	0.4915	0.2155	0.6251	0.0478	0.0517
Model 3 vs. Model 4	120	0.9683	0.8788	0.1657	0.0115	0.0594
<i>Squared-error loss</i>						
Model 1 vs. Model 2	60	0.9857	0.3872	0.5903	0.6343	0.0992
Model 1 vs. Model 2	120	0.9999	0.8701	0.5556	0.1435	0.1533
Model 1 vs. Model 3	60	0.2293	0.8752	0.1565	0.1646	0.0547
Model 1 vs. Model 3	120	0.9997	0.3337	0.4512	0.1813	0.1484
Model 2 vs. Model 3	60	0.0776	0.8334	0.1286	0.1124	0.0126
Model 2 vs. Model 3	120	0.0453	0.1580	0.4150	0.2222	0.1510
Model 1 vs. Model 4	60	0.2513	0.2109	0.0881	0.0077	0.0625
Model 1 vs. Model 4	120	0.9992	0.7599	0.3162	0.0650	0.1323
Model 2 vs. Model 4	60	0.0983	0.1984	0.0543	0.0065	0.0362
Model 2 vs. Model 4	120	0.1427	0.4750	0.1326	0.0088	0.1004
Model 3 vs. Model 4	60	0.4931	0.1800	0.2618	0.2986	0.0896
Model 3 vs. Model 4	120	0.7288	0.8762	0.4661	0.4541	0.0956

Results (p-values; robust heteroskedasticity and autocorrelation consistent standard errors) of the modified Diebold-Mariano test for equal accuracy of forecasts. Results are for a rolling-estimation window. The alternative hypothesis is that the rival model yields more accurate forecasts. The parameter h denotes the forecast horizon (in months).