University of Pretoria
Department of Economics Working Paper Series

Electricity Sales and Forecasting of Stock Market Realized Volatility: A State-Level Analysis of the United States

Matteo Bonato
University of Johannesburg
Oguzhan Cepni
Ostim Technical University, Ostim Technical University, Copenhagen Business School
Rangan Gupta
University of Pretoria
Christian Pierdzioch
Helmut Schmidt University
Working Paper: 2025-40
November 2025

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

Electricity Sales and Forecasting of Stock Market Realized Volatility: A State-Level Analysis of the United States

Submission: November 2025

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

Abstract

We study the out-of-sample forecasting value of and state-level and market-wide overall commercial, industrial, and residential electricity sales for monthly state-level (1995−2025) realized stock market volatility (RV) of the United States (U.S.). We control for state-level and market-wide realized moments (leverage, skewness, kurtosis, and tail risks). We estimate our forecasting models using a boosting algorithm, and two alternative statistical learning algorithms (forward best predictor selection and random forests). We find evidence that realized moments have predictive power for subsequent RV at forecast horizons up to one year in some model configurations, while evidence of predictive power of the growth rate of electricity sales, whether measured at state-level or at the market-level, is mixed and mainly concentrated, on average across states, at the short forecast horizon.

JEL Classifications: C22; C53; G10; G17; Q41

Keywords: Stock market; Realized volatility; Electricity sales; Statistical learning; Forecasting

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.

Data-availability statement: The data used to derive the results documented in this research are available from the authors upon reasonable request.

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

†Ostim Technical University, Ankara, Turkiye; University of Edinburgh Business School, Centre for Business, Climate Change, and Sustainability; Department of Economics, Copenhagen Business School, Denmark. Email address: oce.eco@cbs.dk.

‡Corresponding author. 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.

1 Introduction

Da et al. (2017) have highlighted the role of (industrial) electricity usage in forecasting stock returns of the United States (U.S.), as well as that of Japan and the United Kingdom (UK). We build on their research by analyzing the ability of (commercial, industrial, and residential) electricity sales in forecasting stock returns volatility of the U.S. over the monthly period of April 1995 to February 2025.

In this regard, we explore the predictive value of electricity sales for stock returns volatility along three dimensions. First, we forecast stock market volatility at the state-level rather than at the market-wide level. Stock prices for the states are derived from the sub-aggregation of firm-level stock prices within each of the 50 states considered, based on the location of their headquarters. The underlying reason for taking a disaggregated regional perspective is derived from the premise that core business activities of firms often occur close to their headquarters (Pirinsky and Wang, 2006; Chaney et al., 2012) and, hence, equity prices should contain a non-negligible regional component, to the extent that investors overweight local firms in their portfolios (Coval and Moskowitz, 1999, 2001; Korniotis and Kumar, 2013). Understandably then, the forecasting exercise that we undertake in this research should be of immense value to investors, given that accurate forecasts of stock market volatility carry widespread implications for portfolio selection, derivative pricing, and risk management (Poon and Granger 2003; Rapach et al., 2008; Bollerslev et al., 2018).

Second, rather than relying on model-based estimates of conditional variance, as can be derived from generalized autoregressive conditional heteroskedas-

ticity (GARCH) and stochastic volatility (SV) models, we employ a model-free method of computing monthly realized volatility (RV) as the square root of the sum of the available daily-data-based squared returns over a month (Andersen and Bollerslev, 1998). As a side effect, this realized approach also allows us accommodate for the role of other realized moments (leverage, skewness, kurtosis, and lower and upper tail risks) in our predictive framework, given widespread evidence of their importance in forecasting realized volatility of overall, regional and, sectoral stock returns (Mei et al., 2017; Zhang et al.,2021; Bonato et al., 2022, 2023a, forthcoming; Somani et al., forthcoming), and hence, control for possibly important omitted variables.

Third, because we consider realized moments and electricity sales both for the state and the overall U.S., this leads to inflating the number of predictors in our forecasting models. Hence, besides using the standard ordinary least squares (OLS) estimator, we rely on a boosting algorithm, and we conduct an in-depth analysis of the characteristics of the boosted forecasting models. In addition, we compare the boosting algorithm with two alternative statistical learning algorithms. Such a comparison helps to put the boosting results into perspective, but also renders it possible to synthesize the different algorithms by means of a model-averaging approach. We consider as a first alternative algorithm a forward best predictor selection algorithm, which is conceptually simpler than the boosting approach. As a second algorithm, we consider random forests, which yield forecasting models that are structurally more complex than the boosted forecasting models, but which also accommodate for the possibility of nonlinearity and potential interaction effects among the various predictor variables that

can enter into our forecasting models.

At this stage, it is important to discuss the underlying theoretical foundation of our analysis, and to put our research into perspective relative to earlier research. In much significant earlier research, electricity usage data have long been used to capture business cycles or the general state of the economy (Jorgenson and Griliches, 1967; Burnside et al., 1995, 1996; King and Rebello, 2000; Comin and Gertler, 2006). Electricity usage data are available in high quality because such data are accurately measured and reported due to the highly regulated electric utilities subjected to extensive disclosure requirements. Moreover, electricity usage data inform about the general state of the economy more or less in real time because electricity is difficult to store and the vast majority of modern (consumption and production) activities involves electricity usage in one way or the other (Payne, 2010; Pirlogea and Cicea, 2012; Da et al., 2017; Shahbaz et al., 2017; Doruk, 2024). Why then can data on electricity usage be useful for asset pricing? The standard present value model of asset pricing (Shiller, 1981a, b) predicts that asset market volatility depends on two factors: the variability of cash flows and the discount factor. Given that business cycles or changing economic conditions affect the volatility of variables that reflect future cash flows through fluctuations in economic uncertainty (Bernanke, 1983; Ludvigson et al., 2021) and by impacting the discount factor (Schwert, 1989), one can, in general, hypothesize a (negative) predictive relationship between electricity usage, in particular industrial sales reflecting the state of the business cycle, and equity market volatility.

While Lu et al. (2024) have recently highlighted the role of energy consump-

tion, including commercial, industrial, and residential electricity sales, in forecasting volatility of the aggregate S&P 500 index using various dimension reduction techniques,[1] to the best of our knowledge, ours is the first study to forecast state-level stock market volatility by utilizing the role of (year-on-year) growth rates of electricity sales, over and above realized moments (not considered by Lu et al., 2024), based on various statistical learning algorithms. Nationally aggregated data tends to overlook the heterogeneous nature of the comprising states, thus, potentially obscuring the true dynamics, unique characteristics, and variations within specific groups of states involving the nexus between stock returns volatility and electricity sales. In the process, our paper adds to the burgeoning literature on forecasting the U.S. state-level stock market volatility based on the information content of a wide range of state and national predictors, such as business applications (Bonato et al., 2023a), disaggregated oil shocks (Salisu et al., 2024), economic conditions and its volatility, energy-market and economic policy uncertainties (Salisu et al., 2025, forthcoming a; Candila et al., forthcoming), housing-price and its sentiment sentiment and attention (Salisu et al., forthcoming b), and even climate risks (Bonato et al., 2023b).[2]

In terms of earlier studies on forecasting state-level stock returns volatility of the U.S., we need to discuss how our research differs from a somewhat

[1] In-sample evidence for volatility, in addition to returns, at the aggregate- and industry-level due to industrial usage of electricity, based on a higher-order nonparametric causality model, has been reported by Bonato et al. (2018). See also Apergis and Payne (2014) for an earlier study providing evidence of bidirectional causal relationship between electricity consumption and stock prices in the Organization for Economic Cooperation and Development (OECD) countries.

[2] In this regard, by using commercial and residential electricity sales, which are impacted relatively more by weather conditions in comparison to industrial electricity sales, we are able to incorporate indirectly the role of disaster events emanating from the growing concerns of the association of physical risks with asset prices in the climate-finance literature (Giglio et al., 2021).

related paper by Salisu et al. (2025) mentioned above. This paper utilized the GARCH-mixed data sampling (MIDAS) framework to forecast the daily volatility of state-level stock returns of the U.S., based on a corresponding weekly economic conditions index (ECI), and its volatility. These indexes of economic conditions, as developed by Baumeister et al. (2024), apply a mixed-frequency dynamic factor model (DFM), to combine weekly, monthly, and quarterly data on mobility measures, labor market indicators, real economic activity (including electricity consumption), expectations measures, financial indicators, and household indicators. Besides highlighting the role of the volatility of the ECI, capturing economic uncertainty, in being a strong forecaster of state-level equity market volatility, these authors also depict the importance of the level of the ECI in producing forecasting gains for stock returns volatility of the states. While the ECI provides a broad high-frequency metric of business cycles for the US states, the fact that it contains data of various sectors and multiple frequencies, it, unlike electricity sales used by us, cannot necessarily be considered a real-time proxy for the state of the economy. This, in turn, is likely to over-emphasize its role in the out-of-sample predictability of the stock returns volatility of the U.S. states, more so because the univariate predictor-based GARCH-MIDAS approach that Salisu et al. (2025) use could possibly be suffering from a omitted-variable-bias by excluding the importance of state and market moments.

By analyzing for the first time the role of electricity sales for state-level U.S. stock market volatility, we also add to the extant literature on forecasting equity returns volatility of the U.S. based on a wide array of linear and nonlinear econometric models and economic, financial, and behavioral predictor variables.

Though beyond the scope of our reserarch, the reader is referred to Salisu et al. (2022) and Segnon et al. (2023) for comprehensive reviews of this literature.

In order to get to our empirical findings, we organize the rest of this research as follows. In Section 2, we provide a description of the data we use in our study, while we outline in Section 3 our methods. In Section 4, we present our empirical results. In Section 5, we conclude.

2 The Data

We employ daily stock log-returns returns for the 50 states of the U.S., as well as the same for the S&P 500, capturing the market-wide stock returns, to derive our monthly estimates of realized volatility and other realized moments, which we discuss in detail below. The state-level stock market indexes, and the S&P 500 index, are derived from the Bloomberg terminal, which, in turn, creates these indexes by taking the capitalization-weighted index of equities domiciled in a given state. The rationale behind this approach is grounded in the notion that the core business activities of firms often take place near their headquarters, influenced by the economic dynamics of that particular state.

As far as our dependent variable is concerned, we use the classical estimator of RV, i.e., the square root of the sum of squared daily returns (Andersen and Bollerslev, 1998), given as

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

where $r_{t,i}$ denotes the daily $M \times 1$ returns vector, and $i = 1, ..., M$ is the number of

daily returns over month t.

We plot summary statistics of the natural logarithm of the the state-level RV in Figure 1, where the solid line denotes the cross-state mean and the the boundaries of the shaded area denote the maximum and minimum across states in each month. It is evident from eyeballing Figure 1 that the RVs display a substantial variation across time and across states.

– Figure 1 about here. –

As outlined in Section 1, we control for daily-data-based realized moments, given their importance in the realized volatility literature, both at the state- and market-level. The list of realized moments is given by: realized skewness, $RSKEW$, realized kurtosis, $RKURT$, and realized upside and downside tail risks, TR_{up} and TR_{down}, besides a leverage effect, LEV, which is basically the value of negative realized returns which occurs on a particular month, and zero otherwise.

As in Amaya et al. (2015), we use $RSKEW$ to capture the asymmetry of the returns distribution, while $RKURT$ captures extremes. We compute $RSKEW$ as:

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

and $RKURT$ as:

$$RKURT_t = \frac{M\sum_{i=1}^{M} r_{t,i}^4}{RV_t^2}, \tag{3}$$

Finally, we consider the tail risk estimator by Hill (1975) to derive our realized upside and downside tail risks. We define $X_{t,i}$ as the set of reordered daily returns

on month t, $r_{t,i}$ in such a way that

$$X_{t,i} \geq X_{t,j} \text{ for } i < j. \tag{4}$$

Then, we derive the monthly positive tail risk estimator (TR_{up}) as:

$$TR_t^{up} = \frac{1}{k} \sum_{i=1}^{k} \ln(X_{t,i}) - \ln(X_{t,k}). \tag{5}$$

The (monthly) negative tail risk estimator (TR_{down}) is obtained as:

$$TR_t^{down} = \frac{1}{k} \sum_{i=n-k}^{n} \ln(X_{t,i}) - \ln(X_{t,n-k}) \tag{6}$$

where k is the observation denoting the chosen α (= 5%) tail interval.

We plot summary statistics on the state-level realized moments in Figure 2. The solid lines denote the cross-state mean and the the boundaries of the shaded area denote the maximum and minimum across states in each month.

– Figure 2 about here. –

We now turn to our main focus in terms of predictor variables, i.e., the electricity sales in the commercial ($COMM$), industrial ($INDUS$), and residential ($RESID$) sectors for the 50 states and the overall U.S., again derived from the Bloomberg terminal. Electricity sales basically correspond to the monthly sales in megawatt-hour (MWh) of electricity to ultimate customers in these three sectors. In line with Da et al. (2017), to remove seasonality, we work with the year-on-year growth rates of the sales, and also use a publication lag of two months, i.e., in month t, a forecaster has access to electricity sales data from

month $t-2$.

– Figure 3 about here. –

We plot summary statistics of the growth rates of state-level electricity sales in Figure 3, where the solid line in the panels for the state-level data denotes the cross-state mean and the the boundaries of the shaded area denote the maximum and minimum across states in each month. The cross-state variation of the growth rates of commercial and industrial electricity sales clearly is larger than the cross-state variation of the growth rate of residential sales.

In order to differentiate a market-wide from a state-level predictor variable, we add "-M" to the name of a predictor variable to denote a market-wide predictor variable. For example, LEV denotes a state-level leverage effect, while LEV-M denotes a market-wide leverage effect, and $COMM$ denotes the growth rate of state-level commercial electricity sales, while $COMM$-M denotes the corresponding market-wide growth rate.

Based on data availability of the variables under consideration at the time of writing of this paper, our analysis covers the effective sample period (that is, the sample period that obtains after transforming the data as described in the preceding paragraph) from April, 1995 to February, 2025.

3 Methods

3.1 Forecasting Models

Our baseline forecasting model, which we simply call the RV-model, is given by the following equation:

$$RV\text{-model} : RV_{s,t+h} = \beta_0 + \beta_1 RV_{s,t} + u_{s,t+h}, \tag{7}$$

where $RV_{s,t+h}$ denotes the average state-level, s, realized volatility over the forecast horizon, h, and β_0 and β_1 denote the coefficients to be estimated, while $u_{s,t+h}$ denotes the state-level disturbance term. We compute the average state-level realized volatility for $h > 1$ using data for the periods $t+1, ..., t+h$. Given its simplicity, we estimate the forecasting model given in Equation (7) by the ordinary least squares (OLS) technique. We consider four forecast horizons ranging from one month to one year. Hence, we set $h = 1, 3, 6, 12$.

In order to bring the data closer to normality, and to make sure that forecasts of RV do not take on negative values, we use the natural logarithm of RV for estimating the forecasting model given in Equation (7), and in order to ensure comparability across models, also for all other forecasting models that we study in our research. For forecast evaluation, however, we convert the forecasts back to anti-logs, where we account for the usual Jensen-Ito term.

We next modify Equation (7) to include a vector of state-level realized moments, $M_{s,t}$. As realized moments we consider a leverage effect, realized skewness, realized kurtosis, and the realized positive and realized negative tail risks.

We call the resulting forecasting model the RV-M model. This forecasting model is given by the following equation:

$$RV\text{-M model}: \quad RV_{s,t+h} = \beta_0 + \beta_1 RV_{s,t} + \beta_2 M_{s,t} + u_{s,t+h}, \tag{8}$$

where β_2 denotes an appropriately dimensioned vector of coefficients to be estimated.

As yet another extension, we add to Equation (8) a vector of market-wide realized moments, MM_t. The resulting RV-MM forecasting model is given by the following equation:

$$RV\text{-MM model}: \quad RV_{s,t+h} = \beta_0 + \beta_1 RV_{s,t} + \beta_2 M_{s,t} + \beta_3 MM_t + u_{s,t+h}, \tag{9}$$

where β_3 denotes a vector of coefficients to be estimated.

In order to explore the role of the growth rates of state-level electricity sales, we extend the forecasting model given in Equation (9) to include a vector of the growth rates of state-level electricity sales, E_t. The resulting RV-MM-E forecasting model is given by the following equation:

$$RV\text{-MM-E model}: \quad RV_{s,t+h} = \beta_0 + \beta_1 RV_{s,t} + \beta_2 M_{s,t} + \beta_3 MM_t + \beta_4 E_{s,t} + u_{s,t+h}, \tag{10}$$

where β_4 denotes a vector of coefficients to be estimated.

Finally, we extend the forecasting model given in Equation (10) a vector of growth rates of market-wide electricity sales, $EM_{s,t}$, resulting in the following

RV-MM-EM forecasting model:

$$RV\text{-MM-EM model}: \quad RV_{s,t+h} = \beta_0 + \beta_1 RV_{s,t} + \beta_2 M_{s,t} + \beta_3 MM_t + \beta_4 E_{s,t} + \beta_5 EM_{s,t} + u_{s,t+h}, \tag{11}$$

where β_5 denotes a vector of coefficients to be estimated.

3.2 Boosting Algorithm

Given that the forecasting models outlined in Equations (8) to (11) feature several predictor variables, we use a component-wise functional gradient descent boosting algorithm to estimate them. Here, we outline in a non-technical way the main elements of the boosting algorithm, closely following the exposition in Bonato et al. (2025). A technically-minded reader can find detailed descriptions and derivations in the research by, for example, Bühlmann (2006) and Bühlmann and Hothorn (2007).

In order to set the stage for our description of the boosting algorithm, we introduce some basic notation. We define the vector of predictor variables as x_t, and $f(x_t)$ as the corresponding prediction function (dropping the state-level subindex to simplify the notation), that is, the right-hand side of our forecasting models. In addition, we let $F(RV_{t+h}, f(x_t)$ denote the standard L2 loss function, and we define the empirical risk as $R = \sum_{t=1}^{T} F(RV_{t+h}, f(x_t)$, where T denotes the latest period of time for which data are available when a forecasting model is to be estimated.[3]

[3]When we estimate our forecasting model over a recursively expanding estimation window, the parameter T successively increases as we move the estimation window forward in time until we reach the end of the sample period. Similarly, when we consider a rolling-estimation window, the parameter T increases over time, but in addition the starting point for the summation

The boosting algorithm minimizes R over f which, unlike in the case of the OLS technique, does not only require estimation of the vector of parameters, β, but also a decision as to which predictor variables from the vector x_t to include in a forecasting model. The basic idea motivating the component-wise functional gradient descent boosting algorithm is to solve this minimization problem by using the fact that the OLS residuals of the forecasting models equal the negative of the derivative of the L2 loss function with respect to f. Hence, upon using every predictor variable as a component, also known in the boosting literature as a base learner, one proceeds by estimating by the OLS technique univariate regression models of the negative of the gradient vector, $\partial F/\partial f$, on all base learners separately. These regression models give predictions that are estimates of the negative gradient vector. It then is straightforward to identify, in terms of the residual sum of squares, the base learner that best fits the negative gradient vector. The best-fitting base learner is used to update the function, f, in small steps.

Next, one uses the updated function, f, to compute a new negative gradient vector. Based on this new negative gradient vector, one identifies a new base learner. The new base learner gives another update of f. This sequential process of selecting base learners results in the construction of what is called in the boosting literature a strong learner. Only those base learners (predictor variables) enter the strong learner that the boosting algorithm selects while descending along the gradient of R. Once the updating process stops, one obtains a final strong learner, that is, a forecasting model that has been optimized by the

successively increases as well. In order to keep the notation simple, we focus on the case of a recursive-estimation window.

boosting algorithm in a data-driven way. Moreover, because only selected base learners enter the strong learner, the boosting algorithm can be interpreted also as a model-shrinkage technique that helps a researcher to identify a parsimonious forecasting model.

Model-selection criteria can be used to determine when the updating process that underlies the component-wise functional gradient descent boosting algorithm stops. In the boosting literature, the following four model-selection criteria are commonly used: the Akaike Information Criterion (AIC) (trace), the AIC (active set), the gMDL (trace) criterion, and the the gMDL (active set) criterion. Here, MDL is the abbreviation for minimum description length, and the active set refers to the number of base learners used to construct a strong learner.

It is a well-known result in the boosting literature, and the results of our forecasting experiments confirm this result, that the trace-based model-selection criteria imply a larger number of updating iterations than the active-set based model-selection criteria. Hence, the trace-based model-selection criteria should result in more complex forecasting models than the active-set-based model-selection criteria. Similarly, the AIC-based model-selection criteria typically produce more complex forecasting models than the gMDL-based model-selection criteria.

3.3 Forecast Evaluation

We use statistics based on the mean-absolute forecast error (MAFE) and the root-mean-squared forecast error (RMSFE) to evaluate the out-of-sample predictive performance of our forecasting models. Specifically, we assess the relative

predictive performance of our forecasting models by computing, for every state, s, the forecasting gain (in percent) as

$$\text{MAFE-FG}_s = 100 \times \left(\frac{MAFE_{s,B}}{MAFE_{s,R}} - 1 \right), \tag{12}$$

$$\text{RMSFE-FG}_s = 100 \times \left(\frac{RMSFE_{s,B}}{RMSFE_{s,R}} - 1 \right), \tag{13}$$

where B denotes a benchmark forecasting model and R denotes a rival forecasting model. In case the MAFE (RMSFE) ratio exceeds unity, the rival forecasting model outperforms the benchmark forecasting model. A forecasting gain (loss), in percentages, is indicated by a positive (negative) value of the MAFE-FG (RMSFE-FG) statistic.

In order to study the statistical significance of a difference in the predictive performance of a benchmark and a rival forecasting model, we use two approaches. As our first approach, we use the Clark and West (2007) test statistic, which is a test of the null hypothesis that a benchmark forecasting model and a rival forecasting model have equal predictive performance. The one-sided alternative hypothesis is that the rival model performs better than the benchmark model.

The Clark-West (CW) test is a test of nested models. Hence, in order to use this test, we have to treat our forecasting models as nested models, which is justified because we add in a step-by-step way additional predictor variables to our forecasting models as we move from the simple RV forecasting model to the more complex RV-MM-EM forecasting model. We are aware of the fact, however, that statistical testing often is complicated by the complex structure of forecast-

ing models obtained from applying statistical learning techniques. For example, the RV-MM-E forecasting model is not necessarily a strictly nested version of the RV-MM-EM forecasting model in case the growth rate of the market-wide electricity sales drives the growth rate of state-level electricity sales out of a forecasting model. Similarly, in one of our extensions, we study random forests as an alternative statistical learning technique. Random forests are highly nonlinear, non-parametric estimators, and this clearly complicates statistical testing of differences in predictive performance across forecasting models.

We, therefore, also use a second approach that fully makes use of the cross-state dimension of our data. Specifically, we estimate by the OLS technique a regression model with the MAFE-FG (RMSFE-FG) statistic on the left-hand side and an intercept coefficient on the right-hand side. The regression model is of the following format:

$$\text{MAFE-FG}_s = \alpha_{MAFE} + e_s, \tag{14}$$

$$\text{RMSFE-FG}_s = \alpha_{RMSFE} + e_s, \tag{15}$$

where e_s denotes some state-specific disturbance term. A statistically significant positive intercept coefficient, α, indicates that, in the cross-section of states, the forecasting gains from using the rival rather than the benchmark forecasting model are positive and statistically significant. We use this regression model to test the null hypothesis that the intercept coefficient is equal to zero, against the one-sided alternative hypothesis that the intercept coefficient is positive.

In sum, we do not rely on a single test statistic to assess forecasting gains, but rather use different statistics (the MAFE-FG and RMSFE-FG statistics and

the CW test). In addition, we shall study in Section 4 whether the relative performance of our forecasting models is stable when we use alternative statistical learning techniques to estimate our forecasting models.

It is also interesting to study the potential economic benefits a forecaster can derive from applying the forecasting models. To this end, we consider, like Bollerslev et al. (2018), a mean-variance investor who allocates wealth between a risk-free money market account and a stock-market investment, given a constant Sharpe ratio and a given risk-aversion parameter. We then compute the expected utility (EU) using the out-of-sample volatility forecasts from our forecasting models for the rival forecasting model and the benchmark forecasting model. Equipped with the results for expected utility, we compute the utility forecasting gain as follows:

$$\text{Utility-FG}_s = 100 \times \left(\frac{EU_{s,R}}{EU_{s,B}} - 1 \right), \tag{16}$$

for every state, s. It should be noted that, as compared to the MAFE-FG and RMSFE-FG statistics, the roles of the rival and the benchmark forecasting models in the nominator and denominator on the right-hand-side of Equation (16) are reversed. Finally, we estimate the following cross-state regression model:

$$\text{U-FG}_s = \alpha_U + e_s. \tag{17}$$

Equipped with the estimation result, we test the null hypothesis that the intercept coefficient, α_U, is equal to zero, against the one-sided alternative hypothesis that the intercept coefficient is positive.

3.4 Computational Issues

We use the R language and environment for statistical computing (R Core Team, 2025) and the R add-on package "mboost" (version 2.9-11; Hofner et al., 2014; Hothorn et al., 2010; Bühlmann and Hothorn, 2007) to implement the boosting algorithm. We fix the maximum number of iterations at 500, but as the results that we summarize in Table 1 (see Section 4.1) demonstrate, the number of iterations typically is much smaller in our forecasting experiment. In addition, a common choice in the boosting literature is to set the learning rate to 0.1, and we follow this convention.

In order to setup our forecasting experiments, we use a recursively expanding estimation window to estimate our forecasting models, but we shall also present results for a rolling-estimation window (see Section 4.3). We implement the recursive-estimation-window approach by using the first half of the sample period to initialize the estimations. We then use the boosting algorithm to estimate our forecasting models. Based on the estimated boosted forecasting models, we compute out-of-sample forecasts. We then add another observation at the end of the recursive-estimation window to reestimate our forecasting models. We continue in this way until we reach the end of the sample period.

4 Empirical Results

4.1 Characteristics of the Boosted Forecasting Models

As mentioned in Section 3, the boosting algorithm can be implemented using alternative model-selection criteria. We, therefore, start our empirical analysis by studying which one of the four different model-selection criteria that we consider in our empirical analysis yields the best out-of-sample forecasting performance for our various forecasting models. To this end, we use the MAFE and RMSFE statistics. Specifically, we study for every combination of forecasting model and state which model-selection criterion minimizes the MAFE and RMSFE statistics.

We summarize the results in Table 1, where we report for every forecasting model the number of states for which a model-selection criteria yields the best performance. We do not report results for the RV forecasting model, because we estimate this model by the OLS technique. Two results emerge. First, the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criteria clearly dominate the AIC$_{trace}$ and AIC$_{actset}$ model-selection criteria for the majority of states, and for all forecasting models. Second, the gMDL$_{actset}$ model-selection criterion outperforms the MDL$_{trace}$ model-selection criterion, where the gMDL$_{actset}$ model-selection criterion performs particularly well under the RMSFE statistic.

– Table 1 about here. –

The results that we report in Figures 4 are in line with these results. We report in Figure 4 the cross-state distribution of the number of iterations, averaged across the out-of-sample period, the boosting algorithm needs to minimize the empirical risk function, given a model-selection criterion. Evidently, the boosting

algorithm stops iterating after fewer iterations when we use the active-set-based model-selection criteria rather than the trace-based model-selection criteria. In addition, the two gMDL-based model-selection criteria imply that the boosting algorithm takes fewer iterations than when we use the AIC-based model-selection criteria. We observe this pattern four all four forecast horizons.

– Figure 4 and 5 about here. –

In Figure 5 we report, for every base-learner, its variable importance (VIMP) as defined as the contribution of a base-learner to the reduction of the empirical risk function, accumulated across boosting iterations. We report the VIMP statistic, expressed in percent, for the RV-MM-EM forecasting model, because this model contains the largest number of candidate predictor variables. We compute the VIMP statistic for all states, and average the VIMP statistic for every state over the out-of-sample period. The results in Figure 5, for the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criteria, represent the distribution of the VIMP statistic across states. The VIMP statistic demonstrates that RV is the single most important predictor variable of the subsequent RV. The second most important predictor variable, especially at the short ($h = 1$) and intermediate ($h = 3$) forecast horizons, in terms of the VIMP statistic is the market-wide leverage effect (we add "-M" to the name of a predictor variable to denote market-wide predictor variable), although this predictor variable is far less important than the RV predictor variable.

– Figure 6 about here. –

A small numerical value of variable importance does not necessarily imply that a predictor variable is not included in the boosted forecasting model. Hence,

we plot in Figure 6 how often the boosting algorithm includes a predictor variable in the boosted forecasting model in the out-of-sample forecasting period, again for the RV-MM-EM forecasting model and the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criteria. We find that, across all states, the RV predictor and market-wide leverage effects are often included in the boosted forecasting model, the latter mainly at the short and intermediate forecast horizons. The state-level leverage effect and realized kurtosis also are relatively often included in the boosted forecasting model, but mainly at the short forecast horizon only. Another predictor variable that stands out, at the intermediate and long forecast horizons, is the growth rate of market wide commercial electricity sales. As indicated by the circles in the figure, which point to "outlier" states, the boosted forecasting models also feature the other state-level and market-wide realized moments and the growth rates of the other electricity-sales predictor variables for some states, especially under the gMDL$_{trace}$ model-selection criterion. For example, under the gMDL$_{trace}$ model-selection criterion, the growth rate of industrial electricity sales tends to gain in importance as the length of the forecast horizon increases. As expected, the gMDL$_{actset}$ model-selection criterion is more restrictive in this regard and tends to yield more parsimonious forecasting models with fewer predictor variables.

At a more disaggregated level, it is interesting to analyze time series that show in how many states which growth rate of electricity sales is included in the out-of-sample forecasting period in the boosted RV-MM-EM forecasting model. We plot these time series in Figure 7. Apparently, the gMDL$_{trace}$ model-selection criterion is less restrictive in including the electricity-sales predictor variables

in a forecasting model than the gMDL$_{actset}$ model-selection criterion. The growth rate of commercial market-wide electricity sales is included in the forecasting model in several states, but the importance of this predictor variable declines over time.[4]

– Figure 7 about here. –

In order to further illustrate the role played by the growth rates of electricity sales, we plot in Figures 8 and 9 wordclouds. In order to compute the wordclouds, we record for every state how often the different growth rates of electricity sales are included in the forecasting model under the gMDL$_{trace}$ and the gMDL$_{actset}$ model-selection criteria during the out-of-sample period. We then sum up across the different categories of electricity sales to capture the total frequency of inclusion of electricity sales in the boosted RV-MM-EM forecasting model for a state. The worldclouds illustrate this frequency.[5]

– Figures 8 and 9.bout here. –

Under the gMDL$_{trace}$ model-selection criterion (Figure 8), the wordcloud shows that the growth rate of electricity sales, aggregated across the six different categories, plays an important role (that is, relative to other states) at the short forecast horizon in the states New Mexico, Nevada, Rhode Island, Alaska, Idaho, and Oklahoma. When we consider the intermediate and long forecast horizons, we can add states like Arkansas, Idaho, Kanas, Mississippi, Maine, Wyoming,

[4]In Figure A1 at the end of the paper (Appendix), we plot for the out-of-sample period time series of the number of states for which the boosting algorithm includes at least one subcategory of the growth rates of electricity sales in the forecasting model.

[5]We use the R-add-on package "wordcloud2" (version 0.2.1; Lang and Chien, 2028) to compute the worldclouds.

and Vermont to this list. Because the gMDL$_{actset}$ model-selection criterion is more restrictive with regard to the inclusion of predictor variables in a forecasting model than the gMDL$_{trace}$ model-selection criterion, the wordclouds tend to depict fewer states under the gMDL$_{actset}$ model-selection criterion (Figure 9). For the short forecast horizon, the wordcloud depicts mainly New Mexico and, to a lesser extent, Kansas and Rhode Island. New Mexico and Alaska tend to dominate the wordclouds at the intermediate and long forecast horizons, but we can also include in this list Nevada, Mississippi, Maine, Rhode Island, West Virginia, and Vermont.

4.2 Forecasting Gains of the Boosted Forecasting Models

We start our analysis of the performance of the competing forecasting models by studying the MAFE-FG and RMSFE-FG statistics, where a statistic larger than zero indicates that the rival model performs better than the corresponding benchmark model. As rival and benchmark models, we consider the following model combinations: RV versus RV-M, RV-M versus RV-MM, RV-MM versus RV-MM-E, and RV-MM-E versus RV-MM-EM, that is, we include, step-by-step, additional predictor variables in the array of predictor variables from which the boosting algorithm can select predictor variables. In this way, we study whether the additional predictor variables help the rival model to go beyond the benchmark model in terms of predictive performance. We focus on the gMDL$_{actset}$ model-selection criterion (and to a lesser extent on the gMDL$_{trace}$ model-selection criterion), which gives the best MAFE and RMSFE statistics for the large majority of states (see Table 1), and we use box-and-whisker plots to summarize the

cross-state distribution of the MAFE-FG and RMSFE-FG statistics.

– Figures 10 and 11 about here. –

We summarize the results in Figures 10 and 11. Three results stand out. First, the RV-M forecasting model, that is, the model that results when the boosting algorithm can select predictor variables among the state-level realized moments, often fares better in terms of the MAFE-FG and RMSFE-FG statistics than the RV-model, which is estimated by the OLS technique and only features the contemporaneous RV as a predictor variable. We observe that this good performance of the RV-M forecasting model somewhat strengthens as the length of the forecast horizon increases. Second, adding market-wide realized moments to the array of predictor variables, so that the rival model is the RV-MM forecasting model, improves forecasting performance relative to the RV-M forecasting model for several states, especially at the short and intermediate forecast horizons, and under the RMSFE-FG statistic also for the long forecast horizon. Third, the RV-MM-E forecasting model, which obtains by adding the growth rates of the state-level electricity sales to the array of predictors, improves predictive performance for roughly half of the states at the short forecast horizon and at the intermediate ($h = 6$) forecast horizon. The sharp peak of the box-and-whisker plots at zero, however, indicate that the forecasting gains from using the growth rates of state-level electricity sales as additional predictor variables are moderate in most cases. This is also true for the growth rates of market-wide electricity sales, that is, the RV-MM-EM forecasting model. Moreover, the cross-state distributions of the MAFE-FG and the RMSFE-FG statistics are tilted to the left at the intermediate and long forecast horizons.

– Figure 12 about here. –

We next use the Clark and West (2007) test statistic to study the statistical significance of the potential gains in predictive performance from the competing forecasting models. The box-and-whisker plots that we summarize in Figure 12 display the cross-state distribution of the p-values (based on heteroscedasticity and autocorrelation consistent standard errors; Zeileis (2004) and Zeileis et al. (2020)) of this test statistic for the gMDL$_{actset}$ model-selection criterion (the results for the gMDL$_{trace}$ model-selection criterion are similar and are not reported for brevity). The emerging pattern of test results is largely in line with the results of our analysis of the MAFE-FG and RMSFE-FG statistics. Using state-level realized moments often leads to a rejection of the null hypothesis of an equal predictive performance when the RV forecasting model is the benchmark model. For market-wide realized moments, the test statistic also yields, albeit to a lesser extent than for state-level realized moments, statistically significant results for several states, mainly at the intermediate forecast horizons. Finally, we find evidence against the null hypothesis of an equal predictive performance for the RV-MM and the RV-MM-E forecasting models for approximately half of the states when we study an intermediate forecast horizon, $h = 6$. The test results for the RV-MM and RV-MM-E forecasting models, are only significant for a few states.

– Table 2 about here. –

In Table 2, we summarize the results of the cross-state regressions of the MAFE-FG and RMSFE-FG statistics on a constant. We report the p-values of a test of the null hypothesis that the constant is zero, against the alternative hypothesis

that the constant is positive, where we use heteroscedasticity and autocorrelation consistent standard errors. The results for the RV-M forecasting model are statistically significant relative to the RV forecasting model, except at the short forecast horizon. In turn, the RV-MM forecasting model yields statistically significant test results when compared with the RV-M forecasting model at the short and intermediate forecasting horizons for the MAFE-FG statistic, and at all forecasting horizons for the RMSFE-FG statistic. The RV-MM-E and RV-MM-EM forecasting models only yield statistically significant test results at the short forecasting horizon, but not for the MAFE-FG statistic and the gMDL$_{trace}$ model-selection criterion.

4.3 Boosting Results for a Rolling-Estimation Window

We next present results for a rolling-estimation window. To this end, we estimate the boosted forecasting models on a rolling-estimation window that covers 50% of the data. As compared to the recursive-estimation window, the rolling-estimation window features a constant number of observations, as we delete one observation at the beginning of the estimation window when we add a new observation at its end. We summarize the results (that is, the p-values) of the Clark-West test statistic in Figure 13, again for the gMDL$_{actset}$ model-selection criterion. The results for the rolling-estimation window confirm the results for the recursive-estimation window that we report in Figure 12. We find the strongest evidence against the null hypothesis of an equal predictive performance when we compare the RV-M forecasting model with the RV forecasting model. We further find that the test results for the RV-MM forecasting model, when compared with the RV-

M forecasting model, strengthen somewhat at the two short forecast horizons ($h = 1, 3$).

– Figure 13 and Table 3 about here. –

In Table 3, we summarize the results of the cross-state regressions of the MAFE-FG and RMSFE-FG statistics for a rolling-estimation window. The RV-M forecasting model yields several statistically significant test results, especially at the intermediate and long forecast horizons, when we compare this model with the RV forecasting model. The test results are statistically insignificant in three cases at the short forecast horizon. As for the RV-MM forecasting model, this model gives statistically significant test results when compared with the RV-M forecasting model mainly at the short forecast horizons ($h = 1, 3$). Finally, three test results for the RV-MM-E forecasting model are statistically significant at the short forecast horizon, but only one test result is statistically significant for the RV-MM-EM forecasting model.

4.4 Results for a Forward Best Predictor Selection Algorithm

The next step of our analysis is to estimate our forecasting models by means of an alternative statistical learning algorithm. A forward stepwise best predictor selection algorithm is the first alternative statistical learning algorithm that we consider (see Hastie et al. (2009), Chapter 3, for a textbook exposition). In order to setup this algorithm, we start with a forecasting model that features a constant as the only predictor variable. Estimation is done by the OLS technique.[6]

[6]We use the R-add-on package "leaps" (version 3.2; see Lumley, 2024, based on Fortran code by Miller) for estimation.

We then proceed by estimating, again by the OLS technique, all forecasting models that feature one additional predictor variable. For example, in case of the RV-M model, we estimate forecasting models that feature one of the state-level realized moments. We identify the one forecasting model that minimizes the in-sample residual sum of squares. This forecasting model then is the basis for the next estimation round, in which we estimate all models that feature two predictor variables, that is, the predictor variable selected in the first step plus one of the other predictor variables (in the case of the RV-M forecasting model, this is one of the remaining state-level realized moments). Again, we identify the model that minimizes the in-sample residual sum of squares, and then we proceed, using this model as a basis, to a model that features three predictor variables. In this way, we obtain a sequence of forecasting models. We identify the "best" forecasting model in this sequence as the one that (i) maximizes the adjusted R2 statistic, (ii) minimizes the Bayesian Information Criterion (BIC), or, (iii) minimizes Mallow's Cp criterion. We carry out the estimations for a recursive-estimation window and, thereby, compute sequences of "best" forecasting models. For example, we compute, for every model-selection criterion, sequences of optimal variants of the RV-MM-EM forecasting model, that is, models that can feature all or a subset of the predictor variables among which we can choose to build the RV-MM-EM forecasting model. Finally, we select the "best" forecasting model among the three "best" forecasting models identified by means of the adjusted R2 statistic, the BIC, and Mallow's Cp criterion. As witnessed by the results that we summarize in Table A1 at the end of the paper (Appendix), the clear winner of this competition is the forecasting model selected by means of

the BIC.

– Table 4 about here. –

In this way, we obtain recursively estimated sequences of optimal RV (again estimated by means of the OLS technique), RV-M, RV-MM, RV-MM-E, and RV-MM-EM forecasting models, which we can compare by means of the cross-state regressions of the MAFE-FG and RMSFE-FG statistics (Table 4).[7] While the test results for the comparison of the RV-MM with the the RV-M forecasting model are still significantly significant at the short and intermediate forecast horizons ($h = 1, 3$), the latter does not deliver forecasting gains that are, on average across the states, statistically significantly different from zero relatively to the RV-model. The test results for the RV-MM-E, and RV-MM-EM forecasting models indicate that the growth rates of electricity sales do not contribute to predictive performance in the cross-section of states beyond the performance that is already achieved by the respective benchmark models. Hence, on balance, and especially so for the growth rates of electricity sales, the results are weaker than those for the boosting algorithm.

– Figure 14 about here. –

Figure 14 plots for the RV-MM-EM model and the BIC model-selection criterion how often the predictor variables are included in the forecasting model. Resembling the results we obtain for the boosting algorithm, we find that RV is a top predictor variable, and that the market-wide leverage effect is often included in the forecasting model mainly at the short and intermediate forecast horizons.

[7] We summarize the cross-state distribution of the MAFE-FG and RMSFE-FG statistics at the end of the paper (Appendix). The reader is referred to Figures A2 and A3.

Across the states, the realized kurtosis tends to play a more important role than the state-level leverage effect, while the growth rate of market-wide commercial electricity sales is another relatively often selected predictor variable. The growth rate of industrial electricity sales tends to enter the forecasting model more often as the length of the forecast horizon increases.

4.5 Results for Random Forests

Random forests are a widely studied "off-the-shelf" non-linear and non-parametric ensemble statistical learning estimator (see Breiman, 2001). As compared to the statistical learning algorithms that we have studied so far, random forests have the advantages that they account in a fully data-driven way for potential non-linear patterns in the data as well as potential interaction effects between the predictors. A random forest consists of many individual regression trees, which, in turn, consist of a root and several nodes and branches (see, Breiman et al., 1984). These nodes and branches partition the space of the predictors in a recursive and binary way into non-overlapping regions, which are formed in a top-down way by applying a search-and-split algorithm that gives, at each level of a regression tree, the best splitting predictor variable and a corresponding splitting point. A recursive application of this search-and-split algorithm along the branches of a regression tree yields finer and finer partitions of the predictor space.

While finer partitions produce increasingly granular forecasts of RV, the complex hierarchical structure of a regression tree easily results in overfitting and data-sensitivity problems, which then hamper prediction performance. A ran-

dom forest is constructed in a way so as to overcome these problems as it formed by an ensemble of many regression trees. This ensemble is built by means of a bootstrap algorithm. To this end, we first compute a large number of bootstrap samples, where sampling is with replacement. Second, we estimate a random regression tree on every bootstrap sample. The specific feature of a random regression tree, as opposed to a conventional regression tree, is that tree growing is done using a random subset of the predictor variables for splitting. In this way, the effect of influential predictors on tree building is reduced. Third, we combine the random regression trees to a random forest and compute a forecast of RV by averaging across the individual random regression trees, which, in turn, stabilizes forecasts.[8]

– Table 5 about here. –

In Table 5, we summarize the results for the cross-state regressions of forecasting gains.[9] The RV-M forecasting model, when compared with the RV forecasting model, yields statistically significant test results, where the evidence is relatively strong at the short and the two intermediate forecast horizons. For the RV-MM forecasting model, only one of the test results is statistically significant (RMSFE-FG, $h = 1$). For the RV-MM-E (RV-MM-EM) forecasting model, in turn, the test results are statistically significant, when the model is compared with the RV-MM (RV-MM-E), at the short forecast horizon.

[8]We use the R add-on package “”randomForestSRC” (version 3.3.3) to estimate random forests. See Ishwaran and Kogalur (2025). We fix the maximum number of trees at 1,000, sample with replacement, and use otherwise the default parameters (e.g., node size) of the package for estimation of random forests.

[9]We summarize the results for random forests in Figure A4 (for the MAFE-FG statistic) and in Figure A5 (for the RMSFE-FG statistic) at the end of the paper (Appendix).

– Figure 15 about here. –

In order to study variable importance, we estimate random forests on the full sample and then compute the VIMP statistic as the difference between the (out-of-bag) prediction error under a perturbed predictor variable and the original predictor variable. We do this for every tree and then average across all random trees that form a random forest (Breiman, 2001). The box-and-whisker plots we plot in Figure 15 visualize the distribution of the VIMP statistic we obtain in this way across the states. The contemporaneous realization of RV stands out in terms of the VIMP statistic at all four forecast horizons. Two other relatively important predictor variables are the state-level and market-wide leverage effects, mainly at the short and intermediate forecast horizons. The growth rates of electricity sales appear to gain somewhat in importance at the long forecast horizon, especially as the growth rates of commercial electricity sales and industrial electricity sales are concerned.

4.6 Comparing the Algorithms and a Synthesis

Our results are based on three different statistical learning algorithms. The boosting algorithm is our main algorithm, but we also have presented results for a forward best predictor selection algorithm and random forests. The results for these three algorithm in part reinforce each other, and in other parts the algorithms produce diverging results. An important question, therefore, is whether we can somehow rank the three algorithms. In the context of our forecasting experiments, such a ranking clearly must be based on the relative out-of-sample predictive performance of the three statistical learning algorithms.

In order to develop such a ranking, we use the MAFE-FG and RMSFE-FG statistics. Specifically, we choose the benchmark forecasting model from one algorithm, and the very same rival forecasting model from another algorithm. In this way, we can compare the performance of the boosting algorithm as applied to estimate, for example, the RV-MM model with the performance of one of the other two algorithms, also applied to estimate the RV-MM model, where we evaluate the predictive performance of both algorithms by means of either the MAFE or der RMSFE statistic. We setup such a comparison for every combination of two of the three algorithms, and for all states. We then estimate the cross-state regression model for the MAFE-FG the and RMSFE-FG statistics to compare any two algorithms.

The cross-state regression model that we use to compare our algorithms differs in one respect from the cross-state regression model we have laid out in Section 3.3 to compare forecasting models, which is based on a one-sided test based on the alternative hypothesis that the rival forecasting model has a better predictive performance than the benchmark forecasting model. A one-sided alternative hypothesis makes perfect sense in case we are interested in whether, for example, the RV-MM-E forecasting model outperforms, on average across the states, the RV-MM forecasting model. We now, however, are interested in the question whether the predictive performance of a given forecasting model differs across two statistical learning algorithms, and, if so, which statistical learning algorithm yields the better performance for a given forecasting model. For this reason, we modify the alternative hypothesis that we use to study the results of estimating the cross-state regression model. Specifically, we now test the null

hypothesis of equal predictive performance of two algorithms, given a forecasting model, against the alternative that the two algorithms under scrutiny differ with respect to their predictive performance (two-sided test). We, therefore, report in Table 6 the estimated intercept coefficient (rather than its p-value), because sign of the estimated coefficient informs which algorithm performs better on average across states. As for the interpretation of the estimated intercept coefficient, it is useful to remember that the MAFE-FG and RMSFE-FG statistics are expressed in percent.

– Table 6 about here. –

When we compare the boosting algorithm with random forests, we find that these two statistical learning algorithms do not deliver a statistically different predictive performance at the short forecasting horizon, while the boosting algorithm performs better than random forests in terms of out-of-sample forecasting gains for the intermediate and long forecast horizons. Only for the RV-M model is the predictive performance of the two algorithms not statistically significantly different. The boosting algorithm also outperforms the forward best predictor selection algorithm, irrespective of whether we study the MAFE-FG or the RMSFE-FG statistic, and at all four forecasting horizons. Finally, random forests perform better than the forward best predictor selection algorithm at the short and intermediate forecast horizons. The differences between these two statistical learning algorithms get smaller, and in three cases statistically insignificant, at the long forecast horizon.

– Table 7 about here. –

Against the background of the results that we summarize in Table 6, it is interesting to consider a synthesis of the three algorithms. To this end, we use a simple model-averaging approach. Specifically, we form the average of the forecasts we obtain from applying the boosting algorithm (gMDL$_{actset}$), the forward best predictor selection algorithm (BIC), and random forests. We then run our cross-state regressions of forecasting gains and apply a one-sided test to test the null hypothesis that the intercept coefficient is equal to zero, against the one-sided alternative hypothesis that the intercept coefficient is positive. We summarize the results in Table 7. The results indicate that we can reject the null hypothesis for the RV forecasting model in a comparison with the RV-M forecasting model, while the test results for $h = 1, 3$ are significant when we compare the RV-M forecasting model with the RV-MM forecasting model. The test results for the comparison of the RV-MM forecasting model with the RV-MM-E forecasting model and for the comparison of the RV-MM-E forecasting model with the RV-MM-EM forecasting model are significant at the short forecast horizon ($h = 1$), lending support to the notion that the growth rates of electricity sales contribute to forecasting gains at the short forecast horizon, but not beyond.

4.7 Results for Utility Forecasting Gains

In order to shed light on the utility gains that a forecaster can derive from applying the competing forecasting models, we summarize in Table 8 results of the U-FG cross-state regressions.[10] The results show that we can reject the

[10]We estimate the regression models as a least trimmed squares robust regression model to account for occasional outliers, where we fix the percentage of squared residuals whose sum is to be minimized to 90%. We use the R add-on package "robustbase" (version 0.99-6). See

one-sided null hypothesis when we compare the RV with the RV-M forecasting model for all three algorithms for $h = 1, 3$, and for random forests (the boosting algorithm) also for $h = 6$ ($h = 6, 12$). Moreover, we reject the null hypothesis for the short and the two intermediate forecast horizons when we compare the RV-M forecasting model with the RV-MM forecasting model. Hence, realized state-level and realized market-wide moments clearly yield utility forecasting gains when the forecast horizon is not too long. For the pair RV-MM versus RV-MM-E, the test results are significant for the boosting algorithm at $h = 3, 6$, and for random forests for $h = 1$. For the pair RV-MM-E versus RV-MM-EM, the test results are significant for the boosting algorithm at $h = 3, 6$ when we consider the gMDL$_{trace}$ model-selection criterion, and at $h = 3$ when we consider the gMDL$_{actset}$ model-selection criterion. For random forests, we observe test results for the pair RV-MM versus RV-MM-E at $h = 1$, and for the pair RV-MM-E versus RV-MM-EM at $h = 1, 3$. The test results are all insignificant for the pairs of forecasting models involving the growth rates of electricity sales when we consider the best forward predictor selection algorithm. Taken together, we find some evidence that the growth rates of electricity sales may yield utility forecasting gains at the short and/or intermediate forecast horizons, but clearly not at the long forecast horizon.

– Tables 8 and 9 about here. –

When we use a model-averaging approach to estimate the U-FG cross-state regressions, we obtain the results that we summarize in Table 9. The test results are statistically significant at all four forecast horizons for the pair RV versus

Maechler et al. (2025).

RV-M, and for the short and intermediate forecast horizons for the pair RV-M versus RV-MM. For the pair RV-MM versus RV-MM-E all test results are statistically insignificant, while for the pair RV-MM-E versus RV-MM-EM only the test result at $h = 3$ is statistically insignificant. Hence, we find some weak evidence that using the growth rates of electricity sales yields utility gains at an intermediate forecast horizon, but the test results that we obtain when we setup the U-FG cross-state regressions by means of model-averaging approach are weaker than the test results that we report in Tables 8 for the scenario in which we analyze the utility gains separately for the three statistical learning algorithms.

5 Summary and Concluding Remarks

We have used a boosting algorithm as well as a forward best predictor selection algorithm and random forests to study whether the growth rates of electricity sales have out-of-sample predictive value for the subsequently realized state-level volatility of stock returns, where we have controlled for state-level and market-wide realized moments. The broad picture that arises from our results is that, in those configurations of our forecasting experiment for which we find evidence of forecasting gains from extending a forecasting model to include the growth rates of electricity sales, measured at the level of an individual state or at the level of the market, such forecasting gains mainly are concentrated, in the cross-section of states, at the short forecast horizon. Depending on the statistical learning algorithm being used, we find some evidence that utility forecasting gains may also arise at intermediate forecast horizons. On balance, however, the

evidence is mixed. We find stronger evidence that the predictive value of realized moments, either measured at the level of an individual state or at the level of the market, and utility forecasting gains, extend, in some model configurations, to the intermediate and even to the long forecast horizon.

In terms of variable importance, we have found that the contemporaneous realized state-level volatility plays a dominant role, followed by the market-wide leverage effect. We further have found that the importance of the growth rates of electricity sales as well as the inclusion of these predictors in the forecasting models tend to strengthen on average across states as the length of the forecast horizon increases, especially as far as the growth rates of commercial and industrial electricity sales are concerned. The details of the results, however, vary somewhat across the statistical learning algorithms. For the boosting algorithm, we have also identified, and visualized by means of wordclouds, the states for which the growth rates of electricity sales are relatively often included in the boosted forecasting model.

Moreover, the results of our cross-state regression models have shown that the boosting algorithm and random forests perform equally well at the short forecast horizon, but the boosting algorithm yields a better performance on average, except for the RV-M forecasting model, at the intermediate and long forecast horizons. Both the boosting algorithm and random forests perform better for several configurations of the forecasting experiment than the forward best predictor selection algorithm, where the performance of the best predictor selection algorithm relative to random forests tends to improve as the length of the forecast horizon increases. Results of a simple model-averaging approach that

combines the forecasts from the three statistical learning algorithms have lent further support to the notion that the growth rates of electricity sales contribute to forecasting gains at the short forecast horizon.

The evidence that electricity sales may have some predictive value at a short forecast horizon in the cross-section of states indicates that, in future research, is interesting to study whether the growth rates of electricity sales help to predict the cross-sectional (that is, cross-state) realized volatility of stock returns. Results of such research could be useful for improving portfolio diversification strategies in that the mixture of stocks from firms residing in high-volatility and low-volatility states could be optimized to improve the risk-returns profile of a portfolio or to better control its risk profile.

For future research, expanding our study to a cross-country analysis involving both advanced and emerging stock markets, for example, within Europe and otherwise, contingent on the availability of electricity sales data, would be invaluable to generalize our findings.

References

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? Journal of Financial Economics, 118(1), 135–167.

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.

Apergis, N., and Payne, J.E. (2014). An exploratory analysis of electricity consumption and stock prices: evidence from OECD countries. Energy Sources, Part B: Economics, Planning, and Policy, 9(1), 70–78.

Baumeister, C., Leiva-León, D., and Sims, E. (2024). Tracking weekly state-level economic conditions. Review of Economics and Statistics, 106(2), 483–504.

Bernanke, B.S. (1983). Irreversibility, uncertainty, and cyclical investment. Quarterly Journal of Economics, 98(1), 85–106.

Bollerslev, T., Hood, B., Huss, J., and Pedersen, L.H. (2018). Risk everywhere: Modeling and managing volatility. Review of Financial Studies, 31(7), 2729–2773.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2022). Forecasting realized volatility of international REITs: The role of realized skewness and realized kurtosis. Journal of Forecasting, 41(2), 303–315.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2023a). 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. (2023b). Climate risks and state-level stock market realized volatility. Journal of Financial Markets, 66(C), 100854.

Bonato, M., Demirer, R., and Gupta, R. (2018). The predictive power of industrial electricity usage revisited: evidence from non-parametric causality tests. OPEC Energy Review, 42(2), 93–106.

Bonato, M., Gupta, R., Pierdzioch, C. (Forthcoming). Do Shortages Forecast Aggregate and Sectoral U.S. Stock Market Realized Variance? Evidence from a Century of Data. Journal of Empirical Finance.

Bonato, M., Gupta, R., Pierdzioch, C., and Polat, O. (2025). ESG uncertainty and forecasting realized volatility of gold returns: A boosting approach. In Bourghelle, D. Grandin, P., Jawadi, F., and Rozuin, P. (eds), Financial Market Dynamics: Essays in Honor of Pascal Alphonse (forthcoming).

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Breiman, L. Friedman, J., Olshen, R.A., and Stone, C.J. (1984). Classification and regression trees. Wadsworth, Belmont, California, USA.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. Annals of Statistics, 34(2), 559–583.

Bühlmann, P., and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting (with discussion). Statistical Science, 22(4), 477–505.

Burnside, A.C., Eichenbaum, M.S., and Rebelo, S.T. (1995). Capacity utilization and returns to scale. NBER Macroeconomics Annual, 10, 67–110.

Burnside, A.C., Eichenbaum, M.S., and Rebelo, S.T. (1996). Sectoral Solow residuals. European Economic Review, 40(3–5), 861–869.

Candila, V., Cepni, O., Gallo, G.M., and Gupta, R. (Forthcoming). Combination of GARCH-MIDAS forecasts of US state-level volatilities: The role of local and global economic policy uncertainty. International Journal of Forecasting.

Chaney, T., Sraer, D., and Thesmar, D. (2012). The collateral channel: how real estate shocks affect corporate investment. American Economic Review, 102(6), 2381–2409.

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.

Comin, D., and Gertler, M. (2006). Medium-term business cycles. American Economic Review, 96(3), 523–551.

Coval, J.D., and Moskowitz, T.J. (1999). Home bias at home: local equity preference in domestic portfolios. Journal of Finance, 54(6), 2045–2073.

Coval, J.D., and Moskowitz, T.J. (2001). The geography of investment: Informed trading and asset prices. Journal of Political Economy, 199(4), 811–841.

Da, Z., Huang, D., and Yun, H. (2017). Industrial electricity usage and stock returns. Journal of Financial and Quantitative Analysis, 52(1), 37–69.

Doruk, Ö. T. (2024). The link between electricity consumption and stock market during the pandemic in Türkiye: a novel high-frequency approach. Environmental Science and Pollution Research, 31(11), 17311–17323.

Friedman, J., Tibshirani, R., and Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1), 1–22.

Giglio, S., Kelly, B., and Stroebel, J. (2021). Climate finance. Annual Review of Financial Economics, 13, 15–36.

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

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. Annals of Statistics, 3(5), 1163–1173.

Hofner. B., Mayr. A., Robinzonov, N., and Schmid, M. (2014). Model-based boosting in R: A hands-on tutorial using the R package mboost. Computational Statistics, 29(1), 3–35.

Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M., and Hofner, B. (2010). Model-based boosting 2.0. Journal of Machine Learning Research, 11(71), 2109–2113.

Ishwaran H., and Kogalur, U. B. (2025). Fast unified random forests for survival, regression, and classification (RF-SRC). R package version 3.3.3. URL: https://cran.r-project.org/web/packages/randomForestSRC.

Jorgenson, D.W., and Griliches, Z. (1967). The explanation of productivity change. Review of Economic Studies, 34(3), 249–283.

King, R.G., and Rebello, S.T. (2000). Resuscitating real business cycles. In: Handbook of Macroeconomics, J. Taylor and M. Woodford, eds. Vol 1. (Part B), Amsterdam, Netherlands: North-Holland, 927–1007.

Korniotis, G.M., and Kumar, A. (2013). State-level business cycles and local return predictability. Journal of Finance, 68(3), 1037–1096.

Lang D., and Chien, G. (2018). wordcloud2: Create word cloud by 'htmlwidget'. R package version 0.2.1, URL: https://CRAN.R-project.org/package=wordcloud2.

Lu, F., Ma, F., and Bouri, E. (2024). Stock market volatility predictability: New evidence from energy consumption. Nature: Humanities and Social Sciences Communication, 11, Article No. 1624.

Ludvigson, S.C., Ma, S., and Ng, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? American Economic Journal: Macroeconomics, 13(4), 369–410.

Lumley, T., based on Fortran code by A. Miller (2024). leaps: Regression subset selection. R package version 3.2. URL: https://CRAN.R-project.org/package=leaps.

Maechler, M.Rousseeuw, P., Croux, C. Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T. Koller, M., c("Eduardo", "L. T.") Conceicao, and di Palma, M.A. (2025). robustbase: Basic robust statistics. R package version 0.99-6. URL: http://CRAN.R-project.org/package=robustbase.

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, 153–159.

Payne, J.E. (2010). A survey of the electricity consumption-growth literature. Applied Energy, 87(3), 723–731.

Pirinsky, C., and Wang, Q. (2006). Does corporate headquarters location matter for stock returns? Journal of Finance, 61(4), 1991–2015.

Pirlogea, C., and Cicea, C. (2012). Econometric perspective of the energy consumption and economic growth relation in European Union. Renewable and Sustainable Energy Reviews, 16(8), 5718–5726.

Poon, S-H., and Granger, C.W.J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 47–539.

R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: 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, 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., Gupta, R., and Ogbonna, A.E. (2022). A moving average heterogeneous autoregressive model for forecasting the realized volatility of the US stock market: Evidence from over a century of data. International Journal of Finance and Economics, 27(1), 384–400.

Salisu, A.A., Gupta, R., Cepni, O., and Caraiani, P. (2024). Oil shocks and state-level stock market volatility of the United States: a GARCH-MIDAS approach. Review of Quantitative Finance and Accounting, 63(4), 1473–1510.

Salisu, A.A., Liao, W., Gupta, R., and Cepni, O. (2025). Economic conditions and predictability of US stock returns volatility: Local factor versus national factor in a GARCH-MIDAS Model. Journal of Forecasting, 44(4), 1441–1466.

Salisu, A.A., Gupta, R., and Cepni, O. (forthcoming b). Housing market variables and predictability of state-level stock market volatility of the United States: Evidence from a GARCH-MIDAS approach. Quarterly Review of Economics and Finance.

Salisu, A.A., Ogbonna, A.E., Gupta, R., and Cepni, O. (forthcoming a). Energy market uncertainties and US state-level stock market volatility: A GARCH-MIDAS approach. Financial Innovation.

Schwert, G.W. (1989). Why does stock market volatility change over time? Journal of Finance, 44(5), 1115–1153.

Segnon, M., Gupta, R., and Wilfling, B. (2023). Forecasting stock market volatility with regime switching GARCH-MIDAS: The role of geopolitical risks. International Journal of Forecasting, 40(1), 29−43.

Shahbaz, M., Sarwar, S., Chen, W., and Malik, M.N. (2017). Dynamics of electricity consumption, oil price and economic growth: Global perspective. Energy Policy, 108(C), 256−270.

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.

Somani, D., Gupta, R., Karamakar, S., and Plakandaras, V. (Forthcoming). Supply bottlenecks and machine learning forecasting of international stock market volatility. Finance Research Letters.

Zeileis, A., K'oll, S., Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. Journal of Statistical Software, 95(1), 1−36. URL:

Zeileis, A. (2004), Econometric computing with HC and HAC covariance matrix estimators. Journal of Statistical Software, 11(10), 1−17.

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, 105614.

Table 1: Boosting algorithm and optimal model selection criterion

Model	RV-M	RV-MM	RV-MM-E	RV-MM-EM	RV-M	RV-MM	RV-MM-E	RV-MM-EM
			MAFE				RMSFE	
			$h = 1$				$h = 1$	
AIC_{trace}	7	5	3	5	6	3	4	2
AIC_{actset}	5	5	3	6	1	3	5	7
gMDL_{trace}	18	18	21	16	24	24	13	13
gMDL_{actset}	20	22	23	23	19	20	28	28
			MAFE				RMSFE	
			$h = 3$				$h = 3$	
AIC_{trace}	7	7	5	4	2	2	2	2
AIC_{actset}	7	8	4	3	4	4	2	2
gMDL_{trace}	16	9	13	11	20	11	6	7
gMDL_{actset}	20	26	28	32	24	33	40	39
			MAFE				RMSFE	
			$h = 6$				$h = 6$	
AIC_{trace}	6	6	4	2	3	2	1	2
AIC_{actset}	5	7	4	3	1	1	2	0
gMDL_{trace}	10	10	6	10	10	10	7	9
gMDL_{actset}	29	27	36	35	36	37	40	39
			MAFE				RMSFE	
			$h = 6$				$h = 6$	
AIC_{trace}	8	11	9	6	4	3	5	2
AIC_{actset}	9	5	4	2	4	3	1	1
gMDL_{trace}	16	13	13	12	13	11	7	8
gMDL_{actset}	21	22	24	30	34	34	37	39

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. Using data for every state, the boosting algorithm is implemented for every forecasting model by using the four different model-selection criteria. For every state, every forecasting model, and every forecast horizon, h, the out-of-sample MAFE (RMSFE) statistic is computed and the best model-selection criterion is identified as the one that minimizes the MAFE (RMSFE) statistic. The numbers summarized in the table represent the number of states for which a model-selection criteria yields the best model in terms of the MAFE (RMSFE) statistic. The numbers in the columns of the table, for a given forecast horizon, can exceed the total number of states (50) in case two or more model-selection criteria minimize the MAFE (RMSFE) statistic.

Table 2: Boosting and cross-state regressions of forecasting gains

Panel A: MAFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
		gMDL$_{trace}$		
RV vs. RV-M	0.0518	0.0000	0.0000	0.0001
RV-M vs. RV-MM	0.0003	0.0000	0.0121	0.5100
RV-MM vs. RV-MM-E	0.1262	0.9081	1.0000	0.9911
RV-MM-E vs. RV-MM-EM	0.4593	1.0000	1.0000	1.0000
		gMDL$_{actset}$		
RV vs. RV-M	0.1058	0.0001	0.0000	0.0001
RV-M vs. RV-MM	0.0153	0.0000	0.1060	0.6049
RV-MM vs. RV-MM-E	0.0373	0.9247	0.6889	0.9427
RV-MM-E vs. RV-MM-EM	0.0244	0.9993	1.0000	1.0000

Panel B: RMSFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
		gMDL$_{trace}$		
RV vs. RV-M	0.3401	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0002	0.0000	0.0030	0.0277
RV-MM vs. RV-MM-E	0.0080	0.2937	0.9675	0.9308
RV-MM-E vs. RV-MM-EM	0.0051	0.9991	1.0000	1.0000
		gMDL$_{actset}$		
RV vs. RV-M	0.4937	0.0001	0.0000	0.0000
RV-M vs. RV-MM	0.0029	0.0000	0.0009	0.0160
RV-MM vs. RV-MM-E	0.0126	0.2942	0.2978	0.7514
RV-MM-E vs. RV-MM-EM	0.0222	0.8991	1.0000	1.0000

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive. The p-value for the intercept term is obtained using heteroscedasticity and autocorrelation consistent standard errors.

Table 3: Boosting, a rolling-estimation window, and cross-state regressions of forecasting gains

Panel A: MAFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
		gMDL$_{trace}$		
RV vs. RV-M	0.0575	0.0000	0.0000	0.0393
RV-M vs. RV-MM	0.0011	0.0000	0.2591	0.8251
RV-MM vs. RV-MM-E	0.0754	0.9773	1.0000	1.0000
RV-MM-E vs. RV-MM-EM	0.6059	1.0000	1.0000	1.0000
		gMDL$_{actset}$		
RV vs. RV-M	0.3685	0.0000	0.0000	0.2751
RV-M vs. RV-MM	0.0000	0.0000	0.3188	0.6785
RV-MM vs. RV-MM-E	0.1450	0.7989	0.9874	0.9652
RV-MM-E vs. RV-MM-EM	0.6983	1.0000	1.0000	0.9997

Panel B: RMSFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
		gMDL$_{trace}$		
RV vs. RV-M	0.4253	0.000	0.0000	0.0000
RV-M vs. RV-MM	0.0000	0.000	0.1954	0.3045
RV-MM vs. RV-MM-E	0.0180	0.706	0.9951	1.0000
RV-MM-E vs. RV-MM-EM	0.0522	1.000	1.0000	1.0000
		gMDL$_{actset}$		
RV vs. RV-M	0.7151	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0000	0.0000	0.3410	0.0621
RV-MM vs. RV-MM-E	0.0677	0.1594	0.7426	0.7978
RV-MM-E vs. RV-MM-EM	0.2046	0.9985	1.0000	0.9576

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive. The p-value for the intercept term is obtained using heteroscedasticity and autocorrelation consistent standard errors.

Table 4: Forward best predictor selection algorithm and cross-state regressions of forecasting gains

Panel A: MAFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
			BIC	
RV vs. RV-M	0.9882	0.4280	0.5471	0.4756
RV-M vs. RV-MM	0.0005	0.0039	0.4992	0.6099
RV-MM vs. RV-MM-E	0.4176	0.9408	1.0000	0.9944
RV-MM-E vs. RV-MM-EM	0.9920	1.0000	1.0000	1.0000

Panel B: RMSFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
			BIC	
RV vs. RV-M	0.9944	0.9630	0.9671	0.8434
RV-M vs. RV-MM	0.0000	0.0525	0.9220	0.1095
RV-MM vs. RV-MM-E	0.4260	0.9760	1.0000	0.9865
RV-MM-E vs. RV-MM-EM	0.9653	1.0000	1.0000	1.0000

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive. The p-value for the intercept term is obtained using heteroscedasticity and autocorrelation consistent standard errors.

Table 5: Random forests and cross-state regressions of forecasting gains

Panel A: MAFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
RV vs. RV-M	0.0173	0.0064	0.0031	0.4306
RV-M vs. RV-MM	0.8789	0.9530	0.9726	1.0000
RV-MM vs. RV-MM-E	0.0017	0.3618	0.9705	0.8821
RV-MM-E vs. RV-MM-EM	0.0000	0.9999	1.0000	0.6599

Panel B: RMSFE-FG regression model

Models	$h=1$	$h=3$	$h=6$	$h=12$
RV vs. RV-M	0.0186	0.0040	0.0002	0.0461
RV-M vs. RV-MM	0.0429	0.4706	0.8021	1.0000
RV-MM vs. RV-MM-E	0.0076	0.3772	0.7460	0.7206
RV-MM-E vs. RV-MM-EM	0.0012	0.1029	1.0000	0.9577

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive. The p-value for the intercept term is obtained using heteroscedasticity and autocorrelation consistent standard errors.

Table 6: Comparing statistical learning algorithms in terms of forecasting gains

Panel A: MAFE-FG regression model

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
		Boosting vs. random forests		
RV-M	0.7962	0.1680	0.1374	-0.9627
RV-MM	-0.2833	-1.2815***	-0.6399	-2.6738***
RV-MM-E	0.0690	-0.9753**	-1.3523**	-3.1344***
RV-MM-EM	0.5758	-1.1846**	-1.3319**	-1.8898**
		Best predictor selection vs. boosting		
RV-M	1.9361***	1.3637***	1.7013***	1.1425***
RV-MM	1.7669***	2.0603***	1.9160***	1.1240***
RV-MM-E	1.9593***	2.1693***	2.9487***	1.7222***
RV-MM-EM	2.2421***	3.1372***	5.6231***	3.5816***
		Best predictor selection vs. random forests		
RV-M	2.7634***	1.5125***	1.8160***	0.1019
RV-MM	1.4842**	0.7356	1.2245***	-1.6284***
RV-MM-E	2.0421***	1.1506**	1.4701***	-1.5302**
RV-MM-EM	2.8469***	1.8945***	4.1343***	1.5272**

Panel B: RMSFE-FG regression model

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
		Boosting vs. random forests		
RV-M	1.1503	0.3746	-0.3993	-1.3846*
RV-MM	0.0295	-1.3372***	-1.2024***	-3.0056***
RV-MM-E	0.0884	-1.2798***	-1.5642***	-3.1846***
RV-MM-EM	0.0838	-1.0077**	-1.3793**	-2.7236***
		Best predictor selection vs. boosting		
RV-M	2.8981***	2.4888***	4.0039***	3.2037***
RV-MM	3.0664***	3.9543***	4.7356***	3.4184***
RV-MM-E	3.3058***	4.2110***	5.8312***	4.1619***
RV-MM-EM	3.6256***	4.7595***	8.3882***	6.2256***
		Best predictor selection vs. random forests		
RV-M	4.3299**	2.8839***	3.5769***	1.7193**
RV-MM	3.3003**	2.5895***	3.4610***	0.2535
RV-MM-E	3.5941**	2.9058***	4.1534***	0.7688
RV-MM-EM	3.9139**	3.7261***	6.8823***	3.2376***

The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the two-sided alternative hypothesis is that the intercept term is different from zero. The gMDL$_{actset}$ is used to implement the boosting algorithm. The BIC model-selection criterion is used to implement the forward best predictor selection algorithm. *** (**, *) denotes statistical significance at the 1% (5%, 10%) level, where inference is based on heteroscedasticity and autocorrelation consistent standard errors.

Table 7: Model averaging and cross-state regressions of forecasting gains

Panel A: MAFE-FG regression model

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0003	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0376	0.0002	0.7803	1.0000
RV-MM vs. RV-MM-E	0.0001	0.6410	0.9982	0.9549
RV-MM-E vs. RV-MM-EM	0.0000	1.0000	1.0000	1.0000

Panel B: RMSFE-FG regression model

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0206	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0000	0.0000	0.2849	0.8534
RV-MM vs. RV-MM-E	0.0000	0.2948	0.9255	0.8891
RV-MM-E vs. RV-MM-EM	0.0006	0.9552	1.0000	1.0000

The forecasting models are estimated on a recursively expanding estimation window by means of the boosting algorithm (gMDL$_{actset}$), the forward best predictor selection algorithm (BIC), and random forests, and the forecasts are then averaged across the three algorithms. The initial training period covers the first 50% of the data. The MAFE-FG and RMSFE-FG statistics are computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive. The p-value for the intercept term is obtained using heteroscedasticity and autocorrelation consistent standard errors.

Table 8: Cross-state regressions of utility gains

Panel A: Boosting (gMDL$_{trace}$)

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0000	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0000	0.0000	0.0000	0.0636
RV-MM vs. RV-MM-E	0.9721	0.0299	0.0154	0.9965
RV-MM-E vs. RV-MM-EM	0.2017	0.0001	0.0062	0.9969

Panel B: Boosting (gMDL$_{actset}$)

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0000	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0000	0.0000	0.0038	0.2341
RV-MM vs. RV-MM-E	0.9894	0.0496	0.0424	0.9759
RV-MM-E vs. RV-MM-EM	0.9431	0.0274	0.7103	0.9994

Panel C: Best forward predictor selection (BIC)

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0000	0.0527	0.8314	0.9493
RV-M vs. RV-MM	0.0024	0.0012	0.0857	0.9982
RV-MM vs. RV-MM-E	0.9999	0.2633	0.9803	1.0000
RV-MM-E vs. RV-MM-EM	1.0000	0.0047	0.9976	0.9996

Panel D: Random forests

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0918	0.0012	0.0012	0.9990
RV-M vs. RV-MM	0.0269	0.0012	0.0000	0.9311
RV-MM vs. RV-MM-E	0.0865	0.9843	0.9590	0.8053
RV-MM-E vs. RV-MM-EM	0.0893	0.0017	0.2243	0.0235

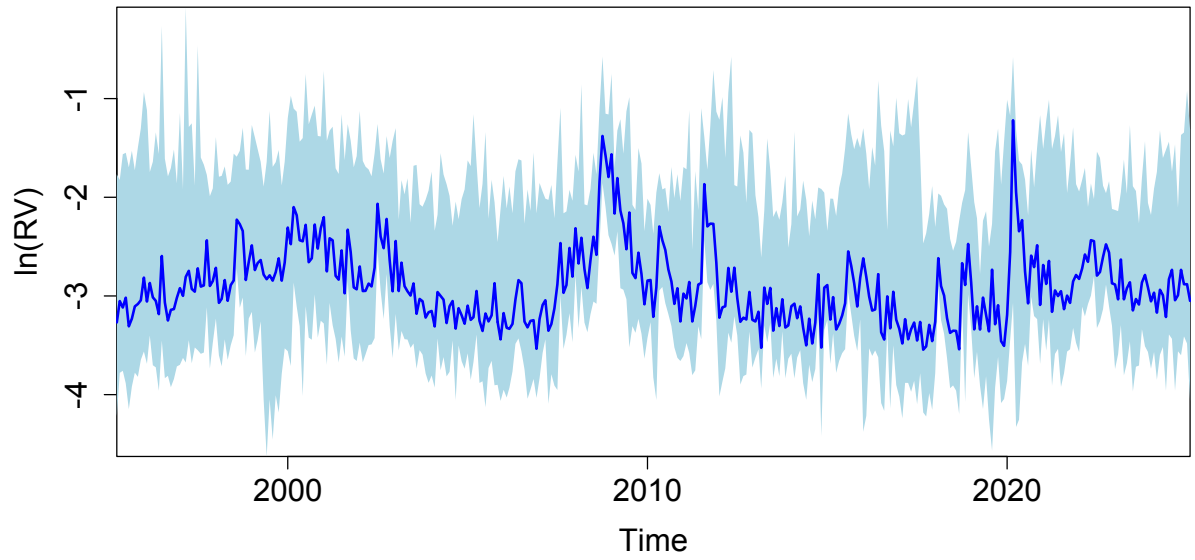
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. The U-FG statistic is computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive.

Table 9: Cross-state regressions of utility gains (model averaging)

Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RV vs. RV-M	0.0000	0.0000	0.0000	0.0000
RV-M vs. RV-MM	0.0005	0.0000	0.0000	0.7532
RV-MM vs. RV-MM-E	0.2749	0.9581	0.5781	0.9434
RV-MM-E vs. RV-MM-EM	0.5675	0.0000	0.7881	0.6925

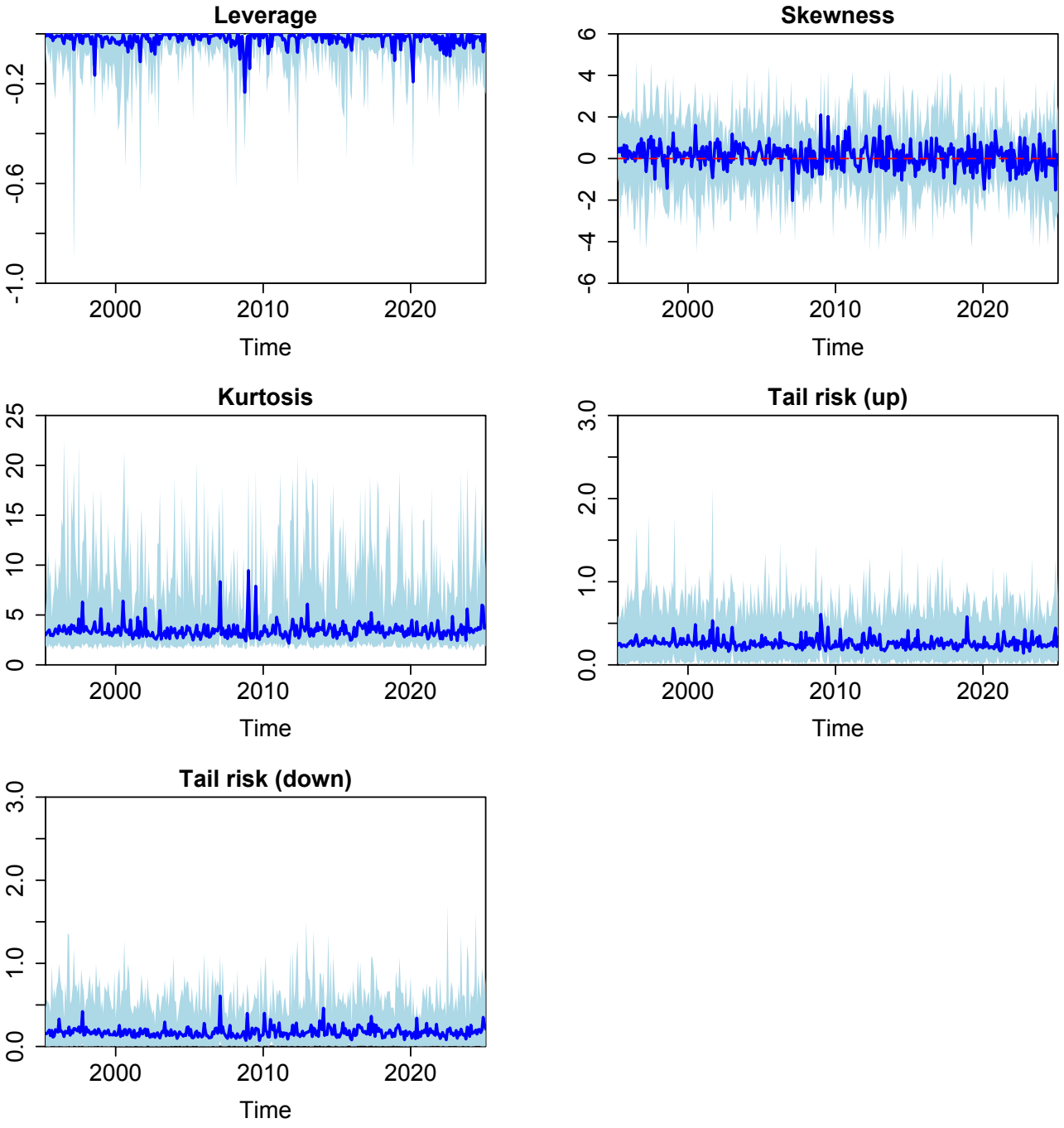
The forecasting models are estimated on a recursively expanding estimation window by means of the boosting algorithm (gMDL$_{actset}$), the forward best predictor selection algorithm (BIC), and random forests, and the forecasts are then averaged across the three algorithms. The initial training period covers the first 50% of the data. The U-FG statistic is computed for the 50 states, and then a regression model of the statistics on an intercept term is estimated. The null hypothesis is that the intercept term is zero, while the one-sided alternative hypothesis is that the intercept term is positive.

Figure 1: State-level RVs



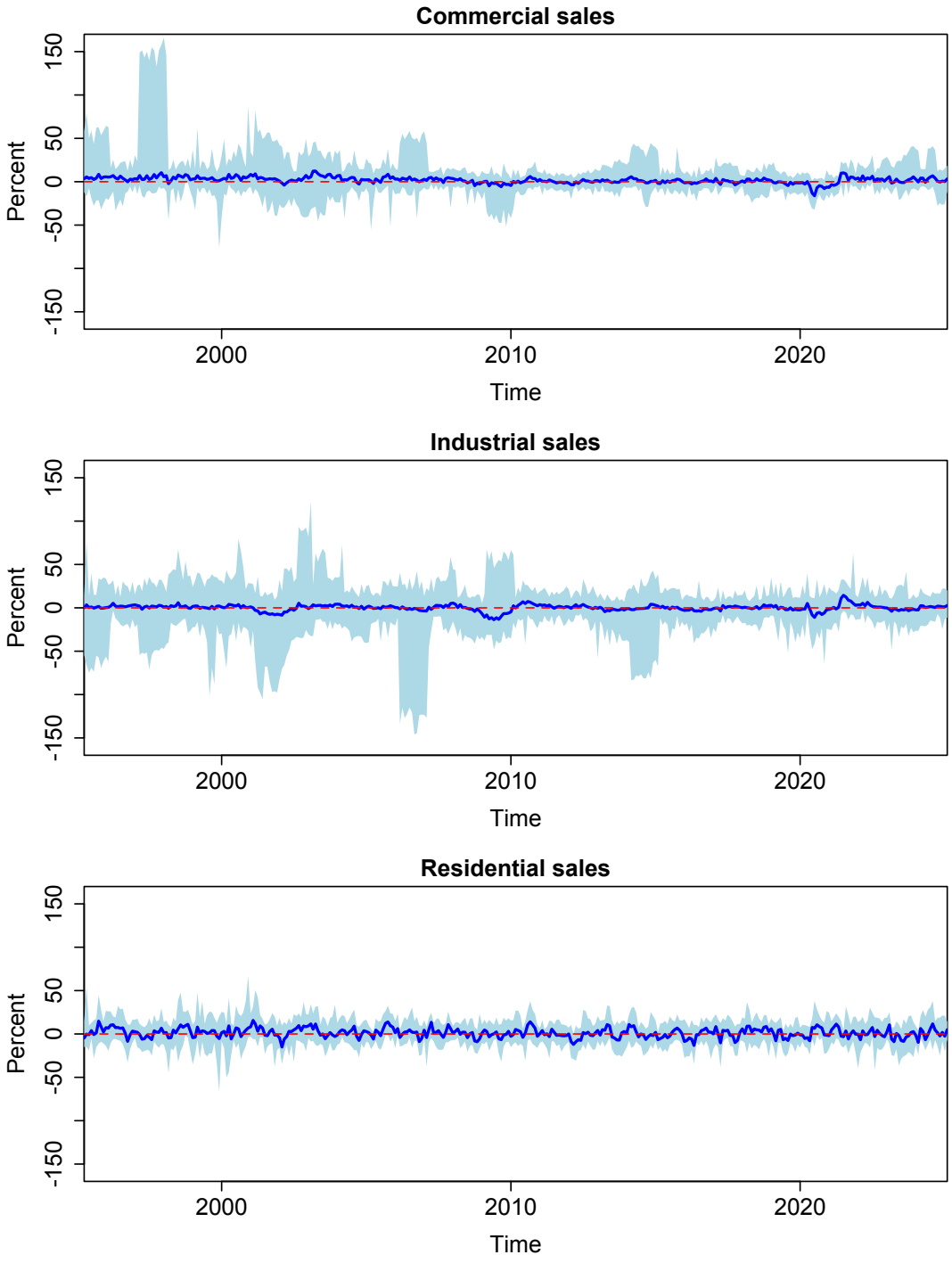
The solid line denotes the cross-state mean and the boundaries of the shaded area denote the maximum and minimum across states in each month.

Figure 2: State-level realized moments



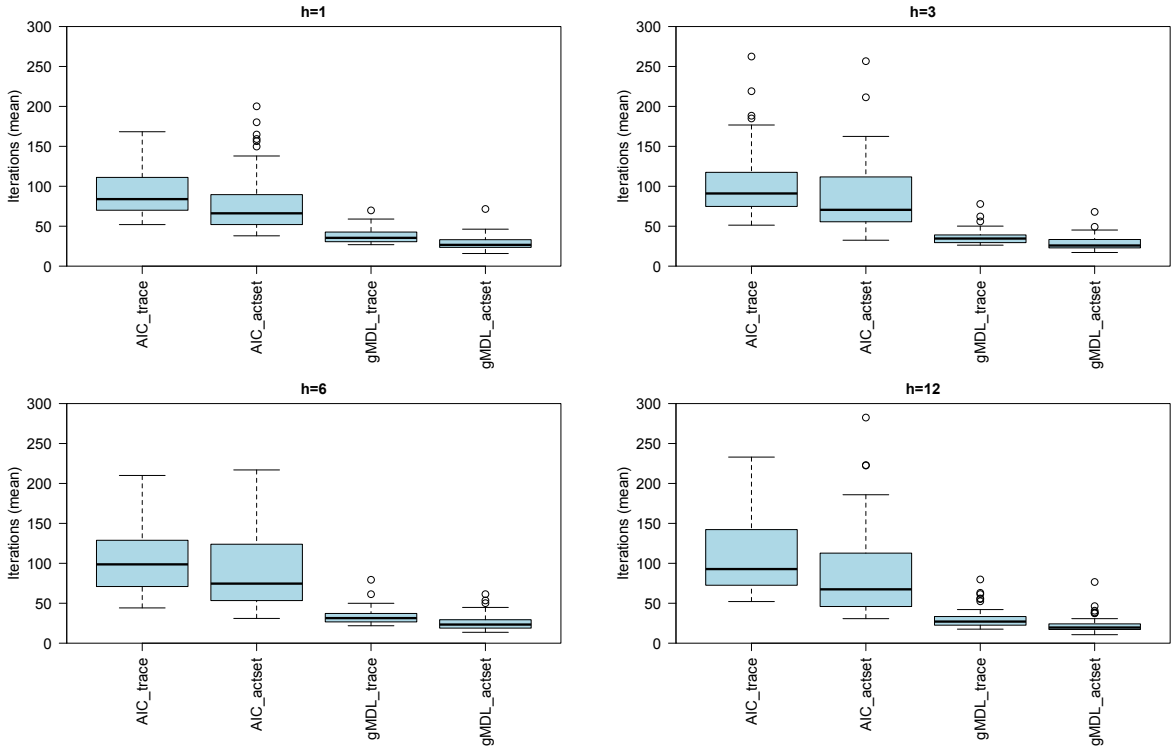
The solid line denotes the cross-state mean and the boundaries of the shaded area denote the maximum and minimum across states in each month.

Figure 3: State-level growth rates of electricity sales



The solid line denotes the cross-state mean and the boundaries of the shaded area denote the maximum and minimum across states in each month.

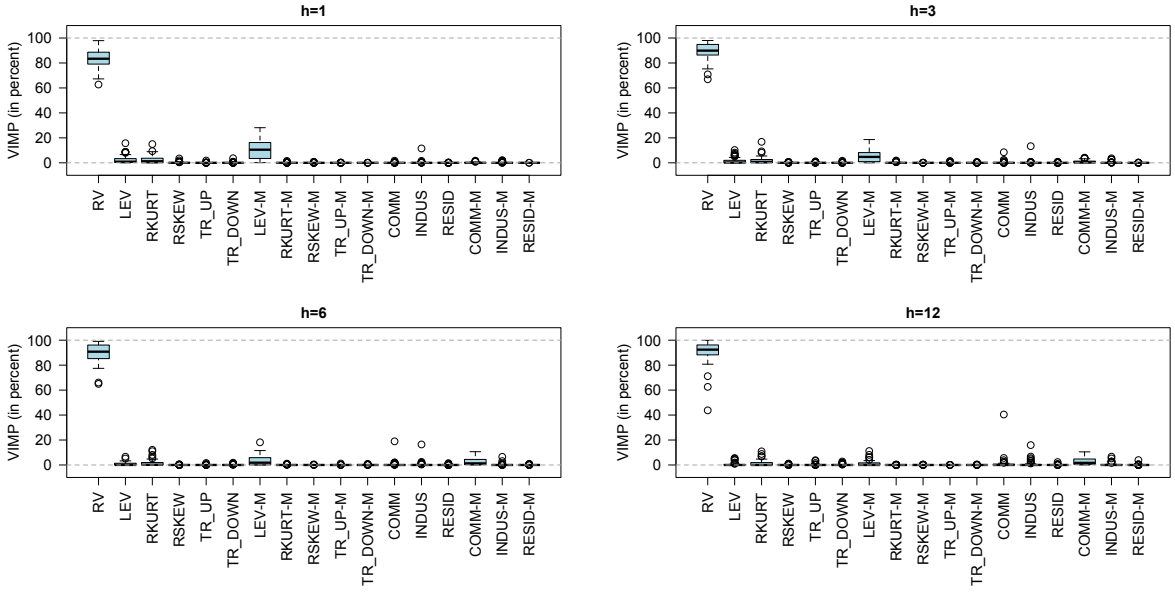
Figure 4: Boosting algorithm and optimal number of iterations (RV-MM-EM)



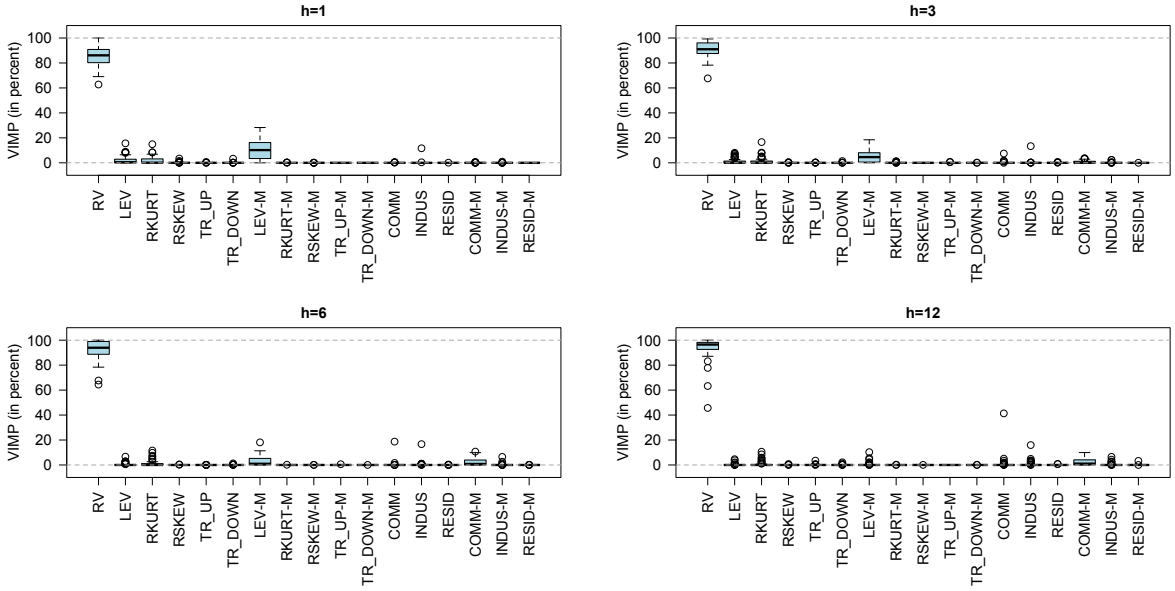
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, the optimal number of iterations are recorded for every forecasting model, every model selection criterion, and all 50 states. The box-and-whisker plots visualize the cross-state distribution of the optimal number of iterations, averaged across the out-of-sample period, where the solid horizontal line represents the median, the shaded area represents the interquartile range, the two whiskers denote the boundaries of 1.5 times the interquartile range, and circles represent points beyond the whiskers.

Figure 5: Boosting algorithm and importance of predictors (RV-MM-EM)

Panel A: gMDL$_{trace}$



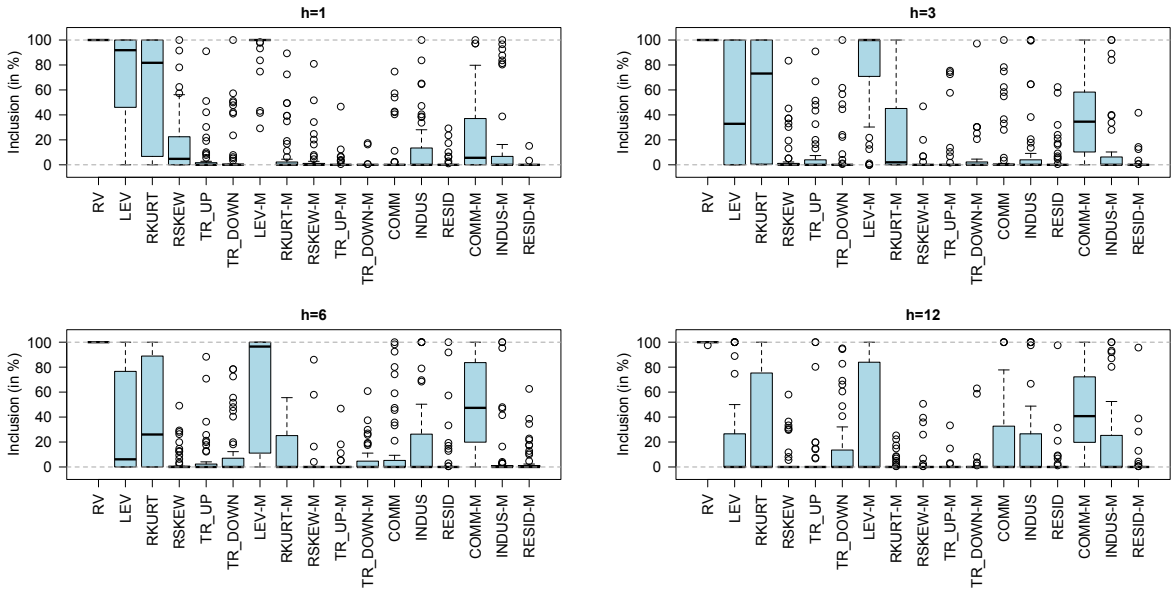
Panel B: gMDL$_{actset}$



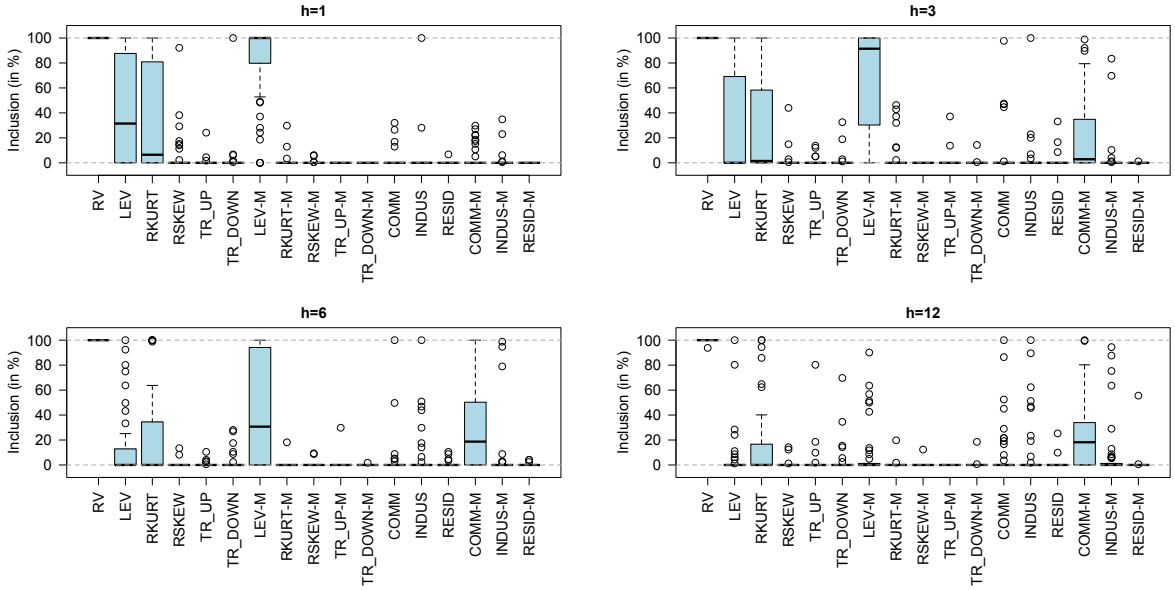
The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For every out-of-sample period, variable importance (VIMP) is recorded for the gMDL$_{trace}$ (gMDL$_{actset}$) model-selection criteria. The VIMP statistic is defined as the contribution of a base-learner to the reduction of the empirical risk function, accumulated across boosting iterations. The VIMP statistic is expressed in percent. The resulting VIMP statistic is averaged over the out-of-sample period. The box-and-whisker plots visualizes the distribution of the VIMP statistic across states, where the solid horizontal line represents the median, the shaded area represents the interquartile range, the two whiskers denote the boundaries of 1.5 times the interquartile range, and circles represent points beyond the whiskers.

Figure 6: Boosting algorithm and selection of predictors (RV-MM-EM)

Panel A: gMDL$_{trace}$



Panel B: gMDL$_{actset}$

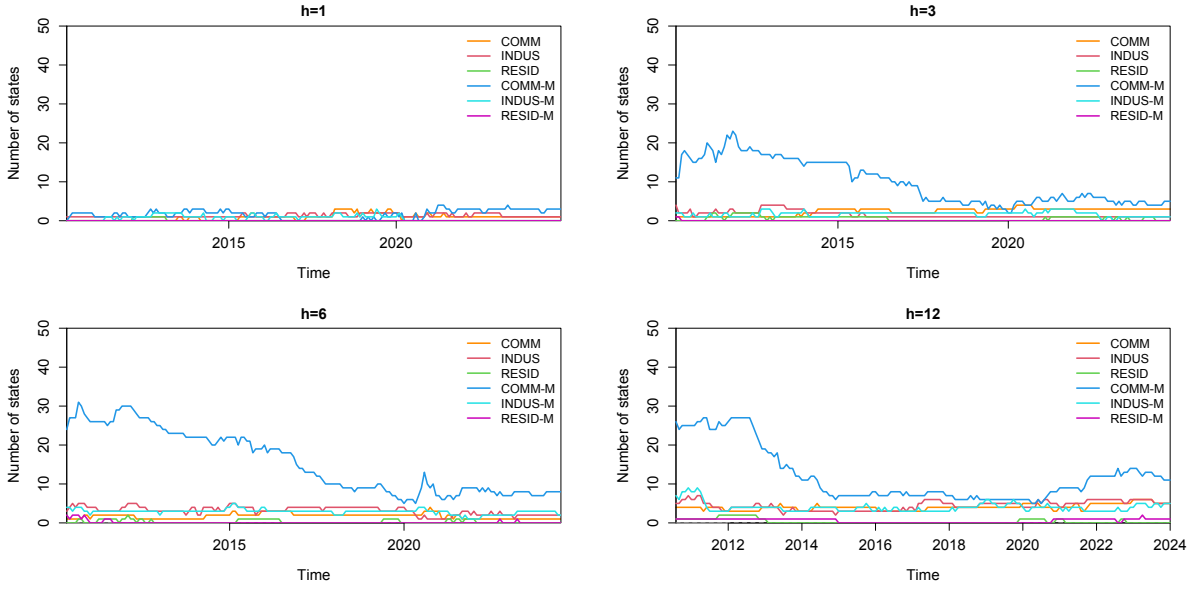


The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, the inclusion of predictors in the forecasting model is recorded for the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criterion, and every state. The box-and-whisker plots visualize how often (in percent), across all 50 states, a predictor is included in the forecasting model in the out-of-sample period, where the solid horizontal line represents the median, the shaded area represents the interquartile range, the two whiskers denote the boundaries of 1.5 times the interquartile range, and circles represent points beyond the whiskers.

Figure 7: Boosting algorithm and time paths of inclusion of the growth rate of electricity sales (RV-MM-EM)

Panel A: gMDL$_{trace}$

Panel B: gMDL$_{actset}$



The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, the inclusion of the subcategories of the growth rate of electricity sales in the forecasting model is recorded for the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criterion, and every state. It then is recorded for every state and every subcategory whether this subcategory is included in the forecasting model. Finally, the number of states for which a subcategory is included in the forecasting model is recorded and plotted over time.

Figure 8: Importance of electricity sales in the RV-MM-EM model (gMDL$_{trace}$)

Panel A: $h = 1$

Panel B: $h = 3$

Panel C: $h = 6$

Panel D: $h = 12$

The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For every state, the inclusion of the different growth rates of electricity sales in the forecasting model during the out-of-sample period is recorded under the gMDL$_{trace}$ model-selection criterion. Then the sum across categories of electricity sales is computed to capture the total frequency of inclusion of electricity sales in the boosted RV-MM-EM forecasting model for a state. The worldclouds illustrate this frequency.

Figure 9: Importance of electricity sales in the RV-MM-EM model (gMDL$_{actset}$)

Panel A: $h = 1$

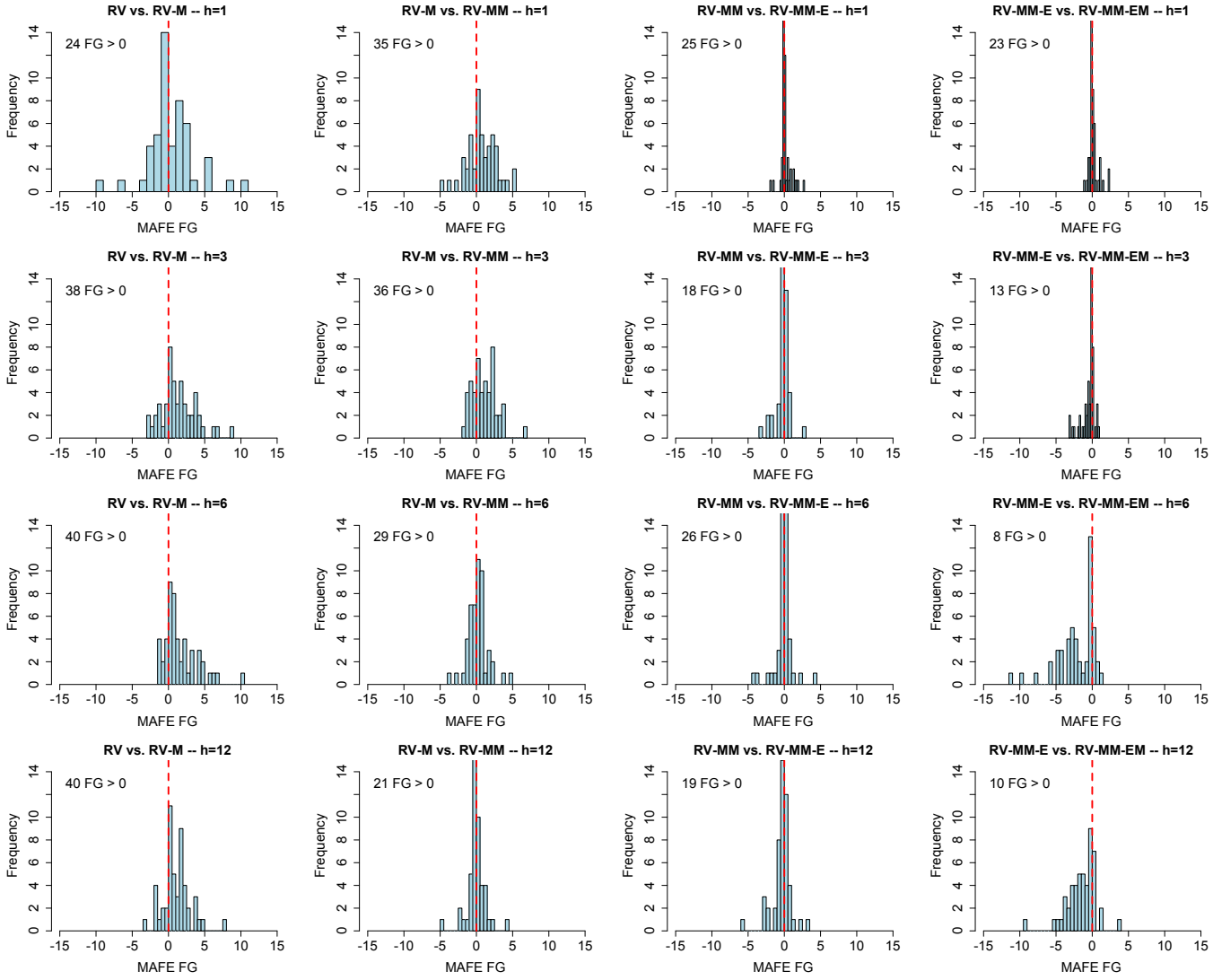
Panel B: $h = 3$

Panel C: $h = 6$

Panel D: $h = 12$

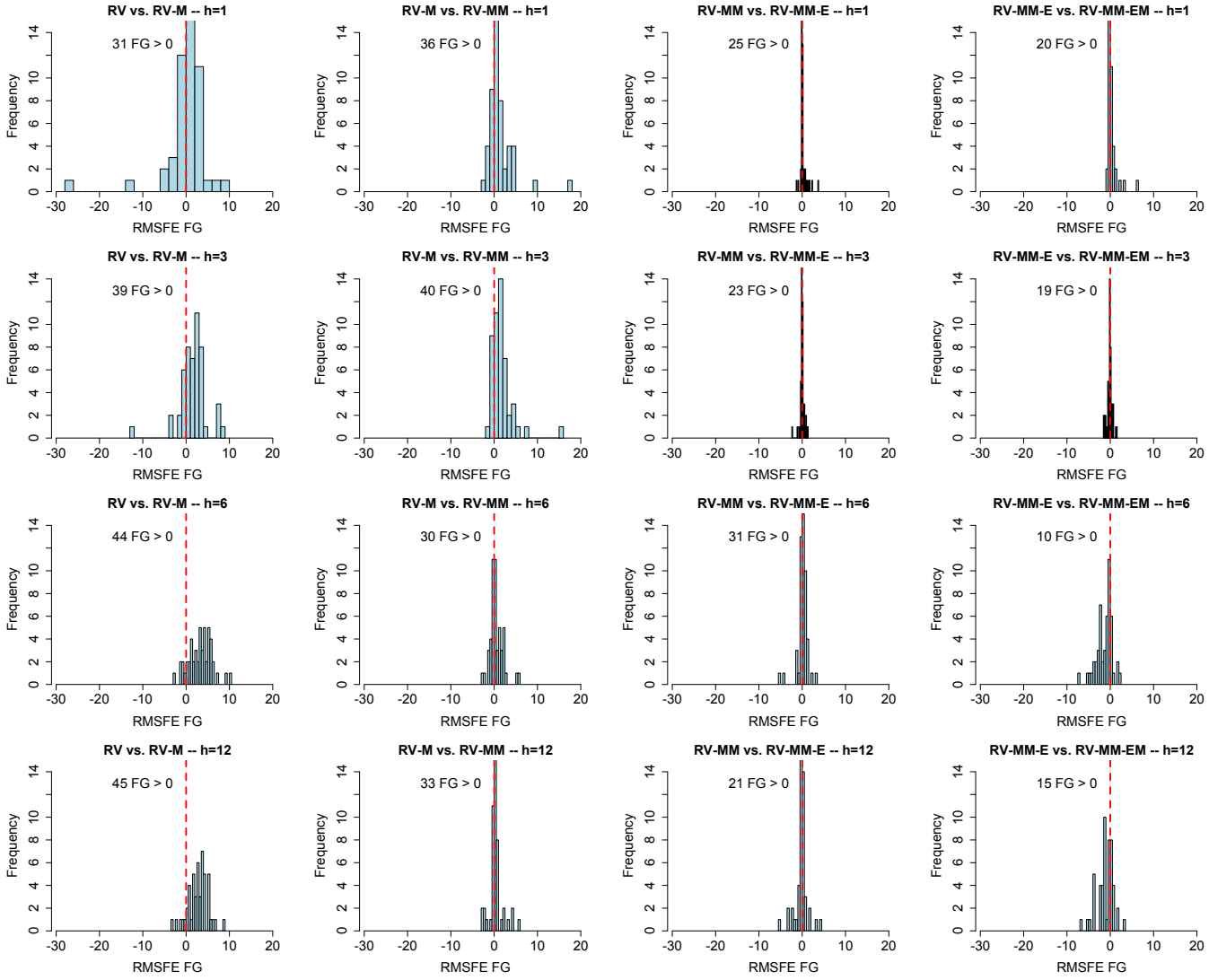
The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For every state, the inclusion of the different growth rates of electricity sales in the forecasting model during the out-of-sample period is recorded under the gMDL$_{actset}$ model-selection criterion. Then the sum across categories of electricity sales is computed to capture the total frequency of inclusion of electricity sales in the boosted RV-MM-EM forecasting model for a state. The worldclouds illustrate this frequency.

Figure 10: Boosting algorithm and the MAFE-FG statistic (gMDL$_{actset}$)



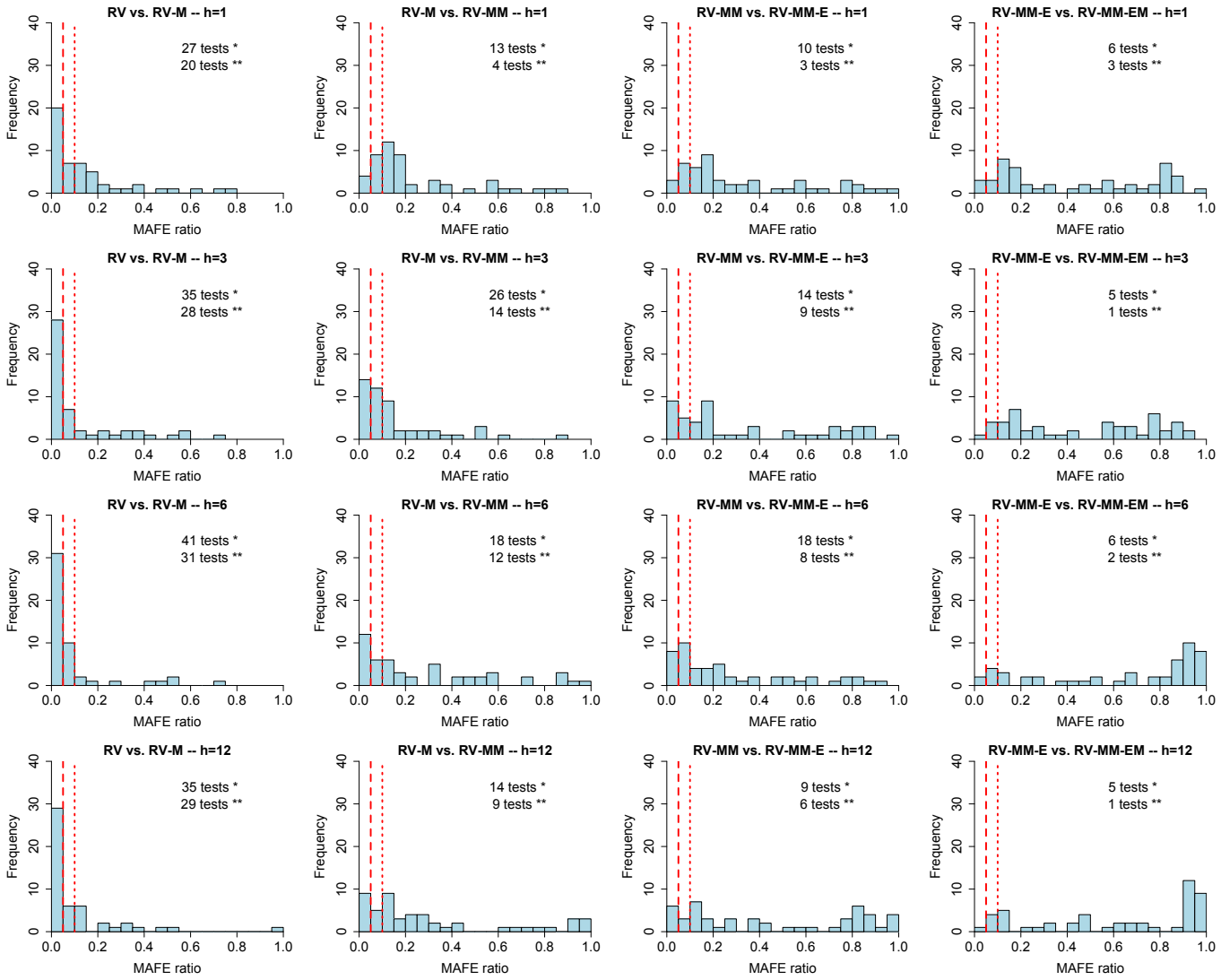
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the MAFE-FG statistic across the 50 states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the MAFE-FG statistic than the benchmark (first) model.

Figure 11: Boosting algorithm and the RMSFE-FG statistic (gMDL$_{actset}$)



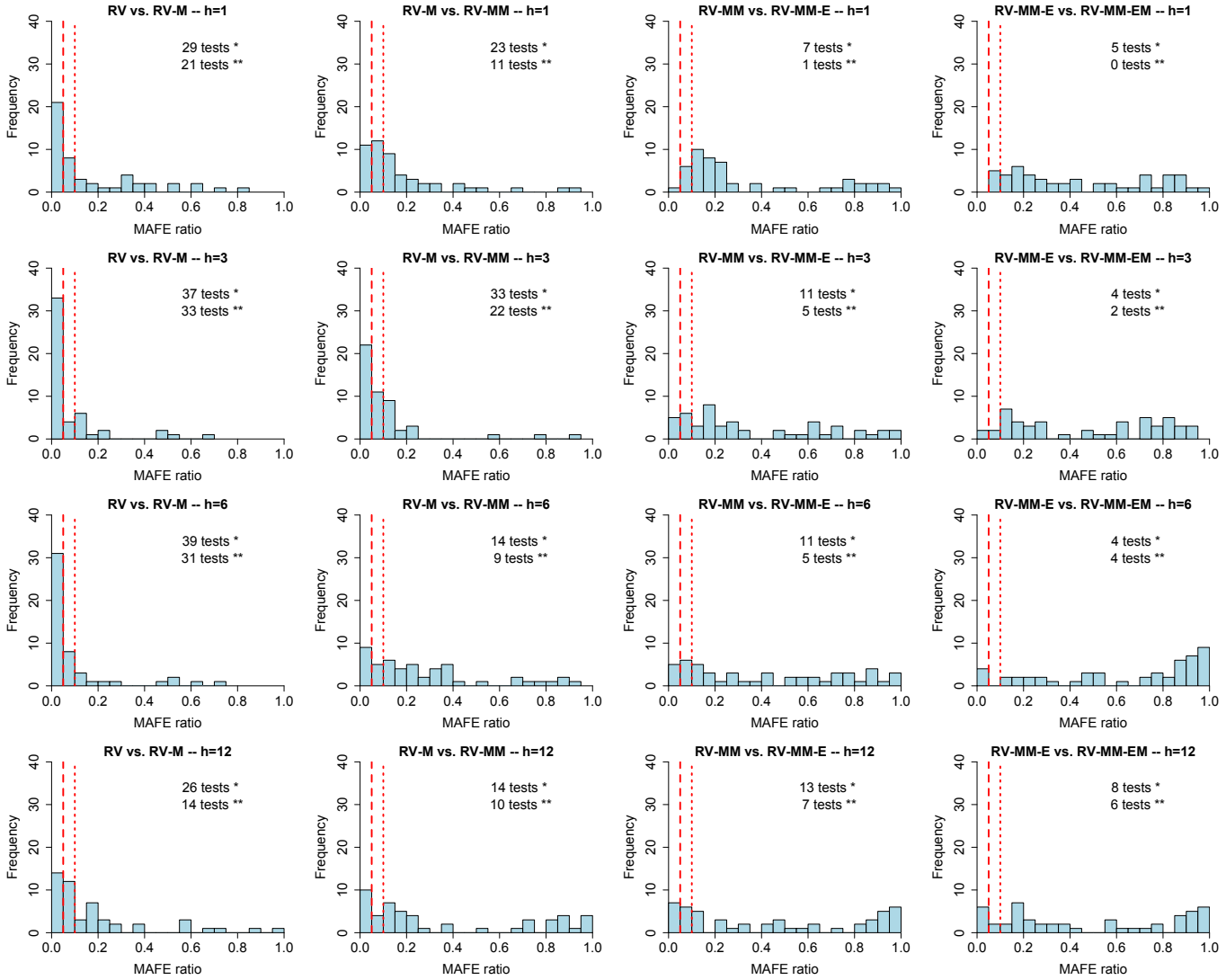
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for 50 state. The barplots visualize the distribution of the RMSFE-FG statistic across the 50 states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the RMSFE-FG statistic than the benchmark (first) model.

Figure 12: Boosting algorithm and the CW statistic (gMDL_{actset})



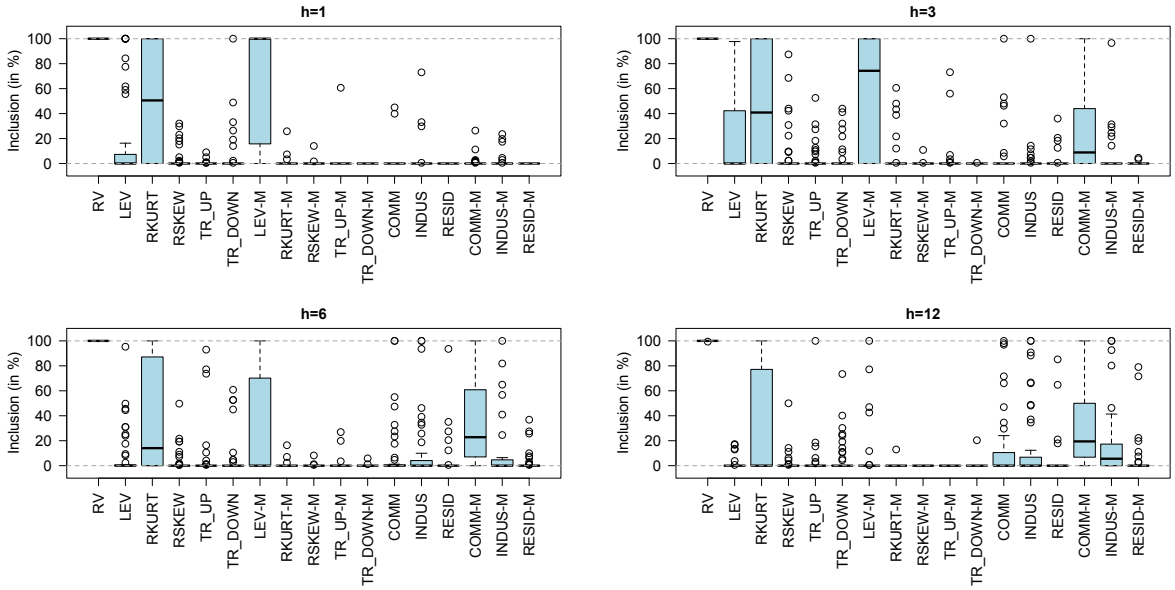
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the p-value (obtained using heteroscedasticity and autocorrelation consistent standard errors) of the CW test across the 50 states. The dashed vertical red lines indicate the 10% and 5% level of significance. The null hypothesis is that the benchmark (first) model and the rival (second) model perform equally well, while the one-sided alternative hypothesis is that the rival model performs better than the benchmark model.

Figure 13: Boosting, a rolling-estimation window, and the CW statistic (gMDL_{actset})



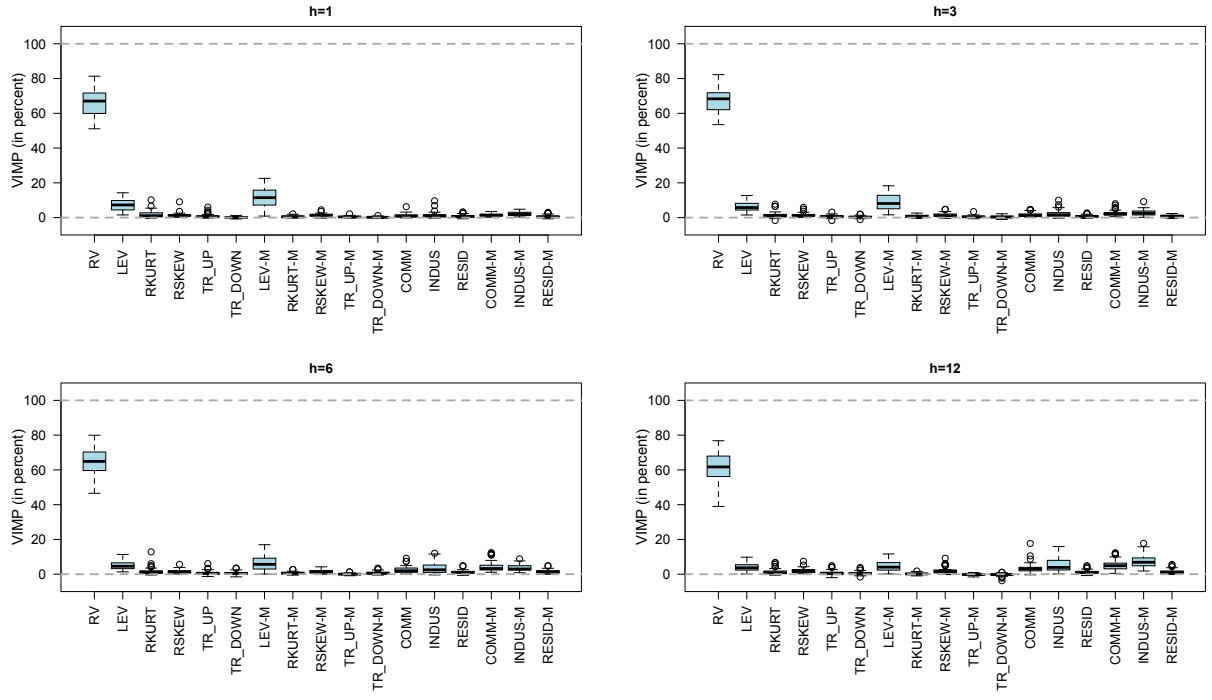
The forecasting models are estimated on a rolling-estimation window that covers 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the p-value (obtained using heteroscedasticity and autocorrelation consistent standard errors) of the CW test across the 50 states. The dashed vertical red lines indicate the 10% and 5% level of significance. The null hypothesis is that the benchmark (first) model and the rival (second) model perform equally well, while the one-sided alternative hypothesis is that the rival model performs better than the benchmark model.

Figure 14: Forward best predictor selection and selection of predictors



The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, the inclusion of predictors in the forecasting model is recorded for the BIC model-selection criterion, and all states. The box-and-whisker plots visualize how often (in percent), across all states, a predictor is included in the forecasting model in the out-of-sample period, where the solid horizontal line represents the median, the shaded area represents the interquartile range, the two whiskers denote the boundaries of 1.5 times the interquartile range, and circles represent points beyond the whiskers.

Figure 15: Random forests and variable importance



The forecasting models are estimated on the full sample of data. Variable importance is computed by calculating for every tree in the random forest the difference between the prediction errors that result when a predictor is perturbed and the original predictor. The result is averaged across all trees in the random forest, and then expressed in percent. The box-and-whisker plots visualize the distribution of variable importance across the states. The dashed vertical red line indicates a ratio of unity. A ratio that exceeds unity indicates that the rival (second) model performs better in terms of the RMSFE statistic than the benchmark (first) model.

Appendix

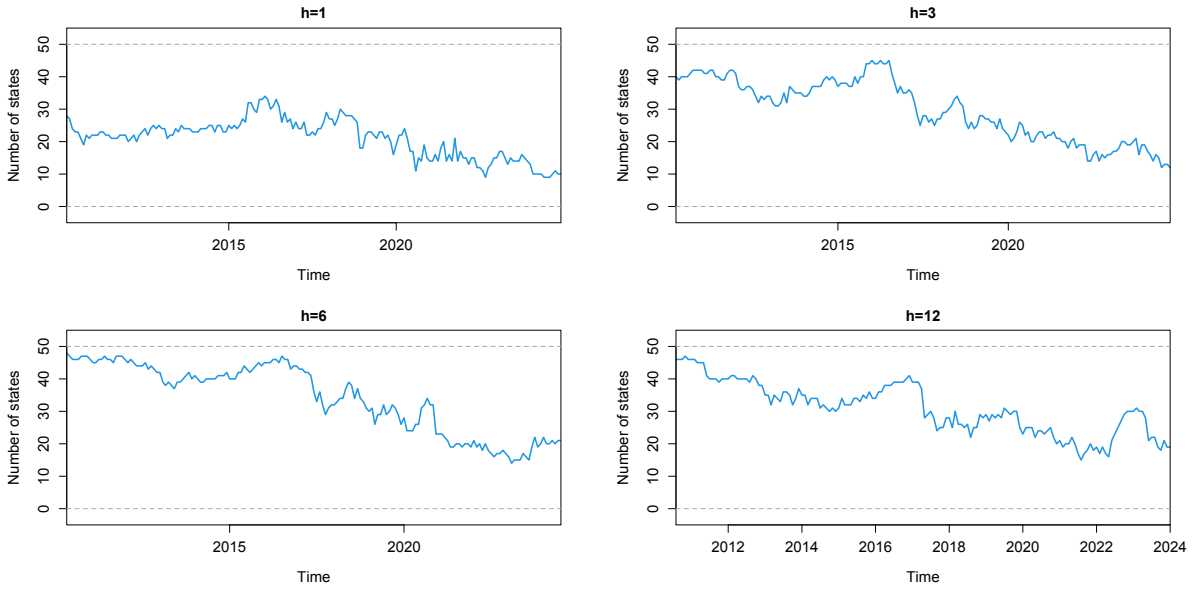
Table A1: Forward best predictor selection algorithm and optimal model selection criterion

Model	RV-M	RV-MM	RV-MM-E	RV-MM-EM	RV-M	RV-MM	RV-MM-E	RV-MM-EM
			MAFE				RMSFE	
			$h=1$				$h=1$	
Adj. R2	11	5	6	12	13	5	6	5
BIC	19	34	34	29	26	39	39	37
Cp	23	11	10	9	14	6	5	8
			MAFE				RMSFE	
			$h=3$				$h=3$	
Adj. R2	13	10	6	3	8	4	3	2
BIC	28	27	38	40	33	40	40	42
Cp	12	13	6	7	12	6	7	6
			MAFE				RMSFE	
			$h=6$				$h=6$	
Adj. R2	12	13	7	4	12	8	7	4
BIC	31	24	29	42	33	35	37	44
Cp	8	13	14	4	6	7	6	2
			MAFE				RMSFE	
			$h=6$				$h=6$	
Adj. R2	12	10	12	7	11	8	5	6
BIC	28	33	33	36	30	34	37	38
Cp	12	8	5	7	12	9	8	6

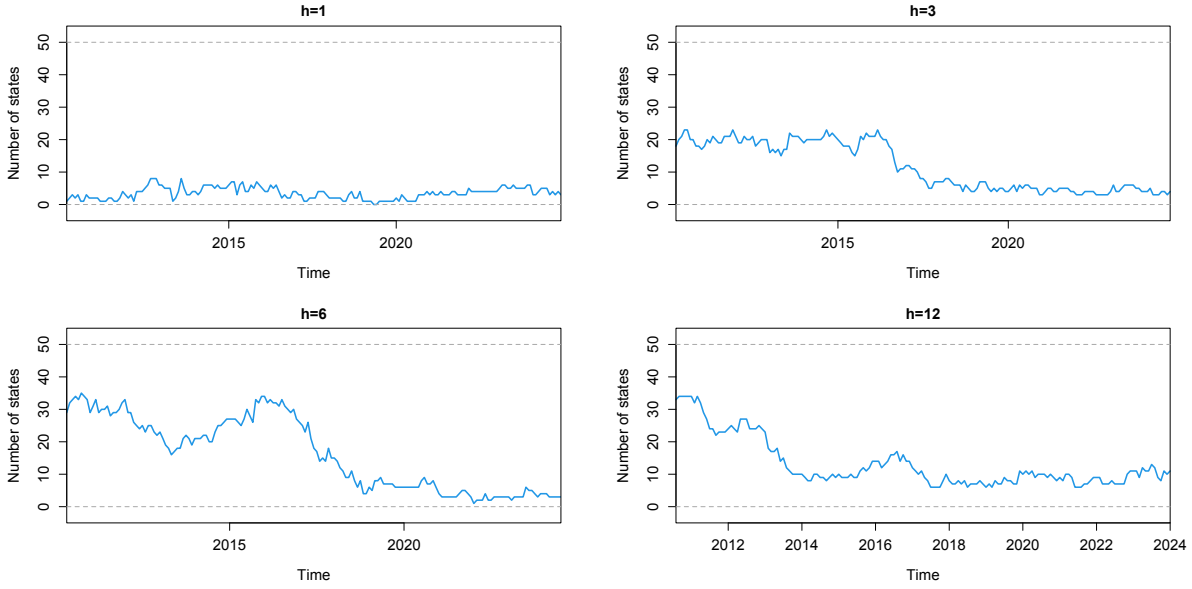
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. Using data for every state, the forward best prediction selection algorithm is implemented for every forecasting model by using the three different model-selection criteria. For every state, every forecasting model, and every forecast horizon, h, the out-of-sample MAFE (RMSFE) statistic is computed and the best model-selection criterion is identified as the one that minimizes the MAFE (RMSFE) statistic. The numbers summarized in the table represent the number of states for which a model-selection criteria yields the best model in terms of the MAFE (RMSFE) statistic. The numbers in the columns of the table, for a given forecast horizon, can exceed the total number of states in case two or more model-selection criteria minimize the MAFE (RMSFE) statistic.

Figure A1: Boosting algorithm and time paths of inclusion of the growth rate of electricity sales (RV-MM-EM)

Panel A: gMDL$_{trace}$

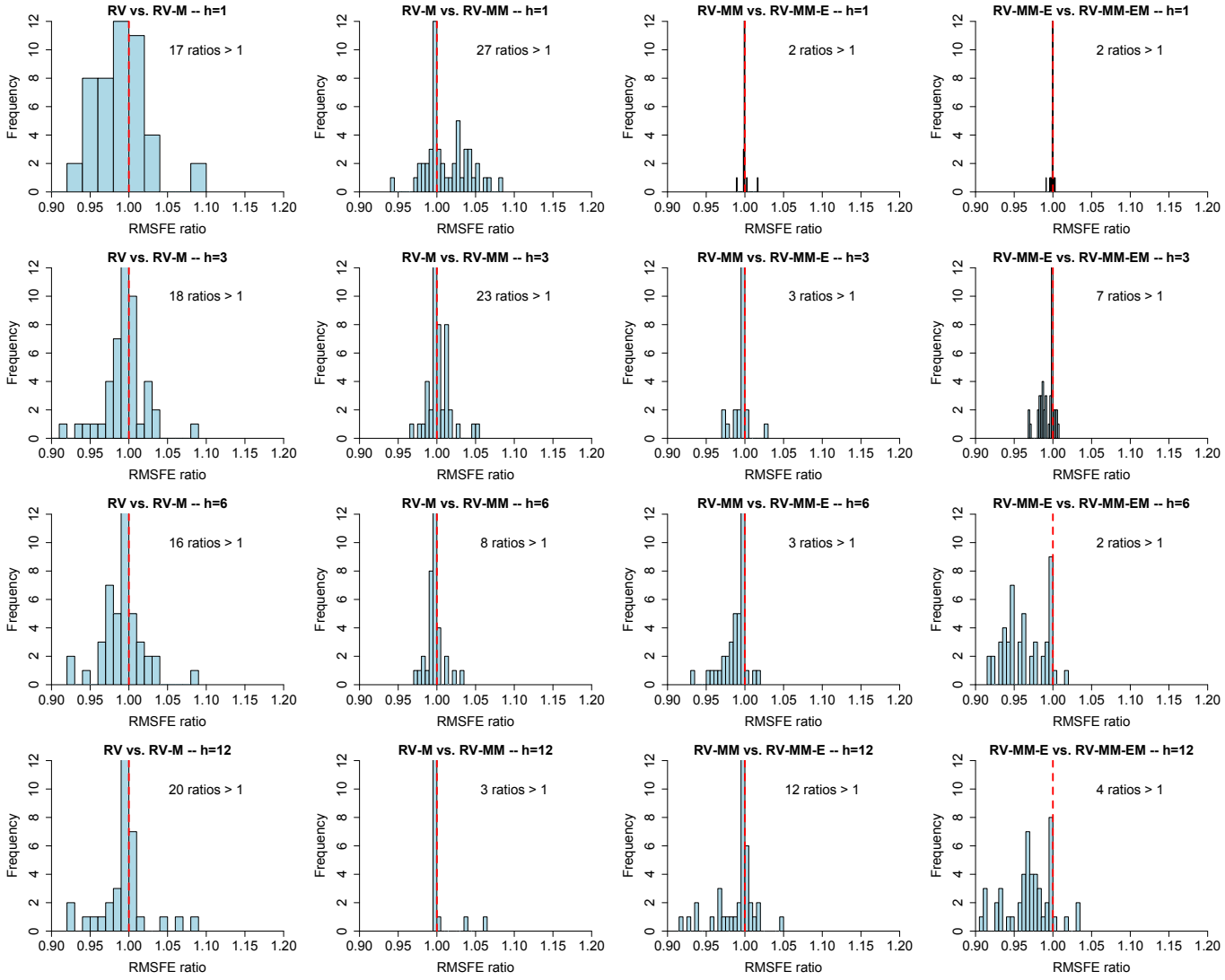


Panel B: gMDL$_{actset}$



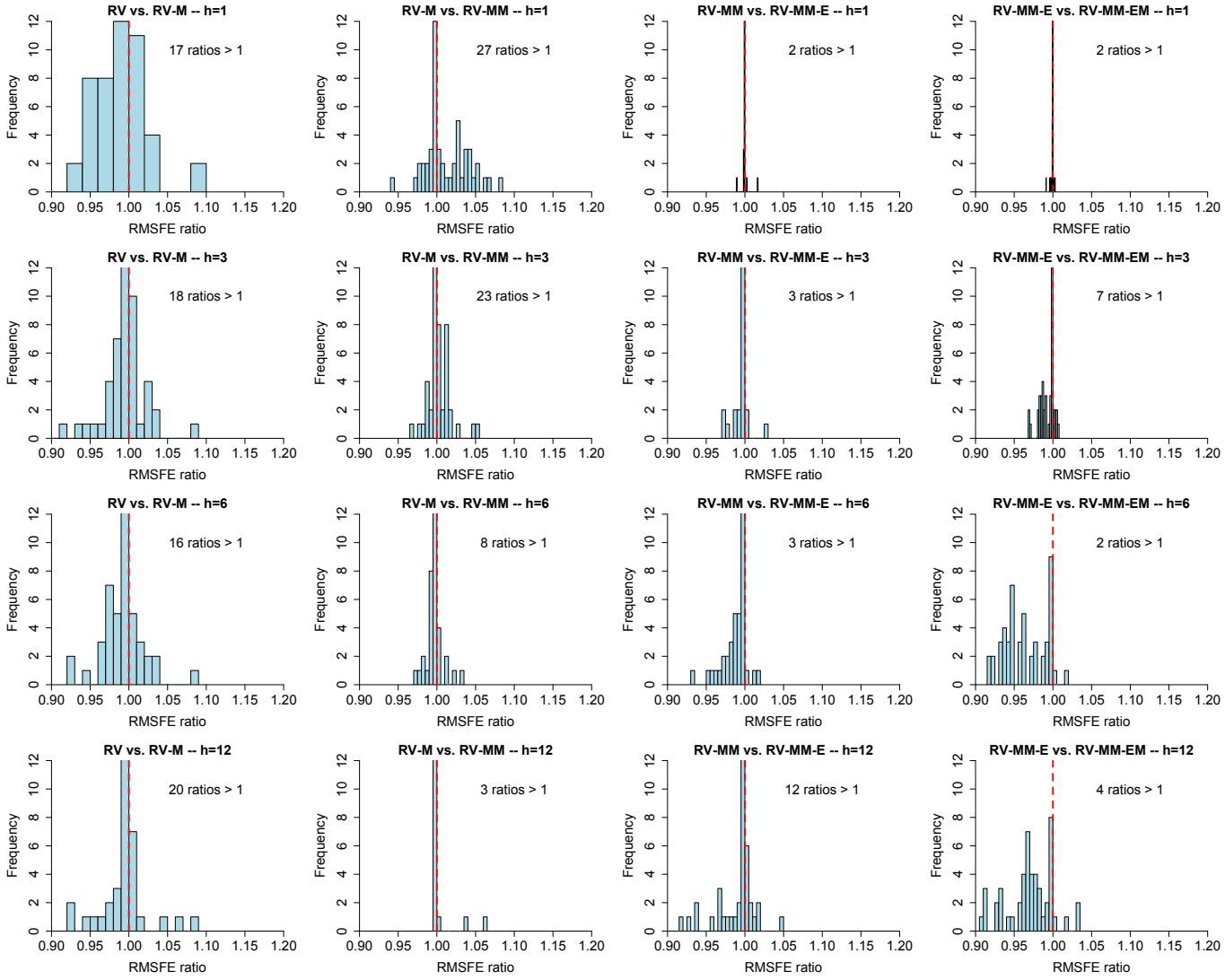
The RV-MM-EM forecasting model is estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, the inclusion of the subcategories of the growth rate of electricity sales in the forecasting model is recorded for the gMDL$_{trace}$ and gMDL$_{actset}$ model-selection criterion, and every state. It then is recorded for every state whether at least one of the subcategory is included in the forecasting model. Finally, the number of states for which at least one of the subcategory is included in the forecasting model is recorded and plotted over time.

Figure A2: Best predictor selection algorithm and the MAFE-FG statistic (BIC criterion)



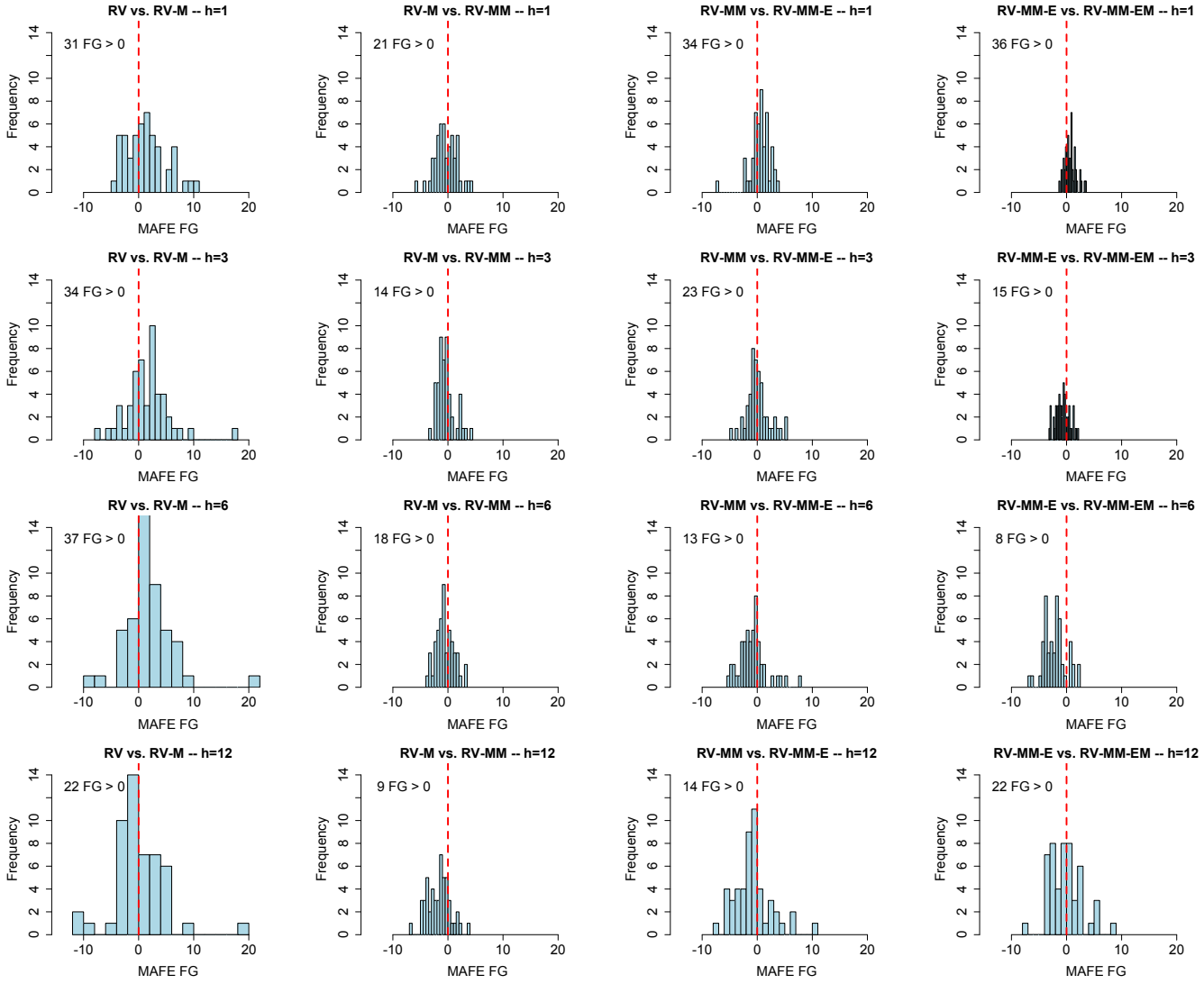
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the MAFE-FG statistic across the states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the MAFE-FG statistic than the benchmark (first) model.

Figure A3: Best predictor selection algorithm and the RMSFE-FG statistic (BIC criterion)



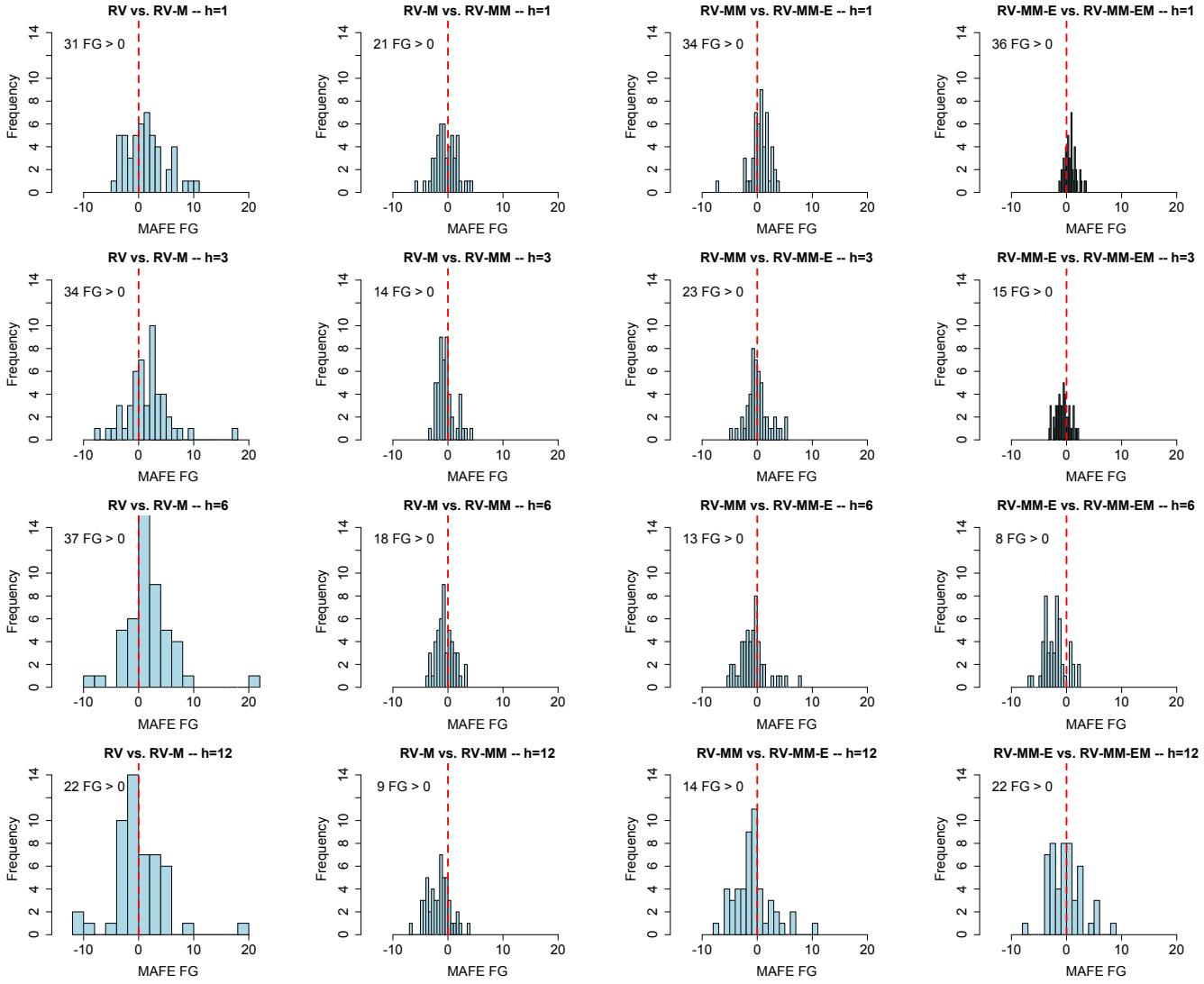
The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the RMSFE-FG statistic across the states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the RMSFE-FG statistic than the benchmark (first) model.

Figure A4: Random forests and MAFE-FG statistic



The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for every state. The barplots visualize the distribution of the MAFE-FG statistic across the states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the MAFE-FG statistic than the benchmark (first) model.

Figure A5: Random forests and the RMSFE-FG statistic



The forecasting models are estimated on a recursively expanding estimation window. The initial training period covers the first 50% of the data. For the out-of-sample period, forecast errors are recorded for the states. The barplots visualize the distribution of the RMSFE-FG statistic across the states. The dashed vertical red line indicates forecasting gains of zero. A ratio that exceeds zero indicates that the rival (second) model performs better in terms of the RMSFE-FG statistic than the benchmark (first) model.