Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors

Barış Soybilgen[1] · Ege Yazgan[1] 📷

Abstract

In this study, we nowcast quarter-over-quarter US GDP growth rates between 2000Q2 and 2018Q4 using tree-based ensemble machine learning models, namely, bagged decision trees, random forests, and stochastic gradient tree boosting. To solve the ragged edge problem and reduce the dimension of the data set, we adopt a dynamic factor model. Dynamic factors extracted from 10 groups of financial and macroeconomic variables are fed to machine learning models for nowcasting US GDP. Our results show that tree-based ensemble models usually outperform linear dynamic factor models. Factors obtained from real variables appear to be more influential in machine learning models. The impact of factors derived from financial and price variables can only become important in predicting GDP after the great financial crisis of 2008–9, reflecting the effect extra loose monetary policies implemented in the period following the crisis.

Keywords Bagging · Boosting · Dynamic Factor Model · Machine Learning · Nowcasting · Random forests

1 Introduction

When there is no problem in measuring the present state of variables of interest, forecasting concentrates only on predicting the future. For example, in weather forecasting, where we know exactly what the weather is today, we only need to forecast the future. However, in other fields, such as economics, we have missing information about the economy's current state, as important macroeconomic aggregates can only be measured with a considerable time delay. Hence, for these

✉ Ege Yazgan
ege.yazgan@bilgi.edu.tr

Barış Soybilgen
baris.soybilgen@bilgi.edu.tr

[1] Istanbul Bilgi University, Istanbul, Turkey

variables, forecasting tasks should also be focused on predicting the present as well as the future and the recent past.[1]

The publication delays of key economic macro indicators pose serious problems for policymakers and all who need to monitor the economy in real time and form their decisions based on timely information. For example, US GDP, which is the key variable describing the overall state of the economy as a whole, is published quarterly. Its first (or advance) estimate is released with almost a month's delay after the end of the corresponding quarter.[2] This delay may be extended up to six months in other countries. Even a delay of just one month constitutes a significant lag in information flow for those who need timely information for their decision-making process. In the absence of timely availability of GDP figures, decision-makers interested in monitoring the overall state of the economy in real time must rely on other indicators that are related to GDP but published with a shorter delay or with no delay at all. These variables can be used to extract information on the current state of economic activity well before the advance estimate of GDP is released. Giannone et al. (2008) developed a joint multivariate nowcasting model to perform this task. By putting an emphasis on forecasting to the present, they call it "nowcasting". This nowcasting model is a particular case of a large class of dynamic factor models (DFMs) estimated by principal components, first introduced by Stock and Watson (2002a), Stock and Watson (2002b), and using the Kalman filter to update predictions, and it is designed to handle the irregularities of real-time data, such as mixed frequencies and non-synchronicity of data releases. The model estimates the unobserved factors that drive the data and produces a forecast of each economic and financial series that it incorporates. Whenever the actual release for a series departs from the model's forecast, this is considered as "news" and affects the nowcast of GDP growth. The DFM proposed by Giannone et al. (2008) and its successors have been used to produce successful nowcasts for a variety of countries ranging from developed economies to emerging markets (for a review of the literature, see Bok et al. 2018; Banbura et al. 2013).

After deriving dynamic factors from a data set that contains high-frequency variables, most studies in the literature link these dynamic factors with quarterly GDP using a linear model. An obvious alternative to linear models is machine learning models. These have recently grown in popularity for macroeconomic forecasting but have rarely been used in the context of nowcasting. To the best of our knowledge, there have been only a handful of studies utilizing machine learning algorithms for nowcasting. In one of the early studies to adopt this approach, Cornec and Mikol (2011) used linear/quadratic discriminant analysis, decision trees, and support vector machines to nowcast the direction of French GDP by utilizing a quarterly data set including business surveys. They showed that linear discriminant analysis outperformed other machine learning and benchmark models in their exercise. In another study, Biau (2010) utilized a quarterly data set that contained only soft data to nowcast Euro Area GDP with a random forest. Their results

---

[1] The need for forecasting the near past is also due to time delay in data releases.

[2] The publication of this first advanced print is complemented by subsequent revisions, namely, the second and third estimates for the quarter in question, in the following months.

indicated that the predictive performance of random forests is not very satisfactory. However, selecting variables according to the importance metric of random forests and then using these variables in a linear model provided good nowcasting accuracy. In a similar study, Richardson et al. (2018) adopted a large-scale quarterly data set, including 500 domestic and international variables. They nowcast New Zealand GDP with various machine learning algorithms such as K-nearest-neighbor regressions, boosted trees, elastic nets, lasso regression, ridge regression, support vector machines, and neural network models. According to their results, the majority of machine learning models outperformed autoregressive models, factor models, and Bayesian vector autoregressive models. Among machine learning models, support vector machines and neural networks had the best performance. By adopting quarterly data sets, Cornec and Mikol (2011), Biau (2010), and Richardson et al. (2018) seemed to ignore the nonsynchronicity of data releases that causes missing values at the end of the data set. However, this is an important issue in nowcasting that should also be taken into account. In contrast to previous studies, Loermann and Maas (2019) used the large-scale monthly data set of McCracken and Ng (2016) (FRED-MD) to nowcast US GDP with a feedforward artificial neural network model. They adopted autoregressive moving average models to fill in missing values at the end of the sample. Even though they used a monthly data set, they conducted one prediction in each quarter. According to their results, their machine learning approach outperformed DFMs and performed as well as a survey by professional forecasters. In a similar line of study, Soybilgen (2020) also used the monthly FRED-MD data set to nowcast US business cycle states with a neural network model. Instead of feeding the data set directly into the machine learning model, Soybilgen (2020) first used a DFM to reduce the dimension of the data set and deal with the nonsynchronicity of the data releases. He then fed these dynamic factors into the machine learning model. His results also indicated that machine learning models provide improvements over regular linear models. In the present paper, we also adopt a similar two-step approach for nowcasting US GDP.

In this study, we nowcast US GDP between 2000Q2 and 2018Q4 using decision-tree-based ensemble machine learning models, namely, bagged decision trees, random forests, and stochastic gradient tree boosting. We also use the large-scale data set of FRED-MD, including more than 100 financial and macro variables. Instead of feeding these variables directly into our machine learning models, we reduce the dimension of the data set using the DFM proposed by Giannone et al. (2008). This helps us both to fill in missing monthly data at the end of the sample in a straightforward manner and to reduce the dimension of the data set.[3] However unlike previous nowcasting studies such as those by Bańbura and Rünstler (2011), Barhoumi et al. (2010), D'Agostino et al. (2012), Matheson (2010), and many others, we do not derive dynamic factors using the whole data set. Instead, we first divide the data set into 10 groups of variables and then derive factors from each group of variables.

---

[3] In a comprehensive study, Coulombe et al. (2019) further showed that combining machine learning models with factors improves the forecasting performance of the models significantly in a data-rich environment.

In our real-time nowcasting exercise that takes account of both historical data availability and data revisions, our tree-based ensemble machine learning models mostly outperform linear DFMs. In the first GDP predictions for the reference quarter, the performance difference between machine learning models and linear models is small. However, when additional data for the reference quarter become available, the prediction performance of tree-based models improves significantly compared to that of linear models. We estimate all models using both a rolling window and an expanding window. Our results indicate that machine learning models estimated using a rolling window have better prediction performance compared to models estimated recursively. We also compare predictions of our models against those of GDPNow, which is a well-known nowcasting model. We show that machine learning models outperform GDPNow slightly when nowcasting at the start of the reference quarter, but GDPNow performs better than machine learning models when nowcasting at the end of the reference quarter. We also analyze which factors are more important when predicting US GDP. Our results show that factors obtained from real variables have much more impact than factors obtained from financial and price variables. However, for random forests and bagged decision trees, the influence of factors derived from financial and price variables increases only after the great financial crisis of 2008-9.

The remainder of this paper is organized as follows: Sect. 2 introduces the data set; Sect. 3 describes the methodology; Sect. 4 presents the empirical results, and Sect. 5 concludes the paper.

2 The data Set

The large-scale data set used in this paper to obtain dynamic factors is based on the FRED-MD monthly database provided by McCracken and Ng (2016). FRED-MD consists of 10 groups of variables: (1) output and income; (2) labor market; (3) housing; (4) consumption, orders, and inventories; (5) money and credit; (6) interest rate; (7) prices; (8) stock market; (9) yield spread; and (10) exchange rate. We use vintage data starting from January 2000 until December 2018.[4] Owing to discontinuities in some old series, the introduction of some newly updated series, and some other data collection problems, FRED-MD vintage data do not have the same number of variables for each period. Variables and their period of use are listed in Appendix 1. Furthermore, all variables are transformed appropriately to ensure stationarity. Their applied transformations are also shown in Appendix 1. For vintage GDP data, we use the data set obtained from the Archival Federal Reserve Economic Data (ALFRED) system.

---

[4] In total, we use 228 different vintage data sets to carry out our nowcasting exercise, and each data set starts from July 1960.

3 Methodology

In this study, we use tree-based ensemble machine learning algorithms that incorporate dynamic factors as explanatory variables. Our reason for adopting dynamic factors instead of using the full data set is that a large number of irrelevant and noisy variables can reduce the prediction performance of models. Using a DFM allows us to reduce the dimension of the data set by eliminating most of the noise that it contains. Furthermore, the DFM of Giannone et al. (2008) can solve the ragged/jagged edge data problem[5] by utilizing a Kalman smoother.

3.1 Dynamic Factor Model

Let us define $x_{t_m} = (x_{1,t_m}, x_{2,t_m}, \ldots, x_{n,t_m})'$, $t_m = 1, 2, \ldots, T_m$ as n monthly standardized series transformed via the Mariano and Murasawa (2003) approximation with t_m being the monthly time index and T_m representing the final month in the monthly data set. Our factor model has the following representation:

$$x_{t_m} = \Lambda f_{t_m} + \epsilon_{t_m}; \quad \epsilon_{t_m} \sim \mathbb{N}(0, \Sigma_{\epsilon_{t_m}}), \tag{1}$$

where Λ is an $n \times r$ matrix of factor loadings for standardized and filtered monthly variables, ϵ_t is the idiosyncratic component, and $f_{t_m} = (f_{1,t_m}, f_{2,t_m}, \ldots, f_{r,t_m})'$ represents the unobserved common factors following a vector autoregression process (VAR) as follows:

$$f_{t_m} = \sum_{i=1}^{p} A_i f_{t_m - i} + B\eta_t; \quad \eta_{t_m} \sim \mathbb{N}(0, I_q), \tag{2}$$

where B is an $r \times q$ matrix of full rank q with $q \leqslant r$, $A_1, A_2, \ldots, A_p$ are $r \times r$ matrices of autoregressive coefficients, and η_t is the q-dimensional vector of common shocks, which follows a white-noise process.

In this study, we use the two-step estimation approach which, as shown by Doz et al. (2011), is able to extract common factors in case of missing values at the end of the sample. In the first step, initial factors and consistent estimates of the parameters are obtained. In the second step, updated estimates of the common factors are obtained with Kalman filtering techniques using the consistent estimates of the parameters. The two-step estimation procedure can be summarized as follows:

1. We extract the first r principal components from the balanced part of the data set where all observations are present[6] and obtain the initial factor estimates, $\tilde{f}_{t_m}$.
2. Using the initial factor estimates, we estimate factor loadings, $\hat{\Lambda}$, and the covariance matrix of the idiosyncratic component, $\hat{\Sigma}_{\epsilon_{t_m}}$, in Eq. (1).

---

[5] The missing data at the end of the sample period due to the publication lag is called the "ragged/jagged edge" problem in the literature. To predict GDP, we first need to forecast the missing values of monthly variables at the end of the sample.

[6] In other words, we ignore the ragged/jagged part of the data set.

3. Similarly, we obtain estimated matrices of autoregressive coefficients, $\hat{A}_1, \hat{A}_2, ..., \hat{A}_p$, and estimated $\hat{B}$ from Eq. 2.

4. The ragged edge part of the data set are incorporated into the procedure by assigning an extremely large value to the variance of the idiosyncratic component where there is missing observations and replacing missing values in x_{t_m} with arbitrary values. In this way, Kalman filter puts no weight on missing observations while computing the factors (Giannone et al. 2008).

5. As Eqs. 1 and 2 can be cast in state-space form using the consistent estimates $(\hat{\Sigma}, \hat{B}, \hat{A}, \hat{\Lambda})$, factors can be re-estimated using one run of Kalman filter and Kalman smoother while incorporating the unbalanced part of the data set.

3.2 Linear Models

To link monthly factors with quarter-over-quarter (QoQ) GDP growth rates, we obtain quarterly factors from their monthly counterparts by extracting those that correspond to the last month of each quarter. Let us assume that f_{t_m}, $t_m = 1, 2, \ldots, T_m$, starts at the first month of a quarter. Then its quarterly counterpart can be represented as $f_{t_q^f}, t_q^f = 1, 2, \ldots, T_q^f$,, where $T_m = T_q^f/3$, with t_q^f being the quarterly time index for factors and T_q^f representing the sample length of quarterly factors. Similarly, let us define QoQ GDP growth rates y_{t_q}, $t_q = 1, 2, \ldots, T_q$, with t_q being the quarterly time index for y and T_q representing the sample length of QoQ GDP growth rates. We can obtain accurate nowcasts of y owing to the fact that $T_q^f > T_q$.

We use two linear models as benchmarks. In the first model, two factors are derived from the whole data set using the $r = 2, q = 2, p = 1$ specification,[7] where r, q, and p represent the static factors, the dynamic factors, and the number of lags, respectively, as introduced in Sect. 3.1. The QoQ GDP growth rates y_{t_q} can be modeled by the quarterly factors $\hat{f}_{i,t_q|t_m}$, which are extracted from the monthly factors estimated using all the information up to T_m:

$$y_{t_q} = c + \sum_{i=1}^{2} \beta_i \hat{f}_{i,t_q|t_m} + \zeta_t, \tag{3}$$

where ζ_t is the error term. Then, the h_q-steps-ahead predictions of quarterly GDP growth rates using parameters derived from Eq. (3) via OLS, $\hat{y}_{t_q+h_q|t_q}$, are obtained as follows:

---

[7] There are various ways to determine the number of factors, such as using the information criteria proposed by Bai and Ng (2002) and Bai and Ng (2007), the best being an ex-post specification grid search with cross-validation. However, we follow a simple but frequently used approach and adopt the specification of Giannone et al. (2008), which is quite appropriate for these kinds of nowcasting exercises. Soybilgen (2020) also tests various DFM specifications in a similar nowcasting exercise and shows that the $r = 2, q = 2, p = 1$ specification produces overall good results.

$$\hat{y}_{t_q+h_q|t_q} = \hat{c} + \sum_{i=1}^{2} \hat{\beta}_i \hat{f}_{i,t_q+h_q|t_m}. \tag{4}$$

In the second model, two factors are extracted from each group in the data set using the $r = 2, q = 2, p = 1$ specification, and we obtain 20 factors in total. As in Eq. (3), y_{t_q} is linked to factors using a single-equation model:

$$y_{t_q} = c + \sum_{i=1}^{20} \alpha_i \hat{f}_{i,t_q|t_m} + \varsigma_t, \tag{5}$$

where ς_t is the error term. Then, h_q-steps-ahead predictions of quarterly GDP growth rates using parameters derived from Eq. (5) via OLS, $\hat{y}_{t_q+h_q|t_q}$, are obtained as follows:

$$\hat{y}_{t_q+h_q|t_q} = \hat{c} + \sum_{i=1}^{20} \hat{\alpha}_i \hat{f}_{i,t_q+h_q|t_m}. \tag{6}$$

3.3 Tree-Based Machine Learning Models

In this study, we use tree-based ensemble machine learning models, namely, bagged decision trees, random forests, and boosted decision trees. As in Eq. (5), we feed 20 factors into our machine learning models, since this approach gives us more information when analyzing which factors are important for predicting GDP.

3.3.1 Bagged Decision Trees and Random Forests

Classification and regression trees (CART), proposed by Breiman et al. (1984), work by dividing the feature space into mutually exclusive rectangular regions that minimize an objective function, namely, the residual sum of squares (RSS) for regression trees, while fitting a simple model in each region. As searching for the best optimal partition that minimizes the objective function is computationally infeasible, the recursive binary splitting strategy, which is a top-down greedy approach, is adopted. In this greedy strategy, we recursively split the space into two distinct regions by finding the best variable k and split point s that minimize the RSS until a stopping criterion is reached, such as the minimum number of observations in each region.

For ease of notation, let us assume that the feature space is partitioned into M regions $R_1, R_2, \ldots, R_M$ and that $\mathbf{f} = \{f_1, f_2, \ldots, f_{20}\}$ denotes the 20 factors used as predictors in our model. Following (Hastie et al. 2009, p. 307), the regression tree can be represented using the following piecewise-constant model:

$$g(\mathbf{f}) = \sum_{m=1}^{M} c_m \mathbb{I}(\mathbf{f} \in R_m), \tag{7}$$

where $\mathbb{I}$ is the indicator function, which is 1 when the arguments evaluate to true and 0 otherwise. The best estimator that minimizes the RSS is $\hat{c}_m = \text{average}\{y_{t_q} : f_{t_q} \in R_m\}$. Regions are defined recursively by finding the split variable k and split point s that solve

$$\min_{k,s} \left[\min_{c_1} \sum_{f_{t_q} \in R_1(k,s)} (y_{t_q} - c_1)^2 + \min_{c_2} \sum_{f_{t_q} \in R_1(k,s)} (y_{t_q} - c_2)^2 \right], \tag{8}$$

where $R_1(k,s) = \{f|f_k < s\}$ and $R_2(k,s) = \{f|f_k \geq s\}$. The optimal solution of the minimization problem (8) is obtained as $\hat{c}_1 = \text{average}(y_{t_q} : f_{t_q} \in R_1)$ and $\hat{c}_2 = \text{average}(y_{t_q} : f_{t_q} \in R_2)$.

Even though decision trees are extremely easy to interpret, they tend to make poor and noisy predictions in many cases compared to more advanced machine learning models. Breiman (2001) introduced random forests, which are a type of an ensemble decision tree model, as a technique with low variance and high prediction performance. Random forests are based on bagging (bootstrap aggregating) of decision trees. For bagged decision trees, we first obtain B bootstrapped training sets from original data and then fit a decision tree to each bootstrapped training set. This procedure is known to reduce variance while increasing the prediction performance of decision trees. However, if fitted bagged decision trees are too strongly correlated with each other, this procedure may fail to yield the desired improvement. Random forests solve this problem by allowing only a random sample of variables to be considered in each split. In this way, bagged fitted trees are dissociated from each other.

Following (Hastie et al. 2009, p. 588), let $b = 1, \ldots, B$ denote the number of bootstrap iterations. The random forest algorithm can then be summarized as follows:

1. Obtain the bootstrapped data from the original data covering the time span up to T_q, $t_q = 1, \ldots, T_q$.
2. Using the bootstrapped data obtained in step 1, estimate a regression tree $\hat{g}_{RF}^{(b)}(f)$ by just considering p factors at random from 20 factors when determining the best variable/split point for each terminal node of tree until the minimum node size $n_{\min}$ is reached.
3. Repeat steps 1 and 2 B times.

After obtaining B decision trees using the above procedure, h_q-steps-ahead predictions of QoQ GDP growth rates are calculated as the average value of B trees as follows:

$$\hat{y}_{t_q+h_q|t_q} = \frac{1}{B} \sum_{b=1}^{B} \hat{g}_{RF}^{(b)}(\hat{f}_{t_q+h_q|t_m}). \tag{9}$$

If $m = 20$ in the random forest procedure, then we obtain bagged decision trees $\hat{g}_{BG}^{(b)}(\hat{f})$. Similar to Eq. (9), we obtain the predictions as

$$\hat{y}_{t_q+h_q|t_q} = \frac{1}{B}\sum_{b=1}^{B}\hat{g}_{BG}^{(b)}(\hat{f}_{t_q+h_q|t_m}). \tag{10}$$

3.3.2 Boosted Decision Trees

Boosting is a general approach that turns weak learners into strong learners in a sequential way instead of separately as in random forests. It is mostly used in the context of decision trees. After an initial estimate, each tree is fitted to the residual of the previous estimate, and this fitted tree is then used to update the current estimate according to a learning parameter. To predict GDP growth rates, we use stochastic gradient tree boosting with squared errors as loss function, following Friedman (2001) and Friedman (2002).

In gradient boosting, regression trees are fitted to pseudo-residuals, with negative gradients, instead of actual residuals, since this simplifies the optimization process. Friedman (2001) also used different learning rates for each region of a decision tree for higher prediction performance. In stochastic gradient tree boosting, Friedman (2002) further improved the model by using only a part of the training set drawn at random without replacement in each iteration.

Following Friedman (2002) and (Hastie et al. 2009, p. 361), let us define $m = 1, \ldots, M$ as the number of boosting iterations, $\{y_{t_q}, f_{t_q}\}_1^{T_q}$ as the original training set, $\{y_{\pi(t_q)}, f_{\pi(t_q)}\}_1^{\tilde{T}_q}$ as the fraction of the original training set randomly selected without replacement, and λ as the learning parameter. Then the gradient tree boosting algorithm can be summarized as follows:

1. Initialize $h_0(\boldsymbol{f}) = \text{argmin}_\gamma \sum_{t_q=1}^{T_q} L(y_{t_q}, \gamma)$.

2. Use the fraction of the training set, $\{y_{\pi(t_q)}, f_{\pi(t_q)}\}_1^{\tilde{T}_q}$.

3. For $\pi(t_q) = 1, 2, \ldots, \tilde{T}_q$ compute $r_{\pi(t_q),m} = \left[\dfrac{\partial L\left(y_{\pi(t_q)}, h(f_{\pi(t_q)})\right)}{\partial h(f_{\pi(t_q)})} \right]_{h=h_{m-1}}$.

4. Fit a regression tree to the targets $r_{\pi(t_q),m}$, giving terminal regions $R_{j,m}, j = 1, 2, \ldots, J_m$.

5. For $j = 1, 2, \ldots, J_m$, compute $\gamma_{j,m} = \text{argmin}_\gamma \sum_{f_{\pi(t_q)} \in R_{j,m}} L\left(y_{\pi(t_q)}, h_{m-1}(f_{\pi(t_q)}) + \gamma\right)$.

6. Update $h_m(\boldsymbol{f}) = h_{m-1}(\boldsymbol{f}) + \lambda \sum_{j=1}^{J_m} \gamma_{j,m} \mathbb{1}(\boldsymbol{f} \in R_m)$.

7. Repeat steps 2, ..., 6 M times.

8. Derive the final model $\hat{h}(\boldsymbol{f}) = h_M(\boldsymbol{f})$.

After the final model $\hat{h}(\boldsymbol{f})$ has been obtained, h_q-steps-ahead predictions of QoQ GDP growth rates are calculated as $\hat{y}_{t_q+h_q|t_q} = \hat{h}(\hat{f}_{t_q+h_q|t_m})$. In the above algorithm,

the complexity of the regression trees can be controlled by adjusting the minimum node size of the terminal nodes, n_{min}, and the maximum depth allowed for each tree, C. Higher C and lower n_{min} produce more complex regression trees, which may cause overfitting of the procedure to occur more quickly. Furthermore, the learning rate λ controls the learning speed of the algorithm. A small value of λ indicates a slow learning process, so we may need a large number M to have a low bias error. On the other hand, a higher λ may cause the model to memorize data quickly, so we need to use a lower number of iterations.

4 Empirical Results

4.1 Nowcasting Performance

We estimate our models between January 2000 and December 2018 using vintage data that take account of data revisions. In each month, we produce predictions for both the current quarter and the next quarter. We assume that each prediction is computed at the end of the month and replicate historical data availability accordingly. As a result, we have six predictions for each reference quarter.

For example, let us assume that we are in January 2001. To replicate the information set of a forecaster in January 2001, we use the vintage data set of January 2001. This data set covers the period between July 1960 and December 2000. To predict the current and next quarters, we also need the period between January 2001 and June 2001. In the first step, we obtain common monthly factors between July 1960 and June 2001 using the DFM. Next, we extract quarterly factors from the monthly factors by collecting each monthly factor that corresponds to the last month of each quarter. As a result, our quarterly factors include the period between 1960Q3 and 2001Q2. In January 2001, we have GDP data until 2000Q4. To estimate the models, we regress quarterly factors on QoQ GDP growth rates between 1960Q3 and 2000Q4. After estimating the models, we calculate nowcasts of GDP data for 2001Q1 and 2001Q2 using quarterly factors of 2001Q1 and 2001Q2. As a result, we obtain the first nowcast of the QoQ GDP growth rate for 2001Q2 in January 2001. The first nowcast for the reference quarter is performed with very little information about the reference quarter, so it is expected to be one of the worst nowcasts. Following the same procedure, the second nowcast of the QoQ GDP growth rate for 2001Q2 is estimated in February 2001 using the vintage data set of February 2001. The third, fourth, fifth, and sixth nowcasts are also estimated in the same manner. Finally, when the advance/first estimate for 2001Q2 is announced by the US Bureau of Economic Analysis (BEA) in July 2001, we can evaluate the nowcasting accuracy of model predictions.

By using root mean square errors (RMSEs), we evaluate the accuracy of the ith nowcasts produced by our models between 2000Q2 and 2018Q4:

$$RMSE_i = ((1/n) \sum_{t_q=2000Q2}^{2018Q4} (y_{t_q} - \hat{y}_{t_q}^{(i)})^2)^{1/2}; \quad i = 1, 2, ..., 6, \tag{11}$$

where $\hat{y}_{t_q}^{(1)}$ denotes the first nowcast, $\hat{y}_{t_q}^{(2)}$ denotes the second nowcast, and so on. n is the number of nowcasts. The RMSE is not always easy to interpret since it gives disproportionally more weight to outliers, so we also use mean absolute errors (MAEs) to evaluate the nowcasting accuracy of models. The MAE of the ith nowcast is calculated as

$$MAE_i = (1/n) \sum_{t_q=2000Q2}^{2018Q4} |y_{t_q} - \hat{y}_{t_q}^{(i)}|; \quad i = 1, 2, ..., 6. \tag{12}$$

We estimate our models using both an expanding estimation window and a rolling estimation window. In the expanding estimation window, we increase the monthly data set by one month in each nowcasting iteration. In the rolling estimation window, we keep the time span of our monthly data set fixed. For example, our estimation period in January 2000 is between July 1960 and December 1999. In February 2000, we use the estimation period between July 1960 and January 2000 for the expanding estimation window and the estimation period between August 1960 and January 2000 for the rolling estimation window.

During the nowcasting exercise, we choose optimal hyperparameters once using the initial estimation window.[8] We perform a grid search using threefold cross-validation three times.[9] For bagged decision trees, we choose hyperparameters that minimize the RMSE according to the number of trees, $B = 100, 200, \ldots, 1000$. The optimal number of trees is 1000. For random forests, we choose hyperparameters that minimize the RMSE according to the number of trees and the number of variables considered at each split, $B = 100, 200, \ldots, 1000$ and $p = 2, 4, 6, \ldots, 18$. For random forests, the optimal numbers of trees and variables are 500 and 12, respectively. For gradient boosting machines, we make a grid search over the following parameters: the maximum depth of each tree, which controls the complexity of the trees, $C = 1, 2, 4, 6, 8, 10$; the learning rate $\lambda = 2^{-8}, 2^{-7}, \ldots, 2^{-1}$; the number of boosting iterations, $M = 10 \times 2^1, 10 \times 2^2, 10 \times 2^3, \ldots, 10 \times 2^8$; and the minimum number of observations in the terminal nodes of the trees, $n_{\min} = 2, 4, \ldots, 20$.[10] The optimal parameter combination for stochastic gradient tree boosting is $C = 2$, $\lambda = 0.03125$, $M = 80$, and $n_{\min} = 2$.[11]

---

[8] We could also choose optimal hyperparameters in each nowcasting iteration, but this greatly increases the computational burden.

[9] We also choose hyperparameters using fivefold cross-validation and reach similar results.

[10] Grid search for stochastic gradient tree boosting is a difficult process. To be sure that we conduct the grid search correctly, we try out hundreds of different parameter combinations ex-post. Results show that our optimal parameter combination produces one of the best results, and there are many parameter combinations that produce similar results. Therefore, results are also robust to parameter selection.

[11] We use the randomForest package in R for estimating random forest and bagged decision trees and the gbm package in R for estimating stochastic gradient tree boosting. We also calculate permutation importance metrics for BDT and RF using the randomForest package in Sect. 4.3. For GBM, we use our

Tables 1 and 2 present RMSEs and MAEs of both machine learning and benchmark models for each prediction i calculated according to Eqs. (11) and (12), respectively. RW refers to the random walk model. 2FLM and 20FLM represent the linear dynamic factor models presented in Eqs. (4) and (6), respectively. BDT, RF, and GBM refer to bagged decision trees, random forests, and stochastic gradient tree boosting, respectively.

In the first predictions for the reference quarter, 20FLM has the highest forecasting performance in terms of RMSE. However, BDT and RF have the best nowcasting performance in terms of MAE. These results indicate that machine learning models are producing volatile nowcasts when there is little information available for the reference quarter. Starting from the second predictions for the reference quarter, all machine learning models consistently outperform competing models. Among machine learning models, the highest nowcasting performance is given by RF, followed in turn by BDT and GBM. The nowcasting performance of machine learning models improves significantly as more information becomes available for the target reference quarter.

In Tables 1 and 2, the last row presents the average RMSEs and MAEs, respectively, of the models. The average RMSE of RF is approximately 28% lower than those of 2FLM and RW. Furthermore, the average RMSE of RF is 5.8% lower than those of 20FLM. BDT and GBM also have lower average RMSEs than all competing models. In terms of MAEs, BDT and RF nowcast QoQ GDP growth rates by nearly 0.18 percentage points better than 2FLM and 0.02 percentage points better than 20FLM on average.

In Tables 1 and 2, we present the prediction performance of the models for the whole period. However, it can be seen from studies in the literature that the forecasting performance of models is usually unstable, and the ranking of models can change over time (see, e.g., Stock and Watson 2003, 2004; Kuzin et al. 2013) Therefore, we present the prediction performance of 2FLM, 20FLM, BDT, RF, and GBM[12] by calculating five years rolling windows of RMSEs in Figure 1 as follows:[13]

$$RMSE_{t,i} = ((1/n) \sum_{t_q=2000Q2+t}^{2005Q1+t} (y_{t_q} - \hat{y}_{t_q}^{(i)})^2)^{1/2}; \quad t = 0,1,2,...,T \quad \& \quad i = 1,2,...,6.$$

(13)

In the first predictions for the reference quarter, the RMSEs of the machine learning models and 20FLM move very closely together until August 2008. After prediction errors due to the great financial crisis, the RMSEs of all the models start

---

Footnote 11 continued

own code for calculating the out of bag permutation importance metric as it is not available in the package. If not otherwise stated, default parameters in packages are not altered.

[12] We omit RW for brevity. Results for RW are available upon request.

[13] We also perform same type of calculations using MAEs as follows: $MAE_{t,i} = (1/n) \sum_{t_q=2000Q2+t}^{2005Q1+t} |y_{t_q} - \hat{y}_{t_q}^{(i)}|; \quad t = 0,1,2,...,T \quad \& \quad i = 1,2,...,6.$ However, we put these results on Appendix 2 as they show similar results.

Table 1 RMSEs of the models estimated with an expanding window for successive nowcasting horizons between 2000Q2 and 2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
First prediction	0.626	0.667	0.506	0.535	0.510	0.538
Second prediction	0.679	0.672	0.499	0.496	0.482	0.498
Third prediction	0.699	0.654	0.493	0.476	0.465	0.487
Fourth prediction	0.507	0.600	0.463	0.430	0.423	0.444
Fifth prediction	0.517	0.533	0.405	0.375	0.361	0.399
Sixth prediction	0.538	0.457	0.367	0.337	0.334	0.358
Average	0.594	0.597	0.455	0.442	0.429	0.454

RW, random walk model; 2FLM, linear dynamic factor model using 2 factors; 20FLM, linear dynamic factor model using 20 factors; BDT, bagged decision trees; RF, random forests; GBM stochastic gradient tree boosting

Table 2 MAEs of the models estimated with an expanding window for successive nowcasting horizons between 2000Q2 and 2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
First prediction	0.446	0.558	0.376	0.367	0.363	0.391
Second prediction	0.489	0.566	0.366	0.356	0.352	0.374
Third prediction	0.499	0.545	0.367	0.343	0.342	0.363
Fourth prediction	0.389	0.475	0.348	0.321	0.322	0.333
Fifth prediction	0.392	0.425	0.320	0.287	0.284	0.303
Sixth prediction	0.405	0.379	0.287	0.269	0.268	0.286
Average	0.437	0.491	0.344	0.324	0.322	0.342

For abbreviations see Table 1

to increase. In the first predictions between 2009 and 2014, 20FLM gives higher prediction performance than the machine learning models, with RF being the second-best model. This indicates that 20FLM produces more accurate forecasts than other models during the financial crisis. When the crisis period is dropped from the RMSE calculations at the end of 2014, RF and BDT perform better than 20FLM. Furthermore in the last two years of the sample period, all of the machine learning models outperform 20FLM. Except for a short period between 2010 and 2012, 2FLM is the worst model. In the second and third predictions for the reference quarter, the results are very similar to those for the first predictions. The main differences are as follows: 20FLM performs much worse than the machine learning models until August 2008, and 2FLM is the worst model for the whole sample.

In the fourth, fifth, and sixth predictions for the reference quarter, new patterns began to emerge. We predict the current quarter in these predictions. In the fourth predictions for the reference quarter, RF and BDT outperform 20FLM in most of the sample. RF even has higher nowcasting power between 2009 and 2014. Before 2008
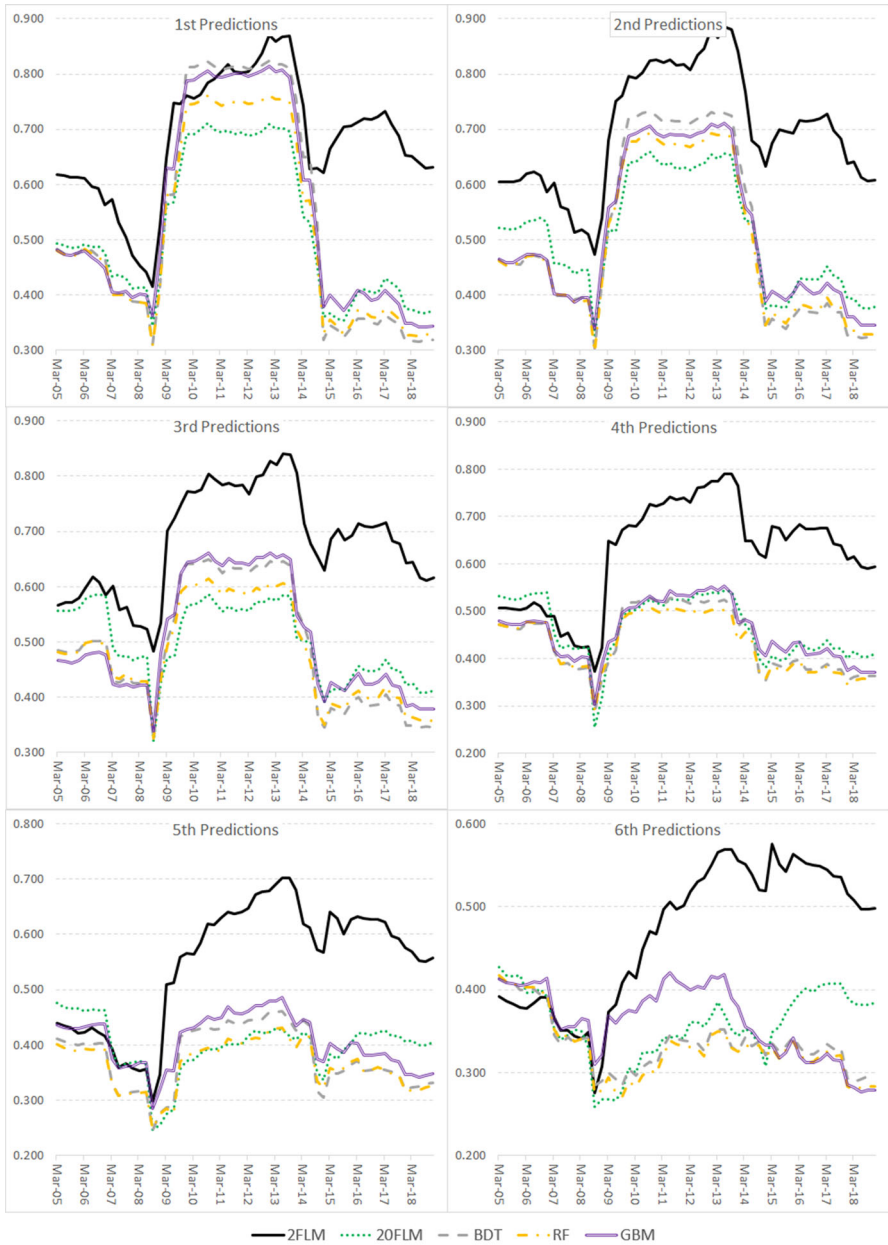
Fig. 1 Five-year rolling average RMSEs of the models estimated with an expanding window for successive nowcasting horizons between 2005Q1 and 2018Q4. For abbreviations see Table 1

and after 2016, all of the machine learning models are able to beat 20FLM. In the fifth predictions for the reference quarter, BDT and RF cannot outperform 20FLM between 2009 and 2014, but the nowcasting performance of RF is very similar to that of 20FLM. In all other periods, the machine learning models generally beat 20FLM. In the sixth predictions for the reference quarter, RF and BDT perform similarly to 20FLM or outperform it slightly until the end of 2014. Interestingly, after 2014, 20FLM performs much worse than the machine learning models. In the fourth, fifth, and sixth predictions, even though 2FLM performs pretty decently until the financial crisis, its prediction performance deteriorates rapidly after the crisis.

In the presence of instabilities induced by structural breaks, rolling window estimation can improve the forecasting performance of models compared to expanding window estimation. Tables 3 and 4 present RMSEs and MAEs of both machine learning and benchmark models estimated with a rolling window. The results show that rolling window estimation improves the prediction performance of both linear and machine learning models. On average, 2FLM enjoys the most significant improvement, followed by BDT, RF, and GBM. However, the prediction performance of 20FLM increases slightly. In Tables 3 and 4, RF has lower RMSEs and MAEs than all of the benchmark models. As in the case of expanding window estimation, BDT and RF outperform all other benchmark models starting from the second predictions for the reference quarter. The average RMSEs of RF and BDT are now 8.6% and 7.5% lower than that of 20FLM, and the average MAEs of RF and BDT are 0.033 percentage points better than that of 20FLM. Rolling window estimation appears to help in improving the nowcasting performance of machine learning models significantly compared to the performance of 20FLM.

Finally by calculating 5 years rolling windows of RMSE, we present the prediction performance of 2FLM, 20FLM, BDT, RF, and GBM estimated using a rolling window in Fig. 2. It seems that Figs. 1 and 2 exhibit very similar results.

Table 3 RMSEs of the models estimated with a rolling window for successive nowcasting horizons between 2000Q2 and 2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
First prediction	0.626	0.631	0.498	0.503	0.491	0.519
Second prediction	0.679	0.635	0.494	0.474	0.468	0.484
Third prediction	0.699	0.616	0.487	0.453	0.447	0.470
Fourth prediction	0.507	0.564	0.457	0.406	0.403	0.433
Fifth prediction	0.517	0.496	0.403	0.348	0.346	0.385
Sixth prediction	0.538	0.418	0.371	0.323	0.321	0.342
Average	0.594	0.560	0.452	0.418	0.413	0.439

For abbreviations see Table 1

Table 4 MAEs of the models estimated with a rolling window for successive nowcasting horizons between 2000Q2 and 2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
First prediction	0.446	0.516	0.369	0.353	0.350	0.370
Second prediction	0.489	0.524	0.361	0.338	0.336	0.353
Third prediction	0.499	0.505	0.362	0.331	0.330	0.344
Fourth prediction	0.389	0.442	0.348	0.307	0.305	0.325
Fifth prediction	0.392	0.391	0.316	0.268	0.273	0.289
Sixth prediction	0.405	0.343	0.288	0.252	0.254	0.267
Average	0.437	0.453	0.341	0.308	0.308	0.325

For abbreviations see Table 1

4.2 Tree-Based Machine Learning Models versus GDPNow

We compare our machine learning models against benchmark models and show that our proposed models beat those models. We also want to compare our models against a well-known nowcasting model watched by market participants to further show that our proposed models can be of value to a user. Probably the most well-documented and well-known nowcasting model for the US economy is Atlanta FED's GDPNow. GDPNow uses an indirect approach for nowcasting GDP by first predicting its subcomponents with linear factor models and Bayesian vector autoregression approaches and then aggregating them (Higgins 2014). Instead, we use a direct approach for nowcasting GDP.[14] As GDPNow only nowcasts the current quarter, we can only test the fourth, fifth, and sixth predictions of our models against GDPNow nowcasts. We obtain historical nowcasts of GDPNow between 2011Q3 and 2018Q4.[15] Table 5 presents RMSEs and MAEs of both machine learning models and GDPNow.[16] Table 6 shows the total number of reference quarters in which a machine learning model predicts actual GDP better than GDPNow together with N, the total number of reference quarters available for the ith prediction.[17]

For fourth predictions, the machine learning models outperform GDPNow in terms of MAEs. Furthermore, BDT, RF, and GDPNow perform very similarly in terms of RMSEs. Out of 29 quarters, RF and BDT beat GDPNow more than 50% of the time. Overall, BDT and RF yield better performance for fourth predictions compared to GDPNow. For fifth predictions, GDPNow outperforms machine learning models slightly in terms of both MAE and RMSE. GDPNow predicts actual

---

[14] An indirect nowcasting approach may also yield better results than a direct approach. However, comparing indirect and direct nowcasting approaches is not within the scope of this study.

[15] GDPNow has started producing nowcasts in the second half of 2011 (Higgins 2014).

[16] As we have already shown that rolling window estimation yields better results, we present results only for rolling window estimation. Results for expanding window estimation are available upon request.

[17] GDPNow did not produce the fourth prediction for the period of 2012Q2. Therefore, $N = 29$ for the fourth predictions and $N = 30$ for others.
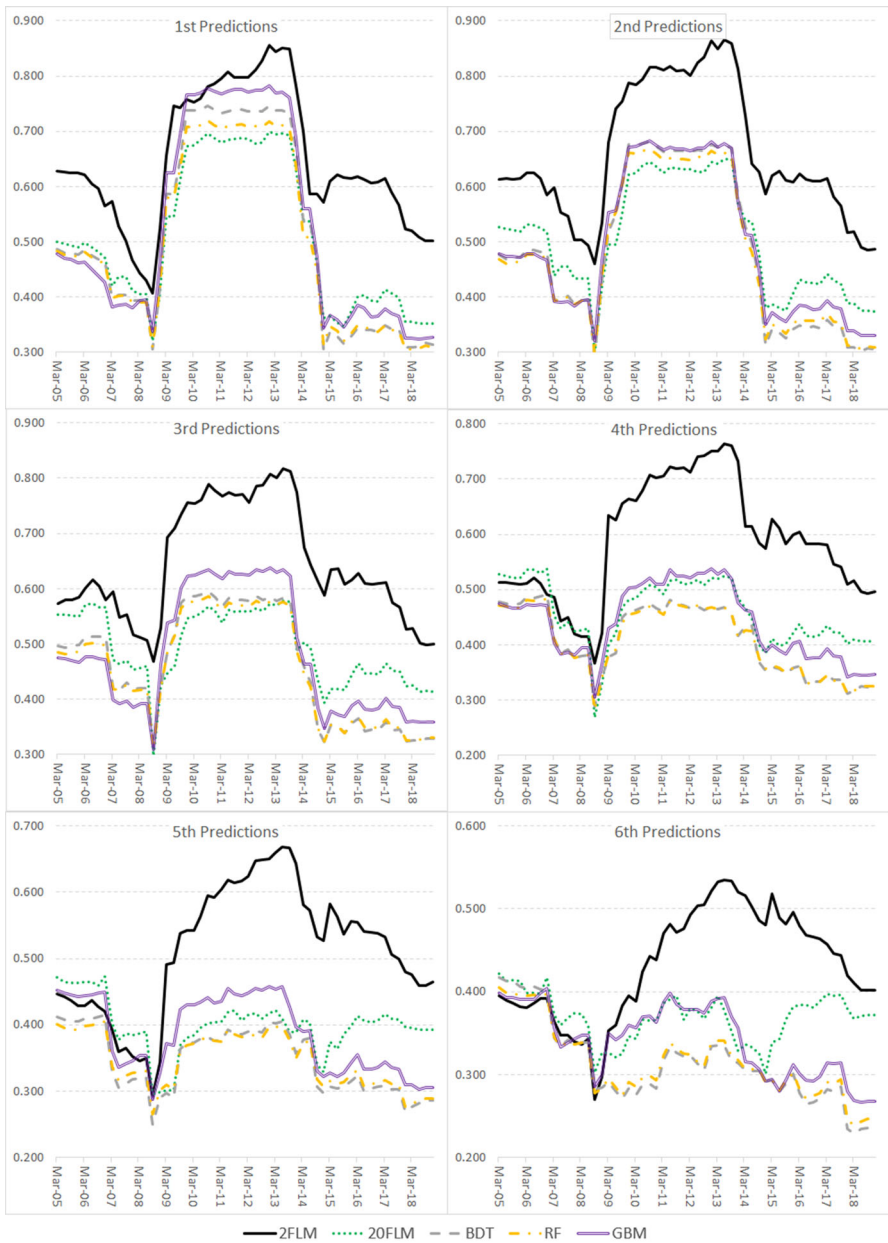
Fig. 2 5 years rolling average RMSEs of the models estimated with a rolling window for successive nowcasting horizons between 2005Q1 and 2018Q4. For abbreviations see Table 1

Table 5 MAEs and RMSEs of machine learning models and Atlanta FED's GDPNow for successive nowcasting horizons between 2011Q3 and 2018Q4

	RMSE				MAE			
	BDT	RF	GBM	GDPNow	BDT	RF	GBM	GDPNow
Fourth prediction	0.090	0.089	0.108	0.090	0.226	0.223	0.242	0.259
Fifth prediction	0.078	0.080	0.088	0.058	0.223	0.226	0.227	0.213
Sixth prediction	0.064	0.067	0.074	0.041	0.188	0.198	0.212	0.154
Average	0.077	0.078	0.090	0.063	0.212	0.216	0.227	0.208

For abbreviations see Table 1. Note that the machine learning models are estimated using a rolling window

Table 6 Total number of reference quarters in which a machine learning model predicts actual GDP better than GDPNow between 2011Q3 and 2018Q4

	BDT	RF	GBM	N
Fourth prediction	16	18	14	29
Fifth prediction	13	14	16	30
Sixth prediction	17	16	13	30
Total	46	48	43	89

For abbreviations see Table 1. N denotes the total number of reference quarters available for the ith prediction. Note that the machine learning models are estimated using a rolling window

QoQ GDP growth rates by 0.01 percentage points better than BDT according to the MAEs. Out of 30 quarters, GBM beats GDPNow 53% of the time, but RF and BDT only perform better 47% and 43% of cases, respectively. For sixth predictions, GDPNow also outperforms machine learning models in both metrics. However, BDT and RF perform better in 54% and 52% of the reference quarters, respectively.

When we take account of all prediction horizons, according to the average MAE, GDPNow nowcasts QoQ GDP growth rates by 0.004 percentage points better compared to BDT and 0.008 percentage points better compared to RF. However, BDT and RF outperform GDPNow in a greater number of reference quarters. Our results indicate that machine learning models can be an alternative to GDPNow for tracking the current state of the economy, especially at the start of the reference quarter.

4.3 Importance of Variables

In this subsection, we analyze which variables have more importance in predicting US GDP with tree-based ensemble models. We use permutation importance metrics to calculate the importance of a variable in our tree-based ensemble models. Permutation importance for a variable is calculated as follows

1. Calculate the prediction accuracy of a baseline model.
2. Re-calculate the prediction accuracy of the same model when the target variable is permutated.
3. Finally, take the difference between two prediction accuracies calculated in steps 1 and 2 to obtain the permutation importance of that particular variable.
4. Repeat steps 2 and 3 to calculate the permutation importance for all variables. We use the out-of-bag sample to calculate the prediction accuracy of a baseline model.

Figure 3 presents the average importance metric of dynamic factors for BDT, RF, and GBM estimated using both a rolling window and an expanding window. To calculate the average importance, we calculate permutation importance metrics for each model in each nowcasting period and then take the simple average of importance metrics calculated over all periods between January 2000 and December 2018. As a result, this produces an importance metric for each factor over the whole nowcasting period.

In all cases, the most important dynamic factors are obtained from real variables. For the output and income group, the first factor has the greatest influence on the models among all other factors, and the influence of the second factor from this group is mostly negligible. The first factor of the output and income group seems to capture nearly all information related to GDP, and the second factor does not contain any additional information. The first factor from the consumption, orders, and inventories group and the first factor from the labor market group are the second and third most important variables, respectively. For RF and BDT, the second factors of these two groups also have important influences on the models compared to factors derived from groups of financial variables. Interestingly, factors derived from financial variables and prices are mostly unimportant. Expanding window estimation and rolling window estimation mostly provide similar results. The most significant difference between them is that importance of the first factor from the output and income group is higher in expanding window estimation than in rolling window estimation, which indicates that output and income variables become less important in recent years.

Overall while groups of real variables constitute almost all important factors, output and income group being the most influential among them, financial and price variables carry very little information for our models. This seems to be an important result which demands further investigation. Therefore, additionally we analyze these results by calculating time-varying aggregate influence metrics of factors derived from real variables (the output and income; the consumption, orders, and inventories; the housing; and the labor market groups) versus factors derived from financial and price variable groups over the whole forecasting period. Let us define $\text{IMP}^r_{t_q,i}$ as the importance value of real factor i at time t_q and $\text{IMP}^{fp}_{t_q,l}$ as the importance value of financial and price factors l at time t_q. Then the aggregate importance of real factors, $\text{AGGIMP}^r_{t_q,i}$, and the aggregate importance of financial and price factors, $\text{AGGIMP}^{fp}_{t_q,l}$, can be calculated as follows:
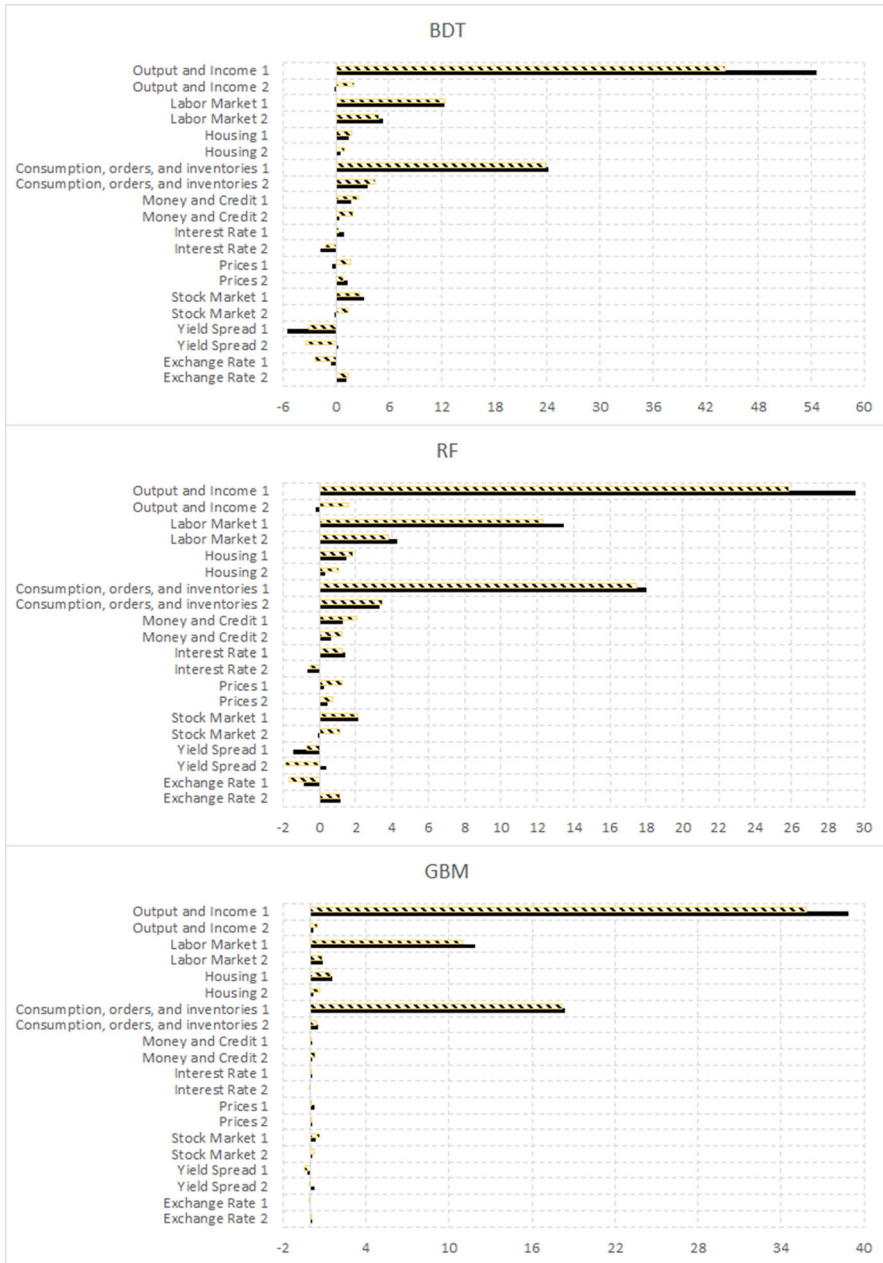
Fig. 3 Average permutation importance values of factors covering the period between January 2000 and December 2018. For abbreviations see Table 1
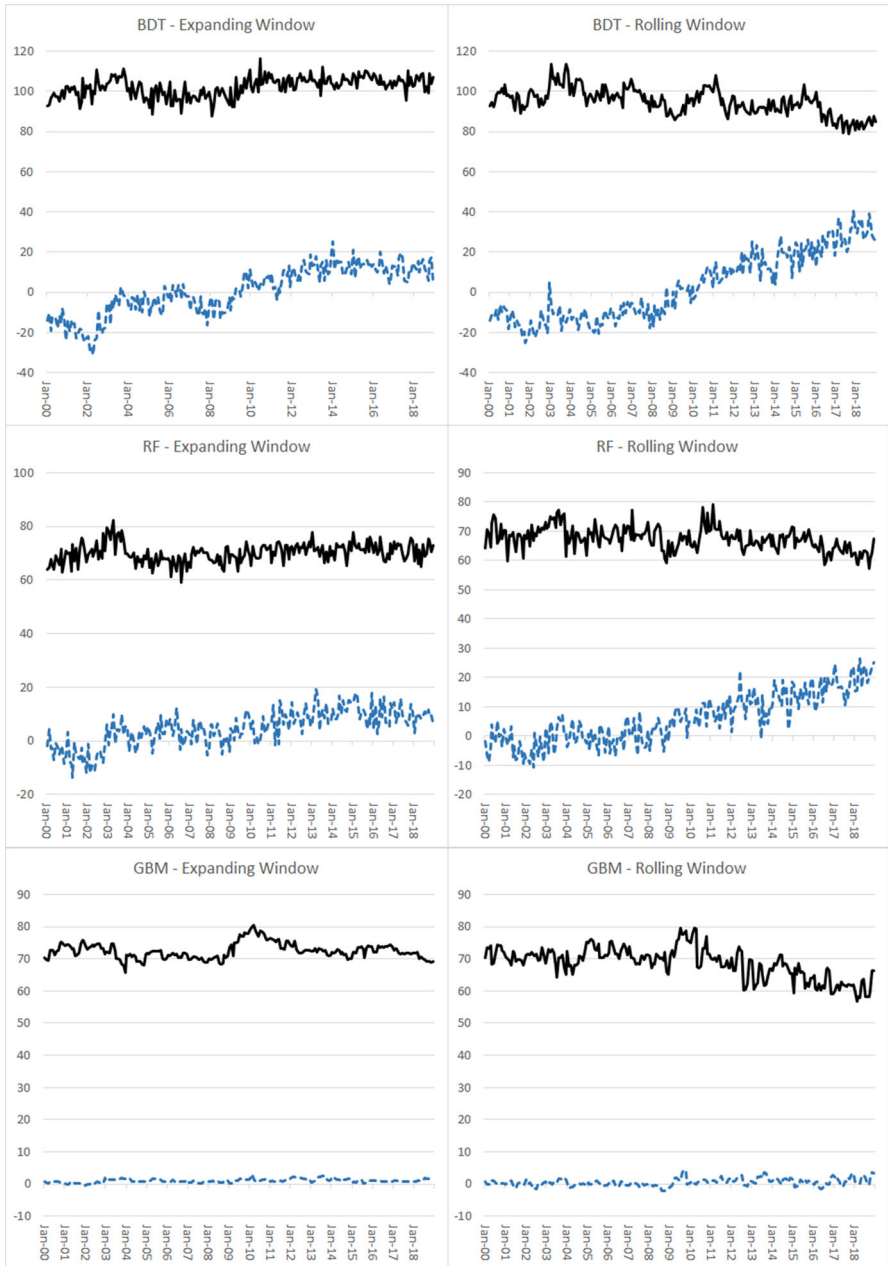
Fig. 4 Time-varying aggregate permutation importance values of real factors and financial and price factors between January 2000 and December 2018. For abbreviations see Table 1

$$\mathrm{AGGIMP}^{r}_{t_q,i} = \sum_{i=1}^{N^r} \mathrm{IMP}^{r}_{t_q,i}, \qquad i = 1, 2, \ldots, N^r, \tag{14}$$

$$\mathrm{AGGIMP}^{fp}_{t_q,i} = \sum_{i=1}^{N^{fp}} \mathrm{IMP}^{fp}_{t_q,i}, \qquad i = 1, 2, \ldots, N^{fp}, \tag{15}$$

where N^r and N^{fp} denote the number of real factors and the number of financial and price factors, respectively.

Figure 4 presents the time-varying aggregate importance of real factors against financial and price factors between January 2000 and December 2018. As expected, real factors dominate financial and price factors in all models. For BDT, the influence of financial and price factors is negative until the great financial crisis. Only in the period after the crisis, the influence of financial and price factors becomes positive. For BDT estimated using an expanding window, the aggregate importance metric of financial and price factors reaches 20% and then fluctuates around 10%. For BDT estimated using a rolling window, the aggregate importance metric of financial and price factors steadily increases to 40%, while the influence of real factors decreases slightly. For RF estimated using a rolling window, the aggregate importance metric of financial and price factors increases steadily after the great financial crisis. For RF estimated using an expanding window, the aggregate importance of financial and price factors increases slightly after the crisis and then fluctuates around 10%. For GBM estimated recursively, we do not see much difference in the relative importance of variables. For GBM estimated using a rolling window, the relative importance of real variables decreases gradually after 2010.

Overall, while the real variables appear to be dominant over the sample, especially prior to the great financial crisis, the importance of financial variables has significantly increased after the crisis. Apparently, ultra expansionary monetary policy measures to stimulate the economy, implemented by the FED following the crisis, have led the information content of financial variables become more important in predicting real GDP.

5 Conclusion

In this study, we used bagged decision trees, random forests, and stochastic gradient tree boosting to nowcast US GDP between January 2000 and December 2018. We used a large-scale data set containing more than 100 financial and macroeconomic variables. Instead of feeding this data set directly to machine learning models, we first extracted dynamic factors from 10 groups of financial and macroeconomic variables. Using a dynamic factor model as an intermediate step solved both the ragged data problem of nowcasting and reduces the dimension of the data set. We estimated our machine learning models using both a rolling window and an expanding window. Finally, we tested which variables are more influential for tree-based ensemble models.

Our results show that tree-based ensemble models beat linear models most of the time. The performance of machine learning models especially increases when more data for the reference quarter become available. Our results also indicate that random forests and bagged decision trees outperform linear models more significantly after the great financial crisis. We also show that tree-based ensemble models estimated with a rolling window have better nowcasting performance than models estimated recursively. A comparison of the results from our machine learning models with those from Atlanta FED's GDPNow, which is a well-known nowcasting model, shows that the machine learning models outperform GDPNow at the start of the reference quarter, but GDPNow performs better than our proposed models at the end of the reference quarter.

Finally, our results indicate that factors obtained from real variables have more impact on the models than factors obtained from financial or price variables, but the influence of factors extracted from financial and price variables increases after the great financial crisis. In this regard, the importance gained by financial variables in predicting the real GDP in the period following the crisis can be interpreted as a sign of effectiveness of the extra loose monetary policy implemented by the Fed in boosting economic growth. Repeating this study for the period after the recent Covid-19 crisis may constitute useful future research to see the effect of the current extra loose monetary policy.

Compliance with ethical standards

Appendix 1: Description of the Data Set

Group[a]	Description	Transformation[b]	Period
1	Real personal income	1	All
1	Real personal income excluding current transfer receipts	1	All
4	Real personal consumption expenditures (chain-type quantity index)	1	2003:12–
4	Real manufacturing and trade industries sales	1	All
4	Real retail and food services sales	1	All
1	Industrial production index	1	All
1	Industrial production: Final products and nonindustrial supplies	1	All
1	Industrial production: Final products (market group)	1	All
1	Industrial production: Consumer goods	1	All
1	Industrial Production: Durable consumer goods	1	2002:11–
1	Industrial production: Nondurable consumer goods	1	2002:11–
1	Industrial production: Business equipment	1	2002:11–

Group[a]	Description	Transformation[b]	Period
1	Industrial production: Materials	1	All
1	Industrial production: Durable materials	1	2002:11–
1	Industrial production: Nondurable materials	1	2002:11–
1	Industrial production: Manufacturing (SIC)	1	All
1	Industrial Production: Residential utilities	1	2002:11–
1	Industrial production: Fuels	1	2002:11–
1	Capacity utilization: Manufacturing (SIC)	2	All
2	Civilian labor force	1	All
2	Civilian employment level	1	All
2	Civilian unemployment rate	2	All
2	Average (mean) duration of unemployment	2	All
2	Number of civilians unemployed for less than 5 Weeks	1	All
2	Number of civilians unemployed for 5–14 weeks	1	All
2	Number of civilians unemployed for 15 weeks and over	1	All
2	Number of civilians unemployed for 15–26 weeks	1	All
2	Number of civilians unemployed for 27 weeks and over	1	All
2	Initial claims	1	All
2	All employees: Total nonfarm payrolls	1	All
2	All employees: Goods-producing industries	1	All
2	All employees: Mining and logging: Mining	1	All
2	All employees: Construction	1	All
2	All employees: Manufacturing	1	All
2	All employees: Durable goods	1	All
2	All employees: Nondurable goods	1	All
2	All employees: Service-providing industries	1	All
2	All employees: Trade, transportation and utilities	1	2003:05–
2	All employees: Wholesale trade	1	All
2	All employees: Retail trade	1	All
2	All employees: Financial activities	1	All
2	All employees: Government	1	All
2	Average weekly hours of production and nonsupervisory employees: Goods-producing	2	All
2	Average weekly overtime hours of production and nonsupervisory employees: Manufacturing	2	All
2	Average weekly hours of production and nonsupervisory employees: Manufacturing	2	All
3	Housing Starts: Total: new rivately owned housing units started	1	All
3	Housing starts in Northeast census region	1	All
3	Housing starts in Midwest census region	1	All
3	Housing starts in South census region	1	All
3	Housing starts in West census region	1	All
3	New private housing units authorized by building [ermits	1	All

Group[a]	Description	Transformation[b]	Period
3	New private housing units authorized by building permits in the Northeast census region	1	All
3	New private housing units authorized by building permits in the Midwest census region	1	All
3	New private housing units authorized by building permits in the South census region	1	All
3	New private housing units authorized by building permits in the West census region	1	All
4	Manufacturers' new orders: Durable goods	1	All
4	New orders for nondefense capital goods	1	All
4	Value of manufacturers' unfilled orders for durable goods industries	1	All
4	Total business inventories	1	All
4	Total business: Inventories to sales ratio	2	All
5	M1 money stock	1	All
5	M2 money stock	1	All
5	Real M2 money stock	1	All
5	St. Louis adjusted monetary base	1	All
5	Total reserves of depository institutions	1	All
5	Reserves of depository institutions. nonborrowed	1	All
5	Commercial and industrial loans. all commercial banks	1	All
5	Real estate loans. all commercial banks	1	All
5	Total nonrevolving credit owned and securitized. outstanding	1	All
5	Nonrevolving consumer credit to personal income	1	All
8	S&P 500	1	All
8	S&P 500 industries	1	All
8	S&P dividend yield	1	All
8	S&P PE ratio	1	All
6	Effective federal funds rate	2	All
6	3 Month AA financial commercial paper rate	2	All
6	3 Month treasury bill: Secondary market rate	2	All
6	6 Month treasury bill: Secondary market rate	2	All
6	1 Year treasury constant maturity rate	2	All
6	5 Year treasury constant maturity rate	2	All
6	10 Year treasury constant maturity rate	2	All
6	Moody's Seasoned Aaa Corporate Bond Yield	2	All
6	Moody's Seasoned Baa Corporate Bond Yield	2	All
9	3 Months commercial paper minus federal funds rate	0	All
9	3 Months treasury bill minus federal funds rate	0	All
9	6 Months treasury bill minus federal funds rate	0	All
9	1 year treasury constant maturity minus federal funds rate	0	All
9	5 years treasury constant maturity minus federal funds rate	0	All
9	10 years treasury constant maturity minus federal funds rate	0	All

Group[a]	Description	Transformation[b]	Period
9	Moody's seasoned Aaa corporate bond minus federal funds rate	0	All
9	Moody's seasoned baa corporate bond minus federal funds rate	0	All
10	Switzerland/U.S. foreign exchange rate	1	All
10	Japan/U.S. foreign exchange rate	1	All
10	U.S./U.K. foreign exchange rate	1	All
10	Canada/U.S. foreign exchange rate	1	All
7	Producer price index by commodity for finished goods	3	–2016:01
7	Producer price Index by commodity for finished consumer goods	3	–2016:01
7	Producer price index by commodity intermediate materials: Supplies and components	3	–2016:01
7	Producer price index by commodity for crude materials for further processing	3	–2016:01
7	Crude oil prices: West texas intermediate (WTI) – cushing, Oklahoma	3	All
7	Producer price index by commodity metals and metal products: Primary nonferrous metals	3	All
7	Consumer price index for all urban consumers: All Items	3	All
7	Consumer price index for all urban consumers: Apparel	3	All
7	Consumer price index for all urban consumers: Transportation	3	All
7	Consumer price index for all urban consumers: Medical care	3	All
7	Consumer price index for all urban consumers: Commodities	3	All
7	Consumer price index for all urban consumers: Durables	3	–2014:11
7	Consumer price index for all urban consumers: Services	3	All
7	Consumer price index for all urban consumers: All items less food	3	All
7	Consumer price index for all urban consumers: All items less shelter	3	All
7	Consumer price index for all urban consumers: All items less medical care	3	All
7	Personal consumption expenditures: Chain-type price index	3	2000:07–
7	Personal consumption expenditures: Durable goods (chain-type price index)	3	2000:07–
7	Personal consumption expenditures: Nondurable goods (chain-type price index)	3	2000:07–
7	Personal consumption expenditures: Services (chain-type price index)	3	2000:07–
2	Average hourly earnings of production and nonsupervisory employees: Goods-Producing	3	All
2	Average hourly earnings of production and nonsupervisory employees: Construction	3	All
2	Average hourly earnings of production and nonsupervisory employees: Manufacturing	3	All
5	MZM money stock	1	All
5		1	All

Group[a]	Description	Transformation[b]	Period
	Consumer motor vehicle loans owned by finance companies. outstanding		
5	Total consumer loans and leases owned and securitized by finance companies. outstanding	1	All
5	Securities in bank credit at all commercial banks	1	All
8	CBOE S&P 100 volatility index: VXO	1	All[c]
7	Producer price index by commodity for finished goods	3	2016:02–
7	Producer price index by commodity for finished consumer goods	3	2016:02-
7	Producer price index by commodity intermediate materials: Supplies and components	3	2016:02–
7	Producer price index by commodity for crude materials for further processing	3	2016:02–
7	Consumer price index for all urban consumers: durables	3	2014:12–

[a]This column shows the group of the variable: 1, output and income; 2, labor market; 3, housing; 4, consumption, orders, and inventories; 5, money and credit; 6, interest rate; 7, prices; 8, stock market; 9, yield spread; 10, exchange rate. [b]This column denotes the following data transformation for a series: 0, no transformation; 1, monthly growth rate; 2, monthly differences; 3, monthly differences of the yearly growth rate. [c] Except for the period between 2004:12 and 2005:07

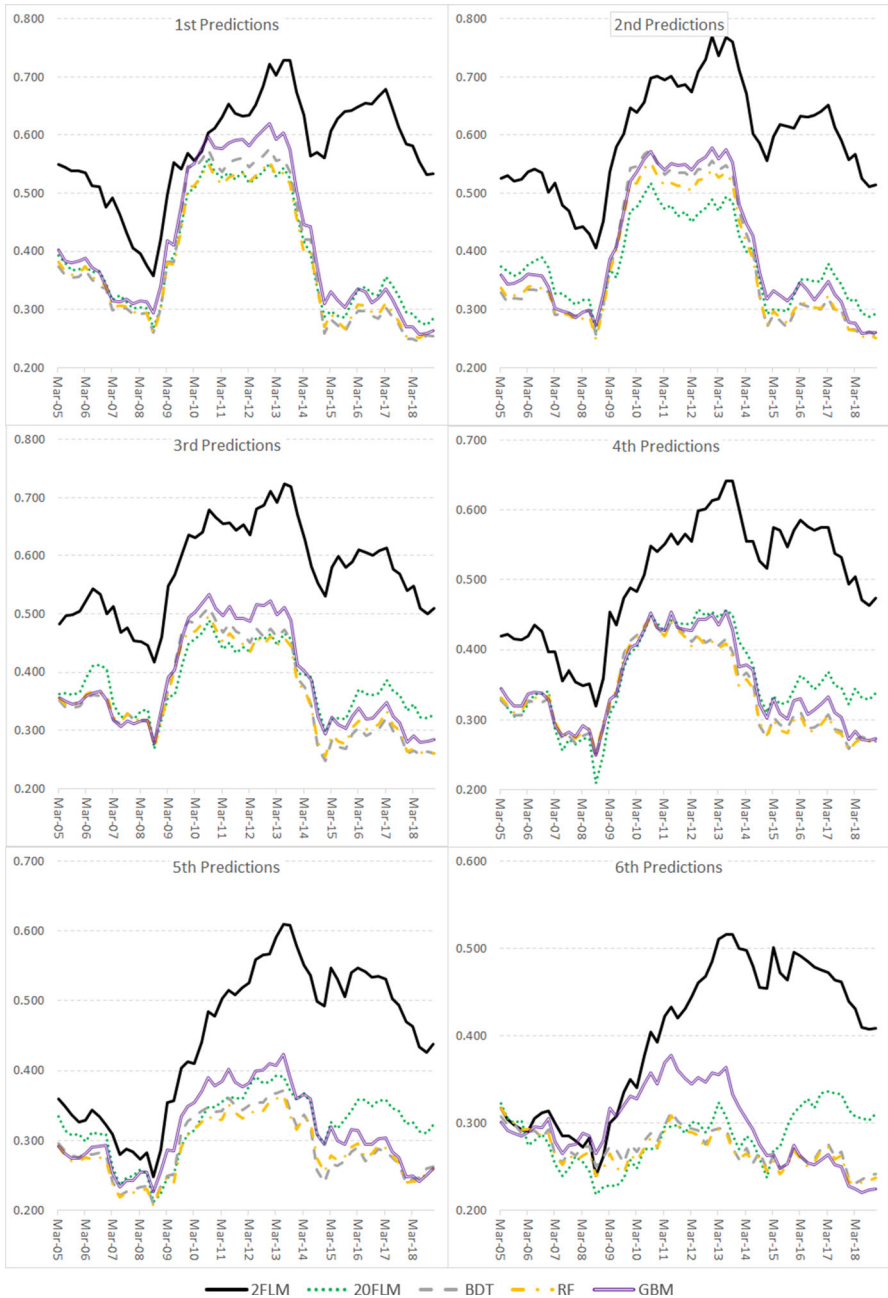## Appendix 2: Five-Year Rolling Averages of MAEs

See Figs. .

Fig. 5 Five-year rolling average MAEs of the models estimated with an expanding window for successive nowcasting horizons between 2005Q1 and 2018Q4. For abbreviations see Table 1
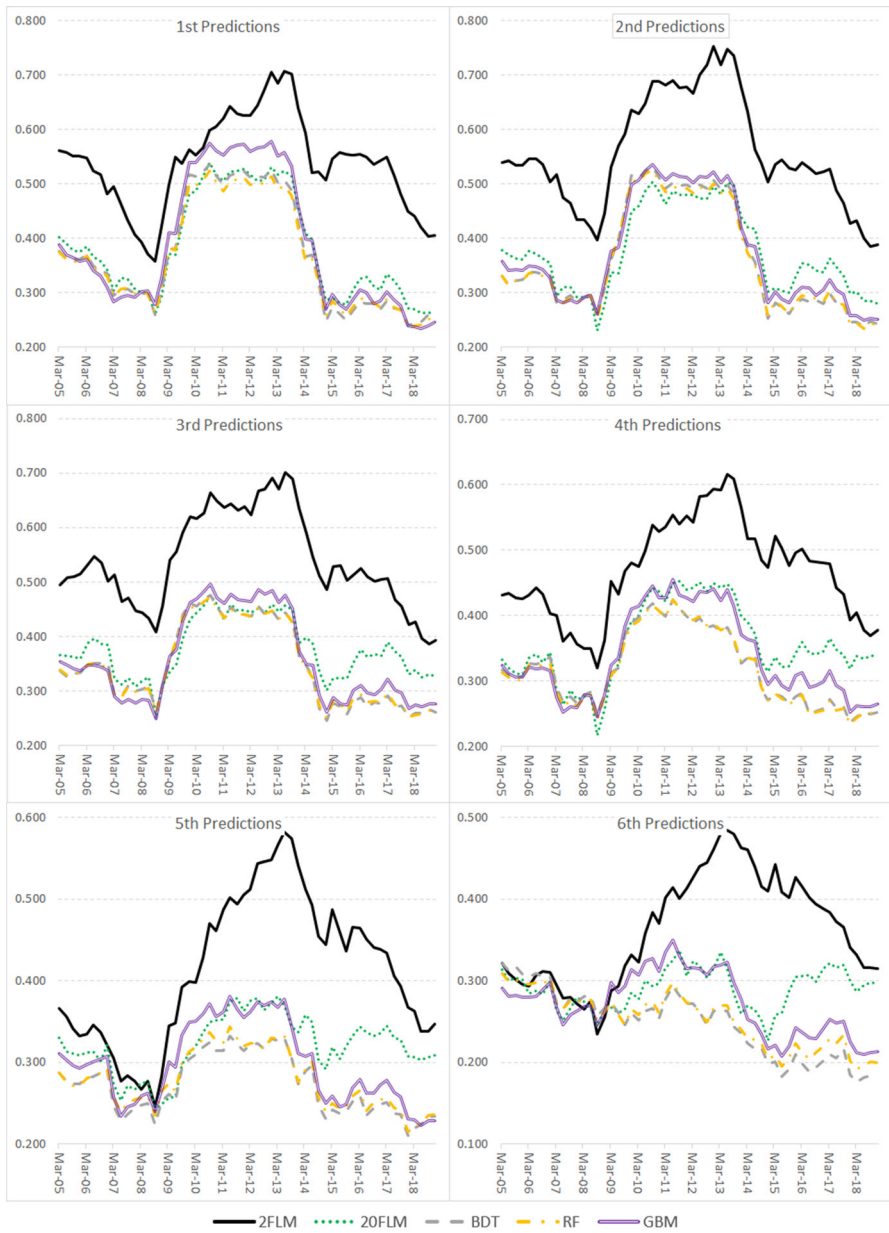
Fig. 6 Five-year rolling average MAEs of the models estimated with a rolling window for successive nowcasting horizons between 2005Q1 and 2018Q4. For abbreviations see Table 1

References

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bai, J., & Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In Handbook of Economic Forecasting (Vol. 2. (Ed.), *G. Elliott and A. Timmermann* (pp. 195–237). Amsterdam: Elsevier-North Holland.

Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, *27*(2), 333–346.

Barhoumi, K., Darné, O., & Ferrara, L. (2010). Are disaggregate data useful for factor analysis in forecasting french gdp? *Journal of Forecasting*, *29*(1–2), 132–144.

Biau, O., et al. (2010). *Euro area gdp forecast using large survey dataset-a random forest approach.* Mimeo: Technical report.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, *10*, 615–643.

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

Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and Regression Trees.* : Wadsworth.

Cornec, M., & Mikol, F. (2011). Nowcasting gdp directional change with an application to french business survey data. Technical report.

Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2019). How is machine learning useful for macroeconomic forecasting? Mimeo.

D'Agostino, A., McQuinn, K., & O'Brien, D. (2012). Nowcasting Irish GDP. *OECD Journal. Journal of Business Cycle Measurement and Analysis*, *2012*(1), 1–11.

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, *164*(1), 188–205.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367–378.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Newyork: Springer.

Higgins, P. C. (2014). Gdpnow: A model for gdp'nowcasting'. Frb atlanta working paper no. 2014-7.

Kuzin, V., Marcellino, M., & Schumacher, C. (2013). Pooling versus model selection for nowcasting gdp with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, *28*(3), 392–411.

Loermann, J., & Maas, B. (2019). Nowcasting US GDP with artificial neural networks. Munich personal repec archive working paper no. 95459.

Mariano, R. S., & Murasawa, Y. (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, *18*(4), 427–443.

Matheson, T. D. (2010). An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys. *Economic Modelling*, *27*(1), 304–314.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Richardson, A., Mulder, T., & Vehbi, T. (2018). Nowcasting New Zealand GDP using machine learning algorithms. Bank indonesia international workshop and seminar on "big data for central bank policies/building pathways for policy making with big data".

Soybilgen, B. (2020). Identifying US business cycle regimes using dynamic factors and neural network models. *Journal of Forecasting*, *39*(5), 827–840.

Stock, J., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.