Computational Economics (2021) 57:247–265
https://doi.org/10.1007/s10614-020-10054-w

Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach

Jaehyun Yoon[1]

Accepted: 20 September 2020 / Published online: 8 October 2020

Abstract

This paper presents a method for creating machine learning models, specifically a gradient boosting model and a random forest model, to forecast real GDP growth. This study focuses on the real GDP growth of Japan and produces forecasts for the years from 2001 to 2018. The forecasts by the International Monetary Fund and Bank of Japan are used as benchmarks. To improve out-of-sample prediction, the cross-validation process, which is designed to choose the optimal hyperparameters, is used. The accuracy of the forecast is measured by mean absolute percentage error and root squared mean error. The results of this paper show that for the 2001–2018 period, the forecasts by the gradient boosting model and random forest model are more accurate than the benchmark forecasts. Between the gradient boosting and random forest models, the gradient boosting model turns out to be more accurate. This study encourages increasing the use of machine learning models in macroeconomic forecasting.

Keywords Macroeconomic forecast · Random forest · Gradient boosting · Machine learning · Real GDP growth

1 Introduction

The ability to forecast macroeconomic variables is highly desirable for the design and implementation of timely policy measures. Among the macroeconomic variables, real GDP growth is one of the most important data. However, forecasting real GDP growth involves complicated calculations, and official data are often available only after at least a one-quarter delay. Due to this delay, policymakers often design and implement policies without knowing the necessary information.

✉ Jaehyun Yoon
j.yoon@aoni.waseda.jp

[1] Faculty of Political Science and Economics, Waseda University, 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan

From this point of view, if available, the accurate forecasting of real GDP growth in advance would be highly valuable.

Forecasting macroeconomic data, such as real GDP growth, is not a simple process. To forecast data, considering the causal relationship between the dependent variable and independent variable, traditional economic forecasting models require predetermined relevant variables to make predictions and often take top-down and theory-driven approaches (Mullainathan and Spiess 2017). This process also requires economic intuition and judgment by forecasters regarding the data and methods used. If there is any flaw in the assumptions made by the forecasters, the models could produce inaccurate predictions.

In contrast to many traditional economic forecasting models, machine learning models mostly deal with pure prediction (Varian 2014). Machine learning models are more flexible than traditional economic forecasting models and can produce predictions without predetermined assumptions or judgments. In fact, in conjunction with technological development and the increase in predictive power, machine learning models have been actively applied in various fields, from forecasting transportation flows to forecasting housing prices. In fact, machine learning methods often perform better than traditional econometric models, as shown by Plakandaras et al. (2015) in the case of forecasting US housing prices. In addition, machine learning models are applied to relatively low-frequency data sets and are shown to produce sound forecasts, as demonstrated in the studies on inflation forecasting by Medeiros et al. (2019) and Inoue and Kilian (2008).

With a focus on forecasting real GDP growth in Japan, this study presents forecasts with machine learning models, specifically a gradient boosting model and a random forest model, and compares their prediction accuracy against the benchmark forecast data published by the Bank of Japan (BOJ) and the International Monetary Fund (IMF) for the years from 2001 to 2018.

This study contributes to the literature in several points. First, this study provides a comparison of the performance of machine learning models on GDP predictions in Japan, which has not been analyzed and covered sufficiently. In specific, this study focuses on gradient boosting and random forest models, as these two models have received great attention due to their outstanding performance at numerous prediction competitions, such as those hosted by Kaggle, and there has also been high demand for comparisons of their forecasting performance. In the past, Biau and D'Elia (2010) used a random forest model to forecast the GDP data of the euro area and found that the machine learning model could produce more accurate predictions than the forecasts produced by a traditional autoregressive model. Jung et al. (2018) predicted real GDP growth in the United States, the United Kingdom, Germany, Spain, Mexico, the Philippines, and Vietnam using elastic net, recurrent neural network, and super-learner models. Tiffin (2016) employed elastic net and random forest models to forecast GDP growth in Lebanon, which provides official GDP growth data only after a 2-year delay. Emsia and Coskuner (2016) used support vector regression to predict the GDP growth of Turkey. However, the prediction of real GDP growth in Japan has not been sufficiently analyzed in the previous literature. Second, this study introduces a machine learning method that produces more accurate predictions of annual real GDP growth in Japan than the forecasts made by two prestigious institutions, the IMF and the BOJ, over a significant period. Lastly, this study presents a cross-validation and hyperparameter tuning process to address forecasting issues, such as overfitting problems, and provides the detailed parameters used in the prediction models, which can serve as a valuable reference for relevant research in the future.

2 Methodology

This study uses two machine learning models: gradient boosting and random forest models. All the models are supervised machine learning models, which means that the models perform analyses based on training data and construct a function to make predictions based on new data.

Using the data from the fourth quarter of 1981 to the second quarter of 2018, the machine learning models predict annual real GDP growth in Japan from 2001 to 2018. The machine learning models are designed to make predictions of annual real GDP growth based on data up to the second quarter of the focal year. For example, for 2001, the machine learning models train and fit their models using data up to the second quarter of that year. This means that the models do not use future data to predict past data.

The response variable for the models is the two-quarters-ahead real annual GDP growth. It should be noted that in cases where the two-quarters-ahead real annual GDP growth is predicted, the two-quarters-ahead real annual GDP is not available for the first quarter of the data; the two-quarters-ahead real annual GDP growth is available only when the third-quarter data are available. For example, for the prediction of the annual real GDP growth of 2002, the models make predictions with data up to the second quarter of 2002. However, the data set for the first quarter of 2002 cannot have two-quarters-ahead real annual GDP growth since the data become available only in the third quarter of 2002. The machine learning models make predictions of the two-quarters-ahead real annual GDP growth for the first-quarter data first. With the forecasted data in the first-quarter data, the models make final predictions of the second-quarter data and then predict the two-quarters-ahead real GDP growth.

All the machine learning algorithms used in this study are implemented with the Scikit-Learn package using Python language.

2.1 Machine Learning Models

2.1.1 Gradient Boosting

The gradient boosting model is an ensemble machine learning model introduced by Friedman (2001). The main idea of the gradient boosting model is to combine multiple weak learners to improve the accuracy and robustness of the final model.

The gradient boosting model starts by making a single leaf and building regression trees. A regression tree is a type of decision tree that is designed to estimate a continuous real-valued function instead of a classifier. The regression tree is constructed through an iterative process that continues to split the data into nodes or branches into smaller groups. Initially, all observations are placed in the same group. The data are then allocated into two partitions, using every possible split on every available predictor. The predictor that splits the tree is that which most clearly separates the observations into two distinct groups and minimizes the residual error, which, in this study, is measured by the Friedman MSE introduced in Friedman (2001).

Based on the error made by the previous tree, the gradient boosting model trains another tree, and it continues to make additional trees in this fashion until the designated number or fit cannot be improved. To avoid overfitting problems, the gradient boosting model uses a learning rate to scale the contribution from the new tree.

Based on Friedman (2001), the algorithm of the gradient boosting model takes the following steps for the input data, $\{(x_i, y_i)\}_{i=1}^{n}$, and a differentiable loss function, $L(y_i, F(x))$, which is a squared regression in this study.

Step 1: Initialize the model with a constant value:

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) \tag{1}$$

where y_i is an observed value, and γ is a predicted value. $F_0(x)$ is the average of the observed values.

Step 2: For m = 1 to M:

(A) Compute

$$\gamma_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad \text{for i} = 1, \ldots, \text{n} \tag{2}$$

(B) Fit a regression tree to the γ_{im} values and create terminal regions R_{jm} for j $= 1 \ldots J_m$

(C) For j $= 1 \ldots J_m$, compute

$$\gamma_{jm} = argmin_\gamma \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma) \tag{3}$$

(D) Update

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_m I(x \in R_{jm}) \tag{4}$$

where ν is the learning rate.

The loss functions used can be customized by setting the learning rate, ν. This feature improves the flexibility of this model while minimizing the overfitting problem by learning from the iterations performed at a slower rate (Hastie et al. 2009).

Step 3: Output:

$$\hat{F}(\mathrm{x}) = F_M(x) \tag{5}$$

After performing all M iterations and updating the $F_m(x)$ function, the final model, $\hat{F}(\mathrm{x})$, approximates the relationship between the independent variables and the dependent variable.

2.1.2 Random Forest

The random forest model, introduced by Breiman (2001), is another ensemble method similar to boosting models. According to Dietterich (2000), the random forest is one of the most successful ensemble models in machine learning. Similar to the gradient boosting model, the random forest model uses regression trees. However, unlike the gradient boosting model, in the random forest model, using bootstrapped data, the regression trees are trained independently, and the output of trees is averaged to produce predictions.

The basic steps of the random forest model are as follows:

Step 1. For m = 1 to M:

(1) Create a bootstrapped sample set, Z of size N, from the training data.
(2) Grow a random forest tree, T_m, for the bootstrapped data by repeating the following steps for each terminal node of the tree until the minimum node size, n_{min}, is reached.

   1. Select x variables at random from the p variables.
   2. Pick the best variable and split point among the x variables.
   3. Split the node into two daughter nodes. The split is decided in such a way that it minimizes MSE, which is calculated as follows:

$$F_0(x) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \gamma\right)^2 \tag{6}$$

   where y_i is an observed value and γ is a predicted value.

In addition to the bootstrapping unique data for each tree predictor, additional randomness is added at each node by randomly assigning a subset of variables to split the nodes. This random process greatly reduces the dependence between individual trees and improves flexibility against a potential overfitting problem. A fully developed tree often leads to an overfitting problem if it fits the model perfectly. In other words, a model with close to perfectly fitting trees may not produce accurate predictions when new data are added. To avoid this problem, a random forest model may prune the trees or limit the number of nodes at the expense of the in-sample fit.

Step 2. Output the ensemble of trees, $\left\{T_m\right\}_{m=1}^{M}$:

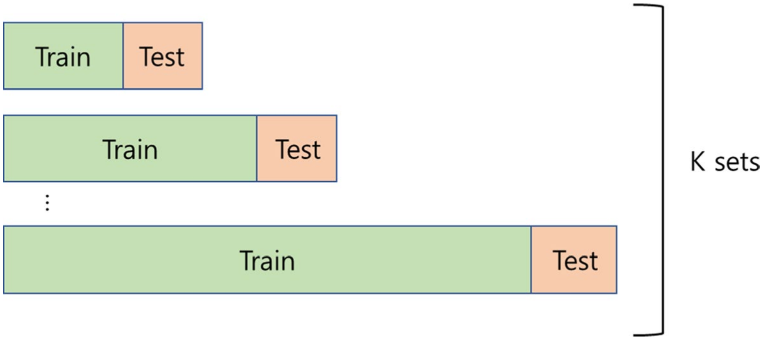

Fig. 1 Cross-validation process

$$\hat{F}_{rf}^{M}(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x). \tag{7}$$

The final output, $\hat{F}_{rf}^{M}(x)$, is calculated by averaging the outputs of all the trees. Averaging over multiple predictions reduces the variance and stabilizes the trees' predictive performance.

2.2 Cross-Validation

The machine learning models used in this study utilize several hyperparameters. This study uses k-fold cross-validation, which is a popular technique for tuning hyperparameters. The k-fold cross-validation separates the training data into k-pieces and separately tests each piece to fit the model. Due to the temporal dependencies among the data, the k-fold cross-validation is designed to set the first k-folds as the training set and the data after the folds as the test set. This ensures that future data are not used to test past data since the forecasting model should exclude all data about events that occur chronologically after the events used to fit the model (Tashman 2000). In this study, following previous literature, including Molinaro et al. (2005), k is set to 10, and the training data are set to 10 subsets to train and fit the model. Figure 1 illustrates the concept behind the cross-validation process used in this study.

Some may argue that the cross-validation process is not needed for the random forest models as the random forest models use trees created from the bagging process. It is true that the out-of-bag process of the random forest model is similar to the cross-validation process, and the cross-validation may not be needed. However, one of the main aims of this study is to compare the performance of the gradient boosting model and the random forest model. To make this comparison as fair as possible, this study applies the cross-validation process to the random forest model. In addition, the out-of-sample data are set to be the same for both the gradient boosting model and the random forest model to ensure a fair comparison of the models.

Table 1 Description of the hyperparameter test

Machine learning model	Hyperparameter	Hyperparameter test set
Gradient boosting	No. of boosting stages	100, 500, 1000
	Learning rate	0.0001, 0.001, 0.01, 0.1, 0.3
	Max. depth of the tree	1, 3, 5, 7, 9, 11, 13, 15, 17, 19
Random forest	No. of trees	100, 500, 1000
	Max. depth of the tree	1, 3, 5, 7, 9, 11, 13, 15, 17, 19

The cross-validation process is designed to select an optimal set of hyperparameters that produces the lowest average mean squared errors based on the tests of 10 subsets. In other words, the set of hyperparameters suggested by the cross-validation process will be used to make forecasts based on the test data set. The hyperparameter tuning strategy that this study uses is grid search, in which all possible combinations of the hyperparameters given are tested (Probst et al. 2019). Regarding the number of predictors, all predictors are considered, and the depth of trees is controlled with the number of splits for both the gradient boosting model and the random forest model. The cross-validation is designed to find a combination of the hyperparameters that minimizes the average of MSE. The hyperparameters determined by the cross-validation are presented in Table 1.

For the forecast of real GDP growth in year x, the same cross-validation will be conducted twice. The first process will be for the forecast of the two-quarters-ahead forecast of year-to-year real GDP growth in the first-quarter data. The second process will be repeated for the second quarter of the data for the final forecast.

2.3 Data

The prediction models use the traditional economic indicators related to national account, employment, monetary, trade, and inflation statistics as the regressors. The inflation variables include the consumer price index and GDP deflator. The national account variables include real government consumption, real private consumption, current account of balance of payments, real annual GDP, real GDP growth (quarter to quarter), real GDP growth (year over year), government balance as share of GDP, gross government debt as percent of GDP, foreign exchange reserves, real stockbuilding, total external debt, and foreign direct investment. The employment variables include total employment and the unemployment rate. The monetary variables include exchange rate against US dollar, exchange rate against euro, and 10-year government bond yields.

The data are chosen based on the availability and previous literature, including Jung et al. (2018) and Richardson et al. (2018). All variables are quarterly data from the fourth quarter of 1981 to the second quarter of 2018. In this study, real GDP growth (year over year) is set as the dependent variable, and other variables are set as independent variables. The number of observations is 147 for each variable. More details on the variables are available in Table 6 in “Appendix”.

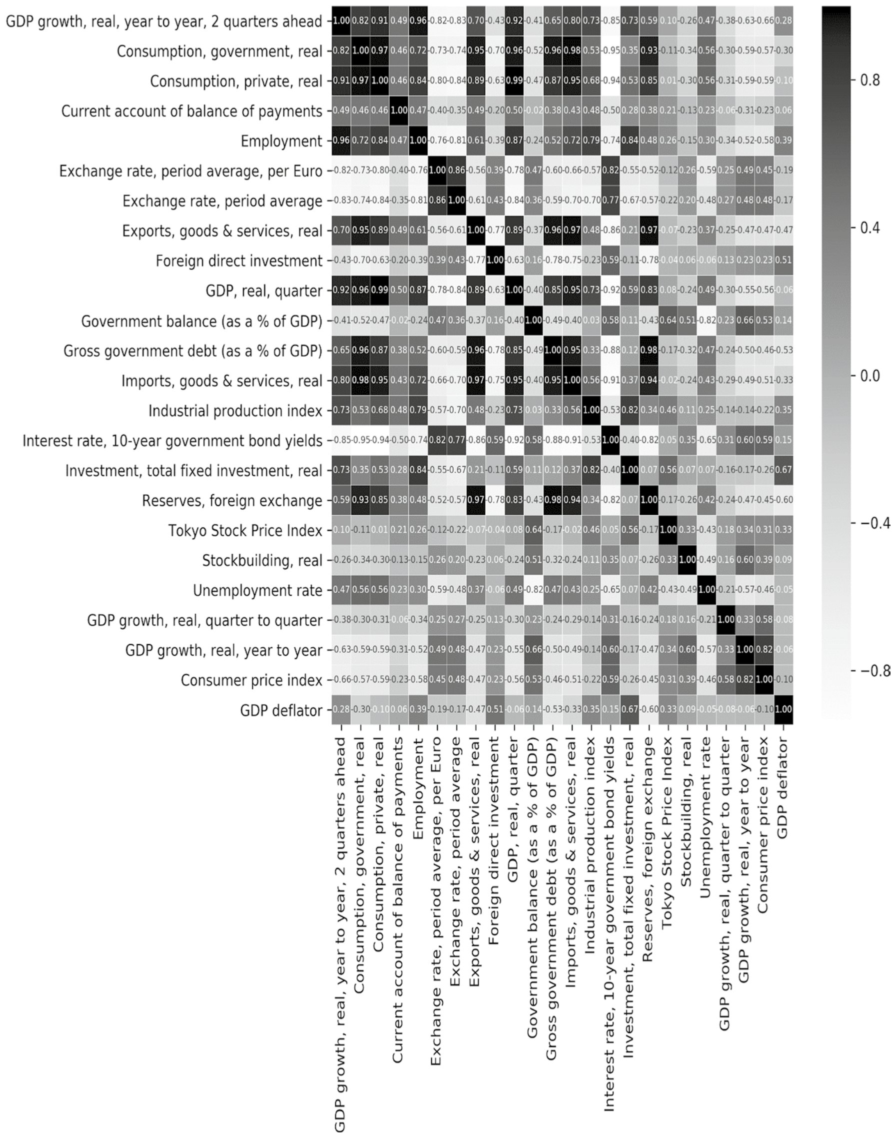

Fig. 2 Correlation matrix of the variables

Figure 2 presents the matrix for the correlation between the dependent variable and the independent variables for the two-quarters-ahead forecast. According to the correlation matrix, correlations among the regressors can be observed. Multiple regressors have correlations with other regressors above 0.5. For traditional linear regression models that focus on the interpretation of the impact of regressors, high correlations may lead to multicollinearity problems; however, ensemble models that focus on prediction, such as the gradient boosting and random forest models, are designed to handle multicollinearity problems using decision trees, which, instead of using all the predictors, choose certain regressors to maximize prediction accuracy and are robust to multicollinearity problems (Sandri and Zuccolotto 2008).

As benchmarks, this study uses the forecast data published by the IMF and the BOJ from 2001 to 2018 to check the performance of the machine learning models used in this study. Although the details of the previous forecast models used by the IMF and BOJ are not available to the public, the results of their models are used as benchmarks in this study, as they are widely accepted and quoted in both the public and private sectors. The forecasts that the IMF and the BOJ publish are the main forecasts on annual real GDP growth in spring and fall. In the case of the BOJ, this study uses the median values of the forecasts by the majority of policy board members.

3 Results

This study presents a method for forecasting the annual real GDP growth[1] of Japan from 2001 to 2018. The machine learning models produce predictions of annual real GDP growth in Japan for each year from 2001 to 2018, using data from the fourth quarter of 1981 up to the second quarter of the year of prediction. For example, for the prediction of annual real GDP growth in Japan in 2018, the machine learning models use data up to the second quarter of 2018.

Table 2 presents the hyperparameters used by the machine learning models, which are selected by the cross-validation process. As Table 2 shows, since the training data receive new data for each new year, the hyperparameters change accordingly to adjust to the new data set.

For benchmark points, this study uses the forecast data generated by the IMF and the BOJ. The IMF and the BOJ provide annual real GDP growth biannually: once in the spring and again in the autumn. Table 3 presents the forecasted real GDP growth of Japan, including that from the machine learning models, the IMF and the BOJ, along with the actual real GDP growth.

Figures 3 and 4 present the graphs of the actual real GDP growth of Japan and those forecasted by the machine learning models, the IMF, and the BOJ.

As shown in Table 3 and Figs. 3 and 4, the machine learning models produce forecasts that are overall more accurate than those produced by the IMF and the BOJ. However, for 2009, a year during which the global economic crisis was ongoing, the machine learning models do not predict the extreme drop in real GDP growth, which was forecasted by the IMF. The actual real GDP growth in 2009 is − 5.42%. The rates forecasted by the gradient boosting and random forest models are − 3.97% and − 2.32%, respectively. The rate by forecasted by the IMF is − 5.37%.

[1] The data on the annual real GDP growth of Japan refer to those published by the World Bank and are obtained from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=JP.

Table 2 Hyperparameters by year

Year	GB			RF	
	Max. depth of the tree	No. of boosting stages	Learning rate	Max. depth of the tree	No. of trees
2001	1	100	0.1	5	1000
2002	1	1000	0.1	5	100
2003	1	500	0.3	9	100
2004	3	100	0.1	9	100
2005	3	100	0.1	9	100
2006	3	100	0.1	5	100
2007	3	500	0.1	5	500
2008	1	1000	0.01	11	100
2009	1	500	0.1	3	1000
2010	1	1000	0.01	7	500
2011	3	500	0.1	9	500
2012	9	1000	0.01	11	500
2013	5	500	0.3	7	500
2014	1	1000	0.1	9	500
2015	1	500	0.1	11	100
2016	1	1000	0.01	7	500
2017	7	100	0.1	9	100
2018	1	500	0.1	7	500

GB and RF are gradient boosting and random forest, respectively

To compare forecast accuracy, mean absolute percentage errors (MAPEs) and root mean squared errors (RMSEs) are calculated for each model and compared. MAPE is a measure strongly preferred and frequently used by both practitioners and academics to assess the accuracy of forecasting models. MAPE is calculated by using the following formula.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{O_i - P_i}{O_i} \right|. \tag{8}$$

P is a predicted value, and O is an observed value. n is the total number of observations. Table 4 presents MAPEs for the machine learning models of this study, the gradient boosting and random forest models, and for the forecasts by the IMF and the BOJ in the spring and fall for each year from 2001 to 2018.

According to Table 4, for the 2001–2018 period, the gradient boosting model appears to have more predictive power than the random forest model. In addition, both the gradient boosting (19.86%) and random forest (67.23%) forecasts are shown to be more accurate than the IMF (103.31% in fall and 183.97% in spring) and the BOJ (86.69% in fall and 111.96% in fall) forecasts. Some may question whether the

Table 3 Actual and predicted real GDP growth of Japan (%)

Year	Actual real GDP growth	Forecasted real GDP growth					
		GB	RF	IMF_F	IMF_S	BOJ_F	BOJ_S
2001	0.41	0.55	0.50	− 0.48	0.60	− 1.05	0.55
2002	0.12	− 0.12	0.04	− 0.55	− 0.97	0.35	− 0.20
2003	1.53	1.43	1.51	1.98	0.82	2.40	1.00
2004	2.20	2.18	2.14	4.42	3.35	3.60	3.10
2005	1.66	1.66	1.67	1.96	0.81	2.20	1.30
2006	1.42	1.38	1.41	2.67	2.80	2.40	2.20
2007	1.65	1.66	1.60	1.95	2.34	1.80	2.10
2008	− 1.09	− 0.96	− 0.85	0.69	1.43	0.10	1.50
2009	− 5.42	− 3.97	− 2.32	− 5.37	− 6.20	− 3.20	− 3.10
2010	4.19	4.27	4.66	2.82	1.90	2.10	1.80
2011	− 0.12	− 0.15	− 1.20	− 0.47	1.40	0.30	0.60
2012	1.50	1.51	1.50	2.22	2.04	1.50	2.30
2013	2.00	2.06	2.05	1.95	1.58	2.70	2.90
2014	0.38	0.28	0.15	0.89	1.35	0.50	1.10
2015	1.22	1.23	1.28	0.59	1.04	1.20	2.00
2016	0.61	0.63	0.62	0.51	0.49	1.00	1.20
2017	1.93	1.96	1.96	1.51	1.25	1.90	1.60
2018	0.79	0.77	0.88	1.14	1.21	1.40	1.60

GB and RF are gradient boosting and random forest, respectively, and represent the forecasts based on the out-of-sample tests. IMF_F and IMF_S are the IMF forecasts made in fall and spring, respectively. BOJ_F and BOJ_S are the BOJ forecasts made in fall and spring, respectively

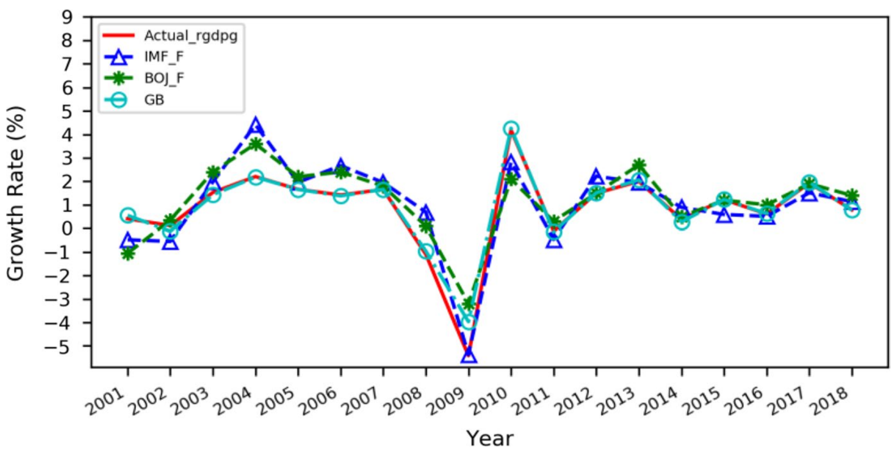

Fig. 3 Actual and predicted real GDP growth in Japan (including the gradient boosting model). *Actual_rgdpg is actual annual real GDP growth of Japan. IMF_F, BOJ_F, and GB are the forecasts made by the IMF in fall, the BOJ in fall and the gradient boosting model (out-of-sample tests)

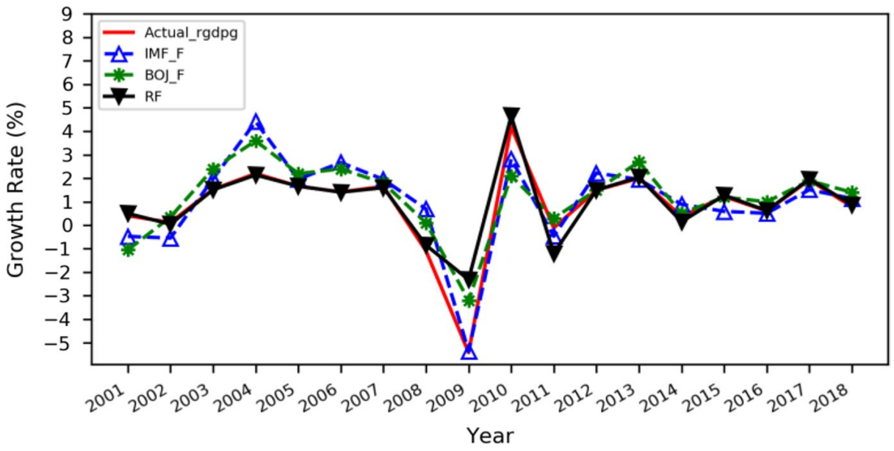

Fig. 4 Actual and predicted real GDP growth in Japan (including the random forest model). *Actual_rgdpg is actual annual real GDP growth of Japan. IMF_F, BOJ_F, and RF are the forecasts made by the IMF in fall, the BOJ in fall and the random forest model (out-of-sample tests)

Table 4 MAPEs for the machine learning models and forecasts by the IMF and the BOJ (%)

Period	GB	RF	IMF_F	IMF_S	BOJ_F	BOJ_S
2001	36.04	23.38	217.18	46.84	358.62	35.47
2001–2002	118.78	43.32	390.08	482.91	277.62	152.48
2001–2003	81.34	29.26	269.81	337.32	204.1	113.17
2001–2004	61.25	22.66	227.43	265.99	168.89	95.03
2001–2005	49.08	18.18	185.51	223.1	141.57	80.39
2001–2006	41.35	15.29	169.28	202.08	129.48	76.14
2001–2007	35.5	13.57	147.68	179.13	112.24	69.12
2001–2008	32.64	14.66	149.62	185.61	111.85	90.12
2001–2009	31.99	19.38	133.09	166.59	103.97	84.86
2001–2010	28.97	18.55	123.04	155.41	98.57	82.08
2001–2011	29.03	102.6	139.73	260.91	122.41	131.14
2001–2012	26.72	94.08	132.15	242.2	112.24	124.7
2001–2013	24.9	87.02	122.16	225.16	106.3	118.57
2001–2014	24.99	85.06	123.27	227.66	101.09	123.91
2001–2015	23.36	79.68	118.5	213.46	94.47	119.88
2001–2016	22.11	74.84	112.11	201.38	92.58	118.45
2001–2017	20.91	70.54	106.78	191.62	87.22	112.49
2001–2018	**19.86**	**67.23**	**103.31**	**183.97**	**86.69**	**111.96**

GB and RF are gradient boosting and random forest, respectively. The first two columns indicate the results from the out-of-sample tests. IMF_F and IMF_S are the IMF forecasts made in fall and spring, respectively. BOJ_F and BOJ_S are the BOJ forecasts made in fall and spring, respectively in-sample forecast models could overfit, and there could be an overfitting problem. Regarding the overfitting problem, Table 7 in "Appendix" presents MAPEs calculated for the in-sample tests. MAPEs for in-sample tests are calculated using the forecast values from the cross-validation process. The average values for the in-sample tests of the gradient boosting model (39.61%) and random forest model (73.13%) for the 2001–2018 period suggest the lack of overfit for the models. In addition, to avoid potential overfitting problems, this study adopts an expanding window method. This method creates a new model for each period using cross-validation and hyperparameter tuning and introduces a certain level of bias into the model to reduce variance. For example, if the model used for the annual real GDP growth of Japan in 2015 is used again for the prediction in 2017, the performance could be significantly low. However, the methodology used in this study creates a new model for the prediction in 2017 using the sample available up to the second quarter of 2017. As a result, the out-of-sample forecasts consistently outperform the those made by the IMF and the BOJ; the MAPEs for the out-of-sample forecasts are lower than those for the IMF and the BOJ. Based on the performance shown by the out-of-sample forecast models, the overfitting problem should not be significant.

RMSE is another measure that is popular among practitioners and academics for assessing the accuracy of forecasting models. RMSE measures the differences between observed and predicated values and is calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(P_i - O_i\right)^2}{n}}. \tag{9}$$

P is a predicted value, and O is an observed value. n is the total number of observations. In the formula, it should be noted that by squaring the difference between the predicted and observed values, the RMSE penalizes large errors.

Table 5 presents the RMSEs for the machine learning models of this study, the gradient boosting and random forest models, and the forecasts by the IMF and the BOJ in spring and fall for each year from 2001 to 2018.

Table 5 presents RMSEs for the predictions of the annual real GDP growth of Japan made by the machine learning models two quarters ahead. The RMSEs for all the machine learning models are lower than those of the IMF and the BOJ for the 2001–2018 period. Based on the RMSEs for the 2001–2018 period, the gradient boosting model appears to have more predictive power than the random forest model. Similar to MAPEs, Table 7 in "Appendix" presents the RMSEs for in-sample tests. The average values for the in-sample tests of the gradient boosting model (0.39) and random forest model (0.57) for the 2001–2018 period show that the overfit should not be significant.

Although the machine learning models made worse forecasts for the crisis year of 2009, both the MAPEs and the RMSEs of the machine learning models suggest that the machine learning models produce predictions that are more accurate than those produced by the IMF and the BOJ for multiple years, including the 2001–2018 period.

Table 5 RMSEs for the machine learning models and the forecasts by the IMF and the BOJ

Period	GB	RF	IMF_F	IMF_S	BOJ_F	BOJ_S
2001	0.15	0.09	0.88	0.19	1.46	0.14
2001–2002	0.20	0.09	0.78	0.78	1.04	0.25
2001–2003	0.17	0.07	0.69	0.75	0.99	0.37
2001–2004	0.15	0.07	1.26	0.87	1.10	0.55
2001–2005	0.13	0.06	1.13	0.87	1.02	0.52
2001–2006	0.12	0.06	1.15	0.97	1.01	0.57
2001–2007	0.11	0.06	1.07	0.94	0.94	0.55
2001–2008	0.12	0.10	1.19	1.25	0.97	1.05
2001–2009	0.50	1.04	1.12	1.21	1.18	1.26
2001–2010	0.47	0.99	1.15	1.36	1.30	1.41
2001–2011	0.45	1.00	1.10	1.37	1.24	1.36
2001–2012	0.43	0.96	1.07	1.32	1.19	1.33
2001–2013	0.41	0.92	1.03	1.28	1.16	1.30
2001–2014	0.40	0.89	1.00	1.26	1.12	1.27
2001–2015	0.39	0.86	0.98	1.21	1.08	1.24
2001–2016	0.37	0.83	0.95	1.18	1.05	1.21
2001–2017	0.36	0.81	0.93	1.15	1.02	1.18
2001–2018	**0.35**	**0.79**	**0.90**	**1.13**	**1.00**	**1.16**

GB and RF are gradient boosting and random forest, respectively. The first two columns indicate the results from the out-of-sample tests. IMF_F and IMF_S are the IMF forecasts made in fall and spring, respectively. BOJ_F and BOJ_S are the BOJ forecasts made in fall and spring, respectively

4 Conclusion

The results of this study advocate the use of machine learning techniques in forecasting macroeconomic data. Based on the customized cross-validation process, the machine learning method employed in this study, which creates gradient boosting and random forest models for the 2001–2018 period, produces forecasts that are more accurate than those made by the IMF and the BOJ. Accuracy is measured by MAPE and RSME.

Traditional econometric models focus on explanations of causal relationships, whereas machine learning models focus on predictions. Machine learning models may not be a good choice for determining the impact of independent variables on the dependent variable or analyzing a causal relationship. However, as shown in this study and in multiple previous studies, machine learning models often show high prediction power.

Since there is no model or methodology that produces the best result for every type of data set, this study contributes to the literature by empirically testing and comparing predictions of real GDP growth in Japan using popular machine learning models based on real data. This study further proposes a recursive method that combines cross-validation and hyperparameter tuning to create accurate models, which can be accurate even with low-frequency macroeconomic data. From this point of view, the method suggested in this study should serve as an effective analysis option for predicting economic variables that could lead to more effective economic policy design and implementation, especially when only low-frequency data are available. Finally, based on the validated result, this study also supports and encourages further research on and use of machine learning models to forecast economic variables and to answer economic questions.

Acknowledgements I would like to express my gratitude to Professor BAAK Saang Joon, Professor KONDO Yasushi, and Professor KONISHI Hideki of Waseda University for their invaluable comments and advice.

Author contribution J.Y. is the sole author.

Funding Not applicable.

Code availability Not applicable.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Availability of Data and Materials Data sources are indicated in the manuscript.

Appendix

See Tables 6 and 7.

Table 6 Description of the variables

Variable	Unit	Scale	Number of observations	Sources
Consumer price index	Index	1990 = 100	147 (4Q of 1981–2Q of 2018)	Ministry of Internal Affairs and Communications of Japan, Oxford Economics
Consumption, government, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
Consumption, private, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
Current account of balance of payments	%		147 (4Q of 1981–2Q of 2018)	Bank of Japan/Ministry of Finance, Oxford Economics
Employment	Person	Thousands	147 (4Q of 1981–2Q of 2018)	Ministry of Internal Affairs and Communications, Oxford Economics
Exchange rate, period average, per euro	Yen per euro		147 (4Q of 1981–2Q of 2018)	Haver Analytics, Oxford Economics
Exchange rate, period average	Yen per US dollar		147 (4Q of 1981–2Q of 2018)	Federal Reserve Board, Oxford Economics
Exports, goods & services, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
External debt, total	US dollar	Millions	147 (4Q of 1981–2Q of 2018)	Ministry of Finance/Federal Reserve Board, Oxford Economics
Foreign direct investment	US dollar	Millions	147 (4Q of 1981–2Q of 2018)	Bank of Japan/Ministry of Finance Japan, Oxford Economics
GDP deflator	Index	1990 = 100	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
GDP, real, quarter	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
GDP growth, real, quarter to quarter	%	Base year = 1990	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics, Author's calculation
GDP growth, real, year over year	%	Base year = 1990	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics, Author's calculation
GDP growth, real, year over year, 2 quarters ahead	%	Base year = 1990	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics, Author's calculation
GDP growth, real, year over year, 3 quarters ahead	%	Base year = 1990	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics, Author's calculation

Table 6 (continued)

Variable	Unit	Scale	Number of observations	Sources
Government balance (as a % of GDP)	%		147 (4Q of 1981–2Q of 2018)	Organization for Economic Cooperation & Development/Cabinet Office of Japan, Oxford Economics
Gross government debt (as a % of GDP)	%		147 (4Q of 1981–2Q of 2018)	Bank of Japan/Cabinet Office of Japan, Oxford Economics
Imports, goods & services, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
Industrial production index	Index	1990 = 100	147 (4Q of 1981–2Q of 2018)	Ministry of Economy, Trade and Industry of Japan, Oxford Economics
Interest rate, 10-year government bond yields	%		147 (4Q of 1981–2Q of 2018)	Ministry of Finance Japan, Oxford Economics
Investment, total fixed investment, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
Reserves, foreign exchange	US dollar	Millions	147 (4Q of 1981–2Q of 2018)	Ministry of Finance
Tokyo Stock Price Index	Index	Jan-04-1968 = 100	147 (4Q of 1981–2Q of 2018)	Financial Times/Nihon Keizai Shinbun (Nikkei), Oxford Economics
Stockbuilding, real	Yen	Billions: 1990 prices	147 (4Q of 1981–2Q of 2018)	Cabinet Office of Japan, Oxford Economics
Unemployment rate	%		147 (4Q of 1981–2Q of 2018)	Ministry of Internal Affairs and Communications of Japan, Oxford Economics

Table 7 MAPEs and RMSEs for the in-sample tests of the machine learning models

Forecast year	MAPE (%)		RMSE	
	GB (in-sample)	RF (in-sample)	GB (in-sample)	RF (in-sample)
2001	80.67	196.11	0.44	0.59
2002	24.04	35.15	0.52	0.78
2003	28.86	39.12	0.30	0.40
2004	27.31	33.94	0.43	0.65
2005	28.06	44.06	0.30	0.42
2006	20.76	23.89	0.31	0.66
2007	8.27	15.92	0.19	0.35
2008	22.35	25.93	0.39	0.50
2009	27.80	48.44	0.35	0.50
2010	61.05	210.95	0.39	0.55
2011	11.36	25.62	0.73	0.83
2012	26.96	46.06	0.21	0.58
2013	9.89	15.39	0.10	0.24
2014	79.98	287.14	0.74	0.87
2015	24.26	24.14	0.73	0.96
2016	9.03	18.61	0.21	0.44
2017	208.40	208.43	0.41	0.50
2018	13.95	17.37	0.26	0.47
Average (2001–2018)	**39.61**	**73.13**	**0.39**	**0.57**

GB and RF are gradient boosting and random forest, respectively. MAPEs and RSMEs are calculated using the forecast values from the cross-validation process used for each forecast year

References

Biau, O., & D'Elia, A. (2010). *Euro area GDP forecasting using large survey datasets: A random forest approach*. Euroindicators working papers.

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

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems. MCS 2000. Lecture notes in computer science* (Vol. 1857, pp. 1–15).

Emsia, E., & Coskuner, C. (2016). Economic growth prediction using optimized support vector machines. *Computational Economics, 48*(3), 453–462.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York: Springer.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of the American Statistical Association, 103*(482), 511–522.

Jung, J.-K., Patnam, M., & Ter-Martirosyan, A. (2018). *An algorithmic crystal ball: Forecasts-based on machine learning*. IMF Working Papers. Washington, D.C.: International Monetary Fund.

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2019). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*,. https://doi.org/10.1080/07350015.2019.1637745.

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics, 21*(15), 3301–3307.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.
Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the U.S. real house price index. *Economic Modelling, 45,* 259–267.
Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(3), e1301.
Richardson, A., Mulder, T., & Vehbi, T. (2018). *Nowcasting New Zealand GDP using machine learning algorithms*. CAMA Working Paper 47/2018. The Australian National University.
Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the Gini variable importance measure in classification trees. *Journal of Computational and Graphical Statistics, 17*(3), 611–628.
Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, *16*(4), 437–450.
Tiffin, A. (2016). *Seeing in the dark: A machine-learning approach to nowcasting in Lebanon*. IMF Working Papers. International Monetary Fund.
Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28.