CEFIS WORKING PAPER SERIES
First Version: February 2020

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

Barış Soybilgen, Istanbul Bilgi University

Ege Yazgan, Istanbul Bilgi University

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

Barış Soybilgen*and Ege Yazgan†

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 problem in nowcasting and improve the prediction performance of machine learning models, 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. We also point out that factors obtained from real variables are more influential for machine learning models but the impact of factors derived from financial and price variables increase after the sub-prime mortgage 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 future. For example in weather forecasting, where we exactly know what the weather today is, we only need to forecast the future. However in other fields, such as economics not all data related today are known today. Important macroeconomic aggregates can only be measured with a considerable time delay. Hence, for these variables,

*Istanbul Bilgi University, baris.soybilgen@bilgi.edu.tr.
†Istanbul Bilgi University, ege.yazgan@bilgi.edu.tr.

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

The publication delays of key economic macro indicators pose serious problems for policy makers and for all who needs to monitor the economy in real time and form their decisions at present, based on the timely information. For example, US GDP, which is the key variable describing the overall state of the economy as a whole, is published at quarterly frequency. Its first (or advance) estimate is released with almost a month delay after the end of the corresponding quarter.[2] This delay may be extended up to six months in other countries. Even a delay such as one month constitutes a significant lag in information flow for those who need timely information in their decision making process. In the absence of a timely availability of GDP figures, the decision makers with an interest in monitoring the overall state of the economy in real time should necessarily rely on other indicators that are related to GDP but published with shorter or 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) develop a joint multivariate nowcasting model to perform this task. By putting an emphasis on forecasting to the present, they call it as "nowcasting". This nowcasting model is a special form of large dynamic factor models estimated by principle components, initially introduced by Stock and Watson (2002a),Stock and Watson (2002b) using Kalman filter to update its predictions and is designed to handle the irregularities of the real time data, such as mixed frequencies and non synchronicity of the data releases. The model estimates the unobserved factors that drive the data and produces a forecast of each economic and financial series which it incorporates. Whenever the actual release for a series departs from the model's forecast, this is considered as a "news" and it affects the nowcast of GDP growth. The dynamic factor model proposed by Giannone et al. (2008) and its successors have been used to produce successful nowcasts for a variety of countries from developed economies to emerging markets.[3]

After deriving dynamic factors from a data set that contains high frequency variables, the literature links these dynamic factors with quarterly GDP using a linear model. An obvious alternative to this linear model is machine learning models. It is recently gaining popularity in macroeconomic forecasting, but it is rarely used in the context of nowcasting. In one of the notable studies, Giusto and Piger (2017) adopt Learning Vector Quantization (LVQ) to

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

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

[3]See Bok et al. (2018) and Banbura et al. (2013) for a review of the literature

nowcast US business cycle states and after identify turning points using non-farm payroll employment, industrial production index, real personal income excluding transfer receipts, and real manufacturing and trade sales. In a similar study, Soybilgen (2020) nowcasts US business cycle states using a neural network model that includes dynamic factors as explanatory variables extracted from a large-scale data set consisted of more than 100 variables. In another study instead of nowcasting business cycle states, Richardson et al. (2018) nowcast New Zeland GDP with a large-scale data set of around 550 variables utilizing various machine learning models such as K-nearest neighbour regressions, boosted trees, elastic net, support vector machines, and neural network models. Finally, Loermann and Maas (2019) nowcast US GDP using a feedforward artificial neural network model with a large-scale data set. Overall these studies showed that ML models provide improvements over regular linear models.[4]

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 McCracken and Ng (2016) 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 dynamic factor model proposed by Giannone et al. (2008). This both helps us the fill missing monthly data at the end of the sample in a straightforward manner and improve the prediction performance of machine learning models.[5] However unlike previous nowcasting studies such as 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 ten groups of variables and then derive factors from each group of variables.

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 dynamic factor models. 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 linear models. We also show that the performance different machine learning models, especially random forests, and linear

[4]In forecasting context, while Makridakis et al. (2018) document lower out-of-sample performance of machine learning models, Chakraborty and Joseph (2017) report the reverse for UK data.

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

models increases after the big 2008 financial crisis. 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 predicting performance compared to models estimated recursively. 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, influence of factors derived from financial and price variables increases only after the 2008 financial crisis.

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

2 The data set

The large-scale data set used in this article 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. Due to the discontinuation of some old series, the introduction of some newly updated series, and some other data collection problems, FRED-MD vintage data in each period does not have the same number of variables. Variables and their period of use are listed in Appendix A. Furthermore, all variables are transformed appropriately to ensure stationary. Their applied transformations are also shown in Appendix A. For vintage GDP data, we use the data set obtained from ArchivaL Federal Reserve Economic Data (ALFRED) system.

3 The methodology

In this study, we use tree-based ensemble machine learning algorithms including dynamic factors as explanatory variables. Our aim in 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 the models. Using a dynamic factor model, we can reduce the dimension of the data set and eliminate the most of the noise from the data set. Furthermore, the DFM of Giannone et al. (2008) can solve the ragged/jagged edge data problem [6] by utilizing a Kalman smoother.

3.1 The dynamic factor model

Let's assume that n monthly series $x_{t_m} = (x_{1,t_m}, x_{2,t_m}, \ldots, x_{n,t_m})', t_m = 1, 2, \ldots, T_m$, which are transformed via Mariano and Murasawa (2003) approximation and then standardized, have the following representation:

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

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

$$f_{t_m} = \Phi(L) f_{t_m - 1} + B\eta_t; \quad \eta_{t_m} \sim \mathbb{N}(0, I_q), \tag{2}$$

where B is an rxq matrix of full rank q with $q \leqslant r$, $\varphi(L)$ is an rxr lag polynomial matrix, and η_t is the q dimensional vector of common shocks that follows a white-noise process.

In this study, we use a two-step estimation approach whose properties are shown byDoz et al. (2011) to obtain common factors. The algorithm is initialized by computing principal components, and parameter estimates of the model is obtained by ordinary least squares. In the second step, updated estimates of the common factors are obtained with the Kalman smoother using the parameters estimated in the first step.

3.2 Linear Models

To link monthly factors with quarter over quarter GDP growth rates, y_{t_q}, we obtain quarterly factors from their monthly counterparts by extracting them that correspond the last month of each quarter. Let's 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 $f_{t_q}, t_q = 1, 2, \ldots, T_q$, where

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

$T_m = T_q/3$.

We use two linear models as benchmark. In the first model, two factors, which are derived from the whole data set using the $r = 2, q = 2, p = 1$ specification[7] are utilized to obtain h_q steps ahead predictions of quarterly GDP growth rates at time t_q, $\hat{y}_{t_q+h_q|t_q}$, as follows:

$$\hat{y}_{t_q+h_q|t_q} = \hat{c} + \sum_{i=1}^{2} \hat{\beta}_i \hat{f}_{i,t_q+h_q|t_q}. \tag{3}$$

In the second model, two factors are extracted from each group j in the data set using the $r = 2, q = 2, p = 1$ specification and we obtain 20 factors in total. Then these 20 factors are used to obtain h_q steps ahead predictions of quarterly GDP growth rates at time t_q, $\hat{y}_{t_q+h_q|t_q}$, as follows:

$$\hat{y}_{t_q+h_q|t_q} = \hat{c} + \sum_{j=1}^{10} \sum_{i=1}^{2} \hat{\beta}_{(i,j)} \hat{f}_{(i,j),t_q+h_q|t_q}. \tag{4}$$

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 equation 4, we feed 20 factors into our machine learning models as 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 searching the feature space to find the best splitting variable and split point via a greedy approach. CART divides the feature space in rectangular regions in an iterative way until a stopping criterion is reached such as the minimum number of observations in each region. In our case of regression trees, each region's prediction is the average of y_{t_q} in that region.

Even though decision trees are extremely easy to interpret, they tend to have poor and noisy

[7] There are various ways to determine the number of factors such as using information criteria proposed by Bai and Ng (2002) and Bai and Ng (2007), the best 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 a quite good in these kinds of nowcasting exercises.

predictions in many cases compared to more advanced machine learning models. Breiman (2001) introduces random forests, which is an ensemble decision tree model, as a technique with low variance and high prediction performance. Random forests are based on bagging (boostrap aggregating) of decision trees. For bagged decision trees, we create K bootstrapped training set from original data, fit a decision tree to each bootstrapped training set, and take simple averages of these fitted K decision trees' predictions. This procedure is known to reduce variance while increasing the prediction performance of decision trees. Even though bagged decision trees work in many cases, they fail to yield desired improvement, when bagged fitted trees are too correlated with each other. Random forests solve this problem by just allowing only a random sample of variables to be considered in each split. In this way, bagged fitted trees are dissociated from each other.

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 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 by following Friedman (2001) and Friedman (2002). In gradient boosting, regression trees are fitted to pseudo residuals, negative gradients, instead of actual residuals. Friedman (2001) also uses different learning rates for each of decision tree's regions for higher prediction performance. In stochastic gradient tree boosting, Friedman (2002) further improves the model by using only a part of the training set drawn at random without replacement in each iteration.

4 Empirical results

4.1 Nowcasting performance

We estimate our models between January 2000 and December 2018 using vintage data that take account data revisions into account. 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, we start predicting 2000Q2 GDP starting from January 2000 until June 2000 for six months, and the advance estimate for 2000Q2 is announced by the US Bureau of Economic Analysis in July 2000.

We evaluate the nowcasting performance of our models between 2000Q2 and 2018Q4 by comparing predictions of the models with the advance estimates of GDP data using root mean square errors (RMSEs). First, we estimate our model using an expanding estimation window. Next, we use a rolling estimation window. We choose optimal hyperparameters using the initial estimation window covering the period between 1960Q3 and 1999Q4. We perform a grid search using three-fold cross-validation three times.[8]

Table 1 presents RMSEs of both machine learning and benchmark models estimated recursively for six successive predictions of the reference quarter as explained above. RW refers to the random walk model. 2FLM and 20FLM represent linear dynamic factor models presented in equation 3 and 4, respectively. BDT, RF, and GBM refer to bagged decision trees, random forests, and stochastic gradient tree boosting, respectively.

Table 1: RMSEs of the Models, Expanding Window, 2000Q2-2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
1st Prediction	0.626	0.667	0.506	0.535	0.510	0.538
2nd Prediction	0.679	0.672	0.499	0.496	0.482	0.498
3rd Prediction	0.699	0.654	0.493	0.476	0.465	0.487
4th Prediction	0.507	0.600	0.463	0.430	0.423	0.444
5th Prediction	0.517	0.533	0.405	0.375	0.361	0.399
6th 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

In the first predictions for the reference quarter, 20FLM has the highest forecasting performance. However, machine learning models beat the remaining two competing models; RW and 2FLM. These results indicate the fact that machine learning models are having a hard time predicting the reference quarter 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 obtained by BT, which is followed by BDT and GBM, respectively. The nowcasting performance of machine learning models improves significantly as more information becomes available for the target reference quarter.

[8]We also choose hyperparameters using a five-fold cross-validation and reach similar results.

The last row of Table 1 presents the average RMSEs of the models. The average RMSE of RF is approximately 28% lower than 2FLM and RW. Furthermore, the average RMSE of RF is 5.8% lower than the average RMSE of 20FLM. BDT and GBM have also lower average RMSE than all competing models.

One interesting result stems from Table 1 is that 2FLM exhibit the worst performing nowcasting performance. It seems that obtaining factors from each group of variables provides much better results than deriving factors from the whole data set.

In Table 1, we presented the prediction performance of the models for the whole period. However, the literature shows that the forecasting performance of models is usually unstable, and the ranking of models can change over time (e.g. Stock and Watson, 2003; Stock and Watson, 2004; Kuzin et al., 2013) Therefore, we present the prediction performance of 2FLM, 20FLM, BDT, RF, and GBM by calculating five years rolling windows of RMSEs in Figure 1.[9]

Figure 1

In the first predictions for the reference quarter, RMSEs of machine learning models and 20FLM move very closely until August 2008. After prediction errors due to the sub-prime mortgage crisis become present, RMSEs of all models start to increase. Between 2009 and 2014, 20FLM have higher prediction performance than machine learning models, and RF is the second best model. It 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 machine learning models outperform 20FLM. Except for a small period between 2010 and 2012, 2FLM is the worst model. In the second and third predictions for the reference quarter, results are very similar to those for the first predictions. The main differences are as follows: 20FLM performs much worse than 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 and after 2016, all machine

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

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 RF's nowcasting performance is very similar to the 20FLM's performance. In all other periods, machine learning models generally beat 20FLM. In the sixth predictions for the reference quarter, RF and BDT perform similar to 20FLM or outperform 20FLM slightly until the end of 2014. Interestingly after 2014, 20FLM perform much worse than machine learning models. In the fourth, fifth, and sixth predictions even though 2FLM performs pretty decently until the crisis, its prediction performance rapidly deteriorates after the financial crisis.

In the presence of instabilities induced by structural breaks, rolling window estimation can improve the forecasting performance of models compared to the expanding window estimation. Table 2 presents RMSEs of both machine learning and benchmark models estimated with a rolling window. 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 Table 2, RF has lower RMSE than all benchmark models at all times. As in Table 1, BDT and GBM outperform all other benchmark models starting from the second predictions for the reference quarter. The average RMSE of RF and BDT is now 8.6% and 7.5% lower than the average RMSE of 20FLM. Rolling window estimation appears to help in improving the nowcasting performance of machine learning models significantly compared to the 20FLM's performance.

Table 2: RMSEs of the Models, Rolling Window, 2000Q2-2018Q4

	RW	2FLM	20FLM	BDT	RF	GBM
1st Prediction	0.626	0.631	0.498	0.503	0.491	0.519
2nd Prediction	0.679	0.635	0.494	0.474	0.468	0.484
3rd Prediction	0.699	0.616	0.487	0.453	0.447	0.470
4th Prediction	0.507	0.564	0.457	0.406	0.403	0.433
5th Prediction	0.517	0.496	0.403	0.348	0.346	0.385
6th 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

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 Figure 2. It seems that Figures 1 and 2 exhibit very similar results.

Figure 2

4.2 Variable importance

In this section, we analyze which variables are more important when predicting US GDP with tree-based ensemble models. For RF and BDT, we follow Breiman (2001) to calculate variable importance. For GBM, we follow Friedman (2001) to calculate relative importance of variables.

Figure 3 presents average importance of dynamic factors for BDT, RF, and GBM estimated using both rolling window and expanding window. To calculate average importance, we first calculate importance values for each factor in each period, and then take simple average of importance values over the whole forecasting sample for each factor. First of all, most important dynamic factors in all cases are derived from groups of real variables such as the consumption, orders, and inventories group, the output and income group, and the labor market group. For the output and income group, the first factor has the highest influence to 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 labor market group and the first factor from the consumption, orders, and inventories 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 to 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 output and income group is higher in expanding window estimation compared to rolling window estimation.

Figure 3

Results for factors derived from financial and price variables are interesting. It seems that they carry very little information for our models. Therefore, we further analyze these results by presenting the aggregate influence of factors derived from real variables namely the consumption, orders, and inventories group, the output and income group, the housing group, and the labor market group versus factors derived from financial and price variable groups over the whole forecasting period.

Figure 4 presents aggregate importance of real factors against financial and price factors. To calculate aggregate importance, we first calculate importance values for each factor in each period and then we aggregate importance values for real factors and financial and price factors. For bagged decision trees, influence of financial and price factors is negative until the sub-prime mortgage crisis. In the period after the crisis, influence of financial and price factors becomes positive. For bagged decision trees estimated using an expanding window, the aggregate importance metric of financial and price factors reaches 20% and then fluctuated around 10%. For bagged decision trees estimated using a rolling window, the aggregate importance metric of financial and price factors steadily increases to 40% while influence of real factors decreases slightly. For random forests estimated using a rolling window, the aggregate importance metric of financial and price factors increases steadily after the latest financial crisis. For random forests estimated using an expanding window, aggregate importance of financial and price factors increases slightly after the crisis and then fluctuated around 10%. Interestingly for GBM estimated recursively, we don't see much difference in relative importance of variables, and for GBM estimated using a rolling window relative importance of financial variables, increases steadily over time.

Figure 4

For random trees and bagged decision trees, it seems that financial variables become more important after the 2007-2009 financial crisis, and this effect is more prominent for models estimated using a rolling window scheme.

5 Conclusion

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

Our results show that tree-based ensemble models beat linear models in most of the time. The performance of machine learning models especially increases when more data for the reference quarter become available. Our results also point out that random forests and bagged decision trees outperform linear models more significantly after the latest financial crisis. We also show that tree-based ensemble models estimated with a rolling window have better nowcasting performance compared to models estimated recursively. 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 influence of factors extracted from financial and price variables increases after the sub-prime mortgage crisis for random forests and bagged decision trees.

References

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

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

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

Bańbura, M. and G. Rünstler (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., O. Darné, and L. Ferrara (2010). Are disaggregate data useful for factor analysis in forecasting french gdp? *Journal of Forecasting 29*(1-2), 132–144.

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (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., J. Friedman, C. Stone, and R. Olshen (1984). *Classification and Regression Trees*. Wadsworth.

Chakraborty, C. and A. Joseph (2017). Machine learning at central banks. Bank of england working paper no. 674.

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

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

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

Giusto, A. and J. Piger (2017). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting 33*(1), 174–184.

Kuzin, V., M. Marcellino, and C. Schumacher (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. and B. Maas (2019). Nowcasting US GDP with artificial neural networks. Munich personal repec archive working paper no. 95459.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2018). The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting 34*(4), 802–808.

Mariano, R. S. and Y. Murasawa (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. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Richardson, A., T. Mulder, and T. Vehbi (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*.

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

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

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

Stock, J. H. and M. W. Watson (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature 41*(3), 788–829.

Figures

Figure 1: 5 Years Rolling Averages of RMSEs, Expanding Window, 2005Q1-2018Q4

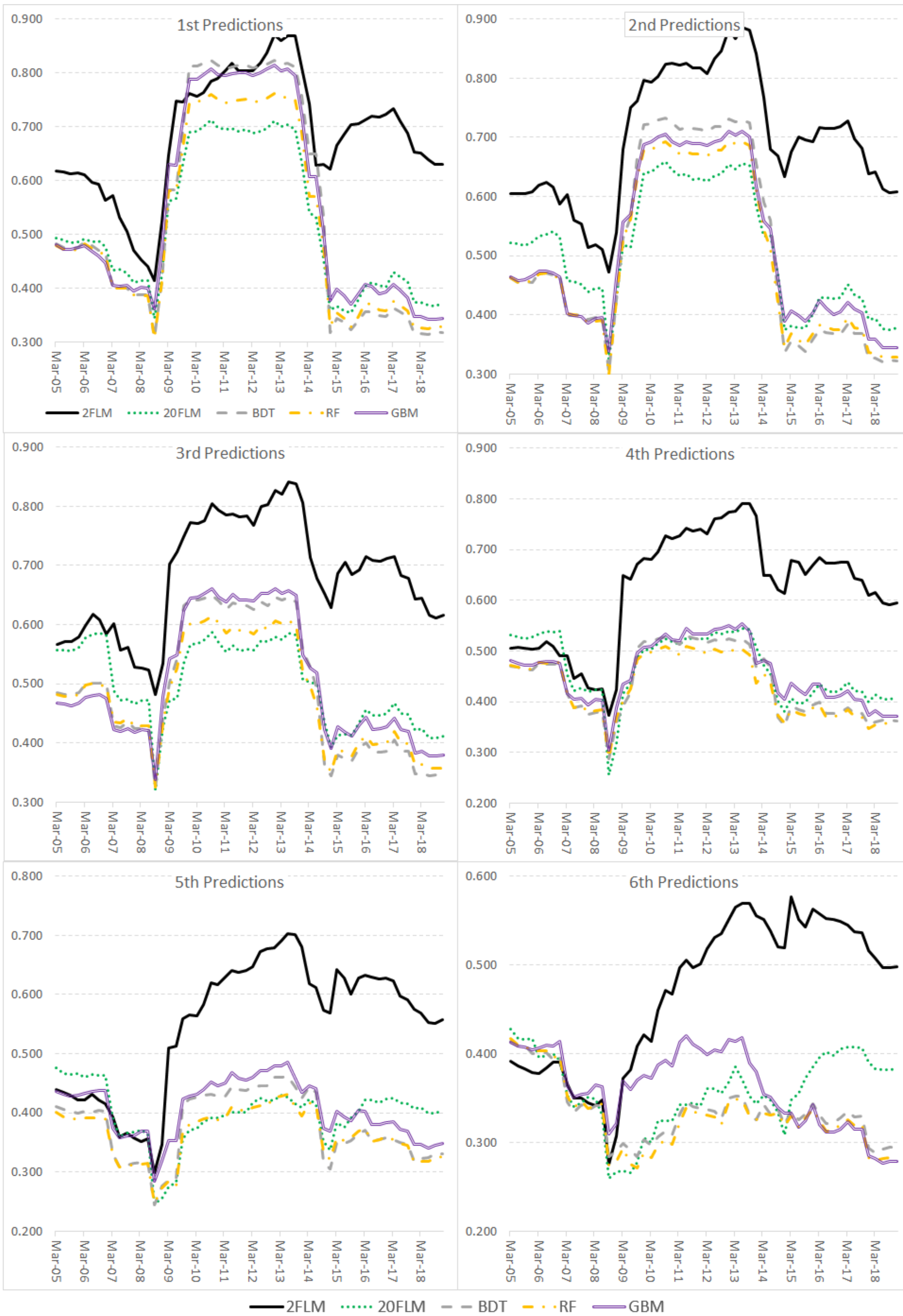


Figure 2: 5 Years Rolling Averages of RMSEs, Rolling Window, 2005Q1-2018Q4

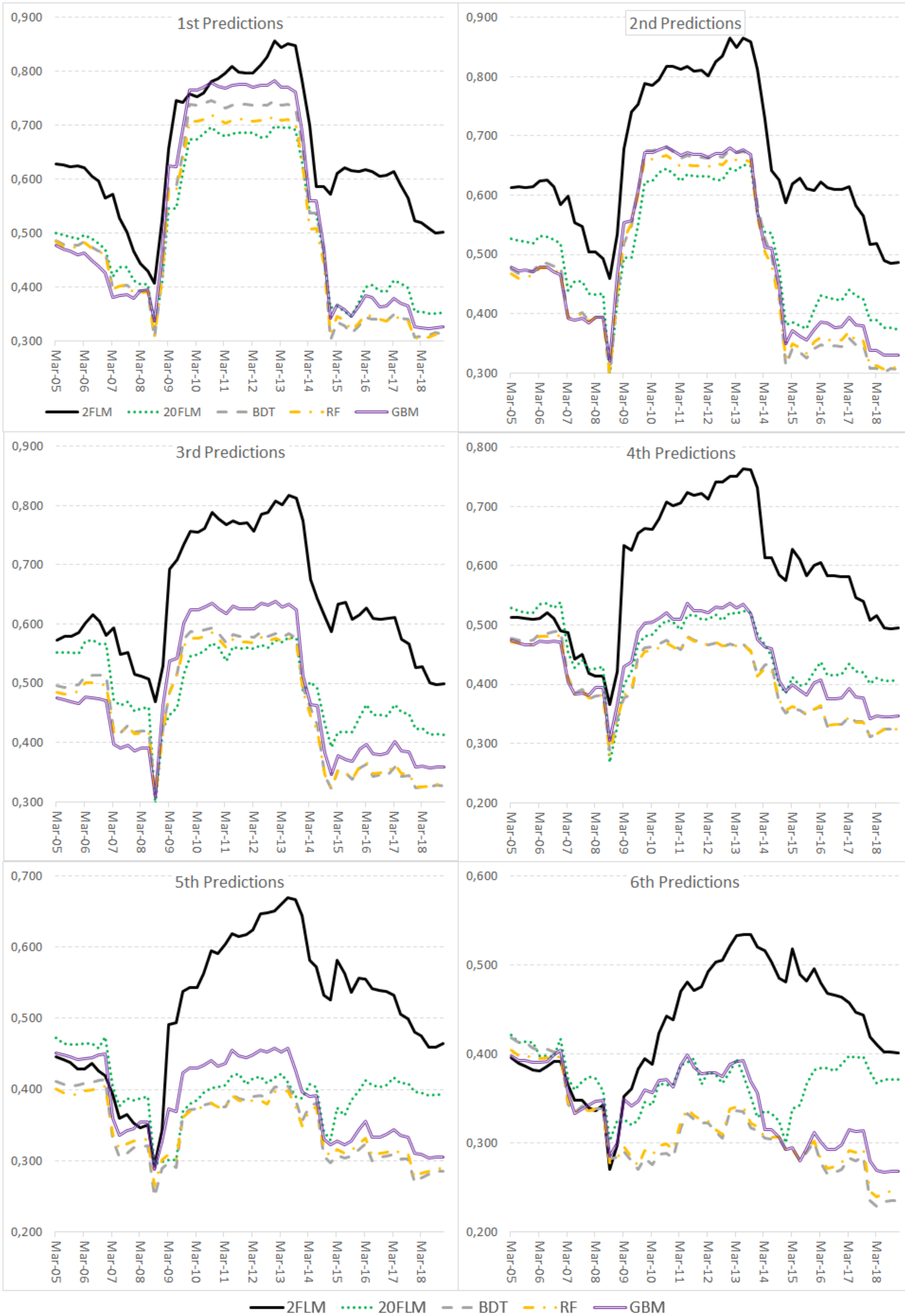


Figure 3: Average Importance of Factors, January 2000-December 2018

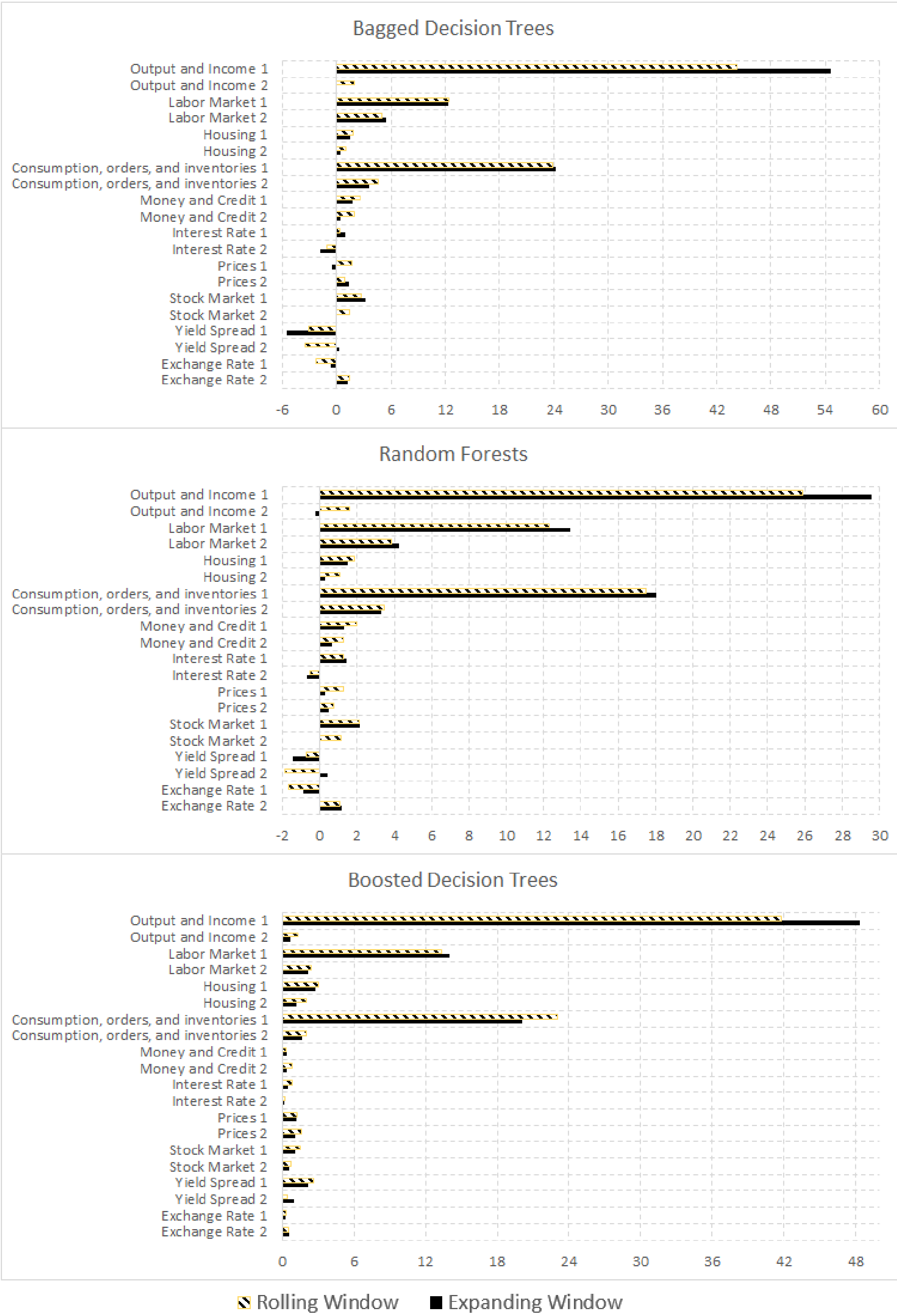
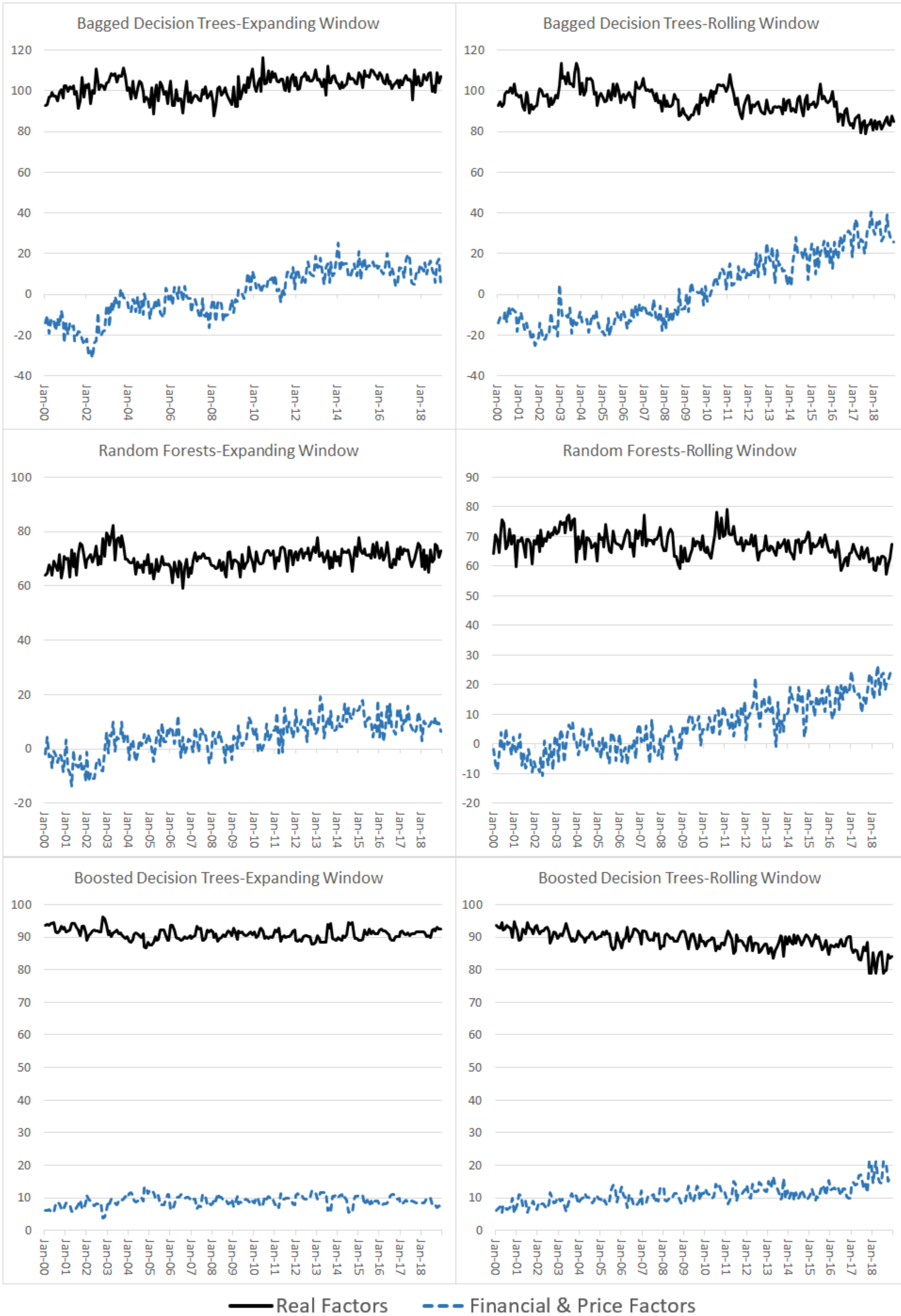


Figure 4: Aggregate Importance of Real Factors and Financial and Price Factors, January 2000-December 2018



Appendix A: the description of the data set

Group	Description	Transformation	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-
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 to 14 Weeks	1	All
2	Number of Civilians Unemployed for 15 Weeks and Over	1	All
2	Number of Civilians Unemployed for 15 to 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 Privately 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 Permits	1	All
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

Code	Description	Transformation	Period
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-Month Commercial Paper Minus Federal Funds Rate	0	All
9	3-Month Treasury Bill Minus Federal Funds Rate	0	All
9	6-Month Treasury Bill Minus Federal Funds Rate	0	All
9	1-Year Treasury Constant Maturity Minus Federal Funds Rate	0	All
9	5-Year Treasury Constant Maturity Minus Federal Funds Rate	0	All
9	10-Year Treasury Constant Maturity Minus Federal Funds Rate	0	All
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	Consumer Motor Vehicle Loans Owned by Finance Companies. Outstanding	1	All
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*
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-

Note: The column "Group" 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, and (10) exchange rate.

The column "Transformation" 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.

* Except the period between 2004:12-2005:07.