

CIRJE-F-1146

**Macroeconomic Forecasting Using Factor Models and
Machine Learning: An Application to Japan**

Kohei Maehashi
School of Engineering, The University of Tokyo

Mototsugu Shintani
The University of Tokyo

March 2020

CIRJE Discussion Papers can be downloaded without charge from:
<http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html>

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan*

Kohei Maehashi[†] and Mototsugu Shintani[‡]

This version: March 2020

Abstract

We perform a thorough comparative analysis of factor models and machine learning to forecast Japanese macroeconomic time series. Our main results can be summarized as follows. First, factor models and machine learning perform better than the conventional AR model in many cases. Second, predictions made by machine learning methods perform particularly well for medium to long forecast horizons. Third, the success of machine learning mainly comes from the nonlinearity and interaction of variables, suggesting the importance of nonlinear structure in predicting the Japanese macroeconomic series. Fourth, while neural networks are helpful in forecasting, simply adding many hidden layers does not necessarily enhance its forecast accuracy. Fifth, the composite forecast of factor models and machine learning performs better than factor models or machine learning alone, and machine learning methods applied to principal components are found to be useful in the composite forecast.

JEL Classification Number: C11, C32, C38, C45, C53.

Keywords: forecast, big data, factor model, machine learning, lasso, ridge, elastic net, regression tree, bagging, random forests, boosting, neural network, composite forecast, Bayesian model averaging.

*The authors are deeply grateful to the seminar participants at the University of Tokyo for their helpful comments and discussions. Shintani gratefully acknowledges the financial support of Grant-in-aid for Scientific Research (17H02510). The views expressed in this paper are those of the authors and do not necessarily reflect the official views of any institutions.

[†]School of Engineering, the University of Tokyo (E-mail: maehashi@econ.rcast.u-tokyo.ac.jp)

[‡]Faculty of Economics, the University of Tokyo (E-mail: shintani@e.u-tokyo.ac.jp)

1 Introduction

Big data environment and machine learning methods have recently attracted a great deal of attention in economic analysis. When it comes to macroeconomic forecasting, Stock and Watson (1999, 2002, 2012) recommended using common factors, that are extracted from a large number of variables, as predictor variables and their dynamic factor approach has been popularly used among macroeconomists. In an application of this methodology to Japanese economy, Shintani (2005) evaluated the performance of dynamic factor models combined with neural networks by using a balanced panel of 235 monthly series from 1973 to 2000, and claimed the importance of considering the nonlinearity to improve the forecast accuracy. However, machine learning methods other than neural networks were not considered by Shintani (2005). In addition, since the sample period ended in 2000, the possibility of structural changes, such as the one caused by the global financial crisis in 2007-2009, could not be incorporated.

In this paper, we extend the analysis of Shintani (2005) and perform a thorough comparative analysis of factor models and machine learning methods for forecasting Japanese economy by using updated macroeconomic data. In particular, we compare the forecast accuracy of factor models and various types of machine learning methods for 7 target variables by using 219 monthly series from 1973 to 2018. Other than the use of the extended series, our forecast design differs from that of Shintani (2005), in the following three points. First, we employ not only neural networks but also other representative machine learning methods such as the regularized least squares methods, which include lasso, ridge, and elastic net, as well as the ensemble learning based on regression trees, which includes bagging, random forests, and boosting. Second, we consider more general neural networks than the one considered in Shintani (2005) in terms of the choice of activation functions and the depth of

hidden layers. Third, we examine the possibility of forecast improvement from composite forecasts to seek for the best combination of the two cases: the case when the common factors are extracted by principal component analysis and the case when machine learning methods are employed.

Our main results can be summarized as follows. First, factor models and machine learning methods perform better than the conventional AR model in most cases. Second, machine learning methods work particularly well for relatively longer forecast horizons. Third, the success of machine learning methods mainly comes from the nonlinearity and interaction of variables, suggesting the importance of nonlinear structure in predicting the Japanese macroeconomic series. Fourth, while neural networks are useful in forecasting, simply adding many hidden layers does not necessarily enhance the forecast accuracy. Fifth, composite forecasts of factor models and machine learning perform better than factor models or machine learning alone, and machine learning methods applied to principal components are found to be useful in the composite forecast.

There are many related studies on macroeconomic forecasting based on factor models and machine learning. For the purpose of dimension reduction under the many predictors environment, a dynamic factor approach based on principal components regression proposed by Stock and Watson (1999, 2002, 2012) has been used in many applications, including Bai and Ng (2002, 2006), Artis et al. (2005), Boivin and Ng (2006), Kim and Swanson (2014, 2018), to name a few. Kitamura and Koike (2003) and Shintani (2005) applied this principal components regression forecast to Japanese economy. There are also an increasing number of studies which employ the machine learning methods in more recent macroeconomic forecasting applications. For example, Inoue and Kilian (2008) considered lasso and ridge regression and compared their performance with bagging for U.S. inflation forecast. Diebold and Shin (2019) employed lasso, ridge and their extensions for Euro area GDP forecast.

Giannone et al. (2019) applied a Bayesian forecasting framework that covered lasso, ridge and elastic net to Stock and Watson type macroeconomic data, among other series, for the purpose of evaluating the usefulness of sparse modeling. Nakajima and Sueishi (2019) employed a lasso-VAR model for forecasting Japanese macroeconomic series. The usefulness of the ensemble machine learning based on regression trees, such as random forests, was emphasized in Medeiros et al. (2019) and Chen et al. (2019). Bai and Ng (2009) examined the effectiveness of boosting in forecasting inflation, interest rate, industrial production, employment and the unemployment rate by using a large set of U.S. macroeconomic data. Macroeconomic forecasting application of neural networks include Nakamura (2005) for U.S. inflation series, Shintani (2005) for Japanese macroeconomic series, and Cook and Hall (2017) for US unemployment series, among others.

We are not interested in investigating the performance of a single method or single target series. Instead, we compare factor models and 8 different machine learning methods to forecast 7 target macroeconomics variables at horizons from 1 month to 3 years. The research design of our analysis is closely related to Kim and Swanson (2018) and Coulombe et al. (2019) who conducted a horse race analysis using various machine learning methods in forecasting the U.S. macroeconomic variables. However, to the best of our knowledge, our study is the first one to conduct comprehensive forecast comparison of machine learning methods using Japanese macroeconomic data.

The remainder of the paper is organized as follows. Section 2 introduces factor models and machine learning methods which will be considered in this paper. In Section 3, we describe the overviews of our data and explain how we evaluate the performance of competing forecast models. Section 4 demonstrates the main empirical results, followed by the additional analysis of composite forecast in Section 5. Concluding remarks are made in Section 6.

2 Forecasting models

We construct an h -period ahead forecast of the variable of interest y_t using various methods. As a benchmark model to evaluate the forecasting performance, we consider a simple univariate autoregressive (AR) model given by

$$y_{t+h} = \phi(L)y_t + \varepsilon_{t+h}, \quad (1)$$

where $\phi(L) = \sum_{j=1}^p \phi_j L^{j-1}$ is the lag polynomial with a lag operator L and ε_{t+h} is the forecast error. The AR forecast is constructed by running a simple linear regression of y_{t+h} on current and lagged y_t 's with the lag length p selected by minimizing Bayesian Information Criterion (BIC).¹

In the following, we briefly describe factor models and machine learning methods which we use in our analysis.

2.1 Factor models

One of the most frequently used procedures of dimension reduction in macroeconomic forecasting, under the many predictors environment, is a dynamic factor approach based on the principal component analysis proposed by Stock and Watson (1999, 2002, 2012). The fundamental idea of a dynamic factor approach is to extract the unobserved latent driving force from a large number of variables and then to use a relatively small number of factors as predictors for forecasting. Let x_{it} be a potential candidate predictor generated from

$$x_{it} = \lambda_i' F_t^k + e_{it}, \quad (2)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where $F_t^k = (f_{1t}, f_{2t}, \dots, f_{kt})'$ is a $k \times 1$ vector of common factors, λ_i is a $k \times 1$ vector of factor loadings and e_{it} is the idiosyncratic disturbance.

¹In the analysis, intercept terms in the AR model and other linear regression models are always included, but are omitted here for notational simplicity.

Principal components regression (PCR): Stock and Watson’s dynamic factor approach requires the following two steps. First, the common factor F_t^k is computed as the principal component of all predictor variables $\{x_{it}\}_{i=1}^N$ by minimizing $\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \lambda_i' F_t^k)^2$. Second, using the estimated factor, the h -period ahead forecast is constructed by running a principal components regression (PCR) of the form

$$y_{t+h} = \beta_F' F_t^k + \varepsilon_{t+h}, \quad (3)$$

where β_F is a $k \times 1$ vector of coefficients. By construction, principal components are orthogonal to each other and normalized to have a unit variance. In the many predictors environment, choosing the appropriate number of factors k , which should be sufficiently smaller than N , is the crucial issue.² We select k by minimizing the following information criterion proposed by Bai and Ng (2002),

$$IC(k) = \ln V(k) + k \left(\frac{N+T}{NT} \right) \ln C_{NT}^2, \quad (4)$$

where $V(k) = \min_{\{\lambda_i, F_t^k\}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \lambda_i' F_t^k)^2$ and $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.³

Factor augmented autoregression (FAAR): We can also generalize PCR in two directions. One is to add lags of factors and the other is to include other covariates such as lagged dependent variables. To this end, we consider the factor augmented autoregression (FAAR) given by

$$y_{t+h} = \beta_F'(L) F_t^k + \phi(L) y_t + \varepsilon_{t+h}, \quad (5)$$

where $\beta_F(L) = \sum_{j=1}^r \beta_{Fj} L^{j-1}$ and $\phi(L) = \sum_{j=1}^p \phi_j L^{j-1}$ are the lag polynomials. Unlike PCR, FAAR nests the AR forecast as a special case when all the factors are excluded from the regression.

²Here, common factors are ordered so that first factor corresponds to the first principal component, namely f_{1t} , followed by the succeeding principal components.

³This is one of the six variants of information criteria proposed by Bai and Ng (2002). However, our results are not sensitive to other possible choices of information criteria.

2.2 Regularized least squares methods

An alternative popular approach to deal with many predictors is the penalized regression, or regularized least squares which nests lasso, ridge and elastic net. All three methods minimize the objective function

$$\sum_{t=1}^T \left\{ (y_{t+h} - \sum_{i=1}^N \beta_i x_{it})^2 + \lambda J(\beta) \right\}, \quad (6)$$

where λ is the parameter on regularization, but differs in the specification of the penalty term $J(\beta)$ where $\beta = (\beta_1, \beta_2, \dots, \beta_N)'$. For all the regularized least squares in our application, we choose λ by cross-validation.⁴

Lasso: The least absolute shrinkage and selection operator (lasso), first introduced by Tibshirani (1996), corresponds to the one with the penalty given by $J(\beta) = \sum_{i=1}^N |\beta_i|$. The penalty term of lasso is the L_1 norm which implies a kink at 0 in a constrained minimization problem with respect to β . This characteristic results in many coefficient estimates to become exactly zero. In this sense, lasso estimator can be viewed as a variable selection in a sparse modeling.

Ridge: Ridge regression, introduced by Hoerl and Kennard (1970), is characterized by L_2 norm penalty $J(\beta) = \sum_{i=1}^N \beta_i^2$. The coefficients of ridge regression can be very close to zero but not exactly zero unlike the lasso. Hence, ridge estimator is a shrinkage method where coefficients are shrunk in order to prevent overfitting.

Elastic net (EN): Although lasso estimator was designed to improve ridge estimator and it is certainly true particularly when most coefficients of the true model are zeros, ridge performs better than lasso when a correlation between predictors is high (Zou and Hastie, 2005). Against this background, elastic net was proposed to make the most of both lasso and ridge simultaneously. The penalty function of elastic net is given by $J(\beta) = \omega \sum_{i=1}^N |\beta_i| +$

⁴In particular, we follow the standard approach in the literature and employ 5-fold cross-validation.

$(1 - \omega) \sum_{i=1}^N \beta_i^2$, where ω is the additional hyperparameter governing the relative importance of L_1 norm penalty to L_2 norm penalty and it is determined by cross-validation. The elastic net reduces to lasso when $\omega = 1$, and it becomes ridge when $\omega = 0$. In this sense, elastic net embodies both selection and shrinkage features.

In forecasting the U.S. macroeconomic variables, Kim and Swanson (2018) included lasso, ridge and elastic net, along with other machine learning methods, and showed that there is no single best method among the three regularized least squares that works for all cases. For this reason, we also consider all three types of regularized least squares methods in our analysis.

So far, our forecast methods are based on the linear combination of predictor variables. To take account of the possibility of nonlinearity in predictive regression or the effect of interaction among predictors, we also consider two classes of machine learning methods: one is ensemble approach based on regression trees (bagging, random forests, boosting) and the other is neural networks.

2.3 Ensemble machine learning based on regression trees

A decision tree is a useful machine learning method to incorporate nonlinearity. For the purpose of macroeconomic forecasting of continuous target variable y_{t+h} , a regression tree can be used to detect groups of observations which have similar characteristics. Within the tree, nodes are generated step by step. At the initial step, observations of a target variable y_{t+h} are sorted into several nodes using one of the predictor variables $X_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$. Some nodes can take values using the sample average of y_{t+h} conditional on the range of a selected predictor. Nodes without values are again divided by using the range of remaining predictors. This process continues until all the nodes take values using some stopping rule.

Specifically, a regression tree with M terminal nodes can be written as

$$y_{t+h} = \sum_{m=1}^M \theta_m \mathbf{1}_{\{X_t \in R_m\}} + \varepsilon_{t+h}, \quad (7)$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function, R_m is a partition of the space of X_t , and θ_m is the sample average of y_{t+h} conditional on $X_t \in R_m$. The estimation of regression trees is to find the best tree structure to minimize the least square criterion $\sum_{t=1}^T \varepsilon_{t+h}^2$. In order to delineate such a tree numerically, we need to choose sorting variables from X_t and set splitting values at each node. We employ a greedy algorithm of a binary regression tree, presented by Breiman et al. (1984) for the purpose of discovering the locally optimal sorting variable and splitting values at each node.

The regression trees are good at dealing with nonlinearity and interaction of variables. Furthermore, interpreting the determinants of a target variable is relatively simple and the structure of the model can be visibly understood. However, at the same time, regression trees are known to be sensitive to small change in data so out-of-sample forecast performance can become very poor. To avoid the overfitting problem, regression trees are typically employed as the basis of ensemble machine learning methods. In our analysis, we consider the following three types of ensemble procedures; bagging, random forests and boosting.

Bagging: Bagging, first proposed by Breiman (1996), comes from the abbreviation of bootstrap aggregating. Breiman (1996) provided the empirical evidence that bagging improved the forecast accuracy in various situations. Later, Bühlmann and Yu (2002) showed that bagging had the ability to reduce forecast errors for i.i.d. data, while Inoue and Kilian (2008) and Stock and Watson (2012) extended this idea to time series framework.

In bagging, bootstrap samples of original predictor variables $X_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$ and the target variable y_{t+h} are repeatedly generated B times. A regression tree is then applied to each bootstrap sample $X_t^{(b)}$ and $y_{t+h}^{(b)}$ to compute the forecast $\hat{y}_{t+h}^{(b)}$. The forecast from

bagging is then constructed as $B^{-1} \sum_{b=1}^B \hat{y}_{t+h}^{(b)}$, namely, the average of forecasts from individual bootstrap samples. Although the forecast of a regression tree in each bootstrap sample suffers from overfitting and tends to be volatile, averaging forecasts of bootstrap samples diminishes the variation and yields the stable forecast. In our application below, we set the number of bootstrap samples at $B = 10$.

Random forests: A random forest is one derivative of bagging and thus their algorithms are closely related with each other (Breiman, 2001). In order to maintain the stability of forecasts from bagging, regression trees of different bootstrap samples should not be highly correlated. Otherwise, the forecast of a regression tree in each bootstrap sample behaves similarly and averaging may not be effective in reducing the variance of forecast.

In random forests, a dropout procedure is employed to decorrelate regression trees of bootstrap samples (see Hastie et al., 2009). Namely, the number of original predictor variables $X_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$ is reduced by randomly drawing a subset of predictor variables $X_t^* = (x_{1t}^*, x_{2t}^*, \dots, x_{kt}^*)'$ where $k < N$. For a subset X_t^* , a forecast is constructed by bagging, namely by $B^{-1} \sum_{b=1}^B \hat{y}_{t+h}^{(b)}$ where $\hat{y}_{t+h}^{(b)}$ is computed using a bootstrap sample $X_t^{*(b)}$ and $y_{t+h}^{(b)}$. Repeat this procedure for many subsets of predictor variables, and compute the average of forecasts from all different subsets. This procedure reduces the correlation of regression trees between each subset because the structures of regression trees are forced to be different by subsampling. The decorrelated regression tree is expected to generate stability of forecasts. In our analysis, we set the dimension for the subset of predictor variables at $k = N/2$.

Boosting: Boosting is another ensemble machine learning algorithm to resolve the overfitting problem. Boosting was originally introduced by Schapire (1990) and Freund (1995) in classification problems. In the context of macroeconomic forecasting, Bai and Ng (2009) employed boosting in selecting predictors of the factor augmented regression.

In boosting, multiple regression trees are constructed as in the case of bagging. Let

a very simple regression tree be given by $\sum_{m=1}^M \theta_m \mathbf{1}_{\{X_t \in R_m\}}$ and set the initial regression tree by $f_0(X_t) = \eta \sum_{m=1}^M \theta_m \mathbf{1}_{\{X_t \in R_m\}}$. Here, $\eta \in (0, 1)$ is a learning rate which we set at $\eta = 0.1$. Note that the depth of all the regression trees in boosting should be relatively shallow which implies that each base learner $f_s(X_t)$, for $s = 0, 1, \dots, S$, is a weak learner. In each stage, information of forecast errors from the previous trees is utilized in searching for a new regression tree. To be more specific, we employ a gradient boosting tree algorithm with L_2 loss function so that the model at s -th stage is updated using

$$f_s(X_t) = f_{s-1}(X_t) + \eta \sum_{m=1}^{M_s} \theta_{sm} \mathbf{1}_{\{X_t \in R_{sm}\}},$$

where a new regression tree $\sum_{m=1}^{M_s} \theta_{sm} \mathbf{1}_{\{X_t \in R_{sm}\}}$ is estimated for the residual from $(s - 1)$ -th stage, $y_{t+h} - f_{s-1}(X_t)$. Continue updating the model until s reaches a total number of boosting stages, which we set at $S = 100$. See Hastie et al. (2009) and Gu et al. (2019) for more details.

2.4 Neural networks

Neural networks can also incorporate nonlinearity and interaction of variables through a very flexible functional form. The structure of a neural network can be described by three components: input layer, hidden layer and output layer.⁵ Each layer is collegiated by synapses which deliver signals of neurons in the preceding layer to the succeeding one. In our setting, the input layer corresponds to predictor variables $X_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$ so that the number of neurons in the input layer is the same as the dimension of predictors N . The hidden layer converts an output from the preceding layer (including the input layer) through an activation function. Finally, the output layer summarizes the output from the hidden layer.

For a single hidden layer feedforward neural network, the forecasting model can be written

⁵In this paper, we focus on the standard feedforward neural network.

as

$$y_{t+h} = f(X_t) + \varepsilon_{t+h}, \quad (8)$$

where

$$f(X_t) = \sum_{j=1}^q \theta_j \sigma(w'_j X_t + b_j) + b,$$

σ is an activation function and q is the number of hidden units (neurons). The model can be estimated by minimizing the least square criterion with respect to (θ_j, w'_j, b_j, b) for $j = 1, \dots, q$.

Neural network becomes more complex and flexible when we increase the number of units in a hidden layer (wider neural network) or increase the number of hidden layers between input and output layers (deeper neural network). For example, a neural network with two hidden layers can be written as

$$f(X_t) = \sum_{k=1}^{q_2} \theta_k^{(2)} \sigma \left(\sum_{j=1}^{q_1} \theta_{kj}^{(1)} \sigma(w'_j X_t + b_j) + b_k^{(1)} \right) + b^{(2)},$$

while that of three hidden layers can be written as

$$f(X_t) = \sum_{m=1}^{q_3} \theta_m^{(3)} \sigma \left(\sum_{k=1}^{q_2} \theta_{mk}^{(2)} \sigma \left(\sum_{j=1}^{q_1} \theta_{kj}^{(1)} \sigma(w'_j X_t + b_j) + b_k^{(1)} \right) + b_m^{(2)} \right) + b^{(3)},$$

where q_ℓ is the number of units for ℓ -th hidden layer.

Gu et al. (2019) examined the effect of increasing the number of hidden layers in the context of asset pricing and found that deep learning does not necessarily outperform shallow learning. They discussed that this result came from the limited sample size available in economic applications. For this reason, we need to carefully choose the width and depth of hidden layers. Since searching for the best neural network structure by cross-validation is computationally intensive, we follow the approach adopted by Gu et al. (2019) and employ neural networks with their hidden layers ranging from 1 to 5, with the number of

hidden unit q_ℓ in each layer ℓ following the geometric pyramid rule (Masters, 1993). In particular, there are 32 units ($q_1 = 32$) in the single hidden layer neural network. For the 2 hidden layers neural network, there are 32 units in layer 1 ($q_1 = 32$) and 16 units in layer 2 ($q_2 = 16$). Likewise, the combination of units for the 5 hidden layers neural network is $(q_1, q_2, q_3, q_4, q_5) = (32, 16, 8, 4, 2)$.

There are several options for the choice of activation functions. In our analysis, we employ two representative activation functions, namely, sigmoid function given by

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (9)$$

and ReLU (rectified linear unit) function

$$\sigma(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{otherwise} \end{cases}, \quad (10)$$

where z is the input of a hidden layer. Figure 1 describes the mapping of both sigmoid and ReLU function.

While sigmoid function was often employed as an activation function in the past, including Shintani (2005) who used neural networks in forecasting Japanese macroeconomic data, ReLU function has become a more popularly used activation function recently. This is because its learning speed is known to be much faster under the deep neural network environment and it can circumvent the vanishing gradient problem.⁶ We examine these two activation functions because their relative performance in the context of macroeconomic forecasting is not yet known.

⁶The vanishing gradient problem refers to the failure in computing deeper neural network models with some activation functions. For details, see Hochreiter and Schmidhuber (1997) and LeCun et al. (2015).

3 Data and evaluation method

3.1 Data

The data that we used are monthly observations of 219 Japanese macroeconomic time series from January 1973 to June 2018, resulting in the many predictors environment for macroeconomic forecasting ($N = 219$, $T = 546$, $N \times T = 119,574$).⁷ Target variables in our analysis are 7 representative macroeconomic variables, which include index of industrial production (IIP), capacity utilization ratio (UTIL), unemployment rate (UR), real wage (WAGE), real household consumption expenditure (CONS), wholesale price index (WPI), and consumer price index (CPI). Table 1 shows the transformation method of each target variable. It is noteworthy that some of our target variables are transformed using year-on-year change of the nonseasonally adjusted series (WPI and CPI) rather than month-on-month change of the seasonally adjusted series (IIP, UTIL, WAGE, and CONS). A full list of all 219 variables is presented in the Appendix.

3.2 Forecast evaluation

Our forecast methods include factor models (PCR, FAAR), regularized least squares methods (lasso, ridge, elastic net), ensemble machine learning based on regression trees (bagging, random forests, boosting) and neural networks (sigmoid and ReLU). These methods are selected to cover standard factor models and machine learning methods for forecasting and we will examine which method (factor models, machine learning, or their combination) performs the best for macroeconomic forecasting. We consider the following 3 types of forecast specifications, which are similar to the ones used in Kim and Swanson (2018), and evaluate their performance using the AR model as a benchmark.

⁷Our dataset is an updated version of Shintani (2005), which covers 235 macroeconomic series. However, due to the revision and termination of some official statistics, the number of variables has been reduced to 219.

Specification 1 (SP1): Factor models

Common factors are extracted as principal components of the entire set of predictor variables. We determine the optimal number of common factors by minimizing (4), with the maximum number of factors set at 20. Forecast models are then estimated as either PCR or FAAR. Regarding FAAR, the number of lags in the AR part is selected by BIC, with the maximum lag length set at 12. In this specification, machine learning methods are not employed.

Specification 2 (SP2): Machine learning applied to individual series

Forecast models are constructed by using one of 8 machine learning methods directly applied to all individual predictor variables. The 8 machine learning methods are lasso, ridge, elastic net (EN), bagging, random forests (RF), boosting (boost), neural network with sigmoid activation function (sigmoid) and neural network with ReLU activation function (ReLU). In this specification, factor models are not employed.

Specification 3 (SP3): Machine learning applied to common factors

Common factors are extracted as principal components of the entire set of predictor variables as in SP1. The number of factors is set at 20. Forecast models are then constructed by one of 8 machine learning methods, using 20 principal components plus lags of target variables (up to 12 lags). This specification can be interpreted as the combination of factor models and machine learning. This specification generalizes the approach used by Shintani (2005) who first extracted common factors and then applied neural networks to the estimated factors.

In our forecasting experiments, we also take into account the lagged factors for SP1 and SP3, which will be denoted by SP1L and SP3L, respectively. Because a preliminary analysis did not yield improvement from adding more than one lags, we only report the results from

the models with one lagged factor. For this reason, in addition to the AR benchmark, we have a total of 5 specifications: SP1, SP1L, SP2, SP3, and SP3L.

We evaluate the pseudo-out-of-sample forecast performance of each model specification for 7 target variables at 9 different forecast horizons ($h = 1, 2, 3, 6, 12, 18, 24, 30, 36$). For each specification, AR lag lengths, the number of factors, and the hyperparameters for machine learning are repeatedly selected at each point of the forecast. For example, at period $t = R$, we construct the forecast \hat{y}_{R+h} of a target variable y_{R+h} using the information only up to $t = R$ and evaluate the forecast error $y_{R+h} - \hat{y}_{R+h}$. For the next period $t = R + 1$, the model is reestimated using the data up to $t = R + 1$ and forecast value \hat{y}_{R+h+1} is constructed. Therefore, the entire series of common factors are recalculated when new forecasts are constructed. Furthermore, the lag length, the number of factors, and the hyperparameters may be different depending on the point of forecast, even if the model specification is unchanged.

We conduct this forecast exercise using both the rolling and recursive schemes. In the rolling scheme, we adjust the initial point of estimation so that the length of training data window is fixed at 120. In the recursive scheme, the initial point of estimation is set at 1973.01 and the sample size increases as we proceed the forecast point ahead. For instance, to construct a forecast at 2018.06 for $h = 1$, the estimation period is from 2008.06 to 2018.05 in the rolling scheme, while the estimation period is from 1973.01 to 2018.05 in the recursive scheme.

As a measure of forecast performance, we focus on mean square forecast errors (MSFEs) defined by $P^{-1} \sum_{t=R}^{T-h} (y_{t+h} - \hat{y}_{t+h})^2$, where \hat{y}_{t+h} is the forecast value for horizon h by a forecast model, R is the initial sample size in estimating the model, and $P(= T - h - R + 1)$ is the number of forecasts. If the MSFE for model 1 is smaller than the MSFE of model 2, we view that the former outperforms the latter in out-of-sample forecast. To provide some

guidance, we compute the Diebold and Mariano (DM) statistic defined by

$$DM = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}}, \quad (11)$$

where \bar{d} is the mean of the difference between squared forecast errors of model 1 (such as factor model or machine learning method) and those of model 2 (such as benchmark AR model), and $\hat{\sigma}_{\bar{d}}$ is the Newey-West standard error of the difference.⁸

4 Main results

This section presents the results of our pseudo-out-of-sample forecasting experiments. Tables 2 to 6 show the forecast performance of 5 specifications from SP1 to SP3L, in terms of their MSFEs relative to the benchmark AR model. A careful examination of Tables 2 to 6 addresses some of our research questions.

First, with some exceptions, most MSFE ratios are less than one, which indicates the better forecast accuracy of all 5 specifications compared to the benchmark AR model. For example, panel A in Table 4 reports that, for IIP forecast at horizon $h = 1$ under the rolling scheme, the lowest relative MSFE among SP2 is 0.762. In addition, because SP2 involves 8 different machine learning methods, panel B shows that the best machine learning method with the MSFE ratio of 0.762 corresponds to lasso. It is further noteworthy that the DM test statistic rejects the null hypothesis of the same forecast performance between SP2 (namely, lasso) and AR at 5% significance level. On the whole, our forecasting experiments indicate that factor models and machine learning enhance the forecast performance in many occasions, as in line with existing literatures (e.g., see Kim and Swanson, 2018, and Shintani, 2005).

⁸The DM test statistic asymptotically follows the standard normal distribution under the null hypothesis of equal MSFEs for nonnested models. However, we also report the results for nested cases because it can still be used for the purpose of informal guidance rather than the rigorous statistical inference.

Second, predictions by machine learning methods tend to perform well for medium and longer forecast horizons. To see this point, Table 7 summarizes the best pair of specification and method in terms of MSFE, based on the results from Tables 2 to 6. According to this table, when the forecast horizon is short, such as $h = 1, 2$ and 3 , the best specification varies among target variables. Moreover, in some cases (WAGE, WPI and CPI), the best relative MSFE among SP1-SP3L is larger than 1, indicating that neither factor models nor machine learning can improve forecast accuracy over that of an AR model. However, for medium to longer horizons, SP2 becomes the best specification for most cases.

Third, with respect to machine learning, Table 4 implies that allowing for nonlinearity and interaction between variables enhances the forecast performance. The best forecast method under SP2 is, in most cases, ensemble machine learning based on regression trees (bagging, random forests and boosting) and they perform significantly better than the AR model. As discussed in Section 2, the main characteristic of ensemble machine learning based on regression trees is its ability to deal with nonlinearity and interaction of variables. In this regard, Coulombe et al. (2019) claimed the importance of employing nonlinear methods, such as random forests, in forecasting U.S. macroeconomic variables, and conjectured that observed nonlinearity came from uncertainty and financial frictions. Shintani (2005) also emphasized the usefulness of a nonlinear approach in forecasting Japanese macroeconomic data. Although we leave further investigations to future research, our experiments here also suggest the importance of nonlinearity in Japanese economy.

Table 8 lists the best combination of specification, method and window scheme. Aside from three findings above, this table implies that the recursive scheme tends to be a better window scheme for shorter horizons, while the rolling scheme seems to be better for medium and longer horizons. For example, when the forecast horizons are $h = 1, 2$ and 3 , the recursive scheme performs better than the rolling scheme for 16 out of 21 results (76.2%).

However, for horizons $h = 24, 30$ and 36 , the situation is completely opposite and the recursive scheme performs better for only 6 out of 21 cases (28.6%). These results may come from the trade-off between the cases of estimation accuracy and parameter stability. Intuitively, the recursive scheme benefits from having a larger sample size for estimation accuracy when the parameters are stable, while the rolling scheme is expected to adjust quickly in the case of structural breaks (e.g., see Clark and McCracken, 2009, and Pesaran et al., 2011).

Let us now examine further on the performance of the neural network forecast. As mentioned in Section 2, the estimation of neural networks involves the choice of the activation function and the choice of width and depth of the neural network. Hence, we first examine the effect of choosing between sigmoid and ReLU activation functions, and then examine the effect of increasing the number of hidden layers. For the sake of brevity, we only discuss the results for the case of SP2 below.⁹

Table 9 shows the MSFE of ReLU relative to that of sigmoid for all target variables and for all forecast horizons. Overall, the table shows that the ReLU performs better than sigmoid for more occasions. To be more specific, MSFE of ReLU is less than that of sigmoid for 418 out of 630 combinations ($= 7$ variables $\times 9$ horizons $\times 2$ schemes $\times 5$ layers), or 66.3% of the total results.

Given the good performance of ReLU, Table 10 further examines the effect of changing the number of hidden layers on the forecast accuracy.¹⁰ For each variable and forecast horizon, the MSFE is normalized relative to that of the single layer ReLU. According to Table 10, adding more layers to the single layer ReLU does not always enhance the forecast accuracy, particularly when the target variables are IIP, UTIL, CONS and WPI. For these variables, the

⁹Similar results are also obtained for SP3 and SP3L. They are available from the authors upon request.

¹⁰While we only report the result of ReLU in the table, results of sigmoid can be obtained from authors upon request.

multiple layer ReLU is sometimes even worse than the single layer ReLU. On the other hand, if the target variables are UR, WAGE and CPI, multiple layers ReLU generally performs better than single layer ReLU. However, increasing layers does not necessarily lead to the lower MSFE in a monotonic manner. These outcomes are in line with Gu et al. (2019) who found that more hidden layers does not necessarily enhance the asset price forecast accuracy.

5 Composite forecast

In this section, we consider the further possibility of improving the out-of-sample forecast performance by employing composite forecast. Composite forecast, or the forecast combination, is motivated by the fact that, even if a model was selected as the best model, other models may still contain some meaningful information which is not captured by the best model. In our analysis, as the method of composite forecast, we consider the following 4 representative approaches, namely, arithmetic mean (AM), Bates-Granger method (BG), Granger-Ramanathan method (GR), and Bayesian model averaging (BMA).

(1) Arithmetic mean (AM): Suppose there are m types of forecast models. This method is based on a simple average with an equal weight $w_i = 1/m$ on each forecast model in constructing the composite forecast.

(2) Bates-Granger method (BG): Bates and Granger (1969) suggested using an optimal weight given by

$$w_i = \frac{\sigma_i^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \dots + \sigma_m^{-2}}, \quad (12)$$

where σ_i^2 is the MSFE of model $i \in \{1, \dots, m\}$. Composite forecast is then computed as the average of multiple forecasts weighted by w_i .

(3) Granger-Ramanathan method (GR): Granger and Ramanathan (1984) introduced a regression method to compute the optimal weight by running a regression of the realized

value of target variable y_t on multiple forecasts $f_{i,t}$ for model i or

$$y_t = \beta_1 f_{1,t} + \beta_2 f_{2,t} + \cdots + \beta_m f_{m,t} + e_t, \quad (13)$$

with restrictions $\sum_{i=1}^m \beta_i = 1$ and $\beta_i \geq 0$. Composite forecast is then computed as the average of multiple forecasts weighted by $w_i = \beta_i$.

(4) Bayesian model averaging (BMA): BMA has recently become a popular method in constructing composite forecasts. In BMA, the weight on model i is constructed using the posterior probability of the corresponding model. Using Bayes rule, this weight can be computed as

$$w_i = p(M_i|X^T) = \frac{p(X^T|M_i)p(M_i)}{\sum_{j=1}^m p(X^T|M_j)p(M_j)}, \quad (14)$$

where $p(M_i|X^T)$ is the posterior probability of model i , $p(X^T|M_i)$ is the marginal likelihood of model i , $p(M_i)$ is the prior probability of model i , and $X^T = \{X_t\}_{t=1}^T$ is the data. See Koop and Potter (2004) for an example of BMA applied to macroeconomic forecasting using dynamic factor models.

Here, we employ these 4 composite forecasts under 2 different settings. One setting is to combine forecasts from various models (e.g., PCR, FAAR, lasso, and so on) within the same specification (for each of SP1, SP1L, SP2, SP3, SP3L). The other setting is to combine forecasts from various specifications (from SP1 to SP3L).

For the first setting, we construct the composite forecast from various models at each specification. For example, when computing AM for SP2, the output is the simple average of 8 forecasts of 8 machine learning methods. Notice that we conduct this exercise under the rolling and recursive schemes, respectively.

Table 11 extends Table 7 so that comparisons are made not only across the best models from each specification but also across composite forecasts constructed by combining all

the models. Table 11 shows the usefulness of the composite forecasts using GR and BMA. In addition, a direct comparison of relative MSFEs between Table 7 and Table 11 reveals that all the values larger than unity in Table 7 disappears in Table 11. This suggests that composite forecasts can bring about the forecast improvement over the benchmark AR model for shorter horizons, while factor models or machine learning alone cannot.

For the second setting, we construct the composite forecast among different specifications. From such a composite forecast, we can examine which specification type conveys the most significant information necessary to improve the forecast accuracy. Because the composite forecast is designed to capture the signals relevant to the target output, coefficient weights on forecast models can be regarded as the relative importance of that model. Figures 2 to 5 show the coefficient weights based on the GR method, with forecast horizons on the horizontal axis. On one hand, in Figures 2 and 3, the best model in each specification as well as the AR model are combined to construct forecasts. On the other hand, in Figures 4 and 5, composite forecasts of each specification, namely, outputs from the aforementioned first setting, are further combined to construct a second stage composite forecast of all specifications.

Figures 2 and 3 suggest the large relative contribution of SP2 (machine learning only) for many target variables and horizons. This tendency becomes more apparent when the target variables are UR, WAGE, WPI and CPI, or when the horizons are longer. It should also be noted that SP3 and SP3L (machine learning applied to principal components) play an important role in many cases as well, particularly for IIP, UTIL and CONS. Meanwhile, there is some contribution from the benchmark AR model for shorter horizons, but its contribution becomes negligible for longer horizons. The same is true in Figures 4 and 5, where we use a composite forecast of composite forecasts, rather than a composite forecast of the best models.

To summarize the results in this section, we find that a composite forecast of factor models

and machine learning perform better than the forecast based solely on factor models or on machine learning. Furthermore, machine learning methods applied to principal components (SP3 and SP3L) are found to be useful in the composite forecast in terms of their relative contribution, along with factor models (SP1 and SP1L) and machine learning (SP2).

6 Conclusion

In this paper, we discuss various factor models and machine learning methods and empirically apply them in forecasting the Japanese macroeconomic data. We find that factor models and machine learning perform better than the conventional AR models in many occasions, and particularly, machine learning methods work well for longer horizons. Because the success of machine learning often comes from the ensemble approach based on regression trees, we consider the nonlinearity and interaction of variables are significant in forecasting the Japanese macroeconomic series. However, at the same time, deep neural networks, which are assumed to accommodate complex nonlinearity and interaction of variables, do not necessarily enhance the forecast accuracy. Furthermore, not only factor models and machine learning, but also model selection through composite forecast improves the forecast performance, and specifically, their joint application is beneficial for macroeconomic forecasting in Japan.

References

- Artis, M.J., Banerjee, A., Marcellino, M., 2005. Factor forecasts for the UK. *Journal of Forecasting* 24, 279-298.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191-221.
- Bai, J., Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74(4), 1133-1150.
- Bai, J., Ng, S., 2009. Boosting diffusion indices. *Journal of Applied Econometrics* 24, 607-629.
- Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *Journal of the Operational Research Society* 20, 451-468.
- Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? *Journal of Econometrics* 132(1), 169-194.
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 24(2), 123-140.
- Breiman, L., 2001. Random forest. *Machine Learning* 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. Chapman and Hall/CRC, New York, NY.
- Bühlmann, P., Yu, B., 2002. Analyzing bagging. *Annals of Statistics* 30(4), 927-961.
- Chen, J.C., Dunn, A., Hood, K., Driessen, A., Batch, A., 2019. Off to the races: a comparison of machine learning and alternative data for predicting economic indicators. In Abraham, K.G., Jarmin, R.S., Moyer, B., Shapiro, M.D., editors, *Big Data for 21st Century Economic Statistics*. University of Chicago Press, Chicago, IL, forthcoming.
- Clark, T.E., McCracken, M.W., 2009. Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics* 27(4), 441-454.
- Cook, T.R., Hall, A.S., 2017. Macroeconomic indicator forecasting with deep neural networks. Federal Reserve Bank of Kansas City, Research Working Paper, 17-11.
- Coulombe, P.G., Leroux, M., Stevanovic, D., Surprenant, S., 2019. How is machine learning useful for macroeconomic forecasting? Mimeo.

- Diebold, F.X., Shin, M., 2018. Machine learning for regularized survey forecast combination: partially-egalitarian lasso and its derivatives. *International Journal of Forecasting* 35(4), 1679-1691.
- Freund, Y., 1995. Boosting a weak learning algorithm by majority. *Information and Computation* 121(2), 256-285.
- Giannone, D., Lenza, M., Primiceri, G.E., 2019. Economic predictions with big data: the illustration of sparsity. Mimeo.
- Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. *Journal of Forecasting* 3(2), 197-204.
- Gu, S., Kelly, B., Xiu, D., 2019. Empirical asset pricing via machine learning. NBER Working Paper 25398.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning* (2nd ed.). Springer, New York, NY.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Computation* 9(8), 1735-1780.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55-67.
- 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 American Statistical Association* 103(482), 511-522.
- Kim, H.H., Swanson, N.R., 2014. Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. *Journal of Econometrics* 178(2), 352-367.
- Kim, H.H., Swanson, N.R., 2018. Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting* 34(2), 339-354.
- Kitamura, T., Koike, R., 2003. The effectiveness of forecasting methods using multiple information variables. *Monetary and Economic Studies* 21(1), 105-143.

- Koop, G., Potter, S., 2004. Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal* 7(2), 550-565.
- LeCun, Y., Bengio, Y., Hilton, G., 2015. Deep learning. *Nature* 521, 436-444.
- 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 and Economic Statistics*, forthcoming.
- Masters, T., 1993. *Practical Neural Network Recipes in C++*. Academic Press, San Diego, CA.
- Nakajima, Y., Sueishi, N., 2019. Forecasting the Japanese macroeconomy using high-dimensional data. Mimeo.
- Nakamura, E., 2005. Inflation forecasting using a neural network. *Economics Letters* 86(3), 373-378.
- Pesaran, M.H., Pick, A., Timmermann, A., 2011. Variable selection, estimation and inference for multi-period forecasting problem. *Journal of Econometrics* 164(1), 173-187.
- Schapire, R.E., 1990. The strength of weak learnability. *Machine Learning* 5(2), 197-227.
- Shintani, M., 2005. Nonlinear forecasting analysis using diffusion indexes: an application to Japan. *Journal of Money, Credit and Banking* 37(3), 517-538.
- Stock, J.H., Watson, M.W., 1999. Forecasting inflation. *Journal of Monetary Economics*, 44(2), 293-335.
- Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20(2), 147-162.
- Stock, J.H., Watson, M.W., 2012. Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics* 30(4), 481-493.
- Tibshirani, R., 1996. Regression shrinkage and selection via lasso. *Journal of the Royal Statistical Society: Series B* 58(1), 267-288.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* 67(2), 301-320.

Table 1: Forecast target variables

Series	Abbreviation	Y_t
Index of industrial production	IIP	$\ln(X_t^{SA}) - \ln(X_{t-1}^{SA})$
Index of capacity utilization ratio	UTIL	$\ln(X_t^{SA}) - \ln(X_{t-1}^{SA})$
Unemployment rate	UR	X_t^{SA}
Real wage index	WAGE	$\ln(X_t^{SA}) - \ln(X_{t-1}^{SA})$
Real household consumption expenditure	CONS	$\ln(X_t^{SA}) - \ln(X_{t-1}^{SA})$
Wholesale price index	WPI	$\ln(X_t^{OR}) - \ln(X_{t-12}^{OR})$
Consumer price index (less fresh food)	CPI	$\ln(X_t^{OR}) - \ln(X_{t-12}^{OR})$

* SA represents the seasonally adjusted series, whereas OR means the original (not seasonally adjusted) series.

Table 2: Relative MSFE and the best method under SP1

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best method under SP1									
Case of rolling scheme									
IIP	0.901	0.978	0.926	1.048	1.024	1.073**	1.036	1.054**	1.046
UTIL	0.871	0.910	0.927	0.978	1.031*	1.088**	0.974	1.029	1.040
UR	1.073	1.035	0.957	0.804	0.738***	0.706***	0.751***	0.817**	0.757***
WAGE	1.087**	1.084**	1.070*	1.164**	0.988	0.996	1.114**	1.086*	1.022
CONS	0.998	0.978	0.924	0.984	0.950	0.995	0.995	1.021	1.027
WPI	1.168***	1.261**	1.342**	1.495**	1.026	1.176	1.166	1.114**	1.158***
CPI	1.287**	1.229*	1.316**	1.290**	1.140	1.184	1.057	0.762***	0.619***
Case of recursive scheme									
IIP	0.883	0.978	0.896	0.966	0.984	0.944	0.964	0.985	0.998
UTIL	0.865	0.924	0.899	0.957	0.967	0.921*	0.973	0.978	1.003
UR	0.915	0.897	0.833*	0.826	0.794**	0.825**	0.836***	0.862***	0.873***
WAGE	1.158***	1.120***	1.118***	1.049***	0.859*	0.821***	0.792***	0.769***	0.771**
CONS	0.951	0.961	0.966	1.040	1.137	0.924	0.898	1.007	1.037
WPI	1.125	1.122	1.125	1.156	1.211	1.016	1.017	1.014	1.028
CPI	1.521***	1.136	1.089	1.052	2.021**	1.589**	1.241	0.985	1.040
Panel B: Best MSFE method under SP1									
Case of rolling scheme									
IIP	FAAR	PCR	FAAR	PCR	FAAR	FAAR	PCR	PCR	PCR
UTIL	PCR	PCR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR
UR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR
WAGE	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	FAAR	PCR
CONS	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR
WPI	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR
CPI	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR
Case of recursive scheme									
IIP	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	FAAR	PCR
UTIL	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	FAAR	PCR
UR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR
WAGE	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	FAAR	PCR
CONS	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	FAAR
WPI	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR
CPI	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	FAAR	PCR

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP1. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP1 outperforms better than AR model, whereas figures higher than 1 mean the best method under SP1 underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B reports the best (lowest) MSFE method under SP1, corresponding to the relative MSFE in panel A.

Table 3: Relative MSFE and the best method under SP1L

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best method under SP1L									
Case of rolling scheme									
IIP	0.907	0.939	0.947	1.048	1.050	1.081**	1.036	1.054**	1.046
UTIL	0.871	0.857	0.936	1.037	1.053	1.088**	0.974	1.029	1.040
UR	1.251**	0.984	0.942	0.752**	0.734***	0.619***	0.725***	0.778***	0.751***
WAGE	1.182***	1.079**	1.087**	1.202***	0.988	0.996	1.114**	1.097*	1.022
CONS	1.048	0.980	0.987	0.984	1.055	1.020	0.995	1.021	1.027
WPI	1.209**	1.293**	1.361**	1.466	1.026	1.176	1.166	1.114**	1.158***
CPI	1.317	1.241**	1.322***	1.303*	1.140	1.184	1.057	0.762***	0.619***
Case of recursive scheme									
IIP	0.899	0.898	0.901	0.979	0.977	0.944	0.964	1.005	0.998
UTIL	0.873	0.866	0.875	0.949	0.975	0.921**	0.973	1.001	1.003
UR	0.944	0.915	0.771**	0.800	0.760	0.859	0.893	0.941	0.958
WAGE	1.157**	1.140***	1.253***	1.111***	0.901	0.821***	0.792***	0.813***	0.771***
CONS	0.943	0.913	0.914	0.883	0.884	0.993	0.898	0.982*	1.011
WPI	1.136	1.124	1.070	1.044	1.163	1.016	1.017	1.014	1.028
CPI	1.528**	1.365**	1.295**	1.425**	1.432***	1.092	1.241	1.062	1.040
Panel B: Best MSFE method under SP1L									
Case of rolling scheme									
IIP	PCR	FAAR	PCR	PCR	PCR	PCR	PCR	PCR	PCR
UTIL	PCR	FAAR	PCR	PCR	PCR	PCR	PCR	PCR	PCR
UR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR
WAGE	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR
CONS	PCR	FAAR	PCR	FAAR	FAAR	FAAR	PCR	PCR	PCR
WPI	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR
CPI	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR	PCR	PCR
Case of recursive scheme									
IIP	PCR	FAAR	FAAR	PCR	FAAR	PCR	PCR	FAAR	PCR
UTIL	PCR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	FAAR	PCR
UR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR
WAGE	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR
CONS	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	FAAR	PCR
WPI	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR	PCR
CPI	FAAR	FAAR	FAAR	FAAR	FAAR	FAAR	PCR	PCR	PCR

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP1L. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP1L outperforms better than AR model, whereas figures higher than 1 mean the best method under SP1L underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B reports the best (lowest) MSFE method under SP1L, corresponding to the relative MSFE in panel A.

Table 4: Relative MSFE and the best method under SP2

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best method under SP2									
Case of rolling scheme									
IIP	0.762**	0.804	0.888*	0.943	1.012	0.974	0.887*	1.000	0.962
UTIL	0.827	0.829*	0.916	0.961	0.974	0.974	0.860	0.980	0.957
UR	1.345***	1.223**	1.064	0.698***	0.291***	0.207***	0.136***	0.093***	0.098***
WAGE	1.172***	1.143**	0.955	0.738***	0.454***	0.414***	0.483***	0.550***	0.385***
CONS	0.975	0.945	0.959	0.970	0.939	0.943**	0.925***	0.916***	0.868***
WPI	1.480***	1.200**	0.998	0.650**	0.561***	0.592***	0.463***	0.301***	0.377***
CPI	1.180**	0.970	0.912	0.462***	0.534***	0.567***	0.354***	0.203***	0.160***
Case of recursive scheme									
IIP	0.779**	0.898	0.912	0.964**	0.969	0.922	0.902*	0.987	0.981
UTIL	0.770**	0.858*	0.894	0.952***	0.956*	0.879**	0.900*	0.986	0.981
UR	1.029	0.891*	0.779***	0.568***	0.315***	0.195***	0.115***	0.094***	0.083***
WAGE	1.204***	1.188***	0.911	0.651***	0.361***	0.302***	0.331***	0.368***	0.277***
CONS	1.095***	1.035	1.035***	1.021***	0.850	0.886	0.785*	0.801***	0.811***
WPI	1.888***	1.339*	1.143	0.758**	0.558***	0.507***	0.410***	0.292***	0.387***
CPI	1.505***	1.109	1.086	0.713***	0.379***	0.775	0.264***	0.181***	0.160***
Panel B: Best MSFE method under SP2									
Case of rolling scheme									
IIP	lasso	EN	bagging	lasso	bagging	lasso	lasso	lasso	EN
UTIL	EN	EN	bagging	lasso	lasso	EN	lasso	lasso	EN
UR	EN	EN	boost	boost	boost	boost	boost	boost	boost
WAGE	lasso	boost	boost	boost	boost	boost	boost	boost	boost
CONS	EN	boost	lasso	EN	boost	bagging	bagging	bagging	boost
WPI	lasso	boost	boost	boost	boost	boost	boost	boost	boost
CPI	EN	boost	RF	RF	boost	RF	boost	boost	boost
Case of recursive scheme									
IIP	EN	lasso	lasso	lasso	lasso	lasso	lasso	lasso	lasso
UTIL	lasso	lasso	lasso	lasso	lasso	lasso	lasso	lasso	lasso
UR	lasso	lasso	lasso	boost	boost	boost	boost	boost	boost
WAGE	EN	boost	boost	boost	boost	boost	boost	boost	boost
CONS	bagging	boost	bagging	bagging	lasso	lasso	lasso	bagging	boost
WPI	boost	boost	boost	boost	boost	boost	boost	boost	boost
CPI	boost	boost	boost	boost	boost	boost	boost	boost	boost

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP2. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP2 outperforms better than AR model, whereas figures higher than 1 mean the best method under SP2 underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B reports the best (lowest) MSFE method under SP2, corresponding to the relative MSFE in panel A.

Table 5: Relative MSFE and the best method under SP3

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best method under SP3									
Case of rolling scheme									
IIP	0.865	0.915	0.928	1.005	0.994	0.988	0.908	0.971	0.991
UTIL	0.829	0.869	0.939	0.972	0.978	0.972	0.873	0.974	0.988
UR	1.263	1.389	1.235	0.949	0.696	0.557***	0.565***	0.464***	0.475***
WAGE	1.287***	1.170**	1.031	0.952	0.907	0.939*	0.884**	0.928*	0.920**
CONS	0.892**	0.870**	0.879*	0.940*	0.966	0.969	0.996	0.999	0.977
WPI	1.502***	1.426***	1.211	0.715	0.782**	0.937*	0.941	0.958	0.991
CPI	1.848**	1.174*	1.000	0.645***	0.773***	0.799***	0.799***	0.608***	0.444***
Case of recursive scheme									
IIP	0.854	0.953	1.004	0.992**	0.976	0.937	0.947	0.995	0.979**
UTIL	0.832**	0.957	0.975	0.978*	1.013	0.886**	0.931	1.006*	0.988
UR	0.997	1.003	0.998	0.866	0.714	0.689**	0.681***	0.640***	0.590***
WAGE	1.137***	1.139***	1.007**	0.971	0.829	0.764	0.678	0.748**	0.691
CONS	0.969	0.976	0.925	0.935	0.908	0.952	0.889	0.930	0.944
WPI	1.254	1.357	1.330	1.067	0.979	0.983	1.011	1.031	1.095
CPI	1.450**	1.181*	1.146	0.933	0.825	0.722***	0.703***	0.606***	0.585***
Panel B: Best MSFE method under SP3									
Case of rolling scheme									
IIP	bagging	bagging	EN	bagging	EN	lasso	EN	lasso	EN
UTIL	EN	bagging	EN	bagging	lasso	EN	lasso	lasso	lasso
UR	ridge	boost	boost	boost	boost	RF	boost	RF	RF
WAGE	lasso	boost	boost	boost	boost	bagging	RF	bagging	bagging
CONS	lasso	lasso	lasso	lasso	EN	lasso	lasso	bagging	bagging
WPI	lasso	lasso	boost	boost	boost	bagging	boost	boost	boost
CPI	boost	boost	boost	boost	boost	boost	boost	boost	boost
Case of recursive scheme									
IIP	bagging	bagging	lasso	EN	bagging	bagging	lasso	lasso	EN
UTIL	bagging	sigmoid	EN	EN	EN	lasso	sigmoid	lasso	lasso
UR	ridge	ridge	ridge	boost	boost	boost	boost	boost	boost
WAGE	EN	bagging	bagging	bagging	bagging	sigmoid	boost	RF	RF
CONS	EN	lasso	lasso	EN	EN	bagging	lasso	bagging	bagging
WPI	lasso	EN	lasso	bagging	bagging	EN	EN	EN	lasso
CPI	boost	boost	boost	boost	RF	boost	boost	boost	boost

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP3. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP3 outperforms better than AR model, whereas figures higher than 1 mean the best method under SP3 underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B reports the best (lowest) MSFE method under SP3, corresponding to the relative MSFE in panel A.

Table 6: Relative MSFE and the best method under SP3L

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best method under SP3L									
Case of rolling scheme									
IIP	0.868	0.921	0.919	0.988	1.004	0.983	0.915	0.985	0.999
UTIL	0.554	0.522	0.579	0.609	0.622	0.607	0.538	0.606	0.614
UR	1.894*	1.520	1.252	1.023	0.805	0.603***	0.588***	0.475***	0.528***
WAGE	1.294***	1.169***	1.052	0.907	0.906*	0.909*	0.814***	0.933	0.902**
CONS	0.906*	0.874**	0.907*	0.905***	0.968	0.971	0.976	0.956**	0.979
WPI	1.954***	1.736***	1.301**	0.772	0.832**	0.941***	0.932*	0.956	1.021
CPI	1.748***	1.202**	0.989	0.596***	0.758***	0.902**	0.796***	0.566***	0.455***
Case of recursive scheme									
IIP	0.786**	0.971	0.930	0.991*	0.984	0.924	0.925*	0.988	0.998
UTIL	0.783**	0.929	0.915*	0.990**	0.992	0.887**	0.927*	1.010	1.000
UR	1.061	1.009	1.004	0.799	0.655	0.677**	0.678***	0.639***	0.575***
WAGE	1.069**	1.131***	1.016**	0.991	0.817	0.751	0.720	0.747***	0.686
CONS	0.964	0.962	0.918	0.923	0.896	0.927	0.856	0.909	0.913*
WPI	1.266	1.348	1.278	1.109	0.938*	0.929***	0.950	1.004	1.079
CPI	1.334*	1.179*	1.082	0.893	0.712**	0.699**	0.698***	0.606***	0.554***
Panel B: Best MSFE method under SP3L									
Case of rolling scheme									
IIP	RF	lasso	bagging	bagging	lasso	lasso	EN	EN	EN
UTIL	EN	EN	EN	EN	EN	EN	EN	EN	EN
UR	boost	boost	boost	boost	boost	boost	boost	RF	boost
WAGE	boost	boost	boost	boost	RF	boost	boost	bagging	bagging
CONS	lasso	EN	EN	EN	lasso	EN	EN	bagging	bagging
WPI	lasso	lasso	boost	boost	bagging	bagging	boost	boost	EN
CPI	lasso	boost	boost	boost	boost	boost	boost	boost	boost
Case of recursive scheme									
IIP	lasso	bagging	EN	lasso	lasso	lasso	lasso	lasso	lasso
UTIL	lasso	bagging	EN	EN	EN	lasso	EN	lasso	EN
UR	ridge	ridge	ridge	boost	boost	boost	boost	boost	boost
WAGE	EN	bagging	bagging	bagging	lasso	sigmoid	boost	bagging	RF
CONS	EN	lasso	EN	EN	ridge	bagging	EN	bagging	bagging
WPI	lasso	lasso	lasso	bagging	bagging	EN	lasso	lasso	bagging
CPI	lasso	boost	boost	boost	boost	boost	RF	boost	boost

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP3L. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP3L outperforms better than AR model, whereas figures higher than 1 mean the best method under SP3L underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B table reports the best (lowest) MSFE method under SP3L, corresponding to the relative MSFE in panel A.

Table 7: MSFE-best specification and method under SP1-SP3L

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best specification and method under SP1-SP3L									
Case of rolling scheme									
IIP	0.762**	0.915	0.888**	0.943	0.994	0.974	0.887	0.971	0.962
UTIL	0.554	0.522	0.579	0.609	0.622	0.607	0.538	0.606	0.614
UR	1.073	0.984	0.942	0.698***	0.291***	0.207***	0.136***	0.093***	0.098***
WAGE	1.087**	1.079**	0.955	0.738***	0.454***	0.414***	0.483***	0.550***	0.385***
CONS	0.892**	0.870**	0.879*	0.905***	0.939	0.943**	0.925***	0.916***	0.868***
WPI	1.168***	1.200**	0.998	0.650**	0.561***	0.592***	0.463***	0.301***	0.377***
CPI	1.180*	0.870	0.912	0.462***	0.534***	0.567***	0.354***	0.203***	0.160***
Case of recursive scheme									
IIP	0.779**	0.898	0.896	0.964**	0.969	0.922	0.902*	0.985	0.979
UTIL	0.770**	0.858*	0.875	0.949	0.956*	0.879**	0.900*	0.978	0.981
UR	0.915	0.891*	0.771*	0.568***	0.315***	0.195***	0.115***	0.094***	0.083***
WAGE	1.069**	1.120***	0.911	0.651***	0.361***	0.302***	0.331***	0.368***	0.277***
CONS	0.943	0.913	0.914	0.883	0.850	0.886	0.785*	0.801***	0.811***
WPI	1.125	1.122	1.070	0.758**	0.558***	0.507***	0.410***	0.292***	0.387***
CPI	1.334*	1.109	1.086	0.713***	0.379***	0.699**	0.264***	0.181***	0.610***
Panel B: Best MSFE specification and method under SP1-SP3L									
Case of rolling scheme									
IIP	SP2 lasso	SP3 bagging	SP2 bagging	SP2 lasso	SP3 EN	SP2 lasso	SP2 lasso	SP3 lasso	SP2 EN
UTIL	SP3L EN	SP3L EN	SP3L EN	SP3L EN	SP3L EN	SP3L EN	SP3L EN	SP3L EN	SP3L EN
UR	SP1 FAAR	SP1L FAAR	SP1L FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
WAGE	SP1 FAAR	SP1L FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
CONS	SP3 lasso	SP3 lasso	SP3 lasso	SP3L EN	SP2 boost	SP2 bagging	SP2 bagging	SP2 bagging	SP2 bagging
WPI	SP1 FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
CPI	SP2 EN	SP2 boost	SP2 RF	SP2 RF	SP2 boost	SP2 RF	SP2 boost	SP2 boost	SP2 boost
Case of recursive scheme									
IIP	SP2 EN	SP1L FAAR	SP1 FAAR	SP2 lasso	SP2 lasso	SP2 lasso	SP2 lasso	SP1 FAAR	SP3 EN
UTIL	SP2 lasso	SP2 lasso	SP1L FAAR	SP1L FAAR	SP2 lasso	SP2 lasso	SP2 lasso	SP1 FAAR	SP2 lasso
UR	SP1 FAAR	SP2 lasso	SP1L FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
WAGE	SP3L EN	SP1 FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
CONS	SP1L FAAR	SP1L FAAR	SP1L FAAR	SP1L FAAR	SP2 lasso	SP2 lasso	SP2 lasso	SP2 bagging	SP2 boost
WPI	SP1 FAAR	SP1 FAAR	SP1L FAAR	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost	SP2 boost
CPI	SP3L lasso	SP2 boost	SP3L boost	SP2 boost	SP2 boost	SP3L boost	SP2 boost	SP2 boost	SP2 boost

Note: The figure in panel A represents the best (lowest) MSFE among the methods under SP1-SP3L. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method under SP1-SP3L outperforms better than AR model, whereas figures higher than 1 mean the best method under SP1-SP3L underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B reports the best (lowest) MSFE specification and method under each window scheme, corresponding to panel A. For each entry, the first row is specification and the second is method.

Table 8: MSFE-best specification, method and window scheme

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
IIP	SP2	SP1L	SP1	SP2	SP2	SP2	SP2	SP3	SP2
	lasso	FAAR	FAAR	lasso	lasso	lasso	lasso	lasso	EN
	rolling	recursive	recursive	rolling	recursive	recursive	rolling	rolling	rolling
UTIL	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L
	EN	EN	EN	EN	EN	EN	EN	EN	EN
	rolling	rolling	rolling	rolling	rolling	rolling	rolling	rolling	rolling
UR	SP1	SP2	SP1L	SP2	SP2	SP2	SP2	SP2	SP2
	FAAR	lasso	FAAR	boost	boost	boost	boost	boost	boost
	recursive	recursive	recursive	recursive	rolling	recursive	recursive	rolling	recursive
WAGE	Bench	Bench	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	AR	AR	boost	boost	boost	boost	boost	boost	boost
	recursive	recursive	recursive	rolling	recursive	rolling	rolling	recursive	rolling
CONS	SP1L	SP1L	SP1L	SP1L	SP2	SP2	SP2	SP2	SP2
	FAAR	FAAR	FAAR	FAAR	lasso	bagging	bagging	bagging	boost
	recursive	recursive	recursive	recursive	recursive	rolling	rolling	recursive	rolling
WPI	Bench	Bench	Bench	SP2	SP2	SP2	SP2	SP2	SP2
	AR	AR	AR	boost	boost	boost	boost	boost	boost
	recursive	recursive	recursive	rolling	recursive	recursive	rolling	rolling	recursive
CPI	Bench	Bench	Bench	SP2	SP2	SP2	SP2	SP2	SP2
	AR	AR	AR	RF	boost	RF	boost	boost	boost
	rolling	recursive	recursive	rolling	rolling	rolling	recursive	rolling	rolling

Note: This table reports the best (lowest) MSFE specification, method and window scheme. For each entry, the first row is specification, the second is method and the third is window scheme.

Table 9: Relative MSFE of ReLU to sigmoid

	# of layer	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Case of rolling scheme										
IIP	1	0.607***	0.556***	0.679***	0.713***	0.669***	0.610***	0.796***	0.697***	0.772***
	2	0.348***	0.374***	0.429***	0.402***	0.366***	0.337***	0.490***	0.438***	0.525***
	3	0.455***	0.449***	0.674***	0.614***	0.536***	0.510***	0.495***	0.543***	0.574***
	4	0.528***	0.473***	0.714***	0.724**	0.562***	0.656**	0.641***	0.663***	0.586***
	5	1.296	1.196*	1.184*	1.245**	1.300***	1.236**	1.284***	1.059	1.211**
UTIL	1	0.507***	0.467***	0.563***	0.541***	0.558***	0.514***	0.680***	0.603***	0.665***
	2	0.294***	0.251***	0.306***	0.316***	0.274***	0.284***	0.403***	0.350***	0.352***
	3	0.481***	0.535***	0.597***	0.549***	0.458***	0.482***	0.608***	0.527***	0.501***
	4	0.408***	0.424***	0.592***	0.471***	0.513***	0.446***	0.597***	0.529***	0.489***
	5	1.168	1.120	1.172	1.195*	1.404***	1.263**	1.400***	1.261**	1.374***
UR	1	3.282***	3.396***	3.145***	2.904***	2.765***	3.214***	3.253***	3.718***	3.463***
	2	7.199***	6.475***	6.449***	6.528***	5.663***	6.822***	7.531***	6.566***	6.004***
	3	3.981***	4.207***	4.291***	4.074***	4.137***	4.356***	3.350***	4.804***	3.664***
	4	1.669***	1.847***	1.744***	1.944***	1.738***	1.868***	2.036***	2.282***	2.232***
	5	1.194	1.271	1.474*	1.255*	1.093	1.403**	1.487**	1.332*	1.412**
WAGE	1	1.313***	1.224**	1.139*	1.183***	1.057	1.061	1.135	1.171**	1.059
	2	1.190	0.935	1.056	0.996	0.880*	0.834***	0.824***	0.914*	0.751***
	3	1.008	0.860**	0.831***	0.875**	0.835**	0.856**	0.716***	0.834**	0.784***
	4	1.067	0.879*	0.956	0.811**	0.843**	0.936	0.857**	0.906	0.748***
	5	0.817**	0.847*	0.930	0.829**	0.944	1.057	1.030	1.027	0.894
CONS	1	0.989	0.948	0.901	0.824***	0.869	0.828*	0.878***	0.820**	0.909
	2	0.742***	0.717***	0.648***	0.604***	0.618***	0.678***	0.625***	0.694***	0.743***
	3	0.665***	0.775***	0.761***	0.609***	0.663***	0.692***	0.684***	0.725***	0.794***
	4	0.868	0.913	0.820**	0.805***	0.706***	0.790***	0.891	0.845**	0.914
	5	1.373***	1.178***	1.104	1.148**	1.178***	1.339***	1.287***	1.126	1.195**
WPI	1	0.596**	0.772***	0.699***	0.674***	0.747***	0.869**	0.777***	0.734***	0.671***
	2	0.345***	0.471***	0.455***	0.380***	0.488***	0.694***	0.506***	0.491***	0.442***
	3	0.302***	0.327***	0.334***	0.423***	0.500***	0.645***	0.491***	0.472***	0.474***
	4	0.220***	0.300***	0.328***	0.405***	0.565***	0.680**	0.633***	0.536***	0.560***
	5	0.408***	0.570***	0.651***	0.913**	0.940	1.428**	1.308*	1.158	1.277**
CPI	1	2.287***	2.456***	2.168***	1.558***	1.297***	1.420**	1.363***	1.363***	1.278***
	2	0.979	0.684	0.975	0.931	0.706*	0.618***	0.711***	1.093	0.696***
	3	0.638	0.722	0.815	0.907	0.723**	0.842	0.881	0.809	0.735**
	4	0.531	0.620	0.774	0.598	0.959	1.000	0.720	0.558**	0.889
	5	0.574***	0.706**	0.835	0.831***	1.164	1.454**	1.179	1.130	1.217**
Case of recursive scheme										
IIP	1	1.083	1.011	1.095	1.114	0.977	1.034	1.100	0.973	0.985
	2	0.878	0.859*	0.868	0.851**	0.781***	0.830***	0.963	0.841	0.770***
	3	0.954	0.983	0.880*	0.908	0.866	0.904*	0.925	0.883	0.859**
	4	0.888	0.976	0.919	0.905	0.912	0.952	0.888	0.879	0.919
	5	1.183	1.281***	1.308***	1.431***	1.275***	1.431***	1.614***	1.333***	1.489***
UTIL	1	1.011	1.012	0.928	0.978	0.871**	0.979	1.060	0.997	0.988
	2	0.944	0.675***	0.649***	0.803**	0.694***	0.704***	0.795**	0.677***	0.706***
	3	0.927	1.062	0.907	0.869*	0.819**	1.050	0.982	0.855*	0.868**
	4	0.916	1.130	0.851	1.029	0.884	0.895	1.018	1.051	0.888**
	5	1.285**	1.495*	1.326***	1.221**	1.347***	1.462***	1.373***	1.314***	1.503***
UR	1	4.553***	4.552***	3.957***	4.123***	4.918***	4.239***	3.713***	4.060***	4.203***
	2	2.250***	1.792***	1.939***	1.922***	2.067***	2.290***	2.274***	2.608***	2.824***
	3	1.312***	1.272**	1.217	1.322**	1.326***	1.540***	1.562***	2.023***	1.672***
	4	0.552***	0.573***	0.590***	0.544***	0.726**	0.600***	0.727**	0.740**	0.783**
	5	0.237***	0.266***	0.278***	0.300***	0.370***	0.369***	0.461***	0.360***	0.503***
WAGE	1	1.007*	1.035	1.047	0.892**	0.854**	1.186**	1.196**	1.187*	1.254***
	2	0.642***	0.647***	0.762***	0.763***	0.646***	0.914	0.909	0.904	1.018
	3	0.738***	0.778***	0.591***	0.625***	0.505***	0.741***	0.857**	0.931	0.839**
	4	0.975	0.808***	0.731***	0.608***	0.573***	0.828***	0.794***	0.887**	0.910
	5	0.959	0.896	0.917	0.912	1.259***	1.331***	1.099	1.412***	1.401***
CONS	1	0.882*	0.848**	0.830**	0.764***	1.058	1.218***	1.169**	1.070	1.121*
	2	0.572***	0.513***	0.473***	0.409***	0.533***	0.935	0.950	0.870**	0.979
	3	0.615***	0.582***	0.529***	0.581***	0.711***	0.830**	0.776***	1.006	0.902*
	4	0.607***	0.601***	0.848*	0.502***	0.723***	0.972	0.808***	1.109	0.953
	5	1.107	1.260***	1.080	1.139	1.070	1.348***	1.289***	1.149*	1.323***
WPI	1	0.397***	0.446***	0.409***	0.527***	0.962	0.927**	0.905***	0.820***	0.820***
	2	0.118***	0.156***	0.158***	0.267***	0.546***	0.621***	0.610***	0.530***	0.443***
	3	0.078***	0.110***	0.163***	0.306***	0.573***	0.636***	0.540***	0.577***	0.464***
	4	0.272***	0.254***	0.240***	0.484***	0.803**	0.787***	0.621***	0.717***	0.523***
	5	0.738***	0.606***	0.651***	0.920**	1.583***	1.718***	1.544***	1.323***	1.396***
CPI	1	0.525***	0.534***	0.657***	0.518***	1.222	1.385***	1.733***	1.344***	1.297**
	2	0.346***	0.554***	0.272**	0.320***	0.813**	1.083	1.069	0.951	0.729***
	3	0.327***	0.240**	0.478**	0.177***	0.743	1.294***	0.990	0.903	0.865
	4	0.220***	0.279*	0.268*	0.300***	0.699	0.844	1.121	0.953	0.983
	5	0.525**	0.964	0.794*	0.804	0.859**	0.900	1.027	1.160	0.997

Note: The figure in this table represents the MSFE of ReLU relative to that of sigmoid. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test.

Table 10: Relative MSFE of multiple layers ReLU to 1 layer ReLU

	# of layer	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Case of rolling scheme										
	1	—	—	—	—	—	—	—	—	—
IIP	2	0.896	0.919	0.934	0.805**	0.842**	0.915**	0.884	1.001	1.054
	3	0.772	0.885	0.996	1.001	0.961	0.956	0.809	0.955	0.942
	4	0.884	0.914	0.948	0.946	0.886	0.988	0.981	0.938	0.898
	5	1.034	0.994	0.911	1.012	0.982	1.006	0.838	0.913	0.953
	1	—	—	—	—	—	—	—	—	—
UTIL	2	0.963	0.915	0.889	1.078	0.915	0.985	1.012	0.895	0.896
	3	1.000	1.009	0.951	1.077	0.909	0.946	0.974	0.936	0.917
	4	0.966	1.000	1.009	1.027	0.948	0.880	1.082	0.843	0.848
	5	1.026	0.990	0.916	1.053	1.012	0.949	0.937	0.989	0.942
	1	—	—	—	—	—	—	—	—	—
UR	2	0.759**	0.641***	0.681***	0.740**	0.685***	0.690***	0.765**	0.647***	0.685***
	3	0.550***	0.526***	0.587***	0.605***	0.615***	0.581***	0.461***	0.621***	0.495***
	4	0.425***	0.422***	0.439***	0.529***	0.471***	0.444***	0.504***	0.512***	0.530***
	5	0.394***	0.385***	0.495***	0.433***	0.373***	0.442***	0.462***	0.400***	0.441***
	1	—	—	—	—	—	—	—	—	—
WAGE	2	0.968	0.876*	0.985	0.907**	0.922	0.841	0.800	0.834**	0.804**
	3	0.901*	0.830**	0.857***	0.831***	0.902	0.876	0.772	0.799***	0.912
	4	0.895*	0.872**	0.921**	0.792***	0.876**	0.892	0.856	0.819**	0.770*
	5	0.835**	0.792***	0.902**	0.826***	0.829**	0.871	0.844	0.847**	0.823*
	1	—	—	—	—	—	—	—	—	—
CONS	2	0.979	0.984	0.974	0.947	0.978	1.069	0.979	1.077	1.048
	3	0.924	1.055	1.085	0.938	1.034	1.080	0.982	1.087	1.064
	4	0.971	1.034	1.015	0.948	0.985	1.003	1.123**	1.111	1.089
	5	1.030	0.937	0.961	1.009	1.133	1.134	1.067**	1.075	1.022
	1	—	—	—	—	—	—	—	—	—
WPI	2	1.012	0.952	0.984	0.824**	0.964	1.083	0.858*	0.919**	0.947
	3	1.138	0.830	0.882	0.909	0.992	1.121	0.929	0.831**	0.984
	4	0.849***	0.849	0.929	0.984	1.000	1.057	1.015	0.814***	1.029
	5	0.799***	0.848	0.947	0.970	0.872	1.144	1.048	0.922	0.996
	1	—	—	—	—	—	—	—	—	—
CPI	2	0.770***	0.437***	0.710***	0.729***	0.776**	0.586***	0.760**	1.023	0.712***
	3	0.564***	0.433***	0.524***	0.629***	0.629***	0.647***	0.805	0.620***	0.570***
	4	0.438***	0.399***	0.518***	0.642***	0.607***	0.681**	0.581***	0.572***	0.590***
	5	0.507***	0.386***	0.521***	0.543***	0.624***	0.630***	0.560***	0.632***	0.593***
Case of recursive scheme										
	1	—	—	—	—	—	—	—	—	—
IIP	2	0.983	0.942	1.025	0.964	1.057	1.053	1.050	1.071	1.056
	3	1.042	1.069	0.944	0.986	1.093	1.055	1.008	1.108	1.092
	4	1.010	1.084	1.106	1.025	1.209**	1.003	0.896	1.065	1.132
	5	0.948	1.037	0.984	0.977	1.057	1.042	1.165	0.992	1.119
	1	—	—	—	—	—	—	—	—	—
UTIL	2	1.123	0.888	0.876*	1.020	1.045	0.970	0.952	0.905*	0.967
	3	0.966	1.050	1.052	0.950	1.014	1.026	0.997	0.919	1.041
	4	0.969	1.144*	0.958	1.175**	1.021	0.863*	0.986	1.092	1.061
	5	1.060	1.115	1.035	0.908	1.114*	0.950	0.900*	0.939	1.066
	1	—	—	—	—	—	—	—	—	—
UR	2	0.791**	0.627***	0.764*	0.725***	0.642***	0.798***	0.800*	0.820**	0.886
	3	0.500***	0.476***	0.553***	0.559***	0.477***	0.645***	0.739***	0.872	0.753
	4	0.419***	0.515***	0.517***	0.487***	0.550***	0.515***	0.588***	0.594***	0.585***
	5	0.365***	0.380***	0.477***	0.500***	0.486***	0.491***	0.629***	0.449***	0.601***
	1	—	—	—	—	—	—	—	—	—
WAGE	2	1.097	0.939	0.976	1.128	1.052	0.934	0.893	0.886	0.966
	3	1.224**	1.056	0.899	1.024	1.127*	0.895	1.044	1.052	0.972
	4	1.305***	0.954	1.027	1.179*	1.135	0.981	1.087	0.993	1.064
	5	1.123	1.016	0.990	1.219**	1.273***	0.999	1.002	1.138**	1.043
	1	—	—	—	—	—	—	—	—	—
CONS	2	1.007	0.996	0.875**	0.965	0.955	0.999	1.123*	1.029	1.127*
	3	0.984	1.083	0.944	1.049	1.031	1.014	0.969	1.219***	1.048
	4	0.944	1.035	1.053	0.973	0.954	1.126	1.057	1.308***	1.108
	5	1.099	1.246***	1.015	1.167**	1.054	1.111	1.171**	1.123**	1.140**
	1	—	—	—	—	—	—	—	—	—
WPI	2	0.834	0.871	0.910	0.804**	0.819*	1.008	1.028	1.019	0.904
	3	0.848	0.964	0.944	0.929	0.919	0.969	1.057	1.131	0.958
	4	0.854	1.035	0.975	0.863	0.862	1.044	1.023	1.084	0.831
	5	0.893*	0.868	0.974	0.988	0.899	0.978	1.060	1.035	0.992
	1	—	—	—	—	—	—	—	—	—
CPI	2	0.863	0.986	0.724**	0.774***	0.744**	0.814**	0.662***	0.855	0.735***
	3	0.778*	0.721*	0.676**	0.632**	0.735***	0.746**	0.613***	0.793***	0.820*
	4	0.993	0.705	0.611***	0.923	0.739*	0.650***	0.670***	0.728***	0.706**
	5	1.294*	1.516**	1.024	1.059	0.630***	0.601***	0.607***	0.723***	0.731**

Note: The figure in this table represents the MSFE of multiple layers ReLU relative to that of 1 layer ReLU. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test.

Table 11: MSFE-best specification and method including forecast combination under SP1-SP3L

	$h = 1$	$h = 2$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
Panel A: Relative MSFE for the best specification and method under SP1-SP3L									
Case of rolling scheme									
IIP	0.679**	0.784	0.823	0.943	0.990	0.929	0.799**	0.927	0.903
UTIL	0.554	0.522	0.579	0.609	0.622	0.607	0.538	0.606	0.614
UR	0.360	0.700	0.617	0.442*	0.181***	0.183***	0.119***	0.074***	0.083***
WAGE	0.215***	0.675***	0.689***	0.588***	0.404***	0.387***	0.441***	0.456***	0.348***
CONS	0.854**	0.817***	0.825***	0.875***	0.875***	0.853***	0.863***	0.846***	0.811***
WPI	0.847	0.527***	0.554***	0.488***	0.428***	0.334***	0.432***	0.270***	0.377***
CPI	0.352	0.478***	0.567***	0.413***	0.393***	0.491***	0.326***	0.186***	0.147***
Case of recursive scheme									
IIP	0.546**	0.898	0.888	0.964**	0.953	0.898	0.835*	0.985	0.807
UTIL	0.715**	0.787	0.875	0.793	0.869	0.820***	0.874*	0.928	0.912
UR	0.198	0.574	0.537	0.359	0.172***	0.169***	0.093***	0.083***	0.055***
WAGE	0.294***	0.701*	0.632***	0.520***	0.331***	0.281***	0.303***	0.352***	0.227***
CONS	0.895	0.878	0.878**	0.883	0.797**	0.797***	0.764**	0.757***	0.752***
PPI	0.491	0.569**	0.724*	0.729**	0.461***	0.296***	0.378***	0.269***	0.387***
CPI	0.410	0.595**	0.701**	0.601***	0.329***	0.441***	0.239***	0.160***	0.160***
Panel B: Best MSFE specification and method under SP1-SP3L									
Case of rolling scheme									
IIP	SP2	SP1	SP2	SP2	SP2	SP3	SP2	SP3	SP2
	GR	BMA	BMA	lasso	BMA	BMA	GR	GR	GR
UTIL	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L	SP3L
	EN	EN	EN	EN	EN	EN	EN	EN	EN
UR	SP1	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	BMA	GR	GR	GR	GR
WAGE	SP1	SP1	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	GR	GR	GR	GR	GR
CONS	SP3	SP3L	SP3	SP2	SP2	SP2	SP2	SP2	SP2
	GR	GR	GR	BMA	GR	BMA	GR	BMA	GR
WPI	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	BMA	BMA	GR	GR	boost
CPI	SP1	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	GR	GR	GR	GR	GR	GR
Case of recursive scheme									
IIP	SP2	SP1L	SP2	SP2	SP3	SP3	SP2	SP1	SP3L
	GR	FAAR	BMA	lasso	GR	GR	BMA	FAAR	GR
UTIL	SP3L	SP2	SP1L	SP2	SP2	SP2	SP1	SP2	SP2
	GR	GR	FAAR	GR	GR	BMA	BMA	BMA	GR
UR	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	BMA	GR	GR	GR	GR
WAGE	SP1	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	BMA	GR	BMA	BMA	BMA
CONS	SP1	SP3L	SP3L	SP1L	SP3L	SP2	SP2	SP2	SP2
	BMA	GR	GR	FAAR	GR	BMA	GR	BMA	GR
WPI	SP3L	SP3	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	BMA	BMA	BMA	BMA	boost
CPI	SP3L	SP2	SP2	SP2	SP2	SP2	SP2	SP2	SP2
	BMA	BMA	BMA	BMA	GR	BMA	BMA	GR	boost

Note: The figure in panel A represents the best (lowest) MSFE among the methods including composite forecast under SP1-SP3L. MSFE is normalized relative to AR model. Therefore, figures lower than 1 represent the best method including composite forecast under SP1-SP3L outperforms better than AR model, whereas figures higher than 1 mean the best method including composite forecast under SP1-SP3L underperforms worse than AR model. ***, **, * stand for 1%, 5%, 10% significance of Diebold-Mariano test. Panel B table reports the best (lowest) MSFE specification and method including composite forecast under each window scheme, corresponding to the relative MSFE in panel A. For each entry, the first row is specification and the second is method.

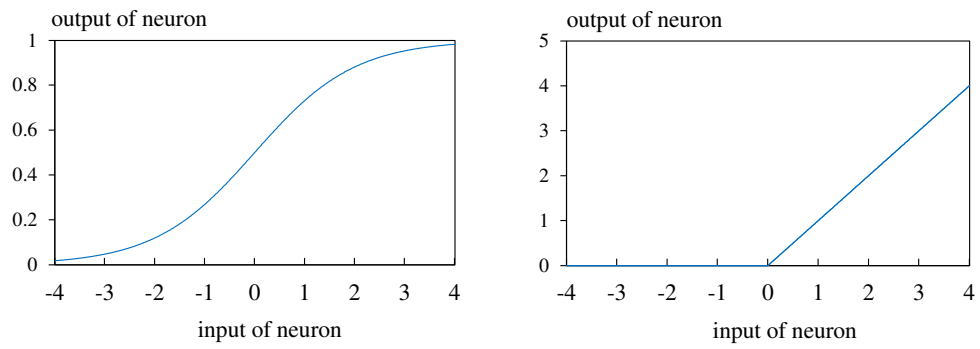


Figure 1: Activation function (left: sigmoid, right: ReLU)

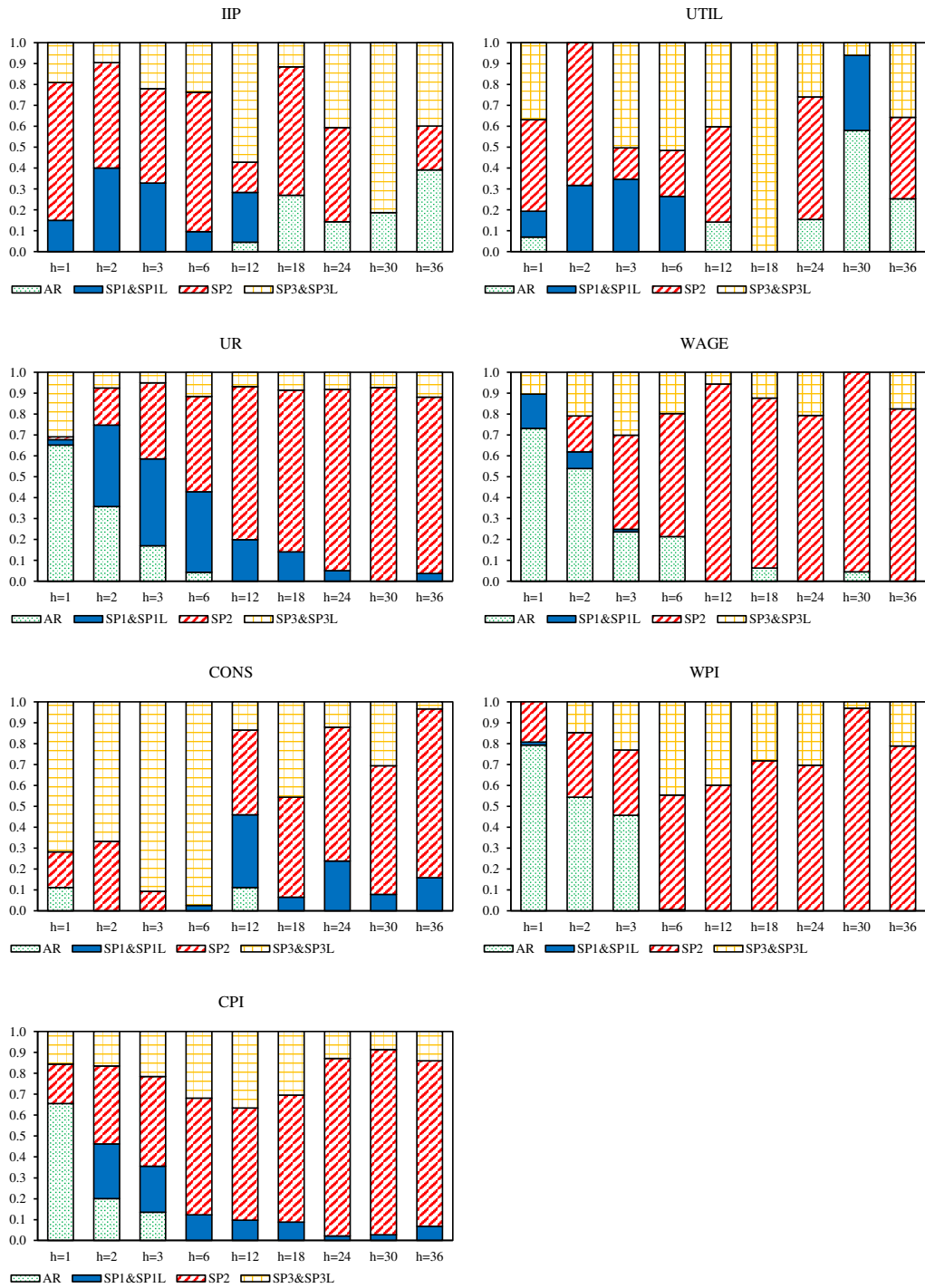


Figure 2: Coefficient weights on GR method for the best model in each specification (rolling)

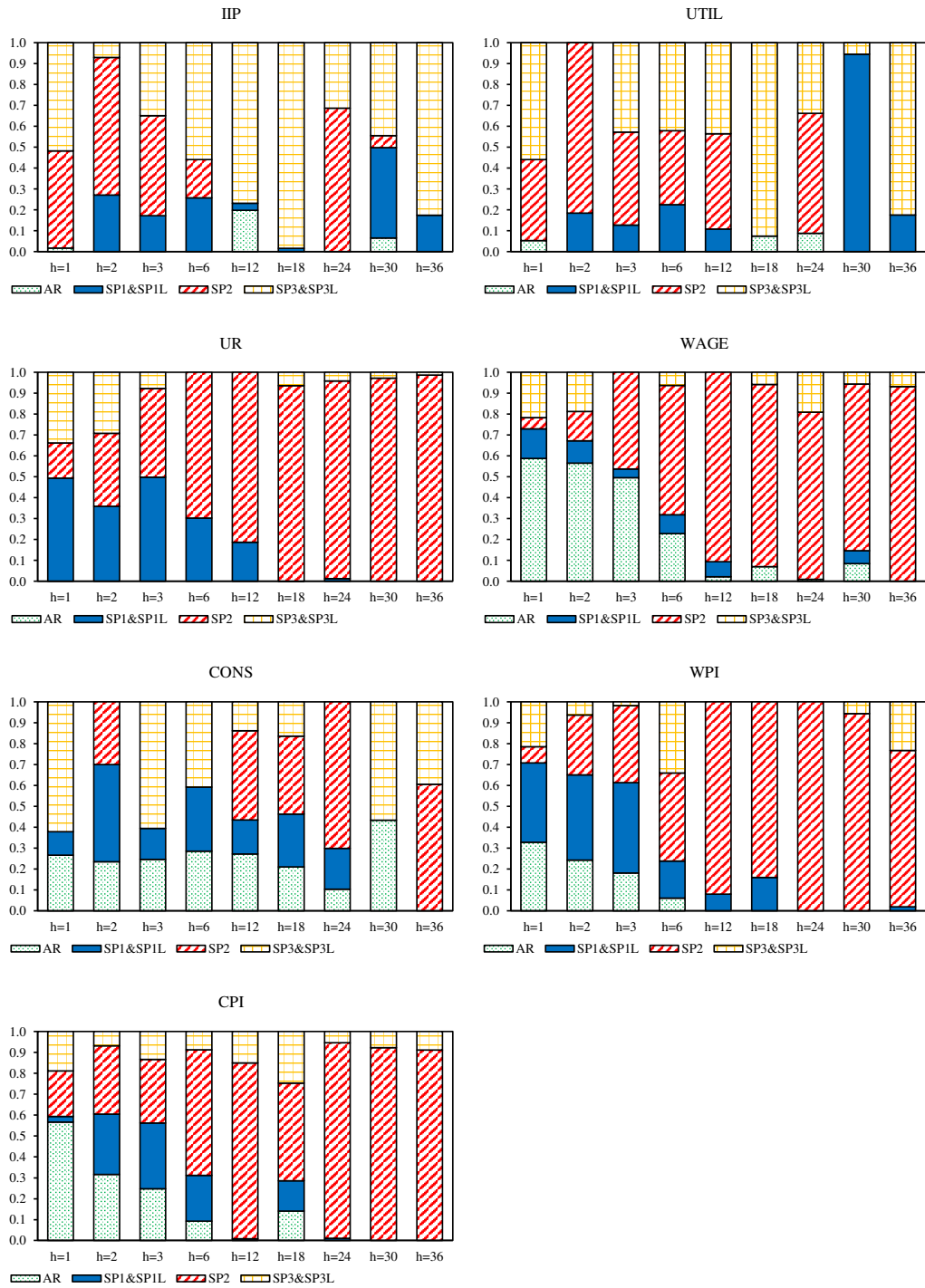


Figure 3: Coefficient weights on GR method for the best model in each specification (recursive)

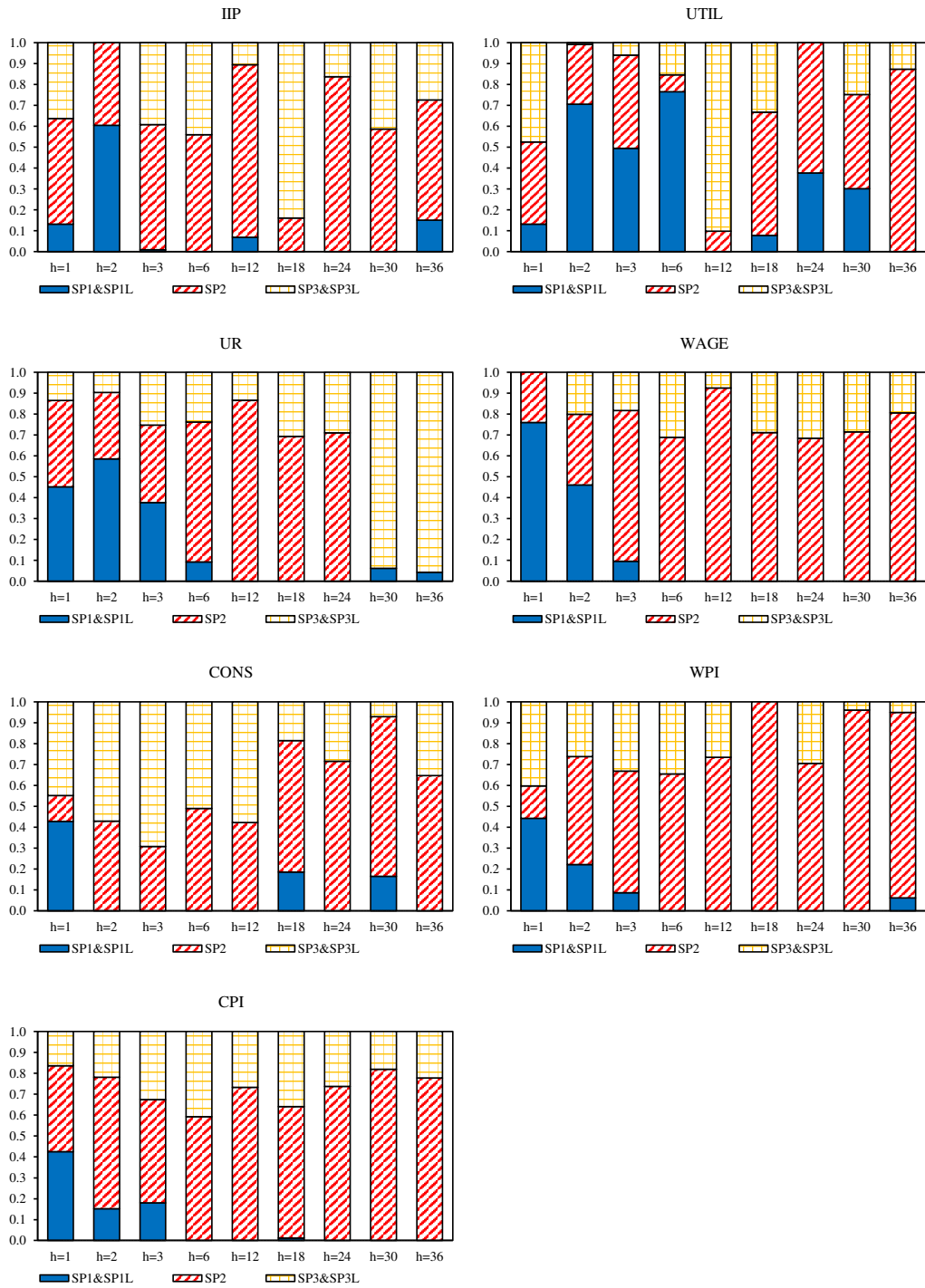


Figure 4: Coefficient weights on GR method for the composite forecast in each specification (rolling)

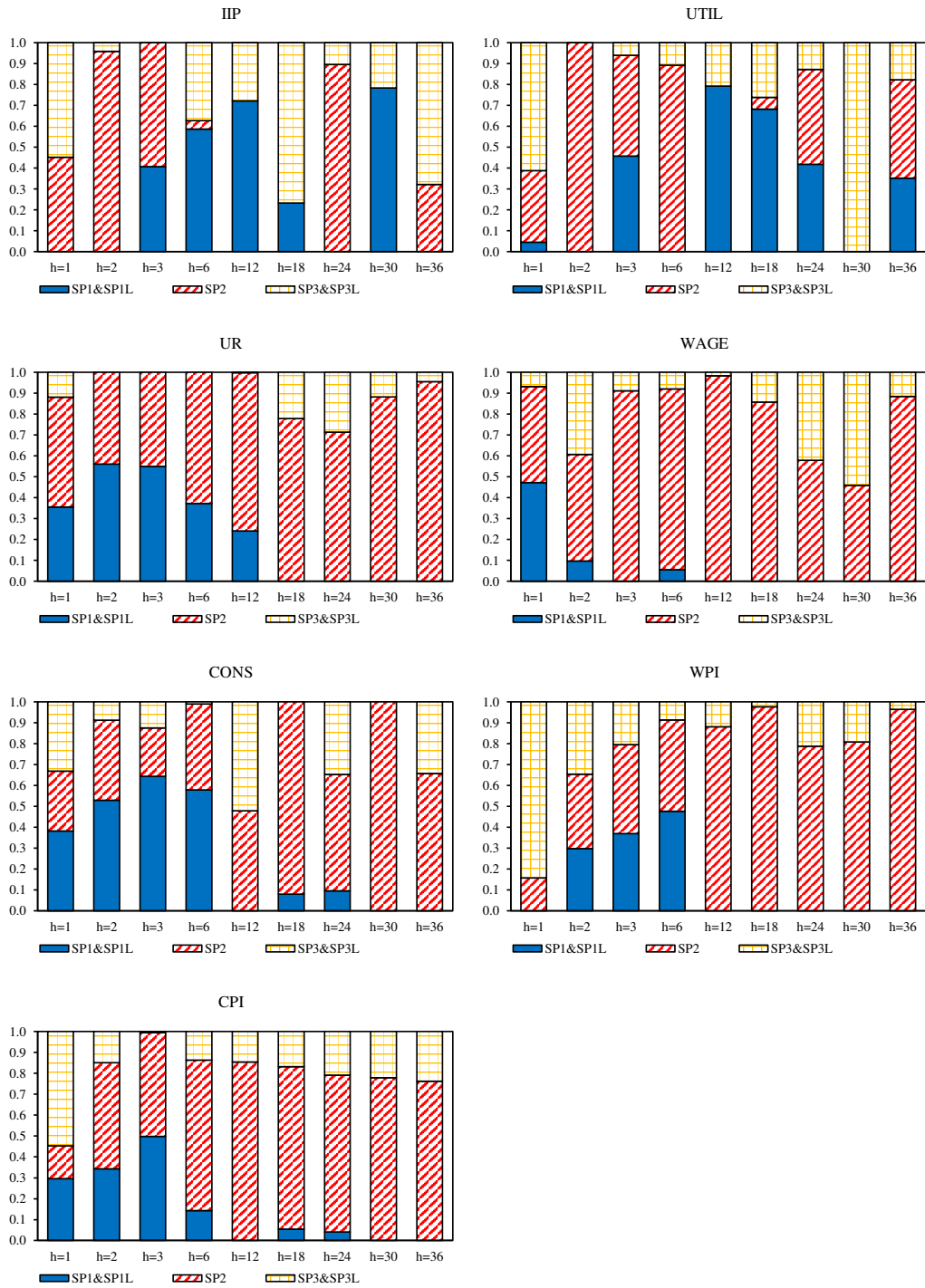


Figure 5: Coefficient weights on GR method for the composite forecast in each specification (recursive)

APPENDIX: DATA DESCRIPTION

This appendix lists the monthly observations of 219 Japanese macroeconomic time series from January 1973 to June 2018, which we used as our “big data” environment. In order to induce stationarity, most of the variables were transformed using the first difference of the log of seasonally adjusted series, with some exceptions. These exceptions include the level of the series itself and the year-on-year change of the original series. How we transformed each data is notified in the “T-Code”¹ column in the table below.

No.	Description	T-Code
	<i>Real Output</i>	
1	Index of Industrial Production (Mining and Manufacturing)	3
2	Index of Industrial Production (Manufacturing)	3
3	Index of Industrial Production (Mining)	3
4	Index of Industrial Production (Iron and Steel)	3
5	Index of Industrial Production (Non-Ferrous Metals)	3
6	Index of Industrial Production (Fabricated Metals)	3
7	Index of Industrial Production (General Machinery)	3
8	Index of Industrial Production (Electrical Machinery)	3
9	Index of Industrial Production (Transport Equipment)	3
10	Index of Industrial Production (Precision Instruments)	3
11	Index of Industrial Production (Ceramics, Clay and Stone Products)	3
12	Index of Industrial Production (Chemicals)	3
13	Index of Industrial Production (Petroleum and Coal Products)	3
14	Index of Industrial Production (Plastic Products)	3
15	Index of Industrial Production (Pulp, Paper and Paper Products)	3
16	Index of Industrial Production (Textiles)	3
17	Index of Industrial Production (Foods and Tobacco)	3
18	Index of Industrial Production (Other Manufacturing)	3
19	Index of Industrial Production (Final Demand Goods)	3
20	Index of Industrial Production (Producer Goods)	3
21	Index of Industrial Production (Producer Goods for Mining and Manufacturing)	3
22	Index of Industrial Production (Producer Goods for Others)	3
23	Index of Producer’s Shipments (Final Demand Goods)	3
24	Index of Producer’s Shipments (Producer Goods)	3
25	Index of Producer’s Shipments (Producer Goods for Mining and Manufacturing)	3
26	Index of Producer’s Shipments (Producer Goods for Others)	3
27	Index of Capacity Utilization Ratio (Manufacturing)	3
28	Index of Capacity Utilization Ratio (Iron and Steel)	3
29	Index of Capacity Utilization Ratio (Non-Ferrous Metals)	3
30	Index of Capacity Utilization Ratio (Fabricated Metals)	3
31	Index of Capacity Utilization Ratio (General Machinery)	3
32	Index of Capacity Utilization Ratio (Electrical Machinery)	3
33	Index of Capacity Utilization Ratio (Transport Equipment)	3
34	Index of Capacity Utilization Ratio (Precision Instruments)	3
35	Index of Capacity Utilization Ratio (Ceramics, Clay and Stone Products)	3
36	Index of Capacity Utilization Ratio (Chemicals)	3
37	Index of Capacity Utilization Ratio (Petroleum and Coal Products)	3
38	Index of Capacity Utilization Ratio (Textiles)	3
39	Index of Capacity Utilization Ratio (Rubber Products)	3
40	Index of Capacity Utilization Ratio (Machinery)	3
41	Index of Tertiary Industry Activity (Total)	3
42	Index of Tertiary Industry Activity (Electricity, Gas, Heat and Water Supply)	3
43	Index of Tertiary Industry Activity (Transport and Communication)	3
44	Index of Tertiary Industry Activity (Transport)	3
45	Index of Tertiary Industry Activity (Wholesale, Retail Trade, Eating and Drinking Places)	3
46	Index of Tertiary Industry Activity (Eating and Drinking Places)	3

continued

¹T-Code 1: the level of the series itself; T-Code 2: the year-on-year change of the original series; T-Code 3: the first difference of the log of the seasonally adjusted series.

continued

No.	Description	T-Code
47	Index of Tertiary Industry Activity (Finance and Insurance)	3
48	Index of Tertiary Industry Activity (Real Estate)	3
49	Index of Tertiary Industry Activity (Services)	3
50	Index of Tertiary Industry Activity (Personal Services)	3
51	Index of Tertiary Industry Activity (Business Services)	3
<i>Inventories</i>		
52	Index of Producer's Inventory Ratio of Finished Goods (Mining and Manufacturing)	3
53	Index of Producer's Inventory Ratio of Finished Goods (Final Demand Goods)	3
54	Index of Producer's Inventory Ratio of Finished Goods (Investment Goods)	3
55	Index of Producer's Inventory Ratio of Finished Goods (Capital Goods)	3
56	Index of Producer's Inventory Ratio of Finished Goods (Construction Goods)	3
57	Index of Producer's Inventory Ratio of Finished Goods (Consumer Goods)	3
58	Index of Producer's Inventory Ratio of Finished Goods (Durable Consumer Goods)	3
59	Index of Producer's Inventory Ratio of Finished Goods (Non-Durable Consumer Goods)	3
60	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods)	3
61	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods for Mining and Manufacturing)	3
62	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods for Others)	3
63	Index of Producer's Inventory of Finished Goods (Mining and Manufacturing)	3
64	Index of Producer's Inventory of Finished Goods (Final Demand Goods)	3
65	Index of Producer's Inventory of Finished Goods (Investment Goods)	3
66	Index of Producer's Inventory of Finished Goods (Capital Goods)	3
67	Index of Producer's Inventory of Finished Goods (Construction Goods)	3
68	Index of Producer's Inventory of Finished Goods (Consumer Goods)	3
69	Index of Producer's Inventory of Finished Goods (Durable Consumer Goods)	3
70	Index of Producer's Inventory of Finished Goods (Non-Durable Consumer Goods)	3
71	Index of Producer's Inventory of Finished Goods (Producer Goods)	3
72	Index of Producer's Inventory of Finished Goods (Producer Goods for Mining and Manufacturing)	3
73	Index of Producer's Inventory of Finished Goods (Producer Goods for Others)	3
<i>Investments</i>		
74	Index of Producer's Shipments (Investment Goods Excluding Transport Equipments)	3
75	Index of Producer's Shipments (Producer Goods)	3
76	Index of Industrial Production (Investment Goods)	3
77	Index of Industrial Production (Capital Goods)	3
78	Index of Industrial Production (Construction Goods)	3
79	Index of Production Capacity (Manufacturing)	3
80	Machinery Orders (Total Excluding Ships)	3
81	Machinery Orders (Private Sector Excluding Volatile Orders)	3
82	Machinery Orders (Manufacturing)	3
83	Machinery Orders (Non-Manufacturing Excluding Volatile Orders)	3
84	Machinery Orders (Government)	3
85	Order Received for Construction (Grand Total)	3
86	Order Received for Construction (Private)	3
87	Order Received for Construction (Public)	3
88	Total Floor Area of Building Construction Started (Grand Total)	3
89	Total Floor Area of Building Construction Started (Mining, Manufacturing and Commercial Use)	3
90	Total Floor Area of Building Construction Started (Mining)	3
91	Total Number of New Housing Construction Started (Total)	3
92	Total Number of New Housing Construction Started (Owned)	3
93	Total Number of New Housing Construction Started (Rented)	3
94	Total Number of New Housing Construction Started (Built for Sale)	3
95	Total Number of New Housing Construction Started (Government Housing Loan Corporation)	3
96	Total Floor Area of New Housing Construction Started (Total)	3
97	Total Floor Area of New Housing Construction Started (Owned)	3
98	Total Floor Area of New Housing Construction Started (Rented)	3
99	Total Floor Area of New Housing Construction Started (Built for Sale)	3
<i>Employment</i>		
100	Index of Non-Scheduled Worked Hours (All Industries - 30 or more persons)	3
101	Index of Non-Scheduled Worked Hours (Manufacturing)	3
102	Index of Total Worked Hours (All Industries - 30 or more persons)	3
103	Index of Total Worked Hours (Manufacturing)	3
104	Ratio of Non-Scheduled to Total Worked Hours (All Industries -30 or more persons)	3
105	Ratio of Non-Scheduled to Total Worked Hours (Manufacturing)	3
106	New Job Offers	3
107	Effective Job Offers	3
108	New Job Offer Rate	3
109	Effective Job Offer Rate	3
110	New Job Offers (Parttime)	3
111	Effective Job Offers (Parttime)	3
112	New Job Offer Rate (Parttime)	3
113	Effective Job Offer Rate (Parttime)	3
114	Index of Regular Workers Employment (All Industries - 30 or more persons)	3

continued

continued

No.	Description	T-Code
115	Index of Regular Workers Employment (All Industries Excluding Services)	3
116	Index of Regular Workers Employment (Mining)	3
117	Index of Regular Workers Employment (Construction)	3
118	Index of Regular Workers Employment (Manufacturing)	3
119	Index of Regular Workers Employment (Electricity, Gas, Heat Supply)	3
120	Index of Regular Workers Employment (Transport and Communication)	3
121	Index of Regular Workers Employment (Wholesale and Retail Trade)	3
122	Index of Regular Workers Employment (Finance and Insurance)	3
123	Index of Regular Workers Employment (Real Estate)	3
124	Index of Regular Workers Employment (Services)	3
125	Number of Unemployment	3
126	Unemployment Rate	1
127	Number of Beneficiaries of Unemployment Insurance (Initial Claimants)	2
128	Number of Beneficiaries of Unemployment Insurance (Total)	2
129	Number of Persons with Unemployment Insurance	2
130	Real Wage Index (Contractual Cash Earnings in All Industries - 30 or more persons)	2
	<i>Consumption</i>	
131	Sales at Department Stores (Total)	2
132	Sales at Department Stores (Per Square Meter Floor Space)	2
133	Index of Sales (Total)	2
134	Index of Sales (Wholesale)	2
135	Index of Sales (Retail)	2
136	Number of New Passenger Car Registrations and Reports (Total)	3
137	Number of New Passenger Car Registrations and Reports (Excluding Cars under 550cc)	3
138	Household Consumption Expenditure (Workers)	2
139	Household Consumption Expenditure (Food)	2
140	Household Disposable Income (Workers)	2
141	Index of Industrial Production (Consumer Goods)	3
142	Index of Industrial Production (Durable Consumer Goods)	3
143	Index of Industrial Production (Non-Durable Consumer Goods)	3
144	Index of Producer's Shipments (Consumer Goods)	3
145	Index of Producer's Shipments (Durable Consumer Goods)	3
146	Index of Producer's Shipments (Non-Durable Consumer Goods)	3
	<i>Firms</i>	
147	Index of Investment Climate (Manufacturing)	3
148	Corporation Tax Revenue	3
149	Suspension of Business Transaction with Bank	3
	<i>Moeny, Stock Price and Interest Rate</i>	
150	Money Stock (M2+CD, Average Outstanding)	2
151	Money Stock (M1, Average Outstanding)	2
152	Monetary Base (Average Outstanding)	2
153	Bank Notes Issued (Average Outstanding)	2
154	Bank Clearings (Number)	3
155	Bank Clearings (Value)	3
156	Nikkei Stock Average 225 Selected Stocks (Average of Month)	3
157	Nikkei Stock Average 500 Selected Stocks	3
158	Stock Price Index (TOPIX)	3
159	Stock Price Average (Tokyo Stock Market First Section)	3
160	Stock Price Index (Fisheries, Agriculture and Forestry)	3
161	Stock Price Index (Mining)	3
162	Stock Price Index (Construction)	3
163	Stock Price Index (Foods)	3
164	Stock Price Index (Textiles)	3
165	Stock Price Index (Pulp and Paper)	3
166	Stock Price Index (Oil and Coal Products)	3
167	Stock Price Index (Rubber Products)	3
168	Stock Price Index (Glass and Ceramics Products)	3
169	Stock Price Index (Iron and Steel)	3
170	Stock Price Index (Non-Ferrous Metals)	3
171	Stock Price Index (Metal Products)	3
172	Stock Price Index (Machinery)	3
173	Stock Price Index (Electrical Machinery)	3
174	Stock Price Index (Transportation Equipment)	3
175	Stock Price Index (Precision Equipment)	3
176	Stock Price Index (Other Products)	3
177	Stock Price Index (Electric and Gas)	3
178	Stock Price Index (Land Transportation)	3
179	Stock Price Index (Marine Transportation)	3
180	Stock Price Index (Air Transportation)	3
181	Stock Price Index (Warehouse and Transport Related)	3
182	Stock Price Index (Communication)	3

continued

continued

No.	Description	T-Code
183	Stock Price Index (Real Estate)	3
184	Stock Price Index (Service)	3
185	Sales Volume (Daily Average, Tokyo Stock Market First Section)	3
186	Sales Value (Daily Average, Tokyo Stock Market First Section)	3
187	Official Discount Rates	1
188	Short-Term Prime Lending Rates	1
189	Long-Term Prime Lending Rates	1
190	Average Contracted Interest Rate on Loans and Discounts (Domestically Licensed Bank)	1
191	Yields on Interest Bearing Government Bonds (10 years)	1
	<i>Price Indexes</i>	
192	Nikkei Commodity Price Index (17 items)	2
193	Nikkei Commodity Price Index (42 items)	2
194	Wholesale Price Index (All Commodities)	2
195	Wholesale Price Index (Manufacturing Industry Products)	2
196	Wholesale Price Index (Raw Materials)	2
197	Wholesale Price Index (Intermediate Materials)	2
198	Wholesale Price Index (Final Goods)	2
199	Wholesale Price Index (Capital Goods)	2
200	Wholesale Price Index (Consumer Goods)	2
201	Wholesale Price Index (Durable Consumer Goods)	2
202	Wholesale Price Index (Non-Durable Consumer Goods)	2
203	Consumer Price Index (General)	2
204	Consumer Price Index (General Excluding Fresh Food)	2
205	Consumer Price Index (General Excluding Fresh Food and Imputed Rent)	2
206	Consumer Price Index (Food)	2
207	Consumer Price Index (Housing)	2
208	Consumer Price Index (Fuel Light and Water Charges)	2
209	Consumer Price Index (Furniture and Household Utilities)	2
210	Consumer Price Index (Clothes and Footware)	2
211	Consumer Price Index (Medical Care)	2
212	Consumer Price Index (Transportation and Communication)	2
213	Consumer Price Index (Reading and Recreation)	2
214	Consumer Price Index (Miscellaneous)	2
	<i>Trade</i>	
215	Terms of Trade Index (All Commodities)	3
216	Quantum Index of Exports (Total)	3
217	Quantum Index of Imports (Total)	3
218	Customs Clearance (Value of Exports, Grand Total)	3
219	Foreign Exchange Rate (Yen per US Dollar, Spot)	3