Forecasting the Japanese macroeconomy using high-dimensional data

Nakajima, Yoshiki

Sueishi, Naoya

(Citation)
The Japanese Economic Review, 73:299-324

(Issue Date)
2020-05-20

(Resource Type)
journal article

(Version)
Accepted Manuscript

Noname manuscript No.
(will be inserted by the editor)

Forecasting the Japanese Macroeconomy Using High-Dimensional Data

Yoshiki Nakajima · **Naoya Sueishi**

Received: date / Accepted: date

Abstract This paper compares several forecasting methods using high-dimensional macroeconomic data from Japan. The diffusion index (DI) model has been widely used to incorporate the information contained in high-dimensional data for forecasting. We propose two selection methods of the number of latent factors in the DI model and compare the DI model with the vector autoregression (VAR) model whose parameters are estimated by lasso-type methods. We find that the DI model tends to be better for short-horizon forecasting, whereas the VAR model tends to be better for long-horizon forecasting. Moreover, we find that the information exploited for forecasting is similar between the DI and VAR models.

JEL Classification Numbers: C38, C53, C55.

Keywords Diffusion index · High-dimensional data · Lasso

This paper is based on the master's thesis of the first author. The views expressed in this paper are the authors' own and do not necessarily reflect the views of the company with whom the first author is associated.

Y. Nakajima
Asset Management One Co., Ltd., Marunouchi, Chiyoda-ku, Tokyo 100-0005, Japan.
E-mail: ynakajima146@gmail.com

N. Sueishi
Graduate School of Economics, Kobe University, 2-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan.
ORCID: https://orcid.org/0000-0001-8757-1655
Tel.: +81-78-803-6827
E-mail: sueishi@econ.kobe-u.ac.jp

1 Introduction

The increased availability of large-scale data in recent years has enabled us to incorporate rich information to forecast macroeconomic series. However, forecasting based on the vector autoregression (VAR) model, which is the traditional and commonly used method of forecasting, has two problems related to high-dimensional data. One is the curse of dimensionality; that is, the number of parameters in VAR models can be much larger than the time series sample size. The other problem is the high correlation among predictors. Macroeconomic data tend to comprise sets of highly correlated variables, which may make ordinary least squares (OLS) estimation unstable.

Since the end of the twentieth century, a number of methods have been proposed to circumvent problems with high-dimensional data in conventional econometric methods. The idea is to reduce the dimensionality of the original data with the least possible loss of information. These methods are broadly classified into two approaches.

The first approach is to reduce the number of predictors in high-dimensional VAR models by penalized least squares estimation. Tibshirani (1996) introduced an L_1-penalized least squares method called "lasso" (least absolute shrinkage and selection operator), and some variants of lasso were subsequently proposed by Zou and Hastie (2005) and Yuan and Lin (2006). These methods make some coefficients of the VAR model exactly zero and, hence, can achieve estimation and variable selection simultaneously. Although lasso-type methods were originally introduced to analyze cross section data, they can also be applied for time series data. Applications in macroeconomics are relatively new and most are limited to analyses of U.S. data. See, for example, Song and Bickel (2011), Li and Chen (2014), and Callot and Kock (2014) for forecasting of monthly macroeconomics indicators. Marsilli (2014) and Uematsu and Tanaka (2019) obtained short-term forecasting (nowcasting) using mixed-frequency data.

The second approach is to use dynamic factor models to exploit a small number of latent variables that may affect a large number of macroeconomic variables. Burns and Mitchell (1946) defined the business cycle as a type of fluctuation found in aggregate economic activity; thus, the idea that a few unobservable common factors have an impact on a large set of observable variables is well rooted in economics. Since the seminal work by Stock and Watson (2002a), dynamic factor models have been widely used in macroeconomics, not only for forecasting (Stock and Watson 2002a,b; Artis et al. 2005; Banerjee et al. 2008) but also for other purposes such as the evaluation of monetary policy (Bernanke et al. 2005; Yamamoto 2019), estimation of dynamic stochastic general equilibrium models (Boivin and Giannoni 2006), development of real-time economic indicators (Altissimo et al. 2010), and nowcasting (Giannone et al. 2008). Applications using Japanese data include those of Shintani (2005), Shibamoto (2007), Hayakawa and Kobayashi (2011), and Chikamatsu et al. (2018).

This paper makes three contributions to the literature of macroeconomic forecasting. First, we apply various high-dimensional models to the Japanese economy in order to forecast important macroeconomic indicators. Making an accurate prediction of the macroeconomy is crucial in order to better understand the country's economic structure on which policymakers and businesses make decisions; however, few studies have explored high-dimensional macroeconomic data in Japan. Although Shintani (2005) showed the benefit of exploiting high-dimensional data for forecasting, he only used a diffusion index (DI) model. Recently, the simulation of Smeekes and Wijler (2018) has shown that penalized least squares methods are robust to model misspecification and perform well even when the underlying data generating process follows a factor model. Our attempt is the first to compare the forecasting performances of the DI model and the high-dimensional VAR model by using Japanese data.

Second, we propose two forecast-oriented methods to select the number of factors in the DI model. Our DI model consists of two equations. The first equation is a factor model used to extract common factors from high-dimensional data. The second equation is a factor-augmented AR model used to predict a series of interests. Because the number of factors is unknown to researchers, it has to be determined in a data-dependent manner. A possible approach is to use the C_p-type criterion of Bai and Ng (2002). However, their criterion tends to select an excessive number of factors because it is designed to find the true number of factors in the factor model and does not consider the forecasting performance of the factor-augmented AR model. We propose cross-validation-based methods to select the number of factors in order to minimize the mean squared forecasting error (MSFE).

Third, we consider an R^2-based approach to investigate the relationship between latent common factors and observable macroeconomic series. One of the problems of factor models is the interpretability of latent factors because they are given by linear combinations of several observable series. Stock and Watson (2002b) and Ludvigson and Ng (2009) studied the economic interpretations of factors by looking into R^2s while regressing each macroeconomic variable on each estimated factor. Our method is different from theirs in that we investigate R^2s while regressing each variable on all factors relevant for forecasting a particular variable. Variables with high R^2 capture the information exploited by the DI model for forecasting.

We have conducted an empirical analysis to investigate the forecasting performances of several high-dimensional models using 127 monthly Japanese macroeconomic data from April 2003 to June 2018. Our main findings are threefold. (1) Incorporating abundant information using high-dimensional data contributes to improvements in the accuracy of forecasting. Moreover, the benefit of using high-dimensional data is larger for longer foresting horizons, although the extent of improvement is relatively small when forecasting leading indicators. (2) The number of factors selected by our selection methods is much smaller than that selected by the criterion of Bai and Ng (2002). This results in a better forecasting performance of our DI model. (3) Our R^2-based approach indicates that the DI model and the high-dimensional VAR model exploit information from similar groups of variables for forecasting if factors of the DI model and variables in the VAR model are selected by our proposed methods. Although informative groups of variables are different depending on the variables to be forecast, they are similar between the DI and VAR models.

The remainder of this paper is organized as follows. Section 2 introduces high-dimensional VAR and DI models and their estimation methods. Section 3 reports the result of the empirical study, and Section 4 concludes.

2 Forecasting Methods

Suppose that we observe an N-dimensional stationary process $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$ for $t = 1, \ldots, T$. Our aim is to predict the value of $y_{i,T+h}$ on the basis of the sample $(y_1, \ldots, y_T)$.

We consider two types of models for forecasting: the VAR model and the DI model. The VAR model is estimated by lasso-type methods for which the size of the regularization parameters must be determined, and the DI model is estimated by the principal component method for which the number of latent factors must be determined. In this section, we introduce our models and propose selection methods for the regularization and tuning parameters of lasso-type methods and the number of latent factors for the principal component method.

2.1 VAR Models and Lasso-Type Estimators

The VAR model is widely used in various areas of macroeconomics to capture the rich dynamics in multiple time series. By using the VAR(P) model, an h-step-ahead direct forecast model is expressed as

$$y_{i,t+h} = \mu_i + \sum_{p=1}^{P} \phi_{p,i}' y_{t+1-p} + \varepsilon_{i,t+h}, \tag{1}$$

where μ_i is an intercept, $\phi_{p,i}$ is an $N \times 1$ vector of coefficients, and $\varepsilon_{i,t+h}$ is an error term. The forecast of $y_{i,T+h}$ based on $(y_1, \ldots, y_T)$ is obtained by

$$\hat{y}_{i,T+h|T,\ldots,1} = \hat{\mu}_i + \sum_{p=1}^{P} \hat{\phi}_{p,i}' y_{T+1-p},$$

where $\hat{\mu}_i$ and $\hat{\phi}_{p,i}$ are estimates of μ_i and $\phi_{p,i}$, respectively. When N is small, the parameters are typically estimated by OLS.

A potential problem of using the VAR model is that the number of parameters to estimate can be very large when the dimension of y_t or the lag length is large. This results in high standard errors of OLS estimates or even the infeasiblity of OLS. Although it is theoretically possible to select a subset of predictors using information criteria such as the Akaike information criterion or the Bayesian information criterion, this approach is computationally too expensive as the total number of subsets grows exponentially large with the number of predictors.

The lasso of Tibshirani (1996) is a computationally efficient method of selecting predictors using penalized least squares estimation. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the

L_1 and L_2 norm, respectively. The lasso for (1) is obtained by solving the following minimization problem:

$$\min_{\mu_i,\phi_{1,i},\dots,\phi_{P,i}} \sum_{t=P}^{T-h} \|y_{i,t+h} - \mu_i - \sum_{p=1}^{P} \phi'_{p,i} y_{t+1-p}\|_2^2 + \lambda \sum_{p=1}^{P} \|\phi_{p,i}\|_1, \tag{2}$$

where $\lambda > 0$ is a regularization parameter. Because of the penalty term, the second term of (2), lasso shrinks coefficients toward zero and sets some of them exactly zero. Hence, a sparse model is obtained. The sparsity of the selected model depends on the value of the regularization parameter. A larger λ selects a sparser model.

Although lasso has been successful in numerous applications, it has some limitations when dealing with high-dimensional macroeconomic data. First, the number of predictors that lasso uniquely selects is limited up to the number of observations. This can be problematic when estimating the VAR model because the number of predictors often exceeds the number of time series observations. Second, when some predictors are highly correlated, which is often the case for macroeconomic variables, lasso tends to select only one of them and does not care which one is selected. This property is undesirable if one is interested in an interpretation of the selected model. Moreover, lasso is inferior to the ridge estimator in terms of forecasting performance when the predictors are highly correlated (Tibshirani 1996).

Zou and Hastie (2005) proposed the elastic net to solve the drawbacks of lasso by combining the L_1 and L_2 penalties. The elastic net minimizes

$$\sum_{t=P}^{T-h} \|y_{i,t+h} - \mu_i - \sum_{p=1}^{P} \phi'_{p,i} y_{t+1-p}\|_2^2 + \lambda \sum_{p=1}^{P} \left((1-\alpha)\|\phi_{p,i}\|_1 + \alpha\|\phi_{p,i}\|_2^2\right),$$

where α $(0 \le \alpha \le 1)$ is a tuning parameter that controls the weight between two penalty functions. Setting $\alpha = 0$ leads to lasso, whereas setting $\alpha = 1$ leads to ridge regression. Zou and Hastie (2005) demonstrated the group effect of the elastic net. Namely, it is able to select variables with high correlation altogether. They also showed that the elastic net can select more predictors than observations owing to the features of the ridge estimator.

Another approach is the group lasso proposed by Yuan and Lin (2006). The group lasso simultaneously includes or excludes a set of variables based on predetermined groups of variables. Although the group lasso is more restrictive than the elastic net in the sense that groups have to be specified in advance, the problem rarely arises in macroeconomic forecasting because macroeconomic variables typically have certain attributes such as output and consumption.

Suppose that y_t can be partitioned into L groups: $y_t = (y_t^{1\prime}, \dots, y_t^{L\prime})'$. Let d_l denote the dimension of y_t^l. Then, (1) can be rewritten as

$$y_{i,t+h} = \mu_i + \sum_{l=1}^{L} \sum_{p=1}^{P} \phi_{p,i}^{l\ \prime} y_{t+1-p}^l + \varepsilon_{i,t+h},$$

where $\phi_{p,i}^l$ is a $d_l \times 1$ vector of coefficients. The group lasso is obtained by minimizing

$$\sum_{t=P}^{T-h} \|y_{i,t+h} - \mu_i - \sum_{l=1}^{L} \sum_{p=1}^{P} \phi_{p,i}^{l\ \prime} y_{t+1-p}^l\|_2^2 + \lambda \sum_{l=1}^{L} \sum_{p=1}^{P} \sqrt{d_l}\|\phi_{p,i}^l\|_2. \tag{3}$$

In the penalty term of the group lasso, $\sqrt{d_l}$ reflects different dimensions between groups and works to avoid favoring groups with higher dimension. Moreover, because all coefficients in the same group are penalized by the L_2 norm $\|\phi^l_{p,i}\|_2$, variables in the same group are either retained or dropped altogether. If all variables are grouped into different groups ($d_l = 1$ for all l), then (3) is equivalent to (2).

So far, we have assumed that the regularization and tuning parameters are given. In practice, however, all the lasso-based methods require an appropriate choice of λ and α. For cross-sectional observations, a standard selection method is the K-fold cross-validation (see, for instance, Arlot and Celisse 2010). In time series applications, however, observations are time-dependent, and randomly splitting the data without taking into account the time dependence may not be appropriate.

To determine the regularization parameter of lasso, we suggest a selection procedure that minimizes the MSFE of rolling window forecasts. We split the sample into two subperiods — training period ($t = 1,\ldots,T_1$) and validation period ($t = T_1+1,\ldots,T$) — and calculate the MSFE in the validation period. That is, for a given value of λ and for $j = 1,\ldots,T-T_1$, we first obtain $\hat{y}^{\lambda}_{i,T_1+j|T_1+j-h,\ldots,j}$, which is an h step-ahead forecast of y_{i,T_1+j} based on the observation $(y_j,\ldots,y_{T_1+j-h})$ when the parameters are estimated by lasso with regularization parameter λ. Then, we calculate

$$\mathrm{MSFE}^{\lambda}_{i,h} = \frac{1}{T-T_1}\sum_{j=1}^{T-T_1}(y_{i,T_1+j} - \hat{y}^{\lambda}_{i,T_1+j|T_1+j-h,\ldots,j})^2$$

for several different values of λ[1] . Finally, the value of λ that minimizes the MSFE is selected. The regularization and tuning parameters of the elastic net and the group lasso are also selected in a similar way.

2.2 DI Models and Principal Component Analysis

Another approach to prediction is to exploit common factors that capture a large part of the variation of high-dimensional data. These common factors are often referred to as the diffusion indices in macroeconomics and, hence, the model is called the DI model.

We use the following DI model:

$$y_t = \Lambda F_t + e_t, \tag{4}$$

$$y_{i,t+h} = \nu_i + \beta'_{F,i}F_t + \sum_{p=1}^{P}\beta_{p,i}y_{i,t+1-p} + u_{i,t+h}, \tag{5}$$

where F_t is an $r\times 1$ vector of unobservable common factors and Λ is an $N\times r$ matrix of factor loadings. It is possible to extend (5) by including lagged factors. The elements of e_t, $N\times 1$ vector of idiosyncratic disturbances, may have weak serial and

[1] Our terminology may be slightly misleading because we actually use observations in both the training and validation periods to obtain some forecasts. For instance, to obtain $\hat{y}^{\lambda}_{i,T|T-h,\ldots,T-T_1}$, we use the observation y_{T-h}, which belongs to the validation period if $h < T-T_1$. In this paper, the term "validation period" means that the MSFE is calculated over the period to determine the regularization parameter.

cross-sectional correlations (Chamberlain and Rothschild 1983). If factors are not included, then (5) reduces to an AR(P) model. An h-step-ahead forecast of $y_{i,T+h}$ based on $(y_1,\dots,y_T)$ is given by

$$\hat{y}_{i,T+h|T,\dots,1} = \hat{\nu}_i + \hat{\beta}'_{F,i}\hat{F}_T + \sum_{p=1}^{P} \hat{\beta}_{p,i} y_{i,T+1-p},$$

where $\hat{\nu}_i$, $\hat{\beta}_{F,i}$ and $\hat{\beta}_{p,i}$ are estimates of coefficients and $\hat{F}_t$ is an estimate of F_t.

The estimation of the DI model is carried out in two steps. Let $Y = (y_1,\dots,y_T)'$ and $F = (F_1,\dots,F_T)'$. In the first step, we estimate Λ and F by the principal component method, which solves

$$\min_{\Lambda,F} \mathrm{tr}\{(Y - F\Lambda')'(Y - F\Lambda')\}$$

subject to the constraint that $F'F/T = I$. Then, $\hat{F}$ is given by $\sqrt{T}$ times eigenvectors corresponding to r largest eigenvalues of YY'. Moreover, Λ is estimated by $\hat{\Lambda} = \hat{F}'Y/T$. In the second step, we replace F by $\hat{F}$ and estimate (5) by OLS. The asymptotic properties of the estimator were investigated by Stock and Watson (2002a) and Bai (2003).

The forecasting performances of the DI model depend on the number of factors included in (5). Let $\hat{\sigma}^2$ be a consistent estimator for $(NT)^{-1}\sum_{t=1}^{T} E[e_t' e_t]$. Bai and Ng (2002) proposed the following C_p-type criterion that balances the trade-off between goodness-of-fit and parsimony of the model:

$$PC(k) = \frac{\mathrm{tr}\{(Y - \hat{F}^k\hat{\Lambda}^{k'})'(Y - \hat{F}^k\hat{\Lambda}^{k'})\}}{NT} + k\hat{\sigma}^2 \frac{N+T}{NT} \ln(\min\{N,T\}), \qquad (6)$$

where $\hat{F}^k$ and $\hat{\Lambda}^k$ are estimated matrices of factors and factor loadings, respectively, when the number of factors is k. The first term on the right-hand side is decreasing in k and represents the goodness-of-fit, whereas the second term is increasing in k and serves as a penalty of including additional factors. The optimum number of factors $\hat{r}$ is then determined as the minimizer of (6). Bai and Ng (2002) showed the consistency of (6); that is, $\hat{r}$ can identify the true number of factors in (4) with probability approaching one.

Selecting the number of factors by (6) is often problematic for two reasons. First, because (6) is designed to identify the true number of factors in (4), $\hat{r}$ is determined irrespective of the predictive power of (5). It is not necessarily optimal to include all true factors in (5) for the purpose of prediction. Second, because the factor loadings are assumed to be independent of t, the DI model fails to incorporate structural breaks in factor loadings. In the presence of a break in factor loadings, the same factor is identified as two different factors before and after the break. Therefore, (6) tends to select $\hat{r}$ that is larger than it actually is; see, for instance, Stock and Watson (2005) and Yamamoto and Tanaka (2015).

We propose two alternative methods to select the number of factors in the DI model. One method, which we refer to as DICV, estimates DI models using different numbers of factors and selects the model that minimizes the MSFE. As in the case of selecting the regularization parameter of lasso, we split the data into two subperiods and select the number of factors that minimizes the MSFE of rolling window forecasts in the validation period.

The other method, which we refer to as DILASSO, is a lasso-based method. We first estimate (4) by including a relatively large number of factors (e.g., $r = 20$). Then, we include all the estimated factors in (5) and estimate the model by imposing the L_1 penalty on $\beta_{F,i}$ and $\beta_{p,i}$. The regularization parameter of the lasso is selected in the same manner as described in the previous subsection.

These approaches are beneficial because they provide the optimum number of factors in terms of the predictive power of the target variable, whereas the criterion of Bai and Ng (2002) selects the number of factors optimum for explaining the original high-dimensional data[2].

3 Empirical Results

3.1 Data

Our dataset comprises 127 monthly macroeconomic time series, which are classified into 11 categories. The series are selected on the basis of Hayakawa and Kobayashi (2011). The sample period is from April 2003 to June 2018.

On the basis of unreported results of the augmented Dickey–Fuller test (Dickey and Fuller 1981) and the Phillips–Perron test (Phillips and Perron 1988), all series are transformed to be stationary. Seasonal adjustment is carried out using X13-ARIMA-SEATS when adjusted series are not available. After all the transformations, each series is standardized to have mean zero and standard deviation unity.

Detailed descriptions of the adopted series are given in Table 1. The series are sourced from Thomson Reuters Datastream (DS), Nikkei NEEDS Financial QUEST (NEEDS), Ministry of Economy, Trade and Industry (METI), Ministry of Health, Labour and Welfare (MHLW), Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Ministry of Internal Affairs and Communications (MIC), and Bank of Japan (BOJ). The series to be forecast are indicated in bold. They are representative macroeconomic indicators selected according to Shibamoto (2007). SA denotes seasonally adjusted series. Numbers in the fourth column indicate transformation codes. The codes represent (1) first difference of the log-transformed series, (2) first difference of the log-transformed series after seasonal adjustment, and (3) first difference of the series.

All the empirical analyses were conducted using R 3.4.2 (R Core Team 2017). Several packages are used for the implementation; *glmnet* for lasso and elastic net (Friedman et al. 2010), *grplasso* for group lasso (Meier 2015), *urca* for unit root tests (Pfaff 2008), and *x12* for seasonal adjustment (Kowarik et al. 2014)[3].

Table 1: List of series

---

[2] Bai and Ng (2008) proposed an alternative method, which reduces the dimension of y_t in (4) in a data-driven way before estimating factors.

[3] The dataset after transformation and implementation codes are available upon request.

Output and Income			
1	**Index of Industrial Production — Mining and Manufacturing (2010=100, SA)**	METI	1
2	Index of Industrial Production — Final Demand Goods (2010=100, SA)	METI	1
3	Index of Industrial Production — Investment Goods (2010=100, SA)	METI	1
4	Index of Industrial Production — Capital Goods (2010=100, SA)	METI	1
5	Index of Industrial Production — Construction Goods (2010=100, SA)	METI	1
6	Index of Industrial Production — Consumer Goods (2010=100, SA)	METI	1
7	Index of Industrial Production — Durable Consumer Goods (2010=100, SA)	METI	1
8	Index of Industrial Production — Nondurable Consumer Goods (2010=100, SA)	METI	1
9	Index of Industrial Production — Producer Goods (2010=100, SA)	METI	1
10	Index of Industrial Production — Producer Goods for Mining and Manufacturing (2010=100, SA)	METI	1
11	Index of Industrial Production — Producer Goods for Others (2010=100, SA)	METI	1
12	**Index of Capacity Utilization Ratio — Manufacturing (2010=100, SA)**	METI	1
13	Index of Production Capacity — Manufacturing (2010=100)	DS	2
14	**Household Disposable Income — Workers (thousand yen)**	DS	2
Employment and Hours			
15	Unemployment Rate (%, SA)	DS	3
16	Number of Unemployed (thousand persons, SA)	DS	1
17	**Employment Index of Regular Workers — All Industries, 30 or More Persons (2015=100, SA)**	MHLW	1
18	Employment Index of Regular Workers — Manufacturing (2015=100, SA)	MHLW	1
19	Active Job Openings-to-Applicants Ratio — Excl. New School Graduates (SA)	MHLW	3
20	Active Job Openings-to-Applicants Ratio — Part-Time (SA)	MHLW	3
21	Effective Job Offers — Excl. New School Graduates (persons, SA)	MHLW	1
22	Effective Job Offers — Part-Time (persons, SA)	MHLW	1
23	New Job Openings-to-Applicants Ratio — Excl. New School Graduates (SA)	MHLW	3
24	New Job Openings-to-Applicants Ratio — Part-Time (SA)	MHLW	3
25	**New Job Offers — Excl. New School Graduates (persons, SA)**	MHLW	1
26	New Job Offers — Part-Time (persons, SA)	MHLW	1
27	**Index of Total Hours Worked — All Industries, 30 or More Persons (2015=100, SA)**	MHLW	1
28	Index of Total Hours Worked — Manufacturing, 30 or More Persons (2015=100, SA)	MHLW	1
29	Index of Nonscheduled Hours Worked — All Industry (2015=100, SA)	MHLW	1
30	Index of Nonscheduled Hours Worked — Manufacturing (2015=100, SA)	MHLW	1
Retail, Manufacturing, and Trade Sales			
31	Sales at Department Stores — Total (million yen)	METI	2
32	Wholesale Sales Value (billion yen)	METI	2
33	Retail Sales Value (billion yen)	METI	2
34	Import Volume Index — Total (2015=100)	DS	2
35	Export Volume Index — Total (2015=100)	DS	2
36	Customs Clearance — Value of Exports, Grand Total (million yen, SA)	NEEDS	1
Consumption			
37	**Real Consumption Activity Index — Total (2011=100, SA)**	DS	1
38	Real Consumption Activity Index — Durable Goods (2011=100, SA)	DS	1
39	Real Consumption Activity Index — Nondurable Goods (2011=100, SA)	DS	1
40	Motor Vehicle New Registrations — Passenger Cars Excl. Below 660 cc	DS	2
Housing Starts and Sales			
41	Total Number of New Housing Construction Started — Total (SA)	MLIT	1
42	Total Number of New Housing Construction Started — Owned (SA)	MLIT	1
43	Total Number of New Housing Construction Started — Rented (SA)	MLIT	1
44	Total Number of New Housing Construction Started — Built for Sale (SA)	MLIT	1
45	**Total Floor Area of New Housing Construction Started — Total (thousand square meters, SA)**	MLIT	1

46	Total Floor Area of New Housing Construction Started — Owned (thousand square meters, SA)	MLIT	1
47	Total Floor Area of New Housing Construction Started — Rented (thousand square meters, SA)	MLIT	1
48	Total Floor Area of New Housing Construction Started — Built for Sale (thousand square meters, SA)	MLIT	1
Inventories and Orders			
49	Index of Producer's Inventory of Finished Goods — Mining and Manufacturing (2010=100, SA)	METI	1
50	Index of Producer's Inventory of Finished Goods — Final Demand Goods (2010=100, SA)	METI	1
51	Index of Producer's Inventory of Finished Goods — Investment Goods (2010=100, SA)	METI	1
52	Index of Producer's Inventory of Finished Goods — Capital Goods (2010=100, SA)	METI	1
53	Index of Producer's Inventory of Finished Goods — Construction Goods (2010=100, SA)	METI	1
54	Index of Producer's Inventory of Finished Goods — Consumer Goods (2010=100, SA)	METI	1
55	Index of Producer's Inventory of Finished Goods — Durable Consumer Goods (2010=100, SA)	METI	1
56	Index of Producer's Inventory of Finished Goods — Nondurable Consumer Goods (2010=100, SA)	METI	1
57	Index of Producer's Inventory of Finished Goods — Producer Goods (2010=100, SA)	METI	1
58	Index of Producer's Inventory of Finished Goods — Producer Goods for Mining and Manufacturing (2010=100, SA)	METI	1
59	Index of Producer's Inventory of Finished Goods — Producer Goods for Others (2010=100, SA)	METI	1
60	Index of Producer's Inventory Ratio of Finished Goods — Mining and Manufacturing (2010=100, SA)	METI	1
61	Index of Producer's Inventory Ratio of Finished Goods — Final Demand Goods (2010=100, SA)	METI	1
62	Index of Producer's Inventory Ratio of Finished Goods — Investment Goods (2010=100, SA)	METI	1
63	Index of Producer's Inventory Ratio of Finished Goods — Capital Goods (2010=100, SA)	METI	1
64	Index of Producer's Inventory Ratio of Finished Goods — Construction Goods (2010=100, SA)	METI	1
65	Index of Producer's Inventory Ratio of Finished Goods — Consumer Goods (2010=100, SA)	METI	1
66	Index of Producer's Inventory Ratio of Finished Goods — Durable Consumer Goods (2010=100, SA)	METI	1
67	Index of Producer's Inventory Ratio of Finished Goods — Nondurable Consumer Goods (2010=100, SA)	METI	1
68	Index of Producer's Inventory Ratio of Finished Goods — Producer Goods (2010=100, SA)	METI	1
69	Index of Producer's Inventory Ratio of Finished Goods — Producer Goods for Mining and Manufacturing (2010=100, SA)	METI	1
70	Index of Producer's Inventory Ratio of Finished Goods — Producer Goods for Others (2010=100, SA)	METI	1
71	**Machinery Orders — Total (billion yen, SA)**	DS	1
72	Machinery Orders — Private Sectors Excl. Ships (billion yen, SA)	DS	1
73	Machinery Orders — Private Sectors Excl. Volatile Orders (billion yen, SA)	DS	1
74	Business Expenditures for New Plant and Equipment at Constant Prices — All Industries (hundred million yen, SA)	DS	1
Stock Prices			

75	Nikkei Stock Average 225 Selected Stocks	DS	1
76	**Tokyo Stock Price Index (TOPIX)**	DS	1
77	**Nikkei Commodity Price Index — 42 items (1970=100)**	DS	2
78	Nikkei Commodity Price Index — 17 items (1970=100)	DS	2
Exchange Rates			
79	**US.Dollar-Yen Spot Rate — Average in the Month (JPY/USD)**	BOJ	1
80	Nominal Effective Exchange Rates (2010=100)	BOJ	1
81	Real Effective Exchange Rates (2010=100)	BOJ	1
Interest Rates			
82	**Newly Issued Government Bonds Yield — 10 Years (% per annum)**	DS	3
83	**Tokyo Interbank Offered Rates (TIBOR) — 3 Months (% per annum)**	DS	3
84	**Interest Rate Spread (normalized, SA)**	DS	3
85	Yield of Interest-Bearing Government Bonds — 10 Years (% per annum)	DS	3
86	Long-Term Prime Lending Rates (% per annum)	DS	3
87	Short-Term Prime Lending Rates (% per annum)	DS	3
88	**Call Rate — Uncollateralized Overnight, Average in the Month (% per annum)**	BOJ	3
89	Call Rate — Uncollateralized Overnight, End of Month (% per annum)	BOJ	3
90	Basic Discount Rate and Basic Loan Rate (% per annum)	BOJ	3
91	Avg. Contracted New Loan and Discount — City Banks (% per annum)	DS	3
92	Avg. Contracted New Loan and Discount — City Banks, Short-Term (% per annum)	DS	3
93	Avg. Contracted New Loan and Discount — City Banks, Long-Term (% per annum)	DS	3
94	Avg. Contracted General Incl. Overdraft Rate — City Banks (% per annum)	DS	3
95	Avg. Contracted Short-Term Rate — City Banks (% per annum)	DS	3
96	Avg. Contracted Long-Term Rate — City Banks (% per annum)	DS	3
Money and Credit Quality Aggregate			
97	Money Supply: M1 (billion yen)	DS	2
98	**Money Supply: M2 (billion yen)**	DS	2
99	Money Supply: M3 (billion yen)	DS	2
100	Money Supply: L (billion yen, SA)	DS	1
101	**Monetary Base — Banknotes in Circulation, Average Amounts Outstanding (hundred million yen)**	BOJ	2
102	Monetary Base — Coins in Circulation, Average Amounts Outstanding (hundred million yen)	BOJ	2
103	Monetary Base — Reserves (hundred million yen)	DS	2
104	Amount of Clearing — Value (million yen)	DS	2
105	Amount of Clearing — Number of Bills (thousand)	DS	2
Price Indices and Wages			
106	Consumer Price Index — General (2005=100)	MIC	2
107	**Consumer Price Index — General, Excl. Fresh Food (2005=100)**	MIC	2
108	Consumer Price Index — General, Excl. Imputed Rent (2005=100)	MIC	2
109	Consumer Price Index — General, Excl. Imputed Rent and Fresh Food (2005=100)	MIC	2
110	Consumer Price Index — Food (2005=100)	MIC	2
111	Consumer Price Index — Housing (2005=100)	MIC	2
112	Consumer Price Index — Fuel, Light and Water Charges (2005=100)	MIC	2
113	Consumer Price Index — Furniture and Household Utensils (2005=100)	MIC	2
114	Consumer Price Index — Clothes and Footwear (2005=100)	MIC	2
115	Consumer Price Index — Medical Care (2005=100)	MIC	2
116	Consumer Price Index — Transportation and Communication (2005=100)	MIC	2
117	Consumer Price Index — Education (2005=100)	MIC	2
118	Consumer Price Index — Reading and Recreation (2005=100)	MIC	2
119	Consumer Price Index — Miscellaneous (2005=100)	MIC	2
120	Producer Price Index — All Commodities (2015=100)	DS	2
121	Export Price Index — Total Average (2015=100)	BOJ	2

122	Import Price Index — Total Average (2015=100)	BOJ	2
123	Terms of Trade Index — All Commodities (2015=100)	DS	2
124	**Real Wage Index — Contractual Cash Earnings in All Industries (2015=100, SA)**	MHLW	1
125	Real Wage Index — Contractual Cash Earnings in Manufacturing (2015=100, SA)	MHLW	1
126	Wage Index — Contractual Cash Earnings in All Industries (2015=100, SA)	MHLW	1
127	Wage Index — Contractual Cash Earnings in Manufacturing (2015=100, SA)	MHLW	1

3.2 Predictive Accuracy

We compare out-of-sample MSFEs of the proposed methods. We use lasso (LASSO), elastic net (ENET), and group lasso (gLASSO) to estimate the VAR model. The number of factors in the DI model is determined by DICV and DILASSO. As a benchmark, we report the MSFE of the AR model, whose lag length is selected by the Bayesian information criterion. Moreover, we report the MSFE of the DI model whose number of factors is selected by (6) (DIPC). The maximum lag order is set to 12 for the AR model and the factor-augmented AR model (5), whereas it is set to 4 for the VAR model. Thus, there are 508 potential predictors for the VAR model. For gLASSO, groups are determined by the 11 categories introduced in Table 1. Because variables with different lags are classified into different groups, there are 44 groups in total.

Our evaluation method of forecasting is similar to that of Nicholson et al. (2017). Now, we explain the case of lasso. We divide the whole sample period into three subperiods: the training period (May 2004 – June 2008), the validation period (July 2008 – June 2013), and the test period (July 2013 – June 2018). Then, we evaluate the performance by the MSFE of the test period. As described in Section 2.1, the regularization parameter of lasso is determined so that the MSFE of the validation period is minimized. Once the regularization parameter is determined, each forecast in the test period is obtained using a rolling window whose size is 60. For instance, the one-year-ahead forecast at July 2013 is obtained by estimating the VAR model using observations from August 2007 to July 2012. The value of the regularization parameter is unchanged to obtain all forecasts in the test period for the same forecasting horizon. Although it may be better to use different values of the regularization parameter, we used the same one because of the computational burden.

The performance of the DI model is also evaluated in a similar way. The number of factors is determined so as to minimize the MSFE of the validation period, and the performance of each method is evaluated by the MSFE of the test period. Note that DICV uses the same number of factors to obtain all forecasts of the same variable. In contrast, DILASSO may use a different number of factors because lasso can select different factors even when the value of the regularization parameter is unchanged throughout the test period.

Strictly speaking, the VAR model and the DI model utilize different information for forecasting because our DI model includes only current factors as predictors, whereas the VAR model can include past values of predictors. Thus, we also compared the performance between VAR(4) and VAR(1) to check whether including past values of predictors improved the MSFE. We do not report the result, which shows that there is no clear difference between VAR(4) and VAR(1) in terms of forecasting performance.

Tables 2–4 report the MSFEs of the test period for different forecasting horizons. The second column of each table reports the MSFE of the AR model in absolute terms. The third to eighth columns report the ratio of the MSFE of each method to that of the AR model. Therefore, values less than one indicate an improvement in forecasting performance over the AR model. The last row shows the number of cases where the forecast of each method outperforms that of the AR model. Asterisks denote leading indicators as defined by the Cabinet Office.

Table 2 shows the result for a one-month horizon ($h = 1$). Overall, the high-dimensional models outperform the benchmark model with a similar degree of improvement, although DIPC works rather poorly. DICV performs relatively better than the lasso-based approaches, indicating the possibility that including all the available information is helpful in predicting a short-term variation.

A closer look at the selected factors shows that DIPC selects 17–19 factors to predict most of the series, which makes the model susceptible to overfitting given the relatively small window size. This is likely caused by the time-independent restriction on the factor loadings and is consistent with the result of Hayakawa and Kobayashi (2011), who reported that the Japanese macroeconomy has 20 factors when the factors are selected by the criterion of Bai and Ng (2002). In contrast, DICV and DI-LASSO select less than 5 factors for most of the series and select more than 10 factors for only a few series.

MSFEs for the one-quarter horizon ($h = 3$) and the one-year horizon ($h = 12$) are reported in Tables 3 and 4, respectively. In general, the relative performances of the DI and VAR models are better in longer horizons. The better longer horizon performances are consistent with earlier studies; for example, Shintani (2005) and Callot and Kock (2014). At the one-year horizon, the VAR-based methods perform no worse than the DI-based methods, indicating that the sparsity assumed in the high-dimensional VAR model may be favorable for longer horizons.

Tables 2–4 also indicate that improvements in the forecasting performances vary between the series to be forecast. Across all the forecasting horizons, the extent of improvement is relatively small for the series belonging to the leading indicators defined by the Cabinet Office. These are the series that move sensitively to the macroeconomic changes, including TOPIX, machinery orders, and housing starts. This result suggests that the benefit of incorporating a large set of information is limited for forecasting leading indicators, albeit not negligible, compared with forecasting other series whose movements are less sensitive to the macroeconomic environment.

3.3 Information Exploited by the VAR and DI Models

The results of Section 3.2 show that forecasting performances are more or less similar for all methods except for DIPC. Now, we investigate the information utilized by each forecasting method.

Table 2 Ratio of MSFE ($h = 1$)

	AR(abs.)	DIPC	DICV	DILASSO	LASSO	ENET	gLASSO
Industrial Production	0.443	1.46	1.118	0.907	0.907	0.969	0.907
Capacity Utilization	0.334	1.786	1.231	0.925	0.927	0.927	0.927
Disposable Income	0.503	1.653	0.899	1.049	1.282	1.279	1.376
Employment Index	0.925	1.338	0.975	0.999	0.989	0.978	1.044
New Job Offers*	0.549	1.434	0.983	1.243	1.263	1.266	1.323
Worked Hours	0.61	0.858	0.644	0.687	0.658	0.657	0.689
Consumption	1.16	1.892	1.052	0.99	0.99	0.99	0.99
Housing Starts*	0.582	1.208	0.85	0.78	0.78	0.78	0.771
Machinery Orders*	0.91	1.172	0.975	1.153	1.118	1.12	1.144
TOPIX*	0.863	1.477	0.84	0.804	0.795	0.795	0.795
Nikkei 42*	0.508	1.432	0.924	1.059	1.209	1.216	1.139
Exchange Rate	1.246	1.188	0.753	0.763	0.699	0.712	0.714
10 Years Govt. Bond	0.459	1.074	0.788	0.735	0.741	0.741	0.741
3 Months TIBOR	0.412	0.544	0.412	0.421	0.404	0.404	0.398
Spread	0.048	0.092	0.081	0.382	1.17	1.143	6.501
Call Rate	0.295	1.042	0.788	0.644	0.619	0.619	0.668
Money Supply: M2*	0.797	1.008	0.807	0.792	0.793	0.821	0.78
Monetary Base	0.303	1.738	1.205	1.109	1.109	1.109	1.108
Consumer Price Index	3.12	0.823	0.656	0.593	0.643	0.594	0.691
Wage Index	0.585	0.965	0.871	1.044	1.312	1.255	1.27
# Outperform	—	5	16	13	13	13	12

Table 3 Ratio of MSFE ($h = 3$)

	AR(abs.)	DIPC	DICV	DILASSO	LASSO	ENET	gLASSO
Industrial Production	0.421	1.89	1.143	1.007	1.670	1.672	1.012
Capacity Utilization	0.280	2.033	1.255	1.062	1.240	1.238	1.064
Disposable Income	0.716	1.226	0.836	0.834	0.834	0.834	0.834
Employment Index	1.004	1.065	0.888	0.824	0.849	0.849	0.849
New Job Offers*	0.670	1.424	0.985	0.989	0.995	0.978	1.053
Worked Hours	0.53	1.288	1.014	0.914	1.132	1.136	1.109
Consumption	1.42	1.495	0.809	0.84	0.89	0.93	0.80
Housing Starts*	0.523	2.065	0.90	0.87	0.87	0.87	0.876
Machinery Orders*	1.39	1.227	0.830	0.809	0.800	0.77	0.796
TOPIX*	0.872	1.292	0.77	0.796	0.784	0.784	0.786
Nikkei 42*	0.721	1.422	0.859	0.923	0.904	0.895	0.901
Exchange Rate	1.506	1.049	0.708	0.673	0.673	0.673	0.673
10 Years Govt. Bond	0.626	1.234	0.737	0.730	0.730	0.730	0.730
3 Months TIBOR	0.363	0.927	0.514	0.482	0.455	0.450	0.437
Spread	0.301	0.511	0.391	0.646	1.41	1.414	1.479
Call Rate	0.305	1.092	0.664	0.623	0.623	0.623	0.636
Money Supply: M2*	0.846	1.125	0.859	0.809	0.809	0.820	0.82
Monetary Base	0.289	1.806	1.275	1.207	1.207	1.202	1.207
Consumer Price Index	3.08	0.624	0.547	0.580	0.574	0.571	0.577
Wage Index	0.915	1.058	0.742	0.893	0.774	0.777	0.77
# Outperform	—	3	16	17	15	15	14

Table 4 Ratio of MSFE ($h = 12$)

	AR(abs.)	DI	DICV	DILASSO	LASSO	ENET	gLASSO
Industrial Production	0.646	1.23	0.891	0.566	0.566	0.566	0.582
Capacity Utilization	0.573	1.404	0.888	0.464	0.464	0.464	0.480
Disposable Income	0.748	0.865	0.823	0.747	0.734	0.735	0.747
Employment Index	0.807	1.279	0.973	0.803	0.842	0.833	0.851
New Job Offers*	0.430	1.996	0.958	0.933	0.932	0.932	0.933
Worked Hours	0.61	1.198	0.968	0.733	0.820	0.822	0.838
Consumption	1.78	1.235	0.757	0.62	0.62	0.62	0.62
Housing Starts*	0.477	1.294	0.98	0.88	0.88	0.88	0.879
Machinery Orders*	1.41	1.274	0.872	0.843	0.842	0.84	0.842
TOPIX*	1.128	1.409	0.83	0.834	0.834	0.834	0.834
Nikkei 42*	0.789	1.198	0.956	0.846	0.846	0.846	0.846
Exchange Rate	1.345	1.447	1.070	0.928	0.930	0.928	0.934
10 Years Govt. Bond	0.634	1.216	0.814	0.785	0.785	0.785	0.799
3 Months TIBOR	0.367	0.782	0.572	0.427	0.427	0.425	0.438
Spread	0.438	1.318	0.908	1.011	1.01	1.011	1.011
Call Rate	0.229	1.459	0.835	0.750	0.750	0.742	0.750
Money Supply: M2*	0.851	1.283	1.008	0.935	0.911	0.857	0.90
Monetary Base	0.362	1.152	1.079	0.958	0.958	0.958	0.937
Consumer Price Index	2.21	0.959	0.812	0.795	0.790	0.773	0.853
Wage Index	0.897	1.027	0.882	0.744	0.744	0.744	0.74
# Outperform	—	3	17	19	19	19	19

Table 5 Average number of nonzero coefficients

	$h=1$			$h=3$			$h=12$		
	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO
Industrial Production	0	153.533	0	21.017	21.25	0	0	0	3.4
Capacity Utilization	0	56.867	0	0.583	0.583	0	0	0	3.533
Disposable Income	2.217	2.283	0	0	0	0	1.933	1.9	0
Employment Index	3.683	9.333	68.617	0.917	0.933	9.9	3.833	115.65	64.467
New Job Offers*	0.133	118.933	53.3	0	212.65	81.35	0.133	0.133	0.25
Worked Hours	8.533	8.983	94.85	27.417	28.45	142	16.983	16.467	59.483
Consumption	0	0	0	1.033	21.45	0	0	0	0
Housing Starts*	0	0	8.967	0	4.917	3.733	0	0	0
Machinery Orders*	3.317	4.617	5.833	0.483	206.183	54.35	0	0	0
TOPIX*	0	0	0	0.017	0.017	0.9	0	0	0
Nikkei 42*	14.333	26.333	101.933	0	51.267	27.7	0	0	0
Exchange Rate	4.733	95.467	52.8	0	0	0	0	31.75	5.55
10 Years Govt. Bond	0	0	0	0	0	0	0	0	1.433
3 Months TIBOR	0.55	0.617	8.5	2.8	32.7	10.867	0	37.55	4.1
Spread	14.217	14.9	65.95	0.333	52.6	9.167	0	0	0
Call Rate	0.5	0.55	144.883	0	5.183	12.067	0	20.467	0
Money Supply: M2*	0	156.45	64.883	0	159.15	1.917	0	200.983	2.867
Monetary Base	0	0	7.817	0	3.65	0	0.05	0.05	61.9
Consumer Price Index	16.067	41.483	139.783	3.85	5.483	9.567	0	105.133	171.1
Wage Index	2.617	22.717	0	0	31.633	0	0	0	0

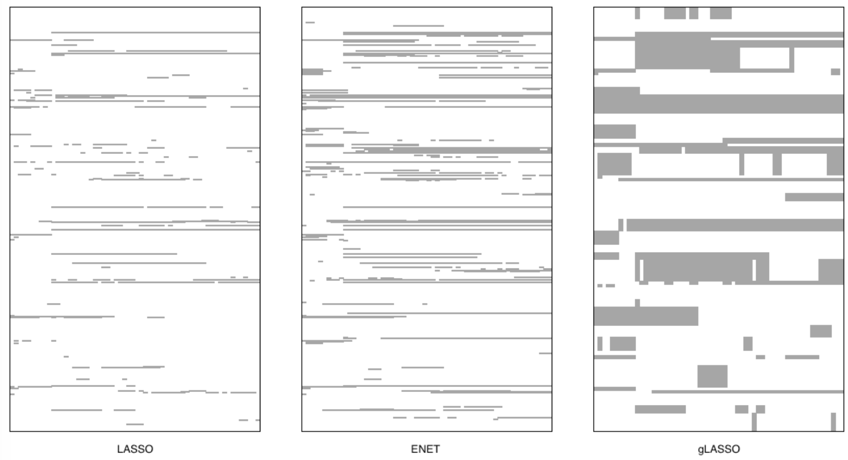

Fig. 1 Visual representations of the variable selection (CPI with $h = 1$)

Table 5 reports the average numbers of nonzero coefficients in the VAR model, where the averages are taken over 60 models used to obtain forecasts in the test period. We find that lasso selects more sparse models partly because of the restriction that the number of nonzero parameters cannot be greater than the number of observations used for estimation. It is also found that the sparsity tends to be high for the longer horizons and leading indicators. This finding is plausible considering the decrease in information for a longer horizon and the limited benefit of exploiting large data to predict a leading indicator.

Figure 1 is a visual representation of the variables selection for forecasting the consumer price index (CPI) with $h = 1$. The vertical axis shows 508 potential predictors, and the horizontal axis shows 60 time points in the test period. The gray grids, which represent nonzero coefficients, are thin and scattered in the case of lasso, indicating that the selected predictors are unstable compared with those of the elastic net and group lasso. This result is consistent with the discussion in Section 2.1 about the grouping feature of the elastic net and group lasso. Similar patterns are also found for the other forecast series and other forecasting horizons.

Next, we look into the frequency of the selected categories. We define a category as selected if at least one variable in the category is selected. Table 6 lists the three most selected categories for predicting CPI, industrial production, and employment index. These variables are hard to predict by a simple AR model or economically important indicators[4]. NA indicates that no category is selected. We observe that although the numbers of selected predictors are quite different among the three methods, the frequently selected categories are similar. For instance, exchange rates is frequently selected by all the methods to predict CPI when $h = 1$. This implies that lasso-based methods utilize similar information for forecasting even though the selected predictors are not necessarily the same.

[4] Because all variables are normalized to have variance unity, a high MSFE of the AR model implies that the variable is hard to predict.

Table 6 Mostly selected categories

	h	1st	2nd	3rd
		LASSO		
CPI	1	Exchange Rates	Stock Price	Price Indices
	3	Output and Income	Price Indices	Money
	12	NA	NA	NA
Industrial Production	1	NA	NA	NA
	3	Consumption	Inventories	Price Indices
	12	NA	NA	NA
Employment Index	1	Output and Income	Housing	Employment
	3	Inventories	Employment	NA
	12	Consumption	Inventories	Interest Rates
		ENET		
CPI	1	Exchange Rates	Stock Price	Housing
	3	Output and Income	Price Indices	Inventories
	12	Exchange Rates	Output and Income	Employment
Industrial Production	1	Money	Inventories	Output and Income
	3	Consumption	Inventories	Price Indices
	12	NA	NA	NA
Employment Index	1	Output and Income	Housing	Consumption
	3	Inventories	Employment	NA
	12	Consumption	Exchange Rates	Stock Price
		gLASSO		
CPI	1	Price Indices	Exchange Rates	Housing
	3	Output and Income	NA	NA
	12	Price Indices	Output and Income	Exchange Rates
Industrial Production	1	NA	NA	NA
	3	NA	NA	NA
	12	Stock Price	Housing	NA
Employment Index	1	Output and Income	Exchange Rates	Housing
	3	Consumption	Inventories	Housing
	12	Consumption	Stock Price	Price Indices

Concerning interpretation of the DI model, we extend the approach of Stock and Watson (2002b), who investigated the R^2 of regressing each variable on each estimated factor. They were interested in the economic interpretation of each factor. Instead, we look into R^2 when the individual series are regressed on a set of factors that are selected by DICV and DILASSO. Thus, we are interested in information exploited by the DI model to predict a particular variable.

Figure 2 shows the R^2s of regressing each series on all the selected factors to predict CPI when $h = 1$. The horizontal axis shows 127 series, which are ordered according to the sequence of Table 1, and the vertical axis shows the R^2s. Because the selected factors are different for different time points, we report the average of the R^2s over 60 cases. A high R^2 means that the variable is relevant to the selected factors, and the information contained in the variable is essential for forecasting CPI. We can find some clusters of variables with high values of R^2. This implies the importance of these groups in forecasting CPI.

Table 7 reports the ranking of categories by the average of R^2s, which is taken over series in the category and 60 time points in the test period. Comparing Tables 6 and 7, we see that categories with high R^2 are similar to the most selected categories by

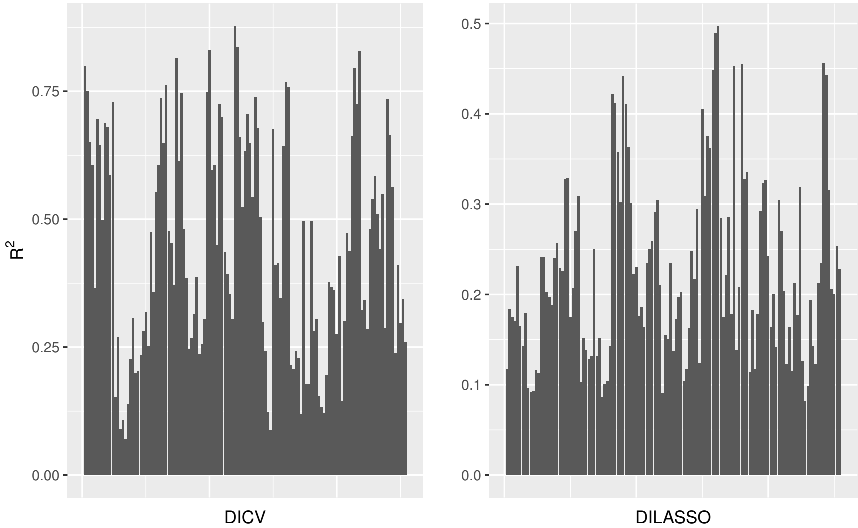

Fig. 2 R^2 between individual series and selected factors (CPI with $h = 1$)

Table 7 High R^2 categories

	h	1st	2nd	3rd
			DICV	
CPI	1	Exchange Rates	Consumption	Output and Income
	3	Consumption	Output and Income	Retail
	12	Output and Income	Consumption	Retail
Industrial Production	1	Exchange Rates	Consumption	Retail
	3	Consumption	Output and Income	Retail
	12	Output and Income	Consumption	Retail
Employment Index	1	Exchange Rates	Consumption	Output and Income
	3	Consumption	Output and Income	Retail
	12	Output and Income	Consumption	Retail
			DILASSO	
CPI	1	Exchange Rates	Housing	Stock Price
	3	Output and Income	Inventories	Housing
	12	Exchange Rates	Stock Price	Housing
Industrial Production	1	NA	NA	NA
	3	Consumption	Housing	Money
	12	NA	NA	NA
Employment Index	1	Output and Income	Consumption	Retail
	3	Housing	Price Indices	Interest Rates
	12	Housing	Interest Rates	Consumption

lasso-based methods. This implies that both the VAR and DI models exploit similar information for forecasting.

4 Conclusion

In this paper, we studied the forecasting performances of various high-dimensional models using 127 monthly Japanese macroeconomic data. We used lasso, elastic net, and group lasso to estimate the high-dimensional VAR model and three different methods to determine the number of factors for estimating the DI model. Overall, using high-dimensional data indeed improves the accuracy of forecasting the Japanese macroeconomic indicators except when the number of factors is determined by the criterion of Bai and Ng (2002). The contradicting performances among DI models are indicative of the importance of selecting factors relevant to forecasting.

This paper also shows that the forecasting performances of high-dimensional models tend to be better in the longer horizons and for variables not belonging to the leading indicators. Moreover, although not conclusive, the comparison between two models suggests an advantage of the DI model in shorter horizons and of the VAR model in longer horizons.

We also compared the VAR and DI models in terms of information utilized for forecasting. To see the information exploited by the DI model, we regressed each observed variable on estimated factors relevant to forecasting and examined which variables have high R^2. The results show that variables with high R^2 are also those that are likely to be selected by lasso, elastic net, and group lasso. This suggests that the information from similar groups of variables is utilized by both the VAR and DI models.

As a final remark, several limitations of this study are noted. The field of high-dimensional data analysis is a rapidly growing area of research, and there are a number of new methods that are not covered in this paper. One interesting extension of this research would be to incorporate nonlinearity, which gives more flexibility to the model (see, for example, Exterkate et al. 2016). There is also a growing interest in the area of forecast combination. If forecasts from different models are combined properly, then even if each model performs poorly, the combination often outperforms the forecast of a best-performing model (Li and Chen 2014; Swanson and Xiong 2018). Combining forecasts from the VAR and DI models may result in further improvement in the forecasting performances.

Acknowledgements We would like to thank Kazuhiko Hayakawa and Masahiko Shibamoto for their advice on creating our dataset. We also thank Shigeyuki Hamori for his comments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Altissimo F, Cristadoro R, Forni M, Lippi M, Veronese G (2010) New Eurocoin: Tracking economic growth in real time. Review of Economics and Statistics 92(4):1024–1034

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Statistics Surveys 4:40–79

Artis MJ, Banerjee A, Marcellino M (2005) Factor forecasts for the UK. Journal of Forecasting 24(4):279–298

Bai J (2003) Inferential theory for factor models of large dimensions. Econometrica 71(1):135–171

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

Bai J, Ng S (2008) Forecasting economic time series using targeted predictors. Journal of Econometrics 146(2):304–317

Banerjee A, Marcellino M, Masten I (2008) Forecasting macroeconomic variables using diffusion indexes in short sample with structural change. In: Rapach DE, Wohar ME (eds) Forecasting in the Presence of Structural Breaks and Model Uncertainty, Emerald Group Publishing, chap 4, pp 149–194

Bernanke BS, Boivin J, Ellasz P (2005) Measuring the effects of monetary policy: A factor-augumented vector autoregressive (FAVAR) approach. Quarterly Journal of Economics 120(1):387–422

Boivin J, Giannoni M (2006) DSGE models in a data-rich environment, NBER Working Paper Series, No. 12772

Burns AF, Mitchell WC (1946) Measuring Business Cycles. NBER, New York

Callot LAF, Kock AB (2014) Oracle efficient estimation and forecasting with the adaptive lasso and the adaptive group lasso in vector autoregressions. In: Essays in Nonlinear Time Series Econometrics, Oxford University Press

Chamberlain G, Rothschild M (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 51(5):1281–1304

Chikamatsu K, Hirakata N, Kido Y, Otaka K (2018) Nowcasting japanese GDPs, Bank of Japan Working Paper Series, No. 18-E-18

Dickey DA, Fuller WA (1981) Likelihood ratio statistic for autoregressive time series with a unit root. Econometrica 49(4):1057–1072

Exterkate P, Groenen PJF, Heij C, van Dijk D (2016) Nonlinear forecasting with many predictors using kernel ridge regression. International Journal of Forecasting 32(3):736–753

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

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

Hayakawa K, Kobayashi Y (2011) Business cycle analysis using a large-scale macro data. In: Quantitative Analysis of Finance and Business Cycle, Minerva Shobo, pp 91–111, in Japanese

Kowarik A, Meraner A, Templ M (2014) Seasonal adjustment with the R packages x12 and x12GUI. Journal of Statistical Software 62(1):1–21

Li J, Chen W (2014) Forecasting macroeconomic time series: LASSO-based approach and their forecast combinations with dynamic factor models. International Journal of Forecasting 30(4):996–1015

Ludvigson SC, Ng S (2009) Macro factors in bond risk premia. Review of Finacial Studies 22(12):5027–5067

Marsilli C (2014) Variable selection in predictive MIDAS models, Banque de France Working Paper

Meier L (2015) grplasso: Fitting User Specified Models with Group Lasso Penalty. R package version 0.4-5

Nicholson WB, Matteson DS, Bien J (2017) VARX-L: Structured regularization for large vector autoregressions with exogenous variables. International Journal of Forecasting 33(3):627–651

Pfaff B (2008) Analysis of Integrated and Cointegrated Time Series with R. Springer

Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75(2):335–346

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, URL https://www.R-project.org/, vienna Austria

Shibamoto M (2007) An analysis of monetary policy shocks in japan: A factor augumented vector autoregressive approach. Japanese Economic Review 58(4):484–503

Shintani M (2005) Nonlinear forecasting analysis using diffusion indexes: An application to japan. Journal of Money, Credit and Banking 37(3):517–538

Smeekes S, Wijler E (2018) Macroeconomic forecasting using penalized regression. International Journal of Forecasting 34(3):408–430

Song S, Bickel PJ (2011) Large vector auto regressions, arXiv:1106.3915

Stock JH, Watson MW (2002a) Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97(460):1167–1179

Stock JH, Watson MW (2002b) Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20(2):147–162

Stock JH, Watson MW (2005) Implications of dynamic factor models for VAR analysis, NBER Working Paper Series, No.11467

Swanson NR, Xiong W (2018) Big data analytics in economics: What have we learned so far, and where should we go from here? Canadian Journal of Economics 51(3):695–746

Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B 58(1):267–288

Uematsu Y, Tanaka S (2019) High-dimensional macroeconomic forecasting and variable selection via penalized regression. Econometrics Journal 22:34–56

Yamamoto Y (2019) Bootstrap inference for impulse response functions in factor-augumented vector autoregressions. Journal of Applied Econometrics 34(2):247–267

Yamamoto Y, Tanaka S (2015) Testing for factor loading structural change under common breaks. Journal of Econometrics 189(1):187–206

Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B 68(1):49–67

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