Forecasting Japanese Macroeconomy Using High-Dimensional Data

by

Yoshiki Nakajima

A Thesis Presented to

Graduate School of Economics

Kobe University

In Partial Fulfillment of the

Requirements for the Degree of

Masters of Arts

Major in Economics

Supervised by

Associate Professor Naoya Sueishi

January 21, 2019

Table of Contents

1. Introduction

In recent years, large-scale data has become increasingly available due to the development of computer technology. When dealing with those high-dimensional data, the traditional statistical methodologies, whose model selection relies on a sequence of tests or the application of information criteria, are faced with the curse of dimensionality because of the exponentially growing cost of calculations (Cock and Kallot, 2014). In addition, the dataset in macroeconomics tends to have a unique structure in that it has a highly-correlated wide variety of variables observed less frequently, resulting in the estimation to be unstable or even infeasible. Conventional methods such as the vector autoregression, therefore, rely on econometricians to select a small number of variables, which inevitably involves an arbitrariness as well as the loss of information.

Supported by the rise of high-performance computer systems, a number of methodologies are proposed to circumvent the problems in conventional econometrics since the end of the twentieth century. The idea is to reduce the dimensionality of the original high-dimensional dataset with the least loss of information possible without the arbitrariness. Those methodologies are broadly classified into two broad categories of approaches. The first one is to reduce the parameter space by including only a subset of variables. Tibshirani (1996) introduced a penalized regression called the lasso, followed by a series of methodologies including the elastic net (Zou and Hastie, 2005) and group lasso (Yuan and Lin, 2006). The applications of those models to the macroeconomic forecast are studied, for example, by Song and Bickel (2011), Li and Chen (2014), and Callot and Kock (2014) for the U.S. but not many for other countries.

Another class of approach is to exploit a small number of latent variables that affect a large number of predictors and is called the dynamic factor model. The idea that a few unobservable common factors have an impact on a large set of observable variables is well rooted in economics, as in Burns and Mitchell's (1946) definition of the business cycle as "a type of fluctuation found in the aggregate economic activity." Since the seminal work by Stock and Watson (2002a), this technique has widely been used in macroeconomics, not only for forecasting but also for other purposes such as the evaluation of monetary policy (Bernanke et al., 2005 for the U.S.; Shibamoto, 2007 for Japan), the estimation of DSGE models (Boivin and Giannone, 2006) and the development of real-time economic indicators (Altissimo et al., 2010 for the Euro Area; Hayakawa and Kobayashi, 2011 for Japan). Several empirical works in macroeconomic forecastings include Stock and Watson (2002a,b) for the macroeconomic indicators in the U. S., Atris et al. (2005) for the U.K., Banerjee et al. (2009) for the Euro Area and Shintani (2005) for Japan.

This thesis makes three contributions. 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 for the better understanding of country's economic structure on which policymakers and businesses make a decision (Li and Chen, 2014), yet few studies have been done to explore the benefit of high-dimensional data in Japan.

Although Shintani (2005) is a leading case in Japan to show the benefit of exploiting the high-dimensional data, it is limited to the diffusion index model, and the focus is purely on forecasting and not on the interpretability. Our attempt is the first of its kind to widely study the forecasting performances of various high-dimensional models to explore the possibility of those models in Japan as well as to provide the interpretation of the results.

Second, we propose two forecast-oriented, data-dependent methods of selecting the number of factors for the dynamic factor model. A commonly used approach in the literature is either to use the criteria proposed by Bai and Ng (2002) or to let econometricians decide it, usually between one and six. The former tend to select the excessive number of factors because it is designed to estimate the number of factors in terms of the relevance to the original dataset, which may not be optimal for forecasting, whereas the latter involves an arbitrary choice. We propose a cross validation-based approach of selecting the number of factors that are optimal for forecasting and without the arbitrariness.

Third, we provide a R^2-based approach of interpreting the relationship between the latent common factors and the original dataset in terms of the predictive power of the dynamic factor model. One of the problems of the dynamic factor models is the interpretability of the latent factors because they are affected by a large number of variables in the original dataset. Stock and Watson (2002b) and Ludvigson and Ng (2009) have studied the economic interpretations of the factors by looking into R^2 but is limited to the general interpretation of individual factors and does not account for the predictive power of those factors. We attempt to extend their approach and provide the interpretability to the common factors.

We have conducted an empirical analysis for the forecasting performances of several high-dimensional models using 127 monthly Japanese macroeconomic data spanning from April 2003 to June 2018. Our main findings are; 1) Incorporating the abundant information contributes to the improvements in the accuracy of forecasting, except for the dynamic factor models with the number of factors selected using Bai and Ng's (2002) criterion. We have also found that the extent of improvement is relatively small when forecasting the leading indicators and the benefit of using high-dimensional data is large for longer horizons. 2) Selecting the number of factors in terms of predictive power improves forecasting performances. The number of selected factors is much smaller in this approach than the one selected by Bai and Ng's (2002) criteria, which performed poorly in forecasting. Our proposed approach performs as good as the other high-dimensional models employed and have outperformed the benchmark model. 3) When factors are estimated in terms of the predictive power, the estimated factors load heavily on the variables that are selected in the lasso and other variable selection methods, indicating that both variable selection and common factor approaches exploit the information from the same groups of variables. The groups shown to be important are different depending on the variables to forecast, yet are mostly consistent across models.

The remainder of this thesis is organized as follows. Section 2 introduces the high-dimensional models along with their backgrounds. Dataset and implementation notes are given in Section 3. Section 4 reports and interprets the results and Section 5 concludes.

2. Models

2.1 VAR models

The vector autoregression (VAR), proposed by Sims (1980), is a widely used tool in various areas of macroeconomics to capture rich dynamics in multiple time series (Stock and Watson, 2001). Assume that we observe N macroeconomic variables $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$, $t = 1,2,...,T$ that follow the stationary process. Then, h-step-ahead direct forecast is expressed in the following VAR form;

$$y_{t+h} = \mu + \sum_{p=0}^{P-1} \Phi_p y_{t-p} + \varepsilon_{t+h} \tag{1}$$

or, equivalently,

$$y_{i,t+h} = \mu_i + \sum_{p=0}^{P-1} \Phi_{p,i} y_{t-p} + \varepsilon_{i,t+h} \ , \tag{2}$$

where μ denotes a N-dimensional vector of intercepts, Φ_p is a $N \times N$ matrix of coefficients at lag p, ε_t represents N-dimensional error term assumed to be i.i.d. with mean zero and covariance matrix of Σ. Subscript i denotes a scaler of i-th element of a y_t, μ and ε_t , or i-th row of Φ_p (such that $\Phi_{p,i}$ is a $1 \times N$ row vector). P and h are the lag order and forecasting horizon, respectively.

In a low-dimensional setting, where the number of independent variables is much smaller than the number of observations, the parameters are estimated by solving the following objective function. Let $\Phi = (\Phi_0, \ldots, \Phi_{P-1})'$ and $\| \cdot \|_2$ denotes L2 norm, then $\hat{\mu}$ and $\hat{\Phi}$ are given such that

$$(\hat{\mu}, \hat{\Phi})' = \underset{\mu, \Phi}{\arg\min} \sum_{t=1}^{T} \| y_{t+h} - \mu - \sum_{p=0}^{P-1} \Phi_p y_{t-p} \|_2^2 \tag{3}$$

or, equivalently,

$$(\hat{\mu}_i, \hat{\Phi}_i)' = \underset{\mu_i, \Phi_i}{\arg\min} \sum_{t=1}^{T} \| y_{i,t+h} - \mu_i - \sum_{p=0}^{P-1} \Phi_{p,i} y_{t-p} \|_2^2 \tag{4}$$

This minimization problem can be solved explicitly by the OLS estimator. Namely, define θ and X as $(\mu_i, \Phi_i)'$ and $(1, y_t', \ldots, y_{t-p}')$, then $\hat{\theta} = (X'X)^{-1} X' y_{t+h}$.

The potential problem of the VAR, however, is that the number of parameters to estimate ($N^2 P$) grows quadratically as the number of variables (N) increases. The result is the high standard errors or noninvertibility of $X'X$, when the number of parameters exceeds the number of observations (T). Although it is theoretically possible to consider all the combinations of the variables and select the model that minimizes some criteria such as AIC or BIC, this approach tends to be computationally too expensive as the combinations grow exponentially ($2^{N^2 P}$). The lasso is an efficient method of selecting variables using penalized regression and is proposed by Tibshirani (1996). Specifically, the lasso estimator imposes L1 norm constraint on the OLS estimator and expressed as follows. For some $s > 0$,

$$(\hat{\mu}_i, \hat{\Phi}_i)' = \arg\min_{\mu_i, \Phi_i} \sum_{t=1}^{T} \|y_{i,t+h} - \mu_i - \sum_{p=0}^{P-1} \Phi_{p,i} y_{t-p}\|_2^2 \tag{5}$$

subject to

$$\sum_{p=0}^{P-1} \|\Phi_{p,i}\|_1 \leq s \ , \tag{6}$$

where $\|\cdot\|_1$ is L1 norm. This equation can be rewritten using a regularization parameter, $\lambda > 0$, to the following.

$$(\hat{\mu}_i, \hat{\Phi}_i)' = \arg\min_{\mu_i, \Phi_i} \sum_{t=1}^{T} \|y_{i,t+h} - \mu_i - \sum_{p=0}^{P-1} \Phi_{p,i} y_{t-p}\|_2^2 + \lambda \sum_{p=0}^{P-1} \|\Phi_{p,i}\|_1 \tag{7}$$

This lasso estimator is known to shrink some coefficients and set others to zero, hence a sparse solution. The sparsity of the model depends on the regularization parameter, λ, and the larger λ selects the more sparse model.

Although the lasso has been successful in many applications, it has some limitations to certain types of dataset. First, the number of predictors that lasso uniquely selects is limited up to the number of observations. This is problematic when dataset contains many variables that exceed the number of observations, which is often the case in the VAR. Secondly, when variables are highly correlated, the lasso tends to select one or only a few of those variables and set coefficients of other variables to zero. Tibshirani (1996) has shown that in the existence of high correlations, the lasso is outperformed by ridge regression, which imposes L2 norm constraint and is proposed by Hoerl and Kennard (1970). In addition to the deterioration of the predictive performances, this makes the relationship between independent variables and dependent variable less clear, which is not favorable when statistician's interest lies in the interpretations of those relationships.

Zou and Hastie (2005) proposed the elastic net to solve the drawbacks of the lasso by combining the lasso with ridge regression to yield the following objective function.

$$\begin{aligned}(\hat{\mu}_i, \hat{\Phi}_i)' = \arg\min_{\mu_i, \Phi_i} &\sum_{t=1}^{T} \|y_{i,t+h} - \mu_i - \sum_{p=0}^{P-1} \Phi_{p,i} y_{t-p}\|_2^2 \\ &+ \lambda \sum_{p=0}^{P-1} \Big((1-\alpha)\|\Phi_{p,i}\|_1 + \alpha\|\Phi_{p,i}\|_2^2\Big),\end{aligned} \tag{8}$$

where α is a tuning parameter that controls the weight of the lasso and ridge and is between 0 and 1. Setting α to be 0 is equivalent to the lasso regression and $\alpha = 1$ to the ridge regression.

Zou and Hastie (2005) have shown the group effect of the elastic net. Namely, it is able to select variables with high correlation altogether. It is also shown that the elastic net can select more variables than observations owing to the features of ridge estimator.

Another approach is the group lasso proposed by Yuan and Lin (2006). This method simultaneously includes or excludes a set of variables based on predetermined groups. Although the group lasso is more restrictive than elastic net in the sense that groups are defined in advance, the problem rarely arises in fields like macroeconomics where the variables have certain attributes such as output and consumption. Suppose that the dataset can be partitioned into L groups and its lags and let d_l represent the dimension of each group with $l = 1,\ldots, L$. Then Eq. (2) is rewritten as

$$y_{i,t+h} = \mu_i + \sum_{p=0}^{P-1} (\Phi_{1p,i}, \ \ldots, \Phi_{Lp,i}) \ (y_{1,t-p}{}', \ldots, y_{L,t-p}{}')' + \varepsilon_{i,t+h} \ , \tag{9}$$

where $\Phi_{lp,i}$ is i-th row of a Φ_{lp} ($N \times d_k$ matrix) with p being the lag and $y_{l.}$ is a $d_l \times 1$ vector. The estimator of the group lasso is then defined as

$$(\hat{\mu}_i, \hat{\Phi}_i)' = \arg\min_{\mu_i, \Phi_i} \sum_{t=1}^{T} \|y_{i,t+h} - \mu_i - \sum_{p=0}^{P-1} \Phi_{p,i} y_{i,t-p}\|_2^2 + \lambda \sum_{p=0}^{P-1} \sum_{l=1}^{L} \sqrt{d_l} \ \|\Phi_{lp,i}\|_2 \ , \tag{10}$$

with $\sqrt{d_l}$ controlling the different dimensions between groups to avoid favoring groups with higher dimension and $\|\Phi_{lp,i}\|_2$ allowing the model to select variables in a block level. When each variable are grouped into a different group ($d_l = 1$ for all l), the Eq. (10) is equivalent to Eq. (7).

All the lasso-based methodologies described above require an appropriate choice of tuning parameters, λ and α, which are unknown in practice. In typical applications where observations are assumed to be independent of each other, a standard method is to use K-fold cross-validation. Namely, a dataset is randomly split into K subsets and parameters are estimated for different values of tuning parameter using $K - 1$ subsets of data. After estimating the parameters from $K - 1$ subsets, predictions are made for the holdout set and compared to the true observation. This step is taken K times, and the tuning parameter with the smallest average prediction error is selected. This approach controls the in-sample and out-of-sample performances and thus avoids the overfitting.

In time series applications, however, observations are time-dependent and randomly splitting the dataset without taking into account this dependence is not appropriate. Following Bańbura et al. (2010), Nicholson et al. (2017) have provided the method of choosing tuning parameters by minimizing the h-step-ahead mean squared forecast error (MSFE). They split the dataset into three subperiods: initialization, tuning parameter selection, and forecast evaluation. The model is first fitted on the data in the initialization period, and h-period-ahead forecast is estimated for the second subperiod using the different values of tuning parameters. This step is repeated until all the forecasts are obtained for the tuning parameter selection period. At the end of this period, the model with the smallest MSFE is chosen and is used for the evaluation period. The final MSFE to be reported is the one from this evaluation period.

2.2 Diffusion index models

Another class of approaches is to exploit the common factors that capture a large part of variances of the high-dimensional dataset. The common factors are often referred to as the diffusion indices in macroeconomics, hence the model "diffusion index model." Geweke (1977) introduced the dynamic factor model as a time-series extension of the factor models. Sargent and Sims (1977), for example, showed that two dynamic factors, extracted from 89 macro-economic variables, can explain a large part of the variances of the original dataset including GNP and wage. These early works on dynamic factor model, however, require time-domain observations to be transformed into a frequency domain in order to extract the dynamic factors. This indirect method of estimation makes the model unable to be applied into forecasting and other purposes (Stock and Watoson, 2011). Although direct estimation of dynamic factors using Kalman filter is explored, for instance, by Engle and Watson (1981, 1983), it is limited to the cases where a dataset is low-dimensional.

Stock and Watson (2002a) proposed the use of principal component to directly estimate the dynamic factors, which can be easily applied for forecasting. Following notations from the previous section, this approach consists of the two forms.

$$y_t = \Lambda F_t + e_t \tag{11}$$

$$y_{i,t+h} = \nu_i + \beta_F' F_t + \sum_{p=0}^{P-1} \beta_p y_{i,t-p} + u_{i,t+h} \quad , \tag{12}$$

where F_t denotes a $r \times 1$ vector of unobservable factors, Λ represents a $N \times r$ factor loading and e_t is a $N \times 1$ idiosyncratic disturbance assumed to have weak serial and cross-sectional correlation. ν, β and u_t are constant, coefficient and error term, respectively.

The estimation are also done in two steps. The first step is to solve F_t and Λ in Eq. (11) such that the following objective function is minimized. The estimated factor, $\hat{F}_t$, is then substituted to Eq. (12) in the second step. With the number of factors, r, fixed, the objective function in the first step is written as;

$$\min_{\Lambda, F_t} \sum_{t=1}^{T} \left((y_t - \Lambda F_t)'(y_t - \Lambda F_t) \right) \ , \tag{13}$$

Connor and Korajczyk (1986) have shown that the principal component estimator is a consistent estimate of F_t under fixed T and $N \to \infty$ when e_t is serially correlated but not cross-sectionally dependent. Stock and Watson (2002a) have proven that this consistency holds for the approximate factor model[1], where e_t is weakly dependent both serially and cross-sectionally,

[1] see Chamberlain and Rothschild (1982) for the precise definition of approximate factor model.

under $N, T \to \infty$. Bai (2003) provides the asymptotic normality of the approximate factor model under the double asymptotic condition ($N, T \to \infty$) and $\sqrt{N}/T \to 0$. Thus, the dynamic factor and factor loadings are estimated as follows.

$$\hat{F} = \sqrt{T} \times \Gamma \tag{14}$$

$$\hat{\Lambda}' = \hat{F}'X/T\ , \tag{15}$$

where $X = (y_1, \ldots, y_T)'$, $F = (F_1, \ldots, F_T)'$, and Γ is a $T \times r$ matrix representing r eigenvectors associated with r largest eigenvalues of XX'/NT.

The forecasting performances of the diffusion index models depend on the number of factors included in Eq. (12). Bai and Ng (2002) proposed the following information criterion that balances the trade-off between including additional factors and parsimoniousness of the model.

$$\begin{aligned} PC(r) = \operatorname{tr}\Big(\frac{(X - \hat{F}_r\hat{\Lambda}_r')'(X - \hat{F}_r\hat{\Lambda}_r')}{NT}\Big) \\ + r\hat{\sigma}^2\frac{N+T}{NT}\ln\big(\min(N, T)\big), \end{aligned} \tag{16}$$

where *tr* denotes matrix trace, $\hat{F}_r$ and $\hat{\Lambda}_r$ are the estimated factors and factor loadings under r factors, respectively. $\hat{\sigma}^2$ is defined as $\operatorname{tr}(e'e/T)$ with $e = (e_1, \ldots, e_T)'$, which can be replaced by $\operatorname{tr}\big((X - \hat{F}_{\mathrm{rmax}}\hat{\Lambda}'_{\mathrm{rmax}})'(X - \hat{F}_{\mathrm{rmax}}\hat{\Lambda}'_{\mathrm{rmax}})/NT\big)$ in practice, where *rmax* is a bounded integer such that $r \leq rmax$. (Bai and Ng, 2002; Mallows, 1973). As in AIC or BIC, the first term of the right-hand side is a monotonically decreasing function of r, representing the loss of information by reducing the dimensionality while the second term monotonically increases serving as a penalty of including additional factors. The optimum number of factors, $\hat{r}$, is then estimated as a minimizer of Eq. (16).

This approach, however, is often problematic for two reasons. First, since the information criteria are designed to capture the variances with fewer factors, $\hat{r}$ is estimated irrespective of the predictive power of Eq. (12), and the same number of factors are used without taking into the account what variable to forecast. Second, since factor loadings are assumed to be independent of t, it fails to incorporate the structural breaks of the factor loadings. In the presence of the structural breaks, the factors are estimated separately for before and after the breaks, leading to the information criteria tending to select $\hat{r}$ that is larger than it actually is. (Stock and Watson, 2005).

Motivated by the discussions of the lasso-based techniques, we propose two alternative techniques for selecting factors. First one is to select r in the same way the tuning parameters are selected in the lasso Namely, we first split data into three subperiods and estimate factors up to *rmax* using the first subperiod. Then, h-step-ahead forecasts for the second subperiod are estimated for different values of r and are compared with the true observations. After obtaining all the predicted value for the second period, the optimum r is selected to minimize the MSFE and is used to predict the third subperiod. The benefit of this approach is that it provides the optimum number of factors in terms of the predictive power of the target variable, while Bai and Ng (2002) provide the number of factors optimum for explaining the original high-dimensional dataset.

Another approach we propose is a lasso-type method of selecting factors and lags. A relatively large number of factors (e.g., $r = 20$) are estimated in the first step, and all the factors are included in the second step as well as the lags of the target variable. The lasso-type objective function is then defined as follows to select a subset of those factors and lags from the set of possible factors and lags.

$$\arg\min_{\nu,\beta} \sum_{t=1}^{T} \| y_{i,t+h} - \nu_i - \beta_F' \hat{F}_t - \sum_{p=0}^{P-1} \beta_p y_{i,t-p} \|_2^2 + \lambda \|\beta\|_1 \ , \tag{17}$$

where $\beta = (\beta_F, \beta_0, \ldots, \beta_{P-1})$. This method encourages the sparsity in the factors and lags, thus enabling the model to select factors with high predictive power as well as allowing for the flexible lag selection. It is noted that selected factors may have little relevance to the original dataset when the last factors are selected, while our first approach guarantees a certain degree of relevance because it selects the first r factors.

3. Data and Implementation

The dataset consists of 127 monthly macroeconomic time series from 11 categories, selected based on Hayakawa and Kobayashi (2011) and sourced from Thomson Reuter Datastream, Nikkei NEEDS Financial Quest, and the ministerial websites. The period spans from April 2003 to June 2018 and is split into three subperiods: April 2003 - June 2008, July 2008 - June 2013, and July 2013 - June 2018. Based on the unreported augmented Dickey-Fuller test (Dickey and Fuller, 1981) and Phillips-Perron test (Phillips and Perron, 1988) for the unit root, all the series are transformed by taking the first difference of the logs except for interest rates, unemployment rate and job openings-to-applicants ratio in which the first difference is taken. Seasonal adjustment method follows X13-ARIMA-SEATS by the US Census Bureau when the adjusted series is not available from the source. After all the transformations, the data is standardized to have mean zero and standard deviation of unity. The detailed descriptions of the series are given in Table 1.

All the empirical analyses were done in R 3.4.2 (R core team, 2017). Several packages are used for the implementation; *glmnet* for the lasso and elastic net (Friedman et al., 2010), *grplasso* for the group lasso (Meier, 2015), *urca* for the unit root tests (Pfaff, 2008) and *x12* for the seasonal adjustment (Kowarik et al., 2014). All of the packages above are publicly available at CRAN[2]. Dataset after transformation and implementation codes are also available at author's GitHub page[3].

---

[2] www.cran.r-project.org

[3] www.github.com/yoshiki146

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 — Non-Durable 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, NSA)	DS	2
14	**Household Disposable Income — Workers (thousand yen, NSA)**	DS	2
Employment and Hours			
15	Unemptoyment Rate (%, SA)	DS	3
16	Number of Unemployment (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 (person, SA)	MHLW	1
22	Effective Job Offers — Part-Time (person, SA)	MHLW	1
23	New Job Openings-to-Applicants Ratio — Excluding 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 (person, SA)**	MHLW	1
26	New Job Offers — Part-Time (person, 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 Non-Scheduled Hours Worked — All industry (2015=100, SA)	MHLW	1
30	Index of Non-Scheduled Hours Worked — Manufacturing (2015=100, SA)	MHLW	1
Retail, Manufacturing and Trade Sales			
31	Sales at Departmnent Stores — Total (million yen, NSA)	METI	2
32	Wholesale Sales Value (billion yen, NSA)	METI	2
33	Retail Sales Value (billion yen, NSA)	METI	2
34	Import Volume Index — Total (2015=100, NSA)	DS	2
35	Export Volume Index — Total (2015=100, NSA)	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 — Non-Durable Goods (2011=100, SA)	DS	1
40	Motor Vehicle New Registrations — Passenger Cars Excl. Below 660c (NSA)	DS	2

Housing Starts and Sales			
41	Total Number of New Housing Construction Started — Built for 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 lnventory 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 lnventory of Finished Goods — Investment Goods (2010=100, SA)	METI	1
52	Index of Producer's lnventory of Finished Goods — Capital Goods (2010=100, SA)	METI	1
53	Index of Producer's lnventory of Finished Goods — Construction Goods (2010=100, SA)	METI	1
54	Index of Producer's lnventory of Finished Goods — Consumer Goods (2010=100, SA)	METI	1
55	Index of Producer's lnventory of Finished Goods — Durable Consumer Goods (2010=100, SA)	METI	1
56	Index of Producer's lnventory of Finished Goods — Nondurable Consumer Goods (2010=100, SA)	METI	1
57	Index of Producer's lnventory of Finished Goods — Producer Goods (2010=100, SA)	METI	1
58	Index of Producer's lnventory of Finished Goods — Producer Goods for Mining and Manufacturing (2010=100, SA)	METI	1
59	Index of Producer's lnventory of Finished Goods — Producer Goods for Others (2010=100, SA)	METI	1
60	Index of Producer's lnventory Ratio of Finished Goods — Mining and Manufacturing (2010=100, SA)	METI	1
61	Index of Producer's lnventory Ratio of Finished Goods — Final Demand Goods (2010=100, SA)	METI	1
62	Index of Producer's lnventory Ratio of Finished Goods — Investment Goods (2010=100, SA)	METI	1
63	Index of Producer's lnventory Ratio of Finished Goods — Capital Goods (2010=100, SA)	METI	1
64	Index of Producer's lnventory Ratio of Finished Goods — Construction Goods (2010=100, SA)	METI	1
65	Index of Producer's lnventory Ratio of Finished Goods — Consumer Goods (2010=100, SA)	METI	1
66	Index of Producer's lnventory Ratio of Finished Goods — Durable Consumer Goods (2010=100, SA)	METI	1

67	Index of Producer's lnventory Ratio of Finished Goods — Nondurable Consumer Goods (2010=100, SA)	METI	1
68	Index of Producer's lnventory Ratio of Finished Goods — Producer Goods (2010=100, SA)	METI	1
69	Index of Producer's lnventory Ratio of Finished Goods — Producer Goods for Mining and Manufacturing (2010=100, SA)	METI	1
70	Index of Producer's lnventory 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 (NSA)	DS	1
76	**Tokyo Stock Price Index (TOPIX; NSA)**	DS	1
77	**Nikkei Commodity Price Index — 42 items (1970=100, NSA)**	DS	2
78	Nikkei Commodity Price Index — 17 items (1970=100, NSA)	DS	2
Exchange Rates			
79	**US.Dollar-Yen Spot Rate — Average in the Month (JPY/USD, NSA)**	BOJ	1
80	Nominal Effective Exchange Rates (2010=100, NSA)	BOJ	1
81	Real Effective Exchange Rates (2010=100, NSA)	BOJ	1
Interest Rates			
82	**Newly Issued Government Bonds Yield — 10 Years (% per annum, NSA)**	DS	3
83	**Tokyo Interbank Offered Rates (TIBOR) — 3 Months (% per annum, NSA)**	DS	3
84	**Interest Rate Spread (normalized, SA)**	DS	3
85	Yield of Interest-Bearing Government Bonds — 10 Years (% per annum, NSA)	DS	3
86	Long-Term Prime Lending Rates (% per annum, NSA)	DS	3
87	Short-Term Prime Lending Rates (% per annum, NSA)	DS	3
88	**Call Rate — Uncollateralized Overnight, Average in the Month (% per annum, NSA)**	BOJ	3
89	Call Rate — Uncollateralized Overnight, End of Month (% per annum, NSA)	BOJ	3
90	The Basic Discount Rate and Basic Loan Rate (% per annum, NSA)	BOJ	3
91	Avg. Contracted New Loan and Discount — City Banks (% per annum, NSA)	DS	3
92	Avg. Contracted New Loan and Discount — City Banks, Short-term (% per annum, NSA)	DS	3
93	Avg. Contracted New Loan and Discount — City Banks, Long-term (% per annum, NSA)	DS	3
94	Avg. Contracted General Incl. Overdraft Rate — City Banks (% per annum, NSA)	DS	3
95	Avg. Contracted Short-term Rate — City Banks (% per annum, NSA)	DS	3
96	Avg. Contracted Long-term Rate — City Banks (% per annum, NSA)	DS	3
Money and Credit Quantity Aggregate			
97	Money Supply: M1 (billion yen, NSA)	DS	2
98	**Money Supply: M2 (billion yen, NSA)**	DS	2
99	Money Supply: M3 (billion yen, NSA)	DS	2
100	Money Supply: L (billion yen, SA)	DS	1
101	**Monetary Base — Banknotes in Circulation, Average Amounts Outstanding (hundred million yen, NSA)**	BOJ	2

102	Monetary Base — Coins in Circulation, Average Amounts Outstanding (hundred million yen, NSA)	BOJ	2
103	Monetary Base — Reserves (hundred million yen, NSA)	DS	2
104	Amount of Clearing — Value (million yen, NSA)	DS	2
105	Amount of Clearing — Number of bills (thousand, NSA)	DS	2
Price Indices and Wages			
106	Consumer Price Index — General (2005=100, NSA)	MIC	2
107	**Consumer Price Index —General, Excl. Fresh Food (2005=100, NSA)**	MIC	2
108	Consumer Price Index — General, Excl. Imputed Rent (2005=100, NSA)	MIC	2
109	Consumer Price Index — General, Excl. Imputed Rent and Fresh Food (2005=100, NSA)	MIC	2
110	Consumer Price Index — Food (2005=100, NSA)	MIC	2
111	Consumer Price Index — Housing (2005=100, NSA)	MIC	2
112	Consumer Price Index — Fuel, Light and Water Charges (2005=100, NSA)	MIC	2
113	Consumer Price Index — Furniture and Household Utensils (2005=100, NSA)	MIC	2
114	Consumer Price Index — Clothes and Footwear (2005=100, NSA)	MIC	2
115	Consumer Price Index — Medical Care (2005=100, NSA)	MIC	2
116	Consumer Price Index — Transportation and Communication (2005=100, NSA)	MIC	2
117	Consumer Price Index — Education (2005=100, NSA)	MIC	2
118	Consumer Price Index — Reading and Recreation (2005=100, NSA)	MIC	2
119	Consumer Price Index — Miscellaneous (2005=100, NSA)	MIC	2
120	Producer Price Index — All Commodities (2015=100, NSA)	DS	2
121	Export Price Index — Total Average (2015=100, NSA)	BOJ	2
122	Import Price Index — Total Average (2015=100, NSA)	BOJ	2
123	Terms of Trade Index — All Commodities (2015=100, NSA)	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

Table 1: The list of time series and their categorizations. The format is; series number, series name, source, and transformation code. (N)SA denotes (non-)seasonal adjustment, and the series to be forecasted are in bold. Source abbreviations are; METI: Ministry of Economy, Trade and Industry; DS: Thomson Reuters Datastream; MHLW: Ministry of Health, Labour and Welfare; NEEDS: Nikkei NEEDS Financial QUEST; MLIT: Ministry of Land, Infrastructure, Transport and Tourism; BOJ: Bank of Japan; MIC: Ministry of Internal Affairs and Communications. Transformation codes represent; 1: first difference of the logs, 2: first difference of the logs after seasonal adjustment, and 3: first difference.

4. Empirical Results

We compare the out-of-sample forecasting performances of the models discussed in Section 2 with the AR model for $h = 1$, 3, and 12 months ahead. The variables to be forecasted are the representative macroeconomic indicators selected according to Shibamoto (2007), and performances are evaluated based on MSFEs in the third subperiod (2013:07-2018:06). All forecasts are estimated directly using a rolling window scheme because the direct forecasting is robust to the model misspecification (Marcellino et al., 2006) and the rolling window is to structural changes by dropping the earlier observations which may follow unrelated data generating process (Clark and McCracken, 2009). Concretely, the first window uses the data between 2008:08-h and 2013:07-h to forecast 2013:07, and the second window uses the data between 2008:09-h and 2013:08-h to forecast 2013:08. This rolling window scheme is repeated until we get all the forecasts and prediction errors for the third subperiod. In the VAR models and two of our proposed diffusion index models, tuning parameters and the number of factors are selected with respect to the MSFEs of the second subperiod. In the diffusion index models (Eq. 12) and the AR model, the maximum lag orders are set to twelve out of which BIC selects the optimum lag order, whereas $P = 4$ for the VAR models[4].

4.1. Predictive Accuracy

Table 2 reports the out-of-sample mean squared forecast error of the aforementioned models. The AR column reports the MSFEs in the absolute term and the rest in the relative term when the MSFE of the AR model for the corresponding row is set to one so that the values of less than one mean the improvement in forecasting performances. Overall, the high-dimensional models outperform the benchmark model with a similar degree of improvement except for the diffusion index model with r selected by Bai and Ng's (2002) information criterion (DI). A closer look into the selected number of factors in DI reveals that the information criterion selects 17-19 factors for each window, which makes the model susceptible to the overfitting, especially when the number of observation in a window is small ($T = 60$). This, as discussed in Section 2, is likely to be caused by the time-independent restriction on the factor loadings and is consistent to Hayakawa and Kobayashi (2011), which reported that the Japanese macroeconomy has twenty factors when Bai and Ng's (2002) criterion is used. When the factors is estimated in terms of the predictive power in our proposed model (DICV and DILASSO), most cases select less than five factors, and more than ten factors are selected in only a few cases. It is also shown that the diffusion index approaches perform slightly better than VAR-based approaches, indicating the possibility that including all the available information is helpful in predicting a short term variation.

The MSFEs for one quarter horizon ($h = 3$) and one year horizon ($h = 12$) are reported in Table 3 and Table 4. In general, the relative performances are better in longer horizons with 19 out of 20 variables outperforming the AR model for $h = 12$ except for DICV model which shows the MSFEs smaller than the benchmark in 17 variables and DI model which continues to perform poorly. The better performances in longer horizon indicate that those high-dimensional models have more benefit when forecasting the longer horizon, which is consistent to earlier studies such as Callot and Kock (2014) for the VAR and DI in the United States and Shintani (2005) for the DI in Japan. At 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 may be favorable for a longer horizon.

---

[4] Unreported analysis for the lasso with $P = 12$ shows the similar result to the one with $P = 4$. We use shorter lag for the computational efficiency and data availability.

	AR(abs.)	DI	DICV	DILASSO	LASSO	ENET	gLASSO
Call Rate	0.295	1.042	**0.788**	**0.643**	**0.619**	**0.619**	**0.668**
Industrial Production	0.443	1.46	1.118	**0.907**	**0.907**	**0.969**	**0.907**
CPI	3.12	**0.823**	**0.656**	**0.595**	**0.643**	**0.594**	**0.691**
3 months TIBOR	0.412	**0.544**	**0.412**	**0.421**	**0.404**	**0.404**	**0.398**
10 years Government Bond	0.459	1.074	**0.788**	**0.737**	**0.741**	**0.741**	**0.741**
Monetary Base	0.303	1.738	1.205	1.109	1.109	1.109	1.108
Money Supply: M2*	0.797	1.008	**0.807**	**0.792**	**0.793**	**0.821**	**0.78**
Exchange Rate	1.246	1.188	**0.753**	**0.767**	**0.699**	**0.712**	**0.714**
Nikkei 42*	0.508	1.432	**0.924**	1.063	1.209	1.216	1.139
Capacity Utilization	0.334	1.786	1.231	**0.925**	**0.927**	**0.927**	**0.927**
Consumption	1.16	1.892	1.052	**0.99**	**0.99**	**0.99**	**0.99**
Spread	0.048	**0.092**	**0.081**	**0.382**	1.17	1.143	6.501
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**
Wage Index	0.585	**0.965**	**0.871**	1.044	1.312	1.255	1.27
Worked Hours	0.61	**0.858**	**0.644**	**0.687**	**0.658**	**0.657**	**0.689**
Employment Index	0.925	1.338	**0.975**	**0.999**	**0.989**	**0.978**	1.044
Disposable Income	0.503	1.653	**0.899**	1.049	1.282	1.279	1.376
New Job Offers*	0.549	1.434	**0.983**	1.243	1.263	1.266	1.323
Numb. Outperform	—	5	16	14	13	13	12

Table 2: Out-of-Sample MSFE of twenty macroeconomic variables for *h=1*. DI represent a diffusion index model with Bai and Ng's (2002) criterion. DICV and DILASSO denote our proposed diffusion index models with *r* selected by *h*-step-ahead forecasting performance and sparse factor/lag structure, respectively. ENET and gLASSO represent the elastic net and group lasso. The column AR reports the absolute level of MSFEs and the rest reports the MSFEs compared to the AR model. The outperforming models ($MSFE < 1$) are in bold. The last row reports the number of cases to outperform the benchmark AR model (out of twenty). Asterisk denotes leading indicator.

Tables 3 and Table 4 have also revealed that the improvements of the forecasting performances vary between the variables. Across all the horizons, the extent of improvement is relatively small for the variables belonging to the leading indicators defined by the Cabinet Office[5]. These are the variables that move sensitively to the macroeconomic changes, including TOPIX, machinery orders and housing starts. The result suggests that incorporating a large set of information in forecasting leading indicators are limited, albeit not negligible, compared to forecasting other variables whose movements are less sensitive to the macroeconomic environment.

[5] www.esri.cao.go.jp/en/stat/di/di-e.html

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

Table 3: Out-of-Sample MSFE of twenty macroeconomic variables for h=3. The AR column reports the absolute level of MSFEs and the rest reports the MSFEs relative to the AR model. The outperforming models (MSFE < 1) are marked bold. The last row reports the number of cases where the examined model outperformed the AR model.

4.2 Interpretations

Although the forecasting performances are more or less similar for the high-dimensional models except for the first diffusion index model (DI), the detailed investigations of how they reach the similar results reveal significant differences among the models. First, we investigate the differences of sparsities among the VAR-based techniques. Tables 5 reports the average number of parameters with a nonzero coefficient per window. We can find that the lasso selects more sparse solutions, partly because of the restriction that the number of nonzero parameters is limited to the number of window size, which is sixty in our empirical application. It is also found that the sparsity tends to be high for the longer horizons and leading indicators. This finding is plausible considering the diminishment of information for a longer horizon and the limited benefit of exploiting large data for predicting a leading indicator.

Next, we explore the relationship between the predictors and the target variable in the VAR models. Figure 1 is a visual representation of the variables selection in the lasso, elastic net and group lasso for CPI at horizon one. We can find that the gray grids, which represent

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

Table 4: Out-of-Sample MSFE of twenty macroeconomic variables for h=12. The column AR reports the absolute level of MSFEs and the rest reports the MSFEs relative to the AR model. The outperforming models (MSFE < 1) are marked bold. The last row reports the number of cases where the examined model outperformed the AR model.

parameters with non-zero coefficient, are thin and scattered in the lasso, indicating that the selections of the parameters are unstable compared to the elastic net and group lasso. The result is consistent with the discussion in Section 2 about the grouping feature of the elastic net and group lasso, which facilitates the model interpretation. An unreported investigation has also found similar patterns for the other variables and other horizons.

To further interpret the mechanism of the forecast, we look into the frequency of selecting the variable in terms of the categories of the variables. Table 6 shows the average ratio of selecting a particular group of variables in the scale of zero to one, with the focus on three variables that are economically important and that the AR performs poorly; CPI, industrial production and employment index. The value of one means that all the variables in that group are selected at every window and for every horizon. In forecasting CPI, the variables categorized as exchange rate and output and income rank among the top three most selected group in all three models, indicating that the information contained in those variables are important in forecasting the future CPI. As for the industrial production, no variables group are placed

	h=1			h=3			h=12		
	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO
Call Rate	0.5	0.55	144.883	0	5.183	12.067	0	20.467	0
IP	0	153.533	0	21.017	21.25	0	0	0	3.4
CPI	16.067	41.483	139.783	3.85	5.483	9.567	0	105.133	171.1
3 months TIBOR	0.55	0.617	8.5	2.8	32.7	10.867	0	37.55	4.1
10 y Govt Bond	0	0	0	0	0	0	0	0	1.433
Monetary Base	0	0	7.817	0	3.65	0	0.05	0.05	61.9
M2	0	156.45	64.883	0	159.15	1.917	0	200.983	2.867
Exchange Rate	4.733	95.467	52.8	0	0	0	0	31.75	5.55
Nikkei 42*	14.333	26.333	101.933	0	51.267	27.7	0	0	0
Capacity Utilization	0	56.867	0	0.583	0.583	0	0	0	3.533
Consumption	0	0	0	1.033	21.45	0	0	0	0
Spread	14.217	14.9	65.95	0.333	52.6	9.167	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
Wage Index	2.617	22.717	0	0	31.633	0	0	0	0
Worked Hours	8.533	8.983	94.85	27.417	28.45	142	16.983	16.467	59.483
Employment Index	3.683	9.333	68.617	0.917	0.933	9.9	3.833	115.65	64.467
Disposable Income	2.217	2.283	0	0	0	0	1.933	1.9	0
New Job Offers*	0.133	118.933	53.3	0	212.65	81.35	0.133	0.133	0.25
All zero: total	9	5	7	11	4	8	17	12	10

Table 5: Average number of parameters with a nonzero coefficient within a window. The last row reports the number of cases in which no parameter is selected at any windows. Asterisk denotes leading indicators.

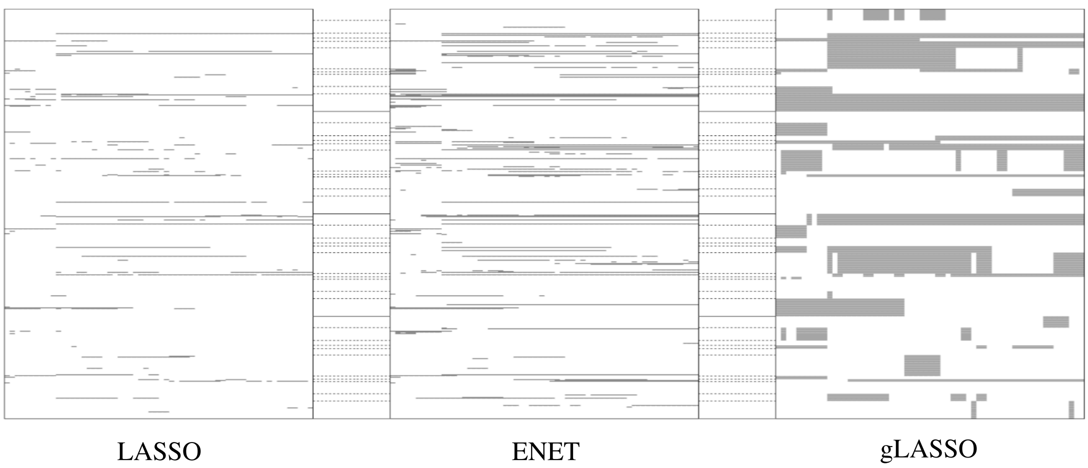

Figure 1: Visual representations of the variable selection for the lasso, elastic net and group lasso. The shadowed grids represent the variables with non-zero coefficients within the window, which corresponds to the horizontal axis. The vertical axis shows 508 potential variables (127 variables and their lags). The dotted and solid lines between the boxes are the boundaries of the variable groups and lags, respectively.

	CPI			Industrial Production			Employment Index		
	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO	LASSO	ENET	gLASSO
Consumption	0.012	0.03	0.124	**0.034**	**0.144**	0	**0.027**	**0.213**	**0.232**
Employment & Hours	0.003	**0.128**	0.156	0.01	0.091	0	0.004	0.05	0.029
Exchange Rates	**0.027**	**0.235**	**0.254**	0.003	0.05	0	0	0.077	**0.133**
Housing Starts & Sales	0.01	0.101	0.243	0.001	0.053	**0.019**	0.005	0.072	0.126
Interest Rates	0.001	0.046	0.021	0.01	0.087	0	0.002	0.057	0.039
Inventories & Orders	0.015	0.089	0.231	**0.021**	**0.15**	0	**0.009**	**0.116**	0.131
Money & Credit Quantity Aggregate	0.006	0.093	0.14	0.016	**0.158**	0	0.001	0.027	0.043
Output & Income	**0.022**	**0.144**	**0.379**	0.006	0.123	0	**0.007**	0.086	**0.133**
Price Indices & Wages	0.021	0.108	**0.308**	**0.019**	0.11	0	0.006	**0.088**	0.096
Retail, Manufacturing & Trade Sales	0.013	0.055	0.11	0.01	0.132	0	0.001	0.076	0.043
Stock Prices	**0.023**	0.108	0.174	0.011	0.098	**0.032**	0	0.085	0.128

Table 6: Average ratio of selecting the variables in a group in the scale between zero and one. The value of one means that all the variables in that group are selected for every window at every horizon, while zero means no variable belonging to the group is selected. Three most selected groups are in bold.

unanimously within the top three most selected groups, possibly resulting in relatively small improvement. Regarding the employment index, the variables related to the consumption is shown to be important, along with inventories and orders and output and income, which rank within top-three in two models and fourth place in the other model.

Concerning the interpretations of the diffusion index models, we extend the approach of Stock and Watson (2002b), in which they investigated the R^2 of the regression of individual series onto each estimated factor, using the full sample period of their dataset. The use of the full sample is appropriate in the sense that they are interested in the economic interpretation of each factor, yet is not suitable when the interest lies in the contribution of the individual series to the predictive power. Instead, we look into the R^2 statistic when the individual series are regressed on a set of factors in a window. Figure 2 reports the R^2 for CPI estimated using the factors of the proposed diffusion index models and averaged over window and horizon. The R^2 statistic close to one means that the estimated factors are relevant to that specific variable and the information contained in that specific variable is essential in forecasting the target variable because the factors are selected in terms of the predictive power in those two approaches. We can find some clusters of variables with similar degrees of R^2, which indicates the importance of those groups of variables in forecasting CPI. Table 7 describes the R^2 statistics averaged by the categories of variables for the variables selected in Table 6. Consumption and output and income are shown to be two of the most important groups for forecasting the employment index, which are also the groups most selected by the VAR models. As for CPI and industrial

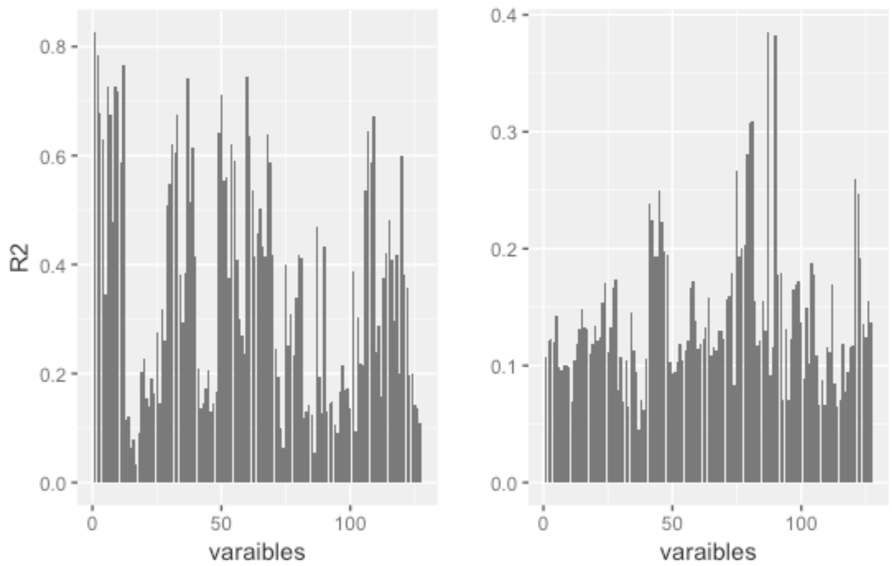

Figure2:R^2 of the regression of r estimated factors on the individual series, averaged over the horizon. The x-axis corresponds to 127 variables in the original dataset. Note that the scale of the y-axes are different.

	CPI		Industrial		Employment Index	
	DICV	DILASS	DICV	DILASS	DICV	DILASS
Consumption	**0.571**	0.071	**0.564**	**0.001**	**0.512**	**0.108**
Employment & Hours	0.213	0.132	0.249	0	0.189	0.07
Exchange Rates	0.39	**0.3**	0.337	0	0.278	0.042
Housing Starts & Sales	0.164	**0.214**	0.265	**0.001**	0.188	**0.118**
Interest Rates	0.172	0.16	0.206	0	0.158	0.091
Inventories & Orders	0.448	0.126	0.398	0	0.298	0.097
Money & Credit Quantity Aggregate	0.212	0.15	0.252	**0.001**	0.192	0.075
Output & Income	**0.584**	0.109	**0.585**	0	**0.512**	**0.127**
Price Indices & Wages	0.357	0.124	0.353	0	0.272	0.082
Retail, Manufactu-ring & Trade Sales	**0.493**	0.099	**0.487**	0	**0.417**	0.088
Stock Prices	0.298	**0.216**	0.284	0	0.232	0.044

Table 7: The R^2 of the regression of individual series on the estimated factors, averaged across categories. The value of one means that the variances within the group are fully explained by the estimated factors for every window at every horizon. The three highest groups are marked bold.

production, although the results are not as clear as the one for the employment index, the groups with high R^2 tend to agree with the most selected groups of the VAR models, indicating that both VAR and diffusion index models exploit the information from the same groups of variables.

5. Conclusion

In this thesis, we have studied the forecasting performances and interpretability of various high-dimensional models using 127 monthly Japanese macroeconomic data. We have employed the lasso, elastic net, and group lasso within the VAR framework and three diffusion index models that are different in how they select the common factors and lags. Overall, our results have shown that using high-dimensional data indeed improves the accuracy of forecasting the Japanese macroeconomic indicators except when the information criterion is used to select the number of factors in the diffusion index model. The contradicting performances within the diffusion index models is indicative of the importance of selecting factors with respect to the predictive power, instead of the relevance to the original dataset. The results have also shown that the forecasting performances are better in longer horizons and for the variables not belonging to the leading indicators, and although not conclusive, the comparison between the two frame-works suggest the advantage of the diffusion index models in the shorter horizon and the lasso-based approach in longer horizon.

In addition to the improvement of the forecasting performances, we have provided the R^2-based approach as an extension of Stock and Watson's (2002b) to interpret the common factors and explain the predictive power contained in the original dataset. The empirical results have shown that the variables with high R^2 are also the ones likely to be selected in the lasso, elastic net, and group lasso, indicating that our proposed method of interpreting the predictive power of the original dataset is valid and that both VAR- and DI-based methods exploit the information from the same groups of variables.

As a final remark, several limitations are noted. The field of high-dimensional/big data analysis is a rapidly growing area of research, and many methodologies are proposed that are not covered in this thesis. One of the interesting extensions of this research would be to incorporate the nonlinearity, which allows more flexibility to the model and is one of the fields with growing attention (for example, Exterkate et al., 2016). It is also known that when models are combined, even the combination of weaker models can often outperform the best performing model (Li and Chen, 2014; Swanson and Xiong, 2018). The combination of the models in this thesis may contribute to the further improvement of forecasting performances. Another direction of the extension would be to forecast the distribution. It is often the case that the econometrician's interest lies in the extreme cases such as a risk of economic crisis. Rossi and Sekhposyan (2014) is a leading case in which they focused on the predictive density of the US output and inflation, yet we are not aware of the application to Japan. The peculiarity of the sample period are also accounted for. The dataset in this research spans approximately 15 years and 183 observations, which is relatively small compared to similar studies. Besides, the second subperiod (2008:07-2013:06) matches the period of the global financial crisis, during when the economic structure may have been different. Incorporating the structural breaks (Stock and Watson, 2009) may lead to further improvement in the forecasting accuracy.

Acknowledgment

I would like to express my sincere gratitude to my advisor Associate Professor Naoya Sueishi for the continuous support and for showing me the intriguing world of econometrics. The dataset for the analysis is created with the help of Professor Kazuhiko Hayakawa. I also thank my family and my late father for always believing in me with unparalleled love.

References

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

Artis, M. J., 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.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71-92.

Banerjee, A., Masten, I., & Massimiliano, M. (2008). *Forecasting macroeconomic variables using diffusion indexes in short samples with structural change*. EUI Working Papers (ECO 2008/17). European University Institute.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, *120*(1), 387-422.

Boivin, J., & Giannoni, M. (2006). *DSGE models in a data-rich environment* (No. w12772). National Bureau of Economic Research.

Burns, A. F., & Mitchell, W. C. (1947). *Measuring business cycles*. National Bureau of Economic Research

Callot, L. A., & Kock, A. B. (2014). Oracle efficient estimation and forecasting with the adaptive lasso and the adaptive group lasso in vector autoregressions. *Essays in Nonlinear Time Series Econometrics*, 238-268.

Chamberlain, G., & Rothschild, M. (1982). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica,* 51(5), 1281-1304

Clark, T. E., & McCracken, M. W. (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, *50*(2), 363-395.

Connor, G., & Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, *15*(3), 373-394.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057-1072.

Engle, R., & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, *76*(376), 774-781.

Exterkate, P., Groenen, P. J., 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.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent Variables in Socio-Economic Models*. in Aigner, D. J., & Goldberger, A. S. (Eds.), Amsterdam: North Holland.

Hayakawa, K., & Kobayashi, Y. (2011). Daikibo Macro Data wo Mochi-ita Keiki Doukou Bunseki [Business Cycle Analysis Using a Large-Scale Macro Data]. In Asako, K., & Watanabe, T. (Eds.), *Finance Keiki-Junkan no Keiryo Bunseki* [Quantitative Analysis of Finance and Business Cycle] (pp. 91-111). Tokyo: Minerva Shobo. In Japanese.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Kowarik, A., Meraner, A., Templ, M., & Schopfhauser, D. (2014). Seasonal Adjustment with the R Packages x12 and x12GUI. *Journal of Statistical Software*, *62*(1), 1-21.

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

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

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, *22*(12), 5027-5067.

Mallows, C. L. (1973). Some comments on Cp. *Technometrics*, *15*(4), 661-675.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1-2), 499-526.

Nicholson, W. B., Matteson, D. S., & 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 Science & Business Media.

Phillips, P. C., & 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*. Vienna. Austria. URL: https://www.R-project.org/.

Sargent, T. J., & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, *1*, 145-168.

Shibamoto, M. (2007). An analysis of monetary policy shocks in Japan: a factor augmented vector autoregressive approach. *The 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*, 517-538.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.

Song, S., & Bickel, P. J. (2011). Large vector auto regressions. *arXiv preprint arXiv:1106.3915*.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, *15*(4), 101-115.

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

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

Stock, J. H., & Watson, M. W. (2005). Implications of Dynamic Factor Models for VAR Analysis (No. w11467). National Bureau of Economic Research.

Stock, J. H., & Watson, M.W. (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability Ch. 7. In Shephard, N., & Castle, J. (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*. Oxford; Oxford University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 267-288.

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.

Watson, M. W., & Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, *23*(3), 385-400.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301-320.