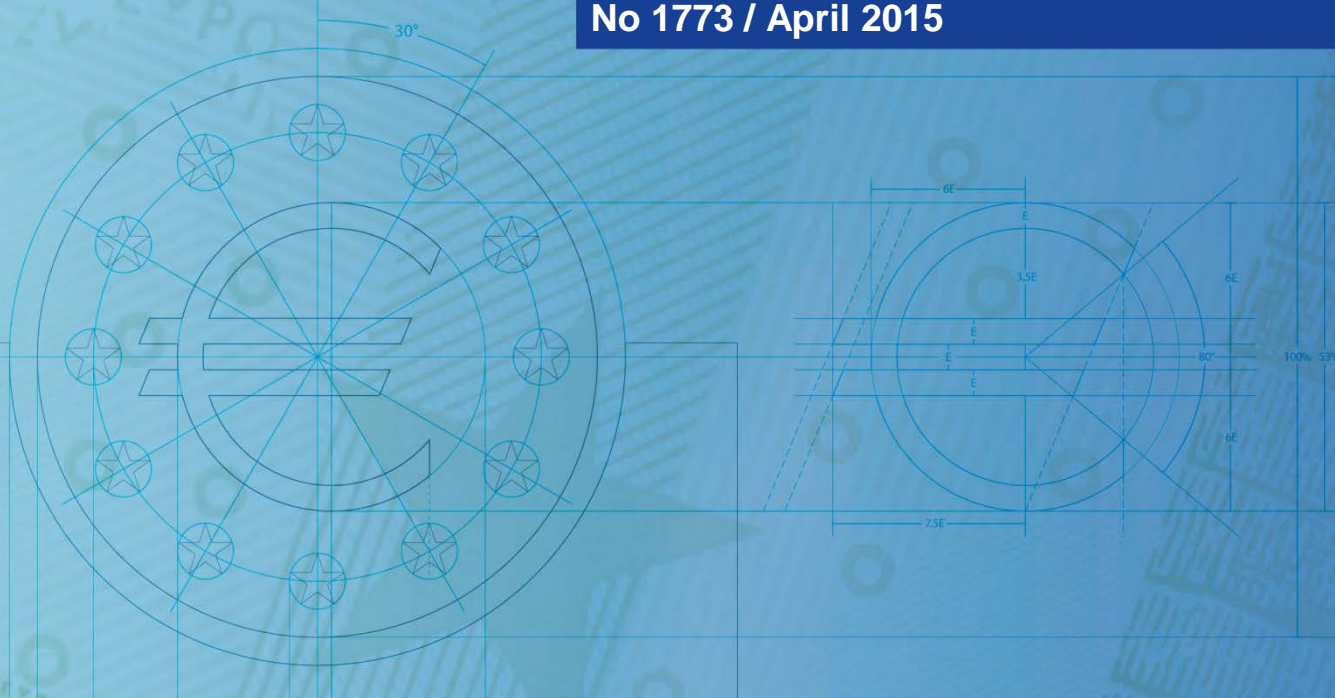# Working Paper Series

Gonzalo Camba-Mendez,
George Kapetanios,
Fotis Papailias,
and Martin R. Weale

An automatic leading indicator, variable reduction and variable selection methods using small and large datasets:

Forecasting the industrial production growth for euro area economies

Gonzalo Camba-Mendez
European Central Bank; e-mail: gonzalo.camba-mendez@ecb.europa.eu

George Kapetanios
Queen Mary, University of London; e-mail: g.kapetanios@qmul.ac.uk

Fotis Papailias
Queen's University Management School; e-mail: f.papailias@qub.ac.uk

Martin R. Weale
Queen Mary, University of London; e-mail: m.weale@qmul.ac.uk

Abstract

This paper assesses the forecasting performance of various variable reduction and variable selection methods. A small and a large set of wisely chosen variables are used in forecasting the industrial production growth for four Euro Area economies. The results indicate that the Automatic Leading Indicator (ALI) model performs well compared to other variable reduction methods in small datasets. However, Partial Least Squares and variable selection using heuristic optimisations of information criteria along with the ALI could be used in model averaging methodologies.

Executive Summary

The issue of forecasting key macroeconomic variables has been constantly on the debate over the past years. In general, methodologies can be divided in two broad categories characterised by the size of the set of explanatory variables taken into consideration. On the one hand, some researchers pursue an strategy where a small number of wisely selected predictors is used (variable selection method). Then, under the assumption of non cross-correlated errors the factor models are estimated by maximum likelihood using the Kalman filter. On the other hand, others suggest forecasting using a large set of predictors where information is summarised using principal components estimation (variable reduction method). The forecasting performance of other variable reduction methods, e.g. Bayesian shrinkage regression or partial least squares (PLS), have also been recently studied.

A difficult issue in both the above categories is the choice of the set of predictors and the actual variables. Recent research suggests that cross-correlation of regressors in large datasets might result in inaccurate forecasts and hence a smaller set is more likely to provide a smaller average forecast error. A key to this problem is the use of variable selection methods.

The purpose of this paper is to evaluate and compare the forecasting ability of (i) variable reduction and (ii) variable selection methods using small and large datasets. We predict the growth rate of the industrial production of the Euro Area (16), France, Germany and Italy. The variable reduction methodologies include: the Bayesian shrinkage regression and the partial least squares. The variable selection methodology employs the following heuristic optimisations: the simulated annealing (SA), the genetic algorithm (GA) and the MC^3 algorithm. Each of these optimisations is used to identify the combination of variables that minimise an information criterion (e.g. BIC, HQ). The benchmark is the Automatic Leading Indicator (ALI) model in the spirit of earlier work by the authors which belongs to the category of small scale dynamic factor models.

Our overall work indicates that among the variable reduction methods using a small set of predictors, the ALI is more likely to provide better forecasts on average. However, the ALI performs well only in the "medium-term" forecasts (4-6 steps ahead) using a large dataset of predictors. The variable reduction methodologies, and especially the PLS model with 1 and 3 factors, are more likely to result in forecasts with smaller errors in the first and 8 to 12 steps ahead. Similarly, the variable selection methods can also be useful for "short-run" (one step ahead) and "long-run" (10-12 steps ahead) predictions. Our results suggest the use of SA and GA as the more robust optimisation methodologies and the BIC as the objective function in the minimisation.

To conclude with, our empirical approach suggests that the models described here should, at least, be considered by researchers interested in model averaging methodologies for forecasting key European macroeconomic variables.

1   Introduction

The issue of forecasting key macroeconomic variables has been constantly on the debate over the past years. In general, methodologies can be divided in two broad categories characterised by the size of the set of explanatory variables taken into consideration. Stock and Watson (1991), Camba-Mendez et al. (2001), Aruoba et al.. (2009), Aruoba and Diebold (2010) and Camacho and Perez-Quiros (2010) are some examples where a small number of wisely selected predictors is used. Then, under the assumption of non cross-correlated errors the factor models are estimated by maximum likelihood using the Kalman filter.

On the other hand, the seminal work by Stock and Watson (2002a) suggests forecasting using a large set of predictors where information is summarised using principal components estimation. The forecasting performance of other variable reduction methods have been recently studied by Del Mol et al. (2006), Kapetanios, Marcellino and Papailias (2012a) among others and include the Bayesian shrinkage regression, partial least squares and others.

A difficult issue in both the above categories is the choice of the set of predictors and the actual variables. Boivin and Ng (2006) suggest that cross-correlation of regressors in large datasets might result in inaccurate forecasts and hence a smaller set is more likely to provide a smaller average forecast error. A key to this problem is the use of variable selection methods. Kapetanios (2007) and Kapetanios, Marcellino and Papailias (2012a) use non-standard optimisation of information criteria in order to identify the appropriate instruments and they forecast EU macroeconomic variables with encouraging results. A recent study by Alvarez et al. (2012) also addresses this issue and compares small and large scale dynamic factor models in the Stock and Watson dataset for the US. However, their approach is limited in that particular class of models.

The purpose of this paper is to evaluate and compare the forecasting ability of (i) variable reduction and (ii) variable selection methods using small and large datasets. We predict the growth rate of the industrial production of the Euro Area (16), France, Germany and Italy. The variable reduction methodologies include: the Bayesian shrinkage regression and the partial least squares. The variable selection methodology employs the following heuristic optimisations: the simulated annealing (SA), the genetic algorithm (GA) and the MC^3 algorithm. Each of these optimisations is used to identify the combination of variables that minimise an information criterion (e.g. BIC, HQ). The benchmark is the Automatic Leading Indicator (ALI) model in the spirit of Camba-Mendez et al. (2001) which belongs to the category of small scale dynamic factor models.

The rest of the paper is organised as follows: Section 2 briefly describes the methodologies, Section 3 is concerned with the forecasting algorithm, Section 4 discusses the results and Section 5 summarises the conclusions.

2 Methodologies

2.1 Variable Selection Methods

We consider the following regression model,

$$y_t = \alpha + \beta^{0\prime} x_t^0 + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where x_t^0 is a k-dimensional vector of stationary predetermined variables. The superscript ^{0} denotes the true regression model. Let the set of all available variables at time t be represented by the N-dimensional vector $x_t = (x_{1,t}, \ldots x_{N,t})'$, where it is currently assumed that the set of variables in x_t^0 is also contained in x_t. The aim of the analysis is to determine x_t^0. Formally, let $\boldsymbol{\mathcal{I}} = (\boldsymbol{\mathcal{I}}_1, \ldots, \boldsymbol{\mathcal{I}}_N)'$ denote a vector of zeros and ones (which we will refer to as string). Let $\boldsymbol{\mathcal{I}}^0$ be the string for which $\boldsymbol{\mathcal{I}}_i^0 = 1$, if $x_{i,t}$ is an element of x_t^0 and zero otherwise. We wish to estimate $\boldsymbol{\mathcal{I}}^0$. Note that in small samples $\boldsymbol{\mathcal{I}}^0$ may not represent the best fitting model for the data at hand.

To do this we consider the use of information criteria to select the variables that go in (1). The generic form of such criteria is usually,

$$IC(\boldsymbol{\mathcal{I}}) = -2L(\boldsymbol{\mathcal{I}}) + C_T(\boldsymbol{\mathcal{I}}), \tag{2}$$

where $L(\boldsymbol{\mathcal{I}})$ is the log-likelihood of the model associated with string $\boldsymbol{\mathcal{I}}$ and $C_T(\boldsymbol{\mathcal{I}})$ is the penalty term associated with the string $\boldsymbol{\mathcal{I}}$. The three most usual penalty terms are $2\tilde{m}(\boldsymbol{\mathcal{I}})$, $ln(T)\tilde{m}(\boldsymbol{\mathcal{I}})$ and $2ln(ln(T))\tilde{m}(\boldsymbol{\mathcal{I}})$ associated with the Akaike (AIC), Bayesian (Schwarz (1978)) (BIC) and Hannan-Quinn (Hanna and Quinn (1979)) (HQ) information criteria. $\tilde{m}(\boldsymbol{\mathcal{I}})$ is the number of free parameters associated with the modelling of the dataset associated with $\boldsymbol{\mathcal{I}}$. Note that, in this case, $\tilde{m}(\boldsymbol{\mathcal{I}}) = \boldsymbol{\mathcal{I}}'\boldsymbol{\mathcal{I}}$. It is straightforward under relatively weak conditions on $x_{j,t}$ and $\epsilon_{j,t}$, and using the results of say, Sin (1996), to show that the string which minimises $IC(.)$ will converge to $\boldsymbol{\mathcal{I}}^0$ with probability approaching one as $T \to \infty$ as long as (i) $C_T(\boldsymbol{\mathcal{I}}) \to \infty$ and (ii) $C_T(\boldsymbol{\mathcal{I}})/T \to 0$.

More specifically, the assumptions needed for the results of Sin (1996) to hold are mild and can be summarised as follows, assuming estimation of the models is undertaken in the context of Gaussian or pseudo maximum likelihood (which in the simplest case, of spherical errors, is equivalent to OLS): (i) Assumption A of Sin (1996) requires measurability, continuity and twice differentiability of the log-likelihood function and a standard identifiability assumption; (ii) A uniform weak law of large numbers for the log-likelihood of each observation and its second derivative; (iii) A central limit theorem for the first derivative of the log-likelihood of each observation. (ii) and (iii) above can be obtained by assuming, e.g., that $x_{j,t}$ are weakly dependent, say, near epoch dependent, processes and $\epsilon_{j,t}$ are martingale difference processes. Hence, it is clear that consistency of model selection as long as the penalty related conditions

hold is straightforwardly obtained. Note that unlike BIC and HQ which consistently estimate the true model in the sense of Sin (1996), AIC is inconsistent, in this sense, since C_T remains bounded, as $T \rightarrow \infty$, contravening the first penalty related condition given in the preceding paragraph. This is the reason why AIC is not included in our calculations. Also, the empirical results in Kapetanios, Marcellino and Papailias (2012a) also suggest the use of BIC or HQ as objective functions. Further, note that in most work dealing with variable selection and information criteria, stationarity of x_t^0 is usually assumed, as is in the preceding analysis, although the analysis may be extended to nonstationary variables.

The problem is of course how to minimise the information criterion. For small dimensional x_t, evaluating the information criterion for all strings may be feasible, as, e.g., in lag order selection. In the case of lag selection the problem is made easier by the fact that there exists a natural ordering of the variables, although in many cases such an ordering may not be the optimal basis for a search algorithm. In the general variable selection case, as soon as N exceeds say 50 or 60 units, this strategy is bound to fail. Since $\mathcal{I}$ is a binary sequence there exist 2^N strings to be evaluated. For example, when $N = 50$ and optimistically assuming that 100000 strings can be evaluated per second, we still need about 357 years for an evaluation of all strings. Clearly this is infeasible.

Although this is a minimisation problem, standard minimisation algorithms do not apply due to the discreteness of the domain over which the objective function (information criterion) needs to be optimised. To overcome this difficulty we use the following heuristic optimisation approaches that include: the simulated annealing, the genetic algorithm and the MC^3.

Since in our case N is very large, N=195 (see Table 1), we should normally compute and compare 2195 IC according to eq.(2). This is clearly not feasible and the alternative approaches analysed below could help the researcher to reduce the computational burden.

2.1.1   Simulated Annealing (SA)

This algorithm provides a local search for the minimum (or maximum) of a function, in our case is eq.(2). The concept is originally based on the manner in which liquids freeze or metals recrystalize in the process of annealing. In an annealing process a melt, initially at high temperature and disordered, is slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered and approaches a 'frozen' ground state. The analogy to an optimisation problem is as follows: the current state of the thermodynamic system is analogous to the current solution to the optimisation problem, the energy equation for the thermodynamic system is analogous to the objective function, and the ground state is analogous to the global optimum. An early application of simulated annealing in econometrics is the work of Goffe et al. (1994) who suggested that simulated annealing could be used to optimise the objective function of various econometric estimators.

Below, we give a description of the algorithm together with the necessary arguments that illustrate its validity in our context. We describe the operation of the algorithm when the domain of the function (information criterion) is the set of binary strings i.e. $\{\boldsymbol{\mathcal{I}} = (\mathcal{I}_1, \ldots, \mathcal{I}_N)' | \mathcal{I}_i \in \{0, 1\}\}$.

Each step of the algorithm works as follows starting from an initial string $\boldsymbol{\mathcal{I}}_0$.

1. Using $\boldsymbol{\mathcal{I}}_i$ choose a neighboring string at random, denoted $\boldsymbol{\mathcal{I}}_{i+1}^*$. We discuss the definition of a neighborhood below.

2. If $IC(\boldsymbol{\mathcal{I}}_i) > IC(\boldsymbol{\mathcal{I}}_{i+1}^*)$, set $\boldsymbol{\mathcal{I}}_{i+1} = \boldsymbol{\mathcal{I}}_{i+1}^*$. Else, set $\boldsymbol{\mathcal{I}}_{i+1} = \boldsymbol{\mathcal{I}}_{i+1}^*$ with probability $e^{(IC(\boldsymbol{\mathcal{I}}_i^*) - IC(\boldsymbol{\mathcal{I}}_{i+1}))/T_i}$ or set $\boldsymbol{\mathcal{I}}_{i+1} = \boldsymbol{\mathcal{I}}_i$ with probability $1 - e^{(IC(\boldsymbol{\mathcal{I}}_i^*) - IC(\boldsymbol{\mathcal{I}}_{i+1}))/T_i}$.

Heuristically, the term T_i gets smaller making it more difficult, as the algorithm proceeds, to choose a point that does not decrease $IC(.)$. The issue of the neighborhood is extremely relevant. What is the neighborhood? Intuitively, the neighborhood could be the set of strings that differ from the current string by one element of the string. But this may be too restrictive. We can allow the algorithm to choose at random, up to some maximum integer (say h), the number of string elements at which the string at steps i and $i+1$ will differ. So the neighborhood is all strings with up to h different bits from the current string. Another issue is when to stop the algorithm. There are a number of alternatives in the literature. We have chosen to stop the algorithm if it has not visited a string with lower $IC(.)$ than the current minimum for a prespecified number of steps (B_v) (Steps which stay at the same string do not count) or if the number of overall steps exceeds some other prespecified number (B_s). All strings visited by the algorithm are stored and the best is chosen at the end rather than the final one.

The simulated annealing algorithm has been proven by Hajek (1998) to converge asymptotically, i.e. as $i \to \infty$, to the maximum of the function as long as $T_i = T_0/ln(i+1)$ for some T_0 for sufficiently large T_0. In particular, for almost sure convergence to the minimum it is required that $T_0 > d^*$. d^* denotes the maximum depth of all local minima of the function $IC(.)$. Heuristically, the depth of a local minimum, $\boldsymbol{\mathcal{I}}_1$, is defined as the smallest number $E > 0$ such that the function exceeds $IC(\boldsymbol{\mathcal{I}}_1) + E$ during its trajectory from[*] this minimum to any other local minimum, $\boldsymbol{\mathcal{I}}_2$, for which $IC(\boldsymbol{\mathcal{I}}_1) > IC(\boldsymbol{\mathcal{I}}_2)$.

This condition needs to be made specific for the problem at hand. We thus need to discuss possible strategies for determining d^* for model searches using information criteria. It is reasonable to assume that the space of models searched via information criteria includes only models with a prespecified maximum number of variables; otherwise problems caused by the lack of degrees of freedom will arise. Then, a possible upper limit for d^* is $2L(\boldsymbol{\mathcal{I}_B}) - 2L(\boldsymbol{\mathcal{I}_A})$ where $L(\boldsymbol{\mathcal{I}_A})$ is the likelihood associated with a regression containing just a constant term

---

[*]A trajectory from $\boldsymbol{\mathcal{I}}_1$ to $\boldsymbol{\mathcal{I}}_2$ is a set of strings, $\boldsymbol{\mathcal{I}}_{11}, \boldsymbol{\mathcal{I}}_{12}, \ldots, \boldsymbol{\mathcal{I}}_{1p}$, such that (i) $\boldsymbol{\mathcal{I}}_{11} \in N(\boldsymbol{\mathcal{I}}_1)$, (ii) $\boldsymbol{\mathcal{I}}_{1p} \in N(\boldsymbol{\mathcal{I}}_2)$ and (iii) $\boldsymbol{\mathcal{I}}_{1i+1} \in N(\boldsymbol{\mathcal{I}}_{1i})$ for all $i = 1, \ldots, p$, where $N(\boldsymbol{\mathcal{I}})$ denotes the set of strings that make up the neighborhood of $\boldsymbol{\mathcal{I}}$.

and $L(\mathcal{I}_B)$ is the likelihood associated with a regression containing the maximum allowable number of variables. Of course, there are many possible sets of variables that contain the maximum allowable number of variables. For this reason we remove the penalty terms and focus on likelihoods. This makes it more likely that $-2L(\mathcal{I}_B)$, for some random $\mathcal{I}_B$ that specifies use of the maximum allowable number of variables, is a lower bound for the optimum value taken by the information criterion.

2.1.2 Genetic Algorithm (GA)

The motivating idea of genetic algorithms is to start with a population of binary strings which then evolve and recombine to produce new populations with 'better' characteristics, i.e. lower values for the information criterion. We start with an initial population represented by a $N \times m$ matrix made up of 0's and 1's. Columns represent strings. m is the chosen size of the population. The theory of genetic algorithms suggests that the composition of the initial population does not matter. Hence, this is generated randomly. Denote this population matrix by P_0. The genetic algorithm involves defining a transition from $\mathbf{P}_i$ to $\mathbf{P}_{i+1}$. Following Kapetanios (2007), the algorithm could be described in the following steps:

1. For $\mathbf{P}_i$ create a $m \times 1$ 'fitness' vector, $\mathbf{p}_i$, by calculating for each column of $\mathbf{P}_i$ its 'fitness'. The choice of the 'fitness' function is completely open and depends on the problem. For our purposes it is the opposite of the information criterion. Normalise $\mathbf{p}_i$, such that its elements lie in $(0,1)$ and add up to 1. Denote this vector by $\mathbf{p}_i^*$. Treat $\mathbf{p}_i^*$ as a vector of probabilities and resample m times out of $\mathbf{P}_i$ with replacement, using the vector $\mathbf{p}_i^*$ as the probabilities with which each string will be sampled. So 'fit' strings are more likely to be chosen. Denote the resampled population matrix by $\mathbf{P}_{i+1}^1$.

2. Perform "cross over" on $\mathbf{P}_{i+1}^1$. To "cross over" we do the following: Arrange all strings in $\mathbf{P}_{i+1}^1$, in pairs (assume that m is even) where the pairings are randomly drawn. Denote a generic pair by $(a_1^\alpha, a_2^\alpha, \ldots, a_N^\alpha)$, $(a_1^\beta, a_2^\beta, \ldots, a_N^\beta)$. Choose a random integer between 2 and $N-1$. Denote this by j. Replace the pair by the following pair: $(a_1^\alpha, a_2^\alpha, \ldots, a_j^\alpha, a_{j+1}^\beta, \ldots, a_N^\beta)$, $(a_1^\beta, a_2^\beta, \ldots, a_j^\beta, a_{j+1}^\alpha, \ldots, a_N^\alpha)$. Perform cross over on each pair with probability p_c. Denote the new population by $\mathbf{P}_{i+1}^2$. Usually p_c is set to some number around 0.5-0.6.

3. Mutate on $\mathbf{P}_{i+1}^2$. This amounts to flipping the bits (0 or 1) of $\mathbf{P}_{i+1}^2$ with probability p_m. p_m is usually set to a small number, say 0.01. After mutation the resulting population is $\mathbf{P}_{i+1}$.

These steps are repeated a prespecified number of times (B_g). Each set of steps is referred to as generation in the genetic literature. If a string is to be chosen, this is the one with maximum fitness. For every generation we store the identity of the string with maximum

'fitness'. Further, this string is allowed to remain intact for that generation. So it gets chosen with probability one in step 1 of the algorithm and does not undergo either cross-over or mutation. At the end of the algorithm the string with the lowest information criterion value over all members of the populations and all generations is chosen. One can think of the transition from one string of maximum fitness to another as a Markov Chain. So this is a Markov Chain algorithm. In fact, the Markov Chain defined over all possible strings is time invariant but not irreducible, as at least the $m - 1$ least fit strings will never be picked. To see this, note that in any population there will be a string with more fitness than that of the $m-1$ worst strings. There has been considerable work on the theoretical properties of genetic algorithms. Hartl and Belew (1990) have shown that with probability approaching one, the population at the n-th generation will contain the global maximum as $n \to \infty$. Perhaps the most relevant result from that work is Theorem 4.1 of Hartl and Belew (1990). This theorem states that as long as (i) the sequence of the maximum fitnesses in the population across generations is monotonically increasing, and (ii) any point in the model space can be reached from any other point by means of mutation and cross-over in a finite number of steps, then the global maximum will be attained as $n \to \infty$. Both these conditions hold for the algorithm described above. The first condition holds by the requirement that the string with the maximum fitness is always kept intact in the population. The second condition holds, since any string of finite length can be obtained from another by cross-over and mutation with non-zero probability in a finite number of steps. For more details on the theory of genetic algorithms see also Morinaka et al. (2001).

2.1.3 MC^3

This algorithm is constructed in a manner similar to simulated annealing. This similarity is, in fact, the main reason why we consider Bayesian methods here. The MC^3 algorithm defines a search path in the model space just like the simulated annealing algorithm we considered in the previous section. As a result we refer to the setup of the previous section to minimise duplication of the exposition. The difference between SA and MC^3 is the criterion used to move from one string to the other at step i. Here, the Bayes factor for string (model) $i + 1$ versus string (model) i is used. This is denoted by $B_{i+1,i}$. The chain moves to the $i + 1$ string with probability $min(1, B_{i+1,i})$. This is again a Metropolis-Hastings type algorithm. The Bayes factor we use, following Fernandez et al. (2001), is given by,

$$
B_{i+1,i} = \left(\frac{g_{0i+1}}{g_{0i+1} + 1} \right)^{k_{i+1}/2} \left(\frac{g_{0i} + 1}{g_{0i}} \right)^{k_i/2} \tag{3}
$$
$$
\left(\frac{\frac{1}{g_{0i}+1} RSS_i + \frac{g_{0i}}{g_{0i}+1} TSS}{\frac{1}{g_{0i+1}+1} RSS_{i+1} + \frac{g_{0i+1}}{g_{0i+1}+1} TSS} \right)^{(T-1)/2},
$$

where RSS_i is the sum of squared residuals of the i-th model, TSS is the sum of the squared deviations from the mean for the dependent variable, k_i is the number of variables in model i and g_{0i} is a model specific constant relating to the prior relative precision. The results of Fernandez et al. (2001) suggest that, for consistent model selection, g_{0i} should be set to $1/T$. This is associated with prior 'a' in the terminology of subsection 4.2 of Fernandez et al. (2001). More details may be found in Fernandez et al. (2001). Our chosen model is the model that minimises the information criterion among all models visited by the MC^3 algorithm. Given the results of Appendix A.3 of Fernandez et al. (2001) concerning the asymptotic equivalence between consistent information criteria and the Bayes factor in (3), we find our approach justified.

2.2 Variable Reduction Methods

2.2.1 Partial Least Squares (PLS)

Partial least squares (PLS) is a relatively new method for estimating regression equations, introduced in order to facilitate the estimation of multiple regressions when there is a large, but finite, amount of regressors.[†] The basic idea is similar to principal component analysis in that factors or components, which are linear combinations of the original regression variables, are used, instead of the original variables, as regressors. A major difference between PC and PLS is that, whereas in PC regressions the factors are constructed taking into account only the values of the x_t variables, in PLS, the relationship between y_t and x_t is considered as well in constructing the factors. PLS regression does not seem to have been explicitly considered for data sets with a very large number of series, i.e., when N is assumed in the limit to converge to infinity.

There are a variety of definitions for PLS and accompanying specific PLS algorithms that inevitably have much in common. A conceptually powerful way of defining PLS is to note that the PLS factors are those linear combinations of x_t, denoted by Υx_t, that give maximum covariance between y_t and Υx_t while being orthogonal to each other. Of course, in analogy to PC factors, an identification assumption is needed, to construct PLS factors, in the usual form of a normalization.

A simple algorithm to construct k PLS factors is discussed among others, in detail, in Helland (1990). Assuming for simplicity that y_t and x_t have been normalized to have zero mean and x_t has been scaled to have unit variance, a simplified version of the algorithm is given below:

1. Set $u_t = y_t$ and $v_{i,t} = x_{i,t}$, $i = 1, ...N$. Set $j = 1$.

[†]Herman Wold and co-workers introduced PLS regression between 1975 and 1982, see, e.g., Wold (1982). Since then it has received much attention in a variety of disciplines, especially in chemometrics, outside of economics.

2. Determine the $N \times 1$ vector of indicator variable weights or loadings $w_j = (w_{1j} \cdots w_{Nj})'$ by computing individual covariances: $w_{ij} = Cov(u_t, \ v_{it}), \ i = 1, ..., N$. Construct the j-th PLS factor by taking the linear combination given by $w_j' v_t$ and denote this factor by $f_{j,t}$.

3. Regress u_t and $v_{i,t}, \ i = 1, ..., N$ on $f_{j,t}$. Denote the residuals of these regressions by $\tilde{u}_t$ and $\tilde{v}_{i,t}$ respectively.

4. If $j = k$ stop, else set $u_t = \tilde{u}_t$, $v_{i,t} = \tilde{v}_{i,t} \ i = 1, .., N$ and $j = j + 1$ and go to step 2.

This algorithm makes clear that PLS is computationally tractable for very large data sets. Once PLS factors are constructed y_t can be modelled or forecasted by regressing y_t on $f_{j,t} \ j = 1, ..., k$. Helland (1988) and Helland (1990) provide a general description of the partial least squares (PLS) regression problem. Helland (1988) shows that the estimates of the coefficients α in the regression of y_t on x_t, as in (1), obtained implicitly via PLS Algorithm and a regression of y_t on $f_{j,t} \ j = 1, ..., k$, are mathematically equivalent to,

$$\hat{\alpha}_{PLS} = V_k (V_k' X' X V_k)^{-1} V_k' X' y, \tag{4}$$

with $V_{k_1} = (X'y \quad X'XX'y \quad \cdots \quad (X'X)^{k-1}X'y)$, $X = (x_1 \cdots x_T)'$ and $y = (y_1 \cdots y_T)'$. Thus, (4) suggests that the PLS factors that result from the PLS Algorithm span the Krylov subspace generated by $X'X$ and $X'y$, resulting in valid approximations of the covariance between y_t and x_t.

2.2.2 Bayesian Shrinkage Regression (BR)

Bayesian regression is a standard tool for providing inference for α in (1) and there exists a large variety of approaches for implementing Bayesian regression. We will provide a brief exposition of this method. A starting point is the specification of a prior distribution for α. Once this is in place standard Bayesian analysis proceeds by incorporating the likelihood from the observed data to obtain a posterior distribution for α which can then be used for a variety of inferential purposes, including, of course, forecasting.

A popular and simple implementation of Bayesian regression results in a shrinkage estimator for α in (1) given by,

$$\hat{\alpha}_{BRR} = (X'X + vI)^{-1}X'y, \tag{5}$$

where $X = (x_1, ..., x_T)'$, $y = (y_1, .., y_T)'$ and v is a shrinkage scalar parameter. The shrinkage estimator (5) shrinks the OLS estimator, given by $(X'X)^{-1}X'y$ towards zero, thus enabling a reduction in the variance of the resulting estimator. This is a major feature of Bayesian regression that makes it useful in forecasting when large data sets are available. This particular implementation of Bayesian regression implies that elements of α are small but different from zero ensuring that all variables in x_t are used for forecasting. In this sense, Bayesian

regression can be linked to other data-rich approaches. When a certain factor structure is assumed in the data, Bayesian regression through (5) will forecast y_t by projecting it on a weighted sum of all N principal components of X, with decaying weights, instead of projecting it on a limited number of r principal components with equal weights as in PC regression; see Del Mol et al. (2006).

2.2.3 Automatic Leading Indicator (ALI)

The last methodology is the automatic leading indicator (ALI) model as introduced by Camba-Mendez et al. (2001). This is a small scale dynamic factor model and includes a two-step procedure where in the first step the factors are extracted and in the second stage a VAR model is estimated and used for forecasting. Consider the following model for the N vector of exogenous variables X,

$$X_t = Bs_t + u_t,$$
$$C\left(L\right)s_t = \eta_t, \tag{6}$$

where B is an (n,k) matrix of unknown parameters, s_t is a k vector of factors that follows a stationary AR(p) process with disturbances η and u_t is an N vector of disturbances. The estimation of the unknown parameters in (6) and the extraction of factors may be combined in the following step-wise fashion. Given knowledge of B, $C(L)$ and the variance matrices of u_t and η_t, (6) can be written in state space form with the Kalman filter used to extract s_t from observation on x_t. Secondly, given the factors s_t, the parameter matrices B, $C(L)$ and the variance matrices of u_t and η_t may be estimated by quasi-maximum likelihood. This step-wise procedure may be iterated until convergence; see Harvey (1993) for further details.

The factors s_t obtained above are then incorporated into a VAR model to forecast y_t as follows,

$$A_y(L)y_t = As(L)s_t + \epsilon_t, \tag{7}$$

where ϵ_t is a zero-mean conditionally homoskedastic and serially uncorrelated error process with positive definite variance matrix uncorrelated with the error processes u_t and η_t. Then given the estimated factors and parameters the model in (7) is estimated via OLS (or maximum likelihood) and the parameter estimates are used in the forecasting exercise as described in the next section. The lags in the AR and VAR model and the choice of factors are determined by a series of tests that include serial correlation and Granger causality; see Camba-Mendez et al. (2001) for more details. It is important to notice here that the ALI is always used with the small set of predictors (see Table 3) as in Camba-Mendez et al. (2001).

2.3 Parameters Setup and Normalisation

For the simulated annealing and genetic algorithms we use the same (default) values as in Kapetanios (2007), i.e. in the simulated annealing $h = 1$, $B_v = 500$, $B_s = 5000$, $T_0 = 10$, in the genetic algorithm $m = 200$, $B_g = 200$, $p_c = 0.6$ and $p_m = 0.1$. We allow a maximum of five hundred iterations. The objective functions to be minimised are the Bayesian and HQ information criteria. In PLS we examine cases with one and three factors and in BR we use $v = 0.5N$ and $v = 2N$ as shrinkages. In ALI we use 1 factor and the initial values are fixed to the unconditional moments.

In all heuristics we have used the data as is, however in ALI, PLS and BR we have normalised the regressors to zero mean and unit variance series.

3 Forecasting and Data Description

We perform a forecasting exercise using the projection method as described in Stock and Watson (2002a). This method, also known as direct approach, is more robust in the presence of possible model mis-specification. The forecasts are given by,

$$\widehat{y}^f_{t+h/T} = \widehat{\beta}^{h\prime} z_{t/T}, \tag{8}$$

where $\widehat{\beta}^h$ is obtained by regressing y_t on z_{t-h} and h denotes the forecast horizon. z_t is a k-dimensional vector of variables and can be equal to x_t or to k-factors series depending on the choice of the estimation method. In the case of the ALI z_t might include the autoregressive components of y_t as well.

At first, we set the max $steps$ ahead, h. Then, we specify the evaluation period, $Eval$, and we omit h observations completely out of the sample. This allows us to end up with a number of $Eval$ forecasts for any given step h. A summary of the pseudo out-of-sample forecasting algorithm follows.

1. Use an initial sample of T_1 observations ($T_1 = T - Eval - h$),

2. With any method described in this section obtain x'_t, $t = 1, 2, ..., T_1$,

3. For $j = 1, 2, ..., h$ steps regress y_t on z'_{t-h} and obtain $\widehat{\beta}^h = \left(\widehat{\beta}^1, ..., \widehat{\beta}^h\right)'$,

4. Calculate the forecasts of $\widehat{y}^f_{t+h}$ using z'_t and $\widehat{\beta}^h$, hence $\widehat{y}^f = \left(\widehat{y}^f_1, ..., \widehat{y}^f_h\right)'$,

5. We repeat the whole procedure increasing the initial sample T_1 to $T_l = T_{l-1} + 1$ until $T_l = T - h$.

At the end of this process we have gathered a number of $Eval$ forecast values for any step h. We consider two looping procedures: (i) a recursive looping where the initial sample

T_1 is augmented by one observation at each time and (ii) a rolling approach where a fixed sample of T_1 length moves across time.

The forecast error is then calculated as,

$$\widehat{e}^f_{t+h} = y_{t+h} - \widehat{y}^f_{t+h},\tag{9}$$

and the statistics of interest can be computed. We are particularly interested in the Mean Absolute Error defined as,

$$MAE = \frac{1}{Eval}\sum_1^{Eval}\left|e^f\right|.\tag{10}$$

All results are presented with the corresponding Diebold and Mariano (1995) test of predictive accuracy where *, ** and *** star signs are used to indicate the predictive accuracy of the alternative model over the benchmark at 10%, 5% and 1% significance levels respectively.

The small dataset of predictors is described in Table 3 and consists of various government bond spreads, the real effective exchange rate, house lending rates, stock market return, money supply and the survey-based economic sentiment indicator for each economy. The dependent variable subject to forecasting is the growth rate of the industrial production of the Euro Area (16), France, Germany and Italy. The data was collected using Macrobond Financial and the dates span from Jan. 1996 to Jan. 2009.

The large dataset of predictors consists of 195 monthly variables (source: Eurostat, PEEIs, the Eurostat labels can be found in the Table 1) spanning from Jan. 1996 to Jan. 2009. The dataset is the same used in Foroni and Marcellino (2011) and it contains a large universe of variables that are potentially useful instruments in forecasting key macroeconomic variables in the Euro Area. Furthermore, in the spirit of Stock and Watson (2002a) we have transformed the series for stationarity using first differences or log differences appropriately (although notice in Table 2 some of the variables remained unchanged). Hence, the resulting data used in the forecast exercise contains growth rates from Feb. 1996 to Jan. 2009 (inclusive).

It is important to notice here that both small and large datasets of predictors are correctly date-aligned and the forecasting exercise uses the same time period in all experiments. The cross-validation (forecasting evaluation) period is set to 84 months, starts in Feb. 2002 and ends in Jan. 2009. The forecast horizon is set to $h = 12$ months.

4 Discussion of Results[‡]

4.1 A Small Dataset of Predictors

Tables 4-7 present the MAE of the benchmark ALI model that uses the small set of predictors and the relative MAE of PLS and BR models using the same dataset.

In Table 4 we present the forecasting exercise for the growth rate of the industrial production for the Euro Area economy. We see that the average MAE of the ALI model across all forecast horizons is equal to 0.007. We see that none of the alternative models outperforms the ALI benchmark. PLS with three factors is slightly better with a relative MAE equal to 0.998 and the Bayesian regression with shrinkage parameter $v = 0.5N$ presents the highest forecast error with relative MAE equal to 1.041. The same can be said for all subsequent steps up to $h = 12$ where the ALI benchmark is should be preferred.

The forecasting results for the growth rate of the industrial production for France can be found in Table 5. As in the EA case, the ALI should be chosen compared to the other variable reduction models using a small dataset. In steps $h = \{1...3, 10\}$ ALI is always better as the relative MAE is constantly above unity. For $h = 4$ to $h = 8$ we see that PLS(1) forecasts are slightly better compared to the benchmark's but not statistically significant. In steps 11 and 12 both BR models are slightly better.

Table 6 describes the forecasting results for Germany's industrial production growth rate. In steps $h = \{1, 3\}$ PLS(3) performs better with a MAE of 0.959 and 0.973 relative to the ALI benchmark. In steps $h = \{2, 4...6\}$ ALI outperforms all other models with an actual MAE of 0.010 across all steps. However in steps $h = \{8, 9, 11, 12\}$ other methods provide slightly better forecasts but not statistically significant. In step $h = 12$ the PLS(3) has a relative MAE of 0.987.

The last case is that of Italy in Table 7. This is another example where ALI is shown to outperform all other methods when a small set of predictors is used. We see that only in step 3 the alternative models, BR(0.5N) and BR(2N) give a relative MAE of 0.981 and 0.973 respectively which, however, is not statistically significant.

4.2 A Large Dataset of Predictors

We continue our study using the variable selection and variable reduction methods as described in Section 2 of the paper with a large dataset of predictors. The benchmark model is still the ALI when using the small dataset.

For the case of the EA we see in Table 8 that in the first step PLS(1)and PLS(3) provide forecasts that are statistically more accurate at 5% and 10% level respectively. In particular,

---

[‡]A cross-comparison of all models is available on request. The qualitative result does not change and hence the results are omitted.

the actual MAE of the ALI is 0.007 and the relative MAE of the above models is 0.940 and 0.884 respectively. For 2, 4, 6, 7 and 8 steps ahead the ALI outperforms all other competing methods with an average MAE of 0.006. PLS(1) and PLS(3) seem to be the most robust alternatives to the benchmark as they are characterised with relative MAE smaller than unity in most of the forecast horizons. This result is statistically significant at least at 10%. ALI loses its forecasting power as we move further ahead. For step $h = 12$ ALI has an actual MAE of 0.08 and PLS(3) has a relative MAE of 0.924.

In Table 9 we see the results for France. Here, the variable reductions methods (using the large dataset) constantly outperform the ALI across all horizons. Better forecasts are provided by SA_{BIC} and GA_{BIC} however they are not statistically significant. PLS(1), PLS(3) and BR(2N) return an average relative MAE of 0.970, 0.965 and 0.983 respectively across all forecast horizons.

Table 10 presents the forecasting exercise for the growth rate of the industrial production for Germany. The evidence here is in favour of the ALI as none of the competing models provides forecasts that are statistically more accurate. If we look at the table we can find cases where GA_{BIC}, PLS(1), PLS(2) and BR(2N) are slightly better however the Diebold and Mariano (1995) test does not indicate any statistical significance.

Moving to last set of empirical results for the case of the industrial production growth for Italy we see again that the variable selection methods can provide better forecasts in some horizons, however these results are not statistically significant. However, the variable reduction methods and, PLS(1), PLS(2) and BR(2N) in particular, return more accurate forecasts at 5% and 1% levels. PLS(1) and PLS(3) provide an average relative MAE of 0.987 and 0.986 respectively.

5   Concluding Remarks

In this paper we approach the issue of forecasting macroeconomic variables using indicators from small and large datasets. The methods we employ include variable reduction and variable selection models. The variables in the latter are selected according to the minimisation of information criteria using some not so standard optimisation techniques that aim to reduce the computational burden. The benchmark model is the Automatic Leading Indicator model of Camba-Mendez et al. (2001) which belongs to the class of the small scale dynamic factor models.

Our overall work indicates that among the variable reduction methods using a small set of predictors, the ALI is more likely to provide better forecasts on average. However, the ALI performs well only in the "medium-term" forecasts (4-6 steps ahead) using a large dataset of predictors. The variable reduction methodologies, and especially the PLS model with 1 and 3 factors, are more likely to result in forecasts with smaller errors in the first and 8 to

12 steps ahead. Similarly, the variable selection methods can also be useful for "short-run" (one step ahead) and "long-run" (10-12 steps ahead) predictions. Our results suggest the use of SA and GA as the more robust optimisation methodologies and the BIC as the objective function in the minimisation.

To conclude with, our empirical approach suggests that the models described here should, at least, be considered by researchers interested in model averaging methodologies for forecasting key European macroeconomic variables.

References

Alvarez, R., Camacho, M., Perez-Quiros, G., 2012. Finite sample performance of small versus large scale dynamic factor models. Discussion Paper 8867, Centre for Economic Policy Research.

Aruoba, B., Diebold, F., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business and Economic Statistics, 27, 417-427.

Aruoba, B., Diebold, F., 2010. Real-time macroeconomic monitoring: Real activity, inflation and interactions. American Economic Review, 100, 20-24.

Ng, S., Boivin, J., 2006. Are more data always better for factor analysis? Journal of Econometrics, 132, 169-194.

Camacho, M., Perez-Quiros, G., 2010. Introducin the Euro-STING: Short Term INdicator of Euro Area Growth. Journal of Applied Econometrics, 25, 663-694.

Camba-Mendez, G., Kapetanios, G., Smith, R. J., Weale, M. R., 2001. An automatic leading indicator of economic activity: forecasting GDP growth for European countries. Econometrics Journal, 4, S56-S90.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253-263.

De Mol, C., Giannone, D., Reichlin, L., 2006. Forecasting with a large number of predictors: is bayesian regression a valid alternative to principal components?. Discussion Paper 5829, Centre for Economic Policy Research.

Fernandez, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for bayesian model averaging. Journal of Econometrics, 100, 381-427.

Forni, M., Hallin, M., Reichlin, L., 2005. The generalized dynamic factor model: one-sided estimation and forecasting. Journal of the American Statistical Association, 100, 830-840.

Foroni, C., Marcellino, M., 2011. A comparison of mixed frequency approaches for modelling euro area macroeconomic variables. Mimeo.

Foroni, C., Marcellino, M., Schumacher, C., 2011. MIDAS and unrestricted MIDAS models. Mimeo.

Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimisation of statistical functions with fimulated annealing. Journal of Econometrics, 60-1, 65-99.

Hajek, B., 1998. Cooling schedules for optimal annealing. Mathematics of Operations Research, 13-2, 311–331.

Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society (Series B), 41, 190-195.

Hartl, H.R.F., Belew, R.K., 1990. A global convergence proof for a class of genetic algorithms. Technical Report,Technical Universiy of Vienna.

Harvey, D., 1993. Time Series Models. Hemel Hempstead: Harvester and Wheatsheaf.

Helland, I.S., 1998. On the structure of partial least squares regression. Communications in Statistics - Simulation and Computation, 17, 581-607.

Helland, I.S., 1990. Partial least squares regression and statistical models. Journal of Statistics, 17, 91-114.

Hoover, K.D., Perez, S.J., 1999. Data mining reconsidered: Encompassing and the general-to-specific approach to specification set. Econometrics Journal, 2, 167-191.

Kapetanios, G., 2006. Variable selection in regression models using non-standard optimisation of information criteria, Computational Statistics and Data Analysis, 52-1, 4-15.

Kapetanios, G., Marcellino, M., Papailias, F., 2012a. Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods. Unpublished Working Paper.

Kapetanios, G., Marcellino, M., Papailias, F., 2012b. Forecasting EU economic activity using financial conditions indexes. Unpublished Working Paper.

Kapetanios, G., Marcellino, M., Papailias, F., 2012c. Business cycles dating for EU economies: an empirical search for the optimal settings. Unpublished Working Paper.

Kapetanios, G., Marcellino, M., Papailias, F., 2012d. Forecasting EU economic activity using summary indicators. Unpublished Working Paper.

Morinaka, Y., Yoshikawa, M., Amagasa, T., 2001. The L-Index: An indexing structure for efficient subsequence matching in time sequence databases. Proceedings of the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining.

Schwarz, G., 1978. Estimating the dimension of the model. Annals of Statistics, 6, 461-464.

Sin, C.Y., White, H., 1996. Information criteria for selecting possibly misspecifed parametric models. Journal of Econometrics, 71-(1-2), 207-225.

Stock, J., Watson, M., 1989. New indexes of coincident and leading economic indicators. NBER Macroeconomic Annual, 351-394.

Stock, J., Watson, M., 1991. A probability model of the coincident economic indicators. In Leading Economic Indicators: New Approaches and Forecasting Records, edited by K. Lahiri and G. Moore. Cambridge University Press.

Stock, J., Watson, M., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 297, 1167-1179.

Stock, J., Watson, M., 2002b. Macroeconomic forecasting using diffusion indexes, Journal of Business & Economics Statistics, 20, 147-162.

Wold, H., 1982. Soft modeling. The basic design and some extensions. In: Joreskog, K. G., Wold, H., Systems under indirect observation, Vol. 2, North-Holland, Amsterdam.

Tables

Labels

#	Label	#	Label	#	Label	#	Label	#	Label	#	Label	#	Label
1	$CP-HI00XEFU$	31	$MIG-COG-IS-PPI$	61	$BS-FS-NY$	91	$BS-CSMCI-B$	121	$B-IS-EPI$	151	$MIG-DCOG-IS-ITND$	181	$EMECB5Y$
2	$CP-HI00XEF$	32	$MIG-DCOG-IS-PPI$	62	$BS-GES-LY$	92	$BS-ESI-I$	122	$B-TO-E36-IS-EPI$	152	$MIG-DCOG-IS-ITT$	182	$EMECB7Y$
3	$CP-HI00XES$	33	$MIG-ING-IS-PPI$	63	$BS-GES-NY$	93	$BS-ICI-BAL$	123	$C-IS-EPI$	153	$MIG-ING-IS-ITD$	183	$EMGBOND$
4	$CP-HI00XE$	34	$MIG-NDCOG-IS-PPI$	64	$BS-MP-NY$	94	$BS-RCI-BAL$	124	$D35-E36-IS-EPI$	154	$MIG-ING-IS-ITND$	184	$FIBOR1Y$
5	$CP-HI00XTB$	35	$MIG-NRG-IS-PPI$	65	$BS-MP-PR$	95	$BS-SCI-BAL$	125	$D-IS-EPI$	155	$MIG-ING-IS-ITT$	185	$FIBOR3M$
6	$CP-HI00$	36	$D35-E36-IS-IMPR$	66	$BS-PT-LY$	96	$RT-LM-UN-T-GT25$	126	$E36-IS-EPI$	156	$MIG-NDCOG-IS-ITD$	186	$FIBOR6M$
7	$CP-HI01$	37	$IS-WSI-F$	67	$BS-PT-NY$	97	$RT-LM-UN-T-LE25$	127	$E-IS-EPI$	157	$MIG-NDCOG-IS-ITT$	187	$BDWU1032R$
8	$CP-HI02$	38	$B-D-IS-WSI$	68	$BS-SFSH$	98	$RT-LM-UN-T-TOT$	128	$MIG-CAG-IS-EPI$	158	$IS-PEI-F-CC11-X-CC113$	188	$BDWU0022R$
9	$CP-HI03$	39	$B-E36-IS-WSI$	69	$BS-SV-NY$	99	$1000-PERS-LM-UN-T-GT25$	129	$MIG-COG-IS-EPI$	159	$IS-IPI-F-CC11-X-CC113$	189	$BDEBDBSIA$
10	$CP-HI04$	40	$B-C-D-IS-WSI$	70	$BS-SV-PR$	100	$1000-PERS-LM-UN-T-LE25$	130	$MIG-DCOG-IS-EPI$	160	$IS-HWI-F$	190	$BDECBXDGA$
11	$CP-HI05$	41	$B-IS-WSI$	71	$BS-UE-NY$	101	$1000-PERS-LM-UN-T-TOT$	131	$MIG-ING-IS-EPI$	161	$C-ORD-IS-IO$	191	$BDECBXDMA$
12	$CP-HI06$	42	$B-TO-E36-IS-WSI$	72	$BS-ICI$	102	$IS-IP$	132	$MIG-NDCOG-IS-EPI$	162	$C-ORD-X-C30-IS-IO$	192	$BDECBXNGA$
13	$CP-HI07$	43	$C-IS-WSI$	73	$BS-IEME$	103	$IS-IP-F-CC1$	133	$MIG-NRG-IS-EPI$	163	$IS-CAR$	193	$BDECBXNOA$
14	$CP-HI08$	44	$D35-E36-IS-WSI$	74	$BS-IEOB$	104	$IS-IP-F-CC2$	134	$G45-IS-EPI$	164	$FOOD-IS-DIT$	194	$BDECBXOLA$
15	$CP-HI09$	45	$D-IS-WSI$	75	$BS-IOB$	105	$IS-IP-F$	135	$B-C-IS-ITD$	165	$IS-DIT$	195	$BDECBXLIA$
16	$CP-HI10$	46	$E36-IS-WSI$	76	$BS-IPE$	106	$B-D-IS-IP$	136	$B-C-IS-ITND$	166	$NFOOD-IS-DIT$		
17	$CP-HI11$	47	$MIG-CAG-IS-WSI$	77	$BS-IPT$	107	$B-C-IS-IP$	137	$B-C-IS-ITT$	167	$NFOOD-X-G473-IS-DIT$		
18	$CP-HI12$	48	$MIG-COG-IS-WSI$	78	$BS-ISFP$	108	$B-IS-IP$	138	$C-IS-ITD$	168	$X-G473-IS-DIT$		
19	$CP-HIE$	49	$MIG-DCOG-IS-WSI$	79	$BS-ISPE$	109	$C-IS-IP$	139	$C-IS-ITND$	169	$M1$		
20	$CP-HIF$	50	$MIG-ING-IS-WSI$	80	$BS-RAS$	110	$C-ORD-IS-IP$	140	$C-IS-ITT$	170	$M2$		
21	$B-D-IS-PPI$	51	$MIG-NDCOG-IS-WSI$	81	$BS-RCI$	111	$D-IS-IP$	141	$C-ORD-IS-ITD$	171	$M3$		
22	$B-E36-IS-PPI$	52	$MIG-NRG-IS-WSI$	82	$BS-REBS$	112	$MIG-CAG-IS-IP$	142	$C-ORD-IS-ITND$	172	$3MI-RT$		
23	$B-C-D-IS-PPI$	53	$BS-CCI-BAL$	83	$BS-REM$	113	$MIG-COG-IS-IP$	143	$C-ORD-IS-ITT$	173	$LTGBY-RT$		
24	$B-IS-PPI$	54	$BS-CEME-BAL$	84	$BS-ROP$	114	$MIG-DCOG-IS-IP$	144	$MIG-CAG-IS-ITD$	174	$EXA-RT-USD$		
25	$B-TO-E36-IS-PPI$	55	$BS-COB-BAL$	85	$BS-RPBS$	115	$MIG-ING-IS-IP$	145	$MIG-CAG-IS-ITND$	175	$EXA-RT-JPY$		
26	$C-IS-PPI$	56	$BS-CPE-BAL$	86	$BS-SABC$	116	$MIG-NDCOG-IS-IP$	146	$MIG-CAG-IS-ITT$	176	$EXA-RT-GBP$		
27	$C-ORD-IS-PPI$	57	$BS-CTA-BAL$	87	$BS-SAEM$	117	$IS-EPI-F$	147	$MIG-COG-IS-ITD$	177	$BDSHRPRCF$		
28	$D-IS-PPI$	58	$BS-BCI$	88	$BS-SARM$	118	$B-D-IS-EPI$	148	$MIG-COG-IS-ITND$	178	$DJES50I$		
29	$E36-IS-PPI$	59	$BS-CSMCI$	89	$BS-SCI$	119	$B-E36-IS-EPI$	149	$MIG-COG-IS-ITT$	179	$EMECB2Y$		
30	$MIG-CAG-IS-PPI$	60	$BS-FS-LY$	90	$BS-SERM$	120	$B-C-IS-EPI$	150	$MIG-DCOG-IS-ITD$	180	$EMECB3Y$		

Table 1: Large dataset of predictors: Labels

Transformations

#	Label	#	Label	#	Label	#	Label	#	Label	#	Label	#	Label
1	$FirstDiff., Logs$	31	$FirstDiff., Logs$	61	$NoChange$	91	$NoChange$	121	$FirstDiff., Logs$	151	$FirstDiff., Logs$	181	$FirstDiff.$
2	$FirstDiff., Logs$	32	$FirstDiff., Logs$	62	$NoChange$	92	$NoChange$	122	$FirstDiff., Logs$	152	$FirstDiff., Logs$	182	$FirstDiff.$
3	$FirstDiff., Logs$	33	$FirstDiff., Logs$	63	$NoChange$	93	$NoChange$	123	$FirstDiff., Logs$	153	$FirstDiff., Logs$	183	$FirstDiff.$
4	$FirstDiff., Logs$	34	$FirstDiff., Logs$	64	$NoChange$	94	$NoChange$	124	$FirstDiff., Logs$	154	$FirstDiff., Logs$	184	$FirstDiff.$
5	$FirstDiff., Logs$	35	$FirstDiff., Logs$	65	$NoChange$	95	$NoChange$	125	$FirstDiff., Logs$	155	$FirstDiff., Logs$	185	$FirstDiff.$
6	$FirstDiff., Logs$	36	$FirstDiff., Logs$	66	$NoChange$	96	$FirstDiff.$	126	$FirstDiff., Logs$	156	$FirstDiff., Logs$	186	$FirstDiff.$
7	$FirstDiff., Logs$	37	$FirstDiff., Logs$	67	$NoChange$	97	$FirstDiff.$	127	$FirstDiff., Logs$	157	$FirstDiff., Logs$	187	$FirstDiff.$
8	$FirstDiff., Logs$	38	$FirstDiff., Logs$	68	$NoChange$	98	$FirstDiff.$	128	$FirstDiff., Logs$	158	$FirstDiff., Logs$	188	$FirstDiff.$
9	$FirstDiff., Logs$	39	$FirstDiff., Logs$	69	$NoChange$	99	$NoChange$	129	$FirstDiff., Logs$	159	$FirstDiff., Logs$	189	$FirstDiff., Logs$
10	$FirstDiff., Logs$	40	$FirstDiff., Logs$	70	$NoChange$	100	$NoChange$	130	$FirstDiff., Logs$	160	$FirstDiff., Logs$	190	$FirstDiff., Logs$
11	$FirstDiff., Logs$	41	$FirstDiff., Logs$	71	$NoChange$	101	$NoChange$	131	$FirstDiff., Logs$	161	$FirstDiff., Logs$	191	$FirstDiff., Logs$
12	$FirstDiff., Logs$	42	$FirstDiff., Logs$	72	$NoChange$	102	$FirstDiff., Logs$	132	$FirstDiff., Logs$	162	$FirstDiff., Logs$	192	$FirstDiff., Logs$
13	$FirstDiff., Logs$	43	$FirstDiff., Logs$	73	$NoChange$	103	$FirstDiff., Logs$	133	$FirstDiff., Logs$	163	$FirstDiff., Logs$	193	$FirstDiff., Logs$
14	$FirstDiff., Logs$	44	$FirstDiff., Logs$	74	$NoChange$	104	$FirstDiff., Logs$	134	$FirstDiff., Logs$	164	$FirstDiff., Logs$	194	$FirstDiff., Logs$
15	$FirstDiff., Logs$	45	$FirstDiff., Logs$	75	$NoChange$	105	$FirstDiff., Logs$	135	$FirstDiff., Logs$	165	$FirstDiff., Logs$	195	$FirstDiff., Logs$
16	$FirstDiff., Logs$	46	$FirstDiff., Logs$	76	$NoChange$	106	$FirstDiff., Logs$	136	$FirstDiff., Logs$	166	$FirstDiff., Logs$		
17	$FirstDiff., Logs$	47	$FirstDiff., Logs$	77	$NoChange$	107	$FirstDiff., Logs$	137	$FirstDiff., Logs$	167	$FirstDiff., Logs$		
18	$FirstDiff., Logs$	48	$FirstDiff., Logs$	78	$NoChange$	108	$FirstDiff., Logs$	138	$FirstDiff., Logs$	168	$FirstDiff., Logs$		
19	$FirstDiff., Logs$	49	$FirstDiff., Logs$	79	$NoChange$	109	$FirstDiff., Logs$	139	$FirstDiff., Logs$	169	$FirstDiff., Logs$		
20	$FirstDiff., Logs$	50	$FirstDiff., Logs$	80	$NoChange$	110	$FirstDiff., Logs$	140	$FirstDiff., Logs$	170	$FirstDiff., Logs$		
21	$FirstDiff., Logs$	51	$FirstDiff., Logs$	81	$NoChange$	111	$FirstDiff., Logs$	141	$FirstDiff., Logs$	171	$FirstDiff., Logs$		
22	$FirstDiff., Logs$	52	$FirstDiff., Logs$	82	$NoChange$	112	$FirstDiff., Logs$	142	$FirstDiff., Logs$	172	$FirstDiff.$		
23	$FirstDiff., Logs$	53	$NoChange$	83	$NoChange$	113	$FirstDiff., Logs$	143	$FirstDiff., Logs$	173	$FirstDiff.$		
24	$FirstDiff., Logs$	54	$NoChange$	84	$NoChange$	114	$FirstDiff., Logs$	144	$FirstDiff., Logs$	174	$FirstDiff.$		
25	$FirstDiff., Logs$	55	$NoChange$	85	$NoChange$	115	$FirstDiff., Logs$	145	$FirstDiff., Logs$	175	$FirstDiff.$		
26	$FirstDiff., Logs$	56	$NoChange$	86	$NoChange$	116	$FirstDiff., Logs$	146	$FirstDiff., Logs$	176	$FirstDiff.$		
27	$FirstDiff., Logs$	57	$NoChange$	87	$NoChange$	117	$FirstDiff., Logs$	147	$FirstDiff., Logs$	177	$FirstDiff., Logs$		
28	$FirstDiff., Logs$	58	$NoChange$	88	$NoChange$	118	$FirstDiff., Logs$	148	$FirstDiff., Logs$	178	$FirstDiff., Logs$		
29	$FirstDiff., Logs$	59	$NoChange$	89	$NoChange$	119	$FirstDiff., Logs$	149	$FirstDiff., Logs$	179	$FirstDiff.$		
30	$FirstDiff., Logs$	60	$NoChange$	90	$NoChange$	120	$FirstDiff., Logs$	150	$FirstDiff., Logs$	180	$FirstDiff.$		

Table 2: Large dataset of predictors: Transformations

		Small Dataset of Predictors		
	Variable	Description	Transformation
		Rates	
1	10yr Government Benchmark Bond Yield	Government Benchmarks	Level
2	10yr Govt Bond/3m US Govt Bond Spread	Government Benchmarks	Level
3	10yr Govt Bond/10yr US Govt Bond Spread	Government Benchmarks	Level
4	2yr Govt Bond/2yr US Govt Bond Spread	Government Benchmarks	Level
5	Corporate Bonds Yield	Corporate Benchmarks	Level
6	Real Effective Exchange Rate	FX Indices, BIS	Level
7	House Lending	Lending for House Purchase	Level
8	Local Equity Index Return	Equity Indices	Growth
9	Local Volatility Index Return	Volatility Indices	Growth
10	Money Supply: M3	Monetary Aggregates	Growth
11	Economic Sentiment Indicator	Economic Sentiment Surveys	Level
		Dependent Variable	
1	Industrial Production (SA) Growth	Industrial Production Index	Growth

Table 3: Small dataset of predictors

						EA Industrial Production Growth Rate						
						Forecast Horizon						
h	1	2	3	4	5	6	7	8	9	10	11	12
						MAE						
ALI	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
						MAE (Relative to ALI)						
PLS(1)	1.020	1.031	1.018	1.050	1.031	1.025	1.029	1.025	1.055	1.046	1.006	1.009
PLS(3)	0.998	1.024	1.051	1.087	1.027	1.052	1.056	0.992	1.080	1.051	1.022	1.063
BR(0.5N)	1.041	1.061	1.019	1.094	1.057	1.046	1.070	1.018	1.057	1.082	1.040	1.031
BR(2N)	1.028	1.049	1.007	1.057	1.026	1.031	1.035	0.977	1.034	1.035	1.016	1.010
						Cross-validation: 84 periods						

Table 4: Forecasting the industrial production growth rate of the EA using a small set of predictors

Notes: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. ***** indicates the 10% significance of Diebold and Mariano (1995) test. ****** indicates the 5% significance of Diebold and Mariano (1995) test. ******* indicates the 1% significance of Diebold and Mariano (1995) test.

						France Industrial Production Growth Rate						
						Forecast Horizon						
h	1	2	3	4	5	6	7	8	9	10	11	12
						MAE						
ALI	0.010	0.010	0.010	0.010	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
						MAE (Relative to ALI)						
PLS(1)	1.005	1.027	1.020	0.998	0.987	0.988	0.996	0.997	0.999	1.021	1.004	0.990
PLS(3)	1.006	1.028	1.030	0.999	1.019	0.990	1.014	0.986	1.006	1.069	1.022	0.993
BR(0.5N)	1.039	1.040	1.020	1.041	1.039	1.009	1.038	0.985	1.057	1.012	0.995	0.998
BR(2N)	1.021	1.030	1.010	1.022	1.010	0.992	1.012	0.978	1.019	1.003	0.993	0.993
						Cross-validation: 84 periods						

Table 5: Forecasting the industrial production growth rate of France using a small set of predictors

h	1	2	3	4	5	6	7	8	9	10	11	12
					Germany Industrial Production Growth Rate							
					Forecast Horizon							
					MAE							
ALI	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.011
					MAE (Relative to ALI)							
PLS(1)	1.002	1.007	0.979	1.014	1.005	1.000	1.004	0.998	0.997	1.029	0.990	0.988
PLS(3)	0.959	1.009	0.973	1.029	1.067	1.052	1.047	0.987	1.041	1.042	1.027	0.987
BR(0.5N)	1.047	1.043	1.029	1.007	1.062	1.102	1.000	1.023	1.077	1.062	1.046	1.005
BR(2N)	1.023	1.027	1.011	1.003	1.039	1.045	0.980	1.000	1.053	1.021	1.010	0.990
					Cross-validation: 84 periods							

Table 6: Forecasting the industrial production growth rate of Germany using a small set of predictors

Notes: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. * indicates the 10% significance of Diebold and Mariano (1995) test. ** indicates the 5% significance of Diebold and Mariano (1995) test. *** indicates the 1% significance of Diebold and Mariano (1995) test.

h	1	2	3	4	5	6	7	8	9	10	11	12
					Italy Industrial Production Growth Rate							
					Forecast Horizon							
					MAE							
ALI	0.011	0.011	0.011	0.011	0.011	0.011	0.012	0.012	0.012	0.012	0.012	0.013
					MAE (Relative to ALI)							
PLS(1)	1.001	1.000	0.999	1.002	1.002	1.002	1.004	1.001	1.004	1.004	0.992	1.001
PLS(3)	1.015	1.006	0.994	1.011	1.026	1.004	1.026	1.003	1.002	1.016	1.012	1.016
BR(0.5N)	1.009	0.998	0.981	1.013	1.031	1.012	1.033	0.999	1.002	1.024	1.030	1.014
BR(2N)	1.004	1.000	0.973	1.007	1.022	1.008	1.027	1.002	1.003	1.019	1.021	1.014
					Cross-validation: 84 periods							

Table 7: Forecasting the industrial production growth rate of Italy using a small set of predictors

Notes: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. ***** indicates the 10% significance of Diebold and Mariano (1995) test. ****** indicates the 5% significance of Diebold and Mariano (1995) test. ******* indicates the 1% significance of Diebold and Mariano (1995) test.

					EA Industrial Production Growth Rate							
					Forecast Horizon							
h	1	2	3	4	5	6	7	8	9	10	11	12
					MAE							
ALI	0.007	0.007	0.007	0.007	0.007	0.007	0.006	0.006	0.007	0.007	0.007	0.008
					MAE (Relative to ALI)							
SA_{BIC}	0.972	1.132	1.131	1.269	1.222	1.205	1.203	1.091	1.117	1.299	1.218	1.088
GA_{BIC}	1.029	1.136	1.051	1.227	1.250	1.184	1.242	1.131	1.188	1.313	1.275	1.185
MC^3_{BIC}	0.975	1.057	1.054	1.261	1.253	1.087	1.104	1.148	1.115	1.027	1.101	1.122
SA_{HQ}	1.238	1.484	1.533	1.770	1.595	1.975	1.634	1.790	1.835	2.070	2.024	1.829
GA_{HQ}	1.142	1.198	1.270	1.471	1.288	1.409	1.284	1.330	1.305	1.436	1.431	1.240
MC^3_{HQ}	0.995	1.106	1.151	1.168	1.229	1.198	1.204	1.177	1.220	1.210	1.229	1.198
PLS(1)	0.940**	1.005	0.937**	1.001	0.978**	1.032	1.007	1.004	0.984**	1.002	0.931**	0.939**
PLS(3)	0.884*	1.006	0.940*	1.009	0.981*	1.036	1.009	0.997	0.978*	0.985*	0.929*	0.924*
BR(0.5N)	1.022	1.136	1.124	1.258	1.206	1.384	1.240	1.243	1.188	1.249	1.168	1.068
BR(2N)	0.934	1.033	0.982	1.095	1.060	1.139	1.080	1.098	1.033	1.060	0.977	0.965
					Cross-validation: 84 periods							

Table 8: Forecasting the industrial production growth rate of the EA using a large set of predictors

Notes: **ALI** denotes the Automatic Leading Indicator model with 1 factor, SA_{BIC} denotes the Simmulated Annealing algorithm that optimises the BIC, GA_{BIC} denotes the Genetic algorithm that optimises the BIC, MC^3_{BIC} denotes the MC^3 algorithm that optimises the BIC, SA_{HQ} denotes the Simmulated Annealing algorithm that optimises the HQ criterion, GA_{HQ} denotes the Genetic algorithm that optimises the HQ criterion, MC^3_{HQ} denotes the MC^3 algorithm that optimises the HQ criterion, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. ***** indicates the 10% significance of Diebold and Mariano (1995) test. ****** indicates the 5% significance of Diebold and Mariano (1995) test. ******* indicates the 1% significance of Diebold and Mariano (1995) test.

h	1	2	3	4	5	6	7	8	9	10	11	12
						MAE						
ALI	0.010	0.010	0.010	0.010	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.012
					MAE (Relative to ALI)							
SA_{BIC}	0.935	1.020	1.093	1.079	1.073	1.088	1.081	1.031	1.005	0.998	1.006	1.015
GA_{BIC}	0.984	0.996	1.048	0.979	1.019	1.025	1.070	1.035	0.980	0.949	1.007	1.044
MC^3_{BIC}	0.940	1.011	1.092	1.060	1.023	1.020	1.054	1.030	0.990	0.970	0.983	1.031
SA_{HQ}	1.184	1.230	1.380	1.535	1.348	1.318	1.252	1.190	1.113	1.355	1.537	1.696
GA_{HQ}	1.098	1.236	1.230	1.310	1.136	1.163	1.373	1.079	1.507	1.623	2.463	2.281
MC^3_{HQ}	0.959	1.089	1.210	1.171	1.063	1.121	1.164	1.092	1.038	1.016	1.002	1.151
PLS(1)	0.957***	1.003	0.974***	0.970***	0.932***	1.003	0.978***	0.973***	0.970***	0.961***	0.955***	0.967***
PLS(3)	0.913***	1.004	0.978***	0.971***	0.938***	1.000	0.990	0.980***	0.946***	0.943***	0.952***	0.960***
BR(0.5N)	0.953***	1.013	1.092	1.155	0.994***	1.124	1.070	1.046	1.072	1.047	1.034	0.945***
BR(2N)	0.928***	0.974***	1.005	1.040	0.942***	1.023	1.000	1.012	0.988**	0.980**	0.981**	0.929***
					Cross-validation: 84 periods							

Table with header: France Industrial Production Growth Rate / Forecast Horizon

Table 9: Forecasting the industrial production growth rate of France using a large set of predictors

Notes: ALI denotes the Automatic Leading Indicator model with 1 factor, SA_{BIC} denotes the Simmulated Annealing algorithm that optimises the BIC, GA_{BIC} denotes the Genetic algorithm that optimises the BIC, MC^3_{BIC} denotes the MC^3 algorithm that optimises the BIC, SA_{HQ} denotes the Simmulated Annealing algorithm that optimises the HQ criterion, GA_{HQ} denotes the Genetic algorithm that optimises the HQ criterion, MC^3_{HQ} denotes the MC^3 algorithm that optimises the HQ criterion, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. ***** indicates the 10% significance of Diebold and Mariano (1995) test. ****** indicates the 5% significance of Diebold and Mariano (1995) test. ******* indicates the 1% significance of Diebold and Mariano (1995) test.

					Germany Industrial Production Growth Rate							
					Forecast Horizon							
h	1	2	3	4	5	6	7	8	9	10	11	12
					MAE							
ALI	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.011
					MAE (Relative to ALI)							
SA_{BIC}	1.016	1.092	0.948	1.020	1.127	1.005	1.036	1.036	0.984	0.993	0.993	1.014
GA_{BIC}	1.004	1.080	0.956	1.115	1.103	0.973	1.067	1.045	1.044	1.044	0.997	1.058
MC^3_{BIC}	1.010	1.063	0.924	1.064	1.125	0.965	1.036	1.028	1.018	1.137	1.075	0.989
SA_{HQ}	1.625	3.160	1.487	1.625	1.619	1.586	1.564	1.756	1.961	1.657	1.619	2.008
GA_{HQ}	1.246	1.236	1.019	1.123	1.360	1.315	1.124	1.195	1.291	1.249	1.182	1.280
MC^3_{HQ}	1.057	1.132	0.996	1.094	1.227	1.064	1.099	1.198	1.044	1.160	1.078	1.052
PLS(1)	1.005	1.005	0.938	0.998	1.008	0.979	1.010	1.011	1.006	0.996	1.007	1.013
PLS(3)	0.945	1.008	0.946	1.004	1.017	0.997	1.017	1.021	1.022	1.000	1.037	0.995
BR(0.5N)	1.114	1.153	0.971	1.155	1.081	1.208	1.108	1.085	1.161	1.154	1.091	1.026
BR(2N)	1.008	1.053	0.939	1.047	1.011	1.046	1.020	1.016	1.064	1.059	1.004	0.988
					Cross-validation: 84 periods							

Table 10: Forecasting the industrial production growth rate of Germany using a large set of predictors

Notes: ALI denotes the Automatic Leading Indicator model with 1 factor, SA_{BIC} denotes the Simmulated Annealing algorithm that optimises the BIC, GA_{BIC} denotes the Genetic algorithm that optimises the BIC, MC^3_{BIC} denotes the MC^3 algorithm that optimises the BIC, SA_{HQ} denotes the Simmulated Annealing algorithm that optimises the HQ criterion, GA_{HQ} denotes the Genetic algorithm that optimises the HQ criterion, MC^3_{HQ} denotes the MC^3 algorithm that optimises the HQ criterion, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. ***** indicates the 10% significance of Diebold and Mariano (1995) test. ****** indicates the 5% significance of Diebold and Mariano (1995) test. ******* indicates the 1% significance of Diebold and Mariano (1995) test.

h	1	2	3	4	5	6	7	8	9	10	11	12
					MAE							
ALI	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.012	0.012	0.012	0.012	0.013
					MAE (Relative to ALI)							
SA_{BIC}	1.008	1.048	1.046	1.040	1.044	0.985	1.037	1.102	1.035	0.990	1.107	1.011
GA_{BIC}	1.002	1.064	1.046	1.032	1.020	0.972	1.020	1.064	1.034	0.988	1.070	0.985
MC^3_{BIC}	0.970	1.032	1.024	1.054	1.009	1.001	1.026	1.078	1.049	1.008	1.026	0.998
SA_{HQ}	1.067	1.136	1.178	1.186	1.090	1.050	1.171	1.319	1.103	1.172	1.053	1.180
GA_{HQ}	1.097	1.069	1.103	1.060	1.038	1.064	1.214	1.066	1.167	1.092	1.068	1.063
MC^3_{HQ}	1.029	1.055	1.030	0.997	1.086	1.014	1.078	1.073	1.042	1.010	1.052	0.962
PLS(1)	0.959***	1.006	0.984***	1.001	0.983***	1.006	0.994	0.976***	1.024	0.984***	0.967***	0.964***
PLS(3)	0.928***	1.010	0.984***	1.012	0.984***	1.003	0.997	0.984**	1.025	0.988***	0.984***	0.936***
BR(0.5N)	1.003	1.068	1.141	1.138	1.055	1.064	1.033	1.092	1.070	1.029	1.018	0.978**
BR(2N)	0.951***	1.014	1.034	1.050	1.006	1.008	1.004	1.032	1.020	0.977***	0.976***	0.962***
					Cross-validation: 84 periods							

Table 11: Forecasting the industrial production growth rate of Italy using a large set of predictors

Notes: ALI denotes the Automatic Leading Indicator model with 1 factor, SA_{BIC} denotes the Simmulated Annealing algorithm that optimises the BIC, GA_{BIC} denotes the Genetic algorithm that optimises the BIC, MC^3_{BIC} denotes the MC^3 algorithm that optimises the BIC, SA_{HQ} denotes the Simmulated Annealing algorithm that optimises the HQ criterion, GA_{HQ} denotes the Genetic algorithm that optimises the HQ criterion, MC^3_{HQ} denotes the MC^3 algorithm that optimises the HQ criterion, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$. * indicates the 10% significance of Diebold and Mariano (1995) test. ** indicates the 5% significance of Diebold and Mariano (1995) test. *** indicates the 1% significance of Diebold and Mariano (1995) test.