CENTRAL BANK OF ICELAND

WORKING PAPERS No. 34

An Embarrassment of Riches: Forecasting Using Large Panels

by
Jana Eklund
Sune Karlsson

May 2007

CENTRAL BANK OF ICELAND

Economics Department

Central Bank of Iceland Working Papers are published by the Economics Department of the Central Bank of Iceland. The views expressed in them are those of their authors and not necessarily the views of the Central Bank of Iceland.

Also available on the Central Bank of Iceland World Wide Web site (http://www.sedlabanki.is)

ISSN: 1028-9445

An Embarrassment of Riches: Forecasting Using Large Panels

Jana Eklund
Bank of England*

Sune Karlsson
Örebro University†

February 7, 2007

Abstract

The problem of having to select a small subset of predictors from a large number of useful variables can be circumvented nowadays in forecasting. One possibility is to efficiently and systematically evaluate all predictors and almost all possible models that these predictors in combination can give rise to. The idea of combining forecasts from various indicator models by using Bayesian model averaging is explored, and compared to diffusion indexes, another method using large number of predictors to forecast. In addition forecasts based on the median model are considered.

Keywords: Bayesian model averaging, Diffusion indexes, Inflation rate.

JEL-codes: C11, C51, C52, C53

*Jana.Eklund@bankofengland.co.uk

†Sune.Karlsson@esi.oru.se

1 Introduction

The number of potential predictors for macroeconomic variables can easily count in the hundreds, e.g. Stock and Watson (2002b) collect some 200 predictor variables for the US economy. Paradoxically, having more information in the form of more predictor variables makes forecasting more difficult. Simply put, the task of formulating a forecasting model and extracting the relevant information from the predictors becomes more complex as the number of possible predictors increases. The increasing availability of data is thus creating new challenges for the forecaster. There are, essentially, two different approaches to address this problem. The first approach builds forecast models based on summaries of the predictor variables, such as principal components, and the second approach is analogous to forecast combination, where the forecasts from a multitude of possible models are averaged.

Each attempts to overcome the shortcomings of the traditional approach of selecting a single forecasting model based on a few predictors. Clearly, using a single model, which by necessity can only incorporate a small subset of the variables, will fail to take account of all information in the data. In addition, by being based on a single model the forecast does not take account of model uncertainty. Basing the forecast model on data summaries in the form of principal components, as in Stock and Watson (2002b), allows information from all the predictors to enter into the forecasts, but not necessarily in an optimal fashion since the summaries of the predictors are created without a reference to the predicted variable. Model averaging, on the other hand, summarizes the different possible *relationships* between the predicted variable and the predictor variables. With appropriately chosen weights, this should lead to more efficient extraction of information. Model averaging also has the advantage of providing robustness against misspecification, and model uncertainty can easily be accounted for if the model averaging is conducted in a Bayesian setting, i.e. the weights are the posterior probabilities of the models. In addition the models and their averages are more easily interpreted than principal components and inference on the importance of individual predictors is available.

The potential benefits of model averaging as a tool for extracting the relevant information from a large set of predictor variables come at the cost of considerable computational complexity. With 100 predictor variables one obtains more than 10^{30} different models just by considering the different possible combinations of the variables, and it is clearly impossible to include all of them in a model averaging exercise. Recent advances in Bayesian computing, utilized by e.g. Jacobson and Karlsson (2004), provide one way forward by identifying the subset of important models as measured by their posterior probability, i.e. the set of models which would receive a non-negligible weight in the forecast combination.

Koop and Potter (2004) apply Bayesian model averaging (BMA) to dynamic factor models by orthogonalizing the predictors, using a transformation to principal components. Koop and Potter conclude that models containing factors do outperform autoregressive models in forecasting, but only narrowly and at short horizons. Also the gains provided by using BMA over forecasting methods based on a single model are more appreciable relative to the small forecasting gains from factor-based models.

The purpose of this paper is to evaluate the forecasting performance of the factor model approach of Stock and Watson (2002b), the Bayesian model averaging approach of Jacobson and Karlsson (2004), and the combined approach of Koop and Potter (2004). Any forecast evaluation is dependent on the choice of variable to forecast and the dataset used. To protect against this, we use three different datasets with two different frequencies, and forecast both inflation and GDP.

In all three cases the forecasts are based on a simple linear model,

$$y_{t+h} = \mathbf{x}_t \boldsymbol{\beta}_h + \varepsilon_{t+h}, \tag{1}$$

where $\mathbf{x}_t$, in the case of Stock and Watson, consists of the first few principal components, possibly augmented with lags of these and lagged values of y_t. In the Bayesian model averaging approach of Jacobson and Karlsson, $\mathbf{x}_t$ is a subset of the regressor variables, possibly including lags of the predictors and y_t, and the forecasts are obtained by averaging over the forecasts from the different models. In the combined approach of Koop and Potter, $\mathbf{x}_t$ contains a selected subset of orthogonalized regressors. There are two features worth noting about this setup, the forecast model depends on the forecast horizon, h, and the forecasts are static, i.e. there is no need to forecast $\mathbf{x}_t$.

The remainder of the chapter is organized as follows. Section 2 presents the forecasting approaches in large panels, Section 3 compares the different forecast methods, and Section 4 concludes.

2 Forecasting methods

2.1 Factor models

The factor based approach to forecasting with large data sets is based on the assumption that the relevant information is captured by a small number of factors common to the predictor variables. The forecasts are constructed using a two-step procedure. First, the method of principal components is used to extract factors from the predictors $\mathbf{x}_t$. In the second step the factors are used to forecast the time series y_{t+h}.

In particular, let y_{t+h} be a scalar series that is being forecast h-periods ahead, and let $\mathbf{x}_t$ be a N-dimensional multiple time series of variables serving as predictors. Now consider the forecasting equation

$$y_{t+h} = \boldsymbol{\beta}\left(\mathrm{L}\right) \mathbf{f}_t + \boldsymbol{\gamma} \mathbf{y}_t + \varepsilon_{t+h}, \qquad t = 1, \ldots, T, \tag{2}$$

where $\mathbf{f}_t$ is a vector of q unobservable common factors and $\mathbf{y}_t$ is a set of $p+1$ variables, such as lags of y_t. Furthermore $\boldsymbol{\beta}(\mathrm{L})$ is a vector lag polynomial and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)'$. Suppose that the observed series $\mathbf{x}_t$ and y_{t+h} in (2) allow for a dynamic factor model with q common dynamic factors $\mathbf{f}_t$

$$x_{it} = \boldsymbol{\lambda}_i(\mathrm{L})\,\mathbf{f}_t + e_{it}, \qquad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{3}$$

where $\boldsymbol{\lambda}_i(\mathrm{L})$ is a lag polynomial vector, and e_{it} is an idiosyncratic disturbance. In addition, it is assumed that

$$\mathrm{E}\left(\varepsilon_{t+h}\middle|\, \mathbf{f}_t, y_t, \mathbf{x}_t, \mathbf{f}_{t-1}, y_{t-1}, \mathbf{x}_{t-1}, \ldots\right) = 0. \tag{4}$$

Assume that the lag polynomial vectors are of order s. The dynamic factor model $(2)-(3)$ can be then restated as

$$y_{t+h} = \mathbf{B}'\mathbf{F}_t + \boldsymbol{\gamma}\mathbf{y}_t + \varepsilon_{t+h}, \tag{5}$$

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \mathbf{e}_t, \tag{6}$$

where $\mathbf{B} = (\boldsymbol{\beta}_0', \ldots, \boldsymbol{\beta}_s')'$, $\mathbf{F}_t = \left(\mathbf{f}_t', \mathbf{f}_{t-1}', \ldots, \mathbf{f}_{t-s}'\right)'$ is a $((s+1)\,q \times 1)$ vector, and the i-th row of $\boldsymbol{\Lambda}$ is $(\lambda_{i0}, \ldots, \lambda_{is})$.

The $(s+1)\,q$ factors $\mathbf{F}_t$ in (6) are estimated using principal components, denoted by $\widetilde{\mathbf{F}}_t$. In the second step, after regressing y_{t+h} on a constant, $\widetilde{\mathbf{F}}_t$, w possible lags of $\widetilde{\mathbf{F}}_t$, p lags of y_t, the general forecasting function becomes

$$\hat{y}_{T+h|T} = \hat{\alpha}_h + \sum_{j=0}^{w} \hat{\mathbf{B}}_{hj}' \widetilde{\mathbf{F}}_{T-j} + \sum_{j=0}^{p} \hat{\gamma}_{hj} y_{T-j}, \tag{7}$$

where $\hat{y}_{T+h|T}$ is the h-step ahead forecast.

Stock and Watson (2002b) use factor models to forecast macroeconomic variables, measuring both real economic activity and prices. The factor model forecast is compared with other forecasting models, such as autoregressive forecast (AR), vector autoregressive forecast and multivariate leading indicator forecast. Stock and Watson (2002b) consider U.S. monthly series with the total number of possible predictors being 215. They find that for real variables factor models with two factors, or autoregressive factor models with two factors improve forecasting performance the most. For price indices the autoregressive factor models forecasts with one factor are preferred. In a recent paper, Boivin and Ng (2005) point out that two researchers can arrive at different forecasts using factor models, because the factors are estimated differently and/or the forecasting equations are specified differently. Boivin and Ng concentrate on the two leading methods in the literature, the dynamic method of Forni, Hallin, Lippi, and Reichlin (2005) and the static method of Stock and Watson (2002a). Boivin and Ng investigate the sensitivity of the estimates of the factors and the forecasts based on factor models to the dynamics of the factors and the specification of the forecasting equation. Their main findings are that unconstrained modelling of the series of interest tends to give more robust forecasts when the data generating process is unknown, and that the methodology of Stock and Watson (2002a) apparently does have these properties.

2.2 Bayesian model averaging

Bayesian model averaging can be used to combine forecasts from the set of models that can be constructed using various combinations of the predictors. The averaging over many different competing models incorporates model as well as parameter uncertainty into conclusions about parameters and predictions. For a BMA overview and literature references see Hoeting, Madigan, Raftery, and Volinsky (1999).

Given a set $\mathfrak{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$ of possible models, prior probabilities of the models, $p(\mathcal{M}_i)$, prior distribution of the parameter in each model, $p(\boldsymbol{\theta}_i | \mathcal{M}_i)$ and likelihoods, $L(\mathbf{y} | \boldsymbol{\theta}_i, \mathcal{M}_i)$ all quantities of interest for model averaging and selection can be obtained by using Bayes rule. The posterior probabilities of the models are given by

$$p(\mathcal{M}_i | \mathbf{y}) = \frac{m(\mathbf{y} | \mathcal{M}_i) p(\mathcal{M}_i)}{\sum_{j=1}^{M} m(\mathbf{y} | \mathcal{M}_j) p(\mathcal{M}_j)} = \left[\sum_{j=1}^{M} \frac{m(\mathbf{y} | \mathcal{M}_j) p(\mathcal{M}_j)}{m(\mathbf{y} | \mathcal{M}_i) p(\mathcal{M}_i)} \right]^{-1}, \tag{8}$$

where $m(\mathbf{y} | \mathcal{M}_i)$ is the marginal likelihood

$$m(\mathbf{y} | \mathcal{M}_i) = \int L(\mathbf{y} | \boldsymbol{\theta}_i, \mathcal{M}_i) p(\boldsymbol{\theta}_i | \mathcal{M}_i) \, \mathrm{d}\boldsymbol{\theta}_i, \tag{9}$$

for model $i = 1, \ldots, M$. The posterior distribution of some quantity of interest, ϕ, when taking account of model uncertainty, is

$$p(\phi | \mathbf{y}) = \sum_{j=1}^{M} p(\phi | \mathbf{y}, \mathcal{M}_j) p(\mathcal{M}_j | \mathbf{y}), \tag{10}$$

which is an average of the posterior distribution under each of the models, weighted by their posterior model probabilities. In particular, the minimum mean squared error forecast is given by

$$\hat{y}_{T+h|T} = \mathrm{E}(y_{T+h} | \mathbf{y}) = \sum_{j=1}^{M} \mathrm{E}(y_{T+h} | \mathbf{y}, \mathcal{M}_j) p(\mathcal{M}_j | \mathbf{y}), \tag{11}$$

where $\mathrm{E}(y_{T+h} | \mathbf{y}, \mathcal{M}_j)$ is the forecast conditional on model $\mathcal{M}_j$. This forecast is a special case of forecast combination with weights w_j

$$\hat{y}_{T+h|T} = \sum_{j=1}^{M} \hat{y}_{T+h,j|T} w_j, \tag{12}$$

where BMA provides optimal weights under the assumptions of the forecasting exercise.

2.3 The parameter prior and the posterior distributions

Consider a linear model with k regressors

$$y_{t+h} = \mathbf{z}_t \boldsymbol{\gamma}_h + \varepsilon_{t+h}, \tag{13}$$

where $\boldsymbol{\gamma}_h = (\alpha_h, \boldsymbol{\beta}_h')'$, is a $k+1$ parameter vector and $\mathbf{z}_t = (\iota, \mathbf{x}_t')$ is a vector of explanatory variables.

A challenging task in BMA and model selection is the specification of the prior distribution for the parameters $\boldsymbol{\gamma}$ in the different models. The posterior model probabilities (8) depend on the prior for the model parameters. Due to the large number of models it is desirable to use priors in an automated fashion. The priors should be relatively uninformative and also robust in the sense that conclusions are qualitatively insensitive to reasonable changes in the priors. A common choice in BMA for the class of the normal linear model is the g-prior of Zellner (1986) for the regression parameters,

$$p\left(\boldsymbol{\gamma} \left| \sigma^2, \mathcal{M}\right.\right) \sim \mathrm{N}_{k+1}\left(\mathbf{0}, c\sigma^2 \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right) \tag{14}$$

that is, the prior mean is set to zero indicating shrinkage of the posterior towards zero and the prior variance is proportional to the data information. Improper priors can be used on the parameters that have identical interpretation across all models. In the case of a linear regression model we can use the usual uninformative prior for the variance,

$$p\left(\sigma^2\right) \propto 1/\sigma^2. \tag{15}$$

These priors lead to a proper posterior on the regression parameters, which are t-distributed with T degrees of freedom,

$$p\left(\boldsymbol{\gamma} \middle| \mathbf{y}\right) \sim \mathrm{t}_{k+1}\left(\boldsymbol{\gamma}_1, S, \mathbf{M}, T\right), \tag{16}$$

where

$$\boldsymbol{\gamma}_1 = \frac{c}{c+1}\hat{\boldsymbol{\gamma}}, \tag{17}$$

is a scaled down version of the least squares estimate, and

$$S = \frac{c}{c+1}\left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\right)'\left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}\right) + \frac{1}{c+1}\mathbf{y}'\mathbf{y}, \tag{18}$$

$$\mathbf{M} = \frac{c+1}{c}\mathbf{Z}'\mathbf{Z}. \tag{19}$$

The marginal likelihood is a multivariate t-distribution

$$m\left(\mathbf{y} \middle| \mathcal{M}\right) \propto (c+1)^{-(k+1)/2} S^{-T/2}. \tag{20}$$

The prior (14) requires only specification of the hyperparameter c. Our prior is similar to the one advocated by Fernández, Ley, and Steel (2001), the essential difference being that they use an improper prior for both the constant term and the variance. In a rather extensive study Fernández, Ley, and Steel (2001) investigate various choices of c. Their recommendation is to set the hyperparameter to

$$c = \begin{cases} N^2 & \text{if } T \leq N^2, \\ T & \text{if } T > N^2. \end{cases} \tag{21}$$

2.3.1 The model prior and the model space

A second challenge arises with the size of the model space. All possible combinations of N potential predictors result in 2^N models. Traversing the complete model space, calculating the posterior probabilities, BMA forecasts and the posterior inclusion probabilities of the variables is thus impractical. A convenient method to identify a set of models with non-negligible posterior model probabilities without examining the full model space, is the reversible jump Markov chain Monte Carlo algorithm, see Green (1995). The details of the algorithm are given as Algorithm 1.

This Markov chain converges to the posterior model probabilities under quite general conditions and provides one way of estimating $p(\mathcal{M}|\mathbf{y})$. The estimated posterior model probabilities (8) are (for obvious reasons) conditional on the set of models visited by the chain. To verify that the Markov chain captures most of the total posterior probability mass the method suggested by George and McCulloch (1997) can be used. This method utilizes two separate Markov chains, each starting at a random model. The secondary chain is run for a predetermined number of steps and is then used to provide a capture-recapture type estimate of the total visited probability for the primary chain.

The number of models that enter the model averaging can be further reduced by imposing restrictions on the high-dimensional model space. Being uninformative about the model space results in all models having equal probability and then unrealistically large models are included in the average. Instead, a model prior that downweights models containing a large number of predictors can be used

$$p(\mathcal{M}_i) \propto \delta^{k_i} (1-\delta)^{N-k_i}, \tag{24}$$

where k_i is number of predictors in a model $\mathcal{M}_i$. Setting $\delta = 0.5$ is equivalent to a constant model prior

$$p(\mathcal{M}_i) = p_i = \frac{1}{M}, \quad i = 1, 2, \ldots, M. \tag{25}$$

2.4 Bayesian model averaging with factor models

Koop and Potter (2004) use Bayesian techniques to select factors in dynamic factor models as well as BMA to average forecasts over model specifications. They consider

Algorithm 1 Reversible jump Markov chain Monte Carlo

Suppose that the Markov chain is at model $\mathcal{M}$, having parameters $\boldsymbol{\theta}_{\mathcal{M}}$, where $\boldsymbol{\theta}_{\mathcal{M}}$ has dimension $\dim(\boldsymbol{\theta}_{\mathcal{M}})$.

1. Propose a jump from model $\mathcal{M}$ to a new model $\mathcal{M}'$ with probability $j(\mathcal{M}'|\mathcal{M})$.

2. Generate vector $\mathbf{u}$ (which can have different dimension than $\boldsymbol{\theta}_{\mathcal{M}'}$) from a specified proposal density $q(\mathbf{u}|\,\boldsymbol{\theta}_{\mathcal{M}}, \mathcal{M}, \mathcal{M}')$.

3. Set $(\boldsymbol{\theta}_{\mathcal{M}'}, \mathbf{u}') = g_{\mathcal{M},\mathcal{M}'}(\boldsymbol{\theta}_{\mathcal{M}}, \mathbf{u})$, where $g_{\mathcal{M},\mathcal{M}'}$ is a specified invertible function. Hence $\dim(\boldsymbol{\theta}_{\mathcal{M}}) + \dim(\mathbf{u}) = \dim(\boldsymbol{\theta}_{\mathcal{M}'}) + \dim(\mathbf{u}')$. Note that $g_{\mathcal{M},\mathcal{M}'} = g^{-1}_{\mathcal{M}',\mathcal{M}}$.

4. Accept the proposed move with probability

$$\alpha = \min\left\{1, \frac{L(\mathbf{y}|\boldsymbol{\theta}_{\mathcal{M}'}, \mathcal{M}')\, p(\boldsymbol{\theta}_{\mathcal{M}'}|\mathcal{M}')\, p(\mathcal{M}')\, j(\mathcal{M}|\mathcal{M}')}{L(\mathbf{y}|\boldsymbol{\theta}_{\mathcal{M}}, \mathcal{M})\, p(\boldsymbol{\theta}_{\mathcal{M}}|\mathcal{M})\, p(\mathcal{M})\, j(\mathcal{M}'|\mathcal{M})} \times \frac{q(\mathbf{u}'|\boldsymbol{\theta}_{\mathcal{M}'}, \mathcal{M}', \mathcal{M})}{q(\mathbf{u}|\boldsymbol{\theta}_{\mathcal{M}}, \mathcal{M}, \mathcal{M}')} \left|\frac{\partial g_{\mathcal{M},\mathcal{M}'}(\boldsymbol{\theta}_{\mathcal{M}}, \mathbf{u})}{\partial(\boldsymbol{\theta}_{\mathcal{M}}, \mathbf{u})}\right|\right\}. \quad (22)$$

5. Set $\mathcal{M} = \mathcal{M}'$ if the move is accepted.

If all parameters of the proposed model are generated directly from a proposal distribution, then $(\boldsymbol{\theta}_{\mathcal{M}'}, \mathbf{u}') = (\mathbf{u}, \boldsymbol{\theta}_{\mathcal{M}})$ with $\dim(\boldsymbol{\theta}_{\mathcal{M}}) = \dim(\mathbf{u}')$ and $\dim(\boldsymbol{\theta}_{\mathcal{M}'}) = \dim(\mathbf{u})$, and the Jacobian is unity. If, in addition, the proposal $q(\mathbf{u}|\boldsymbol{\theta}_{\mathcal{M}}, \mathcal{M}, \mathcal{M}')$ is the posterior $p(\boldsymbol{\theta}_{\mathcal{M}'}|\mathbf{y}, \mathcal{M}')$ then (22) simplifies to

$$\alpha = \min\left\{1, \frac{m(\mathbf{y}|\mathcal{M}')\, p(\mathcal{M}')\, j(\mathcal{M}|\mathcal{M}')}{m(\mathbf{y}|\mathcal{M})\, p(\mathcal{M})\, j(\mathcal{M}'|\mathcal{M})}\right\}. \quad (23)$$

This implies that we do not need to perform steps 2 and 3 of the algorithm. Two types of model changing moves are considered:

1. Draw a variable at random and exclude it from the model if it is already in the model, otherwise add it. This step is attempted with probability p_A.

2. Swap a randomly selected variable in the model for a randomly selected variable outside the model. This step is attempted with probability $1 - p_A$.

a number of different model priors and evaluate their forecasting performance, and the in-sample and out-of-sample performance of model selection, model averaging and factor models. A common set of variables, the constant term and lags of the dependent variable, are included in each model. These are assigned flat priors and marginalized out in the same way as the constant is treated by Fernández, Ley, and Steel (2001). The basic model (13) still applies with suitably transformed dependent and explanatory variables. The priors (14) and (15) are used for the reduced model. To apply the BMA approach on the factor models, the regressors $\mathbf{z}_t$ are transformed to principal components using the orthogonal transformation $\mathbf{W} = \mathbf{ZE}$, where $\mathbf{E}$ is the matrix of eigenvectors of $\mathbf{Z}'\mathbf{Z}$. The model with orthogonal regressors is then

$$y_{t+h} = \mathbf{w}_t \boldsymbol{\zeta}_h + \varepsilon_{t+h}, \tag{26}$$

with $\boldsymbol{\zeta}_h = \mathbf{E}^{-1}\boldsymbol{\gamma}_h$. The prior for the regression coefficients becomes

$$p\left(\boldsymbol{\zeta}\left|\sigma^2, \mathcal{M}\right.\right) \sim \mathrm{N}_{k+1}\left(\mathbf{0}, c\sigma^2\left(\mathbf{E}'\mathbf{Z}'\mathbf{Z}\mathbf{E}\right)^{-1}\right) \tag{27}$$

yielding the posterior

$$p\left(\boldsymbol{\zeta}\right|\mathbf{y}) \sim \mathrm{t}_{k+1}\left(\boldsymbol{\zeta}_1, S, \mathbf{M}, T\right), \tag{28}$$

with $\mathbf{Z}$ and $\boldsymbol{\gamma}$ replaced accordingly by $\mathbf{W}$ and $\boldsymbol{\zeta}$, respectively in the equation (17). The use of orthogonalized regressor has the practical advantage that the computational effort is reduced compared to BMA or Bayesian variable selection with non-orthogonal regressors. Since $\mathbf{W}'\mathbf{W}$ is diagonal the marginal likelihood ratio in (23) simplifies and depends only on the model variables that are unique to either model. The ratios can thus be easily precomputed for the case when the models only differ by one or two variables.

Koop and Potter (2004) focus on forecasting the growth rates of US GDP and inflation using a set of 162 predictors. They conclude that BMA forecasts improve on an AR(2) benchmark forecasts at short, but not at longer horizons and only by a small margin. These findings are attributed to the presence of structural instability and the fact that lags of dependent variable seem to contain most of the information relevant for forecasting. Koop and Potter investigate also the forecasting performance of several model priors. They found that priors, which focus on principal components explaining 99.9% of the variance of the predictors, provide the best results, and that the non-informative prior (25) performs very poorly.

2.5 Median probability model

In addition we take the opportunity to apply a method proposed by Barbieri and Berger (2004). This method does not directly deal with the situation of having many predictors, but it is of some interest and is easily implemented since the posterior model probabilities are available from the BMA exercise.

Barbieri and Berger show that for selection among linear models the optimal predictive model is often the median probability model, which is defined as the model consisting of variables that have overall posterior inclusion probability of at least 1/2. The posterior inclusion probability for a variable i is given by

$$p\left(x_i|\,\mathbf{y}\right) = \sum_{j=1}^{M} \mathit{1}\left(x_i \in \mathcal{M}_j\right) p\left(\mathcal{M}_j|\,\mathbf{y}\right), \tag{29}$$

where $\mathit{1}\left(x_i \in \mathcal{M}_j\right)$ equals one if x_i is included in model $\mathcal{M}_j$ and zero otherwise. It is possible that no variable has a posterior inclusion probability exceeding 1/2. The median probability model is, however, assured to exist in two important cases, one is the problem of variable selection, when any variable can be included or excluded from the model, and the other case is when the models under consideration follow a graphical model structure, for example a sequence of nested models. Barbieri and Berger show that the median probability model will frequently coincide with the highest posterior probability model. One obvious situation is when there is a model with posterior probability higher than 1/2. Other situations include the problem of variable selection under an orthogonal design matrix, certain prior structures and known variance σ^2. In Barbieri and Berger's (2004) experience the median probability model outperforms the maximum probability model in terms of predictive performance. They suggest that the median probability model should routinely be determined and reported as a complement to the maximum probability model.

3 Forecast comparison

We explore the performance of the methods mentioned in the previous section on three different datasets, and compare their performance through the root mean square forecast error (RMSFE). The first dataset is the balanced U.S. monthly dataset of Stock and Watson (2002b) consisting of 146 series from 1960:01 to 1998:12. The second dataset consists of 161 quarterly U.S. time series from 1959Q1 until 2000Q1 and was used in Stock and Watson (2003) and in Koop and Potter (2004). The last dataset is a Swedish dataset comprising of 77 variables including a wide range of indicators of real and monetary aspects of the Swedish economy ranging from 1983Q1 to 2003Q4.[1] We forecast the CPI or the inflation rate for all three datasets, and for the U.S. quarterly data we forecast the GDP growth rate as well. The forecasting model is

$$y_{t+h} = \alpha_h + \mathbf{x}_t \boldsymbol{\beta}_h + \varepsilon_{t+h}, \tag{30}$$

which generalizes (7), (13) and (26) to arbitrary forecasting horizons. The choice of dependent variable as y_{t+h}, instead of y_t, has the great advantage that it does away

[1] See the cited works for the list of the predictors for the US data and Appendix A for the list of the Swedish series and their transformation.

with the need of forecasting the predictors in $\mathbf{x}_t$ when forecasting y_{t+h}. The obvious disadvantage of this choice of dependent variable is that it leads to a different model for each forecast horizon.

Following Stock and Watson the dependent variable and the predictors in the U.S. datasets are transformed into stationary series. In particular, GDP is modelled as being $I(1)$ in logarithms and CPI as $I(2)$ in logarithms. This implies that y_{t+h} for GDP and CPI are transformed as

$$y_{t+h} = a/h \cdot \ln\left(GDP_{t+h}/GDP_t\right), \tag{31}$$

$$y_{t+h} = a/h \cdot \ln\left(CPI_{t+h}/CPI_t\right) - a\ln\left(CPI_t/CPI_{t-1}\right), \tag{32}$$

where $a = 400$ for quarterly data and $a = 1200$ for monthly data.

The Swedish inflation rate is measured as the four-quarter percentage change in the consumer price index and the remaining variables in the dataset are with few exceptions 4 quarter growth rates or 4 quarter log differences. The current level of inflation is included in the set of predictor variables for inflation h-periods ahead. A dummy variable d_t, for the Swedish low inflation regime dated to start in 1992Q1, is always included in the model (30).

The forecasts are constructed for horizons $h = 6$ and $h = 12$ for the monthly data, and for horizons $h = 4$ and $h = 8$ for the quarterly data, respectively. For the U.S. monthly data we evaluate the forecast performance for the period 1989:01 until 1998:12. To investigate the possibility that some forecast work well in some periods but poor during others, we also calculate the forecast for four 30 months sub-periods.

For the factor model forecasts we consider the following variants proposed by Stock and Watson: the forecasts denoted by FM-AR,Lag, based on equation (7), include v estimated factors, w lags of factors and p lags of y_t, where all the lag lengths are determined by the Bayesian information criterion (BIC). The FM-AR forecasts contain no lags of $\hat{\mathbf{F}}_t$, with v and p determined by BIC. Finally, FM forecasts contain only contemporaneous $\hat{\mathbf{F}}_t$, with v selected by BIC. Further, forecasts based on the estimated factors holding the number of factors v fixed are also considered, first determining the lag length of the dependent variable, p, by BIC, and then setting $p = 0$. These are denoted as FM-AR,v, and FM,v, respectively.

Our implementation of BMA in dynamic factor models differs slightly from the implementation in Koop and Potter (2004). We use only contemporaneous values when forming the principal components, in particular we do not include lags of y. We consider forecasts based on two different sets of predictors. The BMA-FM forecasts use only the 20 first principal components as the potential predictors and the BMA-FM-AR augments the set of predictors with p lags of y_t. The use of the 20 first principal components roughly corresponds to the 99.9% prior that Koop and Potter find to work well.

Table 1 Summary of datasets used and settings for forecasting

	U.S. monthly dataset	U.S. quarterly dataset	Swedish quarterly dataset
Total no. of variables, N	146	161	77
Forecast variable, y_t	CPI	CPI, GDP	Inflation
Data span	1960:01-1998:12	1961Q1-2000Q4	1983Q1-2003Q4
Forecast period	1989:01-1998:12	1981Q1-2000Q4	1999Q1-2003Q4
Forecast horizons, h	6,12	4,8	4,8
No. of sub-periods	4	4	2
No. of lags of y_t, p	$[0, 5]$	$[0, 3]$	$[0, 3]$
No. of factors, FM, v	$[1, 4]$	$[1, 4]$	$[1, 4]$
No. of factors, FM-AR, v	$[1, 12]$	$[1, 4]$	$[1, 4]$
No. of factor lags, w	$[0, 2]$	$[0, 2]$	$[0, 5]$
δ in prior (24) in BMA	0.075	0.05	0.1
No. of factors, BMA-FM	20	20	20
δ in prior (24) in BMA-FM	0.5	0.5	0.5
MCMC replicates	5 000 000	5 000 000	5 000 000
Burn-in	50 000	50 000	50 000

The forecasts calculated using the Bayesian approach include forecasts based on the forecast combination from all visited models, forecasts from the 3 models with the highest posterior probabilities, denoted Top1 to Top3, and forecasts based on the median model. For the U.S. monthly data we set the model hyperparameter δ to 0.075 in the usual BMA approach. This corresponds to a prior expected model size of 11 variables. For the BMA-FM approach, $\delta = 0.5$, giving expected size of 10 variables. The value of c is chosen as in (21), i.e. $c = N^2$. The Markov chain is run for 5 000 000 steps with 50 000 steps as burn-in. The parameters defining the forecast experiments are summarized in Table 1.

The results from the different approaches are compared to a benchmark, an AR process with the lag length determined by BIC. In addition we calculate forecasts based on the random walk, i.e. when the forecast of y_{t+h} equals the current value of the dependent variable.

3.1 Results

The results for the transformed U.S. CPI series are reported in Tables 2 and 3, Tables 4 - 7 show the results for the U.S. quarterly data, and the results for the Swedish inflation rate can be found in Tables 8 - 9. The first data column in the tables represents results based on the whole forecasting period and the remaining columns contain results for the sub-periods. The values reported in the tables are the relative RMSFEs

$$\frac{\text{RMSFE}\left(\hat{y}_{T+h|T}\right)}{\text{RMSFE}\left(\hat{y}^{AR}_{T+h|T}\right)}. \tag{33}$$

In general, there is no method that consistently outperforms other methods across all periods, datasets or forecasting horizons. There are, however, some patterns in the results that merit further investigation.

It is important to include lags of the dependent variable when forecasting inflation. Only then is it possible to outperform an AR(p) process. This conclusion is supported in as much as 2/3 of the different periods. Koop and Potter (2004) find that an AR(2) process outperforms factor-based models for longer forecasting horizons. The better predictive performance of an AR(2) can, according to Koop and Potter, likely be explained by the fact that the relevant predictive information is included in the lags of the dependent variable.

One possible exception to this is the US GDP forecasts, where the predictive performance of the FM forecasts is unaffected by allowing for lags of the dependent variable. The GDP forecasts from the sub-periods and the whole period also indicate that it is not clear what the number of included factors should be. The best performing model contains different number of factors across the subperiods. Also, chosing the number of factors included in a model in advance often performs better than using BIC for their determination.

Forecast combination using BMA regularly outperforms the forecasts from models selected by the posterior model probabilities.

The results differ substantially between the Swedish dataset, where the BMA-based methods perform poorly, and the two US datasets. Overall BMA-FM does better than the FM forecasts for the US datasets. Allowing for lags of the dependent variable in BMA-FM-AR improves the forecasts somewhat, but the FM-AR forecasts show a larger improvement.

On the issue of selecting predictors from the original variables, or data summaries such as principal components, the evidence is mixed. The BMA forecasts do better for the monthly data and the BMA-FM forecasts better for the quarterly data.

Comparing the median model with the highest posterior probability model fails to prove its superiority for forecasting. The median model produces smaller RMSFE than Top1 model only in 47% of all cases.

4 Conclusions

This paper compares methods for extracting information relevant for forecasting from a large number of predictors. Factor based models, the Bayesian model averaging approach and the combination of the two are evaluated on US and Swedish data at both monthly and quarterly frequencies. We find that none of the methods is uniformly superior and that no method performs better than, or is outperformed by, a simple AR(p) process.

A possible disadvantage of all the methods considered here is that they are based on linear models that forecast h-steps ahead directly. It is quite possible that these simple models fail to capture all the information contained in the data. In future research, more complicated, and thus more realistic functions, will be considered. This could improve forecast accuracy, but comes at the cost of increased computational complexity as the result.

Table 2 RMSFE relative to an AR(p) for monthly U.S. CPI, 6 months ahead forecast.

	89:1-98:12	89:1-91:6	91:7-93:12	94:1-96:6	96:7-98:12
BMA	1.0911	0.8151	0.8986	1.9560	1.8041
Top 1	1.1247	0.7735	0.8989	2.1858	1.9074
Top 2	1.1300	0.7978	0.9103	2.1334	1.9181
Top 3	1.1148	0.8174	0.8443	2.1195	1.8477
Median model	1.1469	0.7967	0.8989	2.2709	1.8950
BMA-FM-AR	1.2189	1.0005	1.3682	1.7506	1.5350
Top 1	1.2435	1.0155	1.3923	1.8061	1.5742
Top 2	1.2418	1.0065	1.3722	1.8172	1.6308
Top 3	1.2313	0.9941	1.4138	1.7883	1.5304
Median model	1.2348	1.0186	1.3715	1.7645	1.5742
BMA-FM	1.4915	1.3485	1.4502	2.1612	1.7257
Top 1	1.5288	1.3430	1.5996	2.2069	1.7404
Top 2	1.5163	1.3575	1.5199	2.2273	1.6904
Top 3	1.5021	1.3344	1.5422	2.1313	1.7432
Median model	1.5283	1.3430	1.5990	2.2044	1.7404
FM-AR, Lag	0.8859	0.7765	0.8140	1.1341	1.3683
FM-AR	0.8998	0.7764	0.9312	1.1005	1.3103
FM	1.4747	1.3808	1.3082	2.0714	1.7743
FM-AR, $v = 1$	0.9068	0.8074	0.8177	1.1450	1.3864
FM-AR, $v = 2$	0.8998	0.7764	0.9312	1.1005	1.3103
FM-AR, $v = 3$	0.8990	0.7756	0.9278	1.0950	1.3174
FM-AR, $v = 4$	0.9020	0.7766	0.9472	1.0902	1.3065
FM, $v = 1$	1.7330	1.7039	1.4604	2.2050	2.0197
FM, $v = 2$	1.7249	1.6917	1.4193	2.2327	2.0553
FM, $v = 3$	1.6954	1.6390	1.4651	2.2308	1.9917
FM, $v = 4$	1.5412	1.4061	1.2932	2.2986	2.0215
Random walk	2.6683	2.4604	2.8274	3.1155	3.1363
AR, RMSFE	0.0053	0.0083	0.0050	0.0031	0.0031

Table 3 RMSFE relative to an AR(p) for monthly U.S. CPI, 12 months ahead forecast.

	89:1-98:12	89:1-91:6	91:7-93:12	94:1-96:6	96:7-98:12
BMA	1.2098	0.9203	0.9186	1.9561	2.0671
Top 1	1.2541	0.9288	1.0022	1.7705	2.1715
Top 2	1.2765	0.9460	1.0617	2.2253	1.9353
Top 3	1.2147	1.0089	0.8740	1.9053	2.1088
Median model	1.2300	0.9074	0.9534	1.8035	2.1640
BMA-FM-AR	1.2489	1.4686	0.8855	2.2020	1.4674
Top 1	1.2830	1.4890	0.8585	2.4392	1.5491
Top 2	1.2734	1.4655	0.8890	2.2842	1.5670
Top 3	1.3079	1.4671	0.9498	2.2635	1.6436
Median model	1.2769	1.4890	0.8917	2.2423	1.5654
BMA-FM	1.4063	1.6903	1.0391	2.4939	1.4448
Top 1	1.3969	1.6589	1.0378	2.5728	1.3888
Top 2	1.4601	1.6750	1.1153	2.7405	1.4501
Top 3	1.3887	1.7042	0.9846	2.4904	1.4453
Median model	1.3940	1.6551	1.0339	2.5703	1.3893
FM-AR, Lag	0.9472	1.1032	0.5940	1.3414	1.4666
FM-AR	0.9500	1.0960	0.6902	1.3328	1.3079
FM	1.4225	1.7191	1.0590	2.4182	1.4948
FM-AR, $v = 1$	0.9383	1.0845	0.6327	1.3495	1.3760
FM-AR, $v = 2$	0.9230	1.0558	0.6621	1.2532	1.3227
FM-AR, $v = 3$	0.9250	1.0439	0.6244	1.3119	1.4021
FM-AR, $v = 4$	0.9229	1.0570	0.6390	1.2492	1.3690
FM, $v = 1$	1.6768	2.0189	1.3410	2.5525	1.6960
FM, $v = 2$	1.6748	1.9993	1.3237	2.6328	1.7292
FM, $v = 3$	1.6424	1.9722	1.2235	2.8380	1.7192
FM, $v = 4$	1.4618	1.7028	1.0418	2.6525	1.6846
Random walk	2.6668	2.5597	2.5628	4.1441	2.4498
AR, RMSFE	0.0106	0.0108	0.0157	0.0051	0.0076

Table 4 RMSFE relative to an AR(p) for quarterly U.S. CPI, 4 quarters ahead forecast.

	81:1-00:4	81:1-85:4	86:1-90:4	91:1-95:4	96:1-00:4
BMA	0.8118	0.7582	0.8570	0.7386	1.2992
Top 1	0.8814	0.8313	0.9189	0.8344	1.3409
Top 2	0.9429	0.9030	0.9680	0.8794	1.3878
Top 3	0.8782	0.8369	0.9565	0.6481	1.3371
Median model	0.9376	0.9326	0.8691	0.8896	1.3728
BMA-FM-AR	0.8113	0.7122	0.9516	0.8462	1.1386
Top 1	0.8610	0.7881	0.9754	0.8619	1.1308
Top 2	0.8635	0.7488	1.0365	0.8986	1.1863
Top 3	0.8256	0.7370	0.9370	0.8918	1.1373
Median model	0.8899	0.8135	1.0179	0.8619	1.1833
BMA-FM	0.8037	0.7012	0.9480	0.8463	1.1278
Top 1	0.8301	0.7344	0.9708	0.8593	1.1377
Top 2	0.8420	0.7041	1.0480	0.9001	1.1603
Top 3	0.8195	0.7235	0.9346	0.8819	1.1859
Median model	0.8088	0.6732	1.0084	0.8614	1.1378
FM-AR, Lag	0.7029	0.5199	0.9625	0.7350	1.0822
FM-AR	0.7605	0.6579	0.9277	0.7350	1.0822
FM	0.8852	0.8395	1.0013	0.7454	1.1028
FM-AR, $v = 1$	0.8039	0.7461	0.9546	0.5489	1.1170
FM-AR, $v = 2$	0.8116	0.7544	0.9729	0.5163	1.1152
FM-AR, $v = 3$	0.8080	0.7445	0.9843	0.5076	1.1033
FM-AR, $v = 4$	0.7416	0.6219	0.9277	0.7350	1.0822
FM, $v = 1$	0.9676	0.9040	1.0914	0.9737	1.0905
FM, $v = 2$	0.9042	0.8457	1.0223	0.8615	1.0878
FM, $v = 3$	0.8552	0.8290	0.9620	0.5962	1.0741
FM, $v = 4$	0.8383	0.7562	1.0013	0.7454	1.1136
Random walk	1.4910	1.3339	1.6208	2.0157	1.4262
AR, RMSFE	0.0184	0.0289	0.0180	0.0117	0.0077

Table 5 RMSFE relative to an AR(p) for quarterly U.S. CPI, 8 quarters ahead forecast.

	81:1-00:4	81:1-85:4	86:1-90:4	91:1-95:4	96:1-00:4
BMA	0.8469	0.8139	1.0077	0.8152	0.8860
Top 1	1.0033	0.9257	1.2254	1.1174	1.1244
Top 2	0.9542	0.8172	1.3516	1.0963	1.0937
Top 3	0.9726	0.9515	1.1286	0.7770	1.2175
Median model	1.0934	1.0175	1.4188	0.9938	1.3249
BMA-FM-AR	0.7840	0.7343	0.9761	0.7497	0.9600
Top 1	0.8633	0.8264	1.0579	0.7300	1.0673
Top 2	0.8878	0.8926	0.9667	0.6914	0.9918
Top 3	0.8273	0.8023	0.9065	0.8022	1.0280
Median model	0.8194	0.7624	1.0602	0.7175	1.0634
BMA-FM	0.7775	0.7265	0.9689	0.7520	0.9554
Top 1	0.8379	0.7911	1.0564	0.7168	1.0671
Top 2	0.8886	0.8943	0.9611	0.7066	0.9716
Top 3	0.8647	0.8529	0.9557	0.7542	1.0153
Median model	0.8279	0.7899	1.0102	0.7059	1.0671
FM-AR, Lag	0.7368	0.6286	1.1048	0.6695	1.0312
FM-AR	0.7821	0.7276	1.0350	0.6310	1.0310
FM	0.8287	0.7746	1.0765	0.6968	1.0676
FM-AR, $v = 1$	0.8096	0.7539	1.1073	0.5372	1.1077
FM-AR, $v = 2$	0.7942	0.7550	1.0264	0.5092	1.1300
FM-AR, $v = 3$	0.8117	0.7904	1.0201	0.5209	1.0014
FM-AR, $v = 4$	0.7448	0.6703	1.0350	0.6310	1.0310
FM, $v = 1$	0.9387	0.8777	1.1680	0.9187	1.1259
FM, $v = 2$	0.8546	0.8101	1.0164	0.7928	1.1372
FM, $v = 3$	0.7982	0.7801	0.9787	0.5331	1.0014
FM, $v = 4$	0.7814	0.7021	1.0765	0.6968	1.0676
Random walk	1.5620	1.5332	1.6817	1.5093	1.7726
AR, RMSFE	0.0449	0.0757	0.0338	0.0298	0.0173

Table 6 RMSFE relative to an AR(p) for quarterly U.S. GDP, 4 quarters ahead forecast.

	81:1-00:4	81:1-85:4	86:1-90:4	91:1-95:4	96:1-00:4
BMA	1.2063	1.1241	2.0225	1.1881	0.9988
Top 1	1.1099	1.0836	1.7165	0.9905	0.9725
Top 2	1.3137	1.2435	2.2510	1.2081	1.0758
Top 3	1.2749	1.1216	2.4049	1.2816	1.1548
Median model	1.3179	1.1917	2.2045	1.4320	0.9725
BMA-FM-AR	1.0100	0.9199	1.6678	1.0699	0.8370
Top 1	1.0570	0.9495	1.8045	1.1128	0.9399
Top 2	1.0084	0.9092	1.6169	1.0855	0.9616
Top 3	1.0483	0.9517	1.7625	1.0890	0.9312
Median model	1.0215	0.9207	1.7204	1.0653	0.9581
BMA-FM	1.0111	0.9480	1.5730	1.0406	0.8266
Top 1	1.0107	0.9573	1.4718	1.0436	0.8778
Top 2	1.0338	0.9575	1.6618	1.0703	0.8599
Top 3	1.0771	0.9822	1.7482	1.1734	0.7916
Median model	1.0098	0.9572	1.4806	1.0386	0.8662
FM-AR, Lag	1.0031	0.9422	1.6103	1.0319	0.6892
FM-AR	0.9486	0.8780	1.4170	1.0319	0.7901
FM	0.9486	0.8780	1.4170	1.0319	0.7901
FM-AR, $v = 1$	0.9828	0.9901	0.9047	0.9879	0.9620
FM-AR, $v = 2$	0.9323	0.8571	1.3808	1.0319	0.7901
FM-AR, $v = 3$	0.9485	0.8473	1.6461	0.9949	0.8564
FM-AR, $v = 4$	0.9527	0.8575	1.6932	0.9422	0.9254
FM, $v = 1$	0.9828	0.9901	0.9047	0.9879	0.9620
FM, $v = 2$	0.9323	0.8571	1.3808	1.0319	0.7901
FM, $v = 3$	0.9485	0.8473	1.6461	0.9949	0.8564
FM, $v = 4$	0.9527	0.8575	1.6932	0.9422	0.9254
Random walk	1.3654	1.4445	1.1553	1.2466	0.9992
AR, RMSFE	0.0198	0.0325	0.0099	0.0179	0.0097

Table 7 RMSFE relative to an AR(p) for quarterly U.S. GDP, 8 quarters ahead forecast.

	81:1-00:4	81:1-85:4	86:1-90:4	91:1-95:4	96:1-00:4
BMA	1.0405	0.8242	2.4397	1.1525	1.5025
Top 1	1.0523	0.7986	2.3529	1.2145	1.6886
Top 2	1.1186	1.0393	2.5036	1.0022	1.2828
Top 3	1.2311	0.9005	3.0433	1.5464	1.3403
Median model	1.0401	0.8134	2.4967	1.2126	1.2583
BMA-FM-AR	0.9615	0.7761	2.0866	1.1358	1.1387
Top 1	1.0726	0.8735	2.4169	1.2232	1.3094
Top 2	1.0054	0.8168	2.1116	1.1595	1.3516
Top 3	1.0520	0.8435	2.4308	1.2215	1.2274
Median model	1.0428	0.8499	2.2900	1.2102	1.2358
BMA-FM	0.9603	0.7701	2.0662	1.1547	1.1015
Top 1	1.0562	0.8820	2.2669	1.2049	1.2131
Top 2	0.9937	0.7788	2.3552	1.1686	1.1752
Top 3	1.0467	0.8391	2.2653	1.2503	1.2316
Median model	1.0399	0.8352	2.2969	1.2299	1.2157
FM-AR, Lag	1.0480	0.9597	2.2734	1.0671	0.8531
FM-AR	0.9986	0.9264	1.8114	1.0671	0.8531
FM	0.9986	0.9264	1.8114	1.0671	0.8531
FM-AR, $v = 1$	1.0218	1.0313	0.9943	1.0025	1.0042
FM-AR, $v = 2$	0.9608	0.8950	1.4641	1.0671	0.8531
FM-AR, $v = 3$	0.9704	0.8799	1.8158	1.0645	0.8734
FM-AR, $v = 4$	0.9909	0.9141	1.8340	1.0593	0.8706
FM, $v = 1$	1.0218	1.0313	0.9943	1.0025	1.0042
FM, $v = 2$	0.9608	0.8950	1.4641	1.0671	0.8531
FM, $v = 3$	0.9704	0.8799	1.8158	1.0645	0.8734
FM, $v = 4$	0.9909	0.9141	1.8340	1.0593	0.8706
Random walk	1.3643	1.3344	2.6074	1.2800	0.9730
AR, RMSFE	0.0340	0.0557	0.0120	0.0339	0.0151

Table 8 RMSFE relative to an AR(p) for quarterly Swedish inflation rate, 4 quarters ahead forecast.

	99:1-03:4	99:1-01:2	01:3-03:4
BMA	1.8653	1.8252	1.9077
Top 1	2.0525	1.8655	2.2372
Top 2	1.7154	1.5059	1.9164
Top 3	2.2446	2.3128	2.1684
Median model	2.1348	1.9876	2.2833
BMA-FM-AR	3.7170	4.8758	1.7449
Top 1	3.7621	4.9464	1.7314
Top 2	3.9111	5.2129	1.5651
Top 3	3.8674	5.1130	1.6903
Median model	3.8101	5.0274	1.6970
BMA-FM	3.7880	5.0065	1.6604
Top 1	3.9833	5.2662	1.7414
Top 2	4.0304	5.2869	1.8921
Top 3	4.2011	5.5781	1.7559
Median model	4.0192	5.3173	1.7449
FM-AR, Lag	1.2061	1.3896	0.9692
FM-AR	1.2678	1.5859	0.7909
FM	0.7346	0.6250	0.8371
FM-AR, $v = 1$	0.6796	0.5573	0.7909
FM-AR, $v = 2$	0.6678	0.5206	0.7969
FM-AR, $v = 3$	0.6864	0.5148	0.8331
FM-AR, $v = 4$	1.4255	1.7763	0.9038
FM, $v = 1$	0.7346	0.6250	0.8371
FM, $v = 2$	0.6127	0.4593	0.7439
FM, $v = 3$	0.6144	0.4630	0.7442
FM, $v = 4$	0.9607	1.0707	0.8254
Random walk	1.1763	1.2390	1.1046
AR, RMSFE	0.8714	0.8883	0.8541

Table 9 RMSFE relative to an AR(p) for quarterly Swedish inflation rate, 8 quarters ahead forecast.

	99:1-03:4	99:1-01:2	01:3-03:4
BMA	2.4462	1.0398	4.6134
Top 1	2.9144	1.4916	5.2969
Top 2	2.4841	1.3079	4.4822
Top 3	2.6555	1.4714	4.7224
Median model	2.5931	0.9828	4.9694
BMA-FM-AR	2.3955	2.0820	3.1810
Top 1	2.9568	2.2670	4.4709
Top 2	3.1632	2.1864	5.1297
Top 3	3.2356	2.3933	5.0265
Median model	2.5212	2.3200	3.0653
BMA-FM	2.8479	1.8076	4.8192
Top 1	3.1357	1.9874	5.3096
Top 2	2.9875	1.7706	5.1955
Top 3	3.1580	1.9899	5.3608
Median model	2.9583	1.9486	4.9207
FM-AR, Lag	1.0504	0.5611	1.8879
FM-AR	1.0504	0.5611	1.8879
FM	1.1689	0.4917	2.2081
FM-AR, $v = 1$	0.5919	0.4178	0.9481
FM-AR, $v = 2$	0.8329	0.6488	1.2429
FM-AR, $v = 3$	0.7756	0.6011	1.1624
FM-AR, $v = 4$	1.2303	0.6737	2.1957
FM, $v = 1$	0.7033	0.4917	1.1330
FM, $v = 2$	0.6381	0.4496	1.0233
FM, $v = 3$	0.6209	0.4560	0.9694
FM, $v = 4$	1.3026	0.7200	2.3181
Random walk	1.0256	0.8284	1.4808
AR, RMSFE	1.3238	1.6292	0.9223

References

Barbieri, M. M., and J. O. Berger (2004): "Optimal Predictive Model Selection," *The Annals of Statistics*, 32(3), 870–897.

Boivin, J., and S. Ng (2005): "Understanding and Comparing Factor-Based Forecasts," *International Journal of Central Banking*, 1(3), 117–151.

Fernández, C., E. Ley, and M. F. Steel (2001): "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100(2), 381–427.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005): "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100(471), 830–840.

George, E. I., and R. E. McCulloch (1997): "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373.

Green, P. J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82(4), 711–732.

Hoeting, J., D. Madigan, A. E. Raftery, and C. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–417.

Jacobson, T., and S. Karlsson (2004): "Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach," *Journal of Forecasting*, 23(7), 479–496.

Koop, G., and S. Potter (2004): "Forecasting in Dynamic Factor Models Using Bayesian Model Averaging," *Econometrics Journal*, 7(2), 550–565.

Stock, J. H., and M. W. Watson (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167 – 1179.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147 – 162.

——— (2003): "Has the Business Cycle Changed and Why?," in *NBER Macroeconomics Annual 2002*, ed. by M. Gertler, and K. S. Rogoff. The MIT Press.

Zellner, A. (1986): "On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, ed. by P. K. Goel, and A. Zellner, pp. 233–243. North-Holland, Amsterdam.

Appendix A Data

The transformation codes for the time series are

Code	Transformation
1	level
2	4 quarters log difference $(\ln y_t - \ln y_{t-4})$
3	4 quarters growth rate $(y_t - y_{t-4})$
4	4 quarters percentage change $((y_t - y_{t-4})/y_{t-4})$

Table A.1 Financial variables

	Variable	Description	Transf.
1.	GovDebt	Government debt	2
2.	AFGX	Affärsvärlden stock index	2
3.	REPO	Repo rate	1
4.	DISK	Discount rate	1
5.	R3M	3 month money market rate	1
6.	R5Y	5 year government bond rate	1
7.	R10Y	10 year government bond rate	1
8.	GBor	Central government borrowing requirement	1
9.	RsTCW	Short rate (TCW)	1
10.	RlTCW	Long rate (TCW)	1

Table A.2 Exchange rates

	Variable	Description	Transf.
11.	NFX	Effective exchange rate (TCW)	2
12.	RFX	Effective real exchange rate (TCW)	2
13.	USD	SEK/USD exchange rate	2
14.	DEM	SEK/DEM exchange rate	2

Table A.3 Money supply

	Variable	Description	Transf.
15.	M0	Narrow money	2
16.	M3	Broad money	2

Table A.4 Labor costs

	Variable	Description	Transf.
17.	WCSS	Wages incl. social security	2
18.	WgCst	Wages excl. social security	2
19.	WageMM	Hourly wages, mining and manufacturing	2
20.	HLCInd	Hourly labor cost: total industry	2

Table A.5 Population

	Variable	Description	Transf.
21.	PpTot	Total population	2
22.	Pp1664	Share in ages 16-64	2
23.	Pp014	Share in ages 0-14	2
24.	Pp1529	Share in ages 15-29	2
25.	Pp2534	Share in ages 25-34	2
26.	Pp3049	Share in ages 30-49	2
27.	Pp5064	Share in ages 50-64	2
28.	Pp6574	Share in ages 65-74	2
29.	Pp75+	Share 75 and older	2

Table A.6 Labor market variables

	Variable	Description	Transf.
30.	AvJob	# of available jobs	2
31.	LabFrc	# in labor force	2
32.	NLFrc	# not in labor force	2
33.	RelLF	LabFrc/Pp1664	1
34.	Empld	# employed	2
35.	PrvEmp	# privatly employed	2
36.	PubEmp	# publicly employed	2
37.	Av4Wrk	# available for work	2
38.	NA4Wrk	# not available for work	2
39.	NUnemp	# unemployed	2
40.	Unemp	Unemployment	1
41.	U02W	# unemployed < 2 weeks	3
42.	U314W	# unemployed 3 - 14 weeks	3
43.	U1552W	# unemployed 15 - 52 weeks	3
44.	U52W+	# unemployed more than 52 weeks	3
45.	NewJob	New jobs	3

Table A.7 Real activity and Expectations

	Variable	Description	Transf.
46.	IndProd	Industrial production	4
47.	NewCar	New cars	1
48.	NewHouse	New single family houses	1
49.	HourWork	Hours worked	2
50.	GDP	GDP	2
51.	RGDP	Real GDP	2
52.	NAIRU	NAIRU	1
53.	OutGap	Output gap	1
54.	ProdGap	Production gap	1
55.	BCI	Business confidence indicator	1
56.	HExpSWE	Household exp. Swedish economy	1
57.	HExpOwn	Household exp. own economy	1
58.	GDPTCW	TCW-weighted GDP	2

Table A.8 Prices

	Variable	Description	Transf.
59.	InfFor	Foreign CPI (TCW)	4
60.	InfRel	Relative CPI	4
61.	PPP	Real exchange rate	4
62.	Infla	Swedish CPI	4
63.	InfNet	Swedish NPI	4
64.	InfHse	House price index	4
65.	MrtWgh	Weight of mortgage interest in CPI	1
66.	InfUnd	Underlying inflation	4
67.	InfFd	Food component of CPI	4
68.	InfFl	Housing fuel and electricity comp. of CPI	4
69.	InfHWg	Factor price index, housing incl. wages	4
70.	InfCns	Construction cost index	4
71.	InfPrd	Producer price index	4
72.	InfImpP	Import price index	4
73.	InfExp	Export price index	4
74.	InfTCW	TCW-weighted Swedish CPI	4
75.	ExpInf	Households exp. of inflation 1 year from now	1
76.	POilUSD	Oil price, USD	4
77.	POilSEK	Oil price, SEK	4