European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2009/31
DEPARTMENT OF ECONOMICS

FORECASTING LARGE DATASETS WITH BAYESIAN REDUCED RANK MULTIVARIATE MODELS

Andrea Carriero, George Kapetanios and Massimiliano Marcellino

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models

ANDREA CARRIERO,
GEORGE KAPETANIOS
and
MASSIMILIANO MARCELLINO

EUI Working Paper **ECO** 2009/31

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

The author(s)/editor(s) should inform the Economics Department of the EUI if the paper is to be published elsewhere, and should also assume responsibility for any consequent obligation(s).

ISSN 1725-6704

© 2009 Andrea Carriero, George Kapetanios and Massimiliano Marcellino

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models*

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

George Kapetanios
Queen Mary, University of London
g.kapetanios@qmul.ac.uk

Massimiliano Marcellino
European University Institute, IGIER and CEPR
massimiliano.marcellino@eui.eu

This draft: July 2009

Abstract

The paper addresses the issue of forecasting a large set of variables using multivariate models. In particular, we propose three alternative reduced rank forecasting models and compare their predictive performance for US time series with the most promising existing alternatives, namely, factor models, large scale Bayesian VARs, and multivariate boosting. Specifically, we focus on classical reduced rank regression, a two-step procedure that applies, in turn, shrinkage and reduced rank restrictions, and the reduced rank Bayesian VAR of Geweke (1996). We find that using shrinkage and rank reduction in combination rather than separately improves substantially the accuracy of forecasts, both when the whole set of variables is to be forecast, and for key variables such as industrial production growth, inflation, and the federal funds rate. The robustness of this finding is confirmed by a Monte Carlo experiment based on bootstrapped data. We also provide a consistency result for the reduced rank regression valid when the dimension of the system tends to infinity, which opens the ground to use large scale reduced rank models for empirical analysis.

Keywords: Bayesian VARs, factor models, forecasting, reduced rank.

J.E.L. Classification: C11, C13, C33, C53.

*This paper has benefited from several comments and constructive criticisms from Herman van Dijk and two anonymous referees. We also thank seminar participants at Queen Mary University of London, Bank of France, and at the "Large Datasets and Dynamic Factor Models Workshop", October 2007, London.

1 Introduction

Forecasting future developments in the economy is a key element of the decision process in policy making, consumption and investment decisions, and financial planning. While some macroeconomic variables are of particular interest, e.g., GDP growth, inflation or short term interest rates, the attention is more and more focusing on a larger set of indicators, in order to obtain an overall picture of the expected evolution of the economy.

Recently there has been a boost in the developments of econometric methods for the analysis of large datasets, starting with the pioneering work of Forni et al. (2000) and Stock and Watson (2002a, 2002b). The key econometric tool in this context is the factor model, where each of a large set of variables is split into a common component, driven by a very limited number of unobservable factors, and an idiosyncratic component. From a forecasting point of view, the idea is to use the estimated factors for predicting future developments in, possibly, all the many variables under analysis. In practice, factor models have produced fairly accurate forecasts when compared with standard benchmarks, such as AR or VAR based predictions, for several countries and different macroeconomic variables, see e.g. the meta analysis in Eickmeier and Ziegler (2006).

The good performance of factor models has stimulated a search for alternative methods with further enhanced predictive power, see e.g. the overview in Stock and Watson (2006). These can be classified into methods for variable selection, such as LASSO (Tibshirani,1996, De Mol et al. 2006), or boosting (Bai and Ng 2009, Bühlmann, 2006, Lutz and Bühlmann 2006), or bagging (Breiman 1996, Buhlmann and Yu 2002, Inoue and Kilian 2004); Shrinkage estimators, such as ridge regression (De Mol et al. 2006) or Bayesian VARs in the spirit of Doan, Litterman and Sims (1984) (e.g. Banbura et al., 2007); and pooling procedures, where a large set of forecasts from alternative, possibly small scale, models are combined together, see e.g. the survey in Timmermann (2006).

Surprisingly, most existing research has used large datasets only as predictors for a small number of key macroeconomic variables, not considering the issue of forecasting all the series in the dataset itself. As a result, most of the contributions cited above are based on a single equation approach. In this paper we focus on forecasting *all* the variables in a large dataset using *multivariate* models. The focus on the multivariate forecasting performance of the methods distinguishes the results of this paper from many others in the large data set literature, although for completeness we also report results for some key macroeconomics variables.

We propose three additional forecasting methods and evaluate their performance in forecasting a large US macroeconomic dataset, comparing them with the most promising

existing alternatives, namely, large scale Bayesian VARs (BVAR), multivariate boosting (MB), and the Stock and Watson (2002a, b) factor model (SW). The proposed forecasting methods are based on rank reduction. Reduced rank regressions have a long history in the time series literature but have been so far only applied in small models, see e.g. Velu et al. (1986), Reinsel (1983), Reinsel and Velu (1998), Camba-Mendez et al. (2003). RR represents a natural extension of the methods proposed so far in the large dataset literature. Hoogerheide and Van Dijk (2001) show how simultaneous equation models and vector error correction models are restricted instances of a general RR. Also factor models can be obtained as a special case of RR.

The first method considered is classical Reduced Rank Regressions (RR) along the lines of Velu et al. (1986). For the RR regression we provide a novel result on consistency and rate of convergence when the number of variables tends to infinity. The consistency result opens the grounds to using large scale RR for empirical analysis. The second method aims at enhancing further the parameter dimensionality reduction needed in large scale VARs by combining reduced rank restrictions with Bayesian shrinkage. In this setting Geweke (1996) has proposed a model which imposes the rank reduction on the prior as well as on the posterior mean. While Geweke (1996) application focuses on small sized system, for the first time we put the model at work within a large data-set setting. We label this method Bayesian Reduced Rank Regression (BRR). Finally we propose a new method which retains the benefit of rank reduction and shrinkage without paying the high computational cost involved in the BRR. The method, which we label RRP (Reduced Rank Posterior) applies rank reduction on the posterior estimates of a Bayesian VAR and as we shall see it can produce substantial gains in forecast accuracy.

Being multivariate, the proposed reduced rank methods are well suited for medium to large datasets of the dimension typically of interest for central banks, i.e. about 50-60 variables. All the methods are potentially suited to deal with larger datasets, but some of them pose serious computational burdens. In particular, as the number of regressors grows, RR can encounter numerical problems in the estimation of the covariance matrix of the unrestricted residuals, MB involves estimating a growing number of candidate models, while BRR requires simulations involving in each step the inversion of larger matrices. For that very reason in our empirical application we use 52 US macroeconomic variables taken from the dataset provided by Stock and Watson (2005). The series have been chosen in order to represent the main categories of indicators which are relevant for central banks in understanding and forecasting developments in the macroeconomy. Basically, we have discarded from the original dataset of Stock and Watson (2005) those variables containing roughly the same information as others, such as the disaggregated

sectoral data on industrial production and prices. These variables are not of particular interest to be forecasted as they are highly collinear, which may also create serious problems in estimation.

We can anticipate that RRP and BRR produce fairly good forecasts, more accurate than those of competing methods on average across several US macroeconomic variables, when measured in the terms of mean square or mean absolute forecast error. Moreover, they also perform well for key variables, such as industrial production growth, inflation and the short term interest rate. This is encouraging evidence that using shrinkage and rank reduction in combination improves substantially the accuracy of forecasts.

The paper is structured as follows. In Section 2 we describe in more details the forecasting models under comparison, with a special focus on the different types of RR. In Section 3 we present the results of the forecast comparison exercise. In Section 4 we assess their robustness and conduct a Monte Carlo experiment with bootstrapped data. Section 5 concludes.

2 Forecasting Models

We are interested in forecasting the N-vector process $Y_t = (y_{1,t}, y_{2,t}, ..., y_{N,t})'$, where N is large, using a Np-dimensional multiple time series of predictors $X_t = (Y_{t-1}, Y_{t-2}, ...Y_{t-p})'$, observed for $t = 1, ..., T$. The baseline model is therefore a VAR(p):

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + e_t, \tag{1}$$

where means and trends have been removed[1]. Defining $B = (A_1, A_2, ...A_p)'$ equation (1) can be compactly written as:

$$Y_t = B' X_t + e_t. \tag{2}$$

It is convenient to rewrite the VAR in (2) as a multivariate regression:

$$Y = XB + E. \tag{3}$$

In equation (3) the observations are by row, and equations by column, so $Y = (Y_1, ..., Y_T)'$ is a $T \times N$ matrix of dependent variables, $X = (X_1, ..., X_T)'$ is a $T \times M$ matrix of explanatory variables, where $M = Np$.

[1] In our application we transform the variables to stationarity and standardize them prior to estimation and forecasting. The forecasts of the original variables are then computed by inverting the transformation and reattributing means and variances. The transformation is computed rollingly, i.e. by using only the data in each of the rolling samples used for estimation. See Section 3.2 for further details.

The matrix E is the matrix of disturbances, which are assumed to be independent and identically distributed across observations; that is, taking $E = (e_1, e_2, ..., e_T)'$, then $e_i \sim IIDN(0, \Sigma)$. We define r as the rank of the $M \times N$ matrix of coefficients B, where of course $r \leq N$.

We focus on 6 forecasting models: reduced rank regression (RR), Bayesian VARs (BVAR), multivariate boosting (MB), Bayesian reduced rank regression (BRR), reduced rank Posterior (RRP), and factor models (SW).

SW and RR are both based on the idea of reducing dimensionality by imposing a structure which summarizes the information contained in a large set of predictors by focussing on some relevant linear combinations of them. An alternative route to obtain a more parsimonious model might be to impose exclusion restrictions on the predictors. However, excluding some variables from a regression is likely to be relatively ad hoc, unless a coherent statistical framework is adopted to do so. BVAR and MB provide a solution to this problem. Finally, BRR and RRP apply both shrinkage and rank reduction. In the latter case the reduced rank is imposed after the estimation of a BVAR has been performed. In the former case, the rank reduction is imposed on the prior as well as on the posterior mean. Each forecasting model is described in detail in the following six subsections.

2.1 Reduced Rank Regression (RR)

It is often the case that estimation of VAR(p) models results in a large number of insignificant coefficients. Therefore, in order to obtain a more parsimonious model, one might impose rank reduction, i.e. to assume that $rk(B') = r < N$. This is equivalent to the parametric specification:

$$Y_t = \alpha \left(\sum_{i=1}^{p} \beta_i' Y_{t-i} \right) + e_t = \alpha \beta' X_t + e_t, \tag{4}$$

where α and $\beta = (\beta_1', ..., \beta_p')'$ are respectively a $N \times r$ and a $M \times r$ matrices. The model (4) was studied by Velu et al. (1986). Ahn and Reinsel (1988) suggested a more general specification where the rank of the coefficient matrix on each lagged vector of the explanatory variables may differ. However, this generalization creates computational problems in the large N case. Therefore, we focus on (4).

In equation (4), it is assumed that the true rank of the matrices α and β is identical and equal to r which is thus referred to as the rank of the system (4). However, note that the ranks of β_i, $i = 1, ..., p$, need not equal r; in particular, it can be $rk(\beta_i) \leq r$,

$i = 1, ..., p$.

An interesting special case of the RR model (4), which resembles the autoregressive index model of Reinsel (1983), results if $\beta_i = \beta_* K_i$ with $rk(\beta_*) = r$ for some (r, r) matrix K_i which need not be full rank, $i = 1, ..., p$, although $K = (K_1', ..., K_p')'$ is. Hence, $\beta = (I_p \otimes \beta_*)K$ and $\alpha\beta_i' = \alpha_i\beta_*'$, where $\alpha_i = \alpha K_i'$, in which case $\beta_*' y_{t-i}$, $i = 1, ..., p$, may be interpreted as dynamic factors for y_t.

Given the assumed system rank r, Velu et al. (1986) suggested an estimation method for the parameters α and β that may be shown to be quasi-maximum likelihood[2] (see also Reinsel and Velu, 1998). Denote the sample second moment matrices by $S_{YY} = T^{-1}Y'Y$, $S_{YX} = T^{-1}Y'X$, $S_{YX} = S_{XY}'$, and $S_{XX} = T^{-1}X'X$. Hence, the covariance matrix of the unrestricted LS residuals, $S_{YY,X} = S_{YY} - S_{YX}S_{XX}^{-1}S_{XY}$ is the unrestricted quasi-ML estimator of the error process variance matrix. Additionally, let $\{\lambda\}_{t=1}^T$, $\lambda_1^2 \geq \lambda_2^2 \geq ... \geq \lambda_N^2 \geq 0$ denote the ordered squared eigenvalues of the $N \times N$ matrix $S_{YY,X}^{-1/2}S_{YX}S_{XX}^{-1}S_{XY}S_{YY,X}^{-1/2}$ with associated eigenvectors $\{v_i\}_{t=1}^T$ subject to the normalization $v_i'v_j = 1$ if $i = j$ and 0 otherwise, and let $\hat{V} = (v_1, v_2, ..., v_r)$. The quasi-ML estimators for α and β are given by $\hat{\alpha} = S_{YY,X}^{1/2}\hat{V}$ and $\hat{\beta} = S_{XX}^{-1}S_{XY}S_{YY,X}^{-1/2}\hat{V}$, so that $\hat{B}' = S_{YY,X}^{1/2}\hat{V}\hat{V}'S_{YY,X}^{-1/2}S_{XX}^{-1}S_{XY}$.

2.1.1 Some consistency results for RR

This section provides some novel theoretical results on the parameter estimates of the Reduced Rank Regression. In particular we provide consistency and rate of convergence results for the coefficients of an infinite dimensional RR model.[3] The proof does not rely on any assumption of a factor structure, and extends to the case with infinite lags. Before stating the main theorem, we need to state a theorem about the consistency of the underlying infinite dimensional VAR.

We make the following assumptions:

Assumption 1 *(a)* $|\lambda_{\max}(A)| < 1$ *where*

$$A = \begin{pmatrix} A_1 & ... & ... & A_p \\ I & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & I & 0 \end{pmatrix} \tag{5}$$

[2] The method is quasi-maximum likelihood as no Gaussian distribution is assumed.

[3] By infinite dimensional we mean that the dimension of the system (i.e. the number of variables) tends to infinity.

and $|\lambda_{\max}(.)|$ *denotes the maximum eigenvalue of a matrix in absolute value.*

(b) $c_{\max}(A) < \infty$, $r_{\max}(A) < \infty$ *where* $c_{\max}(.)$ *and* $r_{\max}()$ *denote the maximum column and row sum norm of a matrix.*

(c) e_t *is an i.i.d.* $(0, \Sigma_e)$ *sequence with uniformly finite fourth moments and* $c_{\max}(\Sigma_e) < \infty$.

Denote the transpose of the i-th row of $(A_1, A_2, ..., A_p)$ by A^i. We then have the following Theorem on the consistency of the infinite dimensional VAR coefficients:

Theorem 1 *As* N *and* T *diverge, and under assumption 1 ,* $\left\|\hat{A}^i - A^i\right\|^2 = o_p(T^{-a})$ *for all* $i = 1, ..., N$, *and for all* $a < 1/2$, *as long as* $N = o\left((T/\ln(T))^{1/2}\right)$.

Proof. See Appendix. ■

Next, we consider a reduced rank approximation to the VAR model. To keep things general, we consider the case where a singular value decomposition is used to decompose $(A_1, A_2, ..., A_p)$ as $\mathcal{O}\mathcal{K}$ where $\mathcal{O}$ and $\mathcal{K}'$ are $N \times r$ and $Np \times r$ matrices respectively, for some $r < N$. The sample counterpart of this decomposition is given by $(\hat{A}_1, \hat{A}_2, ..., \hat{A}_p) = \hat{\mathcal{O}}\hat{\mathcal{K}}$. Then, we have the following Theorem on the consistency of the infinite dimensional RR coefficients:

Theorem 2 *As* N *and* T *diverge, and under assumption 1 , each element of* $\mathcal{O}$ *and* $\mathcal{K}'$ *is* $o_p(T^{-a+2b})$*-consistent for* $\mathcal{O}$ *and* $\mathcal{K}'$, *for all* $0 < a < 1/2$, *and* $0 < b < 1/4$, $2b < a$, *as long as* $N = o\left(T^b\right)$.

Proof. See Appendix. ■

Note that the above analysis straightforwardly implies that a lag order, $p = p_T$, that tends to infinity is acceptable. In this case, the above result holds as long as $Np_T = o\left((T/\ln(T))^{1/2}\right)$, which means that our results extend to infinite dimensional VAR and RR possibly with infinite lags too. The above consistency result opens the grounds to using large dimensional RR models for empirical analysis.

2.2 Bayesian VAR (BVAR)

Bayesian methods allow to impose restrictions on the data, but also to let the data speak. The exclusion restrictions are imposed as priors, so if some a-priori excluded variable turns out to be relevant in the data, the posterior estimate would contain such information. This provides a way of solving the curse of dimensionality problem without resorting to ad-hoc exclusion of some variables.

In this paper we implement a Normal-Inverted Wishart version of the so-called Minnesota prior of Doan et al. (1984) and Litterman (1986). This version of the prior was proposed by Kadiyala and Karlsson (1997) and allows both to gain substantially in terms of computational efficiency and to avoid the inconvenient assumption of fixed and diagonal residual variance matrix. The use of this prior for forecasting macroeconomic variables with large datasets has been recently advocated by Banbura et al. (2007), who however focus on a smaller set of key macroeconomic variables when evaluating forecasting performance. A version of this prior has also been used by Canova and Ciccarelli (2003) to estimate a panel VAR for the G7 countries. However the number of variables in the model estimated by Canova and Ciccarelli (2003) is considerably smaller with respect to our case, as they consider three variables for each G7 country, for a total of 21 variables.

The Minnesota prior shrinks parameter estimates towards a random walk representation and it has proven to be robustly good in forecasting. In particular, the prior expectations and variances of $A_1, A_2, ..., A_p$ under the Minnesota prior are:

$$E[A_k^{(ij)}] = \begin{cases} 1 & for\ j = i,\ k = 1 \\ 0 & otherwise \end{cases},\ V[A_k^{(ij)}] = \begin{cases} \phi\frac{1}{k^2} & for\ j = i,\ \forall\ k \\ \phi\frac{1}{k^2}\theta\sigma_i^2\sigma_j^{-2} & for\ j \neq i,\ \forall\ k \end{cases}, \tag{6}$$

while the residual variance matrix Σ is fixed and diagonal: $diag(\sigma_1^2, ..., \sigma_N^2)$. The hyperparameter ϕ measures the overall tightness of the prior, and we will return to it later in this subsection. The factor $1/k^2$ is the rate at which prior variance decreases with increasing lag length while the ratio σ_i^2/σ_j^2 accounts for the different scale and variability of the data. Finally, the parameter θ imposes additional shrinkage on the coefficients attached to a regressor when it is not a lag of the dependent variable in a given equation.

Kadiyala and Karlsson (1997) propose a version of this prior which allows to avoid the inconvenient assumption of a fixed and diagonal residual variance matrix and to gain substantially in terms of computational efficiency, at the cost of setting $\theta = 1$. The prior has a Normal-Inverted Wishart form:

$$\Sigma \sim iW(v_0, S_0);\ B|\Sigma \sim N(B_0, \Sigma \otimes \Omega_0), \tag{7}$$

where the parameters v_0, S_0, B_0, Ω_0 are such that the expectation of Σ is equal to the fixed residual covariance matrix of the Minnesota prior, and the prior expectation and variance of B is that of the Minnesota prior (with $\theta = 1$). Moreover, as we forecast after transforming variables to get stationarity, we set $E[A_1^{(ii)}] = 0$ rather than $E[A_1^{(ii)}] = 1$ to

be consistent with the random walk assumption on the original variables. This provides us with the following prior expectations and variances for $A_1, A_2, ..., A_p$:

$$E[A_k^{(ij)}] = 0; \quad V[A_k^{(ij)}] = \phi \frac{1}{k^2} \sigma_i^2 \sigma_j^{-2} \tag{8}$$

The hyperparameter ϕ measures the tightness of the prior: when $\phi = 0$ the prior is imposed exactly and the data do not influence the estimates, while as $\phi \to \infty$ the prior becomes loose and the posterior estimates approach the OLS estimates. The conditional posterior distributions are also of the Normal-Inverted Wishart form:

$$\Sigma|Y \sim IW(\bar{v}, \bar{S}), \; B|\Sigma, Y \sim N(\bar{B}, \Sigma \otimes \bar{\Omega}), \tag{9}$$

where the bar denotes that parameters are those of the posterior distribution. Zellner (1973) shows that integrating out Σ it is possible to obtain the marginal posterior distribution of B, which is a matricvariate t: $B|Y \sim MT(\bar{\Omega}^{-1}, \bar{S}, \bar{B}, \bar{v})$ with posterior mean $\bar{B} = (\Omega_0^{-1} + X'X)^{-1}(\Omega_0^{-1}B_0 + X'Y)$. The posterior mean $\bar{B}$ and the other posterior moments can also be obtained by implementing the prior in the form of dummy variable observations.[4]

2.3 Bayesian Reduced Rank Regression (BRR)

The BVAR and RR described in the previous subsections apply respectively shrinkage and rank reduction. Alternatively we could think of imposing both rank reduction and shrinkage on the VAR.

Bayesian analysis of reduced rank regression has been introduced by Geweke (1996). While Geweke (1996) focuses on small sized system, for the first time we put the model at work within a large data-set setting. As for the reduced rank case, the $M \times N$ matrix of coefficients B is assumed to have rank r, where $r < N$. This rank reduction assumption is equivalent to the parametric specification

$$Y = X\Psi\Phi + E \tag{10}$$

[4]In particular, the prior can be implemented by adding T_d dummy observations Y_d and X_d to the system in (3) so that it becomes $Y^* = X^*B + E^*$, where $Y^* = (Y' \; Y_d')'$ and $E^* = (E' \; E_d')'$ are $(T + T_d) \times N$ matrices, and $X^* = (X' \; X_d')'$ is a $(T + T_d) \times M$ matrix. The dummy observations Y_d and X_d have to be chosen such that their moments coincide with the prior moments $B_0 = (X_d'X_d)^{-1}X_d'Y_d$, $\Omega_0 = (X_d'X_d)^{-1}$, $S_0 = (Y_d - X_dB_0)'(Y_d - X_dB_0)$, and $v_0 = T_d - M - N - 1$. The posterior mean $\bar{B}$ is then given by the OLS estimate of the augmented system (given by the usual formula $\bar{B} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$). For details see Banbura et al. (2007) and Kadiyala and Karlsson (1997).

with Ψ and Φ being respectively $M \times r$ and $r \times N$ matrices. To identify these matrices Geweke (1996) proposes the following normalization[5]:

$$\Phi = [I_r \mid \Phi^*]. \tag{11}$$

Given that normalization a proper prior is:

$$\mid \Sigma \mid^{-(N+v_0+1)} \exp\left[-\frac{1}{2} tr S_0 \Sigma^{-1}\right] \exp\left[-\frac{\tau^2}{2}(tr\Phi^{*\prime}\Phi^* + tr\Psi^\prime\Psi)\right], \tag{12}$$

namely a product of an independent Wishart distribution for Σ with v_0 degrees of freedom and matrix parameter S_0, and independent $N(0, \tau^{-2})$ shrinkage priors for each element of the coefficient matrices Φ^* and Ψ. The conditional posterior distribution of Σ is:

$$\Sigma \mid (\Phi^*, \Psi, X, Y) \sim IW[T + v_0, S_0 + (Y - XB)^\prime(Y - XB)]. \tag{13}$$

The conditional posterior distributions of the coefficients Φ^*,Ψ, are multivariate normals.

In particular, the conditional posterior distribution of Φ^* is:

$$vec(\Phi^*) \mid (\Psi, \Sigma, X, Y) \sim N[\Pi_\Phi \cdot vec(\hat{\Phi}^*),\ \Pi_\Phi], \tag{14}$$

where:

$$\begin{aligned} \hat{\Phi}^* &= (\Psi^\prime X^\prime X \Psi)^{-1}\Psi^\prime X^\prime Y_1 \Sigma^{12}(\Sigma^{22})^{-1} - \Sigma^{12}(\Sigma^{22})^{-1} \\ &\quad + (\Psi^\prime X^\prime X \Psi)^{-1}\Psi^\prime X^\prime Y_2, \end{aligned} \tag{15}$$

$$\Pi_\Phi = [(\Sigma^{22})^{-1} \otimes (\Psi^\prime X^\prime X \Psi)^{-1} + \tau^2 I_{r(N-r)}]^{-1}, \tag{16}$$

and where $Y = [Y_1 \mid Y_2]$ is a partitioning of Y into its first r and last $N - r$ columns and where Σ^{ij} denotes the partitioning of Σ^{-1} into its first r and last $N - r$ rows and columns.

The conditional posterior distribution of Ψ is:

$$vec(\Psi) \mid (\Phi, \Sigma, X, Y) \sim N[\Pi_\Psi \cdot vec(\hat{\Psi}),\ \Pi_\Psi], \tag{17}$$

[5]For a discussion of the role of normalization in reduced rank models see e.g. Kleibergen and van Dijk (1994, 1998) and Hamilton et al. (2007).

where:

$$\hat{\Psi} = \hat{B}[\Phi^{+} + \Phi^{0}\tilde{\Sigma}^{21}(\tilde{\Sigma}^{11})^{-1}], \tag{18}$$

$$\Pi_{\Psi} = [\tilde{\Sigma}^{11} \otimes X'X + \tau^{2} I_{Mr}]^{-1}, \tag{19}$$

and where $\hat{B}$ is the OLS estimator, Φ^{+} is the generalized inverse of Φ, Φ^{0} is column-wise orthogonal to Φ^{+}, and where $\tilde{\Sigma}^{ij}$ denotes the partitioning of $\tilde{\Sigma}^{-1} = ([\Phi^{+}\ \Phi^{0}]'\Sigma[\Phi^{+}\ \Phi^{0}])^{-1}$ into its first r and last $N - r$ rows and columns.

Unconditional posterior distributions can be simulated by using a Gibbs sampling algorithm which draws in turn from (14), (17), and (13).[6] See Geweke (1996) for details.

2.4 Reduced Rank BVAR Posterior (RRP)

The BRR has the shortcoming of being computationally challenging when the assumed rank is high, as the estimation of this model requires simulation involving inversion of Mr -dimensional matrices. We propose a new method which retains the benefit of rank reduction and shrinkage without paying the high computational cost involved in the BRR. The method, which we label RRP (Reduced Rank Posterior) applies rank reduction on the posterior estimates of a Bayesian VAR and, as we shall see, it can produce substantial gains in forecast accuracy.

The implementation of the method is straightforward. First, the system is estimated under the prior distribution described by equation (7), then a rank reduction is imposed as follows. Let $\bar{B}$ be the posterior mean of B and let $\bar{B} = U\Lambda V$ be its singular value decomposition. Collecting the largest r singular values and associated vectors in the matrices $\Lambda^{*} = diag(\lambda_1, \lambda_2, ..., \lambda_r)$, $U^{*} = (u_1, u_2, ..., u_r)$ and $V^{*} = (v_1, v_2, ..., v_r)$ a reduced rank approximation (of rank r) of the posterior mean is given by:

$$\bar{B}_r^{*} = U^{*}\Lambda^{*}V^{*}, \tag{20}$$

which is our RRP estimator.

[6] In our application we draw 1100 draws, we drop the firt 100 as burn in and then we keep each second draw to reduce the autocorrelation inherent to the MCMC scheme. This leaves us with a total of 500 valid draws. Computation time (for producing one forecast) varies with the assumed rank of the system. The longest computation time is required by the system with rank r=10 and it takes about 4 minutes on a 3.33GHz Intel Core Duo processor.

2.5 Multivariate Boosting (MB)

The Minnesota prior reduces the dimensionality of the system by setting (a priori) to zero all but one coefficient in each equation. An alternative method to reach parsimony by eliminating some regressors is boosting. Theoretical results for boosting applied to multivariate models have been developed by Lutz and Bühlmann (2006), while its use for macroeconomic forecasting has been recently advocated by Bai and Ng (2009) within a univariate approach.

Boosting is a procedure that estimates an unknown function $f(X_t)$ as a sum of $\bar{m}$ estimated functions $\Sigma_{m=1}^{\bar{m}}\hat{g}^{(m)}$: $\hat{f}(X_t) = \hat{f}^{(0)} + \xi\Sigma_{m=1}^{\bar{m}}\hat{g}^{(m)}$. The estimated functions $\hat{g}^{(m)}$ are derived using a base learner, which is a fitting procedure based on the minimization of some loss function. The algorithm starts with an empty model and than at each iteration it adds the $\hat{g}^{(m)}$ providing the smallest loss. It is clear that boosting can also be viewed as a variable selection algorithm.

If the function of interest is the conditional mean $f(X_t) = E(Y_t \mid X_t)$, the loss function is the residual sum of squares, and least squares regression is used as fitting procedure, then boosting is basically a stepwise regression which starts with the empty model and adds in each step the most significant covariate. In this case the algorithm works roughly as follows. At each iteration the residual is computed as the difference between the actual data and the fitted value up to that iteration. Then this residual is regressed on all candidate regressors taken individually, and the regressor producing the smallest sum of squared residuals is chosen. The fitted value from this regression ($\hat{g}^{(m)}$) is then added to the cumulative sum before moving on to next iteration.

In this paper we use Multivariate Boosting algorithm with quadratic loss function (L_2 Boosting) and componentwise least square base learner. Let $y_{(i)},\ x_{(i)}, y_{(j)},\ x_{(j)}$ denote the i-th row vectors and j-th column vectors of $Y,\ X$. The algorithm works as follows:

- Step 1. Start with the empty model $\hat{f}_j^{(0)}=\bar{Y}_j,\ j = 1, ..., N$
- Step 2. For $m = 1, ..., \bar{m}$
  - a) Compute the "current" residuals $r_{(i)} = y_{(i)} - \hat{f}_i^{(m-1)},\ i = 1, ..., T.$
  - b) Fit the base learner to $r_{(i)}$ and derive $\hat{g}^{(m)},\ i = 1, ..., T.$
    - Regress the "current" residuals $r_{(i)}$ on each regressor $x_{(j)},\ j = 1, ..., M,$ obtaining $\hat{b}_{(ij)}$
    - For each regressor j and time i compute the loss function $SSR(\hat{b}_{(ij)})$

* Pick the regressor j^* and the sample point i^* which minimized the loss function and set $\hat{g}^{(m)} = \hat{b}_{(i^*j^*)}x_{i^*}$

- Step 3. Update $\hat{f}^{(m)} = \hat{f}^{(m-1)} + \xi\hat{g}^{(m)}$, where ξ is a shrinkage parameter.

The loss function used in step 2 is:

$$L(B) = \frac{1}{2}\sum_{i=1}^{T}(r'_{(i)} - x'_{(i)}B)\Gamma^{-1}(r'_{(i)} - x'_{(i)}B)' \tag{21}$$

with $\Gamma^{-1} = I$.

The base learner used in step 2 fits the linear least squares regression with one selected covariate $x_{(j)}$ and one selected pseudo-response $r'_{(i)}$ so that the loss function in (21) is reduced most:

$$\hat{s}, \hat{t} = \underset{1\leq j\leq M, 1\leq k\leq N}{\arg\min}\{L(B); B_{jk} = \hat{\beta}_{jk}, B_{uv} = 0 \ \forall \ u, v \neq j, k\}$$

Thus, the learner fits one selected element of the matrix B as follows:

$$\hat{\beta}_{jk} = \frac{\sum_{v=1}^{N} r'_v x_j \Gamma^{-1}_{vk}}{x'_j x_j \Gamma^{-1}_{kk}}, \tag{22}$$

$$\hat{B}_{\hat{s}\hat{t}} = \hat{\beta}_{\hat{s}\hat{t}}, \ \hat{B}_{jk} = 0 \ \forall \ jk \neq \hat{s}\hat{t}. \tag{23}$$

Corresponding to the parameter estimate there is a function estimate $\hat{g}_\ell(\cdot)$ defined as follows: for $x = (x_1, ..., x_p)$,

$$\hat{g}_\ell(x) = \begin{cases} \hat{\beta}_{\hat{s}\hat{t}} & for \ \ell = \hat{t}, \\ 0 & otherwise, \end{cases} \qquad \ell = 1, ..., N. \tag{24}$$

The algorithm terminates when the specified final iteration $\bar{m}$ is reached. Lutz and Bühlmann (2006) provide a proof that this procedure can handle cases with infinite N and is able to consistently recover sparse high-dimensional multivariate functions.

The use of the shrinkage parameter ξ has been first suggested by Friedman (2001) and is supported by some theoretical arguments (see Efron et al. 2004, and Bühlmann and Yu 2005). The boosting algorithm depends on ξ but its choice is insensitive as long as ξ is taken to be "small" (i.e. around 0.1). On the other hand, the number

of boosting iterations $\bar{m}$ is a much more crucial parameter. Indeed, $\bar{m}$ is a pivotal quantity regulating the trade-off between parsimony and fit: small values of $\bar{m}$ yield very parsimonious specifications, while as $\bar{m}$ goes to infinity the algorithm approaches a perfect fit. Finally, in our application we slightly depart from the algorithm described by Lutz and Bühlmann (2006), as we always include the first lag of the dependent variable in the model.

2.6 Factor Models (SW)

Finally, a largely used method to overcome the curse of dimensionality problem arising in forecasting with large datasets is using a factor model. In a factor model, the information contained in the predictors X_t is summarized by a set of K factors:

$$X_t = \Gamma F_t + u_t \tag{25}$$

where F_t is a K-dimensional multiple time series of factors and Γ a $N \times K$ matrix of loadings.

The forecast for y_{t+1} given the predictors can be obtained through a two-step procedure, in which in the first step the sample data $\{X_t\}_{t=1}^T$ are used to estimate a time series of factors$\{\hat{F}_t\}_{t=1}^T$ via principal components, and then the forecasts are obtained by projecting $y_{i,t+1}$ onto $\hat{F}_t$ and $y_{i,t}$. Stock and Watson (2002a,b) develop theoretical results for this two-step procedure and show that under a set of moment and rank conditions the mean squared error of the feasible forecast asymptotically approaches that of the optimal infeasible forecast for N and T approaching infinity, see Bai and Ng (2006) for additional details. There are two ways to produce a h-step ahead forecast. First, one can use direct projection of the data onto the space spanned by the factors, i.e. one can produce the h-step ahead forecast as $y_{i,t+h} = \hat{a}_1\hat{F}_t + \hat{a}_2 y_{i,t}$, where $\hat{a}_1$ and $\hat{a}_2$ are the coefficients of a regression of $y_{i,t}$ onto $\hat{F}_{t-h}$ and $y_{i,t-h}$. Alternatively, one can develop a vector time series model for the factors $\hat{F}_t$ and use it to forecast, in turn, $\hat{F}_{t+h}$ and $y_{i,t+h}$. In this paper we use the latter strategy for comparability with the other models.

2.7 Summary of the models

Before moving on to the forecasting exercise it is worth to briefly summarize the main characteristics of the multivariate models under analysis. We have considered six alternative forecasting models, each of them aiming at summarizing in an efficient way the information contained in a large data set. One way to enhance parsimony is to impose

a reduced rank structure on the system (RR). Factor models (SW) can be considered a special case of RR, as they impose a particular type of rank reduction, i.e. a factor structure on the data. A characteristic of these models is that parsimony is obtained by reducing the size of the overall system, but they do not take a precise stance on particular regressors that might or might not be part of the data generating process. Alternatively, one can think of selecting the relevant regressors from a pool of candidates. Methods in this spirit are the BVAR and the MB. In particular, the MB can be considered a pure selection device in which a given regressor is either included or not included in the regression function. A problem of the MB is that it has considerable computational costs, even with a medium-sized dataset. On the other side, the BVAR is a somewhat smoother selection device, as it includes all the regressors, assigning to them different weights depending on the data. Finally, BRR and RRP combine both these strategies, using both rank reduction and Bayesian shrinkage, and as we shall see this leads to gains in forecast accuracy. A shortcoming of the BRR is that it is computationally intensive, so we proposed a new method (RRP) that retains the benefit of rank reduction and shrinkage without paying the high computational cost involved in the BRR.

3 Forecast Comparison

3.1 Data

We analyze the overall performance of the models described in the previous Section in forecasting 52 U.S. macroeconomic time series. The data are monthly observations going from 1959:1 through 2003:12, and are taken from the dataset of Stock and Watson (2005). The series have been chosen in order to represent the main categories of indicators which are relevant for central banks in understanding and forecasting developments in the macroeconomy, trying to be as parsimonious as possible given the computational bounds posed by the estimation of the competing models. In particular, some of the models at hand (RR) can not handle cases in which the time dimension is too short with respect to the cross-sectional dimension (which would be the case given the rolling scheme used for our forecasting exercise), while some others (BRR, MB) would become too computationally intensive. To solve this trade off between economic relevance and parsimony we have removed from the dataset of Stock and Watson (2005) those variables containing roughly the same information of others, such as the disaggregated sectoral data on industrial production and prices. These series contain information collinear to that of their aggregated counterparts, therefore they are both less interesting to forecast,

and very likely to create problems of collinearity.

The time series under analysis represent the typical data-set of interest for central banks, and can be grouped in three broad categories: series related to the real economy, series related to money and prices, and series related to financial markets. In the first group we have series of real output, income, employment, consumption, industrial production, inventories, sales. The second group comprises price indexes and several monetary aggregates. The last group includes interest rates on Treasury bills, exchange rates, and stock indexes.

The series are transformed by taking logarithms and/or differencing so that the transformed series are stationary. Forecasting is performed using the transformed data, then forecasts for the original variables are obtained integrating back. Importantly, we standardize the variables using only the data of the rolling sample used for the estimation, and not the whole sample, so that no information unavailable at the time of the forecast is used. In general, growth rates are used for real quantity variables, first differences are used for nominal interest rates, year on year growth rates for price series. For a detailed summary of the series under analysis and the used transformations see Table 1.

3.2 Forecasting exercise

The forecasting exercise is performed in pseudo real time, using a rolling estimation window of 10 years. Using a rolling scheme is a convenient way to deal with possible sample instability (Pesaran and Timmermann, 2005), and keeping fixed the size of the estimation window shall allow us to use the test proposed by Giacomini and White (2006) for comparing predicting accuracy. In particular, the scheme starts with estimating all the models using data from 1960:1 to 1969:12 (notice one year of data was used in order to compute yearly growth rates for some variables). On the basis of such estimates, forecasts up to 12-step ahead (i.e. for the period 1970:1-1970:12) are produced and stored. Then the estimation window is moved forward one month, becoming 1960:2 to 1970:1, and new forecasts are produced for the period 1970:2-1971:1. The scheme goes on until the forecasts for the period 2003:1-2003:12 are produced. At each point in time all variables are standardized prior to estimation, and then mean and variance are re-attributed to the forecasts accordingly. Importantly, we standardize the variables using only the data of the rolling sample used for the estimation, and not the whole sample, so that no information unavailable at the time of the forecast is used.

The BIC criterion applied to the BVAR for the 52 variables selects one lag both with the rolling samples and with the whole sample. However, this result may be driven by

the high number of parameters to be estimated. To control for this we also applied the BIC to the more parsimonious reduced rank VAR, with rank set to 1, but the selected lag length does not change. To evaluate whether there is any loss from such a short dynamic specification, we also compared the results for the BVAR(1) with those from a BVAR(13), the specification adopted by Banbura et al. (2007) and we found that the gains from using a longer lag specification are minor, if any. Therefore, we have used a one lag specification for all the models.

At each point in time we grid search over the relevant dimensions of the models at hand: for the SW model we search over the number of factors K, for RR we search over the assumed rank r, for BVAR the grid is over the tightness ϕ. For the MB we search over the number of iterations $\bar{m}$ and over the rescaling parameter ξ. For models in which both shrinkage and rank reduction is used, we grid search contemporaneously on both these dimensions.[7] Then, at each point in time we optimize our forecasts by choosing the model which minimized the forecast error for each variable and forecast horizon in the previous 2 years (i.e. 24 periods).

3.3 Comparing predictive accuracy

We will assess predictive accuracy using the multivariate loss function based on the mean squared forecast error proposed by Christoffersen and Diebold (1998). Let $\hat{Y}_{t+h}$ denote the h-step ahead forecast of the vector Y_t, the h-step ahead forecast error is then given by $FE_{t+h} = Y_{t+h} - \hat{Y}_{t+h}$. The multivariate mean square forecast error is given by $E[FE'_{t+h} \cdot W \cdot FE_{t+h}]$, where W is a matrix of weights which accounts for the fact that different series have different volatilities and predictability. We set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. Our measure of forecast accuracy is then given by the trace of the matrix $E[FE'_{t+h} \cdot W \cdot FE_{t+h}]$, which we label Weighted Trace Mean Squared Forecast Error ($WTMSFE$).

We assess predictive accuracy against two different benchmarks, one univariate, and the second multivariate. The univariate benchmark is an autoregressive model with lag length chosen via the BIC information criterion, which we label $AR(p^*)$, where p^* is selected for each individual series separately. The lag length p^* is chosen for each series and at each point in time using the rolling samples and a maximum lag of 13. The

[7] For SW we use $K = 1, 2, 3, 6, 10, 25, 50$ factors, for RR we use rank $r = 1, 2, 3, 6, 10, 25, 50, 52$, for the BVAR we use tightness $\phi = 2.0e-005, 0.0005, 0.002, 0.008, 0.018, 0.072, 0.2, 1,\ 500$, for MB we use $\bar{m} = 2 \times 52 \times 1,\ 2 \times 52 \times 2$ iterations and $\xi = 0.05, 0.1, 0.2$. For RRP we use $\phi = 2.0e-005, 0.0005, 0.002, 0.008, 0.018, 0.072, 0.2, 1,\ 500$ and $r = 1, 2, 3, 6, 10, 25, 50, 52$, for BRR we use $r = 1, 2, 3, 6, 10, 52$ and $\tau = 5, 10, 100$.

multivariate benchmark is the baseline Minnesota prior of Doan et al. (1984) with the standard choice of hyperparameters as in the package Regression Analysis of Time Series (RATS), and we label it $BVAR0$.[8] [9]

Beyond considering the overall performance of the models in forecasting all the series at hand, we also provide results for three key macroeconomic variables, i.e. Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$). We evaluate the accuracy in forecasting an individual variable by using the Mean Squared Forecast Error ($MSFE$). [10] For the individual series, we assess the statistical significance of the differences in the forecasts produced by the various models by using the Giacomini and White (2006) test. This is a test of equal forecasting method accuracy and as such can handle forecasts based on both nested and non-nested models, and regardless from the estimation procedures used in the derivation of the forecasts, including Bayesian methods.

3.4 Results

In this section we present the results of our forecasting exercise. Results against the $AR(p^*)$ benchmark are displayed in Table 2, while results against the $BVAR0$ benchmark are displayed in Table 3. The tables contain 6 panels each corresponding to a different forecast horizons, respectively 1, 2, 3, 6, 9, and 12 month-ahead. The first line of each panel reports the Relative $WTMSFE$ ($RWTMSFE$), i.e. the ratio of the Weighted Trace Mean Squared Forecast Error ($WTMSFE$) of a given model against the $WTMSFE$ of the benchmark. A $RWTMSFE$ below one signals that the model outperforms the benchmark in forecast accuracy. The remaining lines in each panel of the tables focus on some key macroeconomic variables, i.e. Industrial Production (IPS10), CPI Inflation (PUNEW), and the Federal Funds Rate (FYFF). For these variables we report the Relative Mean Squared Forecast Error ($RMSFE$), i.e. the ratio of the $MSFE$ of a given model against the $MSFE$ of the benchmark. Again, a $RMSFE$ below one signals that the model outperforms the benchmark in forecast accuracy. For the individual series the symbols *, **, *** denote respectively rejection at 10%, 5% and 1% level

[8] This is obtained by setting $\phi = 0.2$ in (8).

[9] We have also considered a simple $AR(1)$ and a random walk as benchmarks, but as both these models produce inferior forecasts than the $AR(p^*)$ the results are not reported here but can be found in a previous draft of this paper, available at http://ideas.repec.org/p/qmw/qmwecw/wp617.html

[10] We have also considered a loss function based on absolute rather than squared forecast errors. Results for this case are very similar to those obtained with the squared errors, and therefore we do not report them to save space, but they can be found in a previous draft of this paper, available at http://ideas.repec.org/p/qmw/qmwecw/wp617.html

of the null of equal predictive accuracy according to the Giacomini-White (2005) test. For all the entries in the tables the best models for each horizon are highlighted in bold.

Let us first focus on the overall performance of the models, i.e. the $RWTMSFE$.

For very short horizons (1- and 2- month ahead) there are no models able to beat the $AR(p^*)$ benchmark, while the $BVAR0$ benchmark is outperformed by the BVAR, the RRP and (only for the 2-month ahead) by the BRR. The $AR(p^*)$ is overall a very competitive benchmark, systematically outperforming the $BVAR0$ benchmark for any horizon shorter than 9 month-ahead. On the other side, for longer horizons the $BVAR0$ is slightly better than the $AR(p^*)$.

Overall, among the six models at hand, RRP and BRR produce the best forecasts in terms of $RWTMSFE$, with the BRR working relatively better at short horizons and the RRP at long horizons. At the 3-month ahead horizon the BRR produces gains in $RWTMSFE$ up to 3% (0.97) with respect to the $AR(p^*)$, and up to 12% (0.88) with respect to the $BVAR0$. At long horizons (6- to 12-month ahead) the gains of the RRP range between 13% (0.87) and 16% (0.84) against the $AR(p^*)$ and between 13% (0.87) and 15% (0.85) against the $BVAR0$. Also the BVAR and RR do a good job, but they are both systematically outperformed by either RRP or BRR.

Let us now focus on the prediction of three key macroeconomic variables, i.e. Industrial Production (IPS10), CPI Inflation (PUNEW), and the Federal Funds Rate (FYFF).

Starting with short horizons, at the 1-month ahead horizon the best forecast of industrial production is given by BVAR and BRR, which outperform the $AR(p^*)$ by 10%. The best forecast of inflation is produced by SW and MB, but it is still worse than that of the $AR(p^*)$ benchmark. The best forecast of the federal funds rate is that of the RRP, with a gain of 7% over the $AR(p^*)$ benchmark. At 2- and 3-month ahead BRR produces the best forecast of industrial production, with gains against the $AR(p^*)$ of 19% and 22% respectively. The best forecast for inflation is given by the RPP (and BVAR at 2-month ahead) with gains of 1% and 14%. Finally, the best forecast for the federal funds rate is given again by the RPP with gains of 14% and 11%. A similar pattern but with much higher gains in forecasting all the variables emerges if one looks at the comparison with the $BVAR0$ benchmark.

At longer horizons the RRP systematically and significantly outperforms the $AR(p^*)$ benchmark, with gains that can go up to 35% for industrial production and inflation, and 21% for the federal funds rate. The BRR and BVAR are producing good forecasts but are still below the RRP performance, the RR systematically beats the benchmark but is somewhat far from the RRP performance, while the SW and MB only occasionally beat the benchmark. The results against the BVAR0 show of course a similar ranking,

bur with gains relatively smaller (but still high) in forecasting inflation and industrial production.

To sum up, for very short horizons it is difficult to beat an $AR(p^*)$ benchmark, but the BVAR, the RRP and the BRR can do so for some variables. For intermediate and long horizons the best models are respectively BRR and RRP. The RR and the BVAR produce overall good results, however they are inferior to BRR and RRP. These findings provide encouraging evidence that using shrinkage and rank reduction is useful, and using them *in combination* rather than separately can improve substantially the accuracy of forecasts.

4 Robustness Analysis

In this Section we check the robustness of the results we have obtained so far. In particular, we look at the effects that the variable transformation and standardization might have on the BVAR results, we present subsample results, and we conduct a Monte Carlo simulation based on bootstrapped data.

4.1 The role of variable transformation

In our forecasting exercise the variables are made stationary and standardized before forecasting. The forecasted series are then transformed back by re-attributing mean and variance and by reversing the stationarity transformations. This decision may influence the model performance. In particular, factor models require that variables are stationary and standardized, but the BVARs do not assume so. Actually, the original Litterman (1980) Minnesota prior is specified for a system estimated in levels.

To assess the consequences that variable transformation has on the forecasting performance of the BVAR, we have repeated the forecasting exercise for the BVAR without applying the stationary transformation and the standardization, which we label BVAR(lev). The results are summarized in Table 4.

It turns out that the BVAR(lev) can produce significant forecast improvements for some series and forecast horizons, e.g. for inflation at short horizons, but overall the BVAR specification works better and produces lower $WTMSFE$. A possible explanation for the good performance of the BVAR versus the BVAR(lev) is related to the robustifying effect of differencing in the presence of structural breaks, see e.g. Clements and Hendry (1999).

4.2 Subsample analysis

Next we evaluate the robustness of our results using different subsamples. In order to save space we only present the results based on the $AR(p^*)$ benchmark, which is the most competitive Table 5 contains results obtained using the evaluation sample 1985:1 to 2003:12, while Table 6 refers to the sample 1995:1 to 2003:12.

The pattern emerging from these two tables is similar to that obtained over the whole sample, namely RRP and BRR produce on average the best forecasts, systematically beating the benchmark at long horizons. A new feature is that in these subsamples MB provides very good forecasts for the individual variables 1-month ahead, although only for industrial production it actually beats the benchmark. In addition, all the models perform slightly worse than over the whole sample, signalling that forecasting with multivariate models has become more difficult in the more recent period.

4.3 Monte Carlo evaluation

Finally, we have performed a small Monte Carlo experiment where we compare the alternative models using artificial data. Rather than using an inevitably arbitrary data generation process, we carry out our analysis based on the actual macroeconomic dataset, which is referred to as a 'data based Monte Carlo method' and discussed further in, e.g., Ho and Sørensen (1996). Following this work, we create artificial data by repeatedly bootstrapping the actual dataset. In particular, we use the block bootstrap algorithm described by Politis and Romano (1994), which is designed for block-bootstrapping from stationary data.

We implement the simulation exercise by bootstrapping 100 alternative artificial dataset over the sample 1984:1 to 2003:12. In each month all the models are estimated and forecasts are produced using the same rolling scheme adopted for the actual data. In particular, the first estimation window is 1984:1 to 1994:12 and the first forecast window is 1995:1 to 1995:12, while the last estimation window is 1992:1 to 2002:12 and the last forecast window is 2003:1 to 2003:12. As with the actual data all variables are transformed to obtain stationarity and standardized prior to estimation (using only the data of that particular rolling sample), and then mean and variance are re-attributed to the forecasts accordingly.

Table 7 presents the averages of the $RWTMSFE$ and $RMSFE$ of selected variables over the 100 artificial samples. The simulation results are clearly in line with those obtained using the actual data, namely, RRP produces the best forecasts, followed by the $BVAR$ and RR. This finding confirms that the use of *both* shrinkage *and* rank

reduction produces additional gains with respect to using the two methods separately.

5 Conclusions

In this paper we have addressed the issue of forecasting a large set of variables using multivariate models. In particular, we have proposed three alternative reduced rank forecasting models and compared their predictive performance with the most promising existing alternatives, namely, factor models, large scale Bayesian VARs, and multivariate boosting. We have provided a novel consistency result for the reduced rank regression valid when the dimension of the system tends to infinity; we have proposed a new two-step estimation procedure that applies, in turn, shrinkage and reduced rank restrictions (RRP); and we have implemented in a large data-set context the Bayesian VAR with rank reduction (BRR) proposed by Geweke (1996).

We have found that using shrinkage and rank reduction in combination rather than separately improves substantially the accuracy of forecasts. In particular RRP and BRR produce fairly good forecasts, more accurate than those of competing methods on average across several US macroeconomic variables, and they also perform well for key variables, such as industrial production growth, inflation and the short term interest rate. A small Monte Carlo simulation based on bootstrapped data confirmed these findings.

A natural extension of this study would be to analyze also combinations of the proposed forecasts among themselves and/or with the benchmarks, using fixed or optimal pooling weights. Bayesian model averaging of the different models according to their posterior probabilities is a closely related alternative worth investigating. Finally, the consistency result provided for the RR regression opens the ground to using large scale reduced rank models to also implement structural analysis.

Appendix: Proofs.

Proof of Theorem 1. It is sufficient to prove that for each of the N equations of the VAR model:

$$\left\|\hat{A}^i - A^i\right\|^2 = o_p(T^{-a}) \text{ for all } a < 1/2. \tag{26}$$

To prove (26) we mirror the analysis of Theorems 4 and 5 of An et al. (1982). For simplicity we consider Yule-Walker estimation[11] of $\hat{A}^i$ which is asymptotically equivalent

[11] Yule-Walker estimates are obtained by solving the Yule-Walker equations and replacing population quantities with their sample counterparts. The Yule-Walker equations are relations between the various autocovariances/autocorrelations of autoregressive models. For more details see, e.g. Reinsel (2003).

to OLS estimation. Let γ_i^{fp} and Γ^p denote the vector of covariances between $y_{i,t}$ and $X_{p,t}^p$ and the covariance matrix of $X_{p,t}^p$, respectively and $\hat{\gamma}_i^{fp}$ and $\hat{\Gamma}^p$ their sample counterparts. Further, denote by Γ_{ij}^p, $\hat{\Gamma}_{ij}^p$ γ_{ij}^{fp}, $\hat{\gamma}_{ij}^{fp}$ the i,j-th elements Γ_{ij}^p and $\hat{\Gamma}_{ij}^p$ and the j-th elements of γ_{ij}^{fp} and $\hat{\gamma}_{ij}^{fp}$, respectively. Then, by (25) of An et al. (1982)

$$\Gamma^p\left(\hat{A}^i - A^i\right) = -\left(\hat{\Gamma}^p - \Gamma^p\right)\left(\hat{A}^i - A^i\right) - \left(\hat{\gamma}_i^{fp} - \gamma_i^{fp}\right) - \left(\hat{\Gamma}^p - \Gamma^p\right)A^i \tag{27}$$

Since each $y_{i,t}$ is part of a stationary VAR process by assumption 1(a), and, also taking into account assumption 1(b)-(c), it follows that $y_{i,t}$ satisfies the assumptions of Theorem 5 of An et al. (1982). Define $A^i = (A_1^i, ..., A_{Np}^i)'$and $\hat{A}^i = (\hat{A}_1^i, ..., \hat{A}_{Np}^i)'$. Then, by the proof of Theorem 5 of An et al. (1982) (see the last 3 equations of page 935 of An et al. (1982)), we have

$$\left\|\left(\hat{\Gamma}^p - \Gamma^p\right)\left(\hat{A}^i - A^i\right)\right\|^2 = o_p(1)\sum_{j=1}^{Np}\left(\hat{A}_j^i - A_j^i\right)^2 \tag{28}$$

$$\left\|\hat{\gamma}_i^{fp} - \gamma_i^{fp}\right\|^2 = o_p\left((\ln T/T)^{1/2}\right) \tag{29}$$

and

$$\left\|\left(\hat{\Gamma}^p - \Gamma^p\right)A^i\right\|^2 = o_p\left((\ln T/T)^{1/2}\right) \tag{30}$$

Note that (28)-(30) follow from Theorem 5 of An et al. (1982), if further,

$$\sup_{i,j}\left(\hat{\Gamma}_{ij}^p - \Gamma_{ij}^p\right) = O_p\left((\ln T/T)^{1/2}\right)$$

and

$$\sup_{j}\left(\hat{\gamma}_{ij}^{fp} - \gamma_{ij}^{fp}\right) = O_p\left((\ln T/T)^{1/2}\right)$$

But this follows easily by minor modifications of the proof of Theorem 7.4.3 of Deistler and Hannan (1988) and the uniformity assumption on the fourth and second moments of e_t given in Assumption 1(c). Hence,

$$(1 + o_p(1))\left\|\hat{A}^i - A^i\right\|^2 = o_p\left((\ln T/T)^{1/2}\right) \tag{31}$$

which implies (26) and completes the proof of the theorem. ■

Proof of Theorem 2. We define formally the functions $g_{\mathcal{O}}(.)$ and $g_{\mathcal{K}}(.)$ such that

$$vec(\hat{\mathcal{K}}') = g_{\mathcal{K}}\left(vec(\hat{\mathcal{A}})\right) \tag{32}$$

and

$$vec(\hat{\mathcal{O}}') = g_{\mathcal{O}}\left(vec(\hat{\mathcal{A}})\right) \tag{33}$$

where $\hat{\mathcal{A}} = (\hat{A}_1, \hat{A}_2, ..., \hat{A}_p)$ and $\mathcal{A} = (A_1, A_2, ..., A_p)$. Therefore, $g_{\mathcal{O}}(.)$ and $g_{\mathcal{K}}(.)$ define the singular value decomposition operator. By theorems 5.6 and 5.8 of Chatelin (1983) $g_{\mathcal{O}}(.)$ and $g_{\mathcal{K}}(.)$ are bounded, continuous and differentiable and therefore admit a first order Taylor expansion. Therefore,

$$vec(\hat{\mathcal{K}}') - vec(\mathcal{K}') = \frac{\partial g'_{\mathcal{K}}}{\partial \mathcal{A}}\left(vec(\hat{\mathcal{A}}) - vec(\mathcal{A})\right)\frac{\partial g'_{\mathcal{K}}}{\partial \mathcal{A}} + o_p(T^{-a}) \tag{34}$$

and

$$vec(\hat{\mathcal{O}}') - vec(\mathcal{O}') = \frac{\partial g'_{\mathcal{O}}}{\partial \mathcal{A}}\left(vec(\hat{\mathcal{A}}) - vec(\mathcal{A})\right)\frac{\partial g'_{\mathcal{O}}}{\partial \mathcal{A}} + o_p(T^{-a}) \tag{35}$$

By theorem 1 every element of $\left(vec(\hat{\mathcal{A}}) - vec(\mathcal{A})\right)$ is $o_p(T^{-a})$. The number of columns of $\frac{\partial g'_{\mathcal{K}}}{\partial \mathcal{A}}$ and $\frac{\partial g'_{\mathcal{O}}}{\partial \mathcal{A}}$ are of the order N^2. Thus, each element of $vec(\hat{\mathcal{O}}') - vec(\mathcal{O}')$ and $vec(\hat{\mathcal{K}}') - vec(\mathcal{K}')$ is a linear combination of possibly all elements of $\left(vec(\hat{\mathcal{A}}) - vec(\mathcal{A})\right)$. It then follows that each element of $vec(\hat{\mathcal{O}}') - vec(\mathcal{O}')$ and $vec(\hat{\mathcal{K}}') - vec(\mathcal{K}')$ is $o_p(T^{-a+2b})$-consistent. ■

References

[1] Ahn, S. K., and Reinsel, G. C., 1988, "Nested Reduced-Rank Autoregressive Models for Time Series", Journal of the American Statistical Association, 83, 849–856.

[2] An H.Z, Chen Z.G., Hannan, E.J., 1982, "Autocorrelation, Autoregression and Autoregressive Approximation", The Annals of Statistics, Vol. 10, No. 3, pp. 926-936.

[3] Bai, J. and Ng, S., 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions", Econometrica, 74(4), 1133-1150.

[4] Bai, J. and Ng, S., 2009, "Boosting Diffusion Indices", Journal of Applied Econometrics, (2009).

[5] Banbura M., Giannone D. and Reichlin, L., 2007, "Bayesian VARs with Large Panels", CEPR working paper no.6326

[6] Breiman, L., 1996, "Bagging Predictors", Machine Learning, 36, 105-139.

[7] Buhlmann, P. 2006, "Boosting for High-Dimensional Linear Models", Annals of Statistics 54:2, 559–583.

[8] Buhlmann, P. and B. Yu, 2002, "Analyzing Bagging", Annals of Statistics, 30, 927-961.

[9] Camba-Mendez, G., Kapetanios, G., Smith, R.J., and Weale, M.R., 2003, "Tests of Rank in Reduced Rank Regression Models", Journal of Business and Economic Statistics, 21, 145-155.

[10] Canova, F., and Ciccarelli, M., 2004. "Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model", Journal of Econometrics, Elsevier, vol. 120(2), pages 327-359, June.

[11] Clements M.P. and Hendry D.F., 1999, "Forecasing Nonstationary Economic Time Series", MIT press

[12] De Mol, C., Giannone, D. and Reichlin, L., 2006, "Forecasting Using a Large Number of Predictors. Is Bayesian Regression a Valid Alternative to Principal Components?", ECB Working paper no.700.

[13] Deistler, M. and Hannan, E. J., 1988, "The Statistical Theory of Linear Systems", Wiley.

[14] Doan, T., R. Litterman, and C. A. Sims, 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions", Econometric Reviews, 3, 1-100.

[15] Eickmeier, S. and Ziegler, C., 2006, "How Good are Dynamic Factor Models at Forecasting Output and Inflation? A meta-analytic approach", Discussion Paper Series 1: Economic Studies, 42, Deutsche Bundesbank, Research Centre.

[16] Forni, M., Hallin, M., Lippi, M. & Reichlin, L., 2000, "The Generalized Factor Model: Identification and Estimation", The Review of Economics and Statistics 82(4), 540-554.

[17] Geweke, J., 1996, "Bayesian Reduced Rank Regression in Econometrics", Journal of Econometrics 75, 121–146.

[18] Giacomini R., White H., 2006, "Tests of Conditional Forecast Accuracy", Econometrica, Vol. 74, pp. 1545-1578.

[19] Hamilton, J.D., Waggoner, D.F. and Zha, T. 2007, "Normalization in Econometrics", Econometric Reviews, 26 pp.221–252.

[20] Hoogerheide, L.F., and Van Dijk, H.K., 2001, "Comparison of the Anderson-Rubin Test for Overidentification and the Johansen Test for Cointegration", Econometric Institute report EI 2001-04, Erasmus University Rotterdam.

[21] Inoue, Atsushi and Kilian, Lutz, 2004, "Bagging Time Series Models", CEPR Discussion Paper No. 4333.

[22] Kadiyala, K. R., and S. Karlsson, 1997, "Numerical Methods for Estimation and Inference in Bayesian VAR-Models", Journal of Applied Econometrics, 12(2), 99–132.

[23] Kleibergen, F., van Dijk H.K., 1994, "On the Shape of the Likelihood/Posterior in Cointegration Models", Econometric Theory 10, 514-551

[24] Kleibergen, F., van Dijk H.K., 1998, "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures", Econometric Theory 14, 701-743.

[25] Litterman, R. ,1986, "Forecasting With Bayesian Vector Autoregressions – Five Years of Experience", Journal of Business and Economic Statistics, 4, 25–38.

[26] Lutz, R.W. and Bühlmann, P., 2006, "Boosting for High-multivariate Responses in High-dimensional Linear Regression", Statistica Sinica 16, 471-494

[27] Pesaran, M.H., and Timmermann, A., (2005), "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks", Journal of Econometrics Volume 129, Issues 1-2, Pages 183-217

[28] Politis, D.N., Romano,J.P., 1994, "The Stationary Bootstrap", J. Amer. Statist. Assoc., vol. 89, No. 428, pp. 1303-1313.

[29] Reinsel, G., 1983, "Some Results on Multivariate Autoregressive Index Models", Biometrika, 70, 145–156.

[30] Reinsel, G. C., and Velu, R. P., 1998, "Multivariate Reduced Rank Regression", Lecture Notes in Statistics 136. New York: Springer-Verlag.

[31] Reinsel, G.,C., 2003, "Elements of Multivariate Time Series Analysis", Springer.

[32] Sims, C. A., and T. Zha,1998, "Bayesian Methods for Dynamic Multivariate Models", International Economic Review, 39(4), 949–68.

[33] Stock, J. H. and M. W. Watson, 2002a, "Macroeconomic Forecasting Using Diffusion Indexes", Journal of Business and Economic Statistics, 20, 147-62.

[34] Stock, J. H. and M. W. Watson, 2002b, "Forecasting Using Principal Components from a Large Number of Predictors", Journal of the American Statistical Association, 97, 1167–1179.

[35] Stock, J. H. and M. W. Watson, 2006, "Forecasting with many predictors", Handbook of Economic Forecasting, Elliott, G., Granger, C.W.J., and Timmermann, A. (ed.), NorthHolland

[36] Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso", J.Royal. Statist. Soc B., 58, 267–288.

[37] Timmerman, A., 2006, "Forecast combinations", Handbook of Economic Forecasting, Elliott, G., Granger, C.W.J., and Timmermann, A. (ed.), NorthHolland

[38] Velu, R. P., Reinsel, G. C., and Wichern, D. W., 1986, "Reduced Rank Models for Multiple Time Series", Biometrika, 73, 105–118.

Table 1. Summary of dataset

Code	Series	Transformation
A0M048	EMPLOYEE HOURS IN NONAG. ESTABLISHMENTS (AR, BIL. HOURS)	Monthly Growth Rate
A0M051	PERSONAL INCOME LESS TRANSFER PAYMENTS (AR, BIL. CHAIN 2000 USD)	Monthly Growth Rate
A0M052	PERSONAL INCOME (AR, BIL. CHAIN 2000 USD)	Monthly Growth Rate
A0M057	MANUFACTURING AND TRADE SALES (MIL. CHAIN 1996 USD)	Monthly Growth Rate
A0M059	SALES OF RETAIL STORES (MIL. CHAIN 2000 USD)	Monthly Growth Rate
A0M082	CAPACITY UTILIZATION (MFG)	First Difference
A0M095	RATIO, CONSUMER INSTALLMENT CREDIT TO PERSONAL INCOME (PCT.)	First Difference
A0M224R	REAL CONSUMPTION (AC) A0M224/GMDC	Monthly Growth Rate
CCINRV	CONSUMER CREDIT OUTSTANDING - NONREVOLVING(G19)	Change in Yearly Growth Rate
CES002	EMPLOYEES ON NONFARM PAYROLLS - TOTAL PRIVATE	Monthly Growth Rate
CP90	COMMERCIAL PAPER RATE	First Difference
EXRCAN	FOREIGN EXCHANGE RATE: CANADA (CANADIAN USD PER U.S.USD)	Monthly Growth Rate
EXRJAN	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.USD)	Monthly Growth Rate
EXRSW	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.USD)	Monthly Growth Rate
EXRUK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	Monthly Growth Rate
EXRUS	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)	Monthly Growth Rate
FCLBMC	WKLY RP LG COM BANKS:NET CHANGE COM'L AND INDUS LOANS(BILUSD,SAAR)	No Transf.
FCLNQ	COMMERCIAL AND INDUSTRIAL LOANS OUSTANDING IN 1996 DOLLARS (BCI)	Change in Yearly Growth Rate
FM1	MONEY STOCK: M1 (BILUSD,SA)	Change in Yearly Growth Rate
FM2	MONEY STOCK:M2 (BILUSD,SA)	Change in Yearly Growth Rate
FM2DQ	MONEY SUPPLY - M2 IN 1996 DOLLARS (BCI)	Monthly Growth Rate
FM3	MONEY STOCK: M3 (BILUSD,SA)	Change in Yearly Growth Rate
FMFBA	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MILUSD,SA)	Change in Yearly Growth Rate
FMRNBA	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS (MILUSD,SA)	Change in Yearly Growth Rate
FMRRA	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS (MILUSD,SA)	Change in Yearly Growth Rate
FSDXP	SANDP'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (perc PER ANNUM)	First Difference
FSPCOM	SANDP'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	Monthly Growth Rate
FSPIN	SANDP'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)	Monthly Growth Rate
FSPXE	SANDP'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (perc,NSA)	Monthly Growth Rate
FYAAAC	BOND YIELD: MOODY'S AAA CORPORATE (perc PER ANNUM)	First Difference
FYBAAC	BOND YIELD: MOODY'S BAA CORPORATE (perc PER ANNUM)	First Difference
FYFF	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (perc PER ANNUM,NSA)	First Difference
FYGM3	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(perc PER ANN,NSA)	First Difference
FYGM6	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(perc PER ANN,NSA)	First Difference
FYGT1	INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(perc PER ANN,NSA)	First Difference
FYGT10	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(perc PER ANN,NSA)	First Difference
FYGT5	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(perc PER ANN,NSA)	First Difference
IPS10	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX	Monthly Growth Rate
LHEL	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)	First Difference
LHELX	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	First Difference
LHEM	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	Monthly Growth Rate
LHUR	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS AND OVER (perc,SA)	First Difference
PMDEL	NAPM VENDOR DELIVERIES INDEX (PERCENT)	No Transf.
PMI	PURCHASING MANAGERS' INDEX (SA)	No Transf.
PMNO	NAPM NEW ORDERS INDEX (PERCENT)	No Transf.
PMNV	NAPM INVENTORIES INDEX (PERCENT)	No Transf.
PMP	NAPM PRODUCTION INDEX (PERCENT)	No Transf.
PUNEW	CPI-U: ALL ITEMS (82-84=100,SA)	Change in Yearly Growth Rate
PWCMSA	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	Change in Yearly Growth Rate
PWFCSA	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)	Change in Yearly Growth Rate
PWFSA	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)	Change in Yearly Growth Rate
PWIMSA	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES AND COMPONENTS (SA)	Change in Yearly Growth Rate

Table 2: Relative WTMSFE vs AR(p*) benchmark.

	RR	SW	BVAR	MB	RRP	BRR
Hor: 1						
rel. WTMSFE	1.36	1.70	1.18	2.08	**1.15**	1.22
IPS10	1.10 ***	1.09	**0.90**	0.99	0.97	**0.90**
PUNEW	1.31	**1.08**	1.11 *	**1.08**	1.12 *	1.24 ***
FYFF	1.09 *	1.01	0.94	1.02	**0.93**	0.99
Hor: 2						
rel. WTMSFE	1.17	1.35	1.06	1.67	1.06	**1.05**
IPS10	1.14	1.05	0.86	1.05	0.90	**0.81**
PUNEW	1.08	1.02	**0.99**	1.10	**0.99**	1.05
FYFF	1.01	0.98	0.91	1.01	**0.86 ***	0.95
Hor: 3						
rel. WTMSFE	1.07	1.17	0.99	1.40	0.98	**0.97**
IPS10	1.06	1.03	0.80	1.06	0.81	**0.78 ***
PUNEW	0.93	0.96	0.87 *	1.06	**0.86 ***	0.91
FYFF	1.01	0.99	0.92	1.01	**0.89 ***	0.94
Hor: 6						
rel. WTMSFE	0.94	1.11	0.88	1.09	**0.87**	0.88
IPS10	0.87	1.04	0.69 **	1.01	**0.67 ****	0.71 ***
PUNEW	0.76 ***	0.95	**0.71 *****	1.06	**0.71 *****	0.74 ***
FYFF	1.00	1.12	0.89 *	0.99	**0.83 *****	0.91 ***
Hor: 9						
rel. WTMSFE	0.91	1.13	0.85	1.01	**0.84**	0.87
IPS10	0.82	1.07	0.68 ***	1.02	**0.66 *****	0.72 ***
PUNEW	0.76 *	0.97	**0.67 *****	1.11 *	**0.67 *****	0.69 ***
FYFF	0.97	1.14	0.91 *	1.00	**0.79 *****	0.90 ***
Hor: 12						
rel. WTMSFE	0.90	1.20	0.85	0.97	**0.84**	0.87
IPS10	0.87	1.10	0.68 ***	1.01	**0.65 *****	0.74 ***
PUNEW	0.72 ***	0.99	**0.64 *****	1.06	0.65 ***	**0.64 *****
FYFF	0.95	1.20	0.90 ***	1.00	**0.84 *****	0.91 ***

The Relative WTMSFE is the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the ratio of the MSFE of a given model against the MSFE of the benchmark. The symbols *, **, *** denote respectively rejection at 10%, 5% and 1% level of the null of equal predictive accuacy according to the Giacomini-White (2005) test. Best models are in bold. RR is the Reduced Rank Regression, SW is the Factor Model, BVAR is a Bayesian VAR with Minnesota-type prior, MB is Multivariate Boosting, RRP is Reduced Rank Posterior, BRR is Bayesian Reduced Rank Regression. The forecasting exercise is performed using a rolling window of 10 years. The first estimation window is 1960:1 to 1969:12 and the first forecast window is 1970:1 to 1970:12, while the last estimation window is 1984:1 to 1993:12 and the last forecast window is 2003:1 to 2003:12.

Table 3: Relative WTMSFE vs BVAR0 benchmark.

	RR	SW	BVAR	MB	RRP	BRR
Hor: 1						
rel. WTMSFE	1.13	1.41	0.99	1.73	**0.96**	1.02
IPS10	0.94	0.93	**0.77 *****	0.84	0.82 ***	**0.77 *****
PUNEW	0.87 **	**0.71 *****	0.73 ***	**0.71 *****	0.74 ***	0.82 ***
FYFF	1.02	0.95	0.89 *	0.96	**0.87 *****	0.94
Hor: 2						
rel. WTMSFE	1.02	1.17	0.93	1.45	**0.92**	**0.92**
IPS10	1.04	0.96	0.78 ***	0.95	0.82 ***	**0.74 *****
PUNEW	0.82 ***	0.78 ***	**0.75 *****	0.83 ***	**0.75 *****	0.80 ***
FYFF	1.05	1.01	0.95	1.04	**0.90 *****	0.99
Hor: 3						
rel. WTMSFE	0.97	1.06	0.89	1.27	0.89	**0.88**
IPS10	1.02	1.00	**0.78 *****	1.02	0.78 ***	**0.75 *****
PUNEW	0.81 ***	0.84 ***	0.76 ***	0.92	**0.75 *****	0.79 ***
FYFF	1.04	1.02	0.95	1.05	**0.92**	0.98
Hor: 6						
rel. WTMSFE	0.92	1.08	0.86	1.07	**0.85**	0.87
IPS10	1.00	1.20	0.80 ***	1.17	**0.77 *****	0.82 ***
PUNEW	0.80 ***	1.01	**0.75 *****	1.12	**0.75 *****	0.78 ***
FYFF	0.92 *	1.04	0.82 ***	0.91	**0.77 *****	0.83 ***
Hor: 9						
rel. WTMSFE	0.92	1.15	0.87	1.02	**0.85**	0.89
IPS10	1.04	1.35 *	0.86 ***	1.29	**0.84 *****	0.92
PUNEW	0.88 *	1.12	**0.78 *****	1.28	**0.78 *****	0.80 ***
FYFF	0.92	1.08	0.86 ***	0.95	**0.75 *****	0.85 **
Hor: 12						
rel. WTMSFE	0.94	1.25	0.88	1.01	**0.87**	0.90
IPS10	1.17	1.49 **	0.92 *	1.37 *	**0.88**	0.99
PUNEW	0.91	1.25 **	**0.81 *****	1.33 *	0.82 ***	**0.81 *****
FYFF	0.89 ***	1.12	0.84 ***	0.93	**0.78 *****	0.85 **

The Relative WTMSFE is the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the ratio of the MSFE of a given model against the MSFE of the benchmark. The symbols *, **, *** denote respectively rejection at 10%, 5% and 1% level of the null of equal predictive accuacy according to the Giacomini-White (2005) test. Best models are in bold. RR is the Reduced Rank Regression, SW is the Factor Model, BVAR is a Bayesian VAR with Minnesota-type prior, MB is Multivariate Boosting, RRP is Reduced Rank Posterior, BRR is Bayesian Reduced Rank Regression. The forecasting exercise is performed using a rolling window of 10 years. The first estimation window is 1960:1 to 1969:12 and the first forecast window is 1970:1 to 1970:12, while the last estimation window is 1984:1 to 1993:12 and the last forecast window is 2003:1 to 2003:12.

Table 4: BVAR in Differences and BVAR in Levels vs AR(p*)

	BVAR	BVAR(lev)		BVAR	BVAR(lev)
Hor: 1			**Hor: 6**		
rel. WTMSFE	**1.18**	1.24	**rel. WTMSFE**	**0.88**	1.16
IPS10	**0.90**	0.92	**IPS10**	**0.69**	**0.69**
PUNEW	1.11	**1.02**	**PUNEW**	**0.71**	0.91
FYFF	**0.94**	1.06	**FYFF**	**0.89**	1.03
Hor: 2			**Hor: 9**		
rel. WTMSFE	**1.06**	1.16	**rel. WTMSFE**	**0.85**	1.31
IPS10	0.86	**0.85**	**IPS10**	**0.68**	0.70
PUNEW	0.99	**0.98**	**PUNEW**	**0.67**	0.97
FYFF	**0.91**	1.03	**FYFF**	**0.91**	1.04
Hor: 3			**Hor: 12**		
rel. WTMSFE	**0.99**	1.13	**rel. WTMSFE**	**0.85**	1.50
IPS10	**0.80**	**0.80**	**IPS10**	**0.68**	0.81
PUNEW	**0.87**	0.93	**PUNEW**	**0.64**	1.12
FYFF	**0.92**	1.03	**FYFF**	**0.90**	1.12

The Relative WTMSFE is the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the ratio of the MSFE of a given model against the MSFE of the benchmark. Best models are in bold. BVAR is the Bayesian VAR estimated on stationary standardised data, BVAR(lev) is the same model estimated with the data not transformed. The forecasting exercise is performed using a rolling window of 10 years. The first estimation window is 1960:1 to 1969:12 and the first forecast window is 1970:1 to 1970:12, while the last estimation window is 1984:1 to 1993:12 and the last forecast window is 2003:1 to 2003:12.

Table 5: Relative WTMSFE vs AR(p*) benchmark, sample 1985:1 to 2003:12

	RR	SW	BVAR	MB	RRP	BRR
Hor: 1						
rel. WTMSFE	1.34	1.36	1.28	1.47	**1.19**	1.26
IPS10	1.09	1.08	1.14	**0.84 *****	1.15	0.99
PUNEW	1.30 ***	1.19 ***	1.31 ***	**1.11**	1.18 ***	**1.11 ***
FYFF	1.76 ***	1.79 ***	1.59 ***	**1.01**	1.51 *	1.20
Hor: 2						
rel. WTMSFE	1.23	1.25	1.18	1.32	**1.13**	1.14
IPS10	0.99	1.04	0.97	0.89 *	1.00	**0.88 ***
PUNEW	1.18 **	1.13 **	1.22 ***	1.06	1.10	**1.00**
FYFF	1.37 ***	1.50 ***	1.26	1.06	1.19	**0.96**
Hor: 3						
rel. WTMSFE	1.15	1.16	1.11	1.23	1.08	**1.06**
IPS10	1.01	0.99	0.98	0.93	0.95	**0.89**
PUNEW	1.11	1.17 ***	1.08	1.00	1.00	**0.91 ***
FYFF	1.11	1.25 *	0.99	1.08	0.91	**0.81 ***
Hor: 6						
rel. WTMSFE	1.00	1.19	0.96	1.05	0.93	**0.92**
IPS10	1.01	0.96	0.90	0.95	**0.82 ***	0.88
PUNEW	1.02	1.26 ***	0.88 *	0.93	0.89	**0.79 *****
FYFF	0.96	1.00	0.76 **	1.10	**0.74 *****	0.75 ***
Hor: 9						
rel. WTMSFE	0.96	1.26	0.92	1.02	**0.89**	**0.89**
IPS10	0.98	0.96	0.91	0.96	**0.82 ****	0.89 **
PUNEW	0.99	1.29 ***	0.84	0.97	0.88	**0.79 *****
FYFF	0.85	0.89	**0.69 *****	1.05	**0.69 *****	**0.69 *****
Hor: 12						
rel. WTMSFE	0.93	1.38	0.90	0.99	**0.88**	**0.89**
IPS10	0.98	1.01	0.94	0.98	**0.85 *****	0.91 **
PUNEW	0.95	1.50 ***	0.80 **	0.91 *	0.84 *	**0.77 *****
FYFF	0.79 *	0.83	0.72 ***	1.02	**0.70 *****	0.72 ***

The Relative WTMSFE is the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the ratio of the MSFE of a given model against the MSFE of the benchmark. The symbols *, **, *** denote respectively rejection at 10%, 5% and 1% level of the null of equal predictive accuacy according to the Giacomini-White (2005) test. Best models are in bold. RR is the Reduced Rank Regression, SW is the Factor Model, BVAR is a Bayesian VAR with Minnesota-type prior, MB is Multivariate Boosting, RRP is Reduced Rank Posterior, BRR is Bayesian Reduced Rank Regression. The forecasting exercise is performed using a rolling window of 10 years. The first estimation window is 1974:1 to 1984:12 and the first forecast window is 1985:1 to 1985:12, while the last estimation window is 1992:1 2002:12 and the last forecast window is 2003:1 to 2003:12.

Table 6: Relative WTMSFE vs AR(p*) benchmark, sample 1995:onwards.

	RR	SW	BVAR	MB	RRP	BRR
Hor: 1						
rel. WTMSFE	1.35	1.43	1.25	1.44	**1.17**	1.25
IPS10	1.21	1.11	1.00	**0.90**	1.14	1.01
PUNEW	1.49 ***	1.36 ***	1.34 ***	**1.10**	1.26 ***	1.16 *
FYFF	0.89	1.95	**0.66 *****	0.98 *	0.68 ***	0.77 ***
Hor: 2						
rel. WTMSFE	1.26	1.27	1.18	1.31	**1.13**	1.14
IPS10	1.10	1.02	**0.89**	0.97	0.95	0.92
PUNEW	1.49 ***	1.24 ***	1.49 ***	**1.13**	1.46 ***	1.17
FYFF	1.01	1.83	**0.74 ***	1.14	0.75 *	0.66 ***
Hor: 3						
rel. WTMSFE	1.15	1.21	1.10	1.24	1.08	**1.05**
IPS10	1.06	0.96	0.92	1.04	**0.91**	0.95
PUNEW	1.22	1.32 **	1.26	1.07	1.23	**0.96**
FYFF	0.92	1.51	0.59 ***	1.17	0.62 ***	**0.52 *****
Hor: 6						
rel. WTMSFE	1.02	1.24	0.97	1.07	0.96	**0.93**
IPS10	1.03	1.03	0.85	1.06	**0.80**	0.94
PUNEW	0.89	1.03	0.93	0.93	0.94	**0.77**
FYFF	0.99	1.31	0.59 ***	1.22	0.64 ***	**0.56 *****
Hor: 9						
rel. WTMSFE	0.99	1.33	0.95	1.01	0.92	**0.91**
IPS10	0.99	1.03	0.86	1.04	**0.81**	0.94
PUNEW	0.80	0.96	0.81	0.91	0.85	**0.73**
FYFF	0.90	1.13	**0.60 ****	1.10	0.64 ***	0.61 **
Hor: 12						
rel. WTMSFE	0.98	1.47	0.94	0.98	**0.92**	**0.92**
IPS10	1.02	1.07	0.93	1.04	**0.86**	0.97
PUNEW	0.80 *	0.94	0.76	0.86	0.81	**0.72**
FYFF	0.83	1.00	**0.63 ***	1.01	0.66 *	0.66 *

The Relative WTMSFE is the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the ratio of the MSFE of a given model against the MSFE of the benchmark. The symbols *, **, *** denote respectively rejection at 10%, 5% and 1% level of the null of equal predictive accuacy according to the Giacomini-White (2005) test. Best models are in bold. RR is the Reduced Rank Regression, SW is the Factor Model, BVAR is a Bayesian VAR with Minnesota-type prior, MB is Multivariate Boosting, RRP is Reduced Rank Posterior, BRR is Bayesian Reduced Rank Regression. The forecasting exercise is performed using a rolling window of 10 years. The first estimation window is 1984:1 to 1994:12 and the first forecast window is 1995:1 to 1995:12, while the last estimation window is 1992:1 to 2002:12 and the last forecast window is 2003:1 to 2003:12.

Table 7: Monte Carlo Analysis

	RR	SW	BVAR	RRP
Hor:1				
Avg.rel. WTMSFE	1.28	**1.25**	1.24	1.21
IPS10	1.49	**1.47**	0.97	1.03
PUNEW	1.37	1.25	1.35	**1.17**
FYFF	1.42	**1.38**	1.09	1.25
Hor:2				
Avg.rel. WTMSFE	1.21	1.30	1.18	**1.11**
IPS10	1.37	1.51	1.06	**0.99**
PUNEW	1.22	1.24	1.20	**1.07**
FYFF	1.29	1.33	**1.03**	1.15
Hor:3				
Avg.rel. WTMSFE	1.17	1.34	1.15	**1.06**
IPS10	1.26	1.51	1.06	**0.95**
PUNEW	1.17	1.25	1.16	**1.02**
FYFF	1.21	1.31	**1.03**	**1.03**
Hor:6				
Avg.rel. WTMSFE	1.07	1.41	1.06	**0.95**
IPS10	1.10	1.47	1.05	**0.89**
PUNEW	1.10	1.29	1.04	**0.94**
FYFF	1.09	1.29	1.00	**0.91**
Hor:9				
Avg.rel. WTMSFE	1.02	1.49	1.00	**0.91**
IPS10	1.03	1.49	1.00	**0.89**
PUNEW	1.05	1.28	0.96	**0.91**
FYFF	1.03	1.31	0.95	**0.87**
Hor:12				
Avg.rel. WTMSFE	1.01	1.55	0.99	**0.90**
IPS10	1.00	1.54	0.97	**0.89**
PUNEW	1.03	1.31	0.95	**0.88**
FYFF	1.00	1.33	0.95	**0.86**

The Avg. Relative WTMSFE is average (computed over 100 simulations) of the ratio of the WTMSFE of a given model against the WTMSFE of the benchmark. For the three individual series Industrial Production ($IPS10$), CPI Inflation ($PUNEW$), and the Federal Funds Rate ($FYFF$) the figure reported is the average (computed over 100 simulations) ratio of the MSFE of a given model against the MSFE of the benchmark. Best models are in bold. RR is the Reduced Rank Regression, SW is the Factor Model, BVAR is a Bayesian VAR with Minnesota-type prior, RRP is Reduced Rank Posterior. The forecasting exercise is performed using bootstrapped data on a rolling window of 10 years. The first estimation window is 1984:1 to 1994:12 and the first forecast window is 1995:1 to 1995:12, while the last estimation window is 1992:1 to 2002:12 and the last forecast window is 2003:1 to 2003:12.