Pirschel, Inske; Wolters, Maik

Conference Paper

Forecasting German key macroeconomic variables using large dataset methods

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Forecasting, No. B16-V4

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Forecasting German key macroeconomic variables using large dataset methods*

Inske Pirschel†
University of Kiel and Kiel Institute for the World Economy

Maik H. Wolters‡
University of Kiel and Kiel Institute for the World Economy

February 28, 2014

Abstract

We study the forecasting performance of three alternative large scale approaches for German key macroeconomic variables using a dataset that consists of 123 variables in quarterly frequency. These three approaches handle the dimensionality problem evoked by such a large dataset by aggregating information, yet on different levels. We consider different factor models, a large Bayesian VAR and model averaging techniques, where aggregation takes place before, during and after the estimation of the different models, respectively. We find that overall the large Bayesian VAR provides the most precise forecasts compared to the other large scale approaches and a number of small benchmark models. For some variables the large Bayesian VAR is also the only model producing unbiased forecasts at least for short horizons. While a Bayesian factor augmented VAR with a tight prior also provides quite accurate forecasts overall, the performance of the other methods depends on the variable to be forecast.

Keywords: Large Bayesian VAR, Model averaging, Factor models, Great Recession
JEL-Codes: C53, E31, E32, E37, E47

*We thank Christian Schumacher for very helpful comments and for sharing the dataset used in Schumacher (2007). We further thank Jens Boysen-Hogrefe and participants of the 2013 DIW macroeconometric workshop in Berlin, the 2013 IWH macroeconomic workshop in Halle and seminar participants at IfW Kiel for useful comments.
†inske.pirschel@ifw-kiel.de
‡maik.wolters@ifw-kiel.de

1 Introduction

In forecasting with standard time series methods the following trade-off arises: Given the vast amount of macroeconomic time series all of which potentially contain important information on future macroeconomic dynamics, the forecaster wishes to use as much information as possible to obtain precise forecasts. Estimation and forecasting with large cross-sections, however, may cause huge technical difficulties. As the number of parameters to be estimated in large cross-section models quickly becomes very large, parameter estimates might get imprecise and in-sample overfitting might occur. This can lead to poor out-of-sample forecasts. In some cases the estimation might even be infeasible due to the very limited number of observations in typical macroeconomics applications.

To overcome this curse of dimensionality several large scale time series methods have been proposed. In this paper we study the performance of the three most prominent of these approaches, namely factor models, large Bayesian vector autoregressions and model averaging techniques. These three approaches handle the dimensionality problem evoked by large datasets by aggregating the informational content of the dataset, yet on different levels.

For example, factor models (see e.g. Stock and Watson, 2002a,b; Bernanke and Boivin, 2003; Forni et al., 2000, 2005; Schumacher, 2007) aggregate the information contained in a large number of time series into a small number of static or dynamic factors *prior* to the estimation. These factor time series can be included into standard small scale forecasting models such as autoregressive distributed lag models, vector autoregressions or Bayesian vector autoregressions.

Large Bayesian vector autoregressions (De Mol et al., 2008; Bańbura et al., 2010) on the other hand can handle a large number of time series by applying shrinkage to make estimation feasible. The degree of shrinkage increases with the cross-sectional size of the respective model. With this method the information contained in a large dataset is thus aggregated *during* the estimation process.

By contrast, when using model averaging techniques (see e.g. Bates and Granger, 1969; Stock and Watson, 2003; Timmermann, 2006; Wright, 2009; Faust and Wright, 2009) aggregation takes place *after* the estimation of a large number of small forecasting models which contain only two variables each. The final forecast is computed as a weighted average of the forecasts of all the small forecasting models. Depending on the specification of the weights used to compute the average, there exist several variants of this approach, for example equal weighted averaging and Bayesian model averaging.

Previous literature has mainly focused on evaluating the forecasting performance of one or two of the large scale approaches relative to several small-scale benchmark models or to each other. For example, Bernanke and Boivin (2003) compare forecasts obtained by factor augmented autoregressions (FAAR) and factor augmented vector autoregressions (FAVAR) to those of simple benchmark models and a weighted forecast consisting of either the FAAR and the Federal Reserve's Greenbook projections or the FAVAR and the Greenbook projections. They find that the weighted forecasts are more precise than the FAAR and the FAVAR forecasts, which in turn dominate the forecasts of several simple benchmark models. Faust and Wright (2009) evaluate alternative specifications of factor models as well as equal weighted and

Bayesian model averaging (EWA, BMA) relative to a number of simple benchmark models and the Greenbook projections. They find that model averaging techniques perform better than the factor based forecasting approaches. Bańbura et al. (2010) study the forecasting performance of large Bayesian vector autoregressions (LBVAR) and Bayesian factor augmented vector autoregressions (BFAVAR). They find that the LBVAR generally outperforms the BFAVAR. Wolters (2014) compares the forecasting accuracy of Dynamic Stochastic General Equilibium (DSGE) models to LBVARs and the Fed's Greenbook projections. He finds that weighted forecasts of several DSGE models and LBVARs are more precise than those obtained by individual DSGE models and small Bayesian vector autoregressions (BVAR) and come close to the accuracy of the Greenbook projections at least for medium term horizons. Berg and Henzel (2013) evaluate the relative forecasting performance of BFAVARs, LBVARs and combinations of small BVARs for the Euro area. Their results indicate that the LBVAR performs best but that the forecasts of BFAVARs are only slightly less accurate. As one of the few papers that also evaluate density forecasts, they find that BFAVARS yield more precise density forecasts than the LBVAR. Finally, Schumacher (2007) uses German data and studies the forecasting performance of alternative factor models. He finds that while all factor models clearly outperform a simple benchmark model, the dynamic factor model and the factor model based on subspace algorithms for statespace models dominate the static factor model.

The variety of forecasting performance based rankings of the different models shows that there is no consensus yet which is the most useful method to extract the predictive content from large datasets. With this paper we seek to fill this gap by systematically comparing the forecasting accuracy of the three most prominent large scale approaches.

Beyond that we contribute to the existing literature in the following ways. First, we do not only study the relative performance of different forecasting methods, but we also check their absolute forecasting accuracy. In particular, we test whether the forecasts we obtain with the different models are unbiased. We also report the share of the variance of the forecasted time series that can be explained by each forecasting model. Next, we analyze which large scale forecasting method is suited best to simultaneously predict a larger set of macroeconomic variables. Previous papers—with the exception of Carriero et al. (2011) who forecast all the variables in a large dataset at the same time— only evaluate the forecasts for a small set of key variables which usually include output growth and inflation and in some cases the interest rate and the unemployment rate. In practice however, forecasters might be interested in a larger number of macroeconomic variables. The monthly survey of Consensus Economics among forecasters for example covers about ten variables per country. We therefore think that the advantage of the large scale forecasting models is not only their ability to process the informational content of many time series, but also to provide a coherent forecasting framework that can be used to simultaneously forecast a larger set of core variables. Finally, while the majority of previous work focuses on US macroeconomic time series and a small number of recent papers on the Euro area as a whole, we use a dataset for Germany. This allows us to check whether the results of previous papers are robust to the usage of a large dataset for another country that is smaller and more open than the US and the Euro area.

The dataset that we use consists of 123 variables in quarterly frequency covering a sample period from 1978 until 2013. We include indicators from the following categories: composition of GDP and gross value added by sectors, prices, labor market, financial market, industry, construction, surveys and miscellaneous. Our dataset is a modified and updated version of the one used in Schumacher (2007).

We estimate the three large scale models as well as a number of small benchmark models using a moving window of 15 years of data. The evaluation sample thus ranges from 1993 through 2013. We compute forecasts up to eight quarters ahead for a small set of German key variables, namely output growth, CPI inflation, a short term interest rate and the unemployment rate. To evaluate the relative forecasting performance of the different models we compare root means squared forecasting errors (RMSE), while we compute Mincer-Zarnowitz regressions (see Mincer and Zarnowitz, 1969) to assess the absolute forecasting accuracy of each model. Finally, we extend the analysis to a larger set of eleven macroeconomic variables.

Our results indicate that among the three large scale forecasting approaches the LBVAR shows the best overall forecasting performance followed by the BFAVAR. Both deliver forecasts that are more precise than those obtained by a simple univariate autoregressive (AR) benchmark model for most of the variables of interest. By contrast, for the other factor based models (FAAR and FAVAR) as well as the model averaging techniques (EWA and BMA) and a small BVAR the forecasing performance relative to the AR benchmark model depends heavily on the forecasted variable. We also check whether the models are able to forecast the Great Recession and find that none of them can predict the beginning, depth and length of the recession.

While the LBVAR yields a reduction in RMSEs relative to the AR benchmark between 10 and 15% for a few variables such as CPI inflation and wage growth, the gains in relative forecasting accuracy obtained by the large scale approaches are in most cases much smaller and rarely statistically significant. One explanation for this might be that some of the time series show very little persistence and are thus very difficult to forecast in general. We indicate for which variables this is the case by reporting the R^2 from the Mincer-Zarnowitz regressions. Furthermore, for some variables univariate forecasting models might also be hard to beat because many time series are characterized by common components. This implies that parsimonious univariate models are often sufficient to capture the main information contained in the data. Efficient multivariate modelling therefore becomes a hard task so that improvements of the large data forecasting methods are rather small (see also Carriero et al., 2011; Bernardini and Cubadda, 2014).

The remainder of this paper is structured as follows. Section 2 outlines the different forecasting models, while section 3 describes the dataset that we use and section 4 describes our forecasting approach. In section 5 we evaluate the absolute and relative forecasting performance of the different models, first for a small set of four key variables and then for a larger set of eleven macroeconomic variables, and discuss the results. Finally, section 6 concludes.

2 Forecasting Models

In the following we describe the different forecasting models. Let $\{y_{i,t}\}_{i=1}^{n}$ denote the set of variables to be forecast and $\{x_{j,t}\}_{j=1}^{m}$ the set of possible predictors for the estimation period $t = 1, ..., T$. The total number of variables is given by $n + m = k$. The variables contained in $\{y_{i,t}\}_{i=1}^{n}$ and $\{x_{j,t}\}_{j=1}^{m}$ are in log-levels except for those that are expressed in rates such as the unemployment rate or interest rates which are included in levels. $\{\Delta y_{i,t}\}_{i=1}^{n}$ and $\{\Delta x_{j,t}\}_{j=1}^{m}$ denote the set of annualized quarter-on-quarter growth rates of the variables to be forecast and the possible predictors.[1] Given the information available at time T, we estimate all forecasting models and construct annualized quarter-on-quarter growth rate forecasts $\{\Delta y_{i,T+h}\}_{i=1}^{n}$.[2,3] h is the forecast horizon ranging from one to eight quarters ahead.

2.1 Large Bayesian VAR (LBVAR)

Consider the following VAR $Z_t = c + A_1 Z_{t-1} + ... + A_p Z_{t-p} + \epsilon_t$, where the vector $Z_t = (y_{1,t}, ..., y_{n,t}, x_{1,t}, ..., x_{m,t})'$ contains all the k time series in the dataset, p is the number of lags included in the estimation, c is a k x 1 vector of constants, $A_1, ..., A_p$ are k x k-dimensional parameter matrices and ϵ_t is a k x 1 vector of independently identically distributed white noise error terms with zero mean and covariance matrix Ψ.

We use Bayesian techniques to estimate the model. Since the number of variables that we want to include in the estimation is fairly large ($k = 123$), we follow Bańbura et al. (2010) and choose a prior that shrinks the parameters to be estimated. The degree of shrinkage thereby increases with the size of the cross-section and thus allows the estimation of models where the number of parameters exceeds the number of observations by far. Bańbura et al. (2010) show that this approach is suited well to capture the most important factors in a dataset that is characterized by strong collinearity. This will be the case for our dataset which includes for instance different price indices and business cycle indicators. We implement the Bayesian shrinkage approach by using a version of the Normal inverse Wishart prior (see e.g. Kadiyala and Karlsson, 1997) that is characterized as follows.[4] First, the coefficients $A_1, ..., A_p$ are assumed to be a priori independent and normally distributed. With respect to the constant in the VAR the prior is assumed be diffuse. The moments for the prior distribution of the VAR coefficients are given by:

$$E[(A_\ell)_{ij}] = \begin{cases} \delta_i = 1 \quad \text{or} \quad \delta_i = \mu_i & \text{for} \quad i = j, \ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

[1] To avoid overly complicated notation, variables expressed in rates are included in levels in the Δ terms.

[2] For variables expressed in rates we compute level forecasts which are again included in the Δ terms.

[3] While some of the forecasting models directly yield growth rate forecasts, we obtain log-level or level forecasts from the other models and use these to compute implied quarter-on-quarter growth rate forecasts.

[4] This prior is a natural conjugate prior for our VAR which implies that analytical results are available. In contrast to the widely used Minnesota prior (Litterman, 1986) the Normal inverse Wishart prior allows for correlation between the residuals of different equations of the VAR and does not assume that the residual variance-covariance matrix is fixed and known.

and

$$Var[(A_\ell)_{ij}] = \begin{cases} \frac{\lambda^2}{\ell^2} & \text{for} \quad i = j \\ \frac{\lambda^2 \sigma_i^2}{\ell^2 \sigma_j^2} & \text{otherwise.} \end{cases} \tag{2}$$

where δ_i denotes the prior coefficient mean, $\ell = 1, ..., p$ is the lag length, λ is a hyperparameter governing the importance of the prior beliefs relative to the data and σ_i/σ_j are scale parameters adjusting the prior for the different scale and variability of the data.

According to this prior specification each equation of the VAR is centered around a random walk with drift ($\delta_i = 1$) or an autoregressive process ($\delta_i = \mu_i < 1$), respectively. The prior also incorporates the belief that more recent lags of a variable should provide more reliable information for the estimation. The zero coefficient prior on more recent lags is therefore not imposed as tightly as on less recent lags.

The hyperparameter λ controls the degree of shrinkage. Bańbura et al. (2010) suggest to increase the tightness of the prior with the size of the cross-section to avoid over-fitting. In particular, they propose to set λ so that the LBVAR achieves the same in sample-fit as an unrestricted small VAR without shrinkage. We slightly depart from this approach and set the tightness-parameter λ such that the LBVAR achieves the same in-sample fit as a small BVAR in our key variables $\{y_{i,t}\}_{i=1}^n$. We find that this increases the forecasting performance of the LBVAR considerably.[5] In contrast to Bańbura et al. (2010), we do not set the prior coefficient means equal to zero for stationary variables but rather equal to the sum of coefficient estimates μ_i defined as $\sum_{\iota=1}^{p'} \beta_\iota$ where β_ι denotes the parameter estimates obtained from the simple auxiliary autoregression $Z_{i,t} = d + \sum_{\iota=1}^{p} \beta_\iota Z_{i,t-\iota} + u_t$. In particular, we set $\delta_i = 1$ if $\mu_i \geq 1$ and $\delta_i = \mu_i$ if $\mu_i < 1$. This approach should capture the different degrees of persistence in macroeconomic time series. Finally, in line with Bańbura et al. (2010) we obtain σ_i by computing the standard deviation of the residuals of a univariate autoregression without constant for each of the k variables in the model.

For the estimation we use the variables in log-levels rather than growth rates to not loose information that might possibly be contained in the trends. We set the lag length $p = 4$, however the forecasting performance of the LBVAR proves to be remarkably robust with respect to the number of lags included. Following Bańbura et al. (2010) we implement the prior using dummy variables and augment it to constrain the sum of coefficients of the VAR (see e.g. Sims and Zha, 1998). We re-estimate the model for each period T and compute iterative forecasts of the key variables $y_{i,T+h}$ from which quarter-on-quarter growth rate forecasts $\Delta y_{i,T+h}$ are computed.[6]

2.2 Factor models (FAAR, FAVAR, BFAVAR, DF)

For each of the $i = 1, ..., n$ variables of interest $\Delta y_{i,t}$ the $(k-1)$-dimensional set of predictors is defined as $\Delta X_{j,t} = (\Delta y_{1,t}, ..., \Delta y_{i-1,t}, \Delta y_{i+1,t}, ..., \Delta y_{n,t}, \Delta x_{1,t}, ..., \Delta x_{m,t})$. We standardize $\Delta X_{j,t}$ to have zero mean and unit variance in order to obtain $\Delta X_{j,t}^*$, the standardized predictor matrix.

[5] The unrestricted VAR without shrinkage seems to be overparameterized which yields to a worse forecasting performance than a small BVAR with shrinkage. Since small BVARs have been successfully used for a long time in forecasting (see e.g. Litterman, 1986), we regard them as the more suitable benchmark model.

[6] Since we only want to use information up to the estimation period T this implies that we have to calculate the hyperparameter λ, the prior means for the stationary variables μ_i as well as σ_i for each estimation separately.

Assume that $\Delta X^*_{j,t}$ can be represented by two components which are mutually orthogonal to each other and unobservable, namely the common component χ_t and the idiosyncratic component ξ_t, so we have $\Delta X^*_{j,t} = \chi_t + \xi_t$. The basic idea of factor models is that the information contained in the common component χ_t can be aggregated into a vector of factors F_t of dimension $\kappa \leq (k-1)$ which are able to explain most of the variance of the predictor matrix $\Delta X^*_{j,t}$. The dimension of a large dataset can thus be reduced without loosing valuable information.

In general the common component relates to the factors as: $\chi_t = \sum_{l=0}^{s} \eta_l F_{t-l}$. Depending on the lag structure that is assumed we can distinguish two model variants: the static factor model with $s = 0$ and the dynamic factor model with $s > 0$.

2.2.1 Static Factor Models (FAAR, FAVAR, BFAVAR)

From the standardized set of predictors $\Delta X^*_{j,t}$ we first extract $j = 1, ..., r$ factors $F_{j,t}$ via static principal component analysis. Following Stock and Watson (2002a) we use them to estimate a simple factor augmented autoregression (FAAR) $\Delta y_{i,t+h} = \rho_0 + \rho_1 \Delta y_{i,t} + ... + \rho_p \Delta y_{i,t+1-p} + \gamma_1 F_{1,t} + ... + \gamma_r F_{r,t} + \epsilon_t$ for each $h = 1, ..., 8$ and compute direct forecasts $\Delta y_{i,T+h}$.

As an alternative, we implement the approach proposed by Bernanke et al. (2005) according to which the following factor augmented vector autoregression (FAVAR) is to be estimated: $Z_t = c + B_1 Z_{t-1} + ... + B_p Z_{t-p} + \epsilon_t$ to allow for a more dynamic structure. Following Faust and Wright (2009) we include the variable to be predicted and the factors extracted from the set of predictors in the estimation, i.e. $Z_t = (y_{i,t}, F_{1,t}, ..., F_{r,t})'$.[7] Here the variables are included in log-levels to use information that is possibly contained in the trends. The FAVAR forecasts are computed iteratively, meaning that we compute $Z_{T+1} = c + \hat{B}_1 Z_T + ... + \hat{B}_p Z_{T+1-p}$ which is then iterated forward to obtain Z_{T+h} for $h = 2, ..., 8$.

As a third variant of the static factor model we implement a Bayesian factor augmented vector autoregression (BFAVAR). Here the factor augmented VAR $Z_t = c + B_1 Z_{t-1} + ... + B_p Z_{t-p} + \epsilon_t$ with $Z_t = (y_{i,t}, F_{1,t}, ..., F_{r,t})'$ is not estimated via OLS but rather via the Bayesian approach. The prior is set in a manner analogous to the large Bayesian VAR with the following two exceptions. First, we set the prior coefficient mean for the factors equal to zero to account for the fact that the factors have been extracted from the standardized predictor matrix $\Delta X^*_{j,t}$. Secondly, we set the hyperparameter $\lambda = 0.1$, a standard value in the literature (see fo example Litterman, 1986). As in the FAVAR, we include trending variables in log-levels and compute forecasts iteratively.

For each estimation period T the number of lags used in the FAAR as well as in the FAVAR and the BFAVAR estimation are obtained via the Bayesian information criterion. For the determination of the optimal number of factors r we use the information criterion proposed by Bai and Ng (2002).

[7] We also estimate a FAVAR (and a Bayesian FAVAR as desribed below) that includes a small set of core variables (including the variable to be predicted) and the factors (see e.g. Bernanke and Boivin, 2003; Bańbura et al., 2010). The forecasting performance of this alternative, however, is worse, so that we do not include this model in the main results.

2.2.2 Dynamic Factor Models (DF)

We set up a dynamic factor model in the spirit of Forni et al. (2003) and Forni et al. (2005) as laid out in Schumacher (2007). This implies extracting $j = 1, ..., q$ dynamic factors $\tilde{F}_{j,t}$ from the standardized set of predictors $\Delta X^*_{j,t}$ via dynamic principal component analysis in the frequency domain. Defining $\tilde{F}^*_t = (\tilde{F}'_{j,t}, \tilde{F}'_{j,t-1}, ..., \tilde{F}'_{j,t-s})'$ as vector of contemporaneous and lagged factors with dimension $r = q(s + 1)$, the dynamic factor model can be rewritten as a static factor model $\chi_t = \eta \tilde{F}^*_t$. The factors $\tilde{F}^*_t$ are used to augment a simple autoregression from which direct forecasts are computed. The number of lags of the dependent variable included in the estimation is determined via the Bayesian information criterion. For the number of dynamic factors q we do not apply an information criterion. The reason is that the information criterion proposed by Bai and Ng (2002) to determine the optimal number of static factors for the predictor matrix $\Delta X^*_{j,t}$ results in $r^* = 1$ for all estimation samples. From this we conclude that the optimal number of dynamic factors for the same matrix can only be $q = 1$ as well.

2.3 Model averaging (EWA, BMA)

For each of the $i = 1, ..., n$ variables of interest $\Delta y_{i,t}$ we set up m simple autoregressive distributed lag models $\Delta y_{i,t+h} = \rho_0 + \rho_1 \Delta y_{i,t} + ... + \rho_p \Delta y_{i,t+1-p} + \beta_j \Delta x_{j,t} + \epsilon_{j,t}$ for $j = 1, ..., m$. The general idea of model averaging is to compute a forecast $\Delta y^j_{i,T+h}$ with each model j and aggregate these m model-specific forecasts into one final forecast, i.e. $\Delta y_{i,T+h} = \sum_{j=1}^m \omega_j \Delta y^j_{i,T+h}$, where ω_j denotes the weight given to model-specific forecast $\Delta y^j_{i,T+h}$. The model-specific forecasts are obtained as direct forecasts, analogously for example to the FAAR model discussed above.

According to the specification of the weight ω_j that is attributed to each model-specific forecast two model averaging approaches can be distinguished. The first approach is Equal Weighted Averaging (EWA) as in Stock and Watson (2003, 2004), where the m simple models are estimated via OLS and $\omega_j = \omega = \frac{1}{m}$.

Alternatively, we consider Bayesian Model Averaging (BMA) as laid out in Wright (2009), where each of the model-specific forecasts $\Delta y^j_{i,T+h}$ is weighted with the posterior probability of the respective model $P(M_j)$, i.e. $\omega_j = P(M_j)$.[8] We implement the prior belief that each of the j models is equally likely to be true. Technically, $P(M_j)$ is obtained by dividing the model-specific likelihood $L_j = (1 + \phi)^{-K/2} S_j^{-T/2}$ with $S_j = \sqrt{(\hat{\epsilon}_{j,t})'(\hat{\epsilon}_{j,t}) - (\hat{\epsilon}_{j,t})' X_{j,t}(X'_{j,t} X_{j,t})^{-1} X'_{j,t}(\hat{\epsilon}_{j,t}) \frac{\phi}{1+\phi}}$ and $\hat{\epsilon}_{j,t} = \Delta y_{i,t+h} - B^*_j X_{j,t}$ by the sum of all model likelihoods for a specific horizon h. Here K denotes the number of parameters in the regressor matrix $X_{j,t} = (1, \Delta y_{i,t}, ..., \Delta y_{i,t+1-p}, \Delta x_{j,t})$ and $B^*_j = (\rho^*_0, \rho^*_1, ..., \rho^*_p, \beta^*_j)$ is the coefficient prior mean. For the coefficients of the lagged dependent variable the prior mean is obtained from the auxiliary autoregression $\Delta y_{i,t+h} = \rho^*_0 + \rho*_1 \Delta y_{i,t} + ... + \rho^*_p \Delta y_{i,t+1-p} + \epsilon_{j,t}$, whereas β^*_j is set to zero. The hyperparameter ϕ governs the tightness of the prior. Following Faust and Wright (2009) we set $\phi = 2$. To compute the model-specific forecasts for each variable the model-specific posterior mean of the coefficients $B^{**}_j = \frac{B^{OLS}\phi}{1+\phi} + \frac{B^*}{1+\phi}$ is used as parameter estimate in the autoregressive distributed lag equation. The number of lags p used in the estimation is obtained via the Bayesian information criterion.

[8] The model-specific posterior probability $P(M_j)$ is calculated in each estimation period T for each forecasting horizon h. For simplicity however, we will omit the respective subscripts.

2.4 Small benchmark models (AR, VAR, BVAR)

In order to evaluate the relative forecasting performance of the three large scale approaches we set up a number of small benchmark models which are described below.

2.4.1 Univariate Autoregression (AR)

We compute two types of autoregressive forecasts: direct and recursive. For the direct forecast we estimate univariate autoregressions $\Delta y_{i,t} = c_h + \sum_{j=1}^{p} \rho_{j,h} \Delta y_{i,t-h-j+1} + \epsilon_{t,h}$ for each of the $i = 1, ..., n$ variables of interest separately for each forecasting horizon h. The forecasts of each variable are obtained from the respective model for forecasting horizon h: $\Delta y_{i,T+h} = c_h + \sum_{j=1}^{p} \rho_{j,h} \Delta y_{i,T-j+1}$. For the recursive forecast we estimate the univariate autoregression $\Delta y_{i,t} = c + \sum_{j=1}^{p} \rho_j \Delta y_{i,t-j} + \epsilon_t$ and iterate this equation forward to get forecasts for each variable $\Delta y_{i,T+h}$. The number of lags p included in the estimation is obtained via the Bayesian information criterion in both cases.

2.4.2 Vector Autoregression (VAR)

We estimate an unrestricted vector autoregressive model $Y_t = c + B_1 Y_{t-1} + B_2 Y_{t-2} + ... + B_p Y_{t-p} + \epsilon_t$ where $Y_t = (y_{1,t}, ..., y_{n,t})'$ is a vector containing the variables to be forecast in log-levels. The lag length p is determined via the Bayesian information criterion. The vector of forecasts Y_{T+h} is computed by iterating the model forward.

2.4.3 Small Bayesian Vector Autoregression (BVAR)

We also implement the Bayesian counterpart to the unrestricted VAR in the variables to be forecast. The prior is set in a manner analogous to the large Bayesian VAR with the exception that we set the hyperparameter $\lambda = 0.1$ as in Litterman (1986). To be consistent we chose the lag length to be $p = 4$ as for the large Bayesian VAR. As for the unrestricted VAR the vector of forecasts Y_{T+h} is computed iteratively.

3 Data

Our dataset builds on the dataset used in Schumacher (2007) which we have updated to cover a sample period from 1978Q1 until 2013Q3. It consists of 123 macroeconomic variables in quarterly frequency that can be grouped into the following categories: composition of GDP and gross value added by sectors, prices, labor market, financial market, industry, construction, surveys and miscellaneous. The dataset includes time series of GDP and its components in real terms as well as their corresponding price indices and price adjusted series of gross value added for the main sectors of the German economy. Additionally we consider consumer and producer price indices and a terms of trade series. Among the labor market data that we take into account are employment, unemployment, hours worked, productivity, wages and vacancies. The financial market data contains a number of short- and long-term money market rates and bond yields as well as several German stock market performance indices. The industry and construction data are disaggregated and comprise amongst others production, turnover and new orders. In the surveys category we include sectoral data from the ifo survey on business situation and expectations, stocks and capacity utilization. Finally, we also consider current account data, a raw material world market price index and new passenger car registrations.

Most of the data is obtained via Datastream, while for the remainder our source is the German Federal Statistical Office. The data is seasonally adjusted. Natural logarithms are taken and annualized quarter-on-quarter growth rates are computed for time series not expressed in rates. Following Schumacher (2007) data which is only available for West Germany prior to 1991 is rescaled to the pan-German series to avoid regime shifts.

4 Forecasting Approach

We estimate the forecasting models on a moving window containing 60 observations. For shorter estimation windows the forecasting performance of all models deteriorates. The results obtained for a window of at least 60 observations are very similar to those under a recursive estimation scheme, where one additional quarter of data is added to the estimation sample for each forecasting round. For this reason and because we want to assess the statistical significance of the relative forecasting performance of the respective models using the test of equal unconditional finite-sample predictive ability (Giacomini and White, 2006), which can only be applied to a moving window forecasting scheme, we only report results obtained under he rolling window scheme.

To assess the relative and absolute forecasting performance of the models considered in this paper, we evaluate the forecasts of four key macroeconomic variables: the annualized quarter-on-quarter growth rate of GDP, annualized quarterly harmonized CPI inflation, the three-months money market interest rate and the unemployment rate in percent of the civilian labor force. In section 5.5, we extend the analysis to a larger set of variables to check which of the large data methods is suited best to simultaneously forecast a larger set of core variables.

For the evaluation of the absolute and relative forecasting performance of the different models we focus on two measures. First, we run Mincer-Zarnowitz regressions and check whether the forecasts are unbiased and efficient. This allows us to assess the absolute forecasting accuracy of each model. Secondly, we compute root mean squared prediction errors (RMSE) to compare the relative performance of the different forecasting models.

The forecasts produced by any model should not be systematically higher or lower than the actual value of the variable to be forecast given that the forecast is based on a symmetric loss function. Otherwise the forecast errors would be predictable and the forecasts would be biased.[9] To test whether the forecasts obtained by the different models are optimal in the sense that the forecast errors are unpredictable, we compute Mincer-Zarnowitz regressions (see Mincer and Zarnowitz, 1969). This implies regressing data realizations, y^r_{T+h}, on a constant and the

[9]An exception is the case of an asymmetric loss function that might sometimes be more appropriate for forecasts for fiscal and monetary policy purposes (see e.g. the discussion in Wieland and Wolters, 2013). Yet even if forecasts for fiscal or monetary policy purposes are based on asymmetric loss functions, it is still interesting to check whether forecasts are biased and to assess whether such a bias can be rationalized by assuming a specific asymmetric loss function.

forecasts, y^f_{T+h}:

$$y^r_{T+h} = \alpha + \beta y^f_{t+h}. \tag{3}$$

If the intercept estimate $\hat{\alpha}$ is significantly different from zero, the forecast systematically deviates from the actual data. A slope estimate $\hat{\beta}$ that is significantly different from one indicates that the forecast is inefficient because it systematically over- or underpredicts deviations from the mean. This implies that the residual variance in the regression does not equal the variance of the forecast errors. We run the Mincer-Zarnowitz regressions and estimate Newey-West standard errors with the number of lags equal to the forecast horizon in order to account for serial correlation of overlapping forecasts. To test for forecast bias we conduct F-tests of the joint null hypothesis $\hat{\alpha} = 0$ and $\hat{\beta} = 1$. The R^2 from the Mincer-Zarnowitz regressions shows whether the forecasts contain information about actual future macroeconomic dynamics. It can directly be interpreted as the fraction of the variance in the data that is explained by the forecasts. This fraction will be always below 1 since there are shocks and idiosyncrasies that no economic model can capture.

To compare the relative forecasting accuracy of the different models we compute RMSEs. We report the absolute level of the RMSEs for the simple univariate autoregressive forecast as a benchmark. For the remaining models the RMSEs are computed relative to this benchmark, which implies that values smaller than 1 indicate that the forecast performance of a specific model is better than that of the simple univariate autoregression. To assess the statistical significance of the differences in forecasting performance we conduct a test of equal unconditional finite-sample predictive ability (see Giacomini and White, 2006) using a symmetric loss function. This test can be applied to nested models, meaning that one model can be obtained from another model by imposing certain parameter restrictions, as well as non-nested models. It thus provides a coherent framework for comparing a large number of different forecasting models as is the case in this paper. Asymptotic p-values are computed using Newey-West standard errors to account for serial correlation of the forecast errors.

The theoretical properties of the RMSEs crucially depend on the persistence of the time series. To fix ideas consider the RMSE of a random walk forecast. If a time-series y_t follows a random walk, i.e. $y_t = y_{t-1} + u_t$ with $u_t \sim iidN(0, 1)$, then the forecast $y_{T+h|T}$ is given by y_T and the forecast error for horizon h is given by $e_{T+h|T} = \sum_{j=1}^{h} u_{T+j}$. Hence, the population RMSE is given by $\sqrt{E(e^2_{T+h|T})} = h$ and grows linearly with the forecast horizon. If a time-series however follows an AR(1) process $y_t = \gamma y_{t-1} + u_t$ with $0 < \gamma < 1$, then for a known γ the h-step ahead forecast error is given by $e_{T+h|T} = \sum_{j=1}^{h} \gamma^{j-1} u_{T+j}$ and the population RMSE is given by $\sqrt{E(e^2_{T+h|T})} = \sqrt{(1-\gamma^{2h})/(1-\gamma^2)} \to 1/\sqrt{1-\gamma^2}$ for $h \to \infty$ (see Del Negro and Schorfheide, 2013, for a detailed exposition). Thus, if γ is small the RMSE is relatively flat, while the RMSE increases strongly with horizon h for large γ. For time series with little persistence such as quarterly output growth and inflation the RMSE can therefore expected to be flat, whereas it should be increasing with forecast horizon h for time series which are highly persistent such as interest rates and the unemployment rate.

5 Results

In this section, we evaluate the absolute and relative forecasting performance of the different forecasting models for the four key macroeconomic variables described above. We thus have $\{\Delta y_{i,t}\}_{i=1}^{4} = \{\Delta gdp_t, \Delta cpi_t, i_t^s, u_t\}$. We report results for all forecasting models of section 2, except for those which are strictly dominated by a related alternative. For example, the BFAVAR always performs much better than the FAVAR model, so we do not report results for the latter. From the estimation of the static factor augmented autoregression (FAAR) we find that it is optimal to include only one factor into the estimation, thus the dynamic factor model (DFM) collapses to the static case and we drop it. Finally, the univariate autoregression performs always slightly better when estimated iteratively so we choose this variant as a benchmark instead of the directly estimated variant.

5.1 Output growth

Figure 1 shows the German GDP growth series as well as the forecasts of the AR benchmark and those obtained by one representative variant of each of the three large data methods, namely the LBVAR, the BFAVAR and the BMA.[10] The shaded periods show recessions as dated by the Economic Cycle Research Institute. As can be seen very clearly the GDP growth series shows very little persistence and can thus be expected to be very hard to predict. According to the Bayesian information criterion the optimal number of lags to include in the AR benchmark model alternates between 0 and 1 except for four quarters from 2005Q4 onwards and three quarters from 2008Q2 onwards for which we find $p = 4$. Forecasting models that predict a quick return to average GDP growth rates should therefore be able to predict German GDP growth more precisely.

Panel (a) in table 1 shows the absolute RMSEs for the AR benchmark model and the relative RMSEs for the other forecasting models for GDP growth. The absolute RMSEs of the AR model are quite large and flat over the different forecast horizons. This is in line with what can be expected for forecasts of a time series with a low persistence (see discussion above). Table entries in bold indicate that the null hypothesis of unbiasedness based on the F-test for the two coefficients in the Mincer-Zarnowitz regression (equation 3) cannot be rejected. This is the case for the AR, the BFAVAR and for some horizons also for the LBVAR and the EWA model. The relative RMSEs indicate that the gains in forecasting accuracy for GDP growth obtained by the three large scale approaches are generally at best moderate. The BFAVAR is the only model that is able to beat the univariate autoregression over all forecasting horizons, but the improvements are minor and insignificant. The forecasts of the LBVAR and the EWA are just as accurate as the AR forecasts, while the FAAR, BMA, VAR and the small BVAR perform worse. This is also reflected in figure 1: The AR and BFAVAR forecasts predict a return to the sample average after only one quarter, while the forecasts obtained by the LBVAR need slight additional adjustments over the following quarters. In contrast to that the BMA yields by far the most volatile forecasts.

[10] In this figure and the ones following for the other key variables we plot only two forecasts per year to preserve clarity of the figure.

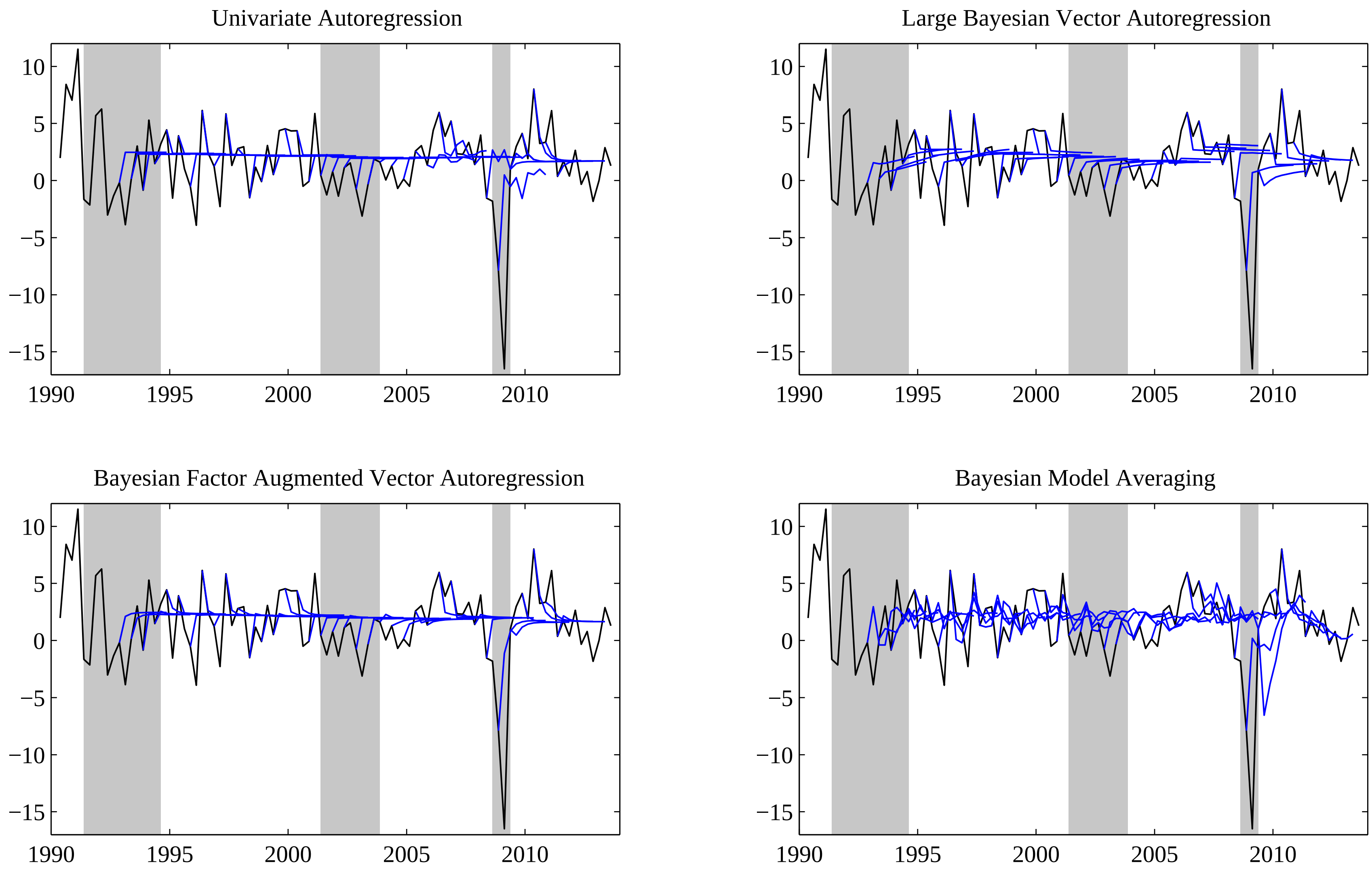

Figure 1: GDP growth forecasts.

Table 1: RMSEs of forecasting models

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	**3.49**	**0.97**	1.05	**0.96**	1.09	**1.00**	1.06	1.01
2	**3.52**	**0.99**	1.07	**0.98**	1.05	**1.00**	1.07	1.03
3	3.55	1.01	1.05●	**0.97**	1.02	1.01	1.10	1.02
4	**3.52**	1.03	1.05•	**0.98**	1.01	1.02	1.12	1.04
8	**3.47**	1.03	**0.98**	**0.99•**	**0.98**	**0.98**	1.11	1.03

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	1.52	**0.88•**	1.03	1.02	0.97•	0.98•	0.92	0.90•
2	1.44	**0.87•**	1.02	1.05	0.93•	0.97•	0.95	**0.92**
3	1.46	0.88●	1.00	1.05	0.95	0.95	1.00	0.93
4	1.51	0.87●	1.03	1.02	0.96	0.94•	1.02	0.92
8	1.61	0.82•	1.09	0.98●	0.93	0.93	1.19	0.94

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	0.38	**0.90●**	0.95	0.98	0.97	0.95•	1.00	0.98
2	0.75	**0.84•**	0.94	0.95	0.93•	0.94•	0.93	0.93•
3	1.07	**0.82●**	0.99	0.94	0.95•	0.96•	0.91	0.93•
4	1.37	0.83•	0.99	0.93	0.95•	0.97•	0.90	0.91•
8	2.28	0.82	1.08	0.86●	1.12	1.04	0.91	0.85•

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	**0.24**	**1.03**	**0.95**	**1.05**	**0.97•**	**0.98●**	**1.26•**	**1.05**
2	**0.44**	**0.98**	**0.99**	**1.03**	**0.92•**	**0.97•**	**1.25•**	**1.04**
3	**0.63**	**0.96**	**1.04**	**1.02**	**0.91•**	**0.98**	**1.25•**	**1.04**
4	**0.81**	**0.96**	**1.08**	**0.99**	**0.90•**	**0.98**	**1.26**	**1.04**
8	**1.32**	**1.01**	**1.19**	**0.97**	**0.92**	**1.07**	**1.40•**	**1.05**

Notes: For the AR the absolute RMSEs are shown, while for the remaining forecasting models the RMSEs are relative to the AR benchmark. The symbols ●, •, •, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively, while bold numbers imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level.

In addition to the large scale approaches and the small benchmark models, we analyze the predictive content of the ifo business climate index for German GDP growth. The ifo index is a leading indicator and often referenced to as the most important benchmark when forecasting German GDP growth (see e.g. Henzel and Rast, 2013).[11] We use the ifo index and the subindex covering business expectations for the next six months and regress GDP growth on a constant and the respective lagged indicator as in Henzel and Rast (2013): $\Delta y_t = \alpha_h + \beta_h \text{ifo}_{t-h} + \epsilon_t$. The forecasts for the different horizons are computed as $\Delta y_{T+h} = \hat{\alpha}_h + \hat{\beta}_h \text{ifo}_T$.

In figure 2 we plot the forecasts and one can see that the ifo indicators lead to more dynamics than the AR forecasts. Table 2 reports the RMSEs relative to the AR benchmark and shows that at least for the one-quarter ahead forecast the ifo expectations index is able to substantially

[11] The ifo index is based on a monthly survey where about 7000 firms report their assessments of the current business situation and their expectations for the next six months. From these two assessments the overall ifo business climate index is calculated.

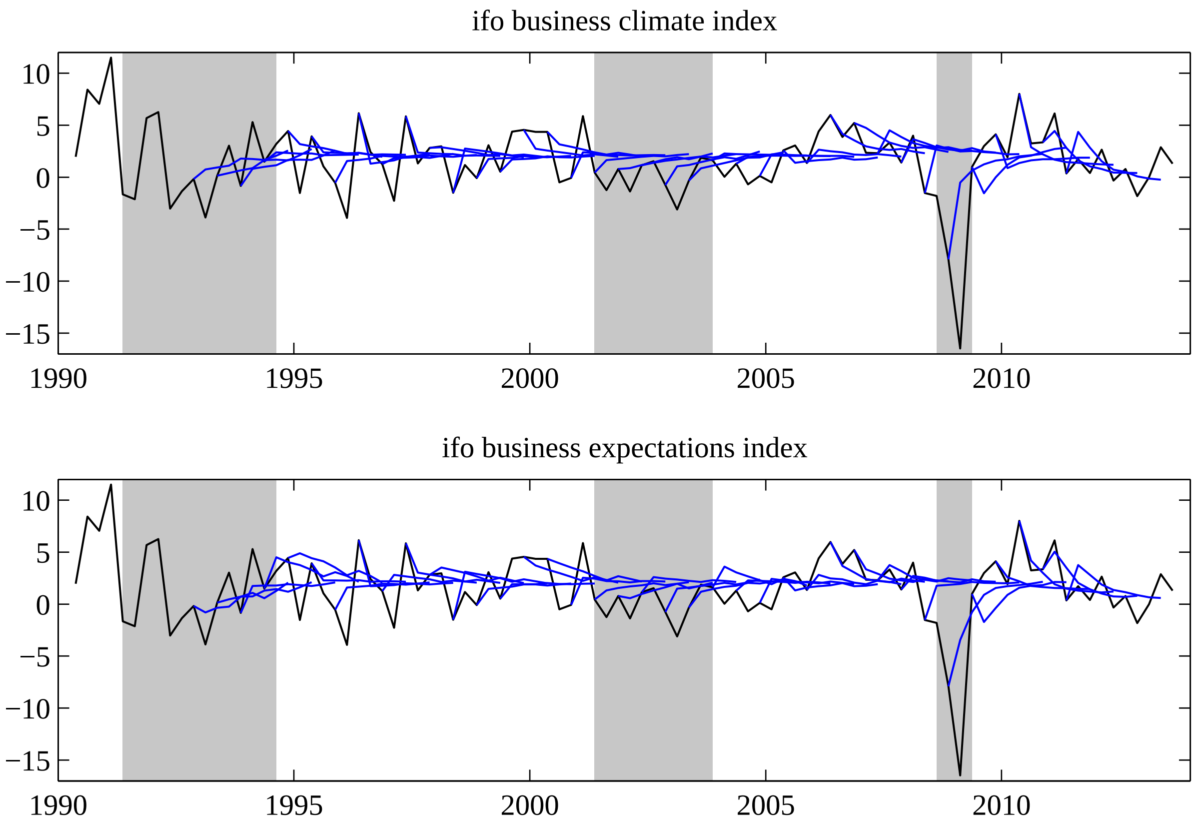

Figure 2: GDP growth forecasts with ifo leading indicators.

reduce the RMSE relative to the AR forecast by 10%. However, the reduction is not statistically significant (p-value: 0.17) and the forecast is biased. The ifo indicator is also included in the large data set used to compute forecasts for the large scale models, yet in a disaggregated version. Apparently, it is not given enough weight to in these models to improve upon the AR forecast.

Table 2: RMSEs for AR and ifo leading indicators

Output growth

horizon	AR	ifo climate	ifo expecations
1	**3.49**	**0.97**	0.89
2	**3.52**	**1.00**	**0.98**
3	3.55	1.00	1.01
4	**3.52**	1.00	1.00
8	**3.47**	**1.00**	**0.99**

Notes: For the AR the absolute RMSEs are shown, while for the remaining forecasting models the RMSEs are relative to the AR benchmark. The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively, while bold numbers imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level.

Given the low persistence of German GDP it does not seem surprising that the forecasts of all models also fail to convey a lot of information about the actual dynamics of this variable. Panel (a) in table 3 shows the R^2 of the Mincer-Zarnowitz regressions for the different models. The entries indicate that not even 10 percent of the variance in the data is explained by the forecasts of any of the different models. This confirms that German GDP growth is extremely difficult to predict due to its low persistence. Adding more information by using a large dataset for the forecasting process apparently only leads to marginal improvements in forecasting accuracy over the simple AR benchmark. The ifo expectations based indicator has more forecasting power for the one quarter ahead forecast. The R^2 for the one-step ahead forecast is 0.24 (not shown in table 3). However, it is close to zero for the higher horizons.

Table 3: R^2 of the Mincer-Zarnowitz regressions

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	0.02	0.03	0.06	0.07	0.01	0.02	0.01	0.01
2	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
3	0.02	0.05	0.01	0.01	0.00	0.02	0.00	0.01
4	0.01	0.11	0.01	0.02	0.00	0.01	0.00	0.01
8	0.02	0.06	0.01	0.00	0.01	0.02	0.01	0.00

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	0.06	0.13	0.05	0.04	0.08	0.06	0.14	0.14
2	0.02	0.04	0.02	0.00	0.05	0.02	0.07	0.05
3	0.01	0.01	0.05	0.00	0.03	0.03	0.03	0.02
4	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
8	0.02	0.03	0.07	0.02	0.04	0.02	0.05	0.04

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	0.95	0.95	0.96	0.96	0.96	0.96	0.95	0.96
2	0.84	0.84	0.86	0.85	0.87	0.86	0.85	0.85
3	0.71	0.68	0.75	0.71	0.75	0.74	0.74	0.72
4	0.58	0.50	0.64	0.56	0.62	0.61	0.62	0.58
8	0.26	0.04	0.27	0.22	0.24	0.27	0.32	0.26

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA	VAR	BVAR
1	0.97	0.97	0.97	0.96	0.97	0.97	0.95	0.96
2	0.89	0.90	0.90	0.89	0.91	0.90	0.84	0.88
3	0.79	0.82	0.78	0.79	0.83	0.80	0.68	0.77
4	0.68	0.72	0.62	0.69	0.74	0.69	0.50	0.65
8	0.29	0.42	0.09	0.40	0.37	0.18	0.02	0.25

The failure of economists to predict the Great Recession of 2008/2009 has led to severe criticism of macroeconomic forecasts and the macroeconomics profession in general. Therefore, in what follows, we study whether the three large data forecasting methods would have been able to forecast the Great Recession. Figure 3 shows the forecasts of the annualized quarter-on-quarter GDP growth rate obtained by the AR benchmark model, the LBVAR, the BFAVAR, the BMA and the two ifo indicators considered above. The forecasts of all six methods look roughly similar. None of the models is able to predict a downturn in GDP in 2008. Once the recession hits, the models further do not predict a further deepening of the recession, but rather a return to positive growth rates after one quarter. The only exception to this is the model based on the ifo expectation index. Business expectations in Germany already dropped largely during the third quarter of 2008, while the largest decreases in GDP growth only occurred in

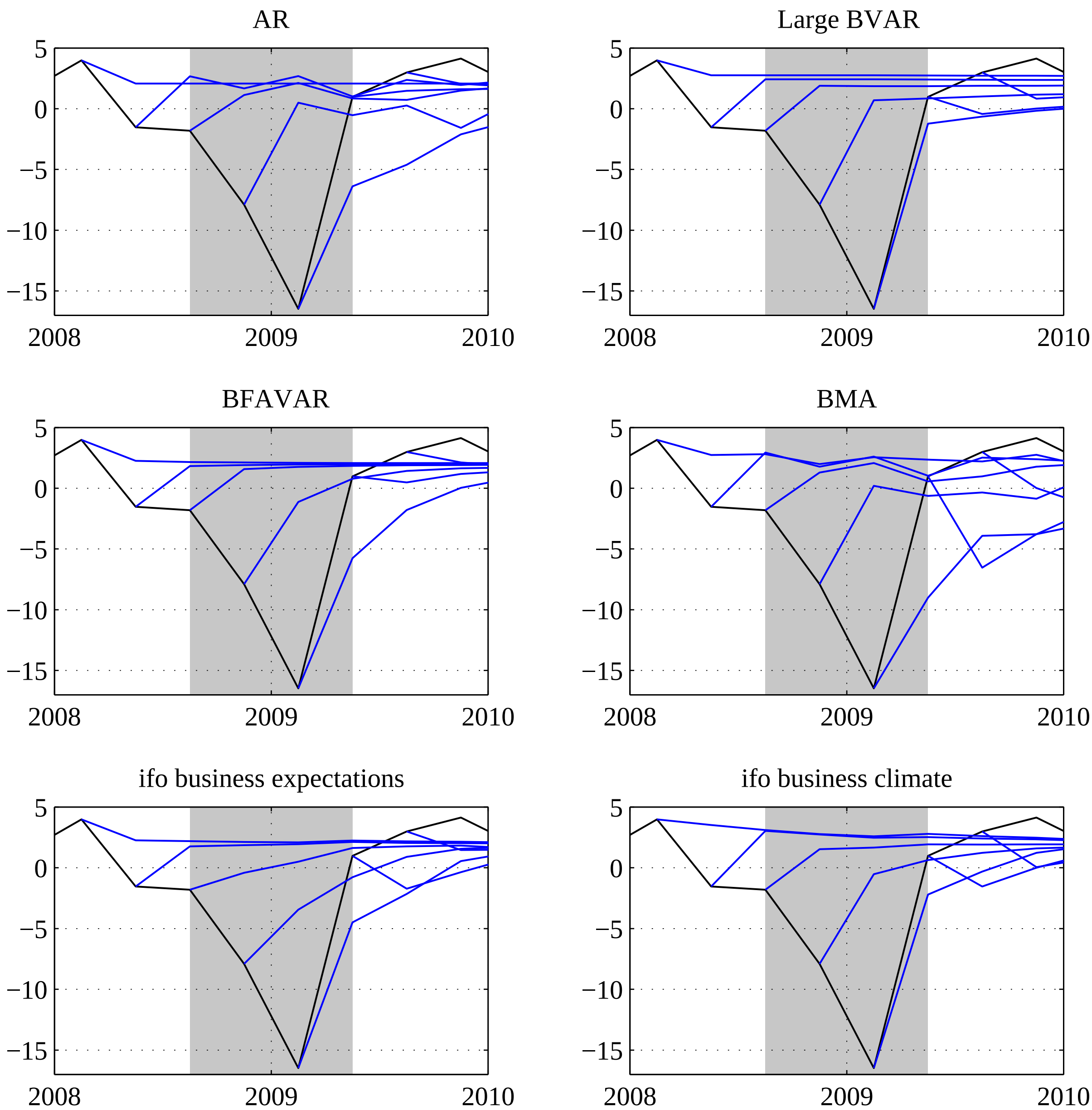

Figure 3: Great Recession GDP growth forecasts.

the fourth quarter of 2008 and the first quarter of 2009. Indeed, we find that the ifo expectation index predicts a negative GDP growth rate of -3.45% for the first quarter of 2009. However, by construction this model can hardly predict a further deepening of the recession. As the forecast is computed as: $\Delta y_{T+h} = \hat{\alpha}_h + \hat{\beta}_h \text{ifo}_T$, the coefficient $\hat{\beta}_h$ would need to increase strongly with the forecasting horizon h to predict a further deepening of the recession. At this point, it would be interesting to study whether regime switching models can account for the important information from the ifo business expectations index before and during recessions more accurately. They would, however, need to detect a regime switch early. We leave this exercise for future research.

As for the turning point of the Great Recession in the first quarter of 2009, again, none of the models is able to predict it. Once the turning point is reached most models underpredict the speed of the recovery. The LBVAR and the ifo business climate models are clear exceptions predicting the first two quarters of the recovery almost exactly. The predictions of these two models for the higher horizons are too low, but still more accurate than the predictions of the other forecasting models.

5.2 Inflation

Figure 4 shows the German CPI inflation rate series as well as the forecasts of the AR benchmark and those obtained by the LBVAR, the BFAVAR and the BMA. The graphs show that German CPI inflation is more persistent than GDP growth, but still shows many spikes, which are presumably hard to predict. The series also includes several changes in the trend which the LBVAR captures quite well. By contrast, the AR and the BFAVAR forecasts return to the previous trend inflation levels. As for GDP growth, the forecasts of the BMA model are more volatile.

Panel (b) in table 1 shows the RMSEs for CPI inflation. The absolute RMSEs shown for the AR model are less than half of those for GDP growth. Still persistence of quarterly inflation is quite low and thus the RMSEs do not increase with the forecast horizon h. The LBVAR significantly outperforms the AR benchmark with reductions in RMSEs ranging between 10 and 20%. It is also the only model that yields unbiased forecasts as indicated by the bold numbers. The small BVAR and both model averaging techniques (BMA and EWA) show smaller RMSEs than the AR model, however the reductions are only significant for a few horizons. By contrast, the factor models cannot beat the AR forecast. While the results for the LBVAR seem encouraging at first sight, we find that the informational content of the forecasts obtained by all models is weak. Panel (b) in table 3 shows the R^2 from Mincer-Zarnowitz regressions. It exceeds 10 percent only for the one quarter ahead forecasts for the large and the small BVAR, while for all other horizons and models the values are close to zero.

5.3 Interest rate

Figure 5 shows the 3-months money market rate series as well as the forecasts of the AR benchmark and those obtained by the LBVAR, the BFAVAR and the BMA. This time series is much more persistent than GDP growth or CPI inflation. Still the forecasting models have to predict some trend changes which might pose a difficulty for most models. In the evaluation sample the interest rate shows an overall downward trend with two temporary increases before the 2001 recession and the recent Great Recession in 2008. The AR, BFAVAR and BMA systematically overpredict the interest rate until about 2005. Interestingly, the LBVAR is able to adjust very quickly to the changes in the trend. Only at turning points the forecasts of the LBVAR are imprecise as the model captures the turning points only after they actually occurred.

Panel (c) in table 1 shows RMSEs for the short term interest rate. The absolute RMSEs for the AR model are much smaller than those for GDP growth and CPI inflation for short horizons but they increase with the forecast horizon reflecting the high persistence of the time series. For $h = 8$ the absolute RMSE of the AR is even higher than that for CPI inflation which indicates that the model has problems to capture the downward trend of the interest rate. The LBVAR, the small BVAR and both model averaging approaches (BMA and EWA) yield significantly

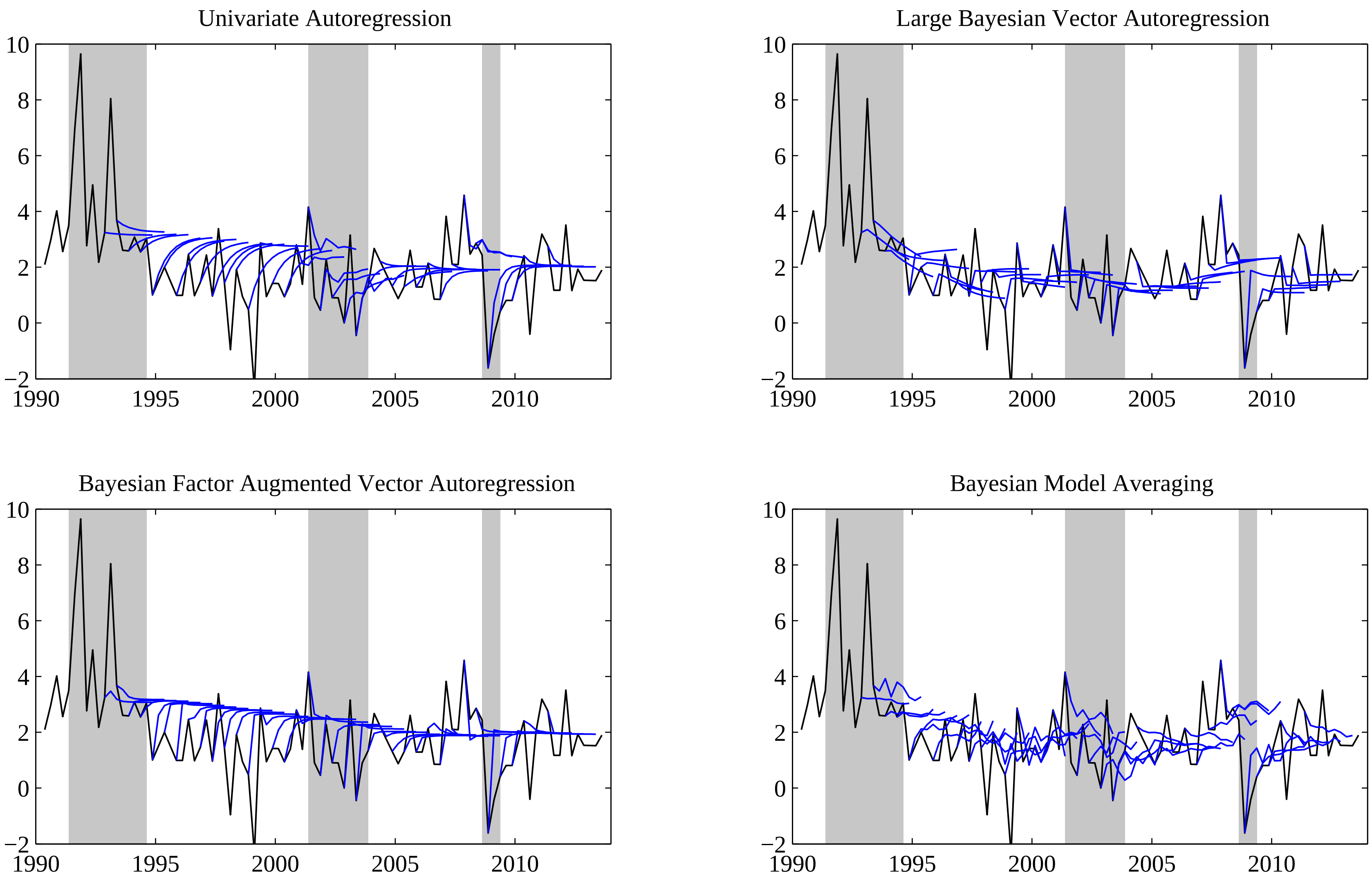

Figure 4: Inflation rate forecasts.

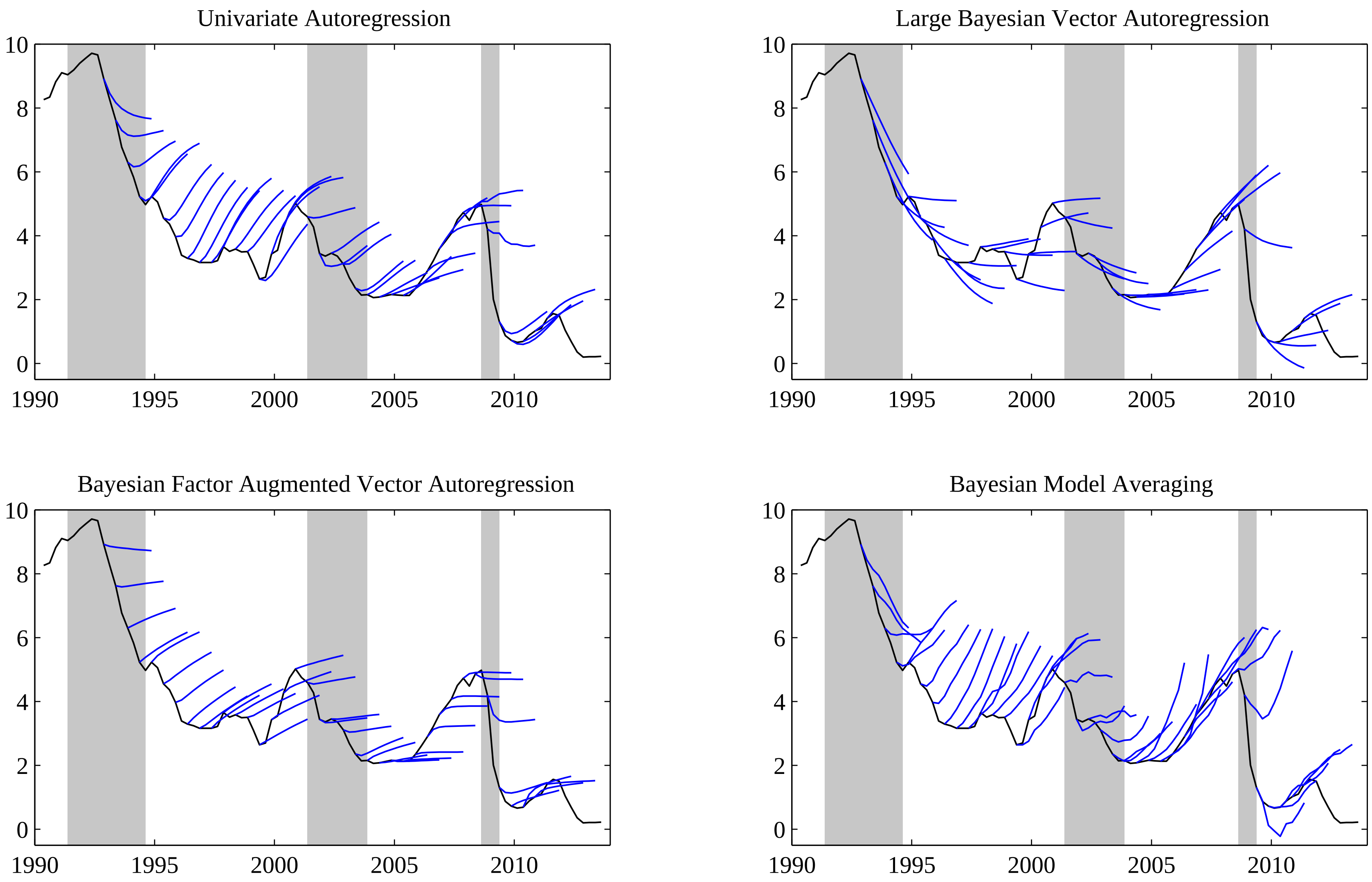

Figure 5: Interest rate forecasts.

better forecasts than the AR. For the LBVAR and the small BVAR the improvement is quite sizable and ranges between 10 and 20%. The forecasts of all other models are also slightly better than the AR forecasts, but not on a statistically significant level. The LBVAR is the only model that produces unbiased forecasts (as indicated by the bold numbers). Panel (c) in table 3 shows that the interest rate forecasts provided for all models are very informative. For $h = 1$ the forecasts of each model can explain more than 90% of the variance in the interest rate series, while for forecasts one year ahead still almost half of the variance can be explained by all model forecasts. However, this is certainly due to the high persistence and the extremely little variation in the interest rate series.

5.4 Unemployment rate

Figure 6 shows the German unemployment rate series as well as the forecasts of the AR benchmark and those obtained by the LBVAR, the BFAVAR and the BMA. The persistence of the unemployment rate series is very high, similar to the short-term interest rate. The unemployment rate, however, does not show one single trend, but increases until 1998, decreases until 2001, increases afterwards until 2005 and falls from there until the end of the sample. Due to these various trend changes, no model systematically over- or underestimates the unemployment rate. The AR and BMA forecasts even get most of the turning points right, while the LBVAR has more difficulties. At least it adjusts relatively quickly to the new trend after a turning point. The BFAVAR predictions resemble those of a random walk, which indicated that the model is not able to capture the most important dynamics in the unemployment rate.

These observations are reflected in the RMSEs in panel (d) of table 1. The absolute RMSEs for the AR are smaller than those for the other three key variables and they increase with the forecast horizon owing to the high persistence of the unemployment rate series. The AR forecasts are quite good as shown in figure 6 and the other models perform just as good, but not better. The exception is the BMA which reduces RMSEs significantly by up to 10% compared to the AR benchmark. One reason might be that the model predicts turning points even better than the AR as indicated in figure 6. The unrestricted VAR produces significantly less accurate forecasts than the AR benchmark. All forecasts are unbiased which reflects that the unemployment rate lies in a similar range in the estimation and the evaluation sample and shows no clear overall trend.

As for the interest rate, we find that the explanatory power of all forecast is extremely high for short forecasting horizons as shown in panel (d) of table 3. This can again be attributed to the high persistence in the unemployment rate series.

5.5 Forecasting a larger number of macroeconomic variables

Often forecasters are interested in the forecasts of a larger number of variables than those contained in the small set of key variables considered so far. In table 4 we show the average absolute RMSEs for the AR benchmark and the average relative RMSEs for all the remaining models for eleven core macroeconomic variables. Besides the four key variables considered thus far, this includes private consumption, machinery and equipment investment, wages, industrial

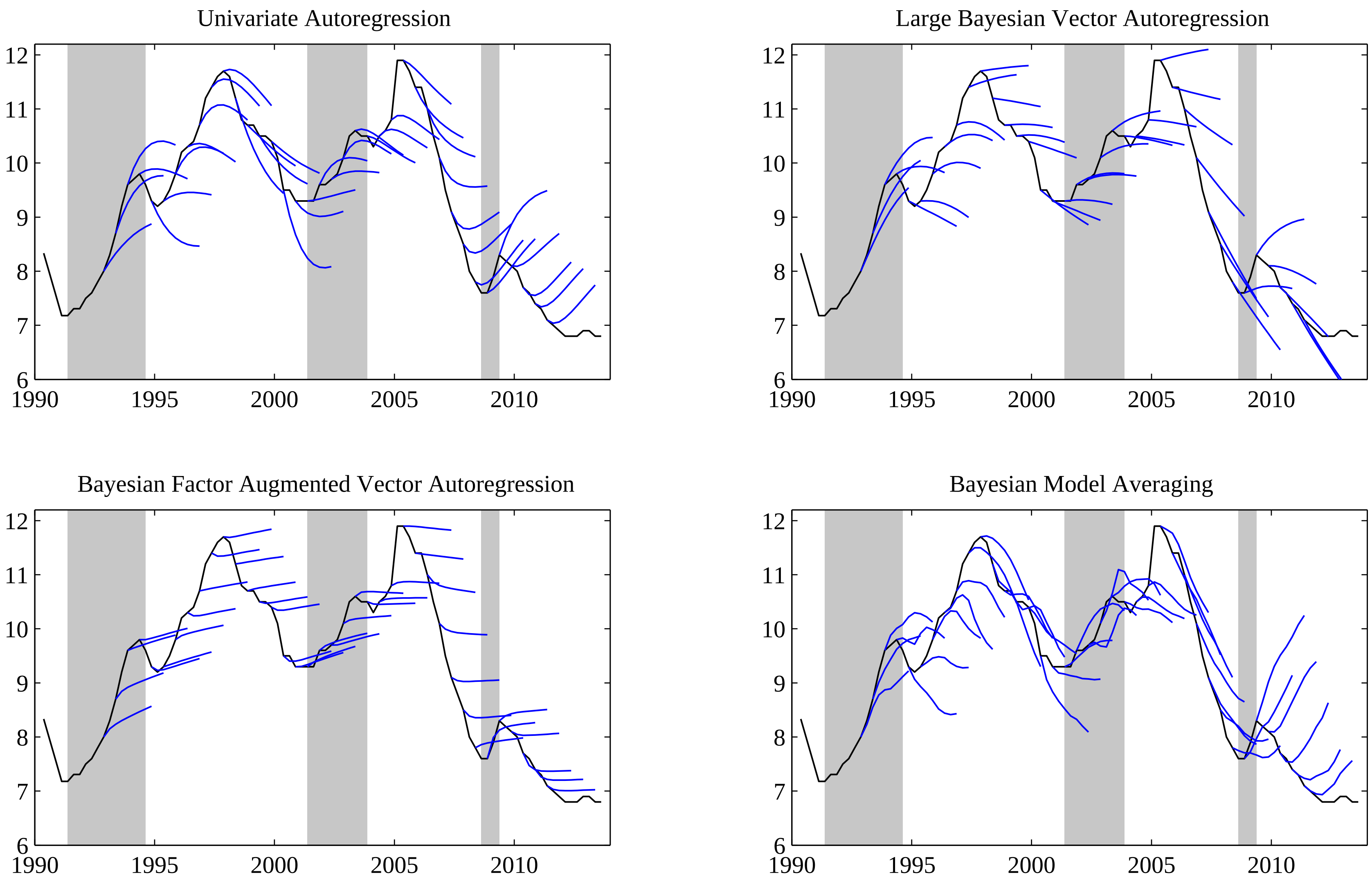

Figure 6: Unemployment rate forecasts.

Table 4: Average RMSE all variables

Average over all 11 variables

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4		0.93	1.01	0.98	0.96	0.97
1-8		0.93	1.03	0.97	0.99	0.98

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	3.52	1.00	1.05	0.98	1.04	1.01
1-8	3.50	1.02	1.04	0.98	1.02	1.00

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	1.48	0.87	1.02	1.03	0.95	0.96
1-8	1.53	0.86	1.03	1.01	0.95	0.96

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	0.89	0.85	0.97	0.95	0.95	0.96
1-8	1.44	0.84	1.01	0.92	1.00	0.98

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	0.53	0.98	1.01	1.02	0.92	0.98
1-8	0.84	0.98	1.09	1.00	0.92	1.01

(e) Private Consumption

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	2.81	0.98	1.04	0.98	0.97	0.98
1-8	2.72	0.99	1.08	0.99	1.00	1.00

(f) Machinery and Equipment Investment

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	13.82	0.95	0.98	0.93	0.98	0.97
1-8	13.78	0.98	0.96	0.94	0.97	0.96

(g) Wages

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	3.16	0.75	1.06	0.92	0.93	0.95
1-8	3.16	0.80	1.12	0.93	0.94	0.96

(h) Industrial Production

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	12.14	0.97	0.97	0.96	0.97	0.98
1-8	12.08	0.99	0.97	0.97	0.97	0.97

(j) PPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	3.79	1.00	0.90	1.00	0.94	0.95
1-8	3.84	1.00	0.96	0.99	0.99	0.98

(k) Long Term Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	0.62	0.95	1.03	0.97	0.97	1.00
1-8	0.90	0.89	1.04	0.91	1.06	1.01

(l) Current Account

horizon	AR	LBVAR	FAAR	BFAVAR	BMA	EWA
1-4	5.70	0.97	1.05	1.02	0.99	0.99
1-8	7.57	0.94	1.08	1.01	1.04	1.00

production, PPI inflation, a ten year interest rate and the current account balance. The average RMSEs are computed over the forecast horizons $h = 1$ until $h = 4$ and until $h = 8$, respectively to asses the relative short- and long-term forecasting performance of the models.

In the upper part of the table, we additionally summarize the average RMSEs over different forecast horizons for each model over all eleven variables. In terms of short term forecasting performance we find that with an average relative RMSE over forecasts up to one year ahead of 0.93 the LBVAR clearly dominates the remaining forecasting models. The FAAR performs worst, while the BFAVAR, the BMA and the EWA show a very similar forecasting accuracy.

For three specific variables dominance of the LBVAR over the AR benchmark is remarkable: when forecasting CPI inflation, the short term interest rate and wages the LBVAR is able to obtain average gains of 12%, 17% and 24%, respectively. There are only two variables, namely output growth and the current account balance, for which the LBVAR cannot outperform the AR benchmark, but yields forecasts that are just as good. The BMA and EWA dominate the average short term forecasting performance of the AR benchmark for nine or ten of the variables, respectively. However, the gains in forecasting accuracy are smaller than those for the LBVAR. For a total of seven of the variables, also the BFAVAR performs better than the AR in the short run. By contrast, the FAAR is dominated by the AR for more than half of the variables under consideration.

If we consider all forecasting horizons $h = 1, ..., 8$, the performance based ranking of the forecasting models remains unchanged. The LBVAR yields the lowest average relative RMSE over all variables, while the BFAVAR, the BMA and the EWA can only slightly outperform the AR benchmark. The FAAR again performs worst.

The LBVAR and the BFAVAR both dominate the remaining forecasting models for eight out of the eleven variables. The superior overall performance of the LBAVAR must therefore clearly be attributed to the outstanding performance of the model for CPI inflation, the short term interest rate and wages. For these variables the average gain in forecasting accuracy ranges between 10 and 18%, whereas the gains obtained by the BFAVAR are at best moderate. Over all forecasting horizons the BMA and EWA outperform the AR for six variables. For some variables the gains are quite sizable (for example for the unemployment rate forecast of the BMA). The FAAR, however, is dominated by the AR benchmark for more than half of the variables under consideration, as measured by the average relative RMSE for forecasts up to two years ahead.

The largest gain over the AR forecast over all forecast horizons and all variables that can be obtained by the three large scale approaches amounts only to 7%. The detailed tables in the appendix, that contain the results for each model for all the 11 variables, show that many variables have a very low persistence and thus their predictable component is only small. Among the variables with higher persistence there might possibly exist high collinearity so that even for these variables multivariate forecasting models can only yield modest gains over simple univariate forecasting models.

6 Conclusion

We have studied three different approaches that are able to use the information from a large dataset to forecast eleven core German macroeconomic time series. Overall, the large Bayesian VAR (LBVAR) performs best and yields more accurate forecasts than the simple univariate autoregression (AR) used as a benchmark for almost all variables we consider. For very few variables such as CPI inflation or wage growth the gains in terms of forecasting accuracy of the LBVAR even amount to a reduction of the root mean squared prediction error of 10 to 20% over the benchmark model. The Bayesian factor augmented VAR (BFAVAR) also yields quite good forecasts, but the gains are considerably smaller. Model averaging techniques can also improve slightly over the AR benchmark in terms of forecasting accuracy, but only for a considerably smaller number of variables.

In general, however, we find that the overall improvements of the large scale approaches upon the univariate benchmark are only modest and strongly depend on the variable to be forecast. One reason for this might be that some time series show very little persistence and are thus very hard to predict by univariate as well as multivariate forecasting models. Yet, even for time series with more persistence, the high collinearity in the large dataset seems to prevent large gains from the large-scale multivariate forecasting models over the simple AR benchmark.

Still, when forecasters are interested in simultaneously predicting a larger number of variables, large-scale forecasting models such as the LBVAR have the advantage that they can be used to coherently forecast many variables. This might be an advantage also when it comes to the interpretation of forecasts. Finally, we have also checked whether any of the forecasting model would have been able to forecast the Great Recession of 2008 and 2009 and find that none of them could predict the downturn of German GDP growth. Yet, once the turning point is reached the LBVAR precisely predicts the recovery, while the remaining methods underpredict the speed of the recovery.

References

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191221.

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

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*, 451–468.

Berg, T. O. and S. R. Henzel (2013). Point and density forecasts for the Euro area using many predictors: are large BVARs really superior? ifo Working Paper No. 155.

Bernanke, B., J. Boivin, and P. Eliasz (2005). Measuring monetary policy: A factor augmented vector autoregressive, (FAVAR) approach. *Quarterly Journal of Economics 120*, 387–422.

Bernanke, B. S. and J. Boivin (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics 50(3)*, 525–546.

Bernardini, E. and G. Cubadda (2014). Macroeconomic forecasting and structural analysis through regularized reduced-rank regression. *International Journal of Forecasting*, forthcoming.

Carriero, A., G. Kapetanios, and M. Marcellino (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics 26*, 735–761.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components? *Journal of Econometrics 146*, 318–328.

Del Negro, M. and F. Schorfheide (2013). DSGE model-based forecasting. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2. Elsevier Publications.

Faust, J. and J. H. Wright (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics 27(4)*, 468–479.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). Do financial variables help forecasting inflation and real activity in the Euro area? *Journal of Monetary Economics 50*, 1243–1255.

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*, 830–840.

Forni, M., M. Hallin, M.Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics 82*, 540–554.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Henzel, S. and S. Rast (2013). Prognoseeigenschaften von Indikatoren zur Vorhersage des Bruttoinlandsprodukts in Deutschland. *ifo Schnelldienst 66 (17),*, 39–46.

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.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics 4*, 25–38.

Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic Forecasts and Expectations*. NBER, New York.

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting 26*, 271–302.

Sims, C. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39(4)*, 949–968.

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*, 1167–1179.

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

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41(3)*, 788–829.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*, 405–430.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 135–196. Amsterdam: North Holland.

Wieland, V. and M. H. Wolters (2013). Forecasting and policy making. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, pp. 239–325. Elsevier Publications.

Wolters, M. H. (2014). Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics*, forthcoming.

Wright, J. (2009). Forecasting U.S. inflation by Bayesian model averaging. *Journal of Forecasting 28*, 131–144.

Appendix

Table 5: Large Bayesian VAR

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
					(a) relative RMSE						
1	0.97	0.88•	1.02	0.92	0.72●	0.94	1.15	0.90•	1.03	1.03	0.98
2	0.99	0.87●	0.95•	0.94	0.73●	0.94	0.96	0.84●	0.98	0.95	0.98
3	1.01	0.88●	0.98	0.98	0.76●	1.00	0.93	0.82●	0.96	0.92	0.96•
4	1.03	0.87●	0.97	0.96	0.78●	1.01	0.95	0.83•	0.96	0.90	0.95●
5	1.04	0.84•	0.97	0.97	0.81•	1.00	0.98	0.82∙	0.96	0.87	0.94●
6	1.04	0.85•	1.03	1.00	0.84∙	1.02	1.00	0.82	0.97	0.84	0.92●
7	1.04	0.85∙	1.02	1.02	0.85	1.00	1.01	0.83	0.99	0.81	0.92●
8	1.03	0.82∙	1.02	1.02	0.86	1.01	1.01	0.82	1.01	0.78∙	0.90●
					(b) p-value Mincer-Zarnowitz regressions						
1	0.46	0.87	0.00	0.68	0.12	0.81	0.17	0.20	0.75	0.49	0.12
2	0.13	0.16	0.01	0.69	0.17	0.61	0.33	0.12	0.61	0.42	0.15
3	0.01	0.03	0.01	0.16	0.08	0.14	0.10	0.07	0.48	0.40	0.15
4	0.00	0.01	0.00	0.03	0.02	0.04	0.04	0.04	0.37	0.40	0.14
5	0.00	0.01	0.00	0.01	0.01	0.03	0.04	0.02	0.29	0.43	0.12
6	0.00	0.00	0.00	0.01	0.00	0.05	0.06	0.01	0.23	0.42	0.11
7	0.00	0.00	0.00	0.01	0.00	0.11	0.08	0.01	0.19	0.37	0.10
8	0.00	0.00	0.00	0.01	0.00	0.24	0.13	0.00	0.16	0.28	0.08
					(c) R^2 Mincer-Zarnowitz regressions						
1	0.03	0.13	0.01	0.13	0.08	0.10	0.18	0.95	0.97	0.94	0.93
2	0.00	0.04	0.00	0.03	0.05	0.01	0.04	0.84	0.90	0.84	0.87
3	0.05	0.01	0.00	0.00	0.01	0.00	0.00	0.68	0.82	0.76	0.81
4	0.11	0.00	0.00	0.04	0.00	0.03	0.00	0.49	0.73	0.68	0.75
5	0.13	0.00	0.01	0.08	0.02	0.03	0.00	0.31	0.63	0.62	0.70
6	0.14	0.01	0.01	0.06	0.06	0.02	0.00	0.18	0.55	0.58	0.68
7	0.12	0.03	0.00	0.07	0.08	0.01	0.00	0.09	0.48	0.55	0.64
8	0.06	0.03	0.00	0.06	0.11	0.00	0.01	0.04	0.42	0.51	0.61

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 6: Static Factor Model (FAAR)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
						(a) relative RMSE					
1	1.05	1.03	1.08	0.96	0.99	0.93	0.92	0.95	0.95	1.02	1.05∙
2	1.07	1.02	0.97	1.02	1.06	0.96●	0.88∙	0.94	0.99	1.04	1.04
3	1.05●	1.00	1.02	0.96	1.14	1.00	0.90∙	0.99	1.04	1.02	1.06
4	1.05∙	1.03	1.07∙	0.96	1.06	0.99	0.92	0.99	1.08	1.02	1.06
5	1.04∙	1.06	1.10●	0.92	1.20	0.97	0.97	1.01	1.12	1.03	1.08∙
6	1.03∙	0.99	1.01	0.96	1.28	0.97	1.01	1.04	1.15	1.04	1.10∙
7	1.02∙	1.06	1.16•	0.97	1.14	0.94	1.04∙	1.07	1.18	1.05∙	1.11
8	0.98	1.09	1.22•	0.97∙	1.10	0.97	1.04	1.08	1.19	1.07•	1.13
						(b) p-value Mincer-Zarnowitz regressions					
1	0.00	0.00	0.00	0.04	0.00	0.13	0.44	0.00	0.95	0.00	0.04
2	0.00	0.00	0.00	0.01	0.00	0.09	0.31	0.00	0.92	0.00	0.07
3	0.00	0.00	0.00	0.17	0.00	0.10	0.28	0.00	0.91	0.00	0.07
4	0.01	0.00	0.00	0.04	0.00	0.08	0.12	0.00	0.87	0.00	0.05
5	0.01	0.00	0.00	0.28	0.00	0.36	0.04	0.00	0.78	0.00	0.03
6	0.02	0.00	0.00	0.31	0.00	0.90	0.00	0.00	0.68	0.00	0.02
7	0.05	0.00	0.00	0.48	0.00	0.92	0.00	0.00	0.57	0.00	0.02
8	0.34	0.00	0.00	0.59	0.00	0.79	0.01	0.00	0.47	0.00	0.01
						(c) R^2 Mincer-Zarnowitz regressions					
1	0.06	0.05	0.21	0.14	0.14	0.14	0.46	0.96	0.97	0.95	0.92
2	0.00	0.02	0.05	0.05	0.01	0.02	0.18	0.86	0.90	0.86	0.86
3	0.01	0.05	0.00	0.03	0.00	0.00	0.03	0.75	0.78	0.80	0.79
4	0.01	0.00	0.00	0.00	0.01	0.02	0.00	0.64	0.62	0.78	0.72
5	0.04	0.00	0.02	0.00	0.02	0.00	0.01	0.50	0.45	0.77	0.66
6	0.01	0.00	0.00	0.00	0.03	0.03	0.08	0.38	0.29	0.75	0.63
7	0.01	0.03	0.02	0.01	0.05	0.06	0.12	0.31	0.18	0.75	0.59
8	0.01	0.07	0.03	0.01	0.15	0.05	0.04	0.27	0.09	0.76	0.55

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 7: Equal Weighted Model Averaging (EWA)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
					(a) relative RMSE						
1	1.00	0.98●	0.97●	1.00	0.94⬤	0.99	0.98●	0.97⬤	0.99	1.01	1.00
2	1.00	0.97	0.95•	0.99	0.95⬤	0.98	0.99	0.98	0.99	1.01	0.99
3	1.01	0.96	0.99	1.01	0.97	0.99	0.97	0.99	0.99	0.99	1.00
4	1.02	0.95	1.02	0.95	0.98	0.98	0.96	0.99	0.99	0.99	0.99
5	1.01	0.98	1.00	0.93	0.97	0.96	0.97	1.00	1.00	1.01	0.99
6	0.99	0.96	1.02	0.97	0.95●	0.99	0.99	1.00	1.01	1.01	1.01
7	0.99	0.95	1.02	0.96	0.94●	0.96	1.04•	1.02	1.03	1.02	1.02
8	0.98	0.94	1.02	0.96•	0.93●	0.99	1.08●	1.04	1.04	1.04	1.03
					(b) p-value Mincer-Zarnowitz regressions						
1	0.10	0.00	0.00	0.06	0.00	0.07	0.59	0.00	0.84	0.00	0.04
2	0.08	0.00	0.01	0.08	0.00	0.07	0.08	0.00	0.86	0.00	0.05
3	0.02	0.00	0.00	0.05	0.00	0.24	0.01	0.00	0.86	0.00	0.04
4	0.03	0.00	0.00	0.06	0.00	0.24	0.01	0.00	0.85	0.00	0.03
5	0.02	0.00	0.00	0.17	0.00	0.69	0.03	0.00	0.83	0.00	0.03
6	0.23	0.00	0.00	0.20	0.00	0.28	0.01	0.00	0.82	0.00	0.02
7	0.32	0.00	0.00	0.65	0.00	0.87	0.00	0.00	0.82	0.00	0.01
8	0.31	0.00	0.00	0.66	0.00	0.75	0.00	0.00	0.84	0.00	0.01
					(c) R^2 Mincer-Zarnowitz regressions						
1	0.02	0.06	0.18	0.06	0.01	0.06	0.38	0.96	0.97	0.95	0.93
2	0.01	0.02	0.01	0.02	0.00	0.00	0.02	0.85	0.90	0.87	0.87
3	0.02	0.03	0.01	0.00	0.02	0.00	0.02	0.72	0.80	0.81	0.81
4	0.01	0.00	0.02	0.00	0.09	0.01	0.02	0.59	0.68	0.79	0.76
5	0.07	0.00	0.01	0.01	0.09	0.00	0.01	0.44	0.55	0.77	0.71
6	0.00	0.00	0.02	0.00	0.08	0.01	0.04	0.35	0.43	0.76	0.68
7	0.00	0.01	0.00	0.01	0.07	0.03	0.10	0.30	0.33	0.75	0.64
8	0.02	0.02	0.00	0.01	0.06	0.00	0.13	0.27	0.22	0.75	0.61

Notes: The symbols ⬤, ●, •, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 8: Bayesian Model Averaging (BMA)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
						(a) relative RMSE					
1	1.09	0.97∙	0.90•	1.04	0.91●	1.01	0.97∙	0.99	0.99	1.11●	1.00
2	1.05	0.94	0.95	0.98	0.96	0.97	0.96∙	1.01	0.98	1.00	0.99
3	1.02	0.97	1.02	1.02	0.98	0.98	0.98	1.01	1.01	0.99	0.99
4	1.01	1.00	1.04	0.96	0.96	0.99	0.97	1.02	0.99	0.98	0.97
5	1.01	1.00	1.01	0.94	0.94	0.97	1.01	1.03	0.97	0.99	0.98
6	0.98	0.97	1.03	0.97	0.94∙	1.01	1.06	1.06	0.94	1.00	1.03
7	0.98	0.97	1.04	0.97	0.93•	0.98	1.12	1.11∙	0.93	1.01	1.09
8	0.98	0.97	1.02	0.98	0.93•	1.00	1.16	1.12∙	0.93	1.04	1.11
						(b) p-value Mincer-Zarnowitz regressions					
1	0.00	0.00	0.00	0.01	0.00	0.01	0.48	0.00	0.75	0.00	0.04
2	0.01	0.00	0.00	0.08	0.00	0.10	0.16	0.00	0.65	0.00	0.06
3	0.02	0.00	0.00	0.03	0.00	0.40	0.01	0.00	0.69	0.00	0.05
4	0.03	0.00	0.00	0.05	0.00	0.05	0.01	0.00	0.74	0.00	0.04
5	0.06	0.00	0.00	0.11	0.00	0.33	0.00	0.00	0.80	0.00	0.03
6	0.31	0.00	0.00	0.22	0.00	0.04	0.00	0.00	0.85	0.00	0.02
7	0.23	0.00	0.00	0.61	0.00	0.53	0.00	0.00	0.90	0.00	0.01
8	0.19	0.00	0.00	0.33	0.00	0.53	0.00	0.00	0.91	0.00	0.01
						(c) R^2 Mincer-Zarnowitz regressions					
1	0.01	0.08	0.23	0.05	0.03	0.07	0.40	0.95	0.97	0.94	0.93
2	0.00	0.05	0.01	0.04	0.00	0.01	0.06	0.84	0.90	0.87	0.87
3	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.72	0.79	0.81	0.82
4	0.00	0.00	0.03	0.00	0.03	0.02	0.01	0.59	0.68	0.78	0.77
5	0.01	0.00	0.01	0.00	0.05	0.00	0.04	0.43	0.59	0.77	0.72
6	0.02	0.00	0.05	0.00	0.06	0.04	0.07	0.30	0.51	0.78	0.67
7	0.01	0.01	0.00	0.00	0.04	0.00	0.05	0.23	0.45	0.78	0.60
8	0.01	0.03	0.00	0.00	0.05	0.00	0.05	0.20	0.37	0.78	0.56

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 9: Bayesian Factor Augmented Vector Autoregression (BFAVAR)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
					(a) relative RMSE						
1	0.96	1.02	1.02	0.90	0.90●	0.93	1.17	0.98	1.05	1.07	1.02
2	0.98	1.05	0.95∙	0.94	0.93●	0.94	0.97	0.95	1.03	0.97	1.02
3	0.97	1.05	0.99	0.95	0.94●	0.98	0.93	0.94	1.02	0.93•	1.02
4	0.98	1.02	0.98	0.91	0.93●	0.99	0.92•	0.93	0.99	0.90●	1.01
5	0.99	1.00	0.98•	0.91	0.94●	0.98	0.95•	0.91∙	0.98	0.87●	1.01
6	0.99	1.00	1.00	0.95	0.94●	1.00	0.98•	0.89∙	0.97	0.85●	1.01
7	0.99∙	0.99	1.00	0.97	0.93●	0.99	0.99	0.88●	0.97	0.83●	1.00
8	0.99∙	0.98•	1.00	0.97	0.93●	1.00	1.00	0.86●	0.97	0.83●	1.00
					(b) p-value Mincer-Zarnowitz regressions						
1	0.23	0.00	0.00	0.82	0.00	0.80	0.89	0.00	0.86	0.00	0.03
2	0.22	0.00	0.00	0.96	0.00	0.59	0.45	0.00	0.80	0.00	0.03
3	0.17	0.00	0.00	0.60	0.00	0.16	0.00	0.00	0.72	0.00	0.02
4	0.14	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.64	0.00	0.02
5	0.20	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.56	0.00	0.02
6	0.22	0.00	0.00	0.19	0.00	0.02	0.01	0.00	0.49	0.00	0.01
7	0.28	0.00	0.00	0.29	0.00	0.04	0.04	0.00	0.44	0.00	0.01
8	0.25	0.00	0.00	0.45	0.00	0.27	0.07	0.00	0.40	0.00	0.01
					(c) R^2 Mincer-Zarnowitz regressions						
1	0.07	0.04	0.00	0.16	0.07	0.12	0.12	0.96	0.96	0.95	0.93
2	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.85	0.89	0.86	0.86
3	0.01	0.00	0.01	0.00	0.02	0.02	0.09	0.71	0.79	0.80	0.81
4	0.02	0.00	0.01	0.11	0.04	0.15	0.13	0.56	0.69	0.75	0.75
5	0.01	0.00	0.01	0.12	0.05	0.20	0.09	0.43	0.60	0.73	0.70
6	0.00	0.01	0.00	0.05	0.05	0.07	0.06	0.33	0.52	0.72	0.67
7	0.00	0.01	0.00	0.03	0.03	0.06	0.04	0.26	0.46	0.71	0.64
8	0.00	0.02	0.00	0.02	0.05	0.02	0.03	0.22	0.40	0.69	0.61

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 10: Small Bayesian Vector Autoregression (BVAR)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
					(a) relative RMSE						
1	1.01	0.93	1.03	0.92∙	0.78●	0.93	1.06	0.92	1.04	1.04	0.99
2	1.05∙	0.95	0.98	0.93	0.88	0.94	0.96	0.91•	1.07	0.95	0.98
3	1.06∙	0.96	1.03	0.93	0.96	0.97	0.97	0.92•	1.10	0.92	0.97
4	1.07•	0.99	1.05	0.90	1.00	0.98	0.99	0.92∙	1.13	0.89∙	0.95∙
5	1.08•	0.98	1.05	0.92	1.05	0.96	1.02	0.91∙	1.15	0.86∙	0.93•
6	1.07•	0.98	1.07∙	0.95	1.08	0.98	1.03	0.91	1.17	0.83•	0.91•
7	1.07•	1.01	1.10•	0.97	1.10	0.97	1.04	0.91	1.20∙	0.82•	0.88•
8	1.05∙	1.03	1.09∙	0.97	1.12	0.99	1.05	0.91	1.22∙	0.82•	0.87●
					(b) p-value Mincer-Zarnowitz regressions						
1	0.06	0.01	0.00	0.38	0.01	0.89	0.52	0.00	0.90	0.06	0.29
2	0.01	0.00	0.00	0.43	0.00	0.55	0.09	0.00	0.79	0.04	0.28
3	0.00	0.00	0.00	0.46	0.00	0.51	0.01	0.00	0.66	0.03	0.24
4	0.00	0.00	0.00	0.41	0.00	0.30	0.00	0.00	0.54	0.02	0.18
5	0.00	0.00	0.00	0.33	0.00	0.50	0.01	0.00	0.41	0.02	0.14
6	0.00	0.00	0.00	0.44	0.00	0.71	0.01	0.00	0.32	0.02	0.12
7	0.00	0.00	0.00	0.42	0.00	0.95	0.02	0.00	0.25	0.01	0.09
8	0.01	0.00	0.00	0.52	0.00	0.92	0.02	0.00	0.19	0.00	0.06
					(c) R^2 Mincer-Zarnowitz regressions						
1	0.02	0.12	0.01	0.15	0.10	0.12	0.29	0.96	0.97	0.94	0.93
2	0.01	0.06	0.02	0.07	0.01	0.02	0.06	0.87	0.88	0.85	0.86
3	0.02	0.03	0.01	0.04	0.00	0.00	0.00	0.74	0.76	0.78	0.81
4	0.03	0.00	0.00	0.02	0.04	0.00	0.01	0.60	0.61	0.73	0.75
5	0.03	0.00	0.00	0.00	0.09	0.00	0.01	0.47	0.47	0.70	0.71
6	0.02	0.00	0.00	0.01	0.11	0.00	0.00	0.37	0.34	0.68	0.70
7	0.02	0.01	0.00	0.00	0.12	0.01	0.00	0.30	0.25	0.67	0.67
8	0.00	0.06	0.00	0.01	0.16	0.01	0.00	0.26	0.17	0.64	0.65

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 11: Vector Autoregression (VAR)

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
(a) relative RMSE											
1	1.29●	1.15∙	1.14	1.16	1.11	1.09	1.33●	1.20●	1.23●	1.23∙	1.11∙
2	1.26●	1.15	1.06	1.19	1.05	1.14	1.21●	1.19∙	1.31∙	1.15	1.18
3	1.25	1.15	1.15●	1.19	1.15	1.16	1.22•	1.20	1.37∙	1.14	1.24
4	1.24	1.18	1.13	1.18	1.26	1.15	1.24•	1.21	1.44∙	1.14	1.20
5	1.33	1.21	1.22∙	1.21	1.40∙	1.18	1.27∙	1.23	1.51∙	1.15	1.18
6	1.28	1.21	1.22∙	1.19	1.51∙	1.13	1.27	1.26	1.59∙	1.14	1.21
7	1.24	1.35	1.26∙	1.13	1.57∙	1.15	1.25	1.30	1.68∙	1.12	1.23
8	1.24	1.46	1.32●	1.09	1.64∙	1.10	1.29	1.33	1.75∙	1.11	1.24
(b) p-value Mincer-Zarnowitz regressions											
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.34	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.23	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.77	0.20	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.16	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.14	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.11	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00
(c) R^2 Mincer-Zarnowitz regressions											
1	0.00	0.07	0.17	0.04	0.01	0.06	0.12	0.94	0.95	0.92	0.92
2	0.01	0.05	0.03	0.02	0.01	0.00	0.00	0.80	0.82	0.78	0.84
3	0.02	0.05	0.00	0.01	0.00	0.00	0.03	0.64	0.62	0.65	0.78
4	0.01	0.02	0.01	0.00	0.02	0.00	0.05	0.50	0.39	0.53	0.73
5	0.06	0.01	0.00	0.01	0.08	0.04	0.03	0.36	0.19	0.43	0.71
6	0.03	0.03	0.00	0.00	0.15	0.00	0.00	0.26	0.06	0.37	0.68
7	0.01	0.00	0.00	0.00	0.14	0.03	0.00	0.19	0.01	0.31	0.65
8	0.02	0.04	0.01	0.01	0.21	0.00	0.00	0.15	0.01	0.26	0.62

Notes: The symbols ●, •, ∙, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively. p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.

Table 12: Univariate Autoregression

horizon	Δgdp_t	Δcpi_t	Δc_t	Δinv_t	Δw_t	Δip_t	Δppi_t	i_t^s	u_t	i_t^l	ca_t
					(a) absolute RMSE						
1	3.49	1.52	3.02	13.45	3.12	12.18	2.98	0.38	0.24	0.31	3.77
2	3.52	1.44	2.80	13.64	3.14	12.35	3.85	0.75	0.44	0.55	5.18
3	3.55	1.46	2.72	13.75	3.16	11.99	4.14	1.07	0.63	0.74	6.31
4	3.52	1.51	2.72	14.42	3.22	12.03	4.17	1.37	0.81	0.89	7.52
5	3.50	1.55	2.73	14.33	3.20	12.30	4.04	1.66	0.97	1.02	8.42
6	3.48	1.56	2.69	13.79	3.18	11.95	3.90	1.90	1.11	1.13	9.04
7	3.46	1.58	2.56	13.46	3.12	12.02	3.85	2.10	1.22	1.24	9.81
8	3.47	1.61	2.56	13.39	3.13	11.80	3.82	2.28	1.32	1.35	10.50
					(b) p-value Mincer-Zarnowitz regressions						
1	0.14	0.00	0.00	0.09	0.00	0.04	0.67	0.00	0.87	0.00	0.03
2	0.08	0.00	0.00	0.05	0.00	0.02	0.07	0.00	0.88	0.00	0.03
3	0.04	0.00	0.00	0.06	0.00	0.12	0.00	0.00	0.87	0.00	0.02
4	0.07	0.00	0.00	0.01	0.00	0.05	0.00	0.00	0.86	0.00	0.02
5	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.00	0.02
6	0.07	0.00	0.00	0.04	0.00	0.15	0.00	0.00	0.83	0.00	0.01
7	0.11	0.00	0.00	0.14	0.00	0.00	0.02	0.00	0.83	0.00	0.01
8	0.07	0.00	0.00	0.17	0.00	0.35	0.09	0.00	0.82	0.00	0.01
					(c) R^2 Mincer-Zarnowitz regressions						
1	0.02	0.06	0.12	0.05	0.00	0.05	0.36	0.95	0.97	0.95	0.93
2	0.00	0.02	0.00	0.02	0.01	0.00	0.02	0.84	0.89	0.87	0.87
3	0.02	0.01	0.00	0.00	0.02	0.00	0.06	0.71	0.79	0.81	0.82
4	0.01	0.00	0.00	0.03	0.05	0.02	0.21	0.58	0.68	0.76	0.76
5	0.03	0.00	0.00	0.05	0.06	0.10	0.17	0.46	0.55	0.74	0.71
6	0.02	0.00	0.00	0.01	0.06	0.02	0.08	0.36	0.45	0.72	0.68
7	0.01	0.01	0.01	0.00	0.07	0.09	0.04	0.31	0.37	0.71	0.64
8	0.02	0.02	0.02	0.00	0.08	0.01	0.02	0.26	0.29	0.70	0.61

Notes: p-values larger than 0.05 imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level, while the R^2 can be interpreted as the fraction of the variance in the data that is explained by the forecasts.