Kiel Working Papers

Kiel Institute
for the World Economy

Forecasting German Key Macroeconomic Variables Using Large Dataset Methods

Inske Pirschel and Maik H. Wolters

No. 1925 | May 2014

Web: www.ifw-kiel.de

Kiel Working Paper No. 1925 | May 2014

Forecasting German Key Macroeconomic Variables Using Large Dataset Methods*

Inske Pirschel and Maik H. Wolters

Abstract:
We study the forecasting performance of three alternative large scale approaches using a dataset for Germany 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 vector autoregression 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 and the Bayesian factor augmented VAR provide the most precise forecasts for a set of eleven core macroeconomic variables, including GDP growth and CPI inflation, and that the performance of these two models is relatively robust to model misspecification. However, our results also indicate that in many cases the gains in forecasting accuracy relative to a simple univariate autoregression are only moderate and none of the models would have been able to predict the Great Recession.

Keywords: Large Bayesian VAR, Model averaging, Factor models, Great Recession

JEL classification: C53, C55, E31, E32, E37, E47

Inske Pirschel
Kiel Institute for the World Economy
Kiellinie 66
24105 Kiel, Germany
inske.pirschel@ifw-kiel.de

Maik H. Wolters
Kiel Institute for the World Economy
Kiellinie 66
24105 Kiel, Germany
maik.wolters@ifw-kiel.de

* We thank Christian Schumacher for sharing the dataset used in Schumacher (2007) and for useful comments and discussions. We further thank Jens Boysen-Hogrefe, Nils Jannsen, Martin Plödt, Tim Schwarzmüller 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.

The responsibility for the contents of the working papers rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular working paper about results or caveats before referring to, or quoting, a paper. Any comments on working papers should be sent directly to the author. Coverphoto: uni_com on photocase.com

1 Introduction

In forecasting with standard time series methods generally the following trade-off arises: given the vast amount of macroeconomic time series all of which potentially contain important information on future macroeconomic dynamics, forecasters wish to use as much information as possible to obtain precise parameter estimates and 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 be imprecise and in-sample overfitting might occur. This can lead to poor out-of-sample forecasts. In some cases 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.

In particular, 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 thereby 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, i.e. 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, such as equal weighted averaging and Bayesian model averaging, for example.

Previous literature has mainly focused on evaluating the forecasting performance of one or two of these large scale approaches relative to several small-scale benchmark models or to each other. For example, Bernanke and Boivin (2003) compare forecasts obtained by a factor augmented autoregression (FAAR) and a factor augmented vector autoregression (FAVAR) to those of several 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 the 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 the model averaging techniques perform better than the factor based forecasting approaches. Bańbura et al. (2010) study the forecasting performance of a large Bayesian vector autoregression (LBVAR) and a Bayesian factor augmented vector autoregression (BFAVAR). According to their results the LBVAR generally outperforms the BFAVAR. Finally, Berg and Henzel (2013) evaluate the relative forecasting performance of the BFAVAR, the LBVAR and combinations of several small VARs for the Euro area in terms of point and density forecasts. Their results indicate that while the LBVAR provides the most accurate point forecasts, the BFAVAR yield the most precise density forecasts.[1]

The variety of forecasting performance based rankings of the different large scale approaches shows that there is no consensus yet about which is the most useful method to extract the predictive content from a large dataset. With this paper we seek to fill this gap by systematically comparing the forecasting accuracy of all 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 obtained with the different models are unbiased. Further, we report the share of the variance of each forecast time series that can be explained by the different forecasting models.

Second, we check to what degree the relative performance of the different forecasting models is robust against model misspecification. To do so, we analyse the accuracy of the forecasts of the different models obtained with the ex post best performing specification of each model.

Third, we investigate 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 forecasts for a small set of key variables which usually include output growth and the inflation rate and in some cases a short-term 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 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,

[1]Going beyond purely reduced form forecasting models, Wolters (2014) compares the forecasting accuracy of Dynamic Stochastic General Equilibrium (DSGE) models to the LBVAR and the Fed's Greenbook projections. He finds that weighted forecasts of several DSGE models and the LBVAR are more precise than those obtained by individual DSGE models and a small Bayesian vector autoregression (BVAR) and come close to the accuracy of the Greenbook projections at least for medium term horizons.

construction, surveys and miscellaneous. Our dataset is a modified and updated version of the dataset used in Schumacher (2007) where the forecasting performance of alternative factor models is studied.

We estimate the three large scale forecasting models as well as a number of small benchmark models using a moving window of 15 years of data. Thus, our evaluation sample ranges from 1993 through 2013. We compute forecasts up to eight quarters ahead for a small set of German key macroeconomic 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. The forecasting models are specified according to various information criteria. As a robustness check we also specify the models based on their ex post forecasting performance. 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 and the BFAVAR show the best overall forecasting performance. Both deliver forecasts that are more precise than those obtained by a simple univariate autoregressive (AR) benchmark for most of the variables of interest. By contrast, for the other factor based models as well as the model averaging techniques the forecasting performance relative to the AR benchmark model depends heavily on the variable to be forecast and the forecasting horizon.

Regarding the models' robustness to misspeficiation, we find that the dynamic factor model (DFM) clearly outperforms all other forecasting models if the forecasts obtained with the ex post best performing specification of each model are compared. However, in the quasi real-time exercise—where the number of lags and factors is chosen based on information criteria or past forecasting accuracy—this performance is unattainable. By contrast, for the LBVAR and the BFAVAR the differences in forecasting accuracy between the ex post optimal specification and the quasi real-time specification are only small.

In general, the gains in forecasting accuracy obtained by the large scale approaches relative to the simple AR benchmark rarely exceed 10% and are in most cases statistically insignificant. 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 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. In section 2 we outline the different forecasting models. In section 3 we describe the dataset that we use, while in section 4 we describe our forecasting approach. In section 5 we evaluate the absolute and relative forecasting performance of the different models for four key variables and check for robustness against model misspecification. In section 6 we extend the analysis to a larger set of of eleven macroeconomic variables and finally, in section 7 we conclude.

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 $t = 1, ..., T$.[2] The total number of variables in our dataset is given by $n + m = k$. We compute annualized quarter-on-quarter growth rates of all variables, denoted by $\{\Delta y_{i,t}\}_{i=1}^{n}$ and $\{\Delta x_{j,t}\}_{j=1}^{m}$, respectively.[3] Given the information available at time T, we estimate all forecasting models and construct forecasts $\{\Delta y_{i,T+h}\}_{i=1}^{n}$ with h being the forecast horizon ranging from one to eight quarters ahead.[4]

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 VAR outlined above. 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 a model 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, as 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.[5] First, the coefficients $A_1, ..., A_p$ are assumed to be a priori independently and normally distributed. With respect to the constant in the VAR the prior is assumed to be diffuse. The moments for the prior distribution of the VAR coefficients are given by:

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

and

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

[2] 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.

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

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

[5] 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.

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 is a scale parameter 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 of the parameters of the VAR which increases with the size of the cross-section to avoid over-fitting. Bańbura et al. (2010) suggest to set the tightness of the prior 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 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.[6] 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 $Z_{i,t} = d + \sum_{\iota=1}^{4} \beta_\iota Z_{i,t-\iota} + u_t$. 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).

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 potential 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 $\Delta X^*_{j,t}$. 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 $\Delta X^*_{j,t}$. With these factors the dimension of a large dataset can thus be reduced without loosing valuable information.

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

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 direct autoregression (FAAR) $\Delta y_{i,t} = \rho_0 + \rho_1 \Delta y_{i,t-h} + ... + \rho_p \Delta y_{i,t-h+1-p} + \gamma_1 F_{1,t-h} + ... + \gamma_r F_{r,t-h} + \epsilon_t$ for each $h = 1, ..., 8$ and compute 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 to allow for a more dynamic structure: $Z_t = c + B_1 Z_{t-1} + ... + B_p Z_{t-p} + \epsilon_t$. 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] The variables to be predicted are included in log-levels to use information that is possibly contained in the trends. The FAVAR forecasts are computed by iterating the recursive system of equations forward.

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 $\Delta X^*_{j,t}$. Second, we set the hyperparameter $\lambda = 0.1$, a standard value in the literature (see fo example Litterman, 1986). As in the FAVAR, we include the 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).

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 a 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 direct autoregression from which forecasts $\Delta y_{i,T+h}$ are computed. The number of lags of the dependent variable included in the estimation is determined via the Bayesian information criterion. For the determination

[7] We also estimate a FAVAR 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.

of the optimal number of dynamic factors q we apply the information criterion proposed by Bai and Ng (2007).

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 direct autoregressive distributed lag models $\Delta y_{i,t} = \rho_0 + \rho_1 \Delta y_{i,t-h} + ... + \rho_p \Delta y_{i,t-h+1-p} + \beta_j \Delta x_{j,t-h} + \epsilon_{j,t}$ 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 the model-specific forecast $\Delta y^j_{i,T+h}$.

According to the specification of the weight ω_j that is attributed to each model-specific forecast we distinguish two model averaging approaches. 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, the posterior probability of model j, $P(M_j)$, is obtained by dividing the model-specific likelihood $L_j = (1+\phi)^{-K/2} S_j^{-T}$ with $S_j = \sqrt{(\hat{\epsilon}_{j,t})'(\hat{\epsilon}_{j,t}) - (\hat{\epsilon}_{j,t})' X_{j,t-h}(X'_{j,t-h}X_{j,t-h})^{-1} X'_{j,t-h}(\hat{\epsilon}_{j,t})\frac{\phi}{1+\phi}}$ and $\hat{\epsilon}_{j,t} = \Delta y_{i,t} - B^*_j X_{j,t-h}$ by the sum of all m model likelihoods for a specific horizon h. Here K denotes the number of parameters in the regressor matrix $X_{j,t-h} = (1, \Delta y_{i,t-h}, ..., \Delta y_{i,t-h+1-p}, \Delta x_{j,t-h})$ 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 $\Delta y_{i,t} = \rho^*_0 + \rho^*_1 \Delta y_{i,t-h} + ... + \rho^*_p \Delta y_{i,t-h+1-p} + \epsilon_{j,t}$, 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.

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 $\Delta y_{i,t} = c + \sum_{j=1}^{p} \rho_j \Delta y_{i,t-h+1-j} + \epsilon_t$ for each of the $i = 1, ..., n$ variables of interest, separately for each forecasting horizon h. The forecasts of

[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 omit the respective subscripts.

each variable for forecasting horizon h are obtained from the respective model: $\Delta y_{i,T+h} = c + \sum_{j=1}^{p} \rho_j \Delta y_{i,T-j+1}$. 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 and compute the vector of forecasts Y_{T+h} 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 the remaining data is directly obtained from 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) we rescale data which is only available for

West Germany prior to 1991 to the pan-German series to avoid regime shifts. A detailed data description is provided in the appendix.

4 Forecasting Approach

We estimate the forecasting models on a moving window containing 60 observations to account for possible structural breaks in the estimation sample. 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 the rolling window scheme.

To assess the relative and absolute forecasting performance of the models, we first 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 6 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 evaluate 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 for each variable i 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, $y^r_{i,t+h}$, on a constant and the forecasts, $y^f_{i,t+h}$:

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

If the intercept estimate $\hat{\alpha}$ is significantly different from zero, the forecasts are systematically larger or smaller than the data realizations. 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

[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.

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 is always smaller than 1 since there are shocks and idiosyncrasies that no economic model can capture.

To assess the relative forecasting accuracy of the different models we compute RMSEs for each variable i and each forecasting horizon h over the evaluation sample T^*: $RMSE_{i,h} = \sqrt{\frac{1}{T^*}\sum_{t=1}^{T^*}(y^r_{i,t+h} - y^f_{i,t+h})^2}$. We report the absolute level of the RMSEs for a simple AR(2) forecast which we use as a benchmark, while for the remaining models the RMSEs are reported relative to this benchmark. This implies that a relative RMSE smaller than 1 indicates that the forecasting performance of a specific model is better than that of the AR(2) benchmark and vice-versa. 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. By contrast, if a time-series 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 for Four Core Variables

In this section we report the results of our forecasting exercise for the following four key macroeconomic variables: output growth, CPI inflation, the short-term interest rate and the unemployment rate. We report results for all forecasting models of section 2, except for those which are strictly dominated by a related alternative from the same model category. For example,

the BFAVAR always performs much better than the FAVAR, so we do not report results for the latter. Since the univariate autoregression always performs slightly better when estimated iteratively, we choose this variant instead of the directly estimated variant.

For each horizon a total of 76 forecasts is computed. As a benchmark, we use an iterative univariate autoregression with a fixed number of lags $p = 2$.

5.1 Output Growth

Figure 1 shows the German GDP growth series as well as the forecasts of the AR 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. It can be clearly seen that 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 estimation of the AR is equal to 1 most of the time.[11] Forecasting models that predict a quick return to the average GDP growth rate should therefore be able to predict German GDP growth more precisely than other models.

Panel (a) in table 1 shows the absolute RMSEs of the AR(2) benchmark and relative RMSEs of the other forecasting models for GDP growth. The absolute RMSEs of the AR(2) 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 low persistence. 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 LBVAR, the BFAVAR and the small BVAR for horizons up to $h = 3$ as well as for the AR and the EWA for two horizons, respectively. For the remaining models, the estimated constant is larger than zero, but the slope parameter is smaller than one (and in some cases even negative) indicating that the forecasts systematically predict less variation than the data actually shows.

The entries in table 1 reveal that the gains in forecasting accuracy for GDP growth obtained by the three large scale approaches are at best moderate and mostly insignificant. Only the BFAVAR is able to significantly improve upon the AR(2) benchmark, at least for the one-step ahead forecast. Among the three large scale approaches, the LBVAR and the BFAVAR yield the most accurate forecasts. However, the small BVAR performs equally well. For horizons up to $h = 2$ or $h = 3$, respectively, the gains of these three models over the AR(2) benchmark amount to more than 5%. The three models also slightly outperform the AR with a variable number of lags, which neither the two model averaging techniques, EWA and BMA, nor the two remaining factor models, FAAR and DFM, and the unrestricted VAR achieve.

These results are also reflected in figure 1: The AR and the LBVAR forecasts predict a return to the sample average after only one quarter, while the forecasts obtained by the BFAVAR need

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

[11] Comparing the autocorrelation function of US and German GDP growth for a sample from 1978-2013 shows that there is significant autocorrelation of up to two lags for US GDP growth, while there is no significant autocorrelation of German GDP growth at all.

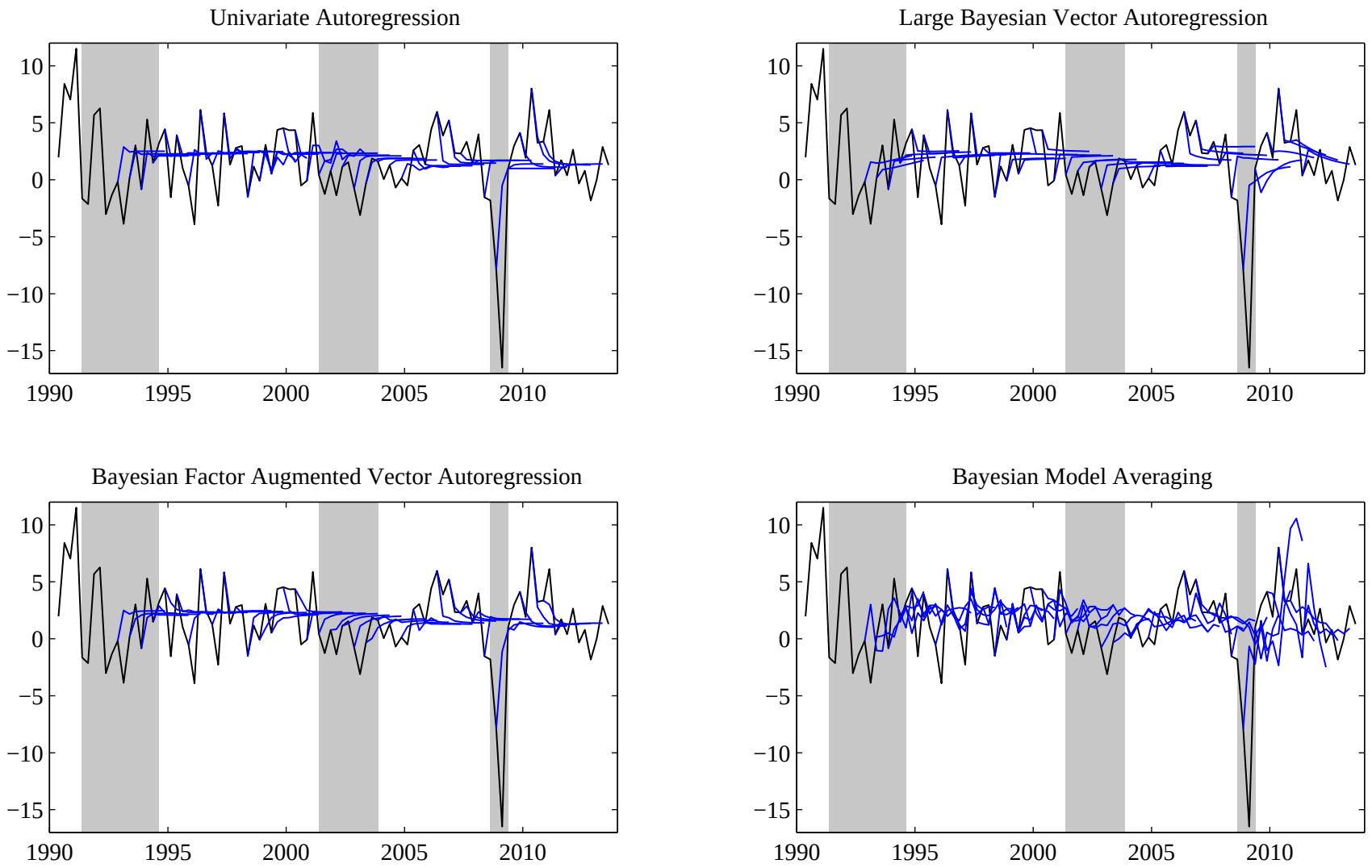

Figure 1: GDP growth forecasts

Table 1: RMSEs of the different forecasting models

(a) Output growth

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	3.58	0.99	**0.93**	0.99	**0.92**•	0.99	0.98	1.01	1.04	**0.92**
2	3.70	0.98	**0.93**	1.11	**0.93**	1.12	0.96	0.95	1.06	**0.91**
3	3.69	0.98	**0.96**	1.09	**0.94**	1.09	**0.96**	0.97	1.08	**0.92**
4	3.62	1.00	0.99	1.10	**0.96**	1.01	1.00	1.03	1.09	**0.95**
8	3.51	**0.99**	1.01	1.18●	**0.97**	1.09	**0.96**	1.04	1.01	**1.00**

(b) CPI Inflation Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	1.49	1.03	**0.89**•	0.97	**0.90**•	0.98	0.99	0.97	**0.96**	**0.89**•
2	1.37	1.01	0.92•	1.05	**0.93**	1.04	0.96●	0.93•	1.01	0.91
3	1.35	1.02	0.95	1.05	0.98	1.03	0.98	0.94	1.07	0.94
4	1.37	1.01	0.98	1.10	0.99	1.09	0.98	0.98	1.11	0.94
8	1.45	0.99	0.94	1.14	0.98	1.12	0.95•	0.96	1.14	0.93

(c) Interest Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.36	1.03	**0.95**	**1.01**	**0.98**	**1.01**	1.00	1.03	1.08	1.01
2	0.70	1.03	**0.89**•	**0.96**	**0.95**	**0.97**	0.97	0.95	1.04	0.96
3	0.99	1.02	**0.88**•	**0.94**	0.94•	**0.96**	0.99	1.01	1.05	0.94•
4	1.25	1.04	**0.89**	**0.99**	0.94	**1.02**	1.03	1.07	1.06	0.92•
8	1.99	1.08	0.94	1.18	0.90	1.13	1.17	1.06	1.06	0.84•

(d) Unemployment Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.24**	**1.00**	**0.99**	**0.99**	**1.00**	**1.00**	**0.97**•	**0.97**•	**1.25**•	**1.04**
2	**0.46**	**1.00**	**0.93**	**0.99**	**0.96**	**1.01**	**0.96**•	**0.96**	**1.28**•	**1.00**
3	**0.68**	**1.00**	**0.91**	**1.02**	**0.94**	**1.05**	**0.99**	**0.98**	**1.32**	**0.99**
4	**0.88**	**1.00**	**0.90**	**1.05**	**0.92**	**1.09**	**1.03**	**1.02**	**1.36**	**0.97**
8	**1.46**	**1.01**	0.98	**1.10**	**0.91**	**1.17**	**1.31**	1.35	1.63	**0.98**

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. Column 2 shows the absolute RMSEs for the AR(2) benchmark model, while all other RMSEs are computed relative to the AR(2) RMSEs. 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.

slight additional adjustment over the following quarters. In contrast to that the BMA yields less systematic and much more volatile forecasts.

Forecasting GDP Growth with the Ifo Business Climate Index. In addition to the large scale approaches and the small benchmark models outlined in section 2, 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).[12] 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): $\Delta y_t = \alpha_h + \beta_h \text{ifo}_{t-h} + \epsilon_{t,h}$.

[12]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.

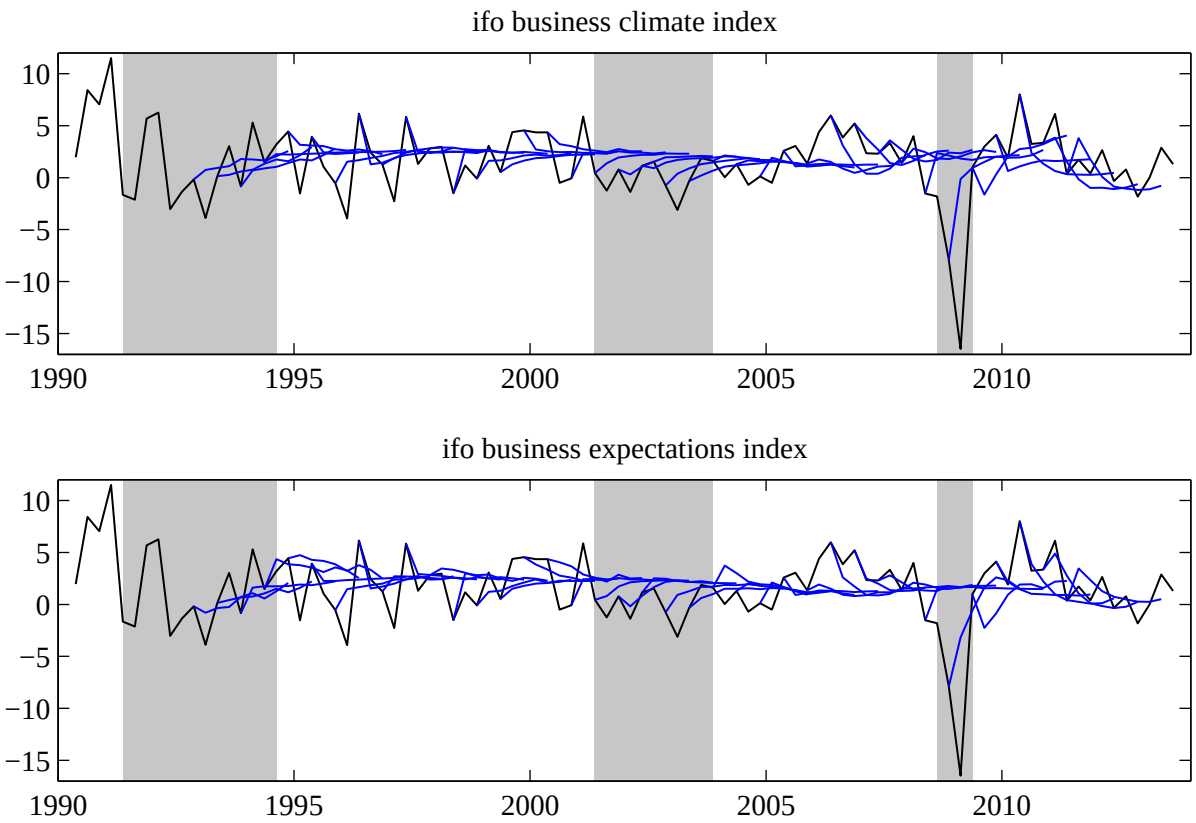

Figure 2: GDP growth forecasts with ifo leading indicators.

The forecasts for the different horizons which are computed as $\Delta y_{T+h} = \hat{\alpha}_h + \hat{\beta}_h \text{ifo}_T$ are shown in figure 2. It can be seen that both ifo indicators lead to more dynamics than the AR forecasts plotted in figure 1. Table 2 reports the RMSEs of the AR and the two ifo indicators relative to the AR(2) benchmark. The entries reveal that, at least for the one-quarter ahead forecast, the ifo expectations index is indeed able to significantly reduce the relative RMSE at the 10% level.[13] However, for higher forecasting horizons the improvements are only minor and insignificant.

Table 2: RMSEs for AR and ifo leading indicators relative to AR(2) benchmark

Output growth

horizon	AR	ifo climate	ifo expecations
1	0.99	**0.94**	**0.89•**
2	0.98	**0.95**	**0.95**
3	0.98	0.97	0.98
4	1.00	**0.97**	0.98
8	**0.99**	1.02	**0.98**

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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.

[13] This 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 considerably upon the AR forecast.

Given the extremely low persistence of German GDP growth it does not seem surprising that the forecasts of all models fail to convey much 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 reveal that, with very few exceptions, not even 10 percent of the variance in the time series is explained by the forecasts of any of the different models. This confirms that German GDP growth is extremely difficult to predict. Adding more information by using a large dataset for the forecasting process apparently only leads to marginal improvements in forecasting accuracy over the AR(2) benchmark. By contrast, 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 0.18 (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	DFM	EWA	BMA	VAR	BVAR
1	0.02	0.07	0.15	0.09	0.16	0.03	0.03	0.01	0.10
2	0.01	0.00	0.03	0.00	0.01	0.00	0.01	0.00	0.03
3	0.02	0.01	0.00	0.02	0.00	0.01	0.00	0.01	0.01
4	0.03	0.06	0.01	0.02	0.00	0.01	0.03	0.01	0.00
8	0.01	0.10	0.02	0.00	0.01	0.03	0.01	0.00	0.01

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.04	0.15	0.12	0.12	0.12	0.06	0.08	0.09	0.18
2	0.03	0.05	0.06	0.03	0.06	0.04	0.07	0.03	0.07
3	0.01	0.02	0.04	0.00	0.04	0.03	0.06	0.01	0.04
4	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02
8	0.01	0.04	0.00	0.00	0.00	0.00	0.01	0.06	0.00

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
2	0.82	0.84	0.82	0.83	0.81	0.84	0.85	0.84	0.84
3	0.66	0.68	0.64	0.67	0.62	0.67	0.65	0.69	0.69
4	0.50	0.49	0.41	0.50	0.38	0.48	0.40	0.55	0.54
8	0.16	0.01	0.00	0.12	0.00	0.09	0.10	0.19	0.22

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.96
2	0.88	0.90	0.89	0.89	0.89	0.89	0.89	0.82	0.88
3	0.76	0.82	0.77	0.80	0.76	0.77	0.78	0.62	0.77
4	0.62	0.74	0.62	0.70	0.61	0.61	0.62	0.41	0.65
8	0.16	0.45	0.14	0.38	0.09	0.00	0.00	0.02	0.27

Forecasting the Great Recession. Several studies, for example Kuzin et al. (2013) and Timmermann and van Dijk (2013), indicate that the performance based ranking of different forecasting models may change considerably during the period of the Great Recession of 2008/2009. Therefore, in what follows, we take a closer look at whether the three large scale forecasting methods would have been able to forecast the slump of German GDP growth during the Great Recession. Figure 3 shows the forecasts of the annualized quarter-on-quarter GDP growth rate

obtained by the AR, the LBVAR, the BFAVAR, the BMA and the two ifo indicators considered above computed for the subsample ranging from 2008Q1 to 2009Q2. Generally, the forecasts of all six models look roughly similar and none of them is able to predict the downturn in GDP growth in 2008. Once the recession hits, the models also fail to predict a further deepening of the recession. The only notable exception is the model based on the ifo expectation index: while the

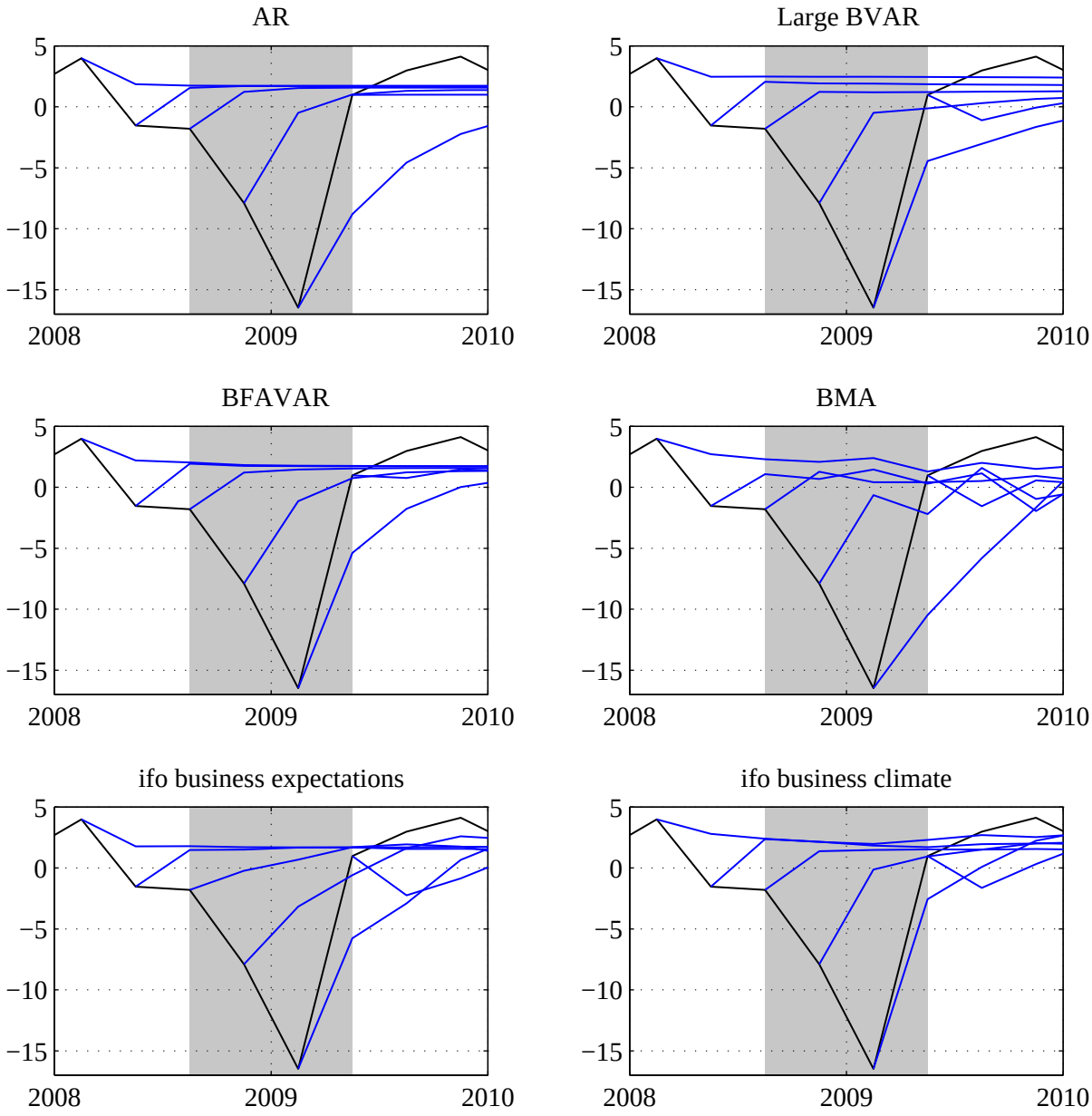

Figure 3: Great Recession GDP growth forecasts

largest decreases in GDP growth occurred only in the fourth quarter of 2008 and the first quarter of 2009, business expectations in Germany already dropped largely during the third quarter of 2008. So, we find that the ifo expectation index indeed predicts a negative GDP growth rate of -3.17% 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 fur-

ther 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[14]. 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. Additionally, once the turning point is reached, the models underpredict the speed of the recovery.

Table 4: Great Recession RMSEs of forecasting models relative to AR(2) benchmark

horizon	AR	LBVAR	BFAVAR	BMA	ifo climate	ifo expecations
1	0.97	0.92	0.90	1.03	0.93	0.82
2	0.92	0.92	0.89	0.91	0.90	0.89
3	0.92	0.94	0.90	0.91	0.90	0.89
4	0.94	0.97	0.93	0.99	0.91	0.90

Notes: The relative RMSEs are calculated for the subsample of the Great Recession. The first forecast is computed in 2008Q1, while the last forecasts is computed in 2009Q2. Significance tests are omitted due to the shortness of the subsample.

The entries in table 4 display the RMSEs of the different forecasting models relative to the AR(2) benchmark for the Great Recession subsample. The ifo expectation based indicator clearly outperforms the other models for the one-step ahead forecast. For higher horizons the models perform more or less equally well.

5.2 Inflation

Figure 4 shows the German CPI inflation rate series as well as the forecasts of the AR 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 will presumably be hard to predict. The series also displays 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 rather than to the new trend. As for GDP growth, the forecasts of the BMA model are considerably more volatile in comparison to the remaining models.

Panel (b) in table 1 shows the RMSEs for CPI inflation. The absolute RMSEs shown for the AR(2) model are less than half of those for GDP growth. Still the persistence of quarterly CPI inflation is quite low and thus the RMSEs do not increase with the forecast horizon h. The LBVAR, the BFAVAR and the small BVAR significantly outperform the AR(2) benchmark for short horizons up to $h = 1$ or $h = 2$, respectively, with reductions in the relative RMSEs of up to 11%. For higher horizons they also provide more precise forecasts, but not on a significant level. These three models are also the only models that yield unbiased forecasts for short horizons up to $h = 2$. For the remaining models, the estimated constant in the Mincer-Zarnowitz regressions

[14]We performed a simple threshold model exercise with the ifo expectation based indicator for the one-step ahead forecast, where we defined the model to be estimated as $\Delta y_{T+h} = \alpha_1 + \beta_1 \text{ifo}_T I(ifo_T \leq \overline{ifo}_T) + \beta_2 \text{ifo}_T I(ifo_T > \overline{ifo}_T) + \epsilon_T$ with $\overline{ifo}_T$ as the average of the ifo expectations based indicator over the sample $t = 1..., T$. However, this did not result in a large gain over the AR(2) benchmark, the relative RMSFE for the one-step ahead forecast is equal to 0.94.

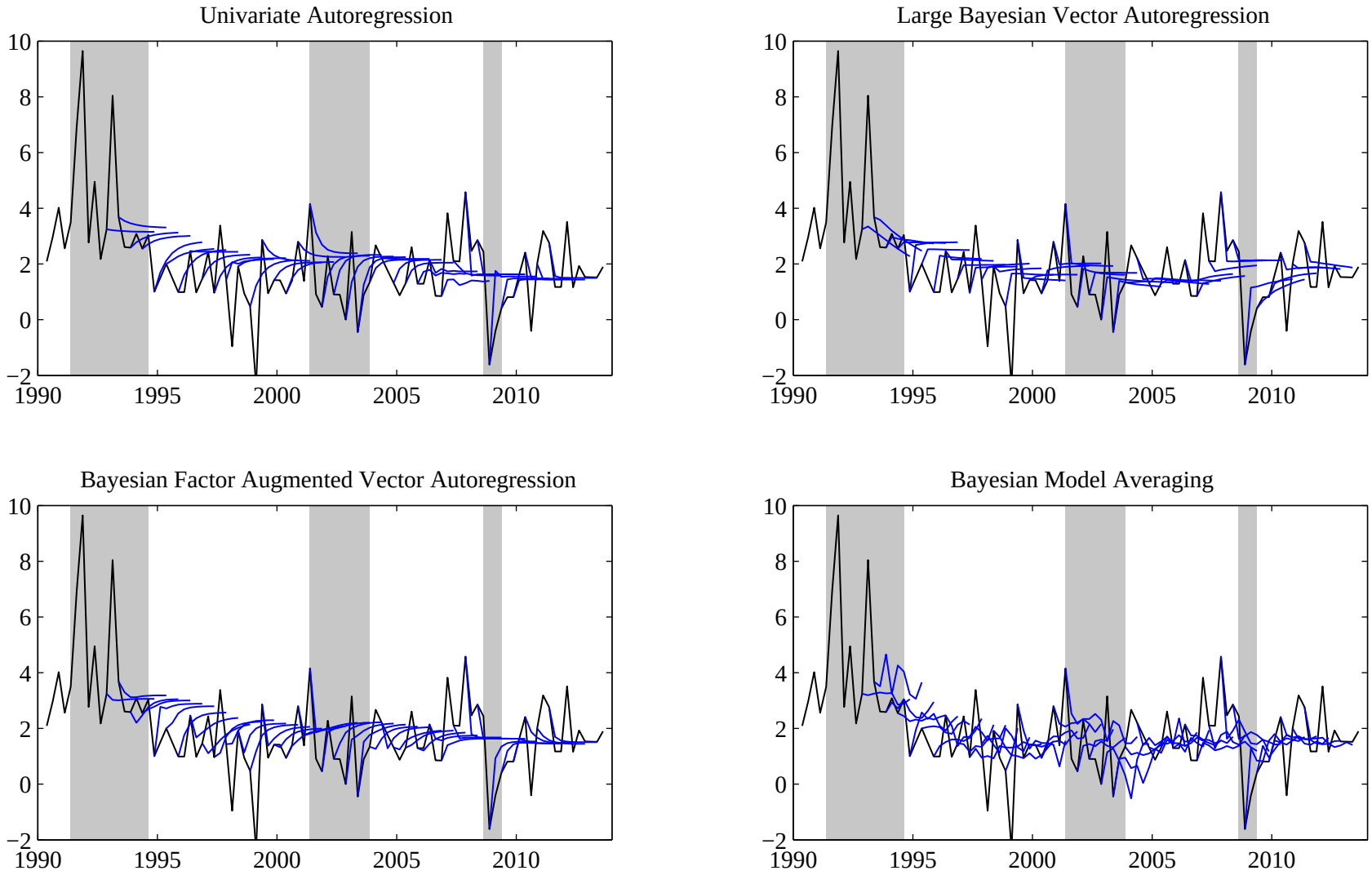

Figure 4: Inflation forecasts

is larger than zero, but the slope parameter is smaller than one. The two model averaging techniques, BMA and EWA, also show smaller RMSEs than the AR(2) benchmark over all forecasting horizons. However the reductions are significant only for some cases. By contrast, the two remaining factor models, FAAR and DFM, as well as the unresticted VAR and the AR with variable number of lags cannot beat the AR(2) benchmark.

While the results for the LBVAR, the BFAVAR and the small BVAR 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 as well as the different factor models, 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 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 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, the model captures the turning points only after they actually occurred.

Panel (c) in table 1 shows the RMSEs for the short term interest rate. The absolute RMSEs for the AR(2) 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(2) is even higher than that for CPI inflation which indicates that the model has problems to capture the downward trend of the interest rate.

Again, the LBVAR, the BFAVAR and the small BVAR yield significantly better forecasts than the AR(2) benchmark for one or several forecast horizons. The forecasts of the other factor models, FAAR and DFM, are also slightly better than the AR(2) forecasts for several forecasting horizons, but not on a statistically significant level. Yet, together with the LBVAR, they are the only models that produce unbiased forecasts for horizons up to $h = 4$. The estimated constants in the Mincer-Zarnowitz regressions increase with the forecast horizon while the estimated slopes decrease with the forecast horizon. Thus, the models are able to predict some of the actual variation in the data for short horizons, while the longer term forecasts equal the average in the estimation sample.

Panel (c) in table 3 shows that the interest rate forecasts of all models are very informative. For $h = 1$ the forecasts of each model can explain 95% of the variance in the interest rate series, while for forecasts one year ahead still almost half of the variance can be explained by most model forecasts. This is certainly due to the high persistence and the extremely small variation in the interest rate series. For long-term forecasts, however, the R^2 drops considerably reflecting that the models have difficulties to capture the overall downward trend in interest rates.

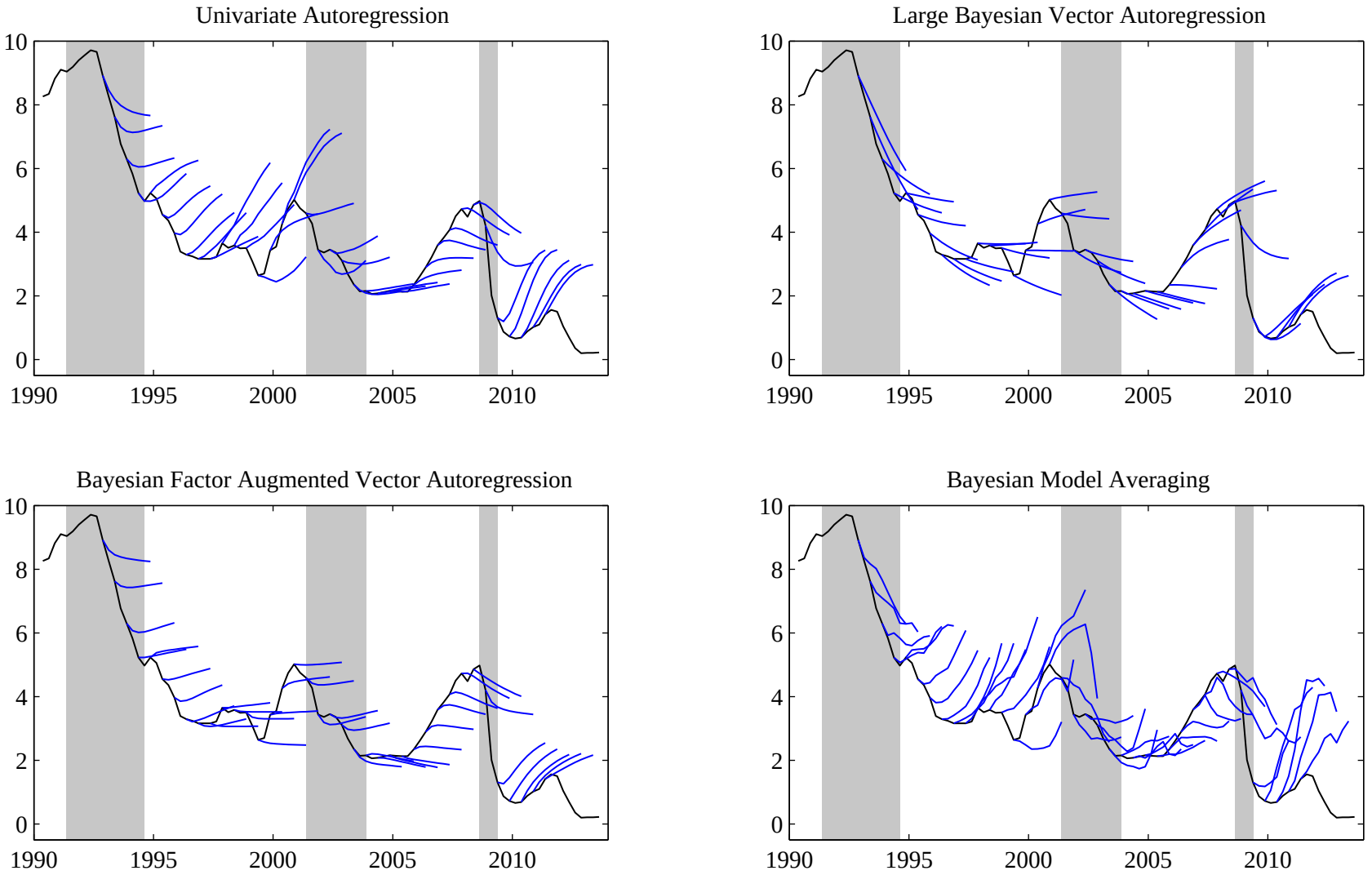

Figure 5: Interest rate forecasts

5.4 Unemployment Rate

Figure 6 shows the German unemployment rate series as well as the forecasts of the AR and those obtained by the LBVAR, the BFAVAR and the BMA. The persistence of the unemployment rate series is very high, similar to that of the short-term interest rate. The unemployment rate, however, does not show one single trend, but increases until 1998, decreases temporarily until 2001, increases again temporarily 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 indicates that the model is not able to capture the most important dynamics in the unemployment rate.

These observations are reflected in the relative RMSEs in panel (d) of table 1. The absolute RMSEs for the AR(2) 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 only exception are the EWA and the BMA which reduce relative RMSEs significantly, at least for the one step ahead forecast. One reason for this might be that the BMA predicts turning points even better than the AR as indicated in figure 6. The unrestricted VAR produces significantly less accurate forecasts than the AR(2) benchmark. With a few exceptions for $h = 8$, 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 forecasts 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. The forecasts of the LBVAR and the BFAVAR for $h = 8$ can even explain about 40% of the data variation, while the other forecasting models have much smaller information content.

5.5 Robustness with Respect to Misspecification

In order to check to what degree possible model misspecification affects the relative forecasting performance of the different models, we repeat the forecasting exercise of the previous section with an *optimized* specification based on ex post information of each forecasting model. So instead of specifying the different models based on various information criteria, in this section, we choose the ex post best performing specification for each model for the evaluation of the relative forecasting performance of the different models.

We report the results of this exercise relative to the AR(2) benchmark, which allows for a direct comparison of each model's performance with its *optimized* specification relative to that based on information criteria as reported in table 1. Thereby we can additionally check which of the different forecasting models yield similarly accurate forecasts in the quasi real-time exercise and with an ex post *optimized* specification and are thus robust against model misspecification.

We obtain the *optimized* specification of each model by computing a variety of different specifications for each model and choosing the one that yields ex post the best forecasting

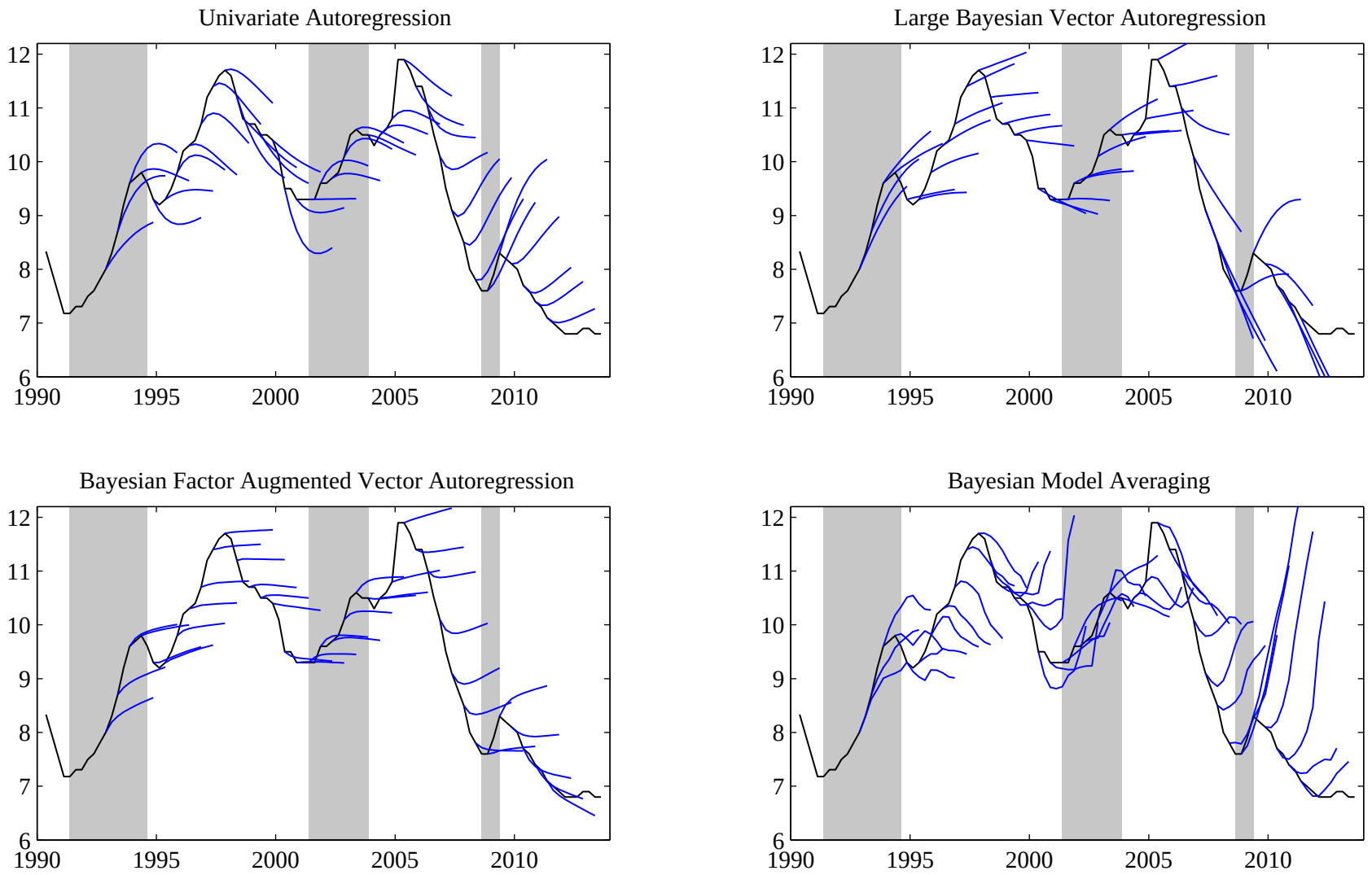

Figure 6: Unemployment rate forecasts

performance. For example, for the FAAR the number of static factors r as well as the number of lags p have to be specified. After defining a range for each of these parameters, i.e. $r = 1, ..., r_{max}$ and $p = 1, ..., p_{max}$, we estimate the FAAR and compute forecats for each possible combination of these parameters. We then choose the specification that yields ex post the highest forecasting accuracy as the *optimized* specification for the FAAR model. Table 5 shows the ranges of the various parameters of the different forecasting models that we consider.

Table 5: Parameter range for forecasting models

parameter	range	forecasting model
number of lags p	1, 2, ..., 4	all forecasting models
degree of shrinkage λ	0.01, 0.02, ..., 0.1	LBVAR
degree of shrinkage ϕ	1, 1.1, ..., 2	BMA
number of static factors r	1, 2, ..., 10	FAAR, FAVAR, BFAVAR
number of dynamic factors q	1, 2, ..., 10	DFM
number of lags of the dynamic factors s	1, 2, ..., 4	DFM

We follow Schumacher (2007) and distinguish the following two approaches: *performance based model selection, time-varying model* (PBTV) and *performance based model selection, constant model* (PBC). With PBTV we divide the evaluation sample into subsamples covering 4 quarters each. For each of these subsamples we select the specification for each forecasting model and for each forecasting horizon that minimizes the respective subsample MSFE. By contrast, with PBC we choose the specification for each model that minimizes the MSFE over the whole evaluation sample for each horizon.

In table 6 we report the results of this exercise for horizons $h = 1$, 4 and 8. As the entries in the upper part of the table indicate, with PBC the relative performance of the different forecasting models is similar to what we found before. The LBVAR and the BFAVAR still provide the most precise forecasts in most cases, however, the gains in accuracy over the remaining large scale approaches are slightly smaller. This picture changes considerably with PBTV, as the entries in the lower part of table 6 reveal. With PBTV the DFM clearly outperforms all remaining models by far. For output growth and the interest rate, the reduction of the relative RMSE obtained by the DFM amounts to more than 50%, at least for the one step ahead forecast. By contrast, the LBVAR which now performs slightly better than the BFAVAR achieves at best a gain of 30%. Moreover, the FAAR is able to yield forecasts that are similarly accurate as those of the LBVAR, while the two model averaging techniques, EWA and BMA, perform slightly less accurate than the BFAVAR.

Comparing the results in table 6 to those reported in table 1 gives rise to the following observations. First, both performance based model selection approaches generally increase the precision of all forecasting models—which of course is not surprising given that these approaches rely on out of sample information. However, while overall PBC leads only to modest gains over the quasi real-time forecasts, the gains obtained with PBTV are very large and, in some cases, even amount to almost 60% (for the DFM, for example). Since the PBC approach chooses the optimal specification separately for each horizon, yet is restricted to a constant model specification, this result clearly indicates that the *optimal* specification of each forecasting

Table 6: RMSEs of forecasting models relative to AR(2) benchmark, PBC and PBTV

Performance based model selection, constant model

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.99**	**0.90•**	0.92	**0.90•**	0.87	0.98•	0.99	1.04	**0.92**
4	**0.97**	0.97	1.02	**0.96**	**0.97**	0.99	1.00	1.09	**0.95**
8	**0.97**	**0.98**	1.00	**0.97**	1.00	**0.96**	**0.96**	1.01	**1.00**

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.97	**0.87●**	0.96	**0.88●**	0.96	0.94•	0.94•	**0.96**	**0.88•**
4	0.98	0.94	1.03	0.93•	1.01	0.97	0.95	1.11	0.94
8	0.98	0.93	1.01	0.93•	0.97	0.95	0.93•	1.14	0.93

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	1.00	**0.82•**	**0.90**	**0.92•**	**0.85**	0.97	0.99	1.02	1.01
4	0.99	**0.85•**	**0.92**	0.89●	**0.85•**	0.99	1.00	1.06	0.92•
8	0.92	0.90	0.99	0.88•	1.01	1.12	1.01	1.06	0.84•

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**1.00**	**0.94**	**0.95**	**0.95**	**0.92**	**0.97●**	**0.96•**	**1.07**	**1.04**
4	**1.00**	**0.91**	**0.94**	**0.86•**	**0.96**	**1.03**	**0.95**	**1.36**	**0.97**
8	**0.94**	**0.94**	**0.91**	**0.87**	**0.95**	**1.25**	**1.27**	1.63	**0.98**

Performance based model selection, time-varying model

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.94●**	**0.83●**	**0.76•**	0.85•	**0.46•**	**0.94●**	**0.94●**	0.97	**0.91**
4	**0.93•**	**0.91•**	**0.89•**	**0.95**	0.76	**0.98**	0.98	1.08	**0.94•**
8	**0.96•**	**0.92•**	**0.92•**	**0.96**	**0.82**	**0.94**	**0.94**	**0.99**	**0.99**

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.94•	**0.78●**	**0.77●**	**0.82●**	0.63•	**0.92●**	**0.90●**	**0.92**	**0.86●**
4	0.95●	**0.79●**	0.84•	**0.90●**	0.68	0.90●	**0.87●**	1.03	0.92
8	0.94●	**0.73●**	**0.77•**	0.91●	0.59	0.90●	0.87●	1.14	0.92

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.93●	**0.73●**	**0.72•**	**0.87●**	0.45	0.90●	0.86●	0.95	1.00
4	0.85●	**0.70●**	**0.73●**	0.85●	0.55●	0.88●	0.88•	0.98	0.89•
8	0.78●	**0.68●**	0.84•	0.83●	0.61	1.02	0.91	1.04	0.80•

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.94**	**0.88**	**0.80•**	**0.90**	**0.54•**	**0.90**	**0.89**	**0.95**	**1.03**
4	**0.85●**	**0.75•**	**0.73●**	**0.80●**	**0.60**	**0.90●**	**0.85●**	**1.22**	**0.94**
8	**0.81•**	**0.81**	**0.76•**	**0.82•**	**0.48•**	**1.18**	**1.12**	1.52	**0.95**

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting 1992Q4. For each horizon a total of 76 forecasts are computed. 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.

model indeed changes over time.

Focussing on the comparison of quasi real-time forecasts to those obtained with PBTV, we can divide the different models into three groups according to their degree of robustness of forecasting performance against model misspecification.

First, for the small BVAR, the BFAVAR, and both model averaging techniques, EWA and BMA, model misspecification has not a large impact on the models' forecasting performance. The optimally specified models reduce the RMSE by only about 10% relative to the quasi real-time RMSEs.

Second, the LBVAR and FAAR are subject to a moderate degree of model misspecification. Their forecasting performance would improve by about 20% if one could optimally specify the models in real-time. For the LBVAR, we find that in the PBTV specification the degree of shrinkage λ varies strongly over time, while in the quasi real-time specification λ is very stable over time. In the same vein, we find that in the PBTV specification the optimal number of static factors r for the FAAR varies largely over time, while the number of factors chosen via the information criterion of Bai and Ng (2002) in the quasi real-time exercise is rather stable.

Third, the quasi real-time forecasts of the DFM are subject to the largest degree of model misspecification. Choosing the *optimal* specification rather than that based on information criteria leads to a considerable improvement the model's forecasting performance. One reason for the large degree of misspecification of the DFM in real-time is that the optimal number of dynamic factors q for each of the four key variables seems to vary substantially for different forecasting horizons. Further, it turns out that the number of dynamic factors q chosen according to the Bai and Ng (2007) information criterion is always considerably smaller than the ex post *optimal* number of dynamic factors.[15]

Real-Time Performance Based Model Selection To check whether performance based model selection can increase the accuracy of the different forecasting models also in a quasi real-time exercise, we specify the models based on past forecasting performance rather than on the various information criteria. We call this approach *performance based model selection, real time* (PBRT). With PBRT, we evaluate the performance of the various specifications of the different forecasting models over a subevaluation sample $t = T - s^{eval} + 1, ..., T$. The best specification of each forecasting model for each horizon is then used to estimate the final model with information up to T and to compute forecasts for $T + h$.

To be consistent, we set the length of the subevaluation sample s^{eval} equal to 4 quarters and estimate the various specifications of the different forecasting models for the subsample evaluation as well as for the final forecast over a rolling window of 60 quarters.[16] In table 7

[15]Recall that any dynamic factor model with q dynamic factors can be rewritten as a static factor model with $r = q(s+1)$ static factors, where s is the number of lags in the dynamic factors. The Bai and Ng (2007) information criterion takes the optimal number of static factors r (obtained with the information criterion proposed by Bai and Ng (2002) with r^{max}) as well as a maximum number of lags s^{max} and a maximum number of dynamic factors q^{max} as given when optimizing over q. By contrast, for the performance based model specification we proceed just the other way around. Specifically, we define q^{max} and s^{max} and chose the best performing q-s-specification. The resulting number of static factors r can thus easily exceed r^{max}. Therefore our approach allows for a lot more dynamics in the dynamic factors which apparently increases the DFM's forecasting performance considerably.

[16]This implies that we cannot start to compute forecasts in 1992Q4 as in the previous analysis. Instead, we

Table 7: RMSEs of different forecasting models relative to AR(2) benchmark, PBRT and IC

Performance based model selection, real-time

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.99**	**0.95**	1.09	**0.90•**	1.19	1.01	1.01	1.07	**0.91**
4	0.99	0.98	1.11•	**0.96**	1.56●	1.00	1.01	1.62•	**0.94**
8	**0.98**	1.03	1.34•	**0.98**	1.75●	1.05	**0.96**	7.66	**1.00**

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	0.99	0.89●	1.03	**0.89●**	1.30●	0.96•	0.96	1.10	**0.92•**
4	1.01	1.03	1.23●	0.99	1.54●	0.99	1.05	1.32●	0.99
8	1.01	1.08	1.50	0.98	1.93●	0.97	0.97	4.36•	1.02

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**1.12**	**0.91**	**1.04**	**0.93**	**1.13**	**0.99**	**1.03**	**1.07**	**0.99**
4	**1.01**	**0.93**	**1.00**	**0.91•**	**1.20**	**1.00**	**1.08**	1.75	**0.94**
8	**0.97**	1.05	1.23	**0.90**	2.39	1.20	1.06	8.77	**0.85**

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**1.03**	**0.97**	**1.08**	**0.93**	**1.03**	**0.99**	**1.01**	**1.09**	**1.02**
4	**1.06**	**0.95**	**1.21**	**0.86•**	**1.26•**	**1.04**	**0.97**	**1.48**	**0.97**
8	**1.20•**	1.27●	**0.99**	**0.91**	**1.21**	1.46	**1.32**	6.22	**0.96**

Information criteria based model selection

(a) Output growth

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**0.98**	**0.92**	0.96	**0.90•**	0.96	**0.99**	1.01	1.03	**0.91**
4	0.97	0.98	1.07	**0.96**	1.00	1.00	1.02	1.10	**0.94**
8	**0.97**	1.01	1.13●	**0.97**	1.02	**0.95**	**1.00**	**0.99**	**0.99**

(b) CPI Inflation Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	1.03	**0.89●**	0.96	**0.89●**	0.96	0.99	0.96	0.99	0.93
4	1.02•	0.99	1.13	0.99	1.11	0.99	1.00	1.18	0.98
8	1.00	0.99	1.17	0.98•	1.11	0.96	0.96	1.29	1.01

(c) Interest Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**1.04**	**0.97**	**0.97**	**0.94**	**1.03**	**1.01**	**1.04**	1.08•	**0.98**
4	**1.08**	**0.99**	**1.04**	**0.93**	**1.09**	**1.07**	**1.14**	1.14	**0.91**
8	1.10	1.08	1.29	**0.91**	1.24	1.25•	1.11	1.19	**0.83•**

(d) Unemployment Rate

horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA	VAR	BVAR
1	**1.00**	**0.97**	**0.99**	**0.97**	**1.00**	**0.97●**	**0.97•**	**1.24•**	**1.01**
4	**1.00**	**0.91**	**1.09**	**0.91**	**1.14**	**1.03**	**1.03**	**1.41**	**0.94**
8	**1.00**	1.01	**1.09**	**0.93**	**1.18**	**1.35**	1.40	1.71	**0.92**

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting 1996Q3. For each horizon a total of 61 forecasts is computed. 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.

we report the results of this exercise, again relative to the AR(2) benchmark. We additionally show the relative RMSEs of the different forecasting models with information criteria based model selection for the same evaluation sample to facilitate the direct comparison of the two approaches.

The results indicate that the reduction of the relative RMSEs found for the PBTV approach is not attainable in real-time. The relative RMSEs of the PBRT exercise are mostly even larger than those obtained with model specifications based on information criteria. Especially for the two factor models, FAAR and DFM, the forecasting precision deteriorates considerably with PBRT. By contrast, the LBVAR and the BFAVAR seem to be quite robust against different model specification approaches. Among all the forecasting models considered, they provide the most accurate forecasts in most cases.

As for why the PBRT approach fails to produce results that are comparable to those which can be achieved with PBTV, there are several possibilities. First, past forecasting performance might not be a good indicator for future forecasting performance if shocks occur, even for samples that are temporally not far apart. Second, the length of the subevaluation sample might simply be too short. Indeed, we find that we can slightly improve the performance of all forecasting models with PBRT by setting $s^{eval} = 8$ rather than $s^{eval} = 4$.[17] Further, for higher horizons the time gap between T and the periods in which the forecasts for the subevaluation sample $t = T - s^{eval} + 1, ..., T$ are estimated might simply be too large so that the model specification that produces the most accurate high-horizon forecasts during the subevaluation sample is no longer optimal for forecasts starting in period T.

6 Forecasting a Larger Number of Macroeconomic Variables

Often forecasters are interested in the forecasts of a larger number of variables than the four key variables considered so far. In table 8, panel (a)-(l), we show the average of the relative RMSEs for the different forecasting models for eleven core macroeconomic variables. Besides those considered thus far, this includes variables such as private consumption, machinery and equipment investment, wages, industrial production, PPI inflation, a ten year interest rate and the current account balance. To asses the relative short- and long-term forecasting performance of the models, the average RMSEs are computed over the forecast horizons $h = 1, ..., 4$ and $h = 1, ..., 8$, respectively. In the upper part of table 8, we additionally summarize the short- and long term average RMSEs for each model over all eleven variables under consideration.

The results confirm our findings from section 5. Of all forecasting models considered, the LBVAR and the BFAVAR obtain the largest gains over the AR(2) benchmark in terms of short- as well as long-term forecasting performance for all eleven variables. Both model averaging techniques, EWA and BMA, also slightly outperform the benchmark for short forecasting horizons, while they perform just as good over all horizons up to $h = 8$. By contrast, the two remaining

choose 1996Q3 as starting point which leaves us with a total of 61 forecasts to evaluate for each forecasting horizon.

[17] Increasing the length of the subevaluation sample always comes at the cost of loosing data for the evaluation of the final forecasts in our exercise. Although this does no play a role in an actual real-time forecasting situation, the number of observations is typically limited here as well which is a clear constraint to the PBRT approach.

Table 8: Average RMSEs relative to AR(2) benchmark

Average over all 11 variables

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	4.70	1.01	0.93	1.01	0.93	1.00	0.96	0.97
1-8	5.22	1.00	0.93	1.05	0.92	1.04	0.98	0.99

(a) Output growth

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	3.65	0.99	0.95	1.07	0.94	1.06	0.98	0.99
1-8	3.61	0.99	0.98	1.14	0.95	1.10	0.97	0.99

(b) CPI Inflation Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	1.39	1.02	0.94	1.04	0.95	1.05	0.98	0.96
1-8	1.41	1.01	0.94	1.07	0.96	1.06	0.97	0.95

(c) Interest Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	0.82	1.03	0.90	0.98	0.95	0.97	1.00	1.02
1-8	1.29	1.05	0.91	1.04	0.93	1.03	1.05	1.05

(d) Unemployment Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	0.57	1.00	0.94	1.01	0.96	1.04	0.99	0.98
1-8	0.92	1.00	0.94	1.05	0.93	1.11	1.09	1.10

(e) Private Consumption

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	2.78	1.06	0.97	1.12	0.98	1.08	1.00	1.01
1-8	2.71	1.04	0.97	1.26	0.98	1.21	1.02	1.02

(f) Machinery and Equipment Investment

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	14.13	1.00	0.91	0.99	0.91	0.99	0.95	0.94
1-8	14.24	1.00	0.94	1.02	0.91	1.01	0.95	0.98

(g) Wages

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	2.81	1.00	0.87	0.97	0.91	0.91	0.94	0.93
1-8	2.89	1.00	0.89	1.03	0.94	0.98	0.96	0.95

(h) Industrial Production

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	15.46	0.96	0.76	0.79	0.76	0.81	0.81	0.83
1-8	18.22	0.80	0.67	0.69	0.66	0.69	0.69	0.70

(j) PPI Inflation Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	3.73	0.99	0.99	0.93	0.99	0.94	0.95	0.94
1-8	3.78	0.99	1.00	0.99	0.99	0.99	0.97	0.99

(k) Long Term Interest Rate

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	0.62	1.04	0.95	1.13	0.94	1.11	1.02	1.00
1-8	0.89	1.05	0.91	1.17	0.88	1.15	1.04	1.02

(l) Current Account

horizon	AR(2)	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1-4	5.70	1.02	1.02	1.11	1.00	1.08	1.01	1.04
1-8	7.49	1.03	1.04	1.14	1.00	1.13	1.03	1.09

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. Column 2 shows the absolute RMSEs for the AR(2) benchmark model, while all other RMSEs are computed relative to the AR(2) RMSEs.

factor models, FAAR and DFM, perform just as good as the AR(2) benchmark for short-term forecasts, but yield less precise forecasts when all horizons up to $h = 8$ are taken into account.

This pattern mostly continues to hold true when analysing each of the eleven variables under consideration separately. In most cases either the LBVAR or the BAVAR perform best, while either the FAAR or the DFM performs worst. However, there are some exceptions. For example, the forecasts of the PPI inflation rate (panel j) obtained by the LBVAR and the BFAVAR are the least accurate ones, at least for short horizons. Further, for the short term interest rate (panel c) the two factor models FAAR and DFM actually perform slightly better than the two model averaging approaches EWA and BMA, at least in the short run.

The average relative RMSEs also reveal that there are considerable differences between the variables under consideration when it comes to the size of the gains in forecasting accuracy of the best performing large scale methods over the AR(2) benchmark. The largest gain amounts to 34% and is obtained by the BFAVAR for industrial production for long forecasting horizons. By contrast, for the current account balance, not even the best performing large scale methods, the LBVAR and the BFAVAR, can actually beat the benchmark. However, in most cases either of the two models (or both) can achieve a moderate reduction in the average relative RMSEs over the AR(2) benchmark.

The detailed tables in the appendix contain the results for each model and all variables. They show that many variables under consideration are characterized by very low persistence and thus their predictable component is presumably 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 univariate forecasting models.

It remains, however, to be explored whether the usage of one coherent forecasting model like a LBVAR is advantageous in terms of the interpretation of the forecasts of a larger number of variables. After all, the other models include different predictors to forecast the different variables, so it might be more difficult to come up with an interpretation of the forecasts that is coherent for the whole set of variables that forecasters are typically interested in.

7 Conclusion

We have studied three different approaches that are able to use the information from a large dataset consisting of 123 variables in quarterly frequency to forecast eleven German core macroeconomic time series. We find that, overall, the large Bayesian vector autoregression (LBVAR) and the Bayesian factor augmented vector autoregression (BFAVAR) perform best and generally yield more accurate forecasts than an AR(2) benchmark model and the other large scale forecasting methods. The LBVAR and the BFAVAR are also the only models that yield unbiased forecasts for a set of four key variables, at least for short horizons.

For the prediction of output growth, we have additionally compared the forecasting performance of the different large scale models to that of different versions of the ifo indicator and we have studied the forecasting performance of the various models during the Great Recession. We find that while the ifo expectations indicator yields some improvement in output growth forecasts over the whole sample, none of the models is able to forecast the Great Recession.

We also checked whether the low forecasting performance of some models is caused by model misspecification. This is indeed the case. In particular, an *optimally* specified dynamic factor model (DFM) could in principle produce the most accurate forecasts among all models considered. However, this performance is unattainable in a real-time forecasting application, whether the model specification is based on information criteria or on past forecasting performance.

Regarding the size of the gains in forecasting accuracy over the AR(2) benchmark obtained with the LBVAR and the BFAVAR in our quasi real-time exercise, we find considerable differences between the various variables. While for industrial production, for example, the reduction of the average relative root mean squared prediction error over forecast horizons up to eight quarters ahead amounts to more than 30%, the gains rarely exceed 10% in most other cases. 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. We additionally check the informational content of the forecasts as measured by the R^2 of Mincer-Zarnowitz regressions and find that it is very weak in most cases, except for highly persistent variables. 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(2) benchmark.

Still, when forecasters are interested in simultaneously predicting a larger number of variables, large-scale forecasting models 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.

References

Bai and Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25*, 52 – 60.

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.

Kuzin, V., M. Marcellino, and C. Schumacher (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics 28*, 392–411.

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.

Timmermann, A. and H. K. van Dijk (2013). Dynamic econometric modeling and forecasting in the presence of instability. *Journal of Econometrics 177*, 131 – 133.

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 A: Data Description

Our dataset consists of 123 macroeconomic time series in quarterly frequency covering the period from 1978Q1 until 2013Q3. Below we provide a detailed description of the dataset that reads as follows:

- Number of the series
- Code of the series (as used in the respective original source, if available)
- Series label
- Source of the series
  - FSO: Federal Statistical Office Germany
    (a) Statistisches Bundesamt, Volkswirtschaftliche Gesamtrechnungen 1970 bis 1991, Fachserie 18 Reihe S.28
    (b) Statistisches Bundesamt, Volkswirtschaftliche Gesamtrechnungen 3. Vierteljahr 2013, Fachserie 18 Reihe 1.3
    (c) Statistisches Bundesamt, Bauhauptgewerbe (query at unit E206)
  - DS: Datastream
  - CS: Schumacher (2007)
- Transformation of the series
  - WG: prior 1991 West-German series rescaled to Pan-German series
  - log: natural logarithm
  - SA: series seasonally adjusted in EViews7[18] (all other series were already seasonally adjusted in the original data source).

[18] Census X12, multiplicative (additive for series with negative numbers), TrendFilter: Auto, no ARIMA, no data transformation

No.	Code prior 1991	Code post 1991	Name of Series	Source prior 1991	Source post 1991	Transformation WG	log	SA
			Use of GDP and gross value added					
1	1.1	BDGDP...D	Real gross domestic product	FSO(a)	DS	x	x	
2	2.3.2	BDCNPER.D	Real private consumption	FSO(a)	DS	x	x	
3	2.3.2	BDCNGOV.D	Real government consumption	FSO(a)	DS	x	x	
4	2.3.5	BDGCMAC.D	Gross fixed capital formation: machinery and equipment	FSO(a)	DS	x	x	
5	2.3.5	BDGCCON.D	Gross fixed capital formation: construction	FSO(a)	DS	x	x	
6	2.3.5	BDGCINT.D	Gross fixed capital formation: other	FSO(a)	DS	x	x	
7	2.3.10	BDEXNGS.D	Exports	FSO(a)	DS	x	x	
8	2.3.10	BDIMNGS.D	Imports	FSO(a)	DS	x	x	
9	2.2	BDVAPAAFE	Gross value added: mining and fishery	FSO(a)	DS	x	x	
10	2.2	BDVAPAECE	Gross value added: producing sector excluding construction	FSO(a)	DS	x	x	
11	2.2	BDVAPACND	Gross value added: construction	FSO(a)	DS	x	x	
12	2.2	BDVAPATFD	Gross value added: wholesale and retail trade, restaurants, hotels, transport	FSO(a)	DS	x	x	
13	2.2		Gross value added: financing and rents*			x	x	
13	2.2	BDVAPAICD/B		FSO(a)	DS			
13	2.2	BDVAPAFID/B		FSO(a)	DS			
13	2.2	BDVAPARED/B		FSO(a)	DS			
13	2.2	BDVAPASTD/B		FSO(a)	DS			
14	2.2		Gross value added: services**			x	x	
14	2.2	BDVAPAAHD/B		FSO(a)	DS			
14	2.2	BDVAPAOSD/B		FSO(a)	DS			
			Prices					
15	JQ0730	BDGDPIPDE	Deflator of GDP	CS	DS	x	x	
16	JQ0059	BDOCMP06E	Deflator of private consumption expenditure	CS	DS	x	x	
17	JQ0060	BDOEXP02E	Deflator of government consumption expenditure	CS	DS	x	x	
18	JQ006	BDGCMAC,B	Deflator of machinery and equipment	CS	DS	x	x	
19	JQ0065	BDIPDCNSE	Deflator of construction	CS	DS	x	x	
20	BDTOTPRCF	BDTOTPRCF	Terms of trade	DS	DS	x	x	x
21	JQ0214	BDEXPPRCF	Export prices	CS	DS	x	x	x
22	JQ0205	BDIMPPRCF	Import prices	CS	DS	x	x	x
23	ECO:DEU:CPIH/Q	BDCONPRCE	Consumer price index	CS	DS	x	x	
24	BDPROPRCF	BDPROPRCF	Producer price index	DS	DS		x	x
			Labor market					
25	2.1.6	2.1.6	Residents	FSO(a)	FSO(b)	x	x	
26	2.1.6	2.1.6	Labour force	FSO(a)	FSO(b)	x	x	
27	2.1.6	2.1.6	Unemployed	FSO(a)	FSO(b)	x	x	
28	2.1.6	2.1.6	Employees and self-employed	FSO(a)	FSO(b)	x	x	
29	2.1.6	2.1.6	Employees	FSO(a)	FSO(b)	x	x	
30	2.1.6	2.1.6	Self-employed	FSO(a)	FSO(b)	x	x	
31	2.1.7	2.1.7	Volume of work, employees and self-employed	FSO(a)	FSO(b)	x	x	
32	2.1.7	2.1.7	Volume of work, employees	FSO(a)	FSO(b)	x	x	
33	2.1.7	2.1.7	Hours, employees and self-employed	FSO(a)	FSO(b)	x	x	
34	2.1.7	2.1.7	Hours, employees	FSO(a)	FSO(b)	x	x	
35	2.1.8	2.1.8	Productivity, per employee	FSO(a)	FSO(b)	x	x	
36	2.1.8	2.1.8	Productivity, per hour	FSO(a)	FSO(b)	x	x	
37	2.1.8	2.1.8	Wages and salaries per employee	FSO(a)	FSO(b)	x	x	
38	2.1.8	2.1.8	Wages and salaries per hour	FSO(a)	FSO(b)	x	x	
39	2.1.4	2.1.4	Wages and salaries, excluding employers social security contributions	FSO(a)	FSO(b)	x	x	
40	2.1.8	2.1.8	Unit labour costs, per production unit	FSO(a)	FSO(b)	x	x	
41	2.1.8	2.1.8	Unit labour costs, per production unit, hourly basis	FSO(a)	FSO(b)	x	x	
42	GS1513	BDUSC,04O(BDUSCC04O)	Vacancies	CS	DS	x	x	
43	US02CC	BDUSCC02Q	Unemployment rate	CS	DS	x		

* Series constructed as average of four real series, each weighted with corresponding nominal series.
**Series constructed as average of two real series, each weighted with corresponding nominal series.

No.	Code prior 1991	Code post 1991	Name of Series	Source prior 1991	Source post 1991	Transformation WG	log	SA
			Financial					
44	BDSU0101,BDSU0304R	BDSU0101,BDSU0304R	Money market rate, overnight deposits	DS	DS			
45	BDSU0104,BDSU0310R	BDSU0104,BDSU0310R	Money market rate, 1 months deposits	DS	DS			
46	BDSU0107,BDSU0316R	BDSU0107,BDSU0316R	Money market rate, 3 months deposits	DS	DS			
47	BDWU0898	BDWU0898	Bond yields with average rest maturity from 1 to 2 years	DS	DS			
48	BDWU0899	BDWU0899	Bond yields with average rest maturity from 2 to 3 years	DS	DS			
49	BDWU0900	BDWU0900	Bond yields with average rest maturity from 3 to 4 years	DS	DS			
50	BDWU0901	BDWU0901	Bond yields with average rest maturity from 4 to 5 years	DS	DS			
51	BDWU0902	BDWU0902	Bond yields with average rest maturity from 5 to 6 years	DS	DS			
52	BDWU0903	BDWU0903	Bond yields with average rest maturity from 6 to 7 years	DS	DS			
53	BDWU8606	BDWU8606	Bond yields with average rest maturity from 7 to 8 years	DS	DS			
54	BDWU8607	BDWU8607	Bond yields with average rest maturity from 8 to 9 years	DS	DS			
55	BDWU8608	BDWU8608	Bond yields with average rest maturity from 9 to 10 years	DS	DS			
56	BDWU001AA	BDWU001AA	Stock prices: CDAX	DS	DS		x	
57	BDWU3141A	BDWU3141A	Stock prices: DAX	DS	DS		x	
58	BDWU035AA	BDWU035AA	Stock prices: REX	DS	DS		x	
			Misc					
59	BDEA4001B	BDEA4001B	Current account: goods trade	DS	DS			
60	BDEA4100B	BDEA4100B	Current account: services	DS	DS			
61	BDEA4170B	BDEA4170B	Current account: factor income	DS	DS			
62	BDEA4220B	BDEA4220B	Current account: transfers	DS	DS			
63	BDHWWAINF	BDHWWAINF	HWWA raw material price index	DS	DS		x	x
64	BDQSLI12G	BDQSLI12G	New car registrations	DS	DS		x	
			Industry					
65		BDUSNA04G	Production: intermediate goods industry	CS	DS	x	x	
66		BDUSNA05G	Production: capital goods industry	CS	DS	x	x	
67		BDUSNI67G	Production: durable and non-durable consumer goods industry	CS	DS	x	x	
68		BDUSNA39G	Production: mechanical engineering	CS	DS	x	x	
69		BDUSNA42G	Production: electrical engineering	CS	DS	x	x	
70		BDUSNA50G	Production: vehicle engineering	CS	DS	x	x	
71		BDSTDCAPG	Export turnover: intermediate goods industry	CS	DS	x	x	x
72		BDSTFINTG	Domestic turnover: intermediate goods industry	CS	DS	x	x	x
73		BDSTFCAPG	Export turnover: capital goods industry	CS	DS	x	x	x
74		BDSTDCONG	Domestic turnover: capital goods industry	CS	DS	x	x	x
75		BDSTFCONG	Export turnover: durable and non-durable consumer goods industry	CS	DS	x	x	x
76		BDSTDMYEG	Domestic turnover: durable and non-durable consumer goods industry	CS	DS	x	x	x
77		BDSTFMYEG	Export turnover: mechanical engineering	CS	DS	x	x	x
78		BDSTDCEOG	Domestic turnover: mechanical engineering	CS	DS	x	x	x
79		BDSTFCEOG	Export turnover: electrical engineering industry	CS	DS	x	x	x
80		BDSTFCEOG	Domestic turnover: electrical engineering industry	CS	DS	x	x	x
81		BDSTFCEOG	Export turnover: vehicle engineering industry	CS	DS	x	x	x
82		BDSTFVEMG	Domestic turnover: vehicle engineering industry	CS	DS	x	x	x
83		BDDBPORDG	Orders received: intermediate goods industry from domestic market	CS	DS	x	x	x
84		BDOBPORDG	Orders received: intermediate goods industry from abroad	CS	DS	x	x	x
85		BDDCPORDG	Orders received: capital goods industry from domestic market	CS	DS	x	x	x
86		BDOCPORDG	Orders received: capital goods industry from abroad	CS	DS	x	x	x
87		BDDCNORDG	Orders received: consumer goods industry from domestic market	CS	DS	x	x	x
88		BDOCNORDG	Orders received: consumer goods industry from abroad	CS	DS	x	x	x
89		BDNODMYEG	Orders received: mechanical engineering industry from domestic market	CS	DS	x	x	x
90		BDNOFMYEG	Orders received: mechanical engineering industry from abroad	CS	DS	x	x	x
91		BDUSC587G	Orders received: electrical engineering industry from domestic market	CS	DS	x	x	x
92		BDUSC588G	Orders received: electrical engineering industry from abroad	CS	DS	x	x	x
93		BDUSC659G	Orders received: vehicle engineering industry from domestic market	CS	DS	x	x	x
94		BDUSC660G	Orders received: vehicle engineering industry from abroad	CS	DS	x	x	x

No.	Code prior 1991	Code post 1991	Name of Series	Source prior 1991	Source post 1991	Transformation WG	log	SA
			Construction					
95	GS17DA	BDUSDA17G	Orders received by the construction sector: building construction	CS	DS	x	x	
96	GS20DA	BDUSDA20G	Orders received by the construction sector: civil engineering	CS	DS	x	x	
97	GS18DA	BDUSDA18G	Orders received by the construction sector: residential building	CS	DS	x	x	
98	GS19DA	BDUSDA19G	Orders received by the construction sector: non-residential building construction	CS	DS	x	x	
99		1.3.1	Man-hours worked in building construction	CS	FSO(c)	x	x	x
100		1.3.1	Man-hours worked in civil engineering	CS	FSO(c)	x	x	x
101		1.3.1	Man-hours worked in residential building	CS	FSO(c)	x	x	x
102		1.3.1	Man-hours worked in industrial building	CS	FSO(c)	x	x	x
103		1.3.1	Man-hours worked in public building	CS	FSO(c)	x	x	x
104		1.4.1	Turnover: building construction	CS	FSO(c)	x	x	x
105		1.4.1	Turnover: civil engineering	CS	FSO(c)	x	x	x
106		BDUSMB36B	Turnover: residential building	CS	DS	x	x	x
107		BDUSMB31B	Turnover: industrial building	CS	DS	x	x	x
108		BDUSMB37B	Turnover: public building	CS	DS	x	x	x
109		BDESPICNG	Production in the construction sector	CS	DS	x	x	x
			Surveys					
110	WGIFOCPAE	BDIFDMPAQ	Business situation: capital goods producers	DS	DS	x		
111	WGIFOCGAE	BDIFDMCAQ	Business situation: producers durable consumer goods	DS	DS	x		
112	WGIFOCOAE	BDIFDMNAQ	Business situation: producers non-durable consumer goods	DS	DS	x		
113	WGIFOCPKE	BDIFDMPKQ	Business expectations for the next 6 months: producers of capital goods	DS	DS	x		
114	WGIFOCGHE	BDIFDMCKQ	Business expectations for the next 6 months: producers of durable consumer goods	DS	DS	x		
115	WGIFOCOKE	BDIFDMNKQ	Business expectations for the next 6 months: producers of non-durable consumer goods	DS	DS	x		
116	WGIFORTHE	BDIFDRSKQ	Business expectations for the next 6 months: retail trade	DS	DS	x		
117	WGIFOWHHE	BDIFDWSKQ	Business expectations for the next 6 months: wholesale trade	DS	DS	x		
118	WGIFOCPCE	BDIFDMPCQ	Stocks of finished goods: producers of capital goods	DS	DS	x		
119	WGIFOCGDE	BDIFDMCCQ	Stocks of finished goods: producers of durable consumer goods	DS	DS	x		
120	WGIFOCOCE	BDIFDNXCQ	Stocks of finished goods: producers of non-durable consumer goods	DS	DS	x		
121	WGIFOUCGQ	BDIFDMPQQ	Capacity utilisation: producers of capital goods	DS	DS	x	x	
122	WGIFOUCRQ	BDIFDMCQQ	Capacity utilisation: producers of durable consumer goods	DS	DS	x	x	
123	WGIFOUCOQ	BDIFDMNQQ	Capacity utilisation: producers of non-durable consumer goods	DS	DS	x	x	

Appendix B: Detailed Results for all Models for Eleven Variables

- Δgdp_t: annualized quarter-on-quarter GDP growth
- Δcpi_t: annualized quarter-on-quarter CPI inflation rate
- Δc_t: annualized quarter-on-quarter private consumption growth
- Δinv_t: annualized quarter-on-quarter investment growth (machinery and equipment investment)
- Δw_t: annualized quarter-on-quarter wage inflation rate
- Δip_t: annualized quarter-on-quarter industrial production growth
- Δppi_t: annualized quarter-on-quarter PPI inflation rate
- i_t^s: 3-months money market rate
- u_t: unemployment rate
- i_t^l: long-term interest rate (bond yields with average rest maturity from 9 to 10 years)
- ca_t: current account balance

Table 9: Univariate Autoregression (AR)

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) RMSE relative to AR(2) benchmark											
1	0.99	1.03	0.98	0.99	1.01	0.98	1.02•	1.03	1.00	1.04	1.00
2	0.98	1.01	1.06	0.99	1.02	1.01	0.99	1.03	1.00	1.04	1.02
3	0.98	1.02	1.09	1.00	0.98	0.97	0.98•	1.02	1.00	1.03	1.03
4	1.00	1.01	1.08	1.00	0.99	0.86	0.96●	1.04	1.00	1.04	1.03
5	0.97	1.00	1.07	1.00	0.99	0.78	0.96●	1.05	1.00•	1.06	1.04
6	0.99	1.01	1.02	1.01	1.00	0.67	1.00	1.06	1.01	1.06	1.04
7	1.00	1.00	1.00	1.00	1.00	0.56	1.01	1.07	1.01	1.06	1.04
8	0.99	0.99	1.00	1.01	1.00	0.54	1.00	1.08	1.01	1.06	1.04
(b) p-value Mincer-Zarnowitz regressions											
1	0.04	0.00	0.00	0.03	0.00	0.00	0.80	0.01	0.79	0.00	0.04
2	0.02	0.00	0.00	0.01	0.00	0.00	0.12	0.01	0.79	0.00	0.03
3	0.01	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.79	0.00	0.03
4	0.01	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.78	0.00	0.02
5	0.01	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.75	0.00	0.02
6	0.03	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.70	0.00	0.02
7	0.05	0.00	0.00	0.01	0.00	0.00	0.26	0.00	0.67	0.00	0.01
8	0.08	0.00	0.00	0.01	0.00	0.19	0.58	0.00	0.62	0.00	0.01
(c) R^2 Mincer-Zarnowitz regressions											
1	0.02	0.04	0.12	0.08	0.01	0.04	0.34	0.95	0.97	0.95	0.93
2	0.01	0.03	0.00	0.04	0.00	0.00	0.03	0.82	0.88	0.86	0.86
3	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.66	0.76	0.79	0.81
4	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.50	0.62	0.73	0.74
5	0.06	0.00	0.02	0.07	0.04	0.08	0.01	0.34	0.47	0.69	0.69
6	0.03	0.00	0.00	0.03	0.05	0.00	0.01	0.24	0.34	0.67	0.67
7	0.02	0.01	0.01	0.01	0.05	0.10	0.00	0.19	0.25	0.67	0.63
8	0.01	0.01	0.00	0.02	0.06	0.02	0.00	0.16	0.16	0.67	0.60

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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: Large Bayesian Vector Autoregression (LBVAR)

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) RMSE relative to AR(2) benchmark											
1	0.93	0.89●	0.99	0.90•	0.84●	0.83	1.11	0.95	0.99	1.02	1.00
2	0.93	0.92•	0.97	0.92	0.90●	0.76	0.96	0.89●	0.93	0.95	1.02
3	0.96	0.95	1.00	0.92	0.86●	0.76	0.94	0.88●	0.91	0.92	1.04
4	0.99	0.98	0.94●	0.92	0.88●	0.70	0.96	0.89	0.90	0.91	1.03
5	0.97	0.95	0.94●	0.92	0.88●	0.62	0.98	0.90	0.91	0.90	1.05
6	1.01	0.95	1.00	0.95	0.90●	0.63	0.99	0.91	0.93	0.88	1.06
7	1.02	0.94	0.97•	0.98	0.92	0.54	1.01	0.93	0.96	0.86	1.07
8	1.01	0.94	0.97	0.98	0.92	0.54	1.03	0.94	0.98	0.84	1.07
(b) p-value Mincer-Zarnowitz regressions											
1	0.44	0.51	0.00	0.64	0.06	0.99	0.55	0.21	0.48	0.47	0.03
2	0.21	0.05	0.04	0.71	0.06	0.44	0.55	0.16	0.31	0.40	0.03
3	0.06	0.01	0.05	0.36	0.04	0.16	0.12	0.12	0.19	0.40	0.03
4	0.01	0.00	0.01	0.05	0.01	0.03	0.06	0.09	0.12	0.41	0.03
5	0.00	0.00	0.02	0.01	0.00	0.02	0.10	0.05	0.07	0.43	0.02
6	0.00	0.00	0.00	0.01	0.00	0.03	0.18	0.03	0.05	0.43	0.02
7	0.00	0.00	0.02	0.00	0.00	0.04	0.21	0.01	0.03	0.38	0.02
8	0.00	0.00	0.01	0.00	0.00	0.17	0.18	0.01	0.02	0.30	0.01
(c) R^2 Mincer-Zarnowitz regressions											
1	0.07	0.15	0.01	0.16	0.05	0.12	0.23	0.95	0.97	0.94	0.93
2	0.00	0.05	0.01	0.06	0.02	0.02	0.05	0.84	0.90	0.84	0.86
3	0.01	0.02	0.02	0.00	0.01	0.00	0.00	0.68	0.82	0.75	0.80
4	0.06	0.00	0.01	0.02	0.00	0.02	0.00	0.49	0.74	0.68	0.73
5	0.10	0.00	0.02	0.08	0.01	0.04	0.00	0.30	0.65	0.61	0.67
6	0.13	0.01	0.00	0.07	0.03	0.03	0.00	0.15	0.58	0.56	0.64
7	0.15	0.02	0.01	0.13	0.05	0.03	0.00	0.06	0.51	0.52	0.59
8	0.10	0.04	0.01	0.11	0.09	0.01	0.00	0.01	0.45	0.48	0.56

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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: Factor Augmented Autoregression (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) RMSE relative to AR(2) benchmark											
1	0.99	0.97	1.02	0.96	0.94	0.79•	0.93	1.01	0.99	1.15•	1.08•
2	1.11	1.05	1.15	1.09	1.05	0.87	0.93	0.96	0.99	1.13●	1.09
3	1.09	1.05	1.16●	1.00	0.90	0.83	0.94	0.94	1.02	1.12	1.11
4	1.10	1.10	1.13●	0.92	0.97	0.67	0.91	0.99	1.05	1.14	1.15
5	1.10	1.05	1.29•	0.98	1.01	0.60	1.02	1.04	1.07	1.15	1.13
6	1.23•	1.04	1.37●	1.05	0.97	0.62	1.04	1.09	1.09	1.15	1.20
7	1.30●	1.19●	1.44●	1.12	1.19	0.57	1.06	1.13	1.10	1.21	1.20
8	1.18●	1.14	1.51•	1.03	1.18•	0.55	1.10	1.18	1.10	1.30•	1.20
(b) p-value Mincer-Zarnowitz regressions											
1	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.68	0.33	0.01	0.18
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.36	0.00	0.20
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.30	0.00	0.19
4	0.00	0.00	0.00	0.01	0.00	0.24	0.07	0.06	0.28	0.00	0.13
5	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.03	0.27	0.00	0.13
6	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.00	0.25	0.00	0.07
7	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.25	0.00	0.04
8	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.25	0.00	0.03
(c) R^2 Mincer-Zarnowitz regressions											
1	0.15	0.12	0.16	0.22	0.12	0.26	0.49	0.95	0.97	0.93	0.91
2	0.03	0.06	0.02	0.09	0.08	0.03	0.21	0.82	0.89	0.82	0.84
3	0.00	0.04	0.00	0.06	0.09	0.02	0.11	0.64	0.77	0.76	0.76
4	0.01	0.00	0.00	0.05	0.02	0.03	0.09	0.41	0.62	0.73	0.66
5	0.00	0.00	0.00	0.02	0.01	0.02	0.04	0.17	0.47	0.73	0.60
6	0.04	0.01	0.01	0.01	0.00	0.02	0.04	0.05	0.33	0.75	0.54
7	0.04	0.00	0.01	0.00	0.00	0.00	0.04	0.00	0.23	0.74	0.49
8	0.02	0.00	0.00	0.03	0.08	0.02	0.00	0.00	0.14	0.70	0.45

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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: 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) RMSE relative to AR(2) benchmark											
1	0.92•	0.90●	0.98	0.90	0.79●	0.83	1.12	0.98	1.00	1.02	1.00
2	0.93	0.93	0.96	0.93	0.93●	0.76	0.96	0.95	0.96	0.94•	1.00
3	0.94	0.98	1.01	0.92	0.95•	0.75	0.93	0.94•	0.94	0.91●	1.00
4	0.96	0.99	0.96●	0.89	0.97•	0.69	0.94	0.94	0.92	0.88●	1.00
5	0.94	0.98	0.96●	0.89	0.97●	0.61	0.96•	0.92	0.90	0.85●	1.00
6	0.97	0.98	1.00	0.91	0.98•	0.62	0.98	0.91	0.90	0.82●	1.00
7	0.98	0.97	0.99•	0.93	0.98●	0.52	1.00	0.91	0.90	0.80●	1.00
8	0.97	0.98	0.98•	0.94•	0.97●	0.53	1.01	0.90	0.91	0.80●	1.00
(b) p-value Mincer-Zarnowitz regressions											
1	0.44	0.92	0.00	0.52	0.25	0.98	0.15	0.18	0.68	0.08	0.13
2	0.29	0.09	0.04	0.85	0.03	0.68	0.71	0.09	0.63	0.05	0.11
3	0.11	0.00	0.02	0.35	0.00	0.11	0.21	0.05	0.57	0.04	0.08
4	0.09	0.00	0.00	0.01	0.00	0.00	0.11	0.03	0.50	0.04	0.06
5	0.13	0.00	0.00	0.00	0.00	0.00	0.16	0.02	0.44	0.04	0.04
6	0.14	0.00	0.00	0.04	0.00	0.01	0.26	0.01	0.39	0.04	0.03
7	0.21	0.00	0.00	0.08	0.00	0.14	0.33	0.01	0.35	0.03	0.02
8	0.27	0.00	0.00	0.25	0.00	0.48	0.33	0.00	0.31	0.01	0.02
(c) R^2 Mincer-Zarnowitz regressions											
1	0.09	0.12	0.00	0.17	0.09	0.11	0.27	0.95	0.97	0.94	0.93
2	0.00	0.03	0.03	0.02	0.00	0.00	0.02	0.83	0.89	0.86	0.86
3	0.02	0.00	0.02	0.01	0.01	0.04	0.01	0.67	0.80	0.79	0.81
4	0.02	0.00	0.00	0.12	0.04	0.13	0.03	0.50	0.70	0.73	0.74
5	0.01	0.00	0.01	0.16	0.06	0.14	0.01	0.35	0.60	0.70	0.70
6	0.01	0.00	0.00	0.08	0.07	0.08	0.01	0.24	0.52	0.69	0.67
7	0.00	0.00	0.00	0.06	0.05	0.03	0.00	0.17	0.45	0.67	0.63
8	0.00	0.00	0.00	0.03	0.07	0.01	0.00	0.12	0.38	0.65	0.60

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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 13: Dynamic Factor Model (DFM)

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) RMSE relative to AR(2) benchmark											
1	1.01	1.00	0.97	0.97	0.88	0.84	0.94	0.99	0.99	1.13•	1.07
2	1.13	1.04	1.07	1.12	0.91	0.91	0.94	0.94	1.00	1.11	1.07
3	1.09	1.03	1.16●	0.98	0.90	0.84	0.95	0.94	1.06	1.08	1.07
4	1.02	1.11	1.13●	0.92	0.96	0.65	0.94	1.00	1.10	1.11	1.09
5	1.08	1.01	1.30•	0.94	0.97	0.61	1.04	1.03	1.14	1.13	1.11
6	1.12●	1.02●	1.32	1.03	0.97	0.62	1.04	1.08	1.17	1.13	1.21
7	1.22●	1.15●	1.27	1.08	1.07	0.55	1.04	1.10	1.18	1.20	1.20
8	1.11	1.14	1.42•	1.02	1.15	0.53	1.01	1.13	1.18	1.28•	1.23
(b) p-value Mincer-Zarnowitz regressions											
1	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.91	0.24	0.01	0.16
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.25	0.00	0.22
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.23	0.00	0.27
4	0.01	0.00	0.00	0.01	0.00	0.32	0.02	0.08	0.22	0.00	0.19
5	0.00	0.00	0.00	0.01	0.00	0.06	0.00	0.03	0.20	0.00	0.14
6	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.18	0.00	0.06
7	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.17	0.00	0.04
8	0.00	0.00	0.00	0.00	0.00	0.09	0.18	0.00	0.16	0.00	0.03
(c) R^2 Mincer-Zarnowitz regressions											
1	0.15	0.11	0.22	0.23	0.18	0.21	0.49	0.95	0.97	0.93	0.91
2	0.01	0.06	0.03	0.08	0.12	0.02	0.20	0.82	0.89	0.83	0.84
3	0.00	0.04	0.00	0.07	0.07	0.02	0.11	0.63	0.76	0.78	0.77
4	0.00	0.00	0.00	0.05	0.00	0.07	0.08	0.38	0.60	0.75	0.68
5	0.00	0.01	0.00	0.04	0.00	0.03	0.03	0.16	0.42	0.76	0.61
6	0.00	0.01	0.01	0.03	0.00	0.04	0.04	0.04	0.27	0.77	0.54
7	0.00	0.00	0.00	0.02	0.00	0.02	0.05	0.00	0.16	0.75	0.50
8	0.00	0.00	0.00	0.05	0.07	0.06	0.04	0.00	0.08	0.71	0.44

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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 14: 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) RMSE relative to AR(2) benchmark											
1	0.98	0.99	0.97	0.97	0.98	0.95	0.98•	1.00	0.97●	1.03	1.00
2	0.96	0.96●	0.97•	0.96•	0.95●	0.82	0.96•	0.97	0.96•	1.02	1.00
3	0.96	0.98	1.05	0.96	0.94●	0.76	0.93•	0.99	0.99	1.01	1.01
4	1.00	0.98	1.03	0.89	0.91●	0.70	0.93	1.03	1.03	1.01	1.01
5	0.95	0.99	1.04	0.89	0.86●	0.59	0.95	1.06	1.09	1.04	1.02
6	0.97	0.97	1.03	0.96	0.97	0.61	0.98	1.09	1.15	1.04	1.06
7	0.97	0.97	1.03	0.96	1.00	0.51	1.02	1.13	1.23	1.06	1.07
8	0.96	0.95•	1.02	0.98	1.05	0.54	1.05	1.17	1.31	1.10	1.09
(b) p-value Mincer-Zarnowitz regressions											
1	0.05	0.01	0.00	0.06	0.00	0.00	0.71	0.01	0.75	0.00	0.04
2	0.05	0.00	0.01	0.05	0.01	0.01	0.34	0.01	0.72	0.00	0.05
3	0.05	0.00	0.00	0.07	0.00	0.16	0.23	0.01	0.67	0.00	0.05
4	0.01	0.00	0.00	0.21	0.00	0.09	0.28	0.00	0.65	0.00	0.04
5	0.14	0.00	0.00	0.18	0.01	0.55	0.33	0.00	0.64	0.00	0.03
6	0.17	0.00	0.00	0.03	0.00	0.45	0.40	0.00	0.53	0.00	0.02
7	0.25	0.00	0.00	0.10	0.00	0.79	0.14	0.00	0.29	0.00	0.02
8	0.29	0.00	0.00	0.08	0.00	0.24	0.02	0.00	0.07	0.00	0.01
(c) R^2 Mincer-Zarnowitz regressions											
1	0.03	0.06	0.16	0.10	0.03	0.03	0.40	0.95	0.97	0.95	0.93
2	0.00	0.04	0.04	0.06	0.01	0.00	0.06	0.84	0.89	0.86	0.87
3	0.01	0.03	0.01	0.00	0.00	0.00	0.00	0.67	0.77	0.79	0.81
4	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.48	0.61	0.75	0.75
5	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.29	0.40	0.72	0.70
6	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.17	0.18	0.70	0.66
7	0.00	0.01	0.00	0.00	0.02	0.02	0.01	0.11	0.03	0.69	0.62
8	0.03	0.00	0.00	0.00	0.04	0.00	0.03	0.09	0.00	0.68	0.58

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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 15: 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) RMSE relative to AR(2) benchmark											
1	1.01	0.97	0.94	0.97	0.92●	0.97	1.00	1.03	0.97●	0.86●	1.01
2	0.95	0.93●	1.00	0.96•	0.93	0.89	0.94	0.95	0.96	1.04	1.04
3	0.97	0.94	1.06	0.96	0.92•	0.79	0.89	1.01	0.98	1.05	1.06
4	1.03	0.98	1.03	0.87	0.93•	0.69	0.94	1.07	1.02	1.04	1.07
5	0.95	1.01	1.07	0.92	0.84●	0.59	0.96	1.10	1.08	1.05	1.07
6	1.01	0.94•	1.04	1.02	0.98	0.61	1.02	1.10	1.16	1.05	1.12
7	0.99	0.91	1.04	1.04	1.03	0.54	1.07	1.09	1.27	1.03	1.16
8	1.04	0.96	1.02	1.08	1.03	0.56	1.11●	1.06	1.35	1.03	1.17
(b) p-value Mincer-Zarnowitz regressions											
1	0.00	0.01	0.00	0.03	0.00	0.00	0.77	0.01	0.63	0.00	0.05
2	0.04	0.00	0.00	0.03	0.01	0.00	0.30	0.01	0.64	0.00	0.11
3	0.03	0.00	0.00	0.03	0.00	0.01	0.47	0.01	0.61	0.00	0.12
4	0.00	0.00	0.00	0.29	0.00	0.08	0.14	0.01	0.56	0.00	0.09
5	0.10	0.00	0.00	0.05	0.02	0.23	0.20	0.00	0.54	0.00	0.08
6	0.01	0.00	0.00	0.00	0.00	0.17	0.05	0.00	0.42	0.00	0.05
7	0.04	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.16	0.00	0.03
8	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.04	0.00	0.02
(c) R^2 Mincer-Zarnowitz regressions											
1	0.03	0.08	0.17	0.11	0.10	0.06	0.38	0.95	0.97	0.96	0.93
2	0.01	0.07	0.01	0.07	0.03	0.02	0.09	0.85	0.89	0.86	0.85
3	0.00	0.06	0.00	0.01	0.01	0.03	0.07	0.65	0.78	0.77	0.78
4	0.03	0.01	0.00	0.04	0.01	0.04	0.01	0.40	0.62	0.71	0.71
5	0.01	0.00	0.00	0.01	0.00	0.05	0.01	0.19	0.41	0.68	0.65
6	0.01	0.00	0.00	0.00	0.01	0.03	0.00	0.10	0.20	0.65	0.60
7	0.04	0.00	0.00	0.00	0.03	0.02	0.01	0.06	0.06	0.63	0.54
8	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.10	0.00	0.63	0.50

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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 16: 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) RMSE relative to AR(2) benchmark											
1	1.26	1.13	1.02	1.26	1.22•	1.12	1.29●	1.26	1.39●	1.34●	1.29●
2	1.49	1.28	1.06	1.49	1.21	1.24	1.33	1.42	1.58•	1.32●	1.45●
3	1.73	1.56	1.22•	1.68	1.41	1.33	1.60	1.65	1.81	1.40●	1.61•
4	1.94	1.92	1.14	1.96	1.74	1.36	1.96	1.91	2.09	1.48●	1.61•
5	2.24	2.31	1.36	2.18	2.07	1.38	2.37	2.18	2.44	1.55●	1.67
6	2.35	2.45	1.42	2.32	2.42	1.34	2.71	2.48	2.83	1.63•	1.80
7	2.41	2.91	1.54	2.25	2.78	1.22	2.97	2.78	3.31	1.72•	1.88
8	2.41	3.30	1.65	2.10	3.11	1.04	3.27	3.03	3.77	1.85	1.88
(b) p-value Mincer-Zarnowitz regressions											
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.21	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.08	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.03	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(c) R^2 Mincer-Zarnowitz regressions											
1	0.04	0.14	0.24	0.08	0.09	0.07	0.23	0.94	0.94	0.91	0.91
2	0.00	0.07	0.04	0.03	0.08	0.00	0.05	0.77	0.73	0.74	0.81
3	0.01	0.03	0.00	0.02	0.01	0.00	0.00	0.56	0.42	0.56	0.70
4	0.00	0.01	0.02	0.01	0.00	0.00	0.02	0.37	0.16	0.41	0.65
5	0.05	0.00	0.01	0.02	0.03	0.05	0.03	0.23	0.03	0.32	0.60
6	0.00	0.04	0.00	0.01	0.03	0.00	0.01	0.14	0.00	0.26	0.54
7	0.00	0.00	0.00	0.00	0.03	0.05	0.00	0.08	0.01	0.22	0.47
8	0.02	0.02	0.04	0.00	0.11	0.00	0.03	0.06	0.02	0.16	0.46

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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 17: 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) RMSE relative to AR(2) benchmark											
1	0.93	0.90•	0.96	0.89	0.84●	0.81	1.06	0.99	1.01	1.07•	1.00
2	0.94	0.94	0.99	0.88	0.94	0.73	0.95	0.97	1.01	1.03	1.01
3	0.96	0.98	1.04●	0.87	0.95	0.72	0.94●	0.98	1.02	1.01	1.02
4	0.99	1.01	0.99	0.85	0.99	0.67	0.96	0.98	1.03	1.00	1.02
5	0.97	1.02	1.00	0.86	1.00	0.59	0.98	0.98	1.05	0.99	1.03
6	1.00	1.00	1.03	0.90	1.02	0.61	0.99	0.98	1.07	0.98	1.03
7	1.02	0.99	1.04•	0.94	1.03	0.51	1.00	0.98	1.10	0.97	1.02
8	1.02	1.01	1.03	0.96	1.05	0.53	1.02	0.98	1.12	0.96	1.02
(b) p-value Mincer-Zarnowitz regressions											
1	0.37	0.07	0.01	0.31	0.03	0.51	0.81	0.01	0.78	0.16	0.17
2	0.11	0.00	0.00	0.36	0.01	0.76	0.37	0.01	0.62	0.11	0.17
3	0.05	0.00	0.00	0.39	0.00	0.74	0.09	0.00	0.47	0.08	0.16
4	0.02	0.00	0.00	0.40	0.00	0.64	0.05	0.00	0.36	0.06	0.13
5	0.03	0.00	0.00	0.39	0.00	0.65	0.07	0.00	0.26	0.06	0.10
6	0.03	0.00	0.00	0.38	0.00	0.57	0.18	0.00	0.19	0.05	0.09
7	0.02	0.00	0.00	0.22	0.00	0.78	0.27	0.00	0.14	0.03	0.07
8	0.02	0.00	0.00	0.20	0.00	0.60	0.21	0.00	0.11	0.01	0.05
(c) R^2 Mincer-Zarnowitz regressions											
1	0.07	0.17	0.02	0.20	0.09	0.16	0.30	0.96	0.97	0.94	0.93
2	0.00	0.08	0.02	0.15	0.03	0.07	0.07	0.84	0.88	0.83	0.86
3	0.00	0.04	0.01	0.11	0.00	0.04	0.01	0.69	0.76	0.74	0.79
4	0.01	0.01	0.00	0.07	0.01	0.01	0.00	0.52	0.63	0.66	0.72
5	0.00	0.00	0.01	0.04	0.03	0.01	0.00	0.36	0.49	0.60	0.66
6	0.00	0.00	0.00	0.04	0.03	0.00	0.00	0.26	0.38	0.57	0.63
7	0.00	0.00	0.00	0.00	0.03	0.02	0.01	0.19	0.29	0.54	0.59
8	0.00	0.02	0.00	0.00	0.06	0.00	0.00	0.16	0.21	0.51	0.56

Notes: All forecasting models are estimated over a rolling window of 60 quarters, starting in 1992Q4. For each horizon a total of 76 forecasts is computed. 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.