Kiel

# Working Papers

Kiel Institute
for the World Economy

# Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?

Inske Pirschel and Maik H. Wolters

**No. 1925** | November 2015

Kiel Working Paper No. 1925 | November 2015

# Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?*

Inske Pirschel and Maik H. Wolters

Abstract:
We study the forecasting performance of three alternative large data forecasting approaches. These three approaches handle the dimensionality problem evoked by a large dataset by aggregating its informational content, 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 forecasting models, respectively. We use a dataset for Germany that consists of 123 variables in quarterly frequency and find that overall the large Bayesian VAR and the Bayesian factor augmented VAR provide the most precise forecasts for a set of 11 core macroeconomic variables. Both considerably outperform the remaining large scale forecasting models in terms of joint forecasting accuracy as measured by the multivariate MSE. Further, we find that the performance of these two models is very robust to the exact specification of the forecasting model.
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.

# 1 Introduction

While forecasters may wish to use as much information as possible to increase the accuracy of their forecasts, the estimation of models with a large number of different time series causes huge technical difficulties as the number of parameters to be estimated quickly becomes very large and in-sample overfitting occurs or estimation becomes even infeasible. To overcome this *curse of dimensionality* several large scale time-series methods have been proposed. The three most prominent of these approaches are factor models, large Bayesian vector autoregressions and model averaging techniques. All of these three approaches handle the dimensionality problem by aggregating the informational content of the large dataset, yet the aggregation takes place on different levels.[1]

In particular, with factor models (see e.g. Stock and Watson, 2002a,b; Bernanke and Boivin, 2003; Forni et al., 2000, 2005) the aggregation of the informational content of a large dataset into a small number of static or dynamic factors takes place *prior* to the estimation of small scale forecasting models such as e.g. autoregressive distributed lag models, vector autoregressions or Bayesian vector autoregressions. These small forecasting models then include the factor time series rather than all the time series of the large dataset (see Banerjee, 2013, for an overview of the factor model approach and a survey on recent papers using factor models for large dataset problems).

Large Bayesian vector autoregressions (De Mol et al., 2008; Bańbura et al., 2010), on the other hand, can be estimated with a large number of time series by applying shrinkage to aggregate the information contained in the large dataset *during* the estimation process. The degree of shrinkage thereby increases with the number of times series included in the respective model.

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) the aggregation of the informational content of the large dataset takes place *after* the estimation of a large number of small scale forecasting models. Here, the final forecast is computed as a weighted average over the individual forecasts of all the small scale forecasting models.

While De Mol et al. (2008) show that there is a theoretical connection between the factor and the shrinkage approach, it is not clear which method to aggregate the informational content of a large dataset performs best from an empirical perspective. In this paper, we therefore systematically study the performance of all three of these alternative large scale approaches using a dataset for Germany that consists of 123 variables in quarterly frequency.

Previous literature has so far only focused on evaluating the forecasting performance of one or two of these large scale approaches relative to several small benchmark models, to the Federal Reserve's Greenbook projections (for US data) or to each other. For example, Bernanke and Boivin (2003) study the performance of a factor augmented autoregressions and vector autoregression relative to the Greenbook forecasts, Faust and Wright (2009) evaluate static factor models as well as model averaging models relative to a number of benchmark models and

[1] An alternative approach to solving this problem are variable selection methods such as targeted predictors (Bai and Ng, 2008), Bayesian variable selection (Korobilis, 2013) or the LASSO approach (Tibshirani, 1996).

the Greenbook projections, Bańbura et al. (2010) study a large Bayesian vector autoregression and a Bayesian factor augmented vector autoregression and Berg and Henzel (2013) focus on the same models, but study euro area instead of US data and additionally evaluate the different models' density forecasts.[2]

By contrast, our analysis includes all three large scale approches outlined above. It focuses on Germany, the largest economy in the euro area, which is considerably smaller but also much more open than the US or the euro area. For Germany, several authors have investigated the forecasting performance of factor models estimated on large datasets relative to small benchmark models (see e.g. Schumacher and Dreger, 2004; Kholodilin and Siliverstovs, 2006; Schumacher, 2007, 2010, 2011). However, so far no comparison of the forecasting performance of factor models to alternative large scale approaches has been provided. Moreover, most existing empirical forecasting applications for Germany focus almost exclusively on forecasting real GDP (see e.g. Drechsel and Scheufele, 2012a,b) rather than a set of key macroeconomic variables as is the case in this study. To our knowledge, the only two exceptions to this are Müller-Dröge et al. (2014) and Buchen and Wohlrabe (2014) who evaluate the forecasts for a larger set of German key macroeconomic variables as well. However, both papers have a different methodological focus than this paper.

With our comprehensive analysis we provide an assessment of the relative joint and also univariate forecasting performance of the different large scale forecasting methods for GDP growth, CPI and PPI inflation, a short- and a long-term interest rate, the unemployment rate, industrial production, real wages, consumption, investment and the current account balance. We deem these 11 variables of special interest to forecasters and policy makers because they are covered, for example, in the monthly survey of *Consensus Economics* among professional forecasters. Moreover, we test whether the forecasts obtained with the different models are unbiased and check whether the relative performance of the different forecasting models is robust against various alternative model specifications.

Our dataset consists of 123 variables in quarterly frequency covering a sample period from 1978 until 2013. We include indicators from the following categories: composition of GDP and gross value added by sectors, prices, labor market, financial market, industry, construction and surveys. Different variants of the three large scale forecasting models as well as a number of small benchmark models are estimated using a moving window of 15 years of data, while the forecasts obtained by the different models are evaluated from 1994 through 2013. To assess the relative (joint) forecasting performance of the different models we compare (multivariate) mean squared forecast errors, while we compute Mincer-Zarnowitz regressions (see Mincer and Zarnowitz, 1969) to test for forecast bias. All forecasting models are specified according to various information criteria. As a robustness check we also specify the models based on their ex post best forecasting performance and implement forecast pooling over a variety of specifications.

Our results indicate that the large Bayesian vector autoregression and the Bayesian factor augmented vector autoregression deliver forecasts that are more precise than those obtained by

[2]Beyond pure reduced form forecasting models, Wolters (2015) compares the forecasting accuracy of a large Bayesian vector autoregression to Dynamic Stochastic General Equilibrium (DSGE) models and the Fed's Greenbook projections.

a univariate autoregressive benchmark or the remaining large scale forecasting models. This holds for both, measures for the joint forecasting performance for the set of 11 variables as well as univariate performance measures for the individual series. We find that in contrast to the remaining factor approaches and the model averaging approaches both models can efficiently exploit the correlation structure between the series of the large dataset to provide relatively accurate forecasts, even for longer forecasting horizons.

With respect to the robustness of the relative forecasting performance of the different models our findings indicate that the forecasting performance of the large Bayesian vector autoregression and the Bayesian factor augmented vector autoregression is very robust to the specific model specification, i.e. the number of lags or factors and the degree of shrinkage. By contrast, the dynamic factor model outperforms all other forecasting models by far if one chooses the ex-post optimal specification. However, in the quasi real-time exercise, where the number of lags and factors is chosen based on information criteria, on past forecasting accuracy or where the forecasts are obtained by pooling over a large set models with different specifications, we find that this performance is unattainable.

Finally, our results indicate that overall the gains in forecasting accuracy obtained by the large scale approaches relative to an autoregressive benchmark are only modest for most variables considered and are in many cases statistically insignificant. We also find that using a large amount of data would not have helped in forecasting the great slump of German GDP growth in 2008 and that a small forecasting model that only includes the ifo business climate index, which is often cited by professional forecasters as *the* single most important predictor for German GDP growth,[3] clearly dominates even the best large scale approaches in terms of short-term GDP growth forecasting performance.

The moderate gains of the large scale approaches can be explained with the extremely low persistence of some of the time series. Moreover, many of the time series seem to be characterized by common components which implies that parsimonious univariate models are often sufficient to capture the most important 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).

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 at the same time. This might be an advantage when it comes to the interpretation of forecasts.

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 (joint) forecasting performance of the different models and check for robustness against model misspecification. Finally, in section 6 we conclude.

[3] The ifo business climate index is based on a monthly survey among about 7000 firms which report their assessments of the current business situation and their expectations for the next six months. From these two assessments the overall ifo index is calculated. The out-of-sample predictive ability of the ifo index for German GDP has been widely studied, see for example Dreger and Schumacher (2005), Kholodilin and Siliverstovs (2006), Abberger (2007), Drechsel and Scheufele (2012b) or Henzel and Rast (2013).

## 2 Forecasting Models

In the following, we provide a brief overview of the different forecasting models. Let $\{y_{i,t}\}_{i=1}^{n}$ denote the set of variables to be forecast in log-levels and $\{x_{j,t}\}_{j=1}^{m}$ the set of possible predictors in log-levels. Variables expressed in rates such as the unemployment rate or interest rates are included in $\{y_{i,t}\}_{i=1}^{n}$ and $\{x_{j,t}\}_{j=1}^{m}$ in levels rather than log-levels. 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. To avoid overly complicated notation, variables expressed in rates are included in levels in the respective $\Delta$ terms as well. 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. 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. For forecasting models that include lags of the dependent variable the number of lags $p$ included in the estimation of each model is obtained via the Bayesian information criterion unless otherwise stated.

### 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. Following Bańbura et al. (2010) we include the variables in log-levels rather than growth rates to not lose information that might possibly be contained in the trends. $c$ is a ($k$x1) vector of constants, $A_1, ..., A_p$ are ($k$x$k$)-dimensional parameter matrices and $\epsilon_t$ is a ($k$x1) vector of independently identically distributed white noise error terms with zero mean and covariance matrix $\Psi$.

We use Bayesian techniques to estimate the large 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 implement a prior that shrinks the parameters of the VAR. This allows for the aggregation of the information contained in the large dataset *during* the estimation process. The degree of shrinkage thereby increases with the size of the cross-section, thus allowing the estimation of a model where the number of parameters exceeds the number of observations by far.

We implement the Bayesian shrinkage approach by using a version of the Normal inverse Wishart prior (see e.g. Kadiyala and Karlsson, 1997) that retains the main principles of the widely used Minnesota prior (Litterman, 1986). According to this prior specification each equation of the VAR is centered around a random walk with drift or an autoregressive process, respectively. In contrast to Bańbura et al. (2010), we do not set $\delta_i$, the prior coefficient means for the first lag of each variable, equal to zero for stationary variables. Instead, we run a univariate autoregression of order $p$ for each of the $k$ elements in $Z_{i,t}$ and set $\delta_i$ equal to the sum of the therewith obtained coefficient estimates defined as $\mu_i = \sum_{\ell=1}^{p} \beta_\ell$ if $\mu_i < 1$. For $\mu_i \geq 1$ we set $\delta_i = 1$. This approach allows us to capture the different degrees of persistence in the dataset.

The shrinkage of the VAR coefficients towards the prior is achieved through the hyperparameter $\lambda$ which enters the prior variance of each coefficent. Bańbura et al. (2010) suggest to set the tightness of the prior, so that the LBVAR achieves the same in-sample fit as an unrestricted small VAR without shrinkage.[4] We slightly depart from this approach and set $\lambda$ such that the LBVAR achieves the same in-sample fit as a small BVAR containing GDP, prices, the unemployment rate and a short-term interest rate, because the respective unrestricted VAR seems to be severely overparameterized.

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 in the estimation (see section 5.3). 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)

Assume that $\Delta X^*_{i,t}$, the standardized set of potential predictors for each variable of interest, can be represented by two components which are mutually orthogonal to each other and unobservable. These are the common component $\chi_{i,t}$ and the idiosyncratic component $\xi_{i,t}$, so that we have $\Delta X^*_{i,t} = \chi_{i,t} + \xi_{i,t}$.

The basic idea of factor models is that the information contained in the common component $\chi_{i,t}$ can be aggregated into a vector of factors $F_{i,t}$ of dimension $\kappa \leq (k-1)$ which are able to explain most of the variance of the predictor matrix $\Delta X^*_{i,t}$. With these factors the dimension of a large dataset can thus be reduced *prior* to the estimation of the forecasting model.

In general the common component relates to the factors as $\chi_{i,t} = \sum_{l=0}^{s} \eta_l F_{i,t-l}$. 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 $\Delta X^*_{i,t}$ we extract the ($r$x1)-dimensional vector of factors $F_{i,t} = (f^1_{i,t}, ..., f^r_{i,t})'$ via static principal component analysis. Following Stock and Watson (2002a) we use these static factors to estimate a factor augmented direct autoregression (FAAR).[5]

Moreover, we implement a factor augmented vector autoregression as proposed by Bernanke et al. (2005) which allows for a more dynamic structure. Following Faust and Wright (2009) we include the variable to be predicted in log-levels and the factors extracted from the set of predictors in the estimation.[6]

We estimate the factor augmented vector autoregression via ordinary least squares (FAVAR)

[4] Of course, this approach is merely an ad-hoc rule of thumb. Alternatively, $\lambda$ could also be chosen to maximize the out-of sample forecasting performance over a pre-sample as for example in Litterman (1986). Giannone et al. (2012) suggest a more sophisticated hierarchical approach to specifying $\lambda$ which relies on maximizing the marginal likelihood, i.e. the density of the data conditional on $\lambda$ after integrating out the uncertainty about the parameters of the VAR. However, since we find that the forecasting performance of the large BVAR is very robust to the exact specification of $\lambda$, we do stick to the rule of thumb.

[5] According to common practice, we chose the direct version of the autoregressive model because the iterated model variant would require the specification of a subsidiary model for the factors $F_{j,t}$ in order to compute forecasts for horizons $h > 1$.

[6] 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 considerably worse, so that we do not include this model in the main results.

as well as with Bayesian techniques (BFAVAR), however the FAVAR performs very poorly so we do not report the results for this model. For the BFAVAR, 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 first lag of the factors $\delta = 0$ to account for the fact that the factors have been extracted from the standardized predictor matrix $\Delta X^*_{i,t}$. Second, we set the hyperparameter $\lambda = 0.2$. For the determination of the optimal number of factors $r$ we use the information criterion $IC_{p2}$ 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, 2005). This implies extracting the ($q$x1)-dimensional vector of dynamic factors $\tilde{F}_{i,t}$ from the standardized set of predictors $\Delta X^*_{i,t}$ via dynamic principal component analysis in the frequency domain. Defining $\tilde{F}^*_{i,t} = (\tilde{F}'_{i,t}, \tilde{F}'_{i,t-1}, ..., \tilde{F}'_{i,t-s})'$ 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 $\chi_{i,t} = \eta \tilde{F}^*_{i,t}$.

The factors $\tilde{F}^*_{i,t}$ are used to augment a direct autoregression, analogously to the FAAR outlined above. For the determination 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 $n$ variables of interest $\Delta y_{i,t}$ we set up $(k-1)$ direct autoregressive distributed lag models $\Delta y_{i,t} = \rho_0 + \rho_1 \Delta y_{i,t-h} + ... + \rho_p \Delta y_{i,t-h+1-p} + \beta_j \Delta x_{j,t-h} + \epsilon_{j,t}$, where $\Delta x_{j,t-h}$ is an element of the $(k-1)$-dimensional set of potential predictors $\Delta X_{i,t}$.

The general idea of model averaging is to compute a forecast $\Delta y^j_{i,t+h}$ with each of the $(k-1)$ models and aggregate the model-specific forecasts *afterwards* into one final forecast, i.e. $\Delta y_{i,t+h} = \sum_{j=1}^{(k-1)} \omega_j \Delta y^j_{i,t+h}$, where $\omega_j$ denotes the weight given to the model-specific forecast $\Delta y^j_{i,t+h}$.

According to the specification of $\omega_j$ we distinguish two model averaging approaches. The first approach is Equal Weighted Averaging (EWA) as in Stock and Watson (2003, 2004), where the $(k-1)$ models are estimated via OLS and $\omega_j = \omega = \frac{1}{(k-1)}$.

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. $\omega_j = P(M_j)$.[7]

## 2.4 Benchmark Model (AR)

In order to evaluate the relative forecasting performance of the three large scale approaches described above we implement a univariate autoregression (AR) $\Delta y_{i,t} = c + \sum_{j=1}^{p} \rho_j \Delta y_{i,t-j} + \epsilon_t$ for each of the variables to be forecast as benchmark.

[7] 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.

# 3 Data

Our dataset builds on the one used in Schumacher (2007) which we have slightly modified and updated to cover a sample from 1978Q1 to 2013Q3. Overall, our dataset consists of 123 macroeconomic variables in quarterly frequency. Series that are available at a higher frequency, e.g. monthly, are converted into quarterly frequency by computing the average over the respective quarter. The data 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. A detailed list of the different series can be found in Appendix A.

Most of the data is obtained via Thomson Reuters Datastream, while the remaining data is directly obtained from the German Federal Statistical Office. We do not account for data revisions in our quasi real-time forecasting exercise, but use the most recent vintage of the data available in December 2013. 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.

# 4 Forecasting Approach

We estimate the various forecasting models on a moving window consisting of 60 observations to account for possible structural breaks in the estimation sample. For the majority of forecasting models, the forecasts are computed by iterating the forecasting models forward, while for the FAAR, DFM and the two model averaging approaches direct forecasts are computed.

The evaluation sample for our pseudo out-of-sample forecasting exercise, denoted by $T = T_0 + 1, ..., T_1$, ranges from 1994Q4 until 2013Q3, thus it contains 76 forecasts for each horizon. Forecast errors are computed as $e_{i,T|T-h} = \Delta y^r_{i,T} - \Delta y^f_{i,T|T-h}$, where $\Delta y^r_{i,T}$ denotes the realized quarter-on-quarter growth rate of variable $i$ in period $T$ and $\Delta y^f_{i,T|T-h}$ denotes the quarter-on-quarter period $T$ growth rate forecast of variable $i$ computed $h$ quarters earlier.

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 (see Mincer and Zarnowitz, 1969) $\Delta y^r_{i,T} = \alpha_{i,h} + \beta_{i,h} \Delta y^f_{i,T|T-h} + \epsilon_{i,T|T-h}$ and conduct F-tests of the joint null hypothesis $\hat{\alpha}_{i,h} = 0$ and $\hat{\beta}_{i,h} = 1$ to check whether the forecasts are unbiased and efficient. This allows us to assess the absolute forecasting accuracy of each model.

Secondly, we compute and analyze (multivariate) mean squared forecasting errors (MSE) to evaluate the relative (joint) predictive ability of the different forecasting models. We report the absolute MSE for the AR forecast which we use as a benchmark, while for the remaining models we report the MSEs relative to this benchmark. Thus a relative MSE smaller than 1 indicates that the forecasting performance of a specific model is more precise than that of the AR benchmark and vice-versa.

To assess the statistical significance of the forecasting performance of the different models for each individual variable relative to the AR benchmark we implement the 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.

Finally, the multivariate root mean squared forecast error, as proposed by Christoffersen and Diebold (1998), is computed as

$$multMSE_h = \frac{1}{T_1 - T_0 - 1} \sum\nolimits_{T=T_0+1}^{T_1} e'_{T|T-h} W e_{T|T-h}, \tag{1}$$

where the (1x$n$)-dimensional vector $e_{T|T-h}$ contains the forecast errors $e_{i,T|T-h}$ for all $n$ variables of interest and $W$ is an ($n$x$n$)-dimensional diagonal weighting matrix.

We follow Carriero et al. (2011) as well as Buchen and Wohlrabe (2014) and specify $W = W^D$ as diagonal matrix with entries being equal to the inverse of the variances of the variables to be forecast. Müller-Dröge et al. (2014) propose to specify $W = W^C$ as the inverse of the sample variance-covariance matrix and we consider this alternative as well. Both versions of the multivariate MSE aim at assessing the joint predictive ability of the different forecasting models, i.e. their suitability to simultaneously forecast a larger set of variables. In the first version the measure has the advantage that it accounts for the fact that variables with a large variance are generally harder to forecast by attributing them a smaller weight. In addition to that, the second version of the measure compensates for possible correlation of the different series.[8]

As for the univariate MSE we report the absolute multivariate MSE for the AR model, while for all remaining models we compute the multivariate MSE relative to this benchmark.

# 5 Results

In this section we report the results of our forecasting exercise. We first focus on the joint forecasting performance of each large scale approach for our set of 11 German key variables. Afterwards, we extend the analysis to the performance for the individual variables, with an emphasis on GDP growth. Finally, we check the robustness of our results.

## 5.1 Forecasting 11 German Key Macroeconomic Variables Jointly

Figure 1 displays the 11 variables that we consider. It can be seen very clearly that there is considerable variation in the degree of persistence of the different variables. For example, German GDP growth shows extremely little persistence and can thus be expected to be very hard to predict. A comparison of the autocorrelation functions of US and German GDP growth for a sample covering 1978-2013 shows that there is significant autocorrelation of up to two lags

[8] The underlying idea is to account for the linear dependence between the different variables that might simultaneously drive their MSEs and thus inflate the measure of joint predictive ability. In principle, this is comparable to the approach of computing the variance of the sum of several random variables where a correction term accounting for the covariance of the pairs of variables is needed as well. Note however, that since we use the inverse of the covariance matrix as correction, the multivariate MSE decreases for positive correlation and increases for negative correlation between the pairs of variables in the dataset.

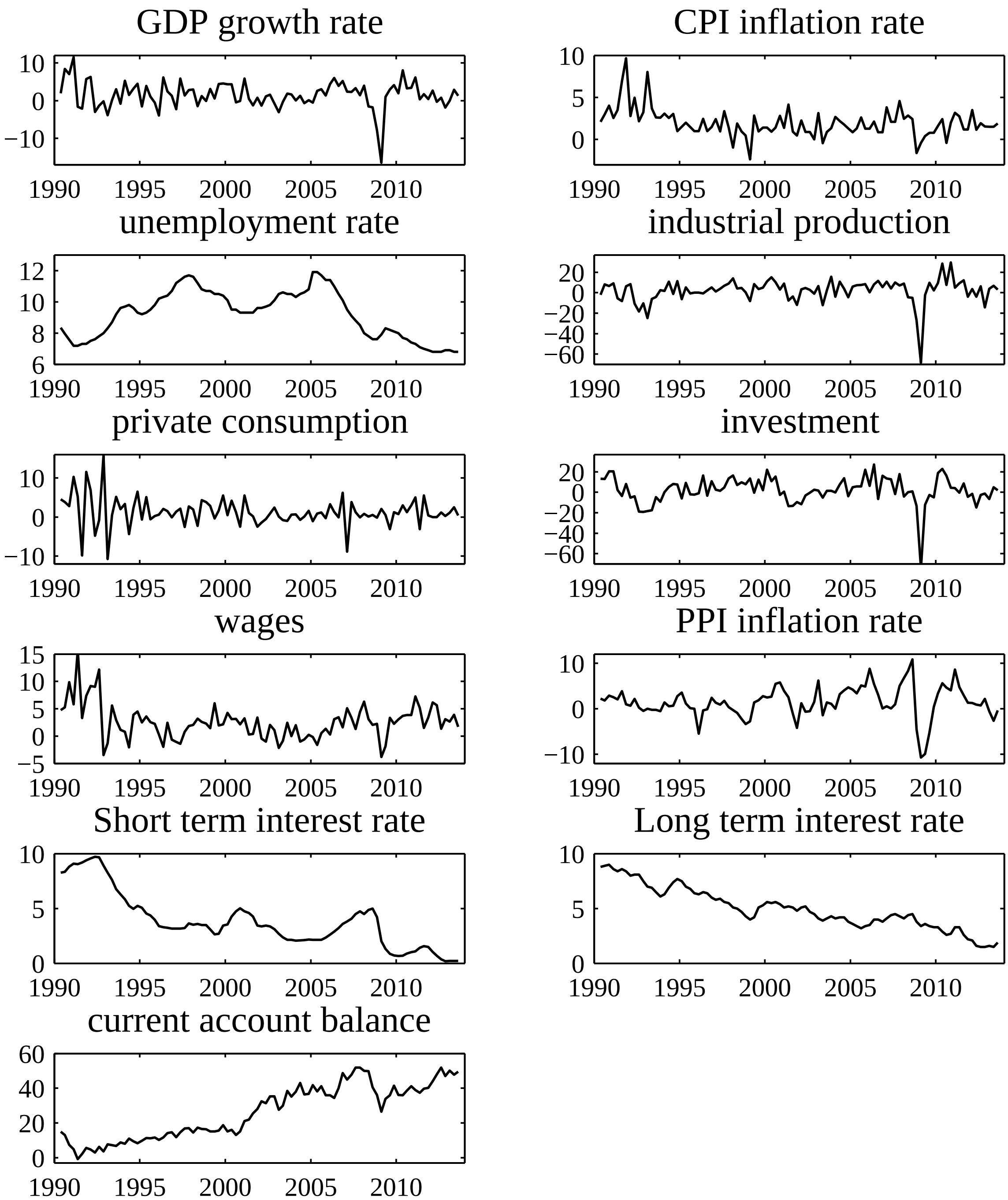


Figure 1: German key macro variables

Notes: The graph shows the 11 German key macroeconomic variables that we consider from 1990 until 2013. For all variables, except those expressed in rates, annualized quarter-on-quarter growth rates are shown. Data sources are listed in Appendix A and Appendix B contains an exact definition of the different variables.

for US GDP growth, while there is no significant autocorrelation at all for German GDP growth. The persistence of industrial production, investment and consumption growth is comparable to that of German GDP growth. Thus, we can expect the different forecasting models to have similar problems in predicting these variables. By contrast, CPI inflation is more persistent than GDP growth, but still shows many spikes, which will presumably be hard to predict as well. The persistence of the unemployment rate series is very high, similar to that of the short- and long-term interest rate and the current account balance. The German unemployment rate does not show a clear overall trend, but instead increases until 1998, decreases until 2001, increases again until 2005 and falls from there until the end of the sample. Predicting these trend changes might pose another difficulty for most forecasting models.

In table 1, we display both versions of the multivariate MSEs of the different large scale approaches relative to the AR benchmark for horizons $h = 1, 4, 8$ as well as the absolute multivariate MSE for the AR. Both measures indicate that the BFAVAR and to a slightly lesser extent the LBVAR provide the most accurate forecasts for all the variables over all forecasting horizons. For short horizons also the remaining large scale approaches are able to improve upon the AR benchmark, though to a different degree. While the FAAR and the DFM perform almost as good as the BFAVAR and the LBVAR, the model averaging techniques do worse. For $h = 8$, the relative performance of the large scale approaches deteriorates considerably. The BFAVAR is the only model that can clearly outperform the AR benchmark according to both multivariate measures considered here.

Table 1: Forecasting a Set of 11 German Key Varaibles

| (a) multivariate MSE, $W = W^D$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 8.03 | 0.84 | 0.85 | 0.80 | 0.87 | 0.95 | 0.94 |
| 4 | 11.16 | 0.86 | 1.02 | 0.82 | 1.00 | 0.89 | 0.91 |
| 8 | 13.52 | 0.90 | 1.35 | 0.86 | 1.28 | 1.09 | 1.10 |
| (b) multivariate MSE, $W = W^C$ | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 11.61 | 0.79 | 0.83 | 0.76 | 0.82 | 0.91 | 0.92 |
| 4 | 15.81 | 0.86 | 1.01 | 0.79 | 1.06 | 0.87 | 0.96 |
| 8 | 18.42 | 1.00 | 1.51 | 0.86 | 1.46 | 1.01 | 1.12 |

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1994Q4 until 2013Q3, thus for each horizon a total of 76 forecasts is computed. The second column shows the absolute multivariate MSEs for the AR benchmark model, while all other MSEs are computed relative to this benchmark. The two measures differ with respect to the weighting matrix $W$ which is a diagonal matrix with the inverse of the series variance as entries ($W^D$ upper panel) and the inverse of the sample covariance matrix ($W^C$ lower panel).

The entries in the lower part of table 1 indicate to what extend the different forecasting models are able to account for possible correlation of the different series. The absolute multivariate MSE of the AR benchmark is considerably higher for the second version of the measure for all horizons indicating that there is negative correlation in the data that could be useful for forecasting. Interestingly, the relative performance of the large scale approaches generally improves for short horizons and deteriorates for longer horizons when comparing the upper and the lower part of the table. This indicates that for short horizons the models are able to make use of the correlation structure in the dataset to provide better forecasts than the AR benchmark. For large horizons, however, this seems no longer to be the case. Especially the factor models, FAAR and DFM, display a very poor longer term joint forecasting performance. A notable exception is the BFAVAR which performs equally well under both versions of the multivariate MSE for all horizons. Apparently, the combination of aggregation information of the large dataset into factors and shrinkage enables the model to efficiently use all the information contained in the dataset, even for longer forecasting horizons.

Overall, our results indicate that for short forecast horizons it does not seem to make a very big difference for the joint forecasting performance of our large scale approaches, whether the information of the large dataset is aggregated before or during the estimation process of the forecasting models, as the factor models and the shrinkage approaches perform similarly well. Aggregation after the estimation process (model averaging approach), however, yields somewhat less precise short-horizon forecasts. For obtaining accurate forecasts for longer horizons using a shrinkage approach seems to be essential to extract the relevant information on the longer-run dynamics of the different variables as evidenced by the very good performance of the BFAVAR and, to a slightly lesser extent, the LBVAR for longer forecast horizons.

### 5.2 Forecasting Performance for the Individual Variables

In table 2, panels (a) - (k), we display the univariate MSEs of the different models for the 11 key variables for horizons $h = 1, 4, 8$.[9] Table entries in bold indicate that the null hypothesis of unbiasedness based on the F-test for the coefficients in the Mincer-Zarnowitz regression cannot be rejected at the 5 % level. The symbols ●, •, •, indicate that the relative MSE is significantly different from one at the 1, 5, or 10% level, respectively.

As can be seen very clearly from the entries in table 2, the BFAVAR and the LBVAR are the best performing models in most cases. However, the size of the gains in accuracy over the AR benchmark as well as the absolute forecasting performance of the different models apparently depend heavily on the specific variable and the respective forecasting horizon.

For *GDP growth* (panel (a) of table 2), the absolute MSEs of the AR benchmark are quite large and flat over the different forecast horizons which is in line with what can be expected for forecasts of a time series with low persistence, (see Del Negro and Schorfheide, 2013, for a detailed exposition). Moreover, the entries in table 2, panel (a) reveal that the gains in forecasting accuracy for German GDP growth obtained by the three large scale approaches are at best moderate and insignificant, while the differences in the relative MSEs between the various forecasting models are rather small. Among the three large scale approaches, the BFAVAR and (to a slightly lesser extend) the LBVAR yield the most accurate forecasts, though for $h = 8$ EWA performs best. In the short-run, the gains of the BFAVAR and the LBVAR over the AR benchmark amount to more than 10%. Yet, for longer horizons, there is almost no improvement upon the AR, which confirms the results of Schumacher (2007, 2010) for factor models and Kholodilin and Siliverstovs (2006) for various alternative leading indicators. Thus,

[9]Figures showing the forecasts are contained in Appendix C.

Table 2: univariate MSEs

(a) GDP growth

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 12.09 | **0.88** | 0.95 | 0.86 | 0.95 | **0.98** | 1.03 |
| 4 | 12.64 | 1.00 | 1.27● | 0.98 | 1.09 | 1.01 | 1.07 |
| 8 | **12.12** | 1.04 | 1.42● | **0.98** | 1.22 | **0.94** | 1.10 |

(b) CPI Inflation Rate

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 1.94 | 0.78● | 0.89 | 0.80• | 0.92 | 0.93● | 0.87● |
| 4 | 1.86 | 0.95 | 1.18 | 0.85● | 1.21 | 0.93 | 0.92 |
| 8 | 2.07 | 0.89 | 1.33 | 0.82● | 1.33 | 0.92 | 0.93 |

(c) Unemployment Rate

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | **0.06** | **0.98** | **0.98** | **0.94** | **0.99** | **0.95●** | **0.95•** |
| 4 | **0.76** | **0.85** | **1.14** | **0.81** | **1.27** | **1.06** | **1.07** |
| 8 | **2.17** | 0.95 | **1.18** | **0.83** | **1.37** | **1.67** | 1.79 |

(d) Industrial Production

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 173.25 | **0.71** | 0.64• | 0.76• | 0.72 | 0.95 | 0.97 |
| 4 | 224.63 | 0.66 | **0.59** | **0.64** | **0.57** | **0.66** | **0.63** |
| 8 | **140.27** | **1.02** | 1.06 | **1.00** | **0.97** | **1.01** | 1.07 |

(e) Private Consumption

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 7.06 | 0.86 | 1.04 | 0.91 | 1.05 | 0.99 | 0.97 |
| 4 | 7.37 | 0.84● | 1.21 | 0.85● | 1.20• | 0.96 | 0.94 |
| 8 | 6.78 | 0.93 | 2.27• | 0.95 | 2.02• | 1.04 | 1.03 |

(f) Machinery and Equipment Investment

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 167.63 | **0.86** | 0.86 | **0.76** | 0.93 | 0.95 | 0.97 |
| 4 | 221.91 | 0.84 | 0.85 | **0.80** | 0.85 | **0.80** | **0.76** |
| 8 | 196.33 | 0.95 | 1.04 | **0.89** | 1.03 | **0.95** | 1.16 |

(g) Wages

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 5.99 | **0.76●** | 0.71 | **0.56●** | 0.64• | **0.91•** | 0.83● |
| 4 | 7.19 | 0.87• | 1.16 | 0.84• | 1.10 | 0.88● | 0.94 |
| 8 | 8.75 | 0.85 | 1.39• | 0.86● | 1.33• | 1.10 | 1.06 |

(h) PPI Inflation Rate

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | **9.22** | **1.20** | **0.82** | **1.03** | **0.85** | **0.91•** | **0.96** |
| 4 | 15.79 | 1.02 | 0.98 | **0.96** | 1.04 | **0.94** | **0.96** |
| 8 | **14.13** | **1.05** | 1.21 | **1.02** | **1.02** | 1.10● | 1.22● |

(i) Short Term Interest Rate

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 0.14 | **0.83•** | **0.88** | **0.73•** | **0.92** | **0.91** | **0.94** |
| 4 | 1.64 | **0.77** | **0.94** | **0.72●** | 0.98 | 1.00 | 1.07 |
| 8 | 4.58 | 0.77 | 1.21 | 0.72● | 1.10 | 1.18 | 0.97 |

(j) Long Term Interest Rate

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | 0.10 | **0.90** | 1.12 | 0.96 | 1.11 | 0.94 | 0.68 |
| 4 | 0.84 | **0.73•** | 1.02 | 0.74● | 1.01 | 0.93 | 0.98● |
| 8 | 1.94 | **0.62•** | 1.49 | 0.63● | 1.46 | 1.07 | 0.94 |

(k) Current Account

| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
|---|---|---|---|---|---|---|---|
| 1 | **15.31** | 0.98 | **1.15** | **1.03** | **1.12** | **0.99** | **1.00** |
| 4 | 60.89 | 1.01 | **1.22** | **0.99** | **1.11** | **0.95** | **1.07** |
| 8 | 114.59 | 1.05 | 1.33 | 0.96 | 1.39 | 1.09 | 1.25 |

Notes: See notes on table 1, first part. The symbols ●, •, ∙, indicate that the relative MSE 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.

adding more information by using a large dataset for the forecasting process of German GDP growth apparently only leads to marginal improvements in forecasting accuracy over the AR benchmark. The results of the Minzer-Zarnowitz regressions reveal that none of the forecasting models is able to provide unbiased forecasts for all forecasting horizons. With the exception of the LBVAR for $h = 1$, the AR and the BFAVAR for $h = 8$ and the EWA for $h = 1, 8$, for the remaining models, the estimated constant $\hat{\alpha}_{i,h}$ is larger than zero, but the estimate of the slope parameter and $\hat{\beta}_{i,h}$ is smaller than one (and in some cases even negative). This indicates that the forecasts systematically predict less variation than the GDP growth series actually shows.

For the German *CPI inflation rate* (panel (b) in table 2), the absolute MSEs for the AR model are much smaller than those for GDP growth. Still the persistence of quarterly CPI inflation is quite low and thus the MSE does not increase much with the forecast horizon $h$. In terms of relative forecasting performance for the CPI inflation rate, only the BFAVAR significantly outperforms the AR benchmark over all forecasting horizons with gains in accuracy ranging between 15% to 20%. For $h = 1$ also the LBVAR and the two model averaging approaches significantly beat the AR benchmark. However, none of the forecasting models is able to yield unbiased forecasts. The estimated constant in the Mincer-Zarnowitz regressions for all models is larger than zero, while the slope parameter is smaller than one (smaller than zero in most cases). This indicates that the forecasts are systematically larger than the actual data which may be attributed to the higher trend inflation in the first part of the estimation sample compared to the evaluation sample. While CPI inflation is more persistent than GDP growth, the informational content of the CPI forecasts obtained by all models is even smaller. The $R^2$ from Mincer-Zarnowitz regressions (not shown in the table) never exceeds 5%.

In contrast to that, for the German *unemployment rate* (panel (c) in table 2) the explanatory power of all forecasts is extremely high, especially for short forecasting horizons. For $h = 1$ the $R^2$ from Mincer-Zarnowitz regressions (not shown in the table) for all forecasting models amounts to 95% or more, while for $h = 4$ it still ranges between 66% and 76%. This must certainly be attributed to the high persistence in the German unemployment rate series, which is also reflected in the small absolute MSE for the AR which increases with the forecasting horizon. Due to the various trend changes in the German unemployment rate series, no model systematically over- or underestimates the unemployment rate. With a few exceptions for $h = 8$, all forecasts are unbiased. However, except for the two model averaging approaches for $h = 1$, no model can significantly outperform the AR benchmark for the prediction of the German unemployment rate.

With some exceptions this also holds for *private consumption* and the *PPI inflation rate* (panel (e) and (h) in table 2). However, while no model is able to yield unbiased forecast for consumption, the forecasts for PPI inflation are unbiased in many cases.

A similar result regarding the relative forecasting performance of the different models can be observed for the German *current account balance* (panel (k) in table 2) and German *machinery and equipment investment* (panel (f) in table 2). We find that for these variables none of the large scale approaches considered can significantly improve upon the AR benchmark which is surprising given the different degrees of persistence of the series.

By contrast, for the *short- and long-term interest rates* (panel (i) and (j) in table 2) as well as for *wages* (panel (g) in table 2) the best performing large scale model, the BFAVAR, almost always significantly outperforms the AR benchmark with sizeable gains in accuracy (45% for wages for $h = 1$).

For *industrial production*, there are big and significant gains in accuracy for the BFAVAR and the FAAR, but only for $h = 1$. For higher horizons the large scale approcches can not significantly improve upon the AR benchmark, though in most cases the forecasts are unbiased.

To sum up, we find that over all variables considered the best performing large scale approaches, namely the LBVAR and the BFAVAR, can clearly improve upon the AR benchmark, especially in the short run. However, the size of the gains in accuracy for each individual variable is highly heterogenous. Moreover, the BFAVAR, the LBVAR and the EWA approach provide genereally more often unbiased forecasts compared to the AR benchmark, the two factor models and the BMA.

**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.

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 during the Great Recession. As pointed out before, the ifo index is a leading indicator and often referenced to as *the* most important benchmark when forecasting German GDP growth (see also Dreger and Schumacher, 2005; Kholodilin and Siliverstovs, 2006; Abberger, 2007; Drechsel and Scheufele, 2012b; Henzel and Rast, 2013). We use the ifo business climate 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): $\Delta y_t = \alpha_h + \beta_h \text{ifo}_{t-h} + \epsilon_{t,h}$.

Figure 2 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, but indicate a relatively quick recovery instead. The only notable exceptions are the one quarter ahead forecasts based on the ifo expectation index and those obtained with the BFAVAR computed in 2008Q4. As business expectations in Germany already dropped largely in 2008Q3, the ifo expectation index predicts a negative GDP growth rate of -3.17% for 2009Q1.[10] The BFAVAR GDP growth forecast is even slightly more pessimistic and amounts to -4.40%. Still, none of the forecasting models is able to predict the turning point of the Great Recession in 2009Q1. Moreover, once the turning point is reached, the models also considerably underpredict

[10] However, by construction this model can hardly predict a further deepening of the recession. Since 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 the further deepening of the recession.

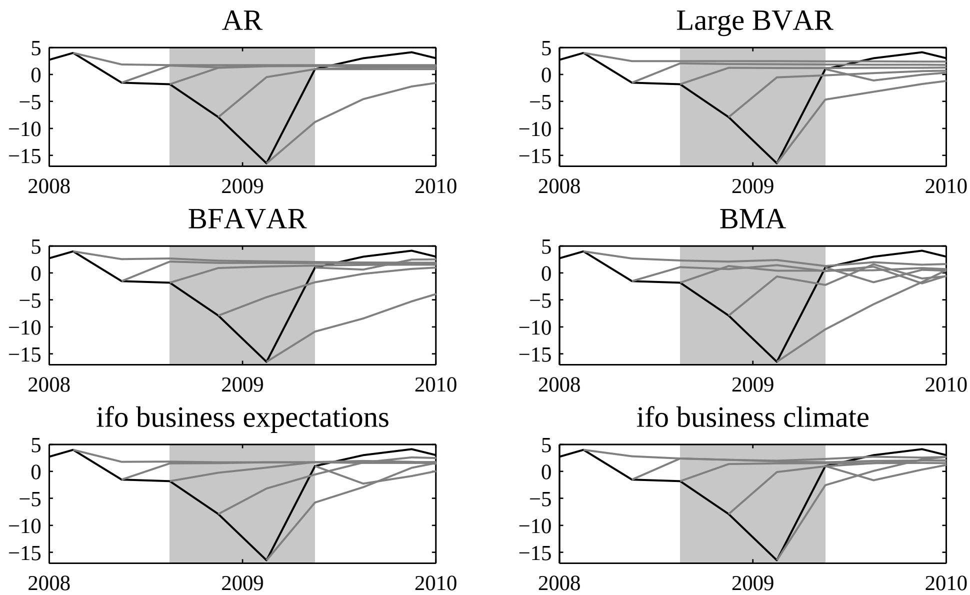


Figure 2: Great Recession GDP growth forecasts

the speed of the recovery for the following quarters.

Given that the performance of the different forecasting models during the Great Recession was more or less equally disappointing, we would not expect that the Great Recession period strongly drives the results reported in table 2, panel (a). Surprisingly, the entries in table 3 which display the MSEs of the different forecasting models relative to the AR benchmark for the pre-Great Recession subsample ranging from 1994Q1 until 2007Q4 indicate that this is not the case. Especially for horizons $h = 4, 8$ the relative performance of the LBVAR and the BFAVAR improves considerably when the Great Recession is excluded from the evaluation sample. In this case both models are able to significantly reduce the relative MSE by approximately 10%. Moreover, the null hypothesis of unbiasedness based on the F-test for the two coefficients in the Mincer-Zarnowitz regression can no longer be rejected for all forecasts obtained by the LBVAR and the BFAVAR.

Table 3: Forecasting German GDP growth, excluding the Great Recession of 2008/2009

| univariate MSE for GDP growth (column 2: absolute; others: relative to AR) | | | | | | | |
|---|---|---|---|---|---|---|---|
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | **6.64** | **0.86** | 0.97 | **0.91** | 0.96 | **0.91** | **0.95** |
| 4 | 7.29 | **0.87**● | 1.36• | **0.88** | 1.10 | **0.92** | 1.03 |
| 8 | 7.07 | **0.90**● | 1.42 | **0.91**• | 1.54• | 0.99 | 1.21 |

Notes: See notes on table 2. In this table the evaluation sample has been adjusted to cover the period from 1994Q4 until 2007Q4.

The ifo business climate index, which performs slightly better than the expectations based index for $h = 1$ (the relative MSEs are equal to 0.84 and 0.94, respectively), yields the most accurate short-run predictions for German GDP growth when we exclude the Great Recession from the evaluation sample. However, the gains upon the best performing large scale approaches are only very small. For $h = 8$ the LBVAR and the BFAVAR clearly beat this important benchmark for the prediction of German GDP growth (the relative MSEs are equal to 1.00 for the ifo business climate and 0.98 for the ifo expectations index, respectively).

### 5.3 Robustness with Respect to Alternative Model Specifications

Next, we want to check the robustness of our results reported thus far with respect to alternative specifications of the different forecasting models. Therefore, we repeat the forecasting exercise of the previous section with an *optimized* specification of each model that is obtained by computing a variety of different specifications for each model and choosing the one that yields ex post the best forecasting performance. For an indication of how robust the model's forecasting performance is against various alternative model specifications, we then check which of the forecasting models yield similarly accurate forecasts with the information criteria based specification and with the ex post optimized specification.

Table 4 shows the ranges of the various parameters of the different forecasting models that we consider for this exercise. For example, for the FAAR the number of static factors $r$ as well

Table 4: Parameter range to determine ex post optimized specification of forecasting models

| parameter | range | forecasting model |
|---|---|---|
| number of lags $p$ | 1, 2, ..., 4 | all forecasting models |
| degree of shrinkage $\lambda$ | 0.01, 0.02, ..., 0.1 | LBVAR |
| degree of shrinkage $\phi$ | 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 |

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 forecasts for each possible combination of these two parameters. We then choose the specification with the ex post highest forecasting accuracy as the optimized specification for the FAAR model.

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 MSE. By contrast, with PBC we choose the specification for each model that minimizes the MSE over the whole evaluation sample for each horizon.

In table 5 we report the results of this exercise for horizons $h = 1$, 4 and 8. Specifically, we display the absolute multivariate MSEs (first version, $W = W^D$) for all 11 key variables for each of the different forecasting models obtained when specified according to the various information criteria (IC) as well as under PBC and PBTV.

Comparing the absolute MSEs reported in table 5 panel (b) and (c) to those reported in

Table 5: Absolute multivariate MSEs with IC, PBC and PBTV

| (a) IC: Information criteria based model selection (quasi real-time forecasting) | | | | | | | |
|---|---|---|---|---|---|---|---|
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 8.03 | 6.74 | 6.82 | 6.41 | 7.01 | 7.59 | 7.51 |
| 4 | 11.16 | 9.60 | 11.39 | 9.20 | 11.18 | 9.92 | 10.10 |
| 8 | 13.52 | 12.14 | 18.26 | 11.62 | 17.32 | 14.76 | 14.89 |
| (b) PBC: Performance based model selection, constant model | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 7.39 | 6.06 | 6.16 | 5.86 | 6.03 | 7.01 | 6.74 |
| 4 | 9.71 | 9.12 | 9.70 | 8.83 | 9.61 | 9.86 | 9.64 |
| 8 | 11.96 | 11.22 | 13.61 | 11.08 | 13.86 | 14.19 | 13.63 |
| (c) PBTV: Performance based model selection, time-varying model | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 6.80 | 5.15 | 4.19 | 4.59 | 2.53 | 6.50 | 6.12 |
| 4 | 8.85 | 7.15 | 6.99 | 7.59 | 4.48 | 8.78 | 8.54 |
| 8 | 10.83 | 8.39 | 9.13 | 9.75 | 5.78 | 12.82 | 11.90 |

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1994Q4 until 2013Q3, thus for each horizon a total of 76 forecasts is computed. The different specification strategies are described in the text.

panel (a) gives rise to the following observations. First, both ex post 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. This indicates that the optimal specification of the various forecasting models changes over time.

Regarding the relative performance of the different forecasting models with PBC the entries in panel (b) indicate that the LBVAR and the BFAVAR again provide the most accurate forecasts for most horizons. However, especially for higher horizons the gains in accuracy of these two models over the AR benchmark are less pronounced with PBC (6% and 7% respectively for $h = 8$) than with IC (10% and 14% respectively for $h = 8$). By contrast, as the entries in panel (c) reveal with PBTV the best performing large scale methods can improve considerably upon the AR benchmark for all forecasting horizons. The DFM, which now clearly outperforms all remaining models by far, achieves a reduction in the absolute multivariate MSE upon the AR benchmark amounting to 60% for $h = 1$ and roughly 50% for higher horizons. The LBVAR and the BFAVAR, which rank lower than the FAAR with PBTV, outperform the AR by 25% and 32%, respectively, for $h = 1$ and 22% and 10%, respectively, for $h = 8$.

Overall the results documented in table 5 allow us to divide the different models into three groups according to the degree of robustness of their forecasting performance against alternative model specifications. First, for the AR benchmark, both model averaging techniques, EWA and BMA, the BFAVAR and (to a slightly lesser extend) the LBVAR we find that the specific model specification does not have a large impact on the models' forecasting performance. For the LBVAR we find that with PBTV the degree of shrinkage $\lambda$ varies strongly over time, while in our quasi real-time specification $\lambda$ is very stable over time. This also applies to the number of lags $p$ included in the estimation. However, the optimally specified model reduces the respective MSEs only very little. This finding is in line with Carriero et al. (2011) who document the robustness of the LBVAR's forecasting performance against the specific choice of $\lambda$ and $p$.

Second, according to our results the forecasting performance of the FAAR depends to a moderate degree on the precise model specification. The FAAR's forecasting performance would improve moderately if one could optimally specify the model in real-time. 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 (see also Schumacher (2007), p. 288).

Third, we show that the accuracy of the quasi real-time forecasts of the DFM depends to a very large degree on the specific model specification. Choosing the optimal specification and allowing for time heterogeneity rather than specifying the model based on information criteria leads to a considerable improvement in the model's forecasting performance. This confirms the findings of Schumacher (2007) who conducts the same analysis for static and dynamic factor models. One reason for the low degree of robustness of the forecasting performance of the DFM to alternative model specifications is certainly that the optimal number of dynamic factors $q$ 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.

**Real-Time Performance Based Model Specification and Forecast Pooling** To check whether the principle of performance based model selection can also increase the accuracy of the different forecasting models when applied 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 ranging from $T - s^{eval} + 1$ until $T$. The best specification of each forecasting model, i.e. the specification that yields the smallest MSE over the subevaluation sample, is then used to estimate the respective model with information up to $T$ and to compute forecasts for $T + h$. We set the length of the subevaluation sample $s^{eval}$ equal to 4 quarters. To be consistent, the various specifications of the different forecasting models for the subsample evaluation as well as for the final forecast are estimated over a rolling window of 60 quarters. This implies that our first subevaluation sample ranges from 1994Q4 until 1995Q3, while the forecasts of the different models for the exercise in this paragraph are evaluated from 1997Q3 until 2013Q3.

Alternatively, we implement forecast pooling, an approach that has been proposed in the literature to overcome the uncertainty related to the selection of the best performing specification of a forecasting model (see or example Kuzin et al., 2013). The basic idea here is, similar to model averaging, to pool over the forecasts obtained with a large set of different specifications of a forecasting model to obtain the final forecast of a variable of interest.[11] We implement two

[11] Conceptionally, the difference between model averaging and pooling lies in the source of uncertainty. While with model averaging there is uncertainty about the predictor variable to include in the estimation, with pooling there is uncertainty with respect to the best performing specification of a model given a set of predictors.

versions of forecast pooling: *unweighted pooling* and *MSE-weighted pooling*. According to the first variant, the final forecast of a variable is obtained by averaging over the various forecasts computed with different specifications of a certain forecasting model. By contrast, with the second variant we use a weighted mean to obtain the final forecast, where the weight is the inverse of the MSE of the respective model specification over the subevaluation sample ranging from $T - s^{eval} + 1$ until $T$.

In table 6, panel (b) - (d) we report the absolute multivariate MSEs that result from this exercise. To facilitate the direct comparison we additionally show the absolute MSEs of the different forecasting models with IC for the same evaluation sample in panel (a). The entries

Table 6: multivariate MSEs with IC, PBRT and Pooling

| (a) IC: Information criteria based model selection | | | | | | | |
|---|---|---|---|---|---|---|---|
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 7.63 | 6.56 | 6.78 | 6.36 | 7.01 | 7.30 | 7.27 |
| 4 | 11.10 | 9.68 | 11.36 | 9.25 | 11.40 | 10.04 | 10.26 |
| 8 | 13.84 | 12.64 | 17.71 | 11.71 | 16.79 | 15.09 | 15.01 |
| (b) PBRT: Performance based model selection, real time | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 7.57 | 7.12 | 7.37 | 5.85 | 9.91 | 7.27 | 7.17 |
| 4 | 10.24 | 14.77 | 12.54 | 10.82 | 17.00 | 10.18 | 10.35 |
| 8 | 12.96 | 33.26 | 14.89 | 14.40 | 22.20 | 14.12 | 13.97 |
| (c) Forecast pooling, unweighted mean | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 7.43 | 6.18 | 6.76 | 6.17 | 6.48 | 7.07 | 6.87 |
| 4 | 10.89 | 10.11 | 11.29 | 9.47 | 12.25 | 9.89 | 10.17 |
| 8 | 13.79 | 13.47 | 17.07 | 11.80 | 19.55 | 14.91 | 14.39 |
| (d) Forecast Pooling, MSE-based mean | | | | | | | |
| horizon | AR | LBVAR | FAAR | BFAVAR | DFM | EWA | BMA |
| 1 | 7.45 | 6.18 | 6.79 | 6.19 | 6.52 | 7.09 | 6.88 |
| 4 | 10.95 | 10.16 | 11.24 | 9.48 | 11.90 | 9.90 | 10.17 |
| 8 | 13.74 | 13.37 | 16.78 | 11.81 | 18.32 | 14.91 | 14.40 |

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1997Q3 until 2013Q3, thus for each horizon a total of 65 forecasts is computed. The different specification strategies are described in the text.

indicate that we can increase the accuracy of all forecasting models for $h = 1$ with either PBRT or forecast pooling compared to IC, though in most cases only by very little. Moreover, for higher horizons there is no improvement upon IC for the LBVAR, the BFAVAR and the DFM with either alternative specification approach.

When comparing the entries in panel (b) - (d), we find that, with the exception of the BFAVAR, in the short-run either unweighted or MSE-weighted forecast pooling works best for all models. This confirms previous findings that pooling is indeed a good alternative to avoid choosing a model specification in real time that does not forecast well (see Kuzin et al. (2013)). By contrast, for $h = 8$ all models for which we find an improvement upon IC perform best with PBRT.

The overall lowest multivariate MSEs for all forecasting horizons are again obtained by the

BFAVAR, with PBRT for $h = 1$ and with IC for higher horizons. The gains in accuracy upon the lowest multivariate MSE of the AR benchmark amount to roughly 20% for $h = 1$ and 10% for higher horizons.

Our results also indicate that the extremely good performance of the DFM with the optimized specification cannot be achieved with any model in real time. Therefore, relying on a single forecasting model that delivers a good performance and is not prone to the pitfall of choosing a specification that does not deliver the most precise forecasts, such as the BFAVAR or the LBVAR, seems to be a good choice for applied forecasters who cannot rely on out-of sample information to specify their forecasting models. Whether or not the small gains in forecasting accuracy over information criteria based model specification obtained with PBRT or forecast pooling justify the additional computational burden that comes with these approaches depends of course on the specific forecasting context at hand.

# 6 Conclusion

We have studied three different approaches to aggregating the informational content of a large dataset for forecasting key macroeconomic variables. We find that, overall, the Bayesian factor augmented vector autoregression and the large Bayesian vector autoregression perform best and generally yield more accurate forecasts than a simple AR benchmark model and other large scale approaches. This holds for both, measures for the joint forecasting performance for a set of 11 core variables as well as univariate performance measures for the individual series.

Our assessment of the joint predictive performance of the large scale approaches reveals that in general for short horizons all large scale approaches are able to efficiently use the correlation structure in the dataset to provide better forecasts than the AR benchmark, while for large horizons this is no longer the case. Here, the combination of aggregating the informational content of the large dataset into factors and shrinkage seems to be the most efficient approach to use all the information in the dataset. We also find that the joint forecasting performance of the BFAVAR and the LBVAR is very robust with respect to the precise model specification, e.g. the number of lags of the dependent variable or the degree of shrinkage.

Regarding the size of the gains in forecasting accuracy over the AR benchmark for the individual series of our set of 11 variables, we find considerable differences for the different variables. While there are large increases in forecasting performance for few variables, the gains in accuracy rarely exceed 10% in most 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. 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 AR 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. Finally, this might also be an advantage when it comes to the interpretation of the forecasts.

## References


Abberger, K. (2007). Forecasting quarter-on-quarter changes of German GDP with monthly business tendency survey results. Ifo Working Paper No. 40.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25*(1), 52–60.

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Banerjee, A. (2013). Editorial introduction to special issue on large data sets. *Oxford Bulletin of Economics and Statistics 75*(1), 1–5.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 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*(1), 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.

Buchen, T. and K. Wohlrabe (2014). Assessing the macroeconomic forecasting performance of boosting, evidence for the United States, the Euro Area, and Germany. *Journal of Forecasting 33*(4), 231–242.

Carriero, A., T. E. Clark, and M. Marcellino (2011). Bayesian VARs: Specification choices and forecast accuracy. *Federal Reserve Bank of Cleveland Working Paper 1112.*

Carriero, A., G. Kapetanios, and M. Marcellino (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics 26*(5), 735–761.

Christoffersen, P. F. and F. X. Diebold (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics 16*(4), 450–458.

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*(2), 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.

Drechsel, K. and R. Scheufele (2012a). Bottom-up or direct? Forecasting German GDP in a data-rich environment. Swiss National Bank Working Papers 16.

Drechsel, K. and R. Scheufele (2012b). The performance of short-term forecasts of the German economy before and during the 2008/2009 recession. *International Journal of Forecasting 28*(2), 428–445.

Dreger, C. and C. Schumacher (2005). Out-of-sample performance of leading indicators for the German business cycle: Single vs. combined forecasts. *Journal of Business Cycle Measurement and Analysis 1*, 71–87.

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*(6), 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*(471), 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*(4), 540–554.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., M. Lenza, and G. E. Primiceri (2012). Prior selection for vector autoregressions. NBER Working Paper 18467.

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.

Kholodilin, K. A. and B. Siliverstovs (2006). On the forecasting properties of the alternative leading indicators for the German GDP: Recent evidence. *Journal of Economics and Statistics 226*(3), 234–259.

Korobilis, D. (2013). Var forecasting using Bayesian variable selection. *Journal of Applied Econometrics 28*(2), 204–230.

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*(3), 392–411.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

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

Müller-Dröge, H. C., T. M. Sinclair, and H. O. Stekler (2014). Evaluating forecasts of a vector of variables: a German forecasting competition. CAMA Working Paper 55.

Pirschel, I. and M. H. Wolters (2014). Forecasting german key macroeconomic variables using large dataset methods. Kiel Working Paper 1925.

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting 26*(4), 271–302.

Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters 107*(2), 95–98.

Schumacher, C. (2011). Forecasting with factor models estimated on large datasets: A review of the recent literature and evidence for German GDP. *Journal of Economics and Statistics 231*(1), 28–49.

Schumacher, C. and C. Dreger (2004). Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models? *Journal of Economics and Statistics 224*(6), 732 – 750.

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*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 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*(6), 405–430.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society 58*(1), 267–288.

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*(2), 131–133.

Wolters, M. H. (2015). Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics 30*(1), 74–96.

Wright, J. (2009). Forecasting U.S. inflation by Bayesian model averaging. *Journal of Forecasting 28*(2), 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: Thomson Reuters 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[13] (all other series were already seasonally adjusted in the original data source).

[13] 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 11 Variables

- $\Delta gdp_t$: annualized quarter-on-quarter GDP growth
- $\Delta cpi_t$: annualized quarter-on-quarter CPI inflation rate
- $\Delta c_t$: annualized quarter-on-quarter private consumption growth
- $\Delta inv_t$: annualized quarter-on-quarter investment growth (machinery and equipment investment)
- $\Delta w_t$: annualized quarter-on-quarter wage inflation rate
- $\Delta ip_t$: annualized quarter-on-quarter industrial production growth
- $\Delta ppi_t$: 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

## Appendix C: Figures of Forecasts for 11 Variables

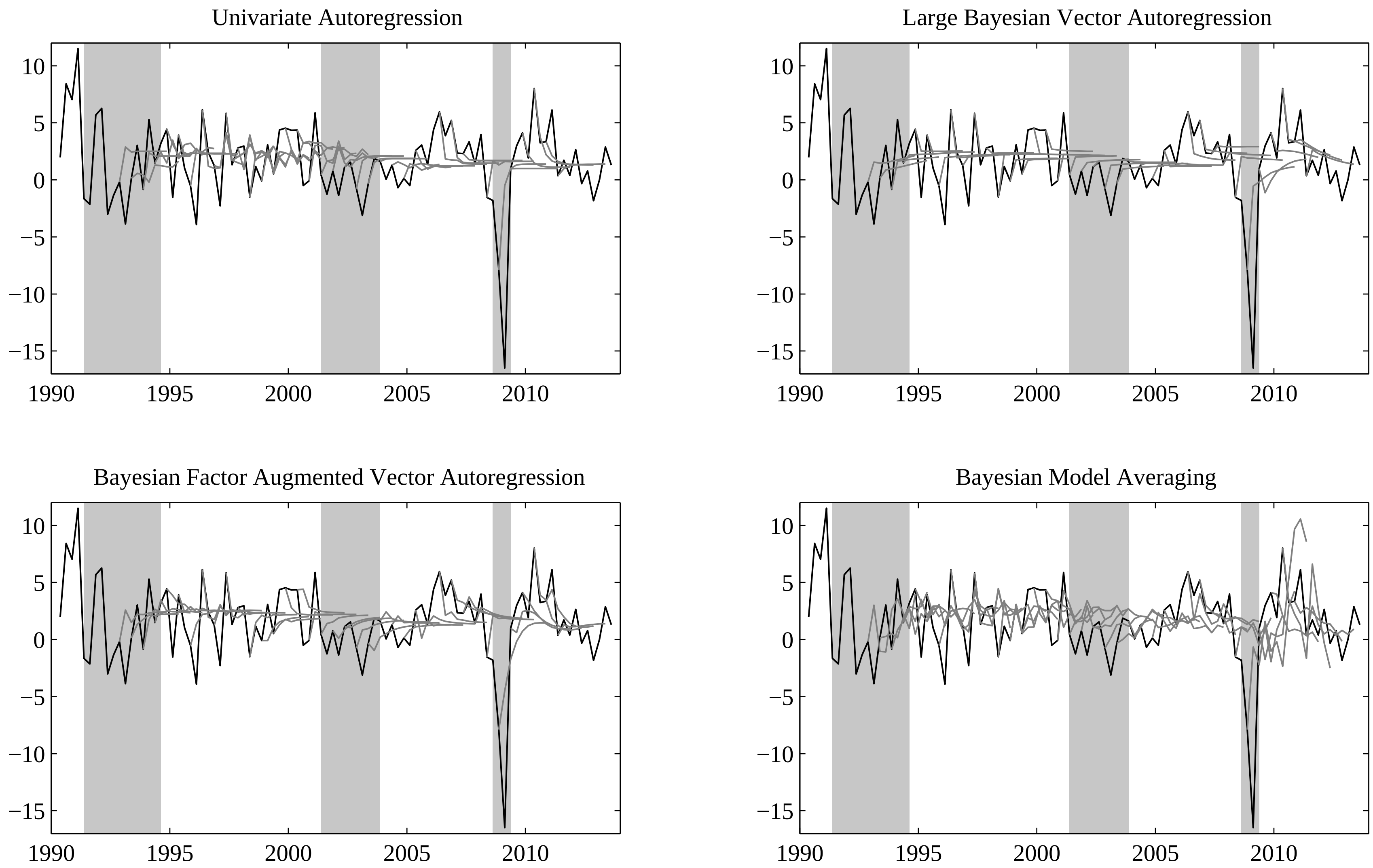


Figure 3: GDP growth forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German GDP growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German GDP growth obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

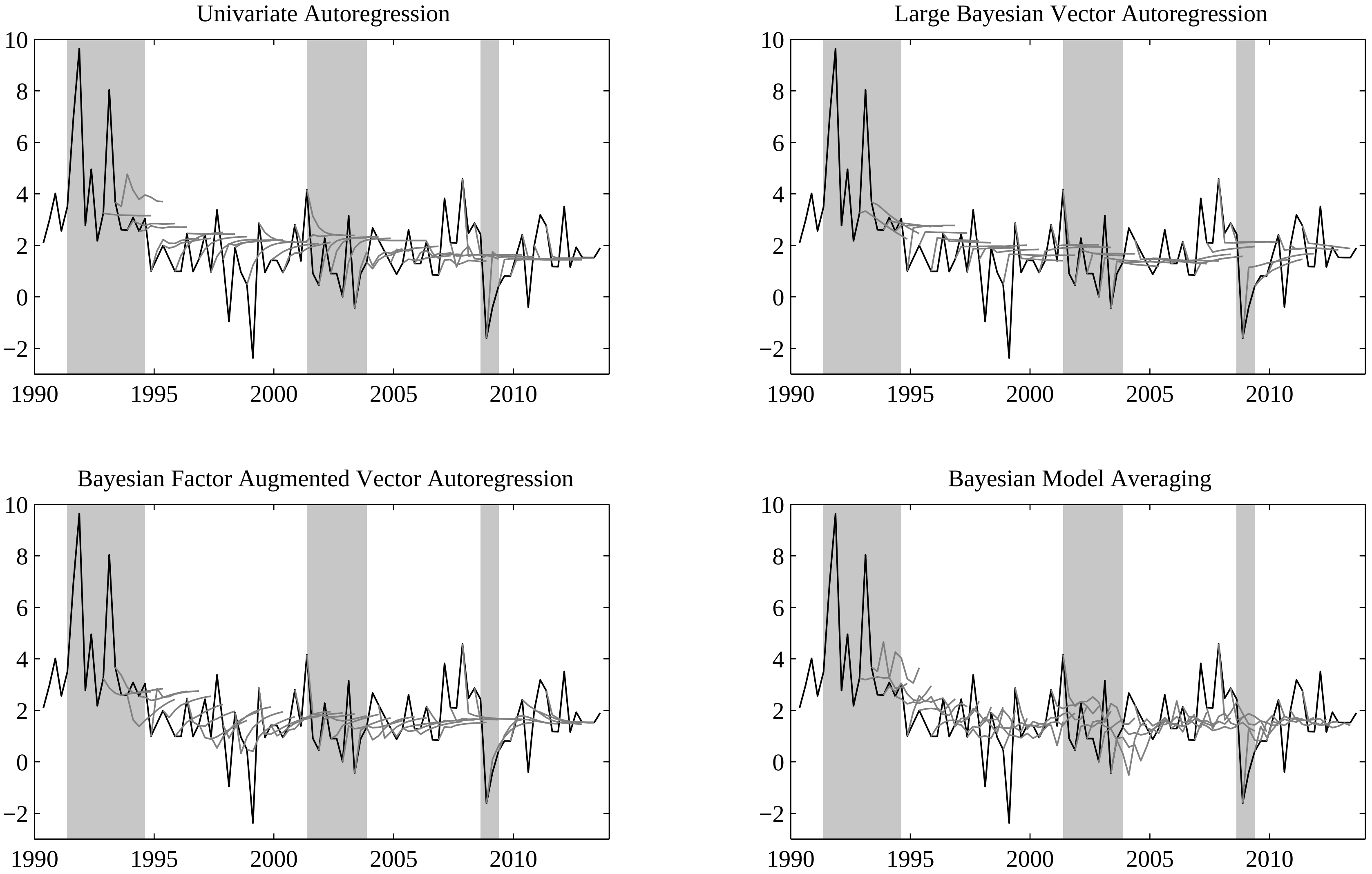


Figure 4: CPI inflation rate forecasts.

Notes: The solid black line displays the CPI inflation rate from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German CPI inflation obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

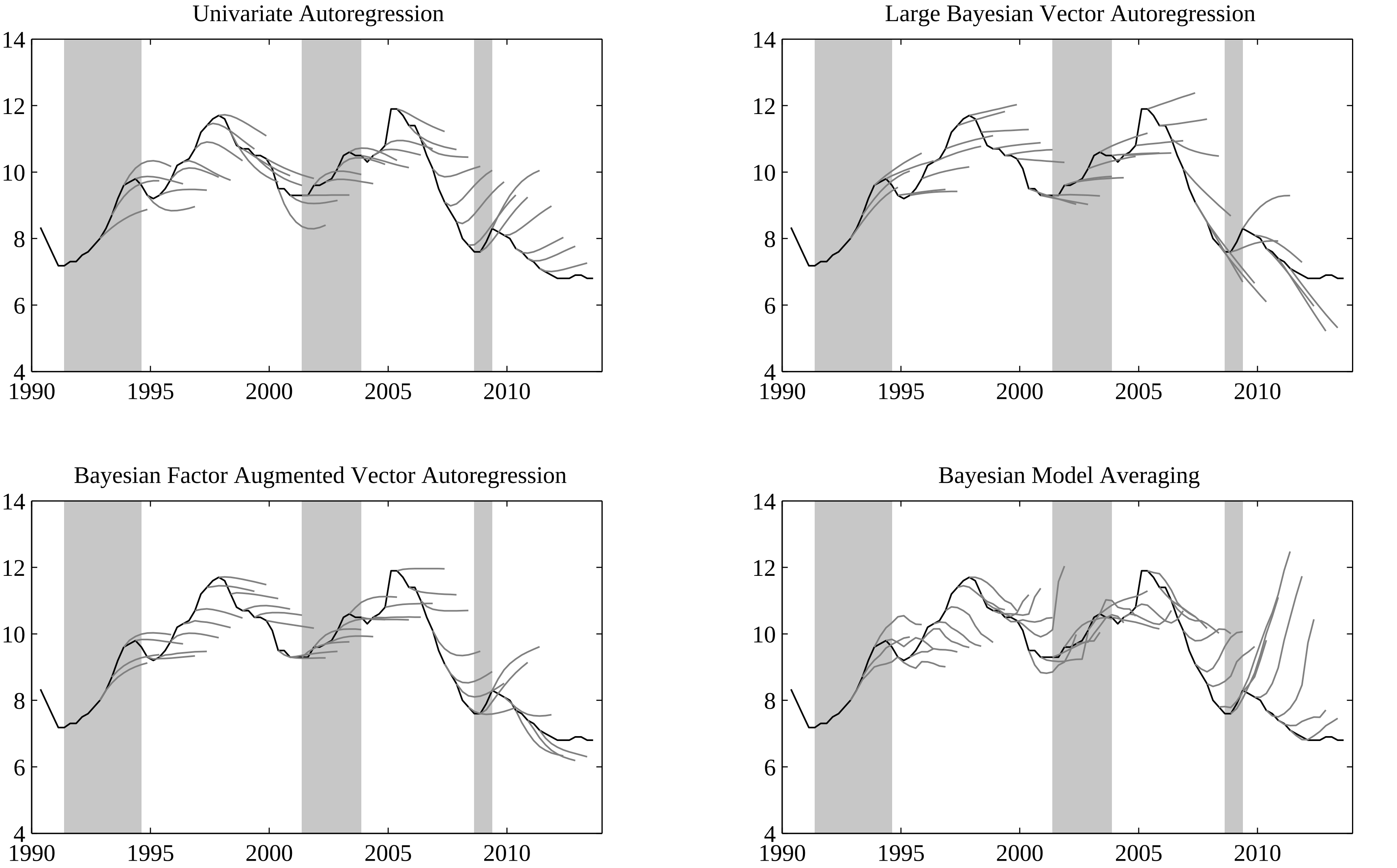


Figure 5: Unemployment rate forecasts.

Notes: The solid black line displays the German unemploament rate from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German unemployment obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

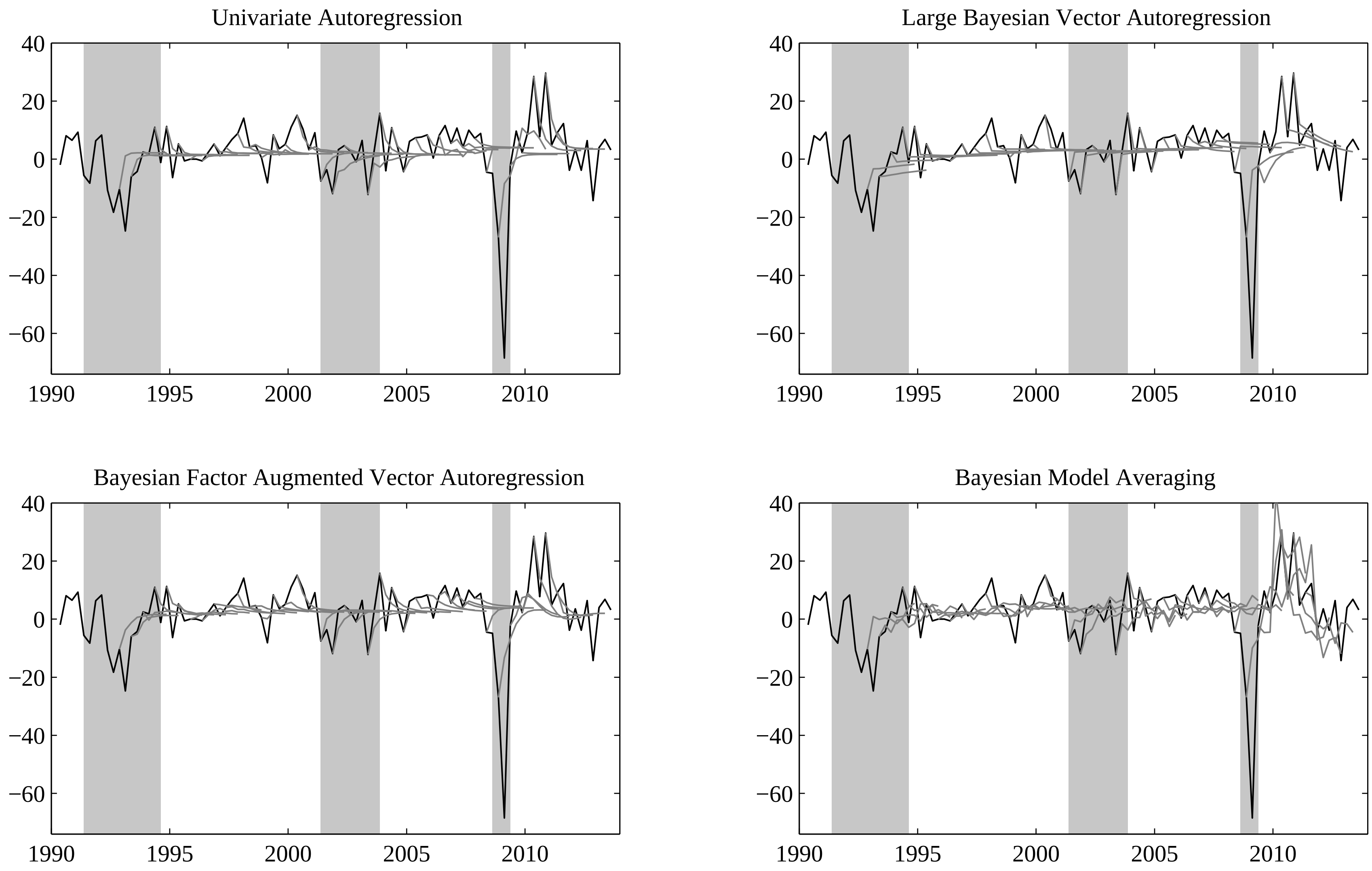


Figure 6: Industrial production growth forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German industrial production growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German industrial production obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

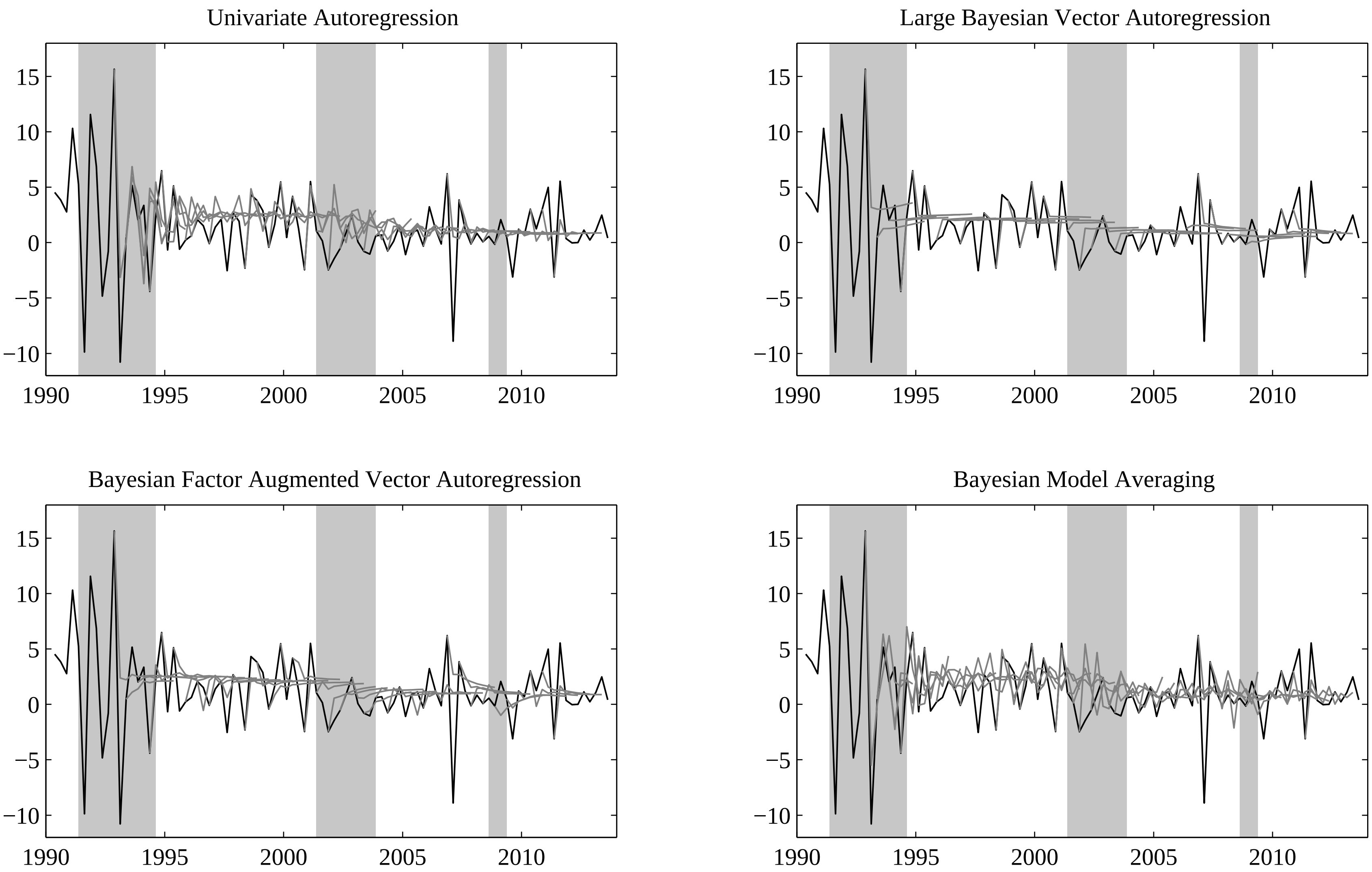


Figure 7: Private Consumption growth forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German private consumption growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German private consumption obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

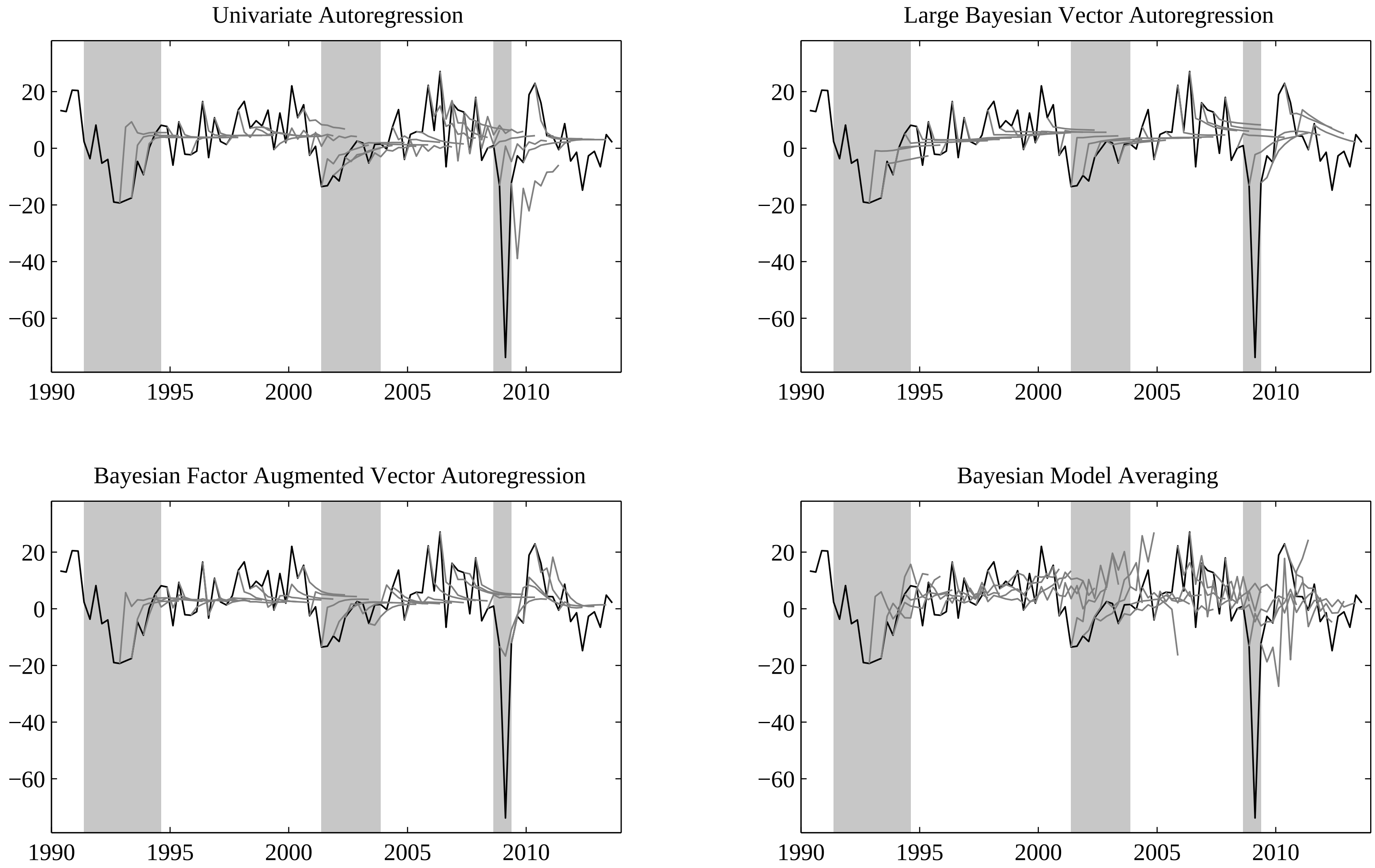


Figure 8: Machinery and Equipment Investment growth forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German machinery and equipment investment growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German machinery and equipment investment obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

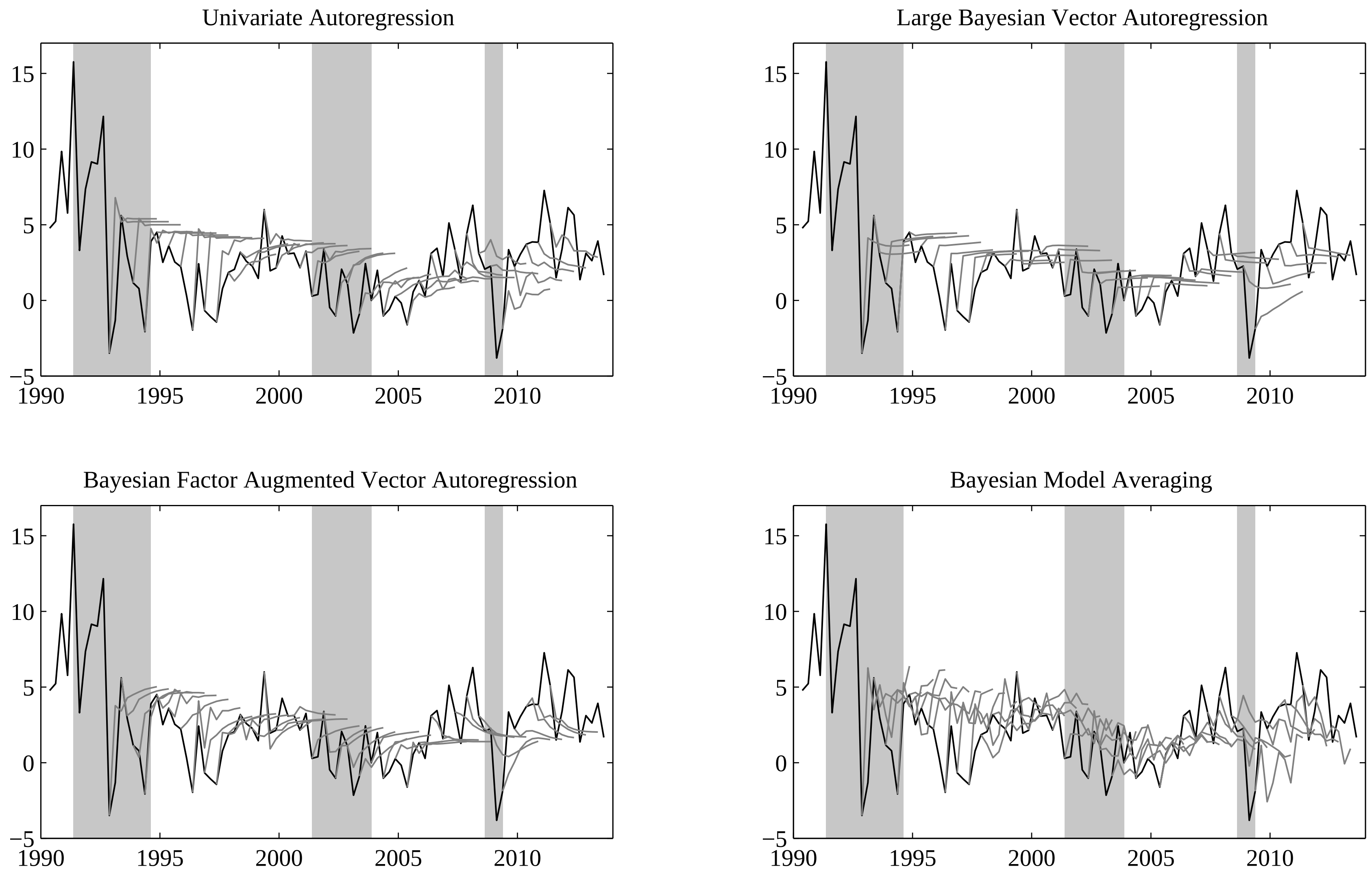


Figure 9: Real wage growth forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German real wage growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for German real wage growth obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

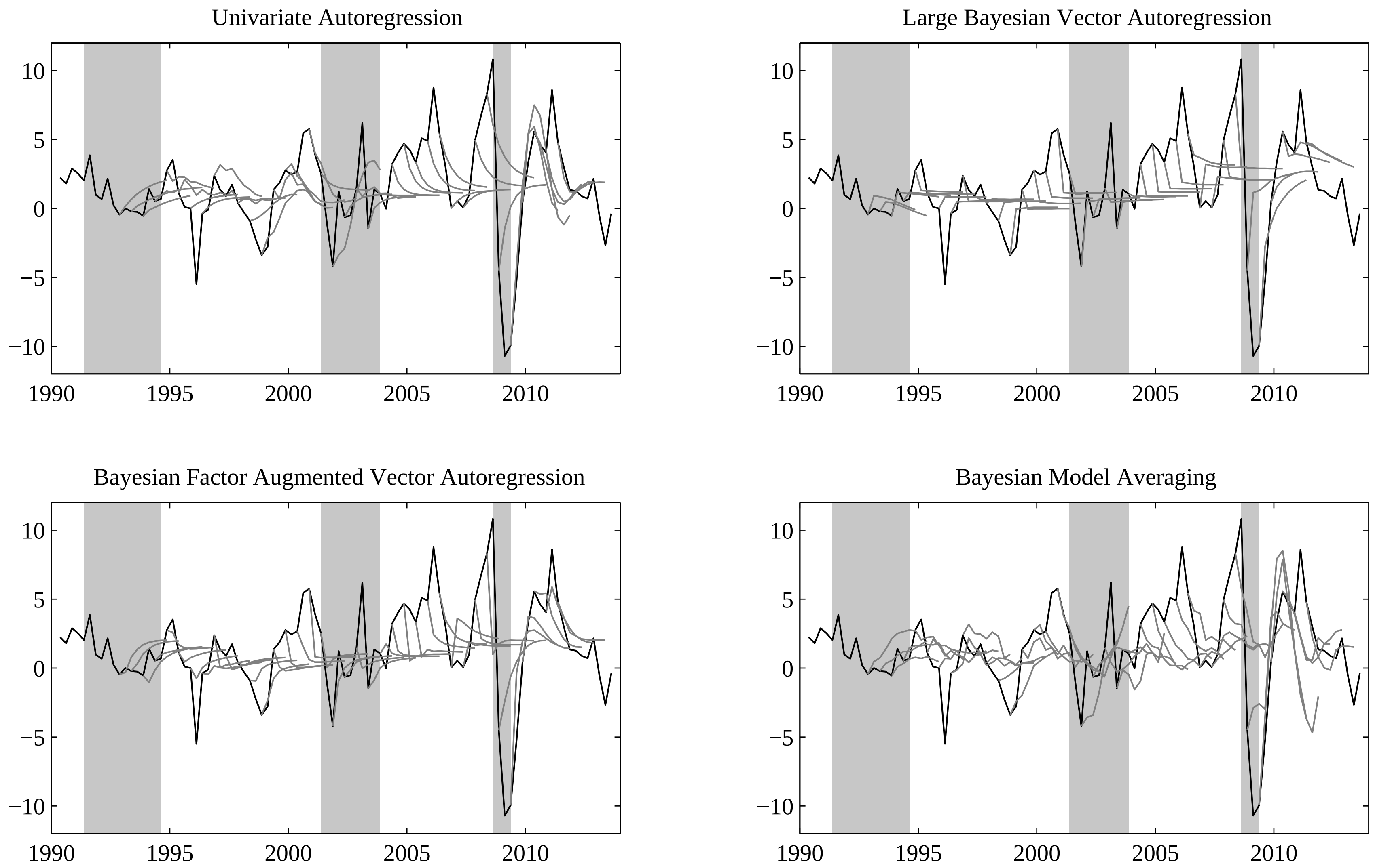


Figure 10: PPI Inflation Rate forecasts.

Notes: The solid black line displays annualized quarter-on-quarter German PPI inflation rate growth from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for the German PPI inflation rate obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

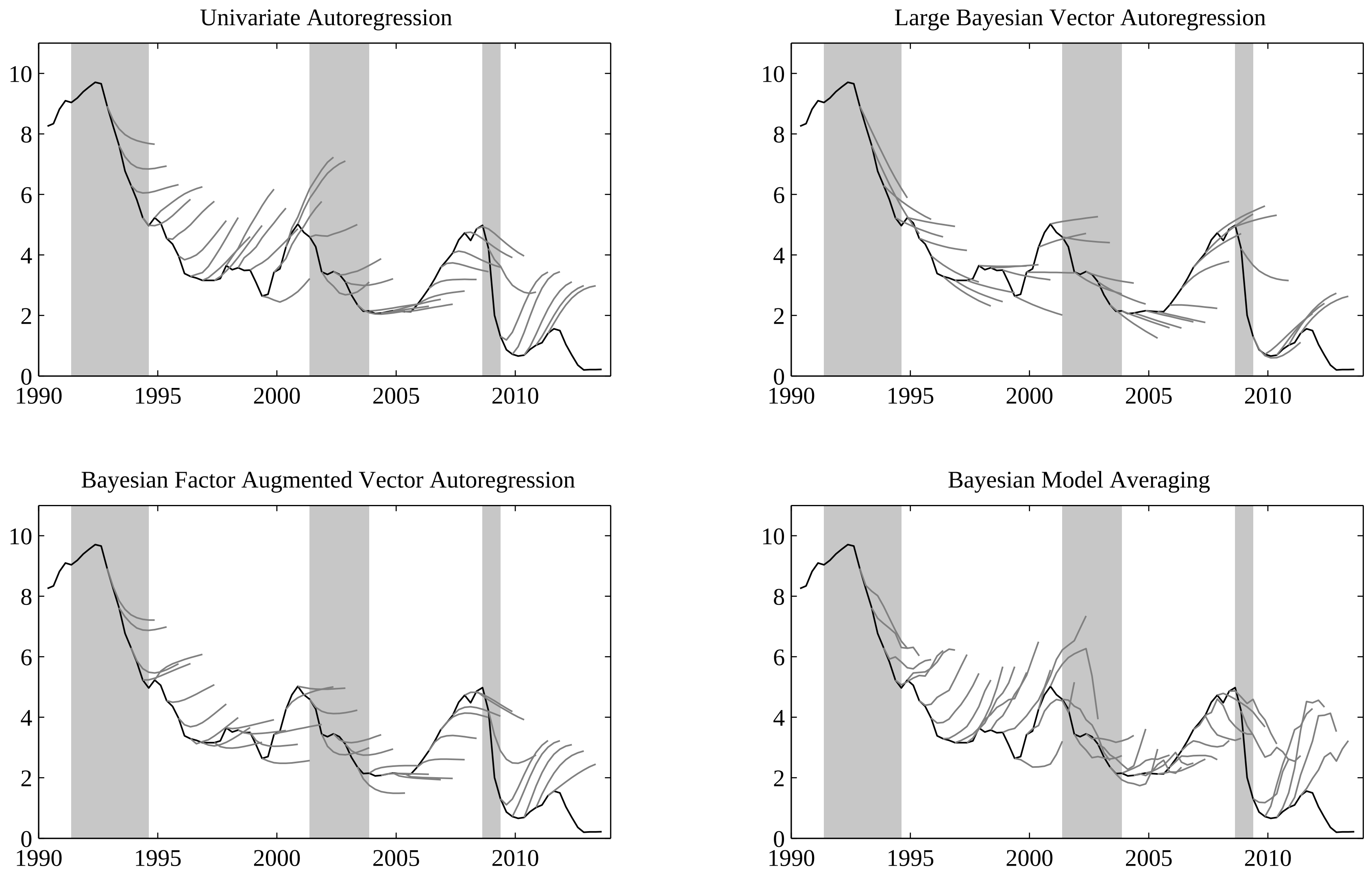


Figure 11: Short Term Interest Rate forecasts.
Notes: The solid black line displays the short term interest rate from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for the short term interest rate obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

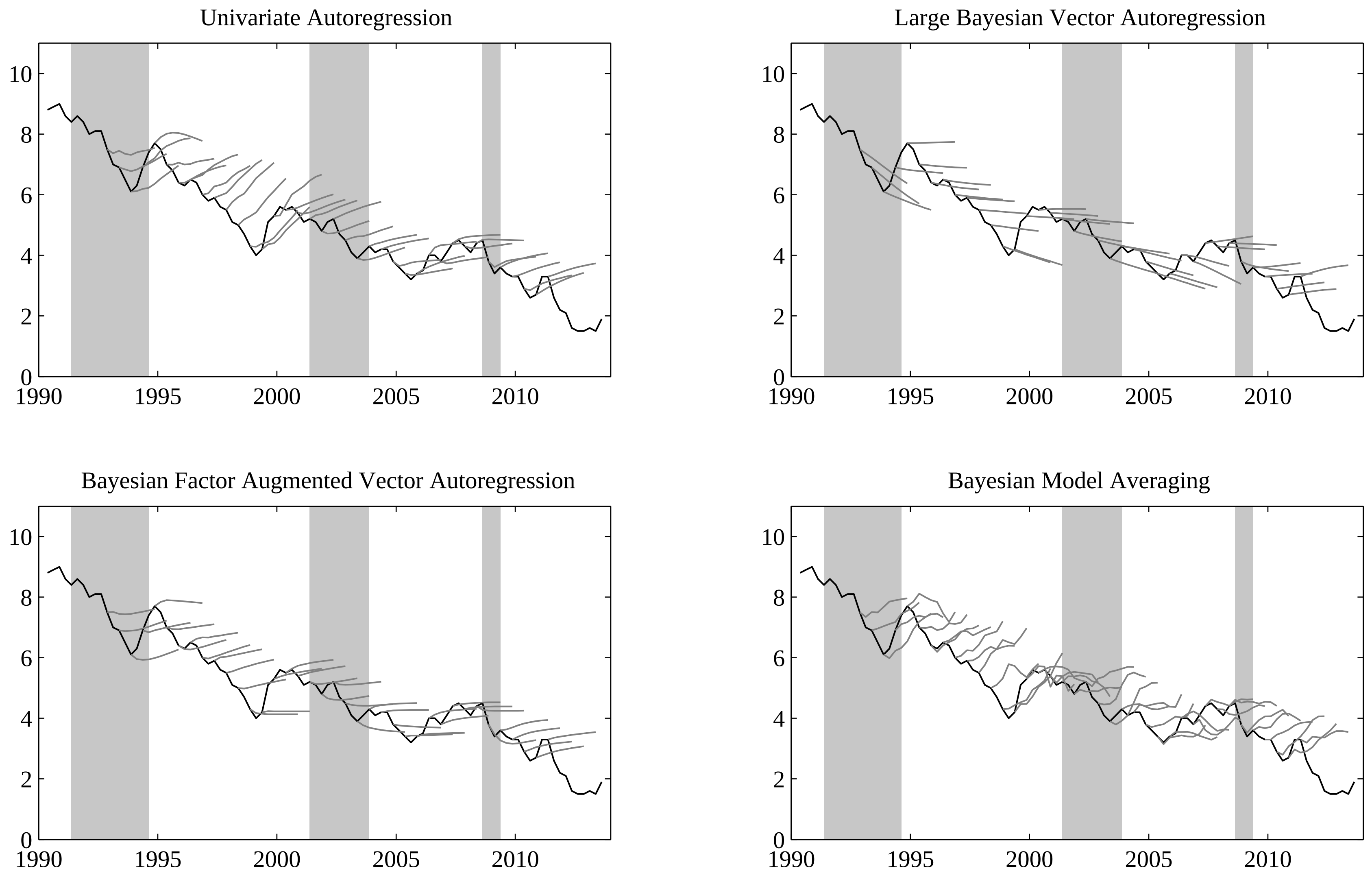


Figure 12: Long Term Interest Rate forecasts.

Notes: The solid black line displays the long term interest rate from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for the long term interest rate obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.

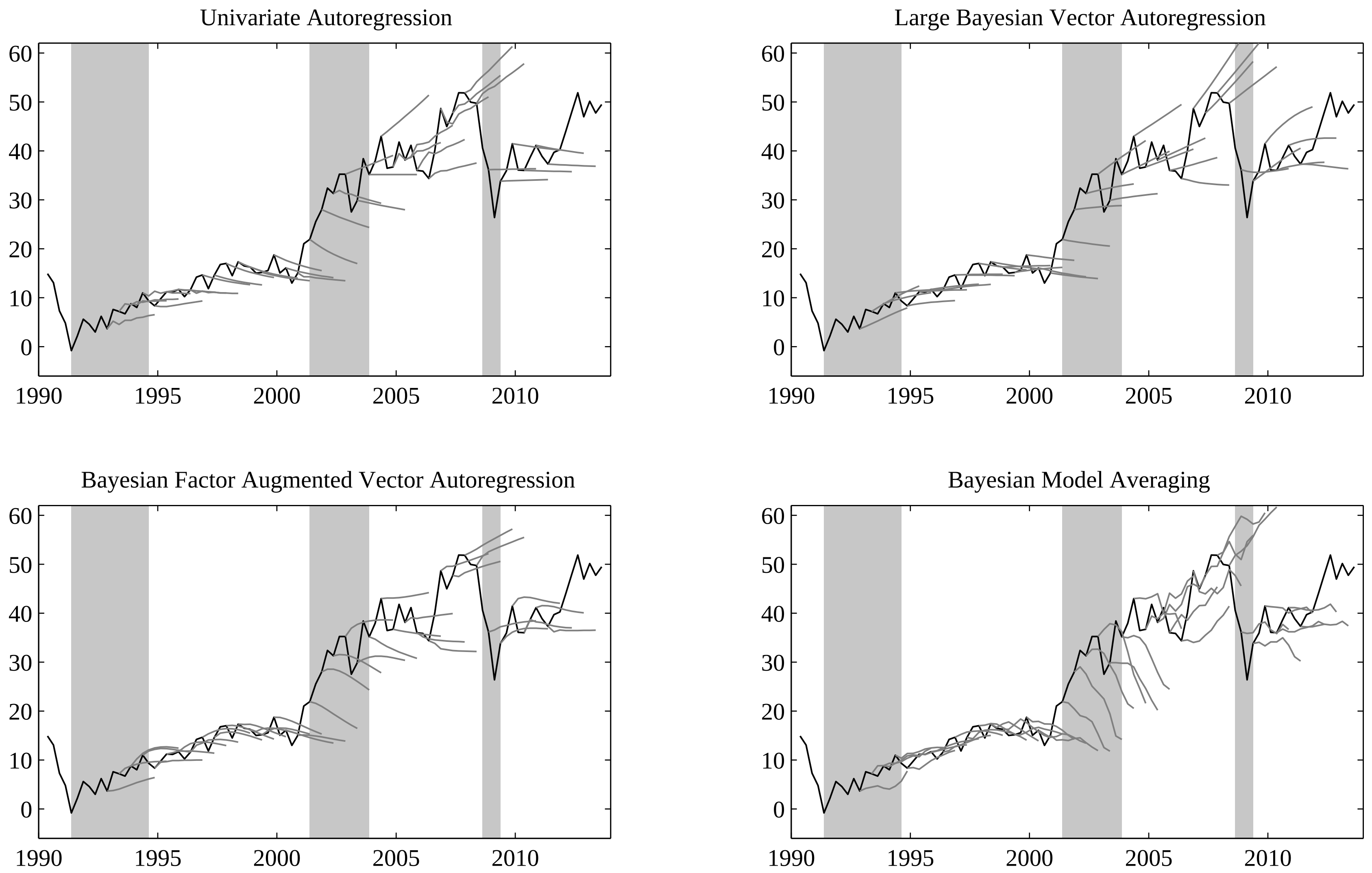


Figure 13: Current Account Balance forecasts.

Notes: The solid black line displays the German current account balance from 1990 until 2013. The grey lines show the $h = 1$ until $h = 8$ quarter ahead forecasts for the German current account balance obtained with the AR benchmark model and one representative variant of each of the large scale forecasting approaches, namely the LBVAR, the BFAVAR and the BMA. The shaded areas indicate recessions as dated by the Economic Cycle Research Institute.