ScienceDirect

International Journal of Forecasting
Volume 31, Issue 4, October–December 2015, Pages 1067-1095

Point and density forecasts for the euro area using Bayesian VARs

Tim O. Berg [a], Steffen R. Henzel [a b]

Show more

Outline | Share Cite

https://doi.org/10.1016/j.ijforecast.2015.03.006
Get rights and content

Abstract

We evaluate variants of the Bayesian vector autoregressive (BVAR) model with respect to their relative and absolute forecast accuracies using point and density forecasts for euro area HICP inflation and GDP growth. We consider BVAR averaging with equal and optimal weights, Bayesian factor augmented VARs (BFAVARs), and large BVARs with ad-hoc, optimal, and estimated hyperparameters. BVAR averaging delivers relatively high RMSEs, but performs better in terms of predictive likelihoods. Large BVARs show the opposite pattern, while BFAVARs perform satisfactorily under both criteria. Continuous ranked probability scores indicate that large BVARs suffer most from extreme observations. Using calibration tests, we detect that most BVARs produce reasonable density forecasts for HICP inflation, but not for GDP growth. In an extensive sensitivity analysis, we show that large BVARs are an excellent choice for certain specifications (recursive estimation, 22 variables, iterative approach, and optimal or estimated hyperparameters), while BFAVARs are competitive under most specifications, and specifically when the cross section is large.

Previous Next

Keywords

Bayesian vector autoregression; Forecasting; Model validation; Large cross section; Euro area

1. Introduction

When forecasting economic outcomes, a large set of indicators is desirable in order to avoid the omitted variable problem. However, forecasting models with large cross sections often suffer from overparameterization, leading to unstable parameter estimates and inaccurate forecasts. In vector autoregressions (VARs), the number of parameters may easily exceed the number of observations, which makes classical estimation infeasible in a data-rich environment.

Traditionally, factor models have been used for handling large cross sections and achieving dimension reduction.[1] In a seminal article, however, Bańbura, Giannone, and Reichlin (2010) argue that VARs can forecast better even when the number of variables is large. They propose Bayesian methods and impose additional information in the form of a Minnesota-type prior in order to shrink the overparameterized VAR towards a parsimonious random walk (see also Carriero et al., 2015, Carriero et al., 2009, D'Agostino et al., 2013, Giannone et al., 2015, Koop, 2013).

There are many possible ways in which a Bayesian VAR (BVAR) can be implemented for forecasting with large datasets. In this paper we build on the work of Bańbura et al. (2010) and evaluate variants of the BVAR which differ in the way in which information is condensed. In particular, we consider BVAR averaging, Bayesian factor augmented VARs (BFAVARs), and large BVARs. We also include the random walk variant and an autoregressive (AR) model as benchmarks. Moreover, we consider a range of specification choices, which affect the performance of each variant. Along with the aggregation weights of the BVAR averaging and the number of factors of the BFAVAR, there are three predominant approaches for determining the degree of shrinkage. First, we follow Bańbura et al. (2010) and select the shrinkage parameter such that the average in-sample fit for our target variables is the same across variants during a training sample period. Second, we obtain the parameter by maximizing the marginal likelihood in each period as per Carriero et al. (2015). Finally, we follow Giannone et al. (2015) and estimate the parameter by hierarchical modeling.

Our paper therefore evaluates all major specification choices previously discussed in the literature. To the best of our knowledge, no one else has yet compared all of these approaches. In a related study, Koop (2013) compares the forecast accuracies of a wide range of alternative prior specifications, including both conjugate and non-conjugate priors. In contrast, we focus on the conjugate Normal Inverse-Wishart (NIW) prior. Besides the fact that Koop (2013) uses US data, while we focus on the euro area, he considers neither BFAVARs with informative priors nor BVAR averaging, which we believe to be interesting modeling approaches. In addition, a particular feature of the analyses by Koop (2013) and Bańbura et al. (2010) is that the authors compare forecasting models of fairly different sizes. For instance, Bańbura et al. (2010) consider systems with 3, 7, 20, or even 131 variables. While these authors focus on the potential benefits of using larger information sets, we aim to reveal possible differences among the competing approaches with respect to the efficient use of a given amount of information. Thus, for each variant of the BVAR, we ensure that forecasts are produced conditional on the same dataset. We evaluate the BVAR variants according to their out-of-sample forecast performances one and four steps ahead. Specifically, we forecast the quarterly change in the euro area harmonized index of consumer prices (HICP) and the real gross domestic product (GDP).

To date, there exist few studies which have evaluated BVAR forecasts for aggregate euro area data (see, e.g., Bańbura et al., 2014, Giannone et al., 2014). Our dataset comprises 44 quarterly macroeconomic and financial indicators for the years 1975–2011. While applications for the US often build on datasets that contain more than one hundred variables, we believe that most countries do not have such large cross sections available.[2] This assumption should at least be true when the time series dimension is required to be large as well. Thus, it is not clear whether conclusions drawn from the specific case of the US will translate to other forecast situations. In our study, we consider a set of indicators that most forecasters would probably label a typical dataset. Moreover, we emphasize at this point that the size of our cross section is also appropriate with respect to all of the variants that we consider. Even for BFAVARs, it has been shown that about 40 series are sufficient to yield a satisfactory forecast accuracy (see Bai and Ng, 2002, Boivin and Ng, 2006). On the other hand, it has been documented that large BVARs achieve good forecast performances with about 20–25 variables (see, e.g., Bańbura et al., 2010, Giannone et al., 2015, Koop, 2013). Thus, our baseline results are derived from a subset of 22 variables, which is similar to that considered in the related literature, whereas all 44 variables are considered as a sensitivity analysis.

We distinguish between BVAR variants based on the accuracy of their point forecasts using root mean squared errors (RMSE), which is appropriate if the loss function of the forecaster depends solely on the forecast error. However, policymakers nowadays also monitor closely the uncertainty that is associated with prospective business cycle and inflation developments. The density forecasts of the Bank of England's Monetary Policy Committee and the *Sveriges*

Riksbank are prominent examples (see, e.g., Boero et al., 2011, Knüppel and Schultefrankenfeld, 2012, Mitchell and Hall, 2005, among others). Thus, we also evaluate density forecasts and rank BVAR variants on the basis of their predictive likelihoods, which is a standard tool in a Bayesian setting (see, e.g., Carriero et al., 2015, Clark, 2011, D'Agostino et al., 2013, Geweke and Amisano, 2010, Giannone et al., 2015, Koop, 2013, among others). However, since recent work by Clark and Ravazzolo (2015) and Ravazzolo and Vahey (2014) has suggested that predictive likelihoods are sensitive to large but infrequent forecast errors, we consider continuous ranked probability scores (CRPS) as an alternative. In addition to this forecasting competition, we also utilize calibration tests for assessing the performances of the density forecasts in absolute terms. Given that Rossi and Sehkposyan (2014) show that the density forecasts generated by BVARs are often not a reasonable description of the actual uncertainty, it is important to determine which specification choices will help to achieve the correct calibration.

Our baseline specification, involving direct-step forecasting and first (log-)differenced data, delivers the following results. Regarding point forecasts, all of the BVARs outperform the random walk variant for HICP inflation at both horizons. The differences in forecast accuracy among these variants are small and often insignificant. For GDP growth, the large BVAR delivers the best forecast one step ahead but cannot improve on the random walk four steps ahead. BFAVARs perform satisfactorily across the different target variables and forecast horizons. Moreover, neither selecting the shrinkage parameter optimally with respect to the marginal likelihood nor estimating it in a hierarchical fashion helps to improve the forecast accuracy compared to the shrinkage procedure of Bańbura et al. (2010). A similar conclusion can be drawn for BVAR averaging. Choosing the aggregation weights optimally according to historical predictive likelihoods does not improve the forecast accuracy compared to an equal weighting scheme.

With respect to density forecasts, we find evidence that an accurate point forecast does not necessarily imply a high predictive likelihood, and *vice versa*. We show that the relative performances of the large BVARs worsen when the predictive likelihood is considered, while that of the BVAR averaging improves. Again, we find that neither choosing the degree of shrinkage optimally nor estimating it in a hierarchical fashion changes this result. When considering CRPS, however, large BVARs show a better relative performance. Such a result suggests that the density forecasts from large BVARs suffer most from extreme observations, but perform well most of the time. Moreover, calibration tests suggest that most BVAR variants produce density forecasts that appear to provide reasonable descriptions of the true distribution for HICP inflation, but we detect poor calibrations of almost all variants with respect to GDP growth.

The results are similar when subsamples are considered or the variables enter in (log-)levels instead of first (log-)differences. In an extensive sensitivity analysis, we also show that large BVARs benefit when a recursive estimation scheme is adopted rather than a rolling one. While the performances of all BVAR variants improve sizably in terms of predictive likelihoods (but not RMSEs), large BVARs profit most from a longer estimation sample. Similarly, all variants gain in terms of predictive likelihoods when four-quarter-ahead forecasts are derived in an iterative fashion, compared to a direct-step approach. Large BVARs benefit most in this case, particularly when the shrinkage parameters are chosen optimally or estimated. However, we also find that, in contrast to BFAVARs and BVAR averaging, large BVARs do not benefit from an enlarged cross section.

The remainder of this paper is organized as follows. In Section 2, we describe our dataset. Section 3 develops the BVAR model and all of the variants that we use to produce out-of-sample forecasts for euro area HICP inflation and GDP growth. In Section 4, we explain our forecast experiment and present the main results. We evaluate the relative performances of the variants based on their point and density forecasts. Moreover, we investigate how good the forecasts are in absolute terms. Sensitivity analyses are provided in Section 5; and finally, Section 6 concludes.

2. Dataset

The full dataset comprises 44 quarterly euro area macroeconomic and financial time series covering the period 1975:1 to 2011:4. In the case of aggregate euro area data, both the cross sectional and time series dimensions represent

approximately the maximum size available at the moment. The dataset includes the overall HICP and real GDP as the series of interest, as well as a collection of indicator variables which cover the following seven categories: national accounts data, price indexes, international data, employment data, surveys, monetary aggregates, and financial data. For the baseline specification, we follow common practice in the forecasting literature and achieve stationarity by (log-)differencing the (non-annualized) data. In a sensitivity analysis, we also consider all variables in (log-)levels, except for those that are already expressed in rates.

In most cases, the series are obtained from the 12th update of the Area-wide Model (AWM) database,[3] which is maintained by the European Central Bank (ECB) and made available by the Euro Area Business Cycle Network (EABCN). The historical series are backdated in a coherent manner using individual country information.[4] The AWM database is the preferred source for both researchers and policymakers who are interested in the euro area. In addition, we also include survey data and monetary aggregates from the Main Economic Indicators (MEI) of the Organisation for Economic Co-operation and Development (OECD), as well as a share price index from Thomson Reuters. A detailed description of the dataset is provided in Appendix A.

The dataset therefore includes all of the variables that a forecaster typically has on her wish list when forecasting euro area HICP inflation and GDP growth. Besides the fact that forecasting euro area data is interesting in its own right, note that, for countries other than the US, no long time series are available for literally hundreds of indicator variables. Hence, our dataset is also typical in the sense that it strikes a balance between the maximum availability of the cross sectional and time series dimensions of the predictors.

However, our baseline results are derived from a subset of only 22 variables, since the related literature has found that BVARs with about 20–25 variables are generally superior to those that include more variables (see, e.g., Bańbura et al., 2010, Giannone et al., 2015, Koop, 2013). In practice, a forecaster may choose a subset of indicators from all of the series available, based on either past experience or sound economic arguments. The latter consideration leads to information sets like the one we consider in our baseline specification. With respect to both the number and type of indicator variables, the baseline dataset is therefore similar to those considered in the related literature. We provide details of its exact composition in Appendix A. Note that we provide a sensitivity analysis using the full dataset.

Finally, at this point we must stress that consistent vintage data are not available, since the euro area did not exist before January 1999. While the absence of real time data is a potential drawback in studies where the objective is to conduct a realistic out-of-sample forecast experiment, we believe that it is not an issue in our case. The forecast experiment in this paper is understood as a model validation exercise that is designed to investigate which variant of the BVAR deals best with the overparameterization problem that is inevitably inherent in models with large cross sections. Evaluating out-of-sample forecasts is an appropriate and established procedure for doing this, since the forecasts reflect all of the sources of error that are typically associated with the modeling of economic outcomes, including parameter uncertainty and model misspecification.

3. Forecasting with a BVAR model

In this section, we develop the BVAR model and the different variants that we use. Moreover, we discuss different specification choices for the shrinkage parameter and the lag length.

3.1. BVAR model

In our baseline specification, we follow Koop (2013) and use the direct-step forecast approach, which entails estimating a different VAR for each horizon and converts the problem of multistep forecasting into one that involves only one-step-ahead forecasting. Specifically, we estimate

$$y_t = c_h + B_{1,h} y_{t-(h-1)-1} + \cdots + B_{p,h} y_{t-(h-1)-p} + u_{t,h}, \tag{1}$$

where y_t is a $n \times 1$ vector of variables, including HICP inflation and GDP growth, among others; c_h is a $n \times 1$ vector of intercepts; $B_{i,h}$ are $n \times n$ matrices of coefficients; $i = 1, \ldots, p$ denotes the lags included; $h \geq 1$ is the forecast horizon; $u_{t,h}$ is a $n \times 1$ vector of normally distributed residual terms with zero mean and covariance matrix Σ_h; and data are available for $t = 1 - (h-1) - p, \ldots, T$. Let us denote $y = (y_1, \ldots, y_T)'$, $x_{t,h} = \left(y'_{t-(h-1)-1}, \ldots, y'_{t-(h-1)-p}, 1\right)'$, $x_h = (x_{1,h}, \ldots, x_{T,h})'$, $B_h = (B_{1,h}, \ldots, B_{p,h}, c_h)'$, and $u_h = (u_{1,h}, \ldots, u_{T,h})'$. The VAR in Eq. (1) can be rewritten as $y = x_h B_h + u_h$. Moreover, let $\beta_h = \mathbf{vec}(B_h)$, with $\mathbf{vec}(\cdot)$ being the column stacking operator and $k = n(1 + np)$. Then, β_h is a $k \times 1$ vector containing all of the coefficients in the model.

In our forecast experiment, we estimate the VAR on up to $n = 44$ variables, including $p = 4$ lags of each variable (and hence, $k = 7788$). However, such a large dimensional system of multivariate regressions cannot be estimated without imposing additional prior beliefs on the parameters. In addition, there is evidence that even VARs with only a handful of variables might benefit from prior information being imposed (see, e.g., Robertson & Tallman, 1999, among others). We follow common practice and use a variant of the Minnesota prior to deal with the dense parameterization of the model. The basic idea is that a random walk with drift is a reasonable description of the data generating process behind most macroeconomic and financial time series when considered in (log-)levels. In addition, the prior captures the belief that own lags are more informative than those of other variables, and more recent lags contain more information than more distant ones. For (log-)differenced data, a white noise process is the appropriate prior, and hence, the VAR is centered around the prior mean $y_{i,t} = c_{i,h} + u_{i,t,h}$ (or $y_{i,t} = c_{i,h} + y_{i,t-(h-1)-1} + u_{i,t,h}$ for (log-)levels), and imposing the white noise prior amounts to shrinking the coefficients towards zero.

In contrast to the original Minnesota prior developed by Litterman, 1980, Litterman, 1986, we do not assume that the residual covariance matrix Σ_h is known and diagonal. Instead, we use a generalized version of the prior proposed by Kadiyala and Karlsson, 1993, Kadiyala and Karlsson, 1997, which allows for correlations among the residual terms. The evidence provided by Bańbura et al. (2010) and Robertson and Tallman (1999) suggests that a generalized Minnesota prior produces accurate forecasts for major macroeconomic series such as GDP growth or inflation, even though the $n(n+1)/2$ distinct elements of Σ_h have to be estimated on top of the k coefficients.

In particular, we consider a conjugate NIW prior of the following form:

$$\Sigma_h \sim \mathrm{IW}(\Psi_h, d_h) \quad \text{and} \quad \beta_h | \Sigma_h \sim \mathrm{N}(b_h, \Sigma_h \otimes \Omega_h), \tag{2}$$

where $\otimes$ denotes the Kronecker product and the elements Ψ_h, d_h, b_h and Ω_h are functions of hyperparameters. The conjugate prior implies a likelihood and posterior that come from the same family of distributions, hence making Bayesian inference feasible even for large values of n.[5]

We follow Bańbura et al. (2010) and implement the prior by constructing the following set of artificial observations:

$$y_h^+ = \begin{bmatrix} \mathrm{diag}(\delta_1 \sigma_{1,h}, \ldots, \delta_n \sigma_{n,h}) / \lambda_h \\ 0_{n(p-1) \times n} \\ \mathrm{diag}(\sigma_{1,h}, \ldots, \sigma_{n,h}) \\ 0_{1 \times n} \end{bmatrix}, \tag{3}$$

$$x_h^+ = \begin{bmatrix} \mathrm{diag}(1, 2, .., p) \otimes \mathrm{diag}(\sigma_{1,h}, \ldots, \sigma_{n,h}) / \lambda_h & 0_{np \times 1} \\ 0_{n \times np} & 0_{n \times 1} \\ 0_{1 \times np} & \epsilon \end{bmatrix},$$

where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix. The hyperparameters δ_i are set equal to either 1, reflecting the prior belief that the variables are characterized by high persistence ((log-)levels), or 0, for variables showing a strong mean reversion ((log-)differences). The hyperparameters $\sigma_{i,h}$ account for the different scale and variability of the series, and are set equal to the standard deviation of a residual from a univariate autoregression for the variable $y_{i,t}$ on a training

sample running from 1975:2 to 1990:1. The lag order is the same as in the VAR ($p = 4$). The hyperparameter ϵ is set to a very small number (10^{-4}), reflecting a diffuse prior for the intercept terms. Finally, the hyperparameter λ_h determines the degree of shrinkage, and hence, the tightness of the prior. As $\lambda_h \to \infty$, the prior becomes uninformative and the posterior expectations coincide with the ordinary least squares (OLS) estimates. For $\lambda_h \to 0$, the posterior equals the prior and the indicator variables do not influence the estimation outcome. Hence, the hyperparameter λ_h is the key parameter in the BVAR, and its selection for each variant is discussed in detail below.

To improve the forecast accuracy of BVARs further, the literature proposes that additional information be imposed in the form of a "sum-of-coefficients" prior (see, e.g. Bańbura et al., 2010, Doan et al., 1984, Giannone et al., 2015, among others). This prior is implemented by generating n artificial observations, and reflects the belief that a no-change forecast is a good forecast at the beginning of a sample period. In particular, we construct:

$$\begin{aligned} y_h^{++} &= \operatorname{diag}\left(\delta_1\mu_1, \ldots, \delta_n\mu_n\right)/\tau_h, \\ x_h^{++} &= \left[1_{1\times p} \otimes \operatorname{diag}\left(\delta_1\mu_1, \ldots, \delta_n\mu_n\right)/\tau_h \; 0_{n\times 1}\right], \end{aligned} \tag{4}$$

where $\tau_h = 10\lambda_h$ and the hyperparameters μ_i capture some prior belief about the average level of the variable $y_{i,t}$. Consistent with the calibration of the $\sigma_{i,h}$ values, we set μ_i equal to the average value of $y_{i,t}$ in the training sample period 1975:2 to 1990:1.[6]

The artificial observations are added on top of the data matrices, which are then used for inference. The augmented regression model reads as

$$y_h^* = x_h^* B_h + u_h^*, \tag{5}$$

where $y_h^* = \left(y', y_h^{+'}, y_h^{++'}\right)'$, $x_h^* = \left(x_h', x_h^{+'}, x_h^{++'}\right)'$, $u_h^* = \left(u_h', u_h^{+'}, u_h^{++'}\right)'$, $u_h^{+'} = \left(u_{1,h}^+, \ldots, u_{T,h}^+\right)$, and $u_h^{++'} = \left(u_{1,h}^{++}, \ldots, u_{T,h}^{++}\right)$. The latter two expressions are the residual terms corresponding to y_h^+ and y_h^{++}, respectively. Adding artificial observations ensures that the total number of observations (actual plus artificial) is always larger than the number of parameters to be estimated.

The posterior of the parameters can be computed in closed form as

$$\Sigma_h \middle| \gamma_h, y \sim \mathrm{IW}\left(\widehat{\Sigma}_h, T+n+2\right) \quad \text{and} \beta_h \middle| \gamma_h, \Sigma_h, y \sim \mathrm{N}\left(\widehat{\beta_h}, \Sigma_h \otimes \left(x_h^{*'} x_h^*\right)^{-1}\right), \tag{6}$$

where $\gamma_h = (\epsilon, \delta_i, \sigma_{i,h}, \lambda_h, \tau_h, \mu_i)'$ collects all of the hyperparameters of the model, and $\widehat{\Sigma}_h$ and $\widehat{\beta}_h$ are the covariance matrix and the coefficients from an OLS regression of y_h^* on x_h^*, respectively.

One advantage of the direct-step forecast approach is that an analytic expression of the posterior can be derived even for multistep forecasts, i.e., the predictive density $p_h\left(y_{T+1,h} \middle| \gamma_h, \beta_h, \Sigma_h, y\right)$ is also available in closed form. Moreover, we exploit the conjugacy of the prior fully, and make use of the fact that, after integrating out Σ_h, draws for the predictions can be obtained directly from their marginal posterior, which follows a multivariate t-distribution (see Koop, 2013). In particular, the predictive mean is given by

$$\mathrm{E}\left(y_{T+1,h} \middle| \gamma_h, y\right) = \left(x_{T+1,h}\widehat{B}_h\right)', \tag{7}$$

while the predictive covariance matrix is

$$\operatorname{var}\left(y_{T+1,h} \middle| \gamma_h, y\right) = \frac{1}{T+n}\left(1 + x_{T+1,h}\left(x_h^{*'} x_h^*\right)^{-1} x_{T+1,h}'\right)\widehat{\Sigma}_h. \tag{8}$$

For most of the BVAR variants, the analytic expressions in Eqs. (7), (8) are used to evaluate point and density forecasts directly, and drawing from the marginal posterior is unnecessary, which makes computation convenient and fast.[7]

We should stress that the direct-step forecast approach is an approximation for $h > 1$, since it rests on the assumption that the error terms in the model are serially uncorrelated, which is not the case. Exact multistep forecasts can be obtained by iteration. However, the direct-step approach is more robust to model misspecification because misspecification will generally inflate the forecast error of iteratively constructed multistep forecasts (see, e.g., Carriero et\special {t4ht=&} al., 2015, for a similar argument). In the sensitivity analysis, we compare the iterative and direct-step approaches. The iterative approach always requires Monte Carlo integration, since the forecasts depend on the coefficients in a nonlinear fashion, and hence, no closed form solution for their marginal posterior is available. Appendix C explains how we obtain forecasts in this case.

3.2. Model variants

We consider the following variants of the BVAR, which provide us with different ways of dealing with the overparameterization problem and condensing the information in the dataset.

Random walk. As a benchmark, we consider the random walk variant. Random walk forecasts are obtained by imposing a dogmatic prior (hence, $\lambda_h \to 0$).[8] Thus, the posterior beliefs are not shaped by the indicator variables, and the random walk is a natural benchmark for investigating whether or not using these series is useful at all for forecasting. The predictive means for HICP inflation and GDP growth are therefore equal to the estimated constant. Note that this term may adapt over time, providing a naïve but competitive benchmark for any forecasting model.

BVAR averaging with equal weights. Recent studies argue that averaging forecasts is a simple and successful method of handling large datasets and improving levels of out-of-sample forecast accuracy (see, e.g., Aiolfi et al., 2011, Barnett et al., 2014, Clark and McCracken, 2010, Henzel and Mayr, 2013, Wolters, 2015, among others). This second variant is motivated by portfolio diversification or hedging arguments, guaranteeing an insurance against large forecast errors. We therefore estimate $m = n - 2$ three-variable BVARs, each of which includes HICP inflation and GDP growth, plus one other indicator variable at a time. We then average the forecasts of the individual models. By segmenting the set of indicators and estimating a battery of parsimonious models, pooling allows us to condense information and avoid parameter proliferation. Therefore, we assume that no shrinkage is needed, and impose an uninformative prior ($\lambda_h \to \infty$).[9] We obtain the predictive densities for HICP inflation and GDP growth by combining the predictive densities of different models as follows:

$$p_h\left(y_{T+1,h}|y, w_{T-1,h}\right) = \sum_{i=1}^{m} w_{T-1,i,h}\, p_{i,h}\left(y_{T+1,h}|y\right), \tag{9}$$

where $w_{T-1,h} = (w_{T-1,1,h} \ldots w_{T-1,m,h})$ contains nonnegative weights that sum to one. Thus, the linear combination of predictive densities is again a predictive density (see Aastveit et al., 2014, Geweke and Amisano, 2011, Geweke and Amisano, 2012, Hall and Mitchell, 2007, among others). In this variant, we fix $w_{T-1,i,h}$ at $1/m$ for all T, i, and h. Such a naïve equal weighting scheme has been found by several authors to perform astonishingly well (see, e.g., Clark and McCracken, 2010, Henzel and Mayr, 2013, among others).

BVAR averaging with optimal weights. The third variant is identical to the second, except that the weights are now chosen using an optimality criterion and computed in real time.[10] In particular, we follow Geweke and Amisano, 2011, Geweke and Amisano, 2012 and calculate the vector $w^*_{T-1,h}$ which maximizes the historical (log) predictive likelihood in each period T:

$$w^*_{T-1,h} = \operatorname{argmax} \sum_{s=1}^{T-1} \log\left[\sum_{i=1}^{m} w_{T-1,i,h}\, p_{i,h}\left(y_s|y\right)\right], \tag{10}$$

where y_s denotes the realized value. The real time optimal weights vector can be computed using a numerical solution algorithm.[11] However, it should be noted that maximizing the historical predictive likelihood does not necessarily guarantee a better out-of sample forecast accuracy than a naïve equal weighting scheme, since the approach introduces an additional estimation uncertainty regarding the weights. Moreover, historical weights might be a poor

approximation of current weights, particularly when structural breaks occur.

BFAVAR with one factor. It is often argued that factor augmented regression models are successful in achieving dimension reductions when forecasting macroeconomic time series (see, e.g., Barhoumi et al., 2013, Forni et al., 2003, Stock and Watson, 2002, Stock and Watson, 2005, Stock and Watson, 2006, Stock and Watson, 2011, among others). The idea is that the bulk of the variation in the indicator variables may be explained by a rather small number of factors that are added to a model with the variables of interest. Here, we consider a BFAVAR that includes HICP inflation, GDP growth and one factor. To estimate the parameters, we follow Bernanke, Boivin, and Eliasz (2005) and use a two-step approach. In the first step, we difference the indicator variables (if necessary) to achieve stationarity, then standardize them by subtracting the mean and dividing by the standard deviation. We then obtain the factor by extracting the first principal component from these standardized series. In the second step, we augment the BVAR by this factor and estimate the system using the Bayesian methods outlined above. As the BFAVAR system consists of only three variables, overparameterization is not a major problem, and little shrinkage is needed.

BFAVAR with three factors. Since the addition of one factor might not be sufficient to capture the dynamics in our dataset, we augment the BVAR from above with three factors, which are obtained by extracting the first three principal components from the standardized series. Given that a VAR with five variables is already large and is likely to suffer from overparameterization, we use Bayesian shrinkage to reduce the dimension of the system further. Note that both BFAVAR variants therefore combine the advantages of the factor approach with those of Bayesian shrinkage.

Large BVAR. In the last variant, we estimate the BVAR on all 22 series at the same time. The literature refers to this variant as a large BVAR (see, e.g., Bańbura et al., 2010, Giannone et al., 2015, Koop, 2013, among others). In order to deal with the dense parameterization of the model, it is estimated using Bayesian shrinkage. Given the large dimensionality of the system, we shrink a lot in this case.

Finally, the BVAR variants are accompanied by univariate $\mathrm{AR}(p_h)$ models for HICP inflation and GDP growth. Such models are the industry standard, and form another competitive benchmark in addition to the random walk variant. We estimate the $\mathrm{AR}(p_h)$ models using an uninformative prior, while the lag length p_h is selected optimally via the Bayesian information criterion (BIC), and may therefore differ across forecast horizons h.[12]

3.3. Selection of hyperparameters

There are various different ways of choosing the degree of shrinkage λ_h, and these will have an impact on the forecast performances of the BVAR variants. As was emphasized by De Mol, Giannone, and Reichlin (2008), one requirement for the hyperparameter λ_h is that the degree of shrinkage must increase with the number of parameters, i.e., $\lambda_h \to 0$ if $n \to \infty$.

Bańbura et al. (2010). One way to determine λ_h rests on the assumption that small VAR systems are parsimonious, and hence, do not suffer from overparameterization. The procedure follows Bańbura et al. (2010) by choosing λ_h such that the in-sample fit of a BVAR is the same as that of a parsimonious VAR. Specifically, we set the degree of shrinkage such that the average in-sample fit of each BVAR variant for HICP inflation and GDP growth during the training sample period 1975:2 to 1990:1 is identical to the average fit of the three-dimensional BVARs used for BVAR averaging. We obtain the desired magnitude of fit by performing a search over a fine grid for λ_h. This procedure ensures that we shrink more as the size of the model increases, and maintains comparability across variants. For that reason, we also apply Bayesian shrinkage to the BFAVAR with one factor.

We should emphasize that this procedure is not optimal, in the sense that the choice of λ_h is not based on any optimality criterion. Moreover, the shrinkage parameter does not evolve over time, but is kept fixed at the value obtained during the training sample period. Carriero et al. (2015) and Giannone et al. (2015) propose different procedures for overcoming these potential drawbacks. Both methods are straightforward to implement, and we

consider them for both BFAVAR variants, as well as for the large BVAR. In this context, we also discuss how the lag length p, which has been fixed at the frequency of the data so far, may be chosen optimally.

Carriero et al. (2015). One possibility is to choose the optimal degree of shrinkage $\lambda^*_{T,h}$ and the optimal lag length $p^*_{T,h}$ at each point in time by maximizing the (log) marginal likelihood of the model:

$$\left(\lambda^*_{T,h}, p^*_{T,h}\right) = \mathbf{argmax}\ \log p_h\left(y\right), \tag{11}$$

while the marginal likelihood $p_h\left(y\right)$ can be obtained in closed form[13] by integrating out the uncertainty about the parameters of the model θ_h, i.e.,

$$p_h\left(y\right) = \int p_h\left(y|\theta_h\right) p_h\left(\theta_h\right) d\theta_h. \tag{12}$$

We optimize over the two-dimensional grid $\lambda_h \in (0.01, 0.02, \ldots, 5)$ and $p_h \in (1, 2, \ldots, 8)$, and in each period choose the pair that maximizes $p_h\left(y\right)$. Note that the optimal value for λ_h decreases when more lags are included. Moreover, we update the remaining hyperparameters in each period accordingly (rolling training sample).

Giannone et al. (2015). Alternatively, λ_h can be treated as an additional unknown parameter and estimated in a hierarchical fashion. This procedure introduces more flexibility, since it also relaxes some of the simplifying choices made for the remaining hyperparameters. Both the main diagonal of the prior covariance matrix Σ_h and the hyperparameter τ_h are also treated as unknown. Thus, this variant requires an additional layer to be added to the prior structure by placing a prior on the hyperparameters—a hyperprior.

It can be shown that the marginal posterior for γ_h, i.e., after integrating out the uncertainty about the model's parameters, is:

$$p_h\left(\gamma_h|y\right) \propto p_h\left(y|\gamma_h\right) \cdot p_h\left(\gamma_h\right), \tag{13}$$

where $p_h\left(y|\gamma_h\right)$ is the marginal likelihood of the model (conditional on γ_h) and $p_h\left(\gamma_h\right)$ is the hyperprior. The latter reflects our confidence about the values of γ_h. We follow common practices and choose relatively diffuse hyperpriors. For λ_h and τ_h, we use Gamma densities with modes of 0.2 and 1 and standard deviations of 0.4 and 1, respectively. The prior on the diagonal elements of Σ_h is an Inverse-Gamma with scale and shape equal to 0.005^2.[14]

Since the joint posterior distribution for the predictive density and γ_h is not available in closed form, we simulate the distribution using a Metropolis–Hastings algorithm, which is outlined in detail in Appendix D. The sampler uses Markov chain Monte Carlo (MCMC) methods and generates γ_h from the marginal posterior with a Metropolis–Hastings update. After the sampler has converged, the forecasts can be drawn from the one-step-ahead predictive density conditional on γ_h. We make use of the fact that the marginal posterior is again available in closed form, conditional on a given draw for the hyperparameters, when forecasts are calculated using the direct-step approach.[15] To obtain a distribution of the forecasts, we draw sequentially a large number of times, and smooth the simulated histograms using a normal kernel function.[16]

In contrast to the procedures of Bańbura et al. (2010) and Carriero et al. (2015), this variant accounts for the uncertainty related to the choice of the hyperparameters, thus affecting the shape of the predictive density. Finally, it should be noted that the lag length is again set equal to four, and parsimony is achieved via the optimal estimation of the hyperparameters.

4. Forecast evaluation

In this section, we evaluate the performances of the BVAR variants in terms of their out-of-sample forecast accuracies. We begin by explaining our forecast experiment and discussing the choice of the shrinkage parameter across variants.

We then present results for the predictive mean and the entire predictive density of HICP inflation and GDP growth. Moreover, we assess how good the forecasts are in absolute terms.

4.1. Forecast experiment

We use each BVAR variant to produce out-of-sample forecasts for HICP inflation and GDP growth for two horizons ($h = 1$ and $h = 4$). For $h = 1$, we start with the sample period from 1976:1 to 1990:4 and generate the predictive density of each variant for 1991:1. We then iterate forward until 2011:3, producing forecasts for 2011:4. The procedure yields a sequence of 84 density forecasts for each variant. For $h = 4$, we start with the sample period from 1975:2 to 1990:1, also producing predictive densities for 1991:1, and iterate accordingly. Thus, the evaluation period runs from 1991:1 to 2011:4 for both horizons. Since our evaluation period is relatively long, we investigate the subsample stability of the results in a sensitivity analysis.

For the baseline specification, we prefer a rolling-window forecast scheme, which always uses the most recent 15 years of data, to a recursive (or expanding) scheme, which uses all of the past observations, since a rolling scheme may be better at handling the parameter instabilities that are likely to be present in aggregate euro area data. Moreover, it is consistent with the conventional view that more recent observations are more informative than those at the very beginning of a sample period. In addition, the statistical tests that we use for comparing forecast performances explicitly build on an asymptotically non-vanishing estimation uncertainty, an assumption that would be violated in an expanding-window forecast scheme (see Amisano and Giacomini, 2007, Giacomini and White, 2006). We then evaluate the sensitivity of the results to this setup and repeat the forecast experiment using a recursive scheme.

The choice of the size of the estimation window is motivated by two competing influences. On the one hand, the window must not be too small, or no meaningful estimation of the BVAR will be possible. On the other hand, the window must not be too large, or the sequence of density forecasts will be too short for inference. In particular, the likelihood ratio tests require relatively large numbers of observations. An estimation window of 15 years seems to account for both concerns well.

4.2. Degree of shrinkage

We document the choice of shrinkage parameters λ_h in the first ($h = 1$) and second ($h = 4$) columns of Table 1. The numbers in the first panel are the result of the calibration procedure proposed by Bańbura et al. (2010), which equates the average in-sample fits for HICP inflation and GDP growth across variants (except for the random walk benchmark, for which $\lambda_h = 0$). As we have emphasized above, we assume that the overparameterization problem does not occur for BVAR averaging, and use an uninformative prior ($\lambda_h = \infty$) in this case. For the BFAVAR with one factor, it turns out that parameter proliferation is not really an issue, and we shrink only a little (0.328 for $h = 1$ and 0.519 for $h = 4$). In the case of the BFAVAR with three factors, the prior becomes tighter, and λ_h is 0.105 and 0.151, respectively. For the large BVAR with all 22 variables, we have to impose a lot more shrinkage, and λ_1 reduces to 0.052 and λ_4 to 0.051.

Table 1. Root mean squared error.

	λ_h		HICP inflation		GDP growth	
	$h = 1$	$h = 4$	$h = 1$	$h = 4$	$h = 1$	$h = 4$
M01: Random walk	0		0.39	0.43	0.62	0.63
M02: BVAR averaging equal weights	∞		0.26	0.33	0.57	0.74
M03: BVAR averaging optimal weights	∞		0.26	0.33	0.52	0.75
M04: BFAVAR one factor	0.328	0.519	0.25	0.29	0.50	0.69

M05: BFAVAR three factors	0.105	0.151	0.26	0.28	0.48	0.65
M06: Large BVAR	0.052	0.051	0.26	0.30	0.45	0.63
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	0.26	0.29	0.52	0.66
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	0.26	0.29	0.48	0.67
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	0.26	0.28	0.44	0.65
M10: BFAVAR one factor, hyperpriors	0.367	0.272	0.26	0.30	0.50	0.66
M11: BFAVAR three factors, hyperpriors	0.314	0.271	0.25	0.30	0.46	0.68
M12: Large BVAR, hyperpriors	0.264	0.203	0.25	0.29	0.44	0.67
M13: $\mathbf{AR}(p_h)$, p_h optimal		n/a	0.27	0.31	0.56	0.64

Notes: the table shows the RMSEs for BVAR variants and an $\mathbf{AR}(p_h)$ model for which p_h is chosen by the Bayesian information criterion (BIC). The evaluation period runs from 1991:1 to 2011:4 and the forecast horizon h is in quarters. For $h = 4$, we use a direct-step forecast. The lower the RMSE is, the better the forecast accuracy of a model. The BVAR variants include the random walk, BVAR averaging with equal weights or optimal weights obtained by maximizing the historical predictive density, the BFAVAR with one and three factor(s), and the large BVAR with 22 variables. In the first panel, the hyperparameter λ_h is set such that the average in-sample fits for HICP inflation and GDP growth of all BVAR variants are the same during the training sample 1975:2 to 1990:1, while $p = 4$. In the second panel, p_h and λ_h are chosen by maximizing the marginal likelihood. We report the average choice for λ_h across periods. In the third panel, we impose a prior on the hyperparameters and estimate the variants in a hierarchical fashion, while $p = 4$. We report the average posterior mean for λ_h across periods.

In the second panel, λ_h is obtained by maximizing the marginal likelihood as per Carriero et al. (2015). Since the optimal choice for the tightness parameter varies over time (though actually only a little), we report the average choice for λ_h across periods. Given that the lag length p_h is determined simultaneously, the numbers are not directly comparable to those in the first panel. It turns out, however, that the optimal choice for p_h in most periods is one or two, which requires less shrinkage than a model with $p = 4$.

In the third panel, we document the average posterior mean for λ_h across periods when a hyperprior is used as per Giannone et al. (2015). The numbers are similar to those in the second panel, although the lag length is set to $p = 4$. However, in the latter case, the remaining hyperparameters are also estimated in real time, and therefore the numbers are hard to compare. In particular, the elements of Σ_h show a substantial degree of variation over time.

4.3. Predictive mean

We first evaluate the BVAR variants and the AR model based on the accuracy of their point forecasts.[17] Let the superscript m denote the predictive mean. We measure the out-of-sample forecast accuracy for model j with respect to variable i at horizon h in terms of the RMSE:

$$\mathrm{RMSE}_{i,h}^{j} = \sqrt{\frac{1}{T_1 - T_0 - H + 1} \sum_{T=T_0+H}^{T_1} \left(y_{i,T|T-h}^{m,j} - y_{i,T} \right)^2}, \tag{14}$$

where $H = 4$ is the maximal forecast horizon, and $T_0 + H = 64$ and $T_1 = 147$ are the start (1991:1) and end (2011:4) of the evaluation period, respectively. For a forecaster with a quadratic loss function, the predictive mean is the optimal forecast and the RMSE is an appropriate measure for discriminating among models (see Weiss, 1996). In Table 1, we report the RMSEs for the BVAR variants, as well as for the AR benchmark model.

With respect to the forecast accuracy of the BVAR variants, the following results emerge from Table 1. First, the random

walk benchmark produces the worst forecast for HICP inflation at both $h = 1$ and $h = 4$. When using a large BVAR, for instance, the forecaster could cut the RMSE by about one third at both horizons, independent of the way in which either the shrinkage parameter or the lag length is chosen. Substantial improvements compared to the random walk forecast are also possible with all other variants.[18] This finding may be either the result of the richer dynamics of the other BVAR variants compared to the random walk benchmark, or due to the additional information contained in the set of indicators. Given that the AR model is also inferior to almost all of the BVAR variants (though only slightly), we believe that the indicator variables do indeed contain valuable information regarding the prospective path of HICP inflation.

Second, we find only small differences in forecast accuracy among the BVAR variants for HICP inflation. Only for $h = 4$ does the BVAR averaging, with both equal and optimal weights, appear to perform worse than the other variants. Overall, the RMSEs range from 0.25 to 0.26 for $h = 1$, and from 0.28 to 0.33 for $h = 4$. In Table E.1 in Appendix E, we provide the results of Giacomini and White (2006) tests of conditional predictive ability for the point forecast. It turns out that the differences among the variants are insignificant in the majority of cases. Apparently, the way in which information is condensed within the BVAR setting has only a limited impact on the forecast accuracy of the predictive mean for HICP inflation. Moreover, the forecast accuracy hardly changes when the lag length, the hyperparameters, or the aggregation weights are chosen optimally.

Third, for GDP growth, all of the other models tend to outperform the random walk benchmark at $h = 1$. In contrast to HICP inflation, the random walk (jointly with one large BVAR variant) delivers the best forecast for GDP growth at $h = 4$, tracked closely by the AR model, suggesting that the set of indicators contains limited information about quarterly changes in the GDP one year ahead. This relatively poor forecast performance of most of the BVAR variants with respect to GDP growth in the longer run is probably due to the fact that the series shows little to no persistence, and thus, unpredictable shocks may account for large parts of the dynamics.

Fourth, the differences among the BVAR variants are somewhat larger than for HICP inflation. The RMSE varies between 0.44 and 0.57 for $h = 1$, and between 0.63 and 0.75 for $h = 4$. The large BVAR with p_h and λ_h optimized and the large BVAR with hyperpriors deliver the best forecast performances at $h = 1$. Both variants perform significantly better than the BVAR averaging (compare Table E.1 in Appendix E). However, the differences in predictive ability between the BVAR variants are insignificant in most cases. Overall, the three large BVAR variants appear to exhibit slight dominance, but the three-factor versions of the BFAVAR are also competitive.

4.4. Predictive density

In the previous subsection, we compared different models in terms of their RMSEs, which is appropriate if the forecaster is concerned only about the accuracy of the predictive mean, and indifferent to the uncertainty surrounding it. In this subsection, we relax this strong assumption and consider the entire density of the forecasts. In particular, we rank the BVAR variants based on predictive likelihoods, which is a convenient way of comparing models in a Bayesian setting and is standard in the literature (see, e.g., D'Agostino et al., 2013, Geweke and Amisano, 2010, Giannone et al., 2015, Koop, 2013, among others). The larger the predictive likelihood, the higher the probability that a variant assigns to the realized value of the variable given the history of the data (y) and the parameters (β_h, Σ_h).

In Table 2, we document the average of the log predictive likelihood of model j with respect to variable i at horizon h, evaluated at the realization $y_{i,T}$:

$$\mathrm{PL}_{i,h}^{j} = \frac{1}{T_1 - T_0 - H + 1} \sum_{T=T_0+H}^{T_1} \log p_h \left(y_{i,T} \middle| y, j, h \right). \tag{15}$$

Table 2 reveals that the random walk produces the worst forecasts for HICP inflation at both horizons, which is in line with our results for the point forecasts. However, we find that the ranking of the BVAR variants changes when the focus shifts from the predictive mean to the entire density of the forecasts. Specifically, the performance of the large BVAR

deteriorates, so that, in most cases, it is now not much better than the AR model, if at all. Despite its quite accurate mean forecast, the large BVAR apparently attaches too low a probability to events that actually occur. Even when the lag length and the shrinkage parameter are chosen optimally, or a prior is imposed on the hyperparameters, the large BVAR variants perform worse than the competing approaches. In contrast, the BFAVAR with three factors maintains its good relative performance when the forecast uncertainty is taken into account. It now delivers the best forecasts at both horizons, while the BFAVAR with only one factor is also competitive. For both BFAVARs, we find that choosing the lag length or the shrinkage parameter in an optimal way does not help to improve the predictive ability. Notably, BVAR averaging, with both equal and optimal weights, which performs worse than the competing approaches in terms of RMSEs, now ranks between the BFAVARs and the large BVARs. Optimizing the aggregation weights with respect to the in-sample forecast accuracy does not improve the out-of-sample predictive ability.

Table 2. Predictive likelihood.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		−1.87	−1.94	−1.96	−2.32
M02: BVAR averaging equal weights	∞		−1.69	−1.65	−1.58	−1.99
M03: BVAR averaging optimal weights	∞		−1.71	−1.66	−1.55	−2.01
M04: BFAVAR one factor	0.328	0.519	−1.59	−1.61	−1.55	−2.02
M05: BFAVAR three factors	0.105	0.151	−1.53	−1.60	−1.53	−2.06
M06: Large BVAR	0.052	0.051	−1.68	−1.78	−1.67	−2.34
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	−1.62	−1.63	−1.60	−2.08
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	−1.65	−1.64	−1.59	−2.10
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	−1.87	−1.77	−1.72	−2.51
M10: BFAVAR one factor, hyperpriors	0.367	0.272	−1.61	−1.58	−1.52	−1.94
M11: BFAVAR three factors, hyperpriors	0.314	0.271	−1.58	−1.65	−1.53	−2.07
M12: Large BVAR, hyperpriors	0.264	0.203	−1.70	−1.78	−1.60	−2.30
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		−1.80	−1.72	−1.61	−1.98

Notes: the table shows the average log predictive likelihood for BVAR variants and an $\mathbf{AR}(p_h)$ model. The higher the average log predictive likelihood, the better the predictive ability of a model. See also the notes to Table 1.

The random walk and the large BVAR jointly display the worst performances for GDP growth at both horizons. As in the case of HICP inflation, the ranking among the BVAR variants is reversed relative to the ranking based on RMSEs. Again, the predictive density of the large BVAR attaches too low a probability to events that actually occur, and taking the uncertainty regarding the prior choice into account does little to overcome this problem. In contrast, both BVAR averaging variants, which produced rather inaccurate mean forecasts, now perform rather well, and appear to dominate many of the other approaches at $h=4$. The best predictive ability at both horizons, however, is displayed by the BFAVAR with one factor when a hyperprior is imposed. Nevertheless, relative to the shrinkage procedure used by Bańbura et al. (2010), the additional gain from estimating the hyperparameters of the BFAVAR is small.

In what follows, we investigate whether the predictive likelihoods reported in Table 2 differ from each other

significantly. We follow Amisano and Giacomini (2007) and construct an average of likelihood ratios for two competing density forecasts j and k:

$$\mathrm{LR}_{i,h}^{j,k} = \frac{1}{T_1 - T_0 - H + 1} \sum_{T=T_0+H}^{T_1} \left(\log p_h \left(y_{i,T} | y, j, h \right) - \log p_h \left(y_{i,T} | y, k, h \right) \right). \tag{16}$$

The test is a conventional likelihood ratio test that is often used for model selection (see, e.g., Giannone et\special {t4ht=&} al., 2015, among others), and is based on the statistic

$$t = \frac{\mathrm{LR}_{i,h}^{j,k}}{\widehat{\sigma} / \sqrt{T_1 - T_0 - H + 1}}, \tag{17}$$

where $\widehat{\sigma}$ is an estimate of the standard deviation of $\mathrm{LR}_{i,h}^{j,k}$. The null hypothesis of equal performances of j and k is rejected at the 5% level whenever $|t| > 1.96$. In the case of rejection, one would choose j (k) whenever $\mathrm{LR}_{i,h}^{j,k}$ is positive (negative). Pairwise likelihood ratio tests for HICP inflation and GDP growth are given in Table 3, Table 4, where a $+$ $(-)$ indicates that the model under consideration, in the column, performs significantly better (worse) than its benchmark model in the row in terms of predictive ability, while a $\cdot$ denotes insignificance.

Table 3. Likelihood ratio tests: HICP inflation.

Benchmark	h	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
M02	1	$\cdot$											
M03		$\cdot$	$\cdot$										
M04		$-$	$\cdot$	$-$									
M05		$-$	$-$	$-$	$\cdot$								
M06		$-$	$\cdot$	$\cdot$	$\cdot$	$+$							
M07		$-$	$\cdot$	$\cdot$	$\cdot$	$+$	$\cdot$						
M08		$\cdot$	$\cdot$	$\cdot$	$\cdot$	$+$	$\cdot$	$\cdot$					
M09		$\cdot$	$+$	$\cdot$	$+$	$+$	$+$	$+$	$+$				
M10		$-$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-$			
M11		$-$	$\cdot$	$-$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-$	$-$	$\cdot$		
M12		$\cdot$	$\cdot$	$\cdot$	$+$	$+$	$\cdot$	$\cdot$	$\cdot$	$-$	$\cdot$	$+$	
M13		$\cdot$	$\cdot$	$\cdot$	$+$	$+$	$\cdot$	$+$	$\cdot$	$\cdot$	$+$	$+$	$\cdot$
M02	4	$-$											
M03		$-$	$\cdot$										
M04		$-$	$\cdot$	$\cdot$									
M05		$-$	$\cdot$	$\cdot$	$\cdot$								
M06		$-$	$\cdot$	$\cdot$	$+$	$+$							
M07		$-$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-$						
M08		$-$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-$	$\cdot$					
M09		$-$	$\cdot$	$\cdot$	$+$	$+$	$\cdot$	$+$	$+$				
M10		$-$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$-$	$\cdot$	$\cdot$	$-$			

M11	−	·	·	·	·	−	·	·	−	·		
M12	−	·	·	+	+	·	+	+	·	+	+	
M13	−	·	·	+	·	·	·	·	·	+	·	·

Notes: the table shows the results of pairwise likelihood ratio tests. + (−) indicates that the model under consideration, in the column, performs significantly better (worse) than its benchmark in the row in terms of predictive ability. · indicates insignificance. The significance level is 5%. The models are: M01 = Random walk; M02 = BVAR averaging equal weights; M03 = BVAR averaging optimal weights; M04 = BFAVAR one factor; M05 = BFAVAR three factors; M06 = Large BVAR; M07 = BFAVAR one factor, p_h and λ_h optimal; M08 = BFAVAR three factors, p_h and λ_h optimal; M09 = Large BVAR, p_h and λ_h optimal; M10 = BFAVAR one factor, hyperpriors; M11 = BFAVAR three factors, hyperpriors; M12 = Large BVAR, hyperpriors; M13 = $\mathrm{AR}(p_h)$, p_h optimal. See also the notes to Table 1.

Table 4. Likelihood ratio tests: GDP growth.

Benchmark	h	**M01**	**M02**	**M03**	**M04**	**M05**	**M06**	**M07**	**M08**	**M09**	**M10**	**M11**	**M12**
M02	1	·											
M03		·	·										
M04		·	·	·									
M05		·	·	·	·								
M06		·	·	·	·	+							
M07		·	·	·	·	·	·						
M08		·	·	·	·	·	·	·					
M09		·	·	·	+	+	·	·	·				
M10		·	·	·	·	·	·	·	·	−			
M11		·	·	·	·	·	·	·	·	−	·		
M12		·	·	·	·	·	·	·	·	·	·	·	
M13		·	·	·	·	·	·	·	·	·	·	·	·
M02	4	·											
M03		·	·										
M04		·	·	·									
M05		·	·	·	·								
M06		·	·	·	·	·							
M07		·	·	·	·	·	−						
M08		·	·	·	·	·	·	·					
M09		+	·	·	·	+	+	+	+				
M10		·	·	·	·	·	·	·	−	−			
M11		·	·	·	·	·	·	·	·	−	·		
M12		·	·	·	·	+	·	+	+	·	+	+	

M13	–	·	·	·	–	–	–	–	–	·	·	–

Notes: see the notes to Table 1, Table 3.

Table 3 reveals that, when forecasting HICP inflation, the random walk and the AR model perform poorly, and none of the BVAR variants are outperformed significantly by these benchmark models. Making use of the set of indicator variables is apparently beneficial. For $h = 1$, we find that both BFAVARs, when combined with a shrinkage procedure as per Bańbura et al. (2010), significantly outperform most of their competitors. Neither optimizing the lag length and tightness of the prior nor estimating the hyperparameters improves the forecast ability further. Similarly, the differences between the two BVAR averaging variants are insignificant. Finally, the large BVAR with p_h and λ_h optimal, as per Carriero et al. (2015), is dominated by nearly all of its competitors at the short horizon, including the large BVAR with the shrinkage procedure as per Bańbura et al. (2010). Similar results are obtained for $h = 4$ as for $h = 1$. Overall, the BFAVARs tend to outperform the other approaches significantly, whereas the large BVAR variants never dominate the competing variants.

For GDP growth, the picture is different. Table 4 shows that none of the BVAR variants are able to outperform either the AR or the random walk significantly. Moreover, the differences among almost all of the BVAR variants are insignificant. As for HICP inflation, the large BVAR with p_h and λ_h optimized appears to deliver a poor forecast for GDP growth, and is significantly worse than the BFAVAR combined with the shrinkage approach of Bańbura et al. (2010). At $h = 4$, GDP growth is apparently difficult to forecast, since the AR is significantly better than most BVAR variants, and we thus conclude that the indicators contain little or no information about GDP growth four quarters ahead. However, we do find that both large BVAR variants with optimally chosen hyperparameters are outperformed significantly by all of the BFAVAR variants, suggesting that neither choosing the lag length and the degree of shrinkage in an optimal fashion nor estimating the hyperparameters helps to improve the forecasting ability of the large BVAR. Overall, BFAVARs deliver good forecast performances across variables and horizons, while the large BVARs are sometimes outperformed by but never outperform any of the other BVAR variants.

As an alternative to evaluating predictive densities, recent studies have proposed the CRPS, which gives a greater reward when a high probability is given to values that are near but not identical to the realization, and attaches less weight to extreme observations than the predictive likelihood (see, e.g., Clark and Ravazzolo, 2015, Ravazzolo and Vahey, 2014, among others).

For a predictive density, the CRPS is obtained by integrating over the squared deviation between its cumulative density function (CDF) and the CDF of the realization:

$$\mathrm{CRPS}_{i,h,T}^{j} = \int_{-\infty}^{\infty} \left(P_h\left(u|y,j,h\right) - \mathbb{I}\left[y_{i,T|h} \leq u\right]\right)^2 du, \quad \text{for } T = T_0 + H, \ldots, T_1, \tag{18}$$

where $P_h\left(\cdot\right)$ denotes the CDF.[19] The CDF of the realization is given by the indicator function $\mathbb{I}\left[\cdot\right]$. The lower the CRPS is, the better the predictive ability of a model. To rank the BVAR variants, we use the average CRPS

$$\overline{\mathrm{CRPS}}_{i,h}^{j} = \frac{1}{T_1 - T_0 - H + 1} \sum_{T=T_0+H}^{T_1} \mathrm{CRPS}_{i,h,T}^{j}. \tag{19}$$

The average CRPS for each variant and horizon is reported in Table 5. For HICP inflation, the random walk is still the worst variant at both horizons, while BVAR averaging and the BFAVARs remain competitive. Unlike the predictive likelihood, the CRPS suggests that using hyperpriors helps to reduce the loss of the large BVAR. Even though both the large BVAR with the shrinkage procedure as per Bańbura et al. (2010) and that with p_h and λ_h optimized are still dominated by the BFAVAR with three factors, the large BVAR of Giannone et al. (2015) shows a better relative performance in terms of CRPS than predictive likelihoods. For GDP growth, the CRPS is even more in favor of the large BVAR. For both horizons, all three variants show relatively low scores, which suggests that the unfavorable relative

performances of these variants in terms of predictive likelihoods are due, at least to some extent, to large but infrequent forecast errors.

Table 5. Continuous ranked probability score.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		0.220	0.244	0.298	0.310
M02: BVAR averaging equal weights	∞		0.142	0.186	0.293	0.371
M03: BVAR averaging optimal weights	∞		0.145	0.185	0.268	0.378
M04: BFAVAR one factor	0.328	0.519	0.133	0.165	0.263	0.354
M05: BFAVAR three factors	0.105	0.151	0.138	0.158	0.251	0.332
M06: Large BVAR	0.052	0.051	0.144	0.171	0.233	0.314
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	0.140	0.167	0.264	0.332
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	0.140	0.161	0.252	0.344
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	0.146	0.159	0.236	0.330
M10: BFAVAR one factor, hyperpriors	0.367	0.272	0.143	0.170	0.261	0.338
M11: BFAVAR three factors, hyperpriors	0.314	0.271	0.136	0.167	0.248	0.355
M12: Large BVAR, hyperpriors	0.264	0.203	0.131	0.164	0.236	0.339
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.147	0.174	0.277	0.317

Notes: the table shows the average CRPS for BVAR variants and an $\mathbf{AR}(p_h)$ model. The lower the average CRPS, the better the predictive ability of a model. See also the notes to Table 1.

4.5. Probability integral transform

While the relative performances of Bayesian forecast models are commonly compared on the basis of predictive likelihoods and CRPS, we also use the probability integral transform (PIT) to assess how good the predictive densities are in absolute terms. In contrast to predictive likelihoods, which are a local measure, since they depend only on the predictive density evaluated at the realization, the PIT provides a non-local assessment of the predictive ability of a forecast model (see Geweke & Amisano, 2010). Given that recent work by Rossi and Sehkposyan (2014) has suggested that the predictive densities generated by BVARs are often not a reasonable description of the actual uncertainty, we investigate which specification choices help to achieve correct calibration.

For a predictive density, the PIT is the value obtained when the corresponding CDF is evaluated at the realization:[20]

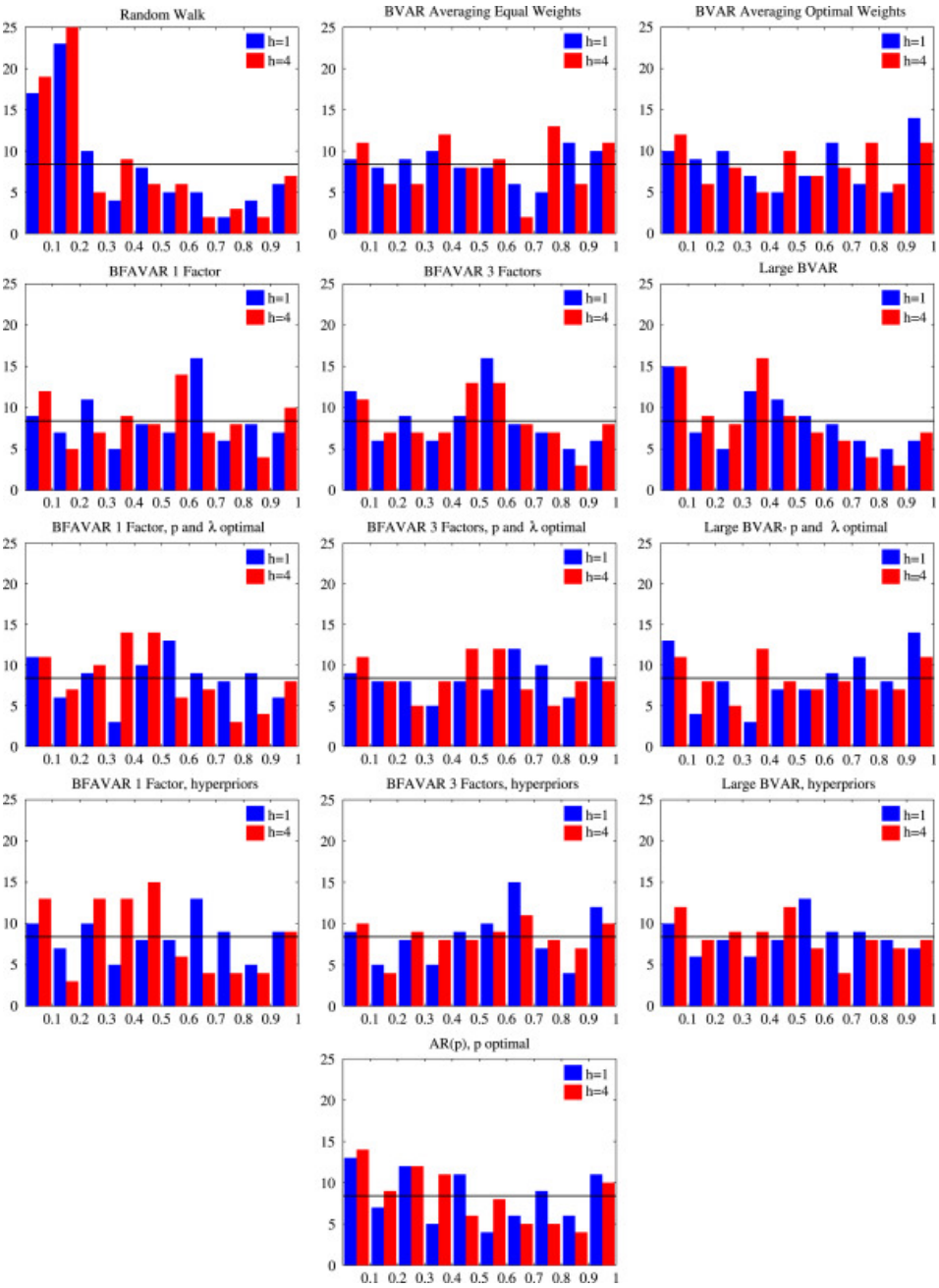
$$z_{i,h,T} = \int_{-\infty}^{y_{i,T}} p_h\left(u|y,j,h\right) du = P_h\left(y_{i,T}|y,j,h\right), \quad \text{for } T = T_0 + H, \ldots, T_1. \tag{20}$$

If the predictive density is consistent with the unobserved true density, $\{z_{i,h,T}\}_{T=T_0+H}^{T_1}$ is independently and identically distributed, $\mathbf{U}(0,1)$, and its CDF is the 45° line.

In practice, we decompose each predictive density into intervals, each of which covers 10% of the probability mass. This is done for any given point in time, and thus the intervals may change over time. We then assign each realization to the corresponding interval. For a predictive density to be a reasonable description of the true distribution of the data, each

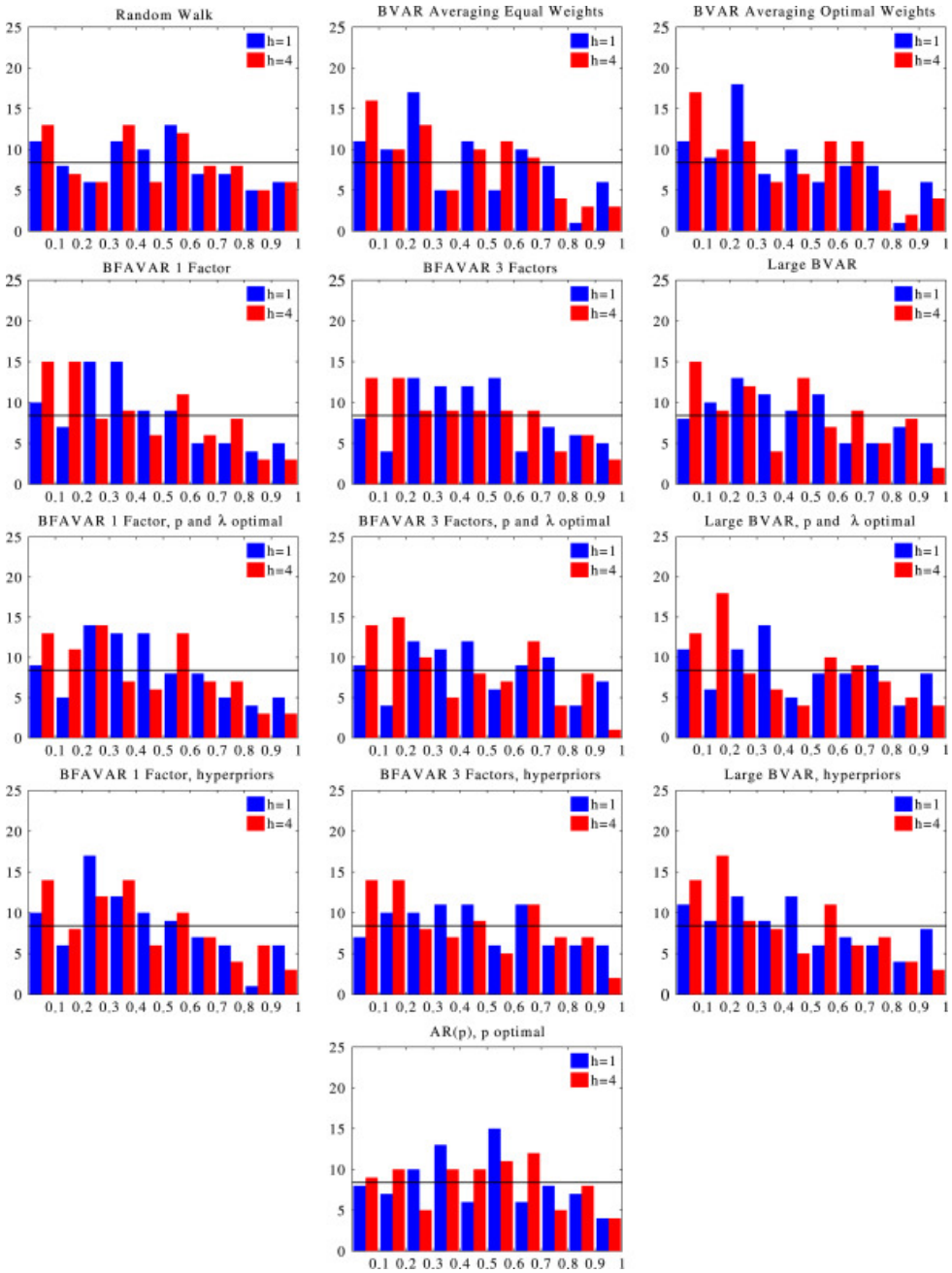
probability band should contain 10% of the realizations (about 8 in our case).

We provide a graphical assessment of the PITs for HICP inflation and GDP growth in Fig. 1, Fig. 2, respectively. Following Rossi and Sehkposyan (2014), we also use the Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) tests to check formally for the uniformity of the distribution of the PITs. Both tests build on the absolute difference between the CDF of a PIT and the CDF of a uniform distribution, i.e., the 45° line. We document their outcomes in Appendix E.



Download : Download full-size image

Fig. 1. Probability integral transform: HICP inflation. Notes: the figure shows the PITs for BVAR variants and an $\mathrm{AR}(p_h)$ model to check for the correct calibration of the predictive density. The horizontal line denotes uniformity. x-axis: probability band; y-axis: number of realizations.



Download : Download full-size image

Fig. 2. Probability integral transform: GDP growth. Notes: the figure shows the PITs for BVAR variants and an $\mathbf{AR}(p_h)$ model to check for the correct calibration of the predictive density. The horizontal line denotes uniformity. x-axis: probability band; y-axis: number of realizations.

In the case of HICP inflation, we obtain support for the conclusions drawn in the previous subsections. Again, the performance of the random walk variant is poor, and Fig. 1 shows that too little weight is given to the lower tail of the distribution. In contrast, most of the BVAR variants, together with the AR model, produce predictive densities that appear to be reasonable descriptions of the true distribution of the data. However, we reject uniformity of the PITs for the large BVAR with shrinkage proposed by Bańbura et al. (2010). Such a result is in line with its unsatisfactory performance in terms of predictive likelihoods and CRPS. The plot of the PITs suggests that the large BVAR attaches too low a probability to very low or even negative inflation rates, while, at the same time, too much weight is given to the upper tails of the distribution of inflation.[21] However, when we take the uncertainty surrounding the choice of the

hyperparameters into account, we cannot reject uniformity for the large BVAR at the 5% level, which suggests that using hyperpriors helps to generate a more realistic density forecast, matching its good relative performance according to CRPS. The BFAVAR with one factor, combined with shrinkage as per Bańbura et al. (2010), appears to be well calibrated at both horizons, whereas optimizing the lag length and the degree of shrinkage, or imposing a hyperprior, leads to a rejection of correct calibration at $h = 4$. Moreover, we again find that all BFAVAR variants with three factors have good predictive abilities for HICP inflation at both horizons.

Given the evidence from the previous subsections, it is not surprising that our findings for GDP growth are different. With only few exceptions, uniformity is rejected by both tests at the 5% level. A notable exception is the BFAVAR with three factors ($h = 1$), where the optimal choice of the hyperparameters helps to achieve a realistic density forecast. Similarly, the large BVAR appears to benefit slightly from optimally chosen hyperparameters. A visual assessment of the plots in Fig. 2 suggests that most BVAR variants appear to generate too little mass in the lower tail of the distribution, since too many realizations seem to fall into the bins covering the first 10%–20% of the probability mass. That is, they underestimate the occurrence of low or even negative GDP growth rates.

5. Sensitivity analysis

In the previous section, we showed how the rankings of the BVARs change as we focus on different target variables, forecast horizons, and loss functions. We also explored which specification choices help to achieve correct calibration. In this section, we assess the sensitivity of our findings with respect to the following features of the forecast experiment. First, we compare the direct-step and iterative forecast approaches for $h = 4$. Second, we consider the full dataset. Third, we use a recursive rather than a rolling estimation scheme. Fourth, we explore subsample stability. Finally, we consider variables in (log-)levels rather than first (log-)differences. The respective tables and figures are all provided in Appendix E.

Iterative approach ($h = 4$). The baseline results are generated using the direct-step forecast approach, which is convenient because the one-step-ahead predictive density is available in closed-form for most variants, and the computational burden is small. However, for $h = 4$, the direct-step approach is an approximation. In this sensitivity check, we compute exact forecasts for $h = 4$ using the iterative approach, which requires Monte Carlo integration for all BVARs, and provide the full set of results. For all other checks, we focus on RMSEs and predictive likelihoods only.

While the accuracy of the predictive mean is similar across the two methods, almost all BVAR variants benefit from the iterative approach in terms of predictive likelihoods. Nevertheless, the large BVAR with shrinkage, as per Bańbura et al. (2010), remains the worst variant for both HICP inflation and GDP growth, at $h = 4$. However, the large BVAR profits substantially from iterative forecasting when p_h and λ_h are optimized or hyperparameters are estimated. In fact, the former shows the highest predictive likelihood for HICP inflation, and hence, we conclude that the large BVARs suffer most from the approximation incorporated in the direct-step approach for $h = 4$. At the same time, it should be noted that the CRPS values of the direct-step and iterative forecast approaches are similar for large BVARs. In regard to the absolute forecast accuracy, the PITs suggest that using hyperpriors helps to produce realistic predictive densities for all variants at $h = 4$.

Full dataset. In our baseline specification, we use a subset of 22 variables from our dataset, neglecting indicators that may potentially be useful for forecasting HICP inflation and GDP growth. In this sensitivity check, we repeat the forecast experiment with the full cross section of 44 series.

While the accuracy of the predictive mean hardly changes as the size of the cross section increases, the variants are all affected differently when we turn to predictive densities. For instance, the accuracies of the BFAVARs and BVAR averaging improve, in terms of predictive likelihoods. The former benefit because a larger cross section helps in the identification of the factors, without increasing the number of parameters to be estimated. The latter benefit from the

larger numbers of density forecasts that enter the prediction pool, making it less vulnerable to extreme predictions. In contrast, the performances of large BVARs worsen, which is consistent with the related literature that finds that BVARs with about 20–25 variables are generally superior to larger systems. Likewise, our results suggest that large BVARs cannot profit from an enlarged information set. It also appears that the additional gains from the larger cross section are more pronounced for HICP inflation than for GDP growth, which is not surprising, since our previous results showed that indicator variables are more relevant for the former than for the latter.

Recursive estimation scheme. Our baseline specification involves a rolling estimation scheme, which is advisable when the data generating process is subject to structural breaks, since the estimated parameters adjust quickly over time. When the underlying process is stable, however, discarding distant observations introduces an additional level of estimation uncertainty. In this sensitivity check, we consider a recursive scheme.

All BVAR variants benefit in terms of predictive likelihoods, but not in terms of RMSEs. Moreover, the predictive abilities of the large BVARs improve more than those of the other variants. The large BVAR with shrinkage as per Bańbura et al. (2010) displays the highest predictive likelihood for HICP inflation at both horizons. However, the variants with λ_h and p_h optimized and with hyperpriors still perform worse than most of their competitors. When GDP growth is considered, the ranking of the BVAR variants is hardly affected. The large BVARs are better than BVAR averaging for $h = 1$, while the BFAVARs with three factors again show the highest predictive likelihood. For $h = 4$, the predictive ability of the large BVARs remains the worst, which may be related to the direct-step approach, as discussed above.

Subsample stability. We focus on the evaluation period 1999:1 to 2011:4, excluding the inflation convergence of the 1990s, which changed the properties of the inflation process. Moreover, we begin with the start of the currency union in January 1999 and mimic the situation encountered by researchers and policymakers when forecasting the euro area in real time.

The relative performances of the BVARs in terms of both RMSEs and predictive likelihoods are not affected when the evaluation period is shortened. The only exception is the random walk variant, which shows a lower RMSE than before with respect to HICP inflation. Given that the predictive mean of that variant adapts slowly over time, it is not surprising that excluding the inflation convergence reduces its forecast error.

(Log-)level specification. In this sensitivity check, we consider all variables in (log-)levels rather than first (log-)differences, and impose the random walk prior ($\delta_i = 1$). Since no closed form solution for the marginal posterior of the predictive density for HICP inflation and GDP growth is available in this case, we evaluate the respective log-levels.

The rankings of the BVARs are not too sensitive to the transformation of the variables. For the HICP and $h = 1$, the differences among the variants in terms of RMSEs remain small, while the BVAR averaging shows an improved relative forecast accuracy for $h = 4$. The differences in RMSEs for GDP are small as well. However, one exception should be noted. While the large BVAR with hyperpriors remains among the best performing variants, the large BVARs with shrinkage as per Bańbura et al. (2010) or with p_h and λ_h optimized perform worse when we use (log-)levels. Turning to predictive likelihoods, we again find that all large BVARs are outperformed by the other variants, while the BFAVARs perform well. Choosing the optimal p_h and λ_h or using hyperpriors does not improve the forecast performance. In fact, the large BVAR with p_h and λ_h optimized has the lowest predictive likelihood for both variables and horizons.

6. Conclusion

Recent articles have suggested that VARs with Bayesian shrinkage produce accurate point and density forecasts even when the number of variables becomes large. However, there are numerous different ways to implement a BVAR, and these specification choices affect the predictive ability. In this paper, we consider BVAR averaging with equal and optimal weights, BFAVARs with one and three factors, and large BVARs. Furthermore, we use the three predominant

approaches to select hyperparameters, and the shrinkage parameter in particular. In order to assess how efficiently the variants process the information contained in a large dataset, we condition our analysis on a given amount of information, and rank the variants according to their out-of-sample forecast accuracies.

We evaluate the predictive mean, and find that all of the BVARs outperform the random walk variant for HICP inflation, with the differences in forecast accuracy between them being small and often insignificant. For GDP growth, the large BVARs deliver the best forecast at $h = 1$, but cannot improve on the random walk at $h = 4$. Moreover, we find that neither setting the shrinkage hyperparameter optimally nor estimating it improves the forecast accuracy compared to the simple shrinkage procedure of Bańbura et al. (2010). Likewise, choosing the aggregation weights for BVAR averaging in an optimal way is not superior to an equal weighting scheme.

We find that an accurate predictive mean does not necessarily imply a high predictive likelihood, and *vice versa*. Hence, model selection based on RMSEs only is misleading if the forecaster is also concerned with the uncertainty. In particular, we show that the relative performances of the large BVARs worsen when the predictive likelihood is considered, while those of BVAR averaging improve. Evaluating the forecast performances based on CRPS, which puts less weight on large but infrequent forecast errors, suggests that large BVARs suffer most from extreme observations, but perform well most of the time. BFAVARs perform satisfactorily across target variables and horizons, while we again find that neither choosing the degree of shrinkage optimally nor estimating it improves the relative forecast accuracy.

Calibration tests suggest that most BVARs produce predictive densities that describe the true distribution reasonably well for HICP inflation, but they detect a poor calibration of all variants with respect to GDP growth. The use of hyperpriors helps to generate more realistic density forecasts. For future research, it might be fruitful to assess the performances of more flexible models. For instance, the introduction of stochastic volatility for accounting explicitly for changes in the shock process could help to achieve better calibrations. Moreover, modeling time-varying parameters may improve the forecast accuracy in unstable environments. The work of Barnett et al. (2014), Carriero et al., 2012, Carriero et al., 2013 and Clark (2011) goes in this direction.

In a sensitivity analysis, we show that the density forecast accuracy of large BVARs in particular depends on the features of the forecast experiment. While all variants benefit from the iterative forecast approach for $h = 4$, large BVARs profit most. When combined with the iterative approach, optimizing or estimating hyperparameters is also beneficial. Moreover, we demonstrate that the relative performances of large BVARs improve considerably in terms of predictive likelihoods when a recursive scheme is used and more observations enter the estimation.

In contrast, large BVARs do not benefit from an enlarged cross section, a result that is in line with the findings of Koop (2013) for US data. For BVARs with a NIW prior, he reports that predictive likelihoods decline for cross sections with 168 variables, while RMSEs are not affected. For the euro area, we show that large BVARs already deliver unfavorable density forecasts for systems of 44 variables, probably due to the fact that fewer time series observations are available. We conclude that deriving density forecasts from VARs based on overly short samples might be dangerous even when Bayesian shrinkage is used. When only a limited time series dimension is at hand, a preselection device, which determines the (subset of) variables that enter the information set, could therefore reduce the cross section and improve the density forecast accuracy. Note that variable preselection is not required for BFAVARs or BVAR averaging, since their parameter space does not increase with the size of the information set.

According to the evidence in this paper, there is no best BVAR variant. In particular, the loss function of the forecaster is relevant for the relative forecast performance. Large BVARs are an excellent choice under certain specifications (a recursive window, 22 variables, or an iterative approach with optimal or estimated shrinkage), while a BFAVAR with three factors is a safe bet, since it is competitive in most setups, and especially when the cross section is large. This last finding is consistent with the recent work of Kim and Swanson (2014), who also find that so-called “hybrid” models, which combine data reduction with shrinkage methods, deliver accurate out-of-sample forecasts.

Acknowledgments

We thank George Athanasopoulos (editor), Peter Zorn, two referees, and participants at the 1st Workshop on High Dimensional Time Series in Macroeconomics and Finance 2013 (Vienna), the 67th European Meeting of the Econometric Society 2013 (Gothenburg), the Annual Meeting of the German Economic Association 2013 (Düsseldorf), the Workshop on Uncertainty and Probabilistic Forecasting 2014 (Heidelberg), the 35th International Symposium on Forecasting 2014 (Rotterdam), and the Ifo Macro Seminar (Munich) for their helpful comments and suggestions. Financial support from the *German Research Foundation* (Grant No. CA 833/2) is gratefully acknowledged.

Appendix A. Description of the dataset

See Table A.1.

Table A.1. The dataset comprises 44 quarterly euro area time series covering the period 1975:1 to 2011:4. The columns are: series number, series mnemonic used by the database, series label, series category, the original data source (AWM = Area-wide Model; OECD: MEI = Organisation for Economic Co-operation and Development: Main Economic Indicators), and the series transformation (1 = log-difference; 2 = difference). The last column indicates whether or not the series is included in the baseline specification (BS). The AWM series come from the database of the Euro Area Business Cycle Network (12th update), while the remaining series are obtained from Datastream.

No.	Mnemonic	Label	Category	Source	Transf.	BS
01	YER	Real GDP	National accounts	AWM	1	×
02	PCR	Real Private Consumption	National accounts	AWM	1	×
03	GCR	Real Government Consumption	National accounts	AWM	1	
04	ITR	Real Gross Investment	National accounts	AWM	1	×
05	XTR	Real Exports of Goods and Services	National accounts	AWM	1	
06	MTR	Real Imports of Goods and Services	National accounts	AWM	1	
07	YFN	GDP at Factor Costs	National accounts	AWM	1	
08	WIN	Compensation to Employees	National accounts	AWM	1	
09	GON	Gross Operating Surplus	National accounts	AWM	1	
10	TIN	Indirect Taxes (net of subsidies)	National accounts	AWM	1	
11	YIN	GDP Income Side	National accounts	AWM	1	
12	NFNYEN	Net Factor Income from Abroad/GDP	National accounts	AWM	2	
13	SAX	Household's Savings Ratio	National accounts	AWM	2	
14	HICP	Overall HICP	Price indexes	AWM	1	×
15	YED	GDP Deflator	Price indexes	AWM	1	×
16	PCD	Private Consumption Deflator	Price indexes	AWM	1	×
17	GCD	Government Consumption Deflator	Price indexes	AWM	1	
18	ITD	Gross Investment Deflator	Price indexes	AWM	1	×
19	XTD	Exports of Goods and Services Deflator	Price indexes	AWM	1	
20	MTD	Imports of Goods and Services Deflator	Price indexes	AWM	1	

21	YFD	GDP at Factor Costs Deflator	Price indexes	AWM	1	
22	YWR	Real World GDP	International	AWM	1	×
23	YWRX	Real World Demand	International	AWM	1	
24	YWD	World GDP Deflator	International	AWM	1	×
25	COMPR	Commodity Prices	International	AWM	1	
26	PCOMU	Non-Oil Commodity Prices	International	AWM	1	×
27	POILU	Oil Prices	International	AWM	1	×
28	LFN	Labor Force	Employment	AWM	1	
29	LNN	Total Employment	Employment	AWM	1	×
30	LEN	Employees	Employment	AWM	1	
31	UNN	Number of Unemployed	Employment	AWM	1	
32	URX	Unemployment Rate	Employment	AWM	2	×
33	LPROD	Labor Productivity	Employment	AWM	1	
34	ULC	Unit Labor Costs	Employment	AWM	1	
35	WRN	Wages	Employment	AWM	1	
36	EKOL2002Q	Composite Leading Indicator	Surveys	OECD: MEI	2	×
37	EKOCS002Q	Consumer Confidence Indicator	Surveys	OECD: MEI	2	×
38	EKQMA027B	M1 Money Stock	Monetary aggregates	OECD: MEI	1	×
39	EKQMA013B	M3 Money Stock	Monetary aggregates	OECD: MEI	1	×
40	STN	Short-Term Interest Rate	Financial	AWM	2	×
41	LTN	Long-Term Interest Rate	Financial	AWM	2	×
42	EMSHRPRCF	Share Price Index	Financial	Thomson Reuters	1	×
43	EEN	Nominal Effective Exchange Rate	Financial	AWM	1	×
44	EXR	Euro per U.S.D. Exchange Rate	Financial	AWM	1	×

Table E.1. Test of equal predictive ability: HICP inflation.

	h	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
M02	1	–											
M03		–	·										
M04		–	·	·									
M05		–	·	·	·								
M06		–	·	·	·	·							
M07		–	·	·	·	·	·						

M08		−	·	·	·	·	·	·					
M09		−	·	·	·	·	·	·	·				
M10		−	·	·	·	·	·	·	·	·			
M11		−	·	·	·	·	·	·	·	·	·		
M12		−	·	·	·	·	·	·	·	·	·	·	
M13		−	·	·	·	·	·	·	·	·	·	·	·
M02	4	−											
M03		−	·										
M04		−	·	·									
M05		−	·	·	·								
M06		−	·	·	·	+							
M07		−	·	·	·	·	·						
M08		−	·	·	·	+	·	−					
M09		−	·	·	−	+	·	−	·				
M10		−	·	·	+	·	·	·	+	·			
M11		−	·	·	·	·	−	+	·	·	·		
M12		−	·	·	·	·	·	·	·	·	−	·	
M13		−	·	·	+	·	·	·	·	+	·	+	·

Notes: the table shows the results of a Giacomini and White (2006) test of equal conditional predictive ability. The test is based on the squared forecast error loss. $+$ $(-)$ indicates that the model under consideration, in the column, performs significantly (at a 5% level) better (worse) than its benchmark in the row, in terms of RMSEs. $\cdot$ indicates insignificance. The models are: M01 = Random walk; M02 = BVAR averaging equal weights; M03 = BVAR averaging optimal weights; M04 = BFAVAR one factor; M05 = BFAVAR three factors; M06 = Large BVAR; M07 = BFAVAR one factor, p_h and λ_h optimal; M08 = BFAVAR three factors, p_h and λ_h optimal; M09 = Large BVAR, p_h and λ_h optimal; M10 = BFAVAR one factor, hyperpriors; M11 = BFAVAR three factors, hyperpriors; M12 = Large BVAR, hyperpriors; M13 = $\mathbf{AR}(p_h)$, p_h optimal. See also the notes to Table 1.

Appendix B. Marginal likelihood

According to Carriero et al. (2015), the analytical expression for the marginal likelihood is:

$$p_h(y) = k_h^{-1} \times \left|\Psi_h + (y - x_h b_h)' \left(I + x_h \Omega_h x_h'\right)^{-1} (y - x_h b_h)\right|^{-\frac{T+d_h}{2}}, \tag{B.1}$$

where

$$k_h = \pi^{\frac{TN}{2}} \times \left|\left(I + x_h \Omega_h x_h'\right)^{-1}\right|^{-\frac{N}{2}} \times \left|\Psi_h\right|^{-\frac{d_h}{2}} \frac{\Gamma_N\left(\frac{d_h}{2}\right)}{\Gamma_N\left(\frac{T+d_h}{2}\right)}, \tag{B.2}$$

and $\Gamma_N(\cdot)$ denotes the N-variate gamma function. See Giannone et al. (2015) for a similar expression.

Appendix C. Monte Carlo integration

The following algorithm can be used to obtain multistep forecasts for the iterative approach. Let H denote the maximal forecast horizon and N the number of draws. Set $h = 1$ in Eqs. (1), (2), (3), (4), (5), (6) and fix γ_1 at a particular value. For $j = 1, \ldots, N$, run a loop over the following steps:

Step 1: Draw $\Sigma_1^j \middle| \gamma_1, y$ and $\beta_1^j \middle| \Sigma_1^j, \gamma_1, y$ from their conditional posterior, given by Eq. (6).

Step 2: Generate a shock sequence $u_{1,T+1}^j, \ldots, u_{1,T+H}^j$ from $u_{1,t} \sim \mathrm{N}\left(0, \Sigma_1^j\right)$.

Step 3: Compute the 1-step-ahead forecast as

$$\hat{y}_{T+1}^j = c_1^j + B_{1,1}^j y_T + \cdots + B_{p,1}^j y_{T-p+1} + u_{1,T+1}^j, \tag{C.1}$$

while the h-step-ahead forecasts are obtained by iteration:

$$\hat{y}_{T+h}^j = c_1^j + B_{1,1}^j \hat{y}_{T+h-1}^j + \cdots + B_{p,1}^j \hat{y}_{T+h-p}^j + u_{1,T+h}^j, \tag{C.2}$$

where $h = 1, \ldots, H$ and $\hat{y}_{T+h}^j = y_{T+h-p}$ for $h \leq p$. □

This algorithm yields $\left(\hat{y}_{T+1}^j, \ldots, \hat{y}_{T+H}^j\right)_{j=1}^N$ as an independent sample of draws from the predictive density. We set $H = 4$ and $N = 10{,}000$. The predictive density is obtained by smoothing the simulated histograms using a normal kernel function.[22]

Appendix D. Markov chain Monte Carlo estimation

This appendix outlines the Metropolis–Hastings algorithm that we use to simulate the joint posterior of the predictive density and the hyperparameters γ_h. Moreover, we assess the convergence of the Markov chain using inefficiency factors.

D.1. Metropolis–Hastings algorithm: direct-step approach

The following algorithm can be used to obtain forecasts for the direct-step approach. Let N_B be the number of burn-in-draws and N_R the number of retained draws. The steps are:

Step 1: Choose a starting point γ_h^0. We use the posterior mode, which is obtained by numerical optimization.[23] For $j = 1, \ldots, N_B + N_R$, run a loop over the following steps.

Step 2: Draw a proposal γ_h^* from a jumping distribution $\mathrm{J}_h\left(\gamma_h^* \middle| \gamma_h^{j-1}\right) = \mathrm{N}\left(\gamma_h^{j-1}, c \cdot \Sigma_{m,h}\right)$, where $\Sigma_{m,h}$ is the inverse of the Hessian computed at the posterior mode, and c is a scaling constant that is chosen to obtain an acceptance ratio of about 20%.

Step 3: Compute the acceptance ratio:

$$r_h = \frac{p_h(\gamma_h^* | y)}{p_h\left(\gamma_h^{j-1} | y\right)}, \tag{D.1}$$

where $p_h(\cdot)$ is the marginal posterior for γ_h. See also Eq. (13).

Step 4: Randomly draw ν from $\mathrm{U}(0, 1)$.

Step 5: Accept or discard the proposal γ_h^* according to the following rule, and update the jumping distribution if necessary:

$$\gamma_h^j = \begin{cases} \gamma_h^* & : \text{if } \nu \leq r_h \\ \gamma_h^{j-1} & : \text{otherwise .} \end{cases} \tag{D.2}$$

If $j \leq N_B$, repeat the previous four steps, otherwise continue.

Step 6: Draw $y_{T+1,h}^j$ from $p_h^j\left(y_{T+1,h}\middle|\gamma_h^j, y\right)$, which is the predictive density conditional on γ_h^j, for which again a closed form solution is available; i.e., the forecasts conditional on γ_h^j can be drawn from their marginal posterior (i.e., after integrating out the uncertainty about Σ_h), which is multivariate-t, with the mean given by Eq. (7) and the covariance matrix by Eq. (8). □

This sampler yields $\left(y_{T+1,h}^j\right)_{j=N_B+1}^{N_B+N_R}$ as an independent sample of draws from the joint predictive density. We set $N_B = N_R = 10{,}000$. The predictive density is obtained by smoothing the simulated histograms using a normal kernel function. The simulations were performed on a standard computer (2.4 GHz, 2 GB RAM) with Matlab, and took about 48 h for the large BVAR and three hours for the BFAVARs per horizon h (baseline specification with 22 variables and 84 forecast periods).

D.2. Metropolis–Hastings algorithm: iterative approach

The following algorithm can be used to obtain multistep forecasts for the iterative approach. Let H denote the maximal forecast horizon, N_B the number of burn-in-draws, and N_R the number of retained draws. Set $h = 1$ in Eqs. (1), (2), (3), (4), (5), (6). The steps are:

Step 1: Choose a starting point γ_1^0. We use the posterior mode, which is obtained by numerical optimization. For $j = 1, \ldots, N_B + N_R$, run a loop over the following steps.

Step 2: Draw a proposal γ_1^* from a jumping distribution $\mathrm{J}_1\left(\gamma_1^*\middle|\gamma_1^{j-1}\right) = \mathrm{N}\left(\gamma_1^{j-1}, c \cdot \Sigma_{m,1}\right)$, where $\Sigma_{m,1}$ is the inverse of the Hessian computed at the posterior mode, and c is a scaling constant that is chosen to obtain an acceptance ratio of about 20%.

Step 3: Compute the acceptance ratio:

$$r_1 = \frac{p_1(\gamma_1^*|y)}{p_1\left(\gamma_1^{j-1}|y\right)}, \tag{D.3}$$

where $p_1(\cdot)$ is the marginal posterior for γ_1. See also Eq. (13).

Step 4: Randomly draw ν from $\mathrm{U}(0, 1)$.

Step 5: Accept or discard the proposal γ_1^* according to the following rule, and update the jumping distribution if necessary:

$$\gamma_1^j = \begin{cases} \gamma_1^* & : \text{if } \nu \leq r_1 \\ \gamma_1^{j-1} & : \text{otherwise .} \end{cases} \tag{D.4}$$

If $j \leq N_B$, repeat the previous four steps, otherwise continue.

Step 6: Draw $\Sigma_1^j\middle|\gamma_1^j, y$ and $\beta_1^j\middle|\Sigma_1^j, \gamma_1^j, y$ from their conditional posterior given by Eq. (6).

Step 7: Generate a shock sequence $u_{1,T+1}^j, \ldots, u_{1,T+H}^j$ from $u_{1,t} \sim \mathrm{N}\left(0, \Sigma_1^j\right)$.

Step 8: Compute the 1-step-ahead forecast as

$$\hat{y}_{T+1}^{j} = c_1^j + B_{1,1}^j y_T + \cdots + B_{p,1}^j y_{T-p+1} + u_{1,T+1}^j, \tag{D.5}$$

while the h-step-ahead forecasts are obtained by iteration:

$$\hat{y}_{T+h}^{j} = c_1^j + B_{1,1}^j \hat{y}_{T+h-1}^j + \cdots + B_{p,1}^j \hat{y}_{T+h-p}^j + u_{1,T+h}^j, \tag{D.6}$$

where $h = 1, \ldots, H$ and $\hat{y}_{T+h}^j = y_{T+h-p}$ for $h \leq p$. □

This sampler yields $\left(\hat{y}_{T+1}^j, \ldots, \hat{y}_{T+H}^j\right)_{j=N_B+1}^{N_B+N_R}$ as an independent sample of draws from the joint predictive density. We set $H = 4$ and $N_B = N_R = 10{,}000$. The predictive density is obtained by smoothing the simulated histograms using a normal kernel function. The simulations took about 120 h for the large BVAR and eight hours for the BFAVARs (baseline specification with 22 variables, 84 forecast periods).

D.3. Convergence

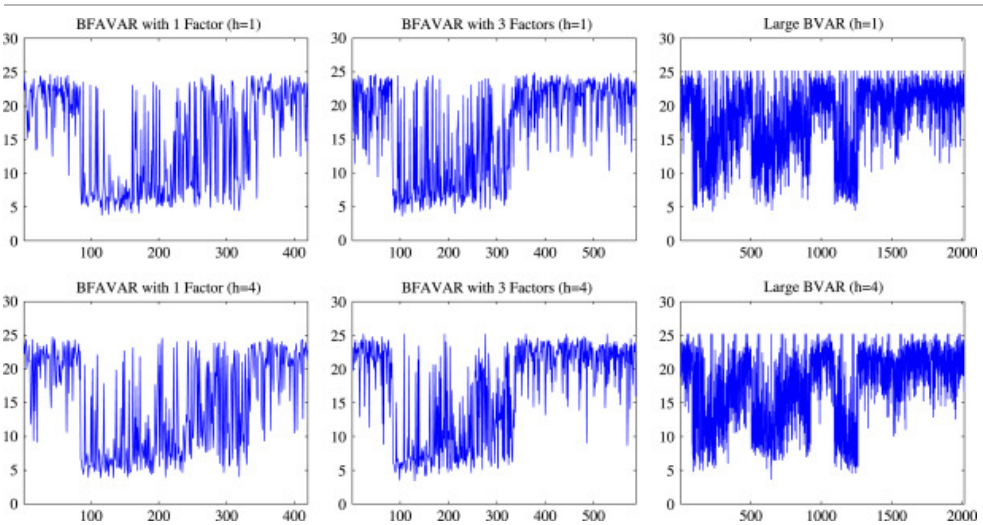
In order to check for the convergence of the Markov chain, we follow Primiceri (2005) and calculate inefficiency factors (IFs) for all of the hyperparameters of each BVAR variant (we have a total of 420 for the BFAVAR with one factor, 588 for the BFAVAR with three factors, and 2016 for the large BVAR). The IF is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the estimated autocorrelation of the chain at lag s. Since independence sampling produces an IF that is equal to one and dependence sampling typically produces an IF that is greater than one, the IF quantifies the relative efficiency loss in the computation of posterior draws from dependent versus independent samples. In practice, values of around 20 are regarded as efficient (see, e.g., Primiceri, 2005, among others), meaning that the econometrician needs to draw 20 times as many draws as from uncorrelated samples.

Following Berg (2015), we calculate the IFs as the inverse of the relative numerical efficiency measure (RNE) of Geweke (1992):

$$\mathrm{RNE} = (2\pi)^{-1} \frac{1}{\mathrm{S}(0)} \int_{-\pi}^{\pi} \mathrm{S}(\omega)\, d\omega, \tag{D.7}$$

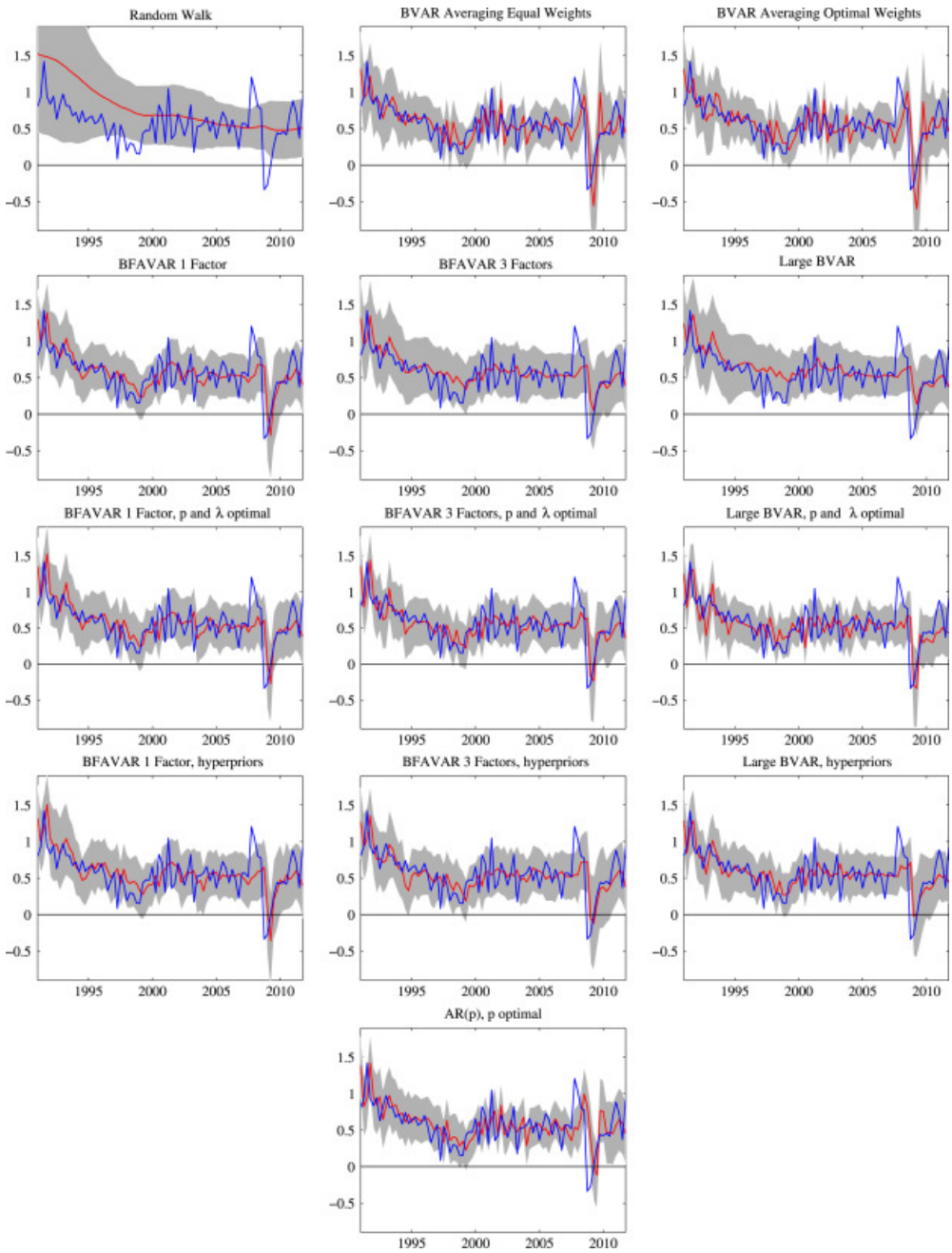
where $\mathrm{S}(\omega)$ denotes the spectral density of the sequence of draws at frequency ω. We estimate the spectral densities by smoothing the periodograms in the frequency domain using a 4% tapered window, as per Primiceri (2005).

In Fig. D.1, we document the IFs for the three BVAR variants at both forecast horizons. The numbers all range between five and 25, providing some evidence that the posterior draws for the hyperparameters come from a converged distribution. Even the large BVAR with more than 2000 hyperparameters shows IFs that are only slightly higher than those for the BFAVARs. All in all, we regard these numbers as satisfactory.



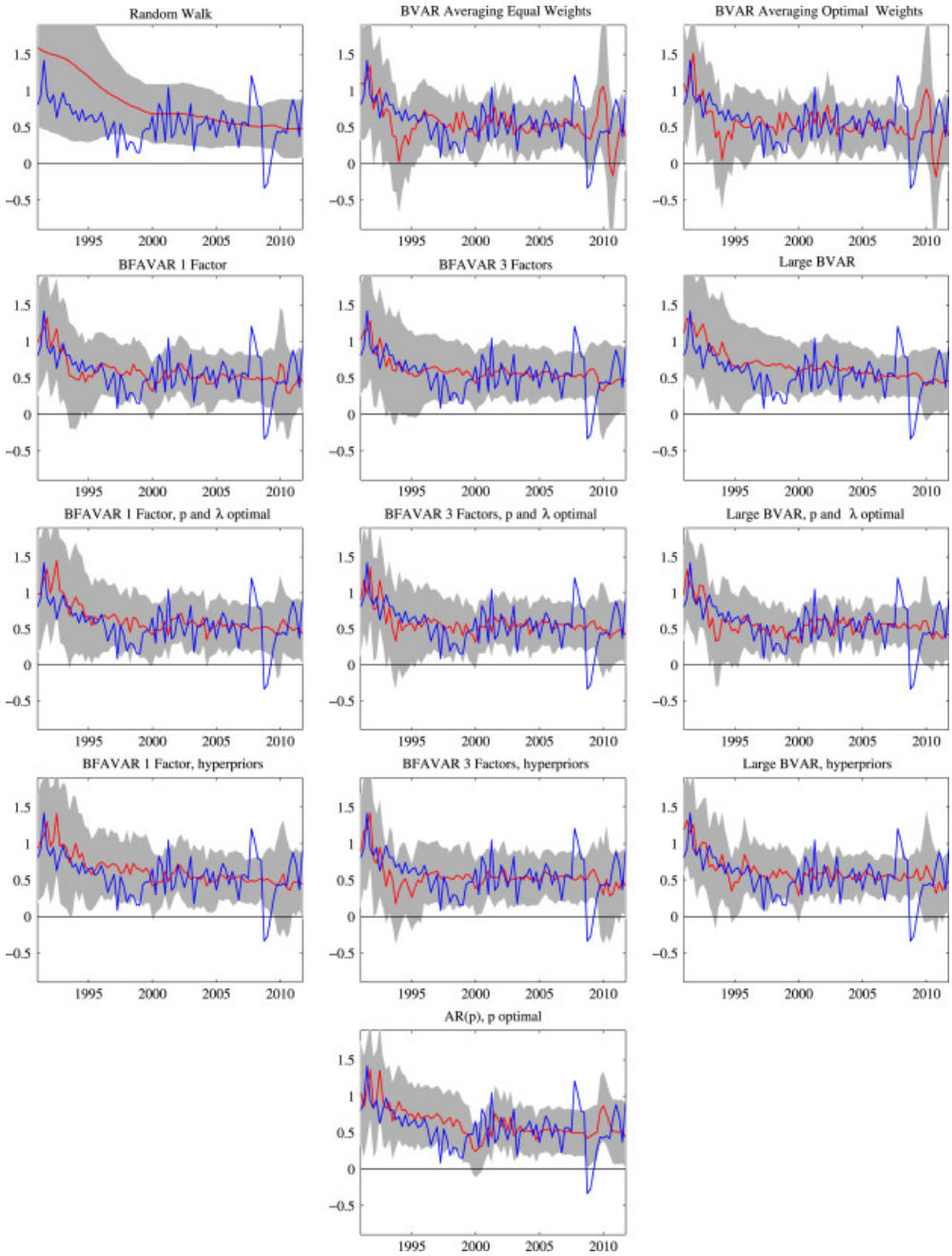
Download : Download full-size image

Fig. D.1. Inefficiency factors. Notes: the figure shows the IFs for all of the hyperparameters of the BVARs to check for convergence in the Markov chain. x-axis: hyperparameter; y-axis: IF.



Download : Download full-size image

Fig. E.1. Forecasts and realizations: HICP inflation ($h = 1$). Notes: the figure shows point forecasts with 90% forecast intervals for BVAR variants and an $\mathrm{AR}(p_h)$ model.



Download : Download full-size image

Fig. E.2. Forecasts and realizations: HICP inflation ($h = 4$). Notes: the figure shows point forecasts with 90% forecast intervals for BVAR variants and an $\mathbf{AR}(p_h)$ model.

Appendix E. Additional tables and figures

In this appendix, we provide additional material for our paper. In particular, we document the results of a Giacomini and White (2006) test of equal conditional predictive ability of the point forecast for HICP inflation and GDP growth. Moreover, we show the outcomes of both Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) tests of uniformity of the probability integral transform (PIT). We also provide tables showing the results of our sensitivity analysis. We report the root mean squared error (RMSE) and predictive likelihood when (a) forecasts are derived from the iterative approach for $h = 4$; (b) the full dataset of 44 variables is considered; (c) a recursive estimation scheme is used; (d) the evaluation period is 1999:1 to 2011:4; and (e) the variables enter the models in (log-)levels instead of first (log-

)differences. For the iterative approach, we also report the continuous ranked probability score (CRPS) and PITs for $h = 4$. Finally, we show time series plots of the forecasts for HICP inflation and GDP growth. See the main text for further details (see Table E.2, Table E.3, Table E.4, Table E.5, Table E.6, Table E.7, Table E.8, Table E.9, Table E.10, Table E.11, Table E.12, Table E.13 and Fig. E.3, Fig. E.4, Fig. E.5).

Table E.2. Test of equal predictive ability: GDP growth.

Benchmark	h	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
M02	1	·											
M03		·	−										
M04		·	·	·									
M05		·	·	·	·								
M06		·	·	·	·	·							
M07		·	·	·	·	·	·						
M08		·	·	·	·	·	·	·					
M09		·	−	−	·	·	·	·	−				
M10		·	·	·	·	·	·	·	·	·			
M11		·	·	·	·	·	·	·	·	·	·		
M12		·	−	−	·	·	·	·	·	·	·	·	
M13		·	·	·	·	·	·	·	·	·	·	·	·
M02	4	·											
M03		·	·										
M04		·	·	·									
M05		·	·	·	·								
M06		·	·	·	·	·							
M07		·	·	·	·	·	·						
M08		·	·	·	·	·	·	·					
M09		·	·	·	·	·	·	·	·				
M10		·	·	·	·	·	·	·	·	·			
M11		+	·	·	·	·	·	·	·	·	·		
M12		·	·	·	·	·	·	·	·	·	·	·	
M13		·	·	·	·	·	·	·	·	·	·	−	·

Notes: See the notes to Table 1, Table E.1.

Table E.3. Anderson–Darling (AD) test of uniformity.

	λ_h	HICP inflation	GDP growth

	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		0.00	0.00	0.02	0.01
M02: BVAR averaging equal weights	∞		0.55	0.63	0.01	0.00
M03: BVAR averaging optimal weights	∞		0.49	0.60	0.01	0.00
M04: BFAVAR one factor	0.328	0.519	0.64	0.54	0.00	0.00
M05: BFAVAR three factors	0.105	0.151	0.21	0.19	0.06	0.00
M06: Large BVAR	0.052	0.051	0.03	0.00	0.02	0.00
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	0.54	0.02	0.01	0.00
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	0.38	0.45	0.27	0.00
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	0.03	0.39	0.11	0.00
M10: BFAVAR one factor, hyperpriors	0.367	0.272	0.71	0.03	0.00	0.00
M11: BFAVAR three factors, hyperpriors	0.314	0.271	0.37	0.52	0.18	0.00
M12: Large BVAR, hyperpriors	0.264	0.203	0.32	0.09	0.04	0.00
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.19	0.01	0.12	0.04

Notes: the table shows the p-values of an AD-test of uniformity of the PITs. See also the notes to Table 1.

Table E.4. Kolmogorov–Smirnov (KS) test of uniformity.

	λ_h		**HICP inflation**		**GDP growth**	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		0.00	0.00	0.18	0.25
M02: BVAR averaging equal weights	∞		0.57	0.86	0.02	0.00
M03: BVAR averaging optimal weights	∞		0.56	0.90	0.01	0.01
M04: BFAVAR one factor	0.328	0.519	0.63	0.69	0.00	0.00
M05: BFAVAR three factors	0.105	0.151	0.18	0.21	0.05	0.03
M06: Large BVAR	0.052	0.051	0.06	0.00	0.05	0.02
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	0.90	0.01	0.00	0.01
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	0.42	0.57	0.63	0.01
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	0.16	0.95	0.15	0.01
M10: BFAVAR one factor, hyperpriors	0.367	0.272	0.97	0.01	0.00	0.00
M11: BFAVAR three factors, hyperpriors	0.314	0.271	0.21	0.77	0.24	0.01
M12: Large BVAR, hyperpriors	0.264	0.203	0.75	0.26	0.09	0.00
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.19	0.02	0.15	0.24

Notes: the table shows the p-values of a KS-test of uniformity of the PITs. See also the notes to Table 1.

Table E.5. Iterative approach ($h = 4$).

	RMSE		Predictive likelihood		CRPS	
	HICP	GDP	HICP	GDP	HICP	GDP
M01: Random walk	0.42	0.64	−1.72	−2.05	0.237	0.317
M02: BVAR averaging equal weights	0.32	0.68	−1.56	−1.77	0.181	0.350
M03: BVAR averaging optimal weights	0.36	0.71	−1.58	−1.83	0.189	0.364
M04: BFAVAR one factor	0.31	0.64	−1.60	−1.90	0.173	0.327
M05: BFAVAR three factors	0.31	0.64	−1.62	−2.04	0.175	0.317
M06: Large BVAR	0.33	0.64	−1.66	−2.13	0.188	0.319
M07: BFAVAR one factor, p_h and λ_h optimal	0.33	0.80	−1.60	−2.03	0.187	0.370
M08: BFAVAR three factors, p_h and λ_h optimal	0.29	0.66	−1.53	−1.96	0.171	0.337
M09: Large BVAR, p_h and λ_h optimal	0.31	0.65	−1.49	−1.95	0.178	0.331
M10: BFAVAR one factor, hyperpriors	0.32	0.68	−1.65	−2.01	0.182	0.338
M11: BFAVAR three factors, hyperpriors	0.26	0.66	−1.56	−1.96	0.164	0.347
M12: Large BVAR, hyperpriors	0.30	0.67	−1.52	−2.01	0.171	0.343
M13: $\mathbf{AR}(p_h)$, p_h optimal	0.32	0.85	−1.90	−2.19	0.159	0.323

Notes: the table shows the RMSEs, average log predictive likelihoods, and CRPS values for BVAR variants and an $\mathbf{AR}(p_h)$ model for $h = 4$ when forecasts are obtained using the iterative approach. See also the notes to Table 1.

Table E.6. Root mean squared error: full dataset.

	λ_h		HICP inflation		GDP growth	
	$h = 1$	$h = 4$	$h = 1$	$h = 4$	$h = 1$	$h = 4$
M01: Random walk	0		0.39	0.43	0.62	0.63
M02: BVAR averaging equal weights	∞		0.26	0.33	0.59	0.75
M03: BVAR averaging optimal weights	∞		0.27	0.31	0.52	0.72
M04: BFAVAR one factor	0.147	0.368	0.24	0.27	0.54	0.66
M05: BFAVAR three factors	0.094	0.170	0.24	0.28	0.52	0.64
M06: Large BVAR	0.040	0.039	0.25	0.29	0.47	0.64
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.291	0.25	0.28	0.57	0.66
M08: BFAVAR three factors, p_h and λ_h optimal	0.371	0.273	0.25	0.28	0.52	0.65
M09: Large BVAR, p_h and λ_h optimal	0.175	0.139	0.26	0.28	0.46	0.66
M10: BFAVAR one factor, hyperpriors	0.357	0.268	0.25	0.27	0.56	0.67
M11: BFAVAR three factors, hyperpriors	0.318	0.250	0.24	0.29	0.53	0.68

M12: Large BVAR, hyperpriors	0.080	0.043	0.25	0.29	0.49	0.66
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.27	0.31	0.56	0.64

Notes: the table shows RMSEs when the full dataset is used. See also the notes to Table 1.

Table E.7. Predictive likelihood: full dataset.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		−2.14	−2.24	−2.26	−2.72
M02: BVAR averaging equal weights	∞		−1.69	−1.65	−1.60	−1.97
M03: BVAR averaging optimal weights	∞		−1.73	−1.63	−1.55	−1.97
M04: BFAVAR one factor	0.147	0.368	−1.49	−1.57	−1.59	−1.99
M05: BFAVAR three factors	0.094	0.170	−1.48	−1.59	−1.56	−2.07
M06: Large BVAR	0.040	0.039	−1.87	−2.00	−1.90	−2.79
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.291	−1.59	−1.60	−1.68	−2.08
M08: BFAVAR three factors, p_h and λ_h optimal	0.371	0.273	−1.66	−1.61	−1.63	−2.10
M09: Large BVAR, p_h and λ_h optimal	0.175	0.139	−2.18	−1.98	−1.99	−3.06
M10: BFAVAR one factor, hyperpriors	0.357	0.268	−1.51	−1.57	−1.62	−1.92
M11: BFAVAR three factors, hyperpriors	0.318	0.250	−1.53	−1.58	−1.59	−2.15
M12: Large BVAR, hyperpriors	0.080	0.043	−1.80	−1.94	−1.96	−3.07
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		−1.80	−1.72	−1.61	−1.98

Notes: the table shows the average log predictive likelihood when the full dataset is used. See also the notes to Table 1, Table 2.

Table E.8. Root mean squared error: recursive estimation scheme.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		0.66	0.72	0.63	0.63
M02: BVAR averaging equal weights	∞		0.27	0.38	0.56	0.66
M03: BVAR averaging optimal weights	∞		0.27	0.35	0.50	0.65
M04: BFAVAR one factor	0.328	0.519	0.27	0.37	0.53	0.67
M05: BFAVAR three factors	0.105	0.151	0.26	0.32	0.48	0.64
M06: Large BVAR	0.052	0.051	0.26	0.31	0.43	0.64
M07: BFAVAR one factor, p_h and λ_h optimal	0.362	0.295	0.29	0.36	0.53	0.66
M08: BFAVAR three factors, p_h and λ_h optimal	0.324	0.261	0.27	0.34	0.46	0.64

M09: Large BVAR, p_h and λ_h optimal	0.151	0.119	0.28	0.31	0.43	0.63
M10: BFAVAR one factor, hyperpriors	0.298	0.241	0.28	0.35	0.53	0.66
M11: BFAVAR three factors, hyperpriors	0.275	0.265	0.28	0.33	0.44	0.66
M12: Large BVAR, hyperpriors	0.266	0.183	0.27	0.32	0.43	0.64
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.28	0.37	0.56	0.64

Notes: the table shows RMSEs when a recursive estimation scheme is used. See also the notes to Table 1.

Table E.9. Predictive likelihood: recursive estimation scheme.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		−1.43	−1.50	−1.71	−1.80
M02: BVAR averaging equal weights	∞		−1.53	−1.44	−1.45	−1.74
M03: BVAR averaging optimal weights	∞		−1.53	−1.41	−1.45	−1.76
M04: BFAVAR one factor	0.328	0.519	−1.55	−1.42	−1.43	−1.78
M05: BFAVAR three factors	0.105	0.151	−1.37	−1.31	−1.37	−1.75
M06: Large BVAR	0.052	0.051	−1.33	−1.25	−1.41	−1.86
M07: BFAVAR one factor, p_h and λ_h optimal	0.362	0.295	−1.59	−1.35	−1.43	−1.74
M08: BFAVAR three factors, p_h and λ_h optimal	0.324	0.261	−1.56	−1.38	−1.38	−1.73
M09: Large BVAR, p_h and λ_h optimal	0.151	0.119	−1.66	−1.35	−1.43	−1.87
M10: BFAVAR one factor, hyperpriors	0.298	0.241	−1.52	−1.35	−1.42	−1.72
M11: BFAVAR three factors, hyperpriors	0.275	0.265	−1.57	−1.39	−1.31	−1.78
M12: Large BVAR, hyperpriors	0.266	0.183	−1.60	−1.40	−1.42	−1.94
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		−1.61	−1.44	−1.46	−1.70

Notes: the table shows the average log predictive likelihood when a recursive estimation scheme is used. See also the notes to Table 1, Table 2.

Table E.10. Root mean squared error: evaluation period 1999:1 to 2011:4.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		0.30	0.31	0.68	0.70
M02: BVAR averaging equal weights	∞		0.29	0.36	0.57	0.79
M03: BVAR averaging optimal weights	∞		0.29	0.36	0.49	0.80
M04: BFAVAR one factor	0.328	0.519	0.27	0.31	0.46	0.72

M05: BFAVAR three factors	0.105	0.151	0.27	0.31	0.49	0.69
M06: Large BVAR	0.052	0.051	0.28	0.30	0.46	0.67
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	0.28	0.31	0.51	0.71
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	0.28	0.31	0.49	0.71
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	0.29	0.30	0.44	0.69
M10: BFAVAR one factor, hyperpriors	0.367	0.272	0.28	0.31	0.47	0.71
M11: BFAVAR three factors, hyperpriors	0.314	0.271	0.28	0.31	0.47	0.73
M12: Large BVAR, hyperpriors	0.264	0.203	0.27	0.31	0.41	0.71
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		0.30	0.32	0.60	0.70

Notes: the table shows RMSEs when the evaluation period is shortened. See also the notes to Table 1.

Table E.11. Predictive likelihood: evaluation period 1999:1 to 2011:4.

	λ_h		HICP inflation		GDP growth	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk	0		−1.91	−2.00	−2.29	−2.83
M02: BVAR averaging equal weights	∞		−1.89	−1.83	−1.60	−2.21
M03: BVAR averaging optimal weights	∞		−1.90	−1.87	−1.53	−2.22
M04: BFAVAR one factor	0.328	0.519	−1.74	−1.81	−1.62	−2.29
M05: BFAVAR three factors	0.105	0.151	−1.68	−1.83	−1.65	−2.38
M06: Large BVAR	0.052	0.051	−1.87	−2.02	−1.84	−2.77
M07: BFAVAR one factor, p_h and λ_h optimal	0.356	0.267	−1.75	−1.84	−1.73	−2.43
M08: BFAVAR three factors, p_h and λ_h optimal	0.316	0.258	−1.80	−1.89	−1.76	−2.44
M09: Large BVAR, p_h and λ_h optimal	0.218	0.171	−2.08	−2.08	−1.86	−3.00
M10: BFAVAR one factor, hyperpriors	0.367	0.272	−1.75	−1.76	−1.58	−2.22
M11: BFAVAR three factors, hyperpriors	0.314	0.271	−1.71	−1.85	−1.67	−2.41
M12: Large BVAR, hyperpriors	0.264	0.203	−1.89	−2.06	−1.67	−2.70
M13: $\mathbf{AR}(p_h)$, p_h optimal	n/a		−2.03	−1.89	−1.80	−2.33

Notes: the table shows the average log predictive likelihood when the evaluation period is shortened. See also the notes to Table 1, Table 2.

Table E.12. Root mean squared error: (log-)level specification.

	λ_h		HICP		GDP	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$

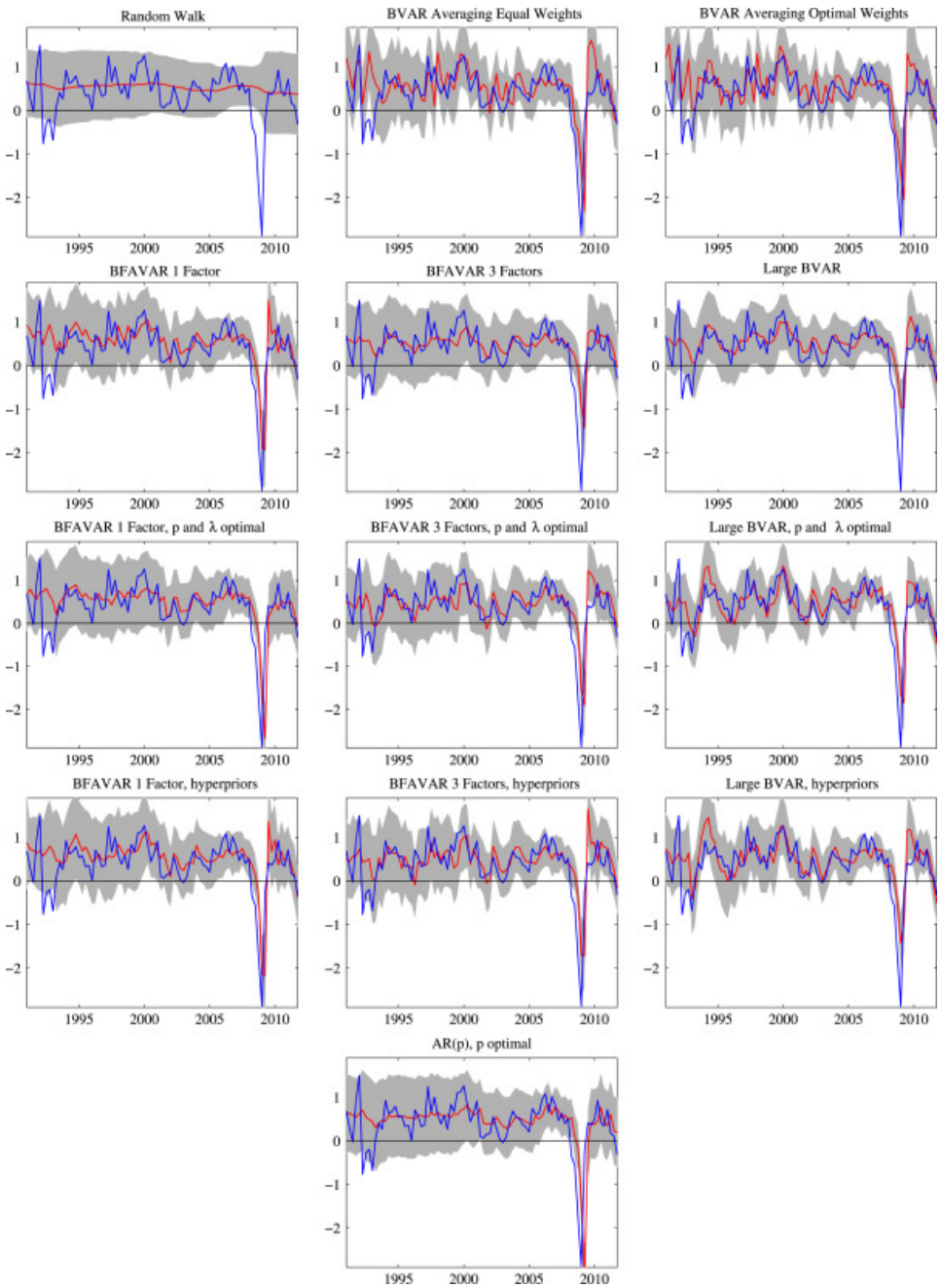
M01: Random walk		0	0.39	1.56	0.62	1.94
M02: BVAR averaging equal weights		∞	0.27	0.97	0.54	1.97
M03: BVAR averaging optimal weights		∞	0.27	0.88	0.50	1.90
M04: BFAVAR one factor	0.253	0.148	0.25	1.08	0.49	1.80
M05: BFAVAR three factors	0.098	0.092	0.26	0.96	0.49	1.81
M06: Large BVAR	0.034	0.041	0.26	0.77	0.53	2.01
M07: BFAVAR one factor, p_h and λ_h optimal	0.918	2.104	0.29	0.90	0.49	1.80
M08: BFAVAR three factors, p_h and λ_h optimal	0.420	0.520	0.26	0.86	0.46	1.71
M09: Large BVAR, p_h and λ_h optimal	0.217	1.443	0.27	0.96	0.53	2.22
M10: BFAVAR one factor, hyperpriors	0.408	0.515	0.25	0.76	0.51	1.94
M11: BFAVAR three factors, hyperpriors	0.350	0.378	0.25	0.84	0.47	1.84
M12: Large BVAR, hyperpriors	0.227	0.579	0.25	0.95	0.45	1.92
M13: $\mathbf{AR}(p_h)$, p_h optimal		n/a	0.28	0.88	0.58	2.03

Notes: the table shows RMSEs when the models are estimated in (log-)levels. See also the notes to Table 1.

Table E.13. Predictive likelihood: (log-)level specification.

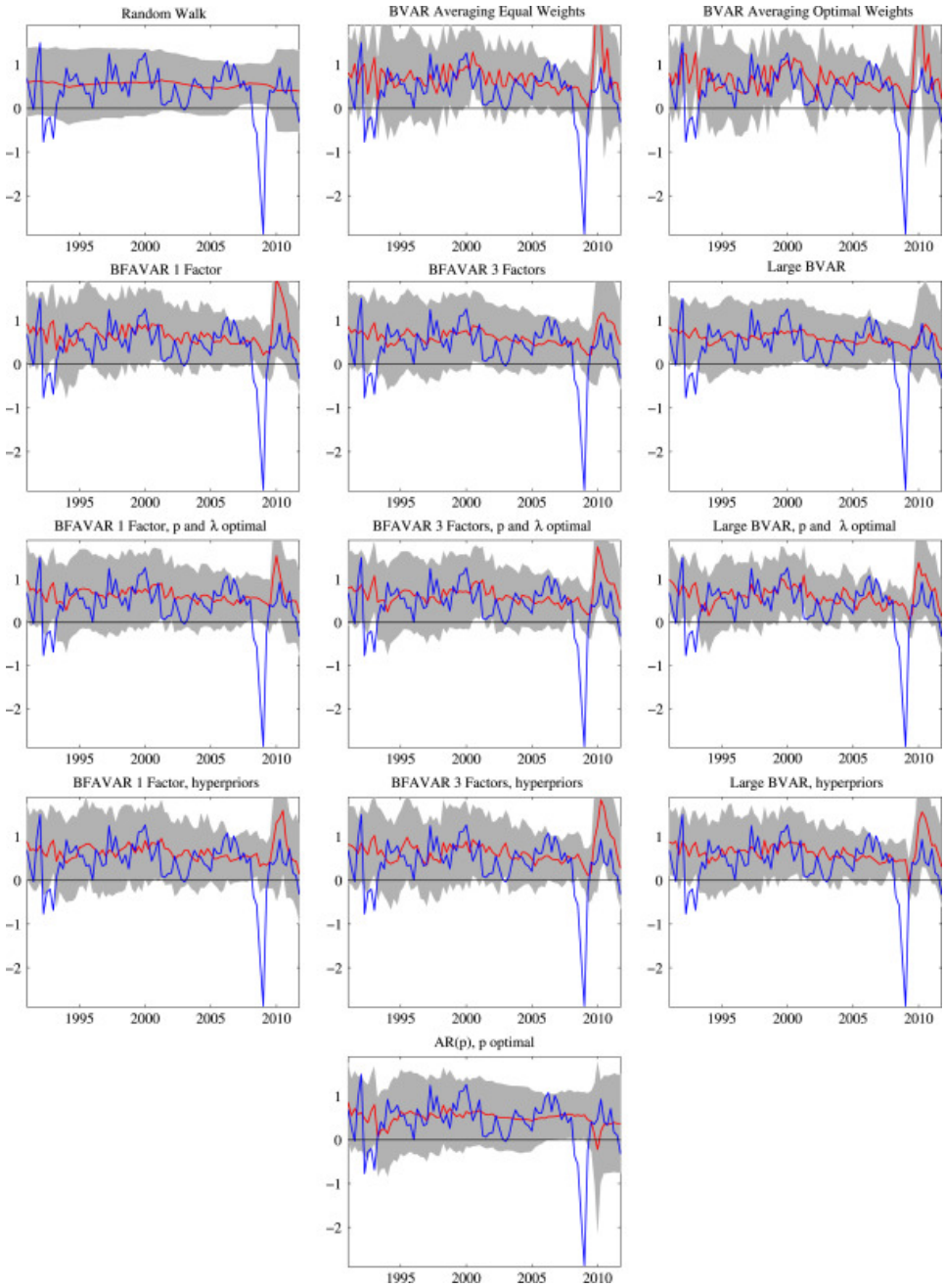
	λ_h		HICP		GDP	
	$h=1$	$h=4$	$h=1$	$h=4$	$h=1$	$h=4$
M01: Random walk		0	−1.87	−2.06	−1.96	−2.74
M02: BVAR averaging equal weights		∞	−1.68	−1.94	−1.50	−2.38
M03: BVAR averaging optimal weights		∞	−1.72	−1.91	−1.49	−2.42
M04: BFAVAR one factor	0.253	0.148	−1.56	−1.67	−1.52	−2.39
M05: BFAVAR three factors	0.098	0.092	−1.52	−1.66	−1.55	−2.35
M06: Large BVAR	0.034	0.041	−1.65	−1.82	−1.80	−2.97
M07: BFAVAR one factor, p_h and λ_h optimal	0.918	2.104	−1.66	−1.83	−1.60	−2.42
M08: BFAVAR three factors, p_h and λ_h optimal	0.420	0.520	−1.63	−1.84	−1.61	−2.44
M09: Large BVAR, p_h and λ_h optimal	0.217	1.443	−1.92	−3.27	−1.86	−4.68
M10: BFAVAR one factor, hyperpriors	0.408	0.515	−1.53	−1.63	−1.53	−2.57
M11: BFAVAR three factors, hyperpriors	0.350	0.378	−1.58	−1.88	−1.54	−2.54
M12: Large BVAR, hyperpriors	0.227	0.579	−1.72	−2.89	−1.64	−4.08
M13: $\mathbf{AR}(p_h)$, p_h optimal		n/a	−1.55	−2.00	−1.59	−2.36

Notes: the table shows the average log predictive likelihood when the models are estimated in (log-)levels. See also the notes to Table 1, Table 2.



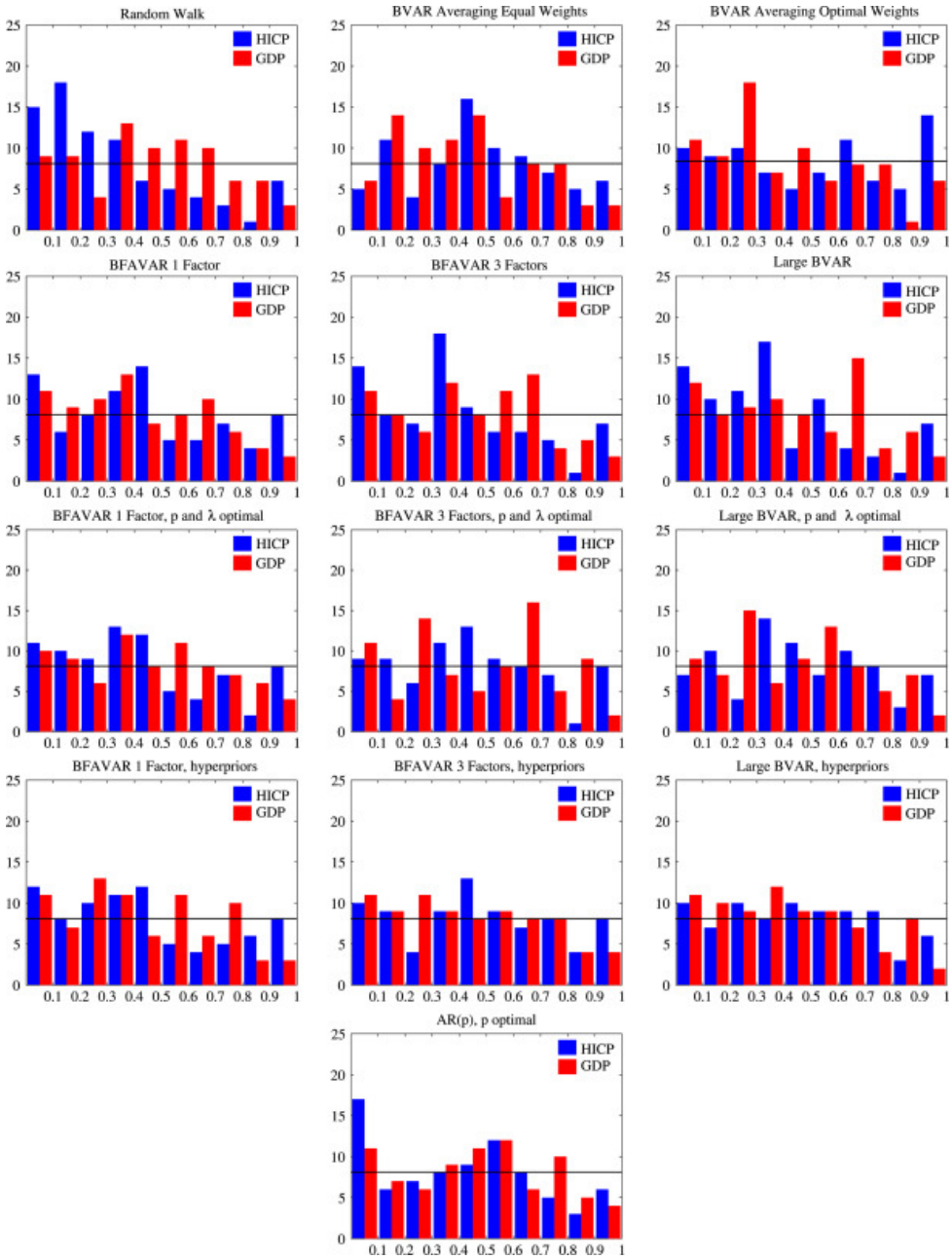
Download : Download full-size image

Fig. E.3. Forecasts and realizations: GDP growth ($h = 1$). Notes: the figure shows point forecasts with 90% forecast intervals for BVAR variants and an $\mathrm{AR}(p_h)$ model.



Download : Download full-size image

Fig. E.4. Forecasts and realizations: GDP growth ($h = 4$). Notes: the figure shows point forecasts with 90% forecast intervals for BVAR variants and an $\mathrm{AR}(p_h)$ model.



Download : Download full-size image

Fig. E.5. Probability integral transform: iterative approach ($h = 4$). Notes: the figure shows the PITs for BVAR variants and an $\mathrm{AR}(p_h)$ model, to check for the correct calibration of the predictive density. The horizontal line denotes uniformity. x-axis: probability band; y-axis: number of realizations.

Recommended articles

References

Aastveit et al., 2014 K.A. Aastveit, K.R. Gerdrup, A.S. Jore, L.A. Thorsrud
Nowcasting GDP in real-time: A density combination approach
Journal of Business and Economic Statistics, 32 (1) (2014), pp. 48-68
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Aiolfi et al., 2011 M. Aiolfi, C. Capistrán, A. Timmermann
Forecast combinations
M.P. Clements, D.F. Hendry (Eds.), Oxford handbook of economic forecasting, Oxford University Press, Oxford (2011)
Google Scholar

Amisano and Giacomini, 2007 G. Amisano, R. Giacomini
Comparing density forecasts via weighted likelihood ratio tests
Journal of Business and Economic Statistics, 25 (2) (2007), pp. 177-190
CrossRef View in Scopus Google Scholar

Bai and Ng, 2002 J. Bai, S. Ng
Determining the number of factors in approximate factor models
Econometrica, 70 (1) (2002), pp. 191-221
CrossRef View in Scopus Google Scholar

Bańbura et al., 2014 Bańbura, M., Giannone, D., & Lenza, M. (2014). *Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections*. Working Paper Series 1733, European Central Bank.
Google Scholar

Bańbura et al., 2010 M. Bańbura, D. Giannone, L. Reichlin
Large Bayesian vector autoregressions
Journal of Applied Econometrics, 25 (1) (2010), pp. 71-92
CrossRef View in Scopus Google Scholar

Barhoumi et al., 2013 K. Barhoumi, O. Darné, L. Ferrara
Testing the number of factors: An empirical assessment for a forecasting purpose
Oxford Bulletin of Economics and Statistics, 75 (1) (2013), pp. 64-79
CrossRef View in Scopus Google Scholar

Barnett et al., 2014 A. Barnett, H. Mumtaz, K. Theodoridis
Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters
International Journal of Forecasting, 30 (1) (2014), pp. 129-143
View PDF View article View in Scopus Google Scholar

Berg, 2015 T.O. Berg
Technology news and the US economy: Time variation and structural changes
Scottish Journal of Political Economy, 62 (2015), pp. 227-263
CrossRef View in Scopus Google Scholar

Bernanke et al., 2005 B.S. Bernanke, J. Boivin, P. Eliasz
Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach
Quarterly Journal of Economics, 120 (1) (2005), pp. 387-422
View in Scopus Google Scholar

Boero et al., 2011 G. Boero, J. Smith, K.F. Wallis
Scoring rules and survey density forecasts

International Journal of Forecasting, 27 (2) (2011), pp. 379-393

View PDF View article View in Scopus ↗ Google Scholar ↗

Boivin and Ng, 2006 J. Boivin, S. Ng

Are more data always better for factor analysis?

Journal of Econometrics, 132 (1) (2006), pp. 169-194

View PDF View article View in Scopus ↗ Google Scholar ↗

Carriero et al., 2012 Carriero, A., Clark, T. E., & Marcellino, M. (2012). *Common drifting volatility in large Bayesian VARs*. CEPR Discussion Papers 8894, C.E.P.R. Discussion Papers.

Google Scholar ↗

Carriero et al., 2013 Carriero, A., Clark, T. E., & Marcellino, M. (2013). *Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility*. CEPR Discussion Papers 9312, C.E.P.R. Discussion Papers.

Google Scholar ↗

Carriero et al., 2015 A. Carriero, T.E. Clark, M. Marcellino

Bayesian VARs: Specification choices and forecast accuracy

Journal of Applied Econometrics, 30 (1) (2015), pp. 46-73

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Carriero et al., 2009 A. Carriero, G. Kapetanios, M. Marcellino

Forecasting exchange rates with a large Bayesian VAR

International Journal of Forecasting, 25 (2) (2009), pp. 400-417

View PDF View article View in Scopus ↗ Google Scholar ↗

Clark, 2011 T.E. Clark

Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility

Journal of Business and Economic Statistics, 29 (3) (2011), pp. 327-341

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Clark and McCracken, 2010 T.E. Clark, M.W. McCracken

Averaging forecasts from VARs with uncertain instabilities

Journal of Applied Econometrics, 25 (1) (2010), pp. 5-29

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Clark and Ravazzolo, 2015 T.E. Clark, F. Ravazzolo

Macroeconomic forecasting performance under alternative specifications of time-varying volatility

Journal of Applied Econometrics, 30 (2015), pp. 551-575

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

D'Agostino et al., 2013 A. D'Agostino, L. Gambetti, D. Giannone

Macroeconomic forecasting and structural change

Journal of Applied Econometrics, 28 (1) (2013), pp. 82-101

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

De Mol et al., 2008 C. De Mol, D. Giannone, L. Reichlin

Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?

Journal of Econometrics, 146 (2) (2008), pp. 318-328

View PDF View article View in Scopus ↗ Google Scholar ↗

Diebold et al., 1998 F.X. Diebold, T.A. Gunther, A.S. Tay

Evaluating density forecasts with applications to financial risk management

International Economic Review, 39 (4) (1998), pp. 863-883

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Diebold et al., 1999 F.X. Diebold, J. Hahn, A.S. Tay

Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange

Review of Economics and Statistics, 81 (4) (1999), pp. 661-673

View in Scopus ↗ Google Scholar ↗

Doan et al., 1984 T. Doan, R. Litterman, C. Sims

Forecasting and conditional projection using realistic prior distributions

Econometric Reviews, 3 (1) (1984), pp. 1-100

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Fagan et al., 2005 G. Fagan, J. Henry, R. Mestre

An area-wide model for the Euro area

Economic Modelling, 22 (1) (2005), pp. 39-59

View PDF View article View in Scopus ↗ Google Scholar ↗

Forni et al., 2003 M. Forni, M. Hallin, M. Lippi, L. Reichlin

Do financial variables help forecasting inflation and real activity in the Euro area?

Journal of Monetary Economics, 50 (6) (2003), pp. 1243-1255

View PDF View article View in Scopus ↗ Google Scholar ↗

Geweke, 1992 J. Geweke

Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments

J.M. Bernardo, J. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian statistics, Oxford University Press (1992), pp. 169-194

CrossRef ↗ Google Scholar ↗

Geweke and Amisano, 2010 J. Geweke, G. Amisano

Comparing and evaluating Bayesian predictive distributions of asset returns

International Journal of Forecasting, 26 (2) (2010), pp. 216-230

View PDF View article View in Scopus ↗ Google Scholar ↗

Geweke and Amisano, 2011 J. Geweke, G. Amisano

Optimal prediction pools

Journal of Econometrics, 164 (1) (2011), pp. 130-141

View PDF View article View in Scopus ↗ Google Scholar ↗

Geweke and Amisano, 2012 J. Geweke, G. Amisano

Prediction with misspecified models

American Economic Review: Papers and Proceedings, 102 (3) (2012), pp. 482-486

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Giacomini and White, 2006 R. Giacomini, H. White
Tests of conditional predicitve ability
Econometrica, 74 (6) (2006), pp. 1545-1578
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Giannone et al., 2014 D. Giannone, M. Lenza, D. Momferatou, L. Onorante
Short-term inflation projections: A Bayesian vector autoregressive approach
International Journal of Forecasting, 30 (3) (2014), pp. 635-644
View PDF View article View in Scopus ↗ Google Scholar ↗

Giannone et al., 2015 D. Giannone, M. Lenza, G.E. Primiceri
Prior selection for vector autoregressions
Review of Economics and Statistics, 97 (2) (2015), pp. 436-451
View in Scopus ↗ Google Scholar ↗

Gneiting and Raftery, 2007 T. Gneiting, A.E. Raftery
Strictly proper scoring rules, prediction, and estimation
Journal of the American Statistical Association, 102 (477) (2007), pp. 359-378
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Gneiting and Ranjan, 2011 T. Gneiting, R. Ranjan
Comparing density forecasts using threshold- and quantile-weighted scoring rules
Journal of Business and Economic Statistics, 29 (3) (2011), pp. 411-422
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Hall and Mitchell, 2007 S.G. Hall, J. Mitchell
Combining density forecasts
International Journal of Forecasting, 23 (1) (2007), pp. 1-13
View PDF View article View in Scopus ↗ Google Scholar ↗

Henzel and Mayr, 2013 S.R. Henzel, J. Mayr
The mechanics of VAR forecast pooling—A DSGE model based Monte Carlo study
North American Journal of Economics and Finance, 24 (1) (2013), pp. 1-24
View PDF View article View in Scopus ↗ Google Scholar ↗

Kadiyala and Karlsson, 1993 K.R. Kadiyala, S. Karlsson
Forecasting with generalized Bayesian vector autoregressions
Journal of Forecasting, 12 (4) (1993), pp. 365-378
Google Scholar ↗

Kadiyala and Karlsson, 1997 K.R. Kadiyala, S. Karlsson
Numerical methods for estimation and inference in Bayesian VAR-models
Journal of Applied Econometrics, 12 (2) (1997), pp. 99-132
View in Scopus ↗ Google Scholar ↗

Kim and Swanson, 2014 H.H. Kim, N.R. Swanson
Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence

Journal of Econometrics, 178 (P2) (2014), pp. 352-367

View PDF View article CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Knüppel and Schultefrankenfeld, 2012 M. Knüppel, G. Schultefrankenfeld

How informative are central bank assessments of macroeconomic risks?

International Journal of Central Banking, 8 (3) (2012), pp. 87-139

View in Scopus ↗ Google Scholar ↗

Koop, 2013 G. Koop

Forecasting with medium and large Bayesian VARS

Journal of Applied Econometrics, 28 (2) (2013), pp. 177-203

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Litterman, 1980 Litterman, R. B. (1980). *A Bayesian procedure for forecasting with vector autoregressions*. Manuscript, Massachusetts Institute of Technology, Department of Economics.

Google Scholar ↗

Litterman, 1986 R.B. Litterman

Forecasting with Bayesian vector autoregressions—Five years of experience

Journal of Business and Economic Statistics, 4 (1) (1986), pp. 25-38

View in Scopus ↗ Google Scholar ↗

Mitchell and Hall, 2005 J. Mitchell, S.G. Hall

Evaluating, comparing and combining density forecasts using the KLIC with an application to the bank of England and NIESR 'fan' charts of inflation

Oxford Bulletin of Economics and Statistics, 67 (s1) (2005), pp. 995-1033

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Pirschel and Wolters, 2014 Pirschel, I., & Wolters, M. H. (2014). *Forecasting German key macroeconomic variables using large dataset methods*. Kiel Working Papers 1925, Kiel Institute for the World Economy.

Google Scholar ↗

Primiceri, 2005 G.E. Primiceri

Time varying structural vector autoregressions and monetary policy

Review of Economic Studies, 72 (3) (2005), pp. 821-852

View in Scopus ↗ Google Scholar ↗

Ravazzolo and Vahey, 2014 F. Ravazzolo, S.P. Vahey

Forecast densities for economic aggregates from disaggregate ensembles

Studies in Nonlinear Dynamics and Econometrics, 18 (4) (2014), pp. 367-381

View in Scopus ↗ Google Scholar ↗

Robertson and Tallman, 1999 J.C. Robertson, E.W. Tallman

Vector autoregressions: Forecasting and reality

Economic Review, 84 (1) (1999), pp. 4-18

View in Scopus ↗ Google Scholar ↗

Rossi and Sehkposyan, 2014 B. Rossi, T. Sehkposyan

Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set

International Journal of Forecasting, 30 (3) (2014), pp. 662-682

View PDF View article View in Scopus ↗ Google Scholar ↗

Stock and Watson, 2002 J.H. Stock, M.W. Watson

Macroeconomic forecasting using diffusion indexes

Journal of Business and Economic Statistics, 20 (2) (2002), pp. 147-161

View in Scopus ↗ Google Scholar ↗

Stock and Watson, 2005 Stock, J. H., & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis*. NBER Working Papers 11467, National Bureau of Economic Research, Inc.

Google Scholar ↗

Stock and Watson, 2006 J.H. Stock, M.W. Watson

Forecasting using many predictors

G. Elliott, C. Granger, A. Timmerman (Eds.), Handbook of economic forecasting, North Holland, Amsterdam (2006)

Google Scholar ↗

Stock and Watson, 2011 J.H. Stock, M.W. Watson

Dynamic factor models

M.P. Clements, D.F. Hendry (Eds.), Oxford handbook of forecasting, Oxford University Press, Oxford (2011)

Google Scholar ↗

Weiss, 1996 A.A. Weiss

Estimating time series models using the relevant cost function

Journal of Applied Econometrics, 11 (5) (1996), pp. 539-560

View in Scopus ↗ Google Scholar ↗

Wolters, 2015 M.H. Wolters

Evaluating point and density forecasts of DSGE models

Journal of Applied Econometrics, 30 (1) (2015), pp. 74-96

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Cited by (18)

Real-time inflation forecasting using non-linear dimension reduction techniques

2023, International Journal of Forecasting

Show abstract ⌄

Macroeconomic transmission of Eurozone shocks to India—A mean-adjusted Bayesian VAR approach

2020, Economic Analysis and Policy

Show abstract ⌄

Forecasting inflation in emerging markets: An evaluation of alternative models

2017, International Journal of Forecasting

Citation Excerpt :

…Although various studies have conducted forecast comparisons across alternative models and time, most of these studies have focused only on one particular country. For instance, Clark and Ravazzolo (2015) and D'Agostino, Gambetti, and Giannone (2013) assess the macroeconomic forecasting performances of alternative models for the United States (US); Barnett, Mumtaz, and Theodoridis (2014) and Groen, Kapetanios, and Price (2009) focus on the United Kingdom (UK); and Berg and Henzel (2015), Caggiano, Kapetanios, and Labhard (2011) and Giannone, Lenza, Momferatou, and Onorante (2014) consider Europe. For EMs, some of the country studies include those of Gupta and Kabundi (2011) for South Africa; Bailliu, Garcés, Kruger, and Messmacher (2003) and Öğünç et al. (2013) for Turkey; Kaya and Yazgan (2014) for Mexico; and Duasa, Ahmad, Ibrahim, and Zainal (2010) for Malaysia.…

Show abstract

The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia

2017, Procedia Computer Science

Show abstract

Financial nowcasts and their usefulness in macroeconomic forecasting

2019, International Journal of Forecasting

Citation Excerpt :

…We also consider specifications that include financial factors, in order to broadly summarize financial conditions without explicitly incorporating many financial variables that would normally entail a tradeoff through a tightening of the priors used in estimation (see e.g. Bańbura et al., 2010). We follow the Bayesian FAVAR approach of Bańbura et al. (2010), Berg and Henzel (2015), and Bernanke, Boivin, and Eliasz (2005); that is, we extract common factors from a large balanced panel of 58 financial variables, which is approximately the same as the smaller of the two datasets used by Andreou et al. (2013), and include these financial factors as additional variables in our VAR.22 Turning to macroeconomic variables, all of the models contain four primary macroeconomic variables of interest: real GDP, CPI inflation, the unemployment rate, and the effective federal funds rate.23…

Show abstract

Bagged neural networks for forecasting Polish (low) inflation

2019, International Journal of Forecasting

Citation Excerpt :

…From an economic perspective, some researchers have responded to the changing relationships among inflation, economic slack and global conditions (e.g. Auer, Borio, & Filardo, 2017; Ciccarelli & Mojon, 2010) by developing flexible Phillips curve models that either utilize non-linear (e.g. Markov switching) methodologies to account for the state-dependent reaction of inflation and its volatility to the determinants of price developments (e.g. González-Rivera & Sun, 2017), or allow the coefficients and the entire model to change over time (e.g. Koop & Korobilis, 2012). Multivariate frameworks frequently include (Bayesian) VAR models, often augmented with common factors (e.g. Berg & Henzel, 2015), time-varying coefficients and shrinkage procedures (e.g. Belmonte, Koop, & Korobilis, 2014), or tight steady-state priors (e.g. Stelmasiak & Szafrański, 2016). Approaches that rely on economic theory employ either the present-value formulation of the hybrid new Keynesian Phillips curve (e.g. Rumler & Valderrama, 2010) or dynamic stochastic equilibrium models (e.g. Edge & Gurkaynak, 2010).…

Show abstract

View all citing articles on Scopus ↗

1 The idea in this body of literature is that the information contained in a large number of indicator variables can be summarized by a rather small number of factors that are added to the variables of interest (see, e.g., Forni et al., 2003, Stock and Watson, 2002, Stock and Watson, 2005, Stock and Watson, 2006, Stock and Watson, 2011, among others).

2 New evidence for Germany using large dataset methods is provided by Pirschel and Wolters (2014).

3 See also Fagan, Henry, and Mestre (2005). The vintage that we use is as of September 2012.

4 Unfortunately, the seasonally adjusted HICP (HICPSA) is not available prior to 1990:1. Hence, we take the unadjusted series (HICP) instead, and apply the Census-X-12-ARIMA program in order to remove the seasonality. The quarterly inflation rates based on our seasonally adjusted index are very similar to those based on the HICPSA over the period 1990:1 to 2011:4; the correlation between them is 0.96.

5 Non-conjugate priors are an alternative to conjugate priors in systems with up to 20 variables, but are not feasible computationally for larger models (see Koop, 2013).

6 It should be noted that the sum-of-coefficients prior has no impact on the forecast accuracy if the series are (log-)differenced, since in that case $\delta_i = 0$, and y_h^{++} and x_h^{++} simply contain rows of zeros. Hence, in our paper, this prior is only relevant when we consider the data in (log-)levels.

7 In fact, when using the direct-step approach, the only variant that involves simulation methods is that in which we estimate the tightness parameter λ_h.

8 In practice, we set $\lambda_h = 10^{-4}$.

9 In practice, we set $\lambda_h = 10^{4}$.

10 It should be noted that this paper defines optimality only with respect to a certain (formal) criterion, it does not refer to the forecasting ability of a variant.

11 We use the Matlab routine “fmincon”.

12 The lag length may also change across periods, but we neglect the t subscript for the sake of brevity.

13 We provide the analytical expression for the marginal likelihood in Appendix B.

14 We have experimented with various different values of the prior for Σ_h and found that this choice had little or no impact on the estimation outcome for most BVAR variants. Only the large BVAR with all 44 variables turned out to be more sensitive to the value chosen. To stabilize inference, we decided to set the value to 0.005^2 (for a similar argument, see Giannone et\special {t4ht=&} al., 2015). In order to ensure comparability, we use the same value for all variants and specifications.

15 Appendix D also explains how to obtain forecasts iteratively in this case.

16 We use the Matlab routine “ksdensity”.

17 Appendix E provides a graphical representation of the point and density forecasts.

18 Table E.1 in Appendix E reveals that these differences in forecast ability are significant at a 5% level according to a Giacomini and White (2006) test of conditional predictive ability.

19 See Gneiting and Raftery (2007) and Gneiting and Ranjan (2011). We calculate the integral using a discrete approximation. The support ranges from -10% to $+10\%$, and the increment size is 0.01 ppt.

20 See Diebold, Gunther, and Tay (1998) and Diebold, Hahn, and Tay (1999).

21 It should be noted that the predictive density for this large BVAR variant is symmetric, but the true distribution of inflation might not be, which may explain the poor performance in part. However, we also find that the BFAVARs perform better despite their symmetric distribution. The differences between the BFAVARs and the large BVAR seem to be mainly the result of an unfavorable location of the density of the latter in certain periods. Fig. E.1, Fig. E.2 reveal that this was the case during the late 1990s and around 2009, when inflation fell to well below average levels.

22 We use the Matlab routine “ksdensity”.

23 We use the Matlab routine “fmincon”.

View Abstract

Copyright © 2015 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

ELSEVIER

All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

RELX™