International Journal of Forecasting 31 (2015) 1067–1095

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Point and density forecasts for the euro area using Bayesian VARs

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

[a] *Ifo Institute, Germany*

[b] *CESifo, Germany*

ARTICLE INFO

Keywords:
Bayesian vector autoregression
Forecasting
Model validation
Large cross section
Euro area

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.

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, Clark, & Marcellino, 2015; Carriero, Kapetanios, & Marcellino, 2009; D'Agostino, Gambetti, & Giannone, 2013; Giannone, Lenza, & Primiceri, 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 par-

* Corresponding author.
E-mail addresses: berg@ifo.de (T.O. Berg), henzel@ifo.de (S.R. Henzel).

[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, Hallin, Lippi, & Reichlin, 2003; Stock & Watson, 2002, 2005, 2006, 2011, among others).

http://dx.doi.org/10.1016/j.ijforecast.2015.03.006
0169-2070/

ticular, 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, Giannone, & Lenza, 2014; Giannone, Lenza, Momferatou, & Onorante, 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 & Ng, 2002; Boivin & 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, Smith, & Wallis, 2011; Knüppel & Schultefrankenfeld, 2012; Mitchell & 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 & 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

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