EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Elena Angelini, Magdalena Lalik,
Michele Lenza, Joan Paredes

Mind the gap: a multi-country BVAR benchmark for the Eurosystem projections

No 2227 / January 2019

Abstract

The Eurosystem staff forecasts are conditional on the financial markets, the global economy and fiscal policy outlook, and include expert judgement. We develop a multi-country BVAR for the four largest countries of the euro area and we show that it provides accurate conditional forecasts of policy relevant variables such as, for example, consumer prices and GDP. The forecasting accuracy and the ability to mimic the path of the Eurosystem projections suggest that the model is a valid benchmark to assess the consistency of the projections with the conditional assumptions. As such, the BVAR can be used to identify possible sources of judgement, based on the gaps between the Eurosystem projections and the historical regularities captured by the model.

JEL codes: C52, C53, E37

Keywords: Multi-country model, cross-checking, conditional forecast, euro area

Non-Technical Summary

The Eurosystem staff macroeconomic projection exercises (BMPE), conducted four times a year, provide a view on the euro area economic outlook to support the ECB Governing Council monetary policy decisions.

The projection exercises follow a bottom-up approach, according to which the projections for the 19 countries of the euro area are prepared, and then aggregated to define the view on the euro area. A variety of sources of information and models (maintained at the European Central Bank and at National Central Banks) are used to prepare the projections. Moreover, as structural econometric models are not well equipped to describe some aspects, the projections are conditioned on a set of so called "technical assumptions", i.e. on the paths of a certain number of variables characterizing the outlook of financial markets, the global economy and country-specific fiscal policy developments. Finally, off-model information that could be useful to determine the euro area outlook is incorporated in the projections via the inclusion of judgemental elements.

This brief description highlights the richness of the projection framework and, at the same time, some of the challenges that it entails. For example, the range of model approaches, which guarantees the robustness of the assessment, also makes it difficult to disentangle the relative role of model and of off-model information in the assessment of the outlook. This, in turn, may complicate the formulation of a narrative of the projections and of the risks surrounding them. In this paper we describe an empirical model developed for cross-checking the consistency of the projections with the technical assumptions. Among other things, the model can be used to identify the main judgmental elements in the projections, by characterizing them as major deviations (i.e. gaps) from the paths implied by historical regularities in the data.

In order to describe the euro area time-series historical regularities, we specify a very general linear model, i.e. a Vector Autoregressive (VAR) model. To mirror the features of the Eurosystem projection exercises, the model has a multi-country dimension, including variables of the four biggest economies of the euro area, i.e. France, Germany, Italy and Spain. As target variables, we include real GDP, real total investment, the harmonized index of consumer prices (HICP), the GDP deflator, wages, loans and lending rates to firms. Then, we include a set of technical assumptions capturing the policy and the global environment in which the euro area operates.

In order to credibly accomplish the task of cross-checking the projections, the model should be able to provide an accurate description of the historical regularities characterizing the economic fluctuations in the different euro area countries. Hence, the bulk of the paper is devoted to an extensive out-of-sample evaluation of point and density forecasts. The evaluation shows that the conditional forecasts from the multi-country VAR model are, in general, more accurate than those of the traditional

empirical benchmarks in the forecasting evaluation literature. We also show that the model, despite being a mechanical tool and only featuring the four biggest countries of the euro area, is able to accurately track the (B)MPE path for GDP and HICP published in real-time. These findings support the idea that the model described in this paper is a valid benchmark to cross-check the Eurosystem projections.

1 Introduction

The Governing Council of the European Central Bank (ECB) draws on several sources of information, when taking interest rate decisions. In particular, the economic analysis, jointly conducted by the staff of the ECB and of the National Central Banks (NCBs) of the Eurosystem, aims to identify the economic shocks driving the business cycle, and embodies a thorough assessment of the inflation dynamics. The economic analysis is made concrete, each quarter, in the Broad Macroeconomic Projections Exercise.

Such exercises bring together in a systematic manner a range of information on current and future economic developments. To reflect the nature of the euro area, which includes 19 sovereign countries with heterogeneous economies, a bottom-up approach is followed, according to which, first, the projections for the individual countries are defined, and then they are aggregated to characterize the euro area outlook. From a methodological standpoint, the exercises are based on conventional (semi-)structural macroeconometric tools (including individual country, multi-country and euro area-wide models; see Fagan and Morgan, 2005).

In order to account for some aspects that the structural econometric models may not be well equipped to describe, the projections are conditioned on a set of so called "technical assumptions", i.e. on the paths of a number of variables characterizing the outlook of financial markets (policy rates, sovereign bond rates, stock prices), the global economy (foreign demand, global prices and exchange rates) and fiscal policy developments. The projection exercises span an horizon of up to three years ahead.

The model outcomes are complemented by the inclusion of the judgmental input of sectoral and country experts, to factor in the insight from relevant sources of information, which are not fully captured by the technical assumptions. Judgmental input can also make up for missing elements or possible miss-specification typical of the economic models, as for example those stemming from the materialization of unconventional economic shocks, which are difficult to capture for models reflecting the historical regularities in the data.

This description of the Eurosystem projection exercises highlights their richness in terms of models, expert judgment and sources of information. However, the broad scope of the exercises, characterized among other things by the use of different models, makes the separation of the model and the judgmental element difficult to accomplish. Such separation may be very relevant for the characterization of the narrative supporting the projections and for the analysis of the main risks surrounding them.

This paper describes an empirical model developed for cross-checking the consistency of the projections with the technical assumptions. Among other things, the model can be used to identify the main judgmental elements in the projections, by characterizing them as major *gaps* with respect to the paths implied by historical regularities in the data. In order to appropriately capture the historical regularities, we choose a very general linear model, i.e. a Vector Autoregressive (VAR) model. To mirror the features of the Eurosystem projection exercises, the model has a multi-country nature, including variables of the four biggest economies of the euro area, i.e. France, Germany, Italy and Spain, which account for about three quarters of the euro area GDP. In more details, for all countries, we include as target variables (i.e. the variables we wish to forecast) real GDP, real total investment, the harmonized index of consumer prices (HICP), the GDP deflator, wages, loans and lending rates to firms. Then, we include a set of variables that are considered as technical assumptions in projection exercises, in particular those capturing the policy and the global environment in which the euro area operates, i.e. the short-term interest rate of the euro area (proxied by a measure of the three-months money market rate), the US dollar/euro exchange rate, the oil price, foreign demand, US GDP and the US short-term interest rate.[1]

The variables (28 target variables and 9 technical assumptions) included in the model are available at the quarterly frequency starting in 1995Q1, and we specify the model in (log-)levels with five lags, to fully account for the dynamics in the relationships across variables. To appropriately handle the estimation of our large and complex model, we employ Bayesian techniques. In practice, we shrink the model parameters toward those of a random walk model, by imposing a Minnesota prior and two priors on the sum-of-coefficients (see Litterman, 1979; Doan et al., 1984; Sims, 1992). De Mol et al. (2008) and Banbura et al. (2010) have shown that, if the variables co-move (as it is typically the case for macroeconomic and financial variables), then the information in the sample still drives the parameter estimates even when the prior beliefs are quite dogmatically imposed, to control for the relevant extent of estimation error incurred in large models. We treat the parameters controlling the informativeness of the prior distributions as random variables, as suggested in Giannone et al. (2015), and we draw from their posterior distribution to account for the source of uncertainty related to the set-up of the prior tightness.

In order to credibly accomplish the task of cross-checking the projections and to identify the main judgemental elements, the model should be able to provide an accurate description of the historical regularities characterizing the economic fluctuations in the different euro area countries and sectors. In other words, only if our BVAR model can be shown to accurately describe the economic fluctuations in the euro area, large gaps of the Eurosystem projections with respect to the model outcomes can be considered as indicative of possible judgemental elements rather

[1]Notice that we align to the conventions of the Eurosystem projection exercises and refer to the included variables as belonging either to the group of the *target* variables, i.e. the variables we want to forecast, or to the group of the *technical assumptions*, i.e. the variables whose future paths the forecasts of the target variables are conditioned upon. This classification is adopted only for the sake of easy reading and has no consequences on the model specification since all the variables are treated as endogenous in the VAR model.

than mere failures of the BVAR model. Hence, the bulk of the paper is devoted to an extensive out-of-sample evaluation of point and density conditional forecasts. Specifically, we evaluate the accuracy of conditional BVAR forecasts (derived by using the algorithm described in Banbura et al., 2015) based on the actual value (i.e. observed ex post) of the future paths of the technical assumptions.[2] Besides validating our measure of the gaps, this exercise also contributes to the literature that evaluates the practices of central bank forecasting, given the relevant role of conditional forecasts in the toolbox of central banks (see, for example, Del Negro and Schorfheide, 2013, Giannone et al., 2014, Fawcett et al., 2015, Iversen et al., 2016 and Domit et al., 2016). We choose to primarily evaluate conditional forecasts based on the actual value of the assumptions (even though we also provide an assessment of the conditional forecasts based on real-time assumptions) because, from a statistical point of view, this is the most appropriate way to assess the accuracy of conditional forecasts (for an extensive discussion of this point, see Faust and Wright, 2008 and Clark and McCracken, 2014).

The conditional forecasts produced by our model for the 28 "target" variables are more accurate, in general, than the unconditional BVAR forecasts, which shows that the model is able to extract the valuable information embedded in the conditional paths. Moreover, the conditional BVAR forecasts also generally improve on the traditional univariate benchmark models of non-forecastability, which suggests that our model is able to accurately forecast the target variables. The improvement in accuracy compared to benchmark models is achieved both in terms of point and density forecasts. The latter are particularly important, given that one of the possible uses of the model is to highlight "large" deviations ("gaps") from historical regularities and, hence, it is important that the model correctly evaluates the uncertainty surrounding the forecasts.

Although meaningful from a statistical point of view, the previous exercise also gives rise to projections that are unfeasible in real-time, given that the future path of the assumptions cannot be observed. To gauge how the model would fare in a context that is closer to real-time, we collect the real-time conditioning assumptions used in the Eurosystem projection exercises in the period 2011Q2 to 2016Q4. Over this relatively short sample, we produce euro area GDP and HICP conditional forecasts by aggregating our forecasts for the four countries in the model, and we compare them with the published Eurosystem projections for the euro area as a whole. Interestingly, the BVAR projections display a similar evolution as the judgmental Eurosystem projections, in spite of the fact that the latter are an aggregate of the whole euro area, while our BVAR includes only the four biggest countries and our projections are produced by using exclusively a mechanical model procedure. This finding, coupled with the forecast accuracy gauged in our out-of-sample

[2]See Herbst and Schorfheide (2012) for the use of conditional forecasts to evaluate the ability of a model to describe the comovement in the data.

evaluation exercises, further corroborates the view that the model described in this paper is a valid benchmark to cross-check the Eurosystem projections. It should also be noticed that, in the cases in which the Eurosystem projections more markedly deviate from the BVAR, in particular for HICP, the Eurosystem projections turn out to be more accurate than the BVAR counterparts.

This paper relates to a large literature on reduced form multi-country models, which includes several alternative model representations such as the Global VARs (GVARs, for a survey see Chudik and Pesaran, 2016), Panel VARs (PVARs, see Canova and Ciccarelli, 2013 for a survey) and dynamic factor models (DFMs, see Stock and Watson, 2002 and Forni et al. 2000). All these model strategies are good representations for large databases if there is a relevant extent of comovement in the variables and should be seen as complements rather than substitutes (see Banbura et al., 2015). We choose Bayesian VARs over the other model techniques because they provide a very natural way to handle potential non-stationarity in the data and require less, possibly ad hoc, specification choices like, for example, the number or the nature of the common factors or global variables and the restrictions on the heterogeneity of the parameters. A model strategy similar to the one in this paper is adopted in Altavilla et al., (2016) and Lenza and Slacalek (2018) to evaluate the effects of structural (standard and non-standard) monetary policy shocks on the economies of the biggest four countries of the euro area. Capolongo and Pacella (2018) show that a multi-country BVAR similar to the one in this paper improves over both a comparable aggregate euro area VAR and on individual country VARs in terms of forecasting accuracy for euro area HICP. The additional contribution of this paper with respect to the literature rests on the analysis of conditional forecasts and on the real-time perspective, both grounded in the practice of the Eurosystem projection exercises.

The remainder of the paper is organized as follows. Section 2 describes the data we use, the multi-country BVAR model and the empirical exercises. Section 3 provides the empirical results and their interpretation. Section 4 concludes.

2 Data, model and empirical exercises

2.1 The Eurosystem projections and the database

The Eurosystem macroeconomic projection exercises are part of the input prepared for the Governing Council's decision-making meetings.[3] The Broad Macroeconomic Projection Exercise (BMPE), in which all the euro area national central banks and the ECB are involved, is carried out twice a year and its outcomes are published in June and December. The ECB Staff Macroeconomic Projection Exercise (MPE) is

[3] See ECB (2001, 2016) for an extensive discussion of the features of the Eurosystem projection exercises.

also carried out twice a year, and its outcomes are published in March and September, alternating with the BMPE. For ease of exposition, from now on we will refer to the projection exercises simply as BMPE.

The published figures include projections for inflation in terms of the Harmonised Index of Consumer Prices (HICP), for the growth rate of real GDP and its main expenditure components, and other important macroeconomic and fiscal variables. The forecast horizon includes the current year and the subsequent two to three years. To reflect the degree of uncertainty attached to such exercises, the projections for inflation and for real GDP growth are published in terms of ranges, with the corresponding midpoints. There are two main steps in the production of the projection exercises. The first step involves the setting of technical assumptions underlying the exercise, covering interest rates, exchange rates, the international environment and fiscal variables. In a second stage, an agreement on a set of macroeconomic projection figures is reached via several iterations involving different layers of ECB and NCB staff. The euro area projections are obtained by aggregation of the individual country projections.

The dataset we use in this paper is largely inspired by the practice of the projection exercises just described, and includes 37 variables. In the model, all variables are treated as endogenous. However, to mirror the convention of the Eurosystem projections, we split the variables in two sets: target variables and assumptions. The target variables are the variables we are interested to forecast. The assumptions are the variables on whose future paths we condition the forecasts of the target variables.

The target variables for the four biggest countries of the euro area (France, Germany, Italy and Spain) in our model are real GDP, real total investment, the Harmonized Index of Consumer Prices (HICP), wages, the GDP deflator, loans and lending rate to firms.[4] The assumptions we use are only a sub-set of those used in the Eurosystem projections and comprise the oil price, the dollar/euro exchange rate, the euro area common short-term interest rate, foreign demand for the four countries, US real GDP and short-term rate. The sample starts in 1995q1 and ends in 2016q3, in the longest available vintage, and all data are available at the quarterly frequency.[5]

We have also collected a real-time database, with the assumptions available in real-time for the projection exercises starting in 2011Q2. The assumptions we include are derived in the following way. The short-term interest rate (three-month EURIBOR) is assumed to evolve in line with the prevailing market expectations, derived from futures rates, at the cut-off date for the projection exercises. The assumptions on the bilateral EUR/USD exchange rate has a purely technical nature in that the exchange rate is assumed to be constant over the projection period at

[4]For wages, we use compensation per employee.

[5]For more details, see the data appendix at the end of the paper.

a level based on an average of recent rates.[6] The future path of oil prices is based on recently observed futures market prices. The path of foreign demand and US GDP is based on modelling and judgemental considerations and is provided by the ECB experts on international economic developments, and the future paths of US short-term rates are also based on market expectations. With some occasional modifications, this type of assumptions is quite standard for central bank forecasting.

2.2 The multi-country BVAR model

This section describes the specification and the estimation method used for the VAR model. The exposition closely follows Giannone et al. (2015). The VAR model for an N-dimensional vector of time-series y_t can be described as:

$$y_t = C + B_1 y_{t-1} + ... + B_p y_{t-p} + \epsilon_t$$

$$\epsilon_t \sim N(0, \Sigma)$$

where B_1 ...B_p are NxN matrices of coefficients on the p lags of the variables, C is an N-dimensional vector of constants and Σ is the covariance matrix of the VAR errors.

In our specification, N=37 and p=5, while the full sample at our disposal ranges from 1995Q1 to 2016Q3. We estimate the model by means of Bayesian techniques, and we impose conjugate prior distributions belonging to the Normal/Inverse-Wishart family. More in detail, the prior for the covariance matrix of the residuals Σ is Inverse-Wishart, while the prior for the autoregressive coefficients is (conditional to Σ) normal.

For the prior on the covariance matrix of the errors, we set the degrees of freedom of the Inverse-Wishart distribution equal to $N+2$, the minimum value that guarantees the existence of the prior mean, and we assume a diagonal scaling matrix Ψ. Notice that we treat Ψ as an hyperparameter, as suggested in Giannone et al. (2015), differently from previous literature that generally fixes the scaling matrix based on sample information.

The baseline prior on the model coefficients is a version of the so-called Minnesota prior (see Litterman, 1979). This prior is centered on the assumption that each variable follows an independent random walk process, possibly with drift, which is a parsimonious yet reasonable approximation of the behaviour of an economic variable. The prior first and second moments for the VAR coefficients are as follows:

[6] This does not in any way constitute a forecast for the future evolution, or an assessment of the appropriate level, of the euro exchange rate.

$$
\begin{aligned}
E\left[(B_s)_{ij}\,|\Sigma\right] &= \begin{cases} 1 & \text{if } i=j \text{ and } s=1 \\ 0 & \text{otherwise} \end{cases} \\
cov\left((B_s)_{ij}\,,(B_r)_{hm}\,|\Sigma\right) &= \begin{cases} \lambda^2 \frac{1}{s^2}\frac{\Sigma_{ih}}{\psi_j/(d-n-1)} & \text{if } m=j \text{ and } r=s \\ 0 & \text{otherwise} \end{cases} ,
\end{aligned}
$$

Notice that the variance of this prior is lower for the coefficients associated with more distant lags, and that coefficients associated with the same variable and lag in different equations are allowed to be correlated. Finally, the key hyperparameter is λ - which controls the scale of all the variances and covariances, and effectively determines the overall tightness of this prior. The terms Σ_{ih}/Ψ_j account for the relative scale of the variables. The prior for the intercept, C, is non-informative.

We complement the Minnesota prior with two priors on the sum of the VAR coefficients, introduced as refinements of the Minnesota prior to further "favor unit roots and cointegration, which fits the beliefs reflected in the practices of many applied macroeconomists" (see Sims and Zha, 1998, p. 958). Loosely speaking, the objective of these additional priors is to reduce the importance of the deterministic component implied by VARs estimated conditioning on the initial observations (see Sims, 1992a and Giannone et al., 2016). The first of these two priors is known as no-cointegration (or, simply, sum-of-coefficients) prior. To understand what this prior entails, we rewrite the VAR equation in an error correction form:

$$
\Delta y_t = C + (B_1 + ... + B_p - I_N)y_{(t-1)} + A_1\Delta y_{t-1} + ... + A_p\Delta y_{(t-p)} + \epsilon_t
$$

where $A_s = -B_{s+1} - ... - B_p$. A VAR in first differences implies the restriction $\Pi = (B_1 + ... + B_p - I_N)$=0. Doan et al. (1984) introduced the no-cointegration prior which centered at 1 the sum of coefficients on own lags for each variable, and at 0 for the sum of coefficients on other variables' lags. This prior also introduces correlation among the coefficients on each variable in each equation. The tightness of this additional prior is controlled by the hyperparameter μ. As μ goes to infinity the prior becomes diffuse while, as it goes to 0, it implies the presence of a unit root in each equation.

The fact that, in the limit, the prior just discussed is not consistent with cointegration motivates the use of an additional prior on the sum of coefficients that was introduced by Sims (1993), and is known as dummy-initial-observation prior. This prior states that a no-change forecast for all variables is a good forecast at the beginning of the sample. The hyperparameter δ controls the tightness of this prior. As δ tends to 0, the prior becomes more dogmatic and all the variables of the VAR are forced to be at their unconditional mean, or the system is characterized by the presence of an unspecified number of unit roots without drift. As such, the dummy-initial observation prior is consistent with cointegration.

Summing up, the setting of these priors depends on the hyperparameters, λ, μ, δ and Ψ, which reflect the informativeness of the prior distributions for the model coefficients. In order to set these parameters, we closely follow the theoretically grounded approach proposed by Giannone et al. (2015). This involves treating the hyper-parameters as additional parameters, in the spirit of hierarchical modelling. As hyper-priors (i.e. prior distributions for the hyperparameters), we use the proper but almost flat distributions proposed in Giannone et al. (2015).[7] In this set-up, the marginal likelihood evaluated at the posterior mode of the hyperparameters is close to its maximum.

Finally, we also assess whether the cross-country interactions embedded in our model are beneficial to forecast the target variables. For this reason, we estimate BVAR models with prior distributions in the independent Normal/Inverse-Wishart class, that allow us to embed prior beliefs which underplay the extent of cross-country interactions. To parameterize the prior distributions, we modify the Minnesota prior described above by setting, for each variables of a specific country, the degree of shrinkage for all the lagged foreign variables to *half* the degree of shrinkage of the lagged domestic variables. Hence, this prior set-up pushes the coefficients of the lagged foreign variables more forcefully toward zero, a priori reducing the extent of cross-country interactions in the model. This procedure also breaks the natural conjugacy in the BVAR set-up, with two main effects.[8] First, the increased computational complexity of the model leads to a relevant slow-down in the estimation algorithms, especially for large models. We address this isse by using the algorithm of Carriero, Clark and Marcellino (2016), which is based on a simple triangularization of the VAR model and allows the simulation of the VAR coefficients equation by equation. Second, the break in conjugacy also implies that the marginal likelihood of the model is not available in closed form, which impedes the use of the algorithm of Giannone et al. (2015) for the set-up of the hyper-parameters describing the prior tigthness. We address this issue by setting the values of the hyper-parameters to the posterior mode of the values obtained by the Giannone et al. (2015)'s methodology, and then we halve the degree of overall shrinkage in the Minnesota prior, for the lagged foreign variables in each of the VAR equations. In the subsequent sections, we will refer to this model as to a "limited spillover" BVAR.

The main use of the model is to produce conditional forecasts. The methodology to produce them is based on Banbura et al. (2015), which in turn is based on the

[7]Specifically, as hyperpriors forλ, μ and δ, we choose Gamma densities with mode equal to 0.2, 1 and 1, the values recommended by Sims and Zha (1998), and standard deviations equal to 0.4, 1 and 1 respectively. Finally, the choice of the hyperprior for each element of the vector $\psi/(d-N-1)$, i.e. the prior mean of the main diagonal of Σ, should be loosely related to the scale of the variables in the model. We pick an Inverse-Gamma with scale and shape equal to $(0.02)^2$ because it seems appropriate for data expressed in annualized log-terms.

[8]See Koop and Korobilis (2010) and Korobilis and Pettenuzzo (2016) for a discussion of the advantages to break the conjugacy in Bayesian VARs with the Minnesota prior.

simulation smoother developed by Durbin and Koopman (2002).

2.3 The empirical exercises

We conduct two exercises to evaluate the ability of our model to provide an accurate cross-checking for the Eurosystem projections. In a first exercise, we evaluate over the sample 2006Q3 to 2016Q3 the forecasts of our target variables conditional on the *actual value* of the nine technical assumptions. This procedure gives rise to forecasts which are not replicable in real-time, because the future paths of the assumptions are unknown, at the stage of producing the forecasts. However, intuitively, this is the appropriate procedure to evaluate the accuracy of conditional forecasts. In fact, Clark and McCracken (2014) show that the traditional measures of forecast accuracy retain their statistical properties when evaluating conditional forecasts only if the latter are based on the true value of the assumptions. We compare our BVAR conditional forecasts to the BVAR unconditional forecasts, to gauge if and to what extent our model is able to extract information from the assumptions, and to the forecasts from a univariate benchmark (an $AR(5)$ model estimated with the same priors described above), to assess the degree of accuracy of our conditional BVAR forecasts. In addition, we also compare our BVAR conditional forecasts with those from a "limited spillover" BVAR model, to gauge whether capturing cross-country spillovers is beneficial for the forecasting accuracy of the model.

While being statistically sound, this exercise does not allow us to infer how our model would fare, in presence of the real-life forecasting error incurred when formulating the assumptions. In order to understand whether the model delivers reliable forecasts also when confronted with realistic forecast errors in the assumptions, in a second exercise, we condition our forecasts on the path of the assumptions used in real-time in the exercises conducted from 2011Q2 until 2016Q4, and we compare them to the BMPE projections on euro area GDP and HICP which were published as outcomes of those exercises.

3 Empirical Results

3.1 Evaluation based on actual assumption paths

Figure 1 reports the results of the evaluation of the BVAR conditional forecasts based on the actual value of the assumptions. In particular, the blue bars refer to the ratio of the mean squared forecast errors (MSFE) of the BVAR conditional forecasts relative either to the BVAR unconditional forecasts (left panel) or to the univariate benchmark model (right panel). Values smaller than one indicate that the conditional BVAR is more accurate than the competing model. The MSFEs are based on one year ahead forecasts and the evaluation sample ranges from 2006Q3 to 2016Q3.[9]

INSERT FIGURE 1

The BVAR conditional forecasts are generally more accurate than the BVAR unconditional forecasts, suggesting that the model is able to extract information from the assumptions. The BVAR conditional forecasts are also generally more accurate than the forecasts from the univariate benchmark model, which is traditionally hard to beat for inflation and real activity in the euro area, over most of the countries and variables. In order to give a more precise statistical account of the gains in forecast accuracy, we also performed the tests of predictive accuracy suggested in Diebold and Mariano (1995). The results of the tests should be considered only as suggestive, given that we compare nested models. Assuming a 10% threshold to characterize statistical significance of the differences in predictive accuracy, compared to the unconditional forecasts the gains in forecast accuracy are generally statistically significant for GDP, HICP, wages (excluding Spain) and lending rates across the four countries. For loans, the gain in forecast accuracy is generally not significant, while for investment and the GDP deflator the gains are significant only for France. In comparison with the autoregressive benchmark, the conditional BVAR forecasts are significantly more accurate for GDP (with the exception of Spain), HICP, lending rates, investment (except Germany) and the GDP deflator (except Germany). Instead, the conditional BVAR forecasts for wages and loans are not statistically different with respect to the autoregressive forecasts, across countries. By and large, also the analysis of the statistical accuracy of the results confirms the view that the conditional BVAR forecasts are generally more accurate than the benchmarks.

Figure 2 retains the same structure as figure 1, but it refers to density forecasts. In particular, it reports the ratio of the (average over the sample) Continuous Ranked Probability Scores (CRPS, see Gneiting and Raftery, 2007) of the conditional BVAR forecasts, over the corresponding measure for the unconditional BVAR and AR forecasts. Once again, a value smaller than one indicates that the conditional BVAR forecasts outperform the benchmarks and the results refer to the horizon of one-year ahead.

INSERT FIGURE 2

As for the point forecasts, the model generally outperforms the benchmarks. This ability to provide an accurate measure of the uncertainty surrounding the forecasts is important because one of the possible uses of the model is to highlight the gaps from the historical regularities, and the accuracy of density forecasts suggests that

[9] The results for the horizons of one quarter ahead and two years ahead give similar qualitative insight.

the model can reliably assess the statistical significance of such gaps. Also for the density forecasts, we perform a more formal analysis, to test the significance of the differences in forecasting accuracy across forecasts. In particular, we test the differences in CRPS, as suggested in Amisano and Giacomini (2007) and Gneiting and Ranjan (2011). The results on the statistical accuracy of the density forecasts are very similar to those for the point forecasts.

The evaluation just discussed covers the full sample under analysis. We also investigate the patterns over time, to assess the stability of the forecasting performance of the conditional BVAR forecasts. In particular, in figure 3 we report the mean squared forecasting errors of the conditional BVAR forecasts and of the AR forecasts, computed over overlapping rolling windows of five years.

INSERT FIGURE 3

Figure 3 shows that, generally, the rolling mean squared errors of the conditional BVAR forecasts lie below those from the autoregressive model. Moreover, for nominal variables (see, for example, wages), in most recent samples the BVAR tends to improve compared to the AR model. This reflects the fact that the multivariate nature of the model and the assumptions help to capture the recent low inflation/low wages environment and also the fact that, for more recent samples, we estimate the VAR model over a longer time-span, improving the estimation accuracy. Notice also that, while the BVAR rolling mean squared errors are generally more stable than those from the autoregressive model, they sometimes exhibit a large volatility which seems to reflect breaks in forecast accuracy. The most relevant cases are HICP and wages in France and Italy. The latter reflect the large errors in the forecasts of these variables that were made in the run-up and during the financial crisis of 2007-2008. When these errors drop from the computation of the rolling mean-squared error, we observe a discrete jump.

Although the out-of-sample performance of our model is generally good, both concerning point and density forecasts, the question still remains whether the multi-country approach we have taken, largely due to the institutional features of the Eurosystem projection exercises, is the most appropriate from the perspective of the forecasting performance of the model. In other words, are the cross-country spillovers captured in our model beneficial for its out-of-sample performance? In order to address this question, we estimate the "limited spillover" multi-country BVAR described at the end section 2.2, which factors in the a priori belief that, in the equation for each variables of a specific country, the lagged foreign variables are less likely to be relevant than the domestic ones. Figure 4 reports the results of the comparison of the forecasting performance of our baseline BVAR conditional forecasts with the "limited spillover" BVAR counterparts. The left panel of figure 4 reports the relative mean squared error, while the right panel report the relative CRPS. In both panels, a value lower than one favours our baseline model over the model with limited spillovers.

INSERT FIGURE 4

Figure 4 shows that, for Germany and France, the baseline model is quite univocally better than the model with more limited cross-country spillovers. For Italy, the evidence goes in a similar direction for what concerns the density forecasts (right panel), while the evidence for the point forecasts is mixed. For Spain, both in terms of point and density forecasts the evidence does not really point in one direction or the other. Interestingly, the Spanish economy is also the one for which the performance of our multi-country model, although still satisfactory, is less strikingly superior to the benchmarks. This result suggests that the Spanish economy is characterized by more idiosyncratic features compared to the other large economies of the euro area and, consequently, is also the one that benefits the less from the multi-country approach.[10]

Overall, our evaluation suggests that the BVAR model provides quite accurate point and density forecasts and, hence, it can be a valid benchmark to assess the consistency of the Eurosystem projections with the conditional assumptions. We now turn to investigate the properties of the model in a context that gets closer to the real-life practice of forecasting.

3.2 Evaluation based on real-time assumption paths

Rather than engaging in a full-fledged analysis of the forecasts as in the previous section, which might be plagued by the short sample for which real-time assumptions are available, we simply compare our conditional BVAR forecasts with a meaningful, real-time, benchmark: the BMPE projections themselves. In particular, we compare an aggregate of our country GDP and HICP forecasts (by using GDP weights) to the BMPE projections for the whole euro area, which have been regularly published. Notably, the yellow shaded area in figure 5 refers to the one-year ahead conditional BVAR density forecasts for GDP (left panel) and HICP (right panel) based on the assumptions available in *real-time* in the context of the BMPE exercises (16th to 84th quantiles of the distribution). In addition, the figure reports the corresponding BMPE projections for the whole euro area (solid black line), the conditional BVAR forecasts based the *actual* value of the assumptions (green dashed line) and the outcomes for the two variables (blue line with dots). The variables are reported in terms of year-on-year growth rates.

INSERT FIGURE 5

In general, while the BVAR forecasts based on real-time assumptions and the BMPE projections are quite accurate for GDP, a sizable positive bias appears for

[10]We also performed a forecasting evaluation, not reported here, in which we included loans among the assumptions, and we found that Spain is the only country to strongly profit from these additional assumptions.

HICP in the exceptional low inflation environment that has characterized the euro area and the global economy in the most recent part of the sample.

Remarkably, despite being produced by a mechanical model procedure without the inclusion of expert judgement and given the differences in the aggregation procedures, the BVAR conditional forecasts based on real-time assumptions capture quite well the features of the Eurosystem projections. Coupled with the evidence on the forecast accuracy in the previous sub-section, this further suggests that the BVAR described in this paper is a valid benchmark against which to cross-check the Eurosystem projections.

Notice also that, in the case of the HICP, the most marked gaps of the BMPE projections from the historical regularities captured by the BVAR conditional forecasts based on real-time assumptions appeared in the context of the exercises conducted between 2012 and 2014. In the occasion of those exercises, the BMPE projections turn out to be closer than the BVAR counterparts to the actual outcomes for HICP and, hence, the gaps prove beneficial for the accuracy of the BMPE projections. At the same time, the BMPE projections and the BVAR conditional forecasts based on real-time assumptions show a positive bias starting with the exercises conducted in mid-2012. The comparison with the BVAR forecasts conditional on the actual value of the assumptions reveals that a substantial part of the bias can be explained by the errors in the technical assumptions.

4 Conclusions

In this paper, we describe a multi-country BVAR for the four largest countries of the euro area and we show that the model provides accurate conditional forecasts of some policy relevant variables in the four countries.

The forecasting accuracy of the model and its ability to mimic the paths of the Eurosystem projections suggest that this model can provide a valid benchmark against which to assess the projections, both in terms of their accuracy and internal consistency.

References

[1] Altavilla, C. D. Giannone and M. Lenza (2016), The Financial and Macroeconomic Effects of the OMT Announcements, *International Journal of Central Banking,* 12(3), 29-57

[2] Amisano, G. and R. Giacomini (2007), Comparing Density Forecasts via Weighted Likelihood Ratio Tests, *Journal of Business and Economic Statistics,* 25, 177-190

[3] Banbura, M., D. Giannone and L. Reichlin (2010), Large Bayesian VARs, *Journal of Applied Econometrics,* 25, 71-92.

[4] Banbura, M., D. Giannone and M. Lenza (2015), Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections, *International Journal of Forecasting,* 31(3), 739-756.

[5] Canova, F., and M. Ciccarelli (2013), Panel vector autoregressive models: a survey, *ECB Working papers 1507.*

[6] Capolongo, A., and C. Pacella (2018), Does a large multi-country model improve the euro area inflation forecasts?, *mimeo, ECARES-ULB.*

[7] Carriero, A., T., Clark and M. Marcellino (2016), Large Vector Autoregressions with Stochastic Volatility and Flexible Priors, *Working Paper 1617, Federal Reserve Bank of Cleveland.*

[8] Chudik, A., and M. H. Pesaran (2016), Theory And Practice of Gvar Modelling, *Journal of Economic Surveys,* 30(1),165-197.

[9] Clark, T. E., and M. W. McCracken (2014), Evaluating Conditional Forecasts from Vector Autoregressions, *Working Papers 2014-25, Federal Reserve Bank of St. Louis.*

[10] Del Negro, M. and F. Schorfheide (2013), DSGE Model-Based Forecasting, *Handbook of Economic Forecasting, Elsevier*, 2-57.

[11] De Mol, C., Giannone, D. and L. Reichlin (2008), Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?, *Econometric Reviews,* 3, 1-100.

[12] Diebold, F. and R. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics,* 13(3), 253-263.

[13] Doan, T., R. Litterman, and C. A. Sims (1984), Forecasting and conditional projection using realistic prior distributions, *Journal of Econometrics,* 146(2), 318-328.

[14] Domit, S., F. Monti and A. Sokol (2016), A Bayesian VAR benchmark for COMPASS, *Bank of England, Staff Working Paper,* 583.

[15] Durbin, J., and S. Koopman (2002), A simple and efficient simulation smoother for state space time series analysis, *Biometrika,* 89(3), 603-615.

[16] European Central Bank (2001), A guide to Eurosystem staff macroeconomic projection exercises.

[17] European Central Bank (2016), A guide to the Eurosystem/ECB staff macroeconomic projection exercises.

[18] Fagan, G., and J. Morgan (2005), Econometric models of the euro-area central banks, *Edward Elgar Publishing.*

[19] Faust, J., and J. Wright (2008), Efficient forecast tests for conditional policy forecasts, *Journal of Econometrics,*146, pp.293-303.

[20] Fawcett, N., L. Korber, R. Masolo and M. Waldron (2015), Evaluating UK Point and Density Forecasts from an Estimated DSGE Model: The Role of Off-Model Information Over the Financial Crisis, *Bank of England, Staff Working Paper,* 538.

[21] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2002), The Generalized Factor Model: Identification And Estimation, *The Review of Economics and Statistics,* 82, 540-554.

[22] Giannone, D., M. Lenza, D. Momferatou and L. Onorante (2014), Short-term inflation projections: A Bayesian vector autoregressive approach, *International Journal of Forecasting,* 30(3), pp. 635-644.

[23] Giannone, D., M. Lenza and G. Primiceri (2015), Prior selection for vector autoregressions *The Review of Economics and Statistics,* 97(2), 436-451.

[24] Giannone, D., M. Lenza and G. Primiceri (2016), Priors for the Long Run, *Journal of the American Statistical Association,* forthcoming.

[25] Gneiting, T. and A. E. Raftery (2007), Strictly Proper Scoring Rules, Prediction and Estimation, *Journal of the American Statistical Association,* 102, 359-378.

[26] Gneiting, T. and R. Ranjan (2011), Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules, *Journal of Business and Economic Statistics,* 29(3), 411-422.

[27] Herbst, E. and F. Schorfheide (2012), Evaluating DSGE model forecasts of comovements, *Journal of Econometrics,* 171(2), pp. 152-166.

[28] Iversen, J., S. Laséen, H. Lundvall and U. Söderström (2016), Real-Time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank, *Sveriges Riksbank Working Paper Series,* 318.

[29] Koop, G. and D. Korobilis (2010), Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, *Foundations and Trends in Econometrics,* 3, 267-358.

[30] Korobilis, D. and D. Pettenuzzo (2016), Adaptive Minnesota Prior for High-Dimensional Vector Autoregressions, *Essex Finance Centre Working Papers,* 18626.

[31] Lenza, M. and J. Slacalek (2018), How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area, *ECB Working papers 2190.*

[32] Litterman, R. (1979), Techniques of Forecasting Using Vector Autoregressions, *Federal Reserve of Minneapolis working paper,* 115.

[33] Sims, C. A. and T. Zha. (1998), Bayesian Methods for Dynamic Multivariate Models, *International Economic Review,* 39, 949-968.

[34] Sims, C. A. (1992), Bayesian Inference for Multivariate Time Series with Trend *mimeo, Princeton University.*

[35] Sims, C. A. (1993), A Nine-Variable Probabilistic Macroeconomic Forecasting Model, *in James H. Stock and Mark W. Watson, eds., Business Cycles, Indicators and Forecasting,* 179-212.

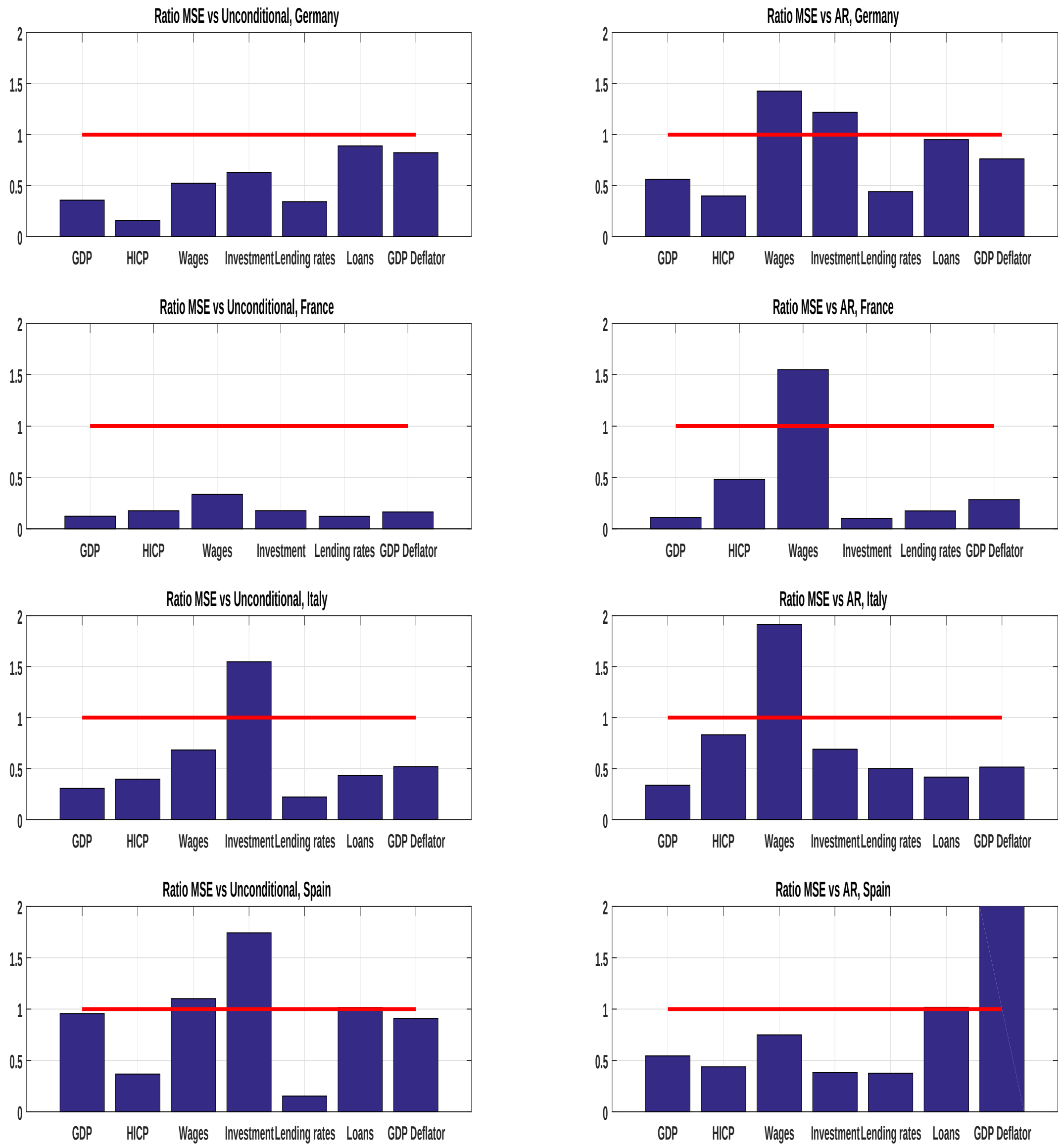
[36] Stock, H. and J. Watson (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics,* 20(2), 147-162.

A Data appendix

No.	Country	Short Name	Description	Source	Target/Assump.
1	Germany	YER-DE	Gross Domestic Product	Eurostat	T
2	Germany	HIC-DE	Harmonised index of consumer prices	Eurostat	T
3	Germany	CEX-DE	Labour compensation per employee	Eurostat	T
4	Germany	ITR-DE	Gross fixed capital formation	Eurostat	T
5	France	YER-FR	Gross Domestic Product	Eurostat	T
6	France	HIC-FR	Harmonised index of consumer prices	Eurostat	T
7	France	CEX-FR	Labour compensation per employee	Eurostat	T
8	France	ITR-FR	Gross fixed capital formation	Eurostat	T
9	Italy	YER-IT	Gross Domestic Product	Eurostat	T
10	Italy	HIC-IT	Harmonised index of consumer prices	Eurostat	T
11	Italy	CEX-IT	Labour compensation per employee	Eurostat	T
12	Italy	ITR-IT	Gross fixed capital formation	Eurostat	T
13	Spain	YER-ES	Gross Domestic Product	Eurostat	T
14	Spain	HIC-ES	Harmonised index of consumer prices	Eurostat	T
15	Spain	CEX-ES	Labour compensation per employee	Eurostat	T
16	Spain	ITR-ES	Gross fixed capital formation	Eurostat	T
17	Germany	TTN-NFCN-LONG-DE	Bank lending rates to firms (non-financial corporations)	ECB	T
18	France	TTN-NFCN-LONG-FR	Bank lending rates to firms (non-financial corporations)	ECB	T
19	Italy	TTN-NFCN-LONG-IT	Bank lending rates to firms (non-financial corporations)	ECB	T
20	Spain	TTN-NFCN-LONG-ES	Bank lending rates to firms (non-financial corporations)	ECB	T
21	Euro Area	POILU	Oil price in US dollar	ECB	A
22	Euro Area	EXR	Nominal exchange rate of euro against US Dollar	ECB	A
23	Euro Area	STN	3 month EURIBOR	ECB	A
24	Germany	U2-LNFC-DE	MFI, Loans to firms (non-financial corporations)	ECB	T
25	Spain	U2-LNFC-ES	MFI, Loans to firms (non-financial corporations)	ECB	T
26	Euro Area	U2-LNFC-ECB	MFI, Loans to firms (non-financial corporations)	ECB	T
27	Italy	U2-LNFC-IT	MFI, Loans to firms (non-financial corporations)	ECB	T
28	Germany	YED-DE	GDP expenditure deflator	Eurostat	T
29	Spain	YED-ES	GDP expenditure deflator	Eurostat	T
30	France	YED-FR	GDP expenditure deflator	Eurostat	T
31	Italy	YED-IT	GDP expenditure deflator	Eurostat	T
32	Germany	WDREX-DE	World demand indicator, non-euro area	ECB	A
33	Spain	WDREX-ES	World demand indicator, non-euro area	ECB	A
34	France	WDREX-FR	World demand indicator, non-euro area	ECB	A
35	Italy	WDREX-IT	World demand indicator, non-euro area	ECB	A
36	US	YER-US	Gross Domestic Product	FRED	A
37	US	STN-US	3 month deposit rate LIBOR	FRED	A

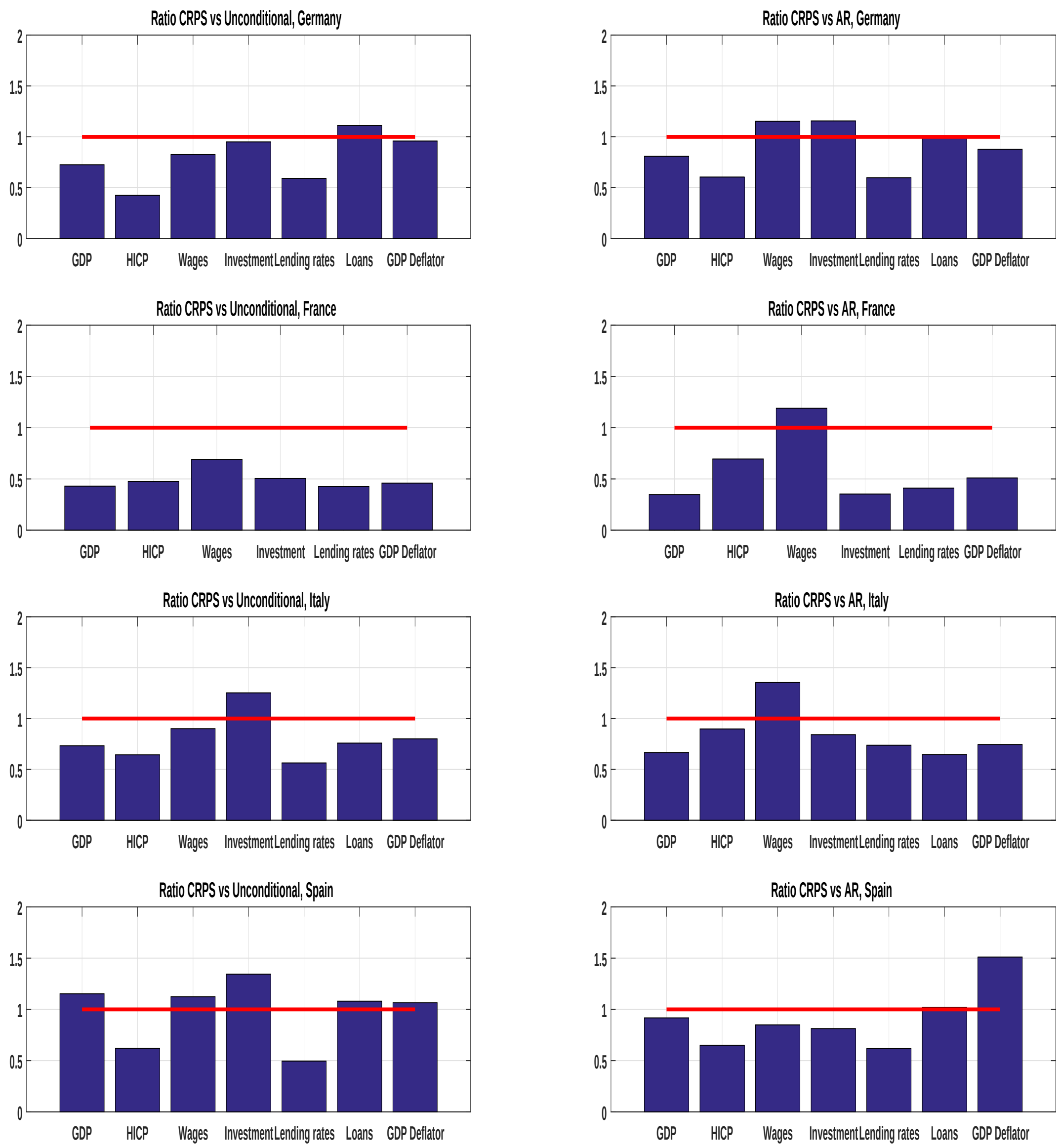
B Figures

Figure 1: Evaluation of point forecasts, versus Unconditional BVAR and AR forecast



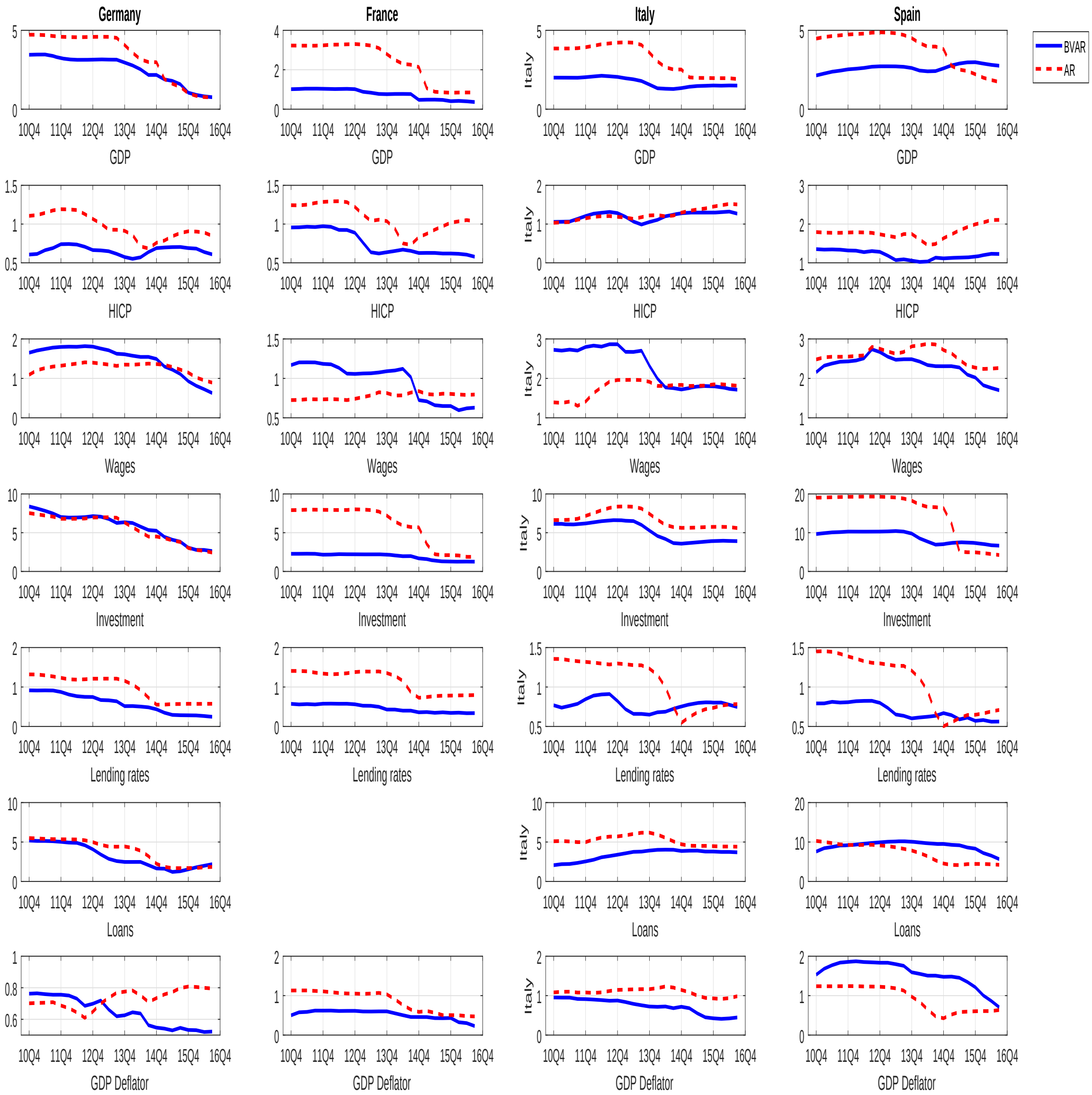
Note: Ratio of Mean Squared Forecast errors. Values smaller than one indicate that the BVAR conditional forecasts out-perform the benchmark. For France, loans not available.

Figure 2: Evaluation of density forecasts, versus Unconditional BVAR and AR forecast



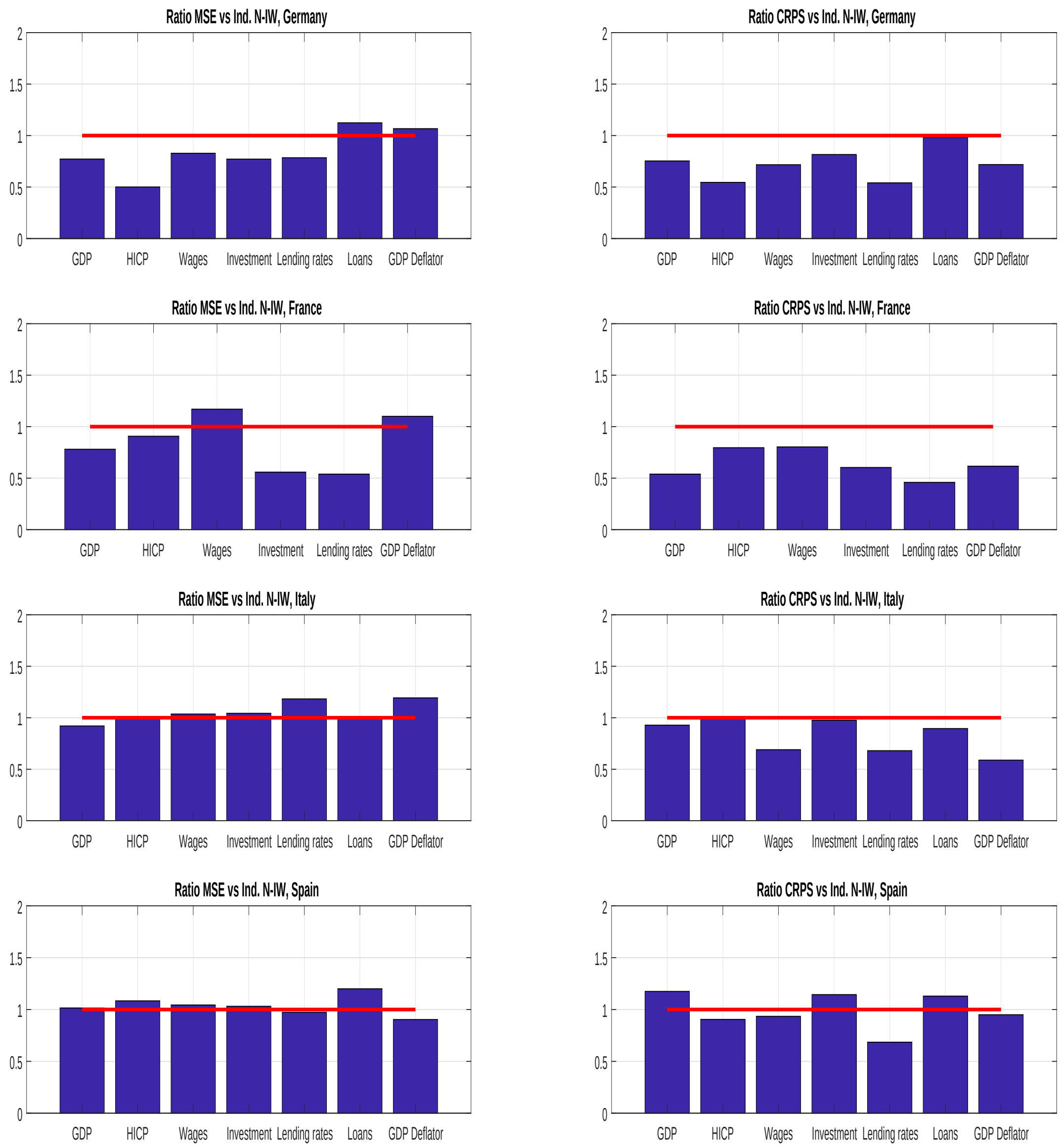
Note: Ratio of average Continuous Ranked Probability Scores. Values smaller than one indicate that the BVAR conditional forecasts out-perform the benchmark. For France, loans not available.

Figure 3: Rolling MSFE for all the target variables



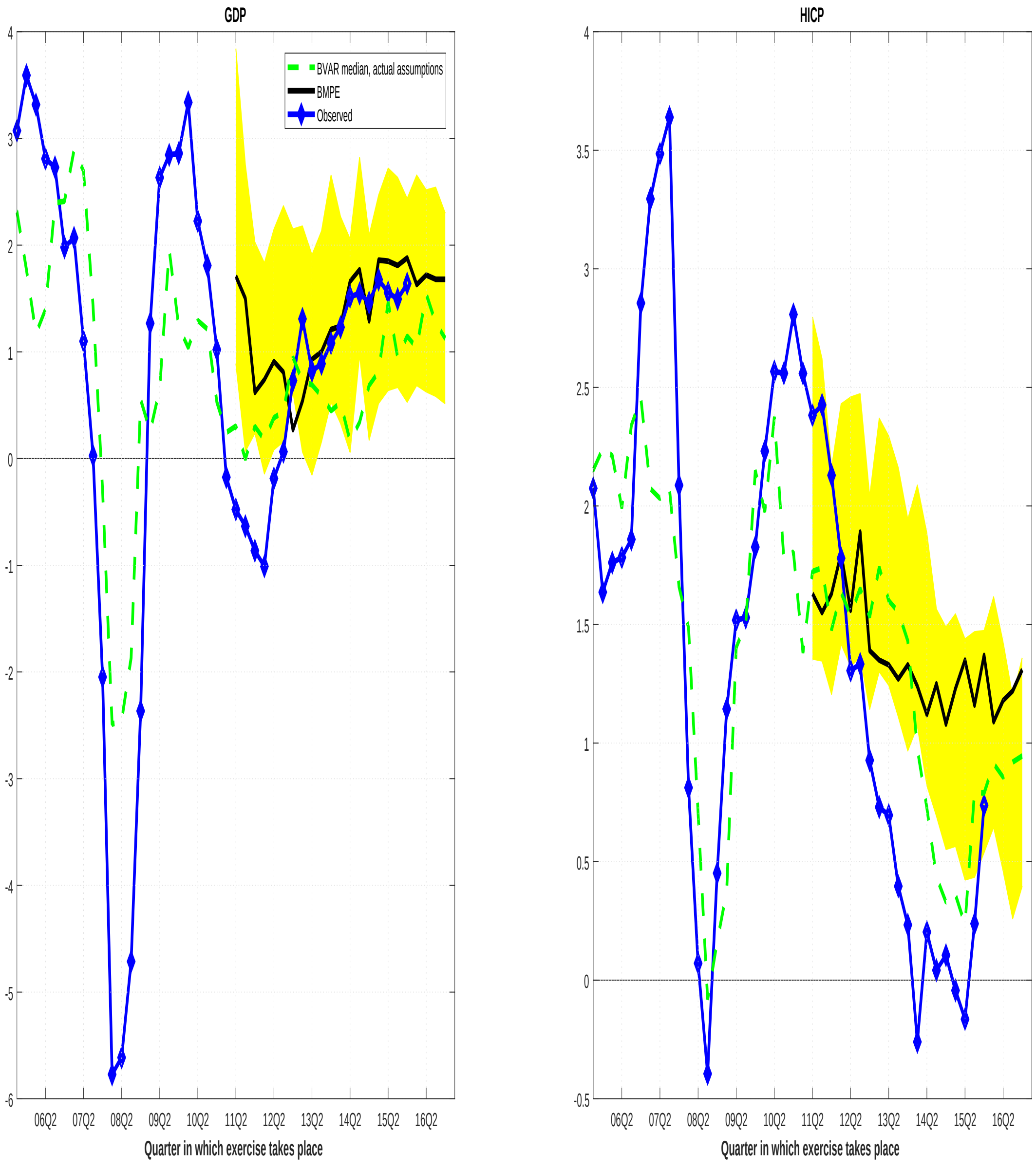
Note: Rolling MSFE, computed over the five years ending in the date indicated on the horizontal axis. BVAR (blue solid line) and AR model (red dashed line). Germany (first column), France (second column), Italy (third column) and Spain (fourth column). For France, loans not available.

Figure 4: Evaluation of point and density forecasts, versus BVAR with Independent Normal-IW prior



Note: Ratio of Mean Squared Forecast errors (left panel) and ratio of average Continuous Ranked Probability Scores (right panel). In both panels, values smaller than one indicate that the BVAR conditional forecasts out-perform the conditional forecasts from the BVAR model with Independent Normal-IW priors that limit the amount of spillovers across countries. For France, loans not available.

Figure 5: BVAR and Eurosystem projections



Note: The horizontal axis reports the date of the exercises in which a projection was produced. The projections refer to an horizon of four quarters ahead (black solid: BMPE and green-dashed: median of the BVAR with assumptions at the actual values). The yellow shaded area refers to the 16th-84th quantiles of the conditional BVAR forecast distribution, evaluated with real-time assumptions. The observed outcomes (blue line with diamonds) refers to four quarters after the date indicated on the horizontal axis. All the variables are reported in terms of year-on-year growth rates.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank. Please address any comments to Elena Angelini, Magdalena Lalik, Michele Lenza, or Joan Paredes.

Elena Angelini
European Central Bank, Frankfurt am Main, Germany; email: elena.angelini@ecb.europa.eu

Magdalena Lalik
European Central Bank, Frankfurt am Main, Germany; email: magdalena.lalik@ecb.europa.eu

Michele Lenza
European Central Bank, Frankfurt am Main, Germany; email: michele.lenza@ecb.europa.eu

Joan Paredes
European Central Bank, Frankfurt am Main, Germany; email: joan.paredes@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-3489-3 ISSN 1725-2806 doi:10.2866/028355 QB-AR-19-008-EN-N