BANK OF ENGLAND

Staff Working Paper No. 583

A Bayesian VAR benchmark for COMPASS

Sílvia Domit, Francesca Monti and Andrej Sokol

January 2016

BANK OF ENGLAND

Staff Working Paper No. 583

A Bayesian VAR benchmark for COMPASS

Sílvia Domit,[(1)] Francesca Monti[(2)] and Andrej Sokol[(3)]

Abstract

We estimate a Bayesian VAR analogue to the Bank of England's DSGE model (COMPASS) and assess their relative performance in forecasting GDP growth and CPI inflation in real time between 2000 and 2012. We find that the BVAR outperformed COMPASS when forecasting both GDP and its expenditure components. In contrast, the performance of these models was similar when forecasting CPI. We also find that, despite underpredicting inflation at most forecast horizons, the BVAR density forecasts outperformed those of COMPASS. Both models overpredicted GDP growth at all forecast horizons, but the BVAR outperformed COMPASS at forecast horizons up to one year ahead. The BVAR's point and density forecast performance is also comparable to that of a Bank of England in-house statistical suite for both GDP and CPI inflation and to the *Inflation Report* projections. Our results are broadly consistent with the findings of similar studies for other advanced economies.

Key words: Forecasting, Bayesian VARs, macro-modelling.

JEL classification: C53, E12, E17.

(1) Bank of England. Email: silvia.domit@bankofengland.co.uk
(2) Bank of England. Email: francesca.monti@bankofengland.co.uk
(3) Bank of England. Email: andrej.sokol@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are grateful to Nicholas Fawcett, Lena Körber, Riccardo Masolo and Matt Waldron for sharing their codes and for insightful comments. We also thank Richard Harrison, Garry Young and an anonymous referee for insightful comments.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

COMPASS is the Bank of England's 'New-Keynesian' dynamic stochastic general equilibrium (DSGE) model. It is used as a central organising macroeconomic model in the MPC's forecasting platform. It provides the basic set of relationships that articulate core macroeconomic mechanisms and provides a disciplining framework by ensuring that forecasts are internally consistent. The MPC's forecasting platform also includes a suite of 50 forecasting models, covering a range of different frameworks and ways of thinking about the economy. Whereas some of these models will articulate economic channels that are omitted from COMPASS, others will generate alternative forecasts for the same variables as COMPASS, providing a cross-check on the central model's forecast.[1]

In this paper, we introduce a new addition to the Bank of England's forecasting toolkit: a Bayesian VAR with the same set of macroeconomic variables as COMPASS. As in COMPASS, we treat the UK as a small open economy and model the rest of the world as exogenous. We then assess the relative performance of the two models in forecasting UK GDP growth and inflation. Comparing COMPASS' dynamic properties and forecasting accuracy with those of more data-driven benchmarks such as VARs can be fruitful for a number of reasons. In particular, DSGE models place a great number of restrictions on the time-series behaviour of the variables they seek to explain and forecast. And their size poses challenges to both estimation and specification analysis, which entails risks for the reliability of their forecasts. We opt for Bayesian rather than classic estimation because both the large number of variables and the limited sample size available make precise classical estimation of an unrestricted VAR unfeasible. Also, Bayesian estimation allows us to easily produce forecast densities. We use a combination of commonly used priors and select the tightness by maximising the marginal likelihood, which tends to improve forecasting performance. To reflect the information available at the time the forecasts would have been produced, both models are re-estimated between 2000 and 2012 using real-time data.

We find that the BVAR's point and density forecasts generally outperformed COMPASS. The BVAR outperformed COMPASS when forecasting both GDP and its expenditure components 1 and 2 years ahead. In contrast, the performance of these models was similar when forecasting CPI inflation. Despite under-predicting inflation at most forecast horizons, the BVAR density forecasts outperformed COMPASS'. Both models over-predicted GDP growth at all forecast horizons, but the BVAR outperformed COMPASS at forecast horizons up to 1 year ahead. Our results are broadly consistent with the findings of similar studies for other advanced economies. Christoffel *et al.* (2011) and Iversen et al. (2014) assess the forecasting performance of various models for the euro area and Sweden, respectively. In line with the findings for the UK, these studies show that forecasts produced by DSGE models are generally outperformed by those produced with medium-scale BVARs. These findings differ from earlier results in the literature, such as Smets and Wouters (2007) and Gurkay-

[1]For more information on the MPC's forecasting platform, see Burgess et al (2013).

nak *et al.* (2013), who found that in the U.S. prior to the crisis, DSGE models appeared to outperform BVARs.[2]

This paper is also closely related to the analysis in Fawcett *et al.* (2015). In that paper the authors compare the real-time forecast performance of COMPASS against statistical and judgemental benchmarks and find that, particularly during the crisis, COMPASS fares worse. We show that the BVAR's forecasting performance is closer to the statistical and judgmental forecasts than are COMPASS' forecasts. However Fawcett *et al.* (2015) point out that when COMPASS is augmented to include a survey measure of short-term GDP growth expectations, which accounts for *timely* off-model information, COMPASS's forecasts are competitive with the judgemental forecasts at all horizons in the post-crisis period.

The paper is organised in four sections. The first introduces the BVAR model, whereas the others compare both point and density forecasts from COMPASS and the BVAR, as well as the *Inflation Report* and the Bank's suite of forecasting models.

2 A Bayesian VAR for the UK economy

2.1 The model

We consider a reduced-form vector autoregression (VAR) model of the observable variables in COMPASS, namely: real GDP, real total consumption, real business investment, real total government expenditure, export volumes, import volumes, export prices, import prices, total hours worked, nominal wages, CPI, nominal exchange rate, nominal interest rate, world demand for imports and world export prices.

In COMPASS, the joint behaviour of these variables is determined by the restrictions imposed by the structure of the model. COMPASS features 5 types of economic agents: households, firms, the government, the rest of the world and the monetary policymaker. There are two types of households: the "unconstrained" households, who have access to financial markets and can accumulate assets, and the "constrained" households, who instead do not have access to financial markets and thus cannot save, but rather spend all their labour income on consumption. Firms operate in monopolistic competition and are subject to costs when adjusting their prices, as in Rotemberg (1982). The monetary authority sets interest rates according to a Taylor-type rule, while an exogenous foreign block drives world demand for imports and world export prices. The model also includes habit formation, investment adjustment costs and wage rigidities.[3]

In the BVAR, on the other hand, we place very few restrictions on the joint behaviour of the series. Except for our priors (see next subsection), the only other restriction we impose is that, as in COMPASS, we treat the UK as a small open econ-

[2] Gurkaynak *et al.* (2013) also suggest that even simpler models such as ARs and small unrestricted VARs can outperform both DSGEs and BVARs if the focus is mainly on forecasting inflation, GDP growth and interest rates. However, in this study we're interested in the forecasting performance for all observables in COMPASS.

[3] For more information on COMPASS, see Burgess et al. (2013).

omy, so the two world variables are treated as block exogenous, such that they are not affected by the UK business cycle. Our empirical specification therefore takes the following form:

$$Y_t = C + A_1 Y_{t-1} + ... + A_p Y_{t-p} + B_1 Z_t + \epsilon_t; \text{ for all } t = 1...T \tag{1}$$

where Y_t is an $n \times 1$ vector of endogenous domestic variables, C is an $n \times 1$ vector of constants, A_1 and A_p are $n \times n$ matrices of coefficients for the first and *p-th* lags of the endogenous variables, Z_t is a $k \times 1$ vector of exogenous world variables, B_1 is an $n \times k$ matrix of coefficients on the contemporaneous exogenous world variables and ϵ_t is an $n \times 1$ vector of iid residuals with mean 0 and covariance matrix Σ.

2.2 Estimation

We opt to estimate the model with Bayesian techniques, as it is well known that working with flat priors can yield poor inference in large dimensional systems. Litterman (1980) and Doan, Litterman, and Sims (1984) show that, combining the likelihood function with some informative prior distributions, improves forecasting performance. The frequentist interpretation of this result is that these priors are successful because they effectively reduce the estimation error, while generating only relatively small biases in the estimates of the parameters.

The BVAR-X model is specified in first differences, with the exception of the nominal interest rate, which is treated as stationary. Whereas estimating the VAR in differences rather than in levels may miss relevant model dynamics (such as potential long run relationships between the variables), Carriero *et al.* (2011) show that modelling the VAR in first differences generally improves forecast accuracy. And it also has the advantage of providing an easier mapping with the variables in COMPASS, with which we aim to be aligned as much as possible.

We use a combination of the conjugate priors most commonly used in the literature: the 'Minnesota' prior introduced by Litterman (1980), the 'sum-of-coefficients' prior, proposed by Doan, Litterman and Sims (1984) and the "dummy-initial- observations" prior (Sims, 1992). The Minnesota prior is based on the assumption that each variable follows a random walk process, possibly with drift, or a white noise process when the variable is stationary. The 'sum-of-coefficients' prior and the "dummy initial observations" priors are refinements to the Minnesota priors intended to improve forecasting performance by reducing the importance of the deterministic component in VARs estimated using initial observations. Both of these priors are implemented adding dummy observations as in Banbura *et al.* (2010). We centre the prior for B, the matrix of coefficients that govern the impact of the exogenous variables on the domestic variables, around zero, and we implement the prior using dummy observations.

We select the informativeness of each of these priors by choosing the tightness of the priors that maximise the marginal likelihood of the model, as suggested in Giannone *et al.* (2015). Conveniently our BVAR-X with conjugate priors maintains the property that the marginal likelihood is available in closed form and can be therefore easily maximised numerically. The marginal likelihood is a measure of out-of-sample forecasting performance of a model, so selecting the tightness of the priors to maximise the marginal likelihood is akin to selecting them according to the one-step-ahead out-of-sample forecasting ability of the model.

This method for prior selection outperforms other commonly used procedures, such as the one described in Litterman (1980), where the tightness of the prior is chosen by maximising the out-of-sample forcasting performance of the model over a pre-sample, and the procedure in Banbura *et al.* (2010), where the priors are chosen by maximising the models' in-sample fit. The procedure in Giannone *et al.* (2015) also addresses the trade-off between model complexity and in-sample fit, as it yields looser priors when the model involves few unknown coefficients relative to the size of the dataset and *viceversa*). This more careful choice of priors might explain why our results differ from earlier studies (for example Smets and Wouters (2007) and Gurkaynak *et al.* (2013)) in finding that our BVAR generally outperforms COMPASS. We are investigating the role of prior selection in driving forecasting performance in ongoing research work.

The lag length for the VAR, (p), is set to 2. We have experimented with different values and have chosen it based on the BVAR's out-of-sample forecasting performance and the model's dynamic properties. Whereas it is possible to choose the lag length optimally during estimation by including it as an argument of the maximisation problem, selecting the lag length somewhat arbitrarily is not unprecedented in the literature (see Giannone *et al.* (2015)). Also, the results in Carriero *et al.* (2011) suggest that shorter lags tend to yield better forecasting performance.

2.3 Data

COMPASS and the BVAR are estimated with real-time data and have comparable information sets. The real-time estimation approach means that each forecast is produced only with information that would have been available at each forecast round. We use the dataset from Fawcett *et al* (2015), which consists of data released about 3-4 weeks-ahead of each *Inflation Report*. The dataset starts in 1987-Q2 and we produce real-time forecasts for the three-year forecast period for every quarterly forecast round between Feb-2000 and Nov-2012.

The off-model constraint quarter forecast based on high-frequency data is imposed as judgement in both the BVAR and COMPASS. The off-model forecasts for the exogenous world variables come from different sources: for COMPASS, these come from an exogenous block attached to the main model; for the BVAR, we use internal forecasts produced by Bank staff. Although the information sets for the exogenous variables are different, this does not drive the difference in forecasting performance between both models. In Appendix A, we consider the case where both models are conditioned on the same set of world forecasts produced by Bank Staff and show that their relative performance is unchanged.

3 Forecast evaluation

3.1 Central forecasts

To assess the forecasting performance of the BVAR, we start by measuring the accuracy of its central forecasts. Figure 1 plots the real-time central forecasts for both COMPASS and the BVAR between 2000 and 2012 against the data outturns[4]. Perhaps the most interesting feature of Figure 1 is the fundamentally different shape of successive GDP growth forecasts in the two models: whereas the BVAR essentially forecasts mean reversion, COMPASS' forecasts feature a pronounced over-shoot before reverting to the mean. This is due to the restriction, characteristic of DSGE model but not imposed in the BVAR, that most shocks[5] be neutral on the *level* of real GDP, such that deviations of GDP *growth* from the mean attributed to them need to over-unwind. Also, all variables in COMPASS have constant long-run growth rates. The long-run growth rate of GDP is set equal to the sample average in each recursively estimated variant. Those sample averages tend to exceed the growth rates observed in the data after the crisis.

To assess the accuracy of these forecasts, for each model we compute the root mean squared forecast errors (RMSFE) in the out-of-sample forecasting period (2000-2012) for each forecast horizon up to three years. Formally, the RMSFE at horizon *h* is given by:

$$RMSFE^h = \sqrt{\frac{1}{P-h+1}\sum_{t=R}^{T-h}\widehat{u}_{t+h}^2} \tag{2}$$

where the forecast error $\widehat{u}_{t+h}$ at forecast origin t and forecast horizon h is defined as the difference between the data and the mean forecast. It is computed based on the first Quarterly National Accounts and on CPI data (which do not get revised).

To assess whether the RMSFEs of the two models are statistically different on average over the out-of-sample period, we use the Diebold-Mariano test, as described in Fawcett *et al.* (2015):

$$DM_h = \frac{1}{\sqrt{P-h+1}}\sum_{t=R}^{T-h}\frac{\widehat{u}_{1,t+h}^2 - \widehat{u}_{2,t+h}^2}{\sqrt{\widehat{\Sigma}}} \tag{3}$$

where $\widehat{\Sigma}$ is an estimate of the long-run variance. Under the null hypothesis H_0 : $E(\widehat{u}_{1,t+h}^2 - \widehat{u}_{2,t+h}^2) = 0$, DM_h converges to a normal distribution.

3.1.1 GDP growth

The BVAR outperformed COMPASS when forecasting aggregate GDP and its main expenditure components. The BVAR significantly outperformed COMPASS for 1-year-

[4] As Del Negro *et al.* (2014), we do not show the forecasts for the policy rate, but we note that for both models the interest rate projection does not violate the zero lower bound constraint on nominal interest rates.

[5] In COMPASS, all but the labour-augmenting productivity shock.

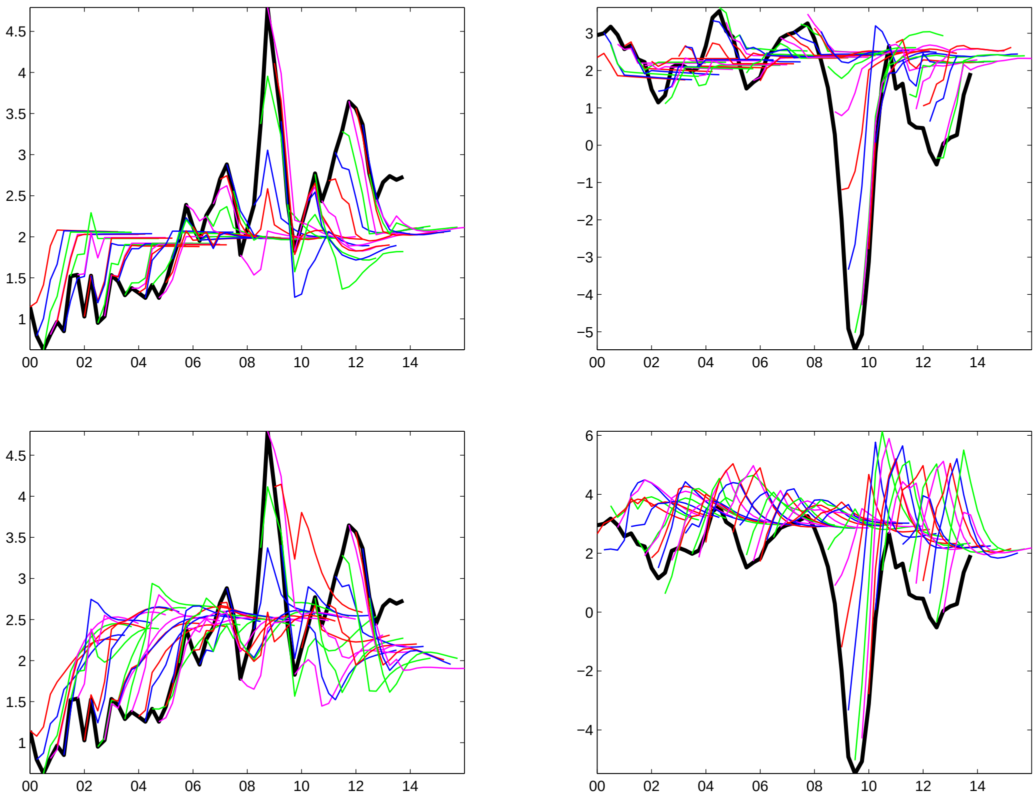

Figure 1: Out-of-sample forecast vintages (coloured lines) vs data (black line), 2000-2012. Annual CPI inflation rate (left column) and Quarterly GDP growth rate (right column) BVAR (top row), COMPASS (bottom row)

ahead annual GDP growth forecasts, with an average RMSFE of just over 2pp (1pp lower than COMPASS') between 2000 and 2012.[6] The BVAR performed better than COMPASS in forecasting GDP at the 1-year forecast horizon throughout the sample, including during the 2008-09 crisis. And the gap between the two models' performance widened in the most recent period, with the BVAR's RMSFE nearly 2pp lower than COMPASS'. The BVAR also outperformed COMPASS at the 2-year horizon, but the difference in RMSFE was smaller and not statistically significant (Figure 2 and Appendix B).

The BVAR's better forecasting performance for GDP growth was mirrored by key expenditure components (Figure 3). The difference was particularly stark for business investment, with the BVAR's average RMSFE (of around 9pp) nearly 50 percent lower than COMPASS' at both the 1 and 2-year ahead horizons.[7] The gap in forecasting performance was particularly large in the earlier part of the sample, with both models performing similarly since the 2008-9 crisis.

[6] The average RMSFE for both models is considerably lower when the 2008-09 figures are excluded.
[7] Although only the 2-year ahead difference was statistically significant, see Appendix B.

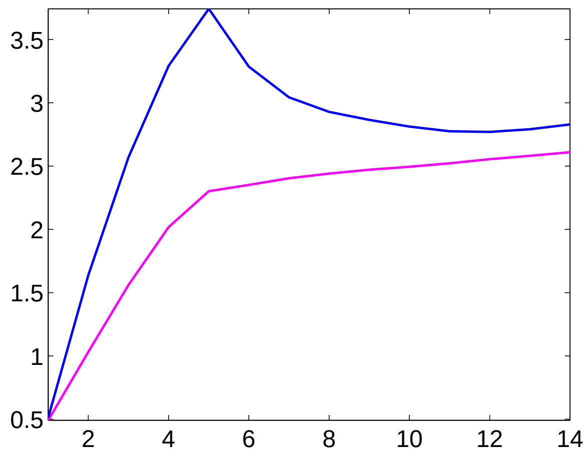

Figure 2: Quarterly GDP growth root mean squared forecast errors at different forecast horizons, 2000-2012. BVAR (pink), COMPASS (blue)

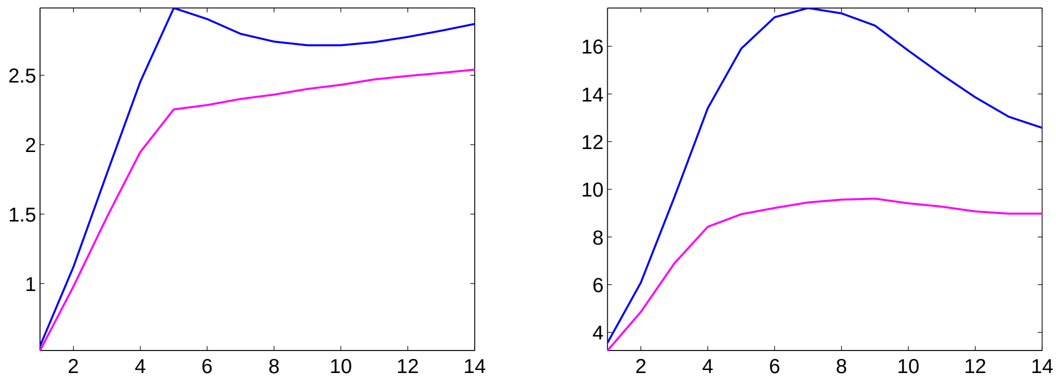

Figure 3: Root mean squared forecast errors at different forecast horizons, 2000-2012. Real private consumption annual growth rate (left column) and Real business investment annual growth rate (right column) BVAR (pink), COMPASS (blue)

3.1.2 Inflation

In contrast to the GDP results, the performance of these models was similar when forecasting CPI inflation. Both COMPASS and the BVAR registered an average absolute forecast error of around 0.9pp for annual CPI inflation at both the 1- and 2-year forecast horizons between 2000 and 2012 (figure 4).

Although COMPASS' and the BVAR's inflation forecasts performed similarly on average between 2000 and 2012, these results changed over time. COMPASS performed worse early on, but better post crisis, with a RMSFE about 0.3pp lower than the BVAR's from 2006 onwards at the two-year forecast horizon.

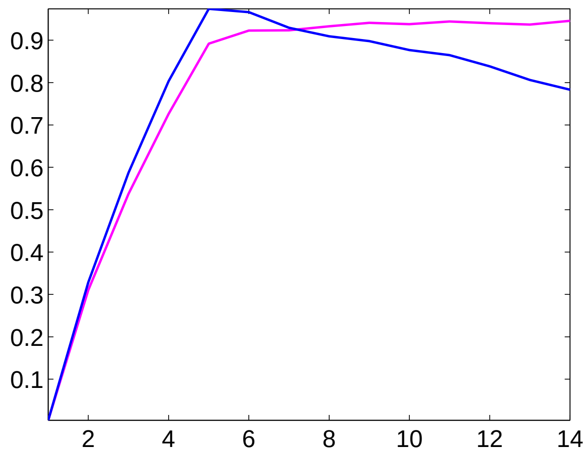

Figure 4: Annual CPI inflation root mean squared forecast errors at different forecast horizons, 2000-2012. BVAR (pink), COMPASS (blue)

3.2 Density forecasts

In this section, we assess the performance of the density forecasts from COMPASS and the BVAR, both relative to each other and to data outturns.[8] Because the forecasts for world variables are imposed exogenously as conditioning paths in the BVAR, with no uncertainty around them. A natural extension would be to allow for uncertainty when imposing conditioning paths, perhaps following the algorithm in Del Negro and Schorfheide (2012).

Forecast densities (or fan charts) are used to describe the degree of uncertainty around the central forecasts. But assessing the performance of a density forecast is less trivial than for point forecasts, because we only observe one realised value of the variable of interest in each period, as opposed to its entire distribution. In this paper, we follow Fawcett *et al.* (2015) and employ the following two methods:

a) To assess the accuracy of each model's density forecast relative to the data, we use probability integral transformations (PITs), which measure the probability of observing a given realised outturn (or lower) in each forecast density. Formally, the PIT for the observation Y at forecast horizon h for a given forecast density at time t is defined as

$$z_{h,t} = \int_{\infty}^{Y_{t+h}^{0}} f_t(Y_{t+h})dY_{t+h} \tag{4}$$

If the probability is close to zero, the fan chart would be 'above' the realised value, suggesting the density over-predicted that variable. Likewise, very high probabilities would indicate under-prediction. If a set of forecast densities offer a good approximation to the true underlying density, then the PITs should be evenly distributed over all the percentiles.

[8]Note that COMPASS' density forecasts are not the same as the MPC's (judgemental) fan charts published in the *Inflation Report*.

b) To compare the accuracy of density forecast from the BVAR with other models, we use average logarithmic scores defined as:

$$\frac{1}{P-h+1}\sum_{t=R}^{T-h}\log f_t(Y_{t+h}) \tag{5}$$

These take a high value if the forecast density assigns a high probability to the actual outturn. We test whether the log scores from various models are statistically different from each other on average over the out-of-sample period by using the likelihood ratio test of Amisano and Giacomini (2007). The test statistic is given by:

$$AG_h = \frac{1}{\sqrt{P-h+1}}\sum_{t=R}^{T-h}\frac{\log f_{1,t}(Y_{t+h}) - \log f_{2,t}(Y_{t+h})}{\sqrt{\widehat{\Sigma}}} \tag{6}$$

where $\widehat{\Sigma}$ is an estimate of the long-run variance. Under the null hypothesis, the two density forecasts $f_{1,t}(.)$ and $f_{2,t}(.)$ perform equally well.

3.2.1 CPI Inflation

Despite under-predicting inflation at most forecast horizons, the BVAR density forecasts outperformed COMPASS'. The concentration of PITs for the BVAR density inflation forecasts at the highest percentiles for most horizons (shown by the bigger dots) indicates that the BVAR underestimated inflation between 2000 and 2012 (Figure 5). The PITs for COMPASS' density inflation forecasts in the same period are concentrated at both the higher and lower percentiles, suggesting that COMPASS underestimated the uncertainty around the central inflation forecast - i.e., its fan charts were too narrow (Figure 5). Overall, logarithmic scores suggest that the BVAR produced more accurate density forecasts for inflation than COMPASS at all horizons between 2000 and 2012 (Figure 6).

3.2.2 GDP growth

Both models over-predicted GDP growth at all forecast horizons, but the BVAR outperformed COMPASS at forecast horizons up to 1 year ahead. PITs for the GDP growth density forecasts of both COMPASS and the BVAR are concentrated at the lowest percentiles across all forecast horizons (large blue dots on figure 7), indicating that both models over-predicted GDP growth between 2000 and 2012. For COMPASS, this is particularly acute at shorter horizons. There is no significant difference in forecasting performance between the two models for forecast horizons of 1 year or longer. But the BVAR GDP forecasts outperformed COMPASS at horizons up to 1 year ahead (Figure 6).

3.3 BVAR performance vis-à-vis other forecasts

This section compares the BVAR's forecasting performance with two other benchmarks: an in-house suite of statistical forecasting models and the Monetary Policy

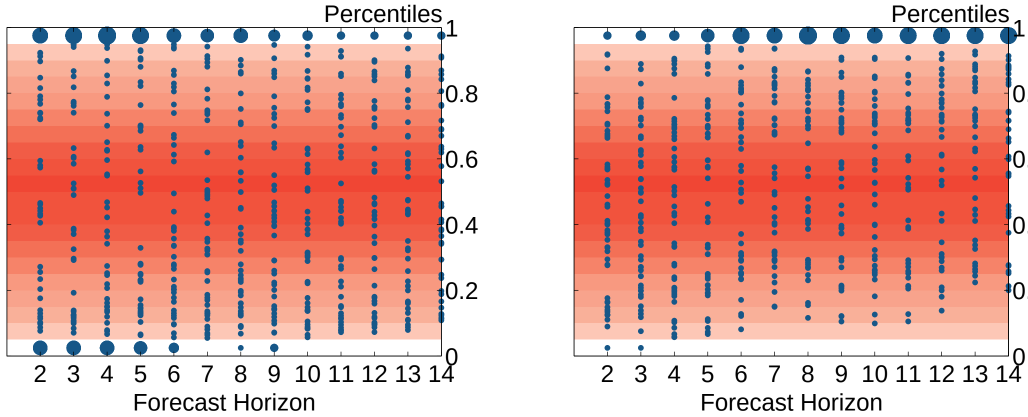

Figure 5: Probability Integral Transforms (PITs) for annual CPI inflation density forecasts, 2000-2012. COMPASS (lhs) versus BVAR (rhs). The percentiles on the y-axis are highlighted with shades of red: the darkest shade is assigned to the 50^{th} percentile and the shades become lighter for less probable areas of the model-implied density. The dark blue dots indicate the actual outturns, while a bigger dots indicating a clustering of the dots in that portion of the model's density.

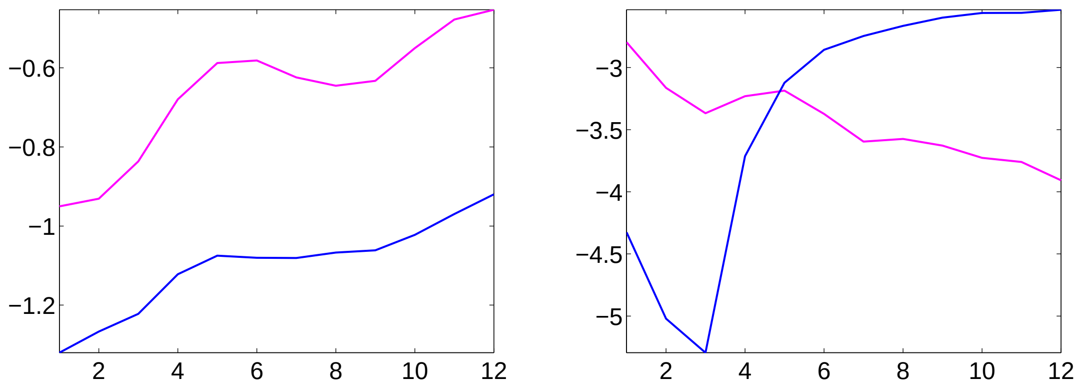

Figure 6: Logarithmic scores for annual CPI inflation (lhs) and annual real GDP growth (rhs) at different forecast horizons, 2000-2012. COMPASS (blue) versus BVAR (pink)

Committe's projections for GDP growth and CPI inflation reported in the *Inflation Report*, as in Fawcett *et al.* (2015). The statistical suite includes a range of univariate and multivariate forecasting models (see Kapetanios *et al.* (2008) for details). The individual forecasts produced by these models are then translated into one forecast using weights based on their predictive likelihoods. The *Inflation Report* projections are the MPC's best collective projections for GDP growth and inflation. They therefore reflect a combination of model-based forecasts and off-model information and judgement. The BVAR and the statistical suite have a small information disadvantage relative to the *Inflation Report* due to issues with the timing of data releases around the publication of the *Inflation Report*.

The BVAR point forecast performance was comparable to the statistical suite for both GDP and CPI inflation. Both models registered similar RMSFEs for 1-year-ahead inflation on average between 2000 and 2012. The BVAR outperformed the statistical

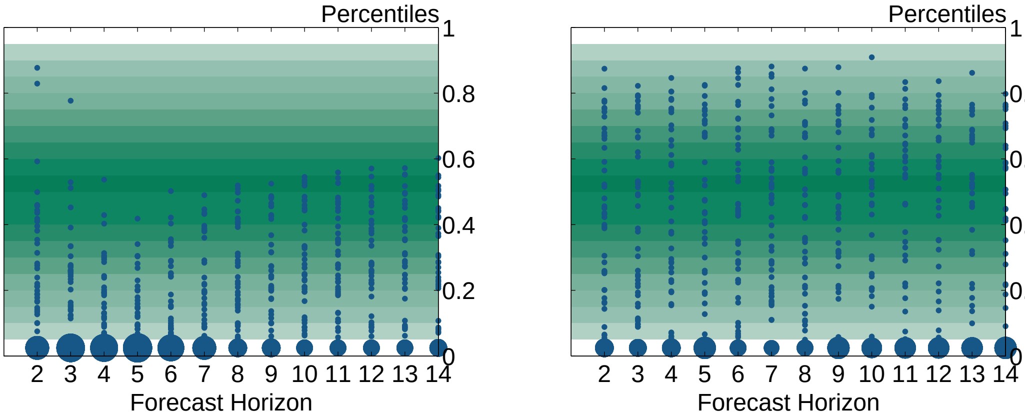

Figure 7: Probability Integral Transforms (PITs) for annual real GDP growth density forecasts, 2000-2012. COMPASS (lhs) versus BVAR (rhs). The percentiles on the y-axis are highlighted with shades of green: the darkest shade is assigned to the 50^{th} percentile and the shades become lighter for less probable areas of the model-implied density. The dark blue dots indicate the actual outturns, while a bigger dots indicating a clustering of the dots in that portion of the model's density.

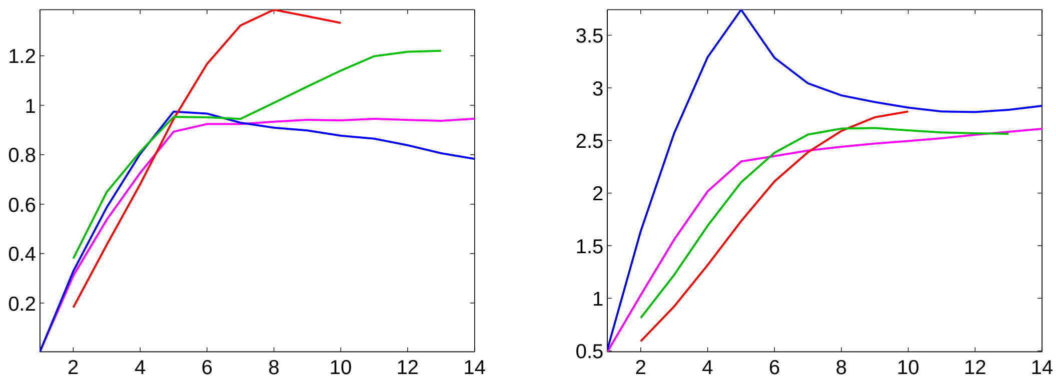

Figure 8: Root mean squared forecast errors for annual CPI inflation (lhs) and annual real GDP growth (rhs) at different forecast horizons, 2000-2012. COMPASS (blue), BVAR (pink), Statistical suite (green) and *Inflation Report* (red)

suite by around 0.2pp at the 2-year forecast horizon, but the difference is not statistically significant (see Appendix B). The Inflation Report was relatively the least accurate around the two-year horizon, but the difference was also not significant (Figure 8, left panel). When forecasting GDP growth both 1 and 2 years ahead, the BVAR, the suite and the Inflation Report performed similarly (Figure 8, right panel). The *Inflation Report* projections had the lowest RMSE at horizons up to two years, but the difference was not statistically significant.

The BVAR density forecasts perform similarly to the statistical suite for CPI, but marginally worse for GDP growth. The statistical suite has the highest log score for the CPI forecast at both 1 and 2 years-ahead horizons, closely followed by the BVAR, but their forecasting performance is not statistically different (Figure 9 and Appendix C). Both perform similarly when forecasting GDP densities 2 years ahead, but the

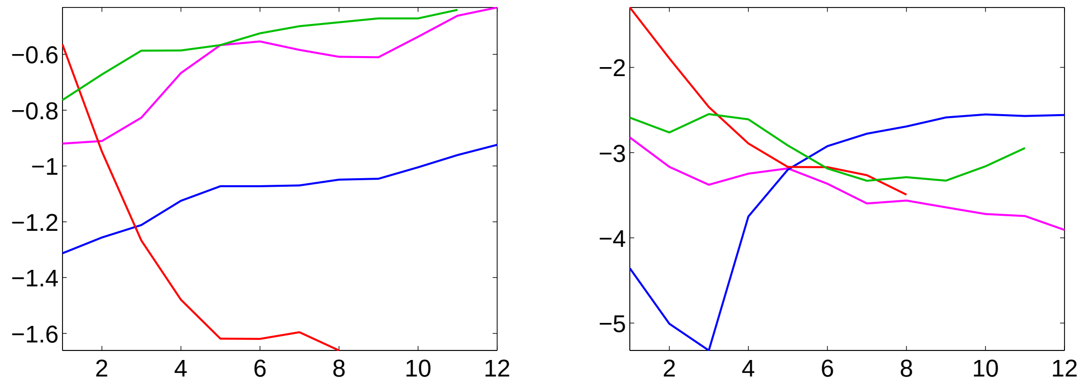

Figure 9: Logarithmic scores for annual CPI inflation (lhs) and annual real GDP growth (rhs) at different forecast horizons, 2000-2012. COMPASS (blue), BVAR (pink), Statistical suite (green) and *Inflation Report* (red)

BVAR is less accurate at the 1 year horizon (Figure 9). The log scores of the *Inflation Report* density forecasts are worse, though not statistically different from the BVAR's, for either inflation or GDP growth.

4 Similar studies for other advanced economies

Other studies have also compared forecasts from DSGE and BVAR models. Iversen *et al.* (2014) investigate the case of the Sveriges Riksbank and explicitly contrast DSGE and BVAR real-time forecasts since 2007, although the mapping between both models (in terms of variables) is not as exact as in our study. They find that the BVAR model forecasts for inflation performed well both in absolute terms and relative to the DSGE model forecasts and the Riksbank's published forecasts. Whereas these results also suggest that BVAR forecasts tend to outperform those of a DSGE model, they differ from our results in that, for Sweden, the BVAR's forecasting performance is superior for inflation, but not GDP.

A second study, by Christoffel *et al.* (2011), examines the forecasting performance of NAWM, the ECB's DSGE model, against Bayesian VAR benchmarks. Their exercise differs from ours in a number of ways. First, the forecast evaluation period starts in 1999 and ends 2006, therefore missing the most recent financial crisis. Second, they do not have a preferred BVAR, but rather assess NAWM against four BVARs which vary in size and type of prior. Finally, the models are re-estimated annually, therefore less frequently than in our study. Despite these differences, they also find that the DSGE model is outperformed by a BVAR benchmark, both in terms of point and density forecasts.

5 Conclusions

This paper focusses on the relative forecasting performance of COMPASS and of its BVAR analogue. Our results show that the BVAR generally outperformed COMPASS when forecasting both GDP and its expenditure components, while the performance of both models was similar when forecasting CPI. We also find that, despite under-predicting inflation at most forecast horizons, the BVAR density forecasts outperformed COMPASS', and that although both models over- predicted GDP growth at all forecast horizons, the BVAR outperformed COMPASS at forecast horizons up to 1 year ahead. These results are broadly consistent with the findings of similar studies for other advanced economies. The BVAR's point forecast performance was also comparable to that of a Bank of England in-house statistical suite for both GDP and CPI inflation and to the *Inflation Report* projections.

Besides probing our baseline assumption of running the forecast evaluation exercises conditioning on the exogenous variables in the BVAR, we think that the most interesting avenue for further research in terms of forecasting performance evaluation is to abandon the Minnesota and sum-of-coefficients priors to explore other sources of priors for the estimation of the BVAR. For example, we're currently experimenting with using COMPASS itself as a source of priors for the BVAR, and how that affects forecasting performance, as in DeJong and Whiteman (1994).

Forecasting performance is not the only relevant dimension along which to compare models. For example, one appealing feature of DSGE models is that they are fully structural, and can thus serve as coherent story-telling devices, which is one reason for their popularity in policy institutions. In order to be able to compare impulse response functions and historical decompositions with those from COMPASS, we are therefore also working on the identification of the BVAR, using a combination of sign and zero restrictions, as in Haberis and Sokol (2014).

References

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

[2] Banbura, M., D. Giannone and L. Reichlin, 2010. "Large Bayesian vector auto regressions," Journal of Applied Econometrics, John Wiley and Sons, vol. 25(1), pages 71-92.

[3] Burgess, S., E. Fernandez-Corugedo, C. Groth, R. Harrison, F. Monti, K. Theodoridis and M. Waldron, 2013. "The Bank of England's Forecasting Platform: COMPASS, MAPS, EASE and the Suite of Models," Bank of England working paper no.471

[4] Carriero, A., T. Clark and M. Marcellino, 2011. "Bayesian VARs: Specification Choices and Forecast Accuracy," CEPR Discussion Papers 8273

[5] Christoffel, K., G. Coenen and A. Warne, 2011. "Forecasting with DSGE Models," in Michael P. Clements and David F. Hendry (eds), Handbook of Economic Forecasting, Oxford University Press, Oxford, Chapter 4, pages 89-127.

[6] DeJong, D. N., and C. H. Whiteman, 1994. "The Forecasting Attributes of Trend- and Difference Stationary Representations for Macroeconomic Time Series," Journal of Forecasting, 13, 279-97.

[7] Del Negro, M., Hasegawa, R. B., and F. Schorfheide, 2014, "Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance," Staff Report No. 695, Federal Reserve Bank of New York.

[8] Del Negro, M., and F. Schorfheide, 2012. "DSGE Model-based Forecasting," SSRN working paper.

[9] Doan, T., R. Litterman, and C. A. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distribution," Econometric Review, vol. 3, pages 1-100.

[10] Diebold, F. X. and R. S. Mariano, 1995. "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, vol. 13(3), pages 253-63.

[11] Fawcett, N., L. Körber, R. M. 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 working paper 538.

[12] Faust, J., and J. H. Wright, 2008. "Efficient forecast tests for conditional policy forecasts.," Journal of Econometrics 146, 293-303.

[13] Giannone, D., M. Lenza and G.E. Primiceri, 2015. "Prior Selection for Vector Autoregressions," The Review of Economics and Statistics, MIT Press, vol. 2(97), pages 436-451.

[14] Gurkaynak, R. S., B. Kisacikoglu and B. Rossi, ”Do DSGE Models Forecast More Accurately Out-of-Sample than VAR Models?” , CEPR Discussion Paper No. DP9576.

[15] Haberis, Alex, and Andrej Sokol, 2014. ”A procedure for combining zero and sign restrictions in a VAR-identification scheme,” Discussion Papers 1410, Centre for Macroeconomics (CFM).

[16] Iversen, J., S. Laseen, H. Lundvall and U. Soderstrom, 2014. “Real-Time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank,” mimeo.

[17] Kapetanios, G., V. Labhard, and S. Price, ”Forecast combination and the Bank of England’s suite of statistical forecasting models,” Economic Modelling, 25, 772-792.

[18] Litterman, R., 1980. “A Bayesian Prodecure for Forecasting with Vector Autoregression,” Massachusetts Institute of Technology, working paper.

[19] Rotemberg, J. J. 1982. Sticky prices in the united states. Journal of Political Economy 90, 11871211.

[20] Sims, C. A. and T. Zha, 1998. “Bayesian Methods for Dynamic Multivariate Models,” International Economic Review, vol. 39(4), pages 949-68.

[21] Smets, F., and R. Wouters, 2007, ”Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” American Economic Review, 97(3): 586-606.

Appendix A: The forecasts for world variables in COMPASS and the BVAR

Because the UK is a small open economy, it makes sense to estimate a VAR model where the world variables are treated as exogenous, such that they are not affected by the UK business cycle. As such, we have a choice of what forecasts to use for these exogenous variables:

1. COMPASS forecasts: One option would be to use the parameters from the exogenous block in COMPASS to produce forecasts for these variables for the BVAR. This would align both information sets, but would have the disadvantage of being more far removed from the quarterly forecast process.

2. Exogenous forecasts: Another option would be to impose the internal staff forecasts for the exogenous world variables as conditioning paths for both COMPASS and the BVAR. This would also yield a 'clean' comparison between both models and would be closest to our regular forecasting process. However, this would require assessing the density forecast performance of COMPASS with conditioning paths.

As a compromise, we compare the BVAR with the world variables forecast from the International Directorate imposed as conditioning paths against COMPASS without any conditioning paths. Figure 10 shows that whereas using the International Directorate's forecasts does give the BVAR an information advantage, the improvement is not large enough to drive the overall point forecast results. And this set up puts us in a better place for when we are able to assess density forecasts with conditioning paths.

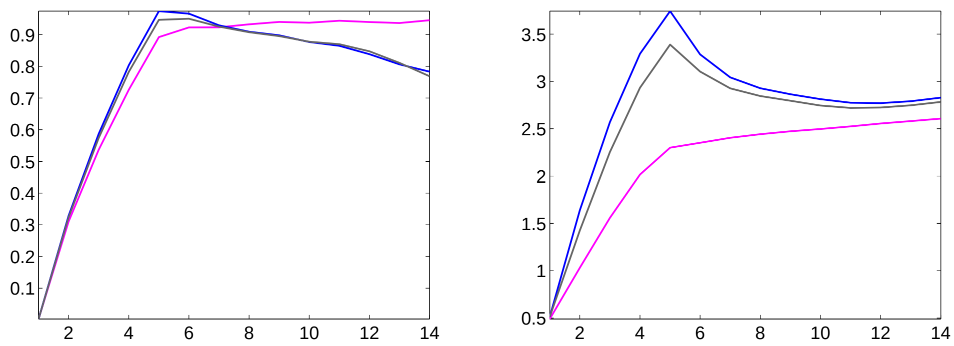

Figure 10: Annual CPI inflation (lhs) and quarterly GDP growth (rhs) root mean squared forecast errors at different forecast horizons, 2000-2012. BVAR (pink), COMPASS (blue), COMPASS with world conditioning path (black).

Appendix B: Diebold-Mariano test statistics

Table 1: BVAR versus COMPASS: Diebold-Mariano Test Statistics for average RMSFEs, 2000-2012(a)

	1 year forecast horizon	2 year forecast horizon
Annual CPI Inflation	-0.60	+0.37
Annual real GDP growth	-2.04**	-1.42
Annual real private consumption growth	-1.93*	-1.43
Annual real business investment growth	-1.60	-1.75*

(a) A negative (positive) number means the BVAR (COMPASS) is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).

Table 2: Annual CPI Inflation - BVAR versus other forecasts: Diebold-Mariano Test Statistics for average RMSFEs, 2000-2012 (a)

	1 year forecast horizon	2 year forecast horizon
Inflation Report	-0.99	-0.95
Statistical Suite	-0.32	-1.29
COMPASS	-0.60	+0.37

(a) A negative number means the BVAR is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).

Table 3: Annual real GDP growth, BVAR versus other forecasts: Diebold-Mariano Test Statistics for average RMSFEs, 2000-2012 (a)

	1 year forecast horizon	2 year forecast horizon
Inflation Report	-0.87	-1.57
Statistical Suite	-0.23	-0.57
COMPASS	-2.04**	-1.42

(a) A negative number means the BVAR is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).

Appendix C: Amisano-Giacomini test statistics

Table 4: BVAR versus COMPASS: Amisano-Giacomini Test Statistics for average logscores, 2000-2012(a)

	1 year forecast horizon	2 year forecast horizon
Annual CPI Inflation	+2.34**	+1.79*
Annual real GDP growth	+0.63	-0.82

(a) A positive (negative) number means the BVAR (COMPASS) is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).

Table 5: Annual CPI Inflation - BVAR versus other forecasts: Amisano-Giacomini Test Statistics for average logscores, 2000-2012 (a)

	1 year forecast horizon	2 year forecast horizon
Inflation Report	+1.27	+1.31
Statistical Suite	-0.83	-1.63
COMPASS	+2.34**	+1.79*

(a) A positive number means the BVAR is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).

Table 6: Annual real GDP growth, BVAR versus other forecasts: Amisano-Giacomini Test Statistics for average logscores, 2000-2012 (a)

	1 year forecast horizon	2 year forecast horizon
Inflation Report	-0.62	-0.05
Statistical Suite	-1.75*	-0.19
COMPASS	+0.63	-0.82

(a) A positive number means the BVAR is more accurate.
** (*) denotes statistical significance at 5 percent (10 percent).