Multivariate Forecasting with BVARs and DSGE Models

Berg, Tim Oliver

Ifo Institute

24 February 2015

Multivariate Forecasting with BVARs and DSGE Models

Tim Oliver Berg*

February 24, 2015

Abstract

In this paper I assess the ability of Bayesian vector autoregressions (BVARs) and dynamic stochastic general equilibrium (DSGE) models of different size to forecast comovements of major macroeconomic series in the euro area. Both approaches are compared to unrestricted VARs in terms of multivariate point and density forecast accuracy measures as well as event probabilities. The evidence suggests that BVARs and DSGE models produce accurate multivariate forecasts even for larger datasets. I also detect that BVARs are well calibrated for most events, while DSGE models are poorly calibrated for some. In sum, I conclude that both are useful tools to achieve parameter dimension reduction.

Keywords: BVARs, DSGE Models, Multivariate Forecasting, Large Dataset, Simulation Methods, Euro Area

JEL-Codes: C11, C52, C53, C55, E37

*Ifo Institute (berg@ifo.de).

1 Introduction

Vector autoregressions (VARs) are useful and thus frequently used tools to forecast macroeconomic time series, in particular when the interest is not only in the outcome of a single variable but the comovement of several major series. However, unrestricted VARs that are estimated by ordinary least squares (OLS) often suffer from their dense parametrization, leading to unstable parameter estimates and hence inaccurate multivariate forecasts. The literature proposes alternative ways to achieve parameter dimension reduction and improve multivariate forecast accuracy. For instance Bańbura, Giannone, and Reichlin (2010) suggest that VARs combined with Bayesian shrinkage (BVARs) can handle large datasets and produce relatively accurate forecasts. The empirical restrictions they impose on the VAR parameters are motivated by the integrating order of the underlying time series, shrinking the overparametrized VAR either towards a parsimonious random walk (for non-stationary data) or a white noise process (for stationary data), while the degree of shrinkage could be selected as in Carriero, Clark, and Marcellino (2015) by maximizing the marginal likelihood or estimated in an hierarchical fashion as suggested by Giannone, Lenza, and Primiceri (2014).

An alternative way to reduce the parameters to be estimated is to set up a dynamic stochastic general equilibrium (DSGE) model which builds on explicit micro foundations and optimizing economic agents. Under certain conditions[1] the state-space representation of a linearized DSGE model can be rewritten as a VAR and estimating such models hence amounts to shrinking a high dimensional vector of VAR parameters towards a lower dimensional vector of structural parameters that can be motivated by economic theory. While DSGE models are particularly useful for structural analyses and policy simulations, they are also widely used for forecasting (see, e.g., Edge and Gürkaynak, 2010; Christoffel, Coenen, and Warne, 2011; Del Negro and Schorfheide, 2013; Smets, Warne, and Wouters, 2014; Wolters, 2015, among others).

So far, the literature on BVARs and DSGE models has predominantly focused on point and density forecasts for individual series but neglected that in particular policymakers are often more interested in the comovement of major variables such as GDP growth and inflation.[2] The contribution of this paper is to extend the scope of the evaluation of BVARs and DSGE models to multivariate forecasts. In particular, I assess the ability of BVARs and DSGE models of different size to forecast comovements of major macroeconomic series in the euro area. Both approaches

[1]If these conditions are not met, the state-space representation has a VARMA form, where the moving average term is not invertible, and a VAR may be a poor approximation (see Franchi and Vidotto, 2013).

[2]Exceptions are Herbst and Schorfheide (2012) for the U.S. and Adolfson, Lindé, and Villani (2007) as well as Christoffel et al. (2011) for the euro area.

are compared to unrestricted VARs in terms of multivariate forecast accuracy measures. The forecast experiment allows me to explore to what extend multivariate forecast accuracy is affected by the way restrictions are imposed on model parameters (empirical, theoretical, or no restrictions) and the number of series included into estimation (from 3 up to 38).

One would expect that BVARs and DSGE models outperform unrestricted VARs in terms of multivariate forecast accuracy, in particular when the size of the systems is relatively large. Yet it is not clear whether DSGE models benefit from an explicit theoretical foundation of their restrictions compared to the purely empirical restrictions of the BVARs. On the one hand, the large number of cross-equation restrictions that emerge from micro foundations could help to produce a more realistic covariance structure between variables and hence improve multivariate forecast accuracy. On the other hand, if these restrictions are too rigid, forecasts could be biased in some direction, inflating forecast errors. For instance, DSGE models impose a common trend on real variables, an assumption that may not be supported in certain samples.

I compare the multivariate forecast accuracy of BVARs to that of three DSGE models, which include progressively larger sets of variables and hence increase in complexity. First, the small-scale closed economy model of An and Schorfheide (2007) that is estimated on three variables: GDP, the GDP deflator, and a short-term interest rate. Second, the medium-scale closed economy model of Smets and Wouters (2007) including in addition consumption, investment, employment, and wages (seven variables). And third, the medium-scale open economy model of Adolfson, Laséen, Lindé, and Villani (2007) which is fit also on exports, imports, the consumption deflator, the investment deflator, an exchange rate, world GDP, the world GDP deflator, and a world interest rate (fifteen variables). These models are frequently used at central banks and policy institutions to conduct simulations and produce forecasts for major macroeconomic series and are thus a natural competitor for BVARs.

Moreover, I add large and extra-large BVARs with, respectively, 23 and 38 variables, including series that could be helpful in forecasting major macroeconomic variables, but are often not included as observables in DSGE models, such as commodity prices, survey data, a long-term interest rate, or share prices. Several authors found that such BVARs are useful for forecasting (see, e.g., Bańbura et al., 2010; Koop, 2013; Giannone et al., 2014; Berg and Henzel, 2014; Pirschel and Wolters, 2014; Carriero et al., 2015, among others).

The forecast models are compared in terms of multivariate point and density forecast accuracy measures as well as event probabilities for different selections of euro area variables during the evaluation period from 1999:1 to 2011:4. Excluding the Great Recession (and the euro crisis) and shortening the evaluation period to 1999:1 to 2007:4 does not affect the model ranking.

Point forecasts are evaluated according to the trace and log determinant statistic of the scaled mean squared error (MSE) matrix. It turns out that the multivariate point forecast accuracy of the unrestricted OLS-VARs declines with the number of series included into estimation, while the accuracy of the BVARs does not. In fact, the large and extra-large BVAR deliver the smallest forecast errors of all models for most variable selections. Therefore I conclude that BVARs can deal with the dimensionality problem and exploit the information in high dimensional datasets. The results are less clear cut for the DSGE models. While the small and medium-scale closed economy models show a mixed performance, the open economy model of Adolfson et al. (2007) dominates the OLS-VARs and performs similar to the BVAR of same size. Thus I conclude that the underperformance of this model compared to the large and extra-large BVAR is not related to its structure but the fact that the largest BVARs include series that are particularly helpful in forecasting macroeconomic variables but are non-modeled in the open economy DSGE model.

Density forecasts are evaluated in terms of log predictive scores, which also reflect the uncertainty that is associated with forecasting macroeconomic outcomes. As policymakers nowadays closely monitor forecast uncertainty, the MSE might not be the relevant loss function. While the satisfactory performance of the large and extra-large BVAR is confirmed when the focus shifts from point to density forecasts, I also obtain that the predictive scores for all three DSGE models are higher than their relative multivariate point forecasts accuracy suggests. I suspect that the theoretical restrictions embedded in these models are helpful in producing a plausible covariance structure between the forecasts. In fact, the open economy DSGE model of Adolfson et al. (2007) outperforms all competitors in terms of predictive scores for most variable selections.

Finally, I provide additional evidence on the ability of BVARs and DSGE models to forecast comovements of major macroeconomic variables in the euro area by conducting an event study. Following Herbst and Schorfheide (2012) I compare model-implied event probabilities to actual frequencies of events. If model probabilities and average frequencies coincide, density forecasts are called well calibrated. In particular, I consider events that are highly relevant for policymakers, as for instance GDP growth, inflation, and the short-term interest rate being above (below) their respective long-run targets. The event study allows me to assess whether the models are able to forecast the directional comovements of these variables, which provides some insights that may not be reflected in MSE or predictive scores. It turns out that the large and extra-large BVAR perform well in terms of multivariate event forecast accuracy for most events. The DSGE models, on the other hand, are poorly calibrated for some events. One possible reason for this deficiency is that the common trend assumption of the DSGE models is too restrictive.

2 Dataset

The dataset contains 38 quarterly euro area macroeconomic series for the period 1984:1 to 2011:4 and covers seven categories labeled, respectively, national accounts, price index, international, employment, survey, monetary aggregate, and financial data. In most cases the series are from the 12th update of the Area-wide Model (AWM) database, which is maintained by the European Central Bank (ECB) and made available by the Euro Area Business Cycle Network.[3] The AWM database is the preferred source for researchers and policymakers alike interested in the euro area since historical series are backdated using individual country information in a consistent manner. Moreover, I add survey data and monetary aggregates from the Main Economic Indicators (MEI) of the Organisation for Economic Co-operation and Development (OECD) and a share price index from the Thomson Reuters Corporation (TRC). A detailed description of the dataset is provided in Table A.1.

I have to stress at this point that consistent vintage data is not available for the entire sample period because the euro area did not exist before January 1999. While the absence of real-time data is a potential drawback in studies where the objective is to mimic a realistic out-of-sample forecast experiment, I believe it is not in this paper. My forecast experiment can be understood as a model validation exercise designed to investigate the ability of BVARs and DSGE models of different size to forecast comovements in major macroeconomic series. Evaluating pseudo out-of-sample forecasts is an appropriate and established procedure to do so since forecasts reflect all sources of error associated with the modeling of economic outcomes, including parameter uncertainty and model misspecification.

In the forecast experiment I use five partitions of this dataset, which include progressively larger sets of variables. The composition of each set is determined by the structure of the corresponding DSGE model:

Small Selection (3 variables) In the smallest specification I work with real GDP, the GDP deflator, and a short-term interest rate. These series are considered in small-scale closed economy DSGE models and always included (see, e.g., An and Schorfheide, 2007, among others).

Medium Selection (7 variables) I add real private consumption, real gross investment, total employment, and real wages. Together with real GDP, the GDP deflator, and the interest rate, these variables are typically included as observables in medium-scale closed economy DSGE models such as the one by Smets and Wouters (2003).

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

Medium-Large Selection (15 variables) The set of variables is extended to an open economy setting by incorporating real exports, real imports, the private consumption deflator, the gross investment deflator, a real effective exchange rate, real world GDP, the world GDP deflator, and a world interest rate.[4] The medium-scale open economy DSGE model of Adolfson et al. (2007) is estimated on these series.

Large Selection (23 variables) I add several series that could be helpful in forecasting major macroeconomic variables, but are often not included as observables in DSGE models: the harmonized index of consumer prices (HICP), the oil price, the unemployment rate, a composite leading indicator, a consumer confidence indicator, the M1 money stock, a long-term interest rate, and a share price index. A similar selection for the U.S. is used in Giannone et al. (2014).

Extra-Large Selection (38 variables) Finally, I consider the full dataset, which includes, among others, a number of additional price indexes and national accounts data. Berg and Henzel (2014) apply their various BVAR specifications to a similar set of variables.

Except for those that are already expressed in rates (e.g. the short-term interest rate) or assumed to be stationary in levels (e.g. the real effective exchange rate), I transform all series into quarterly growth rates, approximated by the first difference of their logarithm. This proceeding is consistent with the underlying assumption of the DSGE models that most variables contain a stochastic trend in the level and need to be differenced to render them stationary. In Figure A.1 I provide time series plots for all variables that are included up to the medium-large selection.

3 Forecasting with BVARs

In this section I develop the BVAR models and discuss all major specification choices. Moreover, I explain how the models are estimated and forecasts obtained.

3.1 BVARs

Consider the following VAR

$$y_t = c + B_1 y_{t-1} + ... + B_p y_{t-p} + u_t, \tag{1}$$

where y_t is a $n \times 1$ vector of endogenous variables; c is a $n \times 1$ vector of intercepts; B_i are $n \times n$ matrices of coefficients; $i = 1, ..., p$ denotes the lags included; u_t is a $n \times 1$ vector of normally

[4]Unfortunately, the AWM database does not have a world interest rate. As Adolfson et al. (2007) I use the federal funds rate from Federal Reserve Economic Data (FRED) as a proxy.

distributed residual terms with zero mean and covariance matrix Σ; and data are available for $t = 1 - p, ..., T$. Let us denote $y = (y_1, ..., y_T)'$, $x_t = \left(y'_{t-1}, ..., y'_{t-p}, 1\right)'$, $x = (x_1, ..., x_T)'$, $B = (B_1, ..., B_p, c)'$, and $u = (u_1, ..., u_T)'$. The VAR in (1) can thus be written as $y = xB + u$. Moreover, let $\beta = \text{vec}(B)$ with $\text{vec}(\cdot)$ being the column stacking operator and $k = n(1 + np)$. Then β is a $k \times 1$ vector containing all coefficients of the model.

In the forecast experiment the VAR is estimated on up to $n = 38$ variables including $p = 2$ lags of each (hence $k = 2,926$).[5] Such a large dimensional system of multivariate regressions is, however, not estimable without imposing additional prior beliefs on the parameters. In addition, there is evidence that even VARs with only few variables might benefit from imposing prior information (see, e.g., Robertson and Tallman, 1999, among others). Hence I follow common practice and use a variant of the Minnesota prior to deal with the dense parameterization of the model. The basic idea is that a white noise process is a reasonable description of the data generating process behind most macroeconomic series once transformed to stationarity. In addition, the prior captures the belief that own lags are more informative than those of other variables and that more recent lags contain more information than more distant ones. The VAR is hence centered around the prior mean $y_{i,t} = c_i + u_{i,t}$ and imposing the white noise prior amounts to shrinking all elements of B_i towards zero.

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

In particular, I consider a conjugate Normal-Inverse-Wishart prior of the following form:

$$\Sigma \sim \text{IW}(\Psi, d) \quad \text{and} \quad \beta|\Sigma \sim \text{N}(b, \Sigma \otimes \Omega), \tag{2}$$

where $\otimes$ denotes the Kronecker product and the elements Ψ, d, b, and Ω are functions of hyperparameters. The conjugate prior implies a likelihood and posterior that come from the same family of distributions and hence makes Bayesian inference feasible even for large n.

Bańbura et al. (2010) recommend to implement the prior by constructing the following set of artificial observations:

[5] The qualitative results are similar for $p = 1$ and $p = 4$.

$$y^{+} = \begin{bmatrix} 0_{np\times n} \\ \text{diag}\left(\sigma_1, ..., \sigma_n\right) \\ 0_{1\times n} \end{bmatrix}, \qquad x^{+} = \begin{bmatrix} \text{diag}\left(1, 2, .., p\right) \otimes \text{diag}\left(\sigma_1, ..., \sigma_n\right)/\lambda & 0_{np\times 1} \\ 0_{n\times np} & 0_{n\times 1} \\ 0_{1\times np} & \epsilon \end{bmatrix}, \tag{3}$$

where diag $(\cdot)$ denotes a diagonal matrix. The σ_i's account for the different scale and variability of the series and are set equal to the standard deviation of a residual from a univariate autoregression for the variable $y_{i,t}$ using the same lag order as in the VAR. The parameter ϵ is set to a small number (10^{-4}), reflecting a diffuse prior for the intercepts. Finally, the parameter λ governs the degree of shrinkage and hence the tightness of the prior. As $\lambda \to \infty$ the prior becomes uninformative and posterior expectations coincide with the ordinary least squares (OLS) estimates. For $\lambda \to 0$ the posterior approaches the dogmatic prior. λ is hence the key parameter in the BVAR and its selection is discussed in detail below.

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

$$y^{*} = x^{*}B + u^{*}, \tag{4}$$

where $y^{*} = \left(y', y^{+'}\right)'$, $x^{*} = \left(x', x^{+'}\right)'$, $u^{*} = \left(u', u^{+'}\right)'$, and $u^{+'} = \left(u_1^{+}, ..., u_T^{+}\right)$. The latter contains the corresponding residual terms to y^{+}.

The conditional posterior distributions of the covariance matrix and coefficients can be computed in closed form as a function of the shrinkage parameter λ:

$$\Sigma|\lambda, y \;\sim\; \text{IW}\left(\hat{\Sigma}, T+n+2\right) \quad \text{and} \quad \beta|\Sigma, \lambda, y \sim \text{N}\left(\hat{\beta}, \Sigma \otimes \left(x^{*'}x^{*}\right)^{-1}\right), \tag{5}$$

where $\hat{\Sigma}$ and $\hat{\beta}$ are the covariance matrix and the coefficients from an OLS regression of y^{*} on x^{*}, respectively.

3.2 Choice for Shrinkage Parameter and MCMC Estimation

Following Giannone et al. (2014) I treat λ as an additional unknown parameter that is estimated in an hierarchical fashion. The approach hence accounts for the uncertainty related to this specification choice and requires to add one more layer to the prior structure by placing a prior on the shrinkage parameter - a hyperprior. It can be shown that the marginal posterior for λ, i.e. after integrating out the posterior uncertainty about the model's parameters, is:

$$p(\lambda|y) \propto p(y|\lambda) \cdot p(\lambda), \tag{6}$$

where $\propto$ denotes proportionality. $p(y|\lambda)$ is the marginal likelihood of the model (conditional on λ) and $p(\lambda)$ is the hyperprior. The former has a well-known analytic form which is provided in Appendix B. The latter needs to be chosen and reflects how confident we are about the values for λ. I follow common practice and choose a diffuse hyperprior. In particular, I consider a Gamma density with mode equal to 0.2 and standard deviation of 0.4.

Since the joint posterior distribution for the parameters and λ is not available in closed form, Giannone et al. (2014) recommend a Metropolis-Hastings algorithm to simulate the distribution. The sampler is a Markov chain Monte Carlo (MCMC) method and generates λ from its marginal posterior with a Metropolis update. After convergence of the sampler, the covariance matrix and coefficients can be drawn from their posterior distribution conditional on λ and forecasts computed in an iterative fashion. In detail, the algorithm works as follows.

Algorithm 1: Metropolis-Hastings sampler to obtain forecasts from BVARs

Let N_B be the number of burn-in-draws and N_R the number of retained draws. The maximal forecast horizon is denoted by H. The steps are:

Step 1: Choose a starting point λ^0. I use the posterior mode, which is obtained by numerical optimization.[6] For $j = 1, ..., N_B + N_R$ run a loop over the following steps.

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

Step 3: Compute the acceptance ratio:

$$r = \frac{p(\lambda^*|y)}{p(\lambda^{j-1}|y)}, \tag{7}$$

where $p(\lambda|y)$ is given by Equation (6).

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

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

[6] I use the Matlab routine *fmincon*.

$$\lambda^j = \begin{cases} \lambda^* & : \text{ if } \quad \nu \leq r \\ \lambda^{j-1} & : \text{ otherwise} \end{cases} \tag{8}$$

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

Step 6: Draw $\Sigma^j | \lambda^j, y$ and $\beta^j | \Sigma^j, \lambda^j, y$ from their conditional posterior in (5).

Step 7: Generate $u^j_{T+1}, ..., u^j_{T+H}$ from $u_t \sim \text{N}\left(0, \Sigma^j\right)$.

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

$$\hat{y}^j_{T+1} = c^j + B^j_1 y_T + ... + B^j_p y_{T-p+1} + u^j_{T+1}, \tag{9}$$

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

$$\hat{y}^j_{T+h} = c^j + B^j_1 \hat{y}^j_{T+h-1} + ... + B^j_p \hat{y}^j_{T+h-p} + u^j_{T+h}, \tag{10}$$

where $h = 1, ..., H$ and $\hat{y}^j_{T+h} = y_{T+h-p}$ for $h \leq p$. $\square$

This algorithm yields $\left(\hat{y}^j_{T+1}, ..., \hat{y}^j_{T+H}\right)^{N_B+N_R}_{j=N_B+1}$ as a sample of forecasts generated from the joint posterior distribution for the parameters and λ, where $N_B = 5,000$ and $N_R = 50,000$.

In order to economize storage space and to reduce the correlation among draws, I thin the chain and keep only one for every 10 draws. The remaining 5,000 draws are used for inference. In Appendix C I provide diagnostics, suggesting that N_B is long enough to let the Markov chain for λ converge to its ergodic random distribution.

4 Forecasting with DSGE Models

In this section I present the main features of the DSGE models that compete with the BVARs. In addition, I describe how the models are estimated and forecasts obtained.

4.1 DSGE Models

The forecast accuracy of the BVARs is compared to the following DSGE models: the small-scale closed economy model of An and Schorfheide (2007) that is estimated on the small selection of the dataset; the medium-scale closed economy model of Smets and Wouters (2007) including the medium selection of the dataset; and the medium-scale open economy model of Adolfson et al. (2007) which is fit to the medium-large selection of the dataset. The models are frequently used at central banks and policy institutions to conduct simulations and produce forecasts for major macroeconomic series, and are therefore a natural competitor for the BVARs. The main features of the three models are as follows.

An and Schorfheide (2007) The first model is a variant of the new Keynesian baseline model, which is developed in Woodford (2003) and Galí (2008). There are numerous applications of the baseline model that differ with respect to the exogenous shocks or the specification of the monetary policy rule. In this paper I use the variant of An and Schorfheide (2007) since they allow for a stochastic growth rate driven by total factor productivity (TFP) growth, so that GDP does not need to be detrended before estimation. Furthermore, Herbst and Schorfheide (2012) show that their model produces reasonable forecasts for U.S. time series. The model consists of three equations: an intertemporal Euler equation, a new Keynesian Phillips curve, and an interest rate rule. The Euler equation is derived from the optimal plans of a representative household, while the Phillips curve results from the optimal decision making of monopolistically competitive firms that face nominal price rigidities. And according to the interest rate rule, monetary policy responds to deviations of inflation and output growth from their respective targets. Finally, the model features exogenous shocks to TFP growth, government spending, and the interest rate.

Smets and Wouters (2007) The second model is a more elaborate version of the baseline model, which features several additional nominal and real frictions as well as exogenous shocks that are needed to produce a reasonable behavior of consumption, investment, employment, and wages. In particular, the model by Smets and Wouters (2007) allows for nominal wage rigidities, consumption habit formation, capital adjustment costs as well as price and wage indexation to past inflation. In contrast to the original version in Smets and Wouters (2003), it also introduces labor-augmenting technology growth, implying that all real variables grow at the same rate in steady-state and again no detrending is needed. Herbst and Schorfheide (2012) found that this model performs well on U.S. data in terms of forecast accuracy.

Adolfson, Laseén, Lindé, and Villani (2007) The third model extends the closed economy setting of Smets and Wouters (2007) by incorporating open economy aspects into it. Adolfson et al. (2007) argue that the structure of their model is rich enough to explain the joint dynamics of a large number of domestic and foreign variables, while Adolfson et al. (2007) also show its ability to produce reasonable forecasts. In particular, the model features the accumulation of net foreign assets, incomplete exchange rate pass-through due to nominal rigidities in the importing and exporting sectors, and a foreign economy. The latter is modeled as an exogenous VAR. In addition, the model includes several domestic and foreign exogenous shocks as well as measurement error in some observables.

Appendix D shows the measurement equations and prior specifications for the three DSGE models. For the complete derivation of the models I refer to the original articles.

4.2 Model Solution and MCMC Estimation

Before forecasts can be obtained from the DSGE models, they have to be solved with numerical methods and estimated with Bayesian techniques (see, e.g., Del Negro and Schorfheide, 2011, 2013; Schorfheide, 2013, among others). In particular, the equilibrium conditions are cast into a system of linear rational expectations difference equations:

$$\Gamma_0 (\theta) \xi_{t-1} + \Gamma_1 (\theta) \xi_t + \Gamma_2 (\theta) \xi_{t+1} = \Pi (\theta) \epsilon_t, \tag{11}$$

where θ is a vector that contains the model parameters; ξ_t is a $r \times 1$ vector of endogenous model variables; Γ_i with $i = 0, 1, 2$ are $r \times r$ coefficient matrices that are functions of θ; ϵ_t is a $q \times 1$ vector of exogenous shocks; Π is a $r \times q$ coefficient matrix that also depends on θ; and data are available for $t = 1, ..., T$. Provided that a unique and non-explosive solution of the system exists for a particular θ, the model variables can be rewritten as a VAR (see, e.g., Klein, 2000; Sims, 2002, among others):

$$\xi_t = \Phi_1 (\theta) \xi_{t-1} + \Phi_\epsilon (\theta) \epsilon_t, \tag{12}$$

where the coefficient matrices Φ_1 ($r \times r$) and Φ_ϵ ($r \times q$) again depend on θ. Let y_t denote a $n \times 1$ vector of observable variables that is linked to the model variables ξ_t as follows

$$y_t = \Psi_0 (\theta) + \Psi_1 (\theta) \xi_t + \eta_t. \tag{13}$$

Ψ_0 is a $n \times 1$ vector containing the steady-state values of y_t conditional on θ and provided that ξ_t is stationary. Ψ_1 is a $n \times r$ coefficient matrix and η_t is a $n \times 1$ vector of normally distributed measurement errors with zero mean and covariance matrix R, which are assumed to be independent of the exogenous shocks ϵ_t.

The transition equations in (12) together with the measurement equations in (13) are known as the state-space representation of the DSGE models. It should be stressed that the dimension of θ is much lower than the number of elements in the system matrices Φ_1, Φ_ϵ, Ψ_0, and Ψ_1. Parsimony in DSGE models is thus achieved by imposing restrictions that come from economic theory, while those in the BVAR were motivated by the integrating order of the data used.

Under the assumption that the exogenous shocks are Gaussian, i.e. $\epsilon_t \sim \mathrm{N} (0, I_q)$, the likelihood function $p (y|\theta)$ for the observable data $y = (y_1, ..., y_T)$ given a value for θ can be evaluated using the standard Kalman filter (see, e.g., Giordani, Pitt, and Kohn, 2011, among others). The posterior distribution for the parameters $p(\theta|y)$ is then obtained by combining the likelihood function and the prior distribution $p(\theta)$:

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta) . \tag{14}$$

Since the relation between θ and the matrices Φ_1, Φ_ϵ, Ψ_0, and Ψ_1 is non-linear, no analytic expression for the posterior distribution is available and a Metropolis-Hasting algorithm is needed to obtain posterior draws. The sampler is similar to that for the BVAR and works as follows.

Algorithm 2: Metropolis-Hastings sampler to obtain forecasts from DSGE models

Let N_B be the number of burn-in-draws and N_R the number of retained draws. The maximal forecast horizon is denoted by H. The steps are:

Step 1: Choose a starting point θ^0. I use the posterior mode, which is obtained by numerical optimization.[7] For $j = 1, ..., N_B + N_R$ run a loop over the following steps.

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

Step 3: Compute the acceptance ratio:

$$r = \frac{p(\theta^*|y)}{p(\theta^{j-1}|y)}, \tag{15}$$

where $p(\theta|y)$ is given by Equation (14).

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

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

$$\theta^j = \begin{cases} \theta^* & : \text{ if } \nu \leq r \\ \theta^{j-1} & : \text{ otherwise} \end{cases} \tag{16}$$

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

Step 6: Draw $\xi^j|\theta^j, y$ from $\mathrm{N}\left(\hat{\xi}^j, P^j\right)$, where $\hat{\xi}^j$ and P^j are computed with the Kalman filter.

[7] I use either the Matlab routine *fmincon* or Chris Sims' *csminwel* depending on the DSGE model.

Step 7: Generate $\epsilon_{T+1}^{j}, ..., \epsilon_{T+H}^{j}$ from $\epsilon_t \sim \mathrm{N}(0, I_q)$ and simulate a path for the model variables using the transition equations in (12):

$$\xi_t^j = \Phi_1\left(\theta^j\right)\xi_{t-1}^j + \Phi_\epsilon\left(\theta^j\right)\epsilon_t^j, \quad t = T+1, ..., T+H \tag{17}$$

Step 8: Generate $\eta_{T+1}^{j}, ..., \eta_{T+H}^{j}$ from $\eta_t \sim \mathrm{N}(0, R)$ and simulate a path for the observable variables using the measurement equations in (13):

$$\hat{y}_t^j = \Psi_0\left(\theta^j\right) + \Psi_1\left(\theta^j\right)\xi_t^j + \eta_t^j, \quad t = T+1, ..., T+H \quad \square \tag{18}$$

This algorithm yields $\left(\hat{y}_{T+1}^{j}, ..., \hat{y}_{T+H}^{j}\right)_{j=N_B+1}^{N_B+N_R}$ as a sample of forecasts generated from the posterior predictive distribution of a DSGE model, where $N_B = 50,000$ and $N_R = 50,000$.

In order to economize storage space and to reduce the correlation among draws, I thin the chain and keep only one for every 10 draws. The remaining 5,000 draws are used for inference. For all computations I use the software package *Dynare*. I inspected the standard convergence diagnostics provided therein to ensure that posterior distributions are well behaved and draws come from the ergodic random distribution.

5 Forecast Evaluation

In this section I explain the forecast experiment and compare the competing models using multivariate forecast accuracy measures.

5.1 Forecast Experiment

In the forecast experiment I work with 11 models. First, I start with an unrestricted VAR that is estimated by OLS.[8] The OLS-VAR is the natural benchmark in a multivariate context and considered up to the medium-large selection of the dataset (OLS-VAR Small, OLS-VAR Medium, OLS-VAR Medium-Large). No forecasts are produced for the large and extra-large data selection since OLS-VARs cannot handle such large cross-sections. Second, I use the BVAR in combination with all five partitions of the dataset (BVAR Small, BVAR Medium, BVAR Medium-Large, BVAR Large, BVAR Extra-Large). And finally, I consider the three DSGE models, again including variables up to the medium-large selection (DSGE Small, DSGE Medium, DSGE Medium-

[8] In fact, I set $\lambda = 10,000$, $N_B = 0$, $N_R = 5,000$ and execute Steps (6) to (8) in Algorithm 1 repeatedly to obtain a forecast distribution for the OLS-VAR.

Large). The forecast experiment hence allows me to assess to what extend multivariate forecast accuracy is affected by the way restrictions are imposed on the model parameters (empirical, theoretical, or no restrictions) and the number of variables included into estimation (from small to extra-large data selection).

For each model I produce forecasts up to $H = 4$. I begin with the estimation sample 1984:1 to 1998:4 and generate forecasts for 1999:1 to 1999:4. I then iterate forward, always updating the estimation sample one quarter at a time, until 2010:4, producing forecasts for 2011:1 to 2011:4.[9] This proceeding yields a sequence of 49 forecasts for each model. The evaluation period hence runs from 1999:1 to 2011:1 for the shortest forecast horizon ($h = 1$) and from 1999:4 to 2011:4 for the longest ($h = 4$). The start of the evaluation period is chosen to coincide with the introduction of the euro. With respect to the end I checked that the model ranking is not too sensitive to the inclusion of the Great Recession (and the euro crisis). In Appendix E I provide point and density forecast accuracy measures for a shortened evaluation period from 1999:h to 2007:h.

5.2 Point Forecasts

In order to investigate the ability of the competing models to produce relatively accurate point forecasts for a set of m variables, I follow common practice and construct a scaled h-step-ahead MSE matrix ($m \times m$) (see, e.g., Adolfson et al., 2007; Christoffel et al., 2011, among others):

$$\Omega_{M,h} = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0}^{T_1} \tilde{e}_{T+h|T} \tilde{e}'_{T+h|T}, \quad h = 1, ..., H, \tag{19}$$

where $T_0 = 60$ and $T_1 = 108$ denote the last observation of the first (1998:4) and final (2010:4) estimation sample, respectively. The scaled h-step-ahead forecast error vector ($m \times 1$) is calculated as $\tilde{e}_{T+h|T} = M^{-1/2} e_{T+h|T}$, with M being a $m \times m$ positive definite scaling matrix and $e_{T+h|T} = \hat{y}_{T+h|T} - y_{T+h}$. The $m \times 1$ vector of point forecasts $\hat{y}_{T+h|T}$ is provided by the mean of the multivariate forecast distribution that is computed from the 5,000 sampled forecasts, while y_{T+h} contains the actual values ($m \times 1$).

Based on this scaled MSE matrix I compute two standard measures of multivariate point forecast accuracy: the trace statistic tr $[\Omega_{M,h}]$ and the log determinant statistic log$|\Omega_{M,h}|$. Moreover, I set M to a diagonal matrix containing the unconditional sample variances of the m variables during the evaluation period from 1999:1 to 2011:4. To understand the differences between

[9] In Berg and Henzel (2014) it is shown that BVARs benefit a lot in terms of predictive scores from recursive estimation schemes that use all available observations up to period T compared to rolling window schemes that discard the most distant observations. Therefore I prefer the former to the latter.

both measures, it is important to note that the log determinant statistic is invariant to this choice of the scaling matrix, while the trace statistic is not (see, e.g., Adolfson et al., 2007, among others).[10] But with M being diagonal, the trace statistic is simply an unweighted average of scaled univariate MSE and hence easy to interpret.

In Table 1 I report the trace and log determinant (in parenthesis) of the scaled MSE matrix for the competing forecast models up to four quarters. The lower the trace or log determinant statistic is for a particular model, the better is its relative multivariate point forecast accuracy with respect to a variable selection. In particular, I discriminate between models based on three different sets of variables: the small selection ($m = 3$), the medium selection ($m = 7$), and the medium-large selection ($m = 15$). Increasing the number of series to be evaluated ensures that results are not too heavily influenced by a single variable. Of course, the number of models that can be taken into account is decreasing in m.

The following picture emerges from the table. First, the multivariate point forecast accuracy of the OLS-VARs clearly deteriorates with the number of variables included into estimation. This conclusion holds true for all forecast horizons and both the small and medium selection of the variables to be evaluated. And with only one exception ($h = 2$ and $m = 3$) both the trace and log determinant statistic also deliver the same ranking of OLS-VARs. The dimensionality problem that arises when unrestricted OLS-VARs become larger apparently dominates the potential benefits of a richer dataset.

Second, imposing restrictions on the parameters in form of a Minnesota-type prior uniformly improves multivariate point forecast accuracy compared to the unrestricted OLS-VARs. The BVARs outperform the OLS-VARs of same size in terms of both accuracy measures at all forecast horizons and for all variable selections. Even for small models it hence seems advisable to shrink the parameter space by adopting an informative prior. Across BVARs I find that multivariate point forecast accuracy does not deteriorate with the size of the dataset. In contrast, the large and extra-large BVARs deliver in most cases much smaller forecast errors than the small, medium, or medium-large specifications. Therefore I conclude that the BVAR can deal with the dimensionality problem and exploit the information contained in high dimensional datasets. However, I also find that with only a few exceptions the large BVAR dominates the extra-large BVAR in terms of multivariate point forecast accuracy, suggesting that the 23 variables may suffice to produce reasonable forecasts and that the additional gains of enlarging the dataset are at best small, if not negative. In a related study Berg and Henzel (2014) reach a similar conclusion when evaluating forecasts for euro area HICP inflation and GDP growth.

[10]Note that tr $[\Omega_{M,h}] =$ tr $\left[M^{-1}\Omega_{I,h}\right]$ and log$|\Omega_{M,h}| =$ log$|\Omega_{I,h}| -$log$|M|$, where $\Omega_{I,h}$ is based on $M = I$.

Table 1: Trace and Log Determinant of Scaled MSE Matrix

Forecast Model	Forecast Horizon h (in Quarters) 1	2	3	4
	Small Selection ($m = 3$)			
OLS-VAR Small	1.93 (−2.90)	2.45 (−1.57)	3.09 (−0.71)	4.28 (0.19)
OLS-VAR Medium	2.08 (−2.85)	2.61 (−1.69)	3.33 (−0.68)	4.58 (0.22)
OLS-VAR Medium-Large	2.61 (−1.92)	3.39 (−0.40)	4.72 (0.53)	6.43 (1.22)
BVAR Small	1.87 (−3.05)	2.32 (−1.74)	2.87 (−0.84)	3.81 (−0.01)
BVAR Medium	1.95 (−2.96)	2.32 (−1.80)	2.93 (−0.84)	3.74 (−0.08)
BVAR Medium-Large	1.78 (−3.04)	2.45 (−1.50)	3.30 (−0.51)	4.15 (0.23)
BVAR Large	1.47 (−3.34)	1.96 (−1.83)	2.44 (−1.10)	3.09 (−0.57)
BVAR Extra-Large	1.72 (−2.98)	2.05 (−1.76)	2.50 (−1.18)	3.15 (−0.60)
DSGE Small	2.74 (−2.12)	3.36 (−0.59)	4.26 (0.45)	5.45 (1.25)
DSGE Medium	1.98 (−2.57)	1.93 (−1.53)	2.22 (−1.08)	2.45 (−0.75)
DSGE Medium-Large	1.94 (−2.96)	2.46 (−1.53)	3.46 (−0.49)	4.77 (0.31)
	Medium Selection ($m = 7$)			
OLS-VAR Medium	6.35 (−6.25)	7.62 (−5.39)	9.48 (−3.87)	11.46 (−3.00)
OLS-VAR Medium-Large	9.11 (−3.83)	9.86 (−2.86)	11.19 (−1.66)	12.84 (−1.33)
BVAR Medium	5.71 (−7.02)	6.66 (−5.84)	8.16 (−4.35)	9.47 (−3.64)
BVAR Medium-Large	5.33 (−6.94)	6.67 (−5.52)	8.05 (−4.14)	9.34 (−3.31)
BVAR Large	4.56 (−7.15)	5.82 (−5.44)	6.77 (−4.30)	7.79 (−3.71)
BVAR Extra-Large	5.17 (−6.26)	5.89 (−5.16)	6.40 (−4.49)	7.28 (−3.71)
DSGE Medium	7.26 (−3.33)	7.65 (−2.96)	8.48 (−2.92)	8.93 (−2.59)
DSGE Medium-Large	5.04 (−7.12)	6.62 (−5.40)	8.53 (−3.90)	10.50 (−3.10)
	Medium-Large Selection ($m = 15$)			
OLS-VAR Medium-Large	17.83 (−14.20)	21.55 (−12.76)	25.56 (−10.76)	28.90 (−10.73)
BVAR Medium-Large	11.73 (−19.31)	15.63 (−17.40)	18.96 (−14.87)	21.64 (−13.72)
BVAR Large	10.33 (−18.52)	13.04 (−16.41)	15.54 (−14.42)	17.78 (−12.91)
BVAR Extra-Large	11.08 (−17.07)	13.34 (−15.54)	15.72 (−13.94)	17.83 (−12.37)
DSGE Medium-Large	11.50 (−19.94)	15.58 (−16.56)	19.35 (−14.04)	23.00 (−12.25)

Notes: this table shows the trace and log determinant (parenthesis) of the scaled MSE matrix for competing forecast models. The evaluation period runs from 1999:h to 2011:h, depending on the forecast horizon h. The MSE matrix is scaled by the unconditional sample variance of the variables during the evaluation period. The lower the trace (or log determinant) is, the better is the multivariate point forecast accuracy of the model with respect to the variable selection. The forecasted variables are: GDP, GDP deflator, interest rate (Small Selection); consumption, investment, employment, wages (Medium Selection); exports, imports, consumption deflator, investment deflator, exchange rate, world GDP, world GDP deflator, federal funds rate (Medium-Large Selection).

Third, the results are less clear cut for the DSGE models. For instance I find that the small DSGE model is outperformed in terms of multivariate point forecast accuracy by both the small BVAR and small OLS-VAR at all horizons. This finding is robust across both accuracy measures and suggests that the structure of the small DSGE model might be too simplistic to generate a plausible comovement of the three series. In contrast, the medium DSGE model does exceptionally well for the small variable selection at all horizons and clearly dominates the medium OLS-VAR. And except for the first horizon the model also outperforms the medium BVAR. However, the forecast accuracy of the medium DSGE model is mixed with respect to the medium selection of the variables. While it is competitive at longer forecast horizons, forecasts are less accurate in the short-run compared to both the medium OLS-VAR and medium BVAR. A possible explanation for the discrepancy of the forecast accuracy of the medium DSGE model between the small and medium variable selection may be the observed difficulty of such models to generate reasonable joint dynamics of consumption and wages (see, e.g., Christoffel et al., 2011; Smets et al., 2014, among others).

The medium-large DSGE model I consider in the forecast competition seems to be less prone to such difficulties though. The model dominates the medium-large OLS-VAR in terms of multivariate point forecast accuracy for all variable selections at all horizons and for both accuracy measures, suggesting that the theoretical restrictions it imposes on the parameters indeed help to produce plausible comovements between the variables and reduce estimation uncertainty. In comparison to the medium-large BVAR, the model performs only slightly worse at all forecast horizons for the small and medium-large variable selection. With respect to the medium selection it is even better than the BVAR in the short-run, but a bit less accurate at longer horizons.

However, the medium-large DSGE model cannot compete with the large and extra-large BVAR. For all variable selections both BVARs substantially outperform the medium-large DSGE model at most horizons and for both accuracy measures. Hence I conclude that the additional information contained in the large and extra-large dataset is particularly helpful in forecasting macroeconomic variables, which is not surprising since those cover several forward-looking series, such as financial or survey indicators. The underperformance of the medium-large DSGE model is hence not related to its structure, but the fact that these series are non-modeled.

5.3 Density Forecasts

In the previous section I compared the competing models with respect to their relative multivariate point forecast accuracy based on the scaled MSE matrix. In the following I extend the competition to the entire multivariate forecast distribution and compute for each variant the

h-step-ahead log predictive score, which is the height of the distribution $p_T(\cdot)$ evaluated at the actual outcome y_{T+h} and summed over all evaluation periods:

$$S_h = \sum_{T=T_0}^{T_1} \log p_T\left(y_{T+h}\right), \quad h = 1, ..., H. \tag{20}$$

Unfortunately, there is no analytic expression for the predictive likelihood $p_T(y_{T+h})$ available and it hence needs to be estimated from the simulated sample of forecasts. Since I evaluate forecasts for up to 15 variables, multivariate kernel density estimation is not a feasible option. Instead I follow Adolfson et al. (2007), assuming that $p_T(\cdot)$ is multivariate normal for all T and estimate the mean forecast vector as well as its covariance matrix for all T and h from the 5,000 sampled forecasts.[11]

The log predictive score for the competing forecast models up to four quarters and the different variable selections is shown in Table 2. The higher the score of a model is, the better is its relative multivariate density forecast accuracy.

In sum, the table suggests that the conclusions of the previous section are largely confirmed when the competition is extended to multivariate density forecasts, with some exceptions for the DSGE models. Once more I find that the forecast accuracy of the OLS-VARs is negatively related to the number of variables included into estimation. While the predictive scores for the small and medium OLS-VAR are rather similar with respect to the small variable selection, the medium-large OLS-VAR is clearly outperformed by both at all forecast horizons. And except for the longest horizon, the medium-large OLS-VAR also shows a much smaller score compared to the medium OLS-VAR for the medium selection, underlining that parameter proliferation can be a severe problem for unrestricted VARs when the size of the dataset is relatively large.

For the density forecasts I also obtain that imposing prior information on the parameters is beneficial with respect to multivariate forecast accuracy. Up to two quarters the BVARs clearly dominate the OLS-VARs of same size in terms of predictive scores for all variable selections. However, the relative advantage of the BVARs is declining with the forecast horizon. For $h = 3$ both approaches deliver similar predictive scores, while for the longest horizon the OLS-VARs show even larger scores for all variable selections, suggesting that relatively accurate multivariate point forecasts do not necessarily imply higher predictive likelihoods. In contrast, the large BVAR again delivers an excellent forecast performance at all horizons for the small variable selection, which confirms that the additional series contained in this dataset are informative for

[11]Warne, Coenen, and Christoffel (2014) propose a Kalman filter method as an alternative to the normal approximation to obtain the predictive likelihood, but find small differences between both approaches.

Table 2: Log Predictive Score

Forecast Model	Forecast Horizon h (in Quarters)			
	1	2	3	4
	Small Selection ($m = 3$)			
OLS-VAR Small	-216.1	-222.5	-222.9	-228.6
OLS-VAR Medium	-220.8	-222.1	-221.9	-228.6
OLS-VAR Medium-Large	-254.6	-251.1	-232.9	-229.8
BVAR Small	-209.8	-217.2	-222.1	-229.8
BVAR Medium	-208.9	-216.7	-225.6	-233.9
BVAR Medium-Large	-207.1	-225.4	-230.0	-235.8
BVAR Large	-202.3	-216.4	-219.6	-219.1
BVAR Extra-Large	-212.1	-224.6	-227.1	-229.6
DSGE Small	-205.4	-217.8	-227.0	-237.7
DSGE Medium	-209.0	-188.5	-185.3	-186.9
DSGE Medium-Large	-185.4	-186.7	-190.2	-197.3
	Medium Selection ($m = 7$)			
OLS-VAR Medium	-524.0	-488.3	-486.1	-494.7
OLS-VAR Medium-Large	-609.1	-532.6	-500.8	-484.5
BVAR Medium	-485.3	-480.0	-494.2	-506.2
BVAR Medium-Large	-484.0	-488.5	-492.6	-501.7
BVAR Large	-485.3	-489.8	-491.2	-486.3
BVAR Extra-Large	-515.3	-495.8	-491.6	-483.1
DSGE Medium	-484.0	-438.8	-432.7	-430.8
DSGE Medium-Large	-409.2	-406.6	-416.5	-425.3
	Medium-Large Selection ($m = 15$)			
OLS-VAR Medium-Large	-1595.4	-1329.3	-1235.8	-1181.7
BVAR Medium-Large	-1194.2	-1223.0	-1247.0	-1252.4
BVAR Large	-1246.8	-1264.5	-1264.5	-1250.5
BVAR Extra-Large	-1302.2	-1277.6	-1271.3	-1235.2
DSGE Medium-Large	-1129.6	-1153.0	-1169.5	-1185.4

Notes: this table shows the log predictive score for competing forecast models. The evaluation period runs from 1999:h to 2011:h, depending on the forecast horizon h. The higher the log predictive score is, the better is the multivariate density forecast accuracy of the model with respect to the variable selection. The forecasted variables are: GDP, GDP deflator, interest rate (Small Selection); consumption, investment, employment, wages (Medium Selection); exports, imports, consumption deflator, investment deflator, exchange rate, world GDP, world GDP deflator, federal funds rate (Medium-Large Selection).

the comovoment of the three variables. With respect to the medium and medium-large selection the performance of the large BVAR is less remarkable though, but still satisfactory. Finally, the table shows that the extra-large BVAR cannot quite confirm its ability to forecast comovements in major macroeconomic series when the focus shifts from point to density forecasts. Since the extra-large BVAR involves estimating a high dimensional covariance matrix (38×38), I believe that posterior uncertainty may be too high for this variant, leading to relatively low predictive likelihoods. In fact, Berg and Henzel (2014) obtain similar evidence on density forecasts for euro area HICP inflation and GDP growth coming from BVARs with 44 variables.

With respect to the DSGE models the most important finding is that their predictive scores are higher than their relative multivariate point forecast accuracy suggests. Therefore I conclude that the theoretical restrictions embedded in these models are helpful in producing a plausible covariance structure between the forecasts. The small DSGE model, which displayed relatively high forecast errors, is not worse than the small OLS-VAR and small BVAR in terms of predictive scores, while the multivariate density forecast accuracy of the medium DSGE model is even better. With only one exception ($h = 1$ and $m = 3$) it shows higher scores than all OLS-VARs and BVARs at all forecast horizons as well as for both variable selections. The medium DSGE model is competitive with the medium-large DSGE model for the small variable selection, but not for the medium selection. In fact, the medium-large DSGE outperforms all competitors at all forecast horizons with respect to the medium selection and except for the longest horizon also for the medium-large selection. The structure of this model is therefore rich enough to produce reasonable comovements between the 15 variables included into estimation.

5.4 Event Forecasts

Next I provide additional evidence on the ability of the competing models to forecast comovements of major macroeconomic series in the euro area by conducting an event study. In particular, I follow Herbst and Schorfheide (2012) and consider two measures of absolute as well as relative multivariate event forecast accuracy.

Let $[y_{i,T+h} \geq a, y_{j,T+h} \geq b]$ be a multivariate event and $p_T\left[\cdot\right]$ the model-implied probability that the event occurs. The absolute multivariate event forecast accuracy is evaluated based on

$$P_h = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0}^{T_1} \left(\ell \left[y_{i,T+h} \geq a, y_{j,T+h} \geq b \right] - p_T \left[y_{i,T+h} \geq a, y_{j,T+h} \geq b \right] \right), \tag{21}$$

where $\ell\left[\cdot\right]$ is an indicator function that takes on unity if the event occurs and zero otherwise. P_h hence measures the difference between the actual frequency of events and the average model-implied probability. If $P_h \approx 0$, the event forecasts are well calibrated. If $P_h < 0$ (> 0), the forecast model overpredicts (underpredicts) the actual frequency of the events.

Furthermore, I assess the relative multivariate event forecast accuracy and rank models according to the scoring rule:

$$\tilde{P}_h = \frac{1}{T_1 - T_0 + 1} \sum_{T=T_0}^{T_1} \left(\ell\left[y_{i,T+h} \geq a, y_{j,T+h} \geq b\right] - p_T\left[y_{i,T+h} \geq a, y_{j,T+h} \geq b\right]\right)^2. \tag{22}$$

While any multivariate event could be used in principle to compute P_h and $\tilde{P}_h$, I shall focus on the small variable selection and consider four events that are particularly relevant for policymakers: First, GDP growth, inflation, and the interest rate are above (below) their respective long-run targets (Event I and II). Second, GDP growth is below (above) and inflation is above (below) their long-run targets (Event III and IV). The long-run targets for GDP growth, inflation, and the interest rate are set to, respectively, 1.5%, 2%, and 3.5% (at annualized rates).[12] Specified in this way, the actual frequency of these events during the evaluation period for $h = 1$ is 10.2%, 16.3%, 44.9%, and 16.3%, respectively. All in all, the event study allows me to assess whether the models are able to forecast the directional comovements of the three variables, which provides some insights that may not be reflected in MSE or predictive scores.

Tables 3 and 4 show both measures of multivariate event forecast accuracy for all four events up to $h = 4$. The lower the mean squared difference is, the better is the relative forecast accuracy of a model with respect to the event, while the closer to zero the mean difference (in parenthesis) is, the better is its absolute forecast accuracy.

From Table 3 we learn that all forecast models overpredict the occurrence of Event I, which suggests that they do not adjust to the fact that the variables tend to decline over the evaluation period. Given that the actual frequency of Event I at $h = 1$ is only 10.2%, absolute deviations of 1.8 percentage points (DSGE Small) and 5.5 percentage points (OLS-VAR Medium) appear large. Such an incorrect calibration of density forecasts would not be reflected in MSE or predictive scores. Consistently, the models also tend to underpredict the likelihood of Event II. However, the absolute deviations from the actual frequency of Event II at $h = 1$ of 16.3% are smaller, varying between 1.0 percentage points (OLS-VAR Small) and 4.1 percentage points (DSGE Small).

Regarding the relative multivariate event forecast accuracy I obtain results that square with previous findings. First, enlarging the size of the OLS-VARs does not improve forecast accuracy.

[12]This choice for the long-run targets can be rationalized as follows. First, long-run real GDP growth is proxied by its sample mean ($\approx 1.5\%$). Second, long-run inflation equals the ECB's target (2%). And third, the long-run interest rate is proxied by long-run nominal GDP growth ($1.5\% + 2\% = 3.5\%$).

Table 3: Event Occurrence and Model-Implied Probability

Forecast Model	Forecast Horizon h (in Quarters) 1	2	3	4
	Event I: GDP Growth (↑) & Inflation (↑) & Interest Rate (↑)			
OLS-VAR Small	0.072 (−0.041)	0.078 (−0.043)	0.079 (−0.065)	0.107 (−0.082)
OLS-VAR Medium	0.073 (−0.055)	0.087 (−0.072)	0.089 (−0.086)	0.107 (−0.108)
OLS-VAR Medium-Large	0.081 (−0.034)	0.094 (−0.065)	0.102 (−0.067)	0.090 (−0.075)
BVAR Small	0.072 (−0.036)	0.079 (−0.037)	0.083 (−0.054)	0.105 (−0.069)
BVAR Medium	0.067 (−0.030)	0.082 (−0.034)	0.089 (−0.049)	0.103 (−0.061)
BVAR Medium-Large	0.073 (−0.020)	0.076 (−0.040)	0.097 (−0.047)	0.092 (−0.056)
BVAR Large	0.073 (−0.029)	0.070 (−0.062)	0.088 (−0.095)	0.078 (−0.111)
BVAR Extra-Large	0.052 (−0.024)	0.068 (−0.048)	0.077 (−0.082)	0.081 (−0.100)
DSGE Small	0.084 (−0.018)	0.088 (−0.009)	0.096 (−0.009)	0.102 (−0.010)
DSGE Medium	0.100 (−0.023)	0.097 (−0.051)	0.100 (−0.049)	0.099 (−0.046)
DSGE Medium-Large	0.068 (−0.048)	0.083 (−0.075)	0.094 (−0.102)	0.104 (−0.129)
	Event II: GDP Growth (↓) & Inflation (↓) & Interest Rate (↓)			
OLS-VAR Small	0.118 (−0.010)	0.145 (−0.010)	0.161 (0.016)	0.181 (0.039)
OLS-VAR Medium	0.118 (−0.016)	0.143 (0.001)	0.167 (0.021)	0.190 (0.043)
OLS-VAR Medium-Large	0.155 (0.021)	0.175 (0.025)	0.145 (0.011)	0.184 (0.004)
BVAR Small	0.118 (−0.013)	0.143 (−0.012)	0.157 (0.012)	0.179 (0.035)
BVAR Medium	0.121 (−0.032)	0.144 (−0.025)	0.158 (−0.003)	0.181 (0.018)
BVAR Medium-Large	0.123 (−0.013)	0.145 (−0.005)	0.145 (−0.004)	0.165 (0.006)
BVAR Large	0.110 (0.028)	0.134 (0.028)	0.120 (0.007)	0.167 (0.011)
BVAR Extra-Large	0.112 (0.039)	0.147 (0.048)	0.138 (0.039)	0.184 (0.046)
DSGE Small	0.130 (−0.041)	0.150 (0.014)	0.158 (0.057)	0.176 (0.092)
DSGE Medium	0.120 (−0.038)	0.141 (−0.032)	0.146 (−0.004)	0.156 (0.022)
DSGE Medium-Large	0.110 (0.013)	0.153 (0.065)	0.166 (0.109)	0.204 (0.146)

Notes: this table shows the mean squared difference as well as the mean difference (parenthesis) between the occurrence of a multivariate event and the model-implied probability for competing forecast models. The events are: GDP growth, inflation, and the interest rate are above ↑ (below ↓) their long-run targets, which are, respectively, 1.5%, 2%, and 3.5% (at annualized rates). The evaluation period runs from 1999:h to 2011:h, depending on the forecast horizon h. The lower the mean squared difference is, the better is the relative multivariate forecast accuracy of a model with respect to the event. The closer to zero the mean difference is, the better is the absolute multivariate forecast accuracy of a model with respect to the event.

Table 4: Event Occurrence and Model-Implied Probability (continued)

Forecast Model	Forecast Horizon h (in Quarters) 1	2	3	4
	Event III: GDP Growth (↑) & Inflation (↓)			
OLS-VAR Small	0.229 (0.017)	0.238 (−0.001)	0.219 (−0.008)	0.247 (−0.028)
OLS-VAR Medium	0.238 (0.040)	0.241 (0.026)	0.226 (0.018)	0.244 (0.010)
OLS-VAR Medium-Large	0.285 (0.050)	0.281 (0.027)	0.244 (0.060)	0.262 (0.065)
BVAR Small	0.226 (0.013)	0.234 (−0.007)	0.219 (−0.019)	0.245 (−0.039)
BVAR Medium	0.232 (0.016)	0.237 (0.002)	0.235 (−0.017)	0.251 (−0.037)
BVAR Medium-Large	0.231 (0.020)	0.222 (−0.005)	0.231 (−0.015)	0.236 (−0.028)
BVAR Large	0.216 (0.033)	0.231 (0.050)	0.225 (0.086)	0.233 (0.098)
BVAR Extra-Large	0.235 (0.046)	0.226 (0.064)	0.230 (0.082)	0.244 (0.087)
DSGE Small	0.246 (0.054)	0.233 (−0.033)	0.225 (−0.093)	0.241 (−0.139)
DSGE Medium	0.224 (0.184)	0.257 (0.162)	0.247 (0.129)	0.242 (0.108)
DSGE Medium-Large	0.267 (0.185)	0.263 (0.186)	0.258 (0.191)	0.261 (0.190)
	Event IV: GDP Growth (↓) & Inflation (↑)			
OLS-VAR Small	0.148 (0.073)	0.144 (0.084)	0.141 (0.080)	0.144 (0.086)
OLS-VAR Medium	0.152 (0.069)	0.146 (0.078)	0.140 (0.067)	0.144 (0.073)
OLS-VAR Medium-Large	0.150 (0.020)	0.145 (0.063)	0.125 (0.039)	0.145 (0.038)
BVAR Small	0.147 (0.073)	0.143 (0.083)	0.142 (0.081)	0.144 (0.086)
BVAR Medium	0.148 (0.072)	0.143 (0.079)	0.142 (0.080)	0.144 (0.084)
BVAR Medium-Large	0.133 (0.045)	0.131 (0.071)	0.133 (0.072)	0.139 (0.075)
BVAR Large	0.134 (0.011)	0.125 (0.019)	0.116 (0.017)	0.131 (0.011)
BVAR Extra-Large	0.151 (−0.014)	0.124 (−0.001)	0.115 (0.000)	0.133 (−0.003)
DSGE Small	0.121 (0.012)	0.130 (0.012)	0.123 (0.018)	0.124 (0.022)
DSGE Medium	0.113 (−0.098)	0.128 (−0.059)	0.125 (−0.060)	0.126 (−0.064)
DSGE Medium-Large	0.138 (−0.079)	0.141 (−0.126)	0.156 (−0.164)	0.169 (−0.191)

Notes: this table shows the mean squared difference as well as the mean difference (parenthesis) between the occurrence of a multivariate event and the model-implied probability for competing forecast models. The events are: GDP growth is above ↑ (below ↓) and inflation is below ↓ (above ↑) their long-run targets, which are, respectively, 1.5% as well as 2% (at annualized rates). The evaluation period runs from 1999:h to 2011:h, depending on the forecast horizon h. The lower the mean squared difference is, the better is the relative multivariate forecast accuracy of a model with respect to the event. The closer to zero the mean difference is, the better is the absolute multivariate forecast accuracy of a model with respect to the event.

Second, the BVARs meet the challenge of dimensionality and exhaust the information contained in larger datasets. In particular, the large and extra-large BVAR perform well in terms of relative multivariate event forecast accuracy. Third, the medium-large DSGE model shows a satisfactory performance for both events and is particularly accurate for the first quarter. However, the small DSGE model is less accurate and dominated by the small OLS-VAR and BVAR for both events at most forecast horizons. Finally, the medium DSGE model also shows a rather poor performance for Event I, but does better than the medium OLS-VAR and BVAR with respect to Event II.

Table 4 shows that most forecast models underpredict the occurrence of Event III and IV. Exceptions include the medium and medium-large DSGE model, which clearly overpredict the likelihood of Event IV at all forecast horizons. At a first glance all models seem to have difficulties in correctly forecasting a situation in which GDP growth is above and inflation is below their respective long-run targets. When inspecting the absolute deviations, however, the problem is less severe. Taking into account that Event III happens quite frequently (44.9% at $h = 1$), deviations between 1.3 percentage points (BVAR Small) and 5.4 percentage points (DSGE Small) are small. For the medium and medium-large DSGE model the divergence between average probabilities and actual frequencies is much larger though. Both models underpredict the occurrence of Event III at $h = 1$ by about 18 percentage points. Despite their relatively accurate multivariate point and density forecasts for the small variable selection, both DSGE models are poorly calibrated for Event III. One possible reason for this deficiency is that the common trend assumption of these models is too restrictive. This finding is also true for Event IV. The medium and medium-large DSGE model overpredict situations in which GDP growth is below and inflation is above their respective long-run targets at $h = 1$ by 9.8 and 7.9 percentage points, respectively. Given that the actual frequency of that event is only 16.3%, these discrepancies again point to a poor calibration of both DSGE models. For all other forecast models the absolute deviations vary from 1.1 percentage points (BVAR Large) to 7.3 percentage points (OLS-VAR Small). In particular, the large and extra-large BVAR are well calibrated for Event IV, even for longer horizons.

With respect to the relative multivariate event forecast accuracy I obtain results that confirm the main conclusions of this paper. For instance the OLS-VARs deliver forecast errors that either increase (Event III) or at least do not decrease (Event IV) with the number of variables included into estimation. Moreover, the table shows that BVARs often dominate OLS-VARs of same size, while the large BVAR performs well for both events at most horizons. And finally, the evidence is mixed for the DSGE models. While all models display a relatively poor performance for Event III, the small and medium DSGE model forecast Event IV accurately.

6 Conclusion

This paper assesses the ability of BVARs and DSGE models of different size to forecast comovements of major macroeconomic variables in the euro area. While the literature as of today has predominantly focused on point and density forecasts for individual series, I extend the scope of the evaluation of BVARs and DSGE models to multivariate forecasts. In particular, I compare both approaches to classical VARs in terms of multivariate point and density forecast accuracy measures as well as event probabilities. The evidence for the evaluation period from 1999:1 to 2011:4 suggests that BVARs and DSGE models produce relatively accurate point and density forecasts even for larger datasets. In fact, the large and extra-large BVAR as well as the medium-scale open economy DSGE model of Adolfson et al. (2007) show an excellent performance in terms of MSE and predictive scores for almost all variable selections. Moreover, I detect that the large and extra-large BVAR produce event probabilities that are commensurable with actual frequencies for most events. The medium-scale open economy DSGE model of Adolfson et al. (2007), on the other hand, is poorly calibrated for some events. One possible reason for this deficiency is that the common trend assumption of this model is too restrictive. Altogether, I conclude that BVARs and DSGE models are both useful tools to achieve parameter dimension reduction and forecast comovements of major macroeconomic variables. For future research it could be fruitful to address the two main deficiencies of DSGE models to improve their multivariate forecast accuracy. First, the common trend assumption could be relaxed. Second, the set of observables could be extended to forward-looking variables such as survey data. Recent work by Smets et al. (2014) goes in this direction.

References

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2007). Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through. *Journal of International Economics 72*(2), 481–511.

Adolfson, M., J. Lindé, and M. Villani (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometric Reviews 26*(2-4), 289–328.

An, S. and F. Schorfheide (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews 26*(2-4), 113–172.

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

Berg, T. O. and S. R. Henzel (2014). Point and Density Forecasts for the Euro Area Using Bayesian VARs. CESifo Working Paper No. 4711.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Christoffel, K., G. Coenen, and A. Warne (2011). Forecasting with DSGE Models. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*, pp. 89–127. New York: Oxford University Press.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics 146*(2), 318–328.

Del Negro, M. and F. Schorfheide (2011). Bayesian Macroeconometrics. In J. Geweke, G. Koop, and H. van Dijk (Eds.), *The Oxford Handbook of Bayesian Econometrics*, pp. 293–389. Oxford: Oxford University Press.

Del Negro, M. and F. Schorfheide (2013). DSGE Model-Based Forecasting. In G. Elliot and A. Timmerman (Eds.), *Handbook of Economic Forecasting*, Volume 24, pp. 57–140. North-Holland: Elsevier.

Edge, R. M. and R. S. Gürkaynak (2010). How Useful are Estimated DSGE Model Forecasts for Central Bankers? *Brookings Papers on Economic Activity 41*(2), 209–259.

Fagan, G., J. Henry, and R. Mestre (2005). An Area-Wide Model for the Euro Area. *Economic Modelling 22*(1), 39–59.

Franchi, M. and A. Vidotto (2013). A Check for Finite Order VAR Representations of DSGE Models. *Economics Letters 120*(1), 100–103.

Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton NJ: Princeton University Press.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, pp. 169–194. Oxford University Press.

Giannone, D., M. Lenza, and G. E. Primiceri (2014). Prior Selection for Vector Autoregressions. forthcoming in *Review of Economics and Statistics*.

Giordani, P., M. Pitt, and R. Kohn (2011). Bayesian Inference for Time Series State Space Models. In J. Geweke, G. Koop, and H. van Dijk (Eds.), *The Oxford Handbook of Bayesian Econometrics*, pp. 61–124. Oxford: Oxford University Press.

Herbst, E. and F. Schorfheide (2012). Evaluating DSGE Model Forecasts of Comovements. *Journal of Econometrics 171*(2), 152–166.

Kadiyala, K. R. and S. Karlsson (1993). Forecasting with Generalized Bayesian Vector Autoregressions. *Journal of Forecasting 12*(4), 365–378.

Kadiyala, K. R. and S. Karlsson (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics 12*(2), 99–132.

Klein, P. (2000). Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model. *Journal of Economic Dynamics and Control 24*(10), 1405–1423.

Koop, G. (2013). Forecasting with Medium and Large Bayesian VARS. *Journal of Applied Econometrics 28*(2), 177–203.

Litterman, R. B. (1980). A Bayesian Procedure for Forecasting with Vector Autoregressions. Manuscript, Massachusetts Institute of Technology, Department of Economics.

Litterman, R. B. (1986). Forecasting With Bayesian Vector Autoregressions - Five Years of Experience. *Journal of Business and Economic Statistics 4*(1), 25–38.

Pirschel, I. and M. H. Wolters (2014). Forecasting German Key Macroeconomic Variables Using Large Dataset Methods. Kiel Working Papers 1925, Kiel Institute for the World Economy.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies 72*(3), 821–852.

Robertson, J. C. and E. W. Tallman (1999). Vector Autoregressions: Forecasting and Reality. *Economic Review 84*(1), 4–18.

Schorfheide, F. (2013). Estimation and Evaluation of DSGE Models: Progress and Challenges. In D. Acemoglu, M. Arellano, and E. Deckel (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Volume 3, pp. 184–230. Cambridge: Cambridge University Press.

Sims, C. A. (2002). Solving Linear Rational Expectations Models. *Computational Economics 20*(1-2), 1–20.

Smets, F., A. Warne, and R. Wouters (2014). Professional Forecasters and Real-Time Forecasting with a DSGE Model. *International Journal of Forecasting 30*(4), 981–995.

Smets, F. and R. Wouters (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association 1*(5), 1123–1175.

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

Warne, A., G. Coenen, and K. Christoffel (2014). Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGEVAR, and VAR Models. CFS Working Paper Series 478, Center for Financial Studies (CFS).

Wolters, M. H. (2015). Evaluating Point and Density Forecasts of DSGE Models. *Journal of Applied Econometrics 30*(1), 74–96.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton NJ: Princeton University Press.

A Dataset

Table A.1: Dataset

No.	Mnemonic	Label	Category	Source	Code*
Small Selection (3 variables)					
01	YER	Real GDP	National Accounts	AWM	1
02	YED	GDP Deflator	Price Index	AWM	1
03	STN	Short-Term Interest Rate	Financial	AWM	3
Medium Selection (7 variables)					
04	PCR	Real Private Consumption	National Accounts	AWM	1
05	ITR	Real Gross Investment	National Accounts	AWM	1
06	LNN	Total Employment	Employment	AWM	1
07	WRN/YED	Real Wages	Employment	AWM	1
Medium-Large Selection (15 variables)					
08	XTR	Real Exports	National Accounts	AWM	1
09	MTR	Real Imports	National Accounts	AWM	1
10	PCD	Private Consumption Deflator	Price Index	AWM	1
11	ITD	Gross Investment Deflator	Price Index	AWM	1
12	EEN/(HICP/YWD)	Real Effective Exchange Rate	Financial	AWM	2
13	YWR	Real World GDP	International	AWM	1
14	YWD	World GDP Deflator	International	AWM	1
15	FEDFUNDS	Federal Funds Rate	International	FRED	3
Large Selection (23 variables)					
16	HICP	Overall HICP	Price Index	AWM	1
17	POILU	Oil Price	International	AWM	1
18	URX	Unemployment Rate	Employment	AWM	3
19	EKOL2002Q	Composite Leading Indicator	Survey	OECD	3
20	EKOCS002Q	Consumer Confidence Indicator	Survey	OECD	3
21	EKQMA027B	M1 Money Stock	Monetary Aggregate	OECD	1
22	LTN	Long-Term Interest Rate	Financial	AWM	3
23	EMSHRPRCF	Share Price Index	Financial	TRC	1
Extra-Large Selection (38 variables)					
24	GCR	Real Government Consumption	National Accounts	AWM	1
25	WIN	Compensation to Employees	National Accounts	AWM	1
26	GON	Gross Operating Surplus	National Accounts	AWM	1
27	TIN	Indirect Taxes	National Accounts	AWM	1
28	SAX	Household's Savings Ratio	National Accounts	AWM	3
29	GCD	Government Consumption Deflator	Price Index	AWM	1
30	XTD	Exports Deflator	Price Index	AWM	1
31	MTD	Imports Deflator	Price Index	AWM	1
32	YWRX	Real World Demand	International	AWM	1
33	YWDX	World Demand Deflator	International	AWM	1
34	COMPR	Commodity Prices	International	AWM	1
35	LFN	Labor Force	Employment	AWM	1
36	LEN	Employees	Employment	AWM	1
37	EKQMA013B	M3 Money Stock	Monetary Aggregate	OECD	1
38	EXR	Euro per U.S.D. Exchange Rate	Financial	AWM	1

* Transformation code: 1 = Log Difference; 2 = Log; 3 = Raw.

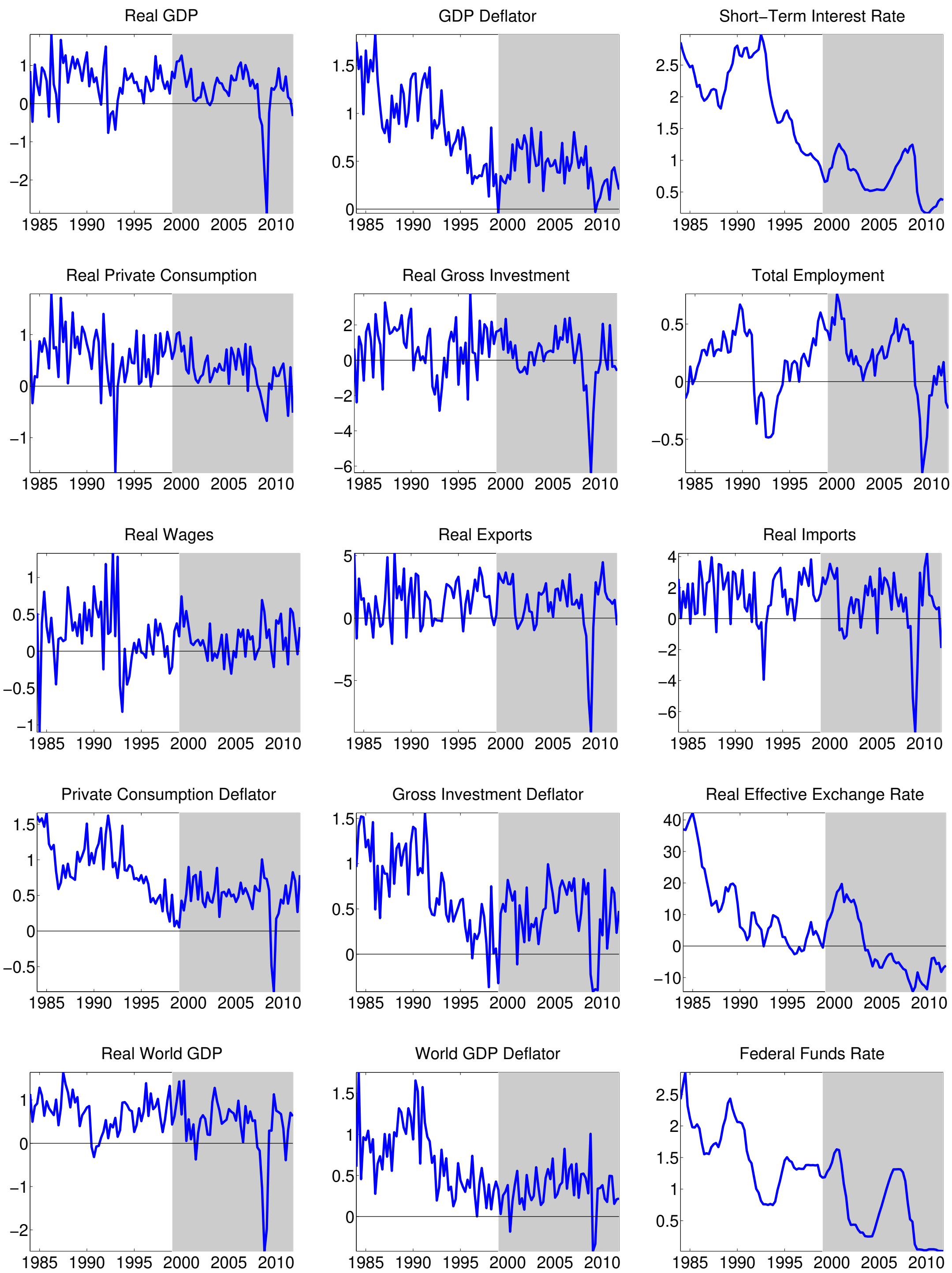

Figure A.1: Dataset. Notes: this figure shows the series included in the medium-large selection of the dataset. The initial estimation sample is 1984:1 to 1998:4, while the evaluation period runs from 1999:1 to 2011:4 (gray shaded area).

B Marginal Likelihood

The marginal likelihood is (see also Carriero et al., 2015):

$$p\,(y|\lambda) = k^{-1} \times |\Psi + (y - xb)' \left(I + x\Omega x'\right)^{-1} (y - xb)\,|^{-\frac{T+d}{2}}, \tag{B.1}$$

where

$$k = \pi^{\frac{TN}{2}} \times |\left(I + x\Omega x'\right)^{-1}|^{-\frac{N}{2}} \times |\Psi|^{-\frac{d}{2}} \times \frac{\Gamma_N\left(\frac{d}{2}\right)}{\Gamma_N\left(\frac{T+d}{2}\right)}, \tag{B.2}$$

and with $\Gamma_N(\cdot)$ denoting the N-variate gamma function.

C Convergence

In order to check for the convergence of the Markov chain, I follow Primiceri (2005) and calculate autocorrelation functions as well as inefficiency factors (IFs) for the shrinkage parameter λ and each BVAR model. Furthermore, I inspect the posterior distribution of λ for its plausibility.

The IF is defined as $\text{IF} = 1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the estimated autocorrelation of the chain at lag s. Since independence sampling produces an IF that is equal to one and dependence sampling typically produces an IF greater than one, the IF quantifies the relative efficiency loss in the computation of posterior draws from dependent versus independent samples. In practice, values around 20 are regarded as efficient (see, e.g., Primiceri, 2005, among others), meaning that the econometrician needs to draw 20 times as many draws as from uncorrelated samples.

The IFs are calculated as the inverse of the relative numerical efficiency measure (RNE) of Geweke (1992):

$$\text{RNE} = (2\pi)^{-1} \frac{1}{\text{S}\,(0)} \int_{-\pi}^{\pi} \text{S}\,(\omega)\, d\omega = \text{IF}^{-1}, \tag{C.1}$$

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

In Figure C.1 I present the autocorrelation function for the shrinkage parameter up to lag 20 for the extra-large BVAR when all observations are used in estimation. The autocorrelation is below 0.1 at the first lag and approximately 0 thereafter, suggesting that the chain mixes well and draws are almost independent. This conclusion is confirmed by the IFs that are documented for all estimation periods. The values vary between 5 and 10, indicating that draws come from the ergodic distribution. The findings are similar for the other BVARs, but not reported here for the sake of brevity. Moreover, I show the posterior distribution of the shrinkage parameter for some BVARs when all observations are used together with the prior distribution in Figure C.2. The posterior distribution is obtained by smoothing the simulated histograms with a normal kernel function. While the hyperprior is dispersed and uninformative, the posterior distributions have plausible shapes. As suggested by De Mol, Giannone, and Reichlin (2008), the size of the shrinkage parameter as well as its posterior uncertainty decline with the number of variables included, meaning that the Minnesota prior becomes tighter the larger the model is.

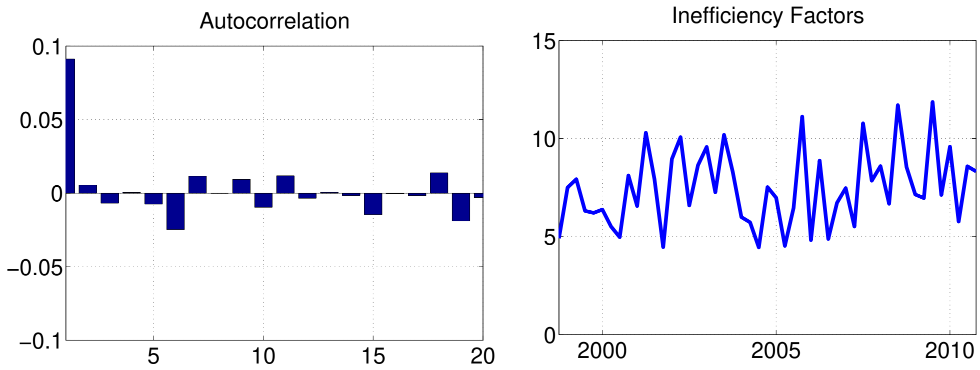

Figure C.1: Convergence Diagnostics for Shrinkage Parameter. Notes: this figure shows the autocorrelation of the posterior draws using data from 1984:1 to 2010:4 (left panel); and inefficiency factors for the period 1998:4 to 2010:4 (right panel). Both statistics are for the extra-large BVAR.

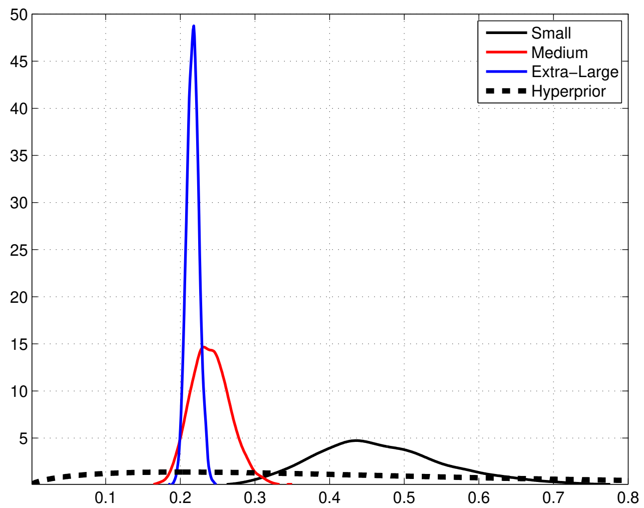

Figure C.2: Prior and Posterior Distribution for Shrinkage Parameter. Notes: this figure shows the prior and posterior distribution of the shrinkage parameter for the small, medium, and extra-large BVAR using data from 1984:1 to 2010:4.

D DSGE Models

This appendix provides the measurement equations and priors for the DSGE models. The measurement equations relate the observable series to the model variables. In order to denote the observables I use the mnemonics shown in Table A.1, while I adopt the original notation for the model parameters and variables. For the Adolfson-Laseén-Lindé-Villani model I add small measurement errors, accounting for approximately 10% of the variance of a series, which stabilizes inference. The priors are also from the original articles, but slightly modified. For instance I add steady-state labor force growth to series that are not measured in per capita terms to avoid detrending before estimation. Moreover, I fix some additional parameters at reasonable values to keep repeated recursive estimation manageable. Note: *StDev* = Standard deviation.

D.1 An-Schorfheide

The measurement equations are

$$\Delta \mathrm{YER}_t = \bar{\gamma} + \bar{e} + \Delta \hat{y}_t + \hat{z}_t \tag{D.1}$$

$$\Delta \mathrm{YED}_t = \pi^A/4 + \hat{\pi}_t \tag{D.2}$$

$$\mathrm{STN}_t = \left(\pi^A + r^A\right)/4 + \bar{\gamma} + \hat{R}_t \tag{D.3}$$

Table D.1: Priors (An-Schorfheide)

Parameter & Description		Density	Mean	StDev
τ	Inverse intertemporal elasticity of substitution	Gamma	2.00	0.50
κ	Slope of Phillips curve	Gamma	0.20	0.10
ψ_2	Monetary policy response to output growth	Gamma	0.50	0.25
ρ_R	Interest rate smoothing	Beta	0.50	0.20
ρ_G	Autocorrelation of government spending shock	Beta	0.80	0.10
ρ_Z	Autocorrelation of TFP growth shock	Beta	0.50	0.20
r^A	Transformation of discount factor (annual)	Gamma	2.00	0.50
π^A	Steady-state inflation (annual)	Gamma	4.00	1.00
$\bar{\gamma}$	Steady-state TFP growth	Normal	0.40	0.10
σ_R	Standard deviation of monetary policy shock	InvGamma	0.25	0.13
σ_G	Standard deviation of government spending shock	InvGamma	1.25	0.66
σ_Z	Standard deviation of TFP growth shock	InvGamma	0.63	0.33
Parameters fixed at constant value				
$\bar{e}$	Steady-state labor force growth		0.20	
ψ_1	Monetary policy response to inflation		1.50	

D.2 Smets-Wouters

The measurement equations are

$$\Delta \text{YER}_t = \bar{\gamma} + \bar{e} + \Delta \hat{y}_t \tag{D.4}$$

$$\Delta \text{YED}_t = \bar{\pi} + \hat{\pi}_t \tag{D.5}$$

$$\text{STN}_t = \bar{r} + \hat{R}_t \tag{D.6}$$

$$\Delta \text{PCR}_t = \bar{\gamma} + \bar{e} + \Delta \hat{c}_t \tag{D.7}$$

$$\Delta \text{ITR}_t = \bar{\gamma} + \bar{e} + \Delta \hat{i}_t \tag{D.8}$$

$$\Delta \text{LNN}_t = \bar{e} + \Delta \hat{e}_t \tag{D.9}$$

$$\Delta \left(\text{WRN/YED}\right)_t = \bar{\gamma} + \Delta \hat{w}_t \tag{D.10}$$

Table D.2: Priors (Smets-Wouters)

Parameter & Description		Density	Mean	StDev
ψ	Elasticity of capital adjustment cost function	Beta	0.50	0.15
Φ	Fixed costs in production	Normal	1.45	0.25
ρ	Interest rate smoothing	Beta	0.75	0.10
$r_{\Delta y}$	Monetary policy response to output growth	Gamma	0.50	0.25
$\bar{\pi}$	Steady-state inflation	Gamma	1.00	0.25
r^Q	Transformation of discount factor	Gamma	0.50	0.12
$\bar{\gamma}$	Steady-state labor augmenting technology growth	Normal	0.40	0.10
$\bar{e}$	Steady-state labor force growth	Normal	0.20	0.50
ρ_i	Autocorrelation of all shocks	Beta	0.50	0.20
μ_i	Moving average term of price & wage mark-up shock	Beta	0.50	0.20
σ_i	Standard deviation of all shocks	InvGamma	0.10	∞
Parameters fixed at constant value				
δ	Depreciation rate		0.025	
g_y	Steady-state government spending to output ratio		0.18	
λ_w	Steady-state labor market mark-up		1.50	
ϵ_w	Kimball labor market aggregator		10	
ϵ_p	Kimball goods market aggregator		10	
φ	Steady-state elasticity of capital adj. cost fct.		5.00	
σ_c	Intertemporal elasticity of substitution		1.50	
h	Habit formation		0.70	
ξ_w	Calvo wages		0.75	
σ_l	Labor supply elasticity		2.00	
ξ_p	Calvo prices		0.75	
ι_w	Indexation wages		0.75	
ι_p	Indexation prices		0.75	
r_π	Monetary policy response to inflation		1.50	
α	Capital share in production		0.30	
ξ_e	Calvo employment		0.50	

D.3 Adolfson-Laseén-Lindé-Villani

The measurement equations are

$$\Delta \text{YER}_t = 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + \Delta \hat{y}_t + \eta_t^{YER} \tag{D.11}$$

$$\Delta \text{YED}_t = 100 \cdot \log \pi + \hat{\pi}_t + \eta_t^{YED} \tag{D.12}$$

$$\text{STN}_t = 100 \cdot \log R + \hat{R}_t \tag{D.13}$$

$$\begin{aligned} \Delta \text{PCR}_t = {} & 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + (1-\omega_c)\left(\Delta c_t + \eta_c \Delta \hat{\gamma}_t^{c,d}\right) \\ & + \omega_c \left(\Delta c_t + \eta_c \left(\Delta \hat{\gamma}_t^{c,d} - \Delta \hat{\gamma}_t^{mc,d}\right)\right) + \eta_t^{PCR} \end{aligned} \tag{D.14}$$

$$\begin{aligned} \Delta \text{ITR}_t = {} & 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + (1-\omega_i)\left(\Delta i_t + \eta_i \Delta \hat{\gamma}_t^{i,d}\right) \\ & + \omega_i \left(\Delta i_t + \eta_i \left(\Delta \hat{\gamma}_t^{i,d} - \Delta \hat{\gamma}_t^{mi,d}\right)\right) + \eta_t^{ITR} \end{aligned} \tag{D.15}$$

$$\Delta \text{LNN}_t = 100 \cdot \log \mu_e + \Delta \hat{e}_t + \eta_t^{LNN} \tag{D.16}$$

$$\Delta \left(\text{WRN/YED}\right)_t = 100 \cdot \log \mu_z + \Delta \hat{w}_t + \eta_t^{WRN} \tag{D.17}$$

$$\Delta \text{XTR}_t = 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + \Delta \hat{y}_t^* - \eta_f \Delta \hat{\gamma}_t^x + \Delta \hat{\tilde{z}}_t^* + \eta_t^{XTR} \tag{D.18}$$

$$\begin{aligned} \Delta \text{MTR}_t = {} & 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + \omega_c \left(\Delta c_t + \eta_c \left(\Delta \hat{\gamma}_t^{c,d} - \Delta \hat{\gamma}_t^{mc,d}\right)\right) \\ & + \omega_i \left(\Delta i_t + \eta_i \left(\Delta \hat{\gamma}_t^{i,d} - \Delta \hat{\gamma}_t^{mi,d}\right)\right) + \eta_t^{MTR} \end{aligned} \tag{D.19}$$

$$\begin{aligned} \Delta \text{PCD}_t = {} & 100 \cdot \log \pi + (1-\omega_c)\left(\Delta c_t + \hat{\pi}_t + \eta_c \Delta \hat{\gamma}_t^{c,d}\right) \\ & + \omega_c \left(\Delta c_t + \hat{\pi}_t^{m,c} + \eta_c \left(\Delta \hat{\gamma}_t^{c,d} - \Delta \hat{\gamma}_t^{mc,d}\right)\right) \\ & - (1-\omega_c)\left(\Delta c_t + \eta_c \Delta \hat{\gamma}_t^{c,d}\right) - \omega_c \left(\Delta c_t + \eta_c \left(\Delta \hat{\gamma}_t^{c,d} - \Delta \hat{\gamma}_t^{mc,d}\right)\right) + \eta_t^{PCD} \end{aligned} \tag{D.20}$$

$$\begin{aligned} \Delta \text{ITD}_t = {} & 100 \cdot \log \pi + (1-\omega_i)\left(\Delta i_t + \hat{\pi}_t + \eta_i \Delta \hat{\gamma}_t^{i,d}\right) \\ & + \omega_i \left(\Delta i_t + \hat{\pi}_t^{m,i} + \eta_i \left(\Delta \hat{\gamma}_t^{i,d} - \Delta \hat{\gamma}_t^{mi,d}\right)\right) \\ & - (1-\omega_i)\left(\Delta i_t + \eta_i \Delta \hat{\gamma}_t^{i,d}\right) - \omega_i \left(\Delta i_t + \eta_i \left(\Delta \hat{\gamma}_t^{i,d} - \Delta \hat{\gamma}_t^{mi,d}\right)\right) + \eta_t^{ITD} \end{aligned} \tag{D.21}$$

$$\left(\text{EEN/}\left(\text{HICP/YWD}\right)\right)_t = \bar{x} + \hat{x}_t \tag{D.22}$$

$$\Delta \text{YWR}_t = 100 \cdot \log \mu_z + 100 \cdot \log \mu_e + \Delta \hat{y}_t^* + \Delta \hat{\tilde{z}}_t^* \tag{D.23}$$

$$\Delta \text{YWD}_t = 100 \cdot \log \pi + \hat{\pi}_t^* \tag{D.24}$$

$$\text{FEDFUNDS}_t = 100 \cdot \log R + \hat{R}_t^* \tag{D.25}$$

Table D.3: Priors (Adolfson-Laseén-Lindé-Villani)

Parameter & Description		Density	Mean	StDev
λ_d	Steady-state domestic mark-up	InvGamma	2.600	∞
$\lambda_{m,c}$	Steady-state import consumption mark-up	InvGamma	2.600	∞
$\lambda_{m,i}$	Steady-state import investment mark-up	InvGamma	2.600	∞
S	Investment adjustment cost parameter	Normal	7.694	1.500
η_i	Substitution elasticity foreign/domestic investment	InvGamma	2.100	1.050
η_f	Substitution elasticity domestic/foreign good	InvGamma	2.100	1.050
μ_z	Steady-state technology growth	Normal	1.004	0.0005
μ_e	Steady-state labor force growth	Normal	1.002	0.0005
$\bar{x}$	Steady-state real exchange rate	Normal	11.75	4
ϕ	Risk premium parameter	InvGamma	0.025	∞
ρ_i	Autocorrelation of all shocks	Beta	0.850	0.100
ρ_R	Interest rate smoothing	Beta	0.800	0.050
$r_{\Delta\pi}$	Monetary policy response to change in inflation	Normal	0.300	0.100
r_x	Monetary policy response to real exchange rate	Normal	0.000	0.050
r_y	Monetary policy response to output	Normal	0.125	0.050
$r_{\Delta y}$	Monetary policy response to output growth	Normal	0.0625	0.050
σ_z	Standard deviation of unit root technology shock	InvGamma	0.450	∞
σ_ϵ	Standard deviation of stationary technology shock	InvGamma	1.500	∞
σ_Υ	Standard deviation of investment-spec. tech. shock	InvGamma	0.450	∞
σ_{z^*}	Standard deviation of asymmetric technology shock	InvGamma	0.850	∞
σ_{ς_c}	Standard deviation of consumption preference shock	InvGamma	0.450	∞
σ_{ς_h}	Standard deviation of labor supply shock	InvGamma	0.450	∞
σ_ϕ	Standard deviation of risk premium shock	InvGamma	0.100	∞
σ_R	Standard deviation of monetary policy shock	InvGamma	0.350	∞
σ_{π^c}	Standard deviation of inflation target shock	InvGamma	0.100	∞
σ_{λ_i}	Standard deviation of all mark-up shocks	InvGamma	0.650	∞

Parameters fixed at constant value

μ	Steady-state money growth	1.010
β	Discount factor	0.999
δ	Depreciation rate	0.013
α	Capital share in production	0.290
A_L	Constant labor disutility	7.500
σ_L	Labor supply elasticity	1.000
λ_w	Wage mark-up	1.050
ω_c	Import share in consumption	0.310
ω_i	Import share in investment	0.550
ρ_π	Autocorrelation of inflation target shock	0.975
g_r	Steady-state government spending to output ratio	0.204
A_q	Cash-money ratio	0.378
σ_q	Money demand parameter	10.620
η_c	Substitution elasticity foreign/domestic consumption	5.000
σ_a	Capital utilization parameter	10^6
ρ_{λ_d}	Autocorrelation of domestic mark-up shock	0.000
λ_x	Steady-state export mark-up	1.000
ν	Steady-state fraction of wages financed in advance	1.000
ξ_w	Calvo wages	0.700
ξ_d	Calvo domestic prices	0.900
$\xi_{m,c}$	Calvo import consumption prices	0.450
$\xi_{m,i}$	Calvo import investment prices	0.750
ξ_x	Calvo export prices	0.650
ξ_e	Calvo employment	0.800
κ_w	Indexation wages	0.500
κ_d	Indexation domestic prices	0.200
$\kappa_{m,c}$	Indexation import consumption prices	0.150
$\kappa_{m,i}$	Indexation import investment prices	0.200
κ_x	Indexation export prices	0.150
b	Habit formation	0.700
r_π	Monetary policy response to inflation	1.700

E Additional Results

Table E.1: Trace and Log Determinant of Scaled MSE Matrix (excl. Great Recession)

Forecast Model	Forecast Horizon h (in Quarters) 1	2	3	4
	Small Selection ($m = 3$)			
OLS-VAR Small	2.71 (−2.15)	2.60 (−1.64)	2.81 (−0.97)	3.61 (−0.06)
OLS-VAR Medium	2.88 (−2.14)	2.71 (−1.81)	2.85 (−1.22)	3.45 (−0.41)
OLS-VAR Medium-Large	3.78 (−0.97)	4.19 (−0.06)	4.04 (0.71)	4.74 (1.20)
BVAR Small	2.63 (−2.26)	2.53 (−1.68)	2.81 (−0.94)	3.53 (−0.10)
BVAR Medium	2.63 (−2.12)	2.49 (−1.64)	2.88 (−0.82)	3.42 (−0.19)
BVAR Medium-Large	2.30 (−2.26)	2.39 (−1.53)	2.89 (−0.57)	3.29 (0.06)
BVAR Large	1.82 (−2.48)	2.23 (−1.20)	3.02 (−0.41)	3.97 (0.20)
BVAR Extra-Large	2.32 (−2.11)	2.25 (−1.35)	2.78 (−0.77)	3.52 (0.10)
DSGE Small	3.12 (−1.50)	3.58 (−0.42)	4.45 (0.40)	5.85 (1.11)
DSGE Medium	3.94 (−0.78)	2.62 (−0.61)	3.07 (−0.09)	3.35 (0.16)
DSGE Medium-Large	2.39 (−2.13)	2.35 (−1.12)	3.35 (−0.17)	3.43 (0.47)
	Medium Selection ($m = 7$)			
OLS-VAR Medium	8.78 (−3.44)	8.39 (−3.73)	9.25 (−3.04)	10.75 (−2.21)
OLS-VAR Medium-Large	13.59 (−0.61)	13.42 (−0.31)	11.12 (0.32)	11.59 (−0.12)
BVAR Medium	7.21 (−4.20)	6.87 (−4.00)	7.90 (−3.27)	9.19 (−2.53)
BVAR Medium-Large	6.91 (−4.16)	7.05 (−3.71)	7.56 (−2.83)	8.48 (−2.26)
BVAR Large	6.16 (−4.20)	6.78 (−2.87)	8.08 (−1.92)	9.78 (−1.27)
BVAR Extra-Large	7.65 (−3.24)	7.01 (−2.32)	7.16 (−1.98)	7.99 (−0.90)
DSGE Medium	10.86 (−0.47)	9.87 (−1.29)	10.74 (−1.37)	10.91 (−1.21)
DSGE Medium-Large	6.61 (−4.38)	7.42 (−3.29)	9.54 (−1.87)	10.79 (−1.60)
	Medium-Large Selection ($m = 15$)			
OLS-VAR Medium-Large	25.14 (−8.32)	24.64 (−7.57)	24.54 (−5.88)	26.92 (−6.89)
BVAR Medium-Large	15.42 (−14.36)	16.19 (−13.16)	18.35 (−10.68)	20.86 (−10.07)
BVAR Large	14.40 (−13.09)	15.55 (−10.94)	19.22 (−8.91)	22.11 (−7.61)
BVAR Extra-Large	15.99 (−11.24)	16.25 (−9.97)	18.87 (−7.93)	20.38 (−6.18)
DSGE Medium-Large	14.92 (−14.63)	18.94 (−12.63)	24.32 (−10.40)	28.81 (−8.97)

Notes: The evaluation period runs from 1999:h to 2007:h. See also notes to Table 1.

Table E.2: Log Predictive Score (excl. Great Recession)

Forecast Model	Forecast Horizon h (in Quarters)			
	1	2	3	4
	Small Selection ($m = 3$)			
OLS-VAR Small	-125.5	-119.6	-116.7	-120.1
OLS-VAR Medium	-127.7	-119.4	-116.1	-118.1
OLS-VAR Medium-Large	-154.8	-137.7	-126.3	-124.4
BVAR Small	-122.9	-118.2	-117.0	-120.7
BVAR Medium	-123.0	-118.2	-118.5	-121.3
BVAR Medium-Large	-124.6	-120.8	-121.3	-122.5
BVAR Large	-125.5	-123.9	-125.0	-124.7
BVAR Extra-Large	-130.2	-122.3	-122.1	-121.8
DSGE Small	-116.2	-115.3	-117.6	-121.3
DSGE Medium	-137.5	-117.6	-116.7	-118.7
DSGE Medium-Large	-115.4	-109.2	-110.1	-111.5
	Medium Selection ($m = 7$)			
OLS-VAR Medium	-320.6	-285.6	-278.5	-282.1
OLS-VAR Medium-Large	-394.6	-319.0	-295.0	-283.5
BVAR Medium	-297.7	-282.3	-281.1	-286.7
BVAR Medium-Large	-305.3	-287.3	-282.6	-282.7
BVAR Large	-310.6	-299.2	-295.4	-289.9
BVAR Extra-Large	-333.4	-297.5	-288.8	-280.1
DSGE Medium	-312.1	-277.3	-273.8	-273.4
DSGE Medium-Large	-261.0	-252.8	-255.6	-255.2
	Medium-Large Selection ($m = 15$)			
OLS-VAR Medium-Large	-1058.1	-781.6	-725.0	-707.1
BVAR Medium-Large	-749.5	-718.6	-729.3	-742.4
BVAR Large	-793.7	-748.5	-736.6	-731.6
BVAR Extra-Large	-836.0	-758.8	-733.5	-716.9
DSGE Medium-Large	-674.1	-672.1	-684.3	-698.1

Notes: The evaluation period runs from 1999:h to 2007:h. See also notes to Table 2.