EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1536 / APRIL 2013

PREDICTIVE LIKELIHOOD COMPARISONS WITH DSGE AND DSGE-VAR MODELS

Anders Warne, Günter Coenen
and Kai Christoffel

In 2013 all ECB publications feature a motif taken from the €5 banknote.

Acknowledgements
This paper was previously circulated with the title "Forecasting with DSGE-VARs". We are particularly grateful to Marta Bańbura who has specified the large Bayesian VAR model we have used in the paper. We are also grateful for discussions with Gianni Amisano (ECB), Michal Andrle (IMF), Jan Brůha (Czech National Bank), Herman van Dijk (Tinbergen Institute), Juha Kilponen (Suomen Pankki), Bartosz Maćkowiak (ECB), Frank Schorfheide (University of Pennsylvania), Mattias Villani (Linköping University), and comments from members of the Working Group on Econometric Modelling, and participants in the Tinbergen Institute workshop on "Recent Theory and Applications of DSGE Models" at Erasmus University Rotterdam and the CEF 2012 conference in Prague. The opinions expressed in this paper are those of the authors and do not necessarily reflect views of the European Central Bank or the Eurosystem. Any remaining errors are the sole responsibility of the authors.

Anders Warne
European Central Bank; e-mail: anders.warne@ecb.europa.eu

Günter Coenen
European Central Bank; e-mail: gunter.coenen@ecb.europa.eu

Kai Christoffel
European Central Bank; e-mail: kai.christoffel@ecb.europa.eu

Address	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	http://www.ecb.europa.eu
Fax	+49 69 1344 6000

ISSN	1725-2806 (online)
EU Catalogue No	QB-AR-13-033-EN-N (online)

Contents

Abstract: This paper shows how to compute the h-step-ahead predictive likelihood for any subset of the observed variables in parametric discrete time series models estimated with Bayesian methods. The subset of variables may vary across forecast horizons and the problem thereby covers marginal and joint predictive likelihoods for a fixed subset as special cases. The basic idea is to utilize well-known techniques for handling missing data when computing the likelihood function, such as a missing observations consistent Kalman filter for linear Gaussian models, but it also extends to nonlinear, nonnormal state-space models. The predictive likelihood can thereafter be calculated via Monte Carlo integration using draws from the posterior distribution. As an empirical illustration, we use euro area data and compare the forecasting performance of the New Area-Wide Model, a small-open-economy DSGE model, to DSGE-VARs, and to reduced-form linear Gaussian models.

Keywords: Bayesian inference, forecasting, Kalman filter, missing data, Monte Carlo integration.

JEL Classification Numbers: C11, C32, C52, C53, E37.

Non-Technical Summary

A dynamic stochastic general equilibrium (DSGE) model, like any other model, is an artificial construct and is therefore always *false*. While this may present a challenge for classical inference, an important aspect of Bayesian inference is precisely that it does *not* need to rely on the assumption that a model is correctly specified. However, the mere fact that a model is misspecified is not particularly interesting per se unless we also have the means to analyse the implications of the specification error. That way, we may be able to separate the features of the model that are sensitive to misspecification from those that appear to be robust. The so-called DSGE-VAR approach, advocated in a series of articles by Del Negro and Schorfheide (2004, 2006, 2009), has been suggested as a tool for studying misspecification of a DSGE model by approximating it with a VAR and allowing the cross-equation restrictions of the DSGE model to be relaxed in a flexible manner.

DSGE-VARs may be indexed by a single parameter which determines the weight on the prior relative to the data. The DSGE model approximation resides at one end of its range, an unrestricted VAR at the other end, and in between these two extremes a large number of models exist. Apart from providing a measure of the degree to which the DSGE model is misspecified, the approach also allows for posterior analysis of the DSGE model parameters, impulse-response analysis, forecast-error-variance decompositions, and so on.

This paper adds to the forecasting literature by making two contributions. First, we present a simple approach to Bayesian estimation of the predictive distribution of a DSGE-VAR model, including its population mean and h-step-ahead covariance matrix. An important purpose of statistical analysis is to not only make sequential forecasts of the future, but also to provide suitable measures of the uncertainty surrounding them. While point forecasts are often of first-order importance, the predictive distribution has during the last few decades been given an increasingly important role in both the theoretical forecasting literature and in empirical forecast comparison and evaluation exercises.

The second and main contribution of our paper concerns the calculation of the predictive likelihood and the suggested approach can be applied to a large family of models, with Gaussian log-linearized DSGE models, DSGE-VARs and BVARs as interesting special cases. Specifically, we show how the marginal and joint h-step-ahead predictive likelihood can be calculated for any *subset* (partition) of the variables over the forecast horizon.

The predictive likelihood for a subset of the observed variables makes it possible to compare forecast accuracy across models that have different conditioning variables but where some of the forecasted variables are shared. A forecaster or policy maker is typically only interested in a limited number of the variables that a multivariate model can predict and a forecast comparison statistic based only on the variables of interest is therefore desirable. A special case is when we are interested in comparing the forecasts of a single variable, such as inflation, across a set of models, whereas larger subsets are required when we are also concerned with predicting

comovements. Although point forecasts may reveal interesting aspects of the models involved, such forecasts are not well suited for model selection, the determination of model weights in a model averaging study, or the construction of optimal prediction pools. The predictive likelihood is a natural tool for dealing with such matters, with the log predictive score—the sum of the log predictive likelihood over the forecast sample—as an extension from single-period comparisons based on the predictive Bayes factor to a multi-period setting.

In our empirical illustration with five linear Gaussian models, the suggested (consistent and unbiased) estimator of the predictive likelihood is compared with a normal predictive likelihood, constructed from the mean vector and the covariance matrix of the predictive distribution. The analysis is an extension of the Christoffel, Coenen, and Warne (2011) study for euro area data and compares the results for the New Area-Wide Model (NAWM), two DSGE-VAR models with the NAWM as prior, a large BVAR, and a multivariate random walk model.

Over the forecast sample 1999Q1 until 2006Q4 we find that the normal density provides a good approximation of the predictive likelihood when examining the density forecasts for the five models. In terms of a model ranking, the log predictive score strongly favors the BVAR model, with the two DSGE-VAR models improving somewhat on the density forecasts of the NAWM, especially at the shorter horizons. The random walk model, on the other hand, is only competitive with the NAWM at the one-step-ahead horizon, especially for the small selection of variables with real GDP growth, GDP deflator inflation, and the short-term nominal interest rate. Moreover, when examining the individual predictive likelihood values it is found that the ranking of models is not determined by one or two outliers, but is broadly supported throughout the forecast sample.

1. Introduction

A dynamic stochastic general equilibrium (DSGE) model, like any other model, is an artificial construct and is therefore always *false*. While this may present a challenge for classical inference, an important aspect of Bayesian inference is precisely that it does *not* need to rely on the assumption that a model is correctly specified. However, the mere fact that a model is misspecified is not particularly interesting per se unless we also have the means to analyse the implications of the specification error. That way, we may be able to separate the features of the model that are sensitive to misspecification from those that appear to be robust. The so-called DSGE-VAR approach, advocated in a series of articles by Del Negro and Schorfheide (2004, 2006, 2009), has been suggested as a tool for studying misspecification of a DSGE model by approximating it with a VAR and allowing the cross-equation restrictions of the DSGE model to be relaxed in a flexible manner.

The idea of using VARs as an alternative modelling device to a DSGE model can be traced back to the literature on indirect inference; see, e.g., Smith (1993). An early attempt to combine DSGE models with Bayesian VARs is Ingram and Whiteman (1994), where the VAR parameters were expressed as a function of the DSGE model parameters. A prior for the DSGE model parameters then implied a prior for the VAR parameters through a first-order Taylor expansion of the mapping. This idea was considerably enriched by Del Negro and Schorfheide (2004), where the prior distribution of the VAR model parameters was determined from the DSGE model by parameterizing the distribution through the implied first and second moments of the DSGE model.

DSGE-VARs may be indexed by a single parameter which determines the weight on the prior relative to the data. The DSGE model approximation resides at one end of its range, an unrestricted VAR at the other end, and in between these two extremes a large number of models exist. Apart from providing a measure of the degree to which the DSGE model is misspecified, the approach also allows for posterior analysis of the DSGE model parameters, impulse-response analysis, forecast-error-variance decompositions, and so on. While these models were first designed to improve forecasting and monetary policy analysis with VARs (Del Negro and Schorfheide, 2004), the extension to a model evaluation toolkit was carried out by Del Negro and Schorfheide (2006), while Del Negro, Schorfheide, Smets, and Wouters (2007) used it to assess the goodness-of-fit of a DSGE model.

This paper adds to the forecasting literature by making two contributions. First, we present a simple approach to Bayesian estimation of the predictive distribution of a DSGE-VAR model, including its population mean and h-step-ahead covariance matrix. Dawid (1984) points out that an important purpose of statistical analysis is to not only make sequential forecasts of the future, but also to provide suitable measures of the uncertainty surrounding them. While point forecasts are often of first-order importance, the predictive distribution has since Dawid's article on the *prequential* approach been given an increasingly important role in both the theoretical

forecasting literature and in empirical forecast comparison and evaluation exercises; see Tay and Wallis (2000) for a survey.

The second and main contribution of our paper concerns the calculation of the predictive likelihood and the suggested approach can be applied to a large family of models, with Gaussian log-linearized DSGE models, DSGE-VARs and BVARs as interesting special cases. Specifically, we show how the marginal and joint h-step-ahead predictive likelihood can be calculated for any *subset* (partition) of the variables over the forecast horizon.

It has long been recognized that using the predictive likelihood is a valid Bayesian approach to model selection (see, e.g., Box, 1980), and the predictive Bayes factor is naturally defined from a ratio of two predictive likelihoods. The predictive likelihood for a subset of the observed variables makes it possible to compare forecast accuracy across models that have different conditioning variables but where some of the forecasted variables are shared. A forecaster or policy maker is typically only interested in a limited number of the variables that a multivariate model can predict and a forecast comparison statistic based only on the variables of interest is therefore desirable. A special case is when we are interested in comparing the forecasts of a single variable, such as inflation, across a set of models, whereas larger subsets are required when we are also concerned with predicting comovements. Although point forecasts may reveal interesting aspects of the models involved, such forecasts are not well suited for model selection, the determination of model weights in a model averaging study, or the construction of optimal prediction pools; see, e.g., Geweke and Amisano (2011, 2012). The predictive likelihood is a natural tool for dealing with such matters, with the log predictive score as an extension from single-period comparisons based on the predictive Bayes factor to a multi-period setting.

We illustrate our contributions to the literature by extending the forecast comparison exercise in Christoffel, Coenen, and Warne (2011)—henceforth CCW—to DSGE-VARs. CCW review forecasting with DSGE models, using the New Area-Wide Model (NAWM; pronounced nŏm) as an example, and their pseudo out-of-sample forecasting exercise covers the period after the introduction of the euro, focusing on partitions of the 12 observed variables in the NAWM (out of 18) that are endogenously determined. CCW compare the forecasts of the NAWM to those obtained from a VAR estimated with maximum likelihood, three BVARs as well as the naïve random walk and sample mean benchmarks. Concerning the BVARs, two classes are considered by CCW: one that is intended for systems with a smaller dimension and that makes use of a prior on the steady state, and one that has been proposed for large data sets; see Bańbura, Giannone, and Reichlin (2010). The random walk and the best performing BVAR model are included in our empirical illustration for comparisons with the NAWM and the DSGE-VAR models.[1]

[1] An interesting approach to Bayesian forecast evaluation in a multivariate setting is suggested by Herbst and Schorfheide (2012). Their goal is to evaluate whether the actual pseudo out-of-sample forecast performance of a model is consistent with the performance that is expected under the predictive distribution of future observations implied by the model. Their evaluation scheme is therefore related to the idea of predictive checks (see, e.g., Geweke, 2005) and the idea of examining whether the predictive distribution of the model is, using the terminology of Dawid (1982), "well calibrated". This contrasts with our approach which compares the accuracy across models.

The paper is organized as follows. Section 2 briefly outlines forecasting with DSGE models. Next, the DSGE-VAR model is discussed in Section 3, before we turn to the issue of forecasting with DSGE-VARs in Section 4. The calculation of the marginal and joint h-step-ahead predictive likelihood for a subset of the variables is discussed in Section 5. In Section 6 we turn to the empirical illustration, beginning with a short introduction to the NAWM before we present the results from estimating DSGE-VARs using the NAWM as a prior. The forecast comparison exercises of the DSGE-VAR relative to the NAWM, a large BVAR and the random walk are discussed in Section 7. Finally, Section 8 summarizes the main findings of the paper.

2. Forecasting with DSGE Models

Let $\theta \in \Theta$ be the vector of parameters of a log-linearized DSGE model that are unknown and are to be estimated. Bayesian inference is based on combining a likelihood function with a prior distribution in order to obtain a posterior distribution of the model parameters conditional on the observed data. The log-linearized DSGE model with rational expectations can be written as:

$$A_{-1}\xi_{t-1} + A_0\xi_t + A_1 E_t\xi_{t+1} = D\eta_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where η_t is a q-dimensional vector with i.i.d. standard normal structural shocks ($\eta_t \sim N(0, I_q)$), while ξ_t is an r-dimensional vector of model variables, defined as deviations from the steady state. The matrices A_i $(r \times r)$, with $i = -1, 0, 1$, and D $(r \times q)$ are functions of θ.

Provided that a unique and convergent solution of the system (1) exists at a particular value of θ (see, e.g., Anderson, 2010, Klein, 2000, or Sims, 2002), we can express the model variables as a VAR system:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1, \ldots, T, \tag{2}$$

where F $(r \times r)$ and B $(r \times q)$ are uniquely determined by θ. The observed variables are denoted by y_t, an n-dimensional vector, which is linked to the model variables ξ_t through the equation

$$y_t = \mu + H'\xi_t + w_t, \quad t = 1, \ldots, T. \tag{3}$$

The measurement error, w_t, is assumed to be i.i.d. Gaussian with mean zero and covariance matrix R, while μ is the population mean (steady state) of y_t conditional on θ provided that ξ_t is stationary. The measurement errors and the shocks η_t are assumed to be independent, while the matrices H and R are also uniquely determined by θ.

The system in (2) and (3) is a state-space model, where equation (2) gives the state or transition equation and (3) the measurement or observation equation. Sargent (1989) was among the first to recognize that linear(ized) rational expectations models can be cast in this form. Provided the number of measurement errors and structural shocks is large enough, we can calculate the likelihood function for the observed data $\mathcal{Y}_T = \{y_1, \ldots, y_T\}$ given a value of θ with the Kalman filter; see, e.g., Durbin and Koopman (2012) for details.

Point and density forecasts are both determined from the predictive density and, for a sequence of future values of the observed variable $y_{T+1}, \ldots, y_{T+h}$, this density can be expressed as

$$p\big(y_{T+1}, \ldots, y_{T+h}|\mathcal{Y}_T\big) = \int_{\theta\in\Theta} p\big(y_{T+1}, \ldots, y_{T+h}|\mathcal{Y}_T, \theta\big)p\big(\theta|\mathcal{Y}_T\big)d\theta, \tag{4}$$

where $p(y_{T+1}, \ldots, y_{T+h}|\mathcal{Y}_T, \theta)$ is a Gaussian, and $p(\theta|\mathcal{Y}_T)$ is the posterior density of θ based on the data available at time T.

From a Bayesian perspective, it may be noticed that for a given model there is no uncertainty about the predictive density and, hence, there is no uncertainty about a point or a density forecast which is determined from it. This can be seen in equation (4) where posterior parameter uncertainty is integrated out and what remains is a deterministic function of the data and the model. In practise, numerical methods typically need to be applied, but the induced simulation uncertainty can be controlled by the econometrician.[2]

If we wish to estimate quantiles, confidence regions (credible sets), or the probability that the variables reach some barrier, then we need a numerical algorithm for computing the predictive density since the integral in (4) cannot be solved analytically. On the other hand, if the forecast study only requires moments from the predictive distribution, then such an algorithm is not needed since the moments can be estimated with high precision using draws from the posterior distribution of the parameters.

A numerical algorithm for evaluating the integral in (4) for ARIMA models was suggested by Thompson and Miller (1986). The basic idea is that M_1 paths for the observed variables over the period $T+1, \ldots, T+h$ are simulated by drawing randomly from the distribution conditional on the parameters for each of the M_2 random draws of θ from its posterior density, yielding a total of $M = M_1M_2$ paths from which the properties of the predictive density can be examined. This so-called *sampling the future* algorithm has been adapted by Adolfson, Lindé, and Villani (2007) to state-space models.

Adolfson et al. also discuss estimation of the population mean and covariance matrix of the predictive distribution. Specifically, the mean of the marginal predictive distribution is

$$E\big[y_{T+i}\big|\mathcal{Y}_T\big] = E_T\big[\mu + H'F^i\xi_{T|T}\big], \quad i = 1, \ldots, h, \tag{5}$$

where E_T denotes the expectation with respect to the posterior of θ at time T, and where $\xi_{T|T}$ is the Kalman filter estimate of ξ_T based on $\mathcal{Y}_T$. Based on the *law of iterated expectations* the forecast uncertainty for y_{T+i} can be decomposed (Rao-Blackwellization) as follows:

$$C\big(y_{T+i}|\mathcal{Y}_T\big) = E_T\big[C\big(y_{T+i}|\mathcal{Y}_T;\theta\big)\big] + C_T\big[E\big(y_{T+i}|\mathcal{Y}_T;\theta\big)\big], \tag{6}$$

[2] This may be contrasted with a frequentist approach to forecasting where a point or a density forecast is conditioned on the unknown parameters of the model, i.e. it is based on the first density term on the right hand side of (4). Once the unknown θ is replaced with a point estimate, the resulting point or density forecast is subject to the estimation uncertainty inherent to the selected estimator and sample period and which cannot be influenced by the econometrician. At the same time, the "true" predictive density, based on the "true" values of the parameters, is deterministic but remains unknown; see Geweke and Whiteman (2006) for discussions on Bayesian forecasting.

where C_T denote the covariance with respect to the posterior of θ at time T. The first term on the right hand side of (6) is given by

$$E_T\big[C\big(y_{T+i}\big|\mathcal{Y}_T;\theta\big)\big] = E_T\big[H'F^iP_{T|T}\big(F^i\big)'H\big] + E_T\left[H'\left(\sum_{j=0}^{i-1}F^jBB'\big(F'\big)^j\right)H\right] + E_T\big[R\big], \quad (7)$$

where $P_{T|T}$ is the covariance matrix of $\xi_T - \xi_{T|T}$. This equation provides the uncertainties due to not observing the current state, future shocks, and future measurement errors, respectively. Similarly, for the second term of the right hand side in (6) we have that

$$C_T\big[E\big(y_{T+i}\big|\mathcal{Y}_T;\theta\big)\big] = C_T\big[\mu + H'F^i\xi_{T|T}\big], \quad (8)$$

which thus reflects the influence of parameter uncertainty on overall forecast uncertainty. The Monte Carlo estimation of these population moments only require draws from the posterior distribution of θ and can therefore be performed without using the sampling the future algorithm. The computation of the joint and marginal predictive likelihood for a subset of the variables can likewise also be performed using only posterior draws. This issue is discussed in Section 5.[3]

3. DSGE-VAR Models

The state-space representation in (2) and (3) can be expressed as a finite-order VARMA representation when the state variables are stationary; see, e.g., Ravenna (2007). If the MA term of the model is invertible the DSGE model has an infinite-order VAR representation, as in Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), and under very specific conditions the VAR representation may even be finite. For models with more shocks and measurement errors than observed variables ($q + \text{rank}(R) > n$), such as the NAWM, the so-called "poor man's invertibility condition" in Fernández-Villaverde et al. (2007)—generally a sufficient condition for the existence of an infinite-order VAR representation—is not satisfied; see Franchi and Paruolo (2012) for details.[4] Nevertheless, a finite order VAR may still serve as a reasonable approximation of a DSGE model.

To setup the DSGE-VAR, we follow Del Negro and Schorfheide (2004) and proceed as follows. The VAR representation of y_t can be written as:

$$y_t = \Phi_0 + \sum_{j=1}^{p}\Phi_j y_{t-j} + \epsilon_t, \quad t = 1,\ldots,T, \quad (9)$$

where $\epsilon_t \sim N_n(0,\Sigma_\epsilon)$. The vector Φ_0 is $n \times 1$, while Φ_j is $n \times n$ for $j = 1,\ldots,p$. We assume that initial values for y_t exists for $t = 0,\ldots,1-p$. Let $X_t = [1\ y'_{t-1}\ \cdots\ y'_{t-p}]'$ be an $(np+1)$-dimensional vector, while the $n \times (np+1)$ matrix $\Phi = [\Phi_0\ \Phi_1\ \cdots\ \Phi_p]$. This means that the

[3] For additional discussions on forecasting with DSGE models, see Del Negro and Schorfheide (2012).

[4] It may be noted that when the state transition matrix F is stable and the system is "square" ($n = q + \text{rank}[R]$), then the poor man's invertibility condition is also necessary. For instance, the Smets and Wouters (2007) model has the same number of observed variables and shocks, zero measurement errors, and a stable F matrix, yet the matrix $(I - B(H'B)^{-1}H')F$ has a unit eigenvalue and the model therefore fails the now both necessary and sufficient invertibility condition; see, Franchi and Paruolo (2012, Corollary 5.3).

VAR can be expressed as

$$y_t = \Phi X_t + \epsilon_t. \tag{10}$$

Del Negro and Schorfheide (2004) suggest that a VAR approximation of the DSGE model can be obtained by replacing the VAR parameters by the implied "estimates" using the population moments conditional on θ. This means that the $n \times 1$ population mean vector for y_t is μ, while the central population autocovariance matrices for (y_t, y_{t-j}) are given by:

$$\Sigma_y^{(j)} = \begin{cases} H'\Sigma_\xi H + R, & \text{if } j = 0, \\ H'F^j\Sigma_\xi H, & \text{for } j = 1, 2, \ldots, \end{cases} \tag{11}$$

with $\Sigma_y^{(-j)} = \Sigma_y^{(j)\prime}$. The state-variable covariance matrix Σ_ξ is obtained via the state equation by solving the Lyapunov equation

$$\Sigma_\xi = F\Sigma_\xi F' + BB'.$$

For small DSGE models this last step may be performed analytically via vectorization while a numerical approach, such as the doubling algorithm, is usually recommended for large or medium-sized models like the NAWM.

Let $\Gamma_{XX}(\theta)$ denote the non-central population covariance matrix $E[X_t X_t'; \theta]$ and $\Gamma_{yX}(\theta)$ the non-central population covariance matrix $E[y_t X_t'; \theta]$ which can be directly determined from (11) and the mean vector μ. The population-based regression of y_t on X_t determines the mapping from the DSGE model parameters to the VAR parameters. Specifically, suppose that $\Gamma_{XX}(\theta)$ is invertible, then

$$\Phi(\theta) = \Gamma_{yX}(\theta)\Gamma_{XX}^{-1}(\theta), \tag{12}$$

$$\Sigma_\epsilon(\theta) = \Gamma_{yy}(\theta) - \Gamma_{yX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{yX}'(\theta), \tag{13}$$

where $\Gamma_{yy}(\theta) = E[y_t y_t'; \theta] = \Sigma_y^{(0)} + \mu\mu'$. The matrices $\Phi(\theta)$ and $\Sigma_\epsilon(\theta)$ are restriction functions that will be used to center the prior distribution of (Φ, Σ_ϵ) conditional on θ and a hyperparameter $\lambda \geq 0$ that measures the deviation of the DSGE-VAR from the VAR approximation of the DSGE model.

The joint prior distribution of the VAR and DSGE model parameters has the following hierarchical structure:

$$p\big(\Phi, \Sigma_\epsilon, \theta|\lambda\big) = p\big(\Phi, \Sigma_\epsilon|\theta, \lambda\big)p(\theta). \tag{14}$$

The conditional prior distribution of the VAR parameters on the right hand side of (14) is centered at the VAR approximation of the DSGE model $(\Phi(\theta), \Sigma_\epsilon(\theta))$, but allows for deviations from the restrictions to account for possible misspecification. The precision of the prior is determined by the hyperparameter λ, which generates a continuum of models that have an unrestricted VAR at one extreme (λ close to zero) and the DSGE model approximation at the other ($\lambda = \infty$). In practise, a grid is used for λ such that a finite number of values

$\Lambda = \{\lambda_1, \ldots, \lambda_N, \infty\}$ are considered and each DSGE-VAR model is indexed by its λ-value, denoted by DSGE-VAR(λ).

Letting vec denote the column stacking operator, the prior of the VAR parameters takes the form

$$\Sigma_\epsilon|\theta,\lambda \sim IW_n\big(\lambda T\Sigma_\epsilon(\theta), \lambda T-(np+1)\big), \tag{15}$$

$$\mathrm{vec}(\Phi)|\Sigma_\epsilon,\theta,\lambda \sim N_{n(np+1)}\big(\mathrm{vec}(\Phi(\theta)), (\lambda T)^{-1}\big[\Gamma_{XX}^{-1}(\theta)\otimes\Sigma_\epsilon\big]\big). \tag{16}$$

The normal-inverted Wishart prior assumed here is proper when $\lambda T \geq n(p+1)+1$. Hence, the domain of λ is restricted to the interval $[(n(p+1)+1)/T, \infty)$ for the DSGE-VAR.

One interpretation of this prior is related to using dummy observations in VARs, as in Lubik and Schorfheide (2006). That is, the prior can be seen as augmenting the VAR model with λT observations of the endogenous variables; see An and Schorfheide (2007), Del Negro et al. (2007), and Lees, Matheson, and Smith (2011) for further discussions on the interpretation of the prior. For the case when $\lambda = \infty$, the VAR parameters are fully determined by the DSGE model and for any given value of θ the population mean and first p population autocovariances of the DSGE-VAR matches exactly those of the DSGE model. In this sense, the DSGE-VAR model with $\lambda = \infty$ may be viewed as a reasonable approximation of the DSGE model.

The posterior density of the VAR parameters conditional on (θ, λ) is proportional to the product of the prior density and the likelihood function for the VAR, denoted by $p(\mathcal{Y}_T|X_1, \Phi, \Sigma_\epsilon, \theta, \lambda)$. Conditional on (θ, λ), the DSGE-VAR prior and likelihood are conjugate and, hence, it follows that the posterior distribution of Φ and Σ_ϵ is also of the normal-inverted Wishart form; see Del Negro and Schorfheide (2004) for details. Notice that the conditional posterior distribution of (Φ, Σ_ϵ) depends on the initial values X_1 via the likelihood function. When we wish to compare marginal likelihoods for DSGE-VAR models to the DSGE model, the information sets need to be the same. This may be handled in the DSGE model by using the sample $t = 1-p, \ldots, 0$ as a training sample for the Kalman filter.

The joint posterior density of the DSGE and VAR model parameters can be factorized as:

$$p\big(\Phi, \Sigma_\epsilon, \theta|\mathcal{Y}_T, X_1, \lambda\big) = p\big(\Phi, \Sigma_\epsilon|\mathcal{Y}_T, X_1, \theta, \lambda\big)p\big(\theta|\mathcal{Y}_T, X_1, \lambda\big). \tag{17}$$

The first term on the right-hand side is given by the product of the conditional posterior densities for $\Phi|\Sigma_\epsilon$ and Σ_ϵ. The second term is the marginal posterior density of θ for a given λ and draws from this density can be obtained as outlined below.

The marginalized likelihood can be determined from the following factorization:

$$p\left(\mathcal{Y}_T\,|X_1, \theta, \lambda\right) = \frac{p\big(\mathcal{Y}_T|X_1, \Phi, \Sigma_\epsilon, \theta, \lambda\big)p\big(\Phi, \Sigma_\epsilon|\theta, \lambda\big)}{p\big(\Phi, \Sigma_\epsilon|\mathcal{Y}_T, X_1, \theta, \lambda\big)}. \tag{18}$$

The first term in the numerator is the likelihood function of the VAR in (9), while the second term is the prior density of the VAR parameters conditional on θ and λ; cf. equations (15) and (16). The denominator is the conditional posterior density of the VAR parameters. An analytical

expression for the marginalized likelihood in (18) is provided in Del Negro and Schorfheide (2004, eq. A.2) and this expression is further streamlined and robustified in Warne (2012, Section 15).

The marginalized likelihood is combined with the prior of the DSGE model parameters to form a posterior kernel that can be utilized by posterior simulators, such as the random walk Metropolis sampler; see, e.g., An and Schorfheide (2007). Once a posterior sample of θ has been simulated from the DSGE-VAR(λ) model, the marginal likelihood can be evaluated. Letting this function be denoted by $p(\mathcal{Y}_T|X_1, \lambda)$, Del Negro and Schorfheide (2004) suggests to select λ such that:

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda} p\big(\mathcal{Y}_T|X_1, \lambda\big). \tag{19}$$

When using DSGE-VAR models for forecasting purposes, an alternative approach to selecting the $\hat{\lambda}$ model would be to average the results over the range of λ. Instead of such model averaging we shall follow the suggestion of Del Negro and Schorfheide (2004) and select a DSGE-VAR model based on the posterior mode estimate of λ. They also suggest that the same criterion can be used to select the lag order. This may often be a good recommendation to follow, but it is also important to check that the DSGE-VAR(∞) model is a good approximation of the DSGE model; see Adolfson, Laséen, Lindé, and Villani (2008) for discussions. We will return to these issues in our empirical illustration.

4. Forecasting with DSGE-VARs

It is straightforward to apply the sampling the future procedure of Thompson and Miller (1986) to a DSGE-VAR model, but one caveat needs to be taken into account. Namely, the integration needs to cover the full parameter space of the DSGE-VAR. This means that the sampling the future algorithm for the DSGE-VAR may be formulated as:

(1) Draw $(\Phi, \Sigma_\epsilon, \theta)$ from $p(\theta|\mathcal{Y}_T, X_1, \lambda)$, $p(\Sigma_\epsilon|\mathcal{Y}_T, X_1, \theta, \lambda)$, and $p(\Phi|\mathcal{Y}_T, X_1, \Sigma_\epsilon, \theta, \lambda)$;
(2) Draw ϵ_{T+i} from $N(0, \Sigma_\epsilon)$ for $i = 1, \ldots, h$;
(3) Simulate a path for $y_{T+1}, \ldots, y_{T+h}$ by feeding the residuals into the VAR system in equation (9);
(4) Repeat steps 2-3 M_1 times for the same $(\Phi, \Sigma_\epsilon, \theta)$;
(5) Repeat steps 1-4 M_2 times.

The above algorithm generates a total of $M = M_1 M_2$ paths from the predictive distribution of the DSGE-VAR model with p lags and hyperparameter λ. For fixed M, we may expect that the quality of the numerical integration approximation is optimized for $M_1 = 1$ since the dependence between paths is likely to be the smallest for this case; see Villani (2001) for a similar algorithm. However, since the inner (M_1) loop is typically faster than the outer (M_2) loop, the quality of the approximation for a fixed computation time can be improved by selecting a value for M_1

which is larger than unity.[5] These arguments are also valid for the DSGE model algorithms discussed in Section 2.

The population mean and covariance matrices from the predictive distribution of y_{T+i}, $i = 1, \ldots, h$, can be estimated without the need for the above simulation algorithm. Let Ψ denote an $np \times np$ matrix with

$$\Psi = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & & 0 & 0 \\ & \ddots & & \\ 0 & & I_n & 0 \end{bmatrix},$$

while J_p is an $np \times n$ matrix with I_n on top and zeros below such that $y_t = J_p' Y_t$, where $Y_t = [y_t' \ \cdots \ y_{t-p+1}']'$. We now rewrite the VAR system for forecasting exercises as:

$$y_{T+i} = J_p' \bar{x}_{T+i} + J_p' \Psi^i Y_T + J_p' \bar{\epsilon}_{T+i}, \quad i = 1, \ldots, h, \tag{20}$$

where

$$\begin{aligned} \bar{x}_{T+i} &= J_p \Phi_0 + \Psi \bar{x}_{T+i-1}, \\ \bar{\epsilon}_{T+i} &= J_p \epsilon_{T+i} + \Psi \bar{\epsilon}_{T+i-1}, \quad i = 1, \ldots, h, \end{aligned}$$

and these np-dimensional vectors are initialized through $\bar{x}_T = \bar{\epsilon}_T = 0$. Notice that this reformulation of the VAR may be used directly in step 3 of the sampling the future algorithm. Furthermore, the population mean of the predictive density of y_{T+i} is given by

$$E\big(y_{T+i} \big| \mathcal{Y}_T, X_1, \lambda\big) = E_T\left[J_p' \bar{x}_{T+i} + J_p' \Psi^i Y_T\right], \quad i = 1, \ldots, h, \tag{21}$$

where E_T denotes the expectation with respect to the posterior of $(\Phi, \Sigma_\epsilon, \theta)$ at time T.

The prediction uncertainty of the DSGE-VAR(λ) can through Rao-Blackwellization be decomposed into two terms: residual or shock uncertainty and parameter uncertainty. That is,

$$C\big(y_{T+i} \big| \mathcal{Y}_T, X_1, \lambda\big) = E_T\big[C\big(y_{T+i} \big| \mathcal{Y}_T, X_1, \Phi, \Sigma_\epsilon, \theta, \lambda\big)\big] + C_T\big[J_p' \bar{x}_{T+i} + J_p' \Psi^i Y_T\big], \tag{22}$$

where C_T denotes the covariance with respect to the posterior of $(\Phi, \Sigma_\epsilon, \theta)$ at time T. Residual uncertainty is determined by the first term on the right-hand side, while the second term represents parameter uncertainty.[6] To develop a simple expression for the first term on the right hand side of (22), let $\bar{\Sigma}_Y^{(i)}$ be defined from the difference equation

$$\bar{\Sigma}_Y^{(i)} = J_p \Sigma_\epsilon J_p' + \Psi \bar{\Sigma}_Y^{(i-1)} \Psi', \quad i = 1, \ldots, h, \tag{23}$$

[5] That is, the inner loop is expected to be faster than the outer unless we already have direct access to a sufficiently large number of draws from $p(\theta|\mathcal{Y}_T, X_1, \lambda)$; draws from the conditional posteriors of the VAR parameters are not time consuming.

[6] The second term on the right-hand side of (22) is simply due to the predictive mean conditional on the parameters being different from the predictive mean.

with the $np \times np$ matrix $\bar{\Sigma}_Y^{(0)} = 0$. It now follows that the residual uncertainty term is:

$$E_T\left[C\left(y_{T+i}\middle|\mathcal{Y}_T, X_1, \Phi, \Sigma_\epsilon, \theta, \lambda\right)\right] = E_T\left[J_p' \bar{\Sigma}_Y^{(i)} J_p\right], \quad i = 1, \ldots, h. \tag{24}$$

To estimate these population moments in practise we average over a sample of S draws from the posterior distribution of $(\Phi, \Sigma_\epsilon, \theta)$. The numerical standard error of these sample moments is negligible, provided that S is sufficiently large and the posterior sampler has converged. Since direct sampling of the VAR parameters conditional on the DSGE model parameters is possible, convergence of the posterior sampler needs to be checked for the θ parameters only.

5. Predictive Likelihood and Log Predictive Score

In the previous sections we have discussed algorithms for computing the predictive distributions of the observed variables in DSGE and DSGE-VAR models as well as the means and covariances of their marginal predictive distributions. A measure of the height of the joint or the marginal predictive density is typically needed when comparing or evaluating density forecasts; see, e.g., Geweke and Amisano (2010). For instance, when comparing density forecasts of a set of models the height of the predictive density at the realized value of the observed variables—the predictive likelihood—is of particular interest.

The use of the predictive likelihood as a valid Bayesian approach to model selection has long been recognized. Box (1980), for example, has emphasized the complementary roles in the model building process of the posterior and predictive distributions, where the former is used for diagnostic checking of the model, while the latter provides a general basis for robustness checks. Moreover, for models with improper priors the predictive likelihood can still be used for model selection provided that the sample being conditioned on is large enough to train the prior to a proper one; see, e.g., Gelfand and Dey (1994), Eklund and Karlsson (2007), and Strachan and van Dijk (2011).

A forecast comparison exercise is naturally cast as a decision problem within a Bayesian setting and therefore needs to be based on a particular preference ordering. Scoring rules can be used to compare the quality of probabilistic forecasts by giving a numerical value using the predictive distribution and an event or value that materializes. A scoring rule is said to be *proper* if a forecaster who maximizes the expected score provides its true subjective distribution; see Winkler and Murphy (1968). If the maximum is unique then the rule is said to be strictly proper.

A widely used scoring rule that was suggested by, e.g., Good (1952) is the log predictive score. Based on the joint predictive density function of $y_{t+1}, \ldots, y_{T+h}$, it can be expressed as

$$S_J(h, m) = \sum_{t=T}^{T+N_h-1} \log p(y_{t+1}, \ldots, y_{t+h}|\mathcal{Y}_t, m), \quad h = 1, \ldots, H, \tag{25}$$

where the J subscript refers to the joint predictive distribution, N_h is the number of time periods the h-step-ahead predictive density is evaluated, $\mathcal{Y}_t$ is the observed data of y_t until period t, and

m is an index for the model. If the scoring rule depends on the predictive density only through the realization of y over the prediction sample, then the scoring rule is said to be *local*. Under the assumption that only local scoring rules are considered, Bernardo (1979) showed that every proper scoring rule is equivalent to a positive constant times the log predictive score plus a real valued function that only depends on the realized data; see Gneiting and Raftery (2007) for a recent survey on scoring rules.

When evaluating the log score with the realized value of y over the prediction sample, the difference between the log score of model m and model k is equal to N_h times the average log predictive Bayes factor of these two models. Hence, a positive value indicates that, on average, model m is better at predicting the variables over the forecast horizon than model k. It is furthermore straightforward to show that the log predictive likelihood of model m is equal to the difference between the log marginal likelihood value when the historical data, $\mathcal{Y}_t$, and the realizations $y_{t+1}, \ldots, y_{t+h}$ are used and the log marginal likelihood value obtained when only the historical data are employed; see, e.g., Geweke (2005). For the realization $\mathcal{Y}_{T+N_1}$ and with $N_h = N_{h-1} - 1$ for $h > 1$, we can rewrite the log predictive score in (25) as

$$S_J(h,m) = \sum_{i=0}^{h-1} \Big[\log p\big(\mathcal{Y}_{T+N_1-i}, m\big) - \log p\big(\mathcal{Y}_{T+i}, m\big) \Big], \quad h = 1, \ldots, H. \tag{26}$$

This means that the log score of model m for one-step-ahead forecasts is equal to the difference between the log marginal likelihood for the full sample $\mathcal{Y}_{T+N_1}$ and the first historical sample $\mathcal{Y}_T$. Moreover, the calculation of the score for h-step-ahead forecasts based on the joint predictive likelihood requires exactly $2h$ marginal likelihood values, where the first h values are based on the samples $\mathcal{Y}_{T+N_1-i}$ and the last h values on $\mathcal{Y}_{T+i}$ for $i = 0, \ldots, h-1$.

A scale for determining how much better (or worse) model m is than model k has been suggested by, e.g., Kass and Raftery (1995) for Bayes factors. In the case of one-step-ahead predictions, this scale may also be applied to the log predictive score where values above 5 may be seen as very strong evidence in favor of model m. Provided that one of these models is assumed to be "true", the translation into posterior probabilities is straightforward, where the case of equal prior probabilities for models m and k and a value of the log score of 5 or more corresponds to a posterior probability of above 99 percent for model m.

It can also be seen that the log predictive likelihood in (25) can be rewritten as a sum of one-step-ahead log predictive likelihoods. Hence, in essence the log score $S_J(h,m)$ covers one-step-ahead forecasts only and is therefore not well suited for a comparison of h-step-ahead forecasts when $h > 1$. When comparing the density forecasts of the NAWM and alternative forecast models, CCW therefore focus on the predictive likelihood of the marginal h-step-ahead forecasts rather than the joint forecasts in (25). The log predictive score can now be expressed as

$$S_M(h,m) = \sum_{t=T}^{T+N_h-1} \log p\big(y_{t+h} | \mathcal{Y}_t, m\big), \quad h = 1, \ldots, H, \tag{27}$$

where the M subscript refers to the marginal predictive distribution.

The relationship between the marginal likelihood and the log predictive score in (27) holds when $h = 1$. For other forecast horizons it is claimed by both CCW (p. 114) and Adolfson et al. (2007, p. 324) that this connection breaks down and, hence, that the marginal likelihood cannot detect if some models perform well on certain forecast horizons while other models do better on other horizons. Furthermore, Adolfson et al. (2007, p. 325) remark that computing $S_M(h, m)$ for $h > 1$ is not an easy task since $p\big(y_{t+h}|\mathcal{Y}_t, m\big)$ does not have a closed form solution and that kernel density estimation from the predictive draws is not practical unless the dimension of y_{t+h} is small. For linear Gaussian models they therefore suggest using a normal approximation of the predictive likelihood based on the mean and the covariance of the marginal predictive distribution.

However, going back one step one realizes that CCW and Adolfson et al. are incorrect since

$$p(y_{t+h}|\mathcal{Y}_t, m) = \frac{p\big(y_{t+h}, \mathcal{Y}_t, m\big)}{p\big(\mathcal{Y}_t, m\big)}, \quad h = 1, \ldots, H. \tag{28}$$

The denominator is the marginal likelihood of model m when using the data $\mathcal{Y}_t$ and the numerator is likewise the marginal likelihood for this model when using the data $(y_{t+h}, \mathcal{Y}_t)$. Hence, the connection between the predictive likelihood and the marginal likelihood remains also for $h > 1$. The problem for calculating the log predictive score in (27) for $h > 1$ therefore concerns the question: it is possible to compute the marginal likelihood for the sample $(y_{t+h}, \mathcal{Y}_t)$?

Suppose we replace the realizations of y_{t+i}, $i = 1, \ldots, h-1$, in $\mathcal{Y}_{t+h}$ with missing observations and apply a valid method for dealing with incomplete-data when evaluating the likelihood function for fixed parameters of model m. This effectively means that we treat missing observations as a method for integrating out variables at certain points in time from the likelihood, and that the marginal likelihood of the model for $(y_{t+h}, \mathcal{Y}_t)$ can thereafter be computed via standard tools.[7] Such an approach can also be used to estimate the marginal likelihood for the data $(y^*_{t+h}, \mathcal{Y}_t)$, where y^*_{t+h} is a subset of the elements of y_{t+h}, as well as for, e.g., the data $(y^*_{t+1}, \ldots, y^*_{t+h}, \mathcal{Y}_t)$. In fact, we may replace data points with missing observations anywhere in the predictive sample $y_{t+1}, \ldots y_{t+h}$ when calculating the likelihood function.

In the case of linear state-space models with Gaussian shocks and measurement errors, the likelihood function can be calculated using a Kalman filter which allows for missing observations; see, e.g., Durbin and Koopman (2012, Chapter 4.10). Once we turn to non-linear, non-normal state-space models a missing observations consistent filter, such as the particle filter (sequential Monte Carlo), may instead be applied when computing the likelihood; see Giordani, Pitt, and Kohn (2011) for a survey on filtering in state-space models, or Durbin and Koopman (2012, Chapter 12) for an introduction to particle filtering.

[7] This idea is related to but also different from data augmentation and other such EM algorithm extensions. For these algorithms, the model is used to replace missing observations with model-based draws of the latent variables and then use complete-data methods to address the incomplete-data problem; see, e.g., Tanner and Wong (1987) and Rubin (1991).

When the joint predictive density for fixed parameters is Gaussian, marginalization can also be conducted directly via the predictive mean and the covariance matrix for given parameters by utilizing well-known properties of the normal distribution. In the case of linear Gaussian models this approach to marginalization is equivalent to the Kalman filter approach, where the latter approach provides a unifying framework and is as parsimonious as possible when dealing with potentially large matrices.

The predictive likelihood for a subset of the variables may now be estimated with, e.g., the harmonic mean estimator; see Gelfand and Dey (1994), the truncated normal version in Geweke (1999, 2005), or the extension to a truncated elliptical in Sims, Waggoner, and Zha (2008).[8] Such estimators ideally require two sets of posterior draws: $\theta_h^{(s)} \in p(\theta|y_{t+h}^*, \mathcal{Y}_t)$, $s = 1, \ldots, S_h$, and $\theta^{(s)} \in p(\theta|\mathcal{Y}_t)$, for $s = 1, \ldots, S$, where the model index m has been suppressed, and where θ denotes all parameters of the model. The marginal predictive likelihood can now be estimated as

$$\begin{aligned}\hat{p}_H(y_{t+h}^*|\mathcal{Y}_t) &= \left[\frac{1}{S_h}\sum_{s=1}^{S_h}\frac{f(\theta_h^{(s)})}{L(y_{t+h}^*|\mathcal{Y}_t,\theta_h^{(s)})p(\theta_h^{(s)}|\mathcal{Y}_t)}\right]^{-1} \\ &= \left[\frac{1}{S_h}\sum_{s=1}^{S_h}\frac{f(\theta_h^{(s)})}{L(y_{t+h}^*|\mathcal{Y}_t,\theta_h^{(s)})L(\mathcal{Y}_t,\theta_h^{(s)})p(\theta_h^{(s)})}\right]^{-1}\left[\frac{1}{S}\sum_{s=1}^{S}\frac{f(\theta^{(s)})}{L(\mathcal{Y}_t,\theta^{(s)})p(\theta^{(s)})}\right],\end{aligned} \tag{29}$$

where $L(y_{t+h}^*|\mathcal{Y}_t,\theta)$ denotes the conditional likelihood and $L(\mathcal{Y}_t,\theta)$ the likelihood. We have also used the fact that $p(\theta|\mathcal{Y}_t) = L(\mathcal{Y}_t,\theta)p(\theta)/p(\mathcal{Y}_t)$, and that the inverse of the marginal likelihood for the $\mathcal{Y}_t$ data is estimated with the second term on the right hand side in the second row of (29). The function $f(\theta)$ is a proper density and therefore integrates to unity.

To avoid having to generate posterior draws for each sample $(y_{t+h}^*, \mathcal{Y}_t)$, it is tempting to replace the draws $\theta_h^{(s)}$ in (29) with the draws $\theta^{(s)}$. If the dimension of y_{t+h}^* is small, this approximation is likely to work well in practise, but unlike the estimator in (29) the resulting predictive likelihood estimator is *not* consistent.

If we insist on using only one set of parameter draws for all forecast horizons when computing the predictive likelihood, we may instead use an importance sampling (IS) estimator; see, e.g., Geweke (2005). With $\theta^{(i)}$ being draws from the importance density $g(\theta)$, a general expression of the IS estimator is

$$\hat{p}_{IS}(y_{t+h}^*|\mathcal{Y}_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{L(y_{t+h}^*|\mathcal{Y}_t,\theta^{(i)})p(\theta^{(i)}|\mathcal{Y}_t)}{g(\theta^{(i)})}. \tag{30}$$

Letting $g(\theta) = p(\theta|\mathcal{Y}_t)$ such that $\theta^{(i)} = \theta^{(s)}$ with $N = S$, the estimator of the predictive likelihood in (30) is simply the average over the S posterior draws $\theta^{(s)}$ of the conditional likelihood, i.e. standard Monte Carlo integration based on the conditional likelihood.

[8] Other methods, such as bridge sampling or cross-entropy with importance sampling may also be considered; see Meng and Wong (1996), Frühwirth-Schnatter (2004), and Chan and Eisenstat (2012).

Under certain conditions (Tierney, 1994), the right hand side of (30) converges almost surely to the expected value of $p(y^*_{t+h}|\mathcal{Y}_t, \theta)$ with respect to $p(\theta|\mathcal{Y}_t)$, i.e. to the predictive likelihood $p(y^*_{t+h}|\mathcal{Y}_t)$.[9] Hence, equipped with the posterior draws $\theta^{(s)}$ and the conditional likelihood, $L(y^*_{t+h}|\mathcal{Y}_t, \theta)$, the predictive likelihood can be consistently estimated directly, without having to compute it from two marginal likelihoods, and without having to sample from the distribution of the parameters conditional on $(y^*_{t+h}, \mathcal{Y}_t)$ for each $h = 1, \ldots, H$.

A further important property of the IS estimator is that it is unbiased (see Chan and Eisenstat, 2012, Proposition 1), while the harmonic mean estimator is not. Furthermore, the latter estimator is sensitive to the choice of $f(\theta)$ and can be difficult to pin down numerically when the dimension of θ is large, while the IS estimator based on the posterior $p(\theta|\mathcal{Y}_t)$ should be less hampered by this. In the case of DSGE models, which are typically tightly parameterized, numerical issues with the harmonic mean should not be a major concern, but for DSGE-VARs and BVAR models the computations need to take all the VAR parameters into account and are therefore likely to be important. For example, in our empirical illustration the DSGE-VAR models have 18 endogenous variables, a constant, varying lag orders plus 45 DSGE model parameters, resulting in $(324p + 234)$ parameters. Furthermore a large BVAR model having the same 18 variables, a constant, and four lags yields a total of 1485 parameters.

The IS estimator is expected to work well in practise when the draws from the importance density cover well enough the parameter region where the conditional likelihood is large. This is typically the case when computing the marginal predictive likelihood with $g(\theta) = p(\theta|\mathcal{Y}_t)$ for shorter forecast horizons or a low dimensional y^*_{t+h} vector, but it is questionable when dealing with the joint predictive likelihood as h becomes large.[10] For such situations it may be useful to consider cross-entropy methods for selecting the importance density optimally, as in Chan and Eisenstat (2012).

For linear Gaussian models the log of the conditional likelihood for y^*_{t+h} is given by

$$\begin{aligned}\log L\big(y^*_{t+h}\big|\mathcal{Y}_t, \theta\big) = &-\frac{n^*}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_{y^*,t+h|t}\right| \\ &-\frac{1}{2}\left(y^*_{t+h} - y^*_{t+h|t}\right)' \Sigma^{-1}_{y^*,t+h|t}\left(y^*_{t+h} - y^*_{t+h|t}\right),\end{aligned} \tag{31}$$

where n^* is the dimension of y^*_{t+h}, $y^*_t = K'y_t$ and $\Sigma_{y^*,t+h|t} = K'\Sigma_{y,t+h|t}K$, with K being an $n \times n^*$ known selection matrix. We here assume that the K matrix is constant across time, but we can also add a time subscript to it and to n^*. For DSGE models the Kalman filter provides

[9] Importance sampling is based on i.i.d. draws from the importance density; see, for instance, Geweke (2005, Chapter 4.2.2) for further details. In the case of DSGE and DSGE-VAR models, the posterior draws are typically obtained via Markov chain Monte Carlo, such as the random walk Metropolis sampler, and are therefore not independent. However, under certain conditions (Tierney, 1994) the estimator in (30) is consistent also when the draws from $g(\theta) = p(\theta|\mathcal{Y}_t)$ are not independent and the same conditions can be used to verify that the harmonic mean estimator in (29) is consistent. In strict terms, the estimator in (30) is not an IS estimator when the iid assumption is violated, but we shall nevertheless use this term also when the draws from the posterior are dependent.

[10] For sufficiently large h the situation resembles the case when the marginal likelihood is computed by averaging the likelihood over the prior draws. Such an estimator typically gives a poor estimate of the marginal likelihood.

us with

$$
\begin{aligned}
y_{t+h|t} &= \mu + H' F^h \xi_{t|t}, \\
\Sigma_{y,t+h|t} &= H' P_{t+h|t} H + R, \\
P_{t+h|t} &= F P_{t+h-1|t} F' + BB', \quad h = 1, \ldots, H,
\end{aligned}
$$

where $\xi_{t|t}$ is the filter estimate of the state variables, and $P_{t|t}$ the corresponding filter estimate of the state variable covariance matrix based on the data $\mathcal{Y}_t$. The matrices (μ, H, R, F, B) are all evaluated for a given value of θ.[11]

Turning to the DSGE-VARs, we can directly make use of equations (20) and (23) to determine $y_{t+h|t}$ and $\Sigma_{y,t+h|t}$. That is,

$$
\begin{aligned}
y_{t+h|t} &= J_p' \bar{x}_{t+h} + J_p' \Psi^h Y_t, \\
\Sigma_{y,t+h|t} &= J_p' \bar{\Sigma}_Y^{(h)} J_p, \quad h = 1, \ldots, H.
\end{aligned}
$$

Adolfson et al. (2007) and CCW approximate the marginal predictive likelihood with a normal density whose mean and covariance for a DSGE model are given by the population expressions discussed in Section 2. In our empirical illustration we shall compare this approximation to the IS estimator, which we treat as the benchmark estimate of the predictive likelihood. This allows us to examine if these estimators differ substantially in practise and, furthermore, makes it possible to directly compare our results with those in CCW.

6. Estimating DSGE-VAR Models Using the NAWM as a Prior

6.1. The New Area-Wide Model of the Euro Area

The NAWM is a micro-founded open-economy model of the euro area designed for use in the ECB/Eurosystem staff projections and for policy analysis.[12] The development of the model has been guided by a principal consideration, namely to provide a comprehensive set of core projection variables, including a number of foreign variables, which, in the form of exogenous assumptions, play an important role in the projections. As a consequence, the scale of the NAWM—compared with a typical DSGE model—is rather large, and it is estimated on 18 macroeconomic time series.

The NAWM features four classes of economic agents: households, firms, a fiscal authority and a monetary authority. Households make optimal choices regarding their purchases of consumption and investment goods, they supply differentiated labor services in monopolistically competitive markets, they set wages as a mark-up over the marginal rate of substitution between consumption and leisure, and they trade in domestic and foreign bonds.

[11] The Kalman filter for missing observations can also be used when we are interested in the joint predictive likelihood for subsets of variables across a sequence of future dates, but we do not provide the details here.

[12] See Christoffel, Coenen, and Warne (2008) for a detailed description of the NAWM's structure, while a shorter presentation of some of its key equations is provided in CCW.

As regards firms, the NAWM distinguishes between domestic producers of tradeable differentiated intermediate goods and domestic producers of three types of non-tradeable final goods: a private consumption good, a private investment good, and a public consumption good. The intermediate-good firms use labor and capital as inputs to produce their differentiated goods, which are sold in monopolistically competitive markets domestically and abroad. Accordingly, they set different prices for domestic and foreign markets as a mark-up over their marginal costs. The final-good firms combine domestic and foreign intermediate goods in different proportions, acting as price takers in fully competitive markets. The foreign intermediate goods are imported from producers abroad, who set their prices in euro, allowing for an incomplete exchange-rate pass-through. A foreign retail firm in turn combines the exported domestic intermediate goods, where aggregate export demand depends on total foreign demand.

Both households and firms face nominal and real frictions, which have been identified as important in generating empirically plausible dynamics. Real frictions are introduced via external habit formation in consumption and through generalized adjustment costs in investment, imports and exports. Nominal frictions arise from staggered price and wage-setting à la Calvo (1983), along with (partial) dynamic indexation of price and wage contracts. In addition, there exist financial frictions in the form of domestic and external risk premia.

The fiscal authority purchases the public consumption good, issues domestic bonds, and levies different types of distortionary taxes. Nevertheless, Ricardian equivalence holds because of the simplifying assumption that the fiscal authority's budget is balanced each period by means of lump-sum taxes. The monetary authority sets the short-term nominal interest rate according to a Taylor-type interest-rate rule, with the objective of stabilizing inflation in line with the ECB's definition of price stability.

The NAWM is closed by a rest-of-the-world block, which is represented by a structural VAR (SVAR) model determining the five foreign variables: foreign demand, foreign prices, the foreign interest rate, foreign competitors' export prices, and the price of oil. The SVAR model does not feature spill-overs from the euro area, in line with the treatment of the foreign variables as exogenous assumptions in the projections.

6.2. Estimation Results

In order to estimate the NAWM, Christoffel, Coenen, and Warne (2008) use time series for 18 macroeconomic variables that have a high degree of importance in the ECB/Eurosystem staff projection exercises. All time series are taken from Update 7 of the AWM database (Fagan, Henry, and Mestre, 2005) except for the time series of extra-euro area trade data (see Dieppe and Warmedinger, 2007, for details on their construction). The estimation sample is given by the period 1985Q1 until 2006Q4, with 1980Q2-1984Q4 serving as training sample. The parameters of the SVAR for the five foreign variables are estimated separately and are kept fixed throughout the estimation of the NAWM. Furthermore, government consumption is specified as an autoregressive process with fixed estimated parameters.

The time series are displayed in Figure 1, where real GDP, private consumption, total investment, exports, imports, the GDP deflator, the consumption deflator, the import deflator, nominal wages, foreign demand, and foreign prices are all expressed as 100 times the first difference of their logarithm. All other variables are expressed in logarithms except for the short-term nominal domestic and foreign interest rates. A number of further transformations are made to ensure that variable measurement is consistent with the properties of the NAWM's balanced-growth path and in line with the underlying assumption that all relative prices are stationary.[13]

First, the sample growth rates of extra-euro area exports and imports as well as foreign demand are matched with the sample growth rate of real GDP. Second, for the logarithm of government consumption we remove a linear trend consistent with the NAWM's steady-state growth rate of 2.0 percent per annum. This trend is assumed to have two components: labor productivity growth of 1.2 percent and labor force growth of 0.8 percent. Third, we take the logarithm of employment and remove a linear trend consistent with a steady-state labor force growth rate of 0.8 percent. Fourth, we construct a measure of the real effective exchange rate from the nominal effective exchange rate, the domestic GDP deflator and foreign prices (defined as a weighted average of foreign GDP deflators) and then remove the sample mean. Finally, competitors' export prices and oil prices (both expressed in the currency basket underlying the construction of the nominal effective exchange rate) are deflated with foreign prices before unrestricted linear trends are removed from the variables.

Christoffel et al. (2008) adopt the empirical approach outlined in Smets and Wouters (2003) and An and Schorfheide (2007) and estimate the NAWM with Bayesian methods. The DSGE-VAR models have been estimated over the same sample as the NAWM with the random walk Metropolis algorithm subject to a Gaussian proposal density. We use the parameter transformations $\phi = g(\theta)$ discussed in, for instance, Warne (2012, Section 6) such that the domain of ϕ is the real line. It is well known that this affects the prior distribution through the Jacobian of the transformation, but not the likelihood function. As in Christoffel et al. (2008), we employ the inverse Hessian from posterior mode estimation of the NAWM as the covariance matrix for the proposal density, but fine-tune it via a positive constant for each pair (λ, p) such that the acceptance rate is about 25-30 percent. Each posterior draw of ϕ is thereafter transformed to θ via $\theta = g^{-1}(\phi)$. From a total of 550,000 posterior draws for each DSGE-VAR model, where the first 50,000 are used as burn-in sample, the marginal likelihood, $p(\mathcal{Y}_T|Y_1, \lambda, p)$, is calculated with the harmonic mean estimator based on the truncated normal density; see Geweke (1999, 2005). All calculations in this paper have been performed with the help of YADA, a Matlab program for Bayesian estimation and evaluation of DSGE and DSGE-VAR models; see Warne (2012) for details.

[13] When deriving the log-linearized equations (1), the NAWM is first cast into stationary form. To this end, all real variables of the NAWM are measured in per-capita terms and scaled by trend labor productivity. The latter variable is assumed to follow a random walk with stochastic drift and defines the model's balanced growth path. Similarly, all nominal variables are normalized with the price of the consumption good. The log-linearization of the model is then performed around its deterministic steady state; see Christoffel et al. (2008) for details.

In our empirical analysis we consider two approaches for selecting a DSGE-VAR model. The first chooses the model with the largest marginal likelihood over all pairs (λ, p), while the second picks p such that the marginal likelihood of the DSGE-VAR(∞) model is close to the marginal likelihood of the DSGE model and then selects λ optimally conditional on p. The estimated marginal likelihood values for a range of DSGE-VAR models, along with the NAWM, are displayed in Figure 2. The lag orders that we consider below range from one to four.

It is noteworthy that for all lag orders $p \geq 2$, the log marginal likelihood at $\lambda = \infty$ is considerably higher than the log marginal likelihood for the NAWM, with increases of 50 units or greater. Compared with Del Negro et al. (2007), who use a model with seven observed variables, the increase in marginal likelihood for the VAR approximation is very large. It should be kept in mind, however, that Del Negro et al. (2007) add an error correction term to their VAR model with the cointegration relations implied by their DSGE model. Consistent with the results shown in Adolfson et al. (2008, Table 2), the error correction form of the DSGE-VAR is likely to have a lower marginal likelihood than the pure DSGE-VAR when the data is not well represented by the cointegration relations.[14] Since the purpose of the current paper is to study the forecasting performance of DSGE-VARs we do not see any need for adding cointegration relations to the model at this stage.

For the DSGE-VAR model with one lag only, the log marginal likelihood is lower at $\lambda = \infty$ than for the NAWM, with a drop of about 11 units. Moreover, the marginal likelihood for the optimal one-lag model ($\hat{\lambda} = 1.25$) is lower than, but close to, the marginal likelihoods for the DSGE-VAR(∞) models with a higher lag order.

From Figure 2 it can be seen that the posterior mode estimates of λ are positively related to the selected lag order. Specifically, when we condition on two lags, the optimal value is $\hat{\lambda} = 2.5$, while for three lags we obtain $\hat{\lambda} = 4$, and when $p = 4$ we get $\hat{\lambda} = 6$. One explanation for the estimated relation between the optimal λ and the lag order is that the lower bound for the range of eligible values, $\lambda_L = (n(p+1)+1)/T$, is increasing in the lag order. Another explanation may be that when the lag order increases, the VAR model needs more dummy observations to better account for the loss of degrees of freedom from the observed sample. In the next section, we shall examine the usefulness of DSGE-VAR models with two lags (first model selection approach) and four lags (second model selection approach) in a forecast comparison exercise.[15]

7. Comparing Forecast Accuracy

The forecast performance of the DSGE-VAR models along with the NAWM, a large BVAR model, and the random walk model is assessed below using a similar procedure as in CCW. The random walk model concerns the transformations y_t for the DSGE model, i.e. a random

[14] In the case of the NAWM, the cointegration relations would be directly related to the balanced-growth property via the common unit-root technology trend.

[15] Although the model $(\lambda = \infty, p = 1)$ obtains a marginal likelihood which is closer to the marginal likelihood of the NAWM than any of the other $(\lambda = \infty, p)$ models, we have nevertheless selected $p = 4$ for this case.

walk for a mixture of first difference and levels variables, and is based on a standard diffuse prior on the covariance matrix of the innovations to the observed variables of the NAWM, while the BVAR model is estimated using the methodology in Bańbura et al. (2010) for the same variables. Appendix A provides details on the posterior properties of the random walk model, while Appendix B gives more information about the BVAR.[16] The parameters of the DSGE and DSGE-VAR models are estimated up to period T when the predictive distribution of periods $T+1,\ldots,T+h$ is to be computed and when T is the fourth quarter of the year. When T corresponds to a quarter $i=1,2,3$, the models are estimated using data until $T-i$, i.e. the DSGE and DSGE-VAR models are reestimated annually. The parameters of the random walk and BVAR models are reestimated every period.

The first pseudo out-of-sample forecasts are computed for 1999Q1—the first quarter after the introduction of the euro—while the final period is 2006Q4. The maximum forecast horizon is eight quarters, yielding 32 quarters with one-step-ahead forecasts and 25 quarters with eight-step-ahead forecasts. We shall only consider forecasts of quarterly growth rates for the variables in first differences, while CCW also study forecasts of annual growth rates for such variables. The forecast comparisons concern both point and density forecasts, where the point forecast comparison covers both univariate and multivariate MSE-based measures, while the density forecasts are summarized with the log predictive score. Compared with CCW, the density forecasts are extended to include the IS estimator, discussed in Section 5, as well as the normal approximation, and to a decomposition of the log predictive likelihood for the normal approximation into forecast uncertainty and forecast error driven terms. The predictive mean and covariance matrix used by the normal approximation are given by equations (5) and (6) for a DSGE model, by equations (21) and (22) for DSGE-VAR models, while the predictive moments for the random walk and BVAR models are shown in Appendix A and B, respectively.

7.1. Point Forecasts

Figure 3 displays the root mean squared forecast errors (RMSEs) for the 12 variables that are endogenously determined in the NAWM.[17] The point forecasts from the NAWM, the two DSGE-VARs, and the BVAR model are computed as the estimated mean of the predictive distribution based on 500 draws of the model parameters from the posterior distribution, where each draw is separated by 1000 draws from the next.[18] To facilitate not only the comparisons with the multivariate point forecasts but also univariate comparisons across variables, the RMSEs in

[16] While CCW consider additional reduced-form models in their forecast comparison exercise, these models were outperformed by the large BVAR and have therefore been excluded from the analyses below.

[17] The remaining six variables are essentially exogenous for the NAWM and are given by the five variables in the foreign SVAR block and real government consumption. Since the parameters that determine the behavior of these variables have been calibrated using data until 2006Q4, the comparisons between the NAWM and the other four models would not be fair for these variables and they are therefore excluded from the analysis; see footnote 4 in CCW for further discussion on this issue.

[18] The mean of the predictive distribution of the DSGE model is given in equation (5), while equation (21) gives the mean for the DSGE-VAR models.

Figure 3 have been scaled with the estimated standard deviations of the variables over the period 1995Q1–2006Q4.

As pointed out by CCW, the univariate RMSE analysis shows that the NAWM fares well when compared with the alternatives. In particular, it forecasts real GDP growth, real export and import growth, import price deflator, employment, the short-term nominal interest rate, and the real effective exchange rate well when compared to the competitors. The variables that the NAWM forecasts badly are nominal wages in particular, but also the GDP deflator, the private consumption deflator and private consumption. Given these poor performance variables it is firstly of particular interest to examine whether the DSGE-VAR models fare better, while at the same time not loosing ground on the variables that the NAWM forecasts comparatively well. Second, we would like to know whether there is any substantial performance difference between the DSGE-VARs.

Turning first to the second issue, it is notable in Figure 3 how close the RMSEs of the two DSGE-VAR models are. For the three variables in levels or log-levels (employment, short-term nominal interest rate, real effective exchange rate) the RMSEs are very close. Although it is possible to detect differences for some of the other variables they are hardly spectacular and, hence, overall the two DSGE-VARs perform similarly. Regarding the first issue, the improved forecast performance when using these models relative to the NAWM is substantial for private consumption, nominal wages, the private consumption deflator, and the GDP deflator. The improvements for nominal wages, in particular, appear to be fairly constant over the forecast horizon.

For the variables where the NAWM performs well from a RMSE perspective, the DSGE-VAR models are often competitive and, in the case of real GDP, they tend to perform somewhat better except at the one-quarter horizon. At the shorter horizons, the NAWM has lower RMSEs for imports and at the longer horizons for total investment, employment and the nominal interest rate. Overall, it therefore seems that the DSGE-VARs improve on the forecast performance relative to the NAWM.

Regarding the point forecasts from the BVAR model it is noteworthy that the RMSEs are generally higher than for the DSGE-VAR models for real GDP, private consumption, total investments, exports and imports. With the exception of exports, the RMSEs for these variables increase with the forecast horizon. The random walk model likewise displays problems forecasting these real variables. For the GDP deflator, we find that the BVAR is competitive at the shorter horizons and that its relative performance weakens with the forecast horizon. In fact, for the remaining variables in Figure 3, the BVAR weakly dominates the other models only for the employment forecasts. The random walk model, on the other hand, is competitive for nominal wages, the short-term nominal interest rate, and the real effective exchange rate. Overall, however, the BVAR and the random walk models tend to be outperformed by, in particular, the DSGE-VAR models.

Table 1 shows the (unscaled) mean forecast errors and, in general, the errors are negative (or close to zero) for private consumption and nominal wages and positive for the private consumption and the GDP deflator for all models except the BVAR. Moreover, it can also be inferred that the mean errors for real wages are negative. For the DSGE-VAR models, negative forecast errors are noticeable at the longer horizons for private consumption but they are not as large in absolute terms as for the NAWM. Negative forecast errors also remain for nominal and real wages, but they are smaller for all horizons and measures. For the price deflators we find that the mean forecast errors are positive and substantially larger for the NAWM than for the DSGE-VARs.

Table 2 provides the percentage share of squared mean errors to the mean squared errors for the same four variables. In particular, the tendency of the NAWM to overpredict private consumption and real wages was discussed in some detail in CCW. It is interesting to note that systematic mean errors explain a large share of the mean squared errors for the NAWM, suggesting that the calibrated steady state of the NAWM along with its balanced-growth assumption is a significant reason for its poor forecast performance for private consumption, real and nominal wages, the GDP deflator, and the private consumption deflator relative to the DSGE-VAR models. It may be recalled that the balanced-growth assumption only affects the prior distribution of the DSGE-VAR parameters and that the weight on the prior relative to the weight on the historical data decreases with the λ hyperparameter, thereby diminishing the influence of this assumption and potential source of misspecification on the forecasts of these models.

Figure 4 displays the individual mean forecast paths for the five models along with the data for quarterly private consumption and nominal wages. It is noteworthy that the forecast paths of the DSGE-VAR models for private consumption are more centered on the actual values than for the NAWM, while an overall downward shift of the paths for the DSGE-VARs relative to the NAWM can be detected for nominal wages. Hence, the lower mean errors for the DSGE-VAR models appear systematically over the forecast sample.

Turning to multivariate measures of point forecast accuracy, such measures are usually based on the (scaled) h-step-ahead MSE matrix

$$\Sigma_M(h) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \tilde{\epsilon}_{t+h|t} \tilde{\epsilon}'_{t+h|t}, \tag{32}$$

where $\tilde{\epsilon}_{t+h|t} = M^{-1/2}\epsilon_{t+h|t}$, and $\epsilon_{t+h|t}$ is the h-step-ahead forecast error for the forecast produced at t. The scaling matrix M is positive definite.

The trace and the log determinant are two statistics that are often used in practise for comparing multivariate point forecast accuracy. The choice of scaling matrix for the forecast errors has a direct impact on the ranking of models when using the trace statistic, while the ranking is invariant to the choice of M for the log determinant statistic. In addition, the trace and log determinant statistics are functions of the eigenvalues of the MSE-matrix, where the largest eigenvalues gives the MSEs of the least predictable linear combinations of the variables while the

smallest eigenvalues are the MSEs of the most predictable linear combinations. Since the trace is equal to the sum of the eigenvalues it follows that this statistic tends to be dominated by the largest eigenvalues, while the determinant is the product of the eigenvalues and is therefore also influenced by the smallest eigenvalues and may even be dominated by them.

As in CCW, the MSE statistics are calculated for three cases. First, we consider a *large selection* with all the 12 variables displayed in Figure 3. The second case covers a *medium selection* with the seven variables studied in Smets and Wouters (2003). That is, real GDP, private consumption, total investment, the GDP deflator, employment, nominal wages, and the short-term nominal interest rate. Finally, we examine a case which may be regarded as being the minimum set of variables relevant to monetary policy. This *small selection* is given by real GDP, the GDP deflator, and the short-term nominal interest rate.

The trace statistics are displayed in Figure 5. The scaling matrix M is diagonal with the estimated variances of the variables over the period 1995Q1–2006Q4 and the scaling is thus the same as for the individual RMSEs in Figure 3. This means that the trace is the sum of the squared scaled RMSEs. Given the univariate analysis it is therefore not surprising that the DSGE-VAR models perform very well for all three selections and over all forecast horizons, while the NAWM is mainly competitive at the longer forecast horizons. Furthermore, it may be noted that the BVAR and the random walk models obtain the largest trace values for longer forecast horizons, while the BVAR has low values for the shorter horizons.

The log determinant statistics are displayed in Figure 6. The BVAR model tends to outperform the other models for the large and medium selection cases, while the NAWM seems to rank at the top for the small selection. The DSGE-VAR models have values close to the BVAR for the larger systems, especially for horizons at and beyond one year. For the smallest system, the DSGE-VAR models also follow the NAWM closely.

To summarize the point forecast results, we find that by relaxing the strict cross-equation restriction of the NAWM through a DSGE-VAR model the forecast performance can be improved, especially for the variables where the NAWM obtains large and systematic forecast errors. Moreover, the evidence is broadly unaffected by the method for selecting the λ hyperparameter and the lag order. For the three multivariate selections, the DSGE-VAR with the overall highest marginal likelihood ($\lambda = 2.5, p = 2$) performs better at the shorter horizons compared with the optimal DSGE-VAR model with four lags, while the situation is reversed at the longer horizons.

The overall ranking of the five models depends on the variables of interest, the forecast horizon, and the choice of MSE statistic, in particular when the univariate RMSE statistic is applied. When using the multivariate statistics, the DSGE-VARs typically outperform the other models for the trace statistic, while the BVAR and the NAWM are competitive for the log determinant statistic. As pointed out by Herbst and Schorfheide (2012), the prediction error loss function underlying the MSE comparisons need not be the relevant loss for policy makers. For example, the use of uncertainty bands in the inflation reports of several central banks (Bank

of Canada, Bank of England, Norges Bank, Sveriges Riksbank, etc) has become instrumental in communicating with the public. In the next section we shall therefore turn our attention to density forecasts and thereby incorporate other features of the predictive distribution, such as prediction uncertainty, into the forecast comparisons.

7.2. Density Forecasts

The log predictive scores based on the IS estimator are reported in Table 3 and depicted in Figure 7 for all variable selections and models. For the NAWM and the two DSGE-VAR models we have used 10,000 posterior draws among the available 500,000 post burn-in draws for each model and time period when calculating the log predictive likelihood. These draws have been selected as draw number $1, 51, 101, \ldots, 499951$ to combine low computational costs with a low correlation between the draws and a sufficiently high estimation accuracy. This procedure yields estimates of the log predictive likelihood that are accurate up to and including the first decimal relative to using all post burn-in draws for these models.[19]

In the case of the random walk model, the predictive likelihood for a selection of variables can be analytically determined and its density is multivariate t; see Appendix A. Direct sampling is possible for the BVAR model through its normal-inverted Wishart posterior and we have used 50,000 draws from its posterior distribution when computing the predictive likelihood with the IS estimator; see Appendix B for details.

When comparing the NAWM with the two DSGE-VAR models, it is noteworthy that the DSGE-VAR model with two lags generally obtains higher log scores for all horizons and variable selections, with values for the four-lag model being close to them, while the NAWM gets lower values. At the longer horizons, the NAWM obtains values that are near the DSGE-VAR values and, in the case of the small selection even slightly higher than the values for the DSGE-VAR models. Hence, it seems that taking misspecification of the NAWM into account through DSGE-VAR models improves forecasting performance, especially at the shorter horizons.

It is also worth pointing out that the random walk model is competitive with the NAWM and the DSGE-VAR models for the one-step-ahead forecasts, especially for the small selection. As the forecast horizon increases, however, the random walk model's performance worsens in comparison with these alternatives.

Compared with the BVAR model, however, the NAWM and the two DSGE-VARs are outperformed for all selections and forecast horizons, except for the small selection at the longer horizons. For example, the difference between the log score of the BVAR and the DSGE-VAR with two lags is at least 27 units for the large selection and 40 units for the medium selection.

[19] The numerical standard errors based on the Newey and West (1987) estimator are small for the IS estimator of the log predictive likelihood. In the case of the DSGE model, the shortest historical sample, and the large selection the standard errors are less than 0.04, for the medium selection they are less than 0.03, and for the small selection less than 0.015. As the length of the historical sample increases, the numerical standard errors decrease. For the DSGE-VAR models, the numerical standard errors are smaller than those for the DSGE model.

Hence, if the log predictive score is employed as a model-selection device it would prefer the BVAR to the other models for all selections and forecast horizons.

To address the issue of how well the normal approximation works for the linear Gaussian models, the log predictive scores for this estimator are displayed in Figure 8. The most prominent feature is how similar these graphs are when compared to those in Figure 7. In fact, the IS estimator and the normal approximation suggest the same ranking of the models for each selection and forecast horizon except for the DSGE-VAR models with the small selection and the eight-step-ahead forecasts. For this particular case, however, the difference in log predictive scores is so small[20] that the models may be viewed as equally good (or bad).

Zooming into the individual density forecasts, the log predictive likelihoods of the NAWM for the large selection are displayed for the IS estimator (solid red line) and the normal approximation (dashed blue line) in Figure 9. It is striking how closely the normal approximation follow the IS estimator. Over the full prediction horizon, the normal approximation underestimates the log predictive score by between 1 and 2.5 units; see also Table 4. It may be noted that this is not a peculiarity of the large selection but is also found for the medium and small selections, although the values tend to be somewhat smaller.

Turning to the log predictive likelihoods of the large selection for the DSGE-VAR with two lags in Figure 10, it is again notable that the differences between the IS estimator and the normal approximation are quite small, where the total difference between them is now in the range of 2 to 4 units. For the medium selection the errors are somewhat larger at the longer horizons, while for the small selection the errors are lower (see Table 4). A similar picture is obtained for the DSGE-VAR model with four lags, but here the approximation errors are nearly half as big as those in the two-lag model and more in line with the small approximation errors for the DSGE model. Still, the normal approximation continues to underestimate the value given by the IS estimator also for the DSGE-VAR models.

The log predictive score based on the IS estimator of the NAWM (dash-dotted grey line) is also shown in Figure 10. For the shorter forecast horizons it is noteworthy that the overall difference in the log predictive score between the NAWM and the DSGE-VAR with two lags is fairly evenly distributed over the sample, supporting the view that the overall difference in the score is a systematic feature rather than the result of a few data points. Moreover, the log predictive likelihoods appear to be positively correlated. As the forecast horizon increases, the two models obtain similar log predictive likelihood values from late 2001 until early 2004, while the NAWM is outperformed by the DSGE-VAR over the earlier and later quarters in the prediction sample.

Turning to the differences between the log score values of the BVAR based on the IS estimator and the normal approximation in Table 4 we find that the latter underestimates the values by

[20] The difference between the log predictive score of the DSGE-VAR with 2 lags and with 4 lags is 0.6 units for the IS estimator and -0.2 units for the normal approximation when evaluating eight-step-ahead forecasts in the 3 variable case.

between 0.5 and 3 units for the large selection of variables, while the errors are smaller for the medium and small selections. In fact, the normal approximation sometimes overestimates the log predictive score, but the absolute size of the error remains quite small. Furthermore, in Figure 11 it can be seen that the BVAR outperforms the NAWM in most quarters over the forecast sample, except a few cases where the BVAR obtains very low values, such as in 2003Q1.

For the random walk model, the normal approximation is somewhat less accurate than for the other models and underestimates the log score by about 6 to 10 units in the case of the large selection. As can be seen from Table 4 the errors become smaller as the dimension of the system decreases from 12 to seven and to three variable, where they are comparable to the errors of the NAWM. At the same time, the approximation errors are quite evenly spread out over the forecast sample, resulting in small errors for the individual predictive likelihood values.

Since the normal approximation provides a good approximation of the actual predictive likelihood for the five models and the three variable selections, we can utilize the analytical form of the normal density to assess if the ranking of models is driven by forecast uncertainty or by forecast errors.[21] The former is given by the expression for the log determinant term and the latter by the quadratic standardized forecast error term based on the population moments of the predictive distribution. That is,

$$D_{t+h|t}(m) = -\frac{\log\left|C\left[y^*_{t+h}|\mathcal{Y}_t, m\right]\right|}{2},$$
$$Q_{t+h|t}(m) = -\frac{\epsilon^*_{t+h|t}(m)'C\left[y^*_{t+h}|\mathcal{Y}_t, m\right]^{-1}\epsilon^*_{t+h|t}(m)}{2},$$

where $\epsilon^*_{t+h|t}(m) = y^*_{t+h} - E[y^*_{t+h}|\mathcal{Y}_t, m]$, m is the model index, $t = T, T+1, \ldots, T+N_h-1$, and $h = 1, \ldots, H$.

The $D_{t+h|t}(m)$ term is depicted in Figure 13 for the five models, eight forecast periods, and each relevant period in the prediction sample for the large selection of variables. Similarly, the $Q_{t+h|t}(m)$ term is displayed in Figure 14. Turning first to the forecast uncertainty term it can be seen that for all models and forecast horizons it is weakly upward sloping over the forecast sample and that the slope is roughly equal across the five models. This indicates that overall forecast uncertainty is slowly decreasing as data are added to the information set. The values for the BVAR model is roughly 5 log units higher in each period than for the second group of models, given by the NAWM and the two DSGE-VARs. The random walk model has the lowest values for all forecast horizons and the difference relative to the second group of models is increasing with the forecast horizon. This latter property is due to the forecast error covariance

[21] It is not possible to perform an exact decomposition of the predictive likelihood based on the IS estimator. With a normal conditional likelihood, this estimator is equal to the *average over the posterior parameter draws* of a constant times (i) the square root of the inverse determinant of the forecast error covariance matrix conditional on the parameters (ii) times the exponent of minus the quadratic standardized forecast errors conditional on the parameters divided by two. Hence, the predictive likelihood is here computed as the log of an average of products and can therefore not be decomposed in a manner that separates the effect of the forecast errors (in (ii)) from the effect of the second moments (in (i)) on the predictive likelihood.

matrix being given by the forecast horizon times a fixed matrix for the random walk model; see equation (A.10) in Appendix A.

Since the log determinant in $D_{t+h|t}(m)$ is equal to the sum of the log of the eigenvalues of the population forecast error covariance matrix, the value of the log determinant term is greater the smaller the eigenvalues of this matrix are. The eigenvalues in turn are small for linear combinations of the variables that, according to the model, are highly predictable.[22] The plots in Figure 13 therefore show that the BVAR model has better predictability in terms of the second moments than the other models, the DSGE-VARs with two and four lags and the NAWM follow as second, third, and fourth, while the random walk comes in last.

Turning to the quadratic standardized forecast error term in Figure 14, it can be seen that the time variation of the log predictive likelihood is due to these forecast errors. This is not surprising since the population covariance matrix changes only smoothly over time while the forecast errors are more volatile. Moreover, the ranking of the models is to some extent reversed, particularly with the BVAR having much larger forecast errors than the other models over the first half of the forecast sample. With the exception of the random walk model, this is broadly consistent with the findings for the point forecasts in Figures 3 and 5.[23] The reversal in rankings for the $Q_{t+h|t}(m)$ term can also be understood from the behavior of second moments, where a given squared forecast error yields a larger value for this term the smaller the uncertainty linked to the forecast is. Nevertheless, when compared with the $D_{t+h|t}(m)$ term the differences between the models are generally smaller for the $Q_{t+h|t}(m)$ term. This suggests that the model ranking based on the log predictive score is primarily determined by the second moment of the predictive distribution in this illustration.

Although these analyses have focused on the large variable selection, the overall findings are also valid for the medium and small selections. With the log predictive score results being mainly driven by the behavior of the log determinant in the $D_{t+h|t}(m)$ term, the eigenvalues and the eigenvectors of the predictive population covariance matrix may also be investigated in more detail to learn which linear combinations of the variables in y_{t+h}^* are the most and the least predictable. Focusing on the large selection at the one-step-ahead horizon, it turns out that the three largest eigenvalues of $C[y_{t+1}^*|\mathcal{Y}_t, m]$ have eigenvectors whose largest values in absolute terms are multiplied by the real effective exchange rate, exports, and imports. Furthermore, for the DSGE and DSGE-VAR models the eigenvectors to the smallest eigenvalues have their largest values in absolute terms on real GDP, employment, and the private consumption deflator, while the corresponding variables for the BVAR and the random-walk are given by employment, the GDP and the private consumption deflators. The largest eigenvalues tend to be linked with the

[22] In fact, the smallest eigenvalue of the population forecast error covariance matrix is the forecast error variance of the most predictable linear combination of the variables in the selection.

[23] It should be kept in mind that the scaled RMSEs and MSEs of the point forecasts are based on the standard deviations of the variables. Since these parameters do not exist for the random walk model they are not appropriate scale factors for this model. In contrast, the forecast error covariance matrix used to standardize the squared forecast errors in the $Q_{t+h|t}(m)$ term is fully model consistent.

real effective exchange rate, exports and imports also for the other forecast horizons, whereas the evidence on the smallest eigenvalues is more varied.

The log determinant terms based on the three smallest and the three largest eigenvalues of the predictive population covariance matrices for the large selection are depicted for all models and forecast horizons in Figures 15 and 16, respectively. As can be seen in these plots, the overall ranking of the models is not affected by these further decompositions of $D_{t+h|t}(m)$. This suggests that the evidence for the log determinant term is not driven by any model being particularly good or bad at forecasting a certain linear combination of the variables. Specifically, the BVAR is better at forecasting its most as well as its least predictable linear combinations of the variables in the large selection relative to the corresponding linear combinations of its competitors.[24] This coincides with the finding in Figures 7 and 8 that the ranking of models is broadly maintained across the large, medium, and small selection of variables.

8. Summary and Conclusions

This paper develops and applies tools for computing and comparing density forecasts based on the predictive likelihood using Bayesian methods. As pointed out by Geweke and Amisano (2010, p. 217), the predictive likelihood

> "...lies at the heart of Bayesian calculus for posterior model probabilities, reflecting the logical positivism of the Bayesian approach: a model is as good as its predictions."

While the calculation of posterior model probabilities requires that the "true" model exists among the set of models under consideration, model selection through the posterior odds ratio remains valid also when all of the models are false.

The predictive likelihood can be applied in a forecast comparison exercise via the log predictive score, but may also be used more generally as a model selection device, to determine weights in a model averaging exercise, or when constructing optimal prediction pools under a Bayesian approach. The paper suggests that the marginal h-step-ahead predictive likelihood can be computed via missing observations techniques when marginalizing the joint predictive density for the full set of variables. As a consequence, the approach makes it possible to calculate the marginal predictive likelihood for a subset of the variables that a model can predict, including a single variable, as well as the joint predictive likelihood for a subset of the variables. Accordingly, marginal and joint density forecasts for shared variables can be compared across models with different dimensions and different conditioning information. In the case of linear time series models with Gaussian innovations, the method is implemented as a missing observations consistent Kalman filter. For nonlinear and non-Gaussian models, a missing observations consistent filter, such as the particle filter, serves the same objective.

[24] The remaining log determinant term (based on the six eigenvalues of the predictive population covariance matrix that are less than the three largest and greater than the three smallest) displays similar properties as the terms for the largest and the smallest eigenvalues.

Once the predictive likelihood at a value of the parameters (the conditional likelihood) can be computed, the paper suggests using Monte Carlo integration over the posterior draws of the parameters to obtain an estimate of the predictive likelihood. This has the interpretation of being an importance sampling (IS) estimator of the predictive likelihood, where the posterior sampler using only historical information represents the importance density. The IS estimator is both consistent and unbiased, while a harmonic mean based estimator of the predictive likelihood would require an additional set of posterior parameter draws to be consistent. However, the IS estimator is less likely to work well in practise when examining joint forecasts spanning a long horizon with many variables. For such a case, the posterior parameter draws are less likely to cover well enough the parameter region where the conditional likelihood is large and it may instead be pertinent to apply a cross-entropy method for selecting the importance density optimally.

In our empirical illustration with five linear Gaussian models, the IS estimator of the predictive likelihood is compared with a normal predictive likelihood, constructed from the mean vector and the covariance matrix of the predictive distribution. The analysis is an extension of the CCW study for euro area data and compares the results for the NAWM, two DSGE-VAR models with the NAWM as prior, a large BVAR, and a multivariate random walk model. The DSGE-VAR models were not included in CCW and are used to relax the strong cross-equation restrictions of the DSGE model.

Over the forecast sample 1999Q1 until 2006Q4 we find that the normal density provides a good approximation of the predictive likelihood when examining the density forecasts for the five models. The value of the predictive likelihood is represented by the IS estimator for all models except the random walk, whose predictive likelihood is multivariate t and can therefore be calculated directly from its analytical expression. In terms of a model ranking, the log predictive score—the sum of the log predictive likelihood over the forecast sample—strongly favors the BVAR model, with the two DSGE-VAR models improving somewhat on the density forecasts of the NAWM, especially at the shorter horizons. The random walk model, on the other hand, is only competitive with the NAWM at the one-step-ahead horizon, especially for the variable selection with real GDP growth, GDP deflator inflation, and the short-term nominal interest rate only. Moreover, when examining the individual predictive likelihood values it is found that the ranking of models is not determined by one or two outliers, but is broadly supported throughout the forecast sample.

When the normal density provides a good approximation of the predictive likelihood, its analytical form can be utilized to assess which feature of the predictive likelihood is driving the ranking of models. Specifically, the normal approximation allows for a simple decomposition of the predictive likelihood into the contributions of forecast uncertainty and forecast errors. The former term is specified via the determinant of the population forecast error covariance matrix, while the latter term is determined through the quadratic standardized forecast errors.

This decomposition suggests that the model ranking in the empirical illustration is primarily influenced by the forecast uncertainty term, while the forecast errors are mainly responsible for the volatility in the predictive likelihood.

Since the determinant of the forecast error covariance matrix is equal to the product of its eigenvalues, the largest (smallest) such eigenvalue gives the forecast error variance of the least (most) predictable linear combination of the variables. An examination of these eigenvalues suggests that the ranking of the models is not determined by any model being able to forecast only its most or its least predictable linear combination of the variables better than the competitors. This is consistent with the finding that the ranking of models is broadly maintained across the large, medium, and small selection of variables.

Appendix A: Posterior Properties of the Random Walk Model

The purpose of this Appendix is to provide technical details on the predictive density of the random walk model with a standard diffuse prior on the residual covariance matrix. An analytical expression of the predictive density is derived and its mean vector and covariance matrix are also determined.

To these ends, let

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{A.1}$$

where the residuals ε_t are assumed to be i.i.d. Gaussian with zero mean and positive definite covariance matrix Ω and y_0 is fixed. The diffuse prior is given by

$$p(\Omega) \propto \left|\Omega\right|^{-(n+1)/2}. \tag{A.2}$$

Stacking the model in (A.1) into $n \times T$ matrices $y = [y_1 \ \cdots \ y_T]$, $x = [y_0 \ \cdots \ y_{T-1}]$, and $\varepsilon = [\varepsilon_1 \ \cdots \ \varepsilon_T]$, the posterior distribution is proportional to the prior times the likelihood, which in natural logarithms can be expressed as

$$\log L\big(y\big|y_0, \Omega\big) + \log p(\Omega) = -\frac{nT}{2}\log(2\pi) - \frac{T+n+1}{2}\log\left|\Omega\right| - \frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\varepsilon\varepsilon'\big]. \tag{A.3}$$

Recognizing that the last two terms on the right hand side of (A.3) form the log of the kernel of the n-dimensional inverted Wishart distribution with location matrix $\varepsilon\varepsilon'$ and T degrees of freedom, we obtain

$$\begin{aligned}\log p\big(\Omega\big|y, y_0\big) = &-\frac{nT}{2}\log(2) - \frac{n(n-1)}{4}\log(\pi) - \log\Gamma_n(T) + \frac{T}{2}\log\left|\varepsilon\varepsilon'\right| \\ &- \frac{T+n+1}{2}\log\left|\Omega\right| - \frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\varepsilon\varepsilon'\big],\end{aligned} \tag{A.4}$$

where

$$\Gamma_n(T) = \prod_{i=1}^{n}\Gamma([T-i+1]/2),$$

for $T \geq n > 0$ with $\Gamma(\cdot)$ being the gamma function. From Bayes theorem it therefore follows that the log marginal likelihood is given by (A.3) minus (A.4), i.e.

$$\log p\big(y|y_0\big) = -\frac{n(2T-n+1)}{4}\log(\pi) + \log\Gamma_n(T) - \frac{T}{2}\log\left|\varepsilon\varepsilon'\right|. \tag{A.5}$$

Normal Approximation of the Marginal Predictive Likelihood

When forecasting with the random walk model it holds that

$$E\big[y_{T+h}\big|y, y_0, \Omega\big] = y_T, \quad h = 1, \ldots, H. \tag{A.6}$$

The forecast error is therefore equal to the accumulation of ε_{T+i} over $i = 1, \ldots, h$, while the forecast error covariance matrix given Ω is

$$C\big(y_{T+h}|y, y_0, \Omega\big) = h\Omega, \quad h = 1, \ldots, H. \tag{A.7}$$

From Rao-Blackwellization we know that the covariance matrix $C(y_{T+h}|y, y_0)$ is equal to the mean of the covariance matrix in (A.7) with respect to the posterior of Ω plus the covariance matrix of the deviation of the mean in (A.6) and its population mean $E[y_{T+h}\big|y, y_0]$. The latter term is zero since the population mean is also y_T, while the former term is given by h times the mean of the posterior of Ω.[25] That is,

$$C\big(y_{T+h}|y, y_0\big) = \frac{h}{T-n-1}\varepsilon\varepsilon'. \tag{A.8}$$

When computing the marginal predictive likelihood with a normal approximation for the full system we therefore make use of the forecast errors $y_{T+h} - y_T$ and the covariance matrix in (A.8).

When forecasting only a subset of the variables we need to take into account how the posterior distribution for the covariance matrix of the corresponding subset of residuals is related to the posterior $p(\Omega|y, y_0)$. Let K be an $n \times n^*$ matrix of columns from I_n which selects $y_t^* = K'y_t$. Similarly, let $K_\perp$ the $n \times (n - n^*)$ matrix which selects the remaining variables from the y_t vector. Define

$$M = \big[K \;\; K_\perp\big], \tag{A.9}$$

i.e. M is an $n \times n$ matrix made up of all the columns of the identity matrix and therefore has a unit determinant while $M^{-1} = M'$. The posterior distribution of $\Omega_M = M'\Omega M$ is an n-dimensional inverted Wishart with location matrix $M'\varepsilon\varepsilon'M$ and T degrees of freedom. Letting $\Omega_K = K'\Omega K$, it follows from, e.g., Bauwens, Lubrano, and Richard (1999, Theorem A.17) that the posterior of Ω_K is an n^*-dimensional inverted Wishart with location matrix $K'\varepsilon\varepsilon'K$ and $T - n + n^*$ degrees of freedom.

With this in mind, the normal approximation of the marginal predictive likelihood for the subset of variables is based on the mean forecast error $y_{T+h}^* - y_T^*$ and the population covariance matrix

$$C\big(y_{T+h}^*|y, y_0\big) = \frac{h}{T-n-1}K'\varepsilon\varepsilon'K. \tag{A.10}$$

Analytical Form of the Marginal Predictive Likelihood

The determination of the marginal predictive likelihood requires an expression for the conditional likelihood function $L(y_{T+h}^*|y, y_0, \Omega)$. From equation (31) and using $y_{T+h|T} = y_T$ and $\Sigma_{y,T+h|T} = h\Omega$ we find that the conditional log-likelihood for the random walk model is given by

$$\log L\big(y_{T+h}^*\big|y, y_0; \Omega\big) = -\frac{n^*}{2}\log(2\pi h) - \frac{1}{2}\log\big|\Omega_K\big| - \frac{1}{2h}\mathrm{tr}\big[\Omega_K^{-1}\varepsilon_h^*\varepsilon_h^{*\prime}\big], \tag{A.11}$$

where $\varepsilon_h^* = y_{T+h}^* - y_T^*$, and the term involving $\log(h)$ is due to $|h\Omega_K| = h^{n^*}|\Omega_K|$. It now remains to multiply the conditional likelihood with the posterior of Ω_K and integrate out Ω_K from the expression.

---

[25] More generally, the posterior distribution of $h\Omega$ is inverted Wishart with location parameter $h\varepsilon\varepsilon'$ and T degrees of freedom.

The product of the conditional log-likelihood of y^*_{T+h} and the posterior of Ω_K is given by:

$$\begin{aligned} p\big(y^*_{T+h}, \Omega_K \big| y, y_0\big) = &\frac{\big|K'\varepsilon\varepsilon' K\big|^{(T-n+n^*)/2}}{(2\pi h)^{n^*/2} 2^{(T-n+n^*)n^*/2} \pi^{n^*(n^*-1)/4} \Gamma_{n^*}(T-n+n^*)} \times \\ &\times \big|\Omega_K\big|^{-(T-n+2n^*+2)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega_K^{-1}\big[K'\varepsilon\varepsilon' K + h^{-1}\varepsilon^*_h \varepsilon^{*\prime}_h\big]\right)\right]. \end{aligned} \tag{A.12}$$

Recognizing that the two terms involving Ω_K is the kernel of an n^*-dimensional inverted Wishart distribution with location matrix $K'\varepsilon\varepsilon' K + h^{-1}\varepsilon^*_h \varepsilon^{*\prime}_h$ and $T-n+n^*+1$ degrees of freedom, it follows that the integral of the density $p\big(y^*_{T+h}, \Omega_K \big| y, y_0\big)$ with respect to Ω_K is equal to the expression in the first term on the right hand side of equation (A.12) times the inverse of the integration constant of the $IW_{n^*}(K'\varepsilon\varepsilon' K + h^{-1}\varepsilon^*_h \varepsilon^{*\prime}_h, T-n+n^*+1)$ distribution. We therefore find that

$$p\big(y^*_{T+h} \big| y, y_0\big) = \frac{\Gamma_{n^*}(T-n+n^*+1)\big|hK'\varepsilon\varepsilon' K\big|^{-1/2}}{\pi^{n^*/2}\Gamma_{n^*}(T-n+n^*)\big|I_{n^*} + \big(hK'\varepsilon\varepsilon' K\big)^{-1}\varepsilon^*_h \varepsilon^{*\prime}_h\big|^{(T-n+n^*+1)/2}}. \tag{A.13}$$

In other words (and as expected), the density of $y^*_{T+h}\big|y, y_0$ is an n^*-dimensional t-distribution with mean y_T, covariance matrix given in (A.10), and $T-n+n^*$ degrees of freedom; see, e.g., Bauwens et al. (1999, Appendix A) for details.[26]

[26] Notice also that $\big|I_{n^*} + \big(hK'\varepsilon\varepsilon' K\big)^{-1}\varepsilon^*_h \varepsilon^{*\prime}_h\big| = 1 + \varepsilon^{*\prime}_h \big(hK'\varepsilon\varepsilon' K\big)^{-1}\varepsilon^*_h$; see, e.g., Magnus and Neudecker (1988, Proof of Theorem 1.9).

Appendix B: Posterior Properties of the Large BVAR Model

The large BVAR is estimated with the methodology suggested in Bańbura et al. (2010) and therefore relies on using dummy observations when implementing the normal-inverted Wishart version of the Minnesota prior. Below we will first present the prior and posterior distribution and thereafter show the relation between the prior parameters and the T_d dummy observations; see also Lubik and Schorfheide (2006).

The VAR representation of y_t is given in equation (10), with $\epsilon_t \sim N_n(0, \Omega)$. Stacking the VAR system as $y = [y_1 \ \cdots \ y_T]$, $X = [X_1 \ \cdots \ X_T]$, and $\epsilon = [\epsilon_1 \ \cdots \ \epsilon_T]$, the log-likelihood is given by

$$\log L\big(y\big|X_1;\Phi,\Omega\big) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log\big|\Omega\big| - \frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\epsilon\epsilon'\big]. \tag{B.1}$$

The normal-inverted Wishart prior for (Φ, Ω) is given by

$$\mathrm{vec}\big(\Phi\big)\big|\Omega \sim N_{n(np+1)}\big(\mathrm{vec}\big(\Phi_\mu\big), \big[\Omega_\Phi \otimes \Omega\big]\big), \tag{B.2}$$

$$\Omega \sim IW_n\big(A, v\big). \tag{B.3}$$

This means that the sum of the log-likelihood and the log prior is given by

$$\begin{aligned}\log f\big(y,\Phi,\Omega\big|X_1\big) = &-\frac{n(T+np+1)}{2}\log(2\pi) - \frac{nv}{2}\log(2) - \frac{n(n-1)}{4}\log(\pi)\\ &- \log\Gamma_n(v) - \frac{n}{2}\log\big|\Omega_\Phi\big| + \frac{v}{2}\log\big|A\big| - \frac{T+n(p+1)+v+2}{2}\log\big|\Omega\big|\\ &- \frac{1}{2}\mathrm{tr}\left[\Omega^{-1}\left(\epsilon\epsilon' + A + \big(\Phi-\Phi_\mu\big)\Omega_\Phi^{-1}\big(\Phi-\Phi_\mu\big)'\right)\right].\end{aligned} \tag{B.4}$$

Using standard "Zellner" algebra, it is straightforward to show that

$$\epsilon\epsilon' + A + (\Phi-\Phi_\mu)\Omega_\Phi^{-1}(\Phi-\Phi_\mu)'\big) = \big(\Phi-\bar\Phi\big)\left(XX' + \Omega_\Phi^{-1}\right)\big(\Phi-\bar\Phi\big)' + S, \tag{B.5}$$

where

$$\begin{aligned}\bar\Phi &= \left(yX' + \Phi_\mu\Omega_\Phi^{-1}\right)\left(XX' + \Omega_\Phi^{-1}\right)^{-1},\\ S &= yy' + A + \Phi_\mu\Omega_\Phi^{-1}\Phi_\mu' - \bar\Phi\left(XX' + \Omega_\Phi^{-1}\right)\bar\Phi'.\end{aligned}$$

Substituting for (B.5) in (B.4), we find that the conjugate normal-inverted Wishart prior gives us a normal posterior for $\Phi|\Omega$ and an inverted Wishart marginal posterior of Ω. Specifically,

$$\mathrm{vec}\big(\Phi\big)\big|\Omega, y, X_1 \sim N_{n(np+1)}\big(\mathrm{vec}\big(\bar\Phi\big), \big[(XX' + \Omega_\Phi^{-1})^{-1} \otimes \Omega\big]\big), \tag{B.6}$$

$$\Omega\big|y, X_1 \sim IW_n\big(S, T+v\big). \tag{B.7}$$

Combining these posterior results with equations (B.4) and (B.5) it follows that the log marginal likelihood is given by

$$\begin{aligned}\log p\big(y\big|X_1\big) = &-\frac{nT}{2}\log(\pi) + \log\Gamma_n(T+v) - \log\Gamma_n(v) - \frac{n}{2}\log\big|\Omega_\Phi\big|\\ &+ \frac{v}{2}\log\big|A\big| - \frac{n}{2}\log\big|XX' + \Omega_\Phi^{-1}\big| - \frac{T+v}{2}\log\big|S\big|.\end{aligned} \tag{B.8}$$

The prior in (B.2) and (B.3) can be implemented through $T_d = n(p+2)+1$ dummy observations by prepending the y $(n \times T)$ and X $(np+1 \times T)$ matrices with the following:

$$\begin{aligned} y_{(d)} &= \begin{bmatrix} \lambda_o^{-1}\text{diag}[\delta \odot \omega] & 0_{n\times n(p-1)} & \text{diag}[\omega] & 0_{n\times 1} & \tau^{-1}\text{diag}[\delta \odot \mu] \end{bmatrix} \\ X_{(d)} &= \begin{bmatrix} 0_{1\times np} & 0_{1\times n} & \gamma^{-1} & 0_{1\times n} \\ \lambda_o^{-1}\big(j_p \otimes \text{diag}[\omega]\big) & 0_{np\times n} & 0_{np\times 1} & \tau^{-1}\big(\imath_p \otimes \text{diag}[\mu]\big) \end{bmatrix}. \end{aligned} \tag{B.9}$$

The vector $\imath_p$ is a p-dimension unit vector, while the $p \times p$ matrix $j_p = \text{diag}[1 \cdots p]$. The hyperparameter $\lambda_o > 0$ gives the overall tightness in the Minnesota prior, the cross-equation tightness is set to unity, while the harmonic lag decay hyperparameter is equal to 2. The hyperparameter $\tau > 0$ handles shrinkage for the sum of coefficients prior on $(I_n - \sum_{i=1}^{p} \Phi_i)$, where $\tau \to 0$ means that the prior on the sum of the lag coefficients approach the case of exact differences, and where shrinkage decreases as τ becomes larger. The n-dimensional vector δ gives the prior mean of the diagonal of Φ_1, ω is a vector of scale parameters for the residuals ϵ_{it}, while μ is a vector that reflects the mean of y_{it}. Finally, γ reflects the overall tightness on Φ_0.

The relationship between the dummy observations and the prior parameters $(\Phi_\mu, \Omega_\Phi, A, v)$ are:

$$\begin{aligned} \Phi_\mu &= y_{(d)}X_{(d)}'\left(X_{(d)}X_{(d)}'\right)^{-1}, & \Omega_\Phi &= \left(X_{(d)}X_{(d)}'\right)^{-1}, \\ A &= \big(y_{(d)} - \Phi_\mu X_{(d)}\big)\big(y_{(d)} - \Phi_\mu X_{(d)}\big)', & v &= T_d - (np+1) + 2. \end{aligned}$$

This guarantees that the prior mean of Ω exists. Letting $y_\star = [y_{(d)}\ y]$ and $X_\star = [X_{(d)}\ X]$, it follows that the posterior parameters

$$\begin{aligned} \bar{\Phi} &= y_\star X_\star'\big(X_\star X_\star'\big)^{-1}, \\ XX' + \Omega_\Phi^{-1} &= X_\star X_\star', \\ S &= \big(y_\star - \bar{\Phi}X_\star\big)\big(y_\star - \bar{\Phi}X_\star\big)'. \end{aligned}$$

In the empirical application, $\tau = 10\lambda_o$, i.e. a relatively loose prior on the sum of the autoregressive matrices. The hyperparameters $\delta_i = 0$ if y_{it} is a first difference variable and $\delta_i = 1$ when y_{it} is a levels variable. The scale parameters ω_i is given by the within-sample residual standard deviation from an AR(p) model for y_{it}, while μ_i is equal to the within-sample mean of y_{it}. The parameter $\varsigma = \gamma^{-1}$ is set to a very small number, which takes care of having an improper prior on Φ_0.

The formula suggested by Bańbura et al. (2010) for selecting λ_o can be expressed as

$$\bar{\lambda}_o\big(\phi\big) = \arg\min_{\lambda_o} \left| \phi - \frac{1}{q}\sum_{j=1}^{q} \frac{\sigma_j^2(\lambda_o)}{\sigma_j^2(0)} \right|,$$

where $\phi \in (0,1)$ is the desired fit, and $\sigma_j^2(\tilde{\lambda}_o)$ is the one-step-ahead mean square forecast error of variable j when $\lambda_o = \tilde{\lambda}_o$. The one-step-ahead within-sample mean square forecast errors used in

the selection scheme are based on the sample 1985Q1–1998Q4. With $\phi = 0.5$, $q = 3$ using real GDP growth, the GDP deflator, and the short-term nominal interest rate, this selection scheme sets $\bar{\lambda}_o = 0.0693$ when $p = 4$.

It should be noted that having an improper prior on Φ_0 technically means that Ω_Φ is singular. This needs to be taken into account when computing, e.g., the log marginal likelihood in (B.8). To deal with this, let

$$X = \begin{bmatrix} \imath_T' \\ Y \end{bmatrix}, \qquad X_{(d)} = \begin{bmatrix} 0_{1\times T_d} \\ Y_{(d)} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \Phi_1 & \dots & \Phi_p \end{bmatrix}, \qquad \Omega_\Phi = \begin{bmatrix} \gamma^2 & 0_{1\times np} \\ 0_{np\times 1} & \Omega_\Gamma \end{bmatrix},$$

where $\imath_T$ is a $T \times 1$ unit vector. The prior for the BVAR is now expressed as

$$\operatorname{vec}(\Gamma)\big|\Omega \sim N_{n^2p}\big(\operatorname{vec}(\Gamma_\mu), [\Omega_\Gamma \otimes \Omega]\big), \tag{B.10}$$

while $p(\Phi_0) = 1$ and the prior of Ω is given by (B.3). Let $Z = y - \Gamma Y$, $\bar{\Phi}_0 = T^{-1} Z \imath_T$, and let

$$D = I_T - T^{-1} \imath_T \imath_T',$$

a $T \times T$ symmetric and idempotent matrix. Through the usual Zellner algebra we have that

$$\epsilon\epsilon' = ZDZ' + (\Phi_0 - \bar{\Phi}_0)\imath_T' \imath_T (\Phi_0 - \bar{\Phi}_0)'.$$

Furthermore, with D being symmetric and idempotent we may define $\tilde{Z} = ZD$, such that $\tilde{y} = yD$, $\tilde{Y} = YD$ and $ZDZ' = \tilde{Z}\tilde{Z}'$. The Zellner algebra now provides us with

$$\tilde{Z}\tilde{Z}' + (\Gamma - \Gamma_\mu)\Omega_\Gamma^{-1}(\Gamma - \Gamma_\mu)' + A = (\Gamma - \bar{\Gamma})\left(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\right)(\Gamma - \bar{\Gamma})' + S,$$

where

$$\begin{aligned} \bar{\Gamma} &= \left(\tilde{y}\tilde{Y}' + \Gamma_\mu \Omega_\Gamma^{-1}\right)\left(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\right)^{-1} \\ S &= \tilde{y}\tilde{y}' + A + \Gamma_\mu \Omega_\Gamma^{-1} \Gamma_\mu' - \bar{\Gamma}\left(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\right)\bar{\Gamma}'. \end{aligned}$$

It can therefore be shown that the normal-inverted Wishart posterior for the VAR parameters is given by

$$\Phi_0\big|\Gamma, \Omega, y, X_1 \sim N_n\big(\bar{\Phi}_0, T^{-1}\Omega\big), \tag{B.11}$$

$$\operatorname{vec}(\Gamma)\big|\Omega, y, X_1 \sim N_{n^2p}\big(\operatorname{vec}(\bar{\Gamma}), \big[(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1})^{-1} \otimes \Omega\big]\big) \tag{B.12}$$

$$\Omega\big|y, X_1 \sim IW_n\big(S, T + v - 1\big). \tag{B.13}$$

Hence, the improper prior on Φ_0 results in a loss of degrees of freedom for the posterior of Ω. Furthermore, the log marginal likelihood is

$$\begin{aligned} \log p\big(y\big|X_1\big) = &-\frac{n(T-1)}{2}\log(\pi) + \log\Gamma_n(T + v - 1) - \log\Gamma_n(v) - \frac{n}{2}\log\big|\Omega_\Gamma\big| \\ &+ \frac{v}{2}\log\big|A\big| - \frac{n}{2}\log(T) - \frac{n}{2}\log\big|\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\big| - \frac{T + v - 1}{2}\log\big|S\big|, \end{aligned} \tag{B.14}$$

where the term $\log(T)$ stems from $T = \imath_T' \imath_T$ and is obtained when integrating out Φ_0 from the joint posterior. The relationship between the dummy observations and the prior parameters is

$$\Gamma_\mu = y_{(d)} Y_{(d)}' \left(Y_{(d)} Y_{(d)}' \right)^{-1}, \qquad \Omega_\Gamma = \left(Y_{(d)} Y_{(d)}' \right)^{-1},$$
$$A = \left(y_{(d)} - \Gamma_\mu Y_{(d)} \right) \left(y_{(d)} - \Gamma_\mu Y_{(d)} \right)', \qquad v = T_d - (np+1) + 2.$$

Letting $\tilde{y}_\star = [y_{(d)}\ \tilde{y}]$ and $\tilde{Y}_\star = [Y_{(d)}\ \tilde{Y}]$, it follows that the posterior parameters

$$\bar{\Gamma} = \tilde{y}_\star \tilde{Y}_\star' \big(\tilde{Y}_\star \tilde{Y}_\star' \big)^{-1},$$
$$\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1} = \tilde{Y}_\star \tilde{Y}_\star',$$
$$S = \big(\tilde{y}_\star - \bar{\Gamma} \tilde{Y}_\star \big) \big(\tilde{y}_\star - \bar{\Gamma} \tilde{Y}_\star \big)'.$$

Normal Approximation of Marginal Predictive Likelihood

Conditional on the parameters and the historical data, the expected value of y_{T+h} is obtained from (20) as

$$E\big[y_{T+h} \big| y, X_1, \Phi, \Omega \big] = J_p' \bar{x}_{T+h} + J_p \Psi^h Y_T, \quad h = 1, \ldots, H. \tag{B.15}$$

Similarly, the covariance matrix of the forecast error for fixed parameters is given by

$$C\big(y_{T+h} \big| y, X_1, \Phi, \Omega \big) = J_p' \bar{\Sigma}_Y^{(h)} J_p, \quad h = 1, \ldots, H, \tag{B.16}$$

where $\bar{\Sigma}_Y^{(h)}$ is given by equation (23) with Ω replacing Σ_ϵ.

The mean of the marginal predictive density of y_{T+h} is computed from (B.15) by taking the expectation with respect to the posterior of (Φ, Ω). It follows that

$$E\big[y_{T+h} \big| y, X_1 \big] = E_T \left[J_p' \bar{x}_{T+h} + J_p' \Psi^h Y_T, \right] \quad h = 1, \ldots, H. \tag{B.17}$$

Next, through Rao-Blackwellization, the covariance matrix of the marginal predictive density is equal to the expected value of (B.16) with respect to the posterior of the parameters plus the covariance of (B.15) with respect to the posterior of the parameters, i.e.

$$C\big(y_{T+h} \big| y, X_1 \big) = E_T \left[J_p' \bar{\Sigma}_Y^{(h)} J_p \right] + C_T \left[J_p' \bar{x}_{T+h} + J_p \Psi^h Y_T \right], \quad h = 1, \ldots, H. \tag{B.18}$$

The normal approximation of the marginal predictive likelihood is now based on using (B.17) as mean and (B.18) as covariance. For subsets of variables we simply take the corresponding elements of the mean and covariances.

TABLE 1: Mean forecast errors for the NAWM, the DSGE-VAR(λ, p), and the BVAR models over the evaluation period 1999Q1–2006Q4.

	DSGE-VAR					DSGE-VAR				
	NAWM	$(2.5;2)$	$(6;4)$	BVAR	RW	NAWM	$(2.5;2)$	$(6;4)$	BVAR	RW
horizon	Private consumption					Nominal wages				
1	−0.12	0.01	0.03	−0.10	−0.02	−0.37	−0.19	−0.21	−0.19	−0.02
2	−0.23	−0.00	0.04	−0.11	−0.02	−0.45	−0.20	−0.22	−0.19	−0.01
4	−0.34	−0.01	0.02	−0.10	−0.04	−0.48	−0.21	−0.21	−0.22	−0.03
8	−0.27	−0.06	−0.06	−0.09	−0.12	−0.47	−0.23	−0.19	−0.22	−0.04
	Consumption deflator					GDP deflator				
1	0.14	0.03	0.03	−0.05	0.00	0.08	0.02	0.02	−0.07	0.00
2	0.25	0.06	0.06	−0.05	0.02	0.14	0.05	0.04	−0.07	0.02
4	0.30	0.07	0.08	−0.05	0.03	0.20	0.08	0.08	−0.07	0.03
8	0.16	0.02	0.06	−0.10	0.02	0.13	0.07	0.10	−0.06	0.05

TABLE 2: Percentage share of squared mean errors in mean squared errors for the NAWM, the DSGE-VAR(λ, p), and the BVAR models over the evaluation period 1999Q1–2006Q4.

	DSGE-VAR					DSGE-VAR				
	NAWM	$(2.5;2)$	$(6;4)$	BVAR	RW	NAWM	$(2.5;2)$	$(6;4)$	BVAR	RW
horizon	Private consumption					Nominal wages				
1	12.6	0.1	1.0	11.2	0.2	45.1	23.4	23.3	24.6	0.1
2	31.7	0.0	1.2	11.3	0.3	73.3	32.2	38.8	25.3	0.0
4	62.9	0.2	0.4	7.8	1.2	79.2	35.7	40.7	25.7	0.6
8	51.7	4.4	5.0	4.4	6.2	78.8	39.8	35.7	16.0	1.3
	Consumption deflator					GDP deflator				
1	21.9	1.9	1.1	5.4	0.0	8.3	0.4	0.4	10.6	0.0
2	55.5	8.5	6.8	4.0	0.5	26.7	3.8	2.9	11.5	0.4
4	65.1	8.0	11.1	3.0	0.9	43.6	11.5	11.7	8.6	0.9
8	45.5	1.1	7.4	8.7	0.3	32.7	8.7	17.1	4.1	1.9

TABLE 3: Log predictive score using the IS estimator over the evaluation period 1999Q1–2006Q4.

horizon	NAWM	DSGE-VAR (2.5; 2)	DSGE-VAR (6; 4)	BVAR	RW
	Large selection (12 variables)				
1	−383.2	−333.0	−347.9	−295.6	−416.4
2	−438.8	−381.9	−395.6	−332.3	−507.0
3	−461.3	−408.0	−419.5	−356.9	−556.2
4	−465.4	−423.3	−431.6	−377.4	−595.0
5	−462.2	−429.0	−436.1	−383.4	−615.9
6	−454.6	−430.6	−437.4	−390.1	−623.6
7	−442.8	−427.1	−432.9	−393.9	−626.7
8	−428.9	−420.1	−424.6	−392.9	−623.9
	Medium selection (seven variables)				
1	−115.6	−82.5	−93.3	−43.1	−129.2
2	−163.9	−124.8	−136.8	−76.7	−191.2
3	−184.7	−149.3	−159.4	−96.3	−225.9
4	−194.3	−163.7	−172.0	−112.1	−251.4
5	−196.3	−171.5	−178.0	−117.8	−266.3
6	−196.7	−176.4	−182.1	−127.4	−273.5
7	−192.0	−176.9	−181.0	−133.1	−278.1
8	−187.7	−176.1	−179.2	−136.0	−279.7
	Small selection (three variables)				
1	−41.6	−32.7	−37.8	−14.8	−43.1
2	−64.6	−54.3	−60.2	−32.8	−69.7
3	−73.3	−64.6	−69.9	−46.3	−81.5
4	−78.4	−71.8	−75.9	−57.1	−92.8
5	−79.8	−75.8	−78.4	−61.9	−99.8
6	−80.5	−77.3	−79.5	−66.5	−101.8
7	−79.3	−78.3	−79.5	−71.8	−104.4
8	−77.6	−78.2	−78.8	−73.2	−105.4

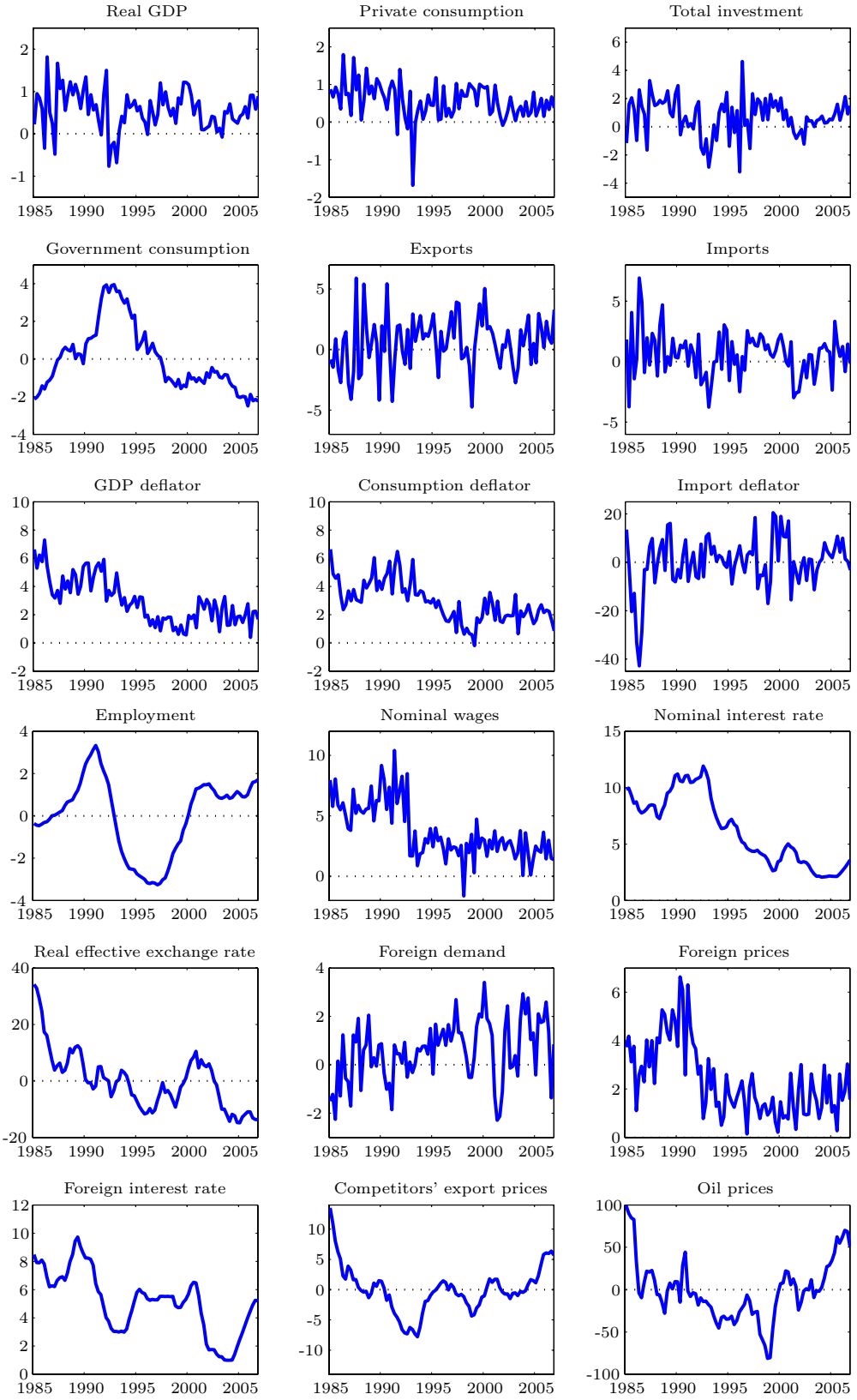
Note: The log predictive likelihood for the random walk model is calculated with its analytical expression; see Appendix A. For the NAWM and the DSGE-VAR models, 10,000 posterior draws have been taken from the available 500,000 post burn-in draws for each time period. The used draws have been selected as draw number $1, 51, 101, \ldots, 499951$. For the BVAR direct sampling is possible and 50,000 posterior draws have been used; see Appendix B.

TABLE 4: Difference between log predictive score using the IS estimator and the normal approximation over the evaluation period 1999Q1–2006Q4.

horizon	NAWM	DSGE-VAR (2.5; 2)	DSGE-VAR (6; 4)	BVAR	RW
	Large selection (12 variables)				
1	1.37	2.47	1.50	2.37	5.98
2	1.39	3.52	1.88	0.25	8.78
3	1.58	3.95	2.17	0.25	9.81
4	1.82	3.68	2.27	2.92	8.81
5	2.12	3.52	2.35	0.69	8.45
6	2.31	3.43	2.10	2.11	8.44
7	2.17	3.39	1.82	2.54	7.86
8	2.13	3.27	1.79	2.46	7.50
	Medium selection (seven variables)				
1	0.93	1.57	0.92	−1.62	3.24
2	1.31	2.68	1.48	−1.28	4.04
3	1.63	3.20	1.95	−1.06	4.36
4	1.81	3.47	2.19	0.06	3.97
5	2.04	3.76	2.29	−1.10	3.82
6	1.96	3.91	2.20	−0.29	3.99
7	1.90	4.13	2.23	−0.63	3.82
8	1.77	4.31	2.21	−0.38	3.72
	Small selection (three variables)				
1	0.70	0.68	0.40	−0.58	0.67
2	0.93	1.08	0.54	−0.50	0.84
3	1.04	1.28	0.72	0.11	1.11
4	1.03	1.31	0.76	0.38	1.00
5	1.00	1.35	0.73	0.34	0.90
6	0.86	1.36	0.64	0.42	1.02
7	0.80	1.38	0.64	0.56	0.98
8	0.75	1.46	0.68	0.76	0.98

Note: See Table 3.

FIGURE 1: The data.



Note: This figure shows the time series of the observed variables used in the estimation of the NAWM. Details on the variable transformations are provided in Christoffel, Coenen, and Warne (2008, Section 3.2) or Section 2.3 in CCW. Inflation and interest rates are reported in annualized percentage terms.

FIGURE 2: Marginal likelihood as a function of λ for different lag orders.

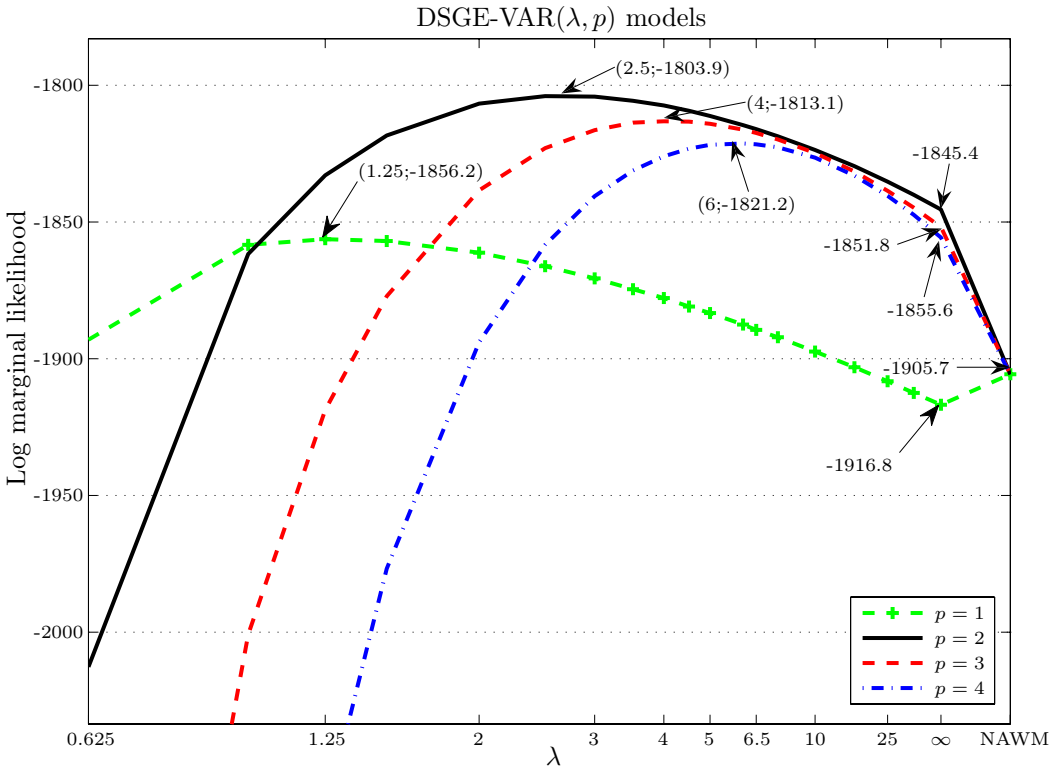


FIGURE 3: Scaled root mean squared forecast errors for 12 variables.

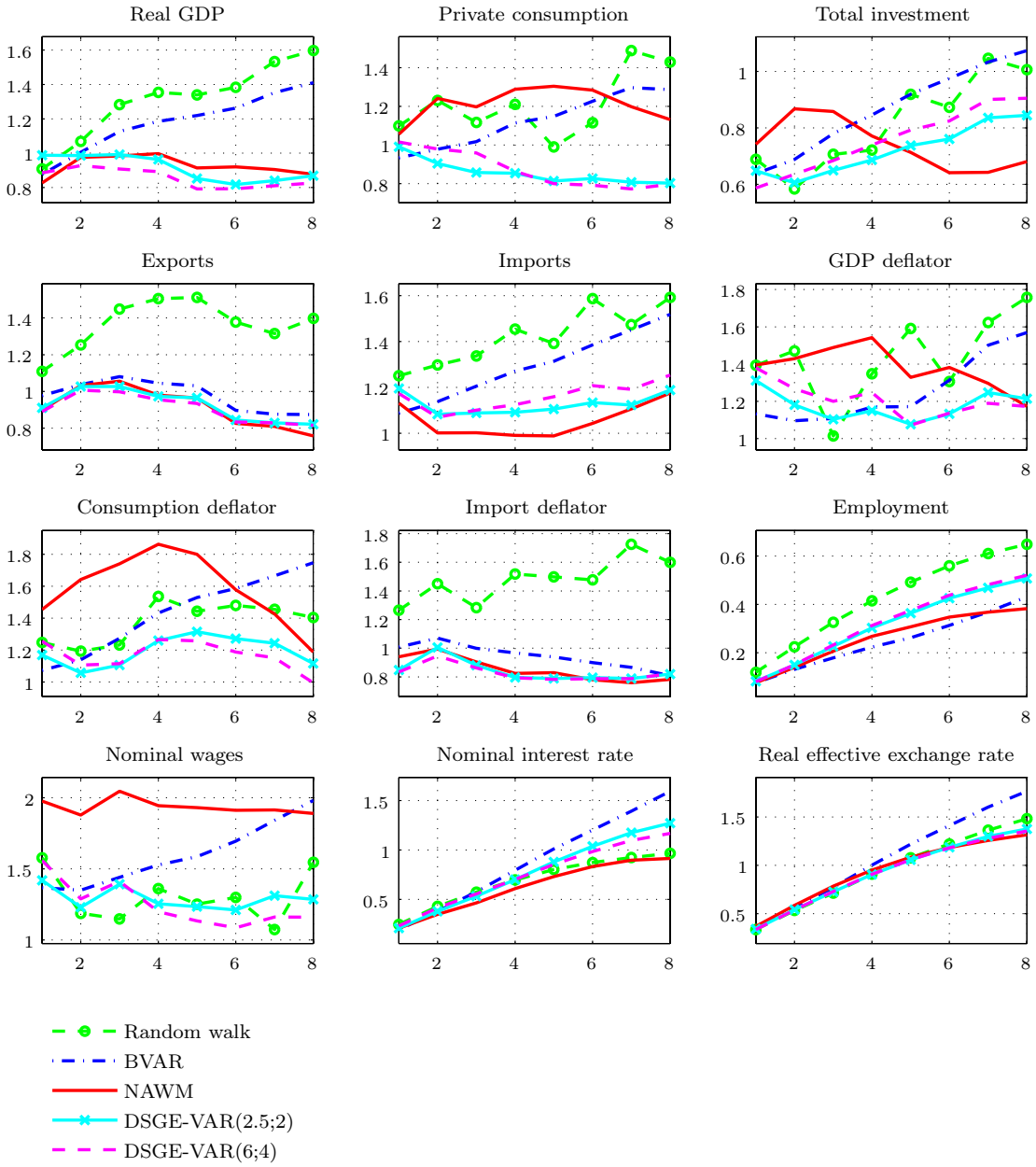


FIGURE 4: Quarterly nominal wages and private consumption mean forecast paths over the forecast evaluation period 1999Q1–2006Q4.

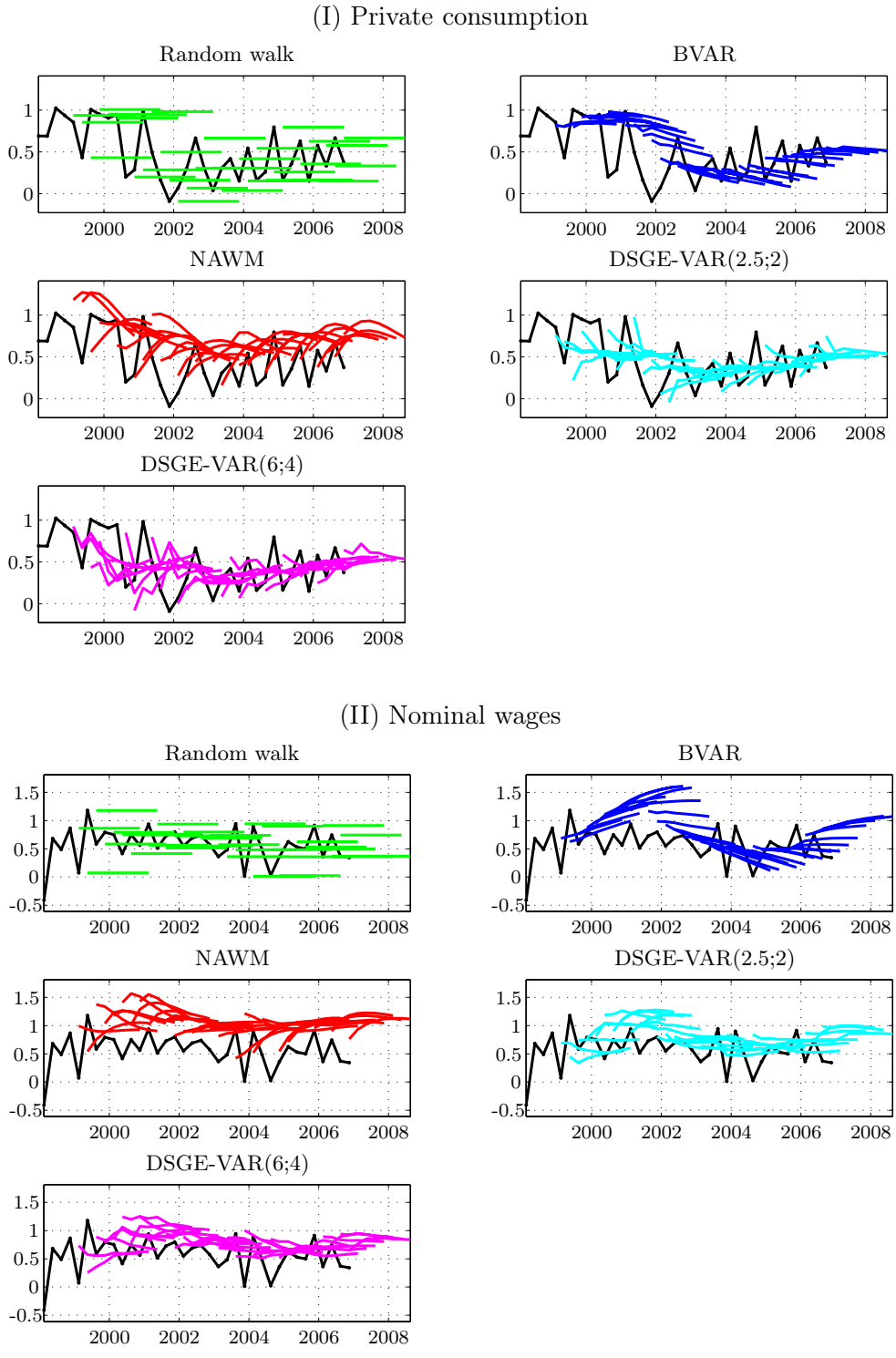


FIGURE 5: Trace statistics of the scaled MSE matrices.

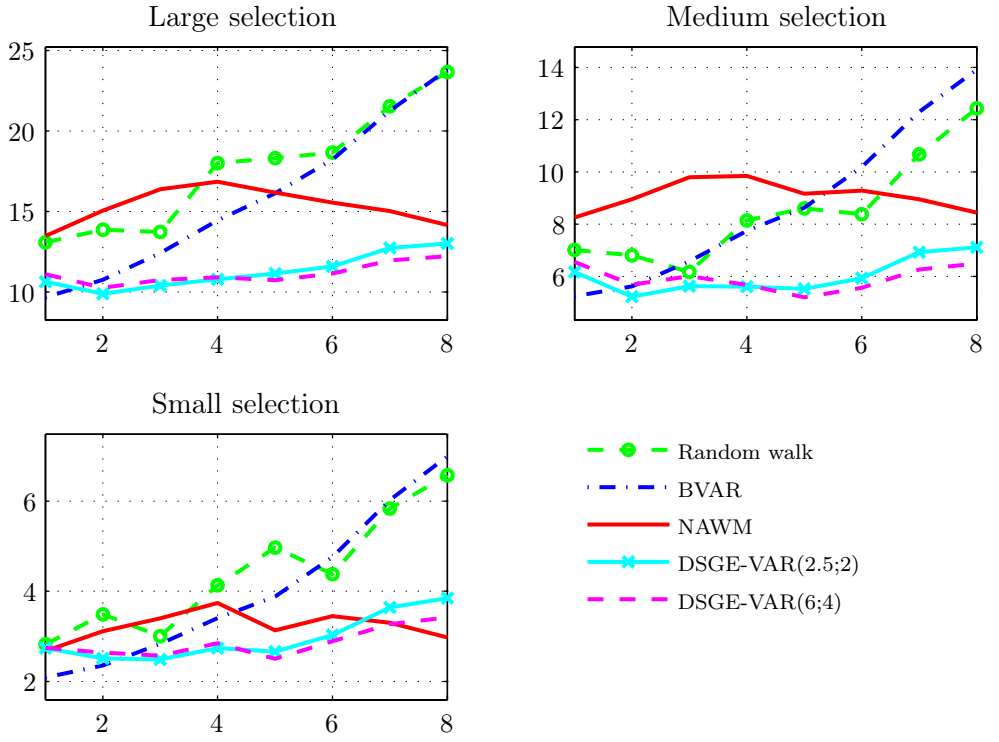


FIGURE 6: Log determinant statistics of the scaled MSE matrices.

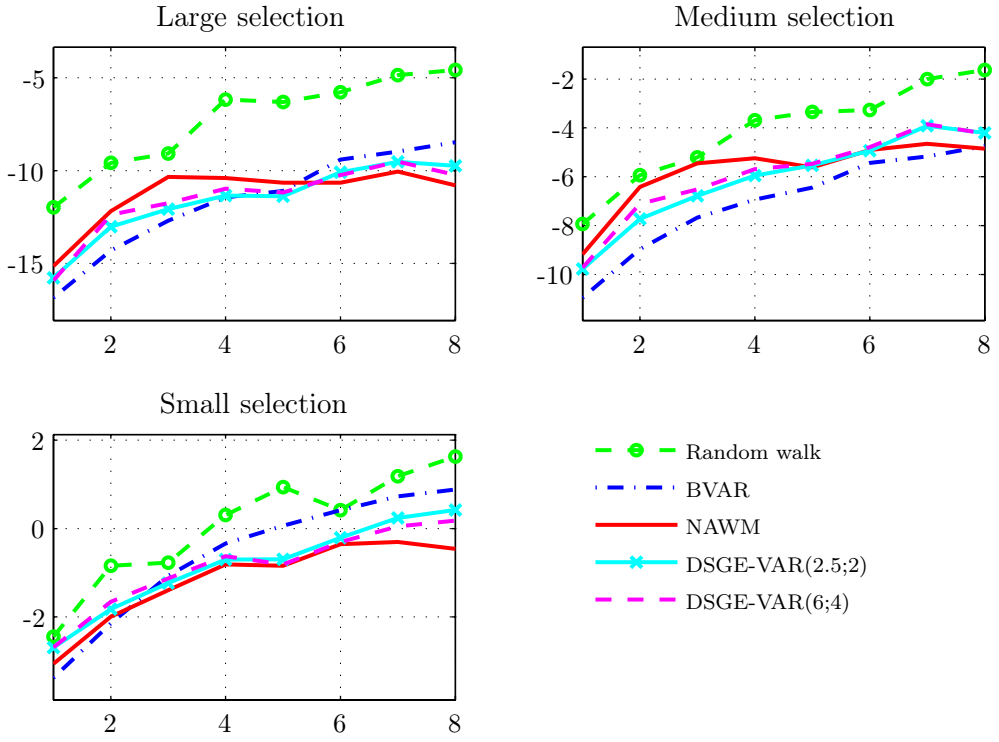
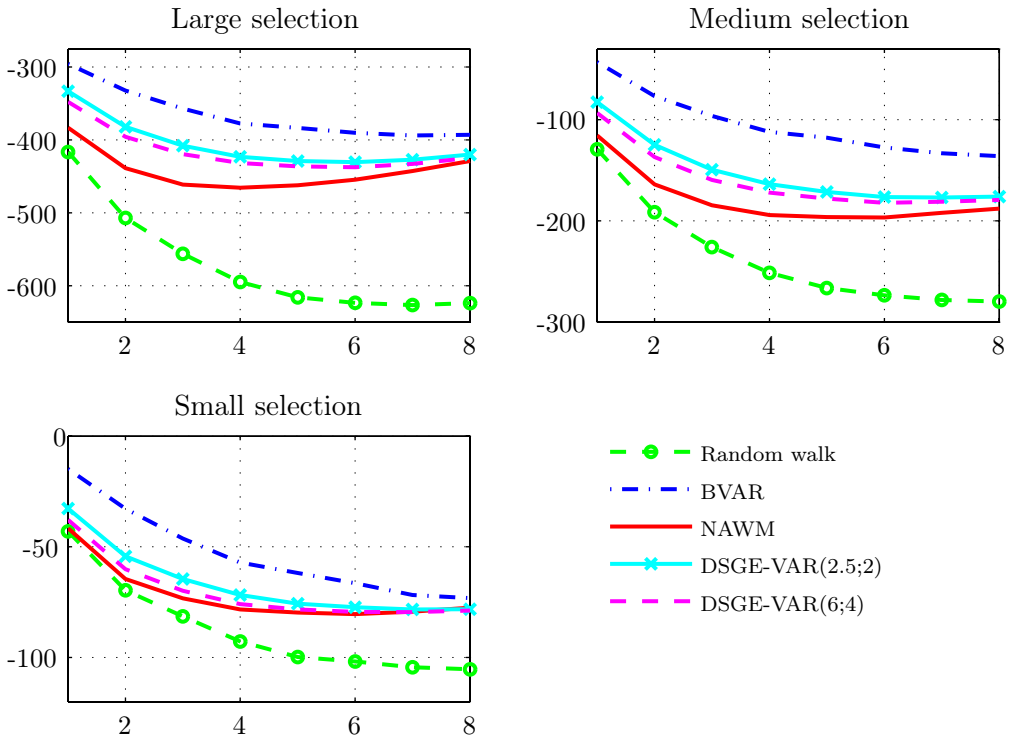


FIGURE 7: Log predictive scores using the IS estimator.



Note: The log predictive likelihood for the random walk model is calculated with its analytical expression.

FIGURE 8: Log predictive scores using the normal approximation.

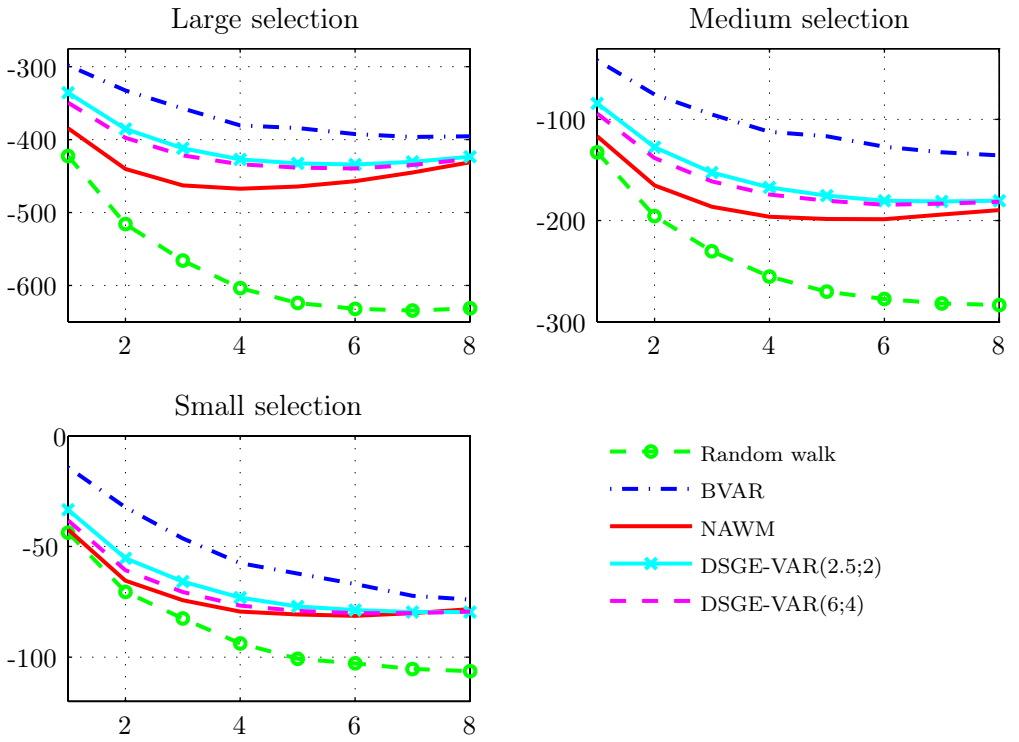


FIGURE 9: Log predictive likelihoods for the large selection of variables using the NAWM.

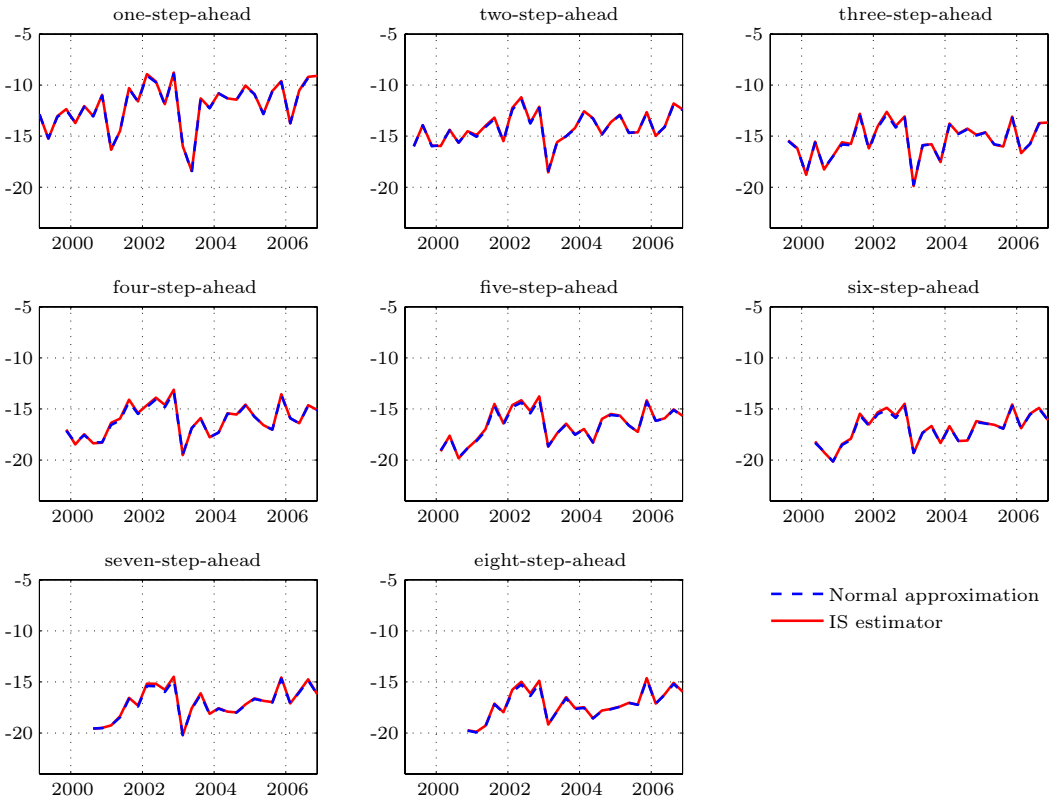


FIGURE 10: Log predictive likelihoods for the large selection of variables using the DSGE-VAR(2) model.

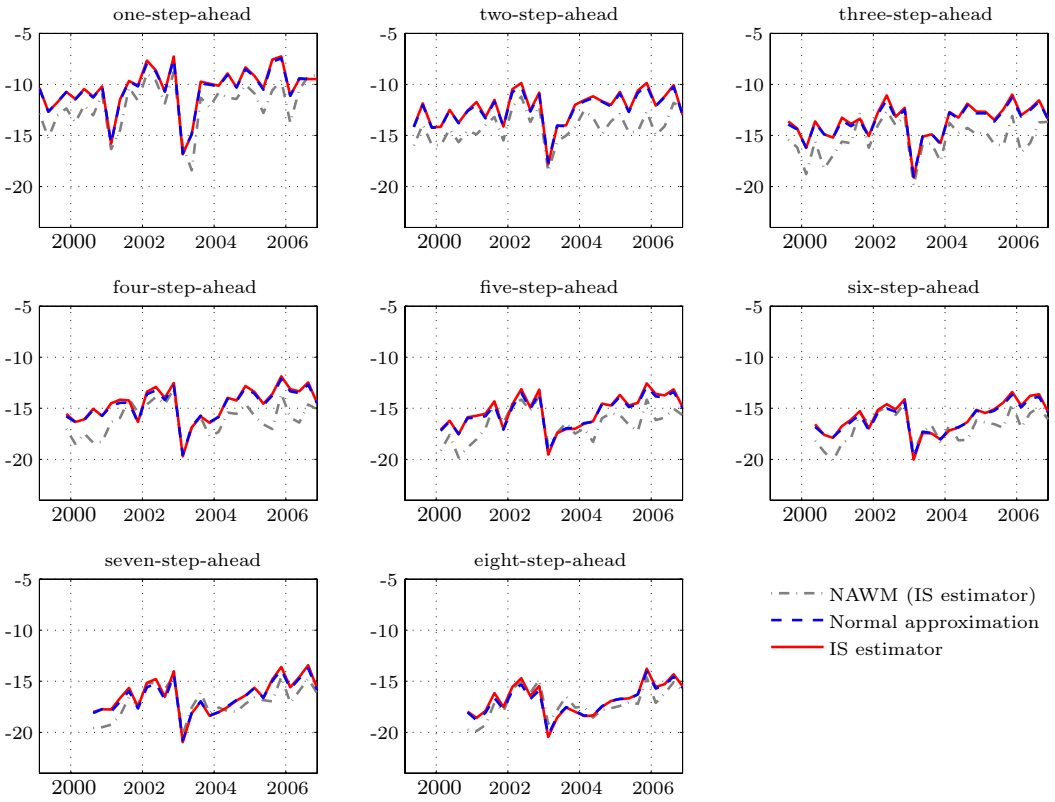


FIGURE 11: Log predictive likelihoods for the large selection of variables using the large BVAR model.

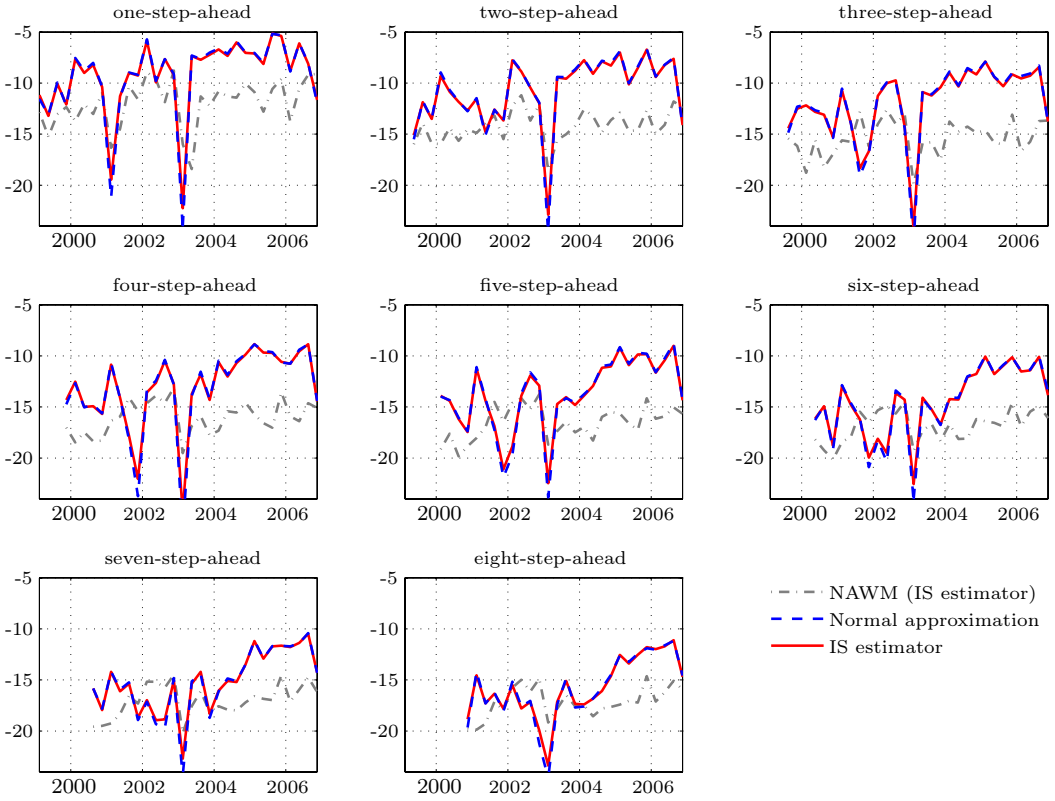


FIGURE 12: Log predictive likelihoods for the large selection of variables using the random walk model.

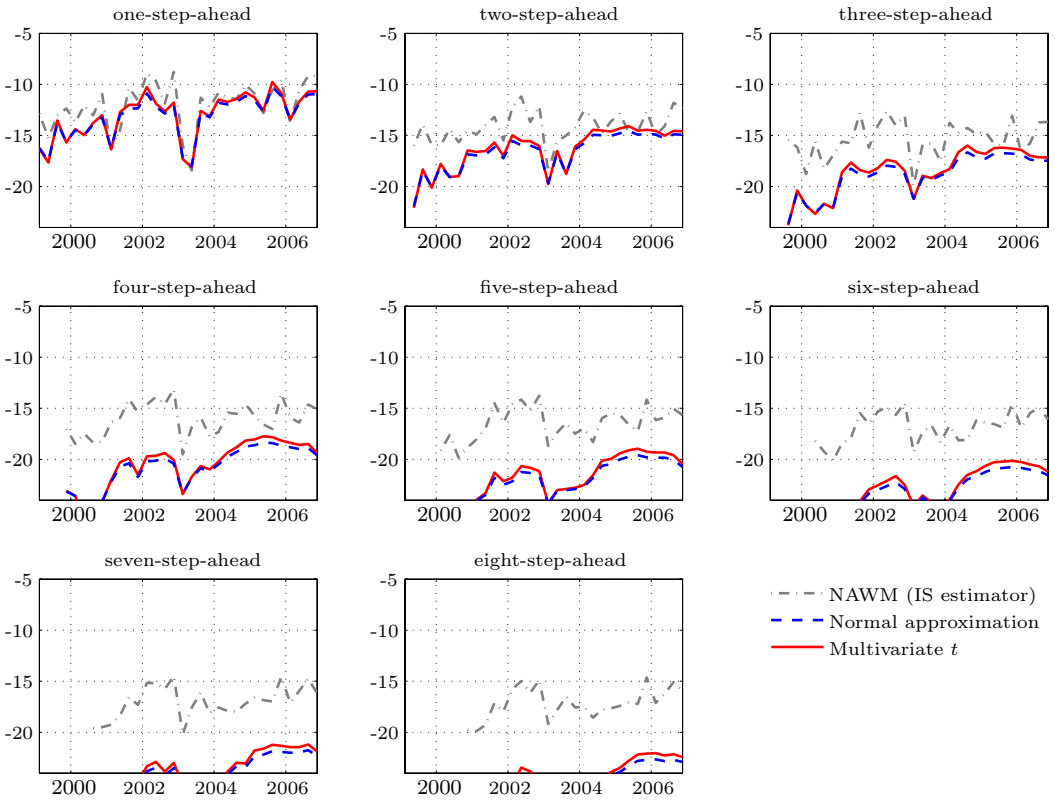


FIGURE 13: The evolution of the log determinant term of the normal density for the large selection of variables.

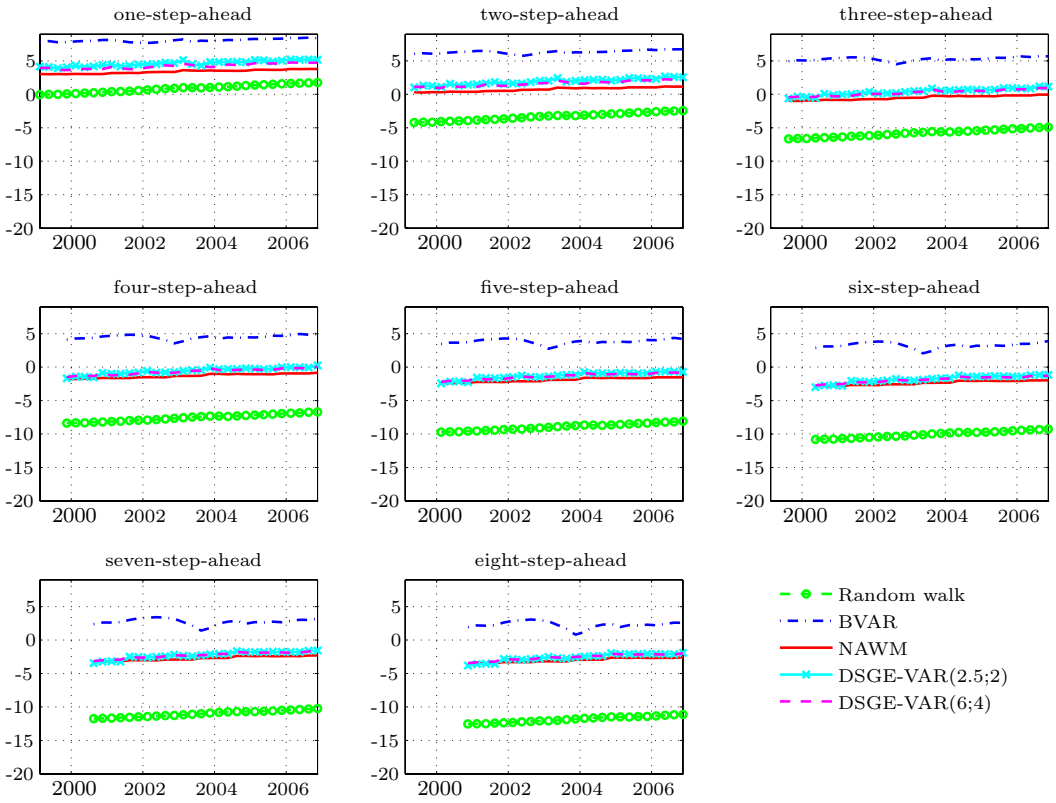


FIGURE 14: The evolution of the quadratic standardized forecast error term of the normal density for the large selection of variables.

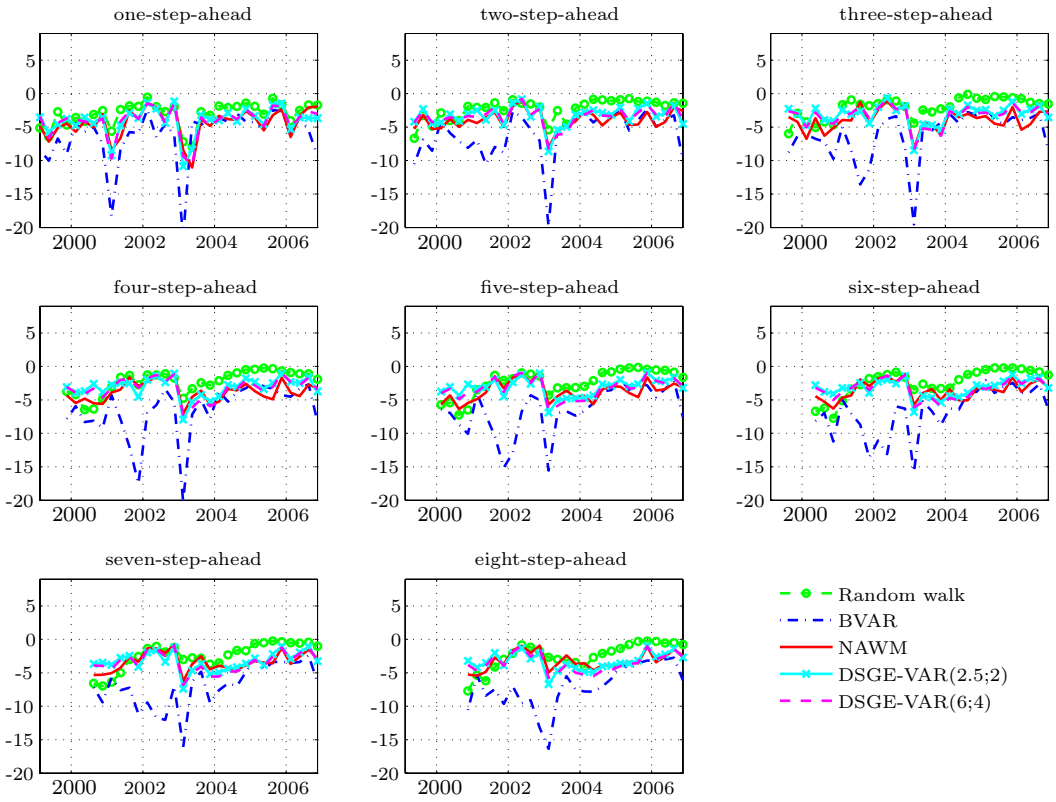


FIGURE 15: The evolution of the log determinant term based on the three smallest eigenvalues of the predictive population covariance matrices.

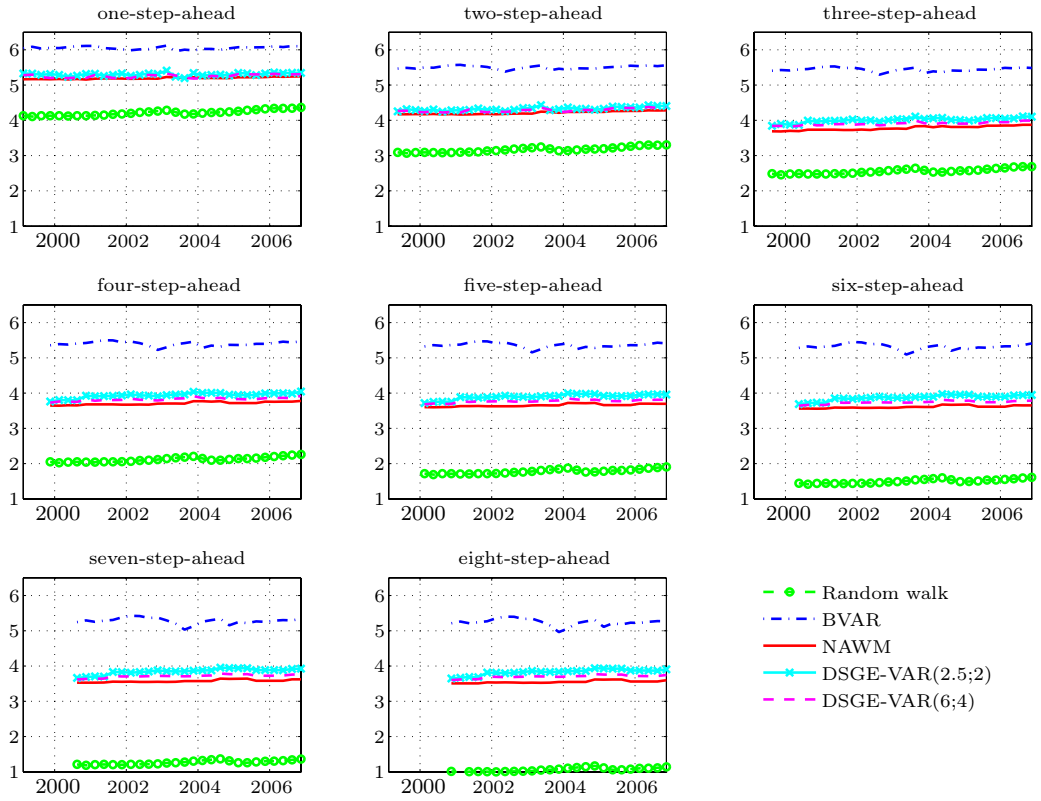
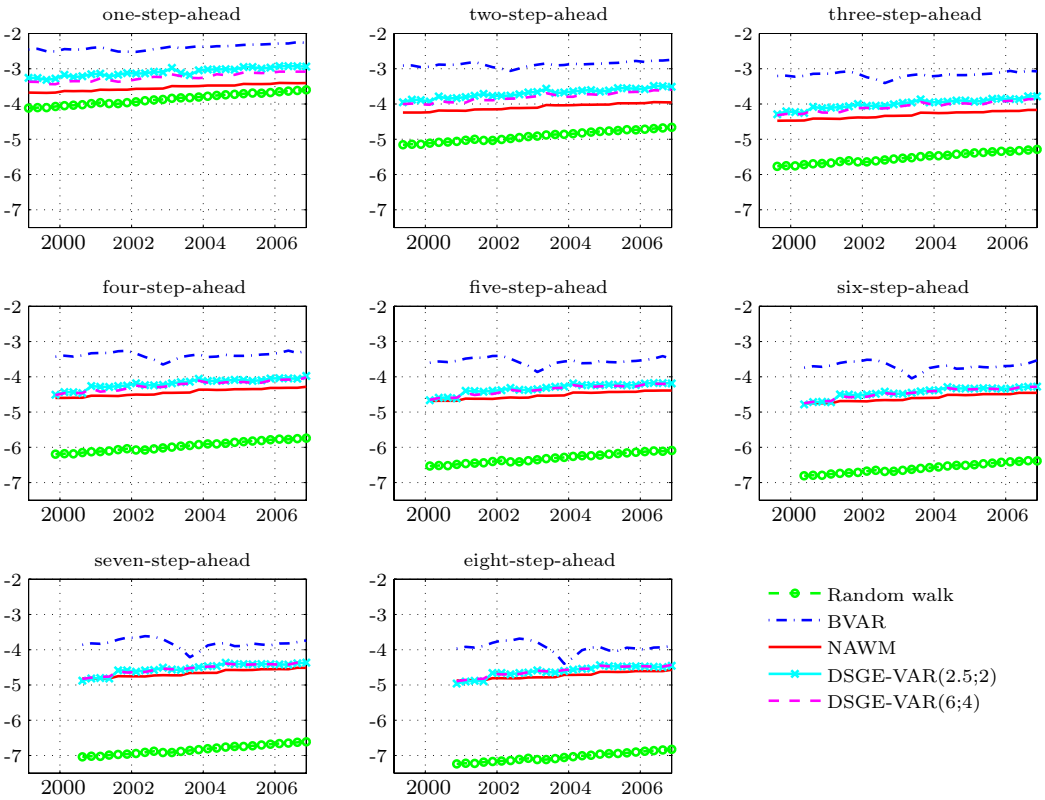


FIGURE 16: The evolution of the log determinant term based on the three largest eigenvalues of the predictive population covariance matrices.



References

Adolfson, M., Laséen, S., Lindé, J., and Villani, M. (2008), "Evaluating an Estimated New Keynesian Small Open Economy Model," *Journal of Economic Dynamics and Control*, 32, 2690–2721.

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

An, S. and Schorfheide, F. (2007), "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172, with discussion, p. 173–219.

Anderson, G. S. (2010), "A Reliable and Computationally Efficient Algorithm for Imposing the Saddle Point Property in Dynamic Models," *Journal of Economic Dynamics and Control*, 34, 472–489.

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

Bauwens, L., Lubrano, M., and Richard, J. F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.

Bernardo, J. M. (1979), "Expected Information as Expected Utility," *The Annals of Statistics*, 7, 686–690.

Box, G. E. P. (1980), "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society Series A*, 143, 383–430.

Calvo, G. A. (1983), "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12, 383–398.

Chan, J. C. C. and Eisenstat, E. (2012), "Marginal Likelihood Estimation with the Cross-Entropy Method," Manuscript, Centre for Applied Macroeconomic Analysis, Australian National University.

Christoffel, K., Coenen, G., and Warne, A. (2008), "The New Area-Wide Model of the Euro Area: A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis," ECB Working Paper Series No. 944.

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

Dawid, A. P. (1982), "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, 77, 605–610.

Dawid, A. P. (1984), "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Series A*, 147, 278–292.

Del Negro, M. and Schorfheide, F. (2004), "Priors from General Equilibrium Models," *International Economic Review*, 45, 643–673.

Del Negro, M. and Schorfheide, F. (2006), "How Good Is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models," *Federal Reserve Bank of Atlanta Economic Review*, 91, 21–37.

Del Negro, M. and Schorfheide, F. (2009), "Monetary Policy Analysis with Potentially Misspecified Models," *American Economic Review*, 99, 1415–1450.

Del Negro, M. and Schorfheide, F. (2012), "DSGE Model-Based Forecasting," Prepared for *Handbook of Economic Forecasting*, Volume 2.

Del Negro, M., Schorfheide, F., Smets, F., and Wouters, R. (2007), "On the Fit of New-Keynesian Models," *Journal of Business & Economic Statistics*, 25, 123–143, with discussion, p. 143–162.

Dieppe, A. and Warmedinger, T. (2007), "Modelling Intra- and Extra-Area Trade Substitution and Exchange Rate Pass-Through in the Euro Area," ECB Working Paper Series No. 760.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, 2nd edition.

Eklund, J. and Karlsson, S. (2007), "Forecast Combinations and Model Averaging using Predictive Measures," *Econometric Reviews*, 26, 329–363.

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

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., and Watson, M. W. (2007), "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

Franchi, M. and Paruolo, P. (2012), "On ABCs (and Ds) of VAR Representations of DSGE Models," Manuscript, University of Rome "La Sapienza".

Frühwirth-Schnatter, S. (2004), "Estimating Marginal Likelihoods for Mixture and Markov Switching Models Using Bridge Sampling Techniques," *Econometrics Journal*, 7, 143–167.

Gelfand, A. and Dey, D. (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society Series B*, 56, 501–514.

Geweke, J. (1999), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–73.

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, John Wiley, Hoboken.

Geweke, J. and Amisano, G. (2010), "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, 26, 216–230.

Geweke, J. and Amisano, G. (2011), "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

Geweke, J. and Amisano, G. (2012), "Prediction and Misspecified Models," *American Economic Review*, 102, 482–486.

Geweke, J. and Whiteman, C. H. (2006), "Bayesian Forecasting," in G. Elliott, C. W. Granger, and A. Timmermann (Editors), *Handbook of Economic Forecasting*, 3–80, North Holland, Amsterdam, volume 1.

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

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

Good, I. J. (1952), "Rational Decisions," *Journal of the Royal Statistical Society Series B*, 14, 107–114.

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

Ingram, B. F. and Whiteman, C. H. (1994), "Supplanting the 'Minnesota' Prior — Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 34, 497–510.

Kass, R. E. and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

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

Lees, K., Matheson, T., and Smith, C. (2011), "Open Economy Forecasting with a DSGE-VAR: Head to Head with the RBNZ Published Forecasts," *International Journal of Forecasting*, 27, 512–528.

Lubik, T. and Schorfheide, F. (2006), "A Bayesian Look at New Open Economy Macroeconomics," in M. L. Gertler and K. S. Rogoff (Editors), *NBER Macroeconomics Annual 2005*, 313–366, MIT Press.

Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.

Meng, X.-L. and Wong, W. H. (1996), "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 6, 831–860.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Ravenna, F. (2007), "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Journal of Monetary Economics*, 54, 2048–2064.

Rubin, D. B. (1991), "EM and Beyond," *Psychometrika*, 56, 241–254.

Sargent, T. J. (1989), "Two Models of Measurement and the Investment Accelerator," *Journal of Political Economy*, 97, 251–287.

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

Sims, C. A., Waggoner, D. F., and Zha, T. (2008), "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146, 255–274.

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

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

Smith, Jr., A. A. (1993), "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Strachan, R. and van Dijk, H. K. (2011), "Divergent Priors and well Behaved Bayes Factors," Tinbergen Institute Discussion Papers 11-006/4, Tinbergen Institute.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540, with discussion, p. 541–550.

Tay, A. S. and Wallis, K. F. (2000), "Density Forecasting: A Survey," *Journal of Forecasting*, 19, 235–254.

Thompson, P. A. and Miller, R. B. (1986), "Sampling the Future: A Bayesian Approach to Forecasting from Univariate Time Series Models," *Journal of Business & Economic Statistics*, 4, 427–436.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728, with discussion, p. 1728–1762.

Villani, M. (2001), "Bayesian Prediction with Cointegrated Vector Autoregressions," *International Journal of Forecasting*, 17, 585–605.

Warne, A. (2012), "YADA Manual — Computational Details," Manuscript, European Central Bank. Available with the YADA distribution.

Winkler, R. L. and Murphy, A. H. (1968), "'Good' Probability Assessors," *Journal of Applied Meteorology*, 7, 751–758.