Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models

Anders Warne, Günter Coenen and Kai Christoffel*

June 26, 2012

Abstract: In this paper we treat the issue of forecasting with DSGE and DSGE-VAR models, with particular attention to Bayesian estimation of the predictive distribution and its mean and covariance. As a novel contribution to the forecasting literature, which extends beyond (log-linearized) DSGE models and DSGE-VARs, we show how the value of the $h$-step-ahead marginal and joint predictive likelihood for a subset of variables can be calculated. This is of interest when attempting to rank models in density forecasting exercises, but it is generally useful since the predictive likelihood is a natural model selection device under a Bayesian approach. The basic idea is to utilize well-known techniques for handling missing data when computing the likelihood function, while the predictive likelihood can thereafter be calculated using either exact (Monte Carlo integration) or approximate methods. This is particularly straightforward for linear Gaussian models since the Kalman filter supports missing data with minor modifications and can thus be used to obtain the value of the likelihood, while the modified harmonic mean estimator and the Laplace approximation may be applied for computing the predictive likelihood.

As an empirical illustration, we use euro area data and compare the forecasting performance of the New Area-Wide Model (NAWM)—a small-open-economy DSGE model developed at the European Central Bank—to DSGE-VARs, which relax the cross-equation restrictions of the former, as well as to reduced-form forecasting models. In earlier work, we have shown that overall the NAWM performs well in comparison to reduced-form models over the forecast sample beginning with the introduction of the euro, but that it has difficulties when it comes to predicting, in particular, wages and private consumption. We therefore examine whether DSGE-VARs can account for the NAWM’s systematic overprediction of both variables and if they can compete with a large BVAR.

Keywords: Bayesian inference, BVAR, DSGE models, DSGE-VAR, forecasting, misspecification, predictive likelihood.

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

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

Note: This paper was previously circulated with the title “Forecasting with DSGE-VARs”. We are particularly grateful to Marta Bańbura who has estimated and computed the forecasts for 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), Frank Schorfheide (University of Pennsylvania), Mattias Villani (Linköping University), and comments from members of the Working Group of Econometric Modelling and the Tinbergen Institute workshop on “Recent Theory and Applications of DSGE Models” at Erasmus University Rotterdam. 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.

* All authors: Directorate General Research, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany; e-mail: anders.warne@ecb.europa.eu, gunter.coenen@ecb.europa.eu, kai.christoffel@ecb.europa.eu.
the assumption that a model is correctly specified.\footnote{Based on earlier work by Phillips (1996), Fernández-Villaverde and Rubio-Ramírez (2004) show that even when all the models being investigated are false, the posterior odds ratio asymptotically favours the model that is closest to the “true” data generating process in the Kullback-Leibler sense. Moreover, parameter point estimates converge to their pseudo-true values.} 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 data generating process to a DSGE model can be traced back to the literature on indirect inference; see Smith (1993) in particular, but also Gouriéroux, Monfort, and Renault (1993). An early attempt of combining 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 that has the DSGE model approximation at one end and an unrestricted VAR at the other end of its range. In between these extreme values, 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.\footnote{On the basis of statistical identification, Consolo, Favero, and Paccagnini (2009) question the appropriateness of using a VAR approximation of the DSGE model. They argue that, for instance, missing variables and time-varying parameters renders the VAR less useful for such purposes and instead they argue for a factor augmented VAR. While a VAR approximation of a DSGE model can certainly be misspecified, the comparison between DSGE and DSGE-VARs based on Bayesian analysis remains valid also when the VAR is misspecified.}

The first contribution of our paper concerns forecasting with DSGE-VARs. 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 \textit{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. In this paper, we present a simple approach to Bayesian estimation of the predictive distribution of a DSGE-VAR, including its population mean, $h$-step-ahead covariance matrix, and a decomposition of the latter into residual and parameter uncertainty.

It has a 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 second and main contribution of our paper concerns the calculation of the predictive likelihood and the underlying idea 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 of the modelled variables over the forecast horizon and the technique can be used for both exact and approximate predictive likelihood methods.

The predictive likelihood for such a subset makes it possible to compare forecast accuracy across models that have different conditioning variables but where the subset being forecasted is shared. Moreover, a subset of the variables that a model can predict is often of greater interest to a forecaster or policy maker than the full set of variables, especially when the models being compared are able to predict a large number of variables and where some or many of them are of marginal interest. 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 concerned with predicting comovements, as in Herbst and Schorfheide (2011). Although a model comparison based on point forecasts may reveal interesting aspects of the models involved, such forecasts are not well suited for model selection or determination of model weights in a model averaging study or when forming optimal prediction pools; see, e.g., Geweke and Amisano (2011). The predictive likelihood is a natural tool for dealing with such matters, with the log predictive score as an extension from the single-period comparisons based on the predictive Bayes factor to a multi-period setting.

We illustrate the above by extending the forecasting 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 nöm) as an example, and their pseudo out-of-sample forecasting exercise covers the period after the introduction of the euro, focusing on the 12 of the 18 observed variables in the NAWM that are endogenously determined. To examine the performance of the NAWM, CCW make use of forecasts obtained from a VAR estimated with maximum likelihood, three BVARs as well as the naïve random walk and sample mean benchmarks. Concerning the BVARs, CCW consider two classes: one that is intended for systems with a smaller dimension and that makes use of a prior on the steady-state (Villani, 2009), 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 DSGE-VAR 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, including an exact and an approximate modified harmonic mean determination as well as two Laplace approximations. 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 examined in Section 7. Finally, Section 8 summarizes the main findings of the paper and concludes.

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 parameter 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 \( \eta_t \) is a \( q \)-dimensional vector with i.i.d. standard normal structural shocks \( \eta_t \sim N(0, I_q) \), while \( \xi_t \) 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 \( \theta \).

Provided that a unique and convergent solution of the system (1) exists at a particular value of \( \theta \) (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 \( \theta \). The observed variables are denoted by \( y_t \), an \( n \)-dimensional vector, which is linked to the model variables \( \xi_t \) 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 \( \mu \) is the population mean (steady state) of \( y_t \) conditional on \( \theta \) provided that \( \xi_t \) is stationary. The measurement errors and the shocks \( \eta_t \) are assumed to be independent, while the matrices \( H \) and \( R \) are also uniquely determined by \( \theta \).

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 $Y_T = \{y_1, \ldots, y_T\}$ given a value of $\theta$ with the Kalman filter; see, e.g., Durbin and Koopman (2001) 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

$$p(y_{T+1}, \ldots, y_{T+h}|Y_T, m) = \int_{\theta \in \Theta} p(y_{T+1}, \ldots, y_{T+h}|Y_T, \theta, m) p(\theta|Y_T, m) d\theta,$$

(4)

where $p(\theta|Y_T, m)$ is the posterior density of $\theta$ based on the data available at time $T$ and the selected model $m$, and $p(y_{T+1}, \ldots, y_{T+h}|Y_T, \theta, m)$ is a normal density.

From our Bayesian perspective, it may be noticed that for a given model there is no uncertainty about the predictive density and, thus, 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.\(^3\)

If we wish to estimate quantiles, confidence regions 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 comparison 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 $\theta$ 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. The authors also discuss estimation of the population mean and covariance matrix of the predictive distribution using only posterior draws of $\theta$ and how this covariance matrix can be decomposed into four terms representing state variable, shock, measurement error, and parameter uncertainty through Rao-Blackwellization; see also CCW.\(^4\)

\(^3\)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 $\theta$ 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.

\(^4\)For a recent Bayesian approach to multivariate forecast evaluation with DSGE models, see Herbst and Schorfheide (2011). For additional discussions on forecasting with DSGE models, see Del Negro and Schorfheide (2012).
The computation of the joint and marginal predictive likelihood for a subset of the variables can also be performed without simulations from the sampling the future algorithm and this task is discussed in Section 5.

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), which guarantees the existence of an infinite-order VAR representation, is not satisfied.\(^5\) Nevertheless, a finite order VAR may still serve as a reasonable approximation of a DSGE model, especially when we are not interested in identifying the structural shocks directly from the VAR model; see, e.g., Chari, Kehoe, and McGrattan (2008) and Christiano, Eichenbaum, and Vigfusson (2007) for discussions.

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:

\[
y_t = \Phi_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \epsilon_t, \quad t = 1, \ldots, T, \tag{5}\]

where \( \epsilon_t \sim N_n(0, \Sigma_\epsilon) \). The vector \( \Phi_0 \) is \( n \times 1 \), while \( \Phi_j \) 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 VAR can be expressed as

\[
y_t = \Phi X_t + \epsilon_t. \tag{6}\]

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 \( \theta \). This means that the \( n \times 1 \) population mean vector for \( y_t \) is \( \mu \), while the central population autocovariance matrices for \( (y_t, y_{t-j}) \) are given by:

\[
\Sigma_y^{(j)} = \begin{cases} 
H'\Sigma_\xi H + R, & \text{if } j = 0, \\
H'F^j\Sigma_\xi H, & \text{for } j = 1, 2, \ldots, 
\end{cases} \tag{7}
\]

\(^5\) The well-known Smets and Wouters (2007) model has the same number of observed variables and shocks, yet it generates unit eigenvalues for the matrix \( [I - B(H' B)^{-1} H']F \) and therefore also fails the poor man’s invertibility condition.
with $\Sigma_y^{(-j)} = \Sigma_{y'}^{(j)}$. The state variable covariance matrix $\Sigma_\xi$ 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 $E[X_tX_t'; \theta]$ and $\Gamma_{yX}(\theta)$ the non-central population covariance matrix $E[y_tX_t'; \theta]$ which can be directly determined from (7) and the mean vector $\mu$. 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

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

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

where $\Gamma_{y\epsilon}(\theta) = E[y_ty_t'; \theta] = \Sigma_y^{(0)} + \mu \mu'$. The matrices $\Phi(\theta)$ and $\Sigma_\epsilon(\theta)$ are restriction functions that will be used to center the prior distribution of $(\Phi, \Sigma_\epsilon)$ conditional on $\theta$ 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:

$$p(\Phi, \Sigma_\epsilon, \theta | \lambda) = p(\Phi, \Sigma_\epsilon | \theta, \lambda) p(\theta).$$

The conditional prior distribution of the VAR parameters on the right hand side of (10) 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 $\lambda$, which generates a continuum of models that have an unrestricted VAR at one extreme ($\lambda$ close to zero) and the DSGE model approximation at the other ($\lambda = \infty$). In practise, a grid is used for $\lambda$ such that a finite number of values $\Lambda = \{\lambda_1, \ldots, \lambda_N\}$ are considered and each DSGE-VAR model is indexed by its $\lambda$-value, denoted by DSGE-VAR($\lambda$).

The prior of the VAR parameters takes the form

$$\Sigma_\epsilon | \theta, \lambda \sim IW_n(\lambda T \Sigma_\epsilon(\theta), \lambda T - (np + 1)),$$

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

The normal-inverted Wishart prior assumed here is proper when $\lambda T \geq n(p+1) + 1$. Hence, the domain of $\lambda$ 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
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 \( \theta \) 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 \( (\theta, \lambda) \) is proportional to the product of the prior density and the likelihood function for the VAR, denoted by \( p(Y_T|X_1, \Phi, \Sigma_\epsilon, \theta, \lambda) \). Conditional on \( (\theta, \lambda) \), the DSGE-VAR prior and likelihood are conjugate and, hence, it follows that the posterior distribution of \( \Phi \) and \( \Sigma_\epsilon \) is also of the normal-inverted Wishart form; see Del Negro and Schorfheide (2004) for details. Notice that the conditional posterior distribution of \( (\Phi, \Sigma_\epsilon) \) depends on the initial values \( X_1 \) via the likelihood function. When we wish to compare marginal likelihoods for the DSGE-VAR to the DSGE model, the information sets need to be the same. This may be handled in the DSGE model by using the sample \( t = 1 \) as a training sample for the Kalman filter, i.e., the likelihood is evaluated only for \( t = 1, \ldots, T \), while the filter also includes the sample \( t = 1 \).

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

\[
p(\Phi, \Sigma_\epsilon, \theta|Y_T, X_1, \lambda) = p(\Phi, \Sigma_\epsilon|Y_T, X_1, \theta, \lambda) p(\theta|Y_T, X_1, \lambda).
\] (13)

The first term on the right-hand side is given by the product of the conditional posterior densities for \( \Phi|\Sigma_\epsilon \) and \( \Sigma_\epsilon \). The second term is the marginal posterior density of \( \theta \) for a given \( \lambda \) and it can be estimated via a marginalized likelihood function, the prior of the DSGE model parameters, and a suitable posterior sampler.

The marginalized likelihood can be determined from the following factorization:

\[
p(Y_T|X_1, \theta, \lambda) = \frac{p(Y_T|X_1, \Phi, \Sigma_\epsilon, \theta, \lambda) p(\Phi, \Sigma_\epsilon|\theta, \lambda)}{p(\Phi, \Sigma_\epsilon|Y_T, X_1, \theta, \lambda)}.
\] (14)

The first term in the numerator is the likelihood function of the VAR in (5), while the second term is the prior density of the VAR parameters conditional on \( \theta \) and \( \lambda \); cf. equations (11) and (12). The denominator is the conditional posterior density of the VAR parameters. An analytical expression for the marginalized likelihood in (14) is provided in Del Negro and Schorfheide (2004, eq. A.2) and this expression is further streamlined and robustified in Warne (2012, Section 15).

Based on the joint conditional posterior density of the VAR parameters, it is straightforward to show that the joint posterior mode of the VAR parameters conditional on \( \theta \) is given by:

\[
\hat{\Phi}(\theta) = \hat{\Phi}(\theta),
\]

\[
\hat{\Sigma}_\epsilon(\theta) = \frac{(1 + \lambda)T}{(1 + \lambda)T + n + 1} \hat{\Sigma}_\epsilon(\theta),
\] (16)
where $\hat{\Phi}(\theta)$ and $(1 + \lambda)T\hat{\Sigma}_\epsilon(\theta)$ are the location parameters of the conditional normal-inverted Wishart posteriors of $\Phi|\Sigma_\epsilon, \theta, \lambda$ and $\Sigma_\epsilon|\theta, \lambda$, respectively; see, e.g., Del Negro and Schorfheide (2004, eq. 28-29).

Posterior mode estimation of $\theta$ via a DSGE-VAR model can be conducted in at least two different ways. The first is to combine the marginalized likelihood function determined from equation (14) with the prior $p(\theta)$ and maximize their product. The resulting marginal posterior mode estimate of $\theta$ is thereafter plugged into (15) and (16) to yield posterior mode values for the VAR parameters.

Alternatively, the conditional posterior mode functions (15) and (16) can be substituted for $\Phi$ and $\Sigma_\epsilon$ in their joint conditional posterior density and this relationship is thereafter multiplied by the marginalized likelihood obtained via (14). The resulting expression forms a concentrated likelihood function which involves the data, $\theta$, and $\lambda$ only; see Warne (2012) for details. Combining this concentrated likelihood with the prior of $p(\theta)$ and maximizing their product gives the joint posterior mode of $(\Phi, \Sigma_\epsilon, \theta)$ via (15) and (16).

Both these posterior mode estimators may be considered when applying a posterior simulator to a DSGE-VAR model. In addition, the combination of the marginalized or the concentrated likelihood with the prior permits direct estimation of an inverse Hessian that may be useful when, e.g., tuning the proposal density for the random-walk Metropolis algorithm. Once a posterior sample has been simulated for the DSGE-VAR($\lambda$), 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 $\lambda$ such that:

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

(17)

An alternative approach would be to average the results from the DSGE-VAR model over the range of $\lambda$. 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 $\lambda$. 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($\infty$) model is a good approximation of the DSGE model; see Adolfsson, Laséen, Lindé, and Villani (2008) for discussions. We will address these issues in our empirical analysis.

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 $p(\Phi|\mathcal{Y}_T, X_1, \Sigma_\epsilon, \theta, \lambda)$;

---

6 This is equivalent to substituting the conditional posterior mode estimate of the VAR parameters into the product of the likelihood function of the data given $(\Phi, \Sigma_\epsilon, \theta)$ and the conditional priors in (11) and (12); see equation (14).
(2) Draw $\epsilon_{T+i}$ 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 (5);
(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 $\lambda$. 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. 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 $\Psi$ denote an $np \times np$ matrix with

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

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:

$$
y_{T+i} = J'_p \bar{x}_{T+i} + J'_p \Psi^i Y_T + J'_p \bar{\epsilon}_{T+i}, \quad i = 1, \ldots, h, \quad (18)
$$

where

$$
\bar{x}_{T+i} = J_p \bar{x}_0 + \Psi \bar{x}_{T+i-1}, \\
\bar{\epsilon}_{T+i} = J_p \bar{\epsilon}_T + \Psi \bar{\epsilon}_{T+i-1}, \quad i = 1, \ldots, h,
$$

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

$$
E(y_{T+i} | \mathcal{Y}_T, X_1, \lambda) = E_T \left[ J'_p \bar{x}_{T+i} + J'_p \Psi^i Y_T \right], \quad i = 1, \ldots, h, \quad (19)
$$

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

---

7 That is, 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 the conditional posteriors of the VAR parameters are not time consuming.
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,

\[ C(y_{T+i}|Y_T, X_1, \lambda) = E_T[C(y_{T+i}|Y_T, X_1, \Phi, \Sigma_e, \theta, \lambda)] + C_T[J_p' \bar{x}_{T+i} + J_p' \Psi Y_T], \tag{20} \]

where \( C_T \) denotes the covariance with respect to the posterior of \((\Phi, \Sigma_e, \theta)\) at time \( T \). Residual uncertainty is determined by the first term on the right-hand side, while the second term represents parameter uncertainty.\(^8\) To develop a simple expression for the first term on the right-hand side of (20), let \( \bar{\Sigma}_Y^{(i)} \) be defined from the difference equation

\[ \bar{\Sigma}_Y^{(i)} = J_p \Sigma_e J_p' + \Psi \Sigma_Y^{(i-1)} \Psi', \quad i = 1, \ldots, h, \tag{21} \]

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

\[ E_T[C(y_{T+i}|Y_T, X_1, \Phi, \Sigma_e, \theta, \lambda)] = E_T[J_p' \bar{\Sigma}_Y^{(i)} J_p], \quad i = 1, \ldots, h. \tag{22} \]

To estimate these population moments in practice we average over a sample of \( N \) draws from the posterior distribution of \( (\Phi, \Sigma_e, \theta) \). With \( N \) being sufficiently large, the numerical standard error of these sample moments is negligible if 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 \( \theta \) 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. The possibility of determining 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). When comparing density forecasts of a set of models the height of the density at the realized value of the observed variables is of particular interest. As Gneiting, Balabdaoui, and Raftery (2007) point out, the assessment of a predictive distribution on the basis of its density and the realized value only—the predictive likelihood—is consistent with the prequential approach of Dawid (1984); see also Geweke (2010).

The use of the predictive likelihood has long been recognized as a valid Bayesian approach to model selection. 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).

\(^8\) The second term on the right-hand side of (20) is simply due to the predictive mean conditional on the parameters being different from the predictive mean.
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 the forecaster maximizes the expected score or utility for an observation drawn from a distribution $D_i$ when the forecaster gives the probabilistic forecast $D_i$ rather than $D_j \neq D_i$. If the maximum is unique then the rule is also said to be strictly proper. Proper scoring rules are important since they encourage the forecaster to be honest.

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

$$S_J(h, m) = \sum_{t=T}^{T+N_h-1} \log p(y_{t+1}, \ldots, y_{t+h}|Y_t, m), \quad h = 1, \ldots, H, \quad (23)$$

where $N_h$ is the number of time periods the $h$-step-ahead predictive density is evaluated, $Y_t$ is the observed data of $y_t$ until period $t$ and $m$ is an index for the model. If the value of the predictive density only depends on the actual realizations of $y$ over the prediction sample, then the scoring rule is said to be local and 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 observed data; see Bernardo and Smith (2000) for general discussions on related issues and Gneiting and Raftery (2007) for a recent survey on scoring rules.

Under the condition that only the realizations of the observed variables over the forecast sample are used in (23), the difference between the log predictive score of model $m$ and model $k$ is equal to $N_h$ times the average log predictive Bayes factor of these two models, where a positive value indicates that, on average, model $m$ is better at predicting the variables over the given sample 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, $Y_t$, 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, Chapter 2.6.2).

For the observations $Y_{T+N_1}$ and with $N_h = N_{h-1} - 1$ for $h > 1$, we can rewrite the log predictive score in (23) as

$$S_J(h, m) = \sum_{i=0}^{h-1} \left[ \log p(Y_{T+N_1-i}, m) - \log p(Y_{T+i}, m) \right], \quad h = 1, \ldots, H. \quad (24)$$

This means that the log predictive score of model $m$ for one-step-ahead forecasts is proportional to the difference between the log-marginal likelihood for the full sample $Y_{T+N_1}$ and the historical sample $Y_T$. 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 $h$ are based on the samples $Y_{T+N_1-i}$ and the last $h$ on $Y_{T+i}$ for $i = 0, \ldots, h-1$. 

- 12 -
It can also be seen that the term within brackets in (24) can be rewritten as a sum of one-step-ahead log predictive likelihoods. Hence, the joint predictive distribution provides a comparison of one-step-ahead forecasts only and is therefore not well suited for a comparison of h-step-ahead forecasts. When comparing the density forecasts of the NAWM and alternative forecast models, CCW therefore focus on the marginal predictive likelihood of the h-step-ahead forecasts rather than the joint predictive likelihood in (23). The log predictive score can now be expressed as

$$S_M(h, m) = \sum_{t=T}^{T+N_h-1} \log p(y_{t+h}|Y_t, m), \quad h = 1, 2, \ldots, H. \quad (25)$$

The relationship between the marginal likelihood and the log predictive score in (25) 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 $p(y_{t+h}|Y_t, m)$ does not have a closed form solution. Since kernel density estimation from the predictive draws is not practical unless the dimension of $y_{t+h}$ is small, they suggest using a normal approximation of the predictive likelihood based on its population moments.

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

$$p(y_{t+h}|Y_t, m) = \frac{p(y_{t+h}, Y_t, m)}{p(Y_t, m)}, \quad h = 1, 2, \ldots, H. \quad (26)$$

The denominator is the marginal likelihood of model $m$ when using the data $Y_t$ and the numerator is likewise the marginal likelihood for this model when using the dataset $(y_{t+h}, Y_t)$. Hence, the connection between the predictive likelihood and the marginal likelihood remains also for $h > 1$. This means that the problem for calculating the log predictive score in (25) for $h > 1$ concerns the question: it is possible to compute the marginal likelihood for the sample $(y_{t+h}, Y_t)$?

Suppose we replace the realizations for $y_{t+i}$, $i = 1, \ldots, h - 1$, with missing observations and apply a valid method for dealing with such data when evaluating the likelihood function of model $m$ for the sample $Y_{t+h}$. 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 for $(y_{t+h}, Y_t)$ can thereafter be computed via standard tools. 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 (2001, Chapter 4.8) or Harvey (1989, Chapter 3.4.7). Once a suitable sampler has been applied to obtain draws from the posterior distribution of the parameters, the marginal likelihood can be estimated with, for instance, the modified harmonic mean.

This idea can also be used to estimate the marginal likelihood for the sample $(y^*_{t+h}, Y_t)$, where $y^*_{t+h}$ is a subset of the elements of $y_{t+h}$, as well as for, e.g., the sample $(y^*_{t+1}, \ldots, y^*_{t+h}, Y_t)$. 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. Moreover, since a VAR model can be written in state-space form the same tool can be used when computing the predictive likelihood for such models.

The predictive likelihood for a subset of the variables may be estimated with, e.g., the modified harmonic mean (Geweke, 1999, 2005, or the extension in Sims, Waggoner, and Zha, 2008) or with bridge sampling techniques; see Meng and Wong (1996), Frühwirth-Schnatter (2004), and references therein. Concerning the former approach, we formally need two sets of posterior draws. Focusing on the data \((y_{t+h}^s, \mathcal{Y}_t)\) the draws are given by \(\theta_h^{(s)} = p(\theta|y_{t+h}^s, \mathcal{Y}_t), s = 1, \ldots, S_h,\) and \(\theta^{(s)} = p(\theta|\mathcal{Y}_t), s = 1, \ldots, S.\) The marginal predictive likelihood can now be estimated as

\[
\hat{p}_H(y_{t+h}^s|\mathcal{Y}_t) = \left[ \frac{1}{S_h} \sum_{s=1}^{S_h} \frac{f(\theta_h^{(s)})}{L(y_{t+h}^s; \mathcal{Y}_t, \theta_h^{(s)})p(\theta_h^{(s)}|\mathcal{Y}_t)} \right]^{-1}
\]

\[
= \left[ \frac{1}{S} \sum_{s=1}^{S} \frac{f(\theta_h^{(s)})}{L(y_{t+h}^s; \mathcal{Y}_t, \theta_h^{(s)})L(\mathcal{Y}_t; \theta^{(s)})p(\theta^{(s)})} \right]^{-1}
\]

\[
\left[ \frac{1}{S} \sum_{s=1}^{S} \frac{f(\theta^{(s)})}{L(\mathcal{Y}_t; \theta^{(s)})p(\theta^{(s)})} \right]
\]

where \(L(y_{t+h}^s; \mathcal{Y}_t, \theta)\) denotes the conditional likelihood, we have used the fact that \(p(\theta|\mathcal{Y}_t) = L(\mathcal{Y}_t; \theta)p(\theta)/p(\mathcal{Y}_t),\) and where the function \(f(\theta)\) is either the truncated normal density, chosen as in Geweke (1999, 2005), or the truncated elliptical density, as in Sims et al. (2008).

Alternatively, we may evaluate the first term on the right hand side of (27) using the posterior draws \(\theta^{(s)} \in p(\theta|\mathcal{Y}_t), s = 1, \ldots, S.\) Formally, this second method is not fully consistent with the modified harmonic mean estimator. However, if we are willing to assume that the posterior density \(p(\theta|y_{t+h}^s, \mathcal{Y}_t)\) is well approximated by \(p(\theta|\mathcal{Y}_t),\) then the second method may be considered. Moreover, this method avoids having to generate a set of posterior draws for each sample \((y_{t+h}^s, \mathcal{Y}_t)\) and is therefore computationally much less expensive than the second method since it only requires posterior draws from the sample \(\mathcal{Y}_t.\) Below we shall follow the second method although, and as pointed out by Geweke and Whiteman (2006), it is not fully consistent with the prequential approach since it does not use all relevant data at \(t + h\) when updating the marginal likelihood for \((y_{t+h}^s, \mathcal{Y}_t)\) from the marginal likelihood for \(\mathcal{Y}_t.\) In contrast, the the first method is consistent with the prequential approach. Nevertheless, when the number of data points that we predict is small the approximation error due to using \(\theta^{(s)}\) instead of \(\theta_h^{(s)}\) in the first term of (27) can also be expected to be small. At the same time, it should be emphasized that the second method should be avoided when predicting a large number of data points, such as for joint \(h\)-step-ahead forecasts with \(h\) being large, since the approximation of \(p(\mathcal{Y}_{t+h})\) is bound to be very poor.9

9 Notice that:

\[
p(\theta|y_{t+h}^s, \mathcal{Y}_t)p(y_{t+h}^s|\mathcal{Y}_t) = p(y_{t+h}^s|\mathcal{Y}_t, \theta)p(\theta|\mathcal{Y}_t).
\]

Hence, the height of the posterior density \(p(\theta|y_{t+h}^s, \mathcal{Y}_t)\) is well approximated by the height of the posterior density \(p(\theta|\mathcal{Y}_t)\) if the ratio \(p(y_{t+h}^s|\mathcal{Y}_t, \theta)/p(y_{t+h}^s|\mathcal{Y}_t)\) is close to unity for all \(\theta \in \Theta\) except when \(p(\theta|\mathcal{Y}_t) \approx 0 \approx p(\theta|y_{t+h}^s, \mathcal{Y}_t).\) However, this also means that \(p(y_{t+h}^s|\mathcal{Y}_t) \approx p(y_{t+h}^s|\mathcal{Y}_t, \theta)\) for all such \(\theta.\)
When the predictive density conditional on the parameters is Gaussian, marginalization with respect to the variables of interest can also be conducted directly from the mean and the covariance matrix of the joint predictive density conditional on the parameters and utilizing the properties of the normal distribution. For large models with long forecast horizons this may involve dealing with large matrices and may therefore be computationally more expensive than using a missing observations approach to marginalization.

While the proposed solution to the problem of how to calculate the log predictive score based on the marginal predictive likelihood for a subset of the observed variables is straightforward, the calculation of marginal likelihoods for large systems, such as those considered below, can be computationally expensive when based on posterior draws. An approximate but computationally inexpensive estimator of the marginal likelihood is the Laplace approximation; see Tierney and Kadane (1986), Gelfand and Dey (1994), and Raftery (1996). It requires that the mode of the log posterior, given by the sum of the log likelihood and the log prior, can be computed and that its Hessian is available.

Letting $\hat{\theta}$ be the posterior mode of $\theta$ and $\hat{\Sigma}$ be minus the Hessian, the Laplace approximation based on the sample $Y_t$ is given by

$$\log \hat{p}_L(Y_t) = \log L(Y_t; \hat{\theta}) + \log p(\hat{\theta}) + \frac{d \log(2\pi) + \log |\hat{\Sigma}|}{2},$$

where $L$ is the likelihood function, $d$ is the dimension of $\theta$, and where the model index $m$ has been suppressed. The third term on the right hand side approximates $-\log p(\hat{\theta}|Y_t)$ with $O(t^{-1})$ accuracy and, hence, the expression in (28) is a reflection of Bayes theorem through what Chib (1995) calls the basic marginal likelihood identify.

Similarly, let $\theta^*$ be the posterior mode when the sample $(y_{t+h}, Y_t)$ is used, with minus the Hessian being denoted by $\Sigma^*$. The Laplace approximation of the marginal predictive likelihood is therefore given by:

$$\log \hat{p}_L(y_{t+h}|Y_t) = \log L(y_{t+h}, Y_t; \theta^*) - \log L(Y_t; \hat{\theta}) + \log p(\theta^*) - \log p(\hat{\theta})$$

$$+ \frac{\log |\hat{\Sigma}| - \log |\Sigma^*|}{2},$$

where the expression takes into account that $\log |A^{-1}| = -\log |A|$ when $A$ has full rank. Gelfand and Dey (1994) refer to (29) as their case (ii) and they note that the approximation has $O(t^{-2})$ accuracy. In other words, the Laplace approximation of the marginal predictive likelihood in (29) is “more accurate” than the Laplace approximation of the marginal likelihood.

Alternatively, the log posterior for the sample $(y_{t+h}, Y_t)$ and its Hessian can be evaluated at the parameter value $\hat{\theta}$ instead of $\theta^*$. This has the advantage that only one posterior mode estimation is required, rather than one plus one for each $y_{t+h}$ ($h = 1, \ldots, H$) that we are interested in. However, the use of $\theta^*$ in (29) ensures that the first derivatives of the log posterior for the sample $(y_{t+h}, Y_t)$ are equal to zero, while the use of $\hat{\theta}$ only ensures that they are, at best, approximately zero. This fact implies an additional error source for the approximation, with the
effect that its accuracy is reduced to $O(t^{-1})$, i.e., the same order of accuracy as the marginal likelihood approximation in (28) has.

Replacing $\theta^\ast$ with $\tilde{\theta}$ in (29) yields the following simplified expression of the log of the marginal predictive likelihood:

$$
\log \tilde{p}_L(y^\ast_{t+h} | \mathcal{Y}_t) = \log L(y^\ast_{t+h}; \mathcal{Y}_t, \tilde{\theta}) + \frac{1}{2} \left( \log |\tilde{\Sigma}_t| - \log |\tilde{\Sigma}_{t+h}| \right),
$$

(30)

where $\tilde{\Sigma}_{t+h}$ and $\tilde{\Sigma}_t$ are minus the Hessians of the log posteriors based on the samples $(y^\ast_{t+h}, \mathcal{Y}_t)$ and $\mathcal{Y}_t$, respectively, evaluated at $\tilde{\theta}$. The first term on the right hand side of (30) is the log of the conditional likelihood, evaluated at the posterior mode using the $\mathcal{Y}_t$ data, and we often expect the marginal predictive likelihood to be dominated by this term. Concerning the second term it may be noted that if the two log determinants are equal, then the Laplace approximation is equal to the $\tilde{\theta}$ plug-in estimator of the predictive likelihood. Moreover, it is straightforward to show that

$$
\tilde{\Sigma}_{t+h} = \tilde{\Sigma}_t - \partial^2 \log L(y^\ast_{t+h}; \mathcal{Y}_t, \tilde{\theta}) / \partial \theta^2. \quad (31)
$$

Hence, we may expect that the second term in (30) is often close to zero. Furthermore, the overall computational cost when using (30) is not reduced by taking the expression on the right hand side of (31) into account unless analytical derivatives are available; see, e.g., Appendix A.

For models with a normal likelihood, the log of the conditional likelihood is given by

$$
\log L(y^\ast_{t+h}; \mathcal{Y}_t, \theta) = -\frac{n^\ast}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_{y^\ast, t+h \mid t}| - \frac{1}{2} \left( y^\ast_{t+h} - y^\ast_{t+h \mid t} \right)' \Sigma_{y^\ast, t+h \mid t}^{-1} \left( y^\ast_{t+h} - y^\ast_{t+h \mid t} \right),
$$

(32)

where $n^\ast$ is the dimension of $y^\ast_{t+h}$, $y^\ast_t = K' y_t$ and $\Sigma_{y^\ast, t+h \mid t} = K' \Sigma_{y^\ast, t+h \mid t} K$, with $K$ being an $n \times n^\ast$ 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 $n^\ast$. For DSGE models the Kalman filter provides us with

$$
y_{t+h \mid t} = \mu + H' F^h \tilde{\xi}_{t \mid t},
$$

$$
\Sigma_{y, t+h \mid t} = H' P_{t+h \mid t \mid H} + R,
$$

$$
P_{t+h \mid t} = FP_{t+h-1 \mid t} F' + BB', \quad h = 1, \ldots, H.
$$

where $\tilde{\xi}_{t \mid t}$ is the filter estimate of the state variables, and $P_{t \mid t}$ the corresponding filter estimate of the state variable covariance matrix based on the data $\mathcal{Y}_t$. The matrices $(\mu, H, R, F, B)$ are all evaluated at the posterior mode of $\theta$.\textsuperscript{10}

\textsuperscript{10} 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; see, e.g., Durbin and Koopman (2001).
Turning to the DSGE-VARs, we can directly make use of equations (18) and (21) to determine \( y_{t+h|t} \) and \( \Sigma_{y,t+h|t} \). That is,

\[
y_{t+h|t} = J_p' \bar{\xi}_{t+h} + J_p' \Psi^h Y_t,
\]

\[
\Sigma_{y,t+h|t} = J_p' \Sigma^{(h)}_{\Phi^h} J_p, \quad h = 1, \ldots, H,
\]

where the \( (\Phi, \Sigma, \theta) \) parameters are evaluated at the posterior mode. As discussed in Section 3, we may either consider a marginalized likelihood or a concentrated likelihood when estimating the posterior mode of \( \theta \) for DSGE-VARs. In our empirical illustration we will focus on the marginalized posterior mode estimate when computing the predictive likelihood via the Laplace approximation.

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. However, the normal approximation requires posterior draws and is therefore computationally more expensive than the Laplace approximation. On the other hand, the normal approximation uses population moments, while the Laplace approximation is based on population moments for fixed parameters. In our empirical illustration we shall make use of both approximations. First, this allows us examine if they differ substantially in practise and, second, makes it possible to directly compare our results with CCW. Since we are interested in how reliable the approximations are for computing the marginal predictive likelihood, we will compare them to the modified harmonic mean estimator.

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; 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. 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.

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 euros, 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 im-
portant in generating empirically plausible dynamics. Real frictions are introduced via external
habit formation in consumption and through generalized adjustment costs in investment, im-
ports 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 stabilising 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 a small set of 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 its construction). The estimation sample is given by
the period 1985Q1 until 2006Q4, with 1980Q2-1984Q4 serving as training sample. The five
variables foreign demand, foreign prices, the foreign interest rate, competitors’ export prices,
and oil prices are modelled using an SVAR, where the estimated parameters are kept fixed
throughout the estimation of the NAWN. 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. 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) adopts the empirical approach outlined in Smets and Wouters (2003) and An and Schorfheide (2007) and estimate the NAWM with Bayesian inference methods. This involves obtaining the posterior distribution of the model’s parameters based on its log-linear state-space representation using the Kalman filter. For the empirical analysis we follow CCW and use YADA, a Matlab program for Bayesian estimation and evaluation of DSGE and DSGE-VAR models; see Warne (2012) for details.

The DSGE-VAR models have been estimated over the same sample as the NAWM using the random-walk Metropolis algorithm with a Gaussian proposal density; see, e.g., An and Schorfheide (2007). We use the parameter transformations \( \phi = g(\theta) \) discussed in, for instance, Warne (2012, Section 6) such that the domain of \( \phi \) 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. We make use of the same covariance matrix as in Christoffel et al. (2008) for the proposal density, but fine-tune the matrix via a positive constant, as discussed by for example An and Schorfheide, for each pair \((\lambda, p)\) such that the acceptance rate is about 25-30 percent. Each posterior draw of \( \phi \) is thereafter transformed to \( \theta \) via \( \theta = g^{-1}(\phi) \). Using 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, \( p(Y_T|Y_1, \lambda) \), is calculated with the modified harmonic mean estimator based on the truncated normal density; see Geweke (1999, 2005).
In our empirical examination of DSGE-VARs we will cover two approaches to model selection. The first simply picks the model with the largest marginal likelihood for \((\lambda, p)\), while the second picks \(p\) such that the marginal likelihood of the DSGE-VAR(\(\infty\)) model is close to the marginal likelihood of the DSGE model and then selects \(\lambda\) 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 80 units or greater. Compared with Del Negro et al. (2007), who use a model with seven observed variables, the increase 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.\(^{11}\) 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 optimal value of \(\hat{\lambda} = 1.25\) for the one lag models is lower but close to the DSGE-VAR(\(\infty\)) models with a higher lag order.

From Figure 2 it can be seen that the posterior mode estimates of \(\lambda\) 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 \(\lambda\) 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 focus on the optimal DSGE-VAR models for two and four lags.

### 7. Comparing Forecast Accuracy

The forecast performance of the DSGE-VAR models along with the NAWM, a large BVAR model, and the random-walk will be assessed below using the same procedure as in CCW. The random-walk model is based on a standard diffuse prior on the covariance matrix of the innovation, while the BVAR model is estimated using the methodology in Bańbura et al. (2010). Appendix A provides details on the posterior properties of the random-walk, while Appendix B

---

\(^{11}\) 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.
gives more information about the BVAR.\textsuperscript{12} The parameters of all models except the random-walk 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 some quarter $i = 1, 2, 3$, the models are estimated using data until $T - i$. Hence, most models are re-estimated annually. For the random-walk model, the covariance matrix is re-estimated 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 length of the maximum forecast horizon is 8 quarters, yielding 32 observations of the one-step-ahead forecasts and 25 of the eight-step-ahead forecasts. Unlike CCW, we shall only consider forecasts of quarterly growth-rates for variables in first differences. 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.\textsuperscript{13}

\section*{7.1. Point Forecasts}

Figure 3 displays the root mean squared forecast errors (RMSE) when forecasting quarterly growth rates of the variables in first differences. The point forecasts from the NAWM and the two DSGE-VAR models are computed as the estimated population mean based on 500 draws of the model parameters from the posterior distribution, where each draw is separated by 1000 draws from the next. For the BVAR we use 500 parameter draws and 500 simulated paths for each parameter draw and compute the mean forecasts from this sample of 250,000 paths from its predictive distribution.\textsuperscript{14}

To facilitate the comparisons with the multivariate point forecast analysis, the RMSEs in Figure 3 have been scaled them with the estimated standard deviations of the variables over the period 1995Q1-2006Q4. Like in CCW, the number of variables is equal to 12 and are the same as the ones we shall consider both for the multivariate point forecasts and the density forecasts. The remaining 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

\textsuperscript{12} While CCW consider additional reduced-form models in their comparison, these models were out-performed by the large BVAR and have therefore been excluded from the analysis.

\textsuperscript{13} An interesting approach to Bayesian forecast evaluation in a multivariate setting is suggested by Herbst and Schorfheide (2011). 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, 2007) 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 different models.

\textsuperscript{14} We have also simulated 500 forecast paths for each parameter draw for the NAWM and the DSGE-VAR models. The difference between the mean forecasts from the 250,000 paths and the estimated population mean using 500 parameter draws is negligible.
and they are therefore excluded from the analysis; see footnote 4 in CCW for further discussion on this issue.

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 where the NAWM fares less well are nominal wages in particular, but also the GDP deflator, the private consumption deflator and private consumption. Given the variables that the NAWM forecasts poorly, it is of particular interest to examine if the DSGE-VAR models fare better, while at the same time not loosing ground on the variables that the NAWM forecasts comparatively well. Moreover, we would like to know if 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 (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. What is more noticeable is the improved forecast performance of these models relative to the NAWM 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, 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.

Looking closer at the forecast errors, Table 1 shows the mean errors of the NAWM and the DSGE-VAR models. Overall, the mean forecast errors are negative (or close to zero) for private consumption and nominal wages and positive for the private consumption and the GDP deflator. Moreover, it can also be inferred that the mean errors for real wages are negative. The tendency of the NAWM to overpredict private consumption and real wages was discussed in some detail in CCW. 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. Figure 4 displays the individual mean forecast paths for the five models along with the data for quarterly private consumption and nominal wages.
Table 2 provides the percentage share of squared mean errors to the mean squared errors for the same four variables. 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 may be 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.

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

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

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 forecast accuracy. The choice of scaling matrix for the forecast errors has a direct impact on the ranking of forecasting models when using the trace statistic, while the ranking is invariant to the choice of $M$ for the log determinant statistic. Furthermore, rankings based on the MSE matrix in (33) are not invariant to non-singular, scale-preserving linear transformations of the variables, while the class of (linear) models is invariant to such isomorphic transformations. This is due to forecast errors being correlated over forecast horizons; see Clements and Hendry (1993). As a consequence, the five models in this paper can have different rankings for forecasts of quarterly and annual log growth rates. In addition, it should be kept in mind that the trace and log determinant statistics are functions of the eigenvalues of the MSE-matrix, where the largest eigenvalues gives the MSE of the least predictable linear combinations of the variables and the smallest the most predictable. Since the trace is equal to the sum of the eigenvalues it follows that this statistic tends to be dominated by the largest eigenvalues (least predictable linear combinations), while the determinant is the product of the eigenvalues and is therefore also influenced by the smallest (most predictable linear combinations) and may even be dominated by them.

As in CCW, the MSE statistics are calculated for three cases. First, all the 12 variables displayed in Figure 3 are used, while the second case focuses on the seven variables 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, the case which may be regarded as being the minimum set of variables relevant to monetary policy is examined. This set 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 statistic is the sum of the squared scaled RMSEs. Given the univariate analysis it is therefore not surprising that
the DSGE-VARs have similar results and perform very well for all three cases, over all forecast horizons.

The log determinant statistic is invariant to the choice of scaling matrix when ranking forecasting models and is displayed in Figure 6. The BVAR model tends to outperform the others for the 12 and 7 variable cases, while the NAWM seems to rank at the top for the 3 variable case. 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-VARs also follow the NAWM closely.

7.2. Density Forecasts

Table 3 shows the share of average parameter uncertainty to average forecast uncertainty for the NAWM along with the DSGE-VAR models for selected horizons. For each estimate of the covariance matrix of the predictive distribution for the \( h \)-step-ahead forecasts we use the population expression in equation (20) for the DSGE-VAR models and equation (15) in CCW for the NAWM and the same 500 parameter draws as were used for the point forecasts in Section 7.1. As expected, parameter uncertainty plays a more important role for the DSGE-VAR models than for the NAWM. Overall, parameter uncertainty plays a very limited role for the NAWM where the highest share is obtained for employment at three percent for eight-quarter-ahead forecasts. For the DSGE-VAR model with two lags the shares reach a maximum of about 19 percent for employment at the same forecast horizon. For the variables in first differences the shares are generally decreasing with the forecast horizon with values around 12-13 percent for the short-term forecasts. Also, the average share of parameter uncertainty at the shorter horizons tend to be lower for the two lag than the four lag model, while the opposite is the case at longer horizons. Still, the differences between the DSGE-VAR models are not huge.

Table 4 reports the log predictive score for all the 18 variables in the NAWM using the three methods for computing the predictive likelihood and for one-step-ahead forecasts only. This corresponds to using equation (24) for \( h = 1 \) under the Laplace approximation and the modified harmonic mean.\(^{15}\) It follows from this equation that the log predictive score is equal to the difference between the log marginal likelihood for the full sample and for the first historical sample, i.e., using data until 1998Q4. In addition, the Laplace approximation is here based on two posterior mode estimates and the statistic is therefore calculated as in equation (29). The normal approximation of the log predictive score instead uses equation (23) since there are not any marginal likelihood values available under this method.

It is striking from the Table that the Laplace approximation values are very close to those of the modified harmonic mean for the NAWM and the two DSGE-VAR models. For the random-walk model this “Laplace” error is large and depends on the number of variables and the number

\(^{15}\) The modified harmonic mean estimator of the log marginal likelihood has been calculated using all 500,000 post burn-in posterior draws, where the burn-in sample was set as the first 50,000 draws. For the BVAR and random-walk models, the log marginal likelihoods are computed from their analytical expressions; see equations (A.5) and (B.8), respectively.
of observations; see equation (A.19) in Appendix A. Moreover, it should be kept in mind that for this model the predictive density is multivariate $t$ with $T - n + n^*$ degrees of freedom and the log predictive score can therefore be computed from an analytical expression.

If the posterior mode estimate based on the sample until 1998Q4 is used when evaluating the log marginal likelihood for the sample until 2006Q4, the value of the log predictive score for the NAWM decreases by about 50 units to $-650.2$. Similarly, for the DSGE-VAR models with two and four lags the log predictive scores drop by about 50 and 30 units, respectively, with log scores of $-600.3$ for the models with two lags and to $-590.3$ for the model with four lags. This underlines that although the Laplace approximation may provide an accurate estimate of the log marginal likelihood when applied properly, it is important to use the relevant posterior mode estimate when using the shortcut in equation (24) for the joint predictive likelihood.

It is also interesting to note that the normal approximation departs less from the modified harmonic mean for the NAWM than for the DSGE-VAR models. One explanation for this may be related to the parameter uncertainty results discussed at the beginning of this Section. The unconditional predictive density may be seen as a “weighted average” of the predictive density conditional on the parameters and the posterior of the parameters. Since the predictive density conditional on the parameters is normal we may expect the unconditional predictive density to be better approximated by a normal when the scale of the posterior density is very small compared with the scale of the predictive density conditional on the parameters.

The values of the log predictive score when using the modified harmonic mean for the 12, 7, and 3 variables cases for the five models are reported in Table 5. 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. These draws have been selected as draw number 1, 51, 101, ..., 499951 to combine low computational costs with a low correlation between the draws and a sufficiently high estimation accuracy. This procedure here yields estimates of the log marginal likelihood that are accurate up to and including the first decimal relative to using all post burn-in draws.

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 the shorter horizons, while the NAWM gets higher values at the longer horizons when examining the 12 and 3 variable cases. In the 7 variable case, the DSGE-VAR with two lags has a higher log score than the NAWM at all examined forecast horizons, but the difference is diminishing with the horizon.

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 in the 3 variable case. As the forecast horizon increases, however, the random-walk model’s performance worsens in comparison with these alternatives.

[Comparison with BVAR]
In Figure 7 the log predictive likelihoods of the NAWM for the 12 variable case is displayed using the modified harmonic mean (solid red line), the normal approximation (dash-dotted green line), and the Laplace approximation (dashed blue line). It is striking how closely the approximation methods follow the modified harmonic mean estimate. It may be noted that this is not an artefact of the 12 variable case but is also found for the 3 and 7 variable cases. When adding up the individual log predictive likelihood estimates for the three methods, the normal approximation is always closer to the modified harmonic mean than the Laplace approximation. Moreover, both approximations over-estimate the value of the log score. For example, the log score using the normal and Laplace approximations are about 11 units greater than the log score under the modified harmonic mean for one-step-ahead forecasts. At the four-step-ahead horizon, the normal approximation has a log score which is about 7 units larger than the modified harmonic mean based score, while the Laplace approximation is roughly 10 units larger. As the horizon increases, these differences appear to become smaller with the normal approximation approaching the modified harmonic mean at a faster rate than the Laplace approximation.

Turning to the log predictive likelihoods in the 12 variable case for the DSGE-VAR with two lags in Figure 8, it is notable that the differences between the approximations and the modified harmonic mean are substantially larger than for the NAWM and that both approximations over-estimate the log predictive likelihood. When accumulating the one-step-ahead log predictive likelihoods, the difference between the normal approximation and the modified harmonic mean is now about 70 units, while the Laplace approximation over-estimates the log score by about 80 units at this horizon. It should be kept in mind that the Laplace approximation of the log predictive likelihood is here based on a single posterior mode estimate per period rather than two such estimates and that this additional source of error may explain at least part of its poor performance when applied to the DSGE-VAR model.

The 3 and 7 variable cases for the log predictive likelihoods show similar behavior as for the 12 variables and similar results are also obtained for the four lag DSGE-VAR model. It may be noted, however, that the approximation errors are somewhat smaller when using the four lag model than the two lag model.

The normal approximation provides more accurate estimates of the log predictive likelihood than the Laplace approximation for the DSGE-VAR models, although not as good as for the NAWM. For the random-walk model, the normal approximation is very accurate and underestimates the log score by about 8-10 units. For this model we know that the predictive density is multivariate $t$ with $T - n + n^*$ degrees of freedom. Hence, it is not surprising that the normal approximation works quite well.

Figures 9–11 display the log scores for the different forecast horizons and cases using the three estimation methods. A common feature of the approximation methods is that they over-estimate the differences between the scores for the DSGE-VAR models and the NAWM, where these errors are particularly important for the Laplace approximation. However, when comparing the two
DSGE-VAR models the overall picture for all estimation methods is that the two lag model has a somewhat higher log score than the four lag model for most comparisons.

Let us finally turn to the log score of the BVAR model. In the case of the normal approximation this model achieves the highest score for all systems and horizons. Compared with the DSGE-VAR models the difference for the 12 variable case is at least 30 units for all horizons. [Discuss Laplace and modified harmonic mean.]

8. **Summary and Conclusions**

[TBW]
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. These properties are used when computing the predictive likelihood using the normal approximation, the Laplace approximation, and an analytical expression with this model.

To this end, let
\[ y_t = y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \]  
(A.1)
where the residuals \( \varepsilon_t \) are assumed to be i.i.d. Gaussian with zero mean and positive definite covariance matrix \( \Omega \) and \( y_0 \) is fixed. The diffuse prior is given by
\[ p(\Omega) \propto |\Omega|^{-(n+1)/2}. \]  
(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
\[ \log L(y; y_0, \Omega) + \log p(\Omega) = -\frac{nT}{2} \log(2\pi) - \frac{T + n + 1}{2} \log |\Omega| - \frac{1}{2} \text{tr}[\Omega^{-1} \varepsilon \varepsilon'] \]  
(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
\[ \log p(\Omega|y, y_0) = -\frac{nT}{2} \log(2) - \frac{n(n-1)}{4} \log(\pi) - \log \Gamma_n(T) + \frac{T}{2} \log |\varepsilon \varepsilon'| \]  
(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.,
\[ \log p(y|y_0) = -\frac{n(2T - n + 1)}{4} \log(\pi) + \log \Gamma_n(T) - \frac{T}{2} \log |\varepsilon \varepsilon'|. \]  
(A.5)

Normal Approximation of Marginal Predictive Likelihood

When forecasting with the random-walk model it holds that
\[ E[y_{T+h}|y, y_0, \Omega] = y_T, \quad h = 1, \ldots, H. \]  
(A.6)
The forecast error is therefore equal to the accumulation of \( \varepsilon_{T+i} \) over \( i = 1, \ldots, h \), while the forecast error covariance matrix given \( \Omega \) is
\[ C(y_{T+h}|y, y_0, \Omega) = h\Omega, \quad h = 1, \ldots, H. \]  
(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 $\Omega$ plus the covariance matrix of the deviation of the mean in (A.6) and its population mean $E[y_{T+h}|y, y_0]$. 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 $\Omega$.\footnote{More generally, the posterior distribution of $h\Omega$ is inverted Wishart with location parameter $h\varepsilon\varepsilon'$ and $T$ degrees of freedom.} That is,

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

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 that selects $y_t^* = K'y_t$. Similarly, let $K_{\perp}$ the the $n \times (n - n^*)$ matrix which selects the remaining variables from the $y_t$ vector. Let

$$M = [K \quad K_{\perp}], \quad \text{(A.9)}$$

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. With $\Omega_M = M'\Omega M$ and the Jacobian of this transformation having unit determinant, it follows that $p(\Omega_M|y, y_0) = p(M'\Omega_M|y, y_0)$. In other words, the posterior distribution of $\Omega_M$ is an $n$-dimensional inverted Wishart with location matrix $M'\varepsilon\varepsilon'M$ and $T$ degrees of freedom. Letting $\Omega_M = K'\Omega K$, it follows from, e.g., Bauwens, Lubrano, and Richard (1999, Theorem A.17) that the posterior of $\Omega_M$ 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

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

**Analytical Form of the Marginal Predictive Likelihood**

The remaining two approaches to computing the marginal predictive likelihood require an expression for the conditional likelihood function $L(y_{T+h}^*; y, y_0, \Omega)$. From equation (32) 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

$$\log L(y_{T+h}^*; y, y_0, \Omega) = -\frac{n^*}{2} \log(2\pi h) - \frac{1}{2} \log |\Omega_K| - \frac{1}{2h} \text{tr}[\Omega_K^{-1}\varepsilon_h\varepsilon_h'K'], \quad \text{(A.11)}$$
where \( \varepsilon_h^* = y_{T+h}^* - y_T^* \), and the term involving \( \log(h) \) is due to \( |h\Omega_K| = h^{n^*} |\Omega_K| \). To determine an analytical expression for the marginal predictive likelihood it remains to multiply the conditional likelihood with the posterior of \( \Omega_K \) and integrate out \( \Omega_K \) from the expression.

More generally, the product of the marginal predictive density of \( y_{T+h}^* \) for fixed \( \Omega_K \) and the posterior of \( \Omega_K \) can be shown to be given by:

\[
p(y_{T+h}^*, \Omega_K | y, y_0) = \frac{|K'\varepsilon \varepsilon'K|^{(T-n+n^*)/2}}{(2\pi h)^{n^*/2}2^{(T-n+n^*)n^*+2}n^{(n^*-1)/4}\Gamma_{n^*}(T-n+n^*)} \times \frac{1}{|\Omega_K|^{-(T-n+2n^*+2)/2}} \exp \left[ -\frac{1}{2} \text{tr} \left( \Omega_K^{-1} \left[ K'\varepsilon \varepsilon'K + h^{-1}\varepsilon_h^* \varepsilon_h''' \right] \right) \right].
\]

(A.12)

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

\[
p(y_{T+h}^* | y, y_0) = \frac{\Gamma_{n^*}(T-n+n^*+1) |hK'\varepsilon \varepsilon'K|^{-1/2}}{\pi^{n^*/2}\Gamma_{n^*}(T-n+n^*) |I_{n^*} + (hK'\varepsilon \varepsilon'K)^{-1}\varepsilon_h^* \varepsilon_h'''|^{(T-n+n^*+1)/2}}.
\]

(A.13)

In other words (and as expected), the density of \( y_{T+h}^* | y, y_0 \) 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.\(^{17}\)

For numerical reasons it may be advised to multiply \( hK'\varepsilon \varepsilon'K \) by, say, \( T^{-1} \). To compensate for this we also need to multiply \( \varepsilon_h^* \varepsilon_h''' \) by \( T^{-1} \), and the denominator by \( T^{n^*/2} \).

**Laplace Approximation of Marginal Predictive Likelihood**

The Laplace approximation of the marginal predictive likelihood in equation (30) requires the posterior mode of \( \Omega \) for the historical sample. It is straightforward to show that it is given by

\[
\tilde{\Omega} = \frac{1}{T + n + 1} \varepsilon \varepsilon'. \tag{A.14}
\]

Letting \text{vech} denote the column stacking operator which only takes the values on and below the diagonal of a square matrix, while \( D_n \) is the \( n^2 \times n(n+1)/2 \) duplication matrix (see Magnus and Neudecker, 1988), it can be shown that minus the Hessian at the mode is given by

\[
- \frac{\partial^2 f(\Omega)}{\partial \text{vech}(\Omega) \partial \text{vech}(\Omega)^T} = \frac{T + n + 1}{2} D_n^T (\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) D_n, \tag{A.15}
\]

where \( f(\Omega) \) is equal to the sum of the log-likelihood and the log-prior in equation (A.3). The \( n(n+1)/2 \) matrix with second partial derivatives of minus the conditional likelihood in (A.11)

\(^{17}\) Notice also that \( |I_{n^*} + (hK'\varepsilon \varepsilon'K)^{-1}\varepsilon_h^* \varepsilon_h'''| = 1 + \varepsilon_h^* (hK'\varepsilon \varepsilon'K)^{-1}\varepsilon_h^* \); see, e.g., Magnus and Neudecker (1988, Proof of Theorem 1.9).
with respect to the unique elements of \( \Omega \), evaluated at \( \tilde{\Omega} \), can be shown to be given by

\[
- \frac{\partial^2 \log L(y_{T+h}^t; y_0, \Omega)}{\partial \text{vech}(\Omega) \partial \text{vech}(\tilde{\Omega})} = \frac{1}{2} D_n \left[ K \tilde{\Omega}^{-1} K' \otimes K \left( (2/h) \tilde{\Omega}_K^{-1} \varepsilon_h' \varepsilon_h + I_n \right) \tilde{\Omega}_K^{-1} K' \right] D_n, \tag{A.16}
\]

where \( \tilde{\Omega}_K = K' \tilde{\Omega} K \). Notice that if the first order partial derivatives of the sum of (A.3) and (A.11) are zero at \( \Omega = \tilde{\Omega} \), then \( \tilde{\Omega}_K^{-1} \varepsilon_h \varepsilon_h' / h = I_n \) and the right hand side of (A.16) simplifies to

\[
D_n'(K\tilde{\Omega}_K^{-1} K' \otimes K\tilde{\Omega}_K^{-1} K') D_n.
\]

The Laplace approximation of the marginal predictive likelihood is now obtained by using equations (A.11), (A.15), and (A.16) in equation (30), with \( \tilde{\Sigma}_{T+h} \) being given by the sum of (A.15) and (A.16). That is,

\[
\log \tilde{p}_L(y_{T+h} | y_0) = -\frac{n}{2} \log(2\pi h) - \frac{1}{2} \log |\tilde{\Omega}_K| - \frac{1}{2h} \text{tr}[\tilde{\Omega}_K^{-1} \varepsilon_h \varepsilon_h'] + \frac{1}{2} \log |D_n'(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) D_n| - \frac{1}{2} \log |D_n'(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) D_n| + \frac{1}{T_n} D_n \left[ K \tilde{\Omega}_K^{-1} K' \otimes K \left( \frac{2}{h} \tilde{\Omega}_K^{-1} \varepsilon_h \varepsilon_h' - I_n \right) \tilde{\Omega}_K^{-1} K' \right] D_n, \tag{A.17}
\]

where \( T_n = T + n + 1 \). We can also determine an analytical expression for the Laplace approximation of the log marginal likelihood. Substituting for (A.3) and (A.16) in (28), we obtain

\[
\log \hat{p}_L(y | y_0) = -\frac{n(2T - n - 1)}{4} \log(2\pi) - \frac{T_n}{2} \log |\tilde{\Omega}| - \frac{n(n + 1)}{4} \log(T_n/2) - \frac{nT_n}{2} - \frac{1}{2} \log |D_n'(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) D_n|.
\]

By making use of Theorem 3.14 in Magnus and Neudecker (1988), it can be shown that

\[
\log |D_n'(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) D_n| = \frac{n(n-1)}{2} \log(2) - (n+1) \log |\tilde{\Omega}|,
\]

where \( \log |\tilde{\Omega}| = -n \log(T_n) + \log |\varepsilon'| \). This means that

\[
\log \hat{p}_L(y | y_0) = -\frac{n(2T - n + 1)}{4} \log(\pi) + \frac{n(2T - n - 1)}{4} \log(T_n/2) + \frac{n}{2} \log(2\pi) - \frac{nT_n}{2} - \frac{T}{2} \log |\varepsilon'|. \tag{A.18}
\]

The approximation error from using the Laplace approximation of the log marginal likelihood is therefore equal to

\[
\log \hat{p}_L(y | y_0) - \log p(y | y_0) = \frac{n(2T - n - 1)}{4} \log(T_n/2) + \frac{n}{2} \log(2\pi) - \frac{nT_n}{2} - \log \Gamma_n(T). \tag{A.19}
\]

It is noteworthy that the approximation error tends to be negative such that the Laplace approximation underestimates the log marginal likelihood. Moreover, for large models and conventional sample sizes the error is huge. For example, a model with \( n = 18 \) and \( T = 100 \), the approximation error is about \(-34.8\), while \( T = 100,000 \) gives an error equal to \(-0.038\). The errors are considerably smaller for the Laplace approximation of the predictive likelihood.
Appendix B: Posterior Properties of the Large BVAR Model

The large BVAR is estimated with the methodology suggested in Banbura 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 \( T_d \) dummy observations; see also Lubik and Schorfheide (2006).

The VAR representation of \( y_t \) is given in equation (6), 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

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

The normal-inverted Wishart prior for \((\Phi, \Omega)\) is given by

\[
\text{vec}(\Phi)|\Omega \sim N_{n(np+1)}(\text{vec}(\Phi_\mu), [\Omega_\Phi \otimes \Omega]), \quad \Omega \sim IW_n(A, v).
\]  
(B.2) 
(B.3)

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

\[
\log f(y, \Phi, \Omega|X_1) = -\frac{n(T + np + 1)}{2} \log(2\pi) - \frac{n}{2} \log(2) - \frac{n(n-1)}{4} \log(\pi)
- \log \Gamma_n(v) - \frac{n}{2} \log |\Omega_\Phi| + \frac{v}{2} \log |A| - \frac{T + n(p + 1) + v + 2}{2} \log |\Omega| \]  
(B.4)

Using standard “Zellner” algebra, it is straightforward to show that

\[
\epsilon' + A + (\Phi - \Phi_\mu)\Omega_\Phi^{-1}(\Phi - \Phi_\mu)' = (\Phi - \tilde{\Phi}) (XX' + \Omega_\Phi^{-1}) (\Phi - \tilde{\Phi})' + S,
\]  
(B.5)

where

\[
\tilde{\Phi} = (yX' + \Phi_\mu\Omega_\Phi^{-1}) (XX' + \Omega_\Phi^{-1})^{-1},
\]

\[
S = yy' + A + \Phi_\mu\Omega_\Phi^{-1}\Phi_\mu - \tilde{\Phi} (XX' + \Omega_\Phi^{-1}) \tilde{\Phi}'.
\]

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

\[
\text{vec}(\Phi)|y, X_1 \sim N_{n(np+1)}(\text{vec}(\tilde{\Phi}), [(XX' + \Omega_\Phi^{-1})^{-1} \otimes \Omega]), \quad \Omega|y, X_1 \sim IW_n(S, T + v).
\]  
(B.6) 
(B.7)

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

\[
\log p(y|X_1) = -\frac{nT}{2} \log(\pi) + \log \Gamma_n(T + v) - \log \Gamma_n(v) - \frac{n}{2} \log |\Omega_\Phi|
+ \frac{v}{2} \log |A| - \frac{n}{2} \log |XX' + \Omega^{-1}| - \frac{T + v}{2} \log |S|.
\]  
(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:

\[
y(d) = \begin{bmatrix} \lambda_0^{-1} & \text{diag}[\delta \odot \omega] & 0_{n \times np(p-1)} & \text{diag}[\omega] & 0_{n \times 1} & \tau^{-1} \text{diag}[\delta \odot \mu] \end{bmatrix}
\]

\[
X(d) = \begin{bmatrix}
0_{1 \times np} & 0_{1 \times n} & \varsigma & 0_{1 \times n} \\
\lambda_0^{-1}(j_p \odot \text{diag}[\omega]) & 0_{np \times n} & 0_{np \times 1} & \tau^{-1}(i_p \odot \text{diag}[\mu])
\end{bmatrix}.
\]

(B.9)

The vector \( i_p \) is a \( p \)-dimensional 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 \( \tau \) becomes larger. The \( n \)-dimensional vector \( \delta \) gives the prior mean of the diagonal of \( \Phi_1 \), \( \omega \) is a vector of scale parameters for the residuals \( \epsilon_{it} \), while \( \mu \) is a vector that reflects the mean of \( y_{it} \). Finally, \( \varsigma \) is a very small number which takes care of having an improper prior on \( \Phi_0 \).

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 differenced variable and \( \delta_i = 1 \) when \( y_{it} \) is a levels variable. The scale parameters \( \omega_i \) is given by the within-sample residual standard deviation from an AR\((p)\) model for \( y_{it} \), while \( \mu_i \) is equal to the within-sample mean of \( y_{it} \).

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

\[
\tilde{\lambda}_o(\phi) = \arg \min_{\lambda_o} \left| \phi - \frac{1}{q} \sum_{j=1}^q \frac{\sigma^2_j(\lambda_o)}{\sigma^2_j(0)} \right|,
\]

where \( \phi \in (0, 1) \) is the desired fit, and \( \sigma^2_j(\lambda_o) \) 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 \( \tilde{\lambda}_o = 0.0693 \) when \( p = 4 \).

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

\[
\Phi_\mu = y(d)X'(d) \left( X(d)X'(d) \right)^{-1}, \quad \Omega_\Phi = \left( X(d)X'(d) \right)^{-1},
\]

\[
A = (y(d) - \Phi_\mu X(d)) (y(d) - \Phi_\mu X(d))', \quad v = T_d - (np + 1) + 2.
\]
This guarantees that the prior mean of $\Omega$ exists. Letting $y_* = [y(d) \ y]$ and $X_* = [X(d) \ X]$, it follows that the posterior parameters

$$
\bar{\Phi} = y_* X'_* (X_* X'_*)^{-1}, \\
XX' + \Omega^{-1}_\Phi = X_* X'_*, \\
S = (y_* - \bar{\Phi} X_*) (y_* - \bar{\Phi} X_*)'.
$$

**Normal Approximation of Marginal Predictive Likelihood**

Conditional on the parameters and the historical data, the expected value of $y_{T+h}$ is obtained from (18) as

$$
E[y_{T+h}|y, X_1, \Phi, \Omega] = J'_p \bar{x}_{T+h} + J'_p \Psi Y_T, \quad h = 1, \ldots, H.
$$

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

$$
C(y_{T+h}|y, X_1, \Phi, \Omega) = J'_p \bar{\Sigma}^{(h)} J_p, \quad h = 1, \ldots, H,
$$

where $\bar{\Sigma}^{(h)}$ is given by equation (21) with $\Omega$ replacing $\Sigma$. The mean of the marginal predictive density of $y_{T+h}$ is computed from (B.10) by taking the expectation with respect to the posterior of $(\Phi, \Omega)$. This means that

$$
E[y_{T+h}|y, X_1] = E_T \left[ J'_p \bar{x}_{T+h} + J'_p \Psi Y_T \right], \quad h = 1, \ldots, H.
$$

Next, through Rao-Blackwellization, the covariance matrix of the marginal predictive density is equal to the expected value of (B.11) with respect to the posterior of the parameters plus the covariance of (B.10) with respect to the posterior of the parameters, i.e.,

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

The normal approximation of the marginal predictive likelihood is now based on using (B.12) as mean and (B.13) as covariance. For subsets of variables we simply take the corresponding elements of the mean and covariances.

**Laplace Approximation of Marginal Predictive Likelihood**

The Laplace approximation of the marginal predictive likelihood in (30) requires that the joint posterior mode of $(\Phi, \Omega)$ for the historical sample is available. It is straightforward to show that this joint mode is given by:

$$
\tilde{\Phi} = \bar{\Phi}, \quad \tilde{\Omega} = \frac{1}{T_{np}} S,
$$

where $T_{np} = T + n(p + 1) + v + 2$. Next, it can be shown that for the parameterization $\beta = [\text{vec}(\Phi)']' \text{vech}(\Omega)'$, minus the Hessian of the log posterior evaluated at the mode is equal to

$$
\tilde{\Sigma}_T = \begin{bmatrix} X_* X'_* \otimes \tilde{\Omega}^{-1} & 0 \\
0 & \frac{T_{np}}{2} D_n' \left[ \tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1} \right] D_n \end{bmatrix}.
$$
Equipped with the minus the Hessian at the mode and the sum of the log-likelihood and the log prior in (B.4), it is straightforward to calculate the value of the log marginal likelihood with the Laplace approximation in equation (28). Specifically, we find that
\[
\log |\Sigma_T| = n \log |X_* X_*'| - (np + 1) \log |\tilde{\Omega}| + \frac{n(n + 1)}{2} \log(T_{np}/2)
+ \frac{n(n - 1)}{2} \log(2) - (n + 1) \log |\tilde{\Omega}|.
\]
Hence,
\[
\log \hat{p}_L(y|X_1) = -\frac{nT}{2} \log(\pi) - \log \Gamma_n(v) + \frac{n(2(T + v) - n - 1)}{4} \log(T_{np}/2) - \frac{nT_{np}}{2} - \frac{n}{2} \log |\Omega_\Phi| + \frac{v}{2} \log |A| - \frac{n}{2} \log |X_* X_*'| - \frac{T + v}{2} \log |S|,
\]
where we have used log $|\tilde{\Omega}| = -n \log(T_{np}) + \log |S|$. From (B.8) it follows that the approximation error is equal to
\[
\log \hat{p}_L(y|X_1) - \log p(y|X_1) = \frac{n(2(T + v) - n - 1)}{4} \log(T_{np}/2) - \frac{nT_{np}}{2} - \log \Gamma_n(T + v).
\]

The conditional likelihood of the BVAR is given by (32) with $y_{T+h|T}$ given by the right hand side of (B.10) and $\Sigma_{y,T+h|T}$ equal to (B.11). To compute the Laplace approximation of the marginal predictive likelihood it remains to determine the $\tilde{\Omega}_{T+h|T}$ matrix in (31). After making considerable use of the matrix differential calculus tricks in Magnus and Neudecker (1988) it can be shown that for $L_h^* = L(y_{T+h}; y, X_1, \Phi, \Omega)$ the following monster appears for $h = 1, 2, \ldots, H$:
\[
-\frac{\partial^2 \log L_h^*}{\partial \text{vec}(\Phi)^2} = \frac{1}{2} \left( \frac{\partial \text{vec}(\Sigma_Y)}{\partial \text{vec}(\Phi)} \right)' \left[ J_p K \Sigma_K^{-1} K' J_p \otimes J_p K \left( 2 \Sigma_K^{-1} \epsilon_h \epsilon_h' - I_n \right) \right]
\times \Sigma_K^{-1} K' J_p \left. \frac{\partial \text{vec}(\Sigma_Y)}{\partial \text{vec}(\Phi)} \right] + \sum_{i=0}^{h-2} M' \left[ \Sigma_Y^{(h-1-i)} \otimes G_{i,h} \right] M
+ \left( \frac{\partial \text{vec}(\Sigma_Y^{(h-1-i)})}{\partial \text{vec}(\Phi)} \right)' \left[ I_{np} \otimes \Psi' G_{i,h} \right] M + M' \left[ I_{np} \otimes G_{i,h} \Psi \right]
\times \Sigma_K^{-1} K' J_p \left. \frac{\partial \text{vec}(\Sigma_Y^{(h-1-i)})}{\partial \text{vec}(\Phi)} \right] + \left( \frac{\partial \tilde{e}_{T+h-1-i}}{\partial \text{vec}(\Phi)} \right)' \left[ I_{np} \otimes \epsilon_h' \Sigma_K^{-1} K' J_p \Psi \right] M
+ M' \left[ I_{np} \otimes (\Psi)' J_p \Sigma_K^{-1} \epsilon_h \right] \frac{\partial \tilde{e}_{T+h-1-i}}{\partial \text{vec}(\Phi)} + \left( \frac{\partial \text{vec}(\epsilon_h^{(h-1-i)})}{\partial \text{vec}(\Phi)} \right)'
\times \left[ I_{np} \otimes J_p K \Sigma_K^{-1} \epsilon_h \epsilon_h' Y_T (\Psi)' \right] C_{np} M + M' C_{np}' \left[ I_{np} \otimes \Psi' Y_T \epsilon_h' \right]
\times \Sigma_K^{-1} K' J_p \left. \frac{\partial \text{vec}(\epsilon_h^{(h-1-i)})}{\partial \text{vec}(\Phi)} \right] + \left( \frac{\partial \epsilon_h^*}{\partial \text{vec}(\Phi)} \right)'
\times \left[ \Sigma_K^{-1} \frac{\partial \epsilon_h^*}{\partial \text{vec}(\Phi)} \right]
\left( \frac{\partial \text{vec}(\Sigma_Y^{(h-1-i)})}{\partial \text{vec}(\Phi)} \right)'
\left[ J_p K \Sigma_K^{-1} \otimes J_p K \Sigma_K^{-1} \epsilon_h \right] \frac{\partial \epsilon_h^*}{\partial \text{vec}(\Phi)}
\times \left( \frac{\partial \epsilon_h^*}{\partial \text{vec}(\Phi)} \right)'
\left[ \Sigma_K^{-1} K' J_p \otimes \epsilon_h' \Sigma_K^{-1} K' J_p \right] \frac{\partial \text{vec}(\Sigma_Y^{(h-1-i)})}{\partial \text{vec}(\Phi)}
\times \left( \frac{\partial \epsilon_h^*}{\partial \text{vec}(\Phi)} \right)'.
\]
Furthermore,

\[
- \frac{\partial^2 \log L^*_h}{\partial \text{vec}(\Phi) \text{vec}(\Omega)'} = \frac{1}{2} \left( \frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Phi)'} \right)' \left[ J_p K \Sigma^{-1}_{K,h} K' p \otimes J_p K (2 \Sigma^{-1}_{K,h} \varepsilon_h \varepsilon_h' - I_{n^*}) \right] \times \Sigma^{-1}_{K,h} K' p \ \frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Omega)'} + \sum_{i=0}^{h-2} M' [I_{np} \otimes G_{i,h} \Psi] \frac{\partial \text{vec}(\Sigma^{(h-1-i)})}{\partial \text{vec}(\Omega)'}
\]  

(B.18)

and

\[
- \frac{\partial^2 \log L^*_h}{\partial \text{vec}(\Omega) \text{vec}(\Omega)'} = \frac{1}{2} \left( \frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Omega)'} \right)' \left[ J_p K \Sigma^{-1}_{K,h} K' p \otimes J_p K (2 \Sigma^{-1}_{K,h} \varepsilon_h \varepsilon_h' - I_{n^*}) \right] \times \Sigma^{-1}_{K,h} K' p \ \frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Omega)'}
\]  

(B.19)

These expression are evaluated at the posterior mode in (B.14) and rely on the following:

\[
G_{i,h} = (\Psi')^i J_p K (I_{n^*} - \Sigma^{-1}_{K,h} \varepsilon_h \varepsilon_h') \Sigma^{-1}_{K,h} K' p \Psi^i, \quad i = 0, 1, \ldots, h - 2,
\]

\[
\Sigma_{K,h} = K' p \Sigma_{K,h} K_p,
\]

\[
\frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Phi)'} = \left( I_{n^2p^2} + C_{np} \right) \left[ \Psi \Sigma^{(h-1)} \otimes I_{np} + [\Psi \otimes \Psi] \frac{\partial \text{vec}(\Sigma^{(h-1)})}{\partial \text{vec}(\Phi)'} \right], \quad h = 1, 2, \ldots, H,
\]

where \( \Sigma^{(0)} = 0 \), \( \frac{\partial \text{vec}(\Sigma^{(0)})}{\partial \text{vec}(\Phi)'} = 0 \), and \( C_{np} \) is the \( n^2p^2 \times n^2p^2 \) commutation matrix such that \( C_{np} \text{vec}(\Psi) = \text{vec}(\Psi') \). Moreover,

\[
\frac{\partial \text{vec}(\Sigma^{(h)})}{\partial \text{vec}(\Omega)'} = \left[ I_{np} \otimes J_p \right] D_n + \left[ \Psi \otimes \Psi \right] \frac{\partial \text{vec}(\Sigma^{(h-1)})}{\partial \text{vec}(\Omega)'} \quad \text{for } h = 1, 2, \ldots, H,
\]

where again \( \frac{\partial \text{vec}(\Sigma^{(0)})}{\partial \text{vec}(\Omega)'} = 0 \). Next,

\[
\frac{\partial \bar{x}_{T+h}}{\partial \text{vec}(\Phi)'} = J_p N + \left[ \bar{x}_{T+h-1} \otimes I_{np} \right] M + \Psi \frac{\partial \bar{x}_{T+h-1}}{\partial \text{vec}(\Phi)'}, \quad h = 1, 2, \ldots, H,
\]

where \( \bar{x}_T = 0 \) and \( \partial \bar{x}_T/\partial \text{vec}(\Phi)' \). Furthermore,

\[
\frac{\partial \text{vec}(\Psi^{(h)})}{\partial \text{vec}(\Phi)'} = \left[ I_{np} \otimes \Psi^{(h-1)} \right] M + \left[ \Psi' \otimes I_{np} \right] \frac{\partial \text{vec}(\Psi^{(h-1)})}{\partial \text{vec}(\Phi)'} \quad \text{for } h = 1, 2, \ldots, H,
\]

where \( \partial \text{vec}(\Psi^{(0)})/\partial \text{vec}(\Phi)' = 0 \). Finally,

\[
\frac{\partial \bar{x}^*_h}{\partial \text{vec}(\Phi)'} = -K' p \frac{\partial \bar{x}_{T+h}}{\partial \text{vec}(\Phi)'} - \left[ Y^*_h \otimes K' p \right] \frac{\partial \text{vec}(\Psi^{(h)})}{\partial \text{vec}(\Phi)'} \quad \text{for } h = 1, 2, \ldots, H,
\]

and the \( n^2p^2 \times n(np + 1) \) and \( n \times n(np + 1) \) matrices

\[
M = \left[ 0_{n^2p^2 \times n} \ (I_{np} \otimes J_p) \right], \quad N = \left[ I_n \ 0_{n \times n^2p} \right].
\]

Notice that \( d \text{vec}(\Psi) = Md \text{vec}(\Phi) \) and \( \Phi_0 = N \text{vec}(\Phi) \).
### Table 1: Mean forecast errors for the NAWM and the DSGE-VAR(λ, p) models over the evaluation period 1999Q1–2006Q4.

<table>
<thead>
<tr>
<th>horizon</th>
<th>DSGE-VAR NAWM (2.5; 2) (6; 4)</th>
<th>DSGE-VAR NAWM (2.5; 2) (6; 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private consumption</td>
<td>Nominal wages</td>
</tr>
<tr>
<td>1</td>
<td>−0.12 0.01 0.03</td>
<td>−0.37 −0.19 −0.21</td>
</tr>
<tr>
<td>2</td>
<td>−0.23 −0.00 0.04</td>
<td>−0.45 −0.20 −0.22</td>
</tr>
<tr>
<td>4</td>
<td>−0.34 −0.01 0.02</td>
<td>−0.48 −0.21 −0.21</td>
</tr>
<tr>
<td>8</td>
<td>−0.27 −0.06 −0.06</td>
<td>−0.47 −0.23 −0.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consumption deflator</th>
<th>GDP deflator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14 0.03 0.03</td>
</tr>
<tr>
<td>2</td>
<td>0.25 0.06 0.06</td>
</tr>
<tr>
<td>4</td>
<td>0.30 0.07 0.08</td>
</tr>
<tr>
<td>8</td>
<td>0.16 0.02 0.06</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>horizon</th>
<th>DSGE-VAR NAWM (2.5; 2) (6; 4)</th>
<th>DSGE-VAR NAWM (2.5; 2) (6; 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private consumption</td>
<td>Nominal wages</td>
</tr>
<tr>
<td>1</td>
<td>12.6 0.1 1.0</td>
<td>45.1 23.4 23.3</td>
</tr>
<tr>
<td>2</td>
<td>31.7 0.0 1.2</td>
<td>73.3 32.2 38.8</td>
</tr>
<tr>
<td>4</td>
<td>62.9 0.2 0.4</td>
<td>79.2 35.7 40.7</td>
</tr>
<tr>
<td>8</td>
<td>51.7 4.4 5.0</td>
<td>78.8 39.8 35.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consumption deflator</th>
<th>GDP deflator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.9 1.9 1.1</td>
</tr>
<tr>
<td>2</td>
<td>55.5 8.5 6.8</td>
</tr>
<tr>
<td>4</td>
<td>65.1 8.0 11.1</td>
</tr>
<tr>
<td>8</td>
<td>45.5 1.1 7.4</td>
</tr>
</tbody>
</table>
Table 3: Percentage share of average parameter uncertainty to average forecast uncertainty.

<table>
<thead>
<tr>
<th>Variable</th>
<th>NAWM</th>
<th>DSGE-VAR(2.5; 2)</th>
<th>DSGE-VAR(6; 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h = 1$</td>
<td>$h = 4$</td>
<td>$h = 8$</td>
</tr>
<tr>
<td>Real GDP</td>
<td>0.9</td>
<td>1.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Private consumption</td>
<td>0.8</td>
<td>1.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Total investment</td>
<td>1.2</td>
<td>1.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Exports</td>
<td>0.8</td>
<td>0.9</td>
<td>0.3</td>
</tr>
<tr>
<td>Imports</td>
<td>0.9</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>GDP deflator</td>
<td>1.8</td>
<td>2.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Consumption deflator</td>
<td>1.8</td>
<td>3.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Import deflator</td>
<td>1.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Employment</td>
<td>0.8</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Nominal wages</td>
<td>2.3</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Nominal interest rate</td>
<td>1.5</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Real exchange rate</td>
<td>0.5</td>
<td>1.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Table 4: Log predictive score of one-step-ahead forecasts for all 18 variables over the evaluation period 1999Q1–2006Q4.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>NAWM</th>
<th>DSGE-VAR(2.5; 2)</th>
<th>DSGE-VAR(6; 4)</th>
<th>BVAR</th>
<th>Random-walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified harmonic mean</td>
<td>−599.9</td>
<td>−551.9</td>
<td>−563.3</td>
<td>−687.8</td>
<td></td>
</tr>
<tr>
<td>Laplace approximation</td>
<td>−599.6</td>
<td>−551.5</td>
<td>−563.4</td>
<td>−668.0</td>
<td></td>
</tr>
<tr>
<td>Normal approximation</td>
<td>−603.8</td>
<td>−567.4</td>
<td>−572.7</td>
<td>−602.9</td>
<td>−674.7</td>
</tr>
</tbody>
</table>

Note: The log predictive score values for the Laplace approximation are calculated as the difference between the log marginal likelihood for the full sample (until 2006Q4) using the posterior mode estimate for this sample and the log marginal likelihood for the first historical sample (until 1998Q4) using the posterior mode estimate for this shorter sample. If the log marginal likelihood for the full sample is instead computed using the posterior mode for the shorter sample, the Laplace approximation obtains a log predictive score of −650.2 for the NAWM, −600.3 and −590.3 for the DSGE-VAR models with two and four lags, respectively. This may be compared with −667.3 for the modified harmonic mean of the NAWM with posterior draws using data until 1998Q4.
Table 5: Log predictive score using the modified harmonic mean over the evaluation period 1999Q1–2006Q4.

<table>
<thead>
<tr>
<th>horizon</th>
<th>NAWM</th>
<th>DSGE-VAR (2.5; 2)</th>
<th>BVAR (6; 4)</th>
<th>Random-walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-395.7</td>
<td>-404.4</td>
<td>-416.6</td>
<td>-416.4</td>
</tr>
<tr>
<td>2</td>
<td>-449.7</td>
<td>-426.5</td>
<td>-443.4</td>
<td>-507.0</td>
</tr>
<tr>
<td>3</td>
<td>-472.5</td>
<td>-443.7</td>
<td>-458.6</td>
<td>-556.2</td>
</tr>
<tr>
<td>4</td>
<td>-474.3</td>
<td>-459.0</td>
<td>-465.7</td>
<td>-595.0</td>
</tr>
<tr>
<td>5</td>
<td>-470.4</td>
<td>-465.6</td>
<td>-467.7</td>
<td>-615.9</td>
</tr>
<tr>
<td>6</td>
<td>-461.3</td>
<td>-469.3</td>
<td>-468.0</td>
<td>-623.6</td>
</tr>
<tr>
<td>7</td>
<td>-449.1</td>
<td>-467.1</td>
<td>-462.8</td>
<td>-626.7</td>
</tr>
<tr>
<td>8</td>
<td>-434.6</td>
<td>-460.2</td>
<td>-452.8</td>
<td>-623.9</td>
</tr>
</tbody>
</table>

| 7 variables |
| 1 | -122.4 | -115.7 | -125.8 | -129.2 |
| 2 | -169.8 | -144.8 | -158.9 | -191.2 |
| 3 | -190.9 | -165.6 | -177.5 | -225.9 |
| 4 | -199.8 | -179.6 | -187.6 | -251.4 |
| 5 | -201.8 | -187.2 | -192.0 | -266.3 |
| 6 | -201.4 | -192.1 | -195.3 | -273.5 |
| 7 | -196.2 | -191.9 | -192.8 | -278.1 |
| 8 | -191.4 | -191.0 | -190.0 | -279.7 |

| 3 variables |
| 1 | -43.5 | -43.9 | -48.6 | -43.1 |
| 2 | -66.6 | -63.5 | -69.6 | -69.7 |
| 3 | -75.3 | -71.6 | -77.4 | -81.5 |
| 4 | -80.4 | -78.9 | -82.4 | -92.8 |
| 5 | -81.8 | -83.0 | -84.3 | -99.8 |
| 6 | -82.2 | -84.3 | -84.8 | -101.8 |
| 7 | -80.8 | -85.2 | -84.4 | -104.4 |
| 8 | -79.0 | -85.0 | -83.3 | -105.4 |

Note: The log predictive likelihood for the random-walk model is calculated with its analytical expression. For the 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, ..., 499951. For the BVAR direct sampling is possible; see Appendix B.
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 $\lambda$ for different lag orders.

DSGE-VAR($\lambda$, $p$) models

- $\lambda = 0.625$,
- $\lambda = 1.25$,
- $\lambda = 2$,
- $\lambda = 3$,
- $\lambda = 4$,
- $\lambda = 6.5$,
- $\lambda = 10$,
- $\lambda = 25$,
- $\lambda = \infty$.

Log marginal likelihood:
- $p = 1$: $-1800$
- $p = 2$: $-1818.3$
- $p = 3$: $-1845.4$
- $p = 4$: $-1855.6$

NAWM:
- $-2000$
- $-1950$
- $-1900$
- $-1850$
- $-1800$
- $-1750$
- $-1700$
- $-1650$
- $-1600$
- $-1550$
- $-1500$
- $-1450$
- $-1400$
- $-1350$
- $-1300$
- $-1250$
- $-1200$
- $-1150$
- $-1100$
- $-1050$
- $-1000$
- $-950$
- $-900$
- $-850$
- $-800$
- $-750$
- $-700$
- $-650$
- $-600$
- $-550$
- $-500$
- $-450$
- $-400$
- $-350$
- $-300$
- $-250$
- $-200$
- $-150$
- $-100$
- $-50$
- $0$
- $50$
- $100$
- $150$
- $200$
- $250$
- $300$
- $350$
- $400$
- $450$
- $500$
- $550$
- $600$
- $650$
- $700$
- $750$
- $800$
- $850$
- $900$
- $950$
- $1000$
- $1050$
- $1100$
- $1150$
- $1200$
- $1250$
- $1300$
- $1350$
- $1400$
- $1450$
- $1500$
- $1550$
- $1600$
- $1650$
- $1700$
- $1750$
- $1800$
- $1850$
- $1900$
- $1950$
- $2000$

Legend:
- $p = 1$
- $p = 2$
- $p = 3$
- $p = 4$
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.

(I) Private consumption

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td></td>
</tr>
</tbody>
</table>

Random walk

NAWM

DSGE-VAR(6;4)

(II) Nominal wages

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
<td></td>
</tr>
</tbody>
</table>

Random walk

NAWM

DSGE-VAR(6;4)
Figure 5: Trace statistics of the scaled MSE matrices.

12 variables

3 variables

7 variables

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

12 variables

3 variables

7 variables
Figure 7: Log predictive likelihoods for 12 variables using the NAWM.

Figure 8: Log predictive likelihoods for 12 variables using the DSGE-VAR(2) model.
Figure 9: Log predictive scores using the modified harmonic mean.

12 variables

7 variables

3 variables

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

Figure 10: Log predictive scores using normal approximation.

12 variables

7 variables

3 variables

- 46 -
Figure 11: Log predictive scores using Laplace approximation.

12 variables

7 variables

3 variables

- Random walk
- NAWM
- DSGE-VAR(2.5;2)
- DSGE-VAR(6;4)


