Marginalized Predictive Likelihood Comparisons with Applications to DSGE, DSGE-VAR, and VAR Models

Anders Warne, Günter Coenen and Kai Christoffel*

April 3, 2013

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

Keywords: Bayesian inference, BVAR, DSGE, DSGE-VAR, forecasting, Kalman filter, missing data, Monte Carlo integration, predictive likelihood.

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

1. Introduction

It has long been recognized that using the predictive likelihood is a valid Bayesian approach to model selection (see, e.g., Box, 1980), and the predictive Bayes factor is naturally defined from a ratio of two predictive likelihoods (Gelfand and Dey, 1994, Kass and Raftery, 1995). In discrete time series analysis, the predictive likelihood for a subset of the observed variables makes it possible to compare forecast accuracy across models that have different conditioning variables but where some of the forecasted variables are shared. A forecaster or policy maker is typically only interested in a limited number of the variables that a multivariate model can predict and a forecast comparison statistic based only on the variables of interest is therefore desirable. A special case is when we are interested in comparing the forecasts of a single variable, such as

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

* 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.

inflation, across a set of models, whereas larger subsets are required if we are also concerned with predicting comovements. Although point forecasts may reveal interesting aspects of the models involved, such forecasts are not well suited for model selection, the determination of model weights in a model averaging study (Eklund and Karlsson, 2007), or the construction of optimal prediction pools (Geweke and Amisano, 2011, 2012). The predictive likelihood is a natural tool for dealing with such matters, with the log predictive score as an extension from single-period comparisons to a multi-period setting.

The determination of the predictive likelihood for a subset of the variables requires that all the variables which are excluded from the subset are integrated out. If the predictive likelihood has a known distribution, such as a multivariate t, then the marginalization problem may be solved through textbook results. However, such cases are rare and typically involve very simple models. As a consequence, the next step is to take into account that the predictive likelihood is equal to the integral over the parameter space of the conditional likelihood times the posterior density of the parameters. By the conditional likelihood we mean the predictive likelihood conditional on a value for the parameters. If the conditional likelihood is based on a distribution where marginalization can be handled analytically, such as for a normal density, then the marginalization problem for the predictive likelihood may be solved at this stage (see, e.g., Andersson and Karlsson, 2008, Karlsson, 2012, or Geweke and Amisano, 2010). What remains to be done once the conditional likelihood has been marginalized is to integrate out the dependence of the parameters.

The main contribution of this paper concerns the calculation of the predictive likelihood for parametric discrete time series models when some of the predicted variables need to be integrated out. Specifically, we show how the h-step-ahead predictive likelihood can be calculated for *any* subset of the variables over the forecast horizon, and the suggested approach can be applied to a large family of models, with Gaussian log-linearized dynamic stochastic general equilibrium (DSGE) models, Bayesian vector autoregressions (BVARs), and DSGE-VARs as interesting special cases. Our approach to marginalization is based on the simple observation that the conditional likelihood for the subset of variables to be predicted is equal to the conditional likelihood for the full set of variables, but where the values of the variables that are excluded from the subset of interest are replaced with missing observations. For linear models with normal conditional likelihoods, the Kalman filter is easily adapted to handle missing data and may therefore be used to marginalize the conditional likelihood. Once we turn to nonlinear, nonnormal time series models that may be cast into state-space form, missing observations consistent filters may likewise be applied to compute the conditional likelihood; see, for instance, Giordani, Pitt, and Kohn (2011).

We illustrate the approach by building onto the forecast comparison exercise in Christoffel, Coenen, and Warne (2011), henceforth CCW. They review forecasting with DSGE models, using

the New Area-Wide Model (NAWM; pronounced nŏm) as an example, and their pseudo out-of-sample forecast exercise covers the period after the introduction of the euro, focusing on three nested partitions of the 12 (out of 18) observed variables that are endogenously determined in the NAWM. A multivariate random walk model with a diffuse prior on the covariance matrix and the best performing BVAR model from CCW, estimated with the methodology developed by Bańbura, Giannone, and Reichlin (2010) for large BVARs, are included in the empirical illustration for comparisons with the NAWM and DSGE-VAR models. Following Adolfson, Lindé, and Villani (2007b), CCW use a normal approximation of the predictive likelihood and we will assess the results from this approximation to those obtained from an estimator of the predictive likelihood based on Monte Carlo integration of the conditional likelihood with respect to the posterior draws.

The remainder of the paper is organized as follows. Section 2 shows how the predictive likelihood can be calculated via the conditional likelihood when we are interested in a subset of the predicted variables. Given a solution to this problem, we thereafter discuss the harmonic mean estimator, importance sampling estimators, and Monte Carlo integration for integrating out the dependence on the model parameters. Section 3 examplifies the approach through linear state-space models with Gaussian innovations, focusing on marginalized h-step-ahead forecasts with log-linearized DSGE models and VAR models. The suggested approach for computing the predictive likelihood for a subset of the variables is thereafter illustrated in a forecast comparison exercise in Section 4. Finally, Section 5 summarizes the main findings of the paper.

2. The Predictive Likelihood

2.1. Notation

To establish notation, let $\theta_m \in \Theta_m$ be a vector of unobserved parameters of a complete model (Geweke, 2005, 2010), indexed by m, while $\mathcal{Y}_T = \{y_1, y_2, \ldots, y_T\}$ is a discrete, real-valued time series for an n-dimensional vector of observables y_t. The observed values of this vector are denoted by y_t^o, while the sample of observations is similarly denoted by $\mathcal{Y}_T^o$. The observables density function for the time series $\mathcal{Y}_T$ is given by $p(\mathcal{Y}_T|\theta_m, m)$, while the likelihood function is denoted by $p(\mathcal{Y}_T^o|\theta_m, m)$. Bayesian inference is based on combining a likelihood function with a prior distribution, $p(\theta_m|m)$, in order to obtain a posterior distribution of the model parameters, $p(\theta_m|\mathcal{Y}_T^o, m)$. From Bayes theorem we know that the posterior is equal to the posterior kernel (the product of the likelihood and the prior) divided by the marginal likelihood, denoted by

$$p\big(\mathcal{Y}_T^o\big|m\big) = \int_{\Theta_m} p\big(\mathcal{Y}_T^o\big|\theta_m, m\big)p\big(\theta_m\big|m\big)d\theta_m. \tag{1}$$

The marginal likelihood is a standard measure of fit in Bayesian analysis and is a joint assessment of how well the prior and likelihood agrees with the data. It is used for computing Bayes factors, posterior odds, and posterior model probabilities (Kass and Raftery, 1995) provided that Lindley's (1957) paradox does not apply (Strachan and van Dijk, 2011).

Point and density forecasts are determined from the predictive density of model m and, for a sequence of future values of the observed variables $\{y_{T+1}, \ldots, y_{T+h}\}$, with $h = 1, \ldots, h^*$, this density can be expressed as

$$p(y_{T+1}, \ldots, y_{T+h} | \mathcal{Y}_T^o, m) = \int_{\Theta_m} p(y_{T+1}, \ldots, y_{T+h} | \mathcal{Y}_T^o, \theta_m, m) p(\theta_m | \mathcal{Y}_T^o, m) d\theta_m. \tag{2}$$

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

2.2. Relationships between the Predictive and the Marginal Likelihood

The predictive likelihood of model m is given by the predictive density in (2) evaluated at the observed values $\{y_{T+1}^o, \ldots, y_{T+h}^o\}$ and it is straightforward to show that it is equal to the ratio of the marginal likelihood for the historical and predictions sample, $\mathcal{Y}_{T+h}^o$, and the marginal likelihood for the historical sample, $\mathcal{Y}_T^o$; see, e.g., Geweke (2005). That is,

$$p(y_{T+1}^o, \ldots, y_{T+h}^o | \mathcal{Y}_T^o, m) = \frac{p(\mathcal{Y}_{T+h}^o | m)}{p(\mathcal{Y}_T^o | m)}, \quad h = 1, \ldots, h^*. \tag{3}$$

The predictive likelihood is therefore the update factor applied to the marginal likelihood for the historical sample that produces the marginal likelihood for the historical and prediction sample. It can be inferred from (3) that the predictive likelihood for model m_i may be greater than for model m_j, yet the marginal likelihoods for model m_i can be smaller than those for model m_j for both samples $\mathcal{Y}_T^o$ and $\mathcal{Y}_{T+h}^o$. Furthermore, the marginal likelihood can be expressed as a product of one-step-ahead predictive likelihoods according to:

$$p(\mathcal{Y}_\tau^o | m) = \prod_{t=1}^{\tau} p(y_t^o | \mathcal{Y}_{t-1}^o, m), \quad \tau = T, T+h,\ h = 1, \ldots, h^*, \tag{4}$$

establishing intimite relationships between marginal and predictive likelihoods.

The use of the predictive likelihood as a valid Bayesian approach to model selection has long been recognized. Box (1980), for example, has emphasized the complementary roles in the model building process of the posterior and predictive distributions, where the former can be used for diagnostic checking, while the latter provides a general basis for robustness checks. For models with improper priors—and therefore potentially subject to Lindley's paradox—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 density; see, e.g., Eklund and Karlsson (2007).

From equations (3) and (4) it follows that the h-step-ahead predictive likelihood can also be expressed as a product of h one-step-ahead predictive likelihoods:

$$p\big(y^o_{T+1},\dots,y^o_{T+h}\big|\mathcal{Y}^o_T,m\big)=\prod_{i=1}^{h}p\big(y^o_{T+i}\big|\mathcal{Y}^o_{T+i-1},m\big).$$

In essence the predictive likelihood in (3), just like the marginal likelihood in (1), therefore covers one-step-ahead forecasts only and is not well suited for a comparison of the forecasting performance for models when we are interested in forecasts at $h>1$. For such cases, the object of interest is the marginalized h-step-ahead predictive likelihood:

$$p\big(y^o_{T+h}\big|\mathcal{Y}^o_T,m\big)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p\big(y_{T+1},\dots,y_{T+h-1},y^o_{T+h}\big|\mathcal{Y}^o_T,m\big)dy_{T+h-1}\cdots dy_{T+1}. \tag{5}$$

The relationship between the ratio of marginal likelihoods and the predictive likelihood in (3) holds (trivially) also for the marginalized predictive likelihood in (5) when $h=1$. For other horizons it is claimed by both Adolfson et al. (2007b, p. 324) and CCW (p. 114) that this connection breaks down. Furthermore, Adolfson et al. (2007b, p. 325) remark that computing the marginalized predictive likelihood in (5) for $h>1$ is not an easy task since $p\big(y^o_{T+h}|\mathcal{Y}^o_T,m\big)$ does not have a closed form solution and that kernel density estimation from predictive draws is not practical unless the dimension of y_{t+h} is small. For linear Gaussian models they therefore suggest using a normal approximation of the predictive likelihood based on the mean and the covariance of the marginalized predictive distribution.

However, going back a step one realizes that both Adolfson et al. (2007b) and CCW are incorrect when they claim that the relationship between the marginal and predictive likelihood breaks down for $h>1$ in (5), since it remains true that

$$p\big(y^o_{T+h}\big|\mathcal{Y}^o_T,m\big)=\frac{p\big(y^o_{T+h},\mathcal{Y}^o_T\big|m\big)}{p\big(\mathcal{Y}^o_T\big|m\big)}. \tag{6}$$

The denominator is the marginal likelihood of model m for the historical sample $\mathcal{Y}^o_T$ and the numerator is likewise the marginal likelihood of this model when using the sample $(y^o_{T+h},\mathcal{Y}^o_T)$. Hence, the connection between the predictive likelihood and the marginal likelihood remains also for $h>1$.

Suppose that the models we wish to compare do not predict the same variables, i.e. y_t depends on m, but that they have a common subset, denoted by $y_{s,t}$, of the observables that they can predict. The dimension of this subset is denoted by n_s, with $n_s\leq n$. For example, we may be concerned with the comparison of density forecasts of inflation from univariate and multivariate models. Hence, rather than integrating out the influence of all observables from certain time periods, as in (5), we would like to integrate out the influence from all observables that are not included in $y_{s,t}$ such that the predictive likelihood only covers $\{y^o_{s,T+1},\dots,y^o_{s,T+h}\}$ conditional on $\mathcal{Y}^o_T$. For this situation, is it also true that the predictive likelihood can be expressed as a ratio of two marginal likelihoods.

These two examples can be treated as special cases of the situation when the dimension of the subset of variables $y_{s,t}$ is time-varying, i.e., when $\dim(y_{s,t}) = n_{s_t} \in \{0, 1, \ldots, n\}$. Letting $\mathcal{Y}_{s,T,h} = \{y_{s,T+1}, \ldots, y_{s,T+h}\}$, the predictive likelihood for this general case is

$$p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, m) = \frac{p(\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T|m)}{p(\mathcal{Y}^o_T|m)}. \tag{7}$$

The problem of computing the marginalized predictive likelihood on the left hand side of (7) can be reformulated into the issue of being able to compute the marginal likelihood for the historical sample and the prediction sample subject to some realizations being ignored, i.e., the numerator on the right hand side. But this suggests that marginalization can be regarded as solving a missing data problem. In other words, provided that we can compute the likelihood function for the historical and prediction sample subject to missing observations we can also compute the marginalized predictive likelihood for any choice $\mathcal{Y}^o_{s,T,h}$ of predicted variables.

The idea is related to data augmentation and other such EM algorithm extensions. For these algorithms, the model is used to replace missing observations with model-based draws of the latent variables and then use complete-data methods to address the incomplete-data problem; see, e.g., Tanner and Wong (1987) and Rubin (1991). In our case, we assume that the observables treated as missing are generated by the same state-space model as $\mathcal{Y}^o_{s,T,h}$.

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

If the conditional likelihood, denoted by $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta_m, m)$, is a normal density, then marginalization can also be conducted directly via the predictive mean and the covariance matrix conditional on the parameters by utilizing well-known properties of the normal distribution. Such an approach to marginalization is considered by Andersson and Karlsson (2008) and is in the case of linear models equivalent to using a Kalman filter for missing data. One advantage with the Kalman filter approach is that it provides a unifying framework for all linear models with a normal conditional likelihood, allowing the same software code to handle different models. Another advantage is that it emphasizes the recursive nature of evaluating the conditional likelihood for various forecast horizons.

2.3. Estimation of the Predictive Likelihood

Once the problem of calculating the conditional likelihood for a subset of the observed variables has been overcome, we proceed with the second step for computing the predictive likelihood. That is, to integrate out the dependence on the parameters. We assume that: (i) parameter

draws from the posterior density based on direct sampling or some Monte Carlo based simulator are available; and (ii) the predictive likelihood does not have a closed form solution for all forecast horizons. The second condition is typically met when we are dealing with a marginalized predictive likelihood, and should a closed form solution exist we would take advantage of it.

The predictive likelihood on the left hand side of (7) may be estimated with, e.g., the harmonic mean (HM) estimator; see Gelfand and Dey (1994), the truncated normal version in Geweke (1999, 2005), or the extension to a truncated elliptical in Sims, Waggoner, and Zha (2008). Other methods, such as bridge sampling or cross-entropy with importance sampling may also be considered; see Meng and Wong (1996), Frühwirth-Schnatter (2004), and Chan and Eisenstat (2013).

HM estimators of the predictive likelihood ideally require two sets of posterior draws: $\theta_h^{(i)} \in p(\theta|\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T)$, $i = 1, \ldots, N_h$, and $\theta^{(j)} \in p(\theta|\mathcal{Y}^o_T)$, for $j = 1, \ldots, N$, where the model index m has been supressed for notational convenience. The predictive likelihood can now be estimated as

$$\hat{p}_{HM}\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T\big) = \left[\frac{1}{N_h}\sum_{i=1}^{N_h} \frac{f\big(\theta_h^{(i)}\big)}{p\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta_h^{(i)}\big)\hat{p}_{HM}\big(\theta_h^{(i)}|\mathcal{Y}^o_T\big)}\right]^{-1}, \tag{8}$$

where the function $f(\theta)$ is a proper density and therefore integrates to unity, and the estimated posterior density is

$$\hat{p}_{HM}\big(\theta_h^{(i)}|\mathcal{Y}^o_T\big) = \frac{p\big(\mathcal{Y}^o_T|\theta_h^{(i)}\big)p\big(\theta_h^{(i)}\big)}{\hat{p}_{HM}\big(\mathcal{Y}^o_T\big)}. \tag{9}$$

The $\theta^{(j)}$ draws are needed to estimate the marginal likelihood in the denominator of (9). It follows from (8) and (9) that the HM estimator of the predictive likelihood is simply a ratio of marginal likelihood estimates:

$$\hat{p}_{HM}\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T\big) = \left[\frac{1}{N_h}\sum_{i=1}^{N_h} \frac{f\big(\theta_h^{(i)}\big)}{p\big(\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T|\theta_h^{(i)}\big)p\big(\theta_h^{(i)}\big)}\right]^{-1} \left[\frac{1}{N}\sum_{j=1}^{N} \frac{f\big(\theta^{(j)}\big)}{p\big(\mathcal{Y}^o_T|\theta^{(j)}\big)p\big(\theta^{(j)}\big)}\right], \tag{10}$$

where $p(\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T|\theta)$ is the product of the conditional likelihood $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta)$ and the likelihood function $p(\mathcal{Y}^o_T|\theta)$.

The posterior draws $\theta_h^{(i)}$ depend on the forecast horizon h as well as on the selected subset of variables. To avoid having to generate posterior draws for each sample $\{\mathcal{Y}^o_{s,T,h}, \mathcal{Y}_T\}$, it is tempting to replace $\theta_h^{(i)}$ in (8) with $\theta^{(j)}$. If the dimension of $\mathcal{Y}_{s,T,h}$ is small, this approximation may work well in practise, but unlike the estimator in (8) the resulting predictive likelihood estimator is *not* consistent.

If we insist on using only one set of parameter draws for all forecast horizons when computing the predictive likelihood, we may instead use an importance sampling (IS) estimator; see, e.g., Geweke (2005). With $\theta^{(k)}$, $k = 1, \ldots, N_k$, being draws from the importance density $g(\theta)$, a

general expression of the IS estimator is

$$\hat{p}_{IS}\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T\big) = \frac{1}{N_k}\sum_{k=1}^{N_k}\frac{p\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T,\theta^{(k)}\big)p\big(\theta^{(k)}|\mathcal{Y}^o_T\big)}{g\big(\theta^{(k)}\big)}. \tag{11}$$

Letting $g(\theta) = p(\theta|\mathcal{Y}^o_T)$ such that $\theta^{(k)} = \theta^{(j)}$ with $N_k = N$, the estimator of the predictive likelihood in (11) is simply the average over the N posterior draws $\theta^{(j)}$ of the conditional likelihood, i.e. standard Monte Carlo integration based on the conditional likelihood. Relying on the idea of Rao-Blackwellization, such an estimator is also suggested for models with a normal conditional likelihood by Andersson and Karlsson (2008); see also Geweke and Amisano (2010).

Under certain regularity conditions (Tierney, 1994), the right hand side of (11) converges almost surely to the expected value of $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T,\theta)$ with respect to $p(\theta|\mathcal{Y}^o_T)$, i.e. to the predictive likelihood $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T)$. Hence, equipped with the posterior draws $\theta^{(j)}$ and the conditional likelihood, $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T,\theta)$, the predictive likelihood can be consistently estimated directly, without having to compute it from two marginal likelihoods, and without having to sample from the distribution of the parameters conditional on $(\mathcal{Y}^o_{s,T,h},\mathcal{Y}^o_T)$ for each forecast horizon $h = 1,\dots,h^*$ and each subset of variables that we are interested in.

A further important property of the IS estimator is that it is unbiased (see Chan and Eisenstat, 2013, Proposition 1), while the HM estimator is not. Furthermore, the latter estimator is sensitive to the choice of $f(\theta)$ and can be difficult to pin down numerically when the dimension of θ is large, while the IS estimator based on the posterior $p(\theta|\mathcal{Y}^o_T)$ should be less hampered by this. In the case of (log-linearized) DSGE models, which are typically tightly parameterized, numerical issues with the HM need not be a major concern, but for BVAR models the computations need to take all the VAR parameters into account and such issues are therefore more likely to be important.

The IS estimator is expected to work well in practise when the draws from the importance density cover well enough the parameter region where the conditional likelihood is large. This is typically the case when computing the marginalized predictive likelihood with $g(\theta) = p(\theta|\mathcal{Y}^o_T)$ for shorter forecast horizons or a low dimensional $\mathcal{Y}_{s,T,h}$ prediction sample, but it is questionable when dealing with the predictive likelihood as the dimension of the prediction sample becomes large. For instance, when $y_{s,t} = y_t$, $t = T+1,\dots,T+h$, and h is sufficiently large, the situation resembles the case when the marginal likelihood is computed by averaging the likelihood over the prior draws, and such an estimator is typically poor. It may then be pertinent to consider cross-entropy methods for selecting the importance density optimally, as in Chan and Eisenstat (2013). With these caveats in mind, the IS estimator with importance density given by the posterior is treated below as the predictive likelihood estimator that is most likely to pinpoint the "true" value. Comparisons with other consistent and unbiased estimators of the predictive likelihood is left for future research.

Finally, importance sampling is based on i.i.d. draws from the importance density. For many parametric time series models, the posterior draws are obtained via Markov chain Monte Carlo,

such as the random walk Metropolis sampler, and are therefore not independent. In strict terms, the estimator in (11) is not an IS estimator when the i.i.d. assumption is violated, but we shall nevertheless use this term also when the draws from the posterior are dependent.

2.4. Normal Approximation of the Predictive Likelihood

It was suggested by Adolfson et al. (2007b) and CCW to approximate the predictive likelihood with a normal density with mean and covariance matrix taken from the predictive density. While such an approximation is not needed when we know how to compute the conditional likelihood, it can nevertheless serve as an analytical tool for enhancing our understanding of the results of a forecast comparison study.

First, the mean and covariance matrix of the predictive density in (2) can be estimated directly from the posterior draws when the mean and covariance matrix of the predicted variables conditional on the historical data and the parameters have analytical expressions. Let these analytically specified moments be denoted by $E[y_{T+1}, \dots y_{T+h}|\mathcal{Y}_T^o, \theta]$ and $C[y_{T+1}, \dots y_{T+h}|\mathcal{Y}_T^o, \theta]$, respectively. The mean of the predictive density is given by

$$E\big[y_{T+1}, \dots y_{T+h}\big|\mathcal{Y}_T^o\big] = E_T\Big[E\big[y_{T+1}, \dots y_{T+h}\big|\mathcal{Y}_T^o, \theta\big]\Big], \tag{12}$$

where E_T denotes the expectation with respect to the posterior $p(\theta|\mathcal{Y}_T^o)$. The covariance matrix can be expressed as

$$C\big[y_{T+1}, \dots y_{T+h}\big|\mathcal{Y}_T^o\big] = E_T\Big[C\big[y_{T+1}, \dots y_{T+h}\big|\mathcal{Y}_T^o, \theta\big]\Big] + C_T\Big[E\big[y_{T+1}, \dots y_{T+h}\big|\mathcal{Y}_T^o, \theta\big]\Big], \tag{13}$$

and C_T denotes the covariance with respect to the posterior. Notice that the covariance matrix of the predictive density is obtained through Rao-Blackwellization, where the first term on the right hand side reflects residual uncertainty and the second term parameter uncertainty.

Second, the normal approximation provides a simple way of decomposing the predictive likelihood into a term reflecting forecast errors and a term driven by forecast uncertainty. The mean and covariance matrix of the predicted variables $\mathcal{Y}_{s,T,h}$ is determined by selecting the proper elements of (12) and (13), respectively. Next, notice that

$$\log \hat{p}_N\big(\mathcal{Y}_{s,T,h}^o\big|\mathcal{Y}_T^o\big) = -\frac{d}{2}\log\big(2\pi\big) + D_{s,T+h} + Q_{s,T+h}, \tag{14}$$

where d is the dimension of the predicted variables $\mathcal{Y}_{s,T,h}$,

$$D_{s,T+h} = -\frac{\log\Big|C\big[\mathcal{Y}_{s,T,h}\big|\mathcal{Y}_T^o\big]\Big|}{2}, \tag{15}$$

$$Q_{s,T+h} = -\frac{\epsilon_{s,T+h}^{o\prime} C\big[\mathcal{Y}_{s,T,h}\big|\mathcal{Y}_T^o\big]^{-1}\epsilon_{s,T+h}^o}{2}, \tag{16}$$

and $\epsilon_{s,T+h}^o$ the the vector of prediction errors for the realizations $\mathcal{Y}_{s,T,h}^o$. The forecast uncertainty term is given by $D_{s,T+h}$, while $Q_{s,T+h}$ gives the impact of the quadratic standardized forecast errors on the normal approximation of the log predictive likelihood. This decomposition may

be of particular interest when the difference between the normal approximation and the IS estimator of the predictive likelihood is small, or the ranking of models is robust across these measures. For such cases, the decomposition may reveal whether forecast uncertainty (15) or forecast errors (16) is responsible for the ranking of models.

3. Linear State-Space Models with Gaussian Innovations

Let us now turn to three models with a normal conditional likelihood. The first case is a log-linearized DSGE model with rational expectations, and the second deals with a VAR model. The first model is structural in the sense that (most of) the parameters have an economic interpretation, and is usually tightly parameterized. VAR models, on the other hand, generally have an abundance of parameters and are nonstructural in the sense that (most of) the parameter do not have an economic interpretation. In addition, we discuss the case of DSGE-VAR models (see Del Negro and Schorfheide, 2004, or Del Negro, Schorfheide, Smets, and Wouters, 2007), which serves as an interesting example of a "marriage" between the structural and nonstructural modelling approaches. We will also make use of DSGE-VAR models in the empirical illustration in Section 4.

3.1. DSGE Models

Since the turn of the century, we have witnessed the development of a new generation of DSGE models that build on explicit micro-foundations with optimizing agents. Major advances in estimation methodology allow the estimation of variants of these models that are able to compete, in terms of data coherence, with more standard time series models, such as VARs; see, among others, the empirical models in Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2003, 2007), and Adolfson, Laséen, Lindé, and Villani (2007a). Efforts have also been undertaken to bring these models to the forecasting arena with promising results; see, for example, CCW, Del Negro and Schorfheide (2012), and references therein.

The log-linearized DSGE model may be written as:

$$A_{-1}\xi_{t-1} + A_0\xi_t + A_1 E_t\xi_{t+1} = D\eta_t, \quad t = 1, 2, \ldots, T, \tag{17}$$

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

Provided that a unique and convergent solution of the system (17) exists at a particular value of θ (see, e.g., Anderson, 2010, Klein, 2000, or Sims, 2002), we can express the model variables as a VAR system:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1, \ldots, T, \tag{18}$$

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

$$y_t = \mu + H'\xi_t + w_t, \quad t = 1, \ldots, T. \tag{19}$$

The measurement errors, w_t, are assumed to be i.i.d. $N(0, R)$, while μ is the population mean (steady state) of y_t conditional on θ provided that $H'\xi_t$ is stationary. The measurement errors and the shocks η_t are assumed to be independent, while the matrices H and R are also uniquely determined by θ.

The system in (18) and (19) is a state-space model, where equation (18) gives the state or transition equation and (19) the measurement or observation equation. Sargent (1989) was among the first to recognize that linear 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 with the Kalman filter; see, e.g., Durbin and Koopman (2012) for details.

Suppose that we are interested in forecasting the subset of observables $\mathcal{Y}_{s,T,h} = y_{s,T+h}$, i.e. the marginal h-step-ahead forecasts for a subset of the observables. The log of the conditional likelihood of the DSGE model is given by

$$\begin{aligned} \log p\big(y^o_{s,T+h} \big| \mathcal{Y}^o_T, \theta\big) &= -\frac{n_s}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_{y_s,T+h|T}\right| \\ -\frac{1}{2}&\left(y^o_{s,T+h} - y_{s,T+h|T}\right)' \Sigma^{-1}_{y_s,T+h|T}\left(y^o_{s,T+h} - y_{s,T+h|T}\right), \end{aligned} \tag{20}$$

where $y_{s,T+h} = S'y_{T+h}$, $y_{s,T+h|T} = S'y_{T+h|T}$, and $\Sigma_{y_s,T+h|T} = S'\Sigma_{y,T+h|T}S$, with S being an $n \times n_s$ known selection matrix $(n_s \leq n)$, and $h = 1, \ldots, h^*$. For convenience we assume that the S matrix is constant over the forecast horizon h, i.e. that we have the same subset of variables for each forecast horizon. The Kalman filter provides us with

$$y_{T+h|T} = \mu + H'F^h\xi_{T|T}, \tag{21}$$

$$\Sigma_{y,T+h|T} = H'P_{T+h|T}H + R, \tag{22}$$

$$P_{T+h|T} = FP_{T+h-1|T}F' + BB', \quad h = 1, \ldots, h^*, \tag{23}$$

where $\xi_{T|T}$ is the filter estimate of the state variables, and $P_{T|T}$ the corresponding filter estimate of the state variable covariance matrix based on the data $\mathcal{Y}^o_T$. The conditional likelihood in (20) may now be combined with the posterior draws of θ to estimate the predictive likelihood of $y^o_{s,T+h}$ for the DSGE model with the IS estimator in (11).

The expressions for the predictive mean and covariance matrix of the observables conditional on the parameters in (21) and (22) may be combined with equations (12) and (13), respectively, and the posterior draws to estimate the mean and covariance matrix of the marginalized predictive density. These moments can thereafter be employed when computing the normal approximation of the predictive likelihood, taking the selection matrix S into account.

3.2. VAR Models

VAR models have played a central role in the development of empirical macroeconomics since the seminal article by Sims (1980). One reason for this success is that they highlight the importance of a multivariate dynamic specification for macroeconomic analysis, letting all observed variables be treated as endogenous. Moreover, they allow for structural interpretations once the underlying structural shocks have been identified from the one-step-ahead forecast errors of the model, popularizing tools such as impulse response functions and forecast error variance decompositions. Furthermore, the models are closely connected with the concepts of cointegration and common trends; see Engle and Granger (1987) and Stock and Watson (1988).

VAR models may also be considered as linear approximations of DSGE models. The state-space representation in (18) and (19) can be expressed as a infinite order VAR representation when the so-called "poor man's invertibility condition" in Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007) is satisfied. As pointed out by Franchi and Paruolo (2012), this condition is typically sufficient for the existence of an infinite order VAR, but is sometimes also necessary.

The VAR model of y_t can be written as:

$$y_t = \Phi_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \epsilon_t, \quad t = 1, \ldots, T, \tag{24}$$

where $\epsilon_t \sim N_n(0, \Sigma_\epsilon)$. The vector Φ_0 is $n \times 1$, while Φ_j is $n \times n$ for $j = 1, \ldots, p$. We assume that initial values for y_t exists for $t = 0, \ldots, 1-p$. Let Ψ denote an $np \times np$ matrix with

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

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

$$y_{T+h} = J_p' \bar{x}_{T+h} + J_p' \Psi^i Y_T + J_p' \bar{\epsilon}_{T+h}, \quad h = 1, \ldots, h^*, \tag{25}$$

where

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

and these np-dimensional vectors are initialized through $\bar{x}_T = \bar{\epsilon}_T = 0$.

The log of the conditional likelihood of the VAR model for the subset of observables $\mathcal{Y}_{s,T,h} = y_{s,T+h}$ is again given by equation (20) when we take into account the parameters of the VAR.

From equation (25) it is straightforward to show that

$$y_{T+h|T} = J_p' \bar{x}_{T+h} + J_p' \Psi^h Y_T^o, \tag{26}$$

$$\Sigma_{y,T+h|T} = J_p' \bar{\Sigma}_Y^{(h)} J_p, \quad h = 1, \ldots, h^*. \tag{27}$$

The $np \times np$ covariance matrix $\bar{\Sigma}_Y^{(h)}$ is defined from the difference equation

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

with the initial value $\bar{\Sigma}_Y^{(0)} = 0$. Based on posterior draws of the VAR parameters, the predictive likelihood can be estimated with the IS estimator. Furthermore, the normal approximation of the predictive likelihood in (14) can be computed for VAR models by utilizing equations (26) and (27) along with the posterior parameter draws and the selection matrix S when computing the predictive mean and covariance matrix in (12) and (13), respectively.

3.3. DSGE-VAR Models

The unobserved parameters of a VAR model are given by $(\Phi_0, \Phi_1, \ldots, \Phi_p, \Sigma_\epsilon)$, provided that the prior distribution of the VAR does not include additional unobserved parameters. BVAR models (see, e.g., Del Negro and Schorfheide, 2011, or Karlsson, 2012) typically include a number of hyperparameters that are calibrated by the researcher and are therefore included in the model index m. A well-known example when the model includes additional parameters through the prior is a DSGE-VAR, where the parameters of the DSGE model appear.

An early attempt to combine DSGE models with 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, λ, which determines the weight on the prior relative to the data. The DSGE model approximation resides at one end of its range ($\lambda = \infty$), an unrestricted VAR at the other end ($\lambda = 0$), and in between these two extremes a large number of models exist. Apart from providing a measure of the degree to which the DSGE model is misspecified, the approach also allows for posterior analysis of the DSGE model parameters, impulse-response analysis, forecast-error-variance decompositions, and so on. While these models were first designed to improve forecasting and monetary policy analysis with VARs, the extension to a model evaluation toolkit was carried out by Del Negro and Schorfheide (2006), while Del Negro et al. (2007) used it to assess the fit of a DSGE model.

When estimating the predictive likelihood for DSGE-VAR models with the IS estimator and the normal approximation, we make use of the expressions provided for VAR models above.

It should be kept in mind that Monte Carlo integration with respect to the parameters of a DSGE-VAR involves all the VAR parameters and all the DSGE model parameters. In principle, it is also possible to integrate out the influence of λ for a finite number of feasible values, but such a model averaging-based approach is not considered below. Rather, we shall condition on a posterior mode estimate of this parameter.

4. Comparing Forecast Accuracy: An Illustration for the Euro Area

In this Section we will illustrate how the predictive likelihood may be used to compare density forecasts for linear models with normal conditional likelihoods. We begin by discussing the set of models used for the exercise. They cover a medium-size DSGE model, developed at the European Central Bank (ECB), which extends the Smets-Wouters model in an open-economy setting, two DSGE-VAR models, a large BVAR model, as well as a multivariate random walk model, and are discussed in 4.1. In 4.2 we thereafter turn our attention to a suitable metric for comparing multi-period density forecasts based on the predictive likelihood, the log predictive score. The last part of the Section, 4.3, presents the forecast sample and summarizes the empirical results of the exercise. All calculations in this paper have been performed with the help of YADA, a Matlab program for Bayesian estimation and evaluation of DSGE and DSGE-VAR models; see Warne (2013) for details.

4.1. Structural and Nonstructural Models

4.1.1. 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 details. The development of this DSGE 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.

In order to estimate the model, Christoffel et al. (2008) use time series for 18 macroeconomic variables and the data are taken from the Area-Wide Model database (Fagan, Henry, and Mestre, 2005) except for the time series of extra-euro area trade data (see Dieppe and Warmedinger, 2007, for details on their construction). The estimation sample is given by the period 1985Q1 until 2006Q4, with 1980Q2-1984Q4 serving as training sample.

The 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 a linear trend consistent with the NAWM's steady-state growth rate of 2.0% per annum is removed. This trend is assumed to have two components: labor productivity growth of 1.2% and labor force growth of 0.8%. Third, for the logarithm of employment a linear trend consistent with a steady-state labor force growth rate of 0.8% is removed. Fourth, a measure of the real effective exchange rate is constructed from the nominal effective exchange rate, the domestic GDP deflator, and foreign prices (defined as a weighted average of foreign GDP deflators) minus its 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.

4.1.2. DSGE-VAR Models with the NAWM as Prior

Christoffel et al. (2008) adopt the empirical approach outlined in Smets and Wouters (2003) and An and Schorfheide (2007) and estimate the NAWM with Bayesian methods. The DSGE-VAR models with the NAWM as prior have been estimated over the same sample with the random walk Metropolis algorithm subject to a Gaussian proposal density.

We consider two approaches for selecting a DSGE-VAR model. The first chooses the model with the largest marginal likelihood over all pairs (λ, p), while the second picks p such that the marginal likelihood of the DSGE-VAR model with $\lambda = \infty$ (the VAR parameters are completely determined from the DSGE model parameters) is the closest to the marginal likelihood of the DSGE model and then selects λ optimally conditional on this p. The log marginal likelihood values for a set of DSGE-VAR models, along with the NAWM, are displayed in Figure 2 and have been calculated with the HM estimator based on the truncated normal density; see Geweke (1999, 2005). We only consider lag orders between one and four.

It is noteworthy that for all lag orders $p \geq 2$, the log marginal likelihood at $\lambda = \infty$ is higher than the log marginal likelihood for the NAWM by a magnitude of 50 log-units or greater. Compared with Del Negro et al. (2007), who use a model with seven observed variables, the increase in marginal likelihood for the VAR approximation is very large. It should be kept in mind, however, that Del Negro et al. (2007) add an error correction term to their VAR model with the cointegration relations implied by their DSGE model. Consistent with the results shown in Adolfson, Laséen, Lindé, and Villani (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. Given the purpose of the current paper, we do not see any need for adding cointegration relations to the DSGE-VAR 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 reduction of about 11 log-units. Moreover, the marginal likelihood for the optimal one-lag model ($\hat{\lambda} = 1.25$) is lower than, but close to, the marginal likelihoods for the DSGE-VAR(∞) models with a higher lag order. Given these discouraging results for the one-lag models, we have opted to drop them from the forecast comparison exercise.

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

With these results in mind, we have decided to allow the DSGE-VAR models $(\lambda, p) = (2.5, 2)$ (first model selection approach) and $(\lambda, p) = (6, 4)$ (second model selection approach) to enter the forecast comparison exercise.

4.1.3. VAR and Random Walk Models

We also consider a Bayesian VAR model for the same observed variables as the NAWM. The usefulness of BVARs of the Minnesota-type for forecasting purposes has long been recognized, as documented early on by Litterman (1986), and such models are therefore natural benchmarks in forecast comparisons. Based on the results reported in CCW, we employ a large BVAR, estimated using the methodology in Bańbura et al. (2010). This approach relies on using dummy observations when implementing the normal-inverted Wishart version of the Minnesota prior. Moreover, the prior mean of the parameters on the first *own* lag of the endogenous variables (diagonal of Φ_1) are either unity, if the variable is measured in log-levels or levels, and zero if it is measured in log first differences. That is, the prior mean supports random walks for all variables in log-levels or levels. In CCW, this large BVAR is referred to as the model with a mixed prior. A more detailed description of this BVAR is found in the online appendix (Appendix B).

The last model we shall consider is a random walk for the vector y_t with the NAWM variables. For this model we make use of a standard diffuse prior for the covariance matrix of the random walk innovations. That is, the vector $y_t - y_{t-1} = \varepsilon_t$ is i.i.d. $N_n(0, \Omega)$, where $p(\Omega) \propto |\Omega|^{-(n+1)/2}$. One advantage of this model is that it allows for an analytical determination of the predictive density. For marginal h-step-ahead forecasts of $y_{s,T+h}$ the predictive density is given by a n_s-dimensional t-distribution with mean $y^o_{s,T}$, covariance matrix

$$\frac{h}{T-n-1}\sum_{t=1}^{T} S'\left(y^o_t - y^o_{t-1}\right)\left(y^o_t - y^o_{t-1}\right)'S,$$

and $T - n + n_s$ degrees of freedom; see the online appendix (Appendix A) for details.

4.2. The Log Predictive Score

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

A widely used scoring rule that was suggested by, e.g., Good (1952) is the log predictive score. Based on the predictive density of $\mathcal{Y}_{s,T,h}$, it can be expressed as

$$S_h(m) = \sum_{t=T}^{T+T_h-1} \log p(\mathcal{Y}_{s,T,h}|\mathcal{Y}_t^o, m), \quad h = 1, \ldots, h^*, \tag{28}$$

where T_h is the number of time periods the h-step-ahead predictive density is evaluated. If the scoring rule depends on the predictive density only through the realization of the variables of interest over the prediction sample, $\mathcal{Y}_{s,T,h}^o$, then the scoring rule is said to be *local*. Under the assumption that only local scoring rules are considered, Bernardo (1979) showed that every proper scoring rule is equivalent to a positive constant times the log predictive score plus a real valued function that only depends on the realized data; see Gneiting and Raftery (2007) for a recent survey on scoring rules.

When comparing the density forecasts of the NAWM, the two DSGE-VAR models, the large BVAR, and the multivariate random walk model we will evaluate the log predictive score in (28) with realizations for different subsets of the observables $\mathcal{Y}_{s,T,h}^o = y_{s,t+h}^o$. Hence, the predictive likelihood for each model and time period is marginalized with respect to the forecast horizon and the variables of interest in the subset. This allows us to utilize the algorithms discussed in Section 3 for the DSGE and VAR models, while the predictive likelihood for the multivariate random walk is, as noted above, analytically determined. Moreover, the log predictive score is optimal in the sense that it uniquely determines the model ranking among all local and proper scoring rules. However, there is no guarantee that it will pick the same model as the forecast horizon or the selected subset of variables changes.

4.3. Density Forecasts

4.3.1. Forecast Sample and Selection of Variables

The first pseudo out-of-sample forecasts are computed for 1999Q1—the first quarter after the introduction of the euro—while the final period is 2006Q4. The maximum forecast horizon is eight quarters, yielding 32 quarters with one-step-ahead forecasts and 25 quarters with eight-step-ahead forecasts. We shall only consider forecasts of quarterly growth rates for the variables in first differences, while CCW also study forecasts of annual growth rates for such variables. The

Kalman filter based forecasts can be adjusted to handle such transformations of the variables; see Warne (2013, Section 12.6.1).

Concerning the selection of variables in the subsets of the observables we follow CCW and exclude the variables which are essentially exogenous in the NAWM. That is, we do not compare density forecasts which include the five foreign variables (foreign demand, foreign prices, foreign interest rate, competitors' export prices, and oil prices) and government consumptions. For the remaining 12 variables we examine three nested subsets. The smallest subset is called the *small selection* and is given by real GDP, the GDP deflator, and the short-term nominal interest rate. This selection may be regarded as the minimum set of variables relevant to monetary policy. The second case covers a *medium selection* with the seven variables studied in Smets and Wouters (2003). In addition to the variables in the small selection, this selection covers private consumption, total investment, employment, and nominal wages. Finally, the *large selection* has 12 variables, given by the medium selection plus exports, imports, the import price deflator, the private consumption deflator, and the real effective exchange rate.

4.3.2. Empirical Results

The log predictive scores based on the IS estimator of the marginal h-step-ahead predictive likelihoood are shown in Figure 3 for all variable selections, forecast horizons, and models. For the NAWM and the two DSGE-VAR models we have used 10,000 posterior draws among the available 500,000 post burn-in draws for each model and time period when calculating the log predictive likelihood. These draws have been selected as draw number $1, 51, \ldots, 499951$ to combine modest computational costs with a small correlation between the draws and a sufficiently high estimation accuracy. This procedure yields estimates of the log predictive likelihood that are accurate up to and including the first decimal. In the case of the NAWM, the numerical standard error of the IS estimator based on the Newey and West (1987) approach is less than 0.04 for the shortest historical sample and the large selection. For the same sample and the medium selection, it is less than 0.03, and for the small selection less than 0.015. As the length of the historical sample increases, the numerical standard errors decrease. Moreover, the numerical standard errors for the DSGE-VAR models are even smaller.

In the case of the random walk model, the predictive likelihood for a selection of variables is multivariate t and can therefore be computed from its analytical expression. Direct sampling is possible for the BVAR model through its normal-inverted Wishart posterior and we have used 50,000 draws from its posterior distribution when computing the predictive likelihood with the IS estimator.

When comparing the NAWM with the two DSGE-VAR models, it is noteworthy that the DSGE-VAR model with two lags generally obtains higher log scores for all horizons and variable selections, with values for the four-lag model being slightly below those for the two-lag model, while the NAWM gets smaller values. At the longer horizons, the NAWM obtains values that are near those of the DSGE-VAR models and, in the case of the small selection, even slightly higher.

Hence, it seems that taking misspecification of the NAWM into account through DSGE-VAR models improves forecasting performance, especially at the shorter horizons.

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

Compared with the BVAR model, however, the NAWM and the two DSGE-VARs are outperformed for all selections and forecast horizons. For example, the difference between the log score of the BVAR and the DSGE-VAR with two lags is at least 27 log-units for the large selection and 40 log-units for the medium selection. The only exception is found for the small selection at the longer horizons, where the differences are about 4-5 log-units for eight-step-ahead forecasts. Nevertheless, if the log predictive score is employed as a model-selection device it would prefer the BVAR to the other models for all selections and forecast horizons. The interested reader is referred to Warne, Coenen, and Christoffel (2013) for additional details on the results.

To address the issue of how well the normal approximation works for these linear Gaussian models, the log predictive scores for this estimator are displayed in Figure 4. The most prominent feature is how similar these graphs are when compared to those in Figure 3. In fact, the IS estimator and the normal approximation suggest the same ranking of the models for each selection and forecast horizon except for the DSGE-VAR models with the small selection and the eight-step-ahead forecasts. For this particular case, however, the difference in log predictive scores is so small that the models may be viewed as equally good (or bad). The differences in log marginal likelihood between the IS estimator and the normal approximation for all models, forecast horizons, and selections of variables are documented in Table 1.

The differences between the IS estimator and the normal approximation of the log predictive score for the NAWM and the two DSGE-VAR models are positive for all forecast horizons and variable selections. The results for the BVAR are mixed with a tendency for positive errors, while for the random walk model the differences are always positive. In terms of log-units, the largest errors for the DSGE related models are obtained for the DSGE-VAR with two lags but are never greater than about 4.3 log-units. For the NAWM, the DSGE-VAR with four lags, and the BVAR the errors are about half the size in magnitude. Furthermore, when comparing the differences between the estimates of the log predictive likelihood for the individual forecast periods and horizons, the overall errors are evenly spread out for the NAWM and the two DSGE-VAR models, while the BVAR tends to display larger deviations between the estimates when the log predictive likelhood values are smaller. Further details are shown in Warne et al. (2013).

Since the normal approximation provides a good approximation of the actual predictive likelihood for the five models and the three variable selections, we can utilize equations (14)–(16) in Section 2.4 to assess if the ranking of the models is driven by forecast uncertainty or by forecast errors. The forecast uncertainty term in equation (15) of the decomposition of the log predictive

likelihood is depicted in Figure 5 for the five models, eight forecast periods, and each relevant period in the prediction sample for the large selection of variables. Analogously, the quadratic standardized forecast error term in equation (16) is displayed in Figure 6 for the same cases. Although the discussion below focuses on the large selection of variables, the overall findings are also valid for the medium and small selections.

Turning first to the forecast uncertainty term in Figure 5 it can be seen that for all models and forecast horizons it is weakly upward sloping over the forecast sample and that the slope is roughly equal across the five models. This indicates that overall forecast uncertainty is slowly decreasing as data are added to the information set. The values for the BVAR model are roughly 5 log-units higher in each period than for the second group of models, given by the two DSGE-VARs and the NAWM. The random walk model has the lowest values for all forecast horizons and the difference relative to the second group of models is increasing with the forecast horizon.

Since the log-determinant in the expression for $D_{s,T+h}$ is equal to the sum of the log of the eigenvalues of the forecast error covariance matrix, the value of the log-determinant term is greater the smaller the eigenvalues of this matrix are. The eigenvalues in turn are small for linear combinations of the variables that, according to the model, are highly predictable. The plots in Figure 5 show that the BVAR model has better predictability in terms of the second moments than the other models. The DSGE-VARs with two and four lags and the NAWM follow as second, third, and fourth, while the random walk comes last.

Turning to the quadratic standardized forecast error term in Figure 6, it can be seen that the time variation of the log predictive likelihood is due to the forecast errors. This is not surprising since the covariance matrix of the predictive distribution changes slowly and smoothly over time while the forecast errors are more volatile. Moreover, the ranking of the models is to some extent reversed, particularly with the BVAR having much larger standardized forecast errors than the other models over the first half of the forecast sample. With the exception of the random walk model, this is broadly consistent with the findings for the point forecasts; see Warne et al. (2013). The reversal in rankings for the forecast error term can also be understood from the behavior of second moments, where a given squared forecast error yields a larger value for this term the smaller the uncertainty linked to the forecast is. Nevertheless, when compared with the forecast uncertainty term the differences between the models are generally smaller for the forecast error term. This suggests that the model ranking based on the log predictive score is primarily determined by the second moments of the predictive distribution in this illustration.

5. Summary and Conclusions

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

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

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

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

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

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

Over the forecast sample 1999Q1–2006Q4 we find that the normal density provides a good approximation of the predictive likelihood when examining the density forecasts for the five models. The "true value" of the predictive likelihood is represented by the IS estimator for all models except the random walk, whose predictive density is multivariate t and is therefore analytically determined. In terms of a model ranking, the log predictive score (the sum of the log predictive likelihood over the forecast sample) strongly favors the BVAR model, with the two DSGE-VAR models improving somewhat on the density forecasts of the NAWM, especially at the shorter horizons. The random walk model, on the other hand, is only competitive with the NAWM at the one-step-ahead horizon, especially for the variable selection with real GDP growth, GDP deflator inflation, and the short-term nominal interest rate only.

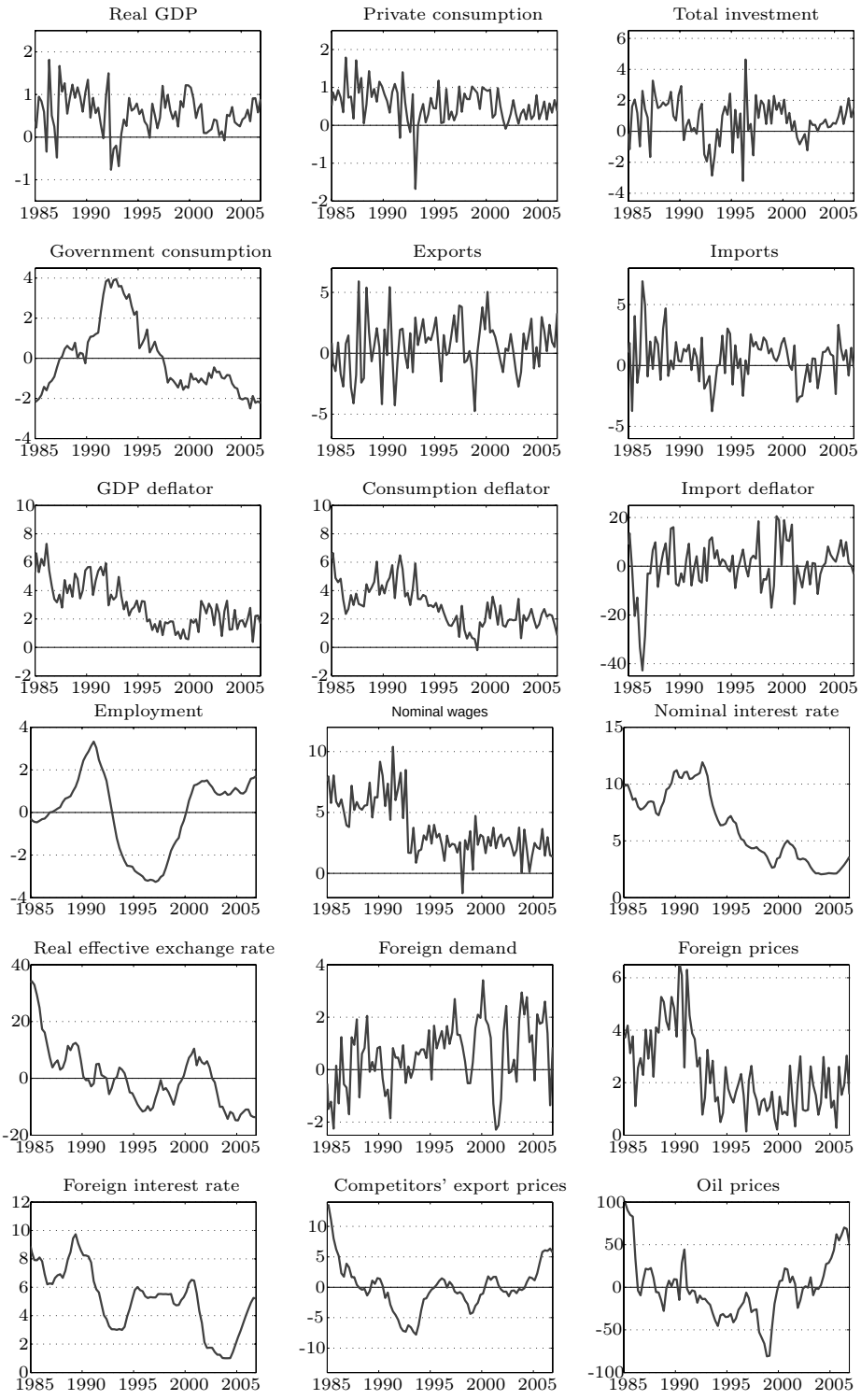
When the error from using a normal approximation of the predictive likelihood relative to the IS estimator is small, then the analytical form of the normal can be utilized to assess which feature of the predictive likelihood is driving the ranking of models. Specifically, the normal approximation allows for a simple decomposition of the predictive likelihood into the contributions of forecast uncertainty and forecast errors. The former term is specified via the determinant of the forecast error covariance matrix, while the latter term is determined through the quadratic standarized forecast errors. This decomposition suggests that the model ranking in the empirical illustration is primarily influenced by the forecast uncertainty term, while the forecast errors are mainly responsible for the volatility in the predictive likelihood.

TABLE 1: Difference between log predictive score using the IS estimator and the normal approximation over the evaluation period 1999Q1–2006Q4.

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

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

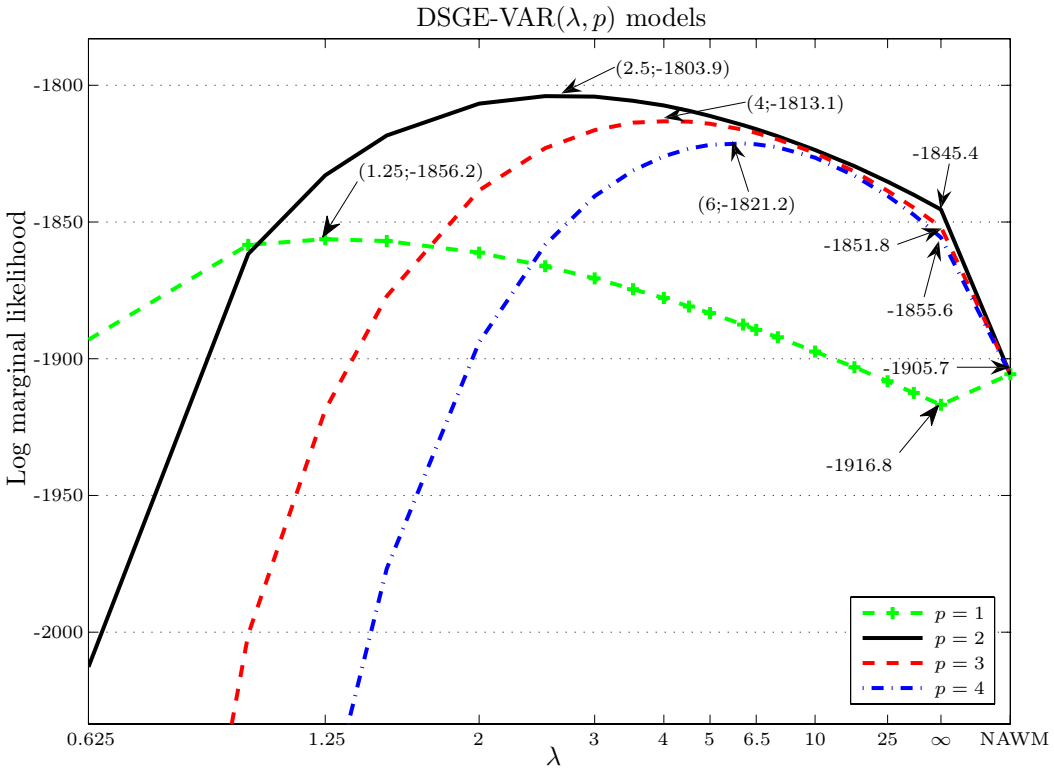
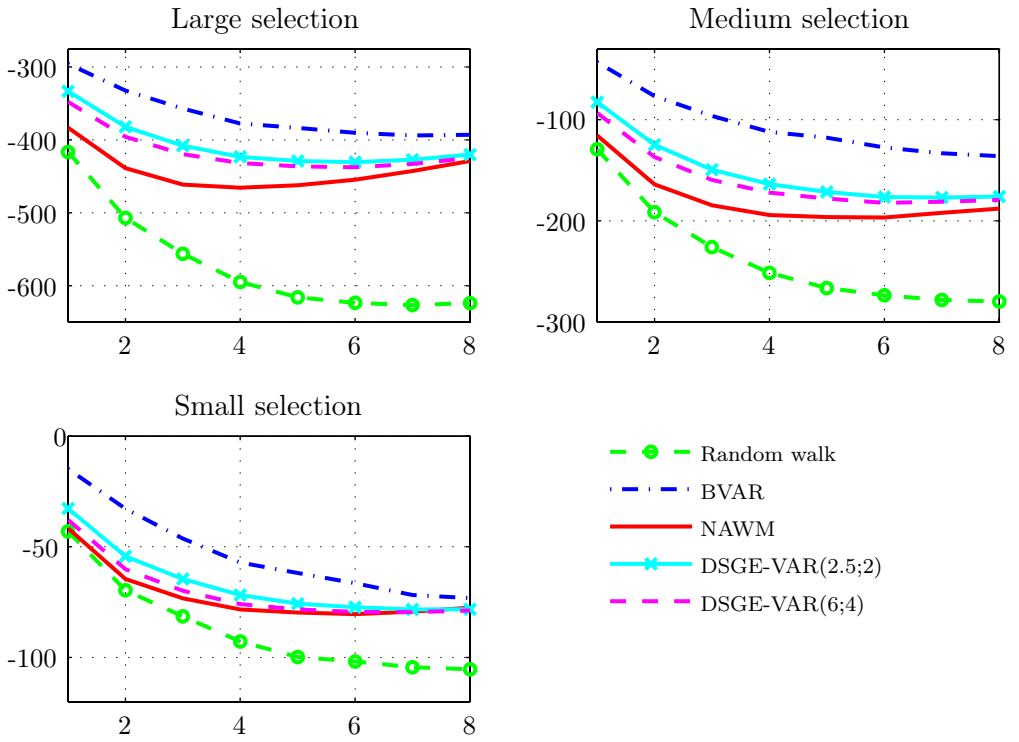


FIGURE 3: Log predictive scores using the IS estimator.



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

FIGURE 4: Log predictive scores using the normal approximation.

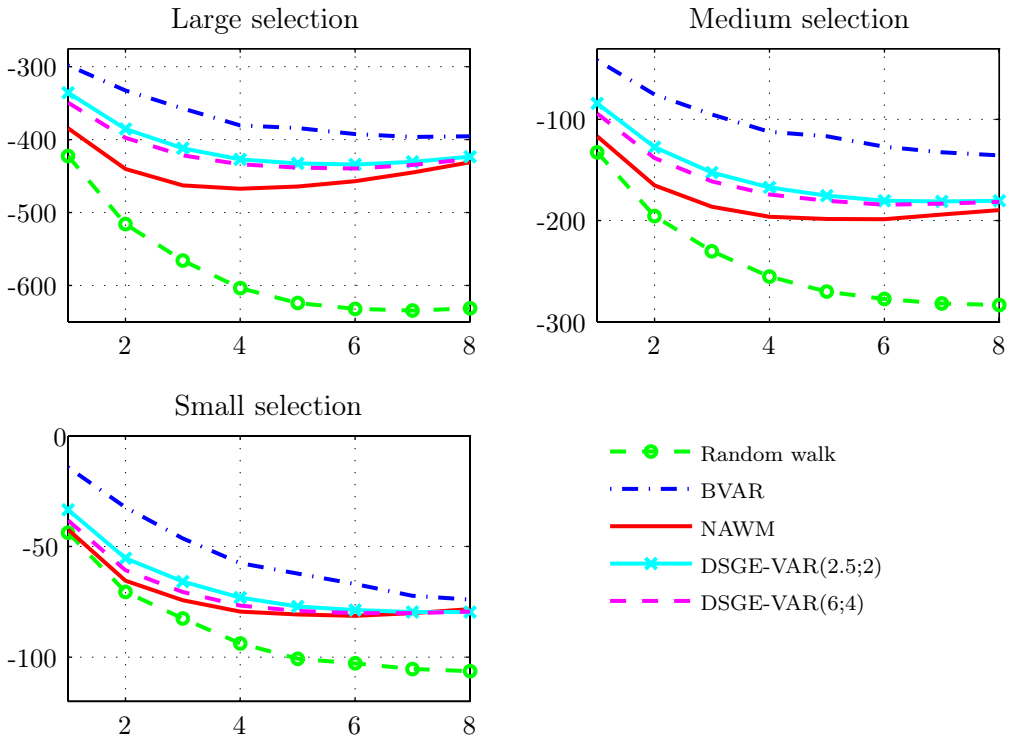


FIGURE 5: The evolution of the forecast uncertainty term of the normal density for the large selection of variables.

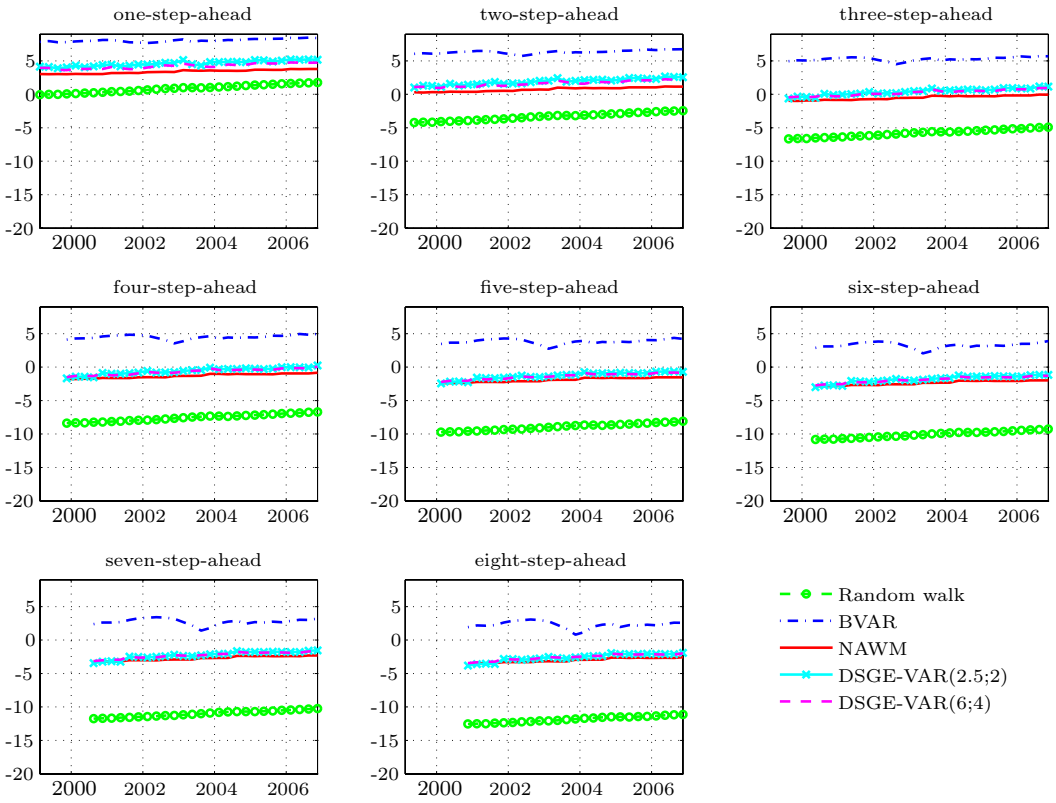
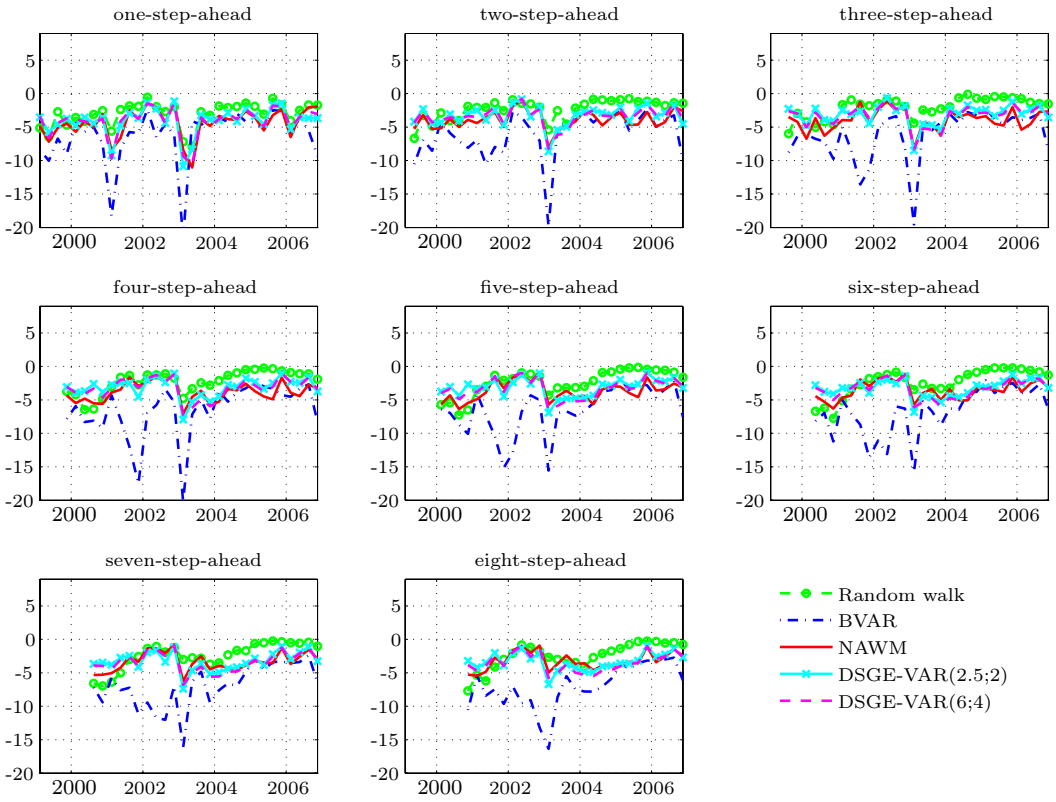


FIGURE 6: The evolution of the quadratic standardized forecast error term of the normal density for the large selection of variables.



References

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

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

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

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

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

Andersson, M. K. and Karlsson, S. (2008), "Bayesian Forecast Combinations for VAR Models," in S. Chib, G. Koop, W. Griffith, and D. Terrell (Editors), *Bayesian Econometrics*, 501–524, Emerald Group Publishing, Bingley, Volume 23 of *Advances of Econometrics*.

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

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

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

Chan, J. C. C. and Eisenstat, E. (2013), "Marginal Likelihood Estimation with the Cross-Entropy Method," *Econometric Reviews*, forthcoming.

Christiano, L. J., Eichenbaum, M., and Evans, C. (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

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

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

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

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

Del Negro, M. and Schorfheide, F. (2011), "Bayesian Macroeconometrics," in J. Geweke, G. Koop, and H. K. van Dijk (Editors), *The Oxford Handbook of Economic Forecasting*, 293–389, Oxford University Press, New York.

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

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

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

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

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

Engle, R. F. and Granger, C. W. J. (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

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

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

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

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

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

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

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

Geweke, J. (2010), *Complete and Incomplete Econometrics Models*, Princeton University Press, Princeton.

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

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

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

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

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

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

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

Karlsson, S. (2012), "Forecasting with Bayesian Vector Autoregressions," Prepared for *Handbook of Economic Forecasting*, Volume 2.

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

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

Lindley, D. V. (1957), "A Statistical Paradox," *Biometrika*, 44, 187–192.

Litterman, R. B. (1986), "Forecasting with Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business & Economic Statistics*, 4, 25–38.

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

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

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

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

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

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

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

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

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

Stock, J. H. and Watson, M. W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

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

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

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

Warne, A. (2013), "YADA Manual — Computational Details," Manuscript, European Central Bank. Available with the YADA distribution.

Warne, A., Coenen, G., and Christoffel, K. (2013), "Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models," Manuscript, European Central Bank.

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